INAUGURAL - DISSERTATION

zur

Erlangung der Doktorwürde

der

Naturwissenschaftlich-Mathematischen Gesamtfakultät

der

Ruprecht - Karls - Universität

Heidelberg

vorgelegt von

Diplom-Mathematiker

Sophon Tunyavetchakit

aus

Bangkok/ Thailand

Tag der mündlichen Prüfung:

Volatility Decomposition and Nonparametric Estimation of Spot Volatility of Models with Poisson Sampling under Market Microstructure Noise

Gutachter: Prof. Dr. Rainer Dahlhaus

Zusammenfassung

Die Schätzung der Volatilität von Hochfrequenzdaten unter Mikrostruktur-Rauschen wurde in den letzten Jahren intensiv untersucht. Im Gegensatz zu den meisten bisherigen Forschungen konzentrieren wir uns in dieser Doktorarbeit auf die Spot-Volatilitätsschätzung eines zeitgeänderten Preismodells, das auf Transaktionszeiten basiert. In diesem Modell wird die Volatilität in zwei Hauptteile, nämlich die Transaktionszeit-Volatilität und die Transaktionsintensität, zerlegt. Diese Teile können anhand gebebener Daten analysiert werden und enthalten wertvolle Informationen. Durch die einzelne Untersuchung der beiden Kurven gewinnen wir mehr Einblicke in die Ursache und die Struktur der Preisschwankungen. Die zentralen methodologischen und theoretischen Beiträge dieser Arbeit sind die Einführung und die theoretische Untersuchung eines neuen Volatilitätschätzers, der auf dieser Volatilitätszerlegung beruht.

Um die Transaktionszeit-Volatilität unter Mikrostruktur-Rauschen zu schätzen, passen wir einen konsistenten Schätzer, der auf der Pre-Averaging-Methode basiert, auf unsere Situation an, um das Rauschen in dem Modell zu berücksichtigen. Die asymptotischen Eigenschaften der beiden Schätzer—der klassische Volatilitätsschätzer und der auf der Zerlegung basierende Volatilitätsschätzer—werden durch einen "infill"-asymptotischen Ansatz untersucht. Wir vergleichen diese Schätzer, um den Vorteil der Volatilitätsfaktorisierung in diesem Transaktionzeit-Modell zu zeigen. Wir stellen fest, dass der alternative den klassischen Schätzer in vielen Fällen im Vergleich hinsichtlicher Konvergenzrate übertrifft. Schließlich untersuchen wir die Leistung unserer Schätzer in endlichen Proben in einer Simulationsstudie. Unsere Analyse realer Daten von hohen liquiden Mitteln zeigt uns einige interessante empirische Phänomene: (i) die U-Form der Spot-Volatilität über einem Handelstag ist hauptsächlich das Feature der Intensität; (ii) die Transaktionszeit-Volatilität ist im Wesentlichen glatter als die Volatilität und die Intensität; (iii) die Auswirkungen des Mikrostruktur-Rauschens auf die Schätzung der Spot-Volatilität ist sehr gering.

Abstract

The estimation of volatility for high-frequency data under market microstructure noise has been extensively studied during recent years. In this work, in contrast to the majority of previous research, we focus on the estimation of spot volatility of a time-changed price-model based on transaction times. In this model, volatility is decomposed into the product of two main curves, namely transaction-time volatility and trading intensity, both of which can be analyzed from data and contain valuable information. By inspecting these two curves individually we gain more insight into the cause and structure of volatility. The main methodological and theoretical contributions of this work are the introduction and theoretical investigation of a new volatility estimator based on this volatility decomposition.

For the estimation of transaction-time volatility under microstructure noise, we adapt a noise-robust estimator based on the pre-averaging technique to our situation in order to cope successfully with the contamination. The asymptotic properties of both estimators—the classical volatility estimator and the alternative volatility estimator (based on the decomposition)—are investigated using an infill asymptotic approach. We compare these estimators in order to see the benefit of factorizing the volatility in this transaction-time model. We find that the alternative estimator outperforms the classical one in many cases in terms of the rate of convergence. Finally, we explore the performance of our estimators in the finite-sample setting in simulations. Our real-data analysis of high-liquid assets reveals several interesting empirical phenomena: (i) the U-shape in the spot volatility over a trading day is primarily the feature of the intensity; (ii) the tick-time volatility is considerably smoother than the clock-time volatility and intensity; (iii) the impact of microstructure noise on spot volatility estimation is very small.

Acknowledgments

First of all, I would like to extend my utmost gratitude to my supervisor, Prof. Dr. Rainer Dahlhaus, for giving me this opportunity to undertake this interesting project and for providing me with the greatest support and advice. Your guidance, teaching, and enthusiasm have always kept me motivated throughout these years. Thank you for all the constructive comments and the fruitful discussions during the course of this project.

I would like to also express my sincere gratitude to, Prof. Dr. Mark Podolskij, for his invaluable discussions during his stay in Heidelberg, and for many useful comments when I attended Zurich Spring School and Aarhus Conference.

I want to thank all my colleagues in the Institute of Applied Mathematics and all the RTG members from the University of Heidelberg and Mannheim for providing me with a friendly and stimulating atmosphere. Special thanks also to my former and current office mate Dr. Christian Schmidt and Nopporn Thamrongrat for their companionship, encouragement and inspiring discussions both intellectually and spiritually. I am also grateful for Nuttarak Sasipong and Dr. Karl Gregory help with my English.

I would also like to take this opportunity to thank both my family and my girlfriend, Rattawan Kullasakboonsri, for their constant support. Without their encouragement, this project would have been impossible.

Lastly, the financial support of Deutsche Forschungsgemeinschaft through the Research Training Group RTG 1953 “Statistical Modeling of Complex Systems and Processes”, the University of Heidelberg, and the Development and Promotion of Science and Technology Talents Project, is gratefully acknowledged.

Contents

Chapter 1

Introduction

The estimation of volatility for high-frequency data under market microstructure noise has been extensively studied during recent years; see Zhang *et al.* [70], Zhang [69], Barndorff-Nielsen *et al.* [11] and Jacod *et al.* [47], among many others, and see also Aït-Sahalia and Jacod [2] for an overview. The majority of these investigations have been in the framework of a diffusion model. In this work, contrary to the majority of previous research, we focus on the estimation of spot volatility of a time-changed price-model based on trading times. In this model, volatility is decomposed into two main factors, namely transaction-time volatility (or tick-time volatility) and trading intensity, both of which can be identified and both of which contain valuable information. The main methodological and theoretical contributions of this dissertation are the introduction and the theoretical investigation of a volatility estimator based on this volatility decomposition.

For the estimation of tick-time volatility under microstructure noise we have to adapt an estimator from diffusion models to our situation which can cope successfully with microstructure noise. Standard volatility estimators have been reported in practice to be non-robust to financial data observed at very high frequencies. Many noise-robust estimators have therefore been introduced in the literature, such as those approaches presented by the authors listed above. Starting with the excellent work of Zhang *et al.* [70], the combination of two different timescales—coarse and fine grids—leads to a consistent estimator for the integrated volatility. This idea is extended later to the multi-timescale estimator that achieves the optimal rate of convergence $n^{-1/4}$; see Zhang [69]. Barndorff-Nielsen *et al.* [11] suggest a flat-top kernel-type estimator, which they call a realized kernel, which combines different lags of autocovariances to eliminate the effects of the microstructure noise contamination. In this work, we apply the pre-averaging technique, which is introduced by Podolskij and Vetter [60] and later extended by Jacod *et al.* [47], in order to construct a noise-robust estimator for the tick-time volatility of our time change model. Recently, the estimator based on a local method of moments, presented by Reiß [64] and Bibinger and Reiß [15], has gotten more attention, as their estimator achieves asymptotic efficiency while the other three approaches above do not. We place our emphasis on the pre-filtering technique since it has the advantages of being intuitive and extendable in a straightforward way to other power variations. In addition to these major approaches, which are mainly

based on additive noise models, a more complex structure such as nonlinear microstructure noise models is also discussed in Dahlhaus and Neddermeyer [25].

Time-changed price-models were first investigated in Clark [22] in connection with finance. In his work, the volume of trades is suggested to be a subordinator of a Brownian motion in order to recover the normality of the distribution of the future cotton price. Afterwards a relationship between asset returns, price fluctuation, and market activities measured by trading volume and numbers of transactions is extensively discussed. Ané and Geman [5] conclude that the number of trades explains the volatility change better than their volume, so they recover the normality of asset returns through this stochastic time change in high-frequency data; see also Jones *et al.* [50], Plerou *et al.* [59] and Gabaix *et al.* [34] for more detailed discussions of this correlation. Due to the existence of various mathematical tools, time-changed Brownian motion is an attractive and tractable process for studying arbitrage-free asset returns, which are shown to be semimartingales (see e.g. Delbaen and Schachermayer [27]). Indeed, having a class of time-changed Brownian motion is satisfactory since it is as large as a class of semimartingale; see Monroe [57]. In recent years, other time change models have been extensively studied, in particular a time-changed Lévy process which allows for a more complex structure in the price models in order to cope with some stylized effects emerging in the real market; for details refer to Carr *et al.* [19] and Carr and Wu [20], and to Belomestny [14] for a statistical treatment of this kind of models.

As mentioned, instead of a classical semimartingale model, we use a diffusion model subordinated by transaction time

$$dX_t = \sigma(t)\ dW_{N_t} \qquad \text{for } t \in [0,T],$$

where $W_\cdot$ is a standard Brownian motion, and the stochastic time change is a point process N_t with a time-varying intensity $\lambda(t)$ representing the accumulated number of trades up to time t. We assume that the drift of the price process is zero, as we are working with short time intervals. Our main objective is the spot volatility of this model, which is expressed by the product of the transaction-time volatility $\sigma^2(\cdot)$ and the intensity $\lambda(\cdot)$, i.e.

$$\sigma^2_{clock}(t) \ = \ \sigma^2(t) \cdot \lambda(t);$$

"volatility per time unit is equal to volatility per transaction multiplied by the average number of transactions per time unit".

This volatility decomposition enables us to construct an alternative volatility estimator by multiplying estimators of $\lambda(\cdot)$ and $\sigma^2(\cdot)$. By inspecting these two estimation curves separately we gain more insight about the cause and structure of volatility. In particular, it is empirically shown that the tick-time volatility estimator of high-liquid stocks is considerably smoother than the clock-time volatility and the trading intensity. Beyond its empirical interpretation, this has an important implication, i.e. we may choose a larger bandwidth, giving us effectively more data, than with the classical volatility estimator. This is reflected, in theory, in a higher rate of convergence of the estimator. A similar

volatility decomposition is obtained by Dahlhaus and Neddermeyer [25]. In their work, in which the models are based on non-linear microstructure noise models, a sequential EM-algorithm on the filtering distributions of the efficient log-prices is used to determine the spot volatility. This estimator will be used as a benchmark in our simulation studies to see the quality of pre-averaging type estimators in finite samples.

We investigate asymptotic properties of the voltility estimators by a type of infill-asymptotics. Specifically, we let the parameters depend on the time span T, similarly to nonparametric regression with time-varying coefficients or locally stationary processes; see Robinson [65] and Dahlhaus [24]. The asymptotic inference is drawn by letting T go to infinity. As the time span T increases, more and more local information of the same structure of parameter curves can be obtained. Therefore, a rescaled time version of the transaction-time model is given as follows:

$$dX_{t,T} = \sigma\left(\frac{t}{T}\right)\frac{1}{\sqrt{T}}\, dW_{N_{t,T}} \qquad \text{for } t \in [0,T],$$

where $N_{t,T}$ is a nonhomogeneous Poisson process admitting an intensity function $\lambda(t/T)$, and $\sigma(t/T)$ is a tick-time volatility function. We have also assumed for simplicity that $N_{\cdot,T}$ and $W_{\cdot}$ are independent; this does not appear to be strictly necessary (in which case some additional asymptotic variance terms will emerge; see our discussions in Chapter 5). As our approach is purely nonparametric, we impose some smoothness conditions, namely Hölder continuity, on the underlying parameter curves $\sigma^2(\cdot)$ and $\lambda(\cdot)$ so that they are required to vary slowly over time. To deal with a measurement error induced by market microstructure noise, we adapt the pre-filtering technique of Jacod *et al.* [47] to our observable transaction-time model

$$Y_{t_i,T} = X_{t_i,T} + \varepsilon_i \qquad \text{for } i = 1, ..., N_T.$$

For instance, a tick-time volatility estimator is given by

$$\widehat{\sigma}^2_{pavg}(u_o) := \frac{T}{NH}\frac{1}{g_2}\sum_{i=i_o-N}^{i_o+N} k\left(\frac{i-i_o}{N}\right)\left(\overline{\triangle Y}_{t_i,T}\right)^2 - \frac{T}{2NH}\frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2}\sum_{i=i_o-N}^{i_o+N} k\left(\frac{i-i_o}{N}\right)\left(Y_{t_i,T} - Y_{t_{i-1},T}\right)^2,$$

where $i_o := \inf\{i : t_i \geq u_o T\}$, for $u_o \in (0,1)$.

The asymptotic results of our pre-averaging volatility estimators are compared in the terms of their asymptotic variance and rate of convergence; this comparison is summarized in Table 4.2. It turns out that the new volatility estimator based on the volatility decomposition outperforms the volatility estimator based on the classical method in many cases, including the case when the tick-time volatility evolves more smoothly than the trading intensity. In particular, an example given at the end of Chapter 4 shows that the alternative estimator can achieve a smaller risk than the lower bound for spot volatility estimation applied to the classical diffusion models which is derived in a minimax sense with respect

to the L_2-loss function; see Munk and Schmidt-Hieber [58]. This highlights the potential of the volatility decomposition in the transaction-time model.

In Chapter 2, we state the mathematical foundations required in the work, in particular nonparametric estimation techniques for deriving spot volatility in the high-frequency framework. In particular, we focus on the kernel-filtering of realized volatility suggested by Kristensen [53] and Fan and Wang [33]. An overview of volatility estimation under market microstructure noise is presented. This noise is indeed the price discrepancy between the efficient log-price and the observed log-price, which is caused by bid-ask spreads, price discreteness, or price formation, etc. Furthermore, some properties of Poisson point processes with links to martingale theory are also given in this chapter for the sake of referring to them later.

In Chapter 3, an introduction to asset price models based on a stochastic time change, which reflects market activities, is provided. The main contribution of this dissertation is the introduction of the new volatility estimator based on the volatility decomposition, which is proven under a quite general assumptions; see Assumptions 3.1.

Chapter 4 is devoted to infill asymptotics for our transaction-time model, which we carry out by applying the rescaling method. The asymptotic properties of both estimators—the classical volatility estimator and the alternative volatility estimator–are compared in Table 4.1 and 4.2 (for the noiseless and noisy model respectively) to demonstrate the benefit of factorizing the volatility. In particular, under the presence of noise, the alternative one outperforms the usual one in many cases in that a faster rate of convergence can be achieved.

A consistent estimator of spot volatility has more to offer than an estimator of the integrated volatility, as it enables the construction of estimators of many functionals of the spot volatility, of which integrated volatility is one example. One of the main drawbacks is its extra smoothing parameter, which needs to be chosen. Chapter 5 discusses the selection of these smoothing parameters and also proposes a method to deal with boundary effects of our statistics. The extension to endogenous time models allowing for a leverage effect between the price process, volatility, and market activities is also briefly discussed. We note that this generalization gives rise to some additional asymptotic variances of the volatility estimators.

In Chapter 6, we examine the performance of our estimators in Monte-Carlo simulations and discuss implications for application. We present some simulation results for our pre-averaging type estimators for a situation with simplified but practical model parameters. From the empirical analysis, we have observed the following on the volatility curve and microstructure noise of ultra-high-frequency data: (i) the typical U-shape over a trading day is mainly the feature of the intensity; (ii) the tick-time volatility estimator is in general smoother than the clock-time volatility and intensity estimators; (iii) the effect of microstructure noise on spot volatility estimation is very small. Moreover, issues concerning the quality of pre-averaging method are also discussed. The overall conclusion is

that the pre-averaging estimator is robust to the different types of noise considered here; however, it performs poorly for small sample sizes due to its slow rate of convergence.

Conclusions are discussed in Chapter 7. Appendices (A), (B) and (C) correspond to the proofs of Chapters 3, 4 and 5, respectively. Throughout this dissertation the following notations are used: $f^{(n)}$ is the n-th derivative of a function f; I_A is the indicator function of A; and an integer part of a real number x is denoted by $\lfloor x \rfloor$. If not otherwise stated, all equalities and inequalities of random expressions are expressed in an almost sure sense.

Chapter 2

Preliminaries

2.1 Brownian Semimartingales and Some Properties

In the last decades financial mathematics has grown very fast, and a lot of models for asset returns have been introduced, both in the discrete- and continuous-time settings. Among many continuous-time models, a standard model for log-price processes is a Brownian semimartingale, which is the solution of the equation

$$dX_t^* = \mu_t^* dt + \sigma_t^* dW_t \quad \text{for } t \in [0, T], \tag{2.1}$$

where μ_t^* and σ_t^* are adapted càdlàg stochastic processes and W_t is a standard Brownian motion. This model is a special case of the more general process called a semimartingale, which is defined as the sum of a local martingale and an adapted process with finite variation on the compact interval[1]. For more details see Jacod and Shiryaev [49] and Protter [62], among many others. The semimartingale has been shown to be an appropriate choice for modeling arbitrage-free asset price returns; see e.g. Delbaen and Schachermayer [27]. The special model (2.1) is attractive and tractable for studying the behavior of price movement of financial products due to its simplified form and the various mathematical tools available for analyzing it. The most important component lying in the Brownian motion part, the volatility σ_t^{*2}, is mainly used in finance and econometrics for risk managing, option pricing, etc., since it explains the uncertainty in the price movement; see an early application of stochastic volatility in option pricing in Hull and White [45]. The process μ_t^* is the drift coefficient, which plays a less important role in the risk analysis, especially when one considers prices over a small period of time, such as a day or a shorter period.

In the high-frequency framework[2] it is well-known that the integrated volatility over a given fixed time period $[0, T]$ can be consistently estimated by the realized volatility (also

[1] It can also be equivalently defined as a good integrator on a class of processes as introduced by Protter [62].

[2] meaning that the observations $X_{t_1}^*, X_{t_2}^*, ..., X_{t_n}^*$ over the fixed time interval $[0, T]$ become more dense as $n \to \infty$; it is therefore an infill asymptotic framework.

called realized variance, realized quadratic variation). That is

$$\sum_{i=1}^{n} \left| X^*_{t_i} - X^*_{t_{i-1}} \right|^2 \xrightarrow{\mathbb{P}} \int_0^T \sigma_s^{*2} ds$$

as the number of subintervals n goes to infinity, where $0 = t_0 < t_1 < ... < t_n = T$ is a deterministic sampling (or a stochastic sampling under some stronger assumptions) with $\max_i |t_i - t_{i-1}| \to 0$ and T fixed. In fact the integrated volatility is the same as the quadratic variation of the model (2.1). Nowadays observing high-frequency data has become easier due to the efficiency of modern technologies. Figure 2.1 shows some examples of transaction data of assets traded on the NASDAQ stock exchange for April 1, 2014.

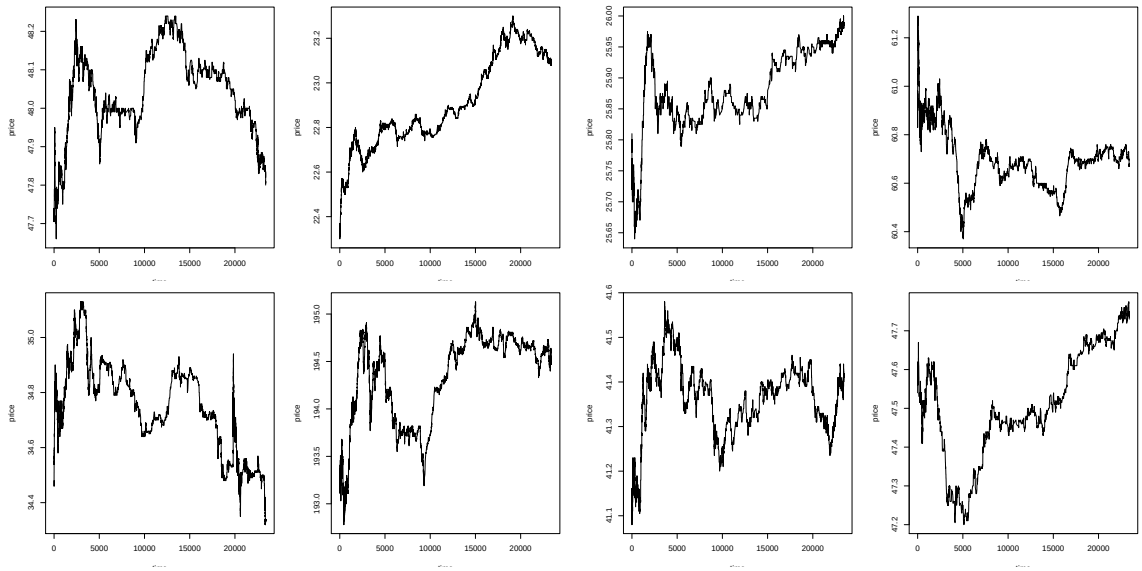


Figure 2.1: Some examples of high-frequency asset prices traded on NASDAQ for April 1, 2014. In the first row (from left to right): C, CSCO, INTC, and JPM and in the second row (from left to right): GM, IBM, MSFT, and VZ. X-axis is in 1-second resolution, i.e. 09:30 AM = 0 and 04:00 PM = 23400.

Asymptotic normality for the realized volatility has also been shown under a variety of assumptions. It is easiest to show in the case of deterministic equidistant observation times, that is when the price process is observed at every time point $t_i = (i \cdot T)/n$. We obtain

$$\sqrt{n} \left\{ \sum_{i=1}^{n} \left| X^*_{t_i} - X^*_{t_{i-1}} \right|^2 - \int_0^T \sigma_s^{*2} ds \right\} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ 2\int_0^T \sigma_s^{*4} ds\right). \tag{2.2}$$

Therefore the resulting limit distribution is a normal variance mixture. To get a feasible version of this central limit theorem we can estimate the stochastic variance by the realized

power variation of order 4 called realized quarticity. This leads to

$$\frac{\sum_{i=1}^{n}\left|X_{t_i}^{*}-X_{t_{i-1}}^{*}\right|^2-\int_0^T\sigma_s^{*2}ds}{\sqrt{\frac{2}{3}\sum_{i=1}^{n}\left|X_{t_i}^{*}-X_{t_{i-1}}^{*}\right|^4}}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)$$

under some conditions on μ_t^* and σ_t^*. See Andersen *et al.* [3], [4], Barndorff-Nielsen and Shephard [7], [8], and the literature therein; the history about the application of the sum of squared increments in finance can also be found in these papers. Due to the popularity of this research area, a vast number of results for this model and its extension are developed in the literature, including realized power variation, bipower or multipower variation, and cases of nonequidistant subdivision; see Barndorff-Nielsen and Shephard [8], [9], [10], and Barndorff-Nielsen *et al.* [13] among many others. In particular, the bi-/multipower variation has the benefit of testing for jumps in the price model. Furthermore, some of these results are generalized to provide functional limit statements; see the excellent work of Jacod and Shiryaev [49] and more recently Jacod and Protter [48]. We also recommend the more accessible work of Podolskij and Vetter [61] for understanding the main issues behind those deep theories and proofs. In particular, this work offers a clear explanation of stable convergence, which is a very useful tool for limit theorems of semimartingales observed at high-frequency. Of course we cannot mention all of the related results in this thesis; nevertheless, most of them can be found in the works of the authors mentioned and in the references therein.

2.2 Poisson Point Processes and Stochastic Intensity Models

A Poisson point process is a fundamentally useful process along with Brownian motion in the theory of continuous-time processes and is widely used for modeling arrivals or occurrences of events of interest. The main interest in this process is in its rate of arrivals, called intensity, which describes the average number of arrivals per unit of time. In many situations it is often useful to consider a Poisson process with a time-varying (or time-dependent) intensity called a nonhomogeneous Poisson process (NHPP). Such a process still has independent increments, but its stationary property does not hold anymore. Formally, given a probability space $(\Omega,\mathcal{F},\mathbb{P})$, let $\{N_t\}_{t\geq 0}$ be a NHPP with respect to the natural filtration $\left\{\mathcal{F}_t^N\right\}_{t\geq 0}$ with a deterministic non-negative real-valued intensity function λ_t, i.e.

i) $N_0=0$ a.s.,

ii) N_t-N_s is independent of $\mathcal{F}_s^N$ for all $0\leq s\leq t$, and

iii) $\mathbb{P}\left(N_t-N_s=k\right)=\exp\left\{-\int_s^t\lambda_l dl\right\}\cdot\frac{\left(\int_s^t\lambda_l dl\right)^k}{k!}$ for $k\in\mathbb{N}$, $0\leq s\leq t$,

where $\mathcal{F}_t^N=\sigma\left(N_s:s\leq t\right)$ contains all information of the process up to time t. It turns out that all realizations of N_t are right-continuous with left limits (càdlàg) which plays a key role in modeling jumps in asset price returns; see Jacod and Protter [48] and Cont

and Tankov [23]. Note that there are other equivalent definitions for Poisson processes which can be found in the literature. The representation used here is more convenient for our purposes. In particular, the associated arrival times $t_i := \inf\{t : N_t \geq i\}$ are stopping times with respect to the natural filtration $\left\{\mathcal{F}_t^N\right\}_{t\geq 0}$ since $\{t_n \leq t\} = \{N_t \geq n\} \in \mathcal{F}_t^N$.

A more general process is the doubly stochastic Poisson process (or conditional Poisson process or Cox process), which allows for stochastic intensity; however, the intensity process is still independent of the Poisson process. More precisely, we first draw a realization of the random intensity λ_t, and once the whole path of λ_t is selected we generate a Poisson process with this intensity. Having such a stochastic intensity is already interesting in practice, but we can go further and extend it to a more general point process with stochastic intensity that need not be Poisson anymore. This can be done by extending Watanabe's characterization of the doubly stochastic Poisson process to a large class of stochastic intensity models—so large that it contains almost all point processes of practical interest; see Brémaud [16, chapter II]. It has been pointed out that there is a connection between this point process and martingale dynamics (it is very helpful since many tools in martingale theory can be applied now). To easily see this relation we look at the compensated Poisson process with a constant intensity $\lambda > 0$. It implies that $N_t - \lambda t$ is a martingale with respect to $\mathcal{F}_t^N$ if N_t is integrable.

The following definition of general stochastic intensity models is the one taken by Brémaud [16]. In fact it is an extension of Watanabe's characterization of doubly stochastic Poisson processes. Throughout this section we will always assume that a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ is given such that $\{\mathcal{F}_t\}_{t\geq 0}$ is right-continuous, i.e. $\cap_{\epsilon>0}\mathcal{F}_{t+\epsilon} = \mathcal{F}_t$, and $\mathcal{F}_0$ contains all $\mathbb{P}-$null sets.

2.1 Definition *(Brémaud [16], Definition D7, p.27) Suppose that a point process N_t is adapted to a filtration $\mathcal{F}_t$, and a non-negative $\mathcal{F}_t$-progressive process λ_t is given such that*

$$\int_0^t \lambda_s ds < \infty \quad \mathbb{P} - a.s., \quad \text{for all } t \geq 0,$$

holds. Then N_t is said to admit the $\mathcal{F}_t$-intensity λ_t if the following condition holds:

$$\mathbb{E}\left[\int_0^\infty c_s \; dN_s\right] \; = \; \mathbb{E}\left[\int_0^\infty c_s \lambda_s \; ds\right], \tag{2.3}$$

for all non-negative $\mathcal{F}_t$-predictable processes c_t.

In most applications the intensity is restricted to predictable, in which case the intensity process is uniquely determined. Indeed, a predictable version of the intensity is always available; see Brémaud [16, Chapter II.4.]. The concept of predictable processes is very important in stochastic integral theory; for our present work it is enough to know that any left-continuous adapted process with right-hand limits (cáglág) is predictable.

There are some essential remarks and results relating to our work when we look at stochastic intensity dynamics, and therefore we summarize them here for the purpose of refering to them later (all the proofs can be found in Brémaud [16]). A review of the whole martingale theory with respect to Poisson point processes can also be looked up in, for example, Kuo [54] and Jacod and Protter [48]. See also a short review in Aalen [1].

2.2 Remark *i) A doubly stochastic Poisson process is a special case of Definition 2.1 in that its intensity λ_t is restricted to being $\mathcal{F}_0$-measurable. Hence, conditionally on $\mathcal{F}_0$, all information about the form of intensity is accessible. In formal language, conditionally on $\mathcal{F}_0$, N_t is an $\mathcal{F}_t$-NHPP.*

ii) By setting the function $c_s = 1_{(0,\tau]}(s)$ in (2.3) where τ is an $\mathcal{F}_t$-stopping time[3]*, we have*

$$\mathbb{E}[N_\tau] = \mathbb{E}\left[\int_0^\infty 1_{(0,\tau]}(s)dN_s\right] = \mathbb{E}\left[\int_0^\infty 1_{(0,\tau]}(s)\lambda_s ds\right] = \mathbb{E}\left[\int_0^\tau \lambda_s ds\right].$$

In particular, for $t \geq 0$

$$\mathbb{E}[N_t] = \mathbb{E}\left[\int_0^t \lambda_s ds\right],$$

which also implies that $N_0 = 0$ a.s.

iii) The corresponding arrival times $t_i := \inf\{t : N_t \geq i\}$ are $\mathcal{F}_t$-stopping times. Moreover, for any $\mathcal{F}_t$-adapted process ϕ_t, ϕ_{t_i} is $\mathcal{F}_{t_i}$-measurable or even $\mathcal{F}_{t_i^-}$-measurable if ϕ_t is (left-)continuous. The stopping filtration at a random time τ is defined by

$$\mathcal{F}_\tau = \sigma\{A \in \mathcal{F} \ : \ A \cap \{\tau \leq t\} \in \mathcal{F}_t, \text{ for all } t \geq 0\};$$

intuitively, it contains all information up to the random time τ. The strict past $\mathcal{F}_{\tau-}$ is defined by

$$\mathcal{F}_{\tau-} = \sigma(A \cap \{t < \tau\}, \ A \in \mathcal{F}_t, \ t \geq 0).$$

Particularly $\mathcal{F}_{\tau-} \subset \mathcal{F}_\tau$. See Brémaud [16, Appendix A2].

iv) The following important properties are listed in Theorem T8, Brémaud [16, p.27]:

a) N_t is non-explosive, i.e. $N_t < \infty$ $\mathbb{P}$-a.s. for $t \geq 0$.

b) $M_t := N_t - \int_0^t \lambda_s ds$ is an $\mathcal{F}_t$-local martingale.

c) If y_t is an $\mathcal{F}_t$-predictable process such that $\mathbb{E}\left[\int_0^t |y_s|\lambda_s ds\right] < \infty$ for all $t \geq 0$, then $\int_0^t y_s dM_s$ is an $\mathcal{F}_t$-martingale.

[3]Since τ is an $\mathcal{F}_t$-stopping time, we have that $\{\tau \leq s\} \in \mathcal{F}_s$ for all s, and therefore $\{\tau \geq s\} \in \mathcal{F}_{s+} = \mathcal{F}_s$ by the right-continuity of the filtration. Thus $c_t = 1_{(0,\tau]}(t)$ is $\mathcal{F}_t$-predictable ($\mathcal{F}_t$-adapted and left-continuous with right limits).

v) According to b), for a sequence of $\mathcal{F}_t$-stopping times $\tau_n \uparrow \infty$, $N_{t\wedge\tau_n} - \int_0^{t\wedge\tau_n} \lambda_s ds$ is an $\mathcal{F}_t$-martingale. Therefore

$$\mathbb{E}\left[N_{t\wedge\tau_n} - N_{s\wedge\tau_n} \;\middle|\; \mathcal{F}_s\right] = \mathbb{E}\left[\int_{s\wedge\tau_n}^{t\wedge\tau_n} \lambda_u du \;\middle|\; \mathcal{F}_s\right].$$

Letting τ_n go to infinity, we obtain

$$\mathbb{E}\left[N_t - N_s \;\middle|\; \mathcal{F}_s\right] = \mathbb{E}\left[\int_s^t \lambda_l dl \;\middle|\; \mathcal{F}_s\right]$$

if λ_t is assumed to be bounded. This yields that independent increments do not exist in this general setting. Moreover, if λ_t is continuous, and therefore predictable, then

$$\lim_{t\to s} \frac{1}{t-s}\mathbb{E}\left[N_t - N_s \;\middle|\; \mathcal{F}_s\right] = \lambda_s \quad \mathbb{P}-a.s.,$$

which reminds us of the classical definition of intensity functions.

vi) The boundedness of λ_t (which we shall always assume in our work) implies that $\mathbb{E}[N_t] < \infty$ and M_t is a square integrable $\mathcal{F}_t$-martingale. Therefore we can avoid working with local martingales. Moreover, by Doob-Meyer's decomposition theorem the compensator of M_t^2, denoted by $< M >_t$, is equal to $\int_0^t \lambda_s ds$. This result is useful as we can now apply many properties of stochastic integration with martingales as integrators such as Itô's isometry or Burkholder-Davis-Gundy's inequalities, see e.g. Kuo [54, Theorem 6.5.8] and Jacod and Protter [48, p.39] respectively.

2.3 Nonparametric Curve Estimation: Spot Volatility Estimator

In this section we will discuss statistical methods used to derive the main parameter in the asset price model (2.1). First it is interesting to have a brief look at some (nonparametric) techniques that are mainly used for estimating underlying curves, particularly in nonlinear regression and density estimation. We begin by thinking of a scatter plot of some data which could have a very complex structure. Our aim is to fit an appropriate curve to it. In nonlinear regression there are two approaches: the first one is the parametric approach, in which we fit a linear (which is normally biased) or a polynomial regression function to the scatter plot. To do this we would need to deal with the choice of the polynomial degree, say p. It is clear that large p can reduce the modeling bias but it can also cause high variability as the number of parameters is too high. In order to avoid the problem of specifying a model, the second approach of nonparametric fitting may be used. A lot of methods are developed in the nonparametric estimation literature, and some of them apply locally on the data set. The most common tools are kernel type estimators, local linear/polynomial fittings, wavelet thresholdings, and spline smoothings. All of these methods have their own advantages over the others, which can be seen in detail in many textbooks; we recommend Härdle and Linton [42] and chapter 2 of Fan and Gijbels [30] for an overview of these

smoothing techniques in the density estimation and regression analysis contexts. Local polynomial fitting in particular is intensively discussed in the latter. Instead of reviewing nonparametric devices in the connection with those topics, we prefer to discuss some of the techniques mentioned above in the volatility curve estimation context.

As mentioned in the beginning of this chapter, the integrated volatility can be consistently measured by the sum of squared increments observed with high-frequency. To extract the time-varying spot volatility σ_t^{*2}, the most intuitive way is to proceed as though the volatility process were constant on a small neighborhood. Then we can filter the realized volatility to get an approximation of the spot volatility at a specific time point $t_o \in [0,T]$, i.e. given high-frequency observations $X_{t_1}^*, ...X_{t_n}^*$ of the model (2.1)

$$\frac{1}{2h} \sum_{t_o-h<t_i\leq t_o+h} \left| X_{t_i}^* - X_{t_{i-1}}^* \right|^2 \approx \sigma^{*2}(t_o)$$

for a small window $h > 0$. By assigning smaller weights for data points that are remote from the interested time point t_o, a lower modeling bias will be obtained. This procedure leads to the kernel-based estimator introduced by Kristensen [53] and Fan and Wang [33], i.e.

$$\widehat{\sigma}_{t_o}^{*2} := \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{t_i - t_o}{h}\right) \left| X_{t_i}^* - X_{t_{i-1}}^* \right|^2, \tag{2.4}$$

where K is a kernel function satisfying regularity conditions, and h is the bandwidth depending on n (recall that the high-frequency framework (or infill asymptotics) is being applied, i.e. we assume that more and more observations on the fixed interval $[0,T]$ are available as n goes to infinity). Since this quantity only estimates the volatility at a single time point t_o, the whole volatility line is therefore obtained by applying this local estimator to a grid of points. Having spot volatility estimates is convenient since we are now able to estimate any functional of the spot volatility such as the former integrated volatility, derivatives of spot volatility (needed in the case of differentiable processes), etc. In order to draw some asymptotic inferences, of course, we need to know more about the structure of this volatility curve. It is natural to assume that the volatility curve satisfies some smoothness condition since we still want to stay in the nonparametric context rather than specifying a parametric model for the volatility. The following asymptotic results are given under assumptions $(A.1) - (A.4)$ and $(K.1)$ in Kristensen [53].

2.3 Theorem *(Kritensen [53], Theorem 3) Suppose that we have equidistant observations, i.e. $t_i = iT/n$. Then for any $a \to 0$ with $a/h \to 0$ we obtain*

$$\sup_{t\in[a,T-a]} \left| \widehat{\sigma}_t^{*2} - \sigma_t^{*2} \right| = O_p\left(h^{\alpha}\right) + O_p\left(\log(n)/\sqrt{nh}\right),$$

and if the bandwidth conditions $nh \to \infty$ and $nh^{2\alpha+1} \to 0$ hold, we get the following asymptotic normality:

$$\sqrt{nh}\left\{\widehat{\sigma}_{t_o}^{*2} - \sigma_{t_o}^{*2}\right\} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0\ , 2\sigma_{t_o}^{*4} \int_{\mathbb{R}} K^2(x)dx\right), \tag{2.5}$$

for $t_o \in (0,T)$, where α is the Hölder exponent of the volatility curve.

The same result is also obtained by Fan and Wang [33, Theorem 1] with a slightly different set of conditions on the volatility process and the bandwidth. They have shown that common volatility models in the literature (such as geometric OU process, Nelson GARCH diffusion process, CIR process) satisfy all these conditions.

As a matter of fact, the main concern in kernel-based methods is the choice of the smoothing window h, since it plays a crucial role in the complexity of the estimation. If the bandwidth is chosen to be small, then the bias will be small. However, the variance will increase since we have fewer data points in each small neighborhood. On the other hand, if the bandwidth is large then the variance will decrease, but we will get a larger bias term. Thus it is necessary to trade-off between these two terms in order to get the most appropriate choice for the smoothing window. Solutions to the bandwidth selection problem have been discussed for a long time, e.g. plug-in and cross-validation methods. The Plug-in method is more restrictive, as we might need to know a priori the smoothness of the curve. A more complicated data-driven bandwidth selection is therefore desirable in many situations, especially when the smoothness/roughness of the curve of interest is unknown; see e.g. cross-validation criterions in Silverman [68] and Härdle [40] for implementations. Kernel estimation has also been applied to extract the stochastic intensity of point processes; see e.g. Ramlau-Hansen [63].

Apart from the bandwidth selection problem, kernel-type estimators have a weakness at their domain boundaries. Since kernel functions are normally chosen to be symmetric functions, the estimators will cause biases at the edges of their domain due to the lack of data points. To overcome this problem one can use one-sided kernel functions or reflection methods; see Zhang and Karunamuni [71] and Fan *et al.* [31]. A more efficient solution that is automatically adapted to boundaries is a local linear fit. This method basically fits a linear model to the scatter plot locally around the time point of interest (in contrast to the kernel-based method which fits only a constant to the local area). For a more detailed discussion about how it can reduce the modeling biases at the edges, we refer the reader to Fan and Gijbels [30]. To overcome boundary effects of spot volatility estimation (2.4) a local linear fit has been briefly discussed in Kristensen [53, section 4] by solving the following problem:

$$\min_{b_0,b_1} \sum_{i=1}^{n} K\left(\frac{t_i - t_o}{h}\right) \left\{\tilde{\sigma}^{*2}_{t_i} - \{b_0 + b_1(t_i - t_o)\}\right\}^2,$$

where $\tilde{\sigma}^{*2}_{t_i}$ is an approximation of the spot volatility at time t_i, leading to the local linear estimate of $\sigma^{2*}_{t_o}$ ($t_o \in [0,T]$)

$$\sum_{i=1}^{n} \frac{1}{h} K^*_c\left(\frac{t_i - t_o}{h}\right) \left[\left(X^*_{t_i} - X^*_{t_{i-1}}\right)^2 + o(1)\right], \tag{2.6}$$

where K^*_c is the so-called equivalent kernel function; see Kristensen [53, eq. (9)] and Fan and Gijbels [30, chapter 3].

Instead of local linear fits, we can also apply higher-order polynomial fitting, which fits the scatter plot better and also reduce the modeling bias. However we will need to deal with the higher order of the polynomial. Another advantage of using local polynomial fitting is the ability to estimate higher-order derivatives of parameter curves, which is useful in, for example, the plug-in method (where the second derivative of the curve normally plays a role in the bandwidth size). In constrast to the parametric approach of polynomial regression, where the degree of polynomials is normally large, local polynomial fitting works locally. We will therefore generally require fewer parameters (controlled by the degree of the polynomial). Typically the degree p is $m + 1$, where m is the smoothness order of the parameter curve; see Fan and Gijbels [30]. As can be seen in the literature, standard volatility processes are normally driven by a Wiener process, which has non-differentiable paths, so a local polynomial of degree $p = 1$ (local linear) should work well for the derivation of spot voaltility curves. For a detailed discussion of local polynomial fitting techniques, see Fan *et al.* [29] and Fan and Gijbels [30].

Other nonparametric methods such as Fourier/wavelet transforms and spline smoothing are also popular and can be applied to estimate spot volatility curves. None of these techniques will be applied or discussed further in this work, so we have omitted the detail and refer the reader to Malliavin and Mancino [55] and Mancino and Sanfelici [56] for a comprehensive treatment of Fourier transforms and to Fan and Wang [32] and Schmidt-Hieber [67] for wavelet thresholdings. We note that wavelet-based methods have shown to be nearly minimax (in rate) for a large class of functions when the smoothness level is unknown.

2.4 Microstructure Noise Models

It has been observed in practice that the naive volatility estimators discussed above are not consistent when the data is observed with high frequency, such as every minute or at a finer resolution; see e.g. Bandi and Russell [6]. This means that the model formation for arbitrage-free price processes is disturbed by some factors in the real asset market. The reason of this incompatibility lies in the so-called market microstructure noise, possibly induced by discreteness of prices, bid-ask spreads or price formation, etc. It produces a noisy component in the arbitrage-free price process (2.1), leading to non-robustness of standard volatility estimators when the sampling interval is too small. Therefore many practitioners prefer to investigate data over longer time intervals in order to obtain unbiased results. Using a coarser grid of data, gathered at, say, a 5-minute or 20-minute resolution, makes the naive estimators more robust. For example, given tick-by-tick data of a liquid asset observed over a trading day $Y^*_{t_1}, ..., Y^*_{t_n}$, one would measure the integrated volatility by

$$\sum \left(Y^*_{\tau_i} - Y^*_{\tau_{i-1}} \right)^2, \quad \text{with } \tau_i = t_{i \cdot K} \text{ for } K \in \mathbb{N}, \tag{2.7}$$

where the sum is taken over $0 < \tau_i \leq T$ with, e.g. $K = 300$ (if the trades are assumed to be exercised nearly every second, 5-minute interval is equal to 300). This approach is quite popular in the early empirical finance literature. However, using such coarse data in

order to avoid microstructure effects is unacceptable from a statistical point of view, since it is unsatisfactory to throw away data, especially such a huge amount of information (for example, if prices are observed at every second and we use data at a 5-minute resolution, then in every 5-minute period we will ignore 299 data points in the analysis); see Zhang *et al.* [70]. As a result, many statistical methods dealing with microstructure noise have been discussed to fully exploit the information hidden in the high-frequency data.

Statistically we can view market microstructure noise as an observation error. For the analysis, observable log-price processes are normally assumed to follow an additive noise model

$$Y^*_{t_i} \;=\; X^*_{t_i} + \varepsilon_i \quad \text{for } i = 1, ..., n, \tag{2.8}$$

where ε_i is assumed to be a noise with

$$\mathbb{E}[\varepsilon_i] = \; 0 \quad \text{and} \quad \text{Var}[\varepsilon_i] = \; \omega^2 < \infty;$$

the latent log-price process X^*_t satisfies the model (2.1) (no other assumption on the distribution of noise is made). Independence between the noise ε and the price process X^* is typically assumed for simplicity of the proofs, but it is not essential. For this overview we will consider only time-equidistant data $t_i = iT/n$ for $i = 1, ..., n$ in order to avoid complications.

To clarify why the standard realized volatility based on high-frequency observations $Y^*_{t_i}$ is an inappropriate quantity for the integrated volatility, we look at

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n}\left(Y^*_{t_i} - Y^*_{t_{i-1}}\right)^2\right] &= \mathbb{E}\left[\sum_{i=1}^{n}\left(X^*_{t_i} - X^*_{t_{i-1}}\right)^2\right] + \; \mathbb{E}\left[\sum_{i=1}^{n}\left(\varepsilon_i - \varepsilon_{i-1}\right)^2\right] \\ &= \int_0^T \sigma^{*2}_t dt \; + \; 2n\omega^2.\end{aligned}$$

This expectation explodes as n goes to infinity (we assume that σ^*_t is at least locally bounded), so that the variablility of the noise term dominates the variability of the latent process. Nevertheless, the standard realized volatility can be used to measure the variance of the microstruture noise ω^2; see Zhang *et al.* [70].

To overcome this non-robustness problem, many methods dealing with microstructure noise have been introduced. Starting with the first idea of Zhou [72], the effect of the contamination is now counteracted by adding a bias correction term into the common realized volatility, leading to an unbiased estimate

$$\sum_{i=1}^{n}\left(Y^*_{t_i} - Y^*_{t_{i-1}}\right)^2 - 2\sum_{i=1}^{n}\left(Y^*_{t_i} - Y^*_{t_{i-1}}\right)\left(Y^*_{t_{i-1}} - Y^*_{t_{i-2}}\right).$$

Unfortunately, this estimate remains inconsistent since the variability is still too high; see also Zumbach *et al.* [73] for the comparison of this estimate with other methods in

simulation studies and applications to real foreign exchange data. The intuitive idea of using higher-lag covariance terms is proposed by Barndorff-Nielsen *et al.* [11].

Now we will demonstrate three main approaches that have been extensively used in the analysis of microstructure noise models in practice. The first approach is from the excellent work of Zhang *et al.* [70], which uses different time scales to construct a consistent estimator for integrated volatility. Their combination of sparse and frequent sampling, which is satisfying from an empirical as well as a statistical point of view, is given by

$$\frac{1}{K}\sum_{k=1}^{K}\left\{\sum_{\tau_i^{(k)}\in G_k}\left(Y^*_{\tau_i^{(k)}}-Y^*_{\tau_{i-1}^{(k)}}\right)^2\right\}-\frac{\bar{n}}{n}\sum_{i=1}^{n}\left(Y^*_{t_i}-Y^*_{t_{i-1}}\right)^2, \tag{2.9}$$

where the original set of grid points $\{t_1,...,t_n\}$ is now partitioned into K subgrids $G_k, k = 1,...,K$. Usually G_k is chosen to be $\left\{\tau_1^{(k)},\tau_2^{(k)},...,\tau_{n_k}^{(k)}\right\} := \{t_k, t_{k+K}, t_{k+2K}, ...t_{k+n_kK}\}$ and $\bar{n} = \sum_{k=1}^{K} n_k/K$. The first term is the average over multiple grids of sparsely observed realized volatility (compare with (2.7)), which reduces the bias and variance of the conventional volatility estimator. The second term is a slight modification of the realized volatility, which deals with the rest of the bias induced by the disturbance ε. As discussed in their work, the best result of the combination of these two different time scales can be attained by choosing the parameter K optimally with respect to mean squared error minimization. In particular the asymptotic properties when $K \to \infty$ as $n \to \infty$ for this estimator and related quantities, such as asymptotic variance estimator and the estimator of the noise spread, are provided in their work. The rate of convergence of order $n^{-1/6}$ is derived for this integrated volatility estimator. Furthermore, many ideas of how to construct robust estimators are fruitfully discussed in their work, both from statistical and practical standpoints.

Unfortunately, Zhang *et al.*'s estimator cannot, as one would expect, reach the optimal convergence rate of $n^{-1/4}$ from the parametric maximum likelihood estimator in Gloter and Jacod [36]. Zhang [69] has therefore improved their two-timescale estimator to a multi-timescale estimator which achieves the optimal rate. In fact, this estimator is a direct extension of the best estimator given by Zhang *et al.* [70], combining different, say H, timescales to build a rate-optimal estimator for the integrated volatility.

Using the idea of autocovariances, the second approach is the realized kernel approach proposed by Barndorff-Nielsen *et al.* [11] and has the form:

$$\gamma_0(Y^*)+\sum_{h=1}^{H}K\left(\frac{h-1}{H}\right)\{\gamma_h(Y^*)+\gamma_{-h}(Y^*)\}, \tag{2.10}$$

where

$$\gamma_h(Y^*) := \sum_{j=1}^{n}(Y^*_{t_i}-Y^*_{t_{i-1}})(Y^*_{t_{i-h}}-Y^*_{t_{i-h-1}})$$

is the realized autocovariance with lag h of the price series Y^*; $K(x)$ is a weight function defined on $[0,1]$ with $K(1) = 0$ and the flat-top property $K(0) = 1$ (the flap-top kernel is in fact needed to eliminate the bias caused by the additional noise). This estimate consists of the standard realized volatility (the first term) and the flat-top kernel smoothing terms (the second term); see Barndorff-Nielsen *et al.* [11]. If $H = 0$ the inconsistent volatility estimate, realized volatility, is recovered; if $H = 1$, it is equivalent to Zhou's estimator, which is unbiased but still inconsistent. Thereby one can reduce the variability of the estimation induced by the noise factor by using many lags of autocovariances. In this way one can construct a consistent estimator that achieves the optimal rate of convergence as found in the multi-timescale setting. This approach is applicable to data with endogenous time points, such as transaction data, which make it more useful in real data analysis. It is noteworthy that the extension of the realized kernel with finite lags to the realized kernel with infinite lags can achieve the efficiency bound given in the parametric version of this problem; see Barndorff-Nielsen *et al.* [11, section 4.5]. However, this idea is of limited relevance in practice as we will not observe a sufficient number of returns to construct virtually infinite lags. For implementation and extensive empirical analyses of this method, we refer to Barndorff-Nielsen *et al.* [12]; in particular, the effect of market microstructure noise on real stock prices is demonstrated via realized kernels.

