DISSERTATION
submitted
to the
Combined Faculty for the Natural Sciences and Mathematics
of
Heidelberg University, Germany
for the degree of
Doctor of Natural Sciences

Put forward by

M.Sc.-Mathematics: Nopporn Thamrongrat
Born in: Nakhon Si Thammarat, Thailand
Oral examination:

Stable Convergence in Statistical Inference and Numerical Approximation of Stochastic Processes

Advisor: Prof. Dr. Mark Podolskij

Abstract

Stable convergence is a type of convergence of random variables, which is stronger than weak convergence but weaker than convergence in probability. It has been used in asymptotic theory of statistics and probability since Rényi originated his work (cf. [91]) in 1963. In this thesis, we study applications of stable convergence in two aspects. First, we show how to estimate the asymptotic (conditional) covariance matrix, which appears in many central limit theorems for stable laws in high-frequency estimation. We employ the idea of subsampling to provide a positive semi-definite estimator for this matrix. We show that our estimator is consistent in both noiseless models and models with additive microstructure noise. This estimate together with stable convergence theorems allow us to make some statistical inferences such as constructing confidence intervals or doing hypothesis testing. Moreover, we provide a decomposition of the leading error terms, from which we are able to get some insights about how to configure the subsampler by optimally choosing its tuning parameters (e.g., the number of subsamples). This leads to a rate of convergence for the subsampler.

Second, we apply stable convergence theorems to show a weak limit theorem for a numerical approximation of Brownian semi-stationary processes studied in [32]. In the original work of [32], the authors propose to use Fourier transformation to embed a given one dimensional Lévy semi-stationary process into a two-parameter stochastic field. For the latter, they use a simple iteration procedure and study the strong approximation error (in L^2 sense) of the resulting numerical scheme given that the volatility process is fully observed. In this work, we give a more precise assessment of the numerical error associated with the Fourier method. We complement their study by analyzing the weak limit of the error process in the framework of Brownian semi-stationary processes, where the drift and the volatility processes need to be numerically simulated.

Zusammenfassung

Stabile Konvergenz ist eine Konvergenzart von Zufallsvariablen, die stärker als die schwache Konvergenz ist, jedoch schwächer als die Konvergenz in Wahrscheinlichkeit. Diese Konvergenz findet Anwendung in der asymptotischen Theorie der Statistik und Wahrscheinlichkeitstheorie, seit Rényi seine Arbeit (siehe [91]) in 1963 verfasste. In dieser Doktorarbeit untersuchen wir zwei Aspekte von Anwendungen der stabilen Konvergenz. Zunächst zeigen wir, wie man die asymptotische (bedingte) Kovarianzmatrix, die in vielen stabilen zentralen Grenzwertsätzen in der Hochfrequenzschätzung vorkommt, schätzt. Dabei verwenden wir die Technik des Subsamplings, um einen positiv semidefiniten Schätzer für diese Matrix bereitzustellen. Wir zeigen, dass unser Schätzer in der Situation eines rauschfreien Modells und eines Modells mit additivem Microstructure-Noise konsistent ist. Durch Anwendung von stabilen Konvergenzsätzen auf diesen Schätzer können wir einige statistische Schlussfolgerungen wie die Konstruktion von Konfidenzintervallen und Hypothesentests ziehen. Darüber hinaus geben wir eine Zerlegung der führenden Fehlerterme, die uns eine optimale Wahl der Tuning-Parameter (die Anzahl der Subsamplings) des Subsamplers ermöglicht. Damit erhalten wir eine Konvergenzrate für den Subsampler.

Im zweiten Teil dieser Arbeit verwenden wir stabile Konvergenzsätze, um einen schwachen Grenzwertsatz für eine numerische Approximation von Brownschen semi-stationären Prozessen, die in [32] behandelt wurden, zu zeigen. In der Arbeit [32] schlagen die Autoren Benth, Eyjolfsson und Veraart vor, Fouriertransformationen zu benutzen, um einen eindimensionalen Lévy semi-stationären Prozess in ein zweiparametriges Zufallsfeld einzubetten. Die Autoren Benth, Eyjolfsson und Veraart nutzen nun eine einfache Iterationsmethode für das Zufallsfeld, um eine Approximation für den Lévy Prozess zu erhalten. Sie untersuchen dann den starken Approximationsfehler (im L^2-Sinne) im Falle, dass der Volatilitätsprozess vollständig beobachtet wird. In unserer Arbeit führen wir in einem von [32] verschiedenen Modell eine detailliertere Betrachtung des Schätzfehlers der Fouriermethode durch. Durch die Analyse von schwachen Grenzwerten des Fehlerprozesses von Brownschen semi-stationären Prozessen erweitern wir die Resultate aus [32]. Außerdem erlauben wir im Gegensatz zu [32], dass der Volatilitätsprozess nicht beobachtet, sondern ebenfalls simuliert wird.

Acknowledgments

Without the support of many people, this thesis would not have been possibly completed. First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Dr. Mark Podolskij, for his invaluable support and patient guidance along the way of my Ph.D. studies. His intelligent ideas and comments made this work possible. Apart from his academic support, how he treats people around him makes me feel that he is not only an advisor, but also a wonderful teacher, a remarkable coach and a very good friend. He trained me to have a positive attitude toward life. His wonderful encouragement helps me to go through problems and accomplish my goals. Among several good and bad times during my education, being his Ph.D. student is one of the best things that occurred to me.

Moreover, let me thank Dr. Bezirgen Veliyev and Assoc. Prof. Dr. Kim Christensen for very fruitful discussion, and Prof. Dr. Rainer Dahlhaus for taking care of me during the last year of my Ph.D. in Heidelberg University. I would like to thank Office of Education Affairs for Thai students in Germany for helping me in all aspects. I would take this opportunity to thank Asst. Prof. Dr. Sanae Rujivan to recommend me to come to study in Heidelberg, and my friend, Sophon Tunyavetchakit, for making my life easier and more comfortable to stay in Germany, Dr. Christian Schmidt and Claudio Heinrich for being my companions throughout my education, and all my friends to make my life so wonderful while staying abroad.

I have been financially supported by the Royal Thai Government Scholarship. Their contributions are gratefully acknowledged. I am also thankful to Deutsche Forschungsgemeinschaft through the Research Training Group (RTG – 1953) for funding the conferences.

Finally, I am deeply grateful to my beloved parents, brother and sister for their constant love, hearty encouragement, and unselfish sacrifice, and all teachers who have taught me all along. I truly believe that all people whom I have not personally mentioned here are aware of my deep appreciation.

Contents

Chapter 1

Introduction

Since the last decade, the advent of financial high-frequency data has led to a huge number of researches in the volatility estimation (see, e.g., [3, 25]). Volatility is a key component in the assessment and prediction of financial risk, be it in asset and derivatives pricing (e.g., [37, 94]), portfolio selection (e.g., [75]), or risk management and hedging (e.g., [66]). High-frequency data are recorded at the tick-by-tick level and store information about the time, price (i.e., a bid-ask quote or transaction price), and size of individual orders and executions.

Suppose a Brownian semimartingale X_t is characterized by the equation

$$X_t = X_0 + \int_0^t a_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s, \tag{1.1}$$

observed at the discrete time $i/n, i = 0, ..., [nt]$ where X_0 is the starting point, $a = (a_t)_{t\geq 0}$ is a predictable and locally bounded drift process, $\sigma = (\sigma_t)_{t\geq 0}$ is an adapted, càdlàg volatility process, while $W = (W_t)_{t\geq 0}$ is a standard Brownian motion. The process $X = (X_t)_{t\geq 0}$ in the expression (1.1) can be interpreted as interest rates or stock prices.

In financial econometrics, a quantity of interest to understand and evaluate financial risk is functionals of volatility:

$$IV(f)_t = \int_0^t f(\sigma_s)\mathrm{d}s, \tag{1.2}$$

i.e. integrated functions of the diffusion coefficient, for some suitable function f. In case of $f(x) = x^2$, then $IV(f)_t$ is a so-called *integrated volatility* or *integrated variance* denoted by $IV_t = \int_0^t \sigma_s^2 \mathrm{d}s$. The econometric challenge is that the objects of interest appearing in (1.2) are latent. Therefore, volatility estimation became a popular field of research in developing robustness and efficiency. It is well-known that the realized variance

$$RV_t^n = \sum_{i=1}^{[nt]} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})^2 \tag{1.3}$$

estimates the integrated volatility $\int_0^t \sigma_s^2 \mathrm{d}s$ discussed by Andersen and Bollerslev [2] and Barndorff-Nielsen and Shephard [23]. To estimate the functionals of volatility in general cases, Barndorff-Nielsen et al. [15] show asymptotic theory for the processes

$$V(f)_t^n = \frac{1}{n} \sum_{i=1}^{[nt]} f\big(\sqrt{n}\Delta_i^n X\big) \tag{1.4}$$

$$V(f,g)_t^n = \frac{1}{n} \sum_{i=1}^{[nt]} f\big(\sqrt{n}\Delta_i^n X\big) g\big(\sqrt{n}\Delta_{i+1}^n X\big), \tag{1.5}$$

under some appropriated functions f and g where $\Delta_i^n X = X_{i/n} - X_{(i-1)/n}$ for $i = 1, \dots, n$. These two processes are called power and bipower variations, respectively. For example, they derive that under some assumptions on $\mathbb{R}^m$-valued function f,

$$V(f)_t^n \xrightarrow{\mathbb{P}} V(f)_t := \int_0^t \rho_{\sigma_s}(f) ds,$$

where $\rho_x(f) = \mathbb{E}[f(xU)]$ for $x \in \mathbb{R}$, $U \sim N(0,1)$ and $\xrightarrow{\mathbb{P}}$ denotes convergence in probability. Furthermore, the authors provide a central limit theorem for the statistic $V(f)_t^n$:

$$\sqrt{n}\Big(V(f)_t^n - V(f)_t\Big) \xrightarrow{d_{st}} MN(0, \Sigma), \tag{1.6}$$

where the notation $\xrightarrow{d_{st}}$ means stable convergence (see, Definition 2.18) and Σ is the $m \times m$ asymptotic conditional covariance matrix, which has elements

$$\Sigma_{ij} = \int_0^t \Big[\rho_{\sigma_s}(f_i f_j) - \rho_{\sigma_s}(f_i)\rho_{\sigma_s}(f_j)\Big] \mathrm{d}s.$$

Note that $V(f)_t^n$ coincides with the realized variance RV_t^n when $f(x) = x^2$.

However, empirical studies show that this usual realized volatility RV_t^n estimates the integrated variance for moderate frequency data such as 5 or 15 min data but no longer works in high frequency data, say 1 minute or less (see, [4, 5, 7, 78]). The source of this error is known as *microstructure noise*. In other words, the observed prices are contaminated by microstructure noise such as bid-ask spreads, price discreteness, and so forth

[36, 73, 93]. From the mathematical point of view, Zhou [102] suggested to econometrics that the observed price process should be written in the form

$$Y_{\frac{i}{n}} = X_{\frac{i}{n}} + \epsilon_{\frac{i}{n}}, \tag{1.7}$$

where X is the latent process in the form of (1.1) and ϵ is the microstructure noise. This paved the way for the next cohort of estimators that were designed to be more resistant to noise, e.g., multiscale estimators introduced by Zhang et al. [100, 101], realized kernel estimators by Barndorff-Nielsen et al. [16] and pre-averaging estimators by Podolskij and Vetter [85] and generalized by Jacod et al. [16]. As such, much progress has been made and today there is no shortage of estimators to provide consistent estimates of volatility functionals in various contexts.

In the meantime, estimation of the asymptotic (conditional) covariance matrix, which appears in many central limit theorems in high-frequency estimation of volatility such as Σ in (1.6), has received less attention. This is surprising, because it can give inference about volatility such as construction of confidence intervals or hypothesis tests. In practice, the asymptotic (co)variance is not trivial to estimate because it often relies on parameters that are substantially more difficult to back out from the available sample of high-frequency data. Moreover, the expression for the asymptotic (co)variance also rests typically and heavily on the properties of the data and it is bound to change depending on these. This is an unpleasant concern with real high-frequency data, which are contaminated by market microstructure noise. While the noise is often assumed to be i.i.d. and independent of the efficient price, there is some empirical and theoretical support for a serially correlated, heteroscedastic and potentially, endogenous noise process at the tick level (e.g., [7, 47, 70, 55]). An estimator of the asymptotic variance designed for i.i.d. and independent noise cannot be expected to give valid inference, if the underlying conditions are violated. It is not trivial to verify the conditions imposed on the noise [58], which makes it more pressing to find estimators that are robust against modeling criteria. Finally, in multivariate analysis, inference would at some stage require an estimate of the asymptotic covariance matrix. Here, the proposed estimator should ideally be positive semi-definite, while, in contrast, some existing estimators of the asymptotic covariance matrix in the high-frequency setting are not assured to be that (see, Section 5.3).

In this work, we propose to use subsampling for assessing the uncertainty embedded in high-frequency estimation of functionals of volatility. Subsampling is based on creating several—properly rescaled—estimates of the parameter(s) of interest using local stretches of sample data and then studying the sampling variation of these. It was originally developed in the context of stationary time series in the long-span domain by Politis and Romano [88]. The term subsampling appeared in the high-frequency literature in Zhang et al. [101], who proposed a two-scale realized variance based on price subsampling. This is different from traditional subsampling and actually does not work for asymptotic variance estimation, because it leads to an overlapping samples problem in the subsampled returns, causing the subsample estimates to be too strongly correlated in large samples. This was pointed out by Kalnina and Linton [69] and Kalnina [67], who propose an alternative inference strategy based on infill return subsampling, which leads to better asymptotic properties. Kalnina [68] extends these ideas to inference about a multivariate parameter, while Ikeda [60] and Varneskov [95] consider subsample estimation of the asymptotic variance of the realized kernel.

The first part of this thesis contains the recent results of Christensen et al. [41]. We apply the idea of subsampling to construct estimators of asymptotic covariance matrices for power and bipower variations, in the forms of (1.4) and (1.5), in both noiseless and noisy markets. Among three main approaches to moderate the effect of noise, i.e. multiscale [100, 101], realized kernel [16] and pre-averaging [16, 85], we choose the pre-averaging one since it allows estimation of other powers of volatility. The idea of this approach is that averaging on a number of $Y_{i/n}$'s near the time point i/n, one can get an estimate, say $\bar{Y}_{i/n}$, which tends to be close to the latent process $X_{i/n}$ because the noise is largely averaged away. Averaging our discrete sample of noisy high-frequency data this way leads to a new set of increments,

$$\Delta\bar{Y}_i^n = \sum_{j=1}^{k_n} w_j^n \Delta_{i+j}^n Y,$$

for a weight function w (see, Subsection 5.1.1 for the details).

Moreover, we show that our estimators are consistent in both settings (with and without noise). This estimate together with stable convergence theorems allow us to make some statistical inferences such as to construct conficence intervals or to do hypothesis testing (see, (4.21)). Furthermore, we provide a decomposition of the leading error terms of the statistic, from which we are able to get some insights about how to configure the subsampler by optimally choosing its tuning parameters (e.g., the number of subsamples). This leads to an optimal rate of convergence for the subsampler; a result that has, to the best of our knowledge, not been derived in earlier work.

There are at least three advantages to use our subsampling technique. First, subsampling is intuitive and relatively easy to compute, because it does not require an extra set of estimators; it uses copies of the original statistic. Second, in the multivariate context, it leads to variance-covariance matrix estimates that are positive semi-definite by construction. Third, subsampling does not explicitly rely on the form of asymptotic variance.

Another application of stable convergence is shown in the following second part of this thesis.

In 2007, Barndorff-Nielsen and Schmiegel introduced a class of spatio-temporal stochastic processes called *ambit fields* in a series of papers [20, 21, 22] in the context of turbulence modelling. However, the ambit fields has found manifold applications in mathematical finance such as energy spot prices [10], power markets [32] and electricity forward markets [11] and in biology (modelling of tumor growth) [19]. In full generality they are described via the formula

$$X_t(x) = \mu + \int_{A_t(x)} g(t,s,x,\xi)\sigma_s(\xi)L(\mathrm{d}s,\mathrm{d}\xi) + \int_{D_t(x)} q(t,s,x,\xi)a_s(\xi)\mathrm{d}s\mathrm{d}\xi \qquad (1.8)$$

where t typically denotes time while x gives the position in space. Furthermore, $A_t(x)$ and $D_t(x)$ are ambit sets, g and q are deterministic weight functions, σ represents the volatility or intermittency field, a is a drift field and L denotes a *Lévy basis*. We recall that a Lévy basis $L = \{L(B) : \ B \in \mathcal{S}\}$, where $\mathcal{S}$ is a δ-ring of an arbitrary non-empty set S such that there exists an increasing sequence of sets $(S_n) \subset \mathcal{S}$ with $\cup_{n\in\mathbb{N}} S_n = S$, is an independently scattered random measure.

In this thesis, attention has been given to a special class of ambit fields, *a Brownian semi-stationary process*, which is defined as

$$X_t = \mu + \int_{-\infty}^{t} g(t-s)\sigma_s W(\mathrm{d}s) + \int_{-\infty}^{t} q(t-s)a_s \mathrm{d}s, \qquad (1.9)$$

where g and q are non-negative deterministic kernels, $(a_t)_{t\in\mathbb{R}}$ and $(\sigma_t)_{t\in\mathbb{R}}$ are adapted càdlàg processes, and W is a two sided Brownian motion. The notion of a semi-stationary process comes from the fact that the process $(X_t)_{t\in\mathbb{R}}$ is stationary whenever $(a_t, \sigma_t)_{t\in\mathbb{R}}$ is stationary and independent of $(W_t)_{t\in\mathbb{R}}$. We note that if the Brownian motion W in (1.9) is replaced with a Lévy process L, then the process X_t is called *Lévy semi-stationary process* instead.

We should point out that a Brownian semi-stationary process of the form (1.9) is not always a semimartingale. For example, the Brownian semi-stationary process with the gamma kernel

$$g(x) = x^{\alpha} \exp^{-\lambda x}$$

for $\alpha \in (\frac{-1}{2}, 0) \cup (0, \frac{1}{2})$ and $\lambda > 0$, which was used to turbulence study in [43], is no longer a semimartingale.

In the past decade, stochastic analysis, probabilistic properties and statistical inference for Brownian (Lévy) semi-stationary processes have been studied in numerous papers. We refer to [8, 9, 13, 14, 29, 22, 33, 43, 51, 82] for the mathematical theory as well as to [12, 83] for recent surveys on theory of ambit fields and their applications.

For practical applications in sciences, numerical approximation of Brownian (Lévy) semi-stationary processes, or, more generally, of ambit fields, is of importance. We remark that due to a moving average structure of a Brownian (Lévy) semi-stationary process (cf. (1.9)), there exists no simple iterative Euler type approximation scheme. For this reason, the authors of [31, 32] have proposed two different embedding strategies to come up with a numerical simulation. The first idea is based on the embedding of a Lévy semi-stationary process into a certain two-parameter stochastic partial differential equation. The second one is based upon a Fourier method, which again interprets a given Lévy semi-stationary process as a realization of a two-parameter stochastic field. We refer to the PhD thesis of Eyjolfsson [48] for a detailed analysis of both methods and their applications to modeling energy markets. We would also like to mention very recent work [39], which investigates numerical simulations of spatio-temporal ambit fields.

The second part of this thesis contains the results from Podolskij and Thamrongrat [84]. We consider the Fourier numerical approach introduced in [32]. The main advantage of this numerical scheme is that it separates the simulation of the stochastic ingredients (σ and W) and the approximation of the deterministic kernel g. This contrasts with a straightforward discretization scheme. The aim of this part is to study the weak limit theory of the numerical scheme associated with the Fourier method proposed in [32, 48]. In the original work [32], the authors discussed the strong approximation error (in the L^2 sense) of the numerical scheme for Lévy semi-stationary processes, where the volatility process $(\sigma_t)_{t\in\mathbb{R}}$ was assumed to be observed. We complement their study by analyzing the weak limit of the error process in the framework of Brownian semi-stationary processes, where the drift and the volatility processes need to be numerically simulated. This obviously gives a more precise assessment of the numerical error associated with the Fourier method.

The thesis is organized as follows. In Chapter 2, we provide some basic definitions and results that are applied in the entire work. In Chapter 3, we prepare a powerful tool, Malliavin calculus, to use in Chapter 4 and 5. In Chapter 4, we show how to construct a positive semi-definite estimator of a asymptotic variance in noiseless models based on a technique of subsampling. Furthermore, we show that our estimator is consistent and a rate of convergence is provided. In Chapter 5, we turn our attention to models with additive microstructure noise. We show that by applying the same subsampling technique, our estimator is robust to noise. At the end of this chapter, we do numerical simulations in order to inspect the finite sample performance of our estimator. Finally, in Chapter 6, which can be studied independently from Chapter 4 and 5, we present a weak limit theorem for a numerical approximation of Brownian semi-stationary processes.

Chapter 2

Preliminaries

2.1 Basic definitions and results

This section states some basic definitions and examples that will be used in the following chapters. Throughout this thesis, all processes are defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Furthermore, we assume that the filtered probability space satisfies the usual assumptions: it is complete, i.e. $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets, and right continuous, i.e. $\mathcal{F}_t = \mathcal{F}_{t+}$ where

$$\mathcal{F}_{t+} = \bigcap_{u>t} \mathcal{F}_u,$$

for all $t \geq 0$.

2.1.1 Skorohod spaces and càdlàg processes

A function f defined on real numbers is called *càdlàg* if it is right continuous and has left limits everywhere. Let $\mathbb{D} = \mathbb{D}[0, T]$ be the set of all càdlàg functions on $[0, T]$, $T > 0$. We introduce a uniform norm on $\mathbb{D}$ by setting

$$\|f\| = \sup_{0\leq t\leq T} |f(t)|.$$

Let Λ denote the set of all strictly increasing, continuous bijections from $[0, T]$ to itself.

Definition 2.1
A *Skorohod metric* on $\mathbb{D}$ is defined by

$$d(f, g) = \inf_{\lambda\in\Lambda} \{\|\lambda - I\| \vee \|f - g \circ \lambda\|\},$$

where $I : [0, T] \to [0, T]$ is the identity function and $\|\cdot\|$ is the uniform norm on $\mathbb{D}$.

The topology generated by the skorohod metric is called the *Skorohod topology* and the space $\mathbb{D}$ is called the *Skorohod space*.

The idea of càdlàg functions allows us to study stochastic theory beyond continuous sample path processes because it admits jumps.

Definition 2.2
A stochastic process $X = (X_t)_{t\geq 0}$ is called càdlàg if almost all its sample paths $t \mapsto X_t$ are càdlàg.

Example 2.3
Classical examples of càdlàg processes are Brownian motion and Poisson processes.

2.1.2 Stopping times and semimartingales

To define a semimartingale, we first present notions of stopping times and local martingales.

Definition 2.4
A random variable $\tau : \Omega \to [a, b]$ is called a *stopping time* with respect to a filtration $\{\mathcal{F}_t \,|\, a \leq t \leq b\}$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \in [a, b]$.

A stopping time is often interpreted as a mechanism to make a decision whether to continue or stop a process at time t on the basis of the information provided by $\mathcal{F}_t$.

Definition 2.5
An $\mathcal{F}_t$-adapted stochastic process X is called a *local martingale* with respect to the filtration $\{\mathcal{F}_t \,|\, a \leq t \leq b\}$ if and only if there exists a sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ such that $(\tau_n)_{n\in\mathbb{N}}$ increases monotonically to b almost surely as $n \to \infty$ and for each n, the stopped process $X_{t\wedge\tau_n}$ is a martingale with respect to $\{\mathcal{F}_t \,|\, a \leq t \leq b\}$.

Note that, from this definition we can easily see that every martingale is local martingale. One can also prove that every bounded local martingale is martingale. The example below is a process which is local martingale but not martingale.

Example 2.6
Let W be a Brownian motion. For each $t > 1/4$, we define a process X_t as

$$X_t = \int_0^t \exp\{W_s^2\} dW_s.$$

Then, the process X is a local martingale. Indeed, these type of processes are always local martingales because they are integrals of continuous processes with respect to Brownian motion (see, [71, Theorem 5.5.2.]). However, it is not martingale since $\mathbb{E}[\exp\{W_t^2\}] = \infty$ for $t > 1/4$ which implies that X_t is not integrable.

We are now ready to define a semimartingale.

Definition 2.7
A stochastic process X defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is called a *semimartingale* if and only if there exist a local martingale process M and a predictable process with finite variation A such that

$$X = X_0 + A + M, \tag{2.1}$$

where X_0 is $\mathcal{F}_0$-measurable and $A_0 = M_0 = 0$. If the process M in the decomposition (2.1) is of the form

$$M_t = \int_0^t \sigma_s dW_s,$$

where σ is a càdlàg process and W is a Brownian motion, we will call X a *Brownian semimartingale*.

We can see that the class of semimartingale contains all the processes of martingales and local martingales.

Example 2.8
(i) All adapted càdlàg processes with finite variation paths are semimartingales.
(ii) Brownian motion and Lévy processes are semimartingales.
(iii) Itô processes, which can be expressed as

$$X_t = X_0 + \int_0^t \sigma_s dW_s + \int_0^t \mu_s ds,$$

where W is a Brownian motion and σ and μ are adapted processes, are semimartingales.

We consider a semimartingale of the form

$$Z_t = Z_0 + \int_0^t \sigma_s dW_s + \int_0^t \mu_s ds + J_t, \tag{2.2}$$

where W is a Brownian motion, σ and μ are adapted processes and J_t denotes the jump part of Z and it is observed on $[0, 1]$ at high frequency, i.e. at $0 = t_0, ..., t_i = i/n, ..., t_n = 1$ with $n \to \infty$. Its quadratic variation (see, [90]),

$$[Z]_t = \int_0^t \sigma_s^2 ds + \sum_{0\leq s\leq t} |\Delta X_s|^2$$

with $\Delta X_s = X_s - X_{s^-}$, is one of the most important unobserved characteristics that we need to know. We call $[Z]_t^c := \int_0^t \sigma_s^2 ds$ the continuous part of the quadratic variation and call $[Z]_t^d := \sum_{0\leq s\leq t} |\Delta X_s|^2$ the discontinuous part of the quadratic variation. Z is said to be a quadratic pure jump semimartingale if $[Z]_t^c = 0$.

2.1.3 Burkholder inequalities

One of the remarkable tools which we often use in the proofs of the main results in this thesis is Burkholder inequality. It helps us to find the bounds of stochastic increments. For the sake of application, we provide two cases separately. The first one is for continuous local martingales and the other case is for discrete martingales. We define $X_t^* = \sup_{s \le t} |X_s|$.

Proposition 2.9 (Burkholder inequality for continuous local martingales ([92], Theorem 4.1))
Let $X = (X_t)_{t \ge 0}$ be a continuous local martingale such that $X_0 = 0$. For all $0 < q < \infty$, there exist positive constants c_q and C_q such that

$$c_q \mathbb{E}[[X]_t^{q/2}] \le \mathbb{E}[(X_t^*)^q] \le C_q \mathbb{E}[[X]_t^{q/2}],$$

where $[X]_t$ denotes the quadratic variation of a process X.

Example 2.10
Let X be an 1-dimensional stochastic process such that $X_t = \int_0^t \sigma_s dW_s$ where σ is a bouned process and W is a Brownian motion. Then, by Burkholder inequality for continuous local martingales we can conclude that

$$\mathbb{E}\Big[\Big| \int_u^t \sigma_s dW_s \Big|^q\Big] \le C_q |t-u|^{q/2},$$

for any $u, q \ge 0$.

Proposition 2.11 (Burkholder inequality for discrete martingales ([44]))
Let $X = (X_n)_{n \ge 0}$ be a discrete martingale such that $X_0 = 0$. For all $1 \le q < \infty$, there exist positive constants c_q and C_q such that

$$c_q \mathbb{E}[(S(X))^{q/2}] \le \mathbb{E}[(X_\infty^*)^q] \le C_q \mathbb{E}[(S(X))^{q/2}],$$

where $S(X) = \sum_{i=1}^{\infty} (X_n - X_{n-1})^2$.

2.2 Modes of convergence

In this section, we present some concepts of stochastic convergence which will be used in this thesis. Throughout this thesis we use the notation O_p in the context of $Z_n = O_p(a_n)$ if and only if for every $\varepsilon > 0$ there exist positive constants C and N such that for every $n > N$, $\mathbb{P}(|Z_n| \le C a_n) > 1 - \varepsilon$. We also use the notation o_p in the sense that $Z_n = o_p(a_n)$ if and only if

$$\frac{Z_n}{a_n} \xrightarrow{\mathbb{P}} 0$$

as $n \to \infty$.

2.2.1 Tightness and convergence in finite dimensional distribution

Tightness and convergence in finite dimensional distribution are elementary tools for showing weak convergence of stochastic processes.

Definition 2.12
A sequence $(X_n)_{n\in\mathbb{N}}$ of random variables is said to be *tight* or *bounded in probability* if for each $\varepsilon > 0$ there exists a positive real number K_ε such that

$$\sup_n \mathbb{P}(|X_n| \geq K_\varepsilon) < \varepsilon.$$

Example 2.13
A sequence $(X_n)_{n\in\mathbb{N}}$ of random variables such that $\sup_n \mathbb{E}[|X_n|^p] < \infty$ for some $p > 0$ is tight. It follows from the Markov inequality.

The next proposition gives conditions of showing tightness of measures.

Proposition 2.14 (Kolmogorov's tightness criteria ([38], Theorem 9.9))
Let (M, ρ) be a complete metric space satisfying the Heine-Borel property and $(S^n)_{n\geq 1}$ be M-valued continuous stochastic processes on $[0, 1]$. The following conditions are satisfied:
(i) there exist $\gamma, \varepsilon > 0$ and $C < \infty$ such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}[\rho(S^n_t, S^n_s)^\gamma] \leq C|t-s|^{1+\varepsilon}$$

for all $0 \leq s, t \leq 1$,
(ii) $\exists m_0 \in M$ such that

$$\sup_{n\in\mathbb{N}} \mathbb{P}(\rho(S^n_0, m_0) > k) = 0$$

as $n \to \infty$. Then the collection of measures, $\{\mu_n = Law(S_n)\}_{n\geq 1}$ on $C([0, 1], M)$ is tight.

Definition 2.15
A sequence of stochastic processes $(X^n)_{n\in\mathbb{N}}$ on $[0, T]$ is said to *converge in finite dimensional distribution* to a stochastic process X if $\forall t_i \in [0, T]$ and $k \in \mathbb{N}$, then

$$(X^n_{t_1}, X^n_{t_2}, ..., X^n_{t_k}) \xrightarrow{d} (X_{t_1}, X_{t_2}, ..., X_{t_k})$$

where $\xrightarrow{d}$ means convergence in distribution.

It has been known that convergence in finite dimensional distribution is necessary for convergence in distribution but not sufficient. However, it can be shown that convergence in finite dimensional distribution together with tightness imply weak convergence [34].

2.2.2 UCP convergence and stable convergence

We say that a sequence of functions $f_n : \mathbb{R}^+ \to \mathbb{R}$ converges *uniformly on compacts* to a function f if and only if for each $t > 0$

$$\sup_{0\leq s\leq t} |f_n(s) - f(s)| \longrightarrow 0$$

as $n \longrightarrow \infty$. This type of convergence is used to define *ucp* convergence.

Definition 2.16
We say that a sequence of stochastic processes $(X^n)_{n\in\mathbb{N}}$ on $[0,T]$ converges to a stochastic process X *uniformly on compacts in probability* or *ucp* denoted by $X^n \overset{ucp}{\Longrightarrow} X$ if for each $\varepsilon > 0$

$$\mathbb{P}(\sup_{0\leq s\leq T} |X^n_s - X_s| > \varepsilon) \longrightarrow 0$$

for $n \longrightarrow \infty$.

We can see that when the stochastic processes X^n and X are nice, e.g. left or right continuous then this definition makes sense since the supremum is confined to measurable countable set of rational times.

Example 2.17
(i). If $(X^n)_{n\in\mathbb{N}}$ is a sequence of càdlàg martingales and X is a stochastic process such that for each $t > 0$

$$\mathbb{E}[|X^n_t - X_t|] \longrightarrow 0,$$

as $n \longrightarrow \infty$, then by Doob's martingale inequality one can show that $X^n \overset{ucp}{\Longrightarrow} X$.
(ii). If M is a semimartingale and $(X^n)_{n\in\mathbb{N}}$ is a sequence of predictable processes converging pointwise to X such that

$$\sup_n |X^n| \text{ is } M\text{-integrable},$$

then by dominated convergence theorem for stochastic integral we can show that $\int X^n dM \overset{ucp}{\Longrightarrow} \int X dM$ (see, [90, Theorem 32]).

Next, we introduce the concept of stable convergence of random variables which plays an important role in this thesis.

Definition 2.18
A sequence $(Y_n)_{n\in\mathbb{N}}$ of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a Polish space (E, Γ) (i.e. a separable, completely metrizable topological space)

is said to *converge stably in law* with limit Y, denoted by $Y_n \xrightarrow{d_{st}} Y$, where Y is defined on extension $(\Omega', \mathcal{F}', \mathbb{P}')$, if and only if for any measurable and bounded random variable Z and any bounded continuous function g it holds that

$$\mathbb{E}[g(Y_n)Z] \longrightarrow \mathbb{E}'[g(Y)Z]$$

as $n \to \infty$.

Note that by this definition the random variables Y_n can also be random processes. Moreover, the stable convergence is stronger than weak convergence but weaker than ucp convergence. We refer to [91] for the original definition of stable convergence and to [65] for further details. We now state some crucial properties of stable convergence in the following part which will be applied to our results.

Proposition 2.19 ([87], Proposition 2.2)
Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of random variables. The following conditions are equivalent:
(i) $Y_n \xrightarrow{d_{st}} Y$
(ii) $(Y_n, Z) \xrightarrow{d} (Y, Z)$ *for any $\mathcal{F}$-measurable variable Z*
(iii) $(Y_n, Z) \xrightarrow{d_{st}} (Y, Z)$ *for any $\mathcal{F}$-measurable variable Z.*

The above proposition gives equivalent definitions for showing stable convergence of random sequences.

Lemma 2.20 (Continuous mapping theorem for stable convergence)
Let $(Y_n)_{n\in\mathbb{N}}$ be a d-dimensional sequence of random variables with $Y_n \xrightarrow{d_{st}} Y$. Let $g : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a continuous function. Then

$$g(Y_n) \xrightarrow{d_{st}} g(Y).$$

Proof. Applying Proposition 2.19, we have that $(Y_n, Z) \xrightarrow{d} (Y, Z)$ for all $\mathcal{F}$-measurable Z. Since g is countinuous, we obtain by the classical continuous mapping theorem that

$$(g(Y_n), Z) \xrightarrow{d} (g(Y), Z).$$

We finish the proof by applying the equivalence of (i) and (ii) in Proposition 2.19. □

For now, it is not obvious that how to use stable convergence in practice. The next proposition will allow us to apply stable convergence to construct confidence intervals. Firstly, we give an auxiliary lemma and present a concept of mixed normal distributions.

Lemma 2.21
Let $(Y_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$ be a d-dimensional sequence of random variables with $Y_n \xrightarrow{d_{st}} Y$ and $V_n \xrightarrow{\mathbb{P}} V$. It holds that

$$(Y_n, V_n) \xrightarrow{d_{st}} (Y, V).$$

Proof. Let Z be $\mathcal{F}$-measurable. Then, Proposition 2.19 implies $(Y_n, V, Z) \xrightarrow{d} (Y, V, Z)$. Moreover, $V_n - V \xrightarrow{\mathbb{P}} 0$ by assumption. Therefore, $(Y_n, V_n, Z) \xrightarrow{d} (Y, V, Z)$. Again using Proposition 2.19, we complete the proof. □

Definition 2.22
We call a random variable $Y = VU$ a *mixed normal distribution* with random conditional variance V^2 denoted by $Y = MN(0, V^2)$ if $V > 0$, $U \sim N(0,1)$ and V and U are independent.

From the statistical point of view, the weak convergence $Y_n \xrightarrow{d} VU = MN(0, V^2)$ is useless if the distribution of V is unknown because confidence intervals are unavailable. More precisely, for a normal distribution with non-deterministic variance V^2, the weak convergence $Y_n \xrightarrow{d} VU = MN(0, V^2)$ does not imply $Y_n/V_n \xrightarrow{d} N(0,1)$ where V_n is an estimator of V (see, [87, p.331]) but the stable convergence does as we shall see in the following proposition.

Proposition 2.23
Let $(Y_n)_{n\in\mathbb{N}}$ be a 1-dimensional sequence of random variables with $Y_n \xrightarrow{d_{st}} Y$. Suppose $Y \sim MN(0, V^2)$ with V being $\mathcal{F}$-measurable. If there exists a 1-dimensional sequence of random variables $(V_n)_{n\in\mathbb{N}}$ such that $V_n \xrightarrow{\mathbb{P}} V$ and $V_n, V > 0$, then

$$\frac{Y_n}{V_n} \xrightarrow{d} N(0,1).$$

Proof. We complete the proof by applying Lemma 2.20 and Lemma 2.21. □

2.3 Limit theorems for semimartingales

In this section, we mainly present asymptotic results for continuous semimartingale processes, namely a law of large numbers and a central limit theorem. These asymptotic properties will be needed in Chapter 4 and 5. The results are based on the work of Barndorff-Nielsen et al. [15]. We also present Jacod's stable central limit theorem at the end of this section to show how to derive limit results for diffusion processes in practice.

We consider a scalar process $X = (X_t)_{t\geq 0}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and adapted to $(\mathcal{F}_t)_{t\geq 0}$ in the form of a continuous Itô semimartingale as expressed by the equation:

$$X_t = X_0 + \int_0^t a_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s, \tag{2.3}$$

where X_0 is the starting point, $a = (a_t)_{t\geq 0}$ is a predictable and locally bounded drift process, $\sigma = (\sigma_t)_{t\geq 0}$ is an adapted, càdlàg volatility process, while $W = (W_t)_{t\geq 0}$ is a standard Brownian motion. We also point out that a and σ are left unspecified, and our results are completely nonparametric (within this class of models). While we do not impose assumptions a priori, we sometimes need an additional regularity condition on σ of the following type.

Assumption (V): σ is of the form:

$$\sigma_t = \sigma_0 + \int_0^t \tilde{a}_s \mathrm{d}s + \int_0^t \tilde{\sigma}_s \mathrm{d}W_s + \int_0^t \tilde{v}_s \mathrm{d}B_s$$
$$+ \int_0^t \int_E \tilde{\delta}(s,x) 1_{\{|\tilde{\delta}(s,x)|\leq 1\}} (\tilde{\mu} - \tilde{\nu})(\mathrm{d}s, \mathrm{d}x) + \int_0^t \int_E \tilde{\delta}(s,x) 1_{\{|\tilde{\delta}(s,x)|> 1\}} \tilde{\mu}(\mathrm{d}s, \mathrm{d}x), \tag{2.4}$$

where σ_0 is its initial value, $\tilde{a} = (\tilde{a}_t)_{t\geq 0}$, $\tilde{\sigma} = (\tilde{\sigma}_t)_{t\geq 0}$ and $\tilde{v} = (\tilde{v}_t)_{t\geq 0}$ are adapted, càdlàg stochastic processes, while $B = (B_t)_{t\geq 0}$ is a standard Brownian motion that is independent of W. Furthermore, $(E, \mathcal{E})$ is a Polish space, $\tilde{\mu}$ is a random measure on $\mathbb{R}_+ \times E$, which is independent of (W, B) and has an intensity measure $\tilde{\nu}(\mathrm{d}s, \mathrm{d}x) = \mathrm{d}s\tilde{F}(\mathrm{d}x)$, where $\tilde{F}$ is a σ-finite measure on $(E, \mathcal{E})$. Also, $\tilde{\delta} : \Omega \times \mathbb{R}_+ \times E \to \mathbb{R}$ is a predictable function and $(S_k)_{k\geq 1}$ is a sequence of stopping times increasing to ∞ such that $|\tilde{\delta}(\omega, s, x)| \wedge 1 \leq \tilde{\psi}_k(x)$ for all (ω, s, z) with $s \leq S_k(\omega)$ and $\int_E \tilde{\psi}_k^2(x) \tilde{F}(\mathrm{d}x) < \infty$ for all $k \geq 1$.

We consider the power variation of the form

$$V(f)^n = \frac{1}{n} \sum_{i=1}^n f\big(\sqrt{n}\Delta_i^n X\big), \tag{2.5}$$

where $f = (f_1, f_2, ..., f_m)'$ is $\mathbb{R}^m$-valued function which satisfies the following assumption.

Assumption (K): The function $h : \mathbb{R} \mapsto \mathbb{R}$ is even and continuously differentiable. Moreover, both h and its derivative h' are at most of polynomial growth.

Note that a function $h : \mathbb{R} \mapsto \mathbb{R}$ is at most of polynomial growth if and only if

$$|h(x)| \leq C(1 + |x|^p),$$

for some positive numbers C and p.

Both Assumptions (**V**) and (**K**) are standard conditions for the validity of central limit theorems for classical high-frequency statistics. The following propositions, which are adapted from [15, Theorem 2.1 and 2.3], describe the limiting properties of power variation.

Proposition 2.24 (Law of large numbers)
Assume that X is a continuous Itô semimartingale as in (2.3) *and the function f is continuous with at most polynomial growth. Then, as $n \to \infty$, it holds that*

$$V(f)^n \xrightarrow{\mathbb{P}} V(f) = \int_0^1 \rho_{\sigma_s}(f) ds, \tag{2.6}$$

where $\rho_x(f) = \mathbb{E}[f(xU)]$ for $x \in \mathbb{R}$ and $U \sim N(0,1)$.

Proposition 2.25 (Central limit theorems)
Assume that X is a continuous Itô semimartingale as in (2.3)*, where the volatility process σ follows Assumption* (**V**) *and Assumption* (**K**) *holds true for each component of $f = (f_1, \dots, f_m)'$. Then, as $n \to \infty$, it holds that*

$$\sqrt{n}\Big(V(f)^n - V(f)\Big) \xrightarrow{d_{st}} MN(0, \Sigma), \tag{2.7}$$

where Σ is the $m \times m$ asymptotic conditional covariance matrix, which has elements

$$\Sigma_{ij} = \int_0^1 \Big[\rho_{\sigma_s}(f_i f_j) - \rho_{\sigma_s}(f_i)\rho_{\sigma_s}(f_j)\Big] \mathrm{d}s.$$

The above propositions provide the necessary foundation for making inference about power variation. It shows that $V(f)^n$ is consistent for $V(f)$. Moreover, the asymptotic distribution of $V(f)^n$ is mixed normal, i.e. it has a random variance-covariance matrix Σ, which is independent from the randomness of the normal distribution. To transform this into a probabilistic statement based on the standard normal distribution, it would be tempting to look at (2.7) and deduce that $\Sigma^{-1/2}\sqrt{n}\Big(V(f)^n - V(f)\Big) \xrightarrow{d} N(0, I_m)$, where I_m is the m-dimensional identity matrix.

We give a practical illustration for our interested quantity $V(f)^n$.

Example 2.26
Let us consider the 1-dimensional case with $f(x) = x^2$. We have that

$$V(f)^n = \sum_{i=1}^{n} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})^2,$$

which is the so-called realized variance (or realized volatility). The the law of large numbers (2.6) and the central limit theorem (2.7) above are translated to

$$V(f)^n \xrightarrow{\mathbb{P}} \int_0^1 \sigma_s^2 ds, \qquad and$$

$$\sqrt{n}(V(f)^n - V(f)) \xrightarrow{d_{st}} MN(0, 2\int_0^1 \sigma_s^4)ds,$$

where $\int_0^1 \sigma_s^4 ds$ is called *the integrated quarticity* (IQ). A feasible estimator of the IQ is the realized quarticity

$$IQ^n = \frac{n}{3}\sum_{i=1}^{n}(X_{\frac{i}{n}} - X_{\frac{i-1}{n}})^4,$$

from which it follows that

$$\frac{\sqrt{n}(V(f)^n - \int_0^1 \sigma_s^2 ds)}{\sqrt{2IQ^n}} \xrightarrow{d} N(0,1).$$

In practice, it is not simple to show stable convergence of processes. The next theorem introduced by Jacod [65] gives an important tool to derive stable limits in general setting for partial sums of triangular arrays.

Let $M = (M_t)_{t\geq 0}$ be a continuous d-dimensional local martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with $M_0 = 0$ and we define $\mathcal{M}_b(M^\perp) = \{$bounded Martingale $N|N_0 = 0$ and the covariation process $\langle M, N\rangle = 0\}$, i.e. the set of all bounded orthogonal martingales to M. We consider a sequence $(Y^n)_{n\geq 1}$ of q-dimensional semimartingales of the form

$$Y_t^n = \sum_{i=1}^{[nt]} X_{in}. \tag{2.8}$$

Theorem 2.27 (Jacod's stable central limit theorem ([65], Theorem IX.7.28))
Let M be a square integrable continuous local martingale and X_{in}'s be $\mathcal{F}_{\frac{i}{n}}$-measurable and square integrable random variables. Suppose that there exist continuous processes F and G with values in $\mathbb{R}^{q\times q}$ and $\mathbb{R}^{q\times d}$, respectively, and a continuous q-dimensional

process B of bounded variation such that

$$\sum_{i=1}^{[nt]} \mathbb{E}[X_{in}|\mathcal{F}_{\frac{i-1}{n}}] \xrightarrow{ucp} B_t \ \forall t, \tag{2.9}$$

$$\sum_{i=1}^{[nt]} \mathbb{E}[X_{in}X_{in}^T|\mathcal{F}_{\frac{i-1}{n}}] - \mathbb{E}[X_{in}|\mathcal{F}_{\frac{i-1}{n}}]\mathbb{E}^T[X_{in}|\mathcal{F}_{\frac{i-1}{n}}] \xrightarrow{\mathbb{P}} F_t \ \forall t, \tag{2.10}$$

$$\sum_{i=1}^{[nt]} \mathbb{E}[X_{in}(M_{\frac{i}{n}} - M_{\frac{i-1}{n}})^T|\mathcal{F}_{\frac{i-1}{n}}] \xrightarrow{\mathbb{P}} G_t \ \forall t, \tag{2.11}$$

$$\sum_{i=1}^{[nt]} \mathbb{E}[\|X_{in}\|^2 \mathbb{1}_{\{\|X_{in}\|>\epsilon\}}|\mathcal{F}_{\frac{i-1}{n}}] \xrightarrow{\mathbb{P}} 0 \ \forall t \ \forall \epsilon > 0, \tag{2.12}$$

$$\sum_{i=1}^{[nt]} \mathbb{E}[X_{in}(N_{\frac{i}{n}} - N_{\frac{i-1}{n}})^T|\mathcal{F}_{\frac{i-1}{n}}] \xrightarrow{\mathbb{P}} 0 \ \forall t \ \forall N \in \mathcal{M}_b(M^{\perp}). \tag{2.13}$$

We further assume that there exist predictable processes u, v and w with values in $\mathbb{R}^{q\times d}$, $\mathbb{R}^{d\times d}$ and $\mathbb{R}^{q\times q}$, respectively, such that

$$\langle M, M^T \rangle_t = \int_0^t u_s u_s^T ds \ \forall t, \tag{2.14}$$

$$G_t = \int_0^t v_s u_s u_s^T ds \ \forall t, \tag{2.15}$$

$$F_t = \int_0^t v_s u_s u_s^T v_s^T + w_s w_s^T ds \ \forall t. \tag{2.16}$$

Then,

$$Y_t^n \xrightarrow{d_{st}} Y_t = B_t + \int_0^t V_s dM_s + \int_0^t W_s dW'_s, \tag{2.17}$$

where W' is a Brownian motion defined on an extension of the original probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, which is independent of $\mathcal{F}$.