The last technique we demonstrate here is introduced by Podolskij and Vetter [60] and later on extended by Jacod *et al.* [47]. This approach relies on the natural idea of using realized volatility based on local averages $\bar{Y}^*_{t_0}, \bar{Y}^*_{t_1}, ...$, where $\bar{Y}^*_{t_i}$ is defined as the average of, say, H data points $Y^*_{t_i}, ..., Y^*_{t_{i+H-1}}$. By doing this the variance of microstructure noise in the pre-averaged terms can be reduced by a factor of $1/H$. Lastly the bias induced by the added noise is taken care of by a bias-correcting term. This pre-averaging estimator is explicitly given by

$$\widehat{C}^n_T := \frac{1}{H}\frac{1}{\int_0^1 g(u)^2 du}\sum_{j=0}^{n-H}\left(\overline{\triangle Y}^*_{t_j}\right)^2 - \frac{1}{2H}\frac{\int_0^1 g^{(1)}(u)^2 du}{\int_0^1 g(u)^2 du}\sum_{i=1}^{n}\left(Y^*_{t_i} - Y^*_{t_{i-1}}\right)^2 \tag{2.11}$$

where

$$\overline{\triangle Y}^*_{t_j} = \sum_{h=0}^{H-1} g\left(\frac{h}{H}\right)\left(Y^*_{t_{j+h}} - Y^*_{t_{j+h-1}}\right)$$

is the average of the price increments weighted by a function g over the block of size H; g is defined on $[0,1]$. This approach has some features which cannot be seen in the multi-timescale and realized kernel approaches: (i) it enables straightforward construction of consistent estimators for other power variations of the process X^*; in particular the asymptotic variance of the estimator (2.11) can be quantified by such a technique in order to get feasible central limit theorems; (ii) together with the concept of bipower (or multipower) realized volatility, it can be used to detect jumps in the price model and to construct jump-robust estimators; see Podolskij and Vetter [60]. Although the pre-averaging estimate can achieve the optimal rate of convergence $n^{-1/4}$, its asymptotic variance is less efficient than that of the realized kernel. In fact, the realized kernel is shown to have the smallest asymptotic variance among these three approaches. We state the asymptotic normality of

the pre-averaging estimator below for the sake of referring to it later, as our method will be based on this technique.

2.4 Theorem *(Jacod et al. [47], Theorem 3.1) Suppose that assumptions (H) and (K) in [47] hold. Let the weighting function g defined on $[0,1]$ satisfy: g is piecewise continuously differentiable with piecewise Lipschitz derivatives $g^{(1)}$, $g(0) = g(1) = 0$ and $\int_0^1 g(u)^2 du > 0$. With $H = \theta \cdot n^{1/2} + o(n^{1/4})$ we get*

$$n^{1/4}\left\{\widehat{C}_T^n - \int_0^T \sigma_t^{*2} dt\right\} \to \int_0^T \gamma_t \, dB_t, \tag{2.12}$$

with the asymptotic variance

$$\begin{aligned}\gamma_t^2 = \frac{4}{\left(\int_0^1 g(u)^2 du\right)^2} \cdot \Bigg[& \int_0^1 \left(\int_\nu^1 g(u)g(u-\nu)du\right)^2 d\nu \cdot \theta\sigma_t^{*4} \\ & + 2\int_0^1 \left(\int_\nu^1 g^{(1)}(u)g^{(1)}(u-\nu)du\right)\left(\int_\nu^1 g(u)g(u-\nu)du\right) d\nu \cdot \frac{\sigma_t^{*2}\omega^2}{\theta} \\ & + \int_0^1 \left(\int_\nu^1 g^{(1)}(u)g^{(1)}(u-\nu)du\right)^2 d\nu \cdot \frac{\omega^4}{\theta^3}\Bigg],\end{aligned}$$

where the above convergence is in the stably-in-law sense (for more details see e.g. Podolskij and Vetter [61]), and B is another standard Brownian motion, being independent of the original space where σ_t^* lives. Moreover, the asymptotic variance can be consistently estimated by Γ_T^n given in their work [47, eq.(3.7)] which leads to a feasible version of the central limit theorem $n^{1/4}\left\{\widehat{C}_T^n - \int_0^T \sigma^{*2}(t)dt\right\} / \sqrt{\Gamma_T^n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ as a result of stable convergences.

Of course we have not stated all conditions for which the asymptotic properties of these three estimators hold. We emphasize that what they all have in common is that their smoothing parameter H plays a crucial role in the estimation. Some solutions of how to select H in practice are therefore discussed in these and related works. Apart from these three main approaches there are many approaches dealing with market microstructure noise in the literature. Recently, the method given by Bibinger and Reiß [15] and Reiß [64] has gotten more attention, as the asymptotic efficiency of parametric volatility estimation can be achieved. We still rely on the pre-averaging procedure because of the advantages mentioned earlier.

In empirical analysis it has been shown that the noise conditions given in model (2.8) are somewhat too restrictive, since the noise sequence exhibits correlation with/dependence on the efficient price process; see Hansen and Lunde [39] and Kalnina and Linton [52]. Nevertheless, all these methods are able to deal with noise structures more general than i.i.d. noise (however, it is much more complicated; see details in their works). The literature investigates not only the effect of additive noise on the realized volatility, but also noise

with more complex structures such as rounding noise, additive noise plus a rounding effect, or nonlinear market microstructure noise; see e.g. Delattre and Jacod [26], Rosenbaum [66] and Dahlhaus and Neddermeyer [25].

As one might notice, none of the estimators (2.9), (2.10), and (2.11) are necessary non-negative, which contradicts the positivity of the spot volatility. Nevertheless, this is likely to be irrelevant in the empirical analysis of high-frequency data, as it has been shown that the noise is sufficiently small such that these three main estimators give satisfactory results; see also our simulation in Chapter 6. Positive estimators for asymptotic variance based on the subsampling method have been proposed by, e.g. Kalnina [51].

Finally, it is noteworthy that all of these noise-robust estimators work quite well even in the noiseless model, i.e. $Y^*_{t_i} = X^*_{t_i}$ for all $i = 1, ..., n$. As a matter of fact, the variability of the estimates is too high compared with the variability of the standard realized volatility due to the slower rate of convergence, $n^{-1/4}$ versus $n^{-1/2}$. Thus, there is no gain in using such estimates in the absence of noise; in other words, it is better to use standard realized volatility when one knows without doubt that the underlying price process is not contaminated or that the noise is almost negligible.

Chapter 3

Time Change Model and Volatility Decomposition

Instead of the classical semimartingale (2.1) which is normally considered in the usual timescale, called calender time or clock time, we want to look at asset price returns relying on a different time clock. This time clock is controlled by a stochastic process that reflects market activities. In this work our main interest is the investigation of the spot volatility (also called instantaneous volatility) of a time-changed price-model based on trading times. This model is a pure jump process which can be interpreted as an asset price model having price changes at every transaction time. The main contributions of this work are the introduction and theoretical investigation of a new volatility estimator based on the volatility decomposition obtained in this time change model. In this chapter we introduce our model, show the volatility decomposition, and discuss the implications for applications.

3.1 Time Change Model

The study of statistical and empirical relationships between stochastic time clock and price processes, particularly the link between market activities, price fluctuations, and asset returns, has been extensively discussed in the financial literature. There is convincing evidence of a correlation between price volatility and the number/volume of trades. The investigation began when the distribution of asset returns over a short time period, such as a day or shorter, was reported to be non-normal[1]. To recover the normality of price distribution many stochastic time change models have been proposed,[2] in particular, a time-changed Brownian motion

$$dX_t = \sigma(t)\, dW_{\mathcal{T}(t)} \qquad \text{for } t \in [0,T], \tag{3.1}$$

[1] We note that the non-normality of asset returns of the standard diffusion model (2.1) can be directly seen when the volatility process is non-deterministic, in which case price processes are mixed normal.

[2] We briefly define a time change as a (possibly) stochastic non-decreasing function $\mathcal{T} : [0,\infty) \to [0,\infty)$ with $\mathcal{T}(0) = 0$ and $\mathcal{T}(t) \to \infty$ as $t \to \infty$. Thus for an arbitrary function/process f and a time change $\mathcal{T}$, we have $f_{\mathcal{T}(t)} = f \circ \mathcal{T}(t)$, where $\circ$ is the mapping composition.

where $W_\cdot$ is a Brownian motion, $\sigma(t)$ is a stochastic process, and $\mathcal{T}(t)$ is a time change process carrying some market infomation which has an impact on asset prices. For instance, $\mathcal{T}(t)$ could represent the accumulated volume up to the time t or it could be the accumulated number of transactions up to time t as well. The former is suggested by Clark [22] from a theoretical and practical perspective that subordinating a Brownian motion by a directing process—in his work the volume of trades—can achieve the normality of the price distribution. This model is therefore applied to real cotton futures price data to show the normality. Since then the link between market activities—measured by the trading volume and the number of trades—and price fluctuation or volatility has been frequently discussed. Ané and Geman [5] conclude that the main reason of the volatility change is rather the number of trades than their size. In particular they recover the normality of asset returns through this stochastic time change in high-frequency data sampling; see more discussions on this topic in Jones *et al.* [50], Plerou *et al.* [59] and Gabaix *et al.* [34].

In the following, asset returns depend on market activities at each time period in the sense that price processes are controlled by a stochastic time clock given in (3.1). Thus prices evolve slowly if there is not much new information, and they evolve faster if there is a lot of news related to the asset—both company-related and general market news. Now $\sigma^2(t)$ represents the price variation per stochastic time clock (in contrast to σ_t^{*2} in the standard diffusion model (2.1), which represents the price fluctuation in the usual timescale, called calendar-time or clock-time volatility). Having the representation (3.1) for asset returns means that identifying the process X_t is the same as identifying $\sigma(t)$ and $\mathcal{T}(t)$. The leverage effect of the volatility and the price process is now introduced in the correlation between the Brownian motion, the tick-time volatility, and the time change. Since $\mathcal{T}(t)$ need not be continuous, the resulting process $W_{\mathcal{T}(t)}$ is allowed to have jumps.

In this work the stochastic time change $\mathcal{T}_t = N_t$ is a point process representing the accumulated number of trades up to time t. Since N_t is a counting process, W_{N_t} is a pure jump process. The use of this kind of process for modeling asset returns is also supported by the work of Geman *et al.* [35] and many others in the time change model literature. Recently a time-changed Brownian motion has been employed to model credit risk, leading to a solution of the first passage problem for a class of processes related to financial modeling; see Hurd [46]. Indeed, modeling arbitrage-free asset price returns with such a model is reasonable since it has been shown that: (i) under a no-arbitrage assumption, asset prices are semimartingales (see Delbaen and Schachermeyer [27]); and (ii) any semimartingale can be transformed into a time-changed Brownian motion (Monroe [57]). Therefore having a class of time-changed Brownian motions is tractable because it is as large as a class of semimartingales. Particularly the standard Brownian semimartingale $\int_0^t \sigma_s^* dW_s$ can be transformed into $B_{\int_0^t \sigma^{*2} ds}$ where $B_\cdot$ is another Wiener process.[3]

Other time change models have been extensively studied in recent years, in particular a time-changed Lévy process, which allows for a more complex structure in the price models

[3]In contrast to the usual definition of time-changed Brownian motion, we do not have $\mathcal{T}_t$ to be a stopping time with respect to $\mathcal{F}_t$ in order to define $W_{\mathcal{T}_t}$. In our framework N_t is a stopping time with respect to the discrete stopping filtration $\mathcal{F}_{t_i}$.

in order to cope with some stylized effects emerging in the financial market; see Geman *et al.* [35], Carr *et al.* [19] and Carr and Wu [20]. For a statistical treatment of such a model see Belomestny [14] and the reference therein.

Our model:

As mentioned above, instead of classical semimartingales we use a time change model based on trading times, more precisely

$$X_t = \int_0^t \sigma_s \, dW_{N_s} \qquad \text{for } t \in [0,T], \tag{3.2}$$

where the directing process N_t is a point process with intensity λ_t reflecting the accumulated number of transactions up to time t, σ_t^2 is called the volatility per transaction time (also called the volatility per tick or the tick-time volatility), and $W(\cdot)$ is a standard Brownian motion. The simplest case is the case in which σ_t and λ_t are non-random and in which $W(\cdot)$ and $N_\cdot$ are independent; in this case, N_t is a nonhomogeneous Poisson process (NHPP). In a more general setting these processes are stochastic and depend on past realizations of X_t and N_t. The leverage effect between $W(\cdot)$ and $N_\cdot$ is replaced by some martingale structure; in fact, $W(\cdot)$ need not be Gaussian (see below).

Clearly, this model is a pure jump model whose jumps occur at every arrival time or transaction time $t_i := \inf\{t \, : \, N_t \geq i\}$ for $i = 1, 2, ...$, since the counting process N_t is a pure jump process having jumps of size 1. The intensity λ_t serves as an arrival rate of transactions per unit of time. We denote by N_T the number of all transactions on the time horizon $[0,T]$. In (3.2) the process σ_t represents the price fluctuation per unit of stochastic time and not the volatility per calendar time (the usual timescale). Nevertheless, we can define volatility per calendar time for this model as a portion of price variation per unit of time, see (3.3) below. In particular, we will show that by looking at this kind of time change model, clock-time volatility appears to be a multiplication of two factors: tick-time volatility and intensity. This result corresponds to the observation that market activities are strongly related to uncertainty of asset prices. We do not consider a drift term in our analysis as we are interested in price oscillations over short time periods such as a day or shorter, and therefore the drift term has less impact on the analysis.

3.2 Volatility Decomposition

We mainly focus on the estimation of spot volatility for the financial transaction-time model (3.2), which is not the function σ_t^2. Rather, σ_t^2 represents the price variation per transaction time, which is different from the meaning of $\sigma_{clock}^2(t)$ in the classical clock-time diffusion model (say $dX_t = \sigma_{clock}(t)dW_t$). We therefore start with a model-independent definition for spot volatility and clarify below its relation to the tick-time volatility σ_t^2. Let us assume that there exists a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, where $(\mathcal{F}_t)_{t\geq 0}$ is an increasing sequence of σ-fields; roughly speaking, it contains all information available

up to and including time t. For an asset price model $\mathbf{X}(t)$ we define

$$\text{vola}_{\text{t}}^2 := \lim_{\Delta\text{t}\to 0} \frac{\mathbb{E}\big[(\mathbf{X}(\text{t}+\Delta\text{t}) - \mathbf{X}(\text{t}))^2 \;\big|\; \mathcal{F}_{\text{t}}\big]}{\Delta\text{t}}\,. \tag{3.3}$$

It is remarkable that in the classical diffusion model vola_{t}^2 does not depend on the point process, and under the assumption that $\sigma_{clock}(t)$ is a right-continuous process with left-limits adapted to $\mathcal{F}_t$ we have $\text{vola}_{\text{t}}^2 = \sigma_{\text{clock}}^2(\text{t})$.[4] Therefore, we use $\sigma_{clock}(t)$ in this work as a synonym for vola_{t}, i.e. we define

$$\sigma_{clock}^2(t) := \text{vola}_{\text{t}}^2.$$

In the transaction-time model considered in this work, we prove below that $\sigma_{clock}^2(t) = \sigma_t^2 \cdot \lambda_t$. We first present the assumptions for this result. To understand our assumptions, note that in (3.2) we do not use the whole process $W(\cdot)$ but only the increments $U_i := W(N_{t_i}) - W(N_{t_{i-1}})$, which we now assume to be a martingale difference sequence.

3.1 Assumption *The X_{t_i} at observation times t_i follow the model $X_{t_i} = X_{t_{i-1}} + \sigma_{t_i} U_i$, where the t_i are the arrival times of a point process N_t. We assume that there exists a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, where $\mathcal{F}_0$ includes all null sets and the filtration $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous such that*

i) N_t is a point process admitting an $\mathcal{F}_t$-intensity λ_t (as in Definition 2.1; see also Brémaud [16, Definition D7]); in particular λ_t is an $\mathcal{F}_t$-progressive process and N_t is adapted to $\mathcal{F}_t$;

ii) σ_t^2 is a non-negative $\mathcal{F}_t$-predictable process; in particular σ_t^2 is $\mathcal{F}_{t-}$-measurable;

iii) U_i is $\mathcal{F}_{t_i}$-measurable for each i with

$$\mathbb{E}\left[U_i \;\middle|\; \mathcal{F}_{t_i-}\right] = 0 \quad \text{and} \quad \mathbb{E}\left[U_i^2 \;\middle|\; \mathcal{F}_{t_i-}\right] = 1.$$

(Remark that for asymptotic consideration we will need a condition on higher moments of this sequences, see Chapter 4.)

This point process N_t is quite general, as one can see in Section 2.2, since it allows for the dynamics of both processes λ_t and N_t as they react to each other. Under all these assumptions, a leverage effect between all processes—the tick-time volatility, the intensity, the point process and the innovations U_i—can be constructed to account for the correlation between market information, volatility, and asset prices. Note that no other condition on

[4] Since $\mathbb{E}\left[(X_{t+\delta} - X_t)^2 \;\middle|\; \mathcal{F}_t\right] = \mathbb{E}\left[\int_t^{t+\delta} \sigma_{clock}^2(s) dW_s \;\middle|\; \mathcal{F}_t\right] = \mathbb{E}\left[\int_t^{t+\delta} \sigma_{clock}^2(s) ds \;\middle|\; \mathcal{F}_t\right]$, the result follows if $\sigma_{clock}^2(\cdot)$ is right-continuous.

the distribution of U_i is required. The martingale difference structure of U_i is satisfied, for example, for an i.i.d. sequence with mean zero and unit variance.

Examples for processes which fulfill these assumptions are given at the end of this section. The natural filtration which satisfies the above conditions is

$$\mathcal{F}_t = \sigma\left(\{N_s : s \leq t\}, \{\lambda_s : s \leq t\}, \{\sigma_s : s \leq t\}, \{U_{N_s} : s \leq t\}\right).$$

3.2 Proposition *Suppose Assumption 3.1 holds. If σ_t and λ_t are continuous processes we have*

$$\sigma^2_{clock}(t) \;=\; \sigma^2_t \cdot \lambda_t. \tag{3.4}$$

Proof. See Appendix A.

Therefore, in the transaction-time model the volatility can be decomposed into the product of two curves which can both be identified from the data. There exists an intuitive interpretation of this decomposition in that the formula reflects the change of time unit. The formula says that

"volatility per time unit is equal to volatility per transaction multiplied by the average number of transactions per time unit".

This formula clearly explains the model for interaction between market activities and volatility of price, i.e. in a more active business day, reflected by a high activity rate, the volatility for the economy is high.

A highlight of the decomposition is the ability to estimate both curves σ^2_t and λ_t separately by various estimates, say $\widehat{\sigma}^2(t)$ and $\widehat{\lambda}(t)$. We use these estimates in two ways:

i) to construct an alternative estimator of $\sigma^2_{clock}(t)$ via

$$\tilde{\sigma}^2_{clock}(t) \;:=\; \widehat{\sigma}^2(t) \cdot \widehat{\lambda}(t);$$

ii) to look at the two curves individually in order to gain more insight about the cause and structure of volatility.

A well-known result in econometrics is that standard volatility estimators are not appropriate in a noisy model when the data is observed with high frequency. To study the effect of so-called market microstructure noise, an additive noise model is introduced, i.e.

$$Y_{t_i} = X_{t_i} + \varepsilon_i, \qquad \text{for } i = 1, ..., N_T.$$

The contamination ε_i is, for example, an i.i.d. noise with $\mathbb{E}[\varepsilon_i] = 0$ and $\mathrm{Var}[\varepsilon_i] < \infty$. For more details about this model and robust volatility estimates we refer to the next chapter (and also to section 2.4).

Now we want to briefly discuss the implications of our model for applications. For real data analysis (and also for our theoretical results in Chapter 4) we mainly use kernel-type estimators, i.e. an estimator of λ_t at time $t_0 \in (0,T)$ is

$$\widehat{\lambda}(t_0, M) := \frac{1}{M} \sum_{i=1}^{N_T} K\left(\frac{t_i - t_0}{M}\right).$$

In order to handle microstructure noise, the pre-averaging technique of Podolskij and Vetter [60], extended by Jacod *et al.* [47], is applied for the estimation of σ_t^2 (adapted to the present model):

$$\begin{aligned}\widehat{\sigma}^2_{pavg}(t_0, m) &:= \frac{1}{mH}\frac{1}{g_2} \sum_{i=i_0-m}^{i_0+m} k\left(\frac{i-i_0}{m}\right) \left(\overline{\triangle Y}_{t_i}\right)^2 \\ &\quad - \frac{1}{2mH} \frac{\sum_{l=1}^{H-1} h^2(\frac{l}{H})}{g_2} \sum_{i=i_0-m}^{i_0+m} k\left(\frac{i-i_0}{m}\right) \left(Y_{t_i} - Y_{t_{i-1}}\right)^2,\end{aligned}$$

where $i_0 := \inf\{i : t_i \geq t_0\}$. The pre-averaging term is

$$\overline{\triangle Y}_{t_i} := \sum_{l=1}^{H-1} g\left(\frac{l}{H}\right) \left(Y_{t_{i+l}} - Y_{t_{i+l-1}}\right)$$

and $h(l/H) := g((l+1)/H) - g(l/H)$. The bandwidth m is the main smoothing parameter of the estimate, while H is the smoothing parameter in the pre-averaging step (a detailed analysis of this pre-filtering estimator is postponed until the next chapter). This leads to a new alternative clock-time estimator based on the volatility decomposition (3.4), given by

$$\tilde{\sigma}^2_{clock,pavg}(t_0, m, M) := \widehat{\sigma}^2_{pavg}(t_0, m) \cdot \widehat{\lambda}(t_0, M)$$

whereas the "classical" pre-averaging clock-time volatility estimate is

$$\begin{aligned}\widehat{\sigma}^2_{clock,pavg}(t_0, M) &:= \frac{1}{MH}\frac{1}{g_2} \sum_{i=1}^{N_T} K\left(\frac{t_i - t_0}{M}\right) \left(\overline{\triangle Y}_{t_i}\right)^2 \\ &\quad - \frac{1}{2MH} \frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2} \sum_{i=1}^{N_T} K\left(\frac{t_i - t_0}{M}\right) \left(Y_{t_i} - Y_{t_{i-1}}\right)^2.\end{aligned}$$

As can be seen in some empirical analyses, e.g. Zumbach *et al.* [73] and Dahlhaus and Neddermeyer [25], tick-time volatility normally varies in a smoother way than the intensity process. Therefore, we have examined this characteristic in some real data examples. In Figures 3.1–3.2 we have applied both volatility estimates to high-frequency data of transaction prices—approximately 25,000 and 30,000 transactions per day—from the NASDAQ stock exchange (MSFT = Microsoft, GM = General Motors). The first row always shows

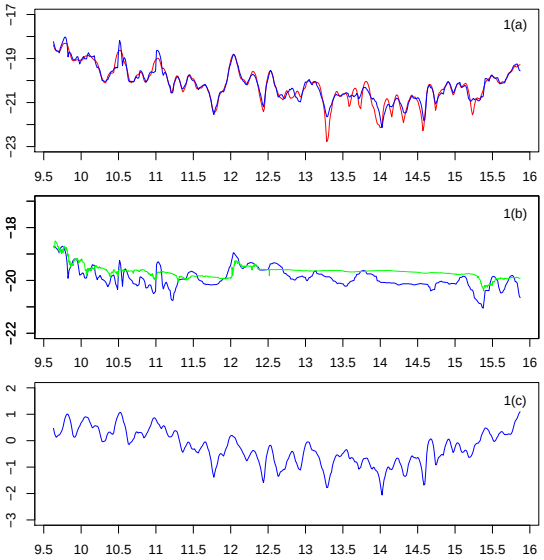


Figure 3.1: Volatility of MSFT on April 1, 2014: Clock-time volatility(a), transaction-time volatility(b), trading intensity(c).

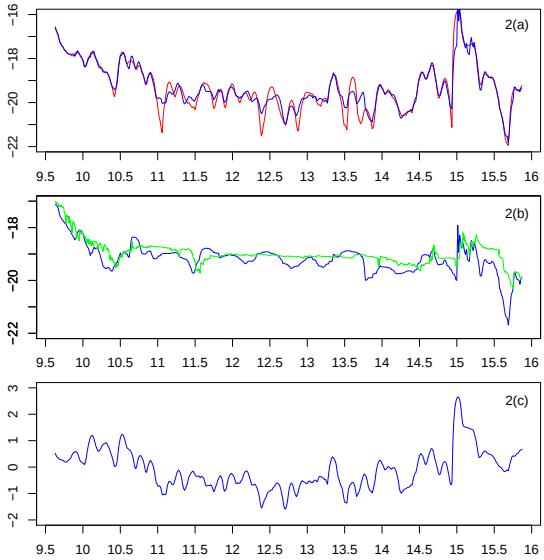


Figure 3.2: Volatility of GM on April 1, 2014: Clock-time volatility(a), transaction-time volatility(b), trading intensity(c).

the logarithm $\log \widehat{\sigma}^2_{clock,pavg}(t,M)$ (red) and $\log \tilde{\sigma}^2_{clock,pavg}(t,m,M)$ (blue), the second row $\log \widehat{\sigma}^2_{pavg}(t,m)$ (blue), and the third row the log trading intensity $\log \widehat{\lambda}(t,M)$ (blue). For these examples we choose $M = 200$, $m = \lfloor 200 \cdot \{\#\text{ of trades}\}/23400 \rfloor$, and $H = 15$; more details about the data, the choice of parameters, and the explanation of the green curves can be found in Chapter 6. Due to the additive relation

$$\log \tilde{\sigma}^2_{clock,pavg}(t,m,M) = \log \widehat{\sigma}^2_{pavg}(t,m) + \log \widehat{\lambda}(t,M)$$

the blue curves in the second and third row sum up to the blue estimator in the first row. From these examples it can be seen that:

i) row (a) shows that the new alternative estimator based on relation (3.4) (blue curve) nicely coincides with the classical clock-time estimator (red);

ii) the tick-time volatility estimator $\log \widehat{\sigma}^2_{pavg}(t,m)$ in row (b) is in general smoother than the clock-time estimator in row (a) and $\log \widehat{\lambda}(t,M)$ in row (c); in particular, the fluctuation of trading intensity in row (c) is the major source of fluctuation of clock-time volatility in row (a);

iii) the decomposition makes it possible to determine to a certain extent the source of volatility changes: for example the peak in 2(a) at 3 PM is mostly due to a peak of trading intensity;

iv) in particular, the curves in 1(a) and 2(a) exhibit the typical U-shape over the trading day. It is notable that this U-shape is mainly a feature of the trading intensity.

In these examples (and more examples in the real data analysis section 6.2) tick-time volatility is considerably smoother than trading intensity. Apart from its practical interpretation, the difference in smoothness of these curves has an important implication for estimation: microstructure noise only affects the smoother curve (b) and not (c). Thus coping with microstructure noise becomes easier in our model with our estimator since we may choose a larger bandwidth with effectively more data than with the classical estimator (red curve in (a)). This is reflected mathematically in a higher rate of convergence of the estimator (see Section 4.5). We mention that in Figures 3.1 and 3.2 we have chosen the same bandwidth for all blue and red estimates for the sake of comparability.

In the following, examples of the time change model based on trading time given in Assumption 3.1 are demonstrated.

3.3 Example *(i) The simplest case is the model in* (3.2) *where σ_t and λ_t are deterministic and $W(\cdot)$ and $N_\cdot$ are independent—i.e. N_t is a NHPP with intensity λ_t and U_i is an i.i.d. normal distributed sequence. Even in this case, the derivation of asymptotic results for the estimates is non-standard since the classical asymptotic setting cannot be applied. We therefore introduce a type of infill asymptotics for this setup in the next chapter.*

(ii) The more general model allowing for stochastic parameters is $dX_t = \sigma_t dL_t$, where L_t is a pure jump process of the form $\sum_{i=1}^{N_t} U_i$. The point process N_t and the sequence of innovations U_i need to satisfy the conditions given in Assumption 3.1, i.e. that N_t has an $\mathcal{F}_t$-intensity λ_t and U_i has a martingale difference structure. The process L_t can be seen as a generalization of a Lévy process without a diffusion part. In fact, a pure jump Lévy process is a compound Poisson process $\sum_{i=1}^{\tilde{N}_t} \tilde{U}_i$ (in this case $\tilde{N}_t$ is a Poisson process and $\tilde{U}_i$ are i.i.d. random variables, whereby Assumption 3.1 is fulfilled).

The leverage effect between all processes can be constructed, for example, by setting $\lambda_t := \alpha_t^2$ with

$$d\alpha_t = a_t dB_t + a_t' dB_t'$$

and

$$d\sigma_t = b_t dB_t + b_t'' dB_t'',$$

where B_t, B_t', and B_t'' are three different $\mathcal{F}_t$-Brownian motions; a_t, a_t', b_t and b_t'' are $\mathcal{F}_t$-adapted processes. The dependence between the price process and the intensity (also the tick-time volatility) lies in the process a_t (also in a_t', b_t and b_t'). For instance, we could model a_t by $da_{t+} = a_t^* \, dX_t$ where a_t^* is an adapted predictable process.

(iii) It is interesting that GARCH-type-models "almost" fit into this framework. For example, let N_t be a point process with right-continuous λ_t satisfying Assumption 3.1, and let

$$\sigma_s^2 := a_0 + a \left\{ X_{t_{j(s)}} - X_{t_{j(s)-1}} \right\}^2 + b\sigma_{t_{j(s)}}^2 \quad \in \mathcal{F}_{s-},$$

where $j(s) := \max\{j : t_j < s\}$ and a_0, a, and b are positive constants. Clearly, this volatility function σ_t^2 is a left-continuous $\mathcal{F}_t$-adapted step function, and therefore predictable.

Since it is not right-continuous we obtain instead of (3.4)*, with the same proof,*

$$\sigma^2_{clock}(t+) = \sigma^2_{t+} \cdot \lambda_t$$

(note that $(\mathcal{F}_t)_{t\geq 0}$ *is right-continuous).*

(iv) In the spirit of the last example we may construct more complex examples where λ_t *is also left-continuous* $\mathcal{F}_t$*-adapted depending both on past arrival times of* $N(t)$ *(e.g. via a structure similar to those in Hawkes-models) and on past log-prices. The volatility* σ_t *may, adding on to the GARCH-structure above, also depend on the intensity of the point process. In that way we could explicitly model the dependence between the price process and the trading intensity.*

Chapter 4

Infill Asymptotics: Spot Volatility Estimation

In order to draw statistical inferences for the transaction-time model (3.2) we introduce a type of infill asymptotics relying on the rescaling method. Our asymptotic arguments are similar to those used in nonparametric regression with time varying coefficients or in modeling locally stationary processes. In this setup a volatility decomposition analogous to (3.4) continues to hold, allowing us to construct an alternative volatility estimator as the product of a tick-time volatility estimator and an intensity estimator. As we will only make assumptions about the smoothness of these parameter functions and not restrict them to any parametric model, our approach is purely nonparametric and is based on a kernel weighting scheme. To handle the effect of market microstructure noise we adapt the so-called pre-averaging procedure for our estimation. A comparison between the estimator based on the classical method and the new estimator based on the factorization of volatility is presented in the last section. It is shown that the alternative estimator outperforms the usual one in many cases, especially when the tick-time volatility is smoother than the clock-time volatility; see Table 4.2.

4.1 Model Setting and Assumptions

In principle, there are two types of asymptotics dealing with statistical inference for asset price models. The first one is the so-called high-frequency framework, in which it is essentially assumed that more and more observations on a fixed time horizon, say $[0, T]$, are obtained. Therefore the time differences between adjacent observations will go to zero; in other words, the frequency of observations tends to infinity. The second one is the low-frequency setup which requires that the time span T go to infinity. For our transaction-time model (3.2) we cannot directly use the first approach relying on the high-frequency assumption since only finite observations over a fixed interval $[0, T]$ are available. For this reason it is necessary to let the time span T go to infinity so as to collect more data to draw inferences. Nevertheless, it is not clear whether increasing T will lead to beneficial asymptotic results, since the parameters σ_t and λ_t vary over time and do not behave in a

periodic way such as would conform to stationarity or ergodicity. Similar problems have been widely discussed in the area of nonparametric regression with time-varying parameters, as by Robinson [65], or later in locally stationary time series as by Dahlhaus [24]. In those works, parameters of interest are rescaled to the unit interval $[0,1]$. For instance, a rescaled version of an AR(1) autoregressive process with a time-varying coefficient is given by $X_{t,T} = a(t/T)X_{t-1,T} + \varepsilon_t$, for $t = 1, ..., T$. In this model, enlarging the sample size T results in the availability of more and more data for estimating local structure, which leads to meaningful asymptotic considerations in this nonstationary AR model; for more details see Dahlhaus [24]. It is worth pointing out that by looking at this rescaled model, only the parameter function $a(\cdot)$, not the process $X_{\cdot,T}$, is observed on a finer and finer grid as $T \to \infty$.

According to the problem of asymptotic construction, we suggest a type of infill asymptotics for this time change model based on trading times. For the sake of simplicity, we assume that log-price processes are of the form

$$dX_{t,T} = \sigma\left(\frac{t}{T}\right)\frac{1}{\sqrt{T}}dW_{N_{t,T}} \quad \text{for } t \in [0,T], \tag{4.1}$$

where $W_\cdot$ is a standard Brownian motion and $N_{t,T}$ is a nonhomogeneous Poisson process (NHPP for short) with a continuous non-negative real-valued intensity function $\lambda(t/T)$; $W_{N_{t,T}}$ is therefore a time-changed Brownian motion. In this model $\sigma^2(\cdot)$ is a real-valued deterministic continuous function called *tick-time volatility*, since it responds to price variation from one trade to the next.

We see that the price process $X_{\cdot,T}$ now depends on the time span T, as the intensity and tick-time volatility depend on T. Our asymptotic inference is indeed drawn by letting the time span T go to infinity in a way similar to the above rescaling idea. That is, the parameters of interest are observed on a finer and finer grid when the time horizon increases. Roughly speaking, we see these parameters as if they were very smooth over small neighborhoods and just slightly changing over time. Throughout this section we will work under the following conditions.

4.1 Assumption *i) The processes $N_{\cdot,T}$ and $W_\cdot$ are independent;*

ii) $\lambda(u)$ and $\sigma^2(u)$ are bounded continuous functions and bounded away from zero uniformly in $u \in [0,1]$.

As our main objective is to find a nonparametric method for deriving the spot volatility of the price process, we must, as is usually done in nonparametric estimation, impose some smoothness restrictions on these parameter curves to ensure that they will change slowly in some sense over time (later on we will require stronger conditions in order to provide limit distributions). The independence between $N_{\cdot,T}$ and $W_\cdot$ enables us to make arguments conditional on the entire process $N_{\cdot,T}$ which greatly simplifies the proofs. In particular, for

consecutive arrival times t_{j-1} and t_j we have

$$X_{t_j,T} - X_{t_{j-1},T} \overset{\text{law}}{=} \sigma\left(\frac{t_j}{T}\right)\frac{1}{\sqrt{T}}U_j, \tag{4.2}$$

since the time-changed Brownian motion $W_{N_{t,T}}$ has the same law as $\sum_{i=1}^{N_{t,T}} U_i$,[1] where U_i are i.i.d. normally distributed variables with zero mean and unit variance and independent of $N_{\cdot,T}$.

A leverage effect between $W_\cdot$ and $N_{\cdot,T}$ could be allowed for in order to explain the correlation between market activities and price processes; see the formation of general models in Assumptions 3.1. A thoroughly asymptotic investigation of this complex transaction-time model under those assumptions is beyond the scope of this work; nevertheless some asymptotic properties are discussed in Section 5.3.

4.2 Remark *i) The point process $N_{t,T}$ can also be characterized by its corresponding arrival times $t_i = \inf\{t \,:\, N_{t,T} \geq i\}$ for $i \in \mathbb{N}$. We denote the number of all arrivals on $[0,T]$ by $N_T := N_{T,T}$. So as to not complicate the notation, we will avoid using the double subscript $t_{i,T}$, although the t_i will always depend on the time span T.*

ii) The model (4.1) can be directly extended to the more general case of independent stochastic processes $\sigma(t/T)$ and $\lambda(t/T)$, where these are also independent of $N_{t,T}$ and W_t. In this situation $N_{t,T}$ becomes a doubly stochastic Poisson process or Cox process. The proofs remain the same except that we might have to change some arguments presented in Appendix B.1.–B.4. from unconditional versions to conditional ones (conditional on $\sigma(\cdot)$ and $\lambda(\cdot)$). A more general model allowing for a form of endogeneity is discussed later.

iii) The square root factor $1/\sqrt{T}$ in (4.1) is necessary for keeping the variance of log-price increments small as the sample size grows, i.e. $\mathbb{E}\left[\left(X_{t_i,T} - X_{t_{i-1},T}\right)^2\right] = O(1/T)$. Otherwise the volatility may be too large and could explode when one considers the integrated volatility over a period of time. This is similar to the classical semimartingale (2.1) in the high-frequency approach, where $\mathbb{E}\left[\left(X^*_{i/n} - X^*_{(i-1)/n}\right)^2\right] = O(1/n)$ if σ^* is bounded, where n is the number of subdivisions on $[0,T]$. Therefore our framework based on the rescaling method is a kind of infill asymptotics.

[1] Considering the characteristic functions of $W_{N_{t,T}}$ and $\sum_{i=1}^{N_{t,T}} U_i$, we have, for $\theta \in \mathbb{R}$,

$$\begin{aligned}\mathbb{E}\left[e^{i\theta\sum_{i=1}^{N_{t,T}} U_i}\right] &= \mathbb{E}\left[\prod_{i=1}^{N_{t,T}} \mathbb{E}\left[e^{i\theta U_i} \;\middle|\; N_{\cdot,T}\right]\right] = \mathbb{E}\left[\exp\left(-\frac{1}{2}\theta^2 N_{t,T}\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[e^{i\theta W_{N_{t,T}}} \;\middle|\; N_{t,T}\right]\right] = \mathbb{E}\left[e^{i\theta W_{N_{t,T}}}\right].\end{aligned}$$

4.3 Proposition *Suppose that $X_{t,T}$ satisfies the model* (4.1)*. The variance of the price increment over $[t_o - bT, t_o + bT]$ is given by*

$$\mathbb{E}\left[(X_{t_o+bT,T} - X_{t_o-bT,T})^2\right] = \int_{u_o-b}^{u_o+b} \sigma^2(u)\lambda(u)du$$

for $t_o \in (0,T)$ (denote $u_o := t_o/T$) and a small window size $b > 0$. Furthermore, if b shrinks to zero we get

$$\frac{1}{2b}\mathbb{E}\left[(X_{t_o+bT,T} - X_{t_o-bT,T})^2\right] \;\to\; \sigma^2(u_o)\cdot\lambda(u_o). \tag{4.3}$$

Proof. See Appendix B.1.

We denote the clock-time volatility (or volatility per calendar time) by

$$\sigma^2_{clock}(\cdot) := \; \sigma^2(\cdot)\lambda(\cdot).$$

This relation shows that a version of the volatility decomposition (3.4) also holds in this rescaled setting. Due to this factorization, all statistics estimating the left-hand side of (4.3) or each factor of the product on the right-hand side are good candidate estimators for the volatility of the model. Throughout this work a nonparametric approach based on a kernel-weighting scheme is applied to construct volatility estimators. It is notable that the clock-time volatility in the sense of vola_t^2, see (3.3), is indeed given by $\sigma^2(t/T)\lambda(t/T)/T$.

As mentioned earlier, it is natural to impose a smoothness condition on the parameter curves in order to derive asymptotic properties, in particular for the derivation of limit distributions. Therefore we will need the following definition called Hölder continuity, cf. Kristensen [53].

4.4 Definition *An m-times differentiable function $f : [0,1] \to \mathbb{R}$ is said to lie in the class $\mathcal{C}^{m,\gamma}[0,1]$, for $0 < \gamma < 1$ and $m \geq 0$, if*

$$|f^{(m)}(\mu+\delta) - f^{(m)}(\mu)| \;\leq\; C\cdot|\delta|^\gamma, \qquad \text{for } |\delta| \to 0 \text{ and a constant } C.$$

Many common parametric volatility models in the literature, such as the Heston, GARCH, CIR, or Chen model, etc., are stochastic processes driven by a Brownian motion. Hence their realizations, being non-differentiable, lie in this class with $m = 0$ and $\gamma < 1/2$; see e.g. Durrett [28]. In our setting the volatility curve $\sigma^2_{clock}(\cdot)$ acts like the one that is driven by a Brownian motion whenever either of the components of the decomposition (4.3) lies in the Hölder class $\mathcal{C}^{0,\gamma}[0,1]$ with $\gamma < 1/2$, while the other component is allowed to be smoother (as a consequence, one may choose larger bandwidth for estimating this curve). Recently other parametric volatility models relying on a fractional Brownian motion have also been applied in price modeling (if the Hurst index H of the fractional Brownian motion is equal to $1/2$, the process becomes a classical Brownian motion). In fact, almost-all trajectories of this process with Hurst index $H \in (0,1)$ lie in the Hölder

class with $m = 0$ and $\gamma < H$. From this viewpoint the Hölder smoothness class seems to be an appropriate choice of smoothness class for our parameter curves.

In the standard diffusion model (2.1), a specific assumption on the structure of the volatility processes σ_t^{*2} is normally made to deal with the approximation error; more precisely, it is assumed to be driven by another Brownian semimartingale, say $d\sigma_t^* = a_t dt + b_t dW_t + \nu_t dV_t$; see Barndorff-Nielsen *et al.* [11] and Jacod *et al.* [47] among others. We here give analogous assumptions for the parameter curves in order to control approximation errors:

4.5 Assumption *i) $\lambda(\cdot)$ lies in $\mathcal{C}^{m',\gamma'}[0,1]$ for $m' = 0, 1, 2$ and $0 < \gamma' < 1$ and is bounded away from zero uniformly in u.*

ii) $\sigma^2(\cdot)$ also lies in $\mathcal{C}^{m,\gamma}[0,1]$ for $m = 0, 1, 2$ and $0 < \gamma < 1$ and is bounded away from zero uniformly in u.

4.2 Transaction Rate Estimator

Now we deal with a rescaled version of an NHPP $N_{t,T}$ for $t \in [0,T]$ that has an intensity function $\lambda(t/T)$. In our application, it represents the accumulated number of trades from a starting time point (market opening time 9:30 AM) up to the time t. A generalization of such processes allowing for stochastic intensity has already been discussed in Section 2.2, but not for the rescaled setting. Formally, we let $\{N_{t,T}\}_{t\in[0,T]}$ be an NHPP with respect to the natural filtration $\left\{\mathcal{F}_{t,T}^N\right\}_{t\in[0,T]}$, $\mathcal{F}_{t,T}^N = \sigma\left(N_{s,T} : s \le t\right)$. The intensity function $\lambda(\cdot)$ is assumed to be a continuous non-negative real-valued function, i.e.

i) $N_{0,T} = 0$ a.s.,

ii) $N_{t,T} - N_{s,T}$ is independent of $\mathcal{F}_{s,T}^N$, for all $s \le t \le T$, and

iii) $\mathbb{P}\left(N_{t,T} - N_{s,T} = k\right) = \exp\left\{-\int_s^t \lambda\left(\frac{l}{T}\right) dl\right\} \cdot \frac{\left(\int_s^t \lambda\left(\frac{l}{T}\right) dl\right)^k}{k!}$, for $k \in \mathbb{N}$, $s \le t \le T$.

Under the boundedness assumption given above we get

$$\mathbb{E}\left[N_{t,T}\right] = \int_0^t \lambda\left(\frac{s}{T}\right) ds \;=\; T \cdot \int_0^u \lambda(\nu) d\nu \;< \infty,$$

where we denote by $u = t/T$ a time point in the rescaled interval $[0,1]$. This also leads to $N_t < \infty$ a.s., and therefore non-explosive.

In fact, the properties and proofs stated below are related to martingale theory, even though one could obtain the same results without using it (since our intensity process is deterministic), but the calculations could be long or cumbersome. Another benefit of using martingale dynamics in our setting is that we can easily extend the model to

the more general case of stochastic intensity models, which are more appropriate in real applications since the rate of arrival varies by time stochastically and simultaneously with the point process. For more details on stochastic-intensity processes we refer the reader to Brémaud [16] and Aalen [1] as well as to Section 2.2.

4.6 Remark *i) Obviously $N_{t,T}$ is not a martingale on $[0,T]$ since its expectation varies by time. Nonetheless we can define a compensated Poisson process*

$$M_{t,T} \;:=\; N_{t,T} - \int_0^t \lambda(s/T)ds,$$

which is now a martingale. The compensator $< M >_{t,T}$ of $M_{t,T}^2$ is given by

$$< M >_{t,T} = \int_0^t \lambda\left(\frac{s}{T}\right) ds.$$

Note that the compensator occurs in Doob-Meyer's decomposition of the submartingale $M_{t,T}^2$ such that $M_{t,T}^2 = \tilde{M}_{t,T} + < M >_{t,T}$ where $\tilde{M}_{t,T}$ is another martingale with respect to $\mathcal{F}_{t,T}^N$. See also Remark 2.2 vi).

ii) It is possible to define a stochastic integral $\int_0^t c_{s,T}\, dM_{s,T}$, where $c_{t,T}$ is a predictable process. In addition, if $\mathbb{E}\left[\int_0^T |c_{s,T}|\lambda(s/T)ds\right] < \infty$, then $\int_0^t c_{s,T} dM_{s,T}$ is a martingale; see Brémaud [16, II.3. Th. T8]. Indeed, for $0 \le s \le t$,

$$\mathbb{E}\left[\int_s^t c_{l,T} dN_{l,T}\right] = \mathbb{E}\left[\int_s^t c_{t,T}\lambda(l/T)dl\right] \overset{(*)}{=} \int_s^t c_{l,T}\lambda(l/T)dl.$$

The last equality $(*)$ holds if $c_{l,T}$ is non-random.

iii) Under the boundedness assumption 4.1 ii) we get

$$0 < \;c_1 \le\; \mathbb{E}\,[t_i - t_{i-1}] \le\; c_2 < \infty,$$

for all $i \in \mathbb{N}$ with constants c_1, c_2, since $1 = \mathbb{E}\left[\int_{t_{i-1}}^{t_i} dN_{t,T}\right] = \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \lambda(t/T)dt\right]$ by the optional sampling theorem.

For the estimation of intensity, we employ a kernel method similar to that used in density estimation. Given a set of observed arrival times (in our application transaction times) $\{t_1, t_2, ..., t_{N_T}\}$ on $(0,T]$, the intensity $\lambda(u_o)$ for $u_o \in (0,1)$ is measured by

$$\widehat{\lambda}(u_o) := \;\frac{1}{\mathfrak{b}T}\sum_{i=1}^{N_T} \mathfrak{K}\left(\frac{t_i - u_o T}{\mathfrak{b}T}\right) \;=\; \frac{1}{\mathfrak{b}T}\int_0^T \mathfrak{K}\left(\frac{t - u_o T}{\mathfrak{b}T}\right) dN_{t,T}, \tag{4.4}$$

where the bandwidth $\mathfrak{b} = \mathfrak{b}(T)$ depending on the time span T satisfies $\mathfrak{b} \to 0$ and $\mathfrak{b}T \to \infty$ as $T \to \infty$. The kernel function $\mathfrak{K}$ (as well as K and k, to be introduced later) fulfills the following conditions:

Condition (K) The kernel function $\mathfrak{K} : \mathbb{R} \to \mathbb{R}^+$ is a continuous, symmetric function such that $\mathfrak{K}(x) = 0$ for $|x| \geq 1$ and $\int_{\mathbb{R}} \mathfrak{K}(x)dx = 1$.

In fact, the kernel function K need not be compactly supported as long as it decays fast enough in order to eliminate the bias effect of remote points; this restriction, however, simplifies the proofs. In the following, asymptotic properties for the intensity estimate are given.

4.7 Theorem *Under Assumption 4.1 we get*

$$\widehat{\lambda}(u_o) \xrightarrow{\mathbb{P}} \lambda(u_o)$$

and

$$\sqrt{\mathfrak{b}T}\left(\widehat{\lambda}(u_o) - \mathbb{E}\widehat{\lambda}(u_o)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \lambda(u_o)\int_{\mathbb{R}} \mathfrak{K}^2(x)dx\right)$$

as $T \to \infty$, for $u_o \in (0,1)$. Moreover if $\lambda(\cdot)$ fulfills Assumption 4.5 i) with $\mathfrak{b}^{2(m'+\gamma')+1}T = o(1)$, then

$$\sqrt{\mathfrak{b}T}\left(\mathbb{E}\widehat{\lambda}(u_o) - \lambda(u_o) - \frac{\mathfrak{b}^2}{2}\lambda^{(2)}(u_o)\int_{\mathbb{R}} x^2\mathfrak{K}(x)dx \cdot I_{\{m'=2\}}\right) = o_p(1),$$

particularly

$$\sqrt{\mathfrak{b}T}\left(\widehat{\lambda}(u_o) - \lambda(u_o) - \frac{\mathfrak{b}^2}{2}\lambda^{(2)}(u_o)\int_{\mathbb{R}} x^2\mathfrak{K}(x)dx \cdot I_{\{m'=2\}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \lambda(u_o)\int_{\mathbb{R}} \mathfrak{K}^2(x)dx\right).$$

Proof. See Appendix B.2.

4.8 Remark *i) The bandwidth condition $\mathfrak{b}^{2(m'+\gamma')+1}T \to 0$ is essential for controlling the asymptotic bias of the estimation. It therefore has a direct impact on the rate of convergence: the smoother the intensity function is, the faster the obtainable rate of convergence.*

ii) We can also apply this estimator to extract the arrival rate of models with random intensity, such as Cox processes or even more general stochastic intensity models. For such cases, martingale dynamics will come into play; see e.g. Ramlau-Hansen [63]. He applies a kernel-based method to obtain an estimator for the time-varying stochastic intensity in the multiplicative intensity model.

4.3 Spot Volatility Estimators: Clock Time vs. Tick Time

To nonparametrically quantify the spot volatility of our time-changed model based on trading times, we adopt the same concept of using quadratic increments as in the classical semimartingale. More precisely, a kernel-filtering of the sum of squared increments will be

applied in this section; compare with (2.4). As a result of the volatility decomposition, we propose two distinct estimators for the spot volatility: the first one is similar to the classical estimator in the standard diffusion model (adapted to our rescaled transaction time model) and the second one is based on the product of the tick-time volatility and intensity estimators. At the end of this section we will give a comparison between these two estimators.