For the purpose of demonstration of how Theorem 2.27 works, let us give a classical example but useful in practice.

Example 2.28 ([87], Example 2.10)
We consider one-dimensional case. Let f, g be real valued continuous functions on $\mathbb{R}$, where g is at most of polynomial growth. We define

$$X_{in} = \frac{1}{\sqrt{n}} f(\sigma_{\frac{i-1}{n}})(g(\sqrt{n}\Delta_i^n W) - \mathbb{E}[g(\sqrt{n}\Delta_i^n W)]), \tag{2.18}$$

where σ is a càdlàg, adapted and bounded process and W is a Brownian motion. We consider the process Y defined by

$$Y_t^n = \sum_{i=1}^{[nt]} X_{in}. \tag{2.19}$$

To see the limiting process Y_t of Y_t^n, we have to check all conditions of (2.9)–(2.13).
For (2.9), since g is of polynomial growth, $\mathbb{E}[X_{in}|\mathcal{F}_{\frac{i-1}{n}}]$ exists and it is 0. Then, we obtain $B = 0$.
For (2.10), a simple computation shows that

$$\sum_{i=1}^{[nt]} \mathbb{E}[X_{in}^2|\mathcal{F}_{\frac{i-1}{n}}] \xrightarrow{\mathbb{P}} F_t := a^2 \int_0^t f^2(\sigma_s)ds,$$

where $a = var(f(U))$ and $U \sim N(0,1)$.
For (2.11), we choose $M = W$. Then, we obtian

$$\sum_{i=1}^{[nt]} \mathbb{E}[X_{in}(W_{\frac{i}{n}} - W_{\frac{i-1}{n}})|\mathcal{F}_{\frac{i-1}{n}}] \xrightarrow{\mathbb{P}} G_t := b \int_0^t f(\sigma_s)ds,$$

where $b = \mathbb{E}[g(U)U]$ and $U \sim N(0,1)$. Therefore,

$$w_s = \sqrt{a^2 - b^2} f(\sigma_s) \text{ and } v_s = bf(\sigma_s).$$

For (2.12), since σ is a bounded process, we have

$$\begin{aligned}
\sum_{i=1}^{[nt]} \mathbb{E}[X_{in}^2 \mathbb{1}_{\{|X_{in}|>\epsilon\}}|\mathcal{F}_{\frac{i-1}{n}}] &\le \frac{1}{\epsilon^2} \sum_{i=1}^{[nt]} \mathbb{E}[X_{in}^4|\mathcal{F}_{\frac{i-1}{n}}] \\
&\le C\frac{1}{n\epsilon^2} \qquad\qquad ,C > 0 \\
&\xrightarrow{\mathbb{P}} 0.
\end{aligned}$$

For (2.13), martingale representation theorem (see, (3.18)) implies that we can write

$$g(\sqrt{n}\Delta_i^n W) - \mathbb{E}[g(\sqrt{n}\Delta_i^n W)] = \int_{\frac{i-1}{n}}^{\frac{i}{n}} \phi_s^n dW_s,$$

for some process ϕ^n. Furthermore, employing the Itô isometry and the fact that $\langle W, N \rangle = 0$ yields

$$\begin{aligned}
\sum_{i=1}^{[nt]} \mathbb{E}[X_{in}(N_{\frac{i}{n}} - N_{\frac{i-1}{n}})|\mathcal{F}_{\frac{i-1}{n}}] &= \frac{1}{\sqrt{n}} f(\sigma_{\frac{i-1}{n}}) \mathbb{E}\Big(\int_{\frac{i-1}{n}}^{\frac{i}{n}} \phi_s^n dW_s \int_{\frac{i-1}{n}}^{\frac{i}{n}} dN_s \Big) \\
&= \frac{1}{\sqrt{n}} f(\sigma_{\frac{i-1}{n}}) \mathbb{E}\Big(\int_{\frac{i-1}{n}}^{\frac{i}{n}} \phi_s^n d\langle W, N \rangle \Big) \\
&= 0.
\end{aligned}$$

Applying Theorem 2.27, we conclude that

$$Y_t^n \xrightarrow{d_{st}} Y_t = b \int_0^t f(\sigma_s) dW_s + \sqrt{a^2 - b^2} \int_0^t f(\sigma_s) dW'_s,$$

where W' is a Brownian motion independent of $\mathcal{F}$.

Chapter 3

Elements of Malliavin calculus

We devote this section to introduction of some concepts and the main theoretical results of Malliavin calculus which will be used later in Chapter 4 and 5. The introduction relies on the approaches of Nualart [77] and Øksendal [80]. The further details and rigorous proofs can also be found in these references. Malliavin calculus allows us to compute derivatives of random variables. We provide some useful tools such as the chain rule and the integration by parts formula. To that end, we present the so-called Clark-Ocone formula which gives explicit form of the martingale representation theorem.

3.1 Isonormal Gaussian processes

The general framework of an isonormal Gaussian process associated with a Hilbert space H is defined here to introduce the Malliavin derivative in the following section.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and H a separable Hilbert space with the inner product and norm denoted by $\langle \cdot\,,\,\cdot \rangle_H$ and $\|.\|_H$, respectively.

Definition 3.1
A centered Gaussian family $W = \{W(h) | h \in H\}$ is called an *isonormal Gaussian process* on $(\Omega, \mathcal{F}, \mathbb{P})$ if

$$\mathbb{E}[W(h)W(g)] = \langle h, g \rangle_H$$

for all $h, g \in H$.

Example 3.2
Let $H := L^2([0,T])$ denote the Hilbert space of square integrable functions $f : [0,T] \to \mathbb{R}$. For each $h \in H$, we define the random variable

$$W(h) = \int_0^T h(t) dB_t$$

where B is a Brownian motion. We have $\mathbb{E}[W(h)] = 0$, and with the Itô isometry, we obtain

$$\mathbb{E}[W(h)W(g)] = \int_0^T h(t)g(t)dt = \langle h, g \rangle_H \,,$$

for all $h, g \in H$. Then, $W = \{W(h), h \in H\}$ is an isonormal Gaussian process.

Example 3.3
Let $H = \mathbb{R}^m$ for some $m \in \mathbb{N}$ and $(e_1, ..., e_m)$ be an orthonormal basis in $\mathbb{R}^m$ with respect to the standard Euclidean inner product. Then, for each $h \in H$, we can write

$$h = \sum_{i=1}^m c_i e_i$$

for coefficient $c_i = \langle e_i, h \rangle$. Let $\{Y_1, ..., Y_m\}$ be a set of i.i.d. random variables such that $Y_i \sim N(0, 1)$ for each i. We set

$$W(h) = \sum_{i=1}^m c_i Y_i.$$

Then, $W = \{W(h), h \in H\}$ is an isonormal Gaussian process.

The linearity and the existence of an isonormal Gaussian process are shown below.

Proposition 3.4 ([77])
Let H be a real separable Hilbert space. Then,

(i) *the mapping $h \mapsto W(h)$ is linear from H to a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$:*

$$\mathbb{E}[|W(\lambda h + \mu g) - \lambda W(h) - \mu W(g)|^2] = 0 \Rightarrow W(\lambda h + \mu g) = \lambda W(h) + \mu W(g) \ a.s.,$$

for any $\lambda, \mu \in \mathbb{R}$ and $h, g \in H$,

(ii) *there exists an H-isonormal Gaussian process W.*

Proof. We will sketch a proof only for (ii) based on an idea of Komolgorov's extension theorem. Since every separable Hilbert space has a countable orthonormal basis, let $(e_i)_{i\in\mathbb{N}}$ be a countable orthonormal basis of H. We denote μ the standard Gaussian measure on $\mathbb{R}$. Define a product space

$$(\Omega, \mathcal{F}, \mathbb{P}) := \Big(\prod_{i=1}^{\infty} \mathbb{R}, \otimes_{i=1}^{\infty} \mathcal{B}(R), \otimes_{i=1}^{\infty} \mu \Big).$$

By construction, the random variables π_n defined by $\pi_n((\omega_i)_{i\in\mathbb{N}}) := \omega_n$ are independent and Gaussian. We complete the proof by defining $W : H \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ such that $W(e_i) = \pi_i$. □

3.2 Malliavin derivative

Malliavin derivative allows us to compute derivatives of random variables. We first introduce a concept of tensor product, which will be used to define the Malliavin derivative.

Let H be a Hilbert space with scalar product $\langle \cdot\, , \cdot \rangle_H$. We define

$$H^{\otimes n} := span\{h_1 \otimes \cdots \otimes h_n \,|\, h_i \in H\},$$

where the tensor product $h_1 \otimes \cdots \otimes h_n$ is multilinear, i.e.

(i) $h_1 \otimes h_2 \cdots \otimes (h_i + g_i) \otimes \cdots \otimes h_n = h_1 \otimes h_2 \cdots \otimes h_i \otimes \cdots \otimes h_n + h_1 \otimes h_2 \cdots \otimes g_i \otimes \cdots \otimes h_n$

(ii) $h_1 \otimes h_2 \cdots \otimes (\lambda h_i) \otimes \cdots \otimes h_n = \lambda(h_1 \otimes h_2 \cdots \otimes h_n)$ for a constant λ.

The scalar product on $H^{\otimes n}$ is induced by

$$\langle e_{i_1} \otimes \cdots \otimes e_{i_n}\, , e_{j_1} \otimes \cdots \otimes e_{j_n} \rangle_{H^{\otimes n}} := \prod_{p=1}^{n} \langle e_{i_p}\, , e_{j_p} \rangle_H \, .$$

Let $C_p^\infty(\mathbb{R}^n)$ denote the space of infinitely differentiable functions such that all derivatives exhibit polynomial growth. The set of smooth random variables is introduced with

$$\mathcal{S} = \Big\{ F = f\big(W(h_1), \ldots, W(h_n)\big) \Big| n \geq 1, h_i \in H \text{ and } f \in C_p^\infty(\mathbb{R}^n) \Big\}.$$

Definition 3.5
The *kth order Malliavin derivative of* $F \in \mathcal{S}$, denoted by $D^k F$, is defined by

$$D^k F = \sum_{i_1, \ldots, i_k = 1}^{n} \frac{\partial^k}{\partial x_{i_1} \cdots \partial x_{i_k}} f\big(W(h_1), \ldots, W(h_n)\big) h_{i_1} \otimes \cdots \otimes h_{i_k}. \tag{3.1}$$

Note that the Malliavin derivative is closable from $(\Omega, \mathcal{F}, \mathbb{P})$ to $L^2(\Omega; H)$ where $L^2(\Omega; H)$ is a class of H-valued random elements Y that are $\mathcal{F}$-measurable and $\mathbb{E}[\|Y\|_H^2] < \infty$. In the setting of $H = L^2([0,T])$, this $L^2(\Omega; H)$ can be identified with $L^2([0,T], \Omega)$. We also note $DF(t) = D_t F$ and for notational ease by $L^2(\Omega)$ we always mean $L^2(\Omega, \mathcal{F}, \mathbb{P})$. We can extend the domain of the derivative D by considering the completion of the set $\mathcal{S}$.

Definition 3.6
Let $F \in \mathcal{S}$, we define the norm $\| \cdot \|_{k,q}$ by

$$\|F\|_{k,q} \equiv \left(\mathbb{E}\big[|F|^q\big] + \sum_{m=1}^{k} \mathbb{E}\big[\|D^m F\|_{H^{\otimes m}}^q\big] \right)^{1/q} .$$

Moreover, we define the Banach space $\mathbb{D}_{k,q}$ to be the completion of the set $\mathcal{S}$ with respect to the norm $\| \cdot \|_{k,q}$.

We have that if $p \leq q$ and $k \leq j$, then $\|F\|_{k,q} \leq \|F\|_{j,q}$ for any $F \in \mathcal{S}$. Therefore, if $k \geq 0$ and $p > q$, then $\mathbb{D}_{k+1,p} \subset \mathbb{D}_{k,q}$.

Example 3.7
Let B be a Brownian motion and $h \in H$

(i) if $F = W(h)$, then $DF = D(W(h)) = h$,

(ii) if $H = L^2([0,T])$ and $F = \int_0^T h(t)dB_t$, then $DF = (h_s)_{s\in[0,T]}$,

(iii) let $(Y_i)_{i\geq 1}$ is i.i.d. with $N(0,1)$-distributed and define

$$S_n = \sum_{i=1}^{n} f(Y_i) \text{ and } Y_i = W(h_i).$$

For $F = S_n$, then $D(S_n) = \sum_{i=1}^{n} f'(Y_i)h_i$.

As in classical calculus, linearity, product rule, chain rule and integration by parts still hold for Malliavin calculus.

Proposition 3.8 ([77])
Let $F, G \in \mathbb{D}_{1,2}$ and $h \in H$. It holds that

(i) *linearity:*

$$D(F+G) = D(F) + D(G), \tag{3.2}$$

(ii) *product rule:*

$$D(FG) = F \cdot D(G) + G \cdot D(F), \tag{3.3}$$

(iii) *chain rule: If $g \in C^1(\mathbb{R})$, then*

$$D(g(F)) = g'(F)DF, \tag{3.4}$$

(iv) *integration by parts formula:*

$$\mathbb{E}[\langle DF, h\rangle_H] = \mathbb{E}[FW(h)]. \tag{3.5}$$

Proof. We will give a proof only for (iv) and we refer to Nualart [77] for the rest. Without loss of generality, we assume that $h = e_1$, $F = f(W(e_1), ..., W(e_n))$ where $e_1, e_2, ..., e_n$

are orthonormal elements of H and $f \in C_p^\infty(\mathbb{R}^n)$. Denote $\varphi(x)$ the density of the standard normal distribution on $\mathbb{R}^n$. Then, we have

$$\begin{aligned}
\mathbb{E}[\langle DF, h\rangle_H] &= \mathbb{E}[\frac{\partial}{\partial x_1} f(W(e_1), ..., W(e_n))] \\
&= \int_{\mathbb{R}^n} \frac{\partial}{\partial x_1} f(x_1, ..., x_n)\varphi(x)dx \\
&= \int_{\mathbb{R}^n} x_1 f(x)\varphi(x)dx \\
&= \mathbb{E}[f(W(e_1), ..., W(e_n))W(e_1)] \\
&= \mathbb{E}[FW(h)].
\end{aligned}$$

□

Therefore, from Proposition 3.8 one can easily show that

$$\mathbb{E}[G\langle DF, h\rangle_H] + \mathbb{E}[F\langle DG, h\rangle_H] = \mathbb{E}[FGW(h)]. \tag{3.6}$$

The integration by parts formula will play an important role along this thesis.

We present the Malliavin calculus for diffusion processes in the next proposition.

Proposition 3.9 ([77], p.124)
If $(X_t)_{t\in[0,1]}$ is a solution of a stochastic differential equation (SDE)

$$dX_t = a(X_t)dt + \sigma(X_t)dB_t,$$

and $a, \sigma \in C^1(\mathbb{R})$ with bounded derivatives, then DX_t is given as the solution of the SDE

$$D_s X_t = \sigma(X_s) \exp\Big(\int_s^t (a' - \frac{1}{2}(\sigma')^2)(X_u)du + \int_s^t \sigma'(X_u)dB_u\Big), \tag{3.7}$$

for $s \le t$, and $D_s X_t = 0$, if $s > t$.

3.3 Wiener chaos

Definition 3.10
Let $H_n(x)$ denote the *nth Hermite polynomial*, which is defined by

$$H_n(x) = (-1)^n \exp(x^2/2)\frac{d^n}{dx^n}(\exp(-x^2/2)), \quad n \ge 1$$

and $H_0(x) = 1$.

The Hermite polynomials are the coefficients for the power expansion of

$$\begin{aligned}F(x,t) &= \exp(xt - \frac{t^2}{2})\\ &= \exp(\frac{x^2}{2})\exp(-\frac{1}{2}(x-t)^2)\\ &= \exp(\frac{x^2}{2})\sum_{n=0}^{\infty}\frac{t^n}{n!}(\frac{d^n}{dt^n}\exp(-\frac{1}{2}(x-t)^2))|_{t=0}\\ &= \sum_{n=0}^{\infty}\frac{t^n}{n!}H_n(x).\end{aligned}$$

For example, $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$ and $(x - \frac{d}{dx})H_n(x) = H_{n+1(x)}$.

The next lemma shows a relation between Hermite polynomials and Gaussian random variables.

Lemma 3.11 ([77], Lemma 1.1.1)
Let X and Y be jointly normally distributed with $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1$. Then, it holds that for all $n, m \geq 0$

$$\mathbb{E}[H_n(X)H_m(Y)] = \begin{cases} 0, & \text{if } n \neq m \\ n!(\mathbb{E}[XY])^n, & \text{if } n = m. \end{cases}$$

Proof. The generating function φ and the characteristic function of joint Gaussian random variables give

$$\mathbb{E}[\varphi(s,X)\varphi(t,Y)] = \mathbb{E}\Big[\exp\Big(sX - \frac{s^2}{2}\Big)\exp\Big(tY - \frac{t^2}{2}\Big)\Big] = \exp(st\mathbb{E}[XY]),$$

for any $s, t \in \mathbb{R}$. Taking the $(n+m)$th partial derivative at $s = t = 0$ yields

$$\mathbb{E}\Big[\frac{\partial^n}{\partial s^n}\varphi(x,X)|_{s=0}\frac{\partial^n}{\partial t^n}\varphi(t,Y)|_{t=0}\Big] = \frac{\partial^n}{\partial s^n}\exp(st\mathbb{E}[XY](s\mathbb{E}[XY]^m))|_{s=t=0}.$$

Then,

$$\mathbb{E}[H_n(X)H_m(Y)] = \begin{cases} 0, & \text{if } n \neq m \\ n!(\mathbb{E}[XY])^n, & \text{if } n = m. \end{cases}$$

□

We will denote the σ-field generated by the random variables $\{W(h), h \in H\}$ by $\mathcal{G}$.

Lemma 3.12 ([77], Lemma 1.1.2)
The random variables $\{\exp(W(h)), h \in H\}$ *form a total subset of* $L^2(\Omega, \mathcal{G}, \mathbb{P})$. *This means if* $X \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ *such that*

$$\mathbb{E}[X \exp(W(h))] = 0$$

for all $h \in H$, *then* $X = 0$ *almost surely.*

Let us now define the spaces

$$\mathcal{H}_n = \overline{span\{H_n(W(h)) \,|\, h \in H \text{ with } \|h\|_H = 1\}} \subset L^2(\Omega, \mathcal{G}, \mathbb{P}). \tag{3.8}$$

These are called the Wiener chaos of order n. We see that $\mathcal{H}_0$ is the set of constants while $\mathcal{H}_1$ coincides with the set of random variables $\{W(h), h \in H\}$. The previous Lemma 3.11 shows that the spaces $\mathcal{H}_n$ and $\mathcal{H}_m$ are orthogonal for $n \neq m$. We present an orthogonal decomposition of $L^2(\Omega, \mathcal{G}, \mathbb{P})$ in the next theorem.

Theorem 3.13 ([77], Theorem 1.1.1)
The space $L^2(\Omega, \mathcal{G}, \mathbb{P})$ *can be decomposed as an orthogonal sum of the subspaces* $\mathcal{H}_n$, $n \geq 0$, *i.e.*

$$L^2(\Omega, \mathcal{G}, \mathbb{P}) = \oplus_{n=0}^{\infty} \mathcal{H}_n. \tag{3.9}$$

The connection of Hermite polynomials and multiple stochastic integrals is provided in the following result.

Proposition 3.14 ([77], Proposition 1.1.4)
Let $H_n(x)$ *be the m-th Hermite polynomial, and let* $h \in H = L^2([0, T])$ *such that* $\|h\|_H = 1$. *Then, it holds that*

$$n! H_n(W(h)) = \int_{[0,T]^n} h(t_1) \cdots h(t_n) dB_{t_1} \cdots dB_{t_n},$$

where the integral in the right hand side of the equation is a multiple stochastic integral (see, [77, Section 1.1.2] for the precise definition).

We say that a function $g : [0, T]^n \to \mathbb{R}$ is symmetric if

$$g(t_{\sigma_1}, ..., t_{\sigma_n}) = g(t_1, ..., t_n)$$

for all permutation σ of the set $\{1, ..., n\}$. As a result, we have the following version of the Wiener chaos expansion.

Theorem 3.15 (Wiener Chaos representations ([77], Theorem 1.1.2))
Any square integrable random variable $F \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ admits the expansion

$$F = \sum_{n=0}^{\infty} I_n(f_n), \tag{3.10}$$

where $I_n(\cdot)$ is the multiple stochastic integral

$$I_n(f_n) = \int_{[0,T]^n} f_n(t_1, t_2, ..., t_n) dB_{t_1} \cdots dB_{t_n}$$

for $f_n \in L^2([0,T])$. Then, it holds that $f_0 = \mathbb{E}[F]$ with the identity mapping I_0. Moreover, we can assume that the functions $f_n \in L^2([0,T]^n)$ are uniquely determined by F and symmetric.

With the Chaos expansion (3.10), we can compute the Malliavin derivative of a square integrable random variable F easily by using the next proposition.

Proposition 3.16 ([80], Theorem 4.18)
Let $F \in \mathbb{D}_{1,2}$ be a square integrable random variable with Wiener chaos representation $\sum_{n=0}^{\infty} I_n(f_n)$. Then, $F \in \mathbb{D}_{1,2}$ if and only if

$$\sum_{n=1}^{\infty} nn! \, \|f_n\|^2_{L^2([0,T]^n)} < \infty \tag{3.11}$$

and in this case

$$D_t F = \sum_{i=1}^{n} n I_{n-1}(f_n(\cdot, t)). \tag{3.12}$$

To compute the Malliavin derivative of conditional expectations shown in Proposition 3.18, it requires the following auxiliary lemma.

Lemma 3.17 ([77], Lemma 1.2.5)
Let $H = L^2(A, \mathcal{A}, m)$ where $(A, \mathcal{A}, m)$ is a separable σ-finite, atomless measure space. Suppose that F is a square integrable random variable with Wiener chaos representation $\sum_{n=0}^{\infty} I_n(f_n)$. Let $G \in \mathcal{A}$. Then

$$\mathbb{E}[F|\mathcal{F}_G] = \sum_{n=0}^{\infty} I_n(f_n \mathbb{1}_G^{\otimes n}), \tag{3.13}$$

where $\mathcal{F}_G$ is the σ-field generated by the random variables $\{W(B) \,|\, B \subset G, B \in \mathcal{A}\}$.

Proof. Without loss of generality, we assume that $F = I_n(f_n)$ and $f_n = \mathbb{1}_{B_1\times\cdots\times B_n}$ where $B_1, ..., B_n$ are mutually disjoint sets of finite measure. Then,

$$\begin{aligned}\mathbb{E}[F|\mathcal{F}_G] &= \mathbb{E}[W(B_1)\cdots W(B_n)|\mathcal{F}_G]\\ &= \mathbb{E}\Big[\prod_{i=1}^{n}(W(B_1\cap G)+W(B_1\cap G^c))|\mathcal{F}_G\Big]\\ &= I_n(\mathbb{1}_{(B_1\cap G)\times\cdots\times(B_n\cap G)}).\end{aligned}$$

□

Proposition 3.18 ([80], Proposition 5.6)
Let $G \in \mathbb{D}_{1,2}$. It holds that for any $t \geq 0$

(i) $\mathbb{E}[G|\mathcal{F}_t] \in \mathbb{D}_{1,2}$ *and*

(ii) $D_s(\mathbb{E}[G|\mathcal{F}_t]) = \mathbb{E}[D_sG|\mathcal{F}_t]\mathbb{1}_{\{s\leq t\}}$.

As a consequence of Proposition 3.18, we can show that the Malliavin derivative at time t of an $\mathcal{F}_s$-adapted stochastic process is $\mathcal{F}_s$-adapted for all t.

Corollary 3.19 ([80], Corollary 5.7)
Let $\sigma_s(w)$ be an $\mathcal{F}_s$-adapted stochastic process. Suppose that $\sigma_s(\cdot) \in \mathbb{D}_{1,2}$ for all s. It holds that

(i) $D_t(\sigma_s(w))$ *is $\mathcal{F}_s$-adapted for all t and*

(ii) *for $t > s$,*

$$D_t(\sigma_s(w)) = 0.$$

Proof. We can complete the proofs for both (i) and (ii) by using Proposition 3.18. That is

$$\begin{aligned}D_t(\sigma_s(w)) &= D_t(\mathbb{E}[\sigma_s(w)|\mathcal{F}_s])\\ &= \mathbb{E}[D_t\sigma_s(w)|\mathcal{F}_s]\mathbb{1}_{[0,s]}(t).\end{aligned}$$

□

3.4 Divergence operator

This section introduces the divergence operator δ which is the adjoint of the Malliavin derivative. The Malliavin Derivative D is an unbounded operator from $L^2(\Omega)$ into $L^2(\Omega; H)$. Furthermore, the domain of D, $\mathbb{D}_{1,2}$, is dense in $L^2(\Omega)$ ([77]). Then, we can define its adjoint δ.

Definition 3.20
The operator δ is the adjoint of D such that the domain of δ denoted by Dom(δ) is the following set:

$$\text{Dom}(\delta) = \{u \in L^2(\Omega; H) | |\mathbb{E}[\langle DF, u\rangle_H]| \le c_u \|F\|_{L^2(\Omega)}, \forall F \in \mathbb{D}_{1,2}\},$$

and if $u \in$ Dom(δ), then $\delta(u)$ is uniquely characterized by the

$$\mathbb{E}[\langle DF, u\rangle_H] = \mathbb{E}[F\delta(u)] \tag{3.14}$$

for all $F \in \mathbb{D}_{1,2}$. δ is called *divergence operator*.

The equation (3.14) is also called *integration by parts formula*. Next, we give examples of elements in Dom(δ) when Hilbert spaces are provided.

Example 3.21
(i) Let $h \in H = L^2([0,T])$. Since $\mathbb{E}[\langle DF, h\rangle_H] = \mathbb{E}[FW(h)]$ for all $F \in \mathbb{D}_{1,2}$ (see, (3.5)),

$$\begin{aligned}\mathbb{E}\Big[\int_0^T D_sF \cdot h(s)ds\Big] &= \mathbb{E}[\langle DF, h\rangle_H] \\ &= \mathbb{E}[FW(h)] = \mathbb{E}\Big[F\int_0^T h(s)dB_s\Big].\end{aligned}$$

Therefore, $h \in$ Dom(δ) and

$$\delta(h) = \int_0^T h(s)dB_s.$$

Moreover, from the above computation, we get another version of the integration by parts formulas

$$\mathbb{E}\Big[F\int_0^T h(s)dB_s\Big] = \mathbb{E}\Big[\int_0^T D_sF \cdot h(s)ds\Big]. \tag{3.15}$$

(ii) Let $u = \sum_i^n F_ih_i$ where $F_i \in \mathcal{S}$ and $h_i \in H$. Employing the linearity (3.2), product rule (3.3) and integration by parts formula (3.5), for any $F \in \mathcal{S}$, we have

$$\begin{aligned}|\mathbb{E}[\langle DF, u\rangle_H]| &= \Big|\mathbb{E}[\langle DF, \sum_{i=1}^n F_ih_i\rangle_H]\Big| = \Big|\mathbb{E}[\sum_{i=1}^n F_i\langle DF, h_i\rangle_H]\Big| \\ &= \Big|\sum_{i=1}^n \mathbb{E}[FF_iW(h_i)] - \mathbb{E}[F\langle DF_i, h_i\rangle_H]\Big| \\ &\le C\|F\|_{L^2(\Omega)},\end{aligned}$$

for some constant C. Then, $u \in \text{Dom}(\delta)$. Furthermore, from the above computation, we obtain

$$\mathbb{E}[\langle DF, u\rangle_H] = \mathbb{E}[F\sum_{i=1}^{n}(F_iW(h_i) - \langle DF_i, h_i\rangle_H)].$$

Therefore, using (3.14), we can conclude that

$$\delta(u) = \sum_{i=1}^{n} F_iW(h_i) - \langle DF_i, h_i\rangle_H.$$

Moreover, we can derive a commutativity relationship between the divergence operator δ and the Malliavin derivative D.

Proposition 3.22 ([77],(1.46))
Let $F \in \mathcal{S}$ and $u = \sum_i^n F_ih_i$ where $F_i \in \mathcal{S}$ and $h_i \in H$. It holds that

$$\langle D(\delta(u)), h\rangle_H = \langle u, h\rangle_H + \delta(\langle Du, h\rangle_H).$$

The next result shows that we can factor out scalar random variables from the divergence.

Proposition 3.23 ([77], Proposition 1.3.3)
Let $F \in \mathbb{D}_{1,2}$ and $u \in$ Dom(δ) such that $Fu \in L^2(\Omega; H)$. Then, $Fu \in$ Dom(δ) and the equality

$$\delta(Fu) = F\delta(u) - \langle DF, u\rangle_H, \tag{3.16}$$

holds ture if the right-hand side of (3.16) *is square integrable.*

Proof. Suppose that J is a smooth random variable with compact support. By using (3.6), we have

$$\begin{aligned}\mathbb{E}[F\delta(Fu)] &= \mathbb{E}[\langle DJ, Fu\rangle_H]\\ &= \mathbb{E}[\langle D(FJ), u\rangle_H - J\langle DF, u\rangle_H]\\ &= \mathbb{E}[FJ\delta(u) - J\langle DF, u\rangle_H]\\ &= \mathbb{E}[J(F\delta(u) - \langle DF, u\rangle_H)].\end{aligned}$$

We are done. □

The following proposition expresses the divergence operator δ in terms of the Wiener chaos decomposition. We will see that the class Dom(δ) coincides with the subspace of $L^2([0,T] \times \Omega)$ under a certain condition.

Proposition 3.24 ([77], Proposition 1.3.7)
Let $u \in L^2([0,T] \times \Omega)$ *with Wiener chaos decomposition*

$$u(t) = \sum_{n=0}^{\infty} I_n(f_n(\cdot, t)).$$

Then, $u \in \mathrm{Dom}(\delta)$ *if and only if*

$$\delta(u) = \sum_{n=0}^{\infty} I_{n+1}(f_n) \tag{3.17}$$

converges in $L^2(\Omega)$.

3.5 Clark-Ocone formula

We recall the well-known martingale representation theorem from Itô calculus that under a certian condition a random variable can be written in terms of an Itô integral with respect to a Brownian motion.

Proposition 3.25 (Martingale representation theorem ([81], Theorem 4.3.4))
Let $0 \leq t \leq T$*,* B *be a Brownian motion on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* $(\mathcal{F}_t)_{t \geq 0}$ *be the filtration generated by this Brownian motion* B*. Let* X *be a martingale with respect to the filtration* $(\mathcal{F}_t)_{t \geq 0}$*. Then, there is an adapted process* ϕ *such that*

$$X_t = X_0 + \int_0^t \phi_u dB_u, \quad 0 \leq t \leq T. \tag{3.18}$$

The process ϕ in (3.18) is implicit. However, if we know that a process X belongs to $\mathbb{D}_{1,2}$, the adapted process ϕ can be identified as the conditional expectation of the Malliavin derivative of X. In other words, we derive an explicit form of the martingale representation theorem.

Proposition 3.26 (The Clark-Ocone formula ([77], Theorem 1.3.14))
Let $F \in \mathbb{D}_{1,2}$ *be* $\mathcal{F}_t$*-measurable. Then,*

$$F = \mathbb{E}[F] + \int_0^t \mathbb{E}[D_u F \mid \mathcal{F}_u] dB_u. \tag{3.19}$$

Proof. [77, Theorem 1.3.14]
Owing to Theorem 3.15, we can write $F = \sum_{n=0}^{\infty} I_n(f_n)$. Using (3.12) and (3.13), we

have

$$\begin{aligned}\mathbb{E}[D_t F|\mathcal{F}_t] &= \sum_{n=1}^{\infty} n\mathbb{E}[I_{n-1}(f_n(\cdot,t))|\mathcal{F}_t] \\ &= \sum_{n=1}^{\infty} nI_{n-1}(f_n(t_1,...,t_{n-1},t)\mathbb{1}_{\{t_1\vee\cdots\vee t_{n-1}<t\}}).\end{aligned}$$

Define $\phi_t = \mathbb{E}[D_t F|\mathcal{F}_t]$. By Proposition 3.24, the integral $\delta(\phi)$ is computed as

$$\delta(\phi) = \sum_{n=1}^{\infty} I_n(f_n) = F - E[F].$$

Since the process ϕ belongs to the class of measurable adapted processes, the integral $\delta(\phi)$ is an Itô integral of ϕ. We complete the proof. □

Chapter 4

Subsampling high-frequency data

In this chapter, we consider a scalar continuous Itô semimartingale $X = (X_t)_{t\geq 0}$ of the form

$$X_t = X_0 + \int_0^t a_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s, \tag{4.1}$$

where X_0 is the starting point, $a = (a_t)_{t\geq 0}$ is a predictable and locally bounded drift process, $\sigma = (\sigma_t)_{t\geq 0}$ is an adapted, càdlàg volatility process, while $W = (W_t)_{t\geq 0}$ is a standard Brownian motion. This process is defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual assumptions and adapted to $(\mathcal{F}_t)_{t\geq 0}$. It could represent the log-price of some financial security.

We are in the high-frequency setting. We suppose that historical data of X is available in the time frame $[0, 1]$. In this interval, we assume that X is recorded at equidistant time points $t_i = i/n$, for $i = 0, 1, \ldots n$, so that $n + 1$ is the total number of observations in the sample. We define the n increments of X as:

$$\Delta_i^n X = X_{i/n} - X_{(i-1)/n}, \quad \text{for } i = 1, \ldots, n. \tag{4.2}$$

The asymptotic theory we derive below is then infill, i.e. we are at some point going to let $n \to \infty$.

In econometrics, the quantity of interest is integrated functions of the diffusion coefficient

$$IV(f)_t = \int_0^t f(\sigma_s)\mathrm{d}s, \tag{4.3}$$

for some suitable function f. When $f(x) = x^2$, it is called integrated volatility.

The challenge is that the objects of interest appearing in (4.3) are latent, but they can be estimated from the available sample in high-frequency framework. A popular statistic, which is well-suited to do this, is the bipower or multipower variation introduced in [24].

It is based on the summation of products of the absolute value of adjacent high-frequency returns. Here, we adopt the more general definition of bipower variation from [15]:

$$V(f,g)^n = \frac{1}{n}\sum_{i=1}^{n-1} f\big(\sqrt{n}\Delta_i^n X\big) g\big(\sqrt{n}\Delta_{i+1}^n X\big), \tag{4.4}$$

where $f = (f_1, \dots, f_m)'$ and $g = (g_1, \dots, g_m)'$ are $\mathbb{R}^m$-valued functions. Note that in Eq. (4.4) the multiplication is understood to be done element-by-element. Indeed, we have already discussed a special case, power variation, in Section 2.3. In [15], the authors show a stochastic limit theorem for the statistic in (4.4).

Proposition 4.1 ([15], Theorem 2.1)
Assume that X is a continuous Itô semimartingale as in (4.1) *and the functions f, g are continuous with at most polynomial growth. Then, as $n \to \infty$, it holds that*

$$V(f,g)^n \xrightarrow{\mathbb{P}} V(f,g) = \int_0^1 \rho_{\sigma_s}(f)\rho_{\sigma_s}(g)ds \tag{4.5}$$

where $\rho_x(f) = \mathbb{E}[f(xU)]$ for $x \in \mathbb{R}$ and $U \sim N(0,1)$.

Furthermore, the authors also provide a central limit theorem for the associated statistic. To be able to acquire a central limit theorem, we recall an additional regularity condition on σ of the following type.

Assumption (V): σ is of the form:

$$\sigma_t = \sigma_0 + \int_0^t \tilde{a}_s \mathrm{d}s + \int_0^t \tilde{\sigma}_s \mathrm{d}W_s + \int_0^t \tilde{v}_s \mathrm{d}B_s$$
$$+ \int_0^t \int_E \tilde{\delta}(s,x) 1_{\{|\tilde{\delta}(s,x)| \le 1\}} (\tilde{\mu} - \tilde{\nu})(\mathrm{d}s, \mathrm{d}x) + \int_0^t \int_E \tilde{\delta}(s,x) 1_{\{|\tilde{\delta}(s,x)| > 1\}} \tilde{\mu}(\mathrm{d}s, \mathrm{d}x), \tag{4.6}$$

where σ_0 is its initial value, $\tilde{a} = (\tilde{a}_t)_{t\ge0}$, $\tilde{\sigma} = (\tilde{\sigma}_t)_{t\ge0}$ and $\tilde{v} = (\tilde{v}_t)_{t\ge0}$ are adapted, càdlàg stochastic processes, while $B = (B_t)_{t\ge0}$ is a standard Brownian motion that is independent of W. Furthermore, $(E, \mathcal{E})$ is a Polish space, $\tilde{\mu}$ is a random measure on $\mathbb{R}_+ \times E$, which is independent of (W, B) and has an intensity measure $\tilde{\nu}(\mathrm{d}s, \mathrm{d}x) = \mathrm{d}s\tilde{F}(\mathrm{d}x)$, where $\tilde{F}$ is a σ-finite measure on $(E, \mathcal{E})$. Also, $\tilde{\delta} : \Omega \times \mathbb{R}_+ \times E \to \mathbb{R}$ is a predictable function and $(S_k)_{k\ge1}$ is a sequence of stopping times increasing to ∞ such that $|\tilde{\delta}(\omega, s, x)| \wedge 1 \le \tilde{\psi}_k(x)$ for all (ω, s, z) with $s \le S_k(\omega)$ and $\int_E \tilde{\psi}_k^2(x)\tilde{F}(\mathrm{d}x) < \infty$ for all $k \ge 1$.

Moreover, we also recall a smoothness property of the functions f and g in (4.4), which we state with a generic function h.

Assumption (K): The function $h : \mathbb{R} \mapsto \mathbb{R}$ is even and continuously differentiable. Moreover, both h and its derivative h' are at most of polynomial growth.

Both Assumption (**V**) and (**K**) are standard conditions for the validity of central limit theorems for classical high-frequency statistics (see, [15]). The following proposition, which is adapted from that paper, then describes the limiting properties of bipower variation.

Proposition 4.2
Assume that X is a continuous Itô semimartingale as in (4.1)*, where the volatility process σ follows Assumption* (**V**) *and Assumption* (**K**) *holds true for each component of $f = (f_1, \ldots, f_m)'$ and $g = (g_1, \ldots, g_m)'$. Then, as $n \to \infty$, it holds that*

$$\sqrt{n}\Big(V(f,g)^n - V(f,g)\Big) \xrightarrow{d_{st}} MN(0,\Sigma). \tag{4.7}$$

Finally, Σ is the $m \times m$ asymptotic conditional covariance matrix, which has elements

$$\begin{aligned}\Sigma_{ij} = \int_0^1 \Big[&\rho_{\sigma_s}(f_i f_j)\rho_{\sigma_s}(g_i g_j) + \rho_{\sigma_s}(f_i)\rho_{\sigma_s}(g_j)\rho_{\sigma_s}(f_j g_i) \\ &+ \rho_{\sigma_s}(f_j)\rho_{\sigma_s}(g_i)\rho_{\sigma_s}(f_i g_j) - 3\rho_{\sigma_s}(f_i)\rho_{\sigma_s}(f_j)\rho_{\sigma_s}(g_i)\rho_{\sigma_s}(g_j)\Big]\mathrm{d}s.\end{aligned} \tag{4.8}$$

Proof. See [15]. □

In fact, the Proposition 4.2 can be extended to non-differentiable functions f and g by replacing Assumption (**K**) with Assumption (**H'**) and (**K'**) from [15]. More precisely,

Assumption (H'): The process $\sigma^2 > 0$.

Assumption (K'): The function f is even and continuously differentiable on the complement B^c of a closed subset $B \subset \mathbb{R}$, and satisfies

$$|y| \leq 1 \Rightarrow \|f(x+y) - f(x)\| \leq C(1 + |x|^p)|y|^r \tag{4.9}$$

for some constants $C > 0$, $p \geq 0$ and $r \in (0,1]$. Furthermore,

a) if $r = 1$, then B has Lebesgue measure 0,

b) if $r < 1$, then B satisfies
for any positive number C and any $N(0, C)$-random variable U, the distance $d(U, B)$ from U to B has a density ϕ_C on $\mathbb{R}_+$, such that

$$\sup_{x\in\mathbb{R}_+, |C|+|C^{-1}|\leq A} \phi_C(x) < \infty \tag{4.10}$$

for all $A < \infty$, and we have

$$x \in B^c, |y| \leq 1 \wedge \frac{d(x, B)}{2} \Rightarrow \begin{cases} \|\nabla f(x)\| \leq \frac{C(1+|x|^p)}{d(x,B)^{1-r}}, \\ \|\nabla f(x+y) - \nabla f(x)\| \leq \frac{C(1+|x|^p)|y|}{d(x,B)^{2-r}}. \end{cases} \tag{4.11}$$

Assumption (**H'**) is a technical condition used in the proof of the Proposition 4.2 for non-differentiable functions f and g, while Assumption (**K'**) puts suitable restrictions on the set, where f and g are not differentiable.

Example 4.3
A classical example, which is used intensively in applied work, is the original bipower variation of [24]. Let

$$f_i(x) = |x|^{q_i} \text{ and } g_i(x) = |x|^{r_i},$$

for $1 \leq i \leq m$ and $q_i, r_i \in (0, 1]$, which do not satisfy Assumption (**K**). However, Assumption (**K'**) is fulfilled. More precisely, the condition (4.9) is obviously satisfied for any $p \geq 0$. After observing that $B = \{0\}$, then it has Lebesgue measure 0 and satisfies (4.10). Since $d(x, B) = d(x, 0) = |x|$, (4.11) is satisfied. Still, if Assumption (**H'**) is also fulfilled, Proposition 4.2 holds. Then,

$$V(f_i, g_i)^n = \frac{1}{n}\sum_{i=1}^{n-1} |\Delta_i^n X|^{q_i} |\Delta_{i+1}^n X|^{r_i}, \quad V(f_i, g_i) = \mu_{q_i}\mu_{r_i} \int_0^1 |\sigma_s|^{q_i+r_i} \mathrm{d}s, \tag{4.12}$$

where $\mu_q = \mathbb{E}[|Z|^q]$ and $Z \sim N(0, 1)$. And Σ has the form:

$$\Sigma_{ij} = \left(\mu_{q_i+q_j}\mu_{r_i+r_j} + \mu_{q_i}\mu_{r_j}\mu_{q_j+r_i} + \mu_{q_j}\mu_{r_i}\mu_{q_i+r_j} - 3\mu_{q_i}\mu_{q_j}\mu_{r_i}\mu_{r_j}\right) \int_0^1 |\sigma_s|^{q_i+q_j+r_i+r_j} \mathrm{d}s. \tag{4.13}$$

The aim of this chapter is to construct an estimator of the asymptotic variance Σ based on a technique of subsampling of high-frequency data. This subsampling was suggested by Kalnina and Linton [69] and Kalnina [67] for the asymptotic variance of the two-scale realized variance, following earlier work in the classical time series literature [88, 89]. Our estimator is not only highly intuitive and simple to implement but also is positive semi-definite by construction which can be use to construct confidence intervals for $V(f,g)$ (see, remark 4.7). We will discuss later in Section 5.3 that some existing estimators fail to be positive definite.

This chapter is arranged as follows. The first section starts with illustrating how to build subsampling estimators for power variation and shows an optimal rate of convergence. Next, we generalize to the case of bipower variation. This can not be directly adapted from the power variation. We provide all proofs of the presented results in the last section.

4.1 Main results

4.1.1 Subsampling for power variation

In order to give an intuition for the general bipower (or multipower) variation case, we shall start with an easier case, the power variation, i.e.

$$V(f,1)^n = V(f)^n = \frac{1}{n}\sum_{i=1}^{n} f\big(\sqrt{n}\Delta_i^n X\big).$$

However, the result in this case is interesting in its own right compared to the general case because of the differences of convergence rates. We refer to Section 2.3 for the asymtotic results of $V(f)^n$.

Given discrete observations $X_0, X_{\frac{1}{n}}, X_{\frac{2}{n}}, \dots, X_1$, we construct the estimator with the following steps.

1. We split up all samples into L subsamples where L divides n. For each $1 \le l \le L$, l-th subsample is composed of the increments $\Delta^n_{(i-1)L+l}X$ where $1 \le i \le n/L$ (see, Figure 4.1).

2. Let $V_l(f)^n$ be the $m \times 1$ dimensional power variation type estimator computed on l-th subsample:

$$V_l(f)^n = \frac{1}{n/L}\sum_{i=1}^{n/L} f(\sqrt{n}\Delta^n_{(i-1)L+l}X).$$

Intuitively, under suitable conditions on L, $V_l(f)^n \xrightarrow{\mathbb{P}} V(f)$ and its asymptotic distribution (more or less) follows from Proposition 4.2, except that its rate of convergence is $(n/L)^{-1/2}$, i.e. $\sqrt{\frac{n}{L}}\Big(V_l(f)^n - V(f)\Big) \xrightarrow{d_{st}} MN(0,\Sigma)$. Moreover, as each subsample is based on non-overlapping increments, the $V_l(f)^n$'s are, asymptotically, conditionally independent. This suggests that by averaging the sum of outer products of $\sqrt{\frac{n}{L}}\Big(V_l(f)^n - V(f)\Big)$, we should get a consistent estimator of Σ.

3. We define

$$\Sigma_n = \frac{1}{L}\sum_{l=1}^{L}\left(\sqrt{\frac{n}{L}}(V_l(f)^n - V(f))\right)\left(\sqrt{\frac{n}{L}}(V_l(f)^n - V(f))\right)', \qquad (4.14)$$

where x' denotes the transpose of x. We can see that Σ_n in (4.14) is positive semi-definite by construction and should be a consistent estimator of Σ.

4. We replace the latent $V(f)$ by its estimator $V(f)^n$ that can be computed from data and get the following feasible version of Σ_n:

$$\hat{\Sigma}_n = \frac{1}{L}\sum_{l=1}^{L}\left(\sqrt{\frac{n}{L}}(V_l(f)^n - V(f)^n)\right)\left(\sqrt{\frac{n}{L}}(V_l(f)^n - V(f)^n)\right)'. \qquad (4.15)$$

This does not affect the asymptotics, because $V(f)^n$ converges much faster than $V_l(f)^n$.

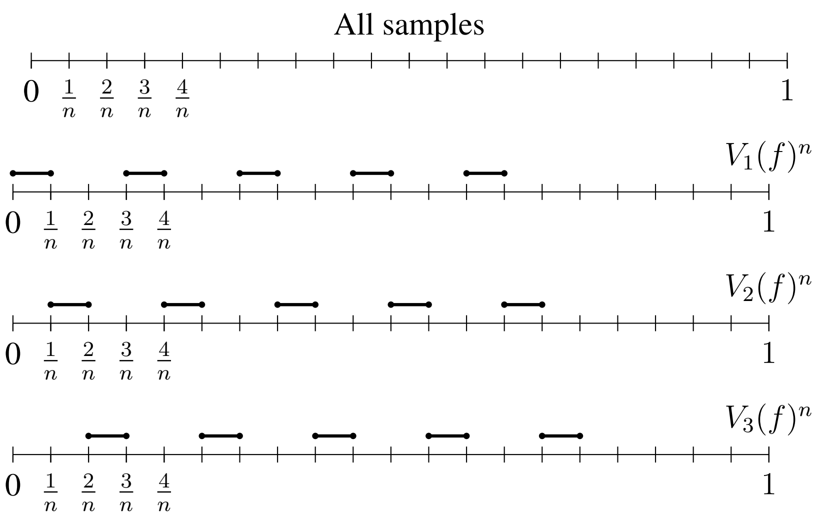


Figure 4.1: The construction of the subsampling estimator for $L = 3$.

For our asymptotic theory, we need to suppose an additional assumption, which is stronger than $(\mathbf{V})$, that the driving terms in both X and σ can be modeled as Brownian semimartingales:

Assumption (H): σ is continuous and follows Assumption (**V**), and each of a, $\tilde{a}$, $\tilde{\sigma}$ and $\tilde{v}$ is continuous of the form in (4.6).

Furthermore, we require a technical Assumption **(M)** that provides the Malliavin smoothness of the random variables appearing in Assumption (**H**).

Assumption (M): We assume that for any $0 \le t \le r \le s$: σ_s, $\tilde{\sigma}_s$, $\tilde{v}_s$, $D_r(\sigma_s)$, $D_r(\tilde{\sigma}_s)$, $D_r(\tilde{v}_s) \in \mathbb{D}_{1,2}$ and

$$\begin{aligned} \mathbb{E}[|D_t(\sigma_s)|^{32}] + \mathbb{E}|D_t(\tilde{\sigma}_s)|^{32}] + \mathbb{E}[|D_t(\tilde{v}_s)|^{32}] \le C, \\ \mathbb{E}[|D_t(D_r(\sigma_s))|^{16}] + \mathbb{E}[|D_t(D_r(\tilde{\sigma}_s))|^{16}] + \mathbb{E}[|D_t(D_r(\tilde{v}_s))|^{16}] \le C. \end{aligned} \tag{4.16}$$

Moreover, $f \in C^3(\mathbb{R})$, while f, f', f'' and f''' exbibit polynomial growth.

Remark 4.4
Assumption **(M)** is not a necessary condition to show consistency of our estimator but it will be used to derive an optimal rate of convergence. We refer to (4.66) for the computation of the Malliavin derivative in this case.

Now, we state the main result which leads to a convergence rate of $\hat{\Sigma}_n$.

Theorem 4.5
Assume that X is a continuous Itô semimartingale as in (4.1)*, where Assumption* (**H**) *holds true, as is Assumption* (**K**) *for each component of $f = (f_1, \dots, f_m)'$. Moreover, we assume that Assumption* (**M**) *is fulfilled. As $n \to \infty$, $L \to \infty$, and $n/L \to \infty$, it holds that*

$$\hat{\Sigma}_n - \Sigma = \underbrace{O_p\left(\frac{1}{\sqrt{L}}\right)}_{\text{CLT}} + \underbrace{O_p\left(\frac{L}{n}\right)}_{\text{blocking}}. \tag{4.17}$$

Proof. See Section 4.2. □

Theorem 4.5 presents the leading errors inherent in $\hat{\Sigma}_n$. The first term, $1/\sqrt{L}$, intuitively follows from a central limit theorem result, because $\hat{\Sigma}_n$ is an empirical mean of L asymptotically, conditionally independent statistics. However, it is not easy to apply this relationship for a formal derivation of the error rate. The second error is more subtle. It comes from freezing the volatility process at the beginning of a subblock of length L/n. If volatility is assumed to be Hölder continuous of order $\alpha \in (0, 1]$, a rough estimate implies an error rate of $(L/n)^\alpha$. However, due to the semimartingale structure of σ, we can improve this to L/n by applying a more refined estimation technique. As such, we should point out that the proof of Theorem 4.5 is much more complex compared to subsampling of i.i.d. observations.

By balancing both errors, we can find the fastest rate of convergence for our method. This requires:

$$L = O\big(n^{2/3}\big), \tag{4.18}$$

such that

$$\hat{\Sigma}_n - \Sigma = O_p\big(n^{-1/3}\big). \tag{4.19}$$

The stable central limit theorem of Proposition 4.2 is also valid for a non-differentiable function f, given that Assumptions **(H')** and **(K')** are satisfied. However, under these weaker conditions, it appears out of reach to derive a convergence rate for $\hat{\Sigma}_n$. We can nonetheless show that $\hat{\Sigma}_n$ still converges in probability to Σ, which is relevant for applied work.

Theorem 4.6
Assume that X is a continuous Itô semimartingale as in (4.1)*, where Assumption* (**V**) *holds true, as are Assumption* (**H'**) *and* (**K'**)*. As $n \to \infty$, $L \to \infty$, and $n/L \to \infty$, it holds that*

$$\hat{\Sigma}_n \xrightarrow{\mathbb{P}} \Sigma. \tag{4.20}$$

Proof. See Section 4.2. □

Remark 4.7
We can combine the consistency of $\hat{\Sigma}_n$ from Theorem 4.6 with the convergence in distribution in (4.7) to obtain a feasible central limit theorem. Applying the properties of stable convergence, we get the following feasible result:

$$\hat{\Sigma}_n^{-1/2}\sqrt{n}\Big(V(f)^n - V(f)\Big) \xrightarrow{d} N(0, I_m), \tag{4.21}$$

which can be used to construct confidence intervals for $V(f)$ or do hypothesis testing. If the convergence had not been stable in law, this result would not follow in general.

4.1.2 Subsampling for bipower variation

In the previous section, we presented a subsampling estimator for the asymptotic conditional covariance matrix of power variation. If we are interested in bipower (or multipower) variation, the theory derived there does not readily apply. This is because the summands in (4.4) are, asymptotically, 1-dependent, which the subsampling approach shown in Figure 4.1 does not adequately capture.

In order to consistently estimate Σ in the bipower case, we use an intuitive blocking approach, which is described next. We define the ith block of high-frequency data by taking:

$$B_i(p) = \Big\{ j : (i-1)p \leq j \leq ip \Big\}, \tag{4.22}$$

where $p \geq 2$ is an integer, and $i \geq 1$.

$B_i(p)$ is composed of adjacent observation time points of $X_{(i-1)p/n}, \ldots, X_{ip/n}$. From this, we can compute p consecutive returns $\Delta^n_{(i-1)p+1}X, \ldots, \Delta^n_{ip}X$. Therefore, $B_i(p)$ plays the role of the interval $[(i-1)/n, i/n]$ for power variation, which was used to compute a single return $\Delta^n_i X$. The only change is that we need to make this interval longer, such that we can consistently estimate the covariance structure of $V(f,g)^n$. As the $B_i(p)$'s are based on non-overlapping increments, it still holds that bipower variations computed from different subsamples are, asymptotically, conditionally independent.

We reset $\hat{\Sigma}_n$ as follows:

$$\hat{\Sigma}_n = \frac{1}{L} \sum_{l=1}^{L} \bigg(\sqrt{\frac{n}{L}} \Big(V_l(f,g)^n - V(f,g)^n \Big) \bigg) \bigg(\sqrt{\frac{n}{L}} \Big(V_l(f,g)^n - V(f,g)^n \Big) \bigg)', \tag{4.23}$$

where, assuming Lp divides n,

$$\begin{aligned} V_l(f,g)^n &= \frac{Lp}{n} \sum_{i=1}^{n/Lp} v_{(i-1)L+l}(f,g)^n, \\ v_i(f,g)^n &= \frac{1}{p-1} \sum_{j,j+1 \in B_i(p)} f\big(\sqrt{n}\Delta^n_j X\big) g\big(\sqrt{n}\Delta^n_{j+1} X\big). \end{aligned} \tag{4.24}$$

Note that n/Lp is the number of blocks assigned to each subsample, and that the subsample statistic $v_i(f,g)^n$ is computed only from data within the ith block $B_i(p)$. As in the above, we definitely require $n \to \infty$, $p \to \infty$, $L \to \infty$, and $n/pL \to \infty$ to prove the asymptotic theory for $\hat{\Sigma}_n$. It turns out, however, we need a slightly stronger condition for the last part to ensure consistency. This is because the rate $\sqrt{\frac{n}{L}}$ in the definition of (4.23) corresponds to the martingale part of $V_l(f,g)^n - V(f,g)^n$, while the statistic $V_l(f,g)^n - V(f,g)^n$ also has a bias term, which is of order Lp/n. Thus, to make the bias negligible with respect to the martingale part, we need $n/Lp^2 \to \infty$. Hence our "minimal" assumptions are based on this condition.

Theorem 4.8
Assume that X is a continuous Itô semimartingale as in (4.1)*, where Assumption* ($\mathbf{H}$) *holds true, as is Assumption* ($\mathbf{K}$) *for each component of $f = (f_1, \ldots, f_m)'$ and $g =$*

$(g_1, \ldots, g_m)'$. *Moreover, Assumption* $(\mathbf{M})$ *is fulfilled. As* $n \to \infty$, $p \to \infty$, $L \to \infty$, *and* $n/Lp^2 \to \infty$, *it holds that*

$$\hat{\Sigma}_n - \Sigma = \underbrace{O_p\left(\frac{1}{\sqrt{L}}\right)}_{\text{CLT}} + \underbrace{O_p\left(\frac{Lp^2}{n}\right)}_{\text{blocking}} + \underbrace{O_p\left(\frac{1}{p}\right)}_{\text{HAC}}. \tag{4.25}$$

Proof. See Section 4.2. □

The first two errors in (4.25) can be interpreted as those in Theorem 4.5, except the second is also affected by the block size p. Meanwhile, the decomposition of $\hat{\Sigma}_n - \Sigma$ in Theorem 4.8 has an extra error of order $O_p(1/p)$. The additional term, which emerges from the computation of the conditional variance of $v_i(f,g)^n$, has an intuitive interpretation. If we recall that in the current setting of bipower variation, the summands in (4.24) (or (4.4)) are asymptotically 1-dependent.

Let us consider the following stylized example. Assume that $(Z_i)_{i\geq 1}$ is a sequence of stationary 1-dependent random variates. Then,

$$\text{var}\left(\frac{1}{\sqrt{p}}\sum_{i=1}^{p} Z_i\right) = \text{var}(Z_1) + 2\frac{(p-1)}{p}\text{cov}(Z_1, Z_2) \to \text{var}(Z_1) + 2\text{cov}(Z_1, Z_2), \tag{4.26}$$

as $p \to \infty$.

This calculation shows that the finite sample variance on the left-hand side is not equal to, but converges towards, the asymptotic variance. The difference, i.e. the bias, is the term $-2\text{cov}(Z_1, Z_2)/p$, which has order $O(1/p)$. This example also helps to illustrate that Theorem 4.8 does not change, and in particular the convergence rate of $\hat{\Sigma}_n$ is unaffected, if we were to compute a higher order multipower variation statistic. Then there would be more covariance terms in (4.26), but the bias in each of them would still be $O(1/p)$.

The fastest rate is again found by balancing the errors, which means taking:

$$L = O(n^{2/5}), \qquad p = O(n^{1/5}), \tag{4.27}$$

for which

$$\hat{\Sigma}_n - \Sigma = O_p(n^{-1/5}). \tag{4.28}$$

Moreover, we note again that the consistency of $\hat{\Sigma}_n$ holds under the weaker assumptions and does not require Assumptions (**K**) and (**M**). We note that the condition $L/p \to \infty$ is necessary to deal with an additional bias term.

Theorem 4.9

Assume that X *is a continuous Itô semimartingale as in* (4.1)*, where Assumption* $(\mathbf{V})$

holds true, as are Assumption (**H'**) *and* (**K'**). *As* $n \to \infty$, $p \to \infty$ $L/p \to \infty$, *and* $n/Lp^2 \to \infty$, *it holds that*

$$\hat{\Sigma}_n \xrightarrow{\mathbb{P}} \Sigma. \tag{4.29}$$

Proof. See Section 4.2. □

To end this section, we should point out that for the power variation estimator covered by Theorem 4.5 in the previous subsection, it follows the work of [15] that there exists another consistent, positive semi-definite estimator of Σ:

$$\hat{S}_n = \frac{1}{2n}\sum_{i=1}^{n-1}\Big(f\big(\sqrt{n}\Delta_i^n X\big) - f\big(\sqrt{n}\Delta_{i+1}^n X\big)\Big)\Big(f\big(\sqrt{n}\Delta_i^n X\big) - f\big(\sqrt{n}\Delta_{i+1}^n X\big)\Big)'. \tag{4.30}$$

$\hat{S}_n$ has a better rate of convergence $n^{-1/2}$ compared to $n^{-1/3}$ derived in the previous section for $\hat{\Sigma}_n$. $\hat{S}_n$ is, therefore, more efficient for power variation, but it does not work for bi- or multipower variation.

4.1.3 Subsampling for truncated bipower variation

In an efficient market, equilibrium prices should adjust instantly to new information about fundamentals. If this leads to a significant revision of the fair value of the asset, the price has to move sharply and, potentially, discretely. This feature of price formation is not captured by the previous setup, where X has continuous sample paths. In this section, we therefore add a jump term to X and develop a framework for jump-robust inference about volatility based on subsampling truncated bipower variation [64, 74]. Accordingly, we assume that:

Assumption (J): X is of the form:

$$\begin{aligned} X_t = X_0 &+ \int_0^t a_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s \\ &+ \int_0^t \int_E \delta(s,x) 1_{\{|\delta(s,x)|\le 1\}} (\mu - \nu)(\mathrm{d}s, \mathrm{d}x) + \int_0^t \int_E \delta(s,x) 1_{\{|\delta(s,x)|>1\}} \mu(\mathrm{d}s, \mathrm{d}x), \end{aligned} \tag{4.31}$$

where X_0, $a = (a_t)_{t\ge 0}$, $\sigma = (\sigma_t)_{t\ge 0}$ and $W = (W_t)_{t\ge 0}$ are defined as in (4.1), while $(E, \mathcal{E})$ is a Polish space, μ is a random measure on $\mathbb{R}_+ \times E$ with compensator $\nu(\mathrm{d}s, \mathrm{d}x) = \mathrm{d}sF(\mathrm{d}x)$, where F is a σ-finite measure on $(E, \mathcal{E})$. Also, $\delta : \Omega \times \mathbb{R}_+ \times E \to \mathbb{R}$ is a predictable function and $(S_k)_{k\ge 1}$ is a sequence of stopping times increasing to ∞ such that $|\delta(\omega, s, x)| \wedge 1 \le \psi_k(x)$ for all (ω, s, z) with $s \le S_k(\omega)$ and $\int_E \psi_k^\beta(x) F(\mathrm{d}x) < \infty$

for all $k \geq 1$ and $\beta \in [0,1)$.

β relates to the activity index of the price jump process. The condition imposed on β implies that the jumps in X are (absolutely) summable, i.e. we restrict attention to jump processes with paths of finite variation, but, possibly, infinite activity.

Although the theory derived here should work with a general f and g, it requires a lot of notation. To develop ideas and maintain a streamlined exposition, we focus on the class of pure bipower variations in this section. The kth coordinate of the truncated bipower variation $\check{V}(q,r)^n$ is therefore:

$$\check{V}(q_k, r_k)^n = \frac{1}{n} \sum_{i=1}^{n-1} |\sqrt{n}\Delta_i^n \check{X}|^{q_k} |\sqrt{n}\Delta_{i+1}^n \check{X}|^{r_k}, \tag{4.32}$$

where $\Delta_i^n \check{X} = \Delta_i^n X \cdot \mathbb{1}_{\{|\Delta_i^n X| \leq u_n\}}$ is the increment after jump-trucation and the threshold level $u_n = \alpha n^{-\breve{\omega}}$ with $\alpha > 0$ and $\breve{\omega} \in (0, 1/2)$. By excluding the largest increments of X, the bipower variation statistic is, asymptotically, merely based on those high-frequency returns that are compatible with a continuous sample path model.

First, we recall the central limit theorem for $\check{V}(q,r)^n$.

Proposition 4.10
*Assume that X is a jump-diffusion process as in Assumption (**J**) and σ follows Assumption (**V**) with $\sigma > 0$. We denote by $s = 1 \wedge \min\{q_k, r_k : q_k > 0, r_k > 0, 1 \leq k \leq m\}$ and $s' = 1 \vee \max\{q_k, r_k : 1 \leq k \leq m\}$. Then, if $\beta \leq s$, $\breve{\omega} > \dfrac{s'-1}{2(s'-\beta)}$, and as $n \to \infty$, it holds that*

$$\sqrt{n}\Big(\check{V}(q,r)^n - V(q,r)\Big) \xrightarrow{d_{st}} \text{MN}(0, \Sigma), \tag{4.33}$$

where the elements of $V(q,r)$ and Σ are given as in (4.12) *and* (4.13).

Proof. See Theorem 13.2.1 and Example 13.2.2 in [64]. □

In the jump-diffusion setting, we define the subsample estimator of Σ as:

$$\hat{\Sigma}_n = \frac{1}{L} \sum_{l=1}^{L} \left(\sqrt{\frac{n}{L}} \Big(\check{V}_l(q,r)^n - \check{V}(q,r)^n \Big) \right) \left(\sqrt{\frac{n}{L}} \Big(\check{V}_l(q,r)^n - \check{V}(q,r)^n \Big) \right)', \tag{4.34}$$

where, assuming Lp divides n,

$$\begin{aligned} \check{V}_l(q_k, r_k)^n &= \frac{Lp}{n} \sum_{i=1}^{n/Lp} v_{(i-1)L+l}(q_k, r_k)^n, \\ v_i(q_k, r_k)^n &= \frac{1}{p-1} \sum_{j, j+1 \in B_i(p)} |\sqrt{n}\Delta_j^n \check{X}|^{q_k} |\sqrt{n}\Delta_{j+1}^n \check{X}|^{r_k}, \end{aligned} \tag{4.35}$$

and $B_i(p)$ is given as in (4.22).

Finally, we are ready to state a consistency result.

Theorem 4.11
Assume that X is a jump-diffusion process as in Assumption (**J**) *and σ follows Assumption* (**V**) *with $\sigma > 0$. Moreover, we require that $\beta \leq s$ and $\breve{\omega} > \frac{s'-1}{2(s'-\beta)}$. Then, as $n \to \infty$, $p \to \infty$, $L/p \to \infty$ and $n/Lp^2 \to \infty$, it holds that*

$$\hat{\Sigma}_n \xrightarrow{\mathbb{P}} \Sigma. \tag{4.36}$$

Proof. See Section 4.2. □

4.2 Proofs

First of all, we assume without loss of generality that the processes $a, \sigma, \tilde{a}, \tilde{\sigma}$ and $\tilde{v}$ are bounded following a standard localization procedure (see, [15]). Precisely, under (4.1) and Assumption (**V**), there exists a sequence of stopping times T_k incresing to ∞ a.s. and constants C_k such that

$$\|a_s\| + \|\sigma_{s-}\| + \|\tilde{a}_s\| + \|\tilde{\sigma}_{s-}\| + \|\tilde{v}_{s-}\| \leq C_k \quad \forall s \leq T_k.$$

We set

$$(\tilde{a}_s^{(k)}, \tilde{\sigma}_s^{(k)}, \tilde{v}_s^{(k)}) = \begin{cases} (\tilde{a}_s, \tilde{\sigma}_s, \tilde{v}_s) & \text{for } s \leq T_k \\ (0,0,0) & \text{for } s > T_k, \end{cases} \tag{4.37}$$

and set

$$\sigma_t^{(k)} = \sigma_0 + \int_0^t \tilde{a}_s^{(k)} \mathrm{d}s + \int_0^t \tilde{\sigma}_s^{(k)} \mathrm{d}W_s + \int_0^t \tilde{v}_s^{(k)} \mathrm{d}B_s,$$

i.e. $\sigma_t^{(k)} = \sigma_{s\wedge T_k}$. We set again $a_s^{(k)} = a_{s\wedge T_k}$ and associate $X^{(k)}$ with $a^{(k)}$ and $\sigma^{(k)}$ by

$$X_t^{(k)} = X_0 + \int_0^t a_s^{(k)} \mathrm{d}s + \int_0^t \sigma_s^{(k)} \mathrm{d}W_s,$$

and similarly $V(f,g)^{n,(k)}$, $V(f,g)_l^{n,(k)}$ and $\hat{\Sigma}_n^{(k)}$ with $X^{(k)}$ by (4.4), (4.24) and (4.23), respectively, and also $V(f,g)^{(k)}$ and $\Sigma^{(k)}$ with $\sigma^{(k)}$ by (4.5) and (4.7). We know that the localizing sequence T_k converges to infinity. Therefore, whenever (4.17), (4.20), (4.25) and (4.29) hold for a sequence of stopped processes, they also hold for the non-stopped process.

We denote by C or C_p (if dependent on a parameter p) a generic constant which might differ from line to line. And, due to the polarization identity

$$\mathrm{cov}(X,Y) = \frac{1}{4}(\mathrm{var}(X+Y) - \mathrm{var}(X-Y)),$$

we can (and shall) assume throughout that $m = 1$, so that all statistics are 1-dimensional.

4.2.1 Proof of Theorem 4.5

We start by observing an important approximation

$$\Delta_i^n X = \underbrace{\int_{\frac{(i-1)}{n}}^{\frac{i}{n}} a_s ds}_{O_p(1/n)} + \underbrace{\int_{\frac{(i-1)}{n}}^{\frac{i}{n}} \sigma_s dW_s}_{O_p(1/\sqrt{n})} .$$

We define

$$\alpha_i^n = \sqrt{n}\sigma_{\frac{i-1}{n}}\Delta_i^n W, \tag{4.38}$$

which is a first order approximation of $\sqrt{n}\Delta_i^n X$. Let us also denote

$$\chi_i^n = f(\alpha_i^n) - \mathbb{E}\left[f\left(\alpha_i^n\right)\Big|\mathcal{F}_{\frac{i-1}{n}}\right].$$

The next lemma can be shown easily by using Burkholder inequality.

Lemma 4.12
Let $p \geq 2$ and h be any function of polynomial growth. Then, we obtain

$$\mathbb{E}[|\alpha_i^n|^p] + \mathbb{E}[|\sqrt{n}\Delta_i^n X|^p] + \mathbb{E}[|h(\alpha_i^n)|^p] + \mathbb{E}[|\chi_i^n|^p] \leq C_p, \tag{4.39}$$

$$\mathbb{E}[|\sqrt{n}\Delta_i^n X - \alpha_i^n|^p] \leq C_p n^{-p/2}. \tag{4.40}$$

In the proofs, we will use Burkholder inequality several times. For any process Y of the form (4.1), we have that for any $p \geq 2$, then

$$\mathbb{E}[|Y_t - Y_s|^p] \leq C_p|t-s|^{p/2} \tag{4.41}$$

(see, Example 2.10 for the details). Following the comments above, the definition of $\hat{\Sigma}_n$ here collapses to:

$$\hat{\Sigma}_n = \frac{1}{L}\sum_{l=1}^{L}\left(\sqrt{\frac{n}{L}}\Big(V_l(f)^n - V(f)^n\Big)\right)^2,$$

To estimate $\hat{\Sigma}_n - \Sigma$, we introduce the following approximations:

$$\Sigma_n = \frac{1}{L}\sum_{l=1}^{L}\left(\sqrt{\frac{n}{L}}\Big(V_l(f)^n - V(f)\Big)\right)^2, \qquad Q_n = \frac{1}{n}\sum_{l=1}^{L}\left(\sum_{i=1}^{n/L}\chi^n_{(i-1)L+l}\right)^2,$$

$$U_n = \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n/L}\big(\chi^n_{(i-1)L+l}\big)^2, \qquad R_n = \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n/L}\mathbb{E}\left[\big(\chi^n_{(i-1)L+l}\big)^2 \mid \mathcal{F}_{\frac{(i-1)L+l-1}{n}}\right].$$

An intuitive explanation for these approximations is as follows. Firstly, $\hat{\Sigma}_n$ is approximated by Σ_n by replacing $V(f)^n$ with its limit $V(f)$. Secondly, since $\alpha^n_i \approx \sqrt{n}\Delta^n_i X$, for each subsample, $V_l(f)^n$ and $V(f)$ can be represented by $\frac{L}{n}\sum_{i=1}^{n/L} f(\alpha^n_{(i-1)L+l})$ and by an average $\frac{L}{n}\sum_{i=1}^{n/L}\mathbb{E}[f(\alpha^n_{(i-1)L+l})|\mathcal{F}_{\frac{i-1}{n}}]$, respectively. Therefore, Q_n estimates Σ_n. Thirdly, we consider the sum of square in U_n instead of the square of sum in Q_n. Next, we approximate U_n by its conditional expectation R_n, which should be the best guess for U_n. Finally, the Riemann sum R_n estimates Σ.

With these notations, it is easy to see that Theorem 4.5 is proven, if we show the following estimates:

Proposition 4.13
Under the conditions of Theorem 4.5, it holds

(i) $\mathbb{E}[|\Sigma_n - Q_n|] \le C\Big(\frac{L}{n} + \frac{1}{\sqrt{L}}\Big)$,

(ii) $\mathbb{E}[|Q_n - U_n|] \le \frac{C}{\sqrt{L}}$,

(iii) $\mathbb{E}[|U_n - R_n|] \le \frac{C}{\sqrt{n}}$,

(iv) $\mathbb{E}[|R_n - \Sigma|] \le \frac{C}{n}$,

(v) $\mathbb{E}[|\hat{\Sigma}_n - \Sigma_n|] \le \frac{C}{L}$.

We prove these estimates in the order (iii), (ii), (iv), (i) and (v) where (i) is the hardest step.

Proof of Proposition 4.13 (iii). This part is almost trivial. Indeed, we note that

$$U_n - R_n = \frac{1}{n}\sum_{i=1}^{n}\Big((\chi^n_i)^2 - \mathbb{E}[(\chi^n_i)^2\,|\mathcal{F}_{\frac{i-1}{n}}]\Big),$$

which is a sum of martingale differences. Moreover, Lemma 4.12 implies $\mathbb{E}[(\chi^n_i)^4] \le C$. Then, it is finished as

$$\mathbb{E}[|U_n - R_n|^2] \le C/n. \tag{4.42}$$

□

Proof of Proposition 4.13 (ii). We observe that

$$Q_n - U_n = \frac{1}{n}\sum_{l=1}^{L} A_l^n$$

where, for each l, we use the notations

$$\begin{aligned} A_l^n &= \left(\sum_{i=1}^{n/L} \chi^n_{(i-1)L+l}\right)^2 - \sum_{i=1}^{n/L} \left(\chi^n_{(i-1)L+l}\right)^2 = \left(S_l^{n/L}\right)^2 - T_l^{n/L} \\ &= \sum_{i,j=1, i\neq j}^{n/L} \chi^n_{(i-1)L+l}\chi^n_{(j-1)L+l}. \end{aligned}$$

Since $A^n_{l_1}$ and $A^n_{l_2}$ are uncorrelated for every $l_1 \neq l_2$, we obtain

$$\mathbb{E}[(Q_n - U_n)^2] = \frac{1}{n^2}\sum_{l=1}^{L} \mathbb{E}[(A_l^n)^2] \leq \frac{C}{n^2}\sum_{l=1}^{L}\left(\mathbb{E}[(S_l^{n/L})^4] + \mathbb{E}[(T_l^{n/L})^2]\right). \tag{4.43}$$

Define $S_l^m = \sum_{i=1}^m \chi^n_{(i-1)L+l}$. We observe that $(S_l^m)_{m=1}^{n/L}$ is a discrete martingale for each fixed l. Then, the discrete Burkholder and Cauchy-Schwarz inequalities and Lemma 4.12 imply

$$\mathbb{E}\left[\left(S_l^{n/L}\right)^4\right] \leq C\mathbb{E}\left[\left(\sum_{i=1}^{n/L} (\chi^n_{(i-1)L+l})^2\right)^2\right] \leq C\left(\frac{n}{L}\right)^2. \tag{4.44}$$

Using Lemma 4.12 again for $T_l^{n/L}$ term, we conclude the proof with

$$\mathbb{E}\left[(Q_n - U_n)^2\right] \leq C/L. \tag{4.45}$$

□

Proof of Proposition 4.13 (iv). We note that

$$\mathbb{E}\left[(\chi^n_{(i-1)L+l})^2 | \mathcal{F}_{\frac{(i-1)L+l-1}{n}}\right] = \rho_{\sigma_{\frac{(i-1)L+l-1}{n}}}(f^2) - \rho^2_{\sigma_{\frac{(i-1)L+l-1}{n}}}(f),$$

where ρ is introduced in Proposition 2.24. Hence R_n is a Riemann approximation of Σ because

$$R_n = \frac{1}{n}\sum_{i=1}^{n} \rho_{\sigma_{\frac{i-1}{n}}}(f^2) - \rho^2_{\sigma_{\frac{i-1}{n}}}(f).$$

Now, we suppress f and define $\tau(x) = \rho_x(f^2) - \rho_x^2(f)$. We note that the mapping

$$\phi(x) := \rho_x(f) = \int_{\mathbb{R}} \frac{f(y)}{\sqrt{2\pi x^2}} \exp\Big(\frac{-y^2}{2x^2}\Big) dy$$

is a smooth function, since f has polynomial growth. Hence τ is also smooth. Then, Taylor's theorem and (4.41) applied to σ imply

$$R_n - \Sigma = \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} \Big[\tau(\sigma_{\frac{i-1}{n}}) - \tau(\sigma_s)\Big] \mathrm{d}s = \sum_{i=1}^n \mu_i^n(1) + \sum_{i=1}^n \mu_i^n(2) + O_p\Big(\frac{1}{n}\Big),$$

where

$$\mu_i^n(1) = -\tau'\big(\sigma_{\frac{i-1}{n}}\big) \int_{\frac{i-1}{n}}^{\frac{i}{n}} \Big(\int_{\frac{i-1}{n}}^{s} \tilde{a}_u \mathrm{d}u\Big) \mathrm{d}s \quad \text{and}$$

$$\mu_i^n(2) = -\tau'\big(\sigma_{\frac{i-1}{n}}\big) \int_{\frac{i-1}{n}}^{\frac{i}{n}} \Big(\int_{\frac{i-1}{n}}^{s} \tilde{\sigma}_u \mathrm{d}W_u + \int_{\frac{i-1}{n}}^{s} \tilde{v}_u \mathrm{d}B_u\Big) \mathrm{d}s,$$

and the error $O_p(1/n)$ is due to the third order of Taylor's theorem. An application of Burkholder inequality and boundedness of $\tilde{a}, \tilde{\sigma}$ and $\tilde{v}$ yields

$$\mathbb{E}[|\mu_i^n(1)|^2] \le \frac{C}{n^4} \text{ and } \mathbb{E}\left[|\mu_i^n(2)|^2\right] \le \frac{C}{n^3}. \tag{4.46}$$

Using martingale difference property, this implies

$$\mathbb{E}\left[\left|\sum_{i=1}^n \mu_i^n(2)\right|^2\right] = \mathbb{E}\left[\sum_{i=1}^n |\mu_i^n(2)|^2\right] \le C\frac{1}{n^2}. \tag{4.47}$$

Now, we are done due to (4.46) – (4.47) and the Cauchy-Schwarz inequality. □

To prove Proposition 4.13 (i), we need a preparation. Let us denote

$$\tilde{V}_l(f)^n = \frac{1}{n/L} \sum_{i=1}^{n/L} f(\alpha^n_{(i-1)L+l}) \text{ and } \hat{V}_l(f)^n = \frac{1}{n/L} \sum_{i=1}^{n/L} \mathbb{E}\left[f(\alpha^n_{(i-1)L+l}) | \mathcal{F}_{\frac{(i-1)L+l-1}{n}}\right].$$

Using the decomposition

$$V_l(f)^n - V(f) = \Big(V_l(f)^n - \tilde{V}_l(f)^n\Big) + \Big(\tilde{V}_l(f)^n - \hat{V}_l(f)^n\Big) + \Big(\hat{V}_l(f)^n - V(f)\Big)$$

together with the identity $(a+b+c)^2 - b^2 = 2a(b+c) + 2cb + a^2 + c^2$, we obtain

$$\Sigma_n - Q_n = D_n^{(1)} + D_n^{(2)} + D_n^{(3)} + D_n^{(4)}$$

where

$$D_n^{(1)} = \frac{2n}{L^2} \sum_{l=1}^{L} \left(V_l(f)^n - \tilde{V}_l(f)^n \right) \left(\tilde{V}_l(f)^n - V(f) \right),$$
$$D_n^{(2)} = \frac{2n}{L^2} \sum_{l=1}^{L} \left(\hat{V}_l(f)^n - V(f) \right) \left(\tilde{V}_l(f)^n - \hat{V}_l(f)^n \right),$$
$$D_n^{(3)} = \frac{n}{L^2} \sum_{l=1}^{L} \left(V_l(f)^n - \tilde{V}_l(f)^n \right)^2,$$
$$D_n^{(4)} = \frac{n}{L^2} \sum_{l=1}^{L} \left(\hat{V}_l(f)^n - V(f) \right)^2.$$

To estimate these terms, we rely on the following preliminary result.

Lemma 4.14
Assume that the conditions of Theorem 4.5 are fulfilled. Then, uniformly in l:

(a) $\mathbb{E}[|\tilde{V}_l(f)^n - \hat{V}_l(f)^n|^2] \leq C\left(\frac{L}{n}\right)$,

(b) $\mathbb{E}[|\hat{V}_l(f)^n - V(f)|^2] \quad \leq C\left(\frac{L}{n}\right)^2$.

(c) $\mathbb{E}[|V_l(f)^n - \tilde{V}_l(f)^n|^2] \leq C\frac{L}{n^2}$.

Proof of lemma 4.14. Part (a) is shown by using the discrete Burkholder inequality as in (4.44). The proof of part (b) is similar to the proof of Proposition 4.13(iv). To prove part (c), we recall condition $(\mathbf{V})$ and write

$$\xi_i^n := \sqrt{n}\Delta_i^n X - \alpha_i^n = \sqrt{n}\left(\int_{\frac{i-1}{n}}^{\frac{i}{n}} a_s ds + \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left(\sigma_s - \sigma_{\frac{i-1}{n}} \right) dW_s \right) = \xi_i^n(1) + \xi_i^n(2).$$

where

$$\xi_i^n(1) = \sqrt{n}\left(a_{\frac{i-1}{n}} \frac{1}{n} + \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left[\tilde{\sigma}_{\frac{i-1}{n}} \left(W_s - W_{\frac{i-1}{n}} \right) + \tilde{v}_{\frac{i-1}{n}} \left(V_s - V_{\frac{i-1}{n}} \right) \right] dW_s \right),$$
$$\xi_i^n(2) = \sqrt{n}\Big(\int_{\frac{i-1}{n}}^{\frac{i}{n}} a_s - a_{\frac{i-1}{n}} ds + \int_{\frac{i-1}{n}}^{\frac{i}{n}} \int_{\frac{i-1}{n}}^{s} \tilde{a}_u du dW_s$$
$$+ \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left[\int_{\frac{i-1}{n}}^{s} \left(\tilde{\sigma}_u - \tilde{\sigma}_{\frac{i-1}{n}} \right) dW_u + \int_{\frac{i-1}{n}}^{s} \left(\tilde{v}_u - \tilde{v}_{\frac{i-1}{n}} \right) dV_u \right] dW_s \Big).$$

The Burkholder and Cauchy-Schwarz inequalities give us the inequalities

$$\mathbb{E}[|\xi_i^n(1)|^4] \leq \frac{C}{n^2}, \tag{4.48}$$

$$\mathbb{E}[|\xi_i^n(2)|^4] \leq C\beta_i^n, \tag{4.49}$$

$$\mathbb{E}[|\xi_i^n|^4] \leq \frac{C}{n^2}, \tag{4.50}$$

where

$$\beta_i^n = \frac{1}{n^4} + \frac{1}{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}} \mathbb{E}\left(|a_s - a_{\frac{i-1}{n}}|^4 + |\tilde{\sigma}_s - \tilde{\sigma}_{\frac{i-1}{n}}|^4 + |\tilde{v}_s - \tilde{v}_{\frac{i-1}{n}}|^4\right) ds.$$

Using Taylor's theorem, we may write $V_l(f)^n - \tilde{V}_l(f)^n = S_l^n(1) + S_l^n(2) + S_l^n(3)$, where

$$S_l^n(1) = \frac{1}{n/L}\sum_{i=1}^{n/L} f'(\alpha^n_{(i-1)L+l})\xi^n_{(i-1)L+l}(1),$$

$$S_l^n(2) = \frac{1}{n/L}\sum_{i=1}^{n/L} f'(\alpha^n_{(i-1)L+l})\xi^n_{(i-1)L+l}(2),$$

$$S_l^n(3) = \frac{1}{n/L}\sum_{i=1}^{n/L} [f'(\eta^n_{i,l}) - f'(\alpha^n_{(i-1)L+l})]\xi^n_{(i-1)L+l},$$

for some $|\eta^n_{i,l} - \alpha^n_{(i-1)L+l}| \leq |\xi^n_{(i-1)L+l}|$. Since f is even, f' is odd which implies the martingale difference property

$$\mathbb{E}[f'(\alpha^n_{(i-1)L+l})\xi^n_{(i-1)L+l}(1)|\mathcal{F}_{\frac{(i-1)L+l-1}{n}}] = 0.$$

Then, the Cauchy-Schwarz inequality, f' being of polynomial growth, Lemma 4.12 and (4.48) imply

$$\mathbb{E}[|S_l^n(1)|^2] = \frac{L^2}{n^2}\sum_{i=1}^{n/L}\mathbb{E}[|f'(\alpha^n_{(i-1)L+l})\xi^n_{(i-1)L+l}(1)|^2] \leq \frac{L}{n^2}. \tag{4.51}$$

Using Cauchy-Schwarz inequality and condition $(\mathbf{V})$ and (4.49), we obtain

$$\begin{aligned}\mathbb{E}\left[|S_l^n(2)|^2\right]^2 &\leq \frac{L^4}{n^4}\mathbb{E}\left[\left(\sum_{i=1}^{n/L}|f'(\alpha^n_{(i-1)L+l})|^2\right)^2\right]\mathbb{E}\left[\left(\sum_{i=1}^{n/L}|\xi^n_{(i-1)L+l}(2)|^2\right)^2\right] \\ &\leq C\frac{L}{n}\sum_{i=1}^{n/L}\beta^n_{(i-1)L+l} \leq C\frac{1}{n^4}.\end{aligned} \tag{4.52}$$

Employing the Cauchy-Schwarz inequality, twice differentiable property of f and using (4.50), we obtain

$$\mathbb{E}\left[|S_l^n(3)|^2\right]^2 \leq C\frac{L}{n^3}\sum_{i=1}^{n/L}\mathbb{E}\left[\left|f'(\eta_{i,l}^n) - f'(\alpha_{(i-1)L+l}^n)\right|^4\right] \leq C\frac{1}{n^4}. \tag{4.53}$$

Then, we finish the proof via (4.51)–(4.53). □

The next result then implies Proposition 4.13(i), and the entire proof is complete.

Lemma 4.15

Under the conditions of Theorem 4.5, we have

(a) $\mathbb{E}[|D_n^{(4)}|] \leq C\frac{L}{n}$,

(b) $\mathbb{E}[|D_n^{(3)}|] \leq C\frac{1}{n}$,

(c) $\mathbb{E}[|D_n^{(1)}|] \leq C\frac{1}{\sqrt{n}}$,

(d) $\mathbb{E}[|D_n^{(2)}|] \leq C\left(\frac{L}{n} + \frac{1}{\sqrt{n}}\right)$.

Proof of Lemma 4.15. We observe that part (a) is an obvious consequence of Lemma 4.14(b).

Concerning parts (b) and (c), we note that Lemma 4.14 implies

$$\mathbb{E}\left[\left(\tilde{V}_l(f)^n - V(f)\right)^2\right] \leq C\frac{L}{n}. \tag{4.54}$$

Then, the Cauchy-Schwarz inequality and the (4.54) yield

$$(\mathbb{E}[|D_n^{(1)}|])^2 \leq C\frac{n}{L^2}\sum_{l=1}^{L}\mathbb{E}[|V_l(f)^n - \tilde{V}_l(f)^n|^2] = C\mathbb{E}[|D_n^{(3)}|].$$

Hence it is enough to show part (b), which follows Lemma 4.14 (c).