4.3.1 Standard Volatility Estimator

Suppose that asset log-returns satisfy the model (4.1). We reiterate that $\sigma^2_{clock}(u_o)$ is the product of $\sigma^2(u_o)$ and $\lambda(u_o)$. Under a given set of observed transaction data $(t_1, X_{t_1,T}), ...,$ $(t_{N_T}, X_{t_{N_T},T})$, we estimate $\sigma^2_{clock}(u_o)$ by

$$\widehat{\sigma}^2_{clock}(u_o) = \sum_{i=1}^{N_T} \frac{1}{b} K\left(\frac{t_i - u_o T}{bT}\right) (X_{t_i,T} - X_{t_{i-1},T})^2 \tag{4.5}$$

for $u_o \in (0,1)$. The bandwidth b depends on the time span T with $b \to 0$ and $bT \to \infty$ as $T \to \infty$, and the kernel function K satisfies condition (K). Obviously b and K may be different from those of the preceding section. In the beginning, we have assumed that the volatility functions are smooth enough that we can proceed as though they were constant on very small segments. Therefore only the data whose arrival times lie in the vicinity of u_o, i.e. in $[u_o - b, u_o + b]$ for a small window size b, will contribute to the estimation of $\sigma^2_{clock}(u_o)$.

Corresponding to the independence between $N_{\cdot,T}$ and $W_\cdot$, it is helpful to look at the following conditional expressions:

$$\begin{aligned}
\mathbb{E}\left[\widehat{\sigma}^2_{clock}(u_o) \;\middle|\; N_{\cdot,T}\right] &= \sum_{i=1}^{N_T} \frac{1}{bT} K\left(\frac{t_i - u_o T}{bT}\right) \sigma^2\left(\frac{t_i}{T}\right) \\
&= \int_0^T \frac{1}{bT} K\left(\frac{t - u_o T}{bT}\right) \sigma^2\left(\frac{t}{T}\right) dN_{\cdot,T},
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}\left[\widehat{\sigma}^2_{clock}(u_o) \;\middle|\; N_{\cdot,T}\right] &= 2\sum_{i=1}^{N_T} \frac{1}{b^2T^2} K^2\left(\frac{t_i - u_o T}{bT}\right) \sigma^4\left(\frac{t_i}{T}\right) \\
&= 2\int_0^T \frac{1}{b^2T^2} K^2\left(\frac{t - u_o T}{bT}\right) \sigma^4\left(\frac{t}{T}\right) dN_{\cdot,T}.
\end{aligned}$$

4.9 Proposition *Under Assumption 4.1, we obtain*

$$\widehat{\sigma}^2_{clock}(u_o) \xrightarrow{\mathbb{P}} \sigma^2_{clock}(u_o),$$

and

$$\sqrt{bT}\left\{\frac{\widehat{\sigma}^2_{clock}(u_o)-\mathbb{E}\left[\widehat{\sigma}^2_{clock}(u_o)\;\middle|\;N_{\cdot,T}\right]}{\sqrt{2\sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}}K^2(x)dx}}\right\}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)$$

as $T\to\infty$, for $u_o\in(0,1)$. The limiting is independent of the Poisson process $N_{\cdot,T}$.

Proof. see Appendix B.3.

4.10 Theorem *Suppose that Assumptions 4.1 and 4.5 are fulfilled, and $b^{2\alpha+1}T=o(1)$ with $\alpha=\min\{m+\gamma,m'+\gamma'\}$. Then the following asymptotic normality holds:*

$$\sqrt{bT}\left\{\widehat{\sigma}^2_{clock}(u_o)-\sigma^2_{clock}(u_o)-BIAS\right\}\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\;3\sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}}K^2(x)dx\right)\tag{4.6}$$

as $T\to\infty$, for $u_o\in(0,1)$. The bias term is given by

$$BIAS=\frac{1}{2}\left(\sigma^2_{clock}(u_o)\right)^{(2)}b^2\int_{\mathbb{R}}x^2K(x)dx\cdot I_{\{m=m'=2\}}.$$

Proof. see Appendix B.3.

4.11 Remark *i) Many techniques for deriving spot volatility have been proposed in the literature of standard diffusion models. Our method is similar to those of Fan and Wang [33] and Kristensen [53] based on the kernel filtering of integrated volatility. It is clear that our asymptotic results based on transaction-time sampling are different from their results; in particular the asymptotic bias and variance derived above rely on the transaction intensity; compare* (4.6) *to* (2.5).

ii) Spot volatility estimators have more to offer than the usual integrated volatility estimators in that they allow for the construction of many consistent estimators for functionals of spot volatility (including the integrated volatility) by using the continuous-mapping theorem. For instance, the unknown component in the asymptotic variance $\sigma^4(u_o)\lambda(u_o)$ can be clearly estimated by the square of $\widehat{\sigma}^2_{clock}(u_o)$ divided by $\widehat{\lambda}(u_o)$. Alternatively, one could construct a consistent estimator for the asymptotic variance by applying the so-called kernel-weighted quarticity, as in the classical semimartingale (see Appendix B.3.):

$$KQ(u_o):=\sum_{i=1}^{N_T}\frac{T}{3b}K\left(\frac{t_i-u_oT}{bT}\right)(X_{t_i,T}-X_{t_{i-1,T}})^4.$$

This implies that

$$\sqrt{bT}\frac{\left\{\widehat{\sigma}^2_{clock}(u_o)-\sigma^2_{clock}(u_o)-\widehat{BIAS}\right\}}{\sqrt{3\cdot KQ(u_o)\int_{\mathbb{R}}K^2(x)dx}}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1),$$

which is a feasible version of the above asymptotic normality, where $\widehat{BIAS}$ is an estimator of the bias term.

4.3.2 Alternative Volatility Estimator

Looking at statistics in clock time or transaction time is an important issue. For example, a high-volatility of a liquid asset could be the effect of the transaction-time volatility or that of the intensity, as can be seen in our volatility factorization formula (4.3). Therefore it is particularly interesting to study the cause and the structure of volatility by evaluating these two factors separately. Since an estimator for the intensity has already been introduced in Section 4.2, we only suggest an estimator for $\sigma^2(u_o)$, which is also based on the kernel-weighting scheme:

$$\begin{aligned}\widehat{\sigma}^2(u_o) &= \frac{T}{H_{1,N}} \sum_{j=i_o-N}^{i_o+N} k\left(\frac{j-i_o}{N}\right) \left(X_{t_j,T} - X_{t_{j-1},T}\right)^2 \\ &= \frac{T}{H_{1,N}} \sum_{j=-N}^{N} k\left(\frac{j}{N}\right) \left(X_{t_{i_o-j},T} - X_{t_{i_o-j-1},T}\right)^2, \end{aligned} \tag{4.7}$$

where $i_o := \inf\{i : t_i \geq u_o T\}$ for $u_o \in (0,1)$, i.e. t_{i_o} is the first arrival time after or at the time point of interest t_o ($u_o = t_o/T$). The second representation of the estimate (4.7) has the advantage that no random elements appear inside the kernel function, which simplifies the proofs. We here define the normalizing factor $H_{n,N} := \sum_{j=-N}^{N} k^n(j/N)$, where the segment length $N = N(T)$ with $N \to \infty$ and $N/T \to 0$ as $T \to \infty$. The existence of the limits $H_{n,N}/N \to \int_{\mathbb{R}} k^n(x)dx$ is assumed for all $n \in \mathbb{N}$. The kernel function k also satisfies Condition (K), i.e. k is a symmetric, smooth, and bounded function with $k(x) = 0$ for $|x| \geq 1$, and $\int_{\mathbb{R}} k(x)dx = 1$.

At first sight, both estimates $\widehat{\sigma}^2_{clock}(\cdot)$ and $\widehat{\sigma}^2(\cdot)$ look very similar, as they are based on the filtering of sums of squares of increments. However, there is a distinction between these two: one is based on tick time and the other is based on clock time. More precisely, in (4.7) the (exactly) N-nearest observed increments from both sides of the considered time point $u_o = t_o/T$ are taken into account so that the influence of the arrival rate is removed, while K in (4.5) uses all increments inside the interval $[t_o - bT, t_o + bT]$. As a matter of fact, the number of transactions/increments over this interval is random and depends on the trading intensity. This is an important concern in empirical analysis since the bandwidth for the volatility estimation of low- and high- liquid stock data should be chosen according to the number of transactions of each stock and trading day.

Likewise, by the independence between $N_{\cdot,T}$ and $W_\cdot$, we first derive the conditional expectation and variance:

$$\mathbb{E}\left[\widehat{\sigma}^2(u_o) \;\middle|\; N_{\cdot,T}\right] = \frac{1}{H_{1,N}} \sum_{j=-N}^{N} k\left(\frac{j}{N}\right) \sigma^2\left(\frac{t_{i_o-j}}{T}\right) \quad \text{and}$$

$$\mathrm{Var}\left[\widehat{\sigma}^2(u_o) \;\middle|\; N_{\cdot,T}\right] = \frac{2}{H^2_{1,N}} \sum_{j=-N}^{N} k^2\left(\frac{j}{N}\right) \sigma^4\left(\frac{t_{i_o-j}}{T}\right).$$

4.12 Proposition *Suppose that Assumption 4.1 is fulfilled. Then*

$$\sqrt{N}\left\{\frac{\widehat{\sigma}^2(u_o)-\mathbb{E}\left[\widehat{\sigma}^2(u_o)\;\middle|\;N_{\cdot,T}\right]}{\sqrt{2\sigma^4(u_o)\int_{\mathbb{R}}k^2(x)dx}}\right\}\xrightarrow{\mathcal{D}}\;\mathcal{N}(0,1),$$

as $T\to\infty$, for $u_o\in(0,1)$, where the normal limiting is independent of the NHPP $N_{\cdot,T}$.

Proof. See Appendix B.3.

It is remarkable that the bias derivation for this estimate is much more complicated than that of $\widehat{\sigma}^2_{clock}(\cdot)$ in the case of higher orders of smoothness, i.e. $\min(m,m')\geq 1$. Specifically, the explicit bias term $BIAS$ in Theorem 4.10 cannot be explicitly stated, even though the kernel function is symmetric and the parameter functions are twice differentiable ($m=m'=2$). This weakness leads to a reduction in the rate of convergence[2] in many cases, particularly when $\min(m,m')\geq 1$ (compare the rate of convergence in Theorem 4.10 with that of Theorem 4.13). Some preliminary results for proving the next theorem are given in Lemma 1 and Corollary 1 in Appendix B.3. The segment conditions (4.8) are given to enable us to neglect some asymptotic bias terms.

4.13 Theorem *Let Assumptions 4.1 and 4.5 hold, i.e. let the volatility per tick-time unit $\sigma^2(\cdot)$ and the intensity $\lambda(\cdot)$ lie in $\mathcal{C}^{m,\gamma}[0,1]$ and $\mathcal{C}^{m',\gamma'}[0,1]$, respectively, for $m,m'=0,1,2$ and $0<\gamma,\gamma'<1$. This implies that*

$$\sqrt{N}\left(\frac{1}{H_{1,N}}\sum_{j=-N}^{N}k\left(\frac{j}{N}\right)\sigma^2\left(\frac{t_{i_o-j}}{T}\right)-\sigma^2(u_o)\right)=\;o_p(1)$$

as

$$\begin{aligned}
N^{1+2\gamma}/T^{2\gamma}&\to 0 \quad \text{for } m=0,\\
N^{3+\gamma^*}/T^{2+\gamma^*}&\to 0 \quad \text{for } m=1 \text{ and } m'=0,\\
N^{3+\gamma}/T^{2+\gamma}&\to 0 \quad \text{for } m=1 \text{ and } m'=1,2,\\
N^{3+\gamma'}/T^{2+\gamma'}&\to 0 \quad \text{for } m=2 \text{ and } m'=0,\\
N^{4}/T^{3}&\to 0 \quad \text{for } m=2 \text{ and } m'=1,2,
\end{aligned}\tag{4.8}$$

with $\gamma^*=\min\{\gamma,\gamma'\}$. Furthermore,

$$\sqrt{N}\left\{\widehat{\sigma}^2(u_o)-\sigma^2(u_o)\right\}\xrightarrow{\mathcal{D}}\;\mathcal{N}\left(0,\;2\sigma^4(u_o)\int_{\mathbb{R}}k^2(x)dx\right)\tag{4.9}$$

as $T\to\infty$, for $u_o\in(0,1)$, where the limit distribution is independent of the point process $N_{\cdot,T}$.

[2]More precisely, $(I_{4,2,1})$ and $(I_{4,2,2})$ in the proof of Theorem 4.13 only vanish if $N^4/T^3\to 0$. This is indeed due to the randomness of time sampling.

Proof. See Appendix B.3.

With respect to the volatility decomposition (4.3), we can construct a new measure for the spot volatility by multiplying $\widehat{\sigma}^2(u_o)$ and $\widehat{\lambda}(u_o)$ together, i.e.

$$\tilde{\sigma}^2_{clock}(u_o) := \widehat{\sigma}^2(u_o) \cdot \widehat{\lambda}(u_o). \tag{4.10}$$

The consistency of the product is evident, as each estimate is consistent, but its asymptotic normality still needs to be examined. We see that the rate of convergence for $\tilde{\sigma}^2_{clock}(\cdot)$ depends on the relation between the convergence rates $\mathfrak{b}T$ and N in Theorems 4.7 and 4.13, respectively. For example, if $N = o(\mathfrak{b}T)$ it is reasonable to look at

$$\begin{aligned}
&\sqrt{N}\left\{\tilde{\sigma}^2_{clock}(u_o) - \sigma^2_{clock}(u_o)\right\} \\
&= \widehat{\lambda}(u_o)\sqrt{N}\left\{\widehat{\sigma}^2(u_o) - \sigma^2(u_o)\right\} \;+\; \sigma^2(u_o)\frac{\sqrt{N}}{\sqrt{\mathfrak{b}T}}\sqrt{\mathfrak{b}T}\left\{\widehat{\lambda}(u_o) - \lambda(u_o) - BIAS_\lambda\right\} \\
&\quad +\; \sigma^2(u_o)\sqrt{N}\cdot BIAS_\lambda \\
&= \widehat{\lambda}(u_o)\sqrt{N}\left\{\widehat{\sigma}^2(u_o) - \sigma^2(u_o)\right\} \;+\; \sigma^2(u_o)\sqrt{N}\cdot BIAS_\lambda \;+\; o(1),
\end{aligned}$$

with

$$BIAS_\lambda = \frac{\mathfrak{b}^2}{2}\lambda^{(2)}(u_o)\int_{\mathbb{R}} x^2\mathfrak{K}(x)dx \cdot I_{\{m'=2\}},$$

and therefore the limit distribution is mainly given by the last theorem.

4.14 Theorem *Let all assumptions in Theorem 4.13 be satisfied. We get*

$$\sqrt{N}\left\{\tilde{\sigma}^2_{clock}(u_o) - \sigma^2_{clock}(u_o) - \tilde{BIAS}\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, V^2) \tag{4.11}$$

with (assume $N/\mathfrak{b}T = c_o > 0$ if N and $\mathfrak{b}T$ are of the same order)

$$\begin{aligned}
V^2 = {} & 2\sigma^4(u_o)\lambda^2(u_o)\int_{\mathbb{R}} k^2(x)dx \cdot I_{\{m=m'=0,\gamma\leq\gamma' \text{ or } m=m'=1,\gamma\leq 2\gamma' \text{ or } m<m' \text{ or } m=m'=2\}} \\
& + c_o\sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}} \mathfrak{K}^2(x)dx \cdot I_{\{m=m'=0,\gamma=\gamma' \text{ or } m=m'=1,\gamma=2\gamma'\}},
\end{aligned}$$

and

$$\sqrt{\mathfrak{b}T}\left\{\tilde{\sigma}^2_{clock}(u_o) - \sigma^2_{clock}(u_o)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, W^2) \tag{4.12}$$

with

$$W^2 = \sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}} \mathfrak{K}^2(x)dx \cdot I_{\{m=m'=0,\gamma>\gamma' \text{ or } m=m'=1,\gamma>2\gamma' \text{ or } m>m'\}}.$$

The additional conditions for the bandwidth $\mathfrak{b}$ and segment length N are given in Theorems 4.7 and 4.13, respectively, and

$$B\tilde{I}AS = \frac{\mathfrak{b}^2}{2}\left(\lambda(u_o)\right)^{(2)}\sigma^2(u_o)\int_{\mathbb{R}} x^2\mathfrak{K}(x)dx \cdot I_{\{m=1,m'=2 \textbf{ or } m=m'=2\}}.$$

Proof. See Appendix B.3.

Until now we have established two different estimators for the spot volatility, namely $\widehat{\sigma}^2_{clock}(\cdot)$ and $\tilde{\sigma}^2_{clock}(\cdot)$, whose results are presented in Theorems 4.10 and 4.14. We discover not only that the asymptotic variances of both estimates are dissimilar, but also that their rates of convergence are unequal in some cases. We therefore compare these two estimators and summarize the results in Table 4.1. For the purpose of comparability, the same kernel functions for $k, \mathfrak{K}$ and K are adopted.

Table 4.1: Comparison of rates of convergence and asymptotic variances between spot volatility estimators $\tilde{\sigma}^2_{clock}(\cdot)$ and $\widehat{\sigma}^2_{clock}(\cdot)$

Case	Conditions			Rate($\tilde{\sigma}^2_{clock}$) vs. Rate($\widehat{\sigma}^2_{clock}$)	Var $\left[\tilde{\sigma}^2_{clock}\right]$ vs. Var $\left[\widehat{\sigma}^2_{clock}\right]$
C1	$m > m'$	-	-		
C2	$m = 0,\ m' = 0$	$\gamma > \gamma'$	-		
C3	$m = 1,\ m' = 1$	$\gamma > 2\gamma'$	-		
C4	$m < m'$	-	$\lambda(u_o) \leq 3/2$	same	smaller
C5	$m = 0,\ m' = 0$	$\gamma = \gamma'$	$\lambda(u_o) \leq (3 - c_o)/2$		
C6	$m = 0,\ m' = 0$	$\gamma < \gamma'$	$\lambda(u_o) \leq 3/2$		
C7	$m = 1,\ m' = 1$	$\gamma = 2\gamma'$	$\lambda(u_o) \leq (3 - c_o)/2$		
C8	$m < m'$	-	$\lambda(u_o) > 3/2$		
C9	$m = 0,\ m' = 0$	$\gamma = \gamma'$	$\lambda(u_o) > (3 - c_o)/2$		
C10	$m = 0,\ m' = 0$	$\gamma < \gamma'$	$\lambda(u_o) > 3/2$	same	larger
C11	$m = 1,\ m' = 1$	$\gamma = 2\gamma'$	$\lambda(u_o) > (3 - c_o)/2$		
C12	$m = 1,\ m' = 1$	$\gamma < 2\gamma'$			
C13	$m = 2,\ m' = 2$			slower	

Table 4.1 displays the performance of the alternative estimator based on the volatility decomposition to that of the classical estimator by comparing their rates of convergence and asymptotic variances. It turns out that the convergence rate of $\tilde{\sigma}^2_{clock}(\cdot)$ is as fast as that of $\widehat{\sigma}^2_{clock}(\cdot)$, and the asymptotic variance of $\tilde{\sigma}^2_{clock}(\cdot)$ is smaller in C1–C7. These cases also include most situations in which the tick-time volatility curve is smoother than the trading intensity curve; see C1–C3. In the next section of microstructure noise models we will see that by decomposing the volatility, not only is the asymptotic variance significantly improved, but also the theoretical rate of convergence.

In addition to the smoothness conditions, an extra restriction on the magnitude of the intensity function is required to determine the goodness of the estimators in C4–C11. The condition $\lambda(\cdot) \leq 3/2$ (or $\lambda(\cdot) \leq (3 - c_o)/2$ for small c_o) is normally satisfied in our transaction data analysis except at the beginning and end of trading days (where the trading activity is high in general; see more details and intensity plots of some real data in Chapter 6). We conjecture that for a time period during which the trading intensity is extremely high, the intensity curve over that period might be ragged, which leads to $m'+\gamma' < m+\gamma$; in particular C8–C11 rarely happen. The major weakness of the alternative estimator appears in cases C12–C13, in which the rate of convergence of $\tilde{\sigma}^2_{clock}(\cdot)$ is worse than that of $\widehat{\sigma}^2_{clock}(\cdot)$. This is due to the fact that in a random-time sampling setting, the bias of $\widehat{\sigma}^2(\cdot)$ cannot be explicitly derived as in the bias derivation of $\widehat{\sigma}^2_{clock}(\cdot)$; thus the rate of convergence decreases.

4.4 Volatility Estimation under Microstructure Noise

One of the major problems of analyzing high-frequency financial data in practice is the discrepancy between efficient log-prices and observed log-prices, which is due to so-called market microstructure noise effects, such as the presence of bid-ask spread bounces, the discreteness of price changes, etc. We refer to Hasbrouck [44] for a thorough review on this topic from an empirical perspective. Since microstructure effects lead to a serial correlation in observed returns, the usual volatility estimates become biased and inconsistent when one observes data with high frequency. This incompatibility has been extensively discussed in, e.g., Bandi and Russell [6] and Zhang *et al.* [70]. See also our short review in Section 2.4.

To account for the effect of microstructure noise we have incorporated an observation error into our model, i.e. we suppose that the observed trading data follows an additive noise model

$$Y_{t_i,T} \;=\; X_{t_i,T} + \varepsilon_i \quad \text{for } i = 1, ..., N_T, \tag{4.13}$$

where ε_i is assumed to be an i.i.d. noise with

$$\mathbb{E}[\varepsilon_i] = \; 0, \quad \mathrm{Var}[\varepsilon_i] = \; \omega^2 < \infty, \quad \text{and} \;\; \mathrm{Var}[\varepsilon_i^2] = \; \theta\omega^4, \; \text{for } \theta \in \mathbb{R}^+.$$

The underlying log-price process $X_{t,T}$ satisfies the model (4.1), and independence between the noise ε and the price process X is assumed. In fact, it has been empirically shown that the assumption about the independence of the noise sequence is justifiable for high-frequency intraday data with transaction sampling, but not for other sampling schemes such as quotation sampling, business-time sampling, 1-minute sampling, etc.; see Hansen and Lunde [39] and Griffin and Oomen [37].

Under the model (4.13) we can see that the volatility estimate $\widehat{\sigma}^2_{clock}(u_o)$ adapted to

the observations $\{Y_{t_i,T}\}_{i=1,...,N_T}$ is now biased, since

$$\begin{aligned}\mathbb{E}\left[\widehat{\sigma}^2_{clock}(u_o)\right] &= \frac{1}{b}\mathbb{E}\left[\sum_{i=1}^{N_T} K\left(\frac{t_i - u_o T}{bT}\right)\left(Y_{t_i,T} - Y_{t_{i-1},T}\right)^2\right] \\ &\to \sigma^2_{clock}(u_o) \;+\; 2T\omega^2\lambda(u_o).\end{aligned}$$

Moreover, it explodes as $T \to \infty$. As usual, this bias can be eliminated by adding a correction term, for example:

$$\widehat{\sigma}^2_{clock}(u_o) + 2\sum_{i=1}^{N_T}\frac{1}{b}K\left(\frac{t_i - u_o T}{bT}\right)\left(Y_{t_{i+1},T} - Y_{t_i,T}\right)\left(Y_{t_i,T} - Y_{t_{i-1},T}\right).$$

Unfortunately, the resulting estimator is still inconsistent. The main reason is a (higher-order) serial correlation in observed returns which keeps the variance of the estimator from vanishing as $T \to \infty$. To solve this problem, a vast literature on estimating volatility using high-frequency data has emerged. We follow the idea of the pre-filtering technique introduced by Podolskij and Vetter [60] and later generalized by Jacod *et al.* [47]. For a detailed discussion of this approach, especially for a comparison of this approach with two other main approaches—the realized kernel approach by Barndorff-Nielsen *et al.* [11] and the multi-timescale approach by Zhang [69]—we refer the reader to their original work.

Hence, under an observed data set $\left\{(t_1, Y_{t_1,T}), ..., (t_{N_T}, Y_{t_{N_T},T})\right\}$ regarded as transaction data, we now estimate the clock-time volatility $\sigma^2_{clock}(\cdot)$ at a time point $u_o \in (0,1)$ by

$$\begin{aligned}\widehat{\sigma}^2_{clock,pavg}(u_o) := \frac{1}{bH}\frac{1}{g_2}\sum_{i=1}^{N_T} K\left(\frac{t_i - u_o T}{bT}\right)\left(\overline{\triangle Y}_{t_i,T}\right)^2 \\ - \frac{1}{2bH}\frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2}\sum_{i=1}^{N_T} K\left(\frac{t_i - u_o T}{bT}\right)\left(Y_{t_i,T} - Y_{t_{i-1},T}\right)^2, \qquad (4.14)\end{aligned}$$

with pre-averaging steps $\overline{\triangle Y}_{t_i,T}$ given by

$$\begin{aligned}\overline{\triangle Y}_{t_i,T} &:= \sum_{l=1}^{H-1} g\left(\frac{l}{H}\right)\left(Y_{t_{i+l},T} - Y_{t_{i+l-1},T}\right) \\ &= -\sum_{l=1}^{H-1}\left\{g\left(\frac{l+1}{H}\right) - g\left(\frac{l}{H}\right)\right\} Y_{t_{i+l},T} \\ &=: -\sum_{l=1}^{H-1} h\left(\frac{l}{H}\right) Y_{t_{i+l},T}.\end{aligned}$$

We define $h(l/H) := g((l+1)/H) - g(l/H)$, where g is another differentiable weight function defined on $[0,1]$ with $g(0) = g(1) = 0$ and which has piecewise Lipschitz continuous

derivatives $g^{(1)}$. The kernel function K and the bandwidth b are similar to those of the preceding section, i.e. K satisfies the condition (K) and b depends on the time span T such that $b \to 0$ and $bT \to \infty$ as $T \to \infty$. Moreover the pre-averaging block size $H = H(b)$ also depends on T such that $H \to \infty$ and $H/bT \to 0$ as $T \to \infty$. We assume that the limits $\sum_{l=1}^{H-1} g(l/H)^n/H$ and $\sum_{l=1}^{H-1} g^{(1)}(l/H)^n/H$ exist and equal $g_n := \int_0^1 g(x)^n dx$ and $g'_n := \int_0^1 g^{(1)}(x)^n dx$, respectively, for $n \in \mathbb{N}$.

The subscript *pavg* stands for the name of the procedure, *pre-averaging*. As its name suggests, we first calculate the average of log returns weighted by a function g over each block of size H, and then we apply a local sum of squares of these averages (similar to the filtering of realized volatility) to construct a spot volatility estimate. By doing this, the variance of the noise is reduced by a factor of $1/H$, as can be seen in the proof; cf. Jacod *et al.* [47]. In particular, this block size H will play a crucial role in this setting along with the main bandwidth size b. Finally, a bias term induced by the additional measurement error will be corrected by the second term of the formula (4.14).

4.15 Theorem *Suppose that Assumptions 4.1 and 4.5 are fulfilled. Let K have bounded first derivatives and let the block size $H = \delta \cdot T^{1/2}$ for $\delta \in (0, \infty)$ and $b^{2\alpha+1}T^{1/2} = o(1)$ with $\alpha = \min\{m + \gamma, m' + \gamma'\}$. Then*

$$\sqrt{bT^{1/2}}\left\{\widehat{\sigma}^2_{clock,pavg}(u_o) - \sigma^2_{clock}(u_o) - BIAS\right\} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \delta\eta_A^2 + \frac{1}{\delta}\eta_B^2 + \frac{1}{\delta^3}\eta_C^2\right)$$

as $T \to \infty$, for $u_o \in (0,1)$, where

$$\begin{aligned}
\eta_A^2 &= 2\sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}} K^2(x)dx,\\
\eta_B^2 &= 4\omega^2\sigma^2(u_o)\lambda(u_o)(g'_2/g_2)\int_{\mathbb{R}} K^2(x)dx,\\
\eta_C^2 &= 2\omega^4\lambda(u_o)(g'_2/g_2)^2\int_{\mathbb{R}} K^2(x)dx, \quad \text{and}\\
BIAS &= \frac{1}{2}\left(\sigma^2(u_o)\lambda(u_o)\right)^{(2)} b^2 \int_{\mathbb{R}} x^2K(x)dx \cdot I_{\{m=m'=2\}}.
\end{aligned}$$

Proof. see Appendix B.4.

In this theorem, the asymptotic consistency of the volatility estimator has been implicitly proven. We see that our result based on the transaction-time model is different from the classical result based on the standard diffusion model in the way that the asymptotic bias and variance rely on the transaction intensity. Compare this with the original result in (2.12) for the integrated volatility estimator.

Due to the concept of volatility decomposition, we now formulate an alternative estimator for the clock-time volatility with the following product

$$\tilde{\sigma}^2_{clock,pavg}(u_o) \ := \ \widehat{\sigma}^2_{pavg}(u_o) \cdot \widehat{\lambda}(u_o). \tag{4.15}$$

For this product, the intensity estimator (4.4) remains the same, since the microstructure noise only affects the price level (or implicitly the tick-time volatility curve) and not the arrival of transactions. Thus, we only need to deal with the estimation of $\sigma^2(\cdot)$ in the presence of noise. This is done by

$$\widehat{\sigma}^2_{pavg}(u_o) := \frac{T}{NH}\frac{1}{g_2}\sum_{i=i_o-N}^{i_o+N} k\left(\frac{i-i_o}{N}\right)\left(\overline{\triangle Y}_{t_i,T}\right)^2 - \frac{T}{2NH}\frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2}\sum_{i=i_o-N}^{i_o+N} k\left(\frac{i-i_o}{N}\right)\left(Y_{t_i,T}-Y_{t_{i-1},T}\right)^2, \tag{4.16}$$

where $i_o := \inf\{i : t_i \geq u_o T\}$. The conditions on the kernel function k and the segment length N are the same as for $\widehat{\sigma}^2(\cdot)$, i.e. $N = N(T) \to \infty$ and $N/T \to 0$ as $T \to \infty$. In fact, this estimator also relies on the pre-averaging approach, and therefore the same conditions for g and h are taken from the preceding clock-time volatility estimator with the block size $H = H(N)$ satisfying $H \to \infty$ and $H/N \to 0$ as $T \to \infty$. Note that we have used the same letters g, h, and H for both estimators in order to not complicate the notation.

4.16 Theorem *Under Assumptions 4.1 and 4.5, the pre-filtering block size $H = \delta \cdot T^{1/2}$ for $\delta \in (0,\infty)$, and k has bounded first derivatives, we obtain*

$$\sqrt{\frac{N}{T^{1/2}}}\left\{\widehat{\sigma}^2_{pavg}(u_o) - \sigma^2(u_o)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \delta\xi_A^2 + \frac{1}{\delta}\xi_B^2 + \frac{1}{\delta^3}\xi_C^2\right),$$

where

$$\begin{aligned}
\xi_A^2 &= 2\sigma^4(u_o)\int_{\mathbb{R}} k^2(x)dx,\\
\xi_B^2 &= 4\omega^2\sigma^2(u_o)(g_2'/g_2)\int_{\mathbb{R}} k^2(x)dx, \quad \text{and}\\
\xi_C^2 &= 2\omega^4(g_2'/g_2)^2\int_{\mathbb{R}} k^2(x)dx,
\end{aligned}$$

under the segment conditions

$$\begin{aligned}
N^{1+2\gamma}/T^{1/2+2\gamma} &\to 0 \quad \text{for } m = 0,\\
N^{3+\gamma^*}/T^{5/2+\gamma^*} &\to 0 \quad \text{for } m = 1 \text{ and } m' = 0,\\
N^{3+\gamma}/T^{5/2+\gamma} &\to 0 \quad \text{for } m = 1 \text{ and } m' = 1,2,\\
N^{3+\gamma'}/T^{5/2+\gamma'} &\to 0 \quad \text{for } m = 2 \text{ and } m' = 0,\\
N^{4}/T^{7/2} &\to 0 \quad \text{for } m = 2 \text{ and } m' = 1,2,
\end{aligned} \tag{4.17}$$

with $\gamma^* = \min\{\gamma,\gamma'\}$ as $T \to \infty$.

Proof. See Appendix B.4.

From (4.15), the alternative estimator $\tilde{\sigma}^2_{clock,pavg}(\cdot)$ is clearly consistent, since it is the product of two consistent estimators. The limit distribution is given below, where the rate of convergence will depend on the convergence rate $\mathfrak{b}T$ and $N/T^{1/2}$ in Theorems 4.7 and 4.16, respectively. For example, for $N/T^{1/2} = o(\mathfrak{b}T)$, we have

$$\begin{aligned}\sqrt{\frac{N}{T^{1/2}}}\left\{\tilde{\sigma}^2_{clock,pavg}(u_o) - \sigma^2_{clock}(u_o)\right\} \\ = \widehat{\lambda}(u_o)\sqrt{\frac{N}{T^{1/2}}}\left\{\widehat{\sigma}^2_{pavg}(u_o) - \sigma^2(u_o)\right\} \;+\; o(1) \;+\; \sigma^2(u_o)\sqrt{\frac{N}{T^{1/2}}}\cdot BIAS_\lambda\end{aligned}$$

with $BIAS_\lambda = \frac{\mathfrak{b}^2}{2}\lambda^{(2)}(u_o)\int_{\mathbb{R}} x^2\mathfrak{K}(x)dx \cdot I_{\{m'=2\}}$. Thereby, the resulting limit distribution is dominated by the limit of $\widehat{\sigma}^2_{pavg}(\cdot)$ given in the last theorem, not that of $\widehat{\lambda}(\cdot)$.

4.17 Theorem *Let all assumptions be satisfied and suppose that the bandwidth $\mathfrak{b}$ and the segment length N fulfill the conditions given in Theorem 4.7 and 4.16, respectively. For $u_o \in (0,1)$ we obtain*

$$\sqrt{\frac{N}{T^{1/2}}}\left\{\tilde{\sigma}^2_{clock,pavg}(u_o) - \sigma^2_{clock}(u_o)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, V^2_{pavg})$$

with

$$\begin{aligned}V^2_{pavg} = \lambda^2(u_o)\left\{\delta\xi^2_A + \frac{1}{\delta}\xi^2_B + \frac{1}{\delta^3}\xi^2_C\right\}\times \\ \times I_{\left\{m=m'=0,\gamma'>\frac{\gamma}{2\gamma+2}\ \textbf{or}\ m=1,m'=0,\gamma'>\frac{\gamma^*+2}{2\gamma^*+8}\ \textbf{or}\ m=2,m'=0,\gamma'>\frac{\sqrt{65}-7}{4}\ \textbf{or}\ m'=1,2\right\}},\end{aligned}$$

and

$$\sqrt{\mathfrak{b}T}\left\{\tilde{\sigma}^2_{clock,pavg}(u_o) - \sigma^2_{clock}(u_o)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, W^2_{pavg}),$$

with

$$\begin{aligned}W^2_{pavg} = \sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}}\mathfrak{K}^2(x)dx \cdot I_{\left\{m=m'=0,\gamma'\le\frac{\gamma}{2\gamma+2}\ \textbf{or}\ m=1,m'=0,\gamma'\le\frac{\gamma^*+2}{2\gamma^*+8}\ \textbf{or}\ m=2,m'=0,\gamma'\le\frac{\sqrt{65}-7}{4}\right\}} \\ + c_1\lambda^2(u_o)\left\{\delta\xi^2_A + \frac{1}{\delta}\xi^2_B + \frac{1}{\delta^3}\xi^2_C\right\}\cdot I_{\left\{m=m'=0,\gamma'=\frac{\gamma}{2\gamma+2}\ \textbf{or}\ m=1,m'=0,\gamma'=\frac{\gamma^*+2}{2\gamma^*+8}\ \textbf{or}\ m=2,m'=0,\gamma'=\frac{\sqrt{65}-7}{4}\right\}},\end{aligned}$$

where $c_1 := bT/(N/T^{1/2})$ if bT and $N/T^{1/2}$ are of the same order.

Proof. See Appendix B.4.

4.18 Remark *i) For the pre-averaging steps, the block size H in* (4.14) *may differ from H in* (4.16) *(both are related to T). For this theoretical investigation, we select $H = O(T^{1/2})$ to balance the rate of convergence of the limit distributions (I), (II), and (III) in order to obtain those asymptotic normality results; see the beginning of the proofs of Theorems 4.15 and 4.16 in Appendix B.4.*

ii) From a practical point of view, all of the unknown components in the asymptotic variances can be estimated by using the existing statistics presented in this section. For instance, the variance ω^2 of the market microstructure noise can be estimated by $\widehat{\sigma}^2_{clock}(u)/(2T\widehat{\lambda}(u))$, since $\widehat{\sigma}^2_{clock}(u) = 2T\omega^2\lambda(u) + o_p(1)$. In the next chapter, we present a method based on least-squares cross-validation to adequately select the smoothing parameters b, $\mathfrak{b}$, N, $H(b)$, and $H(N)$ from the data.

4.5 Comparison

It is well-known that the presence of microstructure noise causes a reduction in the rate of convergence of volatility estimation; compare the convergence speed in (2.2) with that in (2.12). In our noisy model setting, the advantage of the decomposable volatility estimate is *highlighted* since the market microstructure noise does not disturb the arrival of transactions, but rather the price level of $X_{\cdot,T}$; see (4.13). In Chapter 6, our empirical analysis suggests that the tick-time volatility curve is in general less fluctuated than the intensity curve. Therefore coping with microstructure noise becomes easier, as we may choose a larger window for $\widehat{\sigma}^2_{pavg}(\cdot)$ with effectively more data than with $\widehat{\sigma}^2_{pavg,clock}(\cdot)$, which is as rough as the intensity curve. Mathematically, this leads to a higher rate of convergence of the volatility estimator (in some cases even better than the lower bound of the estimation for spot volatility in the standard noisy model, see below).

From this point of view, we now compare the performance of these two different estimates $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ (based on the classical pre-averaging method) and $\tilde{\sigma}^2_{clock,pavg}(\cdot)$ (based on the volatility decomposition). Clearly, the same kernel functions $\mathfrak{K}$, k, and K and weighting function g are used for the purpose of comparability. Before discussing the comparison which is summarized in Table 4.2, we explicitly demonstrate one of those cases in order to emphasize the benefit of volatility factorization in the transaction-time model (other cases can be done similarly). Let $m = m' = 0$ and $\gamma' \leq \gamma/2(\gamma+1)$. According to Theorems 4.17 and 4.15 we have

$$\sqrt{\mathfrak{b}T}\left\{\tilde{\sigma}^2_{clock,pavg}(u_o) - \sigma^2_{clock}(u_o)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, W^2_{pavg})$$

under the bandwidth condition $\mathfrak{b}^{2\gamma'+1}T \to 0$, and

$$\sqrt{bT^{1/2}}\left\{\widehat{\sigma}^2_{clock,pavg}(u_o) - \sigma^2_{clock}(u_o)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\ \delta\eta^2_A + \frac{1}{\delta}\eta^2_B + \frac{1}{\delta^3}\eta^2_C)$$

under another bandwidth condition $b^{2\gamma'+1}T^{1/2} \to 0$. We see that both constraints imply

$$\mathfrak{b}T = o\big(T^{\frac{2\gamma'}{2\gamma'+1}}\big) \quad \text{and} \quad bT^{1/2} = o\big(T^{\frac{\gamma'}{2\gamma'+1}}\big),$$

meaning that the rate of convergence of $\tilde{\sigma}^2_{clock,pavg}(\cdot)$ is much faster than that of $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ in this case.

Moreover, in the standard diffusion model with the presence of noise, a lower bound for spot volatility estimation is derived in a minimax sense with respect to the L_2-loss function. This equals $n^{\frac{-\alpha}{2\alpha+1}}$, given a Hölder-exponent of α for the spot volatility function and where n is the number of subdivisions; see Munk and Schmidt-Hieber [58]. Our estimate $\tilde{\sigma}^2_{clock,pavg}(\cdot)$ exceeds that bound in this particular case and in many other cases; see Table 4.2. Thus, the approach based on volatility decomposition of transaction-time models outperforms previous approaches applied to the standard model in these cases.

Table 4.2: Comparison of rates of convergence and asymptotic variances between spot volatility estimators $\tilde{\sigma}^2_{clock,pavg}$ and $\widehat{\sigma}^2_{clock,pavg}$

Case	Conditions		Rate($\tilde{\sigma}^2_{clock,pavg}$) vs. Rate($\widehat{\sigma}^2_{clock,pavg}$)	Var $\left[\tilde{\sigma}^2_{clock,pavg}\right]$ vs. Var $\left[\widehat{\sigma}^2_{clock,pavg}\right]$
c1	$m=0,\ m'=0$	$\gamma > \gamma'$		
c2	$m=1,\ m'=0$	-		
c3	$m=1,\ m'=1$	$2\gamma' < \gamma$	faster	
c4	$m=2,\ m'=0$	-		
c5	$m=2,\ m'=1$	$\gamma' < 1/2$		
c6	$m=0,\ m'=0$	$\gamma \leq \gamma'$		
c7	$m=0,\ m'=1$	-		
c8	$m=0,\ m'=2$	-	same	smaller, if $\lambda(\cdot) < 1$;
c9	$m=1,\ m'=1$	$\gamma = 2\gamma'$		larger, otherwise.
c10	$m=2,\ m'=1$	$\gamma' = 1/2$		
c11	$m=1,\ m'=1$	$\gamma < 2\gamma'$		
c12	$m=1,\ m'=2$	-		
c13	$m=2,\ m'=1$	$\gamma' > 1/2$	slower	
c14	$m=2,\ m'=2$	-		

Table 4.2 compares the performance of $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ and $\tilde{\sigma}^2_{clock,pavg}(\cdot)$ in terms of their rates of convergence and asymptotic variances. We see that the alternative volatility estimator outperforms the standard estimator in the sense that its rate of convergence is highly improved; see c1–c5. In these cases, the smoothness order of the tick-time volatility is higher than that of the clock-time volatility and intensity[3]. In fact, our data analysis in Chapter 6 and some other empirical analyses, e.g. Dahlhaus and Neddermeyer [25], have

[3]This is clear from above, as we can choose larger window for estimating tick-time volatility than for estimating clock-time volatility.

found that the tick-time volatility curve is in general smoother than the transaction intensity. Thus, we can significantly improve volatility estimation by considering the volatility decomposition. In c6–c10, both estimators possess the same rate of convergence, but the asymptotic variance of $\tilde{\sigma}^2_{clock,pavg}(\cdot)$ turns out to be smaller if the intensity is less than 1; otherwise $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ yields better results.[4] As a result of the bias correction term of $\widehat{\sigma}^2_{pavg}(\cdot)$, also mentioned in the noiseless model (see also C12–C13 in Table 4.1), the alternative estimator has a slower rate of convergence than that of the classical one in cases c11–c14.

[4] In these cases, the resulting limit distribution is dominated by the limit distribution of $\widehat{\sigma}^2_{pavg}(\cdot)$, that is

$$\sqrt{N/T^{1/2}}\left(\tilde{\sigma}^2_{clock,pavg}(\cdot) - \sigma^2(\cdot)\right) = \widehat{\lambda}(\cdot) \cdot \sqrt{N/T^{1/2}}\left(\widehat{\sigma}^2_{pavg}(\cdot) - \sigma^2(\cdot)\right) + o_p(1).$$

Therefore the variance is directly proportional to the rate of arrivals; particularly, the asymptotic variance of this estimate is even less than that of Theorem 4.15 if $\lambda(\cdot) < 1$.

Chapter 5

Other Discussions

5.1 Boundary Effects

As was discussed in Section 2.3, conventional kernel-type estimators will cause biases when estimating points near domain boundaries. Therefore many solutions dealing with this effect in curve estimation have been introduced; see e.g. Fan and Gijbels [30]. In this work we will use the so-called local polynomial fitting technique to eliminate such an edge problem. This method was originally applied in nonparametric regression and later adapted to density estimation. Recently, it has also been employed in volatility curve estimation for standard diffusion models, e.g. by Kristensen [53].

The asymptotic results presented in the previous chapter are only valid for interior time points $u \in (0,1)$, and do not hold for time points near 0 or 1 for the same reason for which boundary problems arise in nonparametric regression.

For example, if $u = 0$, the standard clock-time volatility estimate $\widehat{\sigma}^2_{clock}(u)$ becomes $\frac{1}{b}\sum_{i=1}^{N_T} K\left(\frac{t_i}{bT}\right)\left(X_{t_i,T} - X_{t_{i-1},T}\right)^2$, taking only data from the right-hand side of u into account, which leads to an estimation bias (since the information from the other side is unavailable). Hence, in order to avoid this boundary effect, we apply a local linear fit to our transaction-time model so that biases will be automatically corrected when considering points near borders. We start by creating a local linear estimate (LLE) for $\sigma^2(\cdot)$ at $u_o \in [0,1]$ in the noiseless model (4.1). Let $\mathcal{W}$ be a weight function satisfying condition (K) and $h > 0$ be a bandwidth. We solve the following problem:

$$\min_{a_0(u_o),a_1(u_o)\in\mathbb{R}} \sum_{i=1}^{N_T} \mathcal{W}\left(\frac{u_i - u_o}{h}\right)\left(\widehat{\widehat{\sigma}}^2(u_i) - \{a_0(u_o) + a_1(u_o)(u_i - u_o)\}\right)^2, \tag{5.1}$$

where $\widehat{\widehat{\sigma}}^2(u_i)$, called the response, is an approximation of the tick-time volatility at u_i, which we will define later. Then the LLE is defined by $\widehat{a}_0(u_o)$, which is the solution of the above minimization problem:

$$\begin{bmatrix}\widehat{a}_0(u_o)\\ \widehat{a}_1(u_o)\end{bmatrix} = \begin{bmatrix}S_{T,0}(u_o) & S_{T,1}(u_o)\\ S_{T,1}(u_o) & S_{T,2}(u_o)\end{bmatrix}^{-1}\begin{bmatrix}T_{T,0}(u_o)\\ T_{T,1}(u_o)\end{bmatrix},$$

where

$$S_{T,j}(u_o) = \sum_{i=1}^{N_T} \mathcal{W}\left(\frac{u_i - u_o}{h}\right)(u_i - u_o)^j, \qquad \text{for } j = 0, 1, 2$$

and

$$T_{T,j}(u_o) = \sum_{i=1}^{N_T} \mathcal{W}\left(\frac{u_i - u_o}{h}\right)(u_i - u_o)^j \widehat{\widehat{\sigma}}^2(u_i), \qquad \text{for } j = 0, 1.$$

In particular,

$$\widehat{\sigma}^2_{LLE}(u_o) := \widehat{a}_0(u_o) = \frac{S_{T,2}(u_o)\, T_{T,0}(u_o) \;-\; S_{T,1}(u_o)\, T_{T,1}(u_o)}{S_{T,2}(u_o)\, S_{T,0}(u_o) \;-\; S_{T,1}(u_o)\, S_{T,1}(u_o)}.$$

Furthermore the solution $\widehat{a}_1(u_o)$ can be used to measure the first derivative of the tick-time volatility $\sigma^2(u_o)^{(1)}$; this is one of the special features of the local linear/polynomial fitting method. To have an idea of how this method extends the standard idea of kernel-type estimators, which can be regarded as local constant fits, we look at the following example of the case of local constant approximation, i.e. $a_1(u_o) = 0$. In the noiseless model, one may define the response as

$$\widehat{\widehat{\sigma}}^2(u_i) = \left(X_{t_i,T} - X_{t_{i-1},T}\right)^2 \cdot T, \tag{5.2}$$

since $\mathbb{E}\left[\left(X_{t_i,T} - X_{t_{i-1},T}\right)^2\right] = \sigma^2(t_i/T)/T$. Then by (5.1), we get

$$\begin{aligned}
\widehat{a}_0(u_o) &= \frac{T_{T,0}(u_o)}{S_{T,0}(u_o)} \\
&= \frac{\sum_{i=1}^{N_T} \mathcal{W}\left(\frac{u_i - u_o}{h}\right) \cdot \left(X_{t_i,T} - X_{t_{i-1},T}\right)^2 \cdot T}{\sum_{i=1}^{N_T} \mathcal{W}\left(\frac{u_i - u_o}{h}\right)} \\
&= \frac{\widehat{\sigma}^2_{clock}(u_o)}{\widehat{\lambda}(u_o)}. \qquad \text{(by setting } K(\cdot) = \mathfrak{K}(\cdot) = \mathcal{W}(\cdot), \text{ and } b = \mathfrak{b} = h)
\end{aligned}$$

Therefore, we have recovered the usual kernel estimates given in the previous chapter. We can certainly also employ another choice for the response $\widehat{\widehat{\sigma}}^2(u_i)$ that is more accurate than (5.2). In our analysis, we utilize the kernel-based estimator $\widehat{\sigma}^2(u_i)$ with segment length N given in (4.7) as the response for constructing the LLE, that is

$$\widehat{\widehat{\sigma}}^2(u_i) = \widehat{\sigma}^2(u_i) = \frac{T}{H_{1,N}} \sum_{j=i-N}^{i+N} k\left(\frac{j-i}{N}\right)\left(X_{t_j,T} - X_{t_{j-1},T}\right)^2.$$

Roughly, we will choose N such that $N/T << h$.