Now, we proceed to the proof of part (d). An application of Taylor's theorem and (4.41) for σ permits us to write

$$D_n^{(2)} = E_n + F_n + O_p(L/n) + O_p(1/\sqrt{n}),$$

with

$$E_n = \frac{2n}{L^2}\sum_{l=1}^{L}\left(\sum_{i=1}^{n/L}\phi'(\sigma_{\frac{(i-1)L+l-1}{n}})\int_{\frac{(i-1)L+l-1}{n}}^{\frac{iL+l-1}{n}}\left[\sigma_{\frac{(i-1)L+l-1}{n}} - \sigma_s\right]ds\right) \times \left(\tilde{V}_l(f)^n - \hat{V}_l(f)^n\right),$$

$$F_n = \frac{-n}{L^2}\sum_{l=1}^{L}\left(\sum_{i=1}^{n/L}\int_{\frac{(i-1)L+l-1}{n}}^{\frac{iL+l-1}{n}}\phi''(\sigma_{\frac{(i-1)L+l-1}{n}})\left[\sigma_{\frac{(i-1)L+l-1}{n}} - \sigma_s\right]^2 ds\right) \times \left(\tilde{V}_l(f)^n - \hat{V}_l(f)^n\right).$$

We note that the error $O_p(L/n)$ appears here because of the third order term of Taylor's theorem, i.e.

$$\frac{n}{3L^2}\sum_{l=1}^{L}\left(\sum_{i=1}^{n/L}\int_{\frac{(i-1)L+l-1}{n}}^{\frac{iL+l-1}{n}}\phi'''(\eta_{s,i,l}^n)\left[\sigma_{\frac{(i-1)L+l-1}{n}}-\sigma_s\right]^3 ds\right)\times\left(\tilde{V}_l(f)^n-\hat{V}_l(f)^n\right)$$

for some $\eta_{s,i,l}^n$ satisfying $|\eta_{s,i,l}^n-\sigma_{\frac{(i-1)L+l-1}{n}}|\leq|\sigma_s-\sigma_{\frac{(i-1)L+l-1}{n}}|$, while the error $O_p(1/\sqrt{n})$ occurs due to the boundary integral term around 0 and 1, i.e.

$$\frac{2n}{L^2}\sum_{l=1}^{L}\left(\int_0^{\frac{l-1}{n}}\rho_{\sigma_s}(f)ds+\int_1^{1+\frac{l-1}{n}}\rho_{\sigma_{1+\frac{l-1-L}{n}}}(f)ds\right)\times\left(\tilde{V}_l(f)^n-\hat{V}_l(f)^n\right).$$

Recalling Assumption (**V**), we can rewrite

$$E_n=-(E_n(1)+E_n(2)+E_n(3)+E_n(4)),$$

where

$$E_n(1)=\frac{2n}{L^2}\sum_{l=1}^{L}\left(\sum_{i=1}^{n/L}\phi'(\sigma_{\frac{(i-1)L+l-1}{n}})\frac{L^2}{2n^2}\tilde{a}_{\frac{(i-1)L+l-1}{n}}\right)\left(\tilde{V}_l(f)^n-\hat{V}_l(f)^n\right),$$

$$E_n(2)=\frac{2n}{L^2}\sum_{l=1}^{L}\left(\sum_{i=1}^{n/L}\phi'(\sigma_{\frac{(i-1)L+l-1}{n}})\int_{\frac{(i-1)L+l-1}{n}}^{\frac{iL+l-1}{n}}\left[\tilde{\sigma}_{\frac{(i-1)L+l-1}{n}}(W_s-W_{\frac{(i-1)L+l-1}{n}})\right.\right.$$
$$\left.\left.+\tilde{v}_{\frac{(i-1)L+l-1}{n}}(V_s-V_{\frac{(i-1)L+l-1}{n}})\right]ds\right)\left(\tilde{V}_l(f)^n-\hat{V}_l(f)^n\right),$$

$$E_n(3)=\frac{2n}{L^2}\sum_{l=1}^{L}\left(\sum_{i=1}^{n/L}\phi'(\sigma_{\frac{(i-1)L+l-1}{n}})\int_{\frac{(i-1)L+l-1}{n}}^{\frac{iL+l-1}{n}}\int_{\frac{(i-1)L+l-1}{n}}^{s}(\tilde{a}_u-\tilde{a}_{\frac{(i-1)L+l-1}{n}})duds\right)$$
$$\times\left(\tilde{V}_l(f)^n-\hat{V}_l(f)^n\right),$$

$$E_n(4)=\frac{2n}{L^2}\sum_{l=1}^{L}\left(\sum_{i=1}^{n/L}\phi'(\sigma_{\frac{(i-1)L+l-1}{n}})\int_{\frac{(i-1)L+l-1}{n}}^{\frac{iL+l-1}{n}}\left[\int_{\frac{(i-1)L+l-1}{n}}^{s}(\tilde{\sigma}_u-\tilde{\sigma}_{\frac{(i-1)L+l-1}{n}})dW_u\right.\right.$$
$$\left.\left.+\int_{\frac{(i-1)L+l-1}{n}}^{s}(\tilde{v}_u-\tilde{v}_{\frac{(i-1)L+l-1}{n}})dV_u\right]ds\right)\left(\tilde{V}_l(f)^n-\hat{V}_l(f)^n\right).$$

Using Assumption (**H**), Cauchy-Schwarz inequality and (4.41) for $\tilde{a}$, $\tilde{\sigma}$ and $\tilde{v}$ imply that

$$\mathbb{E}[|E_n(3)|+|E_n(4)|]\leq C\Big(\frac{L}{n}\Big).$$

To deal with the term $E_n(1)$, we first define

$$Q_l^n=\frac{L}{n}\sum_{i=1}^{n/L}\phi'(\sigma_{\frac{(i-1)L+l-1}{n}})\tilde{a}_{\frac{(i-1)L+l-1}{n}}$$

and its limit $Q = \int_0^1 \phi'(\sigma_s)\tilde{a}_s ds$. We observe that

$$E_n(1) = \frac{1}{L}\sum_{l=1}^{L} Q(\tilde{V}_l(f)^n - \hat{V}_l(f)^n) + \frac{1}{L}\sum_{l=1}^{L}(Q_l^n - Q)(\tilde{V}_l(f)^n - \hat{V}_l(f)^n).$$

Therefore,

$$\begin{aligned}
\mathbb{E}[|E_n(1)|] &\leq \mathbb{E}\left[\left|\frac{1}{L}\sum_{l=1}^{L} Q(\tilde{V}_l(f)^n - \hat{V}_l(f)^n)\right|\right] + \mathbb{E}\left[\left|\frac{1}{L}\sum_{l=1}^{L}(Q_l^n - Q)(\tilde{V}_l(f)^n - \hat{V}_l(f)^n)\right|\right] \\
&\leq \mathbb{E}\left[\frac{1}{L^2}Q^2\left(\sum_{l=1}^{L}(\tilde{V}_l(f)^n - \hat{V}_l(f)^n)\right)^2\right]^{1/2} \\
&\quad + \frac{1}{L}\sum_{l=1}^{L}\mathbb{E}[|Q_l^n - Q|^2]^{1/2}\mathbb{E}[|\tilde{V}_l(f)^n - \hat{V}_l(f)^n|^2]^{1/2} \\
&\leq C\left(\frac{1}{L}\left(\mathbb{E}\Big[\Big(\sum_{l=1}^{L}\Big(\frac{1}{n/L}\sum_{i=1}^{n/L}\chi_{i,l}^n\Big)\Big)^4\Big]^{1/4}\right) + \sqrt{\frac{L}{n}}\mathbb{E}[|Q_l^n - Q|^2]^{1/2}\right) \\
&\leq C\left(\frac{1}{n}\mathbb{E}\Big[\Big(\sum_{i=1}^{n}\chi_i^n\Big)^4\Big]^{1/4} + \sqrt{\frac{L}{n}}\mathbb{E}[|Q_l^n - Q|^2]^{1/2}\right) \\
&\leq C\left(\frac{1}{\sqrt{n}} + \frac{L}{n}\right) \qquad (4.55)
\end{aligned}$$

where the martingale difference property of χ_i^n's and Assumption (**H**) are used to obtain the last inequality. The same decomposition as for $E_n(1)$ and the similar techniques applied to the term F_n yield

$$\mathbb{E}[|F_n|] \leq C\left(\frac{1}{\sqrt{n}} + \frac{L}{n}\right).$$

So, for the rest of the proof, we devote to the term $E_n(2)$ only. Here, we assume that $\tilde{v}_s = 0$. Apart from expositional purposes, this is without loss of generality, as the terms involving the product of $\tilde{v}$ and B are much simpler to handle, because W and B are independent. We define

$$G_{i,l}^n = \phi'(\sigma_{\frac{(i-1)L+l-1}{n}})\int_{\frac{(i-1)L+l-1}{n}}^{\frac{iL+l-1}{n}} \tilde{\sigma}_{\frac{(i-1)L+l-1}{n}}(W_s - W_{\frac{(i-1)L+l-1}{n}})ds. \qquad (4.56)$$

Then,

$$E_n(2) = \frac{2n}{L^2}\sum_{l=1}^{L}\left(\sum_{i=1}^{n/L} G_{i,l}^n\right)\left(\tilde{V}_l(f)^n - \hat{V}_l(f)^n\right). \qquad (4.57)$$

We find that

$$(E_n(2))^2 = E_n(2.1) + E_n(2.2), \tag{4.58}$$

where

$$E_n(2.1) = \frac{4n^2}{L^4}\sum_{l=1}^{L}\Big(\sum_{i=1}^{n/L} G^n_{i,l}\Big)^2 (\tilde{V}_l(f)^n - \hat{V}_l(f)^n)^2,$$

$$E_n(2.2) = \frac{4n^2}{L^4}\sum_{l_a\neq l_b}\Big(\sum_{i=1}^{n/L} G^n_{i,l_a}\Big)(\tilde{V}_{l_a}(f)^n - \hat{V}_{l_a}(f)^n)\Big(\sum_{i=1}^{n/L} G^n_{i,l_b}\Big)(\tilde{V}_{l_b}(f)^n - \hat{V}_{l_b}(f)^n).$$

The Burkholder and the Cauchy-Schwarz inequalities imply that

$$\mathbb{E}[|E_n(2.1)|] \leq \frac{C}{n}. \tag{4.59}$$

Recalling (4.58), we will finish with the proof of Lemma 4.15 (d) if we show

$$\mathbb{E}[E_n(2.2)] \leq \frac{C}{n}. \tag{4.60}$$

Proving (4.60) turns out to be rather complicated and involves tools from Malliavin calculus. Note that

$$E_n(2.2) = \frac{8}{L^2}\sum_{l_a\neq l_b}\sum_{i_1,i_2,i_3,i_4=1}^{n/L} G^n_{i_1,l_a}\chi^n_{(i_2-1)L+l_a} G^n_{i_3,l_b}\chi^n_{(i_4-1)L+l_b}. \tag{4.61}$$

Fix l_a and l_b in (4.61). For several choices of i_1, i_2, i_3 and i_4, we will have zero expectations. We recall that expected value of each G^n_i term and χ^n_i term is zero. Hence if we take conditional expectation on the left endpoint of the largest interval and if other three terms are measurable with respect to this point, the expectation will then vanish. Let $i = \max(i_1, i_2, i_3, i_4)$ and j denote the second largest element of these four numbers. If two of these numbers are equal to i, $j = i$ according to our definition. It is easy to observe that if $i - j > 1$, with the conditioning argument above, the expected value will be zero. This restriction on j means that the number of terms with non-zero expectation is smaller than Cn^3/L^3. We will separate the computation in to two cases, i.e.

$$1. \Big(\frac{(i_1-1)L + l_a - 1}{n}, \frac{i_1L + l_a - 1}{n}\Big) \cap \Big(\frac{(i_3-1)L + l_b - 1}{n}, \frac{i_3L + l_b - 1}{n}\Big) \neq \emptyset$$

$$2. \Big(\frac{(i_1-1)L + l_a - 1}{n}, \frac{i_1L + l_a - 1}{n}\Big) \cap \Big(\frac{(i_3-1)L + l_b - 1}{n}, \frac{i_3L + l_b - 1}{n}\Big) = \emptyset.$$

For the first case, without loss of generality, we assume in the following that $i_2 < i_4 < i_1 = i_3$. Next, we will find an upper bound on the expected values of non-zero terms. Since χ_i^n is in $\mathbb{D}_{1,2}$ for each i, the Clark-Ocone formula implies that

$$\chi^n_{(i-1)L+l} = \sqrt{n}\int_{\frac{(i-1)L+l-1}{n}}^{\frac{(i-1)L+l}{n}} \zeta_t^{n,i,l} dW_t$$

where $\zeta_s^{n,i,l} = \frac{1}{\sqrt{n}}\mathbb{E}[D_s(\chi_{(i-1)L+l})|\mathcal{F}_s]$ and $\frac{(i-1)L+l-1}{n} \leq s \leq \frac{(i-1)L+l}{n}$ (see, Proposition 3.26). Then,

$$E_n(2.2) = \frac{8}{L^2}\sum_{l_a\neq l_b}^{L}\sum_{i_1,i_2,i_3,i_4=1}^{n/L}\int_{\frac{(i_3-1)L+l_b-1}{n}}^{\frac{i_3L+l_b-1}{n}}\int_{\frac{(i_1-1)L+l_a-1}{n}}^{\frac{i_1L+l_a-1}{n}} M_n^{l_a,l_b,i_1,i_2,i_3,i_4}(s,u)dsdu. \quad (4.62)$$

where

$$\begin{aligned} M_n^{l_a,l_b,i_1,i_2,i_3,i_4}(s,u) &= \tau_n^{i_1,l_a,i_3,l_b}(W_s - W_{\frac{(i_1-1)L+l_a-1}{n}})(W_u - W_{\frac{(i_3-1)L+l_b-1}{n}}) \\ &\times n\int_{\frac{(i_4-1)L+l_b-1}{n}}^{\frac{(i_4-1)L+l_b}{n}}\Big(\int_{\frac{(i_2-1)L+l_a-1}{n}}^{\frac{(i_2-1)L+l_a}{n}} \zeta_t^{n,i_2,l_a}dW_t\Big)\zeta_r^{n,i_4,l_b}dW_r, \\ \tau_n^{i_1,l_a,i_3,l_b} &= \phi'(\sigma_{\frac{(i_1-1)L+l_a-1}{n}})\tilde{\sigma}_{\frac{(i_1-1)L+l_a-1}{n}}\phi'(\sigma_{\frac{(i_3-1)L+l_b-1}{n}})\tilde{\sigma}_{\frac{(i_3-1)L+l_b-1}{n}}. \end{aligned}$$

In the above expression, we assume that $(i_2-1)L+l_a \leq (i_4-1)L+l_b-1$. Otherwise, we will switch the order of the two integrations. In view of (4.62), estimating $\mathbb{E}[E_n(2.2)]$ amounts to estimating

$$I := \mathbb{E}\left[M_n^{l_a,l_b,i_1,i_2,i_3,i_4}(s,u)\right]$$

and it will be completed if we show $I \leq C(\frac{L}{n^2})$ uniformly in s and u. To accomplish this, we use the integration by parts formula of Malliavin calculus in Proposition 3.8(iv) twice and use the notation

$$C_a = \left[\frac{(i_2-1)L+l_a-1}{n}, \frac{(i_2-1)L+l_a}{n}\right] \text{ and}$$
$$C_b = \left[\frac{(i_4-1)L+l_b-1}{n}, \frac{(i_4-1)L+l_b}{n}\right],$$

we obtain

$$\begin{aligned} I &= n\mathbb{E}\Big[\int_{C_b} D_r\Big(\tau_n^{i_1,l_a,i_3,l_b}(W_s - W_{\frac{(i_1-1)L+l_a-1}{n}})(W_u - W_{\frac{(i_3-1)L+l_b-1}{n}})\Big) \\ &\quad\times\Big(\int_{C_a}\zeta_t^{n,i_2,l_a}dW_t\Big)\zeta_r^{n,i_4,l_b}dr\Big] \\ &= n\mathbb{E}\Big[\int_{C_b}\int_{C_a} D_t\Big(D_r\Big(\tau_n^{i_1,l_a,i_3,l_b}(W_s - W_{\frac{(i_1-1)L+l_a-1}{n}})(W_u - W_{\frac{(i_3-1)L+l_b-1}{n}})\Big) \\ &\quad\times\zeta_r^{n,i_4,l_b}\Big)\zeta_t^{n,i_2,l_a}dtdr\Big] \\ &\equiv n\int_{C_b}\int_{C_a}\mathbb{E}[H_{t,r}^n]\mathrm{d}t\mathrm{d}r. \end{aligned} \quad (4.63)$$

Since $D_t(W_s - W_{\frac{(i-1)L+l-1}{n}}) = \mathbb{1}_{[\frac{(i-1)L+l-1}{n},s]}(t)$ for any t and s, we can rewrite the random variable $H^n_{t,r}$ as the form

$$\begin{aligned}
H^n_{t,r} &= Z_1\big(W_s - W_{t_{i_1-1,l_a-1}}\big)\big(W_u - W_{t_{i_3-1,l_b-1}}\big)\\
&\quad + Z_2\Big(\big(W_s - W_{t_{i_1-1,l_a-1}}\big)\mathbb{1}_{[t_{i_3-1,l_b-1},u]}(t) + \big(W_u - W_{t_{i_3-1,l_b-1}}\big)\mathbb{1}_{[t_{i_1-1,l_a-1},s]}(t)\Big)\\
&\quad + Z_3\Big(\big(W_s - W_{t_{i_1-1,l_a-1}}\big)\mathbb{1}_{[t_{i_3-1,l_b-1},u]}(r) + \big(W_u - W_{t_{i_3-1,l_b-1}}\big)\mathbb{1}_{[t_{i_1-1,l_a-1},s]}(r)\Big)\\
&\quad + Z_4\Big(\mathbb{1}_{[t_{i_3-1,l_b-1},u]}(r)\mathbb{1}_{[t_{i_1-1,l_a-1},s]}(t) + \mathbb{1}_{[t_{i_3-1,l_b-1},u]}(t)\mathbb{1}_{[t_{i_1-1,l_a-1},s]}(r)\Big)\\
&= \sum_{k=1}^{4} H^{n,k}_{t,r}, \qquad (4.64)
\end{aligned}$$

where Z_i are random variables with uniformly bounded second moment and $t_{i,l} = (iL + l)/n$. We observe that:

$$\begin{aligned}
\mathbb{E}[|\zeta^{n,i_2,l_a}_t|^2] \le C,\\
\mathbb{E}[|\zeta^{n,i_4,l_b}_r|^4] \le C.
\end{aligned}$$

To see this fact: the product rule and chain rule of Malliavin derivative properties in Proposition 3.8 yield

$$D_t(\chi^n_{i_2,l_a}) = \sqrt{n} f'(\sqrt{n}\sigma_{\frac{(i_2-1)L+l_a-1}{n}}\Delta^n_{(i_2-1)L+l_a}W)(\sigma_{\frac{(i_2-1)L+l_a-1}{n}}\mathbb{1}_{[\frac{(i_2-1)L+l_a-1}{n},\frac{(i_2-1)L+l_a}{n}]}(t)), \qquad (4.65)$$

since $\frac{(i_2-1)L+l_a-1}{n} \le t$ and both $\mathbb{E}[f(\sqrt{n}\sigma_{\frac{(i_2-1)L+l_a-1}{n}}\Delta^n_{(i_2-1)L+l_a}W)|\mathcal{F}_{\frac{(i_2-1)L+l_a-1}{n}}]$ and $\sigma_{\frac{(i_2-1)L+l_a-1}{n}}$ are $\mathcal{F}_{\frac{(i_2-1)L+l_a-1}{n}}$-measurable. Then, we get

$$\begin{aligned}
\mathbb{E}[|\zeta^{n,i_2,l_a}_t|^2] &= \frac{1}{n}\mathbb{E}[\mathbb{E}^2[D_t(\chi^n_{i_2,l_a})|\mathcal{F}_t]]\\
&\le \frac{C}{n}\mathbb{E}[(D_t(\chi^n_{i_2,l_a}))^2]\\
&\le C.
\end{aligned}$$

The same argument is applied to $\mathbb{E}[|\zeta^{n,i_4,l_b}_r|^4]$. From these facts and Assumption (**M**), we readily deduce that

$$\mathbb{E}[|H^{n,1}_{t,r}|] \le C\frac{L}{n}. \qquad (4.66)$$

On the other hand, the term $\mathbb{1}_{[\frac{(i_1-1)L+l_a-1}{n},s]}(r)$ is different from 0 only for one index i_4 (this consideration also hold for all other indicator functions). Hence plugging $\mathbb{E}\big[H^{n,k}_{t,r}\big]$ back into (4.63), and observing the identity in (4.62), we find that $\mathbb{E}\big[|H^{n,1}_{t,r}|\big]$ is the dominating term. We, therefore, conclude that

$$\mathbb{E}\big[E_n(2.2)\big] \leq \frac{C}{n}.$$

For the second case, without loss of generality we assume that $i_2 < i_3 < i_1 = i_4$. Instead of applying the Malliavin derivative for the two Brownian increments in D_r as in the expression I above, only one Brownian increment is applied in this case, namely

$$H^n_{t,r} = D_t\Big(D_r\Big(\tau^{i_1,l_a,i_3,l_b}_n(W_u - W_{\frac{(i_3-1)L+l_b-1}{n}})\Big)(W_s - W_{\frac{(i_1-1)L+l_a-1}{n}})\zeta^{n,i_4,l_b}_r\Big)\zeta^{n,i_2,l_a}_t dtdr.$$

Again, we can rewrite $H^n_{t,r}$ as the form

$$\begin{aligned}
H^n_{t,r} &= Z_1\big(W_s - W_{t_{i_1-1,l_a-1}}\big)\big(W_u - W_{t_{i_3-1,l_b-1}}\big) \\
&\quad + Z_2\Big(\big(W_s - W_{t_{i_1-1,l_a-1}}\big)\mathbb{1}_{[t_{i_3-1,l_b-1},u]}(t) + \big(W_u - W_{t_{i_3-1,l_b-1}}\big)\mathbb{1}_{[t_{i_1-1,l_a-1},s]}(t)\Big) \\
&\quad + Z_3\big(W_s - W_{t_{i_1-1,l_a-1}}\big)\mathbb{1}_{[t_{i_3-1,l_b-1},u]}(r) \\
&\quad + Z_4\mathbb{1}_{[t_{i_3-1,l_b-1},u]}(r)\mathbb{1}_{[t_{i_1-1,l_a-1},s]}(t) \\
&= \sum_{k=1}^{4} H^{n,k}_{t,r}, \qquad (4.67)
\end{aligned}$$

where Z_i are random variables with uniformly bounded second moment. By considering as previous case, we completes the proof of (4.60), Proposition 4.13(i), and Theorem 4.5.

□

Proof of Proposition 4.13 (v). The identity

$$\sum_{l=1}^{n} V_l(f)^n = LV(f)^n$$

and a simple algebra imply

$$\hat{\Sigma}_n - \Sigma_n = \frac{-n}{L}(V(f)^n - V(f))^2.$$

Similar arguments as in Lemma 4.14 imply that

$$\mathbb{E}[(V(f)^n - V(f))^2] \leq C/n.$$

We complete the proof. □

4.2.2 Proof of Theorem 4.6

We need to show that the quantities which are introduced on the left side of (i)–(v) of Proposition 4.13 all converge to 0 under the weaker assumptions of Theorem 4.6. By observing the proof of Theorem 4.5, we discover that the steps behind (ii), (iii) and (v) do not depend on the stronger Assumption (**H**) and (**M**), nor on the differentiability of the function f. Hence, we can immediately deduce that

$$\mathbb{E}\big[|Q_n - U_n|\big] \to 0, \qquad \mathbb{E}\big[|U_n - R_n|\big] \to 0, \qquad \mathbb{E}\big[|\hat{\Sigma}_n - \Sigma_n|\big] \to 0.$$

On the other hand, (iv) follows from Section 8 (Step 2) in [15]:

$$R_n \xrightarrow{p} \Sigma.$$

We should note that, we can prove Lemma 4.14 (a) and (b) under the assumptions of Theorem 4.6. Consequently, we can show by applying Cauchy-Schwarz inequality that

$$\mathbb{E}[|D_n^{(2)}|] \le C\frac{\sqrt{L}}{\sqrt{n}} \quad \text{and}$$
$$\mathbb{E}[|D_n^{(4)}|] \le C\frac{L}{n}.$$

We found that to compute $\mathbb{E}[|D_n^{(2)}|]$ by using Cauchy-Schwarz inequality we get a worse bound than using Malliavin calculus's theorems.

However, Step 3 and 4 in Section 8 [15] imply that

$$\sup_{1\le l\le L} \mathbb{E}\left[\Big|\sqrt{\frac{n}{L}}\Big(V_l(f)^n - V(f)\Big) - \sqrt{\frac{L}{n}}\sum_{i=1}^{n/L}\chi^n_{(i-1)L+l}\Big|^{1+\epsilon}\right] \to 0.$$

for $\epsilon > 0$ small enough. Moreover, the discrete Burkholder inequality implies that

$$\mathbb{E}\left[\Big|\sqrt{\frac{n}{L}}\Big(V_l(f)^n - V(f)\Big) + \sqrt{\frac{L}{n}}\sum_{i=1}^{n/L}\chi^n_{(i-1)L+l}\Big|^{1+\frac{1}{\epsilon}}\right] \le C.$$

Hence, we find by Hölder inequality that

$$\mathbb{E}\big[|\Sigma_n - Q_n|\big] \to 0,$$

which corresponds to part (i) of Proposition 4.13 and finish the proof. □

4.2.3 Proofs of Theorems 4.8 and 4.9

We can prove Theorem 4.8 by employing the techniques that are used in the proof of Theorem 4.5 and 5.2 from the next chapter. Hence here we only sketch the main parts that enable us to find the convergence rate. We define:

$$\Sigma_n = \frac{1}{L}\sum_{l=1}^{L}\left(\sqrt{\frac{n}{L}}\,(V_l(f,g) - V(f,g))\right)^2, \quad Q_n = \frac{p^2}{n}\sum_{l=1}^{L}\left(\sum_{i=1}^{n/Lp}\chi^n_{(i-1)L+l}\right)^2,$$

$$U_n = \frac{p^2}{n}\sum_{l=1}^{L}\sum_{i=1}^{n/Lp}\left(\chi^n_{(i-1)L+l}\right)^2, \quad R_n = \frac{p^2}{n}\sum_{l=1}^{L}\sum_{i=1}^{n/Lp}\mathbb{E}\Big[\left(\chi^n_{(i-1)L+l}\right)^2 \mid \mathcal{F}_{\frac{(i-1)L+l-1}{n}}\Big],$$

with

$$\eta_i^n = \frac{1}{p-1}\sum_{m\in B_i(p)} f\Big(\sqrt{n}\sigma_{\frac{(i-1)p}{n}}\Delta_m^n W\Big)g\Big(\sqrt{n}\sigma_{\frac{(i-1)p}{n}}\Delta_{m+1}^n W\Big) \text{ and}$$

$$\chi_i^n = \eta_i^n - \mathbb{E}\Big[\eta_i^n \mid \mathcal{F}_{\frac{(i-1)p}{n}}\Big].$$

There exists a $C > 0$, independent of i, such that

$$\mathbb{E}\big[(\eta_i^n)^4\big] \le C \qquad \text{and} \qquad \mathbb{E}\big[(\chi_i^n)^4\big] \le \frac{C}{p^2}, \tag{4.68}$$

where the last inequality holds, because χ_i^n is a sum of 1-dependent random variates. Then, we are done due to the relationship

$$p/n \ll \sqrt{p/n} \ll p/\sqrt{n} \ll 1/\sqrt{L},$$

where $a \ll b$ means that a is much less than b, (the latter follows from $n/Lp^2 \to \infty$) and the following Lemma. We omit the proof because it is similar to the proof of Proposition 5.7.

Lemma 4.16
Assume that the conditions of Theorem 4.8 are fulfilled. Then, we get that

(i) $\mathbb{E}[|\Sigma_n - Q_n|] \le C\left(\frac{Lp^2}{n} + \frac{p}{\sqrt{n}}\right)$,

(ii) $\mathbb{E}[|Q_n - U_n|] \le \frac{C}{\sqrt{L}}$,

(iii) $\mathbb{E}[|U_n - R_n|] \le C\frac{\sqrt{p}}{\sqrt{n}}$,

(iv) $\mathbb{E}[|R_n - \Sigma|] \le C\Big(\frac{p}{n} + \frac{1}{p}\Big)$,

(v) $\mathbb{E}[|\hat{\Sigma}_n - \Sigma_n|] \leq \frac{C}{\sqrt{L}}$.

Lastly, the proof of Theorem 4.9 is analogous to the proofs of Theorems 4.6 and 5.3 and is omitted.

4.2.4 Proof of Theorem 4.11

We denote with X' the continuous part of X and introduce the following approximation of $\hat{\Sigma}_n$:

$$\hat{\Sigma}'_n = \frac{1}{L}\sum_{l=1}^{L}\left(\sqrt{\frac{n}{L}}\Big(V'_l(q,r)^n - V'(q,r)^n\Big)\right)^2,$$

where

$$V'(q,r)^n = \frac{1}{n}\sum_{i=1}^{n-1}|\sqrt{n}\Delta_i^n \check{X}'|^q|\sqrt{n}\Delta_{i+1}^n \check{X}'|^r,$$

$$V'_l(q,r)^n = \frac{Lp}{n}\sum_{i=1}^{n/Lp} v'_{(i-1)L+l}(q,r)^n,$$

$$v'_i(q,r)^n = \frac{1}{p-1}\sum_{j,j+1\in B_i(p)}|\sqrt{n}\Delta_j^n \check{X}'|^q|\sqrt{n}\Delta_{j+1}^n \check{X}'|^r.$$

The proof of Theorem 4.9 implies that $\hat{\Sigma}'_n \xrightarrow{p} \Sigma$. So, it suffices to show that $\hat{\Sigma}_n - \hat{\Sigma}'_n \xrightarrow{p} 0$. Note that

$$\begin{aligned}\hat{\Sigma}_n - \hat{\Sigma}'_n = \frac{1}{L}\sum_{l=1}^{L}&\left(\sqrt{\frac{n}{L}}\Big(V_l(q,r)^n - V'_l(q,r)^n + V'(q,r)^n - V(q,r)^n\Big)\right)\\ &\times\left(\sqrt{\frac{n}{L}}\Big(V_l(q,r)^n - V(q,r)^n + V'_l(q,r)^n - V'(q,r)^n\Big)\right).\end{aligned}$$

For any $j \geq 1$, we set:

$$\bar{\eta}_j^n = |\sqrt{n}\Delta_j^n \check{X}|^q|\sqrt{n}\Delta_{j+1}^n \check{X}|^r - |\sqrt{n}\Delta_j^n \check{X}'|^q|\sqrt{n}\Delta_{j+1}^n \check{X}'|^r.$$

Applying (13.2.21) from [64] with $m = 1+\epsilon$ and $\theta = 0$, we find that:

$$\mathbb{E}\big[|\bar{\eta}_j^n|^{1+\epsilon}|\big] \leq \frac{1}{n^{(1+\epsilon)/2}}\phi_n,$$

uniformly in j, for some sequence ϕ_n going to 0 and $\epsilon \in (0, 1-\beta]$, and under $\breve{\omega} \geq (ms' + \epsilon - 1)/2(ms' - \beta)$. Then, the discrete Hölder inequality implies that

$$\sup_{1 \leq l \leq L} \mathbb{E}\left[\left|\sqrt{\frac{n}{L}}\Big(V_l(q,r)^n - V_l'(q,r)^n\Big)\right|^{1+\epsilon}\right] \to 0$$

and

$$\sup_{1 \leq l \leq L} \mathbb{E}\left[\left|\sqrt{\frac{n}{L}}\Big(V(q,r)^n - V'(q,r)^n\Big)\right|^{1+\epsilon}\right] \to 0$$

Applying the arguments of Lemma 4.14, we also have that

$$\sup_{1 \leq l \leq L} \mathbb{E}\left[\left|\sqrt{\frac{n}{L}}\Big(V_l'(q,r)^n - V'(q,r)^n\Big)\right|^{1+\frac{1}{\epsilon}}\right] + \sup_{1 \leq l \leq L} \mathbb{E}\left[\left|\sqrt{\frac{n}{L}}\Big(V_l(q,r)^n - V(q,r)^n\Big)\right|^{1+\frac{1}{\epsilon}}\right] \leq C.$$

Therefore, again by the Hölder inequality,

$$\mathbb{E}\Big[|\hat{\Sigma}_n - \hat{\Sigma}_n'|\Big] \to 0.$$

As $\epsilon > 0$ can be chosen as small as possible, the proof is complete. □

Chapter 5

Subsampling for the bipower-type pre-averaging estimators

In this chapter, we present how to apply the idea of subsampling method introduced in the previous chapter to use for an underlying model perturbed by an additive noise term, i.e.

$$Y_{\frac{i}{n}} = X_{\frac{i}{n}} + \epsilon_{\frac{i}{n}}, \tag{5.1}$$

where X is defined as in (4.1), while $\epsilon = (\epsilon_t)_{t\geq 0}$ is a microstructure noise explained in the next section. The main goal of this chapter is to construct consistent estimators for the asymptotic variance obtained from a central limit theorem when the model in (5.1) is considered and to see how microstructure noise affects the speed of convergence.

The structure of this chapter is as follows: the first section starts with the introduction of microstructure noise. To construct our estimator, we provide pre-averaging method to get rid of the impact of noise. We also state asymptotic results of bipower variation in the presence of noise, i.e. law of large numbers and a central limit theorem, in this section. The next section presents the main results which we adapt our subsampling to obtain a positive semi-definite estimator and derive a convergent rate. Finally, we devote the last section to the proofs of the results.

5.1 Microsturcture noise

In order to find a consistent estimator for the integrated volatility $IV = \int_0^1 \sigma_s^2 ds$ in the noiseless case, the usual realized volatility of X is the right one as we have shown in Example 2.26. However, empirical studies show that this usual realized volatility explodes when using high frequency data (see, [4, 5, 7, 78]). The source of this error is known as *microstructure noise*. In other words, the observed prices are contaminated by microstructure noise such as bid-ask spreads, price discreteness, and so forth [36, 73, 93].

From the mathematical point of view, Zhou [102] suggested to econometrics that the observed price process should be in the form

$$Y_{\frac{i}{n}} = X_{\frac{i}{n}} + \epsilon_{\frac{i}{n}}, \tag{5.2}$$

where X is the latent process in the form of (4.1) and ϵ is the microstructure noise satisfying the following assumption:

Assumption (N):

(i) ϵ is i.i.d. with $\mathbb{E}[\epsilon_t] = 0$ and $\text{var}(\epsilon_t) = \omega^2$ for all $t \geq 0$.

(ii) ϵ is independent of X.

(iii) The distribution of ϵ is symmetric around 0.

(iv) $\mathbb{E}\big[|\epsilon_t|^s\big] < \infty$ for some $s > 0$.

Let us consider the realized variance of the process Y:

$$\begin{aligned} RV^n &= \sum_{i=1}^{n} (Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}})^2 \\ &= \sum_{i=1}^{n} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})^2 + 2\sum_{i=1}^{n} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})(\epsilon_{\frac{i}{n}} - \epsilon_{\frac{i-1}{n}}) + \sum_{i=1}^{n} (\epsilon_{\frac{i}{n}} - \epsilon_{\frac{i-1}{n}})^2. \end{aligned}$$

Although the limit of the first term is $IV = \int_0^1 \sigma_s^2 ds$, the stochastic orders of the second and the third term are 1 and n, respectively. This means that the observed return process is dominated by the noise term. Hence RV^n is inconsistent for IV.

To handle the market microstructure effects in estimating integrated volatility IV, Zhang et al. [101] introduce two-scale realized volatility (TSRV) method to deal with this problem. This estimator combines the realized volatility sampled from two different time scales. Define

$$[Y, Y]^{(n,K)} = \frac{1}{K} \sum_{i=K}^{n} (Y_{\frac{i}{n}} - Y_{\frac{i-K}{n}})^2,$$

with K being a positive integer. Note that $[Y, Y]^{(n,1)}$ is a usual realized variance. The two-scales realized volatility is given by

$$< \widehat{X, X} >_n^{TSRV} = [Y, Y]^{(n,K)} - 2\frac{n - K + 1}{nK} [Y, Y]^{(n,1)}.$$

Its asymptotic behavior is derived when $K, n \to \infty$. The estimator $< \widehat{X, X} >_n^{TSRV}$ averages the squared returns from sampling every Kth data point and those from every data point. The authors show that the TSRV estimator is consistent, asymptotically unbiased, and normal and found that $K = cn^{2/3}$ is the optimal choice for some $c > 0$, this leads to a stable central limit:

$$n^{1/6}\Big(< \widehat{X, X} >_n^{TSRV} - \int_0^1 \sigma_s^2 ds\Big) \xrightarrow{d_{st}} \int_0^1 \Big(\frac{8}{c^2}\omega^4 + \frac{4c}{3}\sigma_s^4\Big)^{1/2} dW_s', \tag{5.3}$$

where W' is a Brownian motion, independent of the σ-field $\mathcal{F}$.

This TSRV approach was extended to define a multi-scale realized variance estimator introduced by Zhang [100]:

$$< \widehat{X, X} >_n^{MSRV} = \sum_{i=1}^{M} a_i [Y, Y]^{(n,K)},$$

where $M > 2$ and a_i are the weights chosen to get the estimator asymptotically unbiased and to derive the optimal rate of convergence. This new estimator is based on the same subsampling and averaging idea of TSRV estimator but achieves a better optimal rate $n^{-1/4}$. A drawback of this multi-scale approach is that we still do not have a feasible version, which will allow the construction of confidence intervals or hypothesis tests (see, [97]).

Another approach to estimating volatility in the presence of microstructure noise is a realized kernel method proposed by Barndorff-Nielsen et al. [16]. This method is based on a linear combination of autocovariances. For simplicity of calculation, we consider a fixed interval $[0, 1]$ and let $k(x)$ be the non-stochastic weight function defined on $[0, 1]$ and $1/n$ the time gap. The number of observations is n. We define the realized autocovariance of order h with the integer bandwidth $H > 0$, $-H \leq h \leq H$ by

$$\gamma_h(Z) = \sum_{i=1}^{n} (Z_{\frac{i}{n}} - Z_{\frac{i-1}{n}})(Z_{\frac{i-h}{n}} - Z_{\frac{i-h-1}{n}}). \tag{5.4}$$

Note that $\gamma_0(Z)$ is a usual realized variance. The realized kernel estimator is defined as

$$K(Z) = \gamma_0(Z) + \sum_{h=1}^{H} k\Big(\frac{h-1}{H}\Big)(\gamma_h(Z) + \gamma_{-h}(Z)) \tag{5.5}$$

The authors show that the estimator $K(Z)$ is consistent and if $k \in C^2([0, 1])$ with $k(0) = 1$ and $k(1) = 0$, one has a stable limit theorem with the speed of convergence $n^{-1/6}$.

Furthermore, if the additional condition $k'(0) = k'(1) = 0$ holds, we achieve the better rate $n^{-1/4}$. For $H = cn^{1/2}$ we also have a stable central limit of the form

$$n^{1/4}(K(Z) - \int_0^1 \sigma_s^2 ds) \overset{d_{st}}{\to} MN(0, 4ck_1 IQ + \frac{8k_2\omega^2}{c} IV + \frac{4k_3\omega^4}{c^3}),$$

where $k_1 = \int_0^1 k(x)^2 dx$, $k_2 = \int_0^1 k'(x)^2 ds$, $k_3 = \int_0^1 k''(x)^2 dx$ and $IQ = \int_0^1 \sigma_s^4 ds$. Moreover, when $k(x) = 1 - x$, its asymptotic distribution and that of the two-scale estimator coincide. However, we can see from (5.4) that each of $\gamma_h(Z)$, $h \neq 0$, the data outside [0,1] are used. This may lead to some difficulties in practice (see, [97]).

Besides the above two approaches, *pre-averaging* is another intuitive approach introduced by Podolskij and Vetter [86]. In this thesis, we land our strategy to this last approach because it allows estimation of other powers of volatility, which we require in our study. We also refer to [62] for more details of the pre-averaging method. The concept of pre-averaging is explained in the next subsection.

5.1.1 Pre-averaging

We now introduce the idea of pre-averaging based on the works of [62, 85, 86]. The idea is that averaging on a number of $Y_{i/n}$'s near the time point i/n, one can get an estimate, say $\bar{Y}_{i/n}$, which tends to be close to the latent process $X_{i/n}$ because the noise is largely averaged away. To implement pre-averaging, we need some extra notations. We choose a sequence k_n of integers and a scalar $\theta > 0$, such that

$$k_n = \theta\sqrt{n} + o\big(n^{-1/4}\big). \tag{5.6}$$

We also need a weight function $w : \mathbb{R} \mapsto \mathbb{R}$ to do averaging such that

(i) w is continuous on $(0, 1)$ and vanishes outside of it,

(ii) w is piecewise continuously differentiable with a piecewise Lipschitz derivative w',

(iii) $\int_0^1 (w(t))^2 \mathrm{d}t > 0$.

The following numbers and functions are associated with w:

$$\phi_1(s) = \int_0^{1-s} w'(u)w'(u+s)\mathrm{d}u, \quad \phi_2(s) = \int_0^{1-s} w(u)w(u+s)\mathrm{d}u, \quad \text{for } s \in [0,1],$$

$$\phi_3(s) = \int_0^{2-s} w'(u)w'(u+s-1)\mathrm{d}u, \ \phi_4(s) = \int_0^{2-s} w(u)w(u+s-1)\mathrm{d}u, \ \text{for } s \in [0,2],$$

$$\psi_1 = \phi_1(0), \quad \psi_2 = \phi_2(0), \quad \Phi_{ij} = \int_0^1 \phi_i(s)\phi_j(s)\mathrm{d}s, \quad \text{for } i,j = 1,2,$$

$$\psi_1^n = k_n \sum_{j=0}^{k_n} (w_{j+1}^n - w_j^n)^2, \quad \psi_2^n = \frac{1}{k_n}\sum_{j=1}^{k_n}(w_j^n)^2, \tag{5.7}$$

where $w_j^n = w(j/k_n)$.

The pre-averaged return is given as

$$\Delta\bar{Y}_i^n = \sum_{j=1}^{k_n} w_j^n \Delta_{i+j}^n Y = -\sum_{j=0}^{k_n}(w_{j+1}^n - w_j^n)Y_{\frac{i+j}{n}}, \quad \text{for } i = 1,\dots,n-k_n+2. \tag{5.8}$$

An example of a weight function that satisfies the above conditions is $w(x) = (x \wedge (1-x))$. With this setting of the weight function w and supposing that k_n is even, we have

$$\Delta\bar{Y}_i^n = \frac{1}{k_n}\sum_{j=k_n/2}^{k_n-1} Y_{\frac{i+j}{n}} - \frac{1}{k_n}\sum_{j=0}^{k_n/2-1} Y_{\frac{i+j}{n}}. \tag{5.9}$$

Therefore, the use of the term *pre-averaging* is clear by the above expression (5.9).

Note that $\Delta\bar{Y}_i^n$ can be represented as

$$\Delta\bar{Y}_i^n = \int_{\frac{i}{n}}^{\frac{i+k_n}{n}} w_n\Big(s - \frac{i}{n}\Big)dY_s \text{ where } w_n(s) = \sum_i^{k_n} w_j^n \mathbb{1}_{(\frac{j-1}{n},\frac{j}{n}]}(s).$$

We also note that

$$\psi_1^n = \psi_1 + O_p\big(n^{-1/2}\big) \quad \text{and} \quad \psi_2^n = \psi_2 + O_p\big(n^{-1/2}\big), \tag{5.10}$$

which means only ψ_1 and ψ_2 will appear in the asymptotic theory. Still, it is recommendable to use ψ_1^n and ψ_2^n for simulations and empirical work, as it entails better finite sample properties.

Moreover, we have

$$\Delta\bar{X}_i^n = \sum_{j=1}^{k_n} w_j^n \Delta_{i+j}^n X = O_p\Big(\sqrt{\frac{k_n}{n}}\Big) \tag{5.11}$$

and

$$\Delta\bar{\epsilon}_i^n = \sum_{j=1}^{k_n} w_j^n \Delta_{i+j}^n \epsilon = O_p\Big(\sqrt{\frac{1}{k_n}}\Big) \tag{5.12}$$

(see, [97, (3.5)]). We see that the effect of the noise relies on how we choose k_n. Specifically, the noise part has less power on $\Delta\bar{Y}_i^n$ when a bigger k_n is selected. To balance these orders, in this thesis we pick k_n as in (5.6). Therefore, the stochastic orders of the above (5.11) and (5.12) are the same, i.e. $O_p(n^{-1/4})$.

5.1.2 Pre-averaged bipower variation

The appearance of microstructure noise in the model (5.1) leads to additional complications for inference procedures from high-frequency data in general. Hence we limit our attention only to the class of realized bipower variations. That is, in this section we are going to assume that

$$f_i(x) = |x|^{q_i} \text{ and } g_i(x) = |x|^{r_i},$$

where $q_i, r_i \geq 0$. To reflect this change and separate the following results from the previous ones, we write the pre-averaged bipower variation as $V^*(q,r)^n$, where $q = (q_1, \dots, q_m)'$ and $r = (r_1, \dots, r_m)'$ are m-dimensional vectors, whose coordinates index the powers.

The kth coordinate of $V^*(q,r)^n$ is defined as:

$$V^*(q_k, r_k)^n = \frac{1}{n - 2k_n + 2} \sum_{i=1}^{n-2k_n+2} |n^{1/4}\Delta\bar{Y}_i^n|^{q_k} |n^{1/4}\Delta\bar{Y}_{i+k_n}^n|^{r_k}. \tag{5.13}$$

The intuition behind this construction is that pre-averaging induces some autocorrelation (of order k_n) in the pre-averaged price series, which is broken by multiplying pre-averaged returns that are k_n terms apart. In essence, this leads to a lower, effective sample of size $n - 2k_n + 2$. Next, we state results from [85] for a law of large numbers and a central limit theorem of the statistic $V^*(q,r)^n$.

Proposition 5.1 ([85], Theorem 1)
Suppose that $Y_t = X_t + \epsilon_t$ is a noisy diffusion model, where X_t is given by (4.1). *Let q_k and*

r_k be non-negative real numbers where $k \in \{1, ..., m\}$. Assume that $\mathbb{E}[|\epsilon_t|^{2(q_k+r_k)+s}] < \infty$ for some $s > 0$. It holds that

$$V^*(q_k, r_k)^n \xrightarrow{\mathbb{P}} V^*(q_k, r_k) = \mu_{q_k}\mu_{r_k} \int_0^1 \left(\theta\psi_2\sigma_s^2 + \frac{1}{\theta}\psi_1\omega^2\right)^{\frac{q_k+r_k}{2}} ds. \tag{5.14}$$

Moreover, Podolskij and Vetter [85, Theorem 3] also show a central limit theorem for the bipower variation under some conditions:

$$n^{1/4}\Big(V^*(q,r)^n - V^*(q,r)\Big) \xrightarrow{d_{st}} MN(0, \Sigma^*), \tag{5.15}$$

where $q = (q_1, \ldots, q_m)'$ and $r = (r_1, \ldots, r_m)'$ are vectors of even non-negative real numbers, Σ^* is the $m \times m$ conditional covariance matrix of $V^*(q,r)^n$. More precisely, the matrix Σ^* has elements

$$\Sigma^*_{ij} = 2\theta \int_0^1 \int_0^2 h_{ij}\big(\sigma_u, t, f(s)\big) \mathrm{d}s\mathrm{d}u, \tag{5.16}$$

where

$$h_{ij}(x, y, z) = cov(|H_1|^{q_i}|H_2|^{r_i}, |H_3|^{q_j}|H_4|^{r_j}),$$

$x \in \mathbb{R}$, $y = (y_1, y_2)$ is two-dimensional vector, $z = (z_1, z_2, z_3, z_4)$ is four-dimensional vector, $H_1, ..., H_4$ follow a normal distribution with

(i) $\mathbb{E}[H_l] = 0$ and $\mathbb{E}[|H_l|^2] = y_1\omega^2 + y_2x^2$,

(ii) $H_1 \perp H_2$, $H_1 \perp H_4$ and $H_3 \perp H_4$,

(iii) $cov(H_1, H_3) = cov(H_2, H_4) = z_1\omega^2 + z_2x^2$ and $cov(H_2, H_3) = z_3\omega^2 + z_4x^2$,

$t = (\frac{1}{\theta}\psi_1, \theta\psi_2)$ and $f = (f_1, f_2, f_3, f_4)$ is defined for any s in $[0, 2]$ by

$$f_1(s) = \frac{1}{\theta}\phi_1(s), \qquad f_2(s) = \theta\phi_2(s),$$
$$f_3(s) = \frac{1}{\theta}\phi_3(s), \qquad f_4(s) = \theta\phi_4(s),$$

see [85] for more details. In a simpler pre-averaged realized variance case, $(q, r) = (2, 0)$, we have that

$$\Sigma^*(2, 0) = \int_0^1 4\Big(\theta^3\Phi_{22}\sigma_s^4 + 2\theta\Phi_{12}\sigma_s^2\omega^2 + \frac{1}{\theta}\Phi_{11}\omega^4\Big)\mathrm{d}s. \tag{5.17}$$

Thus, pre-averaging slows down the rate of convergence, but $n^{-1/4}$ is nonetheless the fastest rate in noisy diffusion models [52, 53].