Similarly, we can construct other local linear estimates for $\lambda(u_o)$ and $\sigma^2_{clock}(u_o)$, i.e.

$$\widehat{\lambda}_{LLE}(u_o) = \frac{S_{T,2}(u_o)\ \mathfrak{T}_{T,0}(u_o)\ -\ S_{T,1}(u_o)\ \mathfrak{T}_{T,1}(u_o)}{S_{T,2}(u_o)\ S_{T,0}(u_o)\ -\ S_{T,1}(u_o)\ S_{T,1}(u_o)}, \tag{5.3}$$

where

$$\mathfrak{T}_{T,j}(u_o) = \sum_{i=1}^{N_T} \mathcal{W}\left(\frac{u_i - u_o}{h}\right)(u_i - u_o)^j\, \widehat{\widehat{\lambda}}(u_i)$$

with

$$\widehat{\widehat{\lambda}}(u_i) = \widehat{\lambda}(u_i) = \frac{1}{bT}\sum_{j=1}^{N_T} K\left(\frac{t_j - t_i}{bT}\right)$$

and

$$\widehat{\sigma}^2_{clock,LLE}(u_o) = \frac{S_{T,2}(u_o)\ \mathbb{T}_{T,0}(u_o)\ -\ S_{T,1}(u_o)\ \mathbb{T}_{T,1}(u_o)}{S_{T,2}(u_o)\ S_{T,0}(u_o)\ -\ S_{T,1}(u_o)\ S_{T,1}(u_o)}, \tag{5.4}$$

where

$$\mathbb{T}_{T,j}(u_o) = \sum_{i=1}^{N_T} \mathcal{W}\left(\frac{u_i - u_o}{h}\right)(u_i - u_o)^j\, \widehat{\widehat{\sigma}}^2_{clock}(u_i)$$

with

$$\widehat{\widehat{\sigma}}^2_{clock}(u_i) = \widehat{\sigma}^2_{clock}(u_i) = \frac{1}{bT}\sum_{j=1}^{N_T} K\left(\frac{t_j - t_i}{bT}\right)\left(X_{t_j,T} - X_{t_{j-1},T}\right)^2.$$

To account for market microstructure noise, we replace $\widehat{\widehat{\sigma}}^2_{clock}(u_i)$ and $\widehat{\widehat{\sigma}}^2(u_i)$ given above by the noise-robust responses $\widehat{\sigma}^2_{clock,pavg}(u_i)$ and $\widehat{\sigma}^2_{pavg}(u_i)$, respectively. Note that we will use the same symbols $\widehat{\sigma}^2_{clock,LLE}$ and $\widehat{\sigma}^2_{LLE}$, even though the LLEs are equipped with the noise-robust (pre-averaging) responses, in order to not complicate the notation. Likewise, an alternative clock-time volatility estimate based on local linear fitting can be constructed by

$$\tilde{\sigma}^2_{clock,LLE}(u_o) := \widehat{\sigma}^2_{LLE}(u_o) \cdot \widehat{\lambda}_{LLE}(u_o). \tag{5.5}$$

We note that the LLE method entails the choice of an additional tuning parameter, namely the bandwidth h. In our considerations, h is normally chosen such that $h >> \mathfrak{b}$, $h >> b$, and $h >> N/T$ for the corresponding estimators. Under this choice of the bandwidth h, we conjecture that the asymptotic properties of these local linear estimators can be studied via similar arguments to those of LLEs in nonparametric regression and density estimation (see Fan and Gijbels [30] and Cheng [21]). In fact, the bandwidth h needs to be chosen optimally, but this matter is left for future work. This method has the apparent disadvantage of being computational inefficient; in particular one needs to approximate responses (based on pre-filtering) beforehand, which can take a lot of time. One might improve these LLEs by using a fast Fourier transform together with an equivalent kernel. However, we have omitted these developments as our main purpose is not to build a fast computational method, but rather to provide a first step in dealing

with edge effects. For this reason, we will apply such estimates only for a few points near the boundaries to avoid inefficient computation and still use the kernel-based estimates for estimating curves at interior points.

5.2 Bandwidth Selection

One of the most important issues in nonparametric estimation is the determination of appropriate bandwidth choices from data. In fact, we are more concerned with the choice of bandwidths than with the choice of kernel functions. It is well-known that if the tuning parameter is chosen to be small, then the bias will be small, but the variance will increase. On the other hand, if it is large, then the variance will decrease, but we will get a larger bias term. Therefore it is necessary to trade-off between the bias and the variance. A lot of bandwidth selection techniques have been proposed in the literature, e.g. plug-in methods, cross-validation, etc.; see for instance Hart [43] and Härdle [40]. In our situation, we cannot derive the optimal bandwidth choice in terms of the mean squared error (MSE) criterion, since the smoothness order of each curve is unknown. Nevertheless, optimizing tuning parameters without knowledge of the smoothness of the curves to be estimated can be done by cross-validation. This data-driven selection method is well-known and has been developed into many variations; we adopt the technique called least-squares cross-validation in our analysis. The basic idea is to choose a bandwidth from the considered domain such that it optimize a criterion used to measure the distance between the estimator and the true curve. In our consideration, the tuning parameters should be chosen to minimize the integrated squared error (ISE). In fact, we will first need to approximate ISE since it has latent factors. This will be demonstrated later.

In the following, we provide a data-based method for selecting the tuning parameters $\mathfrak{b}$, b, and N from data. This procedure relies on the method mentioned above, that is, the chosen window sizes $\mathfrak{b}_{CV}$, b_{CV}, and N_{CV} will be the minimizers of data-based approximations to

$$\int_I \left\{\widehat{\lambda}(u) - \lambda(u)\right\}^2 du,$$
$$\int_I \left\{\widehat{\sigma}^2_{clock,pavg}(u) - \sigma^2_{clock}(u)\right\}^2 du, \quad \text{and}$$
$$\int_I \left\{\tilde{\sigma}^2_{clock,pavg}(u) - \sigma^2_{clock}(u)\right\}^2 du,$$

respectively, where I is a time domain of interest. Setting $I = [0, 1]$ corresponds to seeking a global bandwidth choice. We can also explore a local one by decreasing the size of the interval, for instance setting $I = [u_o - \delta, u_o + \delta]$ with small δ. To enable us to estimate the parameter curves, we apply this local procedure to a grid of time points over $[0, 1]$ to obtain adaptive bandwidth windows. In the case of symmetric kernel functions, we should carefully choose the interval I to avoid boundary effects.

This idea is similar to that used in Brooks and Marron [18]. That is, we find $\mathfrak{b}$ which

minimizes the following integrated squared error

$$\begin{aligned} ISE_{\widehat{\lambda}}(\mathfrak{b}) &:= \int_I \left\{\widehat{\lambda}(u) - \lambda(u)\right\}^2 du \\ &= \int_I \widehat{\lambda}(u)^2 du - 2\int_I \widehat{\lambda}(u) du \int_I \lambda(u) du + \int_I \lambda(u)^2 du. \end{aligned}$$

The first term can be computed directly from the data and the last term can be neglected since it does not depend on $\mathfrak{b}$. Thus, the second term is the only one that needs to be estimated from the data. Essentially, the optimal bandwidth $\mathfrak{b}_{CV}$ should minimize the cross-validation function

$$CV_{\widehat{\lambda}}(\mathfrak{b}) := \int_I \widehat{\lambda}(u)^2 du - \frac{2}{T}\sum_{u_i \in I} \widehat{\lambda}^{(i)}(u_i), \tag{5.6}$$

that is, $\mathfrak{b}_{CV} = \text{argmin}_{\mathfrak{b}}\, CV_{\widehat{\lambda}}(\mathfrak{b})$ over $\mathfrak{b} \in \Theta_{\mathfrak{b}}$, where

$$\widehat{\lambda}^{(i)}(u_i) := \frac{1}{\mathfrak{b}T} \sum_{j=1, t_j \neq t_i}^{N_T} \mathfrak{K}\left(\frac{t_j - t_i}{\mathfrak{b}T}\right)$$

is the so-called leave-one-out estimator ($u_i = t_i/T$). The range of bandwidths $\Theta_{\mathfrak{b}}$ needs to be appropriately restricted. We can prove that this cross-validation function is an unbiased estimator for $ISE_{\widehat{\lambda}}(\mathfrak{b}) - \int_I \lambda(u)^2 du$ (see Appendix C). Thus, $\mathfrak{b}_{CV}$ should be considerably close to the bandwidth that minimizes $ISE_{\widehat{\lambda}}(\mathfrak{b})$. We conjecture that one can also show the asymptotic optimality for this bandwidth $\mathfrak{b}_{CV}$ by using similar arguments to those used in density estimation or intensity estimation; see e.g. Härdle *et al.* [41] and Brooks and Marron [18]. By asymptotic optimality of $\mathfrak{b}_{CV}$, we refer to the property that

$$\frac{ISE_{\widehat{\lambda}}(\mathfrak{b}_{CV})}{\inf_{\mathfrak{b}\in\Theta_{\mathfrak{b}}} ISE_{\widehat{\lambda}}(\mathfrak{b})} \to 1, \quad \text{a.s.}$$

Since our aim is not to give a comprehensive bandwidth selection study, but rather an illustration of the behavior of both volatility estimators, we do not give a complete investigation of asymptotic optimality for these bandwidth choices.

In much the same way, we can design asymptotically unbiased estimators for the ISEs of $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ and $\tilde{\sigma}^2_{clock,pavg}(\cdot)$ and find the optimal solutions b_{CV} and N_{CV} by minimizing the functions

$$CV_{\widehat{\sigma}^2_{clock,pavg}}(b) := \int_I \left(\widehat{\sigma}^2_{clock,pavg}(u)\right)^2 du - \frac{2}{T}\sum_{u_i \in I} \widehat{\sigma}^{2(i)}_{clock,pavg}(u_i) \cdot \widehat{\sigma}^{2(i)}_{pavg}(u_i) \tag{5.7}$$

(with values of N and $H(N)$ for $\widehat{\sigma}^{2(i)}_{pavg}(u_i)$ and $H(b)$ for $\widehat{\sigma}^{2(i)}_{clock,pavg}(u_i)$ selected beforehand) and

$$CV_{\tilde{\sigma}^2_{clock,pavg}}(N) := \int_I \left(\tilde{\sigma}^2_{clock,pavg}(u)\right)^2 du - \frac{2}{T}\sum_{u_i \in I} \left(\widehat{\sigma}^{2(i)}_{pavg}(u_i)\right)^2 \cdot \widehat{\lambda}^{(i)}(u_i) \tag{5.8}$$

over Θ_b and Θ_N, respectively. Note that in (5.8), the optimal window $\mathfrak{b}_{CV}$ is used for evaluating $\widehat{\lambda}^{(i)}(u_i)$ and $H(N)$ is pre-selected. The optimal block lengths $H(b)$ and $H(N)$ in the pre-averaging steps of $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ and $\tilde{\sigma}^2_{clock,pavg}(\cdot)$ can also be achieved by minimizing $CV_{\widehat{\sigma}^2_{clock,pavg}}(b_{CV})$ and $CV_{\tilde{\sigma}^2_{clock,pavg}}(N_{CV})$ with respect to $H(b)$ and $H(N)$, respectively.

5.3 Endogenous Transaction-Time Models

Up to this point, the asymptotic properties for the volatility estimators were investigated under Assumption (4.1). In particular, the point process $N_{t,T}$ was a nonhomogeneous Poisson process with a non-random intensity function $\lambda(t/T)$, and the tick-time volatility $\sigma(t/T)$ was considered to be a deterministic function as well.

In Remark 4.2 ii) we pointed out that a first step towards having stochastic volatility models is letting $\lambda_{t,T}$ and $\sigma_{t,T}$ be independent random processes, both of which are also independent of the Poisson process $N_{t,T}$ and the innovations U_i (all the proofs remain the same if one carefully constructs arguments conditional on these processes). From the point of view of real application, however, there is a leverage effect between prices, volatility, and news flow. Particularly, the frequency of transactions will depend directly on the information flow, which will influence the price processes. In order to allow for such correlation between those processes, a rescaled version of the transaction-time model similar to (4.1) is determined by assuming that the price process fulfills Assumption 3.1.

Let the price process X observed at transaction time t_i satisfy Assumption 3.1, that is

$$X_{t_i,T} \;=\; X_{t_{i-1},T} + \sigma\left(\frac{t_i}{T}\right)\frac{1}{\sqrt{T}}U_i,$$

where $N_{t,T}$ is a point process having an $\mathcal{F}_{t,T}$-intensity $\lambda_{t,T} = \lambda(t/T)$ and t_i are the corresponding arrival times of $N_{t,T}$. Now $\sigma_{t,T} = \sigma(t/T)$ is an $\mathcal{F}_{t,T}$-predictable process and U_i is $\mathcal{F}_{t_i,T}$-measurable for each i with

$$\mathbb{E}\left[U_i \;\middle|\; \mathcal{F}_{t_i-,T}\right] = \;0 \quad \text{and} \quad \mathbb{E}\left[U_i^2 \;\middle|\; \mathcal{F}_{t_i-,T}\right] = \;1.$$

We also assume that $\mathbb{E}\left[U_i^4 \;\middle|\; \mathcal{F}_{t_i-,T}\right]$ exists. Some examples of such a model are already given in the end of Section 3, including the transaction-time model (4.1). Indeed, this model can be written in continuous time as

$$dX_{t,T} = \;\sigma(t/T)\frac{1}{\sqrt{T}}\,dL_{t,T}, \qquad \text{where} \quad L_{t,T} := \sum_{i=1}^{N_{t,T}} U_i$$

is a pure jump process, which is more general than pure jump Lévy processes. It is worth pointing out that in contrast to other frameworks our setting is purely nonparametric in that we do not give any parametric models to the volatility process. For example, in the continuous GARCH model, the leverage effect is obtained by modeling the volatility σ_t^2 to incorporate this effect—namely via a continuous ARMA model (see Brockwell *et al.* [17]).

It is seen that $X_{t,T}$ is a martingale with respect to the filtration $\mathcal{F}_{t,T}$, where $\mathcal{F}_{t,T}$ consists of all information available up to time t, roughly speaking. Moreover we will assume that realizations of $\lambda(\cdot)$ and $\sigma^2(\cdot)$ fulfill Assumption 4.5, in particular that they are Hölder continuous and bounded away from zero uniformly.

We thus obtain an analogous volatility decomposition

$$\lim_{\delta\to 0}\ \frac{1}{\delta}\mathbb{E}\left[(X_{t+\delta T,T}-X_{t,T})^2\ \middle|\ \mathcal{F}_{t,T}\right]\ =\ \sigma^2(t/T)\lambda(t/T)\ =:\ \sigma^2_{clock}(t/T),$$

so that we can proceed along the same lines of investigation into the asymptotic properties of intensity and volatility estimators as before; however, the proofs become much more difficult, since we need to deal with the correlations between all processes, namely the point process, intensity process, tick-time volatility, and price process. A preparation for such proofs of this general case is provided in our work as some of the asymptotic results being presented are based on martingale theory, in particular Lemma 1 and Corollary 1. We show below that the intensity estimator given in (4.4) can be used in this general setting to measure $\lambda(u_o)$ for $u_o \in (0,1)$. We find that the same asymptotic result as in Theorem 4.7 still holds, that is

$$\sqrt{\mathfrak{b}T}\frac{\left(\widehat{\lambda}(u_o)-\lambda(u_o)-\frac{\mathfrak{b}^2}{2}\lambda^{(2)}(u_o)\int_{\mathbb{R}}x^2\mathfrak{K}(x)dx\cdot I_{\{m'=2\}}\right)}{\sqrt{\lambda(u_o)\int_{\mathbb{R}}\mathfrak{K}^2(x)dx}}\ \xrightarrow{\mathcal{D}}\ \mathcal{N}(0,1)\,; \tag{5.9}$$

see Appendix C. A complete study on the asymptotic behaviors of $\widehat{\sigma}^2_{clock}(\cdot)$ and $\widehat{\sigma}^2(\cdot)$ (or $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ and $\widehat{\sigma}^2_{pavg}(\cdot)$ when microstructure noise is present) is postponed to future work, as this kind of endogeneity is not easy to handle. We note that an additional asymptotic variance term for volatility estimators will occur in this stochastic volatility model due to the endogeneity. We discuss this further in Appendix C.

Chapter 6

Empirical Analysis and Simulation

6.1 Simulation

In this section we illustrate the performance of our volatility estimators in the finite sample setting. For simulation studies, we generate efficient prices from the transaction-time model (4.1) with the following parameter functions:

$$\sigma^2(u) \;=\; \exp(-8 + 0.2\cos(10u\pi)) \tag{6.1}$$

and

$$\lambda(u)(\cdot) \;=\; B_u^2(\cdot) \quad \text{for } u \in (0,1), \tag{6.2}$$

where B_u is another standard Brownian motion on $[0,1]$ which is independent of $W_\cdot$. Note that a realization of the stochastic intensity (6.2) is simulated and kept *fixed* throughout this section[1]. The intensity curve is chosen to be quite rough (since a Brownian motion has always non-differentiable paths), whereas the tick-time volatility varies in a smoother way; this example belongs to case c4 in Table 4.2 from the theoretical results. In fact, these parameter values are set to be reasonable for a real asset. According to the volatility decomposition, the spot volatility curve $\sigma^2_{clock}(\cdot)$ is still non-differentiable since it is the product of $\sigma^2(\cdot)$ and $\lambda(\cdot)$. The trajectory of intensity and the transaction-time (or tick-time) volatility curve are depicted below in Figure 6.1. Row (a) displays the underlying (clock-time) volatility curve on a logarithmic scale. It is therefore the sum of the logarithm of the tick-time volatility in (b) and the log-intensity in (c). We note that in the following, all graphics with the tag (a), (b) and (c) always express plots of (clock-time) volatility, tick-time volatility, and intensity curves, respectively (all on a logarithmic scale). Figure 6.3 shows another set of parameters in which the frequency of transactions remains the same, but for which a realization from a stochastic process is selected as a source for the tick-time volatility.

[1]The realization of the intensity curve is chosen to mimic common transaction intensity as observed in the real market, typically forming a U-shape. With it we can produce an NHPP having around 23,000 transactions per day.

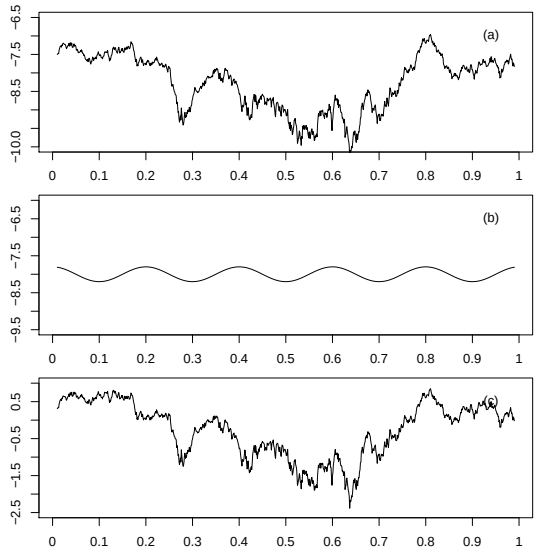


Figure 6.1: Underlying parameter curves given by (6.1) and (6.2) (all in log-scale): (a) clock-time volatility, (b) tick-time volatility (mimicking the case where the true curve is smooth), and (c) intensity.

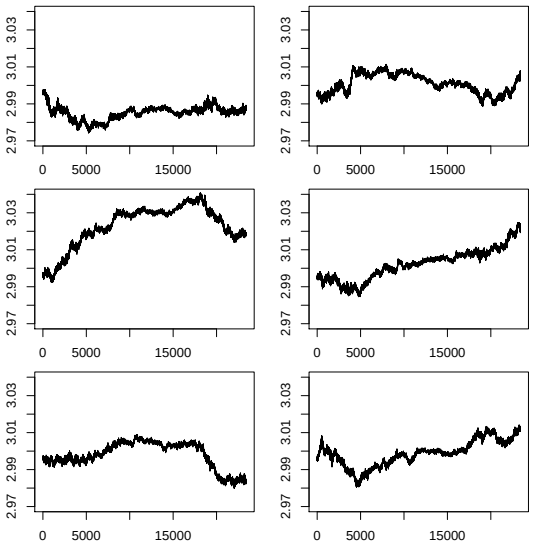


Figure 6.2: Some examples of simulated high-frequency log-price processes with the parameters shown in Figure 6.1, having approximately 23,000 transactions per day.

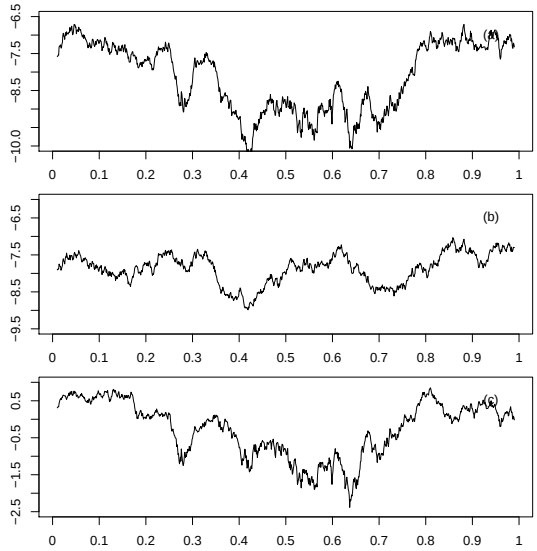


Figure 6.3: Underlying parameter curves (all in log-scale) with fluctuated tick-time volatility and intensity given by (6.2): (a) clock-time volatility, (b) tick-time volatility (as a path of BM, emulating the case where the true curve is fluctuated), and (c) intensity.

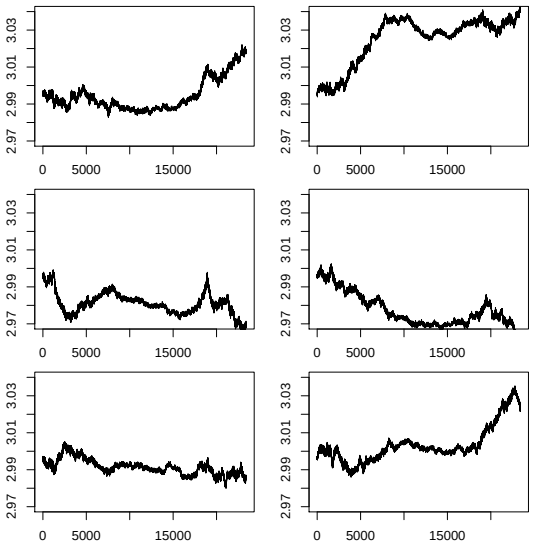


Figure 6.4: Some examples of simulated high-frequency log-price processes with parameters shown in Figure 6.3, having approximately 23,000 transactions per day.

The time span T is chosen to be 23400 corresponding to one-day trading hours in 1-second resolution, i.e. 09:30 AM - 04:00 PM (for example in NYSE and NASDAQ—excluding the pre- and after-market hours). We set the start price $X_{t_0,T} = X_{0,T} = 20$. In our experiments, market microstructure noise is considered to be an additive white noise. In particular we generate

$$Y_{t_i,T} = X_{t_i,T} \ + \ \varepsilon_i, \quad \text{for } i = 1, ..., N_T, \tag{6.3}$$

where ε_i is an i.i.d. Gaussian disturbance sequence with mean zero and a small variance $\omega^2 = 0.0005^2$ (as small as shown in empirical market value; see the next section). The magnitude of this standard deviation will play a crucial role in the quality of volatility estimation when the sample size is small (discussed at the end of this section). Our choice of standard deviation has also been used in many simulation studies for standard Brownian semimartingales under market microstructure noise, particularly in the pre-filtering context[2]. Also, in the end of this section, we will show that our pre-averaging estimators are robust against additive noise with a rounding effect. This will be demonstrated together with results of a benchmark estimator presented by Dahlhaus and Neddermeyer [25], which is based on non-linear microstructure noise model, particularly rounding error.

Under the above settings, transaction data with approximately 23,000 trades per day is generated, which is typical for one-day trading data of highly liquid stocks. Some examples of simulated prices can be seen in Figures 6.2 and 6.4. Having around 23,000 transactions over the time period T means that on average one transaction is executed per second. In fact, it is possible to have many transactions at the same timestamp, i.e. $t_i = ... = t_{i+k}$, for some $i, k \in \mathbb{N}$, with possibly different prices $X_{t_i}, ..., X_{t_{i+k}}$. This is quite common for high-frequency data in the real market since many trades could occur simultaneously within a second.

In this section, the performance of the pre-averaging estimators $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ and $\tilde{\sigma}^2_{clock,pavg}(\cdot)$ on simulated intraday transaction data is assessed according to the numerical bias, relative bias, variance, and mean squared errors over a grid of time points $u = 0.01, 0.02, \ldots, 0.99$ (a total of 99 time points). These results are based on the average of 500 repetitions of the above system (6.3) with the parameters (6.1) and (6.2). The local linear estimators given in Section 5.1 for estimating parameters close to the boundary are also used in our study, however only on a few points at the borders[3].

Although an analytical method for optimal bandwidth selections has already been developed in the previous chapter, it is quite difficult in practice to apply it to every single iteration (in this case 500 iterations) on a personal computer due to the computational time required (especially for such a big data set). We have therefore omitted data-driven

[2]In many empirical analyses of this type of problem, ω is often found to be 0.0005, since the noise-to-signal ratio is normally estimated to be around 10^{-3}, leading to $\omega = 0.0005791914$ when $\sigma^2(\cdot) \approx \exp(-4)$; see Hansen and Lunde [39]. This level of noise is also used by Zhang *et al.* [70] and Jacod *et al.* [47] in their Monte Carlo experiments.

[3]Calculating the LLE estimator at all points is very computationally intensive, an issue that needs to be improved, but which we leave to future work.

bandwidth selection in these experiments, postponing the comprehensive investigation and an improvement of such algorithms to future work.

In our simulations, we therefore examine the behavior of the estimators under many different global bandwidth sizes, ranging from a small window size[4] 0.004 to a large window size 0.014 for $\mathfrak{b}$ (and similarly for b and $N \approx bT$). For example, if $b = 0.01$, it means that only those data points lying in the time frame of size $2bT = 234$ (approx. 4 minutes) around the time point of interest u_o contribute to the intensity estimation at time u_o. We will see that our choices are appropriate in the sense that they correctly estimate the parameter curves and do not over-smooth them[5]. In what follows, we will use the notation $N \sim bT$ (used for the estimation of the tick-time volatility) to indicate that N is varying by time depending on the number of transactions around that time point. More precisely, given a time point u_o, we set N to be the number of transactions on the interval $[u_o - 4bT, u_o + 4bT]$ divided by 8. We did not choose N as $\#\{\text{transactions on } [u_o - bT, u_o + bT]\}/2$, so as not to let the tick-time volatility estimator be greatly influenced by the intensity.

By doing this, we are able to compare the smoothness of two curves: the tick-time volatility estimate and the clock-time volatility estimate, since the segment length is adjusted by time so that it is not too large for time periods of low trading activity and not too small for those of high activity. Now, it can be said that the choice of the bandwidths $\mathfrak{b}, b$, and the segment N are almost of the same level, and as a result that the smoothness of all curves can be compared[6]. Of course, it is advisable to construct an efficient adaptive bandwidth method for finite samples, which automatically adapts to the smoothness/roughness level of the curves.

When studying nonparametric kernel estimation, one can compare many kernel functions to obtain the best one for volatility estimation. However, it is well-known that the choice of kernel functions has much less of an impact on curve estimation than the choice of tuning parameters in nonparametric estimation. Therefore, instead of discussing kernel selection at length, we simply compare two different choices of the kernel function: the triweight and Epanechnikov kernels. The first one assigns more weight to observations nearer to the time point of interest than the second one, which is well-known to be optimal in classical smooth function estimation. Our study shows that the Epanechnikov kernel leads to a reduced mean squared error (MSE) of all estimates at most time points. Moreover the global measure—mean integrated squared error (MISE)—is also smaller under the Epanechnikov kernel. For example, the MISE for $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ with triweight kernel function is 1.970044×10^{-8} whereas with Epanechnikov it is only 1.502683×10^{-8} (all the bandwidths are selected to be the same, in this case $b = 0.008$ and $H = 15$). In what follows, we will choose only the Epanechnikov kernel, i.e. $K(u)$, $\mathfrak{K}(u)$, and $k(u)$ will be $\frac{3}{4}\left(1 - u^2\right) I_{\{|u| \leq 1\}}$. As a weighting function for pre-filtering steps, $g(x)$ is chosen to be

[4] with respect to the number of transactions of data, having around 23,000 transactions on a single trading day.

[5] It is important to remember that the volatility typically has non-differentiable paths, meaning that a too-large kernel window may produce an imprecise result.

[6] This choice is just an approximation used specifically for this simulation study in order to compare the roughness of the curves.

$x(x-1)_{\{0\le x\le 1\}}$.

It is noteworthy that the choice of the block sizes $H(b)$ and $H(N)$ in the pre-averaging steps plays a minor role in the quality of spot volatility curve estimation (the choice of the main smoothing windows b and N are more important); however, a minimum of $H = 5$ is suggested. If $H = 2$, the pre-averaging estimates $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ and $\widehat{\sigma}^2_{pavg}(\cdot)$ will have the same properties as the standard realized volatilities $\widehat{\sigma}^2_{clock}(\cdot)$ and $\widehat{\sigma}^2(\cdot)$, which are non-stable in the noisy models. From our experiments (we do not record all of them here), the choice of $H = 15$ is suitable for the level of noise corresponding to $\omega = 0.0005$. In fact, a larger block size does not yield better results and only decelerates the calculation. From Figures 6.5 and 6.6 we see that the values of the volatility estimates at the time points $u = 0.1, \ldots, 0.9$ tend to almost the same values after $H = 15$ (these results are based on 10 iterations only). In order to derive an optimal choice of the pre-averaging block size H, the relation $H = \delta \cdot T^{1/2}$ obtained in our theoretical discussions might be exploited by minimizing the asymptotic variance to get the optimal δ_{opt}, and then redefining $H = \delta_{opt} \cdot T^{1/2}$ (see Jacod *et al.* [47]). We have so far ignored this issue as it is still unclear how to estimate the unknown parameters properly in small samples.

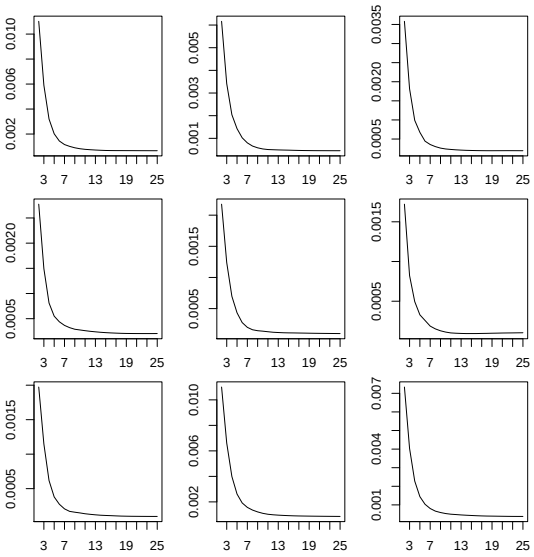


Figure 6.5: Clock-time volatility estimate $\widehat{\sigma}^2_{clock,pavg}(u)$ (y-axis) against block size $H(b)$ (x-axis) for $u = 0.1, 0.2, \ldots, 0.9$ (from upper left to the bottom right). The bandwidth $b = 0.008$ is used.

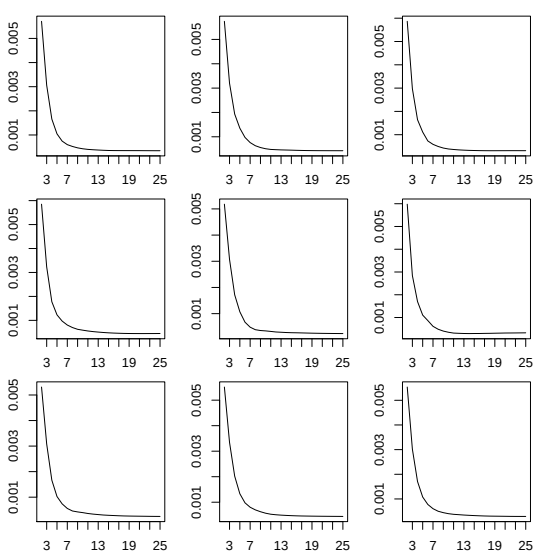


Figure 6.6: Tick-time volatility estimate $\widehat{\sigma}^2_{pavg}(u)$ (y-axis) against block size $H(N)$ (x-axis) for $u = 0.1, 0.2, \ldots, 0.9$ (from upper left to the bottom right). The segment length $N \sim 0.008T$ is used.

In addition to the concern about selecting the block size, the performance of pre-averaging type estimators in small sample data has been discussed in Jacod *et al.* [47]. In particular it has been pointed out that they are slightly biased. More precisely, the second term of $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ (or $\widehat{\sigma}^2_{pavg}(\cdot)$), which is responsible for correcting the bias induced by the contamination, leaves a small bias term (which in large enough samples becomes negligible)

so that, heuristically,

$$\widehat{\sigma}^2_{clock,pavg}(u_o) \approx \sigma^2_{clock}(u_o) + \frac{1}{2bH}\frac{\sum h^2(l/H)}{g_2}\sum_{i=1}^{N_T} K\left(\frac{t_i - u_o T}{bT}\right)\left(X_{t_i,T} - X_{t_{i-1},T}\right)^2$$
$$\approx \sigma^2_{clock}(u_o) + \frac{1}{2H}\frac{\sum h^2(l/H)}{g_2}\sigma^2_{clock}(u_o).$$

Thus, the estimator should be corrected by multiplying a scaling factor, i.e.

$$\widehat{\sigma}^{2\ (adj)}_{clock,pavg}(u_o) := \frac{1}{1 + \frac{1}{2H}\frac{\sum h^2(l/H)}{g_2}} \cdot \widehat{\sigma}^2_{clock,pavg}(u_o) \approx \sigma^2_{clock}(u_o).$$

The tick-time volatility estimate $\widehat{\sigma}^2_{pavg}(\cdot)$ also needs to be adjusted to $\widehat{\sigma}^{2\ (adj)}_{pavg}(\cdot)$ by the same scaling factor to eliminate this small bias. The importance of this adjustment in our finite-sample simulations is depicted in Figure 6.7. We see that the adjusted pre-averaging estimators can effectively diminish this small-sample bias. From now on, all simulated curves that depend on the pre-filtering technique will be automatically adjusted by this factor.

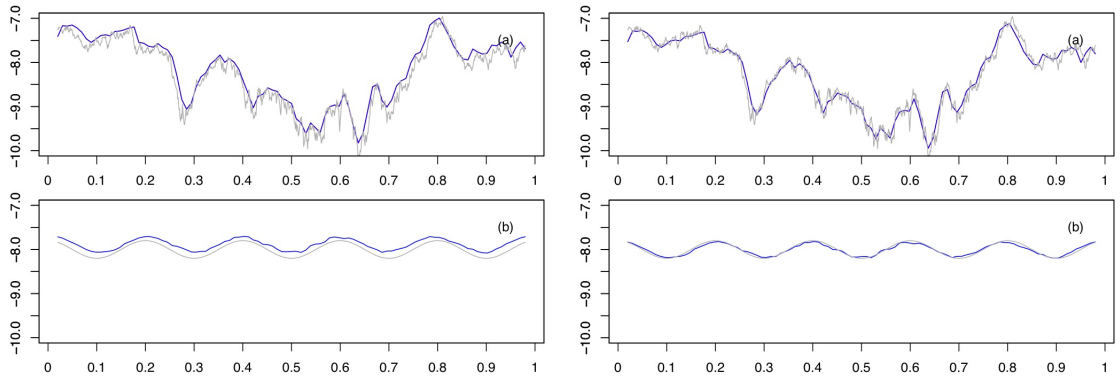


Figure 6.7: Log of pre-averaging estimates $\log \widehat{\sigma}^2_{clock,pavg}(u)$ and $\log \widehat{\sigma}^2_{pavg}(u)$ (left panel, (a) and (b)) vs. and log of bias-adjusted pre-averaging estimates $\log \widehat{\sigma}^{2\ (adj)}_{clock,pavg}(u)$ and $\log \widehat{\sigma}^{2\ (adj)}_{pavg}(u)$ (right panel, (a) and (b)). These results are the averages of 500 simulations based on the window size $b = 0.008$ or $N \sim 0.008T$ (with $H = 15$). The grey curves are the true parameter curves and the blue curves are the corresponding estimates.

According to our experiments, we find that if the bandwidth/segment length is chosen to be too large, the estimation will over-smooth the underlying volatility curve so that only its rough shape can be detected, but not its details or kinks. On the other hand, if the bandwidth is too small, it leads to an inaccurate volatility estimator. In the first step, as we do not know beforehand whether the tick-time volatility is smoother than the clock-time volatility, all window sizes are chosen such that they correspond to the same smoothness level for the sake of comparability of the smoothness/roughness of the two different volatility estimates.

From our experience, we prefer the following choices: $\mathfrak{b} = 0.008$, $b = 0.008$, and $N \sim 0.008T$, as appropriate choices[7] for the tuning parameters (similar results were obtained for the other choices of bandwidth and segment length, and we therefore do not report them here)[8].

To evaluate the precision of the volatility estimation, mean squared errors and mean integrated squared error are calculated, where the integration is discretely approximated. Table 6.1 reports the numerical results for these values of $\widehat{\sigma}^{2\ (adj)}_{clock,pavg}(u)$, $\tilde{\sigma}^{2\ (adj)}_{clock,pavg}(u)$, $\widehat{\lambda}(u)$, and $\widehat{\sigma}^{2\ (adj)}_{pavg}(u)$ at the time points $0.1, \ldots, 0.9$. We see that our estimators perform satisfactorily with respect to the quadratic errors. In fact, we examine these estimates over a finer grid of time points, every 0.01 step, i.e. $u = 0.01, 0.02, \ldots, 0.99$. It turns out that the alternative estimator based on the volatility decomposition outperforms the classical estimator for only 43 out of 99 points. Thus the alternative estimator does not exhibit any superiority over the original estimator in this situation. This may be due to an improper choice of the segment length for estimating $\sigma^2(\cdot)$, which is by construction much smoother than the intensity $\lambda(\cdot)$ and clock-time volatility curve $\sigma^2_{clock}(\cdot)$. The numerical MISE is calculated for each estimate: $\widehat{\sigma}^{2\ (adj)}_{clock,pavg}$, $\tilde{\sigma}^{2\ (adj)}_{clock,pavg}$, $\widehat{\lambda}$, and $\widehat{\sigma}^{2\ (adj)}_{pavg}$, and is equal to 1.267811×10^{-8}, 1.341613×10^{-8}, 7.174735×10^{-3} and 1.887689×10^{-8} respectively.

With a priori knowledge that the tick-time volatility curve is smoother than the intensity curve, one might assign larger segment length $N >> bT$ to enhance the performance of the alternative estimator. To see this, we set $N \sim 0.012T$ and let b and $\mathfrak{b}$ be unchanged.

Table 6.2 shows the Monte Carlo results for the estimators with this set of bandwidths. We see that the numerical MSEs of $\tilde{\sigma}^{2\ (adj)}_{clock,pavg}(u)$ have been improved and are less than those of $\widehat{\sigma}^{2\ (adj)}_{clock,pavg}(u)$ at almost every time point (in fact, at 97 out of 99 time points). Moreover, the numerical MISE of $\tilde{\sigma}^2_{clock,pavg}(\cdot)$ has been reduced (by about 30%). The numerical MISEs are now equal to 1.267811×10^{-8}, 0.936115×10^{-8}, 7.174735×10^{-3}, and 1.263474×10^{-8} for $\widehat{\sigma}^{2\ (adj)}_{clock,pavg}$, $\tilde{\sigma}^{2\ (adj)}_{clock,pavg}$, $\widehat{\lambda}$, and $\widehat{\sigma}^{2\ (adj)}_{pavg}$, respectively.

Hence, the volatility decomposition scheme can improve the volatility estimation in the case that the volatility per tick varies in a smoother way than the trading intensity; this coincides with the theoretical result presented in c1–c5 in Table 4.2.

Figure 6.8 displays the volatility plots (all in log-scale) with these bandwidth choices: $\mathfrak{b} = 0.008, b = 0.008$, and $N \sim 0.012T$. The left column is constructed by a smooth tick-time volatility given in Figure 6.1 (b) and the right one is set up by a fluctuated tick-time volatility given in Figure 6.3 (b). Relying on 500 simulations, the classical volatility estimator $\widehat{\sigma}^{2\ (adj)}_{clock,pavg}(\cdot)$ (red) and the alternative estimator $\tilde{\sigma}^{2(adj)}_{clock,pavg}(\cdot)$ (blue) are almost identical and consistent with the latent parameter curves (grey).

[7] with respect to the amount of data, 23,000 transactions per trading day in this case.

[8] It has been noted above that with this choice of N, $\widehat{\sigma}^2_{pavg}(\cdot)$ uses approximately the same amount of data as considered by $\widehat{\sigma}^2_{clock,pavg}(\cdot)$. If we set $N = bT$ fixed through the line, then it is possible that over some periods of low trading activity (which usually occur in the middle of trading days) the estimator $\widehat{\sigma}^2_{pavg}(\cdot)$ might be too smooth, so that some kinks of the underlying parameter function are blurred. Additionally, this segment length might be too small at the beginning and the end of trading days where the trading activity is quite high; see the next section.

Table 6.1: Numerical bias, relative bias, variance, and mean squared error of (adjusted) volatility estimators and intensity estimator based on the smoothing parameters $\mathfrak{b} = 0.008$, $b = 0.008$, and $N \sim 0.008T$ and $H(b)$ and $H(N)$ are 15.

$\widehat{\sigma}^2_{clock,pavg}(u)$	$u = 0.1$	$u = 0.2$	$u = 0.3$	$u = 0.4$	$u = 0.5$	$u = 0.6$	$u = 0.7$	$u = 0.8$	$u = 0.9$
Bias ($\times 10^{-5}$)	3.347455	3.721725	1.615323	0.834484	0.861053	2.646719	0.487952	5.866782	1.309928
Relative bias	0.061656	0.089387	0.098605	0.044020	0.074896	0.295414	0.044332	0.078642	0.035728
Variance ($\times 10^{-8}$)	1.947030	1.812793	0.571093	0.773642	0.384739	0.365068	0.390233	2.939256	1.185987
MSE ($\times 10^{-8}$)	**2.059085**	1.951305	0.597186	0.780606	**0.392153**	**0.435119**	0.392614	**3.283447**	1.203146
$\tilde{\sigma}^2_{clock,pavg}(u)$	$u = 0.1$	$u = 0.2$	$u = 0.3$	$u = 0.4$	$u = 0.5$	$u = 0.6$	$u = 0.7$	$u = 0.8$	$u = 0.9$
Bias ($\times 10^{-5}$)	3.316245	3.637344	1.512435	0.927749	0.841854	2.688115	0.358666	5.731233	1.228725
Relative bias	0.061081	0.087360	0.092324	0.048940	0.073226	0.300035	0.032586	0.076825	0.033513
Variance ($\times 10^{-8}$)	2.019202	1.596506	0.544116	0.664841	0.427521	0.420792	0.276503	3.542093	1.157029
MSE ($\times 10^{-8}$)	2.129177	**1.728809**	**0.566991**	**0.673448**	0.434608	0.493052	**0.277789**	3.870563	**1.172126**
$\widehat{\lambda}(u)$	$u = 0.1$	$u = 0.2$	$u = 0.3$	$u = 0.4$	$u = 0.5$	$u = 0.6$	$u = 0.7$	$u = 0.8$	$u = 0.9$
Bias	-0.022661	0.057865	0.010821	0.009854	-0.007402	0.066243	-0.024518	0.124287	-0.033980
Relative bias	-0.011463	0.056944	0.018143	0.021300	-0.017685	0.302949	-0.061181	0.068263	-0.025455
Variance	0.006099	0.003218	0.001923	0.001512	0.001364	0.000864	0.001111	0.005601	0.004365
MSE	0.006612	0.006566	0.002040	0.001610	0.001419	0.005252	0.001712	0.021049	0.005520
$\widehat{\sigma}^2_{pavg}(u)$	$u = 0.1$	$u = 0.2$	$u = 0.3$	$u = 0.4$	$u = 0.5$	$u = 0.6$	$u = 0.7$	$u = 0.8$	$u = 0.9$
Bias ($\times 10^{-5}$)	2.030216	1.195927	2.023091	1.232561	2.540906	-0.059015	2.736330	0.318167	1.683519
Relative bias	0.037394	0.028723	0.123496	0.065019	0.221014	-0.006587	0.248608	0.004264	0.045918
Variance ($\times 10^{-8}$)	0.520208	1.325086	1.431340	2.997621	2.462065	5.043301	1.889325	0.909506	0.671999
MSE ($\times 10^{-8}$)	0.561426	1.339389	1.472269	3.012813	2.526627	5.043336	1.964200	0.910518	0.700342

Table 6.2: Numerical bias, relative bias, variance and mean squared error of (adjusted) volatility estimators and intensity estimator based on the smoothing parameters $\mathfrak{b} = 0.008$, $b = 0.008$, and $N \sim 0.012T$ and $H(b)$ and $H(N)$ are 15.

$\widehat{\sigma}^2_{clock,pavg}(u)$	$u = 0.1$	$u = 0.2$	$u = 0.3$	$u = 0.4$	$u = 0.5$	$u = 0.6$	$u = 0.7$	$u = 0.8$	$u = 0.9$
Bias ($\times 10^{-5}$)	3.347455	3.721725	1.615323	0.834484	0.861053	2.646719	0.487952	5.866782	1.309928
Relative bias	0.061656	0.089387	0.098605	0.044020	0.074896	0.295414	0.044332	0.078642	0.035728
Variance ($\times 10^{-8}$)	1.947030	1.812793	0.571093	0.773642	0.384739	0.365068	0.390233	2.939256	1.185987
MSE ($\times 10^{-8}$)	2.059085	1.951305	0.597186	0.780606	0.392153	0.435119	0.392614	3.283447	1.203146
$\tilde{\sigma}^2_{clock,pavg}(u)$	$u = 0.1$	$u = 0.2$	$u = 0.3$	$u = 0.4$	$u = 0.5$	$u = 0.6$	$u = 0.7$	$u = 0.8$	$u = 0.9$
Bias ($\times 10^{-5}$)	3.494410	3.459924	1.621259	0.841284	0.808070	2.679777	0.248889	5.889937	1.135413
Relative bias	0.064362	0.083099	0.098967	0.044379	0.070288	0.299104	0.022612	0.078952	0.030968
Variance ($\times 10^{-8}$)	1.375594	1.110336	0.365442	0.467905	0.283620	0.287664	0.182230	2.495724	0.773142
MSE ($\times 10^{-8}$)	**1.497703**	**1.230047**	**0.391727**	**0.474982**	**0.290150**	**0.359476**	**0.182849**	**2.842637**	**0.786034**
$\widehat{\sigma}^2_{pavg}(u)$	$u = 0.1$	$u = 0.2$	$u = 0.3$	$u = 0.4$	$u = 0.5$	$u = 0.6$	$u = 0.7$	$u = 0.8$	$u = 0.9$
Bias ($\times 10^{-5}$)	2.116333	1.020826	2.187978	1.008003	2.455896	-0.055884	2.432627	0.401883	1.603305
Relative bias	0.038980	0.024517	0.133562	0.053173	0.213620	-0.006237	0.221016	0.005387	0.043730
Variance ($\times 10^{-8}$)	0.349824	0.896734	0.940963	2.072118	1.613509	3.464473	1.216144	0.631533	0.443251
MSE ($\times 10^{-8}$)	0.394613	0.907155	0.988835	2.082278	1.673823	3.464504	1.275321	0.633148	0.468957

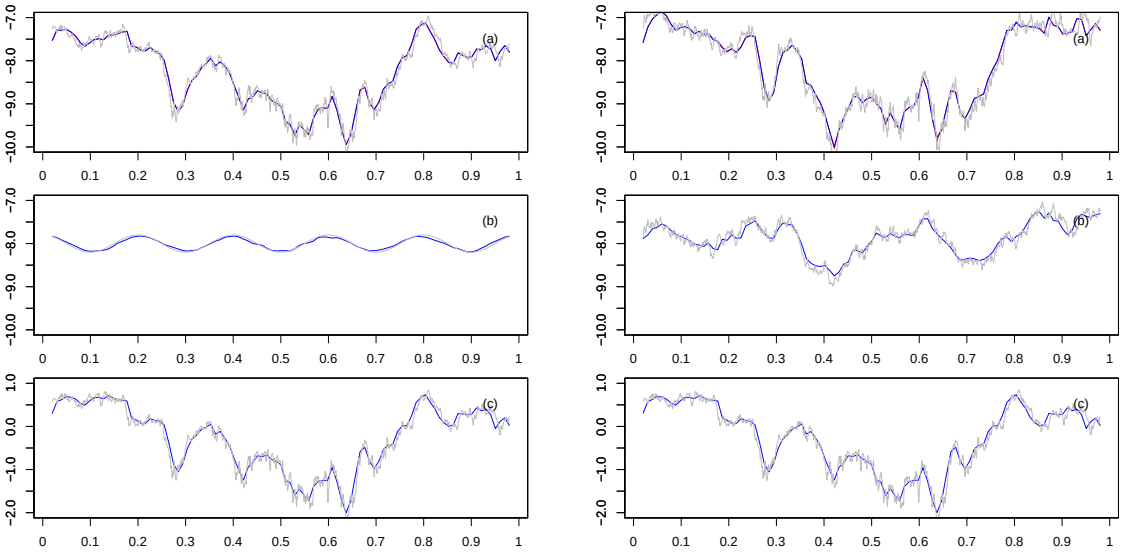


Figure 6.8: The performance of: (a) volatility curve estimators $\log \widehat{\sigma}^{2\ (adj)}_{clock,pavg}(u)$ (red) and $\log \tilde{\sigma}^{2\ (adj)}_{clock,pavg}(u)$ (blue), (b) tick-time volatility estimator $\log \widehat{\sigma}^{2\ (adj)}_{pavg}(u)$ (blue), and (c) intensity curve estimator $\log \widehat{\lambda}(u)$ (blue). These results are based on 500 simulations with $\mathfrak{b} = 0.008$, $b = 0.008$, $N = 0.012T$. The grey lines are the corresponding true parameter curves, where the tick-time volatility is quite smooth in the left column (see Figure 6.1) whereas it is fluctuating in the right column (see Figure 6.3).

Before we end this section, we want to discuss the quality of our volatility estimators in the situation where the standard deviation of additive noise ω varies over 0, 0.00005, 0.0005, 0.001. The last one is regarded as a large effect of disturbance. In Figure 6.9, the pre-averaging estimators $\widehat{\sigma}^{2\ (adj)}_{clock,pavg}(\cdot)$ and $\widehat{\sigma}^{2\ (adj)}_{pavg}(\cdot)$ are drawn together with the standard estimators $\widehat{\sigma}^{2}_{clock}(\cdot)$ and $\widehat{\sigma}^{2}(\cdot)$, which are consistent for the (tick-time) volatility in the noiseless situation. For each experiment, only one iteration of the system is carried out in order to mirror a single real-data application of volatility estimation. Furthermore, the transaction arrivals used here are taken from a real stock: Cisco Systems on April 1, 2014, in order to mimic the real trading intensity. The smoothing parameters and weighting schemes are the same as used in Table 6.1 above.

In the noiseless setting, i.e. for $\omega = 0$, it is clear that the simple (tick-time) realized volatility performs better than the pre-filtering-based estimator in that $\widehat{\sigma}^{2}(\cdot)$ (orange) is everywhere smoother than $\widehat{\sigma}^{2\ (adj)}_{pavg}(\cdot)$ (blue)[9]. The reason for the flutter exhibited by the

[9] Although the latter is robust even in the absence of noise, it is less imprecise than the former one. One of the main reasons that the estimators based on the pre-averaging technique do not look good enough is due to their slower rate of convergence. Thus, one desires to have much more frequent intraday data, e.g. 100,000 trades or more per day for an individual stock in order to extract the spot volatility properly. However, this amount of data seems to be unusual in practice, as ultra-high-frequency data normally consists of about 20,000–40,000 transactions over a trading day.

pre-averaging type estimators is the increase in the finite-sample variance due to the slow rate of convergence. In fact, it has been shown theoretically that the rate of convergence of the standard realized volatility in the noiseless model is much faster than that of the pre-averaging estimators (other estimation techniques dealing with microstructure noise also have the same convergence rate; see Section 2.4).