We will see later in Section 5.3 that there exists a consistent estimator of the above Σ^* provided in [85]. Unfortunately, this approach does not ensure that the whole covariance matrix estimate is positive definite in finite samples. To fix the defect, we construct our estimator by using the idea of subsampling scheme described in the next section.

5.2 Main results

For building our estimator in the presence of noise, we first imitate the procedure from Subsection 4.1.1 by splitting the full sample of noisy high-frequency data into subsamples using a blocking approach.

For simplicity, we try to keep the same notations as in the previous chapter but use them in the context of noise. We redefine:

$$B_i(p) = \Big\{ j : (i-1)pk_n \leq j \leq ipk_n \Big\}, \tag{5.18}$$

where $p \geq 3$ is an integer and $i \geq 1$.

$B_i(p)$ is composed of adjacent observation time points at the ith block of noisy high-frequency data. We can see that the only change compared to the noiseless setting is that $B_i(p)$ uses a larger block size. This implies that we can do a sufficient amount of averaging within each block in order to diminish the noise, while still preserving enough of an effective sample size to estimate the correlation structure of $V^*(q,r)^n$.

We set

$$\hat{\Sigma}_n^* = \frac{1}{L}\sum_{l=1}^{L}\Bigg(\frac{n^{1/4}}{\sqrt{L}}\Big(V_l^*(q,r)^n - V^*(q,r)^n\Big)\Bigg)\Bigg(\frac{n^{1/4}}{\sqrt{L}}\Big(V_l^*(q,r)^n - V^*(q,r)^n\Big)\Bigg)', \tag{5.19}$$

where, assuming Lpk_n divides n,

$$\begin{aligned} V_l^*(q_k,r_k)^n &= \frac{Lpk_n}{n}\sum_{i=1}^{n/Lpk_n} v_{(i-1)L+l}(q_k,r_k)^n, \\ v_i(q_k,r_k)^n &= \frac{1}{pk_n - 2k_n + 2}\sum_{j,j+k_n-1\in B_i(p)} |n^{1/4}\Delta\bar{Y}_j^n|^{q_k}|n^{1/4}\Delta\bar{Y}_{j+k_n}^n|^{r_k}. \end{aligned} \tag{5.20}$$

Note that the summands $v_i(q_k,r_k)^n$ in the subsample estimates $V_l^*(q_k,r_k)^n$ exploit data solely from $B_i(p)$. Therefore, pre-averaging has to be performed locally within the block, so that there is no overlap in the pre-averaged returns across the various blocks.

Theorem 5.2
Assume that $Y_t = X_t + \epsilon_t$ is a noisy diffusion model, where X_t is given by Eq. (4.1)*, that fulfills Assumption* ($\mathbf{H}$) *and* ($\mathbf{M}$)*. Also, Assumption* ($\mathbf{N}$) *with $s > 3 \vee \max\{2(q_k + r_k) : 1 \leq k \leq m\}$ is true. Let $q = (q_1, \dots, q_m)'$ and $r = (r_1, \dots, r_m)'$ be vectors of even non-negative integers. Then, as $n \to \infty$, $p \to \infty$, $L \to \infty$ and $\sqrt{n}/Lp^2 \to \infty$, it holds that*

$$\hat{\Sigma}_n^* - \Sigma^* = \underbrace{O_p\left(\frac{1}{\sqrt{L}}\right)}_{\text{CLT}} + \underbrace{O_p\left(\frac{Lp^2}{\sqrt{n}}\right)}_{\text{blocking}} + \underbrace{O_p\left(\frac{1}{p}\right)}_{\text{HAC}} \tag{5.21}$$

Proof. See Section 5.4. □

As in the no-noise setting, the minimal assumptions we need to prove consistency are $n \to \infty$, $p \to \infty$, $L \to \infty$ and $\sqrt{n}/Lp^2 \to \infty$. The last condition $\sqrt{n}/Lp^2 \to \infty$ ensures that a bias term of the statistic $V_l^*(q,r)^n - V^*(q,r)^n$ is negligible with respect to its martingale part.

Now, we achieve the best rate

$$\hat{\Sigma}_n^* - \Sigma^* = O_p\big(n^{-1/10}\big), \tag{5.22}$$

by choosing

$$L = O(n^{1/5}) \quad \text{and} \quad p = O(n^{1/10}). \tag{5.23}$$

Thus, the existence of microstructure noise also adversely affects the speed of convergence of $\hat{\Sigma}_n^*$.

In practical work, we need consistency results not only for even non-negative integers q_i and r_i as in Theorem 5.2, but also for any choice of the powers greater than zero. In order to complete this task, we assume some stronger conditions on ϵ as follows:

Assumption (A):

(i) ϵ is distributed symmetrically around zero.

(ii) For any $-1 < a < 0$, we have $\mathbb{E}[|\epsilon|^a] \leq \infty$.

Assumption (A') (Cramer's condition):

$$\lim \sup_{|t| \to \infty} \chi(t) < 1,$$

where χ is the characterisitic function of ϵ.

The Assumption **(A)** is required for powers in $(0, 1]$. Specifically, the associated central limit theorem for the bipower variation depends upon the fact that the normal distribution satisfies both **(A)**(i) and **(A)**(ii). The Assumption **(A')** is needed for non-even powers. It will be used to get rid of the bias in the pre-averaged statistic $|\Delta \bar{Y}_i^n|^q$. Usually, the moments of $|n^{1/4} \Delta \bar{\epsilon}_i^n|$ is approximated by the associated moments of a normal distribution. An expansion of Edgeworth-type, for which **(A')** is a standard assumption, is used in this case to show that this error of the approximation becomes sufficiently small. To summarize, these two conditions are provided in [85] to prove the central limit theorem in (5.15) for any powers. Note that we can again dispense with Assumption (**M**) for consistency.

Theorem 5.3
Assume that $Y_t = X_t + \epsilon_t$ is a noisy diffusion model, where X_t is given by Eq. (4.1)*, σ is continuous and fulfills Assumption* $(\mathbf{V})$ *with $\sigma > 0$, while the noise fulfills Assumption* $(\mathbf{N})$ *with $s > 3 \vee \max\{2(q_k + r_k) : 1 \leq k \leq m\}$ and also Assumption* $(\mathbf{A})$ *and* $(\mathbf{A'})$*. Then, as $n \to \infty$, $p \to \infty$, $L/p \to \infty$ and $\sqrt{n}/Lp^2 \to \infty$, it holds for any $q, r \geq 0$ that*

$$\hat{\Sigma}_n^* \xrightarrow{\mathbb{P}} \Sigma^*. \tag{5.24}$$

Proof. See Section 5.4. □

Remark 5.4
It is straightforward to extend the results in this section to the multipower-type pre-averaging estimators.

5.2.1 Extension to dependent and heteroscedastic noise

The i.i.d. framework on the microstructure noise ϵ is a convenient outset, but it is hard to defend at the tick frequency, both in theory and practice [47]. An intriguing ability of the subsampling estimator $\hat{\Sigma}_n^*$ is that it tends to be robust against the intricate features of the noise process, as long as an associated central limit theorem holds. Kalnina [67] studied subsampling in the presence of both autocorrelated and heteroscedastic noise for the two-scale realized variance. In this subsection, we show how our theoretical results adapt to such models, allowing for more general structure in the noise process.

1. Dependent noise

Autocorrelation in tick-by-tick returns can extend beyond the first lag, depending a bit on how you gather the data [55, 7]. This cannot be captured by independent noise, so we start by weakening this assumption to so-called m-dependent noise. Thus, we now assume that the noise process $(\epsilon_t)_{t \geq 0}$ is stationary and that the random variables $\epsilon_{i/n}$ and

$\epsilon_{j/n}$ are independent, only if $|i-j| > m$. Hautsch et al. [58] prove a central limit theorem for pre-averaging in this setup (based on the estimator $V^*(2,0)^n$). As indicated by their results, the law of large numbers and the central limit theorem for the pre-averaged bipower variation estimator $V^*(q,r)^n$ do not change, except that the noise variance ω^2 has to be replaced by the expression $\rho^2 = \omega^2 + 2\sum_{j=1}^{m} \operatorname{cov}(\epsilon_{i/n}, \epsilon_{(i+j)/n})$, i.e.

$$V^*(q_k, r_k)^n \xrightarrow{p} V^*(q_k, r_k) = \mu_{q_k}\mu_{r_k} \int_0^1 \left(\theta\psi_2\sigma_s^2 + \frac{1}{\theta}\psi_1\rho^2\right)^{\frac{q_k+r_k}{2}} \mathrm{d}s. \tag{5.25}$$

Here, the form of Σ^* changes.[1] Our proposed estimator $\hat{\Sigma}_n^*$ is still consistent though, as the change in the bias caused through replacing ω^2 by ρ^2 is corrected by construction in (5.19). This is because $\hat{\Sigma}_n^*$ imitates the underlying covariation, irrespective of the true microstructure model. This implies that the asymptotic results of Theorem 5.2 and 5.3 are also true for the m-dependent noise model.

2. Heteroscedastic noise

The market microstructure reveals itself, for example, via the bid-ask spread. It has been noticed in many empirical asset price series that such margins are not constant through time, but tend to vary systematically within the day in the form of a U-shape. Thus, in the equity market spreads are typically larger in the morning and afternoon than during the middle of the day. To accommodate this, another setup that has been studied is heteroscedastic noise [70]. Here, we follow the exposition in [62] by assuming that

$$\mathbb{E}\big[\epsilon_t \mid X\big] = 0, \qquad \mathbb{E}\big[\epsilon_t^2 \mid X\big] = \omega_t^2 \quad \text{is càdlàg and } (\mathcal{F}_t)\text{-adapted}, \tag{5.26}$$

while, conditional on X, ϵ_t and ϵ_s are independent for any $t \neq s$. By construction, this model exhibits a time-varying variance structure of the noise, which depends on the efficient price X.[2] Note that these conditions do not contradict unconditional dependence in ϵ. In this case, the consistency result translates to:

$$V^*(q_k, r_k)^n \xrightarrow{p} V^*(q_k, r_k) = \mu_{q_k}\mu_{r_k} \int_0^1 \left(\theta\psi_2\sigma_s^2 + \frac{1}{\theta}\psi_1\omega_s^2\right)^{\frac{q_k+r_k}{2}} \mathrm{d}s. \tag{5.27}$$

Again, the estimator $\hat{\Sigma}_n^*$ automatically adapts to the new environment, which, in particular, implies that the consistency result of Theorem 5.3 is still true. In order to maintain

[1] In particular, Hautsch and Podolskij [58] show that the asymptotic variance of the pre-averaged realized variance $V^*(2,0)^n$ is unchanged, apart from replacing ω with ρ everywhere.

[2] An example of this model is additive, uniform noise plus rounding: $Y_t = \gamma\big[(X_t + u_t)/\gamma\big]$, where $u = (u_t)_{t\geq 0}$ is an i.i.d. $\mathcal{U}([0,\gamma])$-distributed process that is independent of X, and $\gamma > 0$ is a fixed rounding level. In this setting, the conditional variance of the noise process is given by: $\omega_t^2 = \gamma^2\left(\left\{\frac{X_t}{\gamma}\right\} - \left\{\frac{X_t}{\gamma}\right\}^2\right)$, with $\{x\} = x - [x]$ denoting the fractional part of x.

unchanged error rates as in Theorem 5.2, however, we need to also impose identical assumptions on the process $(\omega_t)_{t\geq 0}$ as for the volatility $(\sigma_t)_{t\geq 0}$. This is because the role of both processes are identical in all asymptotic expansions.

5.3 Simulations and empirical work

This section contains the simulation results from Christensen et al. [41], which we conduct a small Monte Carlo study. It takes a closer look at the finite sample properties of covariance matrix estimation by subsampling. Throughout, we restrict attention to the noisy setting and estimation of Σ^*. We examine the ability of our proposed estimator $\hat{\Sigma}^*_n$ to assist in drawing feasible inference about the pre-averaged bipower variation.

The efficient log-price X is simulated as:

$$\begin{aligned} \mathrm{d}X_t &= \sigma_t \mathrm{d}W_t, \\ \mathrm{d}\sigma_t^2 &= \kappa(\sigma^2 - \sigma_t^2)\mathrm{d}t + \xi\sigma_t(\rho \mathrm{d}W_t + \sqrt{1-\rho^2}\mathrm{d}B_t), \end{aligned} \tag{5.28}$$

where W_t and B_t are independent standard Brownian motions, while κ, σ^2, ξ and ρ are parameters. The process adopted for σ_t^2 is a Heston [59] model, which is mean-reverting; features square-root volatility; and accommodates a leverage effect.[3]

To get a version of the model from which we can actually simulate data, we apply a standard Euler approximation to the continuous time formulation in (5.28). We then simulate 10,000 independent sample path realizations of the discretized system of bivariate equations.[4] We use two different sample sizes of n = 2,340 and 23,400. In our empirical investigation, we look at high-frequency equity data from NYSE. With a US stock exchange trading session running from 9:30am to 4:00pm—or 6.5 hours—these sample sizes translate into receiving a new price update every ten and one second(s). Our sample sizes are therefore representative of more frequently traded securities.

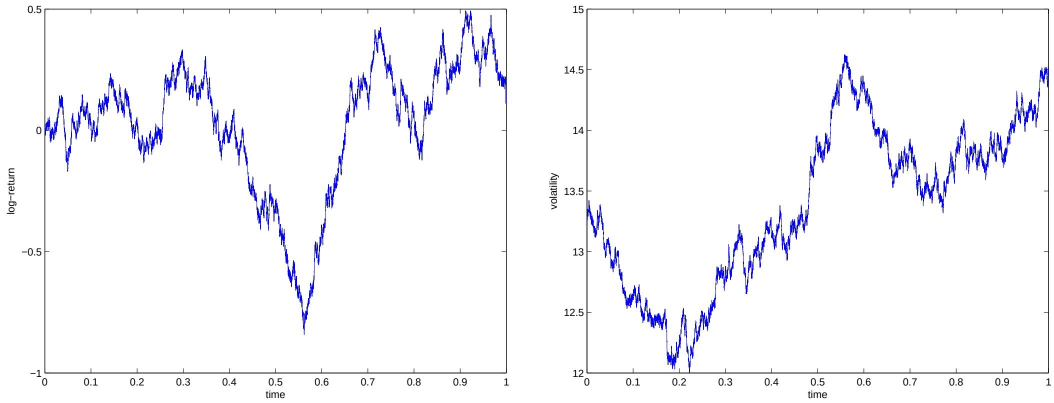
We assume that the parameter values in the volatility equation are $\kappa = 5$, $\sigma^2 = 0.04$, $\xi = 0.50$ and $\rho = -0.50$, which is broadly consistent with prior work [7, 67, 99]. This implies that σ_t is about 20% on an average, annualized basis, but the configuration of the model adopted here can generate a substantial degree of intraday variation in volatility via ξ. An example simulation is provided in Figure 5.1.

[3]The leverage effect describes a negative correlation between an asset's return and volatility [35, 42]. Thus, if a leverage effect is present, one would expect ρ to be negative.

[4]To avoid a systematic effect from an assumed initial condition of volatility, σ_0^2, we restart the variance process in each simulation by drawing at random from its stationary distribution, $\sigma_t^2 \sim \mathrm{Gamma}(2\kappa\sigma^2\xi^{-2}, 2\kappa\xi^{-2})$.

Figure 5.1: An illustration of a simulation from the Heston model.

Panel A: Cumulative log-return. Panel B: Annualized spot volatility.



An autocorrelated and heteroscedastic noise term is added to X. First, we create a set of serially dependent standard normal random variables via an MA(1) filter: $u_{i/n} = u'_{i/n} + \zeta u'_{(i-1)/n}$, where ζ is a parameter and $u'_{i/n} \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(0, \dfrac{1}{1+\zeta^2}\right)$, independent of X. Hence, $u_{i/n} \sim \mathrm{N}(0,1)$ and $\mathrm{cov}(u_{i/n}, u_{(i-1)/n}) = \dfrac{\zeta}{1+\zeta^2}$, so that ζ controls the degree of first-order serial correlation in the noise. In our simulations, we take $\zeta = -0.4$. Second, as in [6], we set $\epsilon_{i/n} = \gamma \dfrac{\sigma_{i/n}}{\sqrt{n}} u_{i/n}$, where γ is the noise-ratio parameter [79]. This formulation implies that, conditional on σ, $\omega_{i/n} = \gamma \dfrac{\sigma_{i/n}}{\sqrt{n}}$ and ensures microstructure noise variation is conditionally heteroscedastic and proportional to the spot volatility of the efficient price. We assume that $\gamma = 0.50$, which is a realistic choice for more liquid assets, see, e.g., [40].[5] Although this noise setting is not formally covered by our theoretical frame, we include it here as a robustness check. To alleviate the impact of noise, we pre-average using the bandwidth $k_n = [\theta\sqrt{n}]$, and we experiment with two choices of the tuning parameter $\theta = 1/3$ and 1. We set the weight function $w(x) = \min(x, 1-x)$, which has been shown to deliver nearly efficient estimates of the integrated variance, when further parametric assumptions are imposed (see, [85]).

[5] We also experimented with a much larger noise-ratio of $\gamma = 2$, as done by [6]. The results were almost identical, albeit slightly worse, than those we report in the main text and, hence, are omitted to conserve space.

5.3.1 A preliminary analysis

We begin by verifying the conjecture made in the introduction of this work, namely that the finite sample properties of existing estimators of Σ^* are poor, which renders a significant fraction of such estimates either nonpositive definite or at least ill-conditioned. To do this, we implement the estimator suggested by Podolskij and Vetter [85]. It is constructed by first defining

$$\tilde{Y}^n_{i,m} = \frac{1}{\sqrt{n}} |n^{1/4} \Delta \bar{Y}^n_m|^{q_i} |n^{1/4} \Delta \bar{Y}^n_{m+k_n}|^{r_i}, \tag{5.29}$$

and setting

$$\chi^n_{ml} = \frac{1}{2} \left[\tilde{Y}^n_{i,m} \left(\tilde{Y}^n_{j,m+l} - \tilde{Y}^n_{j,m+2k_n} \right) + \tilde{Y}^n_{j,m} \left(\tilde{Y}^n_{i,m+l} - \tilde{Y}^n_{i,m+2k_n} \right) \right], \tag{5.30}$$

for any $0 \leq m \leq n - 4k_n + 1$ and $0 \leq l \leq 2k_n$. Then,

$$\tilde{\Sigma}^*_{ij,n} = \frac{2}{\sqrt{n}} \sum_{m=0}^{n-4k_n+1} \sum_{l=0}^{2k_n-1} \chi^n_{ml} \xrightarrow{p} \Sigma^*_{ij}, \tag{5.31}$$

and it follows that $\tilde{\Sigma}^*_n = \left(\tilde{\Sigma}^*_{ij,n} \right) \xrightarrow{p} \Sigma^*$.

Table 5.1 shows several diagnostics that highlight the properties of $\tilde{\Sigma}^*_n$ based on $q = (2,1)'$ and $r = (0,1)'$, i.e. pre-averaged realized variance and $(1,1)$-bipower variation. In the table, we report the outcome from the general noise model, but we also include a comparable i.i.d. noise environment, which we base on setting $\zeta = 0$ in the above and replacing $\sigma_{i/n}$ by $\sqrt{\int_0^1 \sigma^2_s \mathrm{d}s}$ in the noise variance, while noting that Podolskij and Vetter [85] operate under the latter conditions. In addition, the results are also obtained for a scaled Brownian motion (BM), in which volatility has been fixed at its steady-state value of σ^2. The column SV is for the Heston stochastic volatility model, while SPY represents some real high-frequency data that are further commented on in Section 5.3.4.

In Panel A, we report the fraction of the computed $\tilde{\Sigma}^*_n$, which fail to be positive definite, i.e. which have a minimum eigenvalue $\min(\lambda_i) \leq 0$. It suggests that for a small sample of $n = 2,340$ and depending on θ, between 18% – 35% of $\tilde{\Sigma}^*_n$ are nonpositive definite. As expected, these numbers decrease as the sample size increases, but even with a fairly large sample of $n = 23,400$ the failure rate is far from negligible. Moreover, it increases if a longer pre-averaging horizon is employed. As such, this issue therefore has substantial bite in practice, because larger values of θ are typically preferred, when the noise is suspected to violate the i.i.d. assumption [40, 58]. Note that going from constant to stochastic volatility changes the numbers only slightly, so allowing volatility to be time-varying has no discernable impact on the failure rate.

Table 5.1: Proportion of ill-conditioned covariance matrix estimates, $\tilde{\Sigma}_n^*$.

		general noise				i.i.d. noise				
		BM		SV		BM		SV		SPY
$n =$		2,340	23,400	2,340	23,400	2,340	23,400	2,340	23,400	n_{actual}
					2-dimensional setting					
Panel A: Nonpositive definite										
$\theta =$	0.33	0.177	0.094	0.180	0.101	0.178	0.099	0.181	0.111	0.111
	1.00	0.340	0.242	0.343	0.251	0.348	0.241	0.347	0.249	0.233
Panel B: Negative variance										
$\theta =$	0.33	0.141	0.066	0.149	0.072	0.140	0.070	0.146	0.077	0.064
	1.00	0.286	0.199	0.290	0.210	0.292	0.201	0.289	0.206	0.184
Panel C: Condition number ≥ 20										
$\theta =$	0.33	0.073	0.053	0.071	0.058	0.069	0.057	0.071	0.062	0.045
	1.00	0.080	0.096	0.081	0.101	0.079	0.095	0.081	0.100	0.088
					4-dimensional setting					
Panel D: Nonpositive definite										
$\theta =$	0.33	0.454	0.303	0.457	0.305	0.453	0.312	0.458	0.314	0.312
	1.00	0.799	0.577	0.810	0.578	0.809	0.568	0.806	0.589	0.544

Note. We show the proportion of ill-conditioned covariance matrix estimates, when the estimator $\tilde{\Sigma}_n^*$ from [85] is used. $\tilde{\Sigma}_n^*$ is defined in (5.31). The columns report results from a general noise model, where ϵ is autocorrelated and heteroscedastic, and an i.i.d. noise process, plus for a Brownian motion (BM) and stochastic volatility (SV) model for σ. n is the sample size. The simulation design appears in Section 5.3. SPY is based on real high-frequency data, which are further analyzed in Section 5.3.4. n_{actual} denotes the actual sample size, which varies across days, cf. Table 5.2. In Panel A, we report the fraction of $\tilde{\Sigma}_n^*$ estimates that are nonpositive definite, i.e. with a minimum eigenvalue $\min(\lambda_i) \leq 0$. In Panel B, we report how often the linear combination $\omega = [1, -\mu_1^{-2}]'$ of the pre-averaged bipower variation $V^*(q,r)^n$ leads to a negative variance estimate $\omega' \tilde{\Sigma}_n^* \omega \leq 0$. $V^*(q,r)^n$ is defined in (5.13). B implies A, but not vice versa. In Panel C, we compute the fraction of the positive definite $\tilde{\Sigma}_n^*$ estimates that return a condition number $\text{cond}(\tilde{\Sigma}_n^*) \geq 20$. Throughout Panels A – C, the table is based on $q = (2,1)'$ and $r = (0,1)'$, while Panel D reports the updated numbers from Panel A, after changing the estimation problem to $q = (2,1,4,2)'$ and $r = (0,1,0,2)'$.

Turning next to Panel B, we investigate how often the linear combination $\omega' V^*(q,r)^n$ with $\omega = (1, -\mu_1^{-2})'$ results in a negative variance estimate $\omega' \tilde{\Sigma}_n^* \omega \leq 0$. The difference $V^*(2,0)^n - \mu_1^{-2} V^*(1,1)^n$ is often used in applied work, as it provides information about presence of jumps in the price process and permits a statistical test of this hypothesis. Even when the covariance matrix estimate is not positive definite, it could still result in a positive variance estimate $\omega' \tilde{\Sigma}_n^* \omega > 0$, thereby allowing the t-statistic to be computed. While the numbers in Panel B are lower compared to Panel A, they are still high.

In Panel C, we look at those $\tilde{\Sigma}_n^*$ estimates that are positive definite. We compute the percentage of these, which return a condition number $\text{cond}(\tilde{\Sigma}_n^*) \geq 20$.[6] The condition number measures the numerical accuracy of a matrix, and a value above $20 = 10 \times \dim(\tilde{\Sigma}_n^*)$ is generally taken as a sign of an ill-conditioned and nearly singular matrix [54, 57]. As readily seen, we find that about 5% – 10% of the $\tilde{\Sigma}_n^*$ that are deemed ok by a definiteness criteria show signs of being badly scaled.

Lastly, in Panel D we attempt to estimate the joint asymptotic covariance matrix of a 4-dimensional parameter by using $q = (2,1,4,2)'$ and $r = (0,1,0,2)'$. We then report the percentage of the $\tilde{\Sigma}_n^*$ estimates, which are not positive definite, i.e. the numbers can be compared to Panel A. Not surprisingly, increasing the complexity of the problem has a detrimental impact on the estimation errors, and up to 80% of the estimates are now nonpositive definite, making this a devastating issue for inference (e.g., [46, 45, 98, employ a 3-dimensional statistic to test for the parametric form of volatility in diffusion models (both with or without noise)]).

5.3.2 Implementation of the subsampler

We now turn to the subsampler, where we again base our investigation on $V^*(q,r)^n$ using the parameters $q = (2,1)'$ and $r = (0,1)'$. As $\hat{\Sigma}_n^*$ depends on two tuning parameters, p and L, we compute it by varying these across a broad range of values in order to gauge the sensitivity of our estimator to specific choices. We set $p = 3, 5$ and 10, so that the block length of noisy returns before pre-averaging goes from three to ten times the pre-averaging horizon k_n. Moreover, we slice the sample into $L = 5, 10$ and 15 subsamples,

[6] The condition number of an invertible matrix A is defined as $\text{cond}(\text{A}) = ||\text{A}|| \cdot ||\text{A}^{-1}||$, where $|| \cdot ||$ is the L_2 matrix norm. cond(A) can be shown to be the ratio between the largest and smallest singular value of A. It can be interpreted as saying how much small changes in the input matrix get amplified under matrix inversion.

yielding a total of nine combinations of p and L.[7]

Our initial set of simulations suggested that, for small p and L, the raw estimator defined by (5.19) is downward biased, thereby leading to a systematic underestimation of Σ^*. We therefore start by briefly outlining a few corrections that are important in finite samples.

First, in order to center the $V_l^*(q,r)^n$, we should in theory use the unobserved $V^*(q,r)$, which we are forced to replace by a feasible, consistent estimator, i.e. $V^*(q,r)^n$. While this has no impact asymptotically, because $V^*(q,r)^n$ converges much faster than $V_l^*(q,r)^n$, a closer inspection of $\hat{\Sigma}_n^*$ shows that the substitution does entail a standard small sample correction. This implies that $\hat{\Sigma}_n^*$ should be divided by $1 - 1/L$, i.e. the "right" normalization in (5.19) is $L - 1$ and not L.

Second, there is a HAC error associated with p, which—in contrast to the L correction—is more subtle to deal with. The problem originates from the estimation of the autocovariances of the pre-averaged returns, $|n^{1/4}\Delta\bar{Y}_i^n|^{q_i}|n^{1/4}\Delta\bar{Y}_{i+k_n}^n|^{r_i}$, as exemplified by (4.26). As such, it depends on the covariance structure of this series, which, in turn, is a function of several parameters and variables, including spot volatility, the weight function and the variance of the noise process (see, [85]). If we approximate this function, e.g., by assuming that volatility is constant, a detailed calculation (which is omitted here, but available upon request) shows that for the pre-averaged bipower variation estimator $V^*(q_i, r_i)^n$ defined by (5.13), we can roughly correct for the p error by dividing $\hat{\Sigma}_n^*$ with $1 - 1/p$.

It turns out, however, that this is too much, if either q_i or r_i is zero, as it happens for the pre-averaged realized variance, $V^*(2,0)^n$. This is because the summands in (5.13) are, asymptotically, $2k_n$-dependent for non-zero values of both q_i and r_i, while they are only k_n-dependent, if either is zero. Thus, the bias induced by p in the latter is, loosely speaking, half that of the former. This indicates that we ought to divide the elements of $\hat{\Sigma}_n^*$ involving the variance of $V^*(2,0)^n$ and its covariance with $V^*(1,1)^n$ only by $1 - 0.5/p$. This is less appealing, because it would break the positive semi-definiteness property of $\hat{\Sigma}_n^*$. We therefore proceed by using a constant scaling for the entire matrix, and—to strike a balance between the two alternatives—we propose to meet in the middle and rescale $\hat{\Sigma}_n^*$ by $(1 - 0.75/p)$. This choice produces excellent results for the values of p considered in this work, as corroborated by our numerical experiments below, while for $p \geq 10$ the correction is of limited importance and can be ignored.

[7] This implies that for the sample size $n = 2,340$, there are some combinations of p and L, for which there is not enough data to compute the subsampler. Therefore, we restrict attention to $n = 23,400$ in the following. The results for $n = 2,340$, when attainable, are not materially worse, reflecting the slow rates of convergence, and all are available by request.

Third, our theoretical results hinge on n being a multiple of Lpk_n. In practice, where n varies randomly over time and can be very odd, this is an unrealistic assumption, which is almost never satisfied. Instead, for a given choice of k_n, p and L, the maximum number of blocks of length pk_n that can be assigned to each of the L subsample estimates $V_l^*(q_i, r_i)^n$ is:

$$n_{\text{block}} = \left\lfloor \frac{\lfloor n/pk_n \rfloor}{L} \right\rfloor. \tag{5.32}$$

The effective amount of data used to construct the subsampler is therefore often less than the total sample size, i.e. $n_{\text{block}} Lpk_n \leq n$.

We compute $\hat{\Sigma}_n^*$ from the data that fall within the window $[0, n_{\text{block}} Lpk_n/n]$ and subsequently inflate this estimate to cover the whole unit interval. While this entails some loss of information about the underlying variation of the process towards the end of the sample, in our experience this has a very limited influence on the results, unless the data, from which $\hat{\Sigma}_n^*$ is computed, is not representative of the overall level of volatility. In practice, one can minimize this effect by choosing the parameters, such that $n_{\text{block}} Lpk_n$ is close to n.

5.3.3 Results

In Figure 5.2, we plot some kernel smoothed density estimates of the standardized pre-averaged bipower variation, i.e. $n^{1/4}\big(V^*(q_k, r_k)^n - V^*(q_k, r_k)\big)/\sqrt{\hat{\Sigma}_{kk,n}^*}$, where $\hat{\Sigma}_{kk,n}^*$ is the kth diagonal element of our subsampling covariance matrix estimate $\hat{\Sigma}_n^*$. Here, we use $\theta = 1$. The results in Panels A – B are for $V^*(1,1)^n$, while Panels C – D are for $V^*(2,0)^n$. In addition, the left-hand portion of the figure is for $L = 15$ and p changing, while the right-hand part is based on $p = 10$ and for different L. The infeasible result for $V^*(2,0)^n$ replaces the subsampler with the true variance, which is known here (cf.(5.17)).

As the figure shows, the studentized pre-averaging estimators tend to track the asymptotic normal approximation closely across combinations of p and L. The sole exception, appearing in Panel A, is for $V^*(1,1)^n$, when $\hat{\Sigma}_n^*$ is implemented using $p = 3$. Note that if $p = 3$, the effective sample size within a block of noisy high-frequency data is $k_n + 2$ after pre-averaging, whereas the summands of $V^*(1,1)^n$ are $2k_n$-dependent. With such a small value of p, the block length is therefore inadequate to permit estimation of all the required autocovariances. As $|n^{1/4}\Delta\bar{Y}_i^n|^{q_k}|n^{1/4}\Delta\bar{Y}_{i+k_n}^n|^{r_k}$ is strongly positively autocorrelated, this leads to a severe underestimation of the true variation of $V^*(1,1)^n$ and, hence, a pronounced overdispersion of the estimated density. As a practical guide, one should therefore avoid computing $\hat{\Sigma}_n^*$ with $p = 3$, if both q_k and r_k are different from zero, for any $i = 1, \ldots, m$. In comparison, the corresponding graph for $V^*(2,0)^n$ in Panel C is much better scaled, reflecting the lesser dependence inherent in this estimator. Apart from that, the fit tends to improve for larger values of p and L, as expected.[8] Lastly, comparing with the infeasible result in Panels C – D, we note that the estimated densities appear slightly negatively skewed, owing to a modest, positive correlation between $V^*(q_k, r_k)^n$ and $\hat{\Sigma}^*_{kk,n}$.

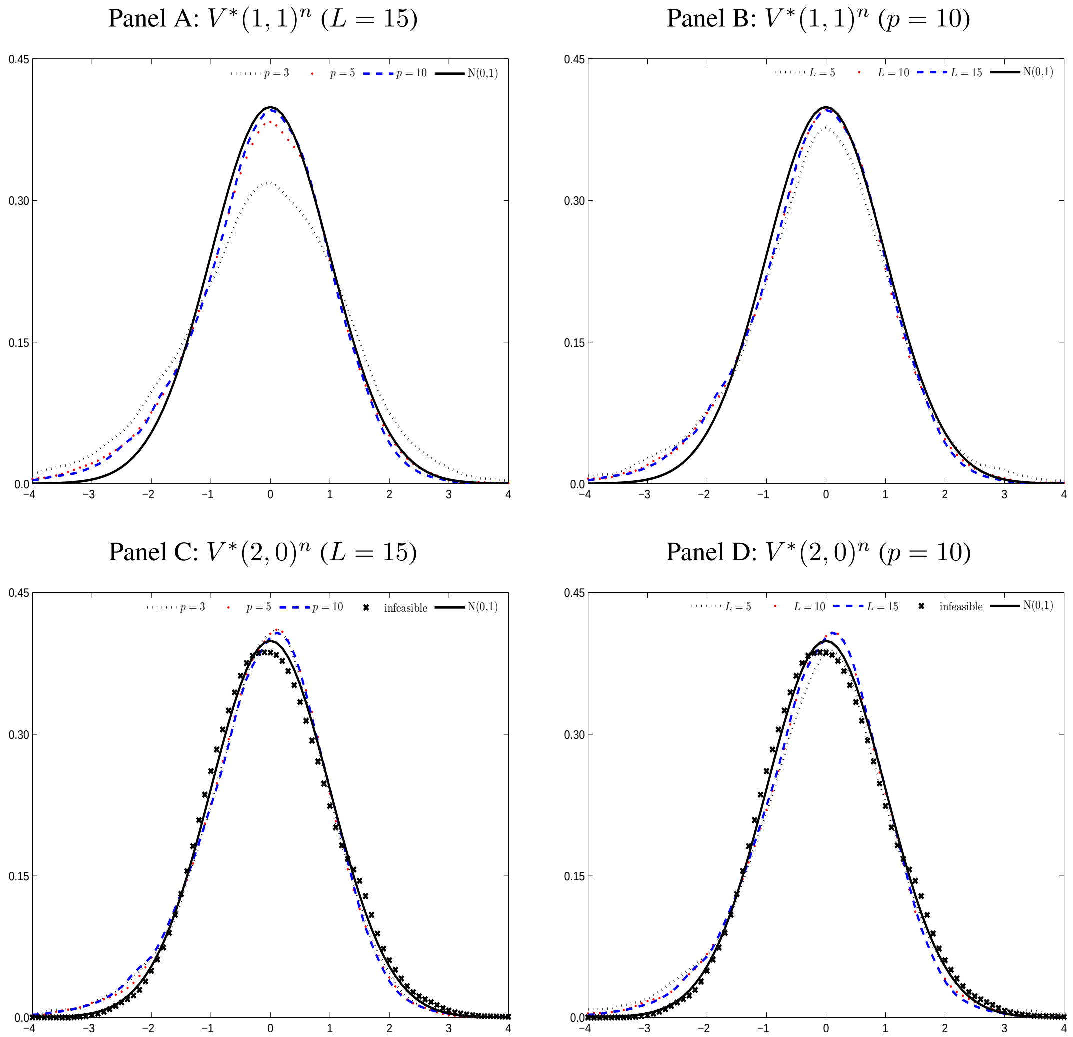
We turn next to Figure 5.3, where we explore how sensitive our findings are to the choice of θ. In this figure, and throughout the remainder of this section, we fix the parameters of $\hat{\Sigma}^*_n$ to $p = 10$ and $L = 15$. As apparent from both panels, the relatively large change in θ has only a minuscule effect on the shape of the estimated density for $V^*(1,1)^n$ and $V^*(2,0)^n$.

In Figure 5.4, using $\theta = 1$, we look at an application that requires one to use information about the full covariance matrix estimate by reporting some results for the linear combination $\omega' V^*(q,r)^n$ with $\omega = (1, -\mu_1^{-2})'$, i.e. $V^*(2,0)^n - \mu_1^{-2} V^*(1,1)^n$. In Table 5.1, we noted it was problematic to standardize this difference with the covariance matrix estimator put forth by Podolskij and Vetter [85], which was often found to be nonpositive definite. The subsampler does not suffer from this issue. In Panel A of Figure 5.4, we therefore plot the time series of the studentized statistic across the simulations runs, using the delta method to conclude that $n^{1/4}\omega'\big(V^*(q,r)^n - V^*(q,r)\big)/\sqrt{\omega'\hat{\Sigma}^*_n\omega} \xrightarrow{d} \mathrm{N}(0,1)$. Panel B inspects the kernel smoothed density estimate of the t-statistic. As evident, the asymptotic distribution theory is a decent description of the actual finite sample variation, although the fit is not perfect. As explained above, $V^*(2,0)^n - \mu_1^{-2} V^*(1,1)^n$ provides information about the presence of jumps in asset prices, and significant positive values would lend support to this hypothesis. Here, the shape of the estimated density implies that the one-sided coverage probabilities of the t-statistic are slightly too large in the right tail. This would render such a hypothesis test mildly conservative, which is preferable in practice (e.g., using the 95% quantile from the standard normal distribution gives a coverage rate of 96.7% in the above figure).

At last, we compare our subsampler $\hat{\Sigma}^*_n$ to an alternative, nonparametric estimator of Σ^*, namely the observed asymptotic variance (AVAR) of [76]. As in our setting, the observed AVAR is based on squared increments (or outer products) of the original statistic(s) computed on smaller streches of high-frequency data, but there are several differences between the construction of the subsampler and observed AVAR. Moreover, there is little guidance on how to select tuning parameters for the latter. We therefore proceed as fol-

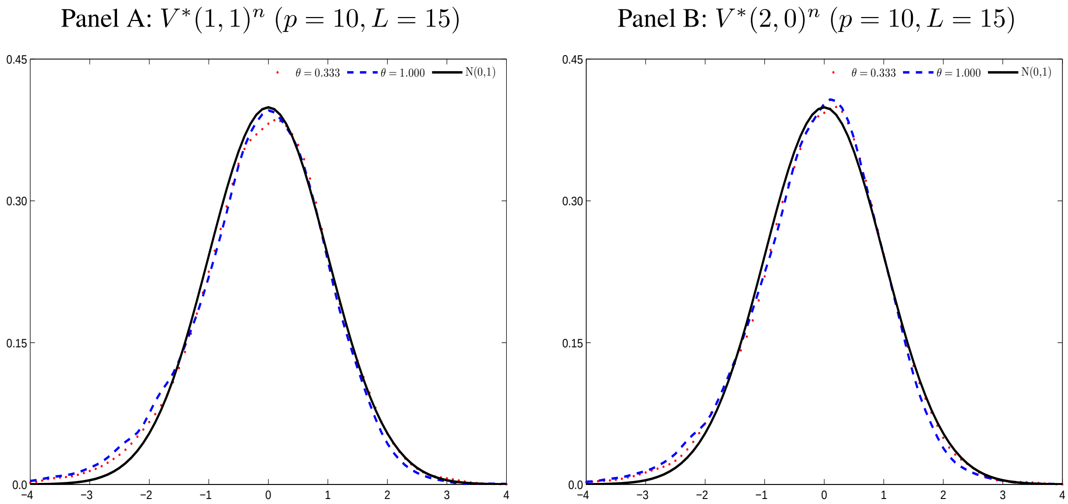
[8] Of course, with a fixed sample size n and pre-averaging window k_n, the parameters p, L, and n_{block} are not free. Thus, everything else is not "fixed", because n_{block} is decreasing for larger values of either p or L (while holding the other fixed).

Figure 5.2: Kernel density estimate of the standardized $V^*(q_k, r_k)^n$: changing p and L.



Note. We show the kernel smoothed density estimates of the standardized pre-averaged bipower variation estimator: $n^{1/4}\big(V^*(q_k, r_k)^n - V^*(q_k, r_k)\big)/\sqrt{\hat{\Sigma}^*_{kk,n}}$, where $\hat{\Sigma}^*_{kk,n}$ is the kth diagonal element of $\hat{\Sigma}^*_n$. Throughout this figure, $n = 23,400$, $\theta = 1$ and we set $k_n = [\theta\sqrt{n}]$ to implement pre-averaging. The simulation data is from a Heston stochastic volatility model, as described in the main text. In the left panel, the subsampler is based on $L = 15$ subsamples, while varying the block length at $p = 3, 5$ or $10 \times k_n$. The right panel is based on $p = 10$, while changing the number of subsamples at $L = 5, 10$ or 15. The $n_{\text{sim}} = 10,000$ simulated t-statistics are smoothed using a Gaussian kernel with optimal bandwidth selection $h = 1.06\hat{\sigma}n_{\text{sim}}^{-1/5}$, where $\hat{\sigma}$ is the sample standard deviation of the data. The infeasible result for $V^*(2,0)^n$ replaces the subsampler with the true variance (cf., (5.17)). The density function of a standard normal random variable (the solid black line) is superimposed as a visual reference.

Figure 5.3: Kernel density estimate of the standardized $V^*(q_k, r_k)^n$: changing θ.



Note. We show the kernel smoothed density estimates of the standardized pre-averaged bipower variation estimator: $n^{1/4}\big(V^*(q_k, r_k)^n - V^*(q_k, r_k)\big)/\sqrt{\hat{\Sigma}^*_{kk,n}}$, where $\hat{\Sigma}^*_{kk,n}$ is the kth diagonal element of $\hat{\Sigma}^*_n$. Throughout this figure, $n = 23,400$, $L = 15$, $p = 10$, and we set $k_n = [\theta\sqrt{n}]$ to implement pre-averaging using $\theta = 1/3$ and $\theta = 1$. The simulation data is from a Heston stochastic volatility model, as described in the main text. The left panel holds the results for $V^*(1,1)^n$, while the right panel is for $V^*(2,0)^n$. The $n_{\text{sim}} = 10,000$ simulated t-statistics are smoothed using a Gaussian kernel with optimal bandwidth selection $h = 1.06\hat{\sigma}n_{\text{sim}}^{-1/5}$, where $\hat{\sigma}$ is the sample standard deviation of the data. The density function of a standard normal random variable (the solid black line) is superimposed as a visual reference.

lows. The sampling grid consists (using their notation) of $B = L$ blocks at the outset. This helps to ensure comparability with $\hat{\Sigma}^*_n$. We then compute the observed AVAR using a two-scale approach, as a linear combination of the K-averaged apparent quadratic covariation with $K_1 = 1$ and $K_2 = 2$; see (24) in [76].[9] A forward half-interval approach is adopted to reduce the impact of edge effects induced by pre-averaging. The outcome is reported in Figure 5.5, where we plot the standardized pre-averaged bipower variation $V^*(2,0)^n$ and $V^*(1,1)^n$ using both $\hat{\Sigma}^*_n$ and the observed AVAR. As apparent, standardization with the subsampler tracks the standard normal curve closer compared to the observed AVAR. Indeed, the standard error of the studentized pre-averaged bipower variation is about 1.05 using $\hat{\Sigma}^*_n$, while it is about 1.20 using the observed AVAR.

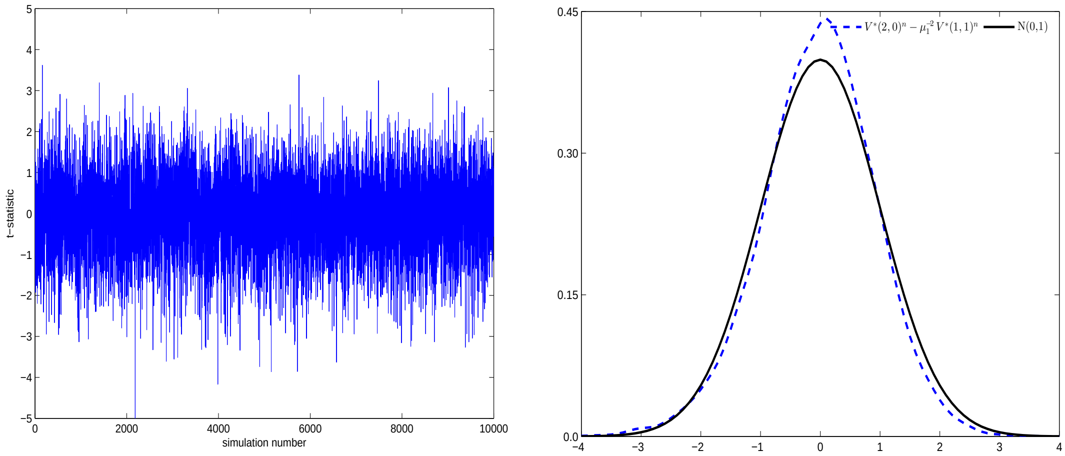
Overall, the simulation results suggest that inference based on $\hat{\Sigma}^*_n$ is fairly robust and

[9]The observed AVAR has a bias, which—although asymptotically negligible—could impair its accuracy in finite samples. The virtue of the two-scale construction, as advocated by Mykland and Zhang [76], is that the bias term cancels out.

Figure 5.4: Properties of the standardized $V^*(2,0)^n - \mu_1^{-2} V^*(1,1)^n$.

Panel A: Point estimate

Panel B: Kernel density estimate



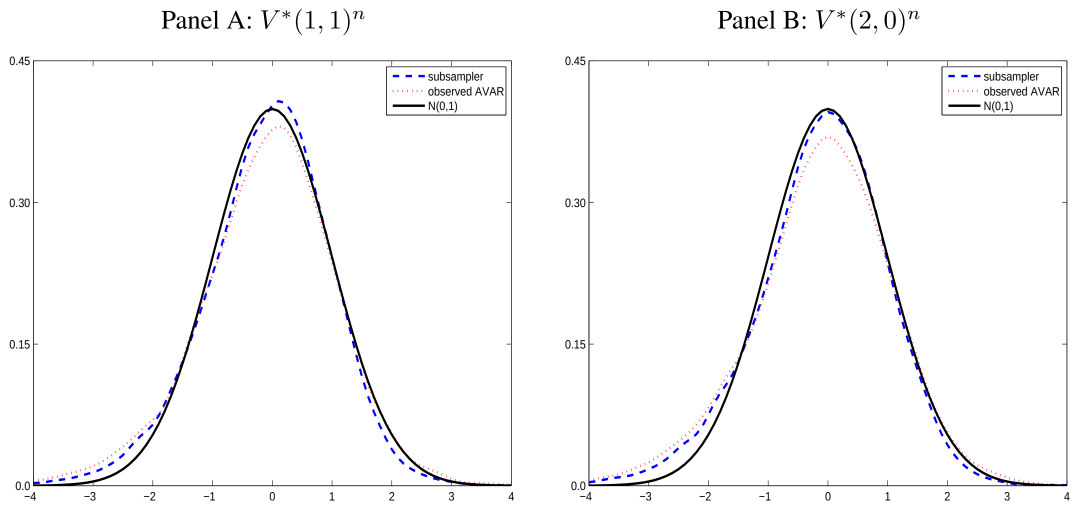
Note. We plot $\omega' V^*(q,r)^n$ with $\omega = (1, -\mu_1^{-2})'$, after it has been standardized by $\hat{\Sigma}_n^*$ based on $p = 10$ and $L = 15$, i.e. $n^{1/4}\big(V^*(2,0)^n - \mu_1^{-2} V^*(1,1)\big)/\sqrt{\omega' \hat{\Sigma}_n^* \omega}$. In Panel A, we plot the point estimates of this t-statistic across simulations, while Panel B displays the corresponding kernel smoothed density estimate. Throughout the figure, $n = 23,400$, and we set $k_n = [\theta\sqrt{n}]$ to implement pre-averaging using $\theta = 1$. The simulation data is from a Heston stochastic volatility model, as described in the main text. In Panel B, the $n_{\text{sim}} = 10,000$ simulated t-statistics are smoothed using a Gaussian kernel with optimal bandwidth selection $h = 1.06\hat{\sigma} n_{\text{sim}}^{-1/5}$, where $\hat{\sigma}$ is the sample standard deviation of data. The density function of a standard normal random variable (the solid black line) is superimposed as a visual reference.

delivers excellent outcomes, even for modest values of its tuning parameters.

5.3.4 Empirical work

Here, we provide a brief illustration of the subsample estimator in the context of some real financial high-frequency data. We analyze tick-data from the Standard and Poor's depository receipts, which is an exchange-traded fund that tracks the performance of the S&P 500 stock index. The shares are listed on several U.S. stock exchanges and trade under the ticker symbol SPY. It is a highly liquid security and provides a good starting point for the subsampler, which is data-intensive. We extracted a transaction price series of the SPY from the TAQ database. The data are recorded at milli-second precision and our complete sample covers the time period from January, 2007 to March, 2011; a total of 1,169 business days. The raw data was filtered for outliers using the recommendations of

Figure 5.5: Comparison of the subsampler and observed asymptotic variance.



Note. We show kernel smoothed density estimates of the standardized pre-averaged bipower variation estimator: $n^{1/4}\left(V^*(q_k,r_k)^n - V^*(q_k,r_k)\right)/\sqrt{\Sigma^*_{kk}}$, where Σ^*_{kk} is the kth diagonal element of Σ^*. We replace Σ^* by the subsampler (based on $L = 15$ and $p = 10$) and the observed asymptotic variance. The latter is computed with $B = 15$, and a two-scale combination of the K-averaged apparent quadratic covariation with $K_1 = 1$ and $K_2 = 2$ using forward half-interval estimators, as explained in [76] around (24). In the figure, $n = 23,400$ and we set $k_n = [\theta\sqrt{n}]$ to implement pre-averaging using $\theta = 1$. The simulation data is from a Heston stochastic volatility model, as described in the main text. The left panel holds the results for $V^*(1,1)^n$, while the right panel is for $V^*(2,0)^n$. The $n_{\text{sim}} = 10,000$ simulated t-statistics are smoothed using a Gaussian kernel with optimal bandwidth selection $h = 1.06\hat{\sigma}n_{\text{sim}}^{-1/5}$, where $\hat{\sigma}$ is the sample standard deviation of the data. The density function of a standard normal random variable (the solid black line) is superimposed as a visual reference.

[40].[10] We also restrict attention to the NYSE trading session, which runs from 9:30am till 4:00pm Eastern Time. Table 5.2 provides a few descriptive statistics that summarize key features of the dataset.

In Panel A of Figure 5.6, we plot the sample autocorrelation function (acf) of the noisy return series, $\Delta_i^n Y$, up to lag 15. There is a pronounced, significantly negative first-order autocorrelation of about -0.35, which is consistent with a bid-ask bounce interpretation of microstructure noise. The acf then increases and turns positive at lag three. The fourth

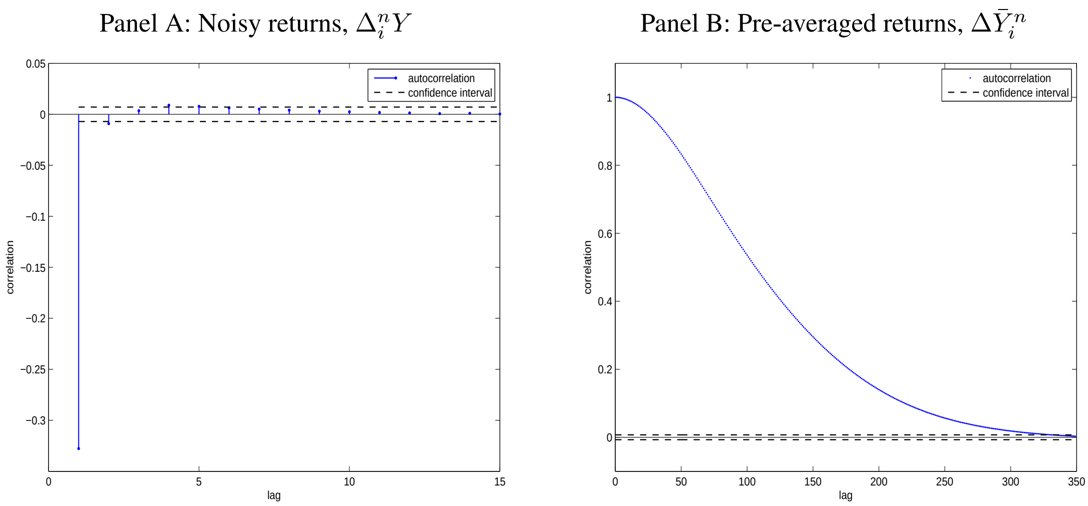
[10] The [40] filter is to a large extent based on the cleaning routines of [17]. The former use a backward-forward matching algorithm to compare a trade to the quote conditions prevailing in the market around the time of the transaction. The latter evaluate each trade against a single preceding bid-ask quote, which may lead to excessive removal of data in fast-moving markets. Apart from that, the filters are identical.

Table 5.2: Summary statistics of the SPY high-frequency data.

Statistic	Sample average	[Min; Max]
r_{oc}	0.003	[-8.254;7.349]
$\hat{\sigma}_{r_{oc}}$	17.037	[3.474;127.923]
n	113.769	[13.127;533.203]
K	320	[115;730]

Note. We report some descriptive statistics of the SPY high-frequency data. r_{oc} is the open-to-close return (in percent), i.e. the difference between the log-price of the last and first transaction of the day. $\hat{\sigma}_{r_{oc}}$ is a realized measure of the standard deviation of r_{oc}. We set $\hat{\sigma}_{r_{oc}} = 100 \times \sqrt{250 \times \widehat{IV}}$, where $\widehat{IV}$ is defined in (5.33). n is the sample size (in 1,000s), while K is the pre-averaging window. The sample period is January, 2007 through March, 2011.

Figure 5.6: Autocorrelation function of SPY return series.



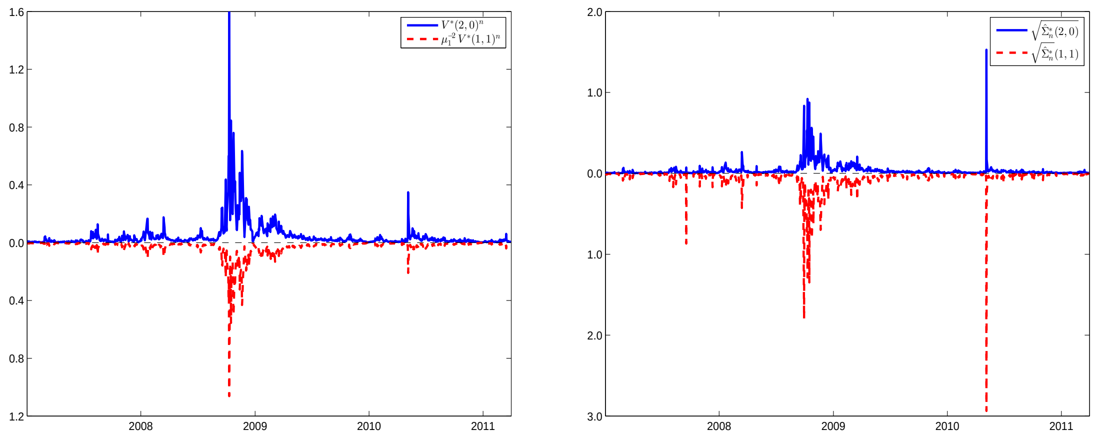
Note. We compute the empirical autocorrelation function (acf) of the SPY returns. Panel A is for the noisy returns (defined in (5.1)), while Panel B is for the pre-averaged returns (defined in (5.8)). The acf is estimated daily and then averaged over time. The sample period covers January, 2007 through March, 2011. The dashed line represents a 95% confidence interval for assessing the white noise null hypothesis.

and fifth autocorrelation actually fall outside the 95% confidence bands based on a white noise null hypothesis. Together with the subsequent monotonic decay of the acf, this indicates that noise operating at the tick-level is not i.i.d., as consistent with prior work [55]. We therefore proceed using a pre-averaging window based on $\theta = 1$, which should be a robust choice in light of the empirical evidence. The acf of the corresponding pre-averaged returns, $\Delta \bar{Y}_i^n$, is presented in Panel B of the figure. As expected, there is a strong dependence in this series up to lag k_n.

Figure 5.7: Time series of integrated variance estimates and standard error.

Panel A: Integrated variance

Panel B: Standard error



Note. In Panel A, we report the time series of the daily $V^*(2,0)^n$ and $\mu_1^{-2} V^*(1,1)^n$ estimates. The series were transformed into measures of the daily integrated variance, as detailed in (5.33). In Panel B, we plot the associated standard errors, based on $\sqrt{\hat{\Sigma}_n^*(2,0)}$ and $\sqrt{\hat{\Sigma}_n^*(1,1)}$. We use $p = 10$ and $L = 15$ to implement the
To facilitate the readability of the figure, the series based on $\mu_1^{-2} V^*(1,1)^n$ and $\sqrt{\hat{\Sigma}_n^*(1,1)}$ are reflected in the x-axis.

We report the resulting time series of $V^*(2,0)^n$ and $\mu_1^{-2} V^*(1,1)^n$ in Panel A of Figure 5.7. Note that the graph for $\mu_1^{-2} V^*(1,1)^n$ has been reflected in the x-axis. The statistics are first computed day-by-day across the whole sample and subsequently updated using (5.14) to provide annualized measures of the integrated variance (assuming 250 trading days p.a., on average), i.e.:

$$\widehat{\mathrm{IV}} = \frac{V^*(2,0)^n}{\theta \psi_2^{k_n}} - \frac{\psi_1^{k_n} \hat{\omega}^2}{\theta \psi_2^{k_n}} \xrightarrow{p} \int_0^1 \sigma_s^2 \mathrm{d}s, \tag{5.33}$$

with an identical transformation of $\mu_1^{-2} V^*(1,1)^n$.

The term $\dfrac{\psi_1^{k_n} \hat{\omega}^2}{\theta \psi_2^{k_n}}$ is a small bias correction that compensates the pre-averaged bipower variation for the residual effect of microstructure noise.[11] $\hat{\omega}^2$ is an estimate of the noise variance, ω^2. There are several estimators, which can serve the role of $\hat{\omega}^2$ [50]. Among these, we adopt the one from Oomen [79], which relies on the first-order autocorrelation of the noisy returns:

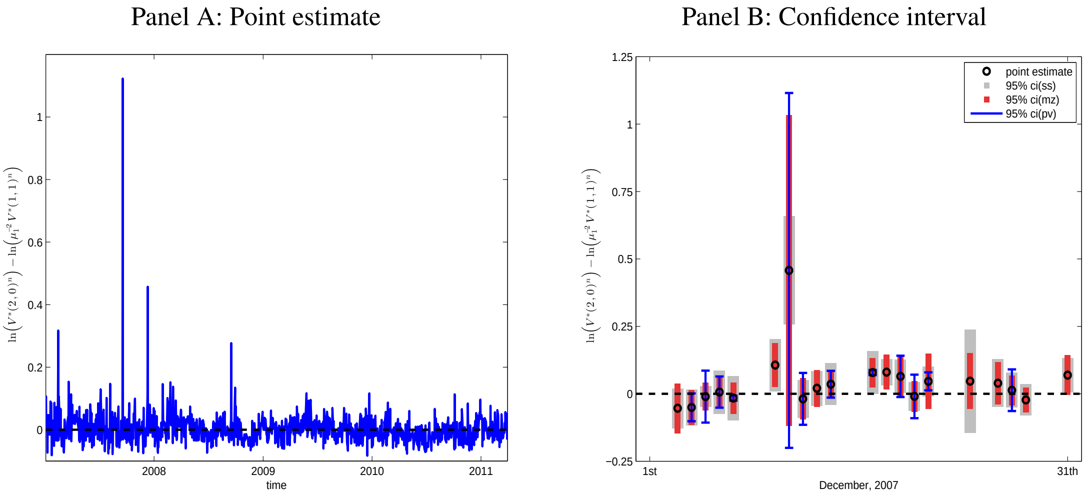
$$\hat{\omega}^2 = -\frac{1}{n-1} \sum_{i=1}^{n-1} \Delta_i^n Y \Delta_{i+1}^n Y \xrightarrow{p} \omega^2. \tag{5.34}$$

We find a high degree of time-variation and persistence in the $\widehat{\text{IV}}$ series. The onset of the financial crisis and—in particular—the unprecedented volatility surrounding the collapse of Lehman Brothers in 2008 stands out visibly. To attach a measure of uncertainty to these estimates, Panel B charts the associated standard error estimate, $\sqrt{\hat{\Sigma}_{11,n}^*}/\theta\psi_2^{k_n}$ and $\sqrt{\hat{\Sigma}_{22,n}^*}/\theta\psi_2^{k_n}$, where the latter are based on the subsampler with $L = 15$ and $p = 10$. As expected, high levels of volatility spill over into the standard errors and tend to decrease estimation accuracy. The apparent outliers showing up in the standard error series in Q3, 2007 and Q2, 2010 correspond to single days with unusual market activity. The first is September 18, 2007, where the Federal Open Market Committee (FOMC) announced an unexpected reduction of its target for the federal funds rate by 50 basis points, while the second is May 6, 2010; the day of the S&P 500 Flash Crash.

Turn next to Figure 5.8, where we conduct inference about $V^*(2,0)^n - \mu_1^{-2} V^*(1,1)^n$. In Panel A, we compute the difference in the logarithms of these numbers, i.e. $\ln\big(V^*(2,0)^n\big) - \ln\big(\mu_1^{-2} V^*(1,1)^n\big)$, which tends to be less volatile compared to the raw statistic. As shown, the majority of the point estimates hover around zero, which is the theoretical limit in diffusion models. There are some notable exceptions though, and in Panel B we examine one of these by zooming in on the month of December, 2007. Alongside the statistic, we here report a two-sided 95% confidence interval. Standard errors were found by applying the delta method (for the function $f(x,y) = \ln(y) - \ln(\mu_1^{-2} x)$) to the joint asymptotic distribution in (5.15) and then replacing the asymptotic variance of the difference by a feasible estimate. In particular, we compare a set of intervals based on the subsampler, $\hat{\Sigma}_n^*$, with those computed from the observed AVAR of [76], which is again computed as explained in the simulation section, and to the [85] estimator, $\tilde{\Sigma}_n^*$. If the latter leads to a negative variance estimate, it is excluded. As consistent with Table 5.1, this is a recurrent problem. Moreover, if all three estimates are well-defined, they are often closely aligned, but both the subsampler and observed AVAR appear less erratic, while $\tilde{\Sigma}_n^*$ is often very narrow

[11] The bias correction in (5.33) is only correct, when the noise is i.i.d. Meanwhile, the estimator of ω^2 we propose in (5.34) is robust to the presence of a heteroscedastic noise process, but it is generally not consistent for ρ^2 from Section 5.2.1, if the noise is autocorrelated. As the current application is merely illustrative, we ignore that issue here.

Figure 5.8: Inference about $\ln\big(V^*(2,0)^n\big) - \ln\big(\mu_1^{-2}V^*(1,1)^n\big)$.



Note. We compute the difference $\ln\big(V^*(2,0)^n\big) - \ln\big(\mu_1^{-2}V^*(1,1)^n\big)$, which shows potential violations of the assumed continuous sample path model. $V^*(2,0)^n$ and $V^*(1,1)^n$ are defined in (5.13). In Panel A, we plot the time series of the daily estimates of this number across the sample, which covers January, 2007 through March, 2011. In Panel B, we add a two-sided 95% confidence interval for the log-difference during the month of December, 2007. The standard errors are found by applying the delta method to the joint asymptotic distribution in (5.15). We replace the asymptotic covariance matrix by the subsampler (wide, grey box), $\hat{\Sigma}_n^*$, the [76] observed asymptotic variance computed as described in the simulation section (narrow, red box), and the estimator proposed in [85] (blue whisker), $\tilde{\Sigma}_n^*$. The former is implemented by setting $p = 10$ and $L = 15$. In both panels, the dashed line represents the limiting value in a pure diffusion model.

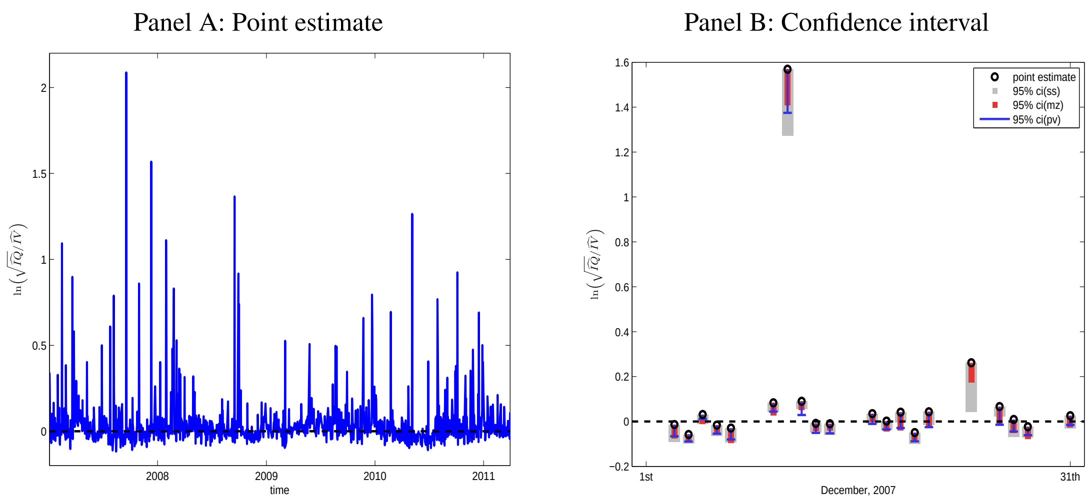
or wide. This is most visible from the big discrepancy on December 11, 2007, marking a day with yet another rate cut by the Fed. On this day, the condition number of $\tilde{\Sigma}_n^*$ is cond$(\tilde{\Sigma}_n^*) = 452.36$, which suggest that the underlying covariance matrix estimate is very fragile. The corresponding figure for the observed asymptotic variance is 64.68, which is again rather high, and indeed it also leads to a very large confidence interval here. Meanwhile, the condition number of the subsampler is more modest at cond$(\tilde{\Sigma}_n^*) = 15.16$, and it generally appears to be the most stable over time.

To end the section, we provide an alternative application, where the subsampler is used to draw inference about the amount of heteroscedasticity in noisy high-frequency data. To do this, we start by computing the statistics $V^*(2,0)^n$ and $V^*(4,0)^n$, i.e. the pre-averaged bipower variation based on the parameter $q = (4,2)'$ (and $r = (0,0)'$). Taking these as input, we appeal to (5.14) by forming the estimate:

$$\widehat{\mathrm{IQ}} = \frac{\mu_4^{-1} V^*(4,0)^n}{(\theta \psi_2^{k_n})^2} - \frac{2\psi_2^{k_n}\psi_1^{k_n}\hat{\omega}^2}{(\theta\psi_2^{k_n})^2}\widehat{\mathrm{IV}} - \frac{(\psi_1^{k_n}\hat{\omega}^2)^2}{(\theta^2\psi_2^{k_n})^2} \xrightarrow{p} \int_0^1 \sigma_s^4 \mathrm{d}s, \qquad (5.35)$$

which converges to the so-called integrated quarticity. We then exploit that $\sqrt{\widehat{\mathrm{IQ}}}/\widehat{\mathrm{IV}} \xrightarrow{p} \sqrt{\int_0^1 \sigma_s^4 \mathrm{d}s} / \int_0^1 \sigma_s^2 \mathrm{d}s \geq 1$, with equality if and only if σ is constant. Thus, an estimated ratio far above one suggests there is significant variation in volatility within the day, while a ratio close to one means σ can be regarded, as if it was approximately constant. This type of statistic has been exploited in earlier work to test for the parametric form of volatility [46, 98].

Figure 5.9: Inference about $\ln\left(\sqrt{\widehat{\mathrm{IQ}}}/\widehat{\mathrm{IV}}\right)$.



Note. We compute the log-ratio $\ln\left(\sqrt{\widehat{\mathrm{IQ}}}/\widehat{\mathrm{IV}}\right)$, which measures the degree of heteroscedasticity in σ within the day. $\widehat{\mathrm{IV}}$ and $\widehat{\mathrm{IQ}}$ are defined in Eqs. (5.33) and (5.35). In Panel A, we plot the time series of the daily estimates of this number across the sample, which covers January, 2007 through March, 2011. In Panel B, we add a left-sided 95% confidence interval for the log-ratio during the month of December, 2007 (the upper end of the interval extending to $+\infty$ is not shown). The standard errors are found by applying the delta method to the joint asymptotic distribution in (5.15). We replace the asymptotic covariance matrix by the subsampler (wide, grey box), $\hat{\Sigma}_n^*$, the [76] observed asymptotic variance computed as described in the simulation section (narrow, red box), and the estimator proposed in [85] (blue whisker), $\tilde{\Sigma}_n^*$. The former is implemented by setting $p = 10$ and $L = 15$. In both panels, the dashed line represents the limiting value in a constant volatility model.

The outcome of this exercise is collected in Figure 5.9. In Panel A, we plot the time series of the estimated log-ratio, i.e. $\ln\Big(\sqrt{\widehat{\mathrm{IQ}}}/\widehat{\mathrm{IV}}\Big)$. We again use a log-transformation in order to improve the scaling of the results and facilitate interpretation of the graphs. The log-ratio should cluster around zero, if volatility is constant. We observe an extreme degree of fluctuation in this statistic over time, and, as anticipated, there are many days, where the log-ratio is large. Still, we also find a decent portion of estimates, which are close to zero. Small negative numbers can be observed as a result of sampling variation. In Panel B, we complement the analysis by looking at the month of December, 2007. We add a left-sided 95% confidence interval for $\ln\Big(\sqrt{\widehat{\mathrm{IQ}}}/\widehat{\mathrm{IV}}\Big)$, where standard errors are again retrieved via the delta method and three estimates of the asymptotic covariance matrix. The interpretation is that on some days, such as the day of the FOMC meeting, volatility is changing a lot, while on others it is not moving much, which is consistent with the findings of [76, Figure 1]. Of course, the latter finding can also arise, if the high-frequency data is not informative enough to discriminate random sampling errors from genuine parameter variation in σ, which could be difficult in times of severe stress in financial markets. In this respect, it is important to acknowledge the limitations of the subsampler, which, albeit consistent, is itself subject to a substantial degree of sampling uncertainty in practice.

5.4 Proofs

5.4.1 Proof of Theorem 5.2

Again as in the noiseless case, we assume without loss of generality that the processes $a, \sigma, \tilde{a}, \tilde{\sigma}$ and $\tilde{v}$ are bounded following a standard localization procedure (see, [15]). We denote by C or C_p (if dependent on a parameter p) a generic constant which may differ from line to line. And, due to the polarization identity, we can (and shall) assume throughout that $m = 1$, so that all statistics are 1-dimensional. We begin by introducing some notation. For $m \geq i$, we define

$$\Delta\bar{Y}^n_{m,i} = \sum_{j=1}^{k_n} w\left(\frac{j}{k_n}\right)\left(\sigma_{\frac{i}{n}}\Delta^n_{m+j}W + \Delta^n_{m+j}\epsilon\right). \tag{5.36}$$

We note that $\Delta\bar{Y}^n_{m,i}$ approximates $\Delta\bar{Y}^n_m$ by evaluating σ at the point i/n. Moreover, we state two auxiliary results from [85], which provide the stochastic order of the statistics $\Delta\bar{S}^n_i$ for the processes, $S = W, X, \epsilon$ or Y for the former and allow us to use the limits ψ_i instead of ψ^n_i for $i = 1, 2$ without influencing the consistency statement for the latter.

Lemma 5.5 ([85], Lemma 1)

Assume that s is a non-negative real number, such that $\mathbb{E}[|\,\epsilon_t\,|^s] < \infty$. Then, for any i and n,

$$\mathbb{E}\Big[|\Delta\bar{Y}^n_{m,i}|^s \mid \mathcal{F}_{\frac{i}{n}}\Big] + \mathbb{E}\Big[|\Delta\bar{Y}^n_i|^s \mid \mathcal{F}_{\frac{i}{n}}\Big] \le Cn^{-s/4}. \tag{5.37}$$

Lemma 5.6 ([85], Lemma 2)

Let $s \ge 0$. Then,

$$\int_0^1 \Big(\theta\psi_2^n\sigma_u^2 + \frac{1}{\theta}\psi_1^n\omega^2\Big)^s \mathrm{d}u - \int_0^1 \Big(\theta\psi_2\sigma_u^2 + \frac{1}{\theta}\psi_1\omega^2\Big)^s \mathrm{d}u = o_p\big(n^{-1/4}\big). \tag{5.38}$$

We use the short form notation $t_{i,l} = (iL+l)pk_n/n$ and introduce some approximations of $\hat{\Sigma}^*_n$ and Σ^* as in Section 4.5:

$$\Sigma_n = \frac{1}{L}\sum_{l=1}^{L}\Bigg(\frac{n^{1/4}}{\sqrt{L}}\Big(V_l^*(q,r)^n - V^*(q,r)\Big)\Bigg)^2, \quad Q_n = \frac{k_n^2p^2}{n^{3/2}}\sum_{l=1}^{L}\Bigg(\sum_{i=1}^{n/Lpk_n}\chi^n_{(i-1)L+l}\Bigg)^2,$$

$$U_n = \frac{k_n^2p^2}{n^{3/2}}\sum_{l=1}^{L}\sum_{i=1}^{n/Lpk_n}\big(\chi^n_{(i-1)L+l}\big)^2, \quad R_n = \frac{k_n^2p^2}{n^{3/2}}\sum_{l=1}^{L}\sum_{i=1}^{n/Lpk_n}\mathbb{E}\big[\big(\chi^n_{(i-1)L+l}\big)^2 \mid \mathcal{F}_{t_{i-1,l-1}}\big],$$

where

$$\eta^n_i = \frac{n^{\frac{q+r}{4}}}{pk_n - 2k_n + 2}\sum_{m,m+k_n-1\in B_i(p)}|\Delta\bar{Y}^n_{m,(i-1)pk_n}|^q|\Delta\bar{Y}^n_{m+k_n,(i-1)pk_n}|^r,$$

and

$$\chi^n_i = \eta^n_i - \mathbb{E}\Big[\eta^n_i \mid \mathcal{F}_{\frac{(i-1)pk_n}{n}}\Big].$$

There exists a $C > 0$, independent of i, such that

$$\mathbb{E}\Big[\big(\eta^n_i\big)^4\Big] \le C \qquad \text{and} \qquad \mathbb{E}\Big[\big(\chi^n_i\big)^4\Big] \le \frac{C}{p^2}, \tag{5.39}$$

where the last inequality holds because of an application of Burkholder inequality and k_n-dependence of the term $\Delta\bar{Y}^n_{m,(i-1)pk_n}$ for each i.

As in the no-noise setting, we complete the proof by showing the following results and the fact that $p/\sqrt{n} \ll \sqrt{p}/n^{1/4} \ll 1/\sqrt{L}$ (which follows $\sqrt{n}/Lp^2 \to \infty$).

Proposition 5.7

Under the conditions of Theorem 5.2, it holds

(i) $\mathbb{E}[|\Sigma_n - Q_n|] \leq C\left(\frac{Lp^2}{\sqrt{n}} + \frac{1}{\sqrt{L}}\right)$,

(ii) $\mathbb{E}[|Q_n - U_n|] \leq \frac{C}{\sqrt{L}}$,

(iii) $\mathbb{E}[|U_n - R_n|] \leq C\frac{\sqrt{p}}{n^{1/4}}$,

(iv) $\mathbb{E}[|R_n - \Sigma^*|] \leq C\Big(\frac{p}{\sqrt{n}} + \frac{1}{p}\Big)$,

(v) $\mathbb{E}[|\hat{\Sigma}_n^* - \Sigma_n|] \leq \frac{C}{\sqrt{L}}$.

We proceed as in the noiseless case, i.e. in the order (iii), (ii), (iv), (i) and (v). However, the proof of (i) is the most complicated one.

Proof of Proposition 5.7 (iii). We observe that

$$U_n - R_n = \frac{k_n^2 p^2}{n^{3/2}} \sum_{i=1}^{n/p_n k_n} (\chi_i^n)^2 - \mathbb{E}\left[(\chi_i^n)^2 | \mathcal{F}_{\frac{(i-1)pk_n}{n}}\right].$$

In view of the martingale difference property as in the proof of Proposition 4.13(iii) and (5.39), we finish the proof with

$$\mathbb{E}[|U_n - R_n|^2] \leq C\frac{k_n^4 p^4}{n^3}\frac{n}{pk_n}\frac{1}{p^2} \leq C\frac{p}{n^{1/2}}. \tag{5.40}$$

□

Proof of Proposition 5.7(ii). Note that

$$Q_n - U_n = \frac{k_n^2 p^2}{n^{3/2}} \sum_{l=1}^{L} A_l^n$$

where

$$A_l^n = \sum_{i,j=1\ i\neq j}^{n/Lpk_n} \chi^n_{(i-1)L+l}\chi^n_{(j-1)L+l} = \left(\sum_{i=1}^{n/Lpk_n} \chi^n_{(i-1)L+l}\right)^2 - \left(\sum_{i=1}^{n/Lpk_n} (\chi^n_{(i-1)L+l})^2\right)$$
$$\equiv \left(S_l^{n/Lpk_n}\right)^2 - T_l^{n/Lpk_n}.$$

Since $A^n_{l_1}$ and $A^n_{l_2}$ are uncorrelated for every $l_1 \neq l_2$, we obtain

$$\mathbb{E}[|Q_n - U_n|^2] = \frac{k_n^4 p^4}{n^3} \sum_{l=1}^{L} \mathbb{E}[(A_l^n)^2] \tag{5.41}$$
$$\leq C\frac{k_n^4 p^4}{n^3} \sum_{l=1}^{L} \left(\mathbb{E}\left[\left(S_l^{n/Lpk_n}\right)^4\right] + \mathbb{E}\left[\left(T_l^{n/Lpk_n}\right)^2\right]\right). \tag{5.42}$$

To estimate the first sum above, we define $S_l^m := \sum_{i=1}^m \chi^n_{(i-1)L+l}$. We observe that $(S_l^m)_{m=1}^{n/Lpk_n}$ is a discrete martingale for each l. Then, the discrete Burkholder and Cauchy-Schwarz inequalities and (5.39) imply

$$\mathbb{E}[(S_l^{n/Lpk_n})^4] \leq C\mathbb{E}\left[\left(\sum_{i=1}^{n/Lpk_n} \left(\chi^n_{(i-1)L+l}\right)^2\right)^2\right] \leq C\Big(\frac{n}{Lp_nk_n}\Big)^2\frac{1}{p^2} \leq C\frac{n}{L^2p^4} \tag{5.43}$$

Using the Cauchy-Schwarz inequality and (5.39) again yield

$$\mathbb{E}[(T_l^{n/Lpk_n})^2] \leq C\frac{n}{L^2p^4}. \tag{5.44}$$

Then, we finish the proof using (5.41), (5.43) and (5.44). □

Proof of Proposition 5.7 (iv). We draw upon the proof of Lemma 8 in [85]. Since h is differentiable, we find that

$$\Sigma^* = 2\theta\int_0^1\int_0^2 h\big(\sigma_u, t, f(s)\big)\mathrm{d}s\mathrm{d}u$$
$$= \frac{2}{\sqrt{n}}\frac{pk_n}{n}\sum_{i=1}^{n/pk_n}\sum_{j=0}^{2k_n-1} h\left(\sigma_{\frac{(i-1)pk_n}{n}}, t_n, f^n\left(\frac{j}{k_n}\right)\right) + O_p\left(\frac{pk_n}{n}\right)$$
$$\equiv R_n' + O_p\left(\frac{pk_n}{n}\right) \tag{5.45}$$

where the function h defined in (5.16), $t_n = \left(\frac{\sqrt{n}}{k_n}\psi_1^n, \frac{k_n}{\sqrt{n}}\psi_2^n\right)$ and

$$f_1^n(s) = \sqrt{n} \sum_{j=0}^{k_n(1-s)} (g_j^n - g_{j+1}^n)(g_{j+sk_n}^n - g_{j+sk_n+1}^n)$$

$$f_2^n(s) = \frac{1}{\sqrt{n}} \sum_{j=0}^{k_n(1-s)} g_j^n g_{j+sk_n}^n.$$

The term $O_p\left(\frac{pk_n}{n}\right)$ in (5.45) comes from Riemann approximation of σ, which dominates the others.

To estimate the term $R_n - R_n'$, we recall that

$$R_n = \frac{k_n^2 p^2}{n^{3/2}} \sum_{i=1}^{n/pk_n} \mathbb{E}\Big[(\chi_i^n)^2 \mid \mathcal{F}_{\frac{(i-1)pk_n}{n}}\Big].$$

For $m \geq l \geq i$, we get

$$\begin{aligned} h\left(\sigma_{\frac{i}{n}}, t_n, f^n\left(\frac{m-l}{k_n}\right)\right) \\ &= \mathbb{E}\Big[|n^{1/4}\Delta\bar{Y}_{m,i}^n|^q |n^{1/4}\Delta\bar{Y}_{m+k_n,i}^n|^r \times |n^{1/4}\Delta\bar{Y}_{l,i}^n|^q |n^{1/4}\Delta\bar{Y}_{l+k_n,i}^n|^r \mid \mathcal{F}_{\frac{i}{n}}\Big] \\ &- \mathbb{E}\Big[|n^{1/4}\Delta\bar{Y}_{m,i}^n|^q |n^{1/4}\Delta\bar{Y}_{m+k_n,i}^n|^r \mid \mathcal{F}_{\frac{i}{n}}\Big] \times \mathbb{E}\Big[|n^{1/4}\Delta\bar{Y}_{l,i}^n|^q |n^{1/4}\Delta\bar{Y}_{l+k_n,i}^n|^r \mid \mathcal{F}_{\frac{i}{n}}\Big]. \end{aligned}$$

Note that the above term vanishes for $m-l \geq 2k_n$. Then, by denoting $N = pk_n - 2k_n + 2$, we find that

$$\begin{aligned} N\mathbb{E}\Big[(\chi_i^n)^2 \mid \mathcal{F}_{\frac{(i-1)pk_n}{n}}\Big] &= h\big(\sigma_{\frac{(i-1)pk_n}{n}}, t_n, f^n(0)\big) + \frac{2}{N}\sum_{j=1}^{2k_n-1}(N-j)h\left(\sigma_{\frac{(i-1)pk_n}{n}}, t_n, f^n\left(\frac{j}{k_n}\right)\right) \\ &= 2\sum_{j=0}^{2k_n-1} h\left(\sigma_{\frac{(i-1)pk_n}{n}}, t_n, f^n\left(\frac{j}{k_n}\right)\right) + O_p(1) + O_p\left(\frac{k_n}{p}\right). \end{aligned}$$

This yields that:

$$\frac{pk_n}{\sqrt{n}}\mathbb{E}\Big[(\chi_i^n)^2 \mid \mathcal{F}_{\frac{(i-1)pk_n}{n}}\Big] = \frac{2}{\sqrt{n}}\sum_{j=0}^{2k_n-1} h\left(\sigma_{\frac{(i-1)pk_n}{n}}, t_n, f^n\left(\frac{j}{k_n}\right)\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{p}\right)$$

uniformly in i. As a result,

$$\mathbb{E}\big[|R_n - R_n'|\big] \le C\bigg(\frac{1}{\sqrt{n}} + \frac{1}{p}\bigg). \tag{5.46}$$

In view of (5.45) – (5.46), the proof is complete. □

To show Proposition 5.7 (i), we need a preparation. Let us denote

$$\tilde{V}_l^*(q,r)^n = \frac{Lpk_n}{n}\sum_{i=1}^{n/Lpk_n} \eta_{(i-l)L+l}^n, \qquad \hat{V}_l^*(q,r)^n = \frac{Lpk_n}{n}\sum_{i=1}^{n/Lpk_n} \mathbb{E}\Big[\eta_{(i-l)L+l}^n \mid \mathcal{F}_{t_{i-1,l-1}}\Big].$$

Then, from the decomposition

$$\begin{aligned} V_l^*(q,r)^n - V^*(q,r) &= \Big(V_l^*(q,r)^n - \tilde{V}_l^*(q,r)^n\Big) + \Big(\tilde{V}_l^*(q,r)^n - \hat{V}_l^*(q,r)^n\Big) \\ &\quad + \Big(\hat{V}_l^*(q,r)^n - V^*(q,r)\Big) \end{aligned}$$

together with the identity $(a+b+c)^2 - b^2 = 2a(b+c) + 2cb + a^2 + c^2$, we obtain

$$\Sigma_n - Q_n = D_n^{(1)} + D_n^{(2)} + D_n^{(3)} + D_n^{(4)},$$

where

$$D_n^{(1)} = \frac{2\sqrt{n}}{L^2}\sum_{l=1}^{L}\Big(V_l^*(q,r)^n - \tilde{V}_l^*(q,r)^n\Big)\Big(\tilde{V}_l^*(q,r)^n - V^*(q,r)\Big),$$

$$D_n^{(2)} = \frac{2\sqrt{n}}{L^2}\sum_{l=1}^{L}\Big(\hat{V}_l^*(q,r)^n - V^*(q,r)\Big)\Big(\tilde{V}_l^*(q,r)^n - \hat{V}_l^*(q,r)^n\Big),$$

$$D_n^{(3)} = \frac{\sqrt{n}}{L^2}\sum_{l=1}^{L}\Big(V_l^*(q,r)^n - \tilde{V}_l^*(q,r)^n\Big)^2,$$

$$D_n^{(4)} = \frac{\sqrt{n}}{L^2}\sum_{l=1}^{L}\Big(\hat{V}_l^*(q,r)^n - V^*(q,r)\Big)^2.$$

To bound these terms, we exploit the following auxiliary Lemma.

Lemma 5.8

Assume that the conditions of Theorem 5.2 are fulfilled. Then, uniformly in l:

(a) $\mathbb{E}[|\tilde{V}_l^*(q,r)^n - \hat{V}_l^*(q,r)^n|^2] \leq C\frac{L}{\sqrt{n}}$,

(b) $\mathbb{E}[|\hat{V}_l^*(q,r)^n - V^*(q,r)|^2] \ \leq C\Big(\frac{L^2p^2}{n} + \frac{1}{\sqrt{n}}\Big)$,

(c) $\mathbb{E}[|V_l^*(q,r)^n - \tilde{V}_l^*(q,r)^n|^2] \leq C\frac{Lp^2}{n}$.

Proof of Lemma 5.8. Part (a) is derived from exploiting the martingale difference property with (5.39). To prove part (b), we start with the decomposition

$$\hat{V}_l^*(q,r)^n - V^*(q,r) = \Big(\hat{V}_l^*(q,r)^n - \check{V}_l^*(q,r)^n\Big) + \Big(\check{V}_l^*(q,r)^n - V^*(q,r)\Big),$$

where

$$\check{V}_l^*(q,r)^n = \mu_q\mu_r\frac{Lpk_n}{n}\sum_{i=1}^{n/Lpk_n}\Big(\theta\psi_2\sigma^2_{t_{i-1,l-1}} + \frac{1}{\theta}\psi_1\omega^2\Big)^{\frac{q+r}{2}}. \tag{5.47}$$

To deal with the second term, we recall Lemma 4.14 (b). Hence the Riemann approximation yields

$$\mathbb{E}\Big[|\check{V}_l^*(q,r)^n - V^*(q,r)|^2\Big] \leq C\frac{L^2p^2}{n}. \tag{5.48}$$

To estimate the first term, let $m \in B_i(p)$. We employ Lemma 4 from [85] to conclude that

$$\begin{aligned}\mathbb{E}\Big[|n^{1/4}\Delta\bar{Y}^n_{m,(i-1)pk_n}|^q|n^{1/4}\Delta\bar{Y}^n_{m+k_n,(i-1)pk_n}|^r \mid \mathcal{F}_{\frac{(i-1)pk_n}{n}}\Big]&\\ = \mu_q\mu_r\Big(\theta\psi_2\sigma^2_{\frac{(i-1)pk_n}{n}} + \frac{1}{\theta}\psi_1\omega^2\Big)^{\frac{q+r}{2}} + o_p\big(n^{-1/4}\big),&\end{aligned} \tag{5.49}$$

uniformly in i and m. Consequently, we find that

$$\mathbb{E}\big[\eta_i^n \mid \mathcal{F}_{t_{i-1,l-1}}\big] = \mu_q\mu_r\Big(\theta\psi_2\sigma^2_{\frac{(i-1)pk_n}{n}} + \frac{1}{\theta}\psi_1\omega^2\Big)^{\frac{q+r}{2}} + o_p\big(n^{-1/4}\big),$$

uniformly in i and m. Using these insights, we can finish part (b) by deducing that

$$\mathbb{E}\Big[|\hat{V}_l^*(q,r)^n - \check{V}_l^*(q,r)^n|^2\Big] \leq \frac{C}{\sqrt{n}}. \tag{5.50}$$

As for the proof of part (c), we proceed as in the proof of noiseless case (see also page 2818 in [85]). Thus, we merely provide a sketch of the main steps for $r = 0$. For any $m \geq i$, we define

$$\xi^n_{m,i}(1) \equiv \sum_{j=1}^{k_n} w\Big(\frac{j}{k_n}\Big)\Bigg(\frac{1}{n}a_{\frac{i}{n}} + \int_{\frac{m+j-1}{n}}^{\frac{m+j}{n}} \Big[\tilde{\sigma}_{\frac{i}{n}}\big(W_s - W_{\frac{i}{n}}\big) + \tilde{v}_{\frac{i}{n}}\big(B_s - B_{\frac{i}{n}}\big)\Big]\mathrm{d}W_s\Bigg)$$

$$\xi^n_{m,i}(2) \equiv \sum_{j=1}^{k_n} w\Big(\frac{j}{k_n}\Big)\Bigg(\int_{\frac{m+j-1}{n}}^{\frac{m+j}{n}} \big(a_s - a_{\frac{i}{n}}\big)\mathrm{d}s + \int_{\frac{m+j-1}{n}}^{\frac{m+j}{n}}\int_{\frac{i}{n}}^{s} \tilde{a}_u \mathrm{d}u\mathrm{d}W_s$$

$$+ \int_{\frac{m+j-1}{n}}^{\frac{m+j}{n}}\Big(\int_{\frac{i}{n}}^{s}\big(\tilde{\sigma}_u - \tilde{\sigma}_{\frac{i}{n}}\big)\mathrm{d}W_u + \int_{\frac{i}{n}}^{s}\big(\tilde{v}_u - \tilde{v}_{\frac{i}{n}}\big)\mathrm{d}B_u\Big)\mathrm{d}W_s\Bigg).$$

We note that $\Delta\bar{Y}^n_m - \Delta\bar{Y}^n_{m,i} = \xi^n_{m,i}(1) + \xi^n_{m,i}(2) \equiv \xi^n_{m,i}$. Assumption (**H**), the Hölder and Burkholder inequalities imply

$$\mathbb{E}\Big[|\xi^n_{m,i}(1)|^4\Big] \leq C\frac{p^2}{n^2}, \tag{5.51}$$

$$\mathbb{E}\Big[|\xi^n_{m,i}(2)|^4\Big] \leq C\frac{p^4}{n^3}, \tag{5.52}$$

$$\mathbb{E}\Big[|\xi^n_{m,i}|^4\Big] \leq C\frac{p^2}{n^2}. \tag{5.53}$$

Now, we let $f(x) = |x|^q$. Taylor's theorem then yields that

$$V^*_l(q,r) - \tilde{V}^*_l(q,r)^n = S^n_l(1) + S^n_l(2) + O_p\big(\frac{p}{\sqrt{n}}\big),$$

where

$$S_l^n(1) = \frac{Lpk_n}{n} \frac{n^{q/4}}{pk_n - 2k_n + 2} \times \sum_{i=1}^{n/Lpk_n} \sum_{m \in B_{(i-1)L+l}(p)} f'\Big(\Delta \bar{Y}^n_{m,((i-1)L+(l-1))pk_n}\Big) \xi^n_{m,(i-1)L+(l-1))pk_n}(1),$$

$$S_l^n(2) = \frac{Lpk_n}{n} \frac{n^{q/4}}{pk_n - 2k_n + 2} \times \sum_{i=1}^{n/Lpk_n} \sum_{m \in B_{(i-1)L+l}(p)} f'\Big(\Delta \bar{Y}^n_{m,((i-1)L+(l-1))pk_n}\Big) \xi^n_{m,(i-1)L+(l-1))pk_n}(2),$$

and the error $O_p\big(p/\sqrt{n}\big)$ occurs due to the differentiability of f and (5.53). In order to bound these terms, we note that Assumption (**N**) implies that $(W, B, \epsilon) \overset{d}{=} -(W, B, \epsilon)$. Also, since $f'\big(\Delta \bar{Y}^n_{m,i}\big)$ is an odd function and $\xi^n_{m,i}(1)$ is an even function of (W, B, ϵ), it follows

$$\mathbb{E}\Big[f'\big(\Delta \bar{Y}^n_{m,ipk_n}\big)\xi^n_{m,ipk_n}(1) \mid \mathcal{F}_{\frac{ipk_n}{n}}\Big] = 0.$$

This property together with the Cauchy-Schwarz inequality, (5.37) and (5.51) mean that

$$\mathbb{E}\big[|S_l^n(1)|^2\big] \leq C\frac{Lp^2}{n}. \tag{5.54}$$

Applying the Cauchy-Schwarz inequality again, combined with (5.37) and (5.52), we also find that

$$\mathbb{E}\big[|S_l^n(2)|^2\big] \leq C\frac{p^2}{n}, \tag{5.55}$$

and with (5.54) – (5.55) at hand, the proof is complete. □

The following results are then sufficient to complete the proof of Theorem 5.2.