As the noise level gets higher, the precision of the realized volatility becomes worse, particularly inconsistent, see Figure 6.9 (ii)–(iv). The orange curve in (b) shows that instead of estimating the underlying tick-time volatility (grey), it detects the variance size ω^2 (a well-known result in the literature of market microstructure noise on high-frequency data). More precisely, from the theoretical viewpoint,

$$\widehat{\sigma}^2(u_o) = \frac{T}{H_{1,N}} \sum_{j=i_o-N}^{i_o+N} k\left(\frac{j-i_o}{N}\right) \left(Y_{t_j,T} - Y_{t_{j-1},T}\right)^2 \xrightarrow{\mathbb{P}} \sigma^2(u_o) \; + \; 2T\omega^2,$$

so it can be roughly said that $\log(\widehat{\sigma}^2(u_o)) \approx \log(2T\omega^2)$ in the long run, which is identical to our simulated value, that is in Figure 6.9 (iii) $\log(2T\omega^2) = -4.448$ for $\omega^2 = 0.0005^2$, or in (iv) $\log(2T\omega^2) = -3.06$ for $\omega^2 = 0.001^2$. From the empirical analyses of high-frequency data that we present in the next section, we will see that microstucture noise of real assets can be very small (quite similar to Figure 6.9 (ii) with $\omega = 0.00005$). We can conclude that the pre-averaging estimates are robust to all levels of noise, but their quality is lacking for small sample data. Therefore, it is recommended to apply the pre-averaging technique only when the noise factor is shown to be significant and cannot be discarded from the model (otherwise, the standard realized volatility performs better).

In practice, the market does not allow price processes to be continuous, but it specifies that the price stay on a certain grid, the fineness of which is determined by the tick size–e.g. rounding to the nearest cent. Therefore, it is natural to include this kind of measurement error into the system. This issue is examined by adding a rounding effect into the additive noise model, i.e. we observe a log-price series[10]

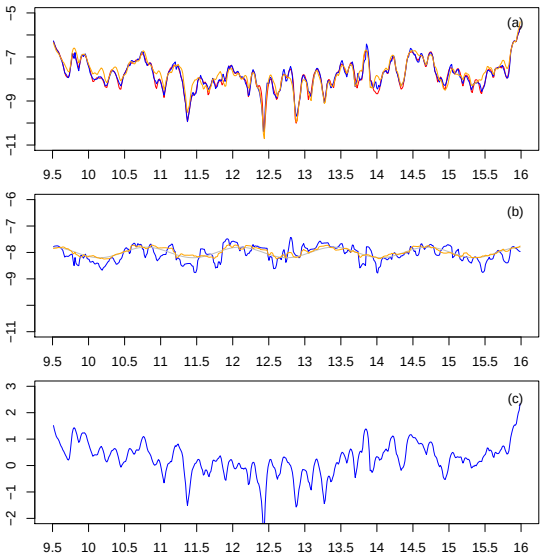
$$Y_{t_i,T} = \; \log\left(\lfloor 100 \cdot (\exp(X_{t_i,T}) + \varepsilon_i)\rfloor/100\right)$$

with $\varepsilon_i \sim \mathcal{N}(0, \omega^2)$. In Figure 6.10, the quality of the pre-averaging estimator in this noisy model is demonstrated[11] and compared with that of a benchmark estimator suggested by Dahlhaus and Neddermeyer [25], denoted by $\widehat{\sigma}^2_{DN}(u)$. This is an online volatility estimator developed for nonlinear market microstructure noise models, particularly under the consideration of rounding noise error.[12]

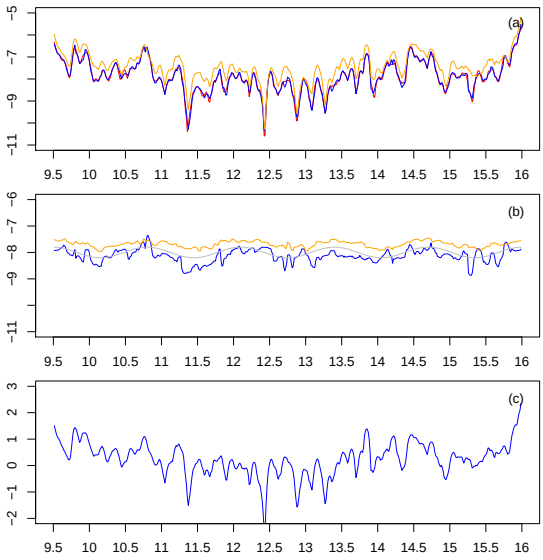
[10] Note that a measurement error is added to the price process, not to the log-price, which is more natural from the real market perspective.

[11] Although we did not prove the consistency of our estimators when noise comes from another source. Nevertheless we will see that the estimation is quite accurate.

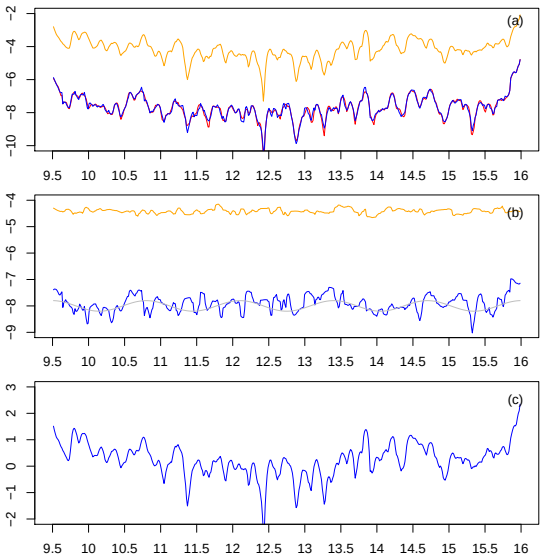
[12] This estimator is constructed by a nonparametric recursive EM algorithm based on an efficient particle filter, which updates the results immediately after the occurrence of a new transaction. The selection of their tuning parameter is also automatic, adapting to the roughness of the estimated curves. We thank the authors for providing the source code of the green estimator used here, which is adapted to our simulations.



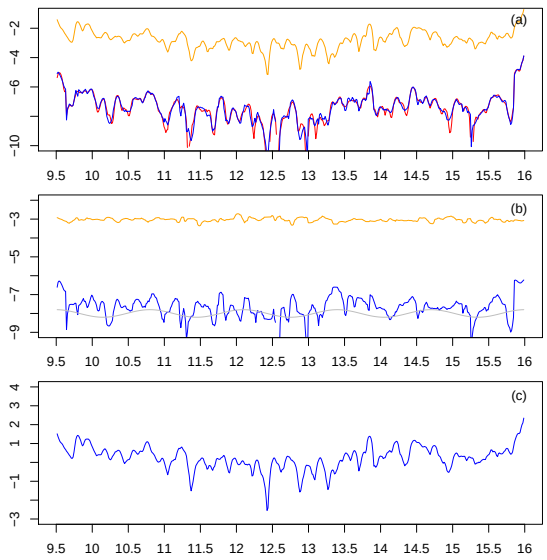
(i) $\omega = 0$



(ii) $\omega = 0.00005$



(iii) $\omega = 0.0005$



(iv) $\omega = 0.001$

Figure 6.9: Comparison between: (a) pre-averaging estimates $\log \widehat{\sigma}^{2\ (adj)}_{clock,pavg}(u)$ (red), $\log \tilde{\sigma}^{2\ (adj)}_{clock,pavg}(u)$ (blue) vs. realized volatility $\log \widehat{\sigma}^2_{clock}(u)$ (orange) in additive noise models, (b) $\log \widehat{\sigma}^{2\ (adj)}_{pavg}(u)$ (blue) vs. $\log \widehat{\sigma}^2(u)$ (orange). In (b), the grey line is the true tick-time volatility curve. In (c), the intensity estimate is derived from real transaction data: Cisco systems on April 1, 2014. The market microstructure noise is assumed to be $\mathcal{N}(0, \omega^2)$.

From Figure 6.10 (i)–(iii), we see that in (b), the pre-averaging estimate (blue) is robust to this kind of noise even when the additive noise factor is of high level. The green estimator is, however, robust to small additive noise with a rounding effect. Whenever the green estimator is justified, it can detect the underlying curve (grey) better than the oscillating blue estimator; see Figure 6.10 (i) and (ii).

This remark emphasizes again that the pre-averaging estimators are robust to many different types of noise, but their performance is poor in finite-sample experiments due to their slow rate of convergence. On the contrary, the Dahlhaus and Neddermeyer estimator is robust and can capture the shape of the volatility better in the rounding plus (small) additive noise models. It is unfortunately non-robust to large variances of noise, see Figure 6.10 (iii). We will see in the next section when discussing the size of market microstructure noise in reality that the volatility plot pattern looks very similar to that in Figure 6.10 (ii). Thus, the smoothness of the green estimator more closely resembles the smoothness of the unknown curve. In particular, this confirms that tick-time volatility is usually much smoother than clock-time volatility and trading intensity in real applications; see the next section.

6.2 Real Data Analysis

In this section, we conduct some empirical analyses of high-freqeuncy financial data provided by the data vendor *QuantQoute TickView*. We analyze the intraday transaction data listed in the Dow Jones Industrial Average (DJIA), having 40 symbols, from April 1–30, 2014.[13] As is typical of transaction data, we do not have the form of the measurement error regarding the construction of the price, which is seen in e.g. 1-minute data (this can be built by using linear interpolation or previous price, etc). It is clear that the number of transactions of each stock on each day is different. Therefore, we list these values for all stocks for April 1, 2014 in Table 6.3 to get an idea of how large they are for common stocks. We see that heavily traded stocks typically have upwards of 15,000 transactions per day; such ultra-high-frequency data is suitable for our real data analysis in order to estimate the volatility curve (see the previous section). Barndorff-Nielsen *et al.* [12] have shown that the size of microstructure noise on each market exchange could differ. Thus, we will mainly investigate the properties of high-frequency data that is traded on only a single exchange, the NASDAQ stock exchange.

[13] A transaction price is recorded whenever a transaction is completed (no matter whether the price has changed or not). Some authors use the term *tick data* for transaction/quotation data, which could lead to confusion since *tick data* normally refers to price data that is recorded when there is a price change of at least one tick (how large a tick depends on the stock and the market). Note that in our terminology, we always use the term *tick time* as a short form of *transaction time*.

Table 6.3: Number of transactions of stocks from DJIA for April 1, 2014, traded on NYSE and NASDAQ.

TICKER	NASDAQ	NYSE	TICKER	NASDAQ	NYSE	TICKER	NASDAQ	NYSE
AA	8561	4287	HD	8296	4205	NKE	4341	2846
AIG	7470	6295	HON	4162	2400	PFE	9324	5530
AXP	6433	3642	HPQ	11384	4672	PG	9851	5843
BA	7793	3749	IBM	11407	5378	PGP	-	-
BAC	8790	4609	INTC	13827	-	T	13964	5365
C	20144	7716	JNJ	10178	5851	TRV	3256	2788
CAT	8204	4294	JPM	16645	5701	UNH	4815	3375
CSCO	35056	-	KO	11483	5523	UTX	7426	3622
CVX	9201	6275	MCD	7056	4146	V	5852	2835
DD	5802	3160	MDLZ	11621	-	VZ	13170	5746
DIS	8561	6616	MMM	4348	3535	WMT	9367	6328
GE	7843	5640	MO	4081	2230	XOM	15479	6585
GM	31044	9423	MRK	9147	4632			
GS	7727	3409	MSFT	25198	-			

time, price, size, exchange, condition, suspicious	
34271000, 129400, 100, Z,	@, 0
34271000, 129400, 400, T,	@, 0
34271000, 129400, 100, B,	F, 0
34271000, 129400, 100, T,	@, 0
34271000, 129400, 900, N,	@, 0
34271000, 129400, 500, B,	@, 0
34271000, 129400, 100, N,	@, 0
34271000, 129450, 100, D,	@, 0
34271000, 129400, 400, N,	@, 0

This is an example of transaction data consisting of the following information: trading time, transaction price, volume, market exchange and sales conditions. In general we clean raw data before analyzing it in the following main steps: i) deleting all pre- and after-market data, i.e. only transactions between 09:30 AM and 04:00 PM are considered; ii) choosing only transactions traded on the NASDAQ stock exchange so as not to mix the influence of microstructure noise from different exchanges; iii) filtering raw data from the outliers, such as price errors, and deleting entries with abnormal sale conditions (see QuantQuote's documentation for full description). Although the accuracy of transaction time is down to milliseconds, the resolution of timestamps is limited to only one second, which leads to the possibility of having multiple consecutive transactions occurring at the same time. Nevertheless, the order of the trades is correctly placed by the data vendor. In the case of multiple trades, which is often the case for liquidly traded equities, these time points will be separated into equally-spaced times, for example t_{10}, t_{11} and t_{12} occurring at time 34210 (= 09:30:10 AM) are adjusted to $t_{10} = 34210$, $t_{11} = 34210.33$ and $t_{12} = 34210.67$.[14] We note again that there are 6.5 market hours in a trading day, which is equal to $T = 23,400$ seconds.

It has been pointed out in practice that market microstructure noise is time-dependent and correlated with the efficient price. However, when one considers data with transaction-time sampling, the noise becomes independent;[15] see Griffin and Oomen [37] and Hansen

[14] We will compare our estimates with the results of the Dahlhaus and Neddermeyer estimator, whose timestamps are adjusted.

[15] Note that this is not true for other time sampling schemes, such as calendar-time sampling (e.g. 1-minute

and Lunde [39]. To confirm this empirical fact for our ultra-high-frequency data in 2014, we look at the autocorrelation function (ACF) of log-returns $Y_{t_i,T}-Y_{t_{i-1},T}$ (not the price itself) of Cisco Systems for 20 consecutive trading days from April 1 - 29, 2014. We note that similar results were obtained for other stocks and are thus not reported here. The ACF plots in Figure 6.11 suggest that the averaged first-order autocorrelation is considerable, and that other higher-order ACFs are mostly (statistically) insignificant. This result supports our independent noise assumption in the model setting, since $\mathrm{Cov}(Y_{t_i,T} - Y_{t_{i-1},T}\ ,\ Y_{t_{i-h},T} - Y_{t_{i-h-1},T}) = 0$ for each $i \in \mathbb{N}$ and $h \geq 2$ if the noise ε_i is an independent sequence. As a consequence, assuming the noise is of an i.i.d. type seems to be reasonable when transaction data are being considered. For more characteristics of microstruture noise from the practical perspective, we refer to Hansen and Lunde [39] and Hasbrouck [44].

To extract the underlying intraday spot volatility from price data, we use the pre-averaging estimators presented in Section 3.2. The distinction between these estimators and the estimators given in the infill-asymptotics section lies in the factor $1/T$, that is

$$\widehat{\sigma}^2_{clock,pavg}(t,M) \ = \ \widehat{\sigma}^2_{clock,pavg}\left(\frac{t}{T}\right)\cdot\frac{1}{T}$$

and

$$\tilde{\sigma}^2_{clock,pavg}(t,m,M) \ = \ \widehat{\sigma}^2_{pavg}(t,m)\cdot\widehat{\lambda}(t,M) \ = \ \widehat{\sigma}^2_{pavg}\left(\frac{t}{T}\right)\cdot\frac{1}{T}\cdot\widehat{\lambda}\left(\frac{t}{T}\right),$$

for $t \in [0,T]$ if we set $M = bT$ and $m = N$. For the sake of comparability, we select on average the same size of smoothing parameters for M and m. In particular, the roughness of the transaction-time volatility can be directly compared to the fluctuation of the (clock-time) volatility and intensity curves.

For the investigation of high-liquid stocks, the choice of tuning parameters are as follows: $M = 250$, $m \sim 250$ and $H = 15$.[16] The capital M can be seen as the window size used for the clock-timescale estimators while the small m is the window size used for the tick-timescale estimation[17]. Therefore, the latter can differ from the former, as the number of transactions over each period of time is random. The choice $M = 250$ is ad hoc, although it is confirmed through our experimentation among many bandwidth choices to be, in our opinion, appropriate, as it does not oversmooth or undersmooth the parameter curves and is still robust against microstructure noise (tested in our experiments). Of course, one still desires to have an adaptive bandwidth approach for volatility estimation which could detect the smoothness of the curve for each time period, as discussed in the preceding section. The weighting functions applied here are the same as in the simulation part, that is $\mathfrak{K}(x)$, $K(x)$, and $k(x)$ are Epanechnikov kernels $\frac{3}{4}\left(1-x^2\right)I_{\{|x|\leq 1\}}$ and the weighting function $g(x)$ in the pre-averaging steps is $g(x) = x(1-x)I_{\{0\leq x\leq 1\}}$.

---

or 5-minute data), quotation-time sampling, or business-time sampling; see Hansen and Lunde [39].

[16] See Section 6.1 for an explanation of the notation $m \sim 250$.

[17] For $\widehat{\sigma}^2_{clock,pavg}(t_o,M)$, all observed pre-averaging terms $\overline{\triangle Y}_{t_i}$ inside the interval $(t_o - M, t_o + M]$ contribute to the estimator at time t_o, whereas exactly m observed pre-averaging terms from the left- and the right-hand side of t_o contribute to the estimator $\widehat{\sigma}^2_{pavg}(t_o,m)$; therefore it does not depend on the transaction intensity.

We do not report the results of all stocks listed in the DJIA but only those of five ultra-high-frequency stocks: Cisco System (CSCO), Citigroup (C), General Motors (GM), Microsoft Corporation (MSFT), and Intel Corporation (INTC). Figures 6.12–6.16 show the analyses of the intraday transaction data for these stocks traded on April 1–4, 2014; the number of transactions are stated in the bracket for each stock and day. As usual, (a) always depicts the plot of logarithm of clock-time volatility estimates $\log \widehat{\sigma}^2_{clock,pavg}(\cdot, M)$ (red) and $\log \tilde{\sigma}^2_{clock,pavg}(\cdot, m, M)$ (blue). The log of tick-time volatility estimate $\log \widehat{\sigma}^2_{pavg}(\cdot, m)$ (blue) are visualized in (b) along with the Dahlhaus and Neddermeyer estimator (green)[18]. The orange line in (b) is the standard tick-time realized volatility $\log \widehat{\sigma}^2(\cdot)$, which is robust only in the noiseless model. In (c), the log of the intensity estimate $\log \widehat{\lambda}(\cdot, M)$(blue) is drawn. Remind that all plots are on a logarithmic scale so that the blue lines in (b) and (c) sum up to the blue line in (a) due to the volatility factorization, i.e.

$$\log \tilde{\sigma}^2_{clock,pavg}(\cdot, m, M) \;=\; \log \widehat{\sigma}^2_{pavg}(\cdot, m) \;+\; \log \widehat{\lambda}(\cdot, M).$$

In order to deal with boundary effects caused by the lack of data at the borders, we apply the local linear fit (over a small number of points at the borders only), which automatically corrects biases.

Our main interest is in the behavior of volatility plots over a trading day. From Figures 6.12–6.16 we see that the volatility decomposition gives us some insight into the cause and structure of volatility. In particular, the results from these four days highlight some interesting issues:

i) row (a) shows that the alternative estimator based on the volatility decomposition (blue curve) nicely coincides with the classical clock-time estimator (red curve). This blue estimator is the sum of the blue estimators in rows (b) and (c) (in log-scale);

ii) the typical volatility smile (U-shape) over a trading day is mainly driven by the trading intensity in row (c). The tick-time volatility is normally high at the beginning of trading days and often decreases at the end of the day, which we see in plot (b);

iii) the tick-time volatility estimator $\log \widehat{\sigma}^2_{pavg}(t, m)$ in row (b) is in general smoother than the clock-time estimator $\log \widehat{\sigma}^2_{clock,pavg}(t, m, M)$ in row (a) and the intensity estimator $\log \widehat{\lambda}(t, M)$ in row (c) for ultra-high-frequency data, i.e. the fluctuation in trading intensity in row (c) is the major source of the fluctuation in clock-time volatility in row (a);

iv) the decomposition enables us to determine to a certain extent the source of volatility changes; for example, in Figure 6.16 the peak in GM on 04/01/2014 at 15 is mostly due to the spike of intensity (i.e. most likely due to some general—not company related—news), or similarly the sudden decrease of the GM on 04/03/2014 at 11.2 is the effect of tick-time volatility;

[18]The resulting curve needs to be corrected in our plot, as this estimator uses a one-sided window in contrast to our two-sided kernel-based estimators. Therefore, the green curve is shifted to the left by a factor depending on the number of trades; more details can be found in their paper [25].

v) in row (b), the green and blue lines have almost the same level of volatility, similarly to their behavior in Figure 6.10 (i) and (ii) of the previous simulation study. This indicates that the market microstructure noise must be very small[19]. In particular, it confirms again that tick-time volatility is usually much smoother than clock-time volatility and trading intensity.

In most of our examples for highly liquid stocks, the tick-time volatility in row (b) is considerably smoother than the trading intensity in row (c) during the days, for example in CSCO between 10 and 13 on 04/01/2014 or between 10 and 12.5 on 04/03/2014. Apart from its practical interpretation, this feature also has an important consequence for estimation: microstructure noise only affects the smoother curve (b) and not (c), i.e. coping with microstructure noise becomes easier since we may choose a larger bandwidth with effectively more data than with the classical estimator (red curve in (a)). Mathematically this is reflected in a higher rate of convergence of the estimator (see Chapter 4 and in particular Table 4.2). We mention again that in these figures, we have chosen on average the same bandwidth size for all blue and red estimators for the sake of comparability.

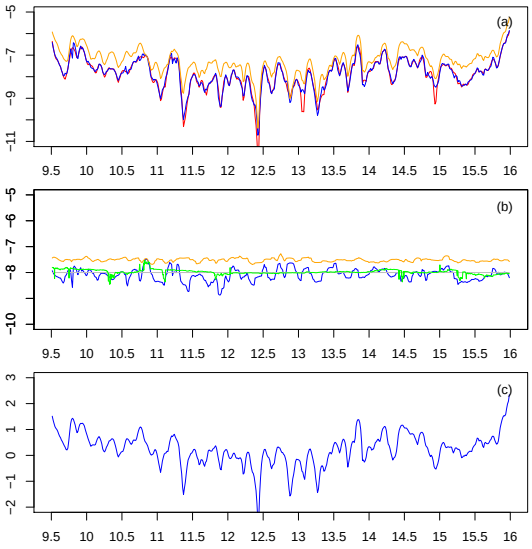
From the asymptotic results, we have seen that market microstructure noise plays a crucial role in the model. In practice, however, at least for the stocks from DJIA in April 2014, the influence of this noise on the spot volatility estimation was shown to be quite small, which coincides with one of the facts stated in Hansen and Lunde [39]. Moreover, on all stocks and days considered here, our pre-filtering-based estimators resulted in only positive values, even though the technique itself does not guarantee this (since the second term of pre-averaging estimators is subtracted from the first term to eliminate the bias induced by noise). This also supports the empirical observation that the standard deviation of the measurement error is very small in practice; if it is large, the second term will dominate the first, leading to negative values of the pre-averaging estimators. In fact, standard deviation of the measurement error is so small that one can hardly distinguish between noise-robust estimates (blue or green curves) and the standard noiseless estimate (orange) of the spot volatility. Despite this fact, the noise cannot be neglected, especially when one considers integrated volatility over a longer time period.

Finally, we present the application of these estimators on less liquid assets. Figure 6.17 visualizes the result of spot volatility estimation of GM traded on the NYSE, which is to some extent different from when it is traded on the NASDAQ; transactions on the NASDAQ occur more frequently than on the NYSE, typically 2–3 times over a trading day.

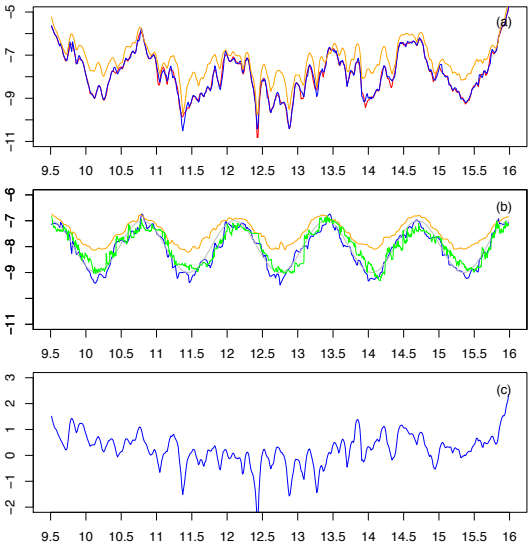
Comparing the results of both exchanges, we see that the estimated clock-time volatility curve on the different exchanges are almost the same. The shapes of both intensity curves

[19] so that all estimates lead to almost the same shape and level, even though they are based on different types of noise models. As was discussed, the quality of pre-averaging estimation (in the sense of variance) is not as good as that of the realized volatility (orange) or that of the Dahlhaus and Neddermeyer estimate (green). We conjecture that some of the large-amplitude oscillation of the blue line in (b) owes to this concern.

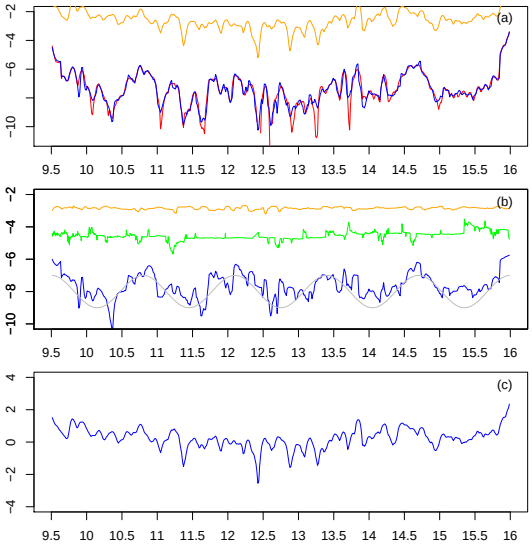
vary in essentially similar ways, whereas the tick-time volatility estimate on the NYSE is quite rough, owing to the poor quality of pre-averaging estimates on small sample sizes. In fact, as the number of transactions on the NYSE are fewer than on the NASDAQ for this particular example, each transaction arising on the NYSE will incorporate the cumulative information of several trades occurring on the NASDAQ, so that the price change at every tick of the latter evolves more slowly than that of the former.



(i) With $\sigma^2(u) = \exp(-8)$.



(ii) With $\sigma^2(u) = \exp(-8 + \cos(10u\pi))$. The model contains noise with small variance $\omega^2 = 0.001^2$.



(iii) With $\sigma^2(u) = \exp(-8 + \cos(10u\pi))$. The model contains noise with large variance $\omega^2 = 0.05^2$.

Figure 6.10: Comparison between: (a) pre-averaging estimates $\log \widehat{\sigma}^{2\ (adj)}_{clock,pavg}(u)$ (red), $\log \tilde{\sigma}^{2\ (adj)}_{clock,pavg}(u)$ (blue) and realized volatility $\log \widehat{\sigma}^2_{clock}(u)$ (orange) in additive plus rounding noise models; (b) $\log \widehat{\sigma}^2_{pavg}(u)$ (blue) vs. $\log \widehat{\sigma}^2(u)$ (orange) vs. $\log \widehat{\sigma}^2_{DN}(u)$ (green). In (c), the intensity estimate is derived from real transaction time of Cisco Systems on April 1, 2014. The additive noise is assumed to be $\mathcal{N}(0, \omega^2)$, and the rounding error is of one cent. The grey line in (b) is the true tick-time volatility curve: for (i) $\sigma^2(u) = \exp(-8)$, and for (ii)-(iii) $\sigma^2(u) = \exp(-8 + \cos(10u\pi))$.

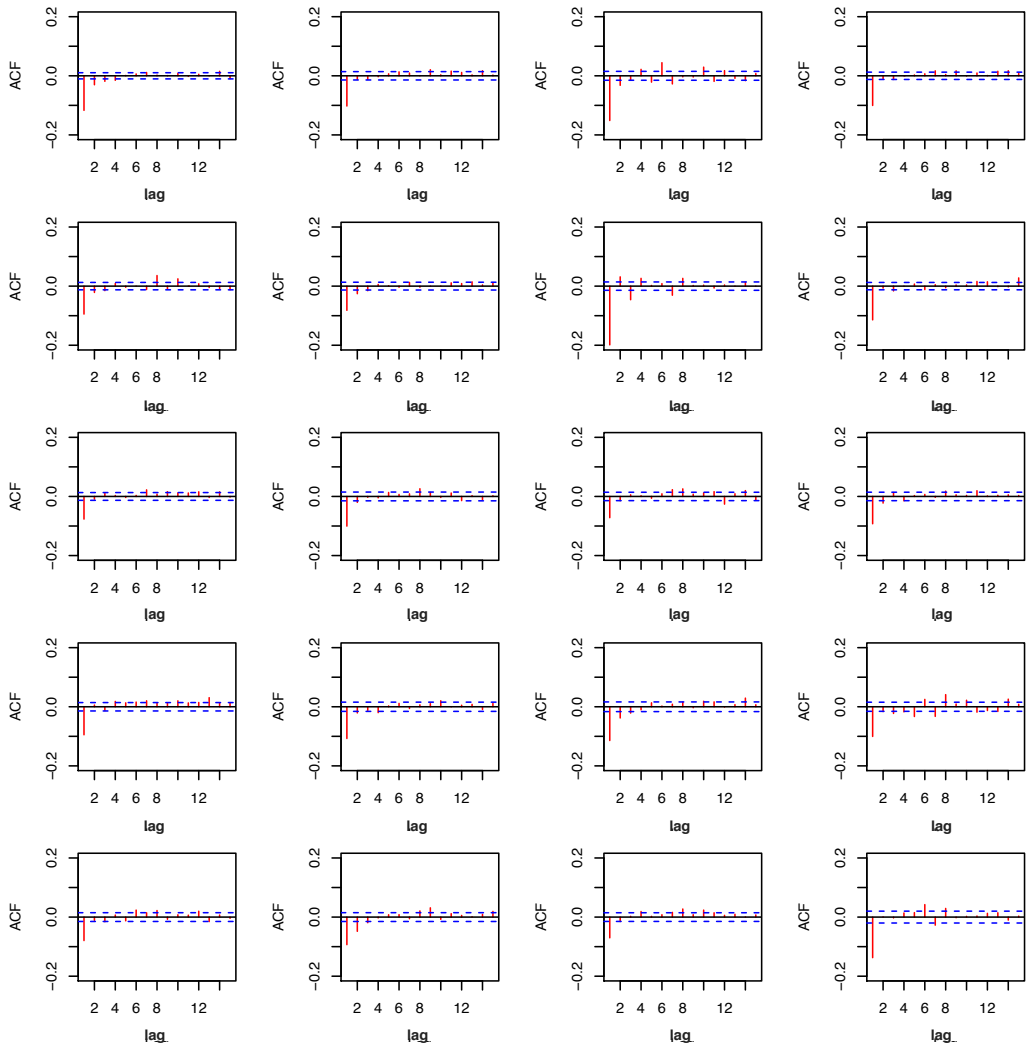
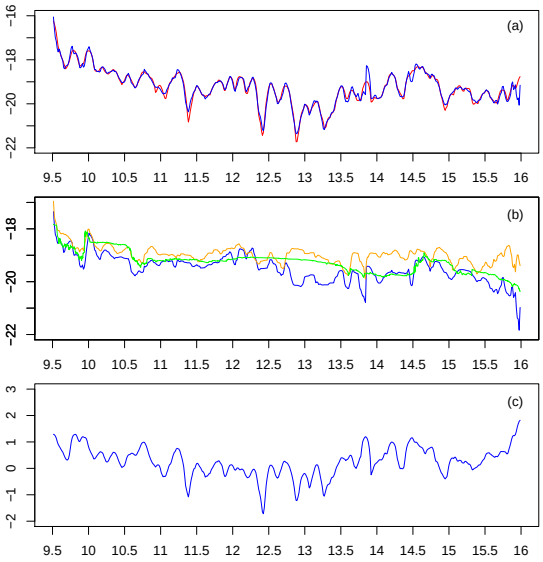
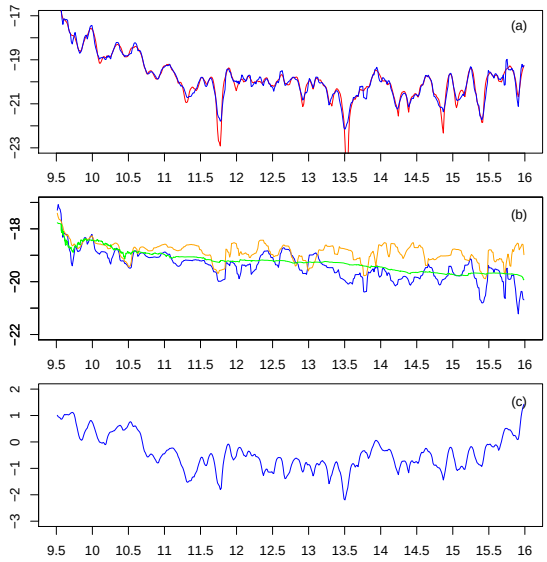


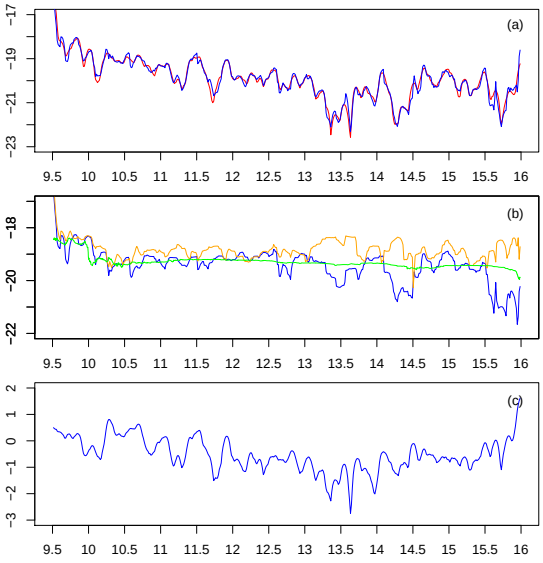
Figure 6.11: ACF of log-returns of CSCO from April 1-29, 2014 (from upper left to bottom right). The blue dashed line represents an approximate 95% confidence interval. These ACF plots suggest that the noise process is significantly correlated with only a lag of one transaction.



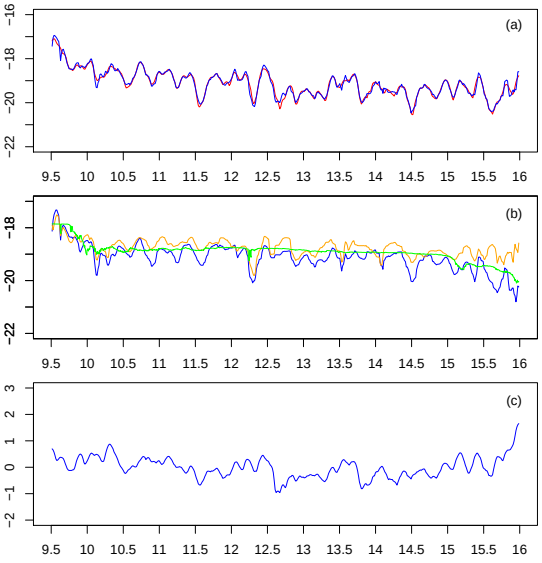
(i) CSCO on 04/01/2014 (35,056 transactions)



(ii) CSCO on 04/02/2014 (19,301 transactions)

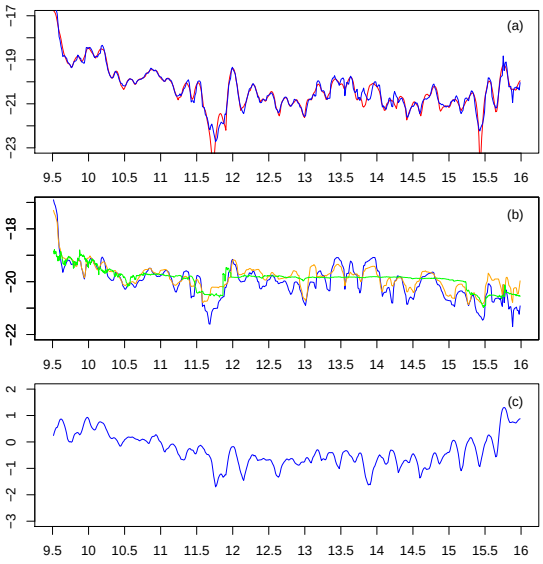


(iii) CSCO on 04/03/2014 (17,030 transactions)

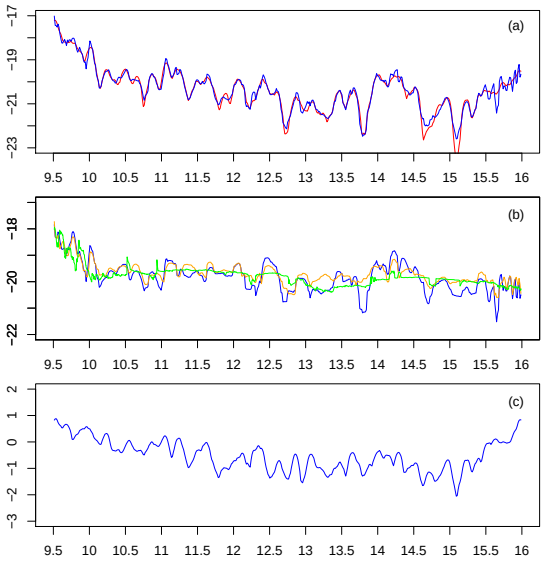


(iv) CSCO on 04/04/2014 (25,266 transactions)

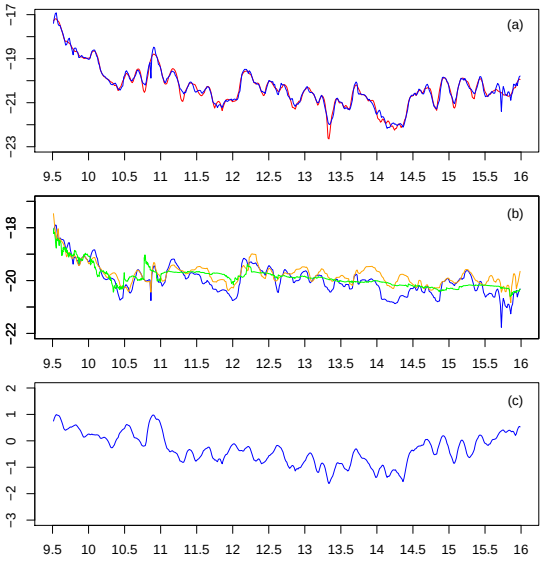
Figure 6.12: Cisco Systems (CSCO) traded on April 1 - 4, 2014 on NASDAQ. The clock-time volatility estimates $\widehat{\sigma}^2_{clock,pavg}(t, M)$ (red) and $\tilde{\sigma}^2_{clock,pavg}(t, m, M)$ (blue) are plotted in (a). In (b), the transaction-time volatility estimate $\widehat{\sigma}^2_{pavg}(t, m)$ (blue), the noiseless tick-time realized volatility (orange) and the Dahlhaus and Neddermeyer estimate (green) are compared. In (c), the intensity curve estimate (blue) is drawn.



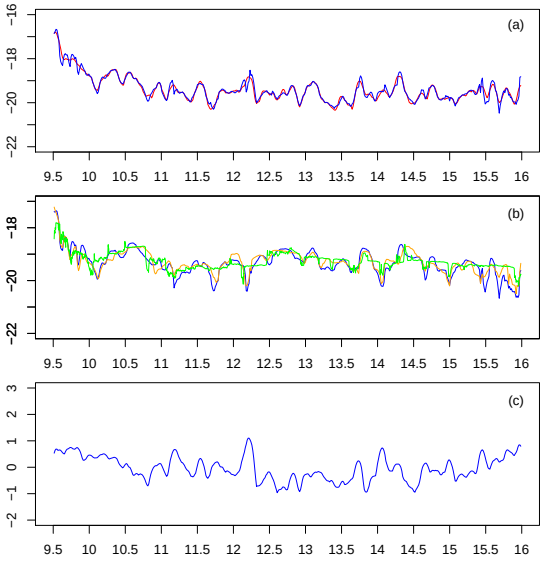
(i) C on 04/01/2014 (20,144 transactions)



(ii) C on 04/02/2014 (16,408 transactions)

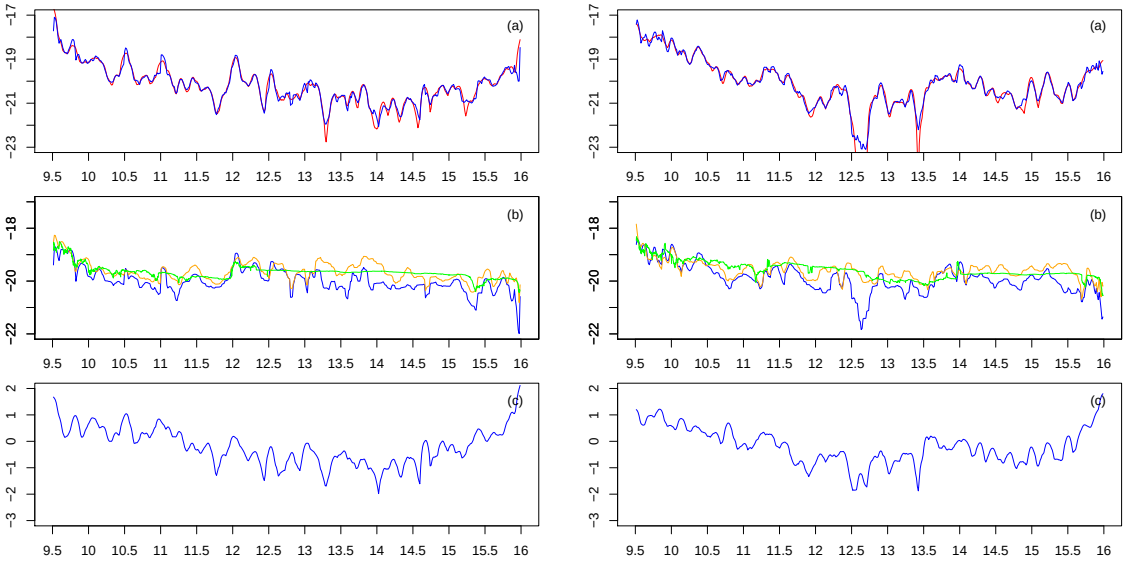


(iii) C on 04/03/2014 (20,050 transactions)



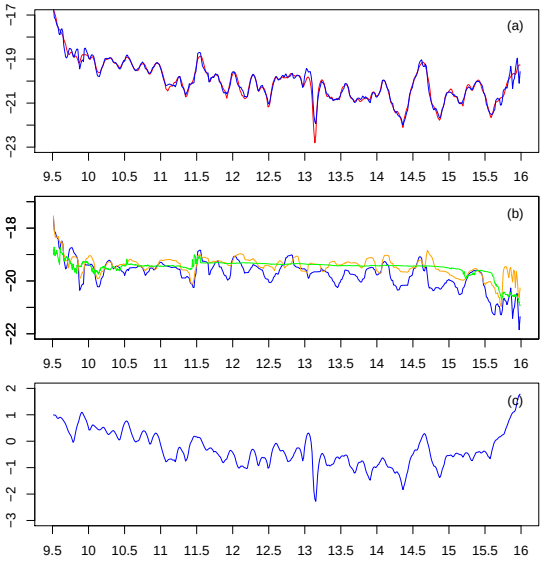
(iv) C on 04/04/2014 (24,723 transactions)

Figure 6.13: Citigroup (C) traded on April 1 - 4, 2014 on NASDAQ. The clock-time volatility estimates $\widehat{\sigma}^2_{clock,pavg}(t,M)$ (red) and $\tilde{\sigma}^2_{clock,pavg}(t,m,M)$ (blue) are plotted in (a). In (b), the transaction-time volatility estimate $\widehat{\sigma}^2_{pavg}(t,m)$ (blue), the noiseless tick-time realized volatility (orange) and the Dahlhaus and Neddermeyer estimate (green) are compared. In (c), the intensity curve estimate (blue) is drawn.

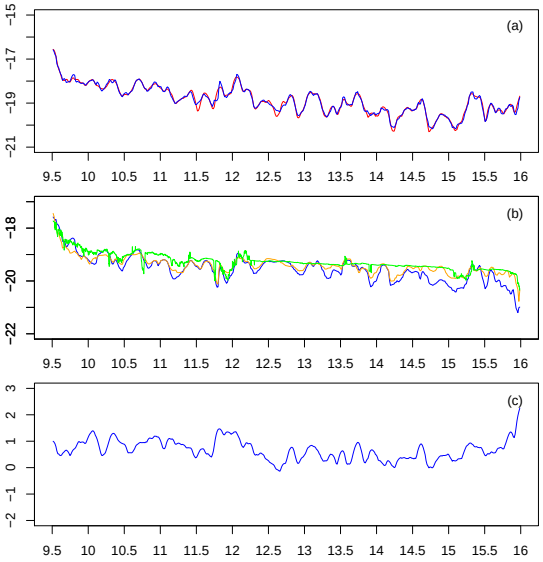


(i) MSFT on 04/01/2014 (25,198 transactions)

(ii) MSFT on 04/02/2014 (24,295 transactions)

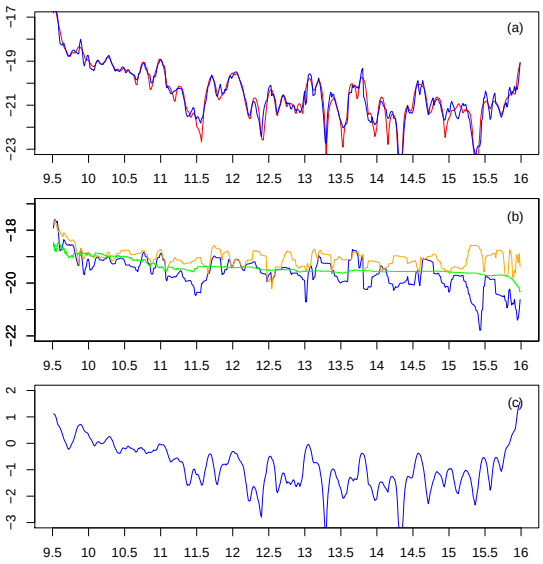


(iii) MSFT on 04/03/2014 (21,687 transactions)

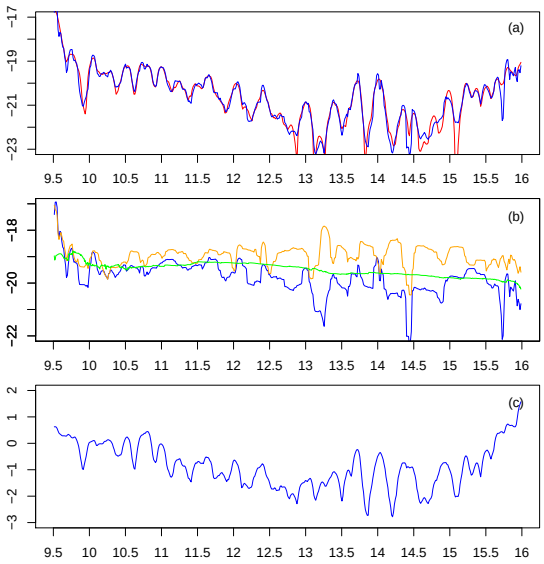


(iv) MSFT on 04/04/2014 (49,521 transactions)

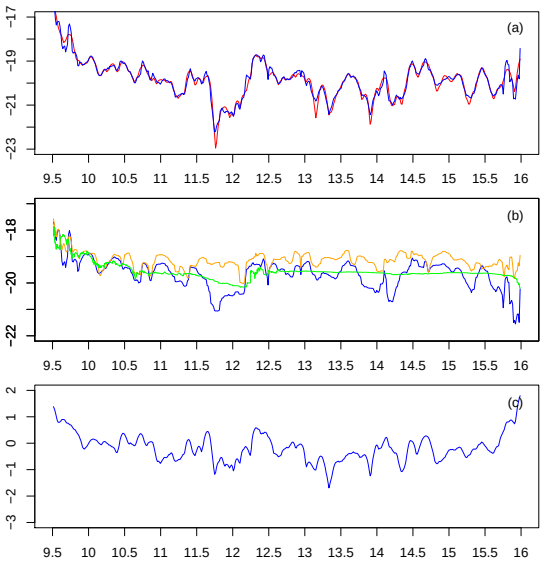
Figure 6.14: Microsoft Corporation (MSFT) traded on April 1 - 4, 2014 on NASDAQ. The clock-time volatility estimates $\widehat{\sigma}^2_{clock,pavg}(t, M)$ (red) and $\widetilde{\sigma}^2_{clock,pavg}(t, m, M)$ (blue) are plotted in (a). In (b), the transaction-time volatility estimate $\widehat{\sigma}^2_{pavg}(t, m)$ (blue), the noiseless tick-time realized volatility (orange) and the Dahlhaus and Neddermeyer estimate (green) are compared. In (c), the intensity curve estimate (blue) is drawn.