Lemma 5.9

Assume that the conditions of Theorem 5.2 are fulfilled. Then, it holds that *(a)* $\mathbb{E}[|D_n^{(4)}|] \le C\Big(\frac{Lp^2}{\sqrt{n}} + \frac{1}{L}\Big)$,

(b) $\mathbb{E}[|D_n^{(3)}|] \le C\frac{p^2}{\sqrt{n}}$,

(c) $\mathbb{E}[|D_n^{(1)}|] \le C\left(\frac{p}{n^{1/4}} + \frac{\sqrt{L}p^2}{\sqrt{n}}\right)$,

(d) $\mathbb{E}[|D_n^{(2)}|] \le C\Big(\frac{1}{\sqrt{L}} + \frac{Lp^{3/2}}{\sqrt{n}}\Big)$.

Proof of Lemma 5.9. Again, part (a) follows from (b). Concerning parts (b) – (c), we apply Lemma 5.8 and find that

$$\mathbb{E}\Big[\Big(\tilde{V}_l^*(q,r)^n - V^*(q,r)\Big)^2\Big] \le C\Big(\frac{L}{\sqrt{n}} + \frac{L^2p^2}{n}\Big). \tag{5.56}$$

Then, the Cauchy-Schwarz inequality and (5.56) yield

$$\big(\mathbb{E}\big[|D_n^{(1)}|\big]\big)^2 \le C\Big(1+\frac{Lp^2}{\sqrt{n}}\Big)\frac{\sqrt{n}}{L^2}\sum_{l=1}^{L}\mathbb{E}\big[|V_l^*(q,r)^n - \tilde{V}_l^*(q,r)^n|^2\big] = C\Big(1+\frac{Lp^2}{\sqrt{n}}\Big)\mathbb{E}\big[|D_n^{(3)}|\big].$$

Hence it is enough to show part (b), which follows from Lemma 5.8 (c).

Regarding part (d), we start by denoting $\phi(x) = \mu_q\mu_r\Big(\theta\psi_2x^2 + \frac{1}{\theta}\psi_1\omega^2\Big)^{\frac{q+r}{2}}$. Note that $\phi(x)$ is a smooth function of x, because both q and r are even. After recalling (5.47), an application of Taylor's theorem and (4.41) for σ implies that

$$D_n^{(2)} = E_n + F_n + G_n + O_p\Big(\frac{Lp^{3/2}}{\sqrt{n}}\Big) + O_p\Big(\frac{p}{n^{1/4}}\Big),$$

where the last error term comes from the boundary integral around 0 and 1,

$$E_n = \frac{2\sqrt{n}}{L^2}\sum_{l=1}^{L}\Bigg(\sum_{i=1}^{n/Lpk_n}\phi'(\sigma_{t_{i-1,l-1}})\int_{t_{i-1,l-1}}^{t_{i,l-1}}\big(\sigma_{t_{i-1,l-1}} - \sigma_s\big)\mathrm{d}s\Bigg) \times \Big(\tilde{V}_l^*(q,r)^n - \hat{V}_l^*(q,r)^n\Big),$$

$$F_n = -\frac{\sqrt{n}}{L^2}\sum_{l=1}^{L}\Bigg(\sum_{i=1}^{n/Lpk_n}\phi''(\sigma_{t_{i-1,l-1}})\int_{t_{i-1,l-1}}^{t_{i,l-1}}\big(\sigma_{t_{i-1,l-1}} - \sigma_s\big)^2\mathrm{d}s\Bigg) \times \Big(\tilde{V}_l^*(q,r)^n - \hat{V}_l^*(q,r)^n\Big),$$

$$G_n = \frac{2\sqrt{n}}{L^2}\sum_{l=1}^{L}\Big(\hat{V}_l^*(q,r)^n - \check{V}_l^*(q,r)^n\Big)\Big(\tilde{V}_l^*(q,r)^n - \hat{V}_l^*(q,r)^n\Big).$$

From the Cauchy-Schwarz inequality, Lemma 5.8(a) and (5.50), we get

$$\mathbb{E}\big[|G_n|\big] \le \frac{C}{\sqrt{L}}.$$

At this stage, we assume $\tilde{v}_s = 0$ as in the noiseless setting. We recall Assumption (**H**), apply (4.41) to $\tilde{a}, \tilde{\sigma}$, and subsequently use Taylor's theorem to conclude that

$$E_n = -E_n(1) - E_n(2) + O_p\bigg(\frac{Lp^{3/2}}{\sqrt{n}}\bigg),$$

where

$$E_n(1) = \frac{2\sqrt{n}}{L^2}\sum_{l=1}^{L}\Bigg(\sum_{i=1}^{n/Lpk_n}\phi'(\sigma_{t_{i-1,l-1}})\frac{L^2k_n^2p^2}{2n^2}\tilde{a}_{t_{i-1,l-1}}\Bigg)\Big(\tilde{V}_l^*(q,r)^n - \hat{V}_l^*(q,r)^n\Big),$$

$$E_n(2) = \frac{2\sqrt{n}}{L^2}\sum_{l=1}^{L}\Bigg(\sum_{i=1}^{n/Lpk_n}\phi'(\sigma_{t_{i-1,l-1}})\int_{t_{i-1,l-1}}^{t_{i,l-1}}\tilde{\sigma}_{t_{i-1,l-1}}\big(W_s - W_{t_{i-1,l-1}}\big)\mathrm{d}s\Bigg)\Big(\tilde{V}_l^*(q,r)^n - \hat{V}_l^*(q,r)^n\Big).$$

For the term $E_n(1)$, we proceed as in the noiseless case. After recalling (5.39) and Lemma 5.8(a), we find that

$$\mathbb{E}\big[|E_n(1)|\big] \le C\Bigg(\frac{p}{n^{1/4}} + \frac{Lp^{3/2}}{\sqrt{n}}\Bigg).$$

Next, the term F_n can be handled in a similar fashion. Thus, we get the estimate

$$\mathbb{E}\big[|F_n|\big] \le C\Bigg(\frac{p}{n^{1/4}} + \frac{Lp^{3/2}}{\sqrt{n}}\Bigg).$$

So, it will be completed if we can show that

$$\mathbb{E}\big[|E_n(2)|\big] \le C\frac{p}{n^{1/4}}. \tag{5.57}$$

Without loss of generality, we assume throughout the remainder of the proof that $r = 0$. We then appeal to the binomial theorem in order to find an expansion of

$$|\Delta\bar{Y}_{m,i}^n|^q = (\Delta\bar{Y}_{m,i}^n)^q = \Big(\sigma_{\frac{i}{n}}\Delta\bar{W}_m^n + \Delta\bar{\epsilon}_m^n\Big)^q,$$

whereby we can separate $\tilde{V}_l^*(q,r)^n - \hat{V}_l^*(q,r)^n$, and hence $E_n(2)$, into $q+1$ terms of the form

$$E_n(2) = \sum_{s=0}^{q} E_n^{(s)}(2),$$

where

$$E_n^{(s)}(2) = \frac{2\sqrt{n}}{L^2} \sum_{l=1}^{L} \left(\sum_{i=1}^{n/Lpk_n} \phi'(\sigma_{t_{i-1,l-1}}) \int_{t_{i-1,l-1}}^{t_{i,l-1}} \tilde{\sigma}_{t_{i-1,l-1}} (W_s - W_{t_{i-1,l-1}}) \mathrm{d}s \right) \times \left(\frac{Lpk_n}{n} \sum_{i=1}^{n/Lpk_n} \chi^n_{(i-1)L+l}(s) \right),$$

$$\chi_i^n(s) = \frac{q!}{s!(q-s)!} \frac{n^{q/4}}{pk_n - k_n + 2} \sum_{m \in B_i(p)} \left(\sigma_{\frac{(i-1)pk_n}{n}} \right)^{q-s} \Big((\Delta \bar{W}_m^n)^{q-s} (\Delta \bar{\epsilon}_m^n)^s - \mathbb{E}\big[(\Delta \bar{W}_m^n)^{q-s} (\Delta \bar{\epsilon}_m^n)^s \big] \Big).$$

As a result, it is sufficient to show that

$$\mathbb{E}\big[|E_n^{(s)}(2)| \big] \le C \frac{p}{n^{1/4}}, \tag{5.58}$$

where s is an arbitrary integer chosen from $0 \le s \le q$. Note the equality:

$$(\Delta \bar{W}_m^n)^{q-s} (\Delta \bar{\epsilon}_m^n)^s - \mathbb{E}\big[(\Delta \bar{W}_m^n)^{q-s} (\Delta \bar{\epsilon}_m^n)^s \big] = \mathbb{E}\big[(\Delta \bar{\epsilon}_m^n)^s \big] \Big((\Delta \bar{W}_m^n)^{q-s} - \mathbb{E}\big[(\Delta \bar{W}_m^n)^{q-s} \big] \Big) + (\Delta \bar{W}_m^n)^{q-s} \Big((\Delta \bar{\epsilon}_m^n)^s - \mathbb{E}\big[(\Delta \bar{\epsilon}_m^n)^s \big] \Big).$$

We then divide $\chi_i^n(s)$, and hence $E_n^{(s)}(2)$, into two parts and denote (by preserving the above order)

$$E_n^{(s)}(2) = \bar{E}_n^{(s)}(2) + \tilde{E}_n^{(s)}(2). \tag{5.59}$$

The term $\bar{E}_n^{(s)}(2)$ can be handled using a decomposition as in (4.58) in the no-noise proof:

$$\big(\bar{E}_n^{(s)}(2) \big)^2 = \bar{E}_n^{(s)}(2.1) + \bar{E}_n^{(s)}(2.2),$$

where $\bar{E}_n^{(s)}(2.1)$ and $\bar{E}_n^{(s)}(2.2)$ are, respectively, composed of squared and mixed terms. We recall that the sequence $n^{s/4}\mathbb{E}\big[(\Delta\bar{\epsilon}_m^n)^s\big]$ is uniformly bounded in m and n. Then, proceeding as in (4.59), we find that

$$\mathbb{E}\Big[\bar{E}_n^{(s)}(2.1)\Big] \leq C\frac{p^2}{\sqrt{n}}. \tag{5.60}$$

Assumption (**M**), and the steps in the Malliavin calculus are also applied in (4.63) – (4.67), mean that

$$\mathbb{E}\Big[\bar{E}_n^{(s)}(2.2)\Big] \leq C\frac{p^2}{\sqrt{n}}. \tag{5.61}$$

For the last term $\tilde{E}_n^{(s)}(2)$, we recall that X and ϵ are independent and $(\Delta\bar{\epsilon}_m^n)^s - \mathbb{E}\big[(\Delta\bar{\epsilon}_m^n)^s\big]$ has mean zero. Then, we decompose $\big(\tilde{E}_n^{(s)}(2)\big)^2$ as in (4.58), and since the mixed terms for different l's are mean zero, we find that

$$\mathbb{E}\Big[\big(\tilde{E}_n^{(s)}(2)\big)^2\Big] \leq C\frac{p^2}{\sqrt{n}}. \tag{5.62}$$

Hence (5.60) – (5.62) lead to (5.58). □

Proof of Proposition 5.7 (v). Using a difference of squares, we can write

$$\begin{aligned}\hat{\Sigma}_n^* - \Sigma_n = \frac{\sqrt{n}}{L^2}\sum_{l=1}^{L}\Big(V^*(q,r) - V^*(q,r)^n\Big)\Big(2\Big(V_l^*(q,r)^n - V^*(q,r)\Big)\\ - \Big(V^*(q,r)^n - V^*(q,r)\Big)\Big).\end{aligned} \tag{5.63}$$

We know that $V^*(q,r) - V^*(q,r)^n = O_p(1/n^{1/4})$ and due to Lemma 5.8, we have

$$\begin{aligned}\frac{n^{1/4}}{\sqrt{L}}\Big(V_l^*(q,r)^n - V^*(q,r)\Big) &= O_p(1)\\ \frac{n^{1/4}}{\sqrt{L}}\Big(V^*(q,r)^n - V^*(q,r)\Big) &= O_p\Big(\frac{1}{\sqrt{L}}\Big).\end{aligned}$$

Plugging in these facts into (5.63) yields

$$\mathbb{E}[|\hat{\Sigma}_n^* - \Sigma_n|] \leq \frac{C}{\sqrt{L}}.$$

□

5.4.2 Proof of Theorem 5.3

The proof is reminiscent to the proof of Theorem 4.6. A careful inspection of the proof of Theorem 5.2 implies that the following steps also hold under the weaker assumptions of Theorem 5.3:

$$\mathbb{E}\big[|Q_n - U_n|\big] \to 0, \quad \mathbb{E}\big[|U_n - R_n|\big] \to 0, \quad \mathbb{E}\big[|R_n - \Sigma^*|\big] \to 0, \quad \mathbb{E}\big[|\hat{\Sigma}_n^* - \Sigma_n|\big] \to 0,$$

Hence given the rates on p and L, it suffices to show that $\mathbb{E}\big[|\Sigma_n - Q_n|\big] \to 0$. A direct consequence of discrete Burkholder inequality implies Lemma 5.8(a) while applying Lemma 4 from [85], we can prove Lemma 5.8(b). Therefore, we can show by applying Cauchy-Schwarz inequality that

$$\mathbb{E}[|D_n^{(2)}|] \leq C\Big(\frac{\sqrt{L}p}{n^{1/4}} + \frac{1}{\sqrt{L}}\Big) \quad \text{and}$$
$$\mathbb{E}[|D_n^{(4)}|] \leq C\Big(\frac{Lp^2}{n} + \frac{1}{L}\Big).$$

Furthermore, from the proof of Lemma 5 ((A.16)–(A.17)) in [85], we can prove that the right-hand side estimate of Lemma 5.8(c) changes from Lp^2/n to $p/\sqrt{n}$, because the estimate in (5.52) is p^2/n^2 instead of p^4/n^3. Then, we finish the proof by an additional condition $L/p \to \infty$, which implies that the right-hand side estimate of Lemma 5.8(a) dominates that of Lemma 5.8(c). □

Chapter 6

Approximation of Brownian semi-stationary processes

In 2007, Barndorff-Nielsen and Schmiegel introduced a class of spatio-temporal stochastic processes called ambit fields in a series of papers [20, 21, 22] in the context of finding flexible stochastic models to describe turbulence. However, manifold applications have been found in mathematical finance such as energy spot prices [10], power markets [32] and electricity forward markets [11] and in biology (modelling of tumor growth) [19]. We refer to (1.8) for the general formula of the ambit fields.

An important purely temporal subclass of ambit fields are the so called *Lévy semi-stationary processes* or in short $\mathcal{LSS}$ processes, which are defined as

$$X_t = \mu + \int_{-\infty}^{t} g(t-s)\sigma_s L(\mathrm{d}s) + \int_{-\infty}^{t} q(t-s)a_s \mathrm{d}s, \tag{6.1}$$

where L is a two-sided one dimensional Lévy motion and the ambit sets are given via $A_t = D_t = (-\infty, t)$. The notion of a semi-stationary process comes from the fact that the process $(X_t)_{t\in\mathbb{R}}$ is stationary whenever $(a_t, \sigma_t)_{t\in\mathbb{R}}$ is stationary and independent of $(L_t)_{t\in\mathbb{R}}$. In the past years, stochastic analysis, probabilistic properties and statistical inference for Lévy semi-stationary processes have been studied in numerous papers. We refer to [8, 9, 13, 14, 29, 22, 33, 43, 51, 82] for the mathematical theory as well as to [96] for an application of $\mathcal{LSS}$ models in electricity prices and to [12, 83] for recent surveys on theory of ambit fields and their applications.

In this thesis, attention is given to a class of ambit fields, a Brownian semi-stationary process or in short $\mathcal{BSS}$ process, i.e. $\mathcal{LSS}$ process driven by Brownian motion. We consider a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\in\mathbb{R}}, \mathbb{P})$, on which all processes are defined. Let a Brownian semi-stationary process be the form

$$X_t = \mu + \int_{-\infty}^{t} g(t-s)\sigma_s W(\mathrm{d}s) + \int_{-\infty}^{t} q(t-s)a_s \mathrm{d}s, \tag{6.2}$$

where g and q are non-negative deterministic kernels, $(a_t)_{t\in\mathbb{R}}$ and $(\sigma_t)_{t\in\mathbb{R}}$ are adapted càdlàg processes, and W is a two sided Brownian motion, i.e.

$$W_t := \begin{cases} B_t, & \text{if } t \geq 0 \\ -B'_{-t}, & \text{if } t < 0, \end{cases}$$

where B is a Brownian motion defined on $\mathbb{R}_{\geq 0}$ and B' is an independent copy of B defined on $\mathbb{R}_+$. We should note that a Brownian semi-stationary process of the form (6.2) is not necessary to be a semimartingale. For example, let $X = (X_t)_{t\in\mathbb{R}}$ be a centered Gaussian process of the form

$$X_t = \int_{-\infty}^{t} g(t-s)\sigma_s W(\mathrm{d}s), \tag{6.3}$$

where the gamma kernel $g(x) = x^\alpha \exp^{-\lambda x}$ for $\alpha \in (-\frac{1}{2}, 0) \cup (0, \frac{1}{2})$ and $\lambda > 0$. Basse [27] show that $g(0+) < \infty$ and $g' \in L^2(\mathbb{R}_+)$ are necessary conditions for being a semimartingale of X. However, in this case $g' \notin L^2(\mathbb{R}_+)$ since g' is not square integrable near 0. Hence if $\sigma \neq 0$, the process X is no longer a semimartingale.

To ensure that the first integral appearing in (6.2) is well-defined, we further assume that

$$\int_{-\infty}^{t} g^2(t-s)\sigma_s^2 \mathrm{d}s < \infty \qquad \text{almost surely} \tag{6.4}$$

for all $t \in \mathbb{R}$ (see, [28, 90]). Note that when $(\sigma_t)_{t\in\mathbb{R}}$ is a square integrable stationary process and $g \in L^2(\mathbb{R}_{\geq 0})$, the above condition (6.4) holds.

A numerical scheme to simulate such Brownian semi-stationary processes in (6.2) is offered in [32, 48]. The authors have introduced a Fourier approximation method and discussed the strong approximation error (in the L^2 sense) of the numerical scheme for Lévy semi-stationary processes. We shall emphasize that in [32, 48] the volatility process $(\sigma_t)_{t\in\mathbb{R}}$ is assumed to be observed. In this work, we complement their study by analyzing the weak limit of the error process in the framework of Brownian semi-stationary processes, where the drift and the volatility processes need to be numerically simulated. This obviously gives a more precise assessment of the numerical error associated with the Fourier method.

The outline of this chapter is organized as follows. In Section 6.1, we describe the Fourier approximation scheme for Brownian semi-stationary processes and present the main results on strong approximation error derived from Benth et al. [32, 48]. Section 6.2 is devoted to our main result, a weak limit theorem associated with a slight modification of the Fourier method.

6.1 A Fourier approximation scheme

We now explain the Fourier approximation scheme based on the works of Benth et al. [32, 48]. First of all we mention that the presence of the drift process $(a_t)_{t\in\mathbb{R}}$ and μ in (6.2) will be essentially ignored in this section. We start with the following assumptions on kernels involved in the description (6.2):

Assumption (A):

(i) The kernel functions g and q have bounded support contained in $[0, \tau]$ for some $\tau > 0$.

(ii) $g, q \in C(\mathbb{R}_{\geq 0})$.

The Assumption (**A**) guarantees that $g \in L^2(\mathbb{R}_{\geq 0})$. Therefore, it satisfies (6.4). In some cases these conditions are rather restrictive and we will also take a look at the case when the kernel function g has a singularity. We will give remarks on them later.

A strategy for simulating a discrete trajectory $X_{t_0}, ..., X_{t_M}$ for some $t_0, ..., t_M \in \mathbb{R}$ by Fourier approximation scheme is to define an even function on $\mathbb{R}$ from the original kernel function g. For any given $\lambda > 0$, we define

$$h(x) := g(|x|) \quad \text{and} \quad h_\lambda(x) := h(x)\exp(\lambda|x|). \tag{6.5}$$

Notice that $g = h$ on $[0, \tau]$ so

$$X_t = \int_{t-\tau}^{t} h(t-s)\sigma_s W(\mathrm{d}s).$$

We introduce the Fourier transform of h_λ via

$$\widehat{h}_\lambda(y) := \int_{\mathbb{R}} h_\lambda(x)\exp(-ixy)\mathrm{d}x. \tag{6.6}$$

Since in general $f \in L^1(\mathbb{R})$ does not imply $\widehat{f} \in L^1(\mathbb{R})$ for example when $f(x) = \mathbb{1}_{[0,1]}(x)$, if we aim to rewrite the function h in the form of Fourier transform, we need to add some conditions on h. Here, we assume that $\widehat{h}_\lambda \in L^1(\mathbb{R})$, the inverse Fourier transform exists and we obtain the identity

$$h(x) = \frac{\exp(-\lambda|x|)}{2\pi} \int_{\mathbb{R}} \widehat{h}_\lambda(y) \exp(ixy) \mathrm{d}y. \tag{6.7}$$

Since the Fourier transform maps $L^1(\mathbb{R})$ functions into the space of continuous functions, we require that $h \in C(\mathbb{R})$. This fact explains the Assumption (**A**)(ii) for the kernel function g. Since h is an even function, the Fourier transform, $\widehat{h}_\lambda$ is also even and for a given number $N \in \mathbb{N}$, we deduce an approximation of h by cutting the tail of the integral off:

$$h(x) = \frac{\exp(-\lambda|x|)}{2\pi} \int_{\mathbb{R}} \widehat{h}_\lambda(y) \cos(yx) \mathrm{d}y \tag{6.8}$$

$$\approx h_N(x) := \exp(-\lambda|x|) \left(\frac{b_0}{2} + \sum_{k=1}^{N} b_k \cos(\frac{k\pi x}{\tau}) \right) \tag{6.9}$$

with

$$b_k = \frac{\widehat{h}_\lambda(k\pi/\tau)}{\tau}. \tag{6.10}$$

Obviously, the above approximation is an L^2-projection onto the linear subspace generated by orthogonal functions $\{\cos(k\pi x/\tau), \sin(k\pi x/\tau)\}_{k=0}^N$, hence we deal with a classical Fourier expansion of the function h (recall that the function h is even by definition, thus the sinus terms do not appear at (6.8)). Now, the basic idea of the numerical approximation method proposed in [32, 48] is based upon the following relationship:

$$\begin{aligned}\int_u^t g(t-s)\sigma_s W(\mathrm{d}s) &\approx \int_u^t h_N(t-s)\sigma_s W(\mathrm{d}s) \\ &= \int_u^t \exp(-\lambda(t-s)) \left\{ \frac{b_0}{2} + \sum_{k=1}^{N} b_k \cos(\frac{k\pi(t-s)}{\tau}) \right\} \sigma_s W(\mathrm{d}s) \\ &= \frac{b_0}{2} \widehat{X}_{\lambda,u}(t,0) + \mathrm{Re} \sum_{k=1}^{N} b_k \widehat{X}_{\lambda,u}(t, \frac{k\pi}{\tau}), \qquad (6.11)\end{aligned}$$

where the complex valued stochastic field $\widehat{X}_{\lambda,u}(t,y)$ is defined via

$$\widehat{X}_{\lambda,u}(t,y) := \int_u^t \exp\{(-\lambda + iy)(t-s)\}\sigma_s W(\mathrm{d}s) \tag{6.12}$$

and $u \in [t-\tau, t]$. In a second step, for a $\delta > 0$ small, we observe the approximation

$$\begin{aligned}\widehat{X}_{\lambda,u}(t+\delta, y) &= \int_u^{t+\delta} \exp\{(-\lambda+iy)(t+\delta-s)\}\sigma_s W(\mathrm{d}s) \qquad (6.13)\\ &= \exp\{(-\lambda+iy)\delta\}\left(\widehat{X}_{\lambda,u}(t,y) + \int_t^{t+\delta} \exp\{(-\lambda+iy)(t-s)\}\sigma_s W(\mathrm{d}s)\right)\\ &\approx \exp\{(-\lambda+iy)\delta\}\left(\widehat{X}_{\lambda,u}(t,y) + \sigma_t(W_{t+\delta} - W_t)\right). \qquad (6.14)\end{aligned}$$

Hence, we obtain a simple iterative scheme for simulating the stochastic field $\widehat{X}_{\lambda,u}(t,y)$ in the variable t. Recall that we assume the drift process a is zero and we wish to simulate the trajectory of $X_{t_0}, \ldots, X_{t_M}$ given the information available at time t_0. It is important to understand the meaning of knowing the information about the involved processes up to time t_0. When the stochastic model for the process $(\sigma_t)_{t\in\mathbb{R}}$ is uncoupled with $(X_t)_{t\in\mathbb{R}}$, then we may use $u = t - \tau$ at (6.11). Indeed, in typical applications such as turbulence and finance this is the case: $(\sigma_t)_{t\in\mathbb{R}}$ is usually modeled via a jump diffusion process driven by a Lévy process, which might be correlated with the Brownian motion W. However, when the process $(X_t)_{t\in\mathbb{R}}$ is itself of a diffusion type, i.e.

$$X_t = \mu + \int_{t-\tau}^t g(t-s)\sigma(X_s)W(\mathrm{d}s) + \int_{t-\tau}^t q(t-s)a(X_s)\mathrm{d}s$$

it is in general impossible to simulate a trajectory of $(X_t)_{t\in\mathbb{R}}$, since for each value t the knowledge of the path $(X_u)_{u\in(t-\tau,t)}$ is required to compute X_t. But, in case we do know the historical path, say, $(X_u)_{u\in[-\tau,0]}$, the simulation of values X_t, $t \geq 0$, becomes possible.

Thus, the numerical simulation procedure is as follows:

(a) Simulate the independent increments $W_{t_i} - W_{t_{i-1}} \sim N(0, t_i - t_{i-1})$ for $i = 1, \ldots, M$.

(b) For each $i = 1, \ldots, M$ and $k = 0, \ldots, N$, simulate $\widehat{X}_{\lambda,u}(t_i, k\pi/\tau)$ from $\widehat{X}_{\lambda,u}(t_{i-1}, k\pi/\tau)$, $W_{t_i} - W_{t_{i-1}}$ and $\sigma_{t_{i-1}}$ by using (6.13).

(c) Simulate X_{t_i} applying steps (a), (b) and (6.11) (with $u = t_0$).

This procedure provides two approximation errors:

(i) The error in N scale from approximation function h by cutting off the tail of the integral of Fourier transform.

(ii) The error in M scale from the discretization error obtained at (6.13).

The main advantage of the numerical scheme described above is that it separates the simulation of the stochastic ingredients (σ and W) and the approximation of the deterministic kernel g (or h). In other words, the stochastic field $\widehat{X}_{\lambda,u}(t,y)$ is simulated via a simple recursive scheme without using the knowledge of g, while the kernel g is approximated via the Fourier transform at (6.8). This is in contrast to a straightforward discretization scheme

$$\int_{t_0}^{t_j} g(t-s)\sigma_s W(\mathrm{d}s) \approx \sum_{i=1}^{j-1} g(t_j - t_i)\sigma_{t_i}(W_{t_{i+1}} - W_{t_i}).$$

This numerical property is useful when considering a whole family of kernel functions $(g_\theta)_{\theta\in\Theta}$, since for any resulting model $X_t(\theta)$ only one realization of the stochastic field $\widehat{X}_{\lambda,u}(t,y)$ needs to be simulated. This can be obviously useful for the simulation of parametric Brownian semi-stationary processes.

We now assess the strong approximation error (in L^2 sense) of the Fourier approximation scheme in both errors, i.e. the approximation of the deterministic kernel and the discretization error. We assume for the moment that the volatility process $(\sigma_t)_{t\in\mathbb{R}}$ is square integrable with bounded second moment. We start with the analysis of the error associated with the approximation of the deterministic kernel g by the function h_N. Then a straightforward computation (e.g., [32, Eq. (4.5)]) implies that

$$\mathbb{E}\left[\Big(\int_{t_0}^{t} \{g(t-s) - h_N(t-s)\}\sigma_s W(\mathrm{d}s)\Big)^2\right] \le C\frac{1-\exp\{-2\lambda(t-t_0)\}}{\lambda}\left(\sum_{k=N+1}^{\infty}|b_k|\right)^2, \tag{6.15}$$

where C is a positive constant and the Fourier coefficients b_k have been defined at (6.10). We remark that

$$\frac{1-\exp\{-2\lambda(t-t_0)\}}{\lambda} \to 2(t-t_0)$$

as $\lambda \to 0$, while

$$\frac{1-\exp\{-2\lambda(t-t_0)\}}{\lambda} \sim \lambda^{-1}$$

as $\lambda \to \infty$. Thus, it is preferable to choose the parameter $\lambda > 0$ large.

Example 6.1

A standard model gamma kernel function g in the context of turbulence (see, [43]) is given via

$$g(x) = x^\alpha \exp(-\bar{\lambda}x)$$

with $\bar{\lambda} > 0$ and $\alpha \in (-\frac{1}{2}, 0) \cup (0, \frac{1}{2})$. Obviously, this gamma kernel violates the statement of the Assumption (**A**) because it has a singularity at $x = 0$ for the values $\alpha \in (-1/2, 0)$ and has unbounded support. Furthermore, the $\mathcal{BSS}$ process with this gamma kernel function is not semimartingale. However, one can easily construct an approximating function g_ε^T which coincides with g on the interval $[\varepsilon, T]$ and satisfies the Assumption (**A**). More precisely, we define

$$g_\varepsilon^T = \begin{cases} g(\varepsilon), & \text{if } 0 \le x \le \varepsilon \\ g(x), & \text{if } \varepsilon \le x \le T \\ \phi(x), & \text{if } T \le x \le T' \\ 0, & \text{if } x > T', \end{cases}$$

where $\phi : [T, T'] \to \mathbb{R}$ is a continuous function such that $\phi(T) = g(T)$ and $\phi(T') = 0$ for some $T' \ge T$. Assuming again the boundedness of the second moment of the process $(\sigma_t)_{t \in \mathbb{R}}$, the approximation error is controlled via

$$\mathbb{E}\left[\Big(\int_{-\infty}^{t} \{g(t-s) - g_\varepsilon^T(t-s)\}\sigma_s W(\mathrm{d}s)\Big)^2\right] \le C\|g - g_\varepsilon^T\|^2_{L^2((0,\varepsilon)\cup(T,\infty))}.$$

Such error can be made arbitrary small by choosing ε small and T large. Clearly, this is a rather general approach, which is not particularly related to a given class of kernel functions g. In a second step one would apply the Fourier approximation method described above to the function g_ε^T. At this stage, it is important to note that the parameter $\lambda > 0$ introduced at (6.5) is naturally restricted through the condition $\lambda < \bar{\lambda}$; otherwise the kernel h_λ would have an explosive behaviour at ∞. Thus, the approximation error discussed at (6.15) can not be made arbitrarily small in λ.

Remark 6.2

The Fourier coefficients b_k can be further approximated under stronger conditions on the function h, which helps to obtain an explicit bound at (6.15). More specifically, when $h \in C^{2n}(\mathbb{R})$ and $h_\lambda^{(2j-1)}(\tau) = 0$ for all $j = 1, \ldots, n$, then it holds that

$$|b_k| \leq Ck^{-2n}.$$

This follows by a repeated application of integration by parts formula (see, [32, Proposition 4.1] for a detailed exposition). In fact, the original work [32] defines another type of smooth interpolation functions h, rather than the mere identity $h(x) = g(|x|)$, to achieve that the relationship $h_\lambda^{(2j-1)}(\tau) = 0$ holds for all $j = 1, \ldots, n$ and some $n \in \mathbb{N}$.

Now, let us turn our attention to the discretization error introduced at (6.13). We assume that $t_0 < \ldots < t_M$ is an equidistant grid with $t_i - t_{i-1} = \Delta t$. According to (6.13) the random variable

$$\eta_j(y) := \sum_{i=1}^{j} \exp\{(-\lambda + iy)(j + 1 - i)\Delta t\}\sigma_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) \tag{6.16}$$

is an approximation of $\widehat{X}_{\lambda,t_0}(t_j, y)$ for any $y \in \mathbb{R}$ whenever the drift process a is assumed to be absent. When $(\sigma_t)_{t\in\mathbb{R}}$ is a weak sense stationary process, a straightforward computation proves that

$$\mathbb{E}[|\widehat{X}_{\lambda,t_0}(t_j, y) - \eta_j(y)|^2] \leq C(t_j - t_0)\Big((\lambda^2 + y^2)(\Delta t)^2 + \mathbb{E}[|\sigma_{t_1} - \sigma_{t_0}|^2]\Big). \tag{6.17}$$

We refer to [32, Lemma 4.2] for a detailed proof.

Example 6.3

Assume that the process $(\sigma_t)_{t\in\mathbb{R}}$ is a continuous stationary Itô semimartingale, i.e.

$$\mathrm{d}\sigma_t = \widetilde{a}_t \mathrm{d}t + \widetilde{\sigma}_t \mathrm{d}B_t,$$

where B is a Brownian motion and $(\widetilde{a}_t)_{t\in\mathbb{R}}$, $(\widetilde{\sigma}_t)_{t\in\mathbb{R}}$ are stochastic processes with bounded second moment. Then the Itô isometry implies that

$$\mathbb{E}[|\sigma_{t_1} - \sigma_{t_0}|^2] \leq C\Delta t.$$

Hence, in this setting Δt becomes the dominating term in the approximation error (6.17).

Combining the estimates at (6.15) and (6.17), we obtain the strong approximation error of the proposed Fourier method, which is the main result of [32] (see, Propositions 4.1 and 4.3 therein).

Proposition 6.4

*Let $t_0 < \ldots < t_M$ be an equidistant grid with $t_i - t_{i-1} = \Delta t$. Assume that condition **(A)** holds and $(\sigma_t)_{t\in\mathbb{R}}$ is a weak sense stationary process. Then the L^2 approximation error associated with the Fourier type numerical scheme is given via*

$$\mathbb{E}\left[\Big|\int_{t_0}^{t_j} g(t_j - s)\sigma_s W(\mathrm{d}s) - \Big(\frac{b_0}{2}\eta_j(0) + \sum_{k=1}^{N} b_k \eta_j(\frac{k\pi}{\tau})\Big)\Big|^2\right] \tag{6.18}$$

$$\leq C\left(\frac{1-\exp\{-2\lambda(t-t_0)\}}{\lambda}\Big(\sum_{k=N+1}^{\infty}|b_k|\Big)^2\right.$$

$$+ (t_j - t_0)\left\{\lambda^2\Big(\frac{|b_0|}{2} + \sum_{k=1}^{N}|b_k|\Big)^2(\Delta t)^2 + (\frac{\pi}{\tau})^2\Big(\sum_{k=1}^{N}k|b_k|\Big)^2(\Delta t)^2\right.$$

$$\left.\left.+\Big(\frac{|b_0|}{2} + \sum_{k=1}^{N}|b_k|\Big)^2\mathbb{E}[|\sigma_{t_1} - \sigma_{t_0}|^2]\right\}\right)$$

for a positive constant C.

6.2 Main results

We consider a Brownian semi-stationary process in (6.2) and assume the condition (6.4). In the previous section, we discussed the strong approximation error (in the L^2 sense) of the numerical scheme for Brownian semi-stationary processes, where the volatility process $(\sigma_t)_{t\in\mathbb{R}}$ is assumed to be observed. In this section, we present the weak limit theory of the numerical scheme associated with the Fourier method proposed in [32, 48]. We complement their study by analyzing the weak limit of the error process in the framework of Brownian semi-stationary processes, where the drift and the volatility processes need to be numerically simulated. This obviously gives a more precise assessment of the numerical error associated with the Fourier method. We shall emphasize that the Fourier approximation scheme investigated in [32, 48] basically ignored the need of simulating the volatility process $(\sigma_t)_{t\in\mathbb{R}}$ in practical applications (the same holds for the drift process $(a_t)_{t\in\mathbb{R}}$).

We again wish to simulate the path $(X_t)_{t\in[t_0,T]}$ for a given terminal time $T > t_0$. We consider t_0 as the starting point, as in the previous section, we fix a time t_0 and assume the knowledge of all processes involved up to that time. Here we propose a numerical scheme for simulating the path $(X_t)_{t\in[t_0,T]}$ which is a slightly modified version of the original Fourier approach. We recall the imposed Assumption (**A**), in particular, the weight functions g and q are assumed to have bounded support contained in $[0,\tau]$. First of all, we further assume some condition on the stochastic process $(a_t,\sigma_t)_{t\in[t_0,T]}$.

Assumption (B):

There exist càdlàg estimators $(a_t^M,\sigma_t^M)_{t\in[t_0,T]}$ of the stochastic process $(a_t,\sigma_t)_{t\in[t_0,T]}$ and the convergence rate $\nu_M \to \infty$ as $M \to \infty$ such that the following functional stable convergence holds:

$$\nu_M\left(a^M - a, \sigma^M - \sigma\right) \xrightarrow{d_{st}} U = (U^1, U^2) \qquad \text{on } D^2([t_0,T]), \tag{6.19}$$

where the convergence is on the space of bivariate càdlàg functions defined on $[t_0,T]$ equipped with the Skorohod topology $D^2([t_0,T])$.

In the following, we will deal with the space of càdlàg processes equipped with the Skorohod topology or with the space of continuous processes equipped with the uniform topology. We remark that the estimators $(a_t^M)_{t\in[t_0,T]}$ and $(\sigma_t^M)_{t\in[t_0,T]}$ might have a different *effective* convergence rate. In this case, we will have either $U_1 \equiv 0$ or $U_2 \equiv 0$.

Before we go to the next step, let us present some examples of convergence at (6.19) to highlight the most prominent results. For simplicity we assume that $a \equiv 0$ in all cases.

Example 6.5

Let us consider a continuous diffusion model for the volatility process σ, i.e.

$$d\sigma_t = \widetilde{a}(\sigma_t)dt + \widetilde{v}(\sigma_t)dB_t, \qquad \sigma_{t_0} = x_0,$$

where B is a Brownian motion possibly correlated with W. We consider an equidistant partition $t_0 = s_0 < s_1 < \ldots < s_M = T$ of the interval $[t_0,T]$ and define the continuous

Euler approximation of σ_t via

$$\sigma_t^M = \sigma_{s_k}^M + \widetilde{a}(\sigma_{s_k}^M)(t - s_k) + \widetilde{v}(\sigma_{s_k}^M)(B_t - B_{s_k}), \qquad t \in [s_k, s_{k+1}].$$

When the functions $\widetilde{a}$ and $\widetilde{v}$ are assumed to be globally Lipschitz and continuously differentiable, it holds that

$$\sqrt{M}(\sigma^M - \sigma) \xrightarrow{d_{st}} U^2 \qquad \text{on } C([t_0, T]),$$

where U^2 is the unique solution of the stochastic differential equation

$$\mathrm{d}U_t^2 = \widetilde{a}'(\sigma_t)U_t^2 \mathrm{d}t + \widetilde{v}'(\sigma_t)U_t^2 \mathrm{d}B_t - \frac{1}{\sqrt{2}}\widetilde{v}\widetilde{v}'(\sigma_t)\mathrm{d}W_t',$$

where W' is a new Brownian motion independent of $\mathcal{F}$. We refer to [63, Theorem 1.2] for a detailed treatment of this result.

Example 6.6

Let us now consider a discontinuous diffusion model for the volatility process σ, i.e.

$$\mathrm{d}\sigma_t = \widetilde{v}(\sigma_{t-})\mathrm{d}L_t, \qquad \sigma_{t_0} = x_0,$$

where L is a purely discontinuous Lévy process. In this framework we study the discretized Euler scheme given via

$$\sigma_{s_{k+1}}^M = \widetilde{v}(\sigma_{s_k}^M)(L_{s_{k+1}} - L_{s_k}), \qquad k = 0, \ldots, M-1.$$

We define the process $U_t^M = \sigma_{[tM]/M}^M - \sigma_{[tM]/M}$. In [61], several classes of Lévy processes L has been studied. For the sake of exposition, we demonstrate the case of a symmetric β-stable Lévy process L with $\beta \in (0, 2)$. Let us assume that $\widetilde{v} \in C^3(\mathbb{R})$. Then, it holds that

$$(M/\log(M))^{1/\beta} U^M \xrightarrow{d_{st}} U^2 \qquad \text{on } D([t_0, T]),$$

where U^2 is the unique solution of the linear equation

$$\mathrm{d}U_t^2 = \widetilde{v}'(\sigma_{t-})U_{t-}^2 \mathrm{d}L_t - \widetilde{v}\widetilde{v}'(\sigma_{t-})\mathrm{d}L_t'$$

and L' is another symmetric β-stable Lévy process (with certain scaling parameter) independent of $\mathcal{F}$. We note that this result does not directly correspond to our condition (6.19) as the discretized process $\sigma_{[tM]/M}$ is used in the definition of U_t^M.

Now, we propose our numerical scheme by following the Fourier type approach from the previous section. We refer to (6.5) the definition of h, h_λ and to (6.6) the definition of the function $\widehat{h}_\lambda$. We replace the Fourier transform approximation proposed at (6.8) by a Riemann sum approximation evaluating the integrand by its left end points of a partition. More specifically, for each fixed N and an equidistant partition, since h is an even function, we introduce the approximation

$$\begin{aligned} h(x) &= \frac{\exp(-\lambda|x|)}{2\pi} \int_{\mathbb{R}} \widehat{h}_\lambda(y) \exp(ixy) \mathrm{d}y \\ &= \frac{\exp(-\lambda|x|)}{2\pi} \int_{\mathbb{R}} \widehat{h}_\lambda(y) \cos(xy) \mathrm{d}y \\ &\approx \widetilde{h}_N(x) := \frac{\exp(-\lambda|x|)}{\pi N} \sum_{k=0}^{c_N} \widehat{h}_\lambda\left(\frac{k}{N}\right) \cos\left(\frac{kx}{N}\right), \end{aligned} \tag{6.20}$$

where c_N is a sequence of numbers in $\mathbb{N}$ satisfying $c_N/N \to \infty$ as $N \to \infty$. The above approximation obviously gives two types of error. The first one comes from the Riemann sum approximation while the other one comes from the tail approximation. Intuitively the former error should dominate the latter. In the following we will also assume that the sequence c_N additionally satisfies the condition

$$N \int_{c_N/N}^{\infty} |\widehat{h}_\lambda(y)| \mathrm{d}y \to 0 \qquad \textit{as } N \to \infty. \tag{6.21}$$

Clearly, such a sequence exists, since $\widehat{h}_\lambda \in L^1(\mathbb{R})$. Condition (6.21) guarantees that the Riemann sum approximation error will dominate.

Remark 6.7

Under some stronger conditions the tail integral at (6.21) can be bounded from above explicitly. Assume that $h \in C^2([-\tau, \tau])$ and has bounded support on $[-\tau, \tau]$. Then a repeated application of integration by parts formula implies

$$\begin{aligned}
\widehat{h}_\lambda(y) &= \int_{\mathbb{R}} h_\lambda(x)\cos(yx)\mathrm{d}x \\
&= \frac{1}{y}\int_{-\tau}^{\tau} h_\lambda(x)\cos(yx)\mathrm{d}\sin(xy) \\
&= \frac{2}{y}h_\lambda(\tau)\sin(\tau y) + \frac{2}{y^2}h'_\lambda(\tau)\cos(\tau y) - \frac{1}{y^2}\int_{-\tau}^{\tau} h''_\lambda(x)\cos(yx)\mathrm{d}x \\
&= -\frac{1}{y^2}\int_{-\tau}^{\tau} h''_\lambda(x)\cos(yx)\mathrm{d}x
\end{aligned}$$

for any $y > 0$ where the last equation comes from the fact that h is continuous and has bounded support on $[-\tau, \tau]$. Thus, for any $u > 0$, we deduce the inequality

$$\begin{aligned}
\int_u^\infty |\widehat{h}_\lambda(y)|\mathrm{d}y &\le C\|h''\|_{L^1}\int_u^\infty y^{-2}\mathrm{d}y \\
&\le C\|h''\|_{L^1}u^{-1}.
\end{aligned}$$

Hence, condition (6.21) holds whenever $N^2/c_N \to 0$ as $N \to \infty$.

Remark 6.8

We remark that the Fourier transform used at (6.8) comes from the L^2 theory. Thus, in contrast to the L^2-distance $\|h - h_N\|_{L^2}$, the limiting behaviour of a standardized version of $h(x) - h_N(x)$ is difficult to study pointwise. This is precisely the reason why we use the Riemann sum approximation instead, for which we will show the convergence of $N(h(x) - \widetilde{h}_N(x))$.

If one can freely choose the simulation rates N and M, the Fourier transform of (6.8) is numerically more preferable. According to the estimate (6.15) and the upper bound for the Fourier coefficient of Remark 6.2 applied for $n = 1$, we readily deduce the rate N^{-1} for the L^2-error approximation connected to (6.8). On the other hand, the effective sample size of the Riemann approximation at (6.20) is c_N. In the setting of the previous remark, the overall Riemann approximation error is $\max(N^{-1}, N/c_N)$. Recalling that $c_N/N \to \infty$, the obtained rate is definitely slower than the one associated with Fourier approximation proposed at (6.8).

Nevertheless, as our aim is to precisely determine the asymptotics associated with the N scale, we will discuss the Riemann approximation approach in the sequel. A statement about the Fourier transform (6.8) will be presented in Remark 6.15.

Recall that

$$\widehat{X}_{\lambda,u}(t,y) = \int_u^t \exp\{(-\lambda + iy)(t-s)\}\sigma_s W(\mathrm{d}s)$$

We now essentially proceed as in the steps (6.11)–(6.13). The first step, it holds that

$$\begin{aligned}\int_u^t g(t-s)\sigma_s W(\mathrm{d}s) &\approx \int_u^t \widetilde{h}_N(t-s)\sigma_s W(\mathrm{d}s)\\ &= \int_u^t \exp(-\lambda(t-s)) \left\{\sum_{k=0}^{c_N} \widetilde{b}_k \cos(\frac{k(t-s)}{N})\right\} \sigma_s W(\mathrm{d}s)\\ &= \mathrm{Re} \sum_{k=0}^{c_N} \widetilde{b}_k \widehat{X}_{\lambda,u}(t, \frac{k}{N}), \end{aligned} \tag{6.22}$$

where $\widetilde{b}_k = \widehat{h}_\lambda(k/N)/(\pi N)$. In the second step, for $\delta > 0$, we obtain the approximation

$$\begin{aligned}\widehat{X}_{\lambda,u}(t+\delta,y) &= \exp\{(-\lambda+iy)\delta\}\left(\widehat{X}_{\lambda,u}(t,y) + \int_t^{t+\delta} \exp\{(-\lambda+iy)(t-s)\}\sigma_s W(\mathrm{d}s)\right)\\ &\approx \exp\{(-\lambda+iy)\delta\}\left(\widehat{X}_{\lambda,u}(t,y) + \int_t^{t+\delta} \exp\{(-\lambda+iy)(t-s)\}\sigma_s^M W(\mathrm{d}s)\right). \end{aligned} \tag{6.23}$$

When the estimator σ^M is assumed to be constant on intervals $[s_{i-1}, s_i)$, $i = 1, \ldots, M$, the last integral at (6.23) can be easily simulated. We remark that this approximation procedure slightly differs from (6.13) as now we leave the exponential term unchanged.

In summary, given that the information up to time t_0 is available, we arrive at the simulated value

$$X_t^{N,M} := \int_{t_0}^{t} \widetilde{h}_N(t-s)\sigma_s^M W(\mathrm{d}s) + \int_{t_0}^{t} q(t-s)a_s^M \mathrm{d}s \tag{6.24}$$

of the random variable

$$X_t^0 = \int_{t_0}^{t} g(t-s)\sigma_s W(\mathrm{d}s) + \int_{t_0}^{t} q(t-s)a_s \mathrm{d}s. \tag{6.25}$$

Note that the drift part of the Brownian semi-stationary process X is estimated in a direct manner, although other methods similar to the treatment of the Brownian part are possible. Now, we wish to study the asymptotic theory for the approximation error $X_t^{N,M} - X_t^0$. Our first result analyzes the limiting behaviour of the function $N(h(x) - \widetilde{h}_N(x))$.

Lemma 6.9

Define the function $\psi_N(x) := N(h(x) - \widetilde{h}_N(x))$*. Let us assume that the condition*

$$\widehat{xh_\lambda(x)} \in L^1(\mathbb{R}), \qquad \widehat{x^2h_\lambda(x)} \in L^1(\mathbb{R}) \tag{6.26}$$

*holds. Then, under Assumption (**A**),* (6.21) *and* (6.26)*, it holds that*

$$\psi_N(x) \to \psi(x) = -\frac{\widehat{h}_\lambda(0)}{2\pi}\exp(-\lambda|x|) \qquad \textit{as } N \to \infty \tag{6.27}$$

for any $x \in \mathbb{R}$*. Furthermore, it holds that*

$$\sup_{N\in\mathbb{N},\, x\in[0,T]} |\psi_N(x)| \le C$$

for any $T > 0$*.*

Proof. See Section 6.3. □

Remark 6.10

If our function h_λ in Lemma 6.9 is a Schwartz function, then it satisfies the condition (6.26). Note that a Schwartz function is a smooth function whose derivatives decay at infinity faster than any power. More precisely, $f : \mathbb{R} \to \mathbb{C}$ is a Schwartz function if and only if f is infinitely differentiable and for all integers $m, n \geq 0$

$$\lim_{x \to \pm\infty} x^m f^n(x) = 0.$$

Remark 6.11

We can also apply Lemma 6.9 to a kernel function with singularity. To see this, we define the kernel function g by

$$g(x) = x^\alpha f(x)$$

where $\alpha \in (-\frac{1}{2}, 0)$ and function $f \in C^1(\mathbb{R}_{\geq 0})$ has bounded support contained in $[0, \tau]$ for some $\tau > 0$ satisfying $f(0) \neq 0$. It is clear that the kernel g has a singularity at $x = 0$. To make Lemma 6.9 applicable to the kernel function g, we first amend it in a neighborhood close to zero by constructing an approximating function

$$g^\varepsilon = \begin{cases} g(\varepsilon), & \text{if } 0 \leq x \leq \varepsilon \\ g(x), & \text{if } x \geq \varepsilon, \end{cases}$$

for some $\varepsilon > 0$. Suppose that the conditions (6.21) and (6.26) hold for corresponding terms of g^ε. Using Lemma 6.9, we have

$$N(h^\varepsilon - h^\varepsilon_N) \to -\frac{\widehat{h}^\varepsilon_\lambda(0)}{2\pi} \exp(-\lambda|x|),$$

as $N \to \infty$ where $h^\varepsilon(.)$, $h^\varepsilon_N(.)$ and $h^\varepsilon_\lambda(0)$ are defined relatively to g^ε as in (6.5). Finally, by helping from the monotone convergence theorem, we can extract the asymptotic behavior of the error when ε converges to zero . That is,

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} N(h^\varepsilon - h^\varepsilon_N) = \frac{\widehat{h}_\lambda(0)}{2\pi} \exp(-\lambda|x|).$$

In order to prepare the main result, we need a further condition on the kernel function g to prove tightness later.

Assumption (C):

(i) The kernel function g has the form

$$g(x) = x^\alpha f(x)$$

for some $\alpha \geq 0$ and function f satisfying $f(0) \neq 0$.

(ii) $f \in C^1(\mathbb{R}_{\geq 0})$ has bounded support contained in $[0, \tau]$.

Notice that the assumption $\alpha \geq 0$ is in accordance with the Assumption (**A**)(ii). Assumption (**C**) implies the following approximation result:

$$\int_0^1 |g(x+\delta) - g(x)|^4 \mathrm{d}x \leq \begin{cases} C\delta^4 & \alpha = 0 \\ C\delta^{\min(4,4\alpha+1)} & \alpha > 0 \end{cases} \tag{6.28}$$

for $\delta \in [0, T]$. The case $\alpha = 0$ is trivial, while the other one follows along the lines of the proof of [51, Lemma 4.1]. As a matter of fact, we also require a good estimate of the left side of (6.28) when the kernel g is replaced by the function ψ_N defined in Lemma 6.9. In the following, we will assume that

$$\sup_{N \in \mathbb{N}} \int_0^1 |\psi_N(x+\delta) - \psi_N(x)|^4 \mathrm{d}x \leq C\delta^{1+\varepsilon} \tag{6.29}$$

for some $\varepsilon > 0$ and $\delta \in [0, T]$.

Remark 6.12

Obviously, as in the case of function g, condition (6.29) would hold if

$$\psi_N(x) = x^\alpha f_N(x),$$

where $f_N \in C^1(\mathbb{R}_{\geq 0})$ with uniformly bounded derivative in $N \in \mathbb{N}$ and x in a compact interval. We can prove condition (6.29) explicitly when the function g is differentiable. Assume that $y\widehat{h}_\lambda(y), y\widehat{h}'_\lambda(y) \in L^1(\mathbb{R}_{\geq 0})$ and c_N is chosen in such a way that the condition

$$N \int_{c_N/N}^\infty |y\widehat{h}_\lambda(y)| \mathrm{d}y \to 0 \qquad \textit{as } N \to \infty$$

is satisfied. As in the proof of Lemma 6.9, we conclude that

$$|\partial_x \partial_y \kappa_x(y)| \leq (|\widehat{h}_\lambda(y)| + |yx\widehat{h}_\lambda(y)| + |y\widehat{h}'_\lambda(y)|)$$

where $\kappa_x(y) = \widehat{h}_\lambda(y)\cos(yx)$ and we deduce that

$$\sup_{x\in[0,T]} |\psi'_N(x)| \leq C\Bigg(N\int_{c_N/N}^{\infty} |y\widehat{h}_\lambda(y)|\mathrm{d}y + N\int_{c_N/N}^{\infty} |\widehat{h}_\lambda(y)|\mathrm{d}y$$
$$+ \sum_{k=0}^{c_N}\int_{k/N}^{(k+1)/N} |\partial_y\kappa_x(\zeta_{k,N}(y))|\mathrm{d}y + \sum_{k=0}^{c_N}\int_{k/N}^{(k+1)/N} |\partial_x\partial_y\kappa_x(\widetilde{\zeta}_{k,N}(y))|\mathrm{d}y\Bigg)$$

for certain values $\zeta_{k,N}(y), \widetilde{\zeta}_{k,N}(y)$ in the interval $(k/N, y)$. Then, due to our integrability conditions, we obtain

$$\sup_{N\in\mathbb{N},\ x\in[0,T]} |\psi'_N(x)| < \infty.$$

Moreover, condition (6.29) is trivially satisfied due to mean value theorem. However, showing (6.29) under Assumption (**C**) seems to be a much harder problem for $\alpha \in (0,1)$.

The next result is the main theorem of this chapter.

Theorem 6.13

*Assume that Assumptions (**A**), (**B**), (**C**),* (6.21)*,* (6.26) *and* (6.29) *hold, and the processes* $(\sigma_t)_{t\in[t_0,T]}$ *and* $(\sigma_t^M)_{t\in[t_0,T]}$ *has finite fourth moment with*

$$\sup_{t\in[t_0,T]} \mathbb{E}[\sigma_t^4] < \infty \quad \textit{and} \quad \sup_{t\in[t_0,T]}\sup_{M\in\mathbb{N}} \mathbb{E}[(\sigma_t^M)^4] < \infty.$$

We also assume that the process $U_t^M = \nu_M(\sigma_t^M - \sigma_t)$ *satisfies*

$$\sup_{t\in[t_0,T]}\sup_{M\in\mathbb{N}} \mathbb{E}[(U_t^M)^4] < \infty. \tag{6.30}$$

Then, we obtain the decomposition

$$X_t^{N,M} - X_t^0 = A_t^{N,M} + B_t^M$$

such that

$$NA^{N,M} \overset{ucp}{\Longrightarrow} A = \frac{\widehat{h}_\lambda(0)}{2\pi} \int_{t_0}^{\cdot} \exp(-\lambda(\cdot - s))\sigma_s W(\mathrm{d}s) \qquad \text{as } N, M \to \infty, \tag{6.31}$$

and

$$\nu_M B^M \xrightarrow{d_{st}} B = \int_{t_0}^{\cdot} g(\cdot - s) U_s^2 W(\mathrm{d}s) + \int_{t_0}^{\cdot} q(\cdot - s) U_s^1 \mathrm{d}s \qquad \text{as } M \to \infty, \tag{6.32}$$

where the stable convergence holds on the space $C([t_0, T])$ equipped with the uniform topology.

Proof. See Section 6.3 □

Remark 6.14

We remark that the stronger conditions (**C**) and (6.29) are not required to prove the finite dimensional version of convergence (6.31) and (6.32).

Theorem 6.13 immediately applies to the weak approximation error analysis. Assume for simplicity that $M = M(N)$ is chosen such that $\nu_M/N \to 1$, so that the Riemann sum approximation error and the simulation error from (6.19) are balanced. We consider a bounded test function $\varphi \in C^1(\mathbb{R})$ with bounded derivative. The mean value theorem implies the identity

$$\varphi(X_t^{N,M}) - \varphi(X_t^0) = \varphi'(\xi_{N,M})(X_t^{N,M} - X_t^0),$$

where $\xi_{N,M}$ is a random value between X_t^0 and $X_t^{N,M}$ with $\xi_{N,M} \xrightarrow{\mathbb{P}} X_t^0$ as $N \to \infty$. By properties of stable convergence, we deduce that $(\xi_{N,M}, N(X_t^{N,M} - X_t^0)) \xrightarrow{d_{st}} (X_t^0, A_t + B_t)$. Hence, given the existence of the involved expectations, we conclude that

$$\mathbb{E}[\varphi(X_t^{N,M})] - \mathbb{E}[\varphi(X_t^0)] = N^{-1}\mathbb{E}'[\varphi'(X_t^0)(A_t + B_t)] + o(N^{-1}). \tag{6.33}$$

We recall that the limit $A_t + B_t$ is defined on the extended probability space $(\Omega', \mathcal{F}', \mathbb{P}')$.

Remark 6.15

The results of Theorem 6.13 may also apply to the original Fourier approximated method proposed in [32, 48]. Let us keep the notation of this section and still denote the approximated value of X_t^0 by $X_t^{N,M}$. Recalling the result of (6.15) (see also Remark 6.2) and assuming that $M = M(N)$ is chosen such that $\sum_{k=N+1}^{\infty} |b_k| \ll \nu_M$, we readily deduce that

$$\nu_M(X_t^{N,M} - X_t^0) \xrightarrow{d_{st}} B_t.$$

Remark 6.16

The results of Theorem 6.13 might transfer to the case of Lévy semi-stationary processes

$$X_t = \mu + \int_{-\infty}^{t} g(t-s)\sigma_s L(\mathrm{d}s) + \int_{-\infty}^{t} q(t-s)a_s \mathrm{d}s$$

under suitable moment assumptions on the driving Lévy motion L (cf. [32]). However, when L is e.g. a β-stable process with $\beta \in (0,2)$, it seems to be much harder to access the weak limit of the approximation error.

6.3 Proofs

6.3.1 Proof of Lemma 6.9

Proof. We recall that

$$\kappa_x(y) = \widehat{h}_\lambda(y)\cos(yx),$$

and denote by $\kappa_x'(y)$ the derivative of $\kappa_x(y)$ with respect to y. The derivative $\kappa_x'(\cdot)$ and $\kappa_x''(\cdot)$ exist because of $xh_\lambda(x), x^2 h_\lambda(x) \in L^1(\mathbb{R})$. Thus,

$$\begin{aligned}\kappa_x'(y) &= -x\widehat{h}_\lambda(y)\sin(xy) + \cos(xy)\widehat{h}_\lambda'(y) \quad \text{and} \\ \kappa_x''(y) &= (-x\widehat{h}_\lambda(y)\sin(xy))' + (\cos(xy)\widehat{h}_\lambda'(y))'.\end{aligned}$$

We recall a well known result from Fourier analysis (e.g., [49, Theorem 8.22]): The condition (6.26) implies that

$$\widehat{h}'_\lambda \in L^1(\mathbb{R}), \qquad \widehat{h}''_\lambda \in L^1(\mathbb{R}). \tag{6.34}$$

Now, observe the decomposition

$$\begin{aligned}
\psi_N(x) &= \frac{N\exp(-\lambda|x|)}{\pi}\sum_{k=0}^{c_N}\int_{k/N}^{(k+1)/N}\left(\kappa_x(y)-\kappa_x(\frac{k}{N})\right)\mathrm{d}y \\
&+ \frac{N\exp(-\lambda|x|)}{\pi}\int_{(c_N+1)/N}^{\infty}\kappa_x(y)\mathrm{d}y \\
&= \frac{N\exp(-\lambda|x|)}{\pi}\sum_{k=0}^{c_N}\int_{k/N}^{(k+1)/N}\left(\kappa_x(y)-\kappa_x(\frac{k}{N})\right)\mathrm{d}y + o(1).
\end{aligned}$$

The approximation follows by the inequality $|\kappa_x(y)| \leq |\widehat{h}_\lambda(y)|$ and condition (6.21). Notice from the above that $\kappa'_x(\cdot), \kappa''_x(\cdot) \in L^1(\mathbb{R}_{\geq 0})$ because of the condition (6.34), we deduce that

$$\begin{aligned}
\psi_N(x) &= \frac{N\exp(-\lambda|x|)}{\pi}\sum_{k=0}^{c_N}\int_{k/N}^{(k+1)/N}\kappa'_x(\frac{k}{N})\left(y-\frac{k}{N}\right)\mathrm{d}y + o(1) \\
&= \frac{\exp(-\lambda|x|)}{2\pi N}\sum_{k=0}^{c_N}\kappa'_x(\frac{k}{N}) + o(1) \\
&\to \frac{\exp(-\lambda|x|)}{2\pi}\int_0^\infty \kappa'_x(y)\mathrm{d}y \qquad \text{as } N\to\infty.
\end{aligned}$$

Next, we are going to show that

$$\int_0^\infty \kappa'_x(y)\mathrm{d}y = -\widehat{h}_\lambda(0). \tag{6.35}$$

Let $b > 0$, the fundamental theorem of calculus implies that

$$\begin{aligned}
\int_0^b \kappa'_x(y)\mathrm{d}y &= \kappa_x(b) - \kappa_x(0) \\
&= \widehat{h}_\lambda(b)\cos(xb) - \widehat{h}_\lambda(0).
\end{aligned}$$

Since $h_\lambda \in L^1(\mathbb{R}_{\geq 0})$, $\widehat{h}_\lambda$ vanishes at infinity. Then, taking $b \to \infty$ yields (6.35). In order to prove the second assertion of the lemma, we observe the inequality

$$|\psi_N(x)| \leq \frac{\exp(-\lambda|x|)}{\pi}\Big(\sum_{k=0}^{c_N}\int_{k/N}^{(k+1)/N} |\kappa_x'(\zeta_{k,N}(y))|\mathrm{d}y + N\int_{c_N/N}^{\infty} |\widehat{h}_\lambda(y)|\mathrm{d}y\Big),$$

where $\zeta_{k,N}(y)$ is a certain value with $\zeta_{k,N}(y) \in (k/N, y)$. Clearly, the second term in the above approximation is bounded in N since it converges to 0. On the other hand, we have that $|\kappa_x'(y)| \leq |x||\widehat{h}_\lambda(y)| + |\widehat{h}_\lambda'(y)|$, and since $\widehat{h}_\lambda, \widehat{h}_\lambda' \in L^1(\mathbb{R}_{\geq 0})$, we readily deduce that

$$\sup_{N\in\mathbb{N},\ x\in[0,T]} |\psi_N(x)| \leq C.$$

This completes the proof of the lemma. □

6.3.2 Proof of Theorem 6.13

Proof. We start by proving the stable convergence in (6.32). Let us first recall a classical result about weak convergence of semimartingales (see, [65, Theorem VI.6.22] or [72]):

Let $(Y_s^n)_{s\in[t_0,T]}$ be a sequence of càdlàg processes such that $Y^n \xrightarrow{d_{st}} Y$ on $D([t_0, T])$ equipped with the Skorohod topology. Then we obtain the weak convergence

$$\int_{t_0}^{\cdot} Y_s^n W(\mathrm{d}s) \Longrightarrow \int_{t_0}^{\cdot} Y_s W(\mathrm{d}s) \qquad \text{on } C([t_0, T])$$

equipped with the uniform topology. This theorem is an easy version of the general result since the integrator W does not depend on n and hence automatically fulfills the P-UT property (see, [65, page 377]). The stable nature of the aforementioned weak convergence follows by joint convergence $(\int_0^{\cdot} Y_s^n W(\mathrm{d}s), Y^n, W) \Longrightarrow (\int_0^{\cdot} Y_s W(\mathrm{d}s), Y, W)$ (cf. [72]). Hence we deduce that

$$\int_0^{\cdot} Y_s^n W(\mathrm{d}s) \xrightarrow{d_{st}} \int_0^{\cdot} Y_s W(\mathrm{d}s) \qquad \text{on } C([t_0, T]) \tag{6.36}$$

equipped with the uniform topology.

It is important to note that this result can not be directly applied to the process B_t^M since this process is not a semimartingale in general. For example when a kernel function g is given by

$$g(x) = x^\alpha \exp(-\lambda x)$$

with $\lambda > 0$ and $\alpha \in (-\frac{1}{2}, 0) \cup (0, \frac{1}{2})$. Thus, we will prove the stable convergence (6.32) by showing the stable convergence of finite dimensional distributions and tightness, we refer these concepts to Section 2.2.

To prove the stable convergence of finite dimensional distribution, we fix $u_1, \dots, u_k \in [t_0, T]$. Due to the condition (6.19), the finite dimensional version of (6.36) and continuous mapping theorem for stable convergence, we conclude the joint stable convergence

$$\left(\left\{\nu_M \int_{t_0}^{u_j} g(u_j - s)\{\sigma_s^M - \sigma_s\}W(\mathrm{d}s)\right\}_{j=1,\dots,k}, \nu_M \int_{t_0}^{\cdot} q(\cdot - s)\{a_s^M - a_s\}\mathrm{d}s\right)$$

$$\xrightarrow{d_{st}} \left(\left\{\int_{t_0}^{u_j} g(u_j - s)U_s^2 W(\mathrm{d}s)\right\}_{j=1,\dots,k}, \int_{t_0}^{\cdot} q(\cdot - s)U_s^1 \mathrm{d}s\right) \tag{6.37}$$

as $M \to \infty$. Here we remark that the stable convergence for the second component indeed holds, since the mapping $F : C([t_0, \tau]) \times D([t_0, T]) \to C([t_0, T])$,

$$F(q, a) = \int_{t_0}^{\cdot} q(\cdot - s)a_s \mathrm{d}s$$

is continuous. Hence, we are left with proving tightness for the first component of the process B_t^M. We fix $u, t \in [t_0, T]$ with $t > u$ and observe the decomposition

$$\nu_M \left(\int_{t_0}^{t} g(t - s)\{\sigma_s^M - \sigma_s\}W(\mathrm{d}s) - \int_{t_0}^{u} g(u - s)\{\sigma_s^M - \sigma_s\}W(\mathrm{d}s)\right)$$

$$= R_M^{(1)}(t, u) + R_M^{(2)}(t, u),$$

where

$$R_M^{(1)}(t,u) = \nu_M \int_u^t g(t-s)\{\sigma_s^M - \sigma_s\} W(\mathrm{d}s) \text{ and}$$

$$R_M^{(2)}(t,u) = \nu_M \int_{t_0}^u \{g(t-s) - g(u-s)\}\{\sigma_s^M - \sigma_s\} W(\mathrm{d}s).$$

Using Burkholder and Cauchy-Schwarz inequalities and (6.30), we have

$$\mathbb{E}[|R_M^{(1)}(t,u)|^4] \le C(t-u) \int_u^t |g(t-s)|^4 \mathrm{d}s.$$

Thus, we conclude that

$$\mathbb{E}[|R_M^{(1)}(t,u)|^4] \le C(t-u)^2. \tag{6.38}$$

Now, using the same methods we conclude that

$$\begin{aligned} \mathbb{E}[|R_M^{(2)}(t,u)|^4] &\le C \int_{t_0}^u |g(t-s) - g(u-s)|^4 \mathrm{d}s \\ &\le C(t-u)^{\min(4,4\alpha+1)}, \end{aligned} \tag{6.39}$$

where we used the inequality (6.28). Thus, applying (6.38), (6.39) and the Kolmogorov's tightness criteria, we deduce the tightness of the first component of the process B_t^M. This completes the proof of (6.32).

Next we prove the the ucp convergence at (6.31). To complete the proof, we first show pointwise convergence at (6.31). We start with the decomposition

$$X_t^{N,M} - X_t^0 = A^{N,M} + B_t^M,$$

where

$$A_t^{N,M} = \int_{t_0}^t \{\widetilde{h}_N(t-s) - g(t-s)\}\sigma_s^M W(\mathrm{d}s),$$

$$B_t^M = \int_{t_0}^t g(t-s)\{\sigma_s^M - \sigma_s\} W(\mathrm{d}s) + \int_{t_0}^t q(t-s)\{a_s^M - a_s\} \mathrm{d}s.$$

Recalling the notation from (6.27), we need to show that

$$\int_{t_0}^{t} \{\psi_N(t-s) - \psi(t-s)\}\sigma_s^M W(\mathrm{d}s) \xrightarrow{\mathbb{P}} 0 \qquad \text{as } N, M \to \infty,$$

for a fixed t. The Itô isometry immediately implies that

$$\sup_{M \in \mathbb{N}} \mathbb{E}\left[\Big|\int_{t_0}^{t} \{\psi_N(t-s) - \psi(t-s)\}\sigma_s^M W(\mathrm{d}s)\Big|^2\right] \le C \int_{t_0}^{t} \{\psi_N(t-s) - \psi(t-s)\}^2 \mathrm{d}s$$

$$\to 0 \qquad \text{as } N \to \infty,$$

which follows by Lemma 6.9 and the dominated convergence theorem. Hence we obtain pointwise convergence at (6.31). Since the limiting process A is continuous, to conclude ucp convergence from pointwise convergence in probability we need to show that

$$\sup_{N,M \in \mathbb{N}} \mathbb{E}[N^4 (A_t^{N,M} - A_u^{N,M})^4] \le C(t-u)^{1+\varepsilon}$$

for $t_0 < u < t$. To this end, we apply the same methods as in (6.38) and (6.39) to deduce the inequality

$$\begin{aligned} &\sup_{N,M \in \mathbb{N}} \mathbb{E}[N^4 (A_t^{N,M} - A_u^{N,M})^4] \\ &\le C\left((t-u)\int_u^t |\psi_N(t-s)|^4 \mathrm{d}s + \int_{t_0}^{u} |\psi_N(t-s) - \psi_N(u-s)|^4 \mathrm{d}s\right) \\ &\le C(t-u)^{1+\varepsilon}, \end{aligned}$$

which follows by Lemma 6.9 and condition (6.29). Therefore, the proof of Theorem 6.13 is completed □

List of Symbols

Σ $m \times m$ conditional covariance matrix of $V(f,g)^n$, 41

σ volatility function, 5

Σ^* $m \times m$ conditional covariance matrix of $V^*(q,r)^n$, 75

Σ_n positive semi-definite estimator of Σ, 44

$\xrightarrow{a.s.}$ almost sure convergence, 12

$\xrightarrow{d_{st}}$ converge stably in law, 17

$\xrightarrow{d}$ convergence in distribution, 15

$\xrightarrow{p}$ convergence in probability, 6

$\xRightarrow{ucp}$ ucp convergence, 16

τ_n sequence of stopping time, 12

$\tilde{a}, \tilde{\sigma}, \tilde{v}$ adapted, càdlàg stochastic process, 19

$V_l(f)^n$ power variation type estimator computed on l-th subsample, 43

$\widehat{h}$ Fourier transform of h, 113

$\widetilde{h}_N(x)$ estimator of $h(x)$, 122

$\{\mathcal{F}_t; 0 \leq t \leq \infty\}$ or $(\mathcal{F}_t)_{t \geq 0}$ filtration, 12

a drift function, 5

B Brownian motion independent of W, 19

$B_i(p)$ the ith block of high-frequency data, 47, 76

BSS Brownian semi-stationary, 112

$C^k(\mathbb{R})$ set of all k times continuous differentiable functions on $\mathbb{R}$, 28

$C_p^\infty(\mathbb{R}^n)$ space of infinitely differentiable functions with polynomial growth, 27

$d(\cdot,\cdot)$ metric, 11

$D^k F$ kth order Malliavin derivative of $F \in \mathcal{S}$, 27

$D_t(\cdot)$ Malliavin derivative at time t, 33

I_m m-dimensional identity matrix, 20

$IV(f)_t$ integrated function of the diffusion coefficient, 5

k_n pre-averaging window, 72

L number of sumsamples, 43

$L^2[0,T]$ Hilbert space of square integrable functions $f : [0,T] \to \mathbb{R}$, 25

$MN(\cdot,\cdot)$ mixed normal distribution, 20

p block size, 48

SDE stochastic differentiable equation, 29

$V(f)$ integrated volatility, 20

$V(f)^n$ power variation, 19

$V(f,g)$ stochastic limit of $V(f,g)^n$, 40

$V(f,g)^n$ bipower variation, 40

$V^*(q,r)$ stochastic limit of $V^*(q,r)^n$, 74

$V^*(q,r)^n$ noisy bipower variation, 74

$V_l^*(q,r)^n$ noisy bipower variation type estimator computed on l-th subsample, 76

$V_l(f,g)^n$ bipower variation type estimator computed on l-th subsample, 47

$v_i(f,g)^n$ subsample statistic computed only from data within the ith block $B_i(p)$, 47

$v_i(q,r)^n$ noisy subsample statistic computed only from data within the ith block $B_i(p)$, 76

W Brownian motion, 5

$X_t^{N,M}$ estimator of BSS X_t^0, 125

Index

Bibliography

[1] Aldous, D.J., Eagleson, G.K.: On mixing and stability of limit theorems. Ann. Probab. **6**(2), 325–331 (1978)

[2] Andersen, T.G., Bollerslev, T: Answering the skeptics: yes, standard volatility models do provide accurate forecasts. Int. Econ. Rev. 39, 885–905 (1998)

[3] Andersen, T. G., Bollerslev, T., Diebold, F. X.: Parametric and nonparametric volatility measurement in Handbook of Financial Econometrics, ed. by L. P. Hansen, and Y. Ait-Sahalia. North-Holland, Amsterdam, pp. 67–138 (2010)

[4] Andersen, T., Terasvirta, T.: Realized volatility. Handbook of Financial Time Series. 555—575 (2009)

[5] Andreou, E., Ghysels, E.: Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation, and Empirical Results. J. Bus. Econ. Stat. **20**(3), 363–376 (2002)

[6] Aït-Sahalia, Y., Jacod, J., Li, J.: Testing for jumps in noisy high frequency data. J. Econometrics. **168**(2) , 207–222 (2012)

[7] Aït-Sahalia, Y., Mykland, P, Zhang, L.: Ultra high frequency volatility estimation with dependent microstructure noise. J. Econometrics. 160:160–175 (2011)

[8] Barndorff-Nielsen, O.E., Benth, F.E., Pedersen, J., Veraart, A.: On stochastic integration for volatility modulated Lévy–driven Volterra processes. Stochastic Process. Appl. **124**(1), 812–847 (2014)

[9] Barndorff-Nielsen, O.E., Benth, F.E., Szozda, B.: On stochastic integration for volatility modulated Brownian-driven Volterra processes via white noise analysis. Infin. Dimens. Anal. Quantum Probab. Relat. Top. **17**(2), 1450011 (2014)

[10] Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.: Modelling energy spot prices by volatility modulated Lévy-driven Volterra processes. Bernoulli. **19**(3), 803–845 (2013)

[11] Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.: Modelling electricity forward markets by ambit fields. Adv. Appl. Probab. **46**(3). (2014)

[12] Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.: Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. To appear in Banach Center Publ. (2014)

[13] Barndorff-Nielsen, O.E., Corcuera, J.M., Podolskij, M.: Multipower variation for Brownian semistationary processes. Bernoulli. **17**(4), 1159–1194 (2011)

[14] Barndorff-Nielsen, O.E., Corcuera, J.M., Podolskij, M.: Limit theorems for functionals of higher order differences of Brownian semi-stationary processes. In: Honor of Yuri Prokhorov, V. (eds.), Shiryaev, A.N., Varadhan, S.R.S., Presman E.L. Prokhorov and Contemporary Probability Theory, vol. 33, pp. 69–96. Springer (2013)

[15] Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J., Podolskij, M., Shephard, N.: A central limit theorem for realized power and bipower variations of continuous semi-martingales. In: From Stochastic Calculus to Mathematical Finance: The Shiryaev Festschrift, Y. Kabanov and R. Lipster and J. Stoyanov, pp. 33–68. Springer-Verlag, Heidelberg (2006)

[16] Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N.: Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. Econometrica. **76**(6), 1481–1536 (2008)

[17] Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N.: Realized kernels in practice: trades and quotes. Econometrics Journal. **12**(3), 1–32 (2009)

[18] Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N.: Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. J. Econometrics. **12**(3), 149–169 (2011)

[19] Barndorff-Nielsen, O.E., Jensen, E.B.V., Jónsdóttir, K.Y., Schmiegel, J.: Spatio-temporal modelling - with a view to biological growth. In: Finkenstdt, B., Held L., IshamV. Statistical Methods for Spatio-Temporal Systems, pp. 47–75. London: Chapman and Hall/CRC (2007)

[20] Barndorff-Nielsen, O.E., Schmiegel, J.: Ambit processes; with applications to turbulence and cancer growth. In: Benth, F.E., Nunno, G.D., Linstrøm, T., Øksendal, B., Zhang, T. (Eds.). Stochastic Anal. Appl. The Abel Symposium 2005, pp. 93–124. Springer, Heidelberg (2007)

[21] Barndorff-Nielsen, O.E., Schmiegel, J.: Time change, volatility and turbulence. In: Sarychev, A., Shiryaev, A., Guerra, M., Grossinho, M.d.R.(Eds.) Proceedings of the Workshop on Mathematical Control Theory and Finance, Lisbon 2007, pp. 29–53. Springer, Berlin (2008)

[22] Barndorff-Nielsen, O.E., Schmiegel, J.: Brownian semistationary processes and volatility/intermittency. In: Albrecher, H., Runggaldier, W., Schachermayer, W. (Eds.) Advanced Financial Modelling. Radon Series Comp. Appl. Math. 8., 1–26 (2009)

[23] Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realised volatility and its use in estimating stochastic volatility models. J. of Roy. Stat. Soc., Series B 64, 253–280 (2002)

[24] Barndorff-Nielsen, O.E., Shephard, N.: Power and bipower variation with stochastic volatility and jumps. J. Financ. Econ. **2**(1), 1–48 (2004)

[25] Barndorff-Nielsen, O.E., Shephard, N.: Variation, jumps, market frictions and high frequency data in financial econometrics in Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Volume III, ed. by R. Blundell, P. Torsten, and W. K. Newey. Cambridge University Press, Cambridge, 328–372 (2007)

[26] Barndorff-Nielsen, O.E., Shephard, N., Winkel, M.: Limit theorems for multipower variation in the presence of jumps. Stoch Process Their Appl. **116**(5), 796–806 (2006)

[27] Basse, A.: Gaussian moving averages and semimartingales. Electron. J. Probab. **13**(39), 1140–1165 (2008)

[28] Basse-O'Connor, A., Graversen, S.-E., Pedersen, J.: A unified approach to stochastic integration on the real line. Theorey Probab. Appl. **58**, 355–380 (2013)

[29] Basse-O'Connor, A., Lachièze-Rey, R., Podolskij, M.: Limit theorems for stationary increments Lévy driven moving average processes. Working paper (2015)

[30] Bennedsen, M., Lunde, A., Pakkanen, M.S.: Discretization of Lévy semistationary processes with application to estimation. CREATES Research Papers, available at arXiv:1407.2754 (2014)

[31] Benth, F.E., Eyjolfsson, H.: Simulation of volatility modulated Volterra processes using hyperbolic stochastic partial differential equations. To appear in Bernoulli (2015)

[32] Benth, F.E., Eyjolfsson, H., Veraart, A.E.D. : Approximating Lévy semistationary processes via Fourier methods in the context of power markets. SIAM J. Financ. Math. **5**(1), 71–98 (2014)

[33] Benth, F.E., Süß, A.: Integration theory for infinite dimensional volatility modulated Volterra processes. Working paper, available at arXiv:1303.7143. To appear in Bernoulli (2013)

[34] Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)

[35] Black, F.: Studies of stock market volatility changes. In Proceedings of the American Statistical Association, Business and Economic Statistics Section. 177–181 (1976)

[36] Black, F.: Noise. J. Finance. **41**(3), 529–543 (1986)

[37] Black, F., Scholes, M.: The pricing of options and corporate liabilities. J. Polit. Econ. **81**(3), 637—654 (1973)

[38] Bruce, K.D.: Rough path analysis. File:rps.tex (2013)

[39] Chen, B., Chong, C., Klüppelberg, C.: Simulation of stochastic Volterra equations driven by space–time Lévy noise. Working paper, available at arXiv:1501.01645 (2015)

[40] Christensen, K., Oomen, R. C. A., Podolskij, M.: Fact or friction: Jumps at ultra high frequency. J. Financ. Econ. 114(3), 576–599 (2014)

[41] Christensen, K., Podolskij, M., Thamrongrat, N., Veliyev, B.: Inference from high-frequency data: A subsampling approach. Working paper (2015)

[42] Christie, A. A.: The stochastic behavior of common stock variances: Value, leverage and interest rate effects. J. Financ. Econ. 10(4), 407–432 (1982)

[43] Corcuera, J.M., Hedevang, E., Pakkanen, M., Podolskij, M.: Asymptotic theory for Brownian semi-stationary processes with application to turbulence. Stochastic Process. Appl. **123**, 2552–2574 (2013)

[44] Davis, B.: On the integrability of the martingale square function. Israel Journal of Mathematics. **8**(2), 187–190 (1970)

[45] Dette, H., Podolskij, M.: Testing the parametric form of the volatility in continuous time diffusion models–A stochastic process approach. J. Econometrics. **143**(1), 56–73 (2008)

[46] Dette, H., Podolskij, M., Vetter, M.: Estimation of integrated volatility in continuous-time financial models with applications to goodness-of-fit testing. Scandinavian Journal of Statistics. **33**(2), 259–278 (2006)

[47] Diebold, F. X., Strasser, G. H.: On the correlation structure of microstructure noise: A financial economic approach. Review of Economic Studies. **80**(4), 1304–1337 (2013)

[48] Eyjolfsson, H.: Approximation methods for ambit fields applied to power markets. PhD thesis, Department of Mathematics, University of Oslo (2013)

[49] Folland, G.B.: Real analysis - modern techniques and their applications. Wiley&Sons, J. (1984)

[50] Gatheral, J., Oomen, R. C. A.: Zero-intelligence realized variance estimation. Finance and Stochastics, **14**(2), 249–283 (2010)

[51] Gärtner, K., Podolskij, M.: On non-standard limits of Brownian semi-stationary processes. Stochastic Process. Appl. **125**(2), 653–677 (2015)

[52] Gloter, A., Jacod, J.: Diffusions with measurement errors. I - local asymptotic normality, ESAIM: Probability and Statistics. **5**, 225–242 (2001)

[53] Gloter, A., Jacod, J.: Diffusions with measurement errors. II - measurement errors, ESAIM: Probability and Statistics. **5**, 243–260 (2001)

[54] Greene, W. H.: Econometric Analysis. Prentice Hall. 7th edition (2011)

[55] Hansen, P.R., Lunde, A.: Realized variznce and market microstructure noise. J. Bus. Econ. Stat. **24**(2), 127–161 (2006)

[56] Hasbrouck, J.: Assessing the quality of a security market: A new approach totransaction-cost measurement, Review of Financial Studies. **6**, 191–212 (1993)

[57] Hautsch, N., Kyj, L. M., Oomen, R. C. A.: A blocking and regularization approach to high dimensional realized covariance estimation. J. Appl. Econom. **27**(4), 625–645 (2012)

[58] Hautsch, N., Podolskij, M.: Pre-averaging based estimation of quadratic variation in the presence of noise and jumps: Theory, implementation, and empirical evidence. J. Bus. Econ. Stat. **31**(2), 165–183 (2013)

[59] Heston, S. L.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies. **6**(2), 327–343 (1993)

[60] Ikeda, S.: A bias-corrected rate-optimal estimator of the covariation matrix of multiple security prices with dependent and endogenous microstructure effect. Working paper, Boston University (2015)

[61] Jacod, J.: The Euler scheme for Lévy driven stochastic differential equations: limit theorems. Ann. Probab. **32**(3), 1830–1872 (2004)

[62] Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M.: Microstructure noise in the continuous case: The pre-averaging approach. Stoch Process Their Appl. **119**(7), 2249–2276 (2009)

[63] Jacod, J., Protter, P.: Asymptotic error distributions for the Euler method for stochastic differential equations. Ann. Probab. **26**(1), 267–307 (1998)

[64] Jacod, J., Protter, P.: Discretization of Processes. Springer-Verlag, Berlin, 2 edn (2012)

[65] Jacod, J., Shiryaev, A.N.: Limit theorems for stochastic processes, 2nd Edition. Springer-Verlag, Berlin (2002)

[66] Jorion, P.: Value at Risk: The New Benchmark for Managing Financial Risk. McGraw-Hill, 3rd edn (2006)

[67] Kalnina, I.: Subsampling high frequency data. J. Econometrics. **161(2)**, 262–283 (2011)

[68] Kalnina, I.: Nonparametric tests of time variation in betas. Working paper, Université de Montréal. (2012)

[69] Kalnina, I., Linton, O.: Inference about realized volatility using infill subsampling. Working paper, London School of Economics. (2007)

[70] Kalnina, I., Linton, O.: Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. J. Econometrics. **147**(1), 47–59 (2008)

[71] Kuo, H.-H.: Introduction to stochastic integration. Springer (2006)

[72] Kurtz, T.G., Protter, P.: Weak limit theorems for stochastic integrals and stochastic differential equations. Ann. Probab. **19**(3), 1035–1070 (1993)

[73] Niederhoffer, V., Osborne, M.F.F.: Market making and reversal on the stock exchange. J. Am. Statist. Assoc. **61**(316), 897–916 (1966)

[74] Mancini, C.: Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. Scandinavian Journal of Statistics. **36**(2), 270–296 (2009)

[75] Markowitz, H. M.:Portfolio selection. J. Finance. **7**(1), 77–91 (1952)

[76] Mykland P. A., Zhang, L.: Assessment of uncertainty in high frequency data: The observed asymptotic variance. Working paper, University of Chicago (2014)

[77] Nualart, D.: The Malliavin calculus and related topics. 2nd Edition. Springer-Varlag, Berlin (2006)

[78] Oomen, R. A. C.: Modelling Realized Variance When Returns Are Serially Correlated. unpublished manuscript, University of Warwick,Warwick Business School. (2002)

[79] Oomen, R. C. A.: Properties of realized variance under alternative sampling schemes. J. Bus. Econ. Stat. 24(2), 219–237 (2006)

[80] Øksendal, B.: An introduction of Malliavin calculus with applications to economics. Technical report, Norwegian school of economics and business administration, Bergen (1996)

[81] Øksendal, B.: Stochastic differential equations. 5th Edition. Springer-Verlag, Heidelberg New York (2003)

[82] Pakkanen, M.: Limit theorems for power variations of ambit fields driven by white noise. Stochastic Process. Appl. **124**(5), 1942–1973 (2014)

[83] Podolskij, M.: Ambit fields: survey and new challenges. To appear in Special Volume of XI Symposium of Probability and Stochastic Processes (eds.) Mena, R., Pardo, J.C., Rivero, V., Bravo, G.U. (2014)

[84] Podolskij, M., Thamrongrat, N.: A weak limit theorem for numerical approximation of Brownian semi-stationary processes. Stochastics of environmental and financial economics: Centre of advanced study, Oslo, Normay. SringerOpen. pp. 101–120 (2015)

[85] Podolskij, M., Vetter, M.: Bipower-type estimation in a noisy diffusion setting. Stochastic Process. Appl. **119**(9), 2803–2831 (2009)

[86] Podolskij, M., Vetter, M.: Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Bernoulli. **15**(3), 634–658 (2009)

[87] Podolskij, M., Vetter, M.: Understanding limit theorems for semimartingales: A short survey. Stat. Neerl. **64** 329–351 (2010)

[88] Politis, D.N., Romano, J.P.: Large sample confidence regions based on subsamples under minimal assumptions. Annals of Statistics. **22**(4), 2031–2050 (1994)

[89] Politis, D.N., Romano, J.P., Wolf, M.: Subsampling. Vol **1**, Springer-Verlag (1999)

[90] Protter, P.: Stochastic integration and differential equations, 2nd Edition Version 2.1. Springer-Verlag (20

[91] Rényi, A.: On stable sequences of events. Sankhyā Ser. A. **25**, 293–302 (1963)

[92] Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion. 2nd Edition. Springer-Verlag (1994)

[93] Roll, R.: A simple model of the implicit bid-ask spread in an efficient market. J. Finance. **39**, 1127–1139 (1984)

[94] Sharpe, W. F.: Capital asset prices: A theory of market equilibrium under conditions of risk. J. Finance. **19**(3), 425—442 (1964)

[95] Varneskov, R.T.: Flat-top realized kernel estimation of quadratic covariation with non-synchronous and noisy asset prices. Working paper, CREATES, Aarhus University (2014)

[96] Veraart, A. E. D., Veraart, L. A. M.: Modelling electricity day-ahead prices by multivariate Lévy semistationary processes. In F. E. Benth, V. A. Kholodnyi, and P. Laurence (Eds.), Quantitative Energy Finance, pp. 157–188. Springer, New york (2014)

[97] Vetter, M.: Estimation methods in noisy diffusion models. PhD thesis, Department of Mathematics, Ruhr-Universität Bochum (2008)

[98] Vetter, M., Dette, H.:Model checks for the volatility under microstructure noise. Bernoulli, 18(4):1421–1447, 2012.

[99] Wang, C. D., Mykland, P. A.: The estimation of leverage effect with high-frequency data. J. Am. Stat. Assoc. 109(505), 197–215 (2014)

[100] Zhang, L.: Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Bernoulli. **12**(6), 1019–1043 (2006)

[101] Zhang, L., Mykland, P.A., Aït-Sahalia, Y.: A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Am. Statist. Assoc. **100**(472), 1394–1411 (2005)

[102] Zhou, B.: High-frequency data and volatility in foreign-exchange rates. J. Bus. Econ. Stat. **14**, 45–52 (1996)