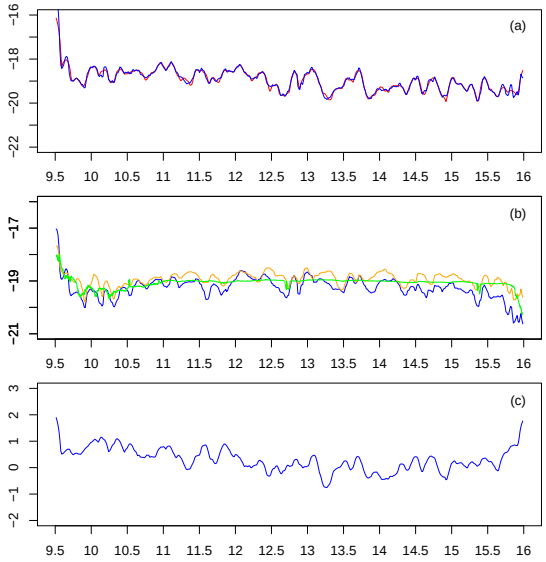
(i) INTC on 04/01/2014 (13,827 transactions)



(ii) INTC on 04/02/2014 (14,017 transactions)

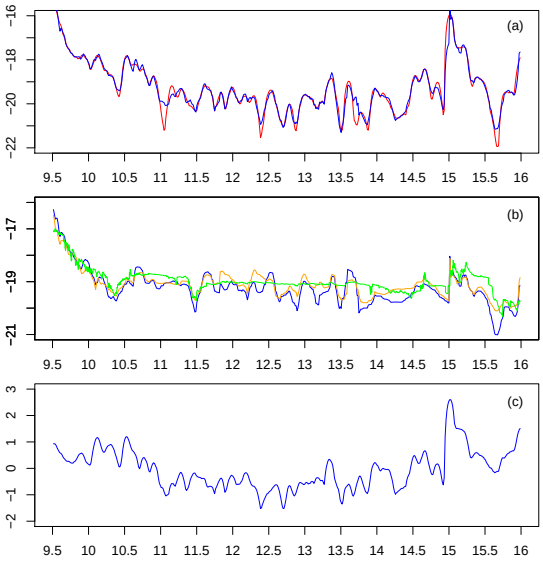


(iii) INTC on 04/03/2014 (22,754 transactions)

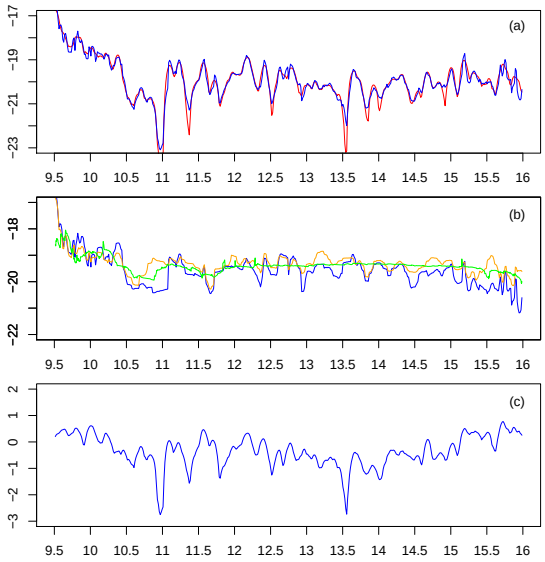


(iv) INTC on 04/04/2014 (35,158 transactions)

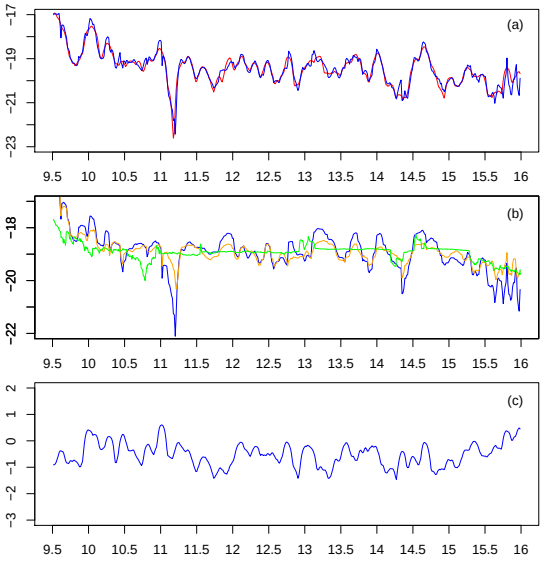
Figure 6.15: Intel Corporation (INTC) traded on April 1 - 4, 2014 on NASDAQ. The clock-time volatility estimates $\widehat{\sigma}^2_{clock,pavg}(t, M)$ (red) and $\tilde{\sigma}^2_{clock,pavg}(t, m, M)$ (blue) are plotted in (a). In (b), the transaction-time volatility estimate $\widehat{\sigma}^2_{pavg}(t, m)$ (blue), the noiseless tick-time realized volatility (orange) and the Dahlhaus and Neddermeyer estimate (green) are compared. In (c), the intensity curve estimate (blue) is drawn.



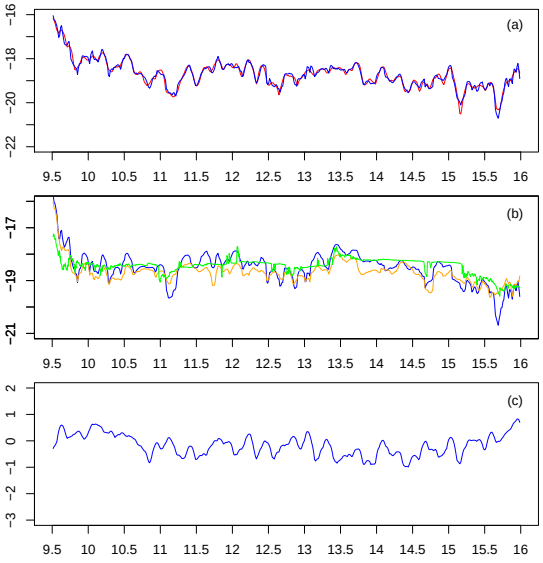
(i) GM on 04/01/2014 (31,044 transactions)



(ii) GM on 04/02/2014 (19,226 transactions)

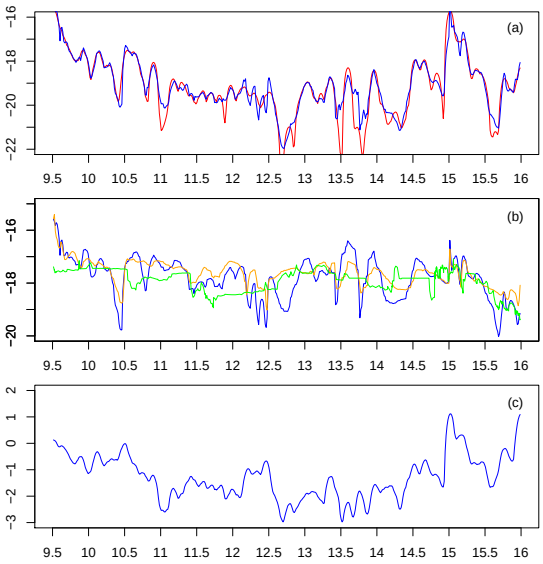


(iii) GM on 04/03/2014 (15,349 transactions)

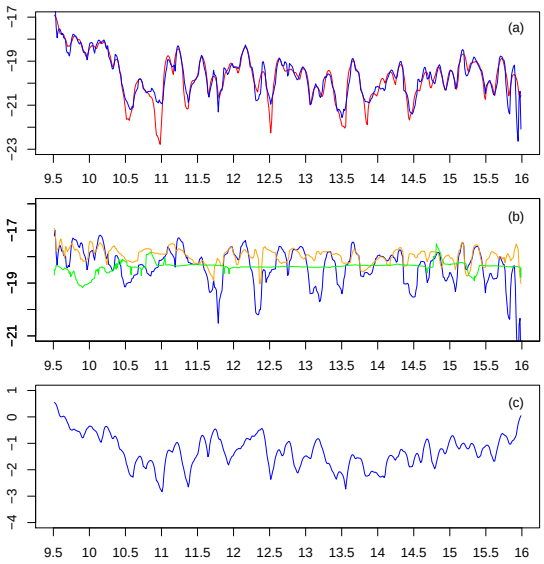


(iv) GM on 04/04/2014 (21,118 transactions)

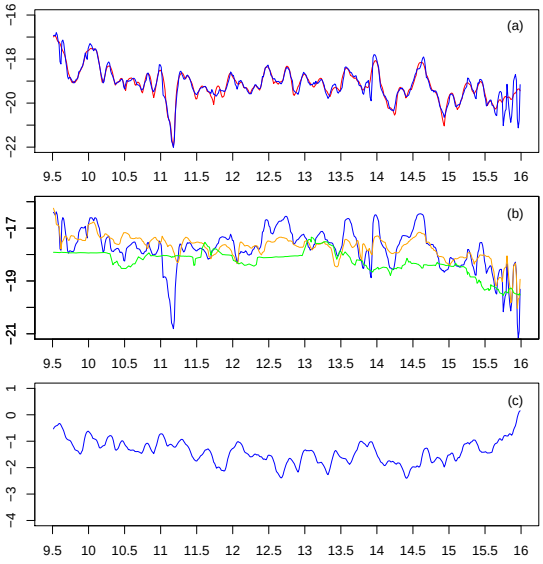
Figure 6.16: General Motors (GM) traded on April 1 - 4, 2014 on NASDAQ. The clock-time volatility estimates $\widehat{\sigma}^2_{clock,pavg}(t, M)$ (red) and $\tilde{\sigma}^2_{clock,pavg}(t, m, M)$(blue) are plotted in (a). In (b), the transaction-time volatility estimate $\widehat{\sigma}^2_{pavg}(t, m)$ (blue), the noiseless tick-time realized volatility (orange) and the Dahlhaus and Neddermeyer estimate (green) are compared. In (c), the intensity curve estimate (blue) is drawn.



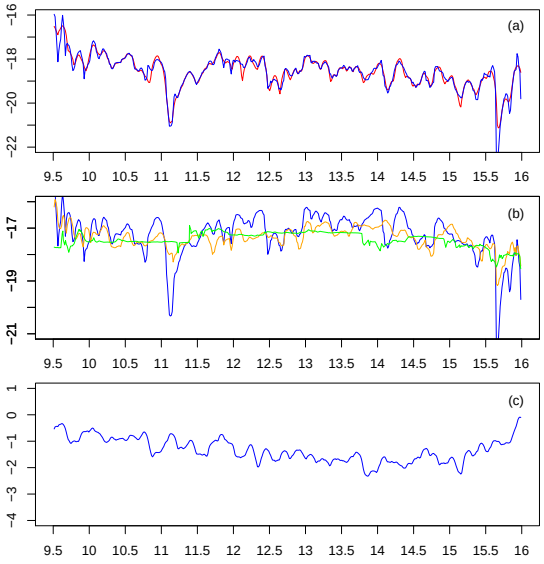
(i) GM on 04/01/2014 (9,423 transactions)



(ii) GM on 04/02/2014 (7,737 transactions)



(iii) GM on 04/03/2014 (6,419 transactions)



(iv) GM on 04/04/2014 (6,830 transactions)

Figure 6.17: General Motors (GM) traded on April 1 - 4, 2014 on NYSE (having less liquidity than NASDAQ). The clock-time volatility estimates $\widehat{\sigma}^2_{clock,pavg}(t, M)$ (red) and $\tilde{\sigma}^2_{clock,pavg}(t, m, M)$ (blue) are plotted in (a). In (b), the transaction-time volatility estimate $\widehat{\sigma}^2_{pavg}(t, m)$ (blue), the noiseless tick-time realized volatility (orange) and the Dahlhaus and Neddermeyer estimate (green) are compared. In (c), the intensity curve estimate (blue) is drawn.

Chapter 7

Conclusion

In this dissertation we have advocated the use of a spot volatility estimator based on a volatility decomposition in a time-changed price-model according to the trading times. In this model clock-time volatility is factorized into the product of two curves, namely tick-time volatility and trading intensity. Both curves can be identified and we have argued that both curves contain valuable information about the original volatility curve. For example, the volatility smile and the increase of volatility at the end of the trading day are in our opinion solely features of trading intensity, while the influence of company related news mainly impacts tick-time volatility or both curves.

A finding which is important in our view is that the tick-time volatility curve is often much smoother than the clock-time volatility curve. This means that the major part of fluctuations in clock-time volatility is due to fluctuations in the trading intensity. These findings also have important implications for statistical inference: microstructure noise does not influence the estimate of trading intensity but influences only the estimator of tick-time volatility. Since this usually is the smoother curve, we may choose a larger bandwidth, resulting in a better rate of convergence and helping the estimator to better cope with microstructure noise. In particular, one may outperform the rate of convergence of the optimal estimator in the classical diffusion model.

For the mathematical investigation of this model we have introduced an infill asymptotic approach and derived the asymptotic properties of the new estimator in the case of a deterministic volatility curve and a deterministic intensity curve of a point process. This can be extended to account for stochastic volatility and intensity processes and also for a leverage effect between price process, price variation, and market activities as discussed in Chapter 5. A comprehensive treatment of this situation seems challenging and is beyond the scope of this paper. From an applied point of view it is probably also of great interest to find proper models where the trading intensity depends on past log-prices and where the volatility has some GARCH-type structure—possibly depending, in addition, on the past intensity of the point process.

Appendices

A. Proofs for Chapter 3

Proof of Proposition 3.2. The continuous-time version of the model under above assumptions is given by $X_t = \sum_{t_i \leq t} \sigma_{t_i} U_i$. For a small $\delta > 0$ we define

$$t_{i_l} := \inf\{t_i : t_i > t\} \qquad \text{and} \qquad t_{i_u} := \sup\{t_i : t_i \leq t + \delta\}$$

to be the first trading time after t and the last trading time before or at time $t + \delta$, respectively. Thus, the conditional variance of the price increment over $(t, t + \delta]$ is given by

$$\begin{aligned}
\mathbb{E}\left[(X_{t+\delta} - X_t)^2 \;\middle|\; \mathcal{F}_t\right] &= \mathbb{E}\left[\left(\sum_{t<t_i\leq t+\delta} \sigma_{t_i} U_i\right)^2 \;\middle|\; \mathcal{F}_t\right] \\
&\overset{(*)}{=} \mathbb{E}\left[\left(\sum_{t<t_i<t_{i_u}} \sigma_{t_i} U_i\right)^2 \;\middle|\; \mathcal{F}_t\right] + \mathbb{E}\left[2\left(\sum_{t<t_i<t_{i_u}} \sigma_{t_i} U_i\right) \mathbb{E}\left[\sigma_{t_{i_u}} U_{i_u} \;\middle|\; \mathcal{F}_{t_{i_u}^-}\right] \;\middle|\; \mathcal{F}_t\right] \\
&\quad + \mathbb{E}\left[\, \mathbb{E}\left[\sigma_{t_{i_u}}^2 U_{i_u}^2 \;\middle|\; \mathcal{F}_{t_{i_u}^-}\right] \;\middle|\; \mathcal{F}_t\right] \\
&= \mathbb{E}\left[\left(\sum_{t<t_i<t_{i_u}} \sigma_{t_i} U_i\right)^2 \;\middle|\; \mathcal{F}_t\right] + \mathbb{E}\left[\sigma_{t_{i_u}}^2 \;\middle|\; \mathcal{F}_t\right] \\
&= \cdots \\
&= \mathbb{E}\left[\sigma_{t_{i_u}}^2 + \sigma_{t_{i_u-1}}^2 + \ldots + \sigma_{t_{i_l}}^2 \;\middle|\; \mathcal{F}_t\right] \\
&= \mathbb{E}\left[\int_t^{t+\delta} \sigma_s^2 \; dN_s \;\middle|\; \mathcal{F}_t\right] \\
&= \mathbb{E}\left[\int_t^{t+\delta} \sigma_s^2 \cdot \lambda_s \; ds \;\middle|\; \mathcal{F}_t\right]
\end{aligned}$$

which implies the assertion due to the dominated convergence theorem, as the processes σ_t and λ_t are continuous over $[0, T]$. The last equality holds by the martingale property of $\int_0^t \sigma_s dM_s$, since $M_t := N_t - \int_0^t \lambda_s ds$ is a martingale and therefore the stochastic integral with M_t as integrator is also a martingale. In $(*)$ we can take $\sigma_{t_{i_u}}$ out of the conditional

expectation, since we have assumed that σ_t is $\mathcal{F}_t$-predictable, so σ_{t_i} is $\mathcal{F}_{t_i-}$-measurable. □

B. Proofs for Chapter 4

B.1. Proofs for Section 4.1

Proof of Proposition 4.3. This proposition is a special case of Proposition 3.2, which is applied to the rescaled time version of the transaction-time model. Since there is no leverage effect between $W_\cdot$ and $N_{\cdot,T}$, we are able to give arguments conditional on the whole process $N_{\cdot,T}$, i.e.

$$\begin{aligned}
\mathbb{E}\left[\left(X_{t_o+bT,T}-X_{t_o-bT,T}\right)^2\right] &= \mathbb{E}\left[\left(\int_{t_o-bT}^{t_o+bT}\sigma\left(\frac{t}{T}\right)\frac{1}{\sqrt{T}}dW_{N_{t,T}}\right)^2\right] \\
&= \mathbb{E}\left[\sum_{t_i\in\mathcal{S}}\sigma^2\left(\frac{t_i}{T}\right)\frac{1}{T}\cdot\mathbb{E}\left[\left(W_{N_{t_i,T}}-W_{N_{t_{i-1},T}}\right)^2\;\middle|\;N_{\cdot,T}\right]\right] \\
&= \mathbb{E}\left[\sum_{t_i\in\mathcal{S}}\sigma^2\left(\frac{t_i}{T}\right)\frac{1}{T}\right] = \mathbb{E}\left[\int_{t_o-bT}^{t_o+bT}\sigma^2\left(\frac{t}{T}\right)\frac{1}{T}dN_{t,T}\right] \\
&= \int_{u_o-b}^{u_o+b}\sigma^2(u)\lambda(u)du,
\end{aligned}$$

where $\mathcal{S}=\{t_i : t_i \text{ arrival times of } N_{\cdot,T} \text{ in } (t_o-bT, t_o+bT]\}$. The last equality is valid under the assumption of the continuity of the tick-time volatility and intensity functions. □

B.2. Proofs for Section 4.2

In what follows, some steps of the proofs are related to martingale theory. Although one could obtain the same results without using martingale theory, the calculations could be long and cumbersome.[1] Another benefit of using martingale theory in our work is that we can extend all the proofs more easily to a more general case of stochastic intensity models, since martingale dynamics are needed in these situations. We define $M_{t,T}:=N_{t,T}-\int_0^t\lambda(s/T)ds$ (clearly, it is a martingale). Then, it is possible to define a stochastic integral $\int_0^t c_{s,T}\,dM_{s,T}$, where $c_{t,T}$ is a predictable process. We can show that

$$\mathbb{E}\left[\int_0^t c_{s,T}\,dM_{s,T}\right]^2 = \mathbb{E}\left[\int_0^t c_{s,T}\lambda(s/T)\,ds\right]; \tag{A.1.}$$

[1] Without martingale tools, we would often need to calculate conditional probability/ density/ expectation, e.g. for $t,s>0$

$$\mathbb{P}\left(t_j-t_{j-1}>t\;\middle|\;t_{j-1}=s\right) = \mathbb{P}\left(N_{t+s,T}-N_{s,T}=0\;\middle|\;t_{j-1}=s\right) = \exp\left(-\int_s^{s+t}\lambda(l/T)dl\right).$$

see e.g. Kuo [54, Chapter 6]. Throughout the proofs, C is used as a generic constant.

Proof of Theorem 4.7. By the isometry (A.1.), we have

$$\begin{aligned}\mathbb{E}\left[\widehat{\lambda}(u_o) - \frac{1}{\mathfrak{b}T}\int_0^T \mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right)\lambda\left(\frac{t}{T}\right)dt\right]^2 &= \frac{1}{\mathfrak{b}^2T^2}\mathbb{E}\left[\int_0^T \mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right)dM_{t,T}\right]^2 \\ &= \frac{1}{\mathfrak{b}^2T^2}\int_0^T \mathfrak{K}^2\left(\frac{t-u_oT}{\mathfrak{b}T}\right)\lambda\left(\frac{t}{T}\right)dt \\ &= \frac{1}{\mathfrak{b}T}\int_{\mathbb{R}} \mathfrak{K}^2(x)\lambda(u_o+x\mathfrak{b})dx \;\to\; 0,\end{aligned}$$

for $\mathfrak{b} \le u_o \le 1-\mathfrak{b}$ with $0 < \mathfrak{b} \le 1/2$. A usual bias calculation gives

$$\begin{aligned}\frac{1}{\mathfrak{b}T}\int_0^T \mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right)\lambda\left(\frac{t}{T}\right)dt - \lambda(u_o) &= \int_{-u_o/b}^{(1-u_o)/b} \mathfrak{K}(x)\lambda(u_o+x\mathfrak{b})dx - \lambda(u_o) \\ &= \int_{\mathbb{R}} \mathfrak{K}(x)\left\{\lambda(u_o+x\mathfrak{b}) - \lambda(u_o)\right\}dx,\end{aligned}$$

which implies the consistency of the intensity estimate $\widehat{\lambda}(\cdot)$, since $\lambda(\cdot)$ is bounded continuous. To show the asymptotic normality we use the limit theorem for triangular arrays. First, we divide $[0,T]$ into M_T equidistant subintervals of a fixed length $\triangle$, i.e. $M_T = \lfloor T/\triangle \rfloor \to \infty$. We define $\triangle_j := j\triangle$ for $j = 0, 1, ..., M_T$ and rewrite

$$\widehat{\lambda}(u_o) = \sum_{i=1}^{M_T}\int_{\triangle_{i-1}}^{\triangle_i} \frac{1}{\mathfrak{b}T}\mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right)dN_{t,T} \;+\; \int_{\triangle_{M_T}}^{T} \frac{1}{\mathfrak{b}T}\mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right)dN_{t,T},$$

which gives

$$\begin{aligned}&\sqrt{\mathfrak{b}T}\left(\widehat{\lambda}(u_o) - \frac{1}{\mathfrak{b}T}\int_0^T \mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right)\lambda\left(\frac{t}{T}\right)dt\right) \\ &\quad= \sum_{i=1}^{M_T}\int_{\triangle_{i-1}}^{\triangle_i} \frac{1}{\sqrt{\mathfrak{b}T}}\mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right)dM_{t,T} \;+\; \int_{\triangle_{M_T}}^{T} \frac{1}{\sqrt{\mathfrak{b}T}}\mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right)dM_{t,T} \\ &\quad=: \sum_{i=1}^{M_T} Z_{i,T} \;+\; R. \qquad (A.2.)\end{aligned}$$

Owing to independent increments of $N_{t,T}$, the random variables $Z_{i,T}$, $i = 1, ..., M_T$, are independent with $\mathbb{E}[Z_{i,T}] = 0$. We can show that i)

$$\begin{aligned}V_{M_T}^2 &:= \sum_{i=1}^{M_T}\mathbb{E}[Z_{i,T}^2] \\ &= \int_0^T \frac{1}{\mathfrak{b}T}\mathfrak{K}^2\left(\frac{t-u_oT}{\mathfrak{b}T}\right)\lambda\left(\frac{t}{T}\right)dt - \int_{\triangle_{M_T}}^{T} \frac{1}{\mathfrak{b}T}\mathfrak{K}^2\left(\frac{t-u_oT}{\mathfrak{b}T}\right)\lambda\left(\frac{t}{T}\right)dt \\ &\to \lambda(u_o)\int_{\mathbb{R}} \mathfrak{K}^2(x)dx \;=:\; V^2,\end{aligned}$$

since

$$\mathbb{E}[Z_{i,T}^2] = \mathbb{E}\left[\int_{\triangle_{i-1}}^{\triangle_i} \frac{1}{\mathfrak{b}T}\mathfrak{K}^2\left(\frac{t-u_oT}{\mathfrak{b}T}\right)\lambda\left(\frac{t}{T}\right)dt\right],$$

and ii) (Lindeberg's condition) for all $\varepsilon > 0$,

$$\frac{1}{V_{M_T}^2}\sum_{i=1}^{M_T}\mathbb{E}\left[Z_{i,T}^2\ I_{\left\{|Z_{i,T}|\geq\varepsilon\cdot V_{M_T}\right\}}\right]$$
$$= \frac{1}{V_{M_T}^2}\sum_{i=1}^{M_T}\mathbb{E}\left[Z_{i,T}^2\ I_{\left\{\left|\int_{\triangle_{i-1}}^{\triangle_i}\mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right)dM_{t,T}\right|\geq\sqrt{\mathfrak{b}T}\cdot\varepsilon\cdot V_{M_T}\right\}}\right] \to\ 0$$

by the dominated convergence theorem, since $\mathbb{E}\left[Z_{i,T}^2\right] < \infty$. Thus $\sum_{i=1}^{M_T} Z_{i,T} \xrightarrow{\mathcal{D}} \mathcal{N}(0, V^2)$ as $T \to \infty$, which implies that (A.2.) $\xrightarrow{\mathcal{D}} \mathcal{N}(0, V^2)$ as the rest term R is asymptotically negligible. Similarly, we derive that

$$\mathbb{E}\left[\widehat{\lambda}(u_o)\right] - \lambda(u_o) = \frac{1}{2}\lambda^{(2)}(u_o)\int_{\mathbb{R}}\mathfrak{b}^2x^2\mathfrak{K}(x)dx\cdot I_{\{m'=2\}}\ +\ O\left(\int_{\mathbb{R}}\mathfrak{K}(x)|x\mathfrak{b}|^{m'+\gamma'}dx\right),$$

since $\lambda(\cdot)$ lies in $\mathcal{C}^{m',\gamma'}$. The assertion of the theorem is therefore verified by the bandwidth restriction $\mathfrak{b}^{2(m'+\gamma')+1}T = o(1)$. □

B.3. Proofs for Section 4.3

Proof of Proposition 4.9. By virtue of the independence between $N_{\cdot,T}$ and $W_\cdot$, we can prove the assertion by arguing below conditionally on $N_{\cdot,T}$.[2] Thereby, we see as if the arrival times $t_i, i = 1, ..., N_T$, were a deterministic time sampling. We begin by investigating the following limit distribution

$$\frac{\widehat{\sigma}_{clock}^2(u_o) - \mathbb{E}\left[\widehat{\sigma}_{clock}^2(u_o)\ \middle|\ N_{\cdot,T}\right]}{\sqrt{\mathrm{Var}\left[\widehat{\sigma}_{clock}^2(u_o)\ \middle|\ N_{\cdot,T}\right]}}\ \xrightarrow{\mathcal{D}}\ \mathcal{N}(0,1). \tag{A.3.}$$

We set

$$c_{t_i,T} := \frac{\frac{1}{bT}K\left(\frac{t_i-u_oT}{bT}\right)\sigma^2\left(\frac{t_i}{T}\right)}{\sqrt{\sum_{j=1}^{N_T}\frac{1}{b^2T^2}K^2\left(\frac{t_j-u_oT}{bT}\right)\sigma^4\left(\frac{t_j}{T}\right)}}.$$

[2]For example, if a statistic $D^T \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ conditional on $N_{\cdot,T}$, where $\mathcal{N}(0,1)$ is independent of $N_{\cdot,T}$, then

$$\mathbb{P}(D^T \leq d)\ =\ \mathbb{E}[\mathbb{P}(D^T \leq d|N_{\cdot,T})]\ \to\ \mathbb{E}[\lim\mathbb{P}(D^T \leq d|N_{\cdot,T})]\ =\ \mathbb{P}(Z \leq d),$$

where Z is a standard normal random variable.

Conditional on $N_{\cdot,T}$, the left-hand side of (A.3.) has the same law as (see also (4.2))

$$\sum_{i=1}^{N_T} c_{t_i,T} \left(\frac{U_i^2 - 1}{\sqrt{2}} \right)$$

with $\frac{U_i^2-1}{\sqrt{2}}$ i.i.d. random variables with mean zero and unit variance and independent of $N_{\cdot,T}$. According to Barndorff-Nielsen and Shephard [10, Corollary 3.1], the limit distribution (A.3.) holds, since $\sum_{i=1}^{N_T} c_{t_i,T}^2 = 1$ and $\max_{i=1,\ldots,N_T} |c_{t_i,T}| \to 0$.[3] To conclude the result, the denominator is shown to be

$$bT \cdot \mathrm{Var}\left[\widehat{\sigma}^2_{clock}(u_o) \;\middle|\; N_{\cdot,T} \right] = \; 2\sigma^4(u_o)\lambda(u_o) \int_{\mathbb{R}} K^2(x)dx + o_p(1), \qquad (*)$$

thus

$$\sqrt{bT} \left\{ \frac{\widehat{\sigma}^2_{clock}(u_o) - \mathbb{E}\left[\widehat{\sigma}^2_{clock}(u_o) \;\middle|\; N_{\cdot,T} \right]}{\sqrt{2\sigma^4(u_o)\lambda(u_o) \int_{\mathbb{R}} K^2(x)dx}} \right\} = \text{(A.3.)} \cdot \sqrt{bT} \frac{\sqrt{2 \sum_{i=1}^{N_T} \frac{1}{b^2T^2} K^2 \left(\frac{t_i - u_o T}{bT} \right) \sigma^4 \left(\frac{t_i}{T} \right)}}{\sqrt{2\sigma^4(u_o)\lambda(u_o) \int_{\mathbb{R}} K^2(x)dx}}$$
$$\xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

For the derivation of (*), it is sufficient to show that

$$\mathbb{E}\left[\sum_{i=1}^{N_T} \frac{1}{bT} K^2 \left(\frac{t_i - u_o T}{bT} \right) \sigma^4 \left(\frac{t_i}{T} \right) - \sigma^4(u_o)\lambda(u_o) \int_{\mathbb{R}} K^2(x)dx \right]^2 \to \; 0,$$

which is justified by

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{N_T} \frac{1}{bT} K^2 \left(\frac{t_i - u_o T}{bT} \right) \sigma^4 \left(\frac{t_i}{T} \right) - \int_0^T \frac{1}{bT} K^2 \left(\frac{t - u_o T}{bT} \right) \sigma^4 \left(\frac{t}{T} \right) \lambda \left(\frac{t}{T} \right) dt \right]^2 \\
&= \; \mathbb{E}\left[\int_0^T \frac{1}{bT} K^2 \left(\frac{t - u_o T}{bT} \right) \sigma^4 \left(\frac{t}{T} \right) dM_{t,T} \right]^2 \\
&= \; \int_0^T \frac{1}{b^2T^2} K^4 \left(\frac{t - u_o T}{bT} \right) \sigma^8 \left(\frac{t}{T} \right) \lambda \left(\frac{t}{T} \right) dt \\
&= \; \frac{1}{bT} \int_{\mathbb{R}} K^4(x) \sigma^8(u_o + xb) \lambda(u_o + xb) dx \; \to \; 0
\end{aligned}$$

[3] due to the boundedness of $\sigma^2(\cdot)$ and $\lambda(\cdot)$,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{N_T} K^2 \left(\frac{t_i - u_o T}{bT} \right) \sigma^4 \left(\frac{t_i}{T} \right) \right] &= \mathbb{E}\left[\int_0^T K^2 \left(\frac{t - u_o T}{bT} \right) \sigma^4 \left(\frac{t}{T} \right) dN_{t,T} \right] \\
&= bT \int_{\mathbb{R}} K^2(x) \sigma^4(u_o + xb) \lambda(u_o + xb) dx,
\end{aligned}$$

therefore $c_1 \cdot bT \leq \sum_{i=1}^{N_T} K^2 \left(\frac{t_i - u_o T}{bT} \right) \sigma^4 \left(\frac{t_i}{T} \right) \leq c_2 \cdot bT$, for some constants c_1 and c_2.

and

$$\left[\int_0^T \frac{1}{bT}K^2\left(\frac{t-u_oT}{bT}\right)\sigma^4\left(\frac{t}{T}\right)\lambda\left(\frac{t}{T}\right)dt \;-\; \sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}}K^2(x)dx\right]^2$$
$$= \left[\int_{\mathbb{R}} K^2(x)\left\{\sigma^4(u_o+xb)\lambda(u_o+xb) \;-\; \sigma^4(u_o)\lambda(u_o)\right\}\right]^2 \;\to\; 0,$$

since $\sigma^2(\cdot)$ and $\lambda(\cdot)$ are bounded continuous. Finally, in order to prove the asymptotic consistency, we still need to provide the negligibility of the estimation bias. Similarly, we have

$$\begin{aligned}
&\mathbb{E}\left[\widehat{\sigma}^2_{clock}(u_o) \;\middle|\; N_{\cdot,T}\right] - \sigma^2_{clock}(u_o)\\
&= \left\{\mathbb{E}\left[\widehat{\sigma}^2_{clock}(u_o) \;\middle|\; N_{\cdot,T}\right] - \mathbb{E}\left[\sigma^2_{clock}(u_o)\right]\right\} \;+\; \left\{\mathbb{E}\left[\sigma^2_{clock}(u_o)\right] - \sigma^2_{clock}(u_o)\right\}\\
&= \int_0^T \frac{1}{bT}K\left(\frac{t-u_oT}{bT}\right)\sigma^2\left(\frac{t}{T}\right)dM_{t,T} \;+\; \int_{\mathbb{R}} K(x)\left\{\sigma^2_{clock}(u_o+xb) - \sigma^2_{clock}(u_o)\right\}dx\\
&= o_p(1).
\end{aligned}$$

□

Proof of Theorem 4.10. We rewrite the left-hand side of (4.6) as

$$\begin{aligned}
\sqrt{bT}\left\{\widehat{\sigma}^2_{clock}(u_o) - \sigma^2_{clock}(u_o)\right\} &= \sqrt{bT}\left\{\widehat{\sigma}^2_{clock}(u_o) - \mathbb{E}\left[\widehat{\sigma}^2_{clock}(u_o) \;\middle|\; N_{\cdot,T}\right]\right\}\\
&\quad + \sqrt{bT}\left\{\mathbb{E}\left[\widehat{\sigma}^2_{clock}(u_o) \;\middle|\; N_{\cdot,T}\right] - \mathbb{E}\left[\widehat{\sigma}^2_{clock}(u_o)\right]\right\}\\
&\quad + \sqrt{bT}\left\{\mathbb{E}\left[\widehat{\sigma}^2_{clock}(u_o)\right] - \sigma^2_{clock}(u_o)\right\}\\
&=: (A) + (B) + (C).
\end{aligned}$$

The first term (A) has already been done in the previous proposition, having the normal limit distribution $\mathcal{N}\left(0,\; 2\sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}}K^2(x)dx\right)$ which is independent of $N_{\cdot,T}$. Next, similar to the proof for the intensity estimate we obtain

$$\begin{aligned}
(B) &= \sqrt{bT}\left\{\int_0^T \frac{1}{bT}K\left(\frac{t-u_oT}{bT}\right)\sigma^2\left(\frac{t}{T}\right)dN_{t,T} - \int_0^T \frac{1}{bT}K\left(\frac{t-u_oT}{bT}\right)\sigma^2\left(\frac{t}{T}\right)\lambda\left(\frac{t}{T}\right)dt\right\}\\
&\xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\; \sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}}K^2(x)dx\right).
\end{aligned}$$

And lastly, the bias term (C) is equal to

$$\begin{aligned}
&\sqrt{bT}\left\{\int_0^T \frac{1}{bT}K\left(\frac{t-u_oT}{bT}\right)\sigma^2\left(\frac{t}{T}\right)\lambda\left(\frac{t}{T}\right)dt - \sigma^2(u_o)\lambda(u_o)\right\}\\
&= \sqrt{bT}\left\{\int_{\mathbb{R}} K(x)\left\{\sigma^2(u_o+xb)\lambda(u_o+xb) - \sigma^2(u_o)\lambda(u_o)\right\}dx\right\}\\
&= \sqrt{bT}\left\{\frac{1}{2}(\sigma^2(u_o)\lambda(u_o))^{(2)}\int_{\mathbb{R}} K(x)x^2b^2dx\cdot I_{\{m=m'=2\}}\right\} + o(1),
\end{aligned}$$

as $b^{2\alpha+1}T \to 0$, with $\alpha = \min\{m+\gamma, m'+\gamma'\}$. Thus, the claim of the theorem is justified by all these terms[4]. □

Proof of Remark 4.11 ii). This proof relies on the same concept of the proof for $\widehat{\sigma}^2_{clock}(\cdot)$. The conditional expectation and variance of $KQ(u_o)$ are given by

$$\mathbb{E}\left[KQ(u_o) \;\middle|\; N_{\cdot,T}\right] = \sum_{i=1}^{N_T} \frac{1}{bT} K\left(\frac{t_i - u_o T}{bT}\right) \sigma^4\left(\frac{t_i}{T}\right) \quad \text{and}$$

$$\mathrm{Var}\left[KQ(u_o) \;\middle|\; N_{\cdot,T}\right] = \frac{32}{3}\frac{1}{b^2T^2} \sum_{i=1}^{N_T} K^2\left(\frac{t_i - u_o T}{bT}\right) \sigma^8\left(\frac{t_i}{T}\right).$$

Direct calculations yield

$$\begin{aligned}\mathrm{Var}\left[KQ(u_o)\right] &= \mathbb{E}\left[\mathrm{Var}\left[\widehat{\sigma}^4_{clock}(u_o) \;\middle|\; N_{\cdot,T}\right]\right] \;+\; \mathrm{Var}\left[\mathbb{E}\left[\widehat{\sigma}^4_{clock}(u_o) \;\middle|\; N_{\cdot,T}\right]\right] \\ &= \frac{35}{3}\frac{1}{b^2T^2}\int_0^T K^2\left(\frac{t - u_o T}{bT}\right)\sigma^8\left(\frac{t}{T}\right)\lambda\left(\frac{t}{T}\right)dt \;\to\; 0.\end{aligned}$$

The bias term vanishes asymptotically by the continuity of parameter functions, that is

$$\begin{aligned}\mathbb{E}\left[KQ(u_o)\right] - \sigma^4(u_o)\lambda(u_o) &= \int_{\mathbb{R}} K(x)\left\{\sigma^4(u_o + xb)\lambda(u_o + xb) - \sigma^4(u_o)\lambda(u_o)\right\}dx \\ &\to\; 0.\end{aligned}$$

□

Proof of Proposition 4.12. By following the arguments of the proof of Proposition 4.9, we get conditional on $N_{\cdot,T}$

$$\begin{aligned}\frac{\widehat{\sigma}^2(u_o) - \mathbb{E}\left[\widehat{\sigma}^2(u_o) \;\middle|\; N_{\cdot,T}\right]}{\sqrt{\mathrm{Var}\left[\widehat{\sigma}^2(u_o) \;\middle|\; N_{\cdot,T}\right]}} &= \sum_{j=-N}^{N} \frac{k\left(\frac{j}{N}\right)\sigma^2\left(\frac{t_{i_o-j}}{T}\right)}{\sqrt{\sum_{j=-N}^{N} k^2\left(\frac{j}{N}\right)\sigma^4\left(\frac{t_{i_o-j}}{T}\right)}}\left(\frac{U_j^2 - 1}{\sqrt{2}}\right) \\ &=: \sum_{j=-N}^{N} C_{j,N}\left(\frac{U_j^2 - 1}{\sqrt{2}}\right) \;\xrightarrow{\mathcal{D}}\; \mathcal{N}(0,1)\end{aligned}$$

[4]We note that the limit of $(A)+(B)$ can be verified by its characteristic function, i.e. for all $t \in \mathbb{R}$

$$\begin{aligned}\mathbb{E}\left[\exp\left(it\left\{(A)+(B)\right\}\right)\right] &= \mathbb{E}\left[\exp\left(it(B)\right)\cdot\mathbb{E}\left[\exp\left(it(A)\right) \;\middle|\; N_{\cdot,T}\right]\right] \\ &\to \mathbb{E}\left[\lim\exp\left(it(B)\right)\cdot\lim\mathbb{E}\left[\exp\left(it(A)\right) \;\middle|\; N_{\cdot,T}\right]\right] \\ &\qquad\qquad \text{(by dominated convergence theorem)} \\ &\to \exp\left(-\frac{t^2a^2}{2}\right)\cdot\lim\mathbb{E}\left[\exp\left(it(B)\right)\right] \;=\; \exp\left(-\frac{t^2(a^2+b^2)}{2}\right),\end{aligned}$$

since the limit of (A) is independent of $\{N_{\cdot,T}\}$. $\mathcal{N}(0,a^2)$ and $\mathcal{N}(0,b^2)$ are the limits of (A) and (B) respectively.

with $(U_j^2 - 1)/\sqrt{2} \sim^{i.i.d.} \mathcal{N}(0,1)$, since $\sum_j C_{j,N}^2 = 1$ and $\max_j |C_{j,N}| \to 0$. Further we look at the denominator

$$\begin{aligned} N \cdot \mathrm{Var}\left[\widehat{\sigma}^2(u_o) \;\middle|\; N_{\cdot,T}\right] &= N \cdot \frac{2}{H_{1,N}^2} \sum_{j=-N}^{N} k^2\left(\frac{j}{N}\right) \sigma^4\left(\frac{t_{i_o-j}}{T}\right) \\ &= N \cdot \frac{2}{H_{1,N}^2} \sum_{j=-N}^{N} k^2\left(\frac{j}{N}\right) \left\{\sigma^4\left(\frac{t_{i_o-j}}{T}\right) - \sigma^2(u_o)\right\} \;+\; N \cdot \frac{2}{H_{1,N}^2} \sum_{j=-N}^{N} k^2\left(\frac{j}{N}\right) \sigma^2(u_o) \\ &= o_p(1) + \; 2\sigma^2(u_o) \int_{\mathbb{R}} k^2(x)dx, \end{aligned}$$

since

$$\begin{aligned} &\mathbb{E}\left[N \cdot \frac{2}{H_{1,N}^2} \sum_{j=-N}^{N} k^2\left(\frac{j}{N}\right) \left|\sigma^4\left(\frac{t_{i_o-j}}{T}\right) - \sigma^2(u_o)\right|\right] \\ &\quad\leq\; C \cdot \frac{N^2}{H_{1,N}^2} \frac{1}{N} \sum_{j=-N}^{N} k^2\left(\frac{j}{N}\right) \mathbb{E}\left[\left|\frac{t_{i_o-j}}{T} - \frac{t_o}{T}\right|^{\gamma}\right] \\ &\quad\leq\; C \cdot \frac{N^{\gamma}}{T^{\gamma}} \frac{N^2}{H_{1,N}^2} \frac{1}{N} \sum_{j=-N}^{N} k^2\left(\frac{j}{N}\right) \left(\frac{|j|}{N}\right)^{\gamma} \;\to\; 0 \qquad \text{(see Lemma 1 below)} \end{aligned}$$

as $N/T \to 0$. This completes the proof. □

For the bias derivation of the tick-time volatility estimate $\widehat{\sigma}^2(\cdot)$ (and also $\widehat{\sigma}^2_{pavg}(\cdot)$ in the next section of microstructure noise models) we will need the following results—Lemma 1 and Corollary 1. In fact, these results are investigated under a general setting for point processes allowing for stochastic intensity. More precisely, given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_{t,T})_{t\in[0,T]}, \mathbb{P})$, a point process $N_{t,T}$ has an $\mathcal{F}_{t,T}$-intensity $\lambda(t/T)$ if the conditions in Definition 2.1 hold, cf. Brémaud [16, D7].

Lemma 1 Suppose the intensity process $\lambda(u)$ is bounded continuous and bounded away from zero uniformly in $u \in [0,1]$, with probability 1. Then, for $j \in \mathbb{N}$ and $0 < l \leq 4$, it implies that

i) $\mathbb{E}\left[\int_{t_i}^{t_{i+j}} \lambda(s/T)ds \;\middle|\; \mathcal{F}_{t_i,T}\right] = j$,

ii) $\mathbb{E}\left[\left(\int_{t_i}^{t_{i+j}} \lambda(s/T)ds\right)^2 \;\middle|\; \mathcal{F}_{t_i,T}\right] = j^2 + j$, and

iii) $\mathbb{E}\left[(t_{i+j} - t_i)^l \;\middle|\; \mathcal{F}_{t_i,T}\right] = O(j^l)$.

Proof. Since $M_{t,T} = N_{t,T} - \int_0^t \lambda(s/T)ds$ is a martingale and the arrival times t_i are stopping times, we get

$$j \;=\; \mathbb{E}\left[N_{t_{i+j},T} - N_{t_i,T} \;\middle|\; \mathcal{F}_{t_i,T}\right] \;=\; \mathbb{E}\left[\int_{t_i}^{t_{i+j}} \lambda(s/T)ds \;\middle|\; \mathcal{F}_{t_i,T}\right]$$

by the optional sampling theorem. Moreover, it is clear that $\tilde{M}_{t,T} := M_{t,T}^2 - \int_0^t \lambda(s/T)ds$ is another martingale, thus

$$\begin{aligned}
0 &= \mathbb{E}\left[\tilde{M}_{t_{i+j},T} - \tilde{M}_{t_i,T} \;\middle|\; \mathcal{F}_{t_i,T}\right] \\
&= \mathbb{E}\left[M_{t_{i+j},T}^2 - M_{t_i,T}^2 \;\middle|\; \mathcal{F}_{t_i,T}\right] - \mathbb{E}\left[\int_{t_i}^{t_{i+j}} \lambda(l/T)dl \;\middle|\; \mathcal{F}_{t_i,T}\right] \\
&= \mathbb{E}\left[\left(M_{t_{i+j},T} - M_{t_i,T}\right)^2 \;\middle|\; \mathcal{F}_{t_i,T}\right] - j \\
&= \mathbb{E}\left[\left(j - \int_{t_i}^{t_{i+j}} \lambda(l/T)dl\right)^2 \;\middle|\; \mathcal{F}_{t_i,T}\right] - j,
\end{aligned}$$

i.e. $\mathbb{E}\left[\left(\int_{t_i}^{t_{i+j}} \lambda(s/T)ds\right)^2 \;\middle|\; \mathcal{F}_{t_i,T}\right] = j^2 + j$. Therefore we have shown i) and ii). In fact, iii) has also been displayed for $l = 1, 2$, since λ is bounded and bounded away from zero. Burkholder-Davis-Gundy's inequality (see Jacod and Protter [48, p.39]) yields

$$\begin{aligned}
\mathbb{E}\left[\left|M_{t_{i+j},T} - M_{t_i,T}\right|^4 \;\middle|\; \mathcal{F}_{t_i,T}\right] &\le C \cdot \mathbb{E}\left[\left(\int_{t_i}^{t_{i+j}} \lambda(s/T)ds\right)^2 \;\middle|\; \mathcal{F}_{t_i,T}\right] \\
&\le \mathbb{E}\left[(t_{i+j} - t_i)^2 \;\middle|\; \mathcal{F}_{t_j,T}\right] \;=\; O(j^2).
\end{aligned}$$

By applying Hölder's inequality, $\mathbb{E}\left[\left|M_{t_{i+j},T} - M_{t_i,T}\right|^3 \;\middle|\; \mathcal{F}_{t_i,T}\right] = O(j^{3/2})$. Hence

$$\begin{aligned}
&\mathbb{E}\left[(t_{i+j} - t_i)^4 \;\middle|\; \mathcal{F}_{t_i,T}\right] \;\le\; C \cdot \mathbb{E}\left[\left(\int_{t_i}^{t_{i+j}} \lambda(s/T)ds\right)^4 \;\middle|\; \mathcal{F}_{t_i,T}\right] \\
&= C \cdot \mathbb{E}\left[\left(M_{t_i,T} - M_{t_{i+j},T} \;+\; j\right)^4 \;\middle|\; \mathcal{F}_{t_i,T}\right] \\
&= C \cdot \Big(\mathbb{E}\left[\left(M_{t_i,T} - M_{t_{i+j},T}\right)^4 \;\middle|\; \mathcal{F}_{t_i,T}\right] + 3j\mathbb{E}\left[\left(M_{t_i,T} - M_{t_{i+j},T}\right)^3 \;\middle|\; \mathcal{F}_{t_i,T}\right] \\
&\quad + 6j^2\mathbb{E}\left[\left(M_{t_i,T} - M_{t_{i+j},T}\right)^2 \;\middle|\; \mathcal{F}_{t_i,T}\right] + 3j^3\mathbb{E}\left[M_{t_i,T} - M_{t_{i+j},T} \;\middle|\; \mathcal{F}_{t_i,T}\right] + j^4\Big) \\
&= O(j^4),
\end{aligned}$$

since $\mathbb{E}\left[M_{t_i,T} - M_{t_{i+j},T} \;\middle|\; \mathcal{F}_{t_i,T}\right] = 0$. The rest can be easily justified by employing Hölder's inequality. □

Corollary 1 Let paths of $\lambda(\cdot)$ lie in $\mathcal{C}^{m',\gamma'}[0,1]$ and satisfy Assumption 4.5 i). For $j \in \mathbb{N}$, we get

i) $\mathbb{E}\left[t_{i+j} - t_i \;\middle|\; \mathcal{F}_{t_i,T}\right] = \frac{j}{\lambda(t_i/T)} - \frac{1}{2}\frac{\lambda^{(1)}(t_i/T)}{\lambda^2(t_i/T)}\frac{j^2}{T}I_{\{m'\neq 0\}} + O\left(\frac{j^{1+m'+\gamma'}}{T^{m'+\gamma'}}I_{\{m'\neq 2\}}\right) + O\left(\frac{j^3}{T^2}I_{\{m'=2\}}\right)$,

ii) $\mathbb{E}\left[(t_{i+j} - t_i)^2 \;\middle|\; \mathcal{F}_{t_i,T}\right] = \frac{j^2+j}{\lambda^2(t_i/T)} \;+\; O\left(\frac{j^{2+\gamma'}}{T^{\gamma'}}I_{\{m'=0\}} + \frac{j^3}{T}I_{\{m'\neq 0\}}\right)$.

Proof. By Lemma 1 we get

$$\mathbb{E}\left[\lambda(t_i/T)(t_{i+j}-t_i)\;\middle|\;\mathcal{F}_{t_i,T}\right] = \underbrace{\mathbb{E}\left[\int_{t_i}^{t_{i+j}}\{\lambda(t_i/T)-\lambda(l/T)\}\,dl\;\middle|\;\mathcal{F}_{t_i,T}\right]}_{=:(\Lambda)} + j.$$

Therefore, the first statement i) can be verified by considering the following cases: (for $\omega \in \Omega$)

a) for $\lambda(\cdot)(\omega) \in \mathcal{C}^{0,\gamma'}[0,1]$,

$$\begin{aligned}(\Lambda) &\le C\cdot\left[\int_{t_i}^{t_{i+j}}\left(\frac{l-t_i}{T}\right)^{\gamma'} dl\;\middle|\;\mathcal{F}_{t_i,T}\right]\\ &\le C\cdot\mathbb{E}\left[\frac{(t_{i+j}-t_i)^{1+\gamma'}}{T^{\gamma'}}\;\middle|\;\mathcal{F}_{t_i,T}\right] \le\; C\cdot\frac{j^{1+\gamma'}}{T^{\gamma'}}; \qquad \text{(see Lemma 1)}\end{aligned}$$

b) for $\lambda(\cdot)(\omega) \in \mathcal{C}^{1,\gamma'}[0,1]$,

$$\begin{aligned}(\Lambda) =&\; -\lambda^{(1)}\left(\frac{t_i}{T}\right)\mathbb{E}\left[\int_{t_i}^{t_{i+j}}\left(\frac{l-t_i}{T}\right)dl\;\middle|\;\mathcal{F}_{t_i,T}\right] + O\left(\mathbb{E}\left[\int_{t_i}^{t_{i+j}}\left(\frac{l-t_i}{T}\right)^{1+\gamma'}dl\;\middle|\;\mathcal{F}_{t_i,T}\right]\right)\\ =&\; -\frac{1}{2}\lambda^{(1)}\left(\frac{t_i}{T}\right)\mathbb{E}\left[\frac{(t_{i+j}-t_i)^2}{T}\;\middle|\;\mathcal{F}_{t_i,T}\right] + O\left(\mathbb{E}\left[\frac{(t_{i+j}-t_i)^{2+\gamma'}}{T^{1+\gamma'}}\;\middle|\;\mathcal{F}_{t_i,T}\right]\right)\\ \overset{(*)}{=}&\; -\frac{1}{2}\frac{\lambda^{(1)}(t_i/T)}{\lambda^2(t_i/T)}\frac{j^2}{T} + O\left(\frac{j^{2+\gamma'}}{T^{1+\gamma'}}\right);\end{aligned}$$

c) for $\lambda(\cdot)(\omega) \in \mathcal{C}^{2,\gamma'}[0,1]$,

$$\begin{aligned}(\Lambda) =&\; -\lambda^{(1)}\left(\frac{t_i}{T}\right)\mathbb{E}\left[\int_{t_i}^{t_{i+j}}\left(\frac{l-t_i}{T}\right)dl\;\middle|\;\mathcal{F}_{t_i,T}\right] + O\left(\mathbb{E}\left[\int_{t_i}^{t_{i+j}}\left(\frac{l-t_i}{T}\right)^{2}dl\;\middle|\;\mathcal{F}_{t_i,T}\right]\right)\\ =&\; -\frac{1}{2}\lambda^{(1)}\left(\frac{t_i}{T}\right)\mathbb{E}\left[\frac{(t_{i+j}-t_i)^2}{T}\;\middle|\;\mathcal{F}_{t_i,T}\right] + O\left(\mathbb{E}\left[\frac{(t_{i+j}-t_i)^{3}}{T^2}\;\middle|\;\mathcal{F}_{t_i,T}\right]\right)\\ \overset{(*)}{=}&\; -\frac{1}{2}\frac{\lambda^{(1)}(t_i/T)}{\lambda^2(t_i/T)}\frac{j^2}{T} + O\left(\frac{j^3}{T^2}\right).\end{aligned}$$

To show $(*)$ in b) and c), we have to apply the result ii) beforehand. Without doing this we can approximate them to only the order $O(j^2/T)$; this order is, however, enough

to show the assertion ii). More precisely, similar to a) - c) without $(*)$ we get

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_{t_i}^{t_{i+j}}\lambda(t_i/T)dl\right)^2-\left(\int_{t_i}^{t_{i+j}}\lambda(l/T)dl\right)^2\;\Big|\;\mathcal{F}_{t_i,T}\right]\\
&=\;\mathbb{E}\left[\underbrace{\int_{t_i}^{t_{i+j}}\{\lambda(t_i/T)-\lambda(l/T)\}\,dl}_{\text{see i) without }(*)}\cdot\underbrace{\int_{t_i}^{t_{i+j}}\{\lambda(t_i/T)+\lambda(l/T)\}\,dl}_{=O(t_{i+j}-t_i)}\;\Big|\;\mathcal{F}_{t_i,T}\right]\\
&=\;O\left(\frac{j^{2+\gamma'}}{T^{\gamma'}}I_{\{m'=0\}}+\frac{j^3}{T}I_{\{m'\neq0\}}\right)
\end{aligned}$$

by Lemma 1. In particular,

$$\begin{aligned}
\mathbb{E}\left[\lambda^2(t_i/T)(t_{i+j}-t_i)^2\;\Big|\;\mathcal{F}_{t_i,T}\right]&=\mathbb{E}\left[\left(\int_{t_i}^{t_{i+j}}\lambda(l/T)dl\right)^2\;\Big|\;\mathcal{F}_{t_i,T}\right]\\
&\quad+\mathbb{E}\left[\left(\int_{t_i}^{t_{i+j}}\lambda(t_i/T)dl\right)^2-\left(\int_{t_i}^{t_{i+j}}\lambda(l/T)dl\right)^2\;\Big|\;\mathcal{F}_{t_i,T}\right]\\
&=j^2+j+O\left(\frac{j^{2+\gamma'}}{T^{\gamma'}}I_{\{m'=0\}}+\frac{j^3}{T}I_{\{m'\neq0\}}\right).
\end{aligned}$$

□

The next lemma is an application of Corollary 1 applied to the case of deterministic intensity. This result will be often used in the coming proofs, and therefore we explicitly state it here; the proof is straightforward.

Lemma 2 Under the same conditions as stated in Corollary 1, if the intensity $\lambda(\cdot)$ is deterministic, we have

i) $\mathbb{E}\left[t_{i_o+j}-t_{i_o}\right]=\frac{j}{\lambda(u_o)}-\frac{1}{2}\frac{\lambda^{(1)}(u_o)}{\lambda^2(u_o)}\frac{j^2}{T}I_{\{m'\neq0\}}+O\left(\frac{|j|^{1+m'+\gamma'}}{T^{m'+\gamma'}}I_{\{m'\neq2\}}\right)+O\left(\frac{|j|^3}{T^2}I_{\{m'=2\}}\right)$,

ii) $\mathbb{E}\left[(t_{i_o+j}-t_{i_o})^2\right]=\frac{j^2+j}{\lambda^2(u_o)}+O\left(\frac{|j|^{2+\gamma'}}{T^{\gamma'}}I_{\{m'=0\}}+\frac{|j|^3}{T}I_{\{m'\neq0\}}\right)$,

for $j\in\mathbb{Z}$, where $i_o:=\inf\{i:t_i\geq u_oT\}$ for a given (fixed) time point $u_o\in(0,1)$. □

Proof of Theorem 4.13. For notational simplicity we will examine only the case of twice-differentiable functions, i.e. $m=m'=2$ (other cases can be examined in much the same way). For this bias derivation, we will have to use Lemmas 1 and 2 again and again.

For $\sigma(\cdot) \in \mathcal{C}^{2,\gamma}$ and $\lambda(\cdot) \in \mathcal{C}^{2,\gamma'}$, we can expand

$$\begin{aligned}
&\sqrt{N}\left(\frac{1}{H_{1,N}}\sum_{j=-N}^{N} k\left(\frac{j}{N}\right)\left\{\sigma^2\left(\frac{t_{i_o-j}}{T}\right)-\sigma^2\left(\frac{t_o}{T}\right)\right\}\right)\\
&\quad=\frac{\sqrt{N}}{H_{1,N}}\sum_{j=-N}^{N} k\left(\frac{j}{N}\right)\times\\
&\qquad\times\left\{(\sigma^2(u_o))^{(1)}\left(\frac{t_{i_o-j}}{T}-\frac{t_o}{T}\right)+\frac{(\sigma^2(u_o))^{(2)}}{2}\left(\frac{t_{i_o-j}}{T}-\frac{t_o}{T}\right)^2+O\left(\left|\frac{t_{i_o-j}}{T}-\frac{t_o}{T}\right|^{2+\gamma}\right)\right\}\\
&\quad=:\ (A)+(B)+(C).
\end{aligned}$$

First, we prove that $(A)=o_p(1)$ by separating it into $(A1)$ and $(A2)$, where

$$\begin{aligned}
(A1)&=\frac{\sqrt{N}}{H_{1,N}}(\sigma^2(u_o))^{(1)}\sum_{j=-N}^{N} k\left(\frac{j}{N}\right)\left(\frac{t_{i_o-j}}{T}-\frac{t_{i_o}}{T}\right)\quad\text{and}\\
(A2)&=\frac{\sqrt{N}}{H_{1,N}}(\sigma^2(u_o))^{(1)}\sum_{j=-N}^{N} k\left(\frac{j}{N}\right)\left(\frac{t_{i_o}}{T}-\frac{t_o}{T}\right).
\end{aligned}$$

Obviously, $(A2)$ is of order $o_p(1)$. We will show below that $\mathbb{E}\left[(A1)^2\right]=o(1)$, which results that $(A1)=o_p(1)$. By the symmetry of the kernel function k we can rewrite

$$(A1)=\frac{\sqrt{N}}{H_{1,N}}(\sigma^2(u_o))^{(1)}\sum_{j=1}^{N} k\left(\frac{j}{N}\right)\left\{\left(\frac{t_{i_o-j}}{T}-\frac{t_{i_o}}{T}\right)+\left(\frac{t_{i_o+j}}{T}-\frac{t_{i_o}}{T}\right)\right\},$$

thus (by denoting $\sigma'_\cdot:=(\sigma^2(u_o))^{(1)}$, $\sigma''_\cdot:=(\sigma^2(u_o))^{(2)}$, $\lambda'_\cdot:=(\lambda(u_o))^{(1)}$ and $\lambda_\cdot:=\lambda(u_o)$)

$$\begin{aligned}
\mathbb{E}\left[(A1)^2\right]&=\frac{N}{H_{1,N}^2}\sigma'^2_\cdot\sum_{i,j=1}^{N} k\left(\frac{i}{N}\right)k\left(\frac{j}{N}\right)\times\\
&\quad\times\left(\mathbb{E}\left[\left(\frac{t_{i_o-i}}{T}-\frac{t_{i_o}}{T}\right)\left(\frac{t_{i_o-j}}{T}-\frac{t_{i_o}}{T}\right)\right]+\mathbb{E}\left[\left(\frac{t_{i_o-i}}{T}-\frac{t_{i_o}}{T}\right)\left(\frac{t_{i_o+j}}{T}-\frac{t_{i_o}}{T}\right)\right]\right.\\
&\quad\left.+\mathbb{E}\left[\left(\frac{t_{i_o+i}}{T}-\frac{t_{i_o}}{T}\right)\left(\frac{t_{i_o-j}}{T}-\frac{t_{i_o}}{T}\right)\right]+\mathbb{E}\left[\left(\frac{t_{i_o+i}}{T}-\frac{t_{i_o}}{T}\right)\left(\frac{t_{i_o+j}}{T}-\frac{t_{i_o}}{T}\right)\right]\right)\\
&=:(I_1)+(I_2)+(I_3)+(I_4).
\end{aligned}$$

Since $N_{t,T}$ has independent increments, the non-overlapping interarrival times are independent, particularly $\mathbb{E}\left[(t_i-t_j)(t_k-t_l)\right]=\mathbb{E}\left[t_i-t_j\right]\mathbb{E}\left[t_k-t_l\right]$ for $l<k\le j<i$. Therefore,

by Lemma 2

$$
\begin{aligned}
(I_2) &= \frac{N}{H_{1,N}^2}\sigma_{\cdot}'^2 \sum_{i,j=1}^{N} k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \mathbb{E}\left[\frac{t_{i_o-i}}{T} - \frac{t_{i_o}}{T}\right] \mathbb{E}\left[\frac{t_{i_o+j}}{T} - \frac{t_{i_o}}{T}\right] \\
&= \frac{N}{T^2}\sigma_{\cdot}'^2 \frac{1}{H_{1,N}^2} \sum_{i,j=1}^{N} k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \times \\
&\quad \times \left\{\frac{-i}{\lambda_{\cdot}} - \frac{1}{2}\frac{\lambda_{\cdot}'}{\lambda_{\cdot}^2}\frac{i^2}{T} + O\left(\frac{i^3}{T^2}\right)\right\} \left\{\frac{j}{\lambda_{\cdot}} - \frac{1}{2}\frac{\lambda_{\cdot}'}{\lambda_{\cdot}^2}\frac{j^2}{T} + O\left(\frac{j^3}{T^2}\right)\right\} \\
&=: (I_{2,1}) + ... + (I_{2,9})
\end{aligned}
$$

with

$$
\begin{aligned}
(I_{2,1}) &= -\frac{N^3}{T^2} \cdot \frac{\sigma_{\cdot}'^2}{\lambda_{\cdot}^2} \frac{N^2}{H_{1,N}^2} \left(\frac{1}{N}\sum_{i=1}^{N} k\left(\frac{i}{N}\right)\frac{i}{N}\right)^2, \\
(I_{2,2}) &= \frac{N^4}{T^3} \cdot \frac{\sigma_{\cdot}'^2}{2} \frac{\lambda_{\cdot}'}{\lambda_{\cdot}^3} \frac{N^2}{H_{1,N}^2} \left(\frac{1}{N}\sum_{i=1}^{N} k\left(\frac{i}{N}\right)\frac{i}{N}\right)\left(\frac{1}{N}\sum_{j=1}^{N} k\left(\frac{j}{N}\right)\frac{j^2}{N^2}\right), \\
(I_{2,3}) &= O\left(\frac{N^5}{T^4} \cdot \frac{N^2}{H_{1,N}^2} \left(\frac{1}{N}\sum_{i=1}^{N} k\left(\frac{i}{N}\right)\frac{i}{N}\right)\left(\frac{1}{N}\sum_{j=1}^{N} k\left(\frac{j}{N}\right)\frac{j^3}{N^3}\right)\right), \\
(I_{2,4}) &= -\frac{N^4}{T^3} \cdot \frac{\sigma_{\cdot}'^2}{2} \frac{\lambda_{\cdot}'}{\lambda_{\cdot}^3} \frac{N^2}{H_{1,N}^2} \left(\frac{1}{N}\sum_{j=1}^{N} k\left(\frac{j}{N}\right)\frac{j}{N}\right)\left(\frac{1}{N}\sum_{i=1}^{N} k\left(\frac{i}{N}\right)\frac{i^2}{N^2}\right), \\
&\qquad\qquad ((I_{2,2}) \text{ cancels out } (I_{2,4})) \\
(I_{2,5}) &= \frac{N^5}{T^4} \cdot \frac{\sigma_{\cdot}'^2}{4} \frac{\lambda_{\cdot}'^2}{\lambda_{\cdot}^4} \frac{N^2}{H_{1,N}^2} \left(\frac{1}{N}\sum_{i=1}^{N} k\left(\frac{i}{N}\right)\frac{i^2}{N^2}\right)^2, \\
(I_{2,6}) &= O\left(\frac{N^6}{T^5} \cdot \frac{N^2}{H_{1,N}^2} \left(\frac{1}{N}\sum_{i=1}^{N} k\left(\frac{i}{N}\right)\frac{i^2}{N^2}\right)\left(\frac{1}{N}\sum_{j=1}^{N} k\left(\frac{j}{N}\right)\frac{j^3}{N^3}\right)\right), \\
(I_{2,7}) &= O\left(\frac{N^5}{T^4} \cdot \frac{N^2}{H_{1,N}^2} \left(\frac{1}{N}\sum_{i=1}^{N} k\left(\frac{i}{N}\right)\frac{i^3}{N^3}\right)\left(\frac{1}{N}\sum_{j=1}^{N} k\left(\frac{j}{N}\right)\frac{j}{N}\right)\right), \\
(I_{2,8}) &= O\left(\frac{N^6}{T^5} \cdot \frac{N^2}{H_{1,N}^2} \left(\frac{1}{N}\sum_{i=1}^{N} k\left(\frac{i}{N}\right)\frac{i^3}{N^3}\right)\left(\frac{1}{N}\sum_{j=1}^{N} k\left(\frac{j}{N}\right)\frac{j^2}{N^2}\right)\right) \quad \text{and} \\
(I_{2,9}) &= O\left(\frac{N^7}{T^6} \cdot \frac{N^2}{H_{1,N}^2} \left(\frac{1}{N}\sum_{i=1}^{N} k\left(\frac{i}{N}\right)\frac{i^3}{N^3}\right)^2\right).
\end{aligned}
$$

In particular, we obtain

$$
(I_2) = (I_{2,1}) + o(1) \tag{A.4.}
$$

if $N^5/T^4 \to 0$ (which is indeed satisfied by our segment length condition $N^4/T^3 \to 0$). We will see later that $(I_{2,1})$ will be eliminated so that $(I_2) = o(1)$. Let us continue to (I_4).

$$\begin{aligned}(I_4) &= \frac{N}{T^2}\cdot\sigma_{\cdot}'^2\frac{1}{H_{1,N}^2}\sum_{i=1}^{N}k^2\left(\frac{i}{N}\right)\mathbb{E}\left[(t_{i_o+i}-t_{i_o})^2\right]\\ &\quad+\frac{N}{T^2}\cdot\sigma_{\cdot}'^2\frac{1}{H_{1,N}^2}\sum_{i\neq j,j<i}k\left(\frac{i}{N}\right)k\left(\frac{j}{N}\right)\mathbb{E}\left[(t_{i_o+i}-t_{i_o})(t_{i_o+j}-t_{i_o})\right]\\ &\quad+\frac{N}{T^2}\cdot\sigma_{\cdot}'^2\frac{1}{H_{1,N}^2}\sum_{i\neq j,j>i}k\left(\frac{i}{N}\right)k\left(\frac{j}{N}\right)\mathbb{E}\left[(t_{i_o+i}-t_{i_o})(t_{i_o+j}-t_{i_o})\right]\\ &=: (I_{4,1})+(I_{4,2})+(I_{4,3}),\end{aligned}$$

where

$$(I_{4,1}) = \frac{N}{T^2}\cdot\sigma_{\cdot}'^2\frac{1}{H_{1,N}^2}\sum_{i=1}^{N}k^2\left(\frac{i}{N}\right)\left\{\frac{i^2+i}{\lambda_{\cdot}^2}+O\left(\frac{i^3}{T}\right)\right\}\ \to\ 0, \qquad \text{(by Lemma 2)}$$

and

$$\begin{aligned}(I_{4,2}) &= \frac{N}{T^2}\cdot\sigma_{\cdot}'^2\frac{1}{H_{1,N}^2}\sum_{i\neq j,j<i}k\left(\frac{i}{N}\right)k\left(\frac{j}{N}\right)\mathbb{E}\left[(t_{i_o+i}-t_{i_o+j})(t_{i_o+j}-t_{i_o})+(t_{i_o+j}-t_{i_o})^2\right]\\ &=: (I_{4,2,1})+(I_{4,2,2}).\end{aligned}$$

Again by the independence of interarrival times and Lemma 2 we get

$$\begin{aligned}(I_{4,2,1}) &= \frac{N}{T^2}\cdot\sigma_{\cdot}'^2\frac{1}{H_{1,N}^2}\sum_{i\neq j,j<i}k\left(\frac{i}{N}\right)k\left(\frac{j}{N}\right)\mathbb{E}\left[t_{i_o+i}-t_{i_o+j}\right]\mathbb{E}\left[t_{i_o+j}-t_{i_o}\right]\\ &= \frac{N}{T^2}\cdot\sigma_{\cdot}'^2\frac{1}{H_{1,N}^2}\sum_{i\neq j,j<i}k\left(\frac{i}{N}\right)k\left(\frac{j}{N}\right)\left\{\mathbb{E}\left[t_{i_o+i}-t_{i_o}\right]\mathbb{E}\left[t_{i_o+j}-t_{i_o}\right]-\left(\mathbb{E}\left[t_{i_o+j}-t_{i_o}\right]\right)^2\right\}\\ &= \frac{N^3}{T^2}\cdot\frac{\sigma_{\cdot}'^2}{\lambda_{\cdot}^2}\frac{N^2}{H_{1,N}^2}\frac{1}{N^2}\sum_{i\neq j,j<i}k\left(\frac{i}{N}\right)k\left(\frac{j}{N}\right)\frac{i}{N}\frac{j}{N}\\ &\quad-\frac{N^3}{T^2}\cdot\frac{\sigma_{\cdot}'^2}{\lambda_{\cdot}^2}\frac{N^2}{H_{1,N}^2}\frac{1}{N^2}\sum_{i\neq j,j<i}k\left(\frac{i}{N}\right)k\left(\frac{j}{N}\right)\frac{j^2}{N^2}+o(1)\\ &=: (I_{4,2,1,1})+(I_{4,2,1,2})+o(1)\end{aligned}$$

as $N^4/T^3 \to 0$ (in the same way as in the derivation of (I_2)). The next term

$$\begin{aligned}(I_{4,2,2}) &= \frac{N}{T^2}\cdot\sigma_{\cdot}'^2\frac{1}{H_{1,N}^2}\sum_{i\neq j,j<i}k\left(\frac{i}{N}\right)k\left(\frac{j}{N}\right)\mathbb{E}\left[(t_{i_o+j}-t_{i_o})^2\right]\\ &= \frac{N}{T^2}\cdot\sigma_{\cdot}'^2\frac{1}{H_{1,N}^2}\sum_{i\neq j,j<i}k\left(\frac{i}{N}\right)k\left(\frac{j}{N}\right)\left\{\frac{j^2+j}{\lambda_{\cdot}^2}+O\left(\frac{j^3}{T}\right)\right\}\\ &= \frac{N^3}{T^2}\cdot\frac{\sigma_{\cdot}'^2}{\lambda_{\cdot}^2}\frac{N^2}{H_{1,N}^2}\frac{1}{N^2}\sum_{i\neq j,j<i}k\left(\frac{i}{N}\right)k\left(\frac{j}{N}\right)\frac{j^2}{N^2}+o(1)\end{aligned}$$

as $N^4/T^3 \to 0$. Evidently, the first term of $(I_{4,2,2})$ wipes out $(I_{4,2,1,2})$. Furthermore, by changing the role of i and j, the same results can be derived for $(I_{4,3})$, and therefore

$$(I_4) = (I_{4,2,1,1}) + (I_{4,3,1,1}) + o(1).$$

Fortunately, we see that the sum of $(I_{4,2,1,1})$ and $(I_{4,3,1,1})$ is exactly the negative of $(I_{2,1})$ in (A.4.), thus

$$(I_2) + (I_4) \;=\; o(1) \qquad \text{as } N^4/T^3 \to 0.$$

Analogously, we can show that $(I_1) + (I_3) = o(1)$, so we can conclude that $(A1)$ is of order $o_p(1)$. Finally, the negligibility of (B) and (C) is done by

$$\begin{aligned}\mathbb{E}\,|(B)| &= \frac{\sqrt{N}}{T^2} \cdot \frac{\sigma_{\cdot}''}{2} \frac{N}{H_{1,N}} \frac{1}{N} \sum_{j=-N}^{N} k\left(\frac{j}{N}\right) \mathbb{E}\left[(t_{i_o-j} - t_o)^2\right] \\ &= \frac{\sqrt{N}}{T^2} \cdot \frac{\sigma_{\cdot}''}{2} \frac{N}{H_{1,N}} \frac{1}{N} \sum_{j=-N}^{N} k\left(\frac{j}{N}\right) \left\{ \frac{j^2}{\lambda_{\cdot}^2} - \frac{j}{\lambda_{\cdot}^2} + O\left(\frac{|j|^3}{T}\right) \right\} \;=\; o(1),\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\,|(C)| &= O\left(\frac{\sqrt{N}}{H_{1,N}} \sum_{j=-N}^{N} k\left(\frac{j}{N}\right) \mathbb{E}\left[\left| \frac{t_{i_o-j}}{T} - \frac{t_o}{T} \right|^{2+\gamma} \right] \right) \\ &= O\left(\frac{N}{H_{1,N}} \frac{\sqrt{N} N^{2+\gamma}}{T^{2+\gamma}} \frac{1}{N} \sum_{j=-N}^{N} k\left(\frac{j}{N}\right) \frac{j^{2+\gamma}}{N^{2+\gamma}} \right) \;=\; o(1).\end{aligned}$$

The assertion of the limit distribution is the consequence of this bias derivation and the limit distribution in Proposition 4.12. □

Proof of Theorem 4.14. The basic idea of how to derive this limit distribution has already been demonstrated in an example before the theorem. Here, we present only the case of $m = 0$ explicitly. Other cases ($m = 1, 2$) can be determined by the same considerations with the corresponding restriction on the window size $\mathfrak{b}$ and the segment length N. Interestingly, the bias term in the case of $m' = 2$ does not vanish.

Case 1: $m = 0$, $m' = 0$. From Theorems 4.7 and 4.13 it is necessary to restrict $\mathfrak{b}$ and N such that

$$N = o\left(T^{\frac{2\gamma}{1+2\gamma}}\right) \quad \text{and} \quad \mathfrak{b}T = o\left(T^{\frac{2\gamma'}{1+2\gamma'}}\right)$$

in order to get those limit distributions. We divide it again into a), b) and c):

a) If $\gamma = \gamma'$ then $N = O(\mathfrak{b}T)$. This gives

$$\begin{aligned}
&\sqrt{N}\left\{\widehat{\sigma}^2(u_o)\widehat{\lambda}(u_o) - \sigma^2(u_o)\lambda(u_o)\right\} \\
&\quad = \underbrace{\widehat{\lambda}(u_o)}_{\xrightarrow{\mathbb{P}}\lambda(u_o)} \sqrt{N}\left\{\widehat{\sigma}^2(u_o) - \sigma^2(u_o)\right\} \;+\; \sigma^2(u_o)\underbrace{\frac{\sqrt{N}}{\sqrt{\mathfrak{b}T}}}_{=\sqrt{c_o}} \cdot \sqrt{\mathfrak{b}T}\left\{\widehat{\lambda}(u_o) - \lambda(u_o)\right\} \\
&\quad \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\; 2\sigma^4(u_o)\lambda^2(u_o)\int_{\mathbb{R}} k^2(x)dx \;+\; c_o\sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}} \mathfrak{K}^2(x)dx\right),
\end{aligned}$$

where the limit distribution of these two summands is attained by the same arguments as in Theorem 4.10.

b) If $\gamma < \gamma'$ then $N = o(\mathfrak{b}T)$. This leads to

$$\begin{aligned}
\sqrt{N}\left\{\widehat{\sigma}^2(u_o)\widehat{\lambda}(u_o) - \sigma^2(u_o)\lambda(u_o)\right\} &= \widehat{\lambda}(u_o) \cdot \sqrt{N}\left\{\widehat{\sigma}^2(u_o) - \sigma^2(u_o)\right\} \\
&\xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\; 2\sigma^4(u_o)\lambda^2(u_o)\int_{\mathbb{R}} k^2(x)dx\right).
\end{aligned}$$

c) If $\gamma > \gamma'$ then $\mathfrak{b}T = o(N)$, which implies that

$$\begin{aligned}
\sqrt{\mathfrak{b}T}\left\{\widehat{\sigma}^2(u_o)\widehat{\lambda}(u_o) - \sigma^2(u_o)\lambda(u_o)\right\} &= \sigma^2(u_o) \cdot \sqrt{\mathfrak{b}T}\left\{\widehat{\lambda}(u_o) - \lambda(u_o)\right\} \\
&\xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\; \sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}} \mathfrak{K}^2(x)dx\right).
\end{aligned}$$

Case 2: $m = 0$, $m' = 1$. Accordingly, $\mathfrak{b}$ and N are restricted to

$$N = o\left(T^{\frac{2\gamma}{1+2\gamma}}\right) \quad \text{and} \quad \mathfrak{b}T = o\left(T^{\frac{2+2\gamma'}{3+2\gamma'}}\right),$$

which yields that $N = o(\mathfrak{b}T)$ always holds. Therefore

$$\sqrt{N}\left\{\widehat{\sigma}^2(u_o)\widehat{\lambda}(u_o) - \sigma^2(u_o)\lambda(u_o)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\; 2\sigma^4(u_o)\lambda^2(u_o)\int_{\mathbb{R}} k^2(x)dx\right).$$

Case 3: $m = 0$, $m' = 2$. The smoothing parameters are restricted to

$$N = o\left(T^{\frac{2\gamma}{1+2\gamma}}\right) \quad \text{and} \quad \mathfrak{b}T = o\left(T^{\frac{4+2\gamma'}{5+2\gamma'}}\right),$$

in particular $N = o(\mathfrak{b}T)$. As a consequence, we get

$$\begin{aligned}
&\sqrt{N}\left\{\widehat{\sigma}^2(u_o)\widehat{\lambda}(u_o) - \sigma^2(u_o)\lambda(u_o) - \sigma^2(u_o)\cdot BIAS_\lambda\right\} \\
&\quad = \widehat{\lambda}(u_o)\cdot\sqrt{N}\left\{\widehat{\sigma}^2(u_o) - \sigma^2(u_o)\right\} \;+\; \sigma^2(u_o)\frac{\sqrt{N}}{\sqrt{\mathfrak{b}T}}\cdot\sqrt{\mathfrak{b}T}\left\{\widehat{\lambda}(u_o) - \lambda(u_o) - BIAS_\lambda\right\} \\
&\quad \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 2\sigma^4(u_o)\lambda^2(u_o)\int_{\mathbb{R}} k^2(x)dx\right).
\end{aligned}$$

With respect to the above conditions on N and $\mathfrak{b}$, the bias term $\sqrt{N}\sigma^2(u_o)BIAS_\lambda$ vanishes asymptotically, therefore

$$\sqrt{N}\left\{\widehat{\sigma}^2(u_o)\widehat{\lambda}(u_o) - \sigma^2(u_o)\lambda(u_o)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ 2\sigma^4(u_o)\lambda^2(u_o)\int_{\mathbb{R}} k^2(x)dx\right).$$

□

B.4. Proofs for Section 4.4

As the proofs in this section are quite long, we have omitted some details that are similar to those of the preceding section in the noiseless situation. As usual, C is used as a generic constant.

Proof of Theorem 4.15. First we will focus on the asymptotic normality of the statistic

$$\begin{aligned}\widehat{\sigma}^2_{clock,pre}(u_o) := \frac{1}{b}\frac{1}{g_2}\sum_{i=0}^{M_T-1} K\left(\frac{t_{iH}-u_oT}{bT}\right)\left(\overline{\triangle Y}_{t_{iH},T}\right)^2 \\ - \frac{1}{2bH}\frac{\sum_{l=1}^{H-1}h^2(l/H)}{g_2}\sum_{i=1}^{N_T} K\left(\frac{t_i-u_oT}{bT}\right)\left(Y_{t_i,T}-Y_{t_{i-1},T}\right)^2\end{aligned}$$

with $M_T := \lfloor N_T/H \rfloor$ being the random number of blocks. The first term consists of non-overlapping blocks of data so that, conditionally on $N_{\cdot,T}$, a central limit theorem for independent triangular arrays can be applied. The second term of $\widehat{\sigma}^2_{clock,pre}(\cdot)$ remains the same as that of $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ and plays no role in the limit distribution; however, it corrects the bias caused by the additive microstructure noise. In the end of the proof, we will show that the distinction between the two estimates $\widehat{\sigma}^2_{clock,pre}(\cdot)$ and $\widehat{\sigma}^2_{clock,pavg}(\cdot)$ is asymptotically negligible so that both have the same limit distribution.

We divide the proof into following parts. Parts $(I)-(III)$ concern the limit distribution, where the rate of convergence will be balanced by the choice of the block size H. Parts $(IV)-(V)$ are related to biases; especially (V) deals with the bias caused by the additional noise.

$$(I)\quad \sqrt{\frac{bT}{H}}\left\{\frac{1}{b}\frac{1}{g_2}\sum_{i=0}^{M_T-1} K\left(\frac{t_{iH}-u_oT}{bT}\right)\left\{\left(\overline{\triangle X}_{t_{iH},T}\right)^2 - \mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2 \;\middle|\; N_{\cdot,T}\right]\right\}\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\eta_A^2),$$

$$(II)\quad \sqrt{bH}\left\{\frac{1}{b}\frac{2}{g_2}\sum_{i=0}^{M_T-1} K\left(\frac{t_{iH}-u_oT}{bT}\right)\left(\overline{\triangle X}_{t_{iH},T}\right)\left(\overline{\triangle \varepsilon}_{iH}\right)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\eta_B^2), \quad \text{and}$$

$$(III)\quad \sqrt{\frac{bH^3}{T}}\left\{\frac{1}{b}\frac{1}{g_2}\sum_{i=0}^{M_T-1} K\left(\frac{t_{iH}-u_oT}{bT}\right)\left\{\left(\overline{\triangle \varepsilon}_{iH}\right)^2 - \mathbb{E}\left[\left(\overline{\triangle \varepsilon}_{iH}\right)^2\right]\right\}\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\eta_C^2).$$

Furthermore, under the condition $H = \delta \cdot T^{1/2}$ for $\delta \in (0,\infty)$ and the condition $b^{2\alpha+1}T^{1/2} \to 0$ with $\alpha = \min\{m+\gamma, m'+\gamma'\}$, the biases are negligible in the limit, i.e.

$$(IV)\quad \sqrt{bT^{1/2}}\left\{\frac{1}{b}\frac{1}{g_2}\sum_{i=0}^{M_T-1}K\left(\frac{t_{iH}-u_oT}{bT}\right)\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2\;\middle|\;N_{\cdot,T}\right]-\sigma^2(u_o)\lambda(u_o)-BIAS\right\}$$
$$=\ o_p(1),\quad \text{and}$$
$$(V)\quad \sqrt{bT^{1/2}}\left(\frac{1}{b}\frac{1}{g_2}\sum_{i=0}^{M_T-1}K\left(\frac{t_{iH}-u_oT}{bT}\right)\mathbb{E}\left[\left(\overline{\triangle\varepsilon}_{iH}\right)^2\right]\right.$$
$$\left.-\frac{1}{2bH}\frac{\sum_{l=1}^{H-1}h^2(l/H)}{g_2}\sum_{i=1}^{N_T}K\left(\frac{t_i-u_oT}{bT}\right)\left(Y_{t_i,T}-Y_{t_{i-1},T}\right)^2\right)\ =\ o_p(1).$$

We set the left-hand side of (I) to be $\sum_{i=0}^{M_T-1}A_{i,T}$, where

$$A_{i,T}=\sqrt{\frac{bT}{H}}\frac{1}{b}\frac{1}{g_2}K\left(\frac{t_{iH}-u_oT}{bT}\right)\left\{\left(\overline{\triangle X}_{t_{iH},T}\right)^2-\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2\;\middle|\;N_{\cdot,T}\right]\right\}.$$

We see that conditional on $N_{\cdot,T}$, $\{A_{i,T}\}_{i=0,\ldots,M_T-1}$ is an independent triangular array, so it is sufficient to show that (see e.g. Hall and Heyde [38, Corollary 3.1])

$$a)\ \sum_{i=0}^{M_T-1}\mathbb{E}\left[A_{i,T}^2\;\middle|\;N_{\cdot,T}\right]\xrightarrow{\mathbb{P}}\eta_A^2\quad\text{and}\quad b)\ \sum_{i=0}^{M_T-1}\mathbb{E}\left[A_{i,T}^4\;\middle|\;N_{\cdot,T}\right]\xrightarrow{\mathbb{P}}0.$$

Corresponding to independent increments of $X_{\cdot,T}$, it gives

$$\sum_{i=0}^{M_T-1}\mathbb{E}\left[A_{i,T}^2\;\middle|\;N_{\cdot,T}\right]$$
$$=\ \sum_{i=0}^{M_T-1}\frac{T}{bH}\frac{1}{g_2^2}K^2\left(\frac{t_{iH}-u_oT}{bT}\right)\left\{\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^4\;\middle|\;N_{\cdot,T}\right]-\left(\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2\;\middle|\;N_{\cdot,T}\right]\right)^2\right\}$$
$$=\ \sum_{i=0}^{M_T-1}\frac{2}{H}\frac{1}{bT}\frac{1}{g_2^2}K^2\left(\frac{t_{iH}-u_oT}{bT}\right)\times$$
$$\times\left\{\sum_{l=1}^{H-1}g^4\left(\frac{l}{H}\right)\sigma^4\left(\frac{t_{iH+l}}{T}\right)+\left(\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\sigma^2\left(\frac{t_{iH+l}}{T}\right)\right)^2\right\}$$
(see (a_1) and (a_2) below)
$$=\ \sum_{i=1}^{N_T}\frac{2}{H^2}\frac{1}{bT}\frac{1}{g_2^2}K^2\left(\frac{t_i-u_oT}{bT}\right)\sigma^4\left(\frac{t_i}{T}\right)\left\{\sum_{l=1}^{H-1}g^4\left(\frac{l}{H}\right)+\left(\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\right)^2\right\}\ +\ o_p(1)$$
(see (a_3) and (a_4) below)

$$= \frac{2}{H^2}\frac{(\sum g^2(l/H))^2}{g_2^2}\sum_{i=1}^{N_T}\frac{1}{bT}K^2\left(\frac{t_i-u_oT}{bT}\right)\sigma^4\left(\frac{t_i}{T}\right) \;+\; o_p(1)$$
$$\xrightarrow{\mathbb{P}} 2\sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}}K^2(x)dx.$$

(a_1) : by independent increments of $X_{\cdot,T}$,

$$\mathbb{E}\left[(\overline{\triangle X}_{t_{iH},T})^4 \;\middle|\; N_{\cdot,T}\right] = \mathbb{E}\left[\left\{\sum_{l=1}^{H-1} g\left(\frac{l}{H}\right)\left(X_{t_{iH+l,T}} - X_{t_{iH+l-1,T}}\right)\right\}^4 \;\middle|\; N_{\cdot,T}\right]$$
$$= \sum_{l=1}^{H-1} g^4\left(\frac{l}{H}\right)\mathbb{E}\left[\left(X_{t_{iH+l,T}} - X_{t_{iH+l-1,T}}\right)^4 \;\middle|\; N_{\cdot,T}\right]$$
$$+ 3\sum_{l\neq l'} g^2\left(\frac{l}{H}\right)g^2\left(\frac{l'}{H}\right)\mathbb{E}\left[\left(X_{t_{iH+l,T}} - X_{t_{iH+l-1,T}}\right)^2 \;\middle|\; N_{\cdot,T}\right]\times$$
$$\times\,\mathbb{E}\left[\left(X_{t_{iH+l',T}} - X_{t_{iH+l'-1,T}}\right)^2 \;\middle|\; N_{\cdot,T}\right]$$
$$= 3\sum_{l=1}^{H-1} g^4\left(\frac{l}{H}\right)\sigma^4\left(\frac{t_{iH+l}}{T}\right)\frac{1}{T^2} + 3\left(\sum_{l=1}^{H-1} g^2\left(\frac{l}{H}\right)\sigma^2\left(\frac{t_{iH+l}}{T}\right)\frac{1}{T}\right)^2$$
$$- \sum_{l=1}^{H-1} g^4\left(\frac{l}{H}\right)\sigma^4\left(\frac{t_{iH+l}}{T}\right)\frac{1}{T^2}.$$

(a_2) :

$$\left(\mathbb{E}\left[(\overline{\triangle X}_{t_{iH},T})^2 \;\middle|\; N_{\cdot,T}\right]\right)^2 = \left(\sum_{l=1}^{H-1} g^2\left(\frac{l}{H}\right)\mathbb{E}\left[\left(X_{t_{iH+l,T}} - X_{t_{iH+l-1,T}}\right)^2 \;\middle|\; N_{\cdot,T}\right]\right)^2$$
$$= \left(\sum_{l=1}^{H-1} g^2\left(\frac{l}{H}\right)\sigma^2\left(\frac{t_{iH+l}}{T}\right)\frac{1}{T}\right)^2.$$

(a_3) : (and (a_4) analogously)

$$H\cdot\sum_{i=0}^{M_T-1}\frac{1}{H}\frac{1}{bT}\frac{1}{g_2^2}K^2\left(\frac{t_{iH}-u_oT}{bT}\right)\sum_{l=1}^{H-1} g^4\left(\frac{l}{H}\right)\sigma^4\left(\frac{t_{iH+l}}{T}\right)$$
$$- \sum_{j=1}^{N_T}\frac{1}{H}\frac{1}{bT}\frac{1}{g_2^2}K^2\left(\frac{t_j-u_oT}{bT}\right)\sum_{l=1}^{H-1} g^4\left(\frac{l}{H}\right)\sigma^4\left(\frac{t_{\lfloor j/H\rfloor\cdot H+l}}{T}\right)$$

$$
\begin{aligned}
&+ \sum_{j=1}^{N_T} \frac{1}{H}\frac{1}{bT}\frac{1}{g_2^2} K^2\left(\frac{t_j - u_o T}{bT}\right) \sum_{l=1}^{H-1} g^4\left(\frac{l}{H}\right) \sigma^4\left(\frac{t_{\lfloor j/H \rfloor \cdot H + l}}{T}\right) \\
&- \sum_{j=1}^{N_T} \frac{1}{H}\frac{1}{bT}\frac{1}{g_2^2} K^2\left(\frac{t_j - u_o T}{bT}\right) \sum_{l=1}^{H-1} g^4\left(\frac{l}{H}\right) \sigma^4\left(\frac{t_j}{T}\right) \\
&=: (*) + (**) \\
&= O\left(\frac{H}{bT}\right) + O_p\left(\left|\frac{H}{T}\right|^{\gamma}\right) \;=\; o_p(1),
\end{aligned}
$$

since $H/bT \to 0$. As K has bounded first derivatives and $\sigma(\cdot) \in \mathcal{C}^{m,\gamma}$,

$$
\begin{aligned}
(*) &\leq C \cdot \sum_{j=1}^{N_T} \frac{1}{H}\frac{1}{bT} \left| K^2\left(\frac{t_{\lfloor j/H \rfloor \cdot H} - u_o T}{bT}\right) - K^2\left(\frac{t_j - u_o T}{bT}\right)\right| \cdot \sum_{l=1}^{H-1} g^4\left(\frac{l}{H}\right) \\
&\leq C \cdot \sum_{j=1}^{N_T} \frac{1}{bT} \left|\left(K^2(...)\right)^{(1)}\right| \left|\frac{t_{\lfloor j/H \rfloor \cdot H} - t_j}{bT}\right| \cdot \frac{1}{H} \sum_{l=1}^{H-1} g^4\left(\frac{l}{H}\right) \;=\; O\left(\frac{H}{bT}\right),
\end{aligned}
$$

(see also Lemma 2), and

$$
\begin{aligned}
(**) &\leq C \cdot \sum_{j=1}^{N_T} \frac{1}{H}\frac{1}{bT} K^2\left(\frac{t_j - u_o T}{bT}\right) \sum_{l=1}^{H-1} g^4\left(\frac{l}{H}\right) \left|\sigma^4\left(\frac{t_{\lfloor j/H \rfloor \cdot H + l}}{T}\right) - \sigma^4\left(\frac{t_j}{T}\right)\right| \\
&\leq C \cdot \frac{1}{H} \sum_{l=1}^{H-1} g^4\left(\frac{l}{H}\right) \cdot \sum_{j=1}^{N_T} \frac{1}{bT} K^2\left(\frac{t_j - u_o T}{bT}\right) \cdot O\left(\left|\frac{H}{T}\right|^{\gamma}\right) \;=\; O_p\left(\left|\frac{H}{T}\right|^{\gamma}\right).
\end{aligned}
$$

We now turn to condition *b*):

$$
\begin{aligned}
&\sum_{i=0}^{M_T - 1} \mathbb{E}\left[A_{i,T}^4 \;\middle|\; N_{\cdot,T}\right] \\
&\quad \leq C \cdot \frac{T^2}{b^2 H^2} \sum_{i=0}^{M_T-1} K^4\left(\frac{t_{iH} - u_o T}{bT}\right) \mathbb{E}\left[\left\{\left(\overline{\triangle X}_{t_{iH},T}\right)^2 - \mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2 \;\middle|\; N_{\cdot,T}\right]\right\}^4 \;\middle|\; N_{\cdot,T}\right] \\
&\quad \leq C \cdot \frac{T^2}{b^2 H^2} \sum_{i=0}^{M_T-1} K^4\left(\frac{t_{iH} - u_o T}{bT}\right) \left(\sum_{l=1}^{H-1} g^2\left(\frac{l}{H}\right)\right)^4 \frac{1}{T^4} \qquad \text{(see (***))} \\
&\quad = o_p(1),
\end{aligned}
$$

since $H/bT \to 0$ and $\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^8 \;\middle|\; N_{\cdot,T}\right] \le C \cdot \left\{\sum_{l=1}^{H-1} g^2\left(\frac{l}{H}\right)\sigma^2\left(\frac{t_{iH+l}}{T}\right)\right\}^4 \frac{1}{T^4}$ holds.

$$
\begin{aligned}
(***) &= \mathbb{E}\left[\left\{\left(\overline{\triangle X}_{t_{iH},T}\right)^2 - \mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2 \;\middle|\; N_{\cdot,T}\right]\right\}^4 \;\middle|\; N_{\cdot,T}\right] \\
&= \mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^8 \;\middle|\; N_{\cdot,T}\right] - 4\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^6 \;\middle|\; N_{\cdot,T}\right] \cdot \mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2 \;\middle|\; N_{\cdot,T}\right] \\
&\quad + 6\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^4 \;\middle|\; N_{\cdot,T}\right] \cdot \left(\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2 \;\middle|\; N_{\cdot,T}\right]\right)^2 \\
&\quad - 4\left(\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2 \;\middle|\; N_{\cdot,T}\right]\right)^4 + \left(\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2 \;\middle|\; N_{\cdot,T}\right]\right)^4 .
\end{aligned}
$$

In order to derive the limit distribution in (II) we set

$$
\sqrt{bH}\left\{\frac{1}{b}\frac{2}{g_2}\sum_{i=0}^{M_T-1} K\left(\frac{t_{iH}-u_oT}{bT}\right)\left(\overline{\triangle X}_{t_{iH},T}\right)\left(\overline{\triangle \varepsilon}_{iH}\right)\right\} \;=:\; \sum_{i=0}^{M_T-1} B_{i,T}.
$$

Likewise we show that (conditional on $N_{\cdot,T}$)

$$
c)\;\sum_{i=0}^{M_T-1}\mathbb{E}\left[B_{i,T}^2 \;\middle|\; N_{\cdot,T}\right] \xrightarrow{\mathbb{P}} \eta_B^2 \quad \text{and} \quad d)\;\sum_{i=0}^{M_T-1}\mathbb{E}\left[B_{i,T}^4 \;\middle|\; N_{\cdot,T}\right] \xrightarrow{\mathbb{P}} 0.
$$

Since

$$
\mathbb{E}\left[\left(\overline{\triangle\varepsilon}_{iH}\right)^2\right] \;=\; \mathbb{E}\left[\left(\sum_{l=1}^{H-1} h\left(\frac{l}{H}\right)\varepsilon_{iH+l}\right)^2\right] \;=\; \omega^2\sum_{l=1}^{H-1} h^2\left(\frac{l}{H}\right),
$$

it implies that

$$
\begin{aligned}
&\mathbb{E}\left[B_{i,T}^2 \;\middle|\; N_{\cdot,T}\right] \\
&= \frac{H}{b}\frac{4}{g_2^2}K^2\left(\frac{t_{iH}-u_oT}{bT}\right)\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2 \;\middle|\; N_{\cdot,T}\right]\cdot\mathbb{E}\left[\left(\overline{\triangle\varepsilon}_{iH}\right)^2 \;\middle|\; N_{\cdot,T}\right] \\
&= \frac{H}{b}\frac{4}{g_2^2}K^2\left(\frac{t_{iH}-u_oT}{bT}\right)\left(\sum_{l=1}^{H-1} g^2\left(\frac{l}{H}\right)\sigma^2\left(\frac{t_{iH+l}}{T}\right)\frac{1}{T}\right)\left(\omega^2\sum_{l=1}^{H-1} h^2\left(\frac{l}{H}\right)\right).
\end{aligned}
$$

(see (a_2))

Hence

$$
\begin{aligned}
&\sum_{i=0}^{M_T-1}\mathbb{E}\left[B_{i,T}^2 \;\middle|\; N_{\cdot,T}\right] \\
&\quad= \frac{4\omega^2}{g_2^2}\cdot H\sum_{i=0}^{M_T-1}\frac{1}{bT}K^2\left(\frac{t_{iH}-u_oT}{bT}\right)\sigma^2\left(\frac{t_{iH+l}}{T}\right)\cdot\sum_{l=1}^{H-1} g^2\left(\frac{l}{H}\right)\cdot\sum_{l=1}^{H-1} h^2\left(\frac{l}{H}\right) \\
&\quad\xrightarrow{\mathbb{P}} 4\omega^2\sigma^2(u_o)\lambda(u_o)(g_2'/g_2)\int_{\mathbb{R}} K^2(x)dx, \qquad \text{(similar to } (a_3))
\end{aligned}
$$

since

$$\sum_{l=1}^{H-1} h^2\left(\frac{l}{H}\right) \;=\; \sum_{l=1}^{H-1}\left\{g\left(\frac{l+1}{H}\right)-g\left(\frac{l}{H}\right)\right\}^2 \;=\; \sum_{l=1}^{H-1}\left\{g^{(1)}\left(\frac{l}{H}\right)\frac{1}{H}+o\left(\frac{1}{H}\right)\right\}^2.$$

Similarly,

$$\begin{aligned}
&\sum_{i=0}^{M_T-1}\mathbb{E}\left[B_{i,T}^4 \;\middle|\; N_{\cdot,T}\right]\\
&=\sum_{i=0}^{M_T-1}\frac{H^2}{b^2}\frac{16}{g_2^4}K^4\left(\frac{t_{iH}-u_oT}{bT}\right)\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^4 \;\middle|\; N_{\cdot,T}\right]\cdot\mathbb{E}\left[\left(\overline{\triangle\varepsilon}_{iH}\right)^4 \;\middle|\; N_{\cdot,T}\right]\\
&\le C\cdot\sum_{i=0}^{M_T-1}\frac{H^2}{b^2}K^4\left(\frac{t_{iH}-u_oT}{bT}\right)\left(\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\frac{1}{T}\right)^2\left(\sum_{l=1}^{H-1}h^2\left(\frac{l}{H}\right)\right)^2 \quad \text{(see } (a_1)\text{)}\\
&\le C\cdot\frac{H^2}{bT}\sum_{i=0}^{M_T-1}\frac{1}{bT}K^4\left(\frac{t_{iH}-u_oT}{bT}\right)\left(\frac{1}{H}\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\right)^2\left(\frac{1}{H}\sum_{l=1}^{H-1}\left(g^{(1)}\left(\frac{l}{H}\right)\right)^2+o\left(\frac{1}{H}\right)\right)^2\\
&\xrightarrow{\mathbb{P}} O\left(\frac{H}{bT}\right)\cdot\lambda(u_o)\int_{\mathbb{R}}K^4(x)dx\cdot\left(\int g^2(x)dx\right)^2\cdot\left(\int g'(x)^2dx\right)^2\xrightarrow{\mathbb{P}}\;0.
\end{aligned}$$

Therefore c) and d) have been proven. We proceed analogously to show the limit distribution (III). We denote the left-hand side of the assertion by $\sum_{i=0}^{M_T-1}C_{i,T}$ and show that

$$e)\;\sum_{i=0}^{M_T-1}\mathbb{E}\left[C_{i,T}^2 \;\middle|\; N_{\cdot,T}\right]\xrightarrow{\mathbb{P}}\eta_C^2 \quad\text{and}\quad f)\;\sum_{i=0}^{M_T-1}\mathbb{E}\left[C_{i,T}^4 \;\middle|\; N_{\cdot,T}\right]\xrightarrow{\mathbb{P}}0.$$

On account of the assumption $\mathbb{E}\left[\varepsilon_i^4\right]=\theta\omega^4$, $\theta\in\mathbb{R}^+$, we have (similar to the calculation of $\mathbb{E}\left[(\overline{\triangle X}_{i,T})^4\right]$)

$$\begin{aligned}
\mathbb{E}\left[(\overline{\triangle\varepsilon}_{iH})^4\right] &= \mathbb{E}\left[\left(\sum_{l=1}^{H-1}h\left(\frac{l}{H}\right)\varepsilon_{iH+l}\right)^4\right]\\
&= \sum_{l=1}^{H-1}h^4\left(\frac{l}{H}\right)\mathbb{E}\left[\varepsilon_{iH+l}^4\right]+3\sum_{l\neq l'}h^2\left(\frac{l}{H}\right)h^2\left(\frac{l'}{H}\right)\mathbb{E}\left[\varepsilon_{iH+l}^2\right]\mathbb{E}\left[\varepsilon_{iH+l'}^2\right]\\
&= (\theta-1)\sum_{l=1}^{H-1}h^4\left(\frac{l}{H}\right)\omega^4+3\left(\sum_{l=1}^{H-1}h^2\left(\frac{l}{H}\right)\right)^2\omega^4,
\end{aligned}$$

thus

$$
\begin{aligned}
&\sum_{i=0}^{M_T-1} \mathbb{E}\left[C_{i,T}^2 \;\middle|\; N_{\cdot,T}\right] \\
&\quad= \sum_{i=0}^{M_T-1} \frac{H^3}{bT}\frac{1}{g_2^2}K^2\left(\frac{t_{iH}-u_oT}{bT}\right)\left\{\mathbb{E}\left[\left(\overline{\triangle\varepsilon}_{iH}\right)^4\right]-\left(\mathbb{E}\left[\left(\overline{\triangle\varepsilon}_{iH}\right)^2\right]\right)^2\right\} \\
&\quad= \sum_{i=0}^{M_T-1} \frac{H^3}{bT}\frac{\omega^4}{g_2^2}K^2\left(\frac{t_{iH}-u_oT}{bT}\right)\left\{(\theta-1)\sum_{l=1}^{H-1}h^4\left(\frac{l}{H}\right)+2\left(\sum_{l=1}^{H-1}h^2\left(\frac{l}{H}\right)\right)^2\right\} \\
&\quad= \frac{1}{H}\cdot H\sum_{i=0}^{M_T-1}\frac{1}{bT}K^2\left(\frac{t_{iH}-u_oT}{bT}\right)\cdot\frac{(\theta-1)\omega^4}{g_2^2}\cdot\left\{\sum_{l=1}^{H-1}\left(g'\left(\frac{l}{H}\right)\right)^4\frac{1}{H}+o\left(\frac{1}{H}\right)\right\} \\
&\qquad+\frac{2\omega^4}{g_2^2}\cdot H\sum_{i=0}^{M_T-1}\frac{1}{bT}K^2\left(\frac{t_{iH}-u_oT}{bT}\right)\cdot\left\{\sum_{l=1}^{H-1}\left(g'\left(\frac{l}{H}\right)\right)^2\frac{1}{H}+o\left(\frac{1}{H}\right)\right\}^2 \\
&\quad\xrightarrow{\mathbb{P}} 2\omega^4\lambda(u_o)(g_2'/g_2)^2\int_{\mathbb{R}}K^2(x)dx,
\end{aligned}
$$

which leads to $e)$. In fact, $f)$ can be established in the same manner as before, therefore omitted. To build the joint distribution of (I), (II) and (III), it suffices to show that

$$
\sum_{i=0}^{M_T-1} a\sqrt{\delta}A_{i,T}+b\frac{1}{\sqrt{\delta}}B_{i,T}+c\frac{1}{\sqrt{\delta^3}}C_{i,T} \;\xrightarrow{\mathcal{D}}\; a\sqrt{\delta}A+b\frac{1}{\sqrt{\delta}}B+c\frac{1}{\sqrt{\delta^3}}C
$$

by Cramer-Wold's theorem, for all $a,b,c\in\mathbb{R}$, where A,B and C are the limits of (I), (II) and (III), respectively. The pre-averaging block size H is chosen to equal $\delta\cdot T^{1/2}$ with $\delta\in(0,\infty)$. Indeed, this joint limit is a direct consequence of $a)-f)$. To sum up, we have shown the first main part of the limit distribution of $\widehat{\sigma}^2_{clock,pre}(u_o)$, i.e.

$$
\begin{aligned}
&\sqrt{bT^{1/2}}\left\{\sum_{i=0}^{M_T-1}\frac{1}{b}\frac{1}{g_2}K\left(\frac{t_{iH}-u_oT}{bT}\right)\left\{\left(\overline{\triangle Y}_{t_{iH},T}\right)^2-\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2\;\middle|\;N_{\cdot,T}\right]-\mathbb{E}\left[\left(\overline{\triangle\varepsilon}_{iH}\right)^2\right]\right\}\right\} \\
&\qquad\xrightarrow{\mathcal{D}}\; \mathcal{N}(0,\;\delta\eta_A^2+\frac{1}{\delta}\eta_B^2+\frac{1}{\delta^3}\eta_C^2),
\end{aligned}
$$

since $\left(\overline{\triangle Y}_{t_{iH},T}\right)=\left(\overline{\triangle X}_{t_{iH},T}\right)+\left(\overline{\triangle\varepsilon}_{t_{iH},T}\right)$.

We carry out the proof by showing the asymptotic biases in (IV) and (V). In order to

derive (IV) we see that

$$
\begin{aligned}
&\sqrt{bT^{1/2}}\left(\frac{1}{b}\frac{1}{g_2}\sum_{i=0}^{M_T-1}K\left(\frac{t_{iH}-u_oT}{bT}\right)\mathbb{E}\left[\left(\overline{\triangle X}_{t_{iH},T}\right)^2\;\middle|\;N_{\cdot,T}\right]-\sigma^2(u_o)\lambda(u_o)-BIAS\right)\\
&=\sqrt{bT^{1/2}}\frac{1}{H}\frac{1}{g_2}\cdot\left(H\sum_{i=0}^{M_T-1}\frac{1}{bT}K\left(\frac{t_{iH}-u_oT}{bT}\right)\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\sigma^2\left(\frac{t_{iH+l}}{T}\right)\right.\\
&\quad\left.-\sum_{j=1}^{N_T}\frac{1}{bT}K\left(\frac{t_j-u_oT}{bT}\right)\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\sigma^2\left(\frac{t_{\lfloor j/H\rfloor\cdot H+l}}{T}\right)\right)\\
&+\sqrt{bT^{1/2}}\frac{1}{H}\frac{1}{g_2}\sum_{j=1}^{N_T}\frac{1}{bT}K\left(\frac{t_j-u_oT}{bT}\right)\cdot\left(\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\sigma^2\left(\frac{t_{\lfloor j/H\rfloor\cdot H+l}}{T}\right)\right.\\
&\quad\left.-\sigma^2\left(\frac{t_j}{T}\right)\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\right)\\
&+\sqrt{bT^{1/2}}\left(\frac{1}{H}\frac{1}{g_2}\sum_{j=1}^{N_T}\frac{1}{bT}K\left(\frac{t_j-u_oT}{bT}\right)\sigma^2\left(\frac{t_j}{T}\right)\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\right.\\
&\quad\left.-\sigma^2(u_o)\lambda(u_o)-\frac{1}{2}\left(\sigma^2(u_o)\lambda(u_o)\right)^{(2)}b^2\int_{\mathbb{R}}x^2K(x)dxI_{\{m=m'=2\}}\right)\\
&=:(i)+(ii)+(iii)\;=\;o_p(1),
\end{aligned}
\tag{A.5.}
$$

since K has first bounded derivatives and $\sigma(\cdot)\in\mathcal{C}^{m,\gamma}$; see also $(a3)$. The last equality (A.5.) is satisfied by the bandwidth condition $b^{2\alpha+1}T^{1/2}\to 0$. More precisely,

$$
\begin{aligned}
(i)&\le C\cdot\sqrt{bT^{1/2}}\frac{1}{H}\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\sum_{j=0}^{N_T}\frac{1}{bT}\left|K\left(\frac{t_{\lfloor j/H\rfloor\cdot H}-u_oT}{bT}\right)-K\left(\frac{t_j-u_oT}{bT}\right)\right|\\
&=O\left(\sqrt{bT^{1/2}}\cdot\frac{H}{bT}\right)\;=\;o(1),
\end{aligned}
$$

$$
\begin{aligned}
(ii)&\le C\cdot\sqrt{bT^{1/2}}\frac{1}{H}\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\sum_{j=1}^{N_T}\frac{1}{bT}\left|K\left(\frac{t_j-u_oT}{bT}\right)\right|\left|\sigma^2\left(\frac{t_{\lfloor j/H\rfloor\cdot H+l}}{T}\right)-\sigma^2\left(\frac{t_j}{T}\right)\right|\\
&=O_p\left(\sqrt{bT^{1/2}}\left|\frac{H}{T}\right|^{\gamma}\right)\;=\;o_p(1),
\end{aligned}
$$

and

$$
\begin{aligned}
(iii) = \sqrt{bT^{1/2}}\Bigg(& \frac{1}{H}\frac{1}{g_2}\sum_{j=1}^{N_T}\frac{1}{bT}K\left(\frac{t_j-u_oT}{bT}\right)\sigma^2\left(\frac{t_j}{T}\right)\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right) \\
& - \mathbb{E}\left[\frac{1}{H}\frac{1}{g_2}\sum_{j=1}^{N_T}\frac{1}{bT}K\left(\frac{t_j-u_oT}{bT}\right)\sigma^2\left(\frac{t_j}{T}\right)\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\right]\Bigg) \\
& + \sqrt{bT^{1/2}}\Bigg(\mathbb{E}\left[\frac{1}{H}\frac{1}{g_2}\sum_{j=1}^{N_T}\frac{1}{bT}K\left(\frac{t_j-u_oT}{bT}\right)\sigma^2\left(\frac{t_j}{T}\right)\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\right] \\
& - \left\{\sigma^2(u_o)\lambda(u_o)+\frac{1}{2}\left(\sigma^2(u_o)\lambda(u_o)\right)^{(2)}b^2\int_{\mathbb{R}}x^2K(x)dx\cdot I_{\{m=m'=2\}}\right\}\Bigg) \\
= & \sqrt{bT^{1/2}}\left(\int_0^T\frac{1}{bT}K\left(\frac{t-u_oT}{bT}\right)\sigma^2\left(\frac{t}{T}\right)dM_{t,T}\right)\cdot\frac{\frac{1}{H}\sum_{l=1}^{H-1}g^2(l/H)}{g_2}+O\left(\sqrt{bT^{1/2}}b^{\min(\gamma,\gamma')}\right) \\
= & \, o_p(1).
\end{aligned}
$$

Finally, we separate (V) into two summands

$$
\begin{aligned}
& \sqrt{bT^{1/2}}\left\{\frac{1}{b}\frac{1}{g_2}\sum_{i=0}^{M_T-1}K\left(\frac{t_{iH}-u_oT}{bT}\right)\mathbb{E}\left[\left(\overline{\triangle\varepsilon}_{iH}\right)^2\right]-\frac{T}{H}\frac{\lambda(u_0)}{g_2}\sum_{l=1}^{H-1}h^2\left(\frac{l}{H}\right)\omega^2\right\} \\
& +\sqrt{bT^{1/2}}\left\{\frac{T}{H}\frac{\lambda(u_0)}{g_2}\sum_{l=1}^{H-1}h^2\left(\frac{l}{H}\right)\omega^2-\frac{T}{2H}\frac{\sum_{l=1}^{H-1}h^2(l/H)}{g_2}\cdot\frac{1}{bT}\sum_{l=1}^{N_T}K\left(\frac{t_i-u_oT}{bT}\right)\left(Y_{t_i,T}-Y_{t_{i-1},T}\right)^2\right\} \\
& \quad =: (iv)+(v).
\end{aligned}
$$

Since $\mathbb{E}\left[\left(\overline{\triangle\varepsilon}_{iH}\right)^2\right]=\sum_{l=1}^{H-1}h^2\left(\frac{l}{H}\right)\omega^2$, direct calculations yield

$$
\begin{aligned}
\mathbb{E}\left[(iv)^2\right] & \le C\cdot bT^{1/2}\left(\sum_{l=1}^{H-1}h^2\left(\frac{l}{H}\right)\right)^2\mathbb{E}\left[\frac{1}{b}\sum_{i=0}^{M_T-1}K\left(\frac{t_{iH}-u_oT}{bT}\right)-\frac{T}{H}\lambda(u_o)\right]^2 \\
& \le C\cdot bT^{1/2}\frac{T^2}{H^2}\left(\sum_{l=1}^{H-1}h^2\left(\frac{l}{H}\right)\right)^2\times \\
& \quad\times\left(\mathbb{E}\left[\frac{H}{bT}\sum_{i=0}^{M_T-1}K\left(\frac{t_{iH}-u_oT}{bT}\right)-\mathbb{E}\left[\frac{H}{bT}\sum_{i=0}^{M_T-1}K\left(\frac{t_{iH}-u_oT}{bT}\right)\right]\right]^2\right. \\
& \quad\left.+\left(\mathbb{E}\left[\frac{H}{bT}\sum_{i=0}^{M_T-1}K\left(\frac{t_{iH}-u_oT}{bT}\right)\right]-\lambda(u_o)\right)^2\right) \\
& = O\left(bT^{1/2}\frac{T^2}{H^2}\frac{1}{H^2}\frac{1}{bT}\right)+O\left(bT^{1/2}\frac{T^2}{H^2}\frac{1}{H^2}b^{2\gamma'}\right) \;=\; o(1).
\end{aligned}
$$

The last term (v) is divided again into the sum of (v_1) and (v_2), where

$$(v_1) := \sqrt{bT^{1/2}}\Bigg[\frac{T}{H}\frac{\lambda(u_0)}{g_2}\sum_{l=1}^{H-1}h^2\left(\frac{l}{H}\right)\omega^2 - \frac{T}{2H}\frac{\sum_{l=1}^{H-1}h^2(l/H)}{g_2}\frac{1}{bT}\sum_{l=1}^{N_T}K\left(\frac{t_i-u_oT}{bT}\right)(\varepsilon_i-\varepsilon_{i-1})^2\Bigg]$$

and

$$(v_2) := -\sqrt{bT^{1/2}}\Bigg[\frac{T}{2H}\frac{\sum_{l=1}^{H-1}h^2(l/H)}{g_2}\frac{1}{bT}\sum_{l=1}^{N_T}K\left(\frac{t_i-u_oT}{bT}\right)\times \times\Big(\left(X_{t_i,T}-X_{t_{i-1},T}\right)^2+2\left(X_{t_i,T}-X_{t_{i-1},T}\right)(\varepsilon_i-\varepsilon_{i-1})\Big)\Bigg].$$

It is easy to show that $\mathbb{E}\left[(v_1)^2\right]=o(1)$ and $\mathbb{E}\left[\,|(v_2)|\,\right]=o(1)$ hold. In summary, we have therefore shown the asymptotic normality for $\widehat{\sigma}^2_{clock,pre}(u_o)$.

To complete the proof, the difference between $\widehat{\sigma}^2_{clock,pavg}(u_o)$ and $\widehat{\sigma}^2_{clock,pre}(u_o)$ needs to be determined; in particular, we show that

$$\sqrt{bT^{1/2}}\left\{\widehat{\sigma}^2_{clock,pre}(u_o)-\widehat{\sigma}^2_{clock,pavg}(u_o)\right\} \;=\; o_p(1).$$

Since the derivative of K is bounded, it is enough to verify that (see also $(a3)$)

$$\begin{aligned}
&\sqrt{bT^{1/2}}\frac{1}{bH}\frac{1}{g_2}\sum_{i=1}^{N_T}K\left(\frac{t_i-u_oT}{bT}\right)\left\{\left(\overline{\triangle Y}_{t_{\lfloor i/H\rfloor\cdot H},T}\right)^2-\left(\overline{\triangle Y}_{t_i,T}\right)^2\right\}\\
&=\sqrt{bT^{1/2}}\frac{1}{bH}\frac{1}{g_2}\sum_{i=1}^{N_T}K\left(\frac{t_i-u_oT}{bT}\right)\Bigg(\left\{\left(\overline{\triangle X}_{t_{\lfloor i/H\rfloor\cdot H},T}\right)^2-\left(\overline{\triangle X}_{t_i,T}\right)^2\right\}\\
&\quad+\left\{\left(\overline{\triangle\varepsilon}_{\lfloor i/H\rfloor\cdot H}\right)^2-\left(\overline{\triangle\varepsilon}_i\right)^2\right\}+2\left\{\left(\overline{\triangle X}_{t_{\lfloor i/H\rfloor\cdot H},T}\right)\left(\overline{\triangle\varepsilon}_{\lfloor i/H\rfloor\cdot H}\right)-\left(\overline{\triangle X}_{t_i,T}\right)\left(\overline{\triangle\varepsilon}_i\right)\right\}\Bigg)\\
&=:(T_1)+(T_2)+(T_3)\;=\;o_p(1). \qquad \text{(A.6.)}
\end{aligned}$$

We perform only the proof of the first term (T_1), since (T_2) and (T_3) can be done in the same manner by employing $\overline{\triangle\varepsilon}_i=-\sum_{l=1}^{H-1}h\left(\frac{l}{H}\right)\varepsilon_{i+l}$; compare with (II) and (III). We will prove below that $\mathbb{E}\left[(T_1)^2\;\middle|\;N_{\cdot,T}\right]=o_p(1)$ which concludes that (A.6.) $=o_p(1)$.

$$\mathbb{E}\left[(T_1)^2 \;\middle|\; N_{\cdot,T}\right] = \frac{T^{1/2}}{bH^2}\frac{1}{g_2^2}\sum_{i,j=1}^{N_T} K\left(\frac{t_i - u_oT}{bT}\right) K\left(\frac{t_j - u_oT}{bT}\right) \times$$

$$\times\Bigg(\mathbb{E}\left[\left(\overline{\triangle X}_{t_{\lfloor i/H\rfloor\cdot H},T}\right)^2\left(\overline{\triangle X}_{t_{\lfloor j/H\rfloor\cdot H},T}\right)^2 \;\middle|\; N_{\cdot,T}\right] - \mathbb{E}\left[\left(\overline{\triangle X}_{t_{\lfloor i/H\rfloor\cdot H},T}\right)^2\left(\overline{\triangle X}_{t_j,T}\right)^2 \;\middle|\; N_{\cdot,T}\right]$$

$$- \mathbb{E}\left[\left(\overline{\triangle X}_{t_i,T}\right)^2\left(\overline{\triangle X}_{t_{\lfloor j/H\rfloor\cdot H},T}\right)^2 \;\middle|\; N_{\cdot,T}\right] + \mathbb{E}\left[\left(\overline{\triangle X}_{t_i,T}\right)^2\left(\overline{\triangle X}_{t_j,T}\right)^2 \;\middle|\; N_{\cdot,T}\right]\Bigg)$$

$=: (\triangle_1) - (\triangle_2) - (\triangle_3) + (\triangle_4).$

The first term is split up into three small terms:

$$\begin{aligned}(\triangle_1) = {} & \frac{T^{1/2}}{bH^2}\frac{1}{g_2^2}\sum_{i=j} K^2\left(\frac{t_i - u_oT}{bT}\right)\mathbb{E}\left[\left(\overline{\triangle X}_{t_{\lfloor i/H\rfloor\cdot H},T}\right)^2 \;\middle|\; N_{\cdot,T}\right] \\ & + \frac{T^{1/2}}{bH^2}\frac{1}{g_2^2}\sum_{|i-j|\geq H} K\left(\frac{t_i - u_oT}{bT}\right) K\left(\frac{t_j - u_oT}{bT}\right)\times \\ & \quad \times \mathbb{E}\left[\left(\overline{\triangle X}_{t_{\lfloor i/H\rfloor\cdot H},T}\right)^2 \;\middle|\; N_{\cdot,T}\right]\cdot\mathbb{E}\left[\left(\overline{\triangle X}_{t_{\lfloor j/H\rfloor\cdot H},T}\right)^2 \;\middle|\; N_{\cdot,T}\right] \\ & \qquad\qquad \text{(since both terms are independent.)} \\ & + \frac{T^{1/2}}{bH^2}\frac{1}{g_2^2}\sum_{0<|i-j|<H} K\left(\frac{t_i - u_oT}{bT}\right) K\left(\frac{t_j - u_oT}{bT}\right)\times \\ & \quad \times \mathbb{E}\left[\left(\overline{\triangle X}_{t_{\lfloor i/H\rfloor\cdot H},T}\right)^2\left(\overline{\triangle X}_{t_{\lfloor j/H\rfloor\cdot H},T}\right)^2 \;\middle|\; N_{\cdot,T}\right] \\ =: {} & (\triangle_{1,1}) + (\triangle_{1,2}) + (\triangle_{1,3}).\end{aligned}$$

Likewise, we expand $(\triangle_2) = (\triangle_{2,1}) + (\triangle_{2,2}) + (\triangle_{2,3})$. It is clear that $(\triangle_{1,1})$ and $(\triangle_{2,1})$ are of smaller order than the others, therefore neglected. Since $(\triangle_{1,2})$ is equal to $(\triangle_{2,2})$, they cancel each other out. Lastly,

$$\begin{aligned}&(\triangle_{1,3}) - (\triangle_{2,3}) \\ &\quad = \frac{T^{1/2}}{bH^2}\frac{1}{g_2^2}\sum_{0<|i-j|<H} K\left(\frac{t_i - u_oT}{bT}\right) K\left(\frac{t_j - u_oT}{bT}\right)\times \\ &\qquad \times \mathbb{E}\left[\left(\overline{\triangle X}_{t_{\lfloor i/H\rfloor\cdot H},T}\right)^2\left\{\left(\overline{\triangle X}_{t_{\lfloor j/H\rfloor\cdot H},T}\right)^2 - \left(\overline{\triangle X}_{t_j,T}\right)^2\right\} \;\middle|\; N_{\cdot,T}\right] \\ &\quad = \frac{2T^{1/2}}{bH^2}\frac{1}{g_2^2}\sum_{i=1}^{N_T}\sum_{\alpha=1}^{H-1} K\left(\frac{t_i - u_oT}{bT}\right) K\left(\frac{t_{i+\alpha} - u_oT}{bT}\right)\times \\ &\qquad \times \mathbb{E}\left[\left(\overline{\triangle X}_{t_{\lfloor i/H\rfloor\cdot H},T}\right)^2\left\{\left(\overline{\triangle X}_{t_{\lfloor (i+\alpha)/H\rfloor\cdot H},T}\right)^2 - \left(\overline{\triangle X}_{t_{i+\alpha},T}\right)^2\right\} \;\middle|\; N_{\cdot,T}\right]\end{aligned}$$

$$= \frac{2T^{1/2}}{bH^2}\frac{1}{g_2^2}\sum_{i=1}^{N_T} K^2\left(\frac{t_i - u_o T}{bT}\right) \times$$

$$\times \mathbb{E}\left[\left(\overline{\triangle X}_{t_{\lfloor i/H\rfloor\cdot H},T}\right)^2 \underbrace{\sum_{\alpha=1}^{H-1}\left\{\left(\overline{\triangle X}_{t_{\lfloor (i+\alpha)/H\rfloor\cdot H},T}\right)^2 - \left(\overline{\triangle X}_{t_{i+\alpha},T}\right)^2\right\}}_{(\clubsuit)} \;\middle|\; N_{\cdot,T}\right] + \; o(1)$$

$$= \frac{2T^{1/2}}{bH^2}\frac{1}{g_2^2}\sum_{i=1}^{N_T} K^2\left(\frac{t_i - u_o T}{bT}\right) o\left(\frac{H^3}{T^2}\right) + o(1) \;=\; o_p(1),$$

since g is differentiable, $g^{(1)}$ is Lipschitz's continuous, and

$$(\clubsuit) = \sum_{\alpha=1}^{H-1}\left\{\left(\overline{\triangle X}_{t_{\lfloor (i+\alpha)/H\rfloor\cdot H},T} - \overline{\triangle X}_{t_{i+\alpha},T}\right)\cdot\left(\overline{\triangle X}_{t_{\lfloor (i+\alpha)/H\rfloor\cdot H},T} + \overline{\triangle X}_{t_{i+\alpha},T}\right)\right\}$$

$$= \sum_{\alpha=1}^{H-1}\sum_{l,l'=1}^{H-(j \mod H)}\left\{h\left(\frac{l+(j \mod H)}{H}\right) - h\left(\frac{l}{H}\right)\right\}\times$$

$$\times\left\{h\left(\frac{l'+(j \mod H)}{H}\right) - h\left(\frac{l'}{H}\right)\right\}\left(X_{t_{\alpha+l,T}} - X_{t_{\alpha+l-1,T}}\right)\left(X_{t_{\alpha+l',T}} - X_{t_{\alpha+l'-1,T}}\right).$$

Thus, we get that $(\triangle_1) - (\triangle_2) = o_p(1)$. Analogously, $(\triangle_4) - (\triangle_3) = o_p(1)$ and hence $(T_1) = o_p(1)$. □

Proof of Theorem 4.16. The proof of this theorem is analogous to that of Theorem 4.15 with an exception of the bias term (IV), which is non-trivial in this case. Therefore, we have omitted its details and give only the outline of the proof. We begin by reformulating the tick-time volatility estimate

$$\widehat{\sigma}^2_{pavg}(u_o) = \frac{T}{NH}\frac{1}{g_2}\sum_{i=-N}^{N} k\left(\frac{i}{N}\right)\left(\overline{\triangle Y}_{t_{i_o+i},T}\right)^2$$

$$- \frac{T}{2NH}\frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2}\sum_{i=-N}^{N} k\left(\frac{i}{N}\right)\left(Y_{t_{i_o+i},T} - Y_{t_{i_o+i-1},T}\right)^2.$$

This formula has a benefit of having non-random elements in the kernel function, which simplifies the proof. Let

$$\widehat{\sigma}^2_{pre}(u_o) := \frac{T}{N}\frac{1}{g_2}\sum_{j=-M}^{M} k\left(\frac{j}{M}\right)\left(\overline{\triangle Y}_{t_{i_o+jH},T}\right)^2$$

$$- \frac{T}{2NH}\frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2}\sum_{i=-N}^{N} k\left(\frac{i}{N}\right)\left(Y_{t_{i_o+i},T} - Y_{t_{i_o+i-1},T}\right)^2,$$

where $M = M(T) = \lfloor N/H \rfloor$.[5] Similar to the previous theorem, we point out that the difference between these two statistics— $\widehat{\sigma}^2_{pre}(u_o)$ and $\widehat{\sigma}^2_{pavg}(u_o)$—is asymptotically negligible, i.e.

$$\sqrt{N/T^{1/2}}\left\{\widehat{\sigma}^2_{pre}(u_o) - \widehat{\sigma}^2_{pavg}(u_o)\right\} = o_p(1).$$

This means seeking the limit distribution of $\widehat{\sigma}^2_{pre}(u_o)$ is sufficient to infer the limit of $\widehat{\sigma}^2_{pavg}(u_o)$. Likewise, we must show the following statements:

$$(I): \ \sqrt{\frac{N}{H}}\left\{\frac{T}{N}\frac{1}{g_2}\sum_{j=-M}^{M} k\left(\frac{j}{M}\right)\left\{\left(\overline{\triangle X}_{t_{i_o+jH},T}\right)^2 - \mathbb{E}\left[\left(\overline{\triangle X}_{t_{i_o+jH},T}\right)^2 \ \middle| \ N_{\cdot,T}\right]\right\}\right\}$$
$$\xrightarrow{\mathcal{D}} \ \mathcal{N}(0, \ \xi_A^2),$$
$$(II): \ \sqrt{\frac{NH}{T}}\left\{\frac{T}{N}\frac{2}{g_2}\sum_{j=-M}^{M} k\left(\frac{j}{M}\right)\left(\overline{\triangle X}_{t_{i_o+jH},T}\right)\left(\overline{\triangle \varepsilon}_{i_o+jH}\right)\right\} \xrightarrow{\mathcal{D}} \ \mathcal{N}(0, \ \xi_B^2),$$
$$(III): \ \sqrt{\frac{NH^3}{T^2}}\left\{\frac{T}{N}\frac{1}{g_2}\sum_{j=-M}^{M} k\left(\frac{j}{M}\right)\left\{\left(\overline{\triangle \varepsilon}_{i_o+jH}\right)^2 - \mathbb{E}\left[\left(\overline{\triangle \varepsilon}_{i_o+jH}\right)^2\right]\right\}\right\} \xrightarrow{\mathcal{D}} \ \mathcal{N}(0, \ \xi_C^2),$$

and the asymptotic biases

$$(IV): \ \sqrt{\frac{N}{H}}\left\{\frac{T}{N}\frac{1}{g_2}\sum_{j=-M}^{M} k\left(\frac{j}{M}\right)\mathbb{E}\left[\left(\overline{\triangle X}_{t_{i_o+jH},T}\right)^2 \ \middle| \ N_{\cdot,T}\right] - \sigma^2(u_o)\right\} \ = \ o_p(1), \quad \text{and}$$
$$(V): \ \sqrt{\frac{NH^3}{T^2}}\left(\frac{T}{N}\frac{1}{g_2}\sum_{j=-M}^{M} k\left(\frac{j}{M}\right)\mathbb{E}\left[\left(\overline{\triangle \varepsilon}_{i_o+jH}\right)^2\right]\right.$$
$$\left. - \frac{T}{2NH}\frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2}\sum_{i=-N}^{N} k\left(\frac{i}{N}\right)\left(Y_{t_{i_o+i},T} - Y_{t_{i_o+i-1},T}\right)^2\right) \ = \ o_p(1).$$

Although the essential idea for the derivation of $(I)-(III)$ is the same as the preceding theorem, we explicitly demonstrate the first claim in order to see how those arguments can be adapted to this tick-time volatility estimator. Let us denote the left-hand side of (I) by

$$\sqrt{\frac{N}{H}}\left\{\frac{T}{N}\frac{1}{g_2}\sum_{j=-M}^{M} k\left(\frac{j}{M}\right)\left\{\left(\overline{\triangle X}_{t_{i_o+jH,T}}\right)^2 - \mathbb{E}\left[\left(\overline{\triangle X}_{t_{i_o+jH},T}\right)^2 \ \middle| \ N_{\cdot,T}\right]\right\}\right\}$$
$$=: \sum_{j=-M}^{M} A_{j,M}.$$

[5] For simplicity, we might assume that $M = N/H$ is an integer, as the result remains the same in asymptotic.

We see that conditional on $N_{\cdot,T}$, $\{A_{j,M}\}_{j=-M,\dots,M}$ are independent. Hence, it suffices to show that

$$a)\ \sum_{j=-M}^{M}\mathbb{E}\left[A_{j,M}^2\ \middle|\ N_{\cdot,T}\right]\xrightarrow{\mathbb{P}}\xi_A^2\quad\text{and}\quad b)\ \sum_{j=-M}^{M}\mathbb{E}\left[A_{j,M}^4\ \middle|\ N_{\cdot,T}\right]\xrightarrow{\mathbb{P}}0.$$

Due to the calculations (a_1) and (a_2) from the last proof, we have

$$\begin{aligned}
&\sum_{j=-M}^{M}\frac{N}{H}\frac{T^2}{N^2}\frac{1}{g_2^2}\cdot k^2\left(\frac{j}{M}\right)\left\{\mathbb{E}\left[\left(\overline{\triangle X}_{t_{i_o+jH,T}}\right)^4\ \middle|\ N_{\cdot,T}\right]-\left(\mathbb{E}\left[\left(\overline{\triangle X}_{t_{i_o+jH},T}\right)^2\ \middle|\ N_{\cdot,T}\right]\right)^2\right\}\\
&\quad=2\sum_{j=-M}^{M}\frac{N}{H}\frac{T^2}{N^2}\frac{1}{g_2^2}\cdot k^2\left(\frac{j}{M}\right)\sum_{l=1}^{H-1}g^4\left(\frac{l}{H}\right)\sigma^4\left(\frac{t_{i_o+jH+l}}{T}\right)\frac{1}{T^2}\\
&\qquad+2\sum_{j=-M}^{M}\frac{N}{H}\frac{T^2}{N^2}\frac{1}{g_2^2}\cdot k^2\left(\frac{j}{M}\right)\left(\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\sigma^2\left(\frac{t_{i_o+jH+l}}{T}\right)\frac{1}{T}\right)^2\\
&\quad=:(a_1^*)+(a_2^*)\ \xrightarrow{\mathbb{P}}\ 2\sigma^2(u_o)\int_{\mathbb{R}}k^2(x)dx,
\end{aligned}$$

since

$$\begin{aligned}
(a_1^*)&=\frac{2}{HN}\frac{1}{g_2^2}\cdot\sum_{j=-M}^{M}k^2\left(\frac{j}{M}\right)\cdot\sum_{l=1}^{H-1}g^4\left(\frac{l}{H}\right)\left\{\sigma^4\left(\frac{t_o}{T}\right)+\left(\sigma^4\left(\frac{t_{i_o+jH+l}}{T}\right)-\sigma^4\left(\frac{t_o}{T}\right)\right)\right\}\\
&=\frac{M}{N}\frac{2}{g_2^2}\cdot\frac{1}{M}\sum_{j=-M}^{M}k^2\left(\frac{j}{M}\right)\cdot\frac{1}{H}\sum_{l=1}^{H-1}g^4\left(\frac{l}{H}\right)\left\{\sigma^4\left(\frac{t_o}{T}\right)+O\left(\left|\frac{jH}{T}\right|^{\gamma}\right)\right\}\\
&=O\left(\frac{M}{N}\right)\ +\ O\left(\frac{1}{H}\cdot\left|\frac{N}{T}\right|^{\gamma}\right)\ \to\ 0,
\end{aligned}$$

and

$$\begin{aligned}
(a_2^*)&=\frac{2}{NH}\frac{1}{g_2^2}\cdot\sum_{j=-M}^{M}k^2\left(\frac{j}{M}\right)\left(\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\sigma^2(u_o)\right)^2\\
&\quad+\frac{2}{NH}\frac{1}{g_2^2}\cdot\sum_{j=-M}^{M}k^2\left(\frac{j}{M}\right)\left(\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\left\{\sigma^2\left(\frac{t_{i_o+jH+l}}{T}\right)-\sigma^2(u_o)\right\}\right)^2\\
&\quad+\frac{2}{NH}\frac{1}{g_2^2}\cdot\sum_{j=-M}^{M}k^2\left(\frac{j}{M}\right)\left(\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\sigma^2(u_o)\right)\left(\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\left\{\sigma^2\left(\frac{t_{i_o+jH+l}}{T}\right)-\sigma^2(u_o)\right\}\right)\\
&=2\sigma^4(u_o)\int_{\mathbb{R}}k^2(x)dx\ +\ O\left(\left|\frac{N}{T}\right|^{2\gamma}\right)\ +\ O\left(\left|\frac{N}{T}\right|^{\gamma}\right).
\end{aligned}$$

Similar calculations imply that $b)$ is of the order $O(H/N)$, particularly $o(1)$. As was pointed out in the beginning, one of the most difficult parts in this proof is the derivation of the

asymptotic bias (IV), which is not obvious in the tick-time volatility estimation. We have

$$\begin{aligned}
&\sqrt{\frac{N}{H}}\left\{\frac{T}{N}\frac{1}{g_2}\sum_{j=-M}^{M}k\left(\frac{j}{M}\right)\mathbb{E}\left[\left(\overline{\triangle X}_{t_{i_o+jH},T}\right)^2\;\middle|\;N_{\cdot,T}\right]-\sigma^2(u_o)\right\}\\
&\quad=\sqrt{\frac{N}{H}}\frac{T}{N}\frac{1}{g_2}\frac{1}{H}\sum_{j=-N}^{N}\left\{k\left(\frac{\lfloor j/H\rfloor}{M}\right)-k\left(\frac{j}{N}\right)\right\}\mathbb{E}\left[\left(\overline{\triangle X}_{t_{i_o+\lfloor j/H\rfloor\cdot H},T}\right)^2\;\middle|\;N_{\cdot,T}\right]\\
&\qquad+\sqrt{\frac{N}{H}}\left\{\frac{T}{N}\frac{1}{g_2}\frac{1}{H}\sum_{j=-N}^{N}k\left(\frac{j}{N}\right)\mathbb{E}\left[\left(\overline{\triangle X}_{t_{i_o+\lfloor j/H\rfloor\cdot H},T}\right)^2\;\middle|\;N_{\cdot,T}\right]-\sigma^2(u_o)\right\}\\
&\quad=:(i)+(ii),
\end{aligned}$$

where

$$(i)\le C\cdot\sqrt{\frac{N}{H}}\frac{T}{N}\frac{1}{H}\sum_{j=-N}^{N}k^{(1)}(...)\left|\frac{\lfloor j/H\rfloor}{M}-\frac{j}{N}\right|\cdot\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\frac{1}{T}\;=\;O\left(\sqrt{\frac{H}{N}}\right)\;=\;o(1)$$

by the boundedness of derivatives $k'(\cdot)$, and

$$\begin{aligned}
(ii)&=\sqrt{\frac{N}{H}}\left\{\frac{T}{N}\frac{1}{g_2}\frac{1}{H}\sum_{j=-N}^{N}k\left(\frac{j}{N}\right)\mathbb{E}\left[\left(\overline{\triangle X}_{t_{i_o+\lfloor j/H\rfloor\cdot H},T}\right)^2\;\middle|\;N_{\cdot,T}\right]-\sigma^2(u_o)\right\}\\
&=\sqrt{\frac{N}{H}}\cdot\left\{\frac{1}{N}\frac{1}{g_2}\frac{1}{H}\sum_{j=-N}^{N}k\left(\frac{j}{N}\right)\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\sigma^2\left(\frac{t_{i_o+\lfloor j/H\rfloor\cdot H+l}}{T}\right)-\sigma^2(u_o)\right\}\\
&=\sqrt{\frac{N}{H}}\cdot\frac{1}{N}\frac{1}{g_2}\frac{1}{H}\sum_{j=-N}^{N}k\left(\frac{j}{N}\right)\sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\times\\
&\quad\times\left(\left\{\sigma^2\left(\frac{t_{i_o+\lfloor j/H\rfloor\cdot H+l}}{T}\right)-\sigma^2\left(\frac{t_{i_o+j}}{T}\right)\right\}+\left\{\sigma^2\left(\frac{t_{i_o+j}}{T}\right)-\sigma^2(u_o)\right\}\right)\\
&=\;O\left(\sqrt{\frac{N}{H}}\cdot\left|\frac{H}{T}\right|^{\gamma}\right)+\underbrace{\sqrt{\frac{N}{H}}\frac{1}{N}\sum_{j=-N}^{N}k\left(\frac{j}{N}\right)\left\{\sigma^2\left(\frac{t_{i_o+j}}{T}\right)-\sigma^2(u_o)\right\}}_{(\clubsuit)}\\
&=o(1)\;+\;o_p(1),
\end{aligned}$$

as the segment length condition (4.17) holds. To obtain (♣), the same proof for the asymptotic bias in Theorem 4.13 is carried over (with a small change in the rate of convergence). Lastly, the derivation of (V) is straightforward. In summary, from (I) to (V) we can conclude that

$$\sqrt{\frac{N}{T^{1/2}}}\left\{\widehat{\sigma}^2_{pre}(u_o)-\sigma^2(u_o)\right\}\;\xrightarrow{\mathcal{D}}\;\mathcal{N}\left(0,\delta\xi_A^2+\frac{1}{\delta}\xi_B^2+\frac{1}{\delta^3}\xi_C^2\right),$$

if the block size H is chosen to be $\delta \cdot T^{1/2}$, with $\delta \in (0, \infty)$. □

Proof of Theorem 4.17. In the case of $m = m' = 0$, it is necessary to restrict the the bandwidth size $\mathfrak{b}$ and the segment length N according to Theorems 4.7 and 4.16, respectively, i.e.

$$N/T^{1/2} = o\big(T^{\frac{\gamma}{1+2\gamma}}\big) \quad \text{and} \quad \mathfrak{b}T = o\big(T^{\frac{2\gamma'}{1+2\gamma'}}\big)$$

in order to obtain those limit distributions. We analyze it by looking at the following situations:

a) If $\gamma' = \frac{\gamma}{2\gamma+2}$ then $N/T^{1/2} = O(\mathfrak{b}T)$. This gives

$$\begin{aligned}
&\sqrt{\mathfrak{b}T}\left\{\widehat{\sigma}^2_{pavg}(u_o)\widehat{\lambda}(u_o) - \sigma^2(u_o)\lambda(u_o)\right\} \\
&= \underbrace{\widehat{\sigma}^2_{pavg}(u_o)}_{\xrightarrow{\mathbb{P}}\sigma^2(u_o)} \cdot \sqrt{\mathfrak{b}T}\left\{\widehat{\lambda}(u_o) - \lambda(u_o)\right\} \;+\; \lambda(u_o)\underbrace{\frac{\sqrt{\mathfrak{b}T}}{\sqrt{N/T^{1/2}}}}_{=\sqrt{c_1}} \cdot \sqrt{N/T^{1/2}}\left\{\widehat{\sigma}^2_{pavg}(u_o) - \sigma^2(u_o)\right\} \\
&\xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\; \sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}} \mathfrak{K}^2(x)dx \;+\; c_1\lambda^2(u_o)\left\{\delta\xi_A^2 + \frac{1}{\delta}\xi_B^2 + \frac{1}{\delta^3}\xi_C^2\right\}\right),
\end{aligned}$$

since the limits of the first and second terms are independent.

b) If $\gamma' > \frac{\gamma}{2\gamma+2}$ then $N/T^{1/2} = o(\mathfrak{b}T)$. This leads to

$$\begin{aligned}
\sqrt{N/T^{1/2}}\left\{\widehat{\sigma}^2_{pavg}(u_o)\widehat{\lambda}(u_o) - \sigma^2(u_o)\lambda(u_o)\right\} &= \widehat{\lambda}(u_o) \cdot \sqrt{N/T^{1/2}}\left\{\widehat{\sigma}^2_{pavg}(u_o) - \sigma^2(u_o)\right\} \\
&\xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\; \lambda^2(u_o)\left\{\delta\xi_A^2 \;+\; \frac{1}{\delta}\xi_B^2 + \frac{1}{\delta^3}\xi_C^2\right\}\right).
\end{aligned}$$

c) If $\gamma' < \frac{\gamma}{2\gamma+2}$ then $\mathfrak{b}T = o(N/T^{1/2})$, which implies that

$$\begin{aligned}
\sqrt{\mathfrak{b}T}\left\{\widehat{\sigma}^2_{pavg}(u_o)\widehat{\lambda}(u_o) - \sigma^2(u_o)\lambda(u_o)\right\} &= \sigma^2(u_o) \cdot \sqrt{\mathfrak{b}T}\left\{\widehat{\lambda}(u_o) - \lambda(u_o)\right\} \\
&\xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\; \sigma^4(u_o)\lambda(u_o)\int_{\mathbb{R}} \mathfrak{K}^2(x)dx\right).
\end{aligned}$$

Other cases—$m = 0$ together with $m' = 1, 2$, or $m = 1, 2$— can be verified by the same arguments with the corresponding restrictions on $\mathfrak{b}$ and N. □

C. Proofs for Chapter 5

Proof of unbiasedness of $CV_{\widehat{\lambda}}(\mathfrak{b})$. To demonstrate the unbiasedness of $CV_{\widehat{\lambda}}(\mathfrak{b})$ for $ISE_{\widehat{\lambda}}(\mathfrak{b}) - \int_I \lambda(u)^2 du$, we need to show that

$$\mathbb{E}\left[\frac{1}{T}\sum_{u_i \in I} \widehat{\lambda}^{(i)}(u_i) - \int_I \widehat{\lambda}(u)\lambda(u)du\right] \;=\; 0.$$

Without loss of generality, we set $I = [0,1]$. We define $\mathfrak{K}_\mathfrak{b}(x) := \mathfrak{K}(x/\mathfrak{b}T)/\mathfrak{b}$, so the expectation of the first term is given by

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{T}\sum_{i=1}^{N_T}\widehat{\lambda}^{(i)}(u_i)\right] = \mathbb{E}\left[\frac{1}{T}\sum_{i=1}^{N_T}\frac{1}{\mathfrak{b}T}\sum_{j=1,t_j\neq t_i}^{N_T}\mathfrak{K}\left(\frac{t_j-t_i}{\mathfrak{b}T}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{T^2}\sum_{j=1}^{N_T}\sum_{i=1}^{N_T}\frac{1}{\mathfrak{b}}\mathfrak{K}\left(\frac{t_j-t_i}{\mathfrak{b}T}\right)I_{\{t_j>t_i\}}\right] + \mathbb{E}\left[\frac{1}{T^2}\sum_{i=1}^{N_T}\sum_{j=1}^{N_T}\frac{1}{\mathfrak{b}}\mathfrak{K}\left(\frac{t_j-t_i}{\mathfrak{b}T}\right)I_{\{t_j<t_i\}}\right] \\
&= 2\mathbb{E}\left[\frac{1}{T^2}\int_0^T\left(\int_0^T\mathfrak{K}_\mathfrak{b}(s-t)I_{\{s>t\}}dN_{t,T}\right)dN_{s,T}\right] \qquad \text{(since } \mathfrak{K} \text{ is symmetric at 0)} \\
&= 2\mathbb{E}\left[\frac{1}{T^2}\int_0^T\left(\int_0^T\mathfrak{K}_\mathfrak{b}(s-t)I_{\{s>t\}}dN_{t,T}\right)\lambda(s/T)ds\right] \\
&\qquad\qquad \text{(as inner integrals are predictable, see remark (2.2))} \\
&= \frac{2}{T^2}\int_0^T\left(\int_0^T\mathfrak{K}_\mathfrak{b}(s-t)I_{\{s>t\}}\lambda(t/T)dt\right)\lambda(s/T)ds \\
&= \frac{1}{T^2}\int_0^T\int_0^T\mathfrak{K}_\mathfrak{b}(s-t)\lambda(s/T)\lambda(t/T)I_{\{s\neq t\}}\,ds\,dt,
\end{aligned}$$

which is equal to $\mathbb{E}\left[\int_0^1\widehat{\lambda}(u)\lambda(u)du\right]$ (except null sets) since

$$\begin{aligned}
\mathbb{E}\left[\int_0^1\widehat{\lambda}(u)\lambda(u)du\right] &= \int_0^1\mathbb{E}\left[\widehat{\lambda}(u)\right]\lambda(u)du \\
&= \int_0^T\mathbb{E}\left[\int_0^T\mathfrak{K}_b(s-t)dN_{s,T}/T\right]\lambda(t/T)\,dt/T \\
&= \frac{1}{T^2}\int_0^T\int_0^T\mathfrak{K}_b(s-t)\lambda(s/T)\lambda(t/T)\,ds\,dt.
\end{aligned}$$

□

Proof of unbiasedness of $CV_{\widehat{\sigma}^2_{clock,pavg}}(\mathfrak{b})$**.** (The proof of the unbiasedness of $CV_{\widehat{\sigma}^2_{clock,pavg}}(N)$ is analogous, therefore omitted.) Below, all equalities with the sign $(*)$ mean they hold only in asymptotic. First, we want to show that

$$\mathbb{E}\left[\frac{1}{T}\sum_{i=1}^{N_T}\widehat{\sigma}^{2(i)}_{clock,pavg}(u_i)\sigma^2(u_i) - \int_0^1\widehat{\sigma}^2_{clock,pavg}(u)\sigma^2_{clock}(u)\right] \overset{(*)}{=} 0. \qquad \text{(A.7.)}$$

We start by calculating the expectation of the first term

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{T}\sum_{i=1}^{N_T}\widehat{\sigma}^{2(i)}_{clock,pavg}(u_i)\sigma^2(u_i)\right]\\
&\overset{(*)}{=}\mathbb{E}\left[\frac{1}{T}\sum_{i=1}^{N_T}\frac{1}{bH}\frac{1}{g_2}\sum_{j=1}^{N_T}K\left(\frac{t_j-t_i}{bT}\right)\left(\overline{\triangle Y}^{(i)}_{t_j,T}\right)^2 I_{\{t_j\neq t_i\}}\cdot\sigma^2\left(\frac{t_i}{T}\right)\right]\\
&\quad-\ \mathbb{E}\left[\frac{1}{T}\sum_{i=1}^{N_T}\frac{1}{2bH}\frac{\sum h^2(l/H)}{g_2}\sum_{j=1}^{N_T}K\left(\frac{t_j-t_i}{bT}\right)\left(Y_{t_j,T}-Y_{t_{j-1},T}\right)^2 I_{\{t_j\neq t_i\}}\cdot\sigma^2\left(\frac{t_i}{T}\right)\right]\\
&=:(A)-(B),
\end{aligned}$$

where $\overline{\triangle Y}^{(i)}_{t_j,T}$ is the leave-one-out pre-averaging term (i.e. every term with the time index t_i does not appear in the formula). Since

$$\mathbb{E}\left[\left(\overline{\triangle Y}^{(i)}_{t_j,T}\right)^2\ \middle|\ N_{\cdot,T}\right]\ =\ \sum_{l=1}^{H-1}g^2\left(\frac{l}{H}\right)\sigma^2\left(\frac{t_{j+l}}{T}\right)\frac{1}{T}I_{\{t_{j+l}\neq t_i\}}\ +\ \omega^2\sum_{l=1}^{H-1}h^2\left(\frac{l}{H}\right),$$

we have

$$\begin{aligned}
(A)&=\mathbb{E}\left[\frac{1}{T}\sum_{i=1}^{N_T}\frac{1}{bH}\frac{1}{g_2}\sum_{j=1}^{N_T}K\left(\frac{t_j-t_i}{bT}\right)\mathbb{E}\left[\left(\overline{\triangle Y}^{(i)}_{t_j,T}\right)^2\ \middle|\ N_{\cdot,T}\right]I_{\{t_j\neq t_i\}}\cdot\sigma^2\left(\frac{t_i}{T}\right)\right]\\
&\overset{(*)}{=}\mathbb{E}\left[\frac{1}{T}\sum_{i=1}^{N_T}\frac{1}{bH}\frac{\sum g^2(l/H)}{g_2}\sum_{j=1}^{N_T}K\left(\frac{t_j-t_i}{bT}\right)\sigma^2\left(\frac{t_j}{T}\right)\frac{1}{T}I_{\{t_j\neq t_i\}}\cdot\sigma^2\left(\frac{t_i}{T}\right)\right]\\
&\quad+\ \mathbb{E}\left[\frac{1}{T}\sum_{i=1}^{N_T}\frac{1}{bH}\frac{\omega^2\sum h^2(l/H)}{g_2}\sum_{j=1}^{N_T}K\left(\frac{t_j-t_i}{bT}\right)I_{\{t_j\neq t_i\}}\cdot\sigma^2\left(\frac{t_i}{T}\right)\right]\\
&=:(A1)\ +\ (A2).
\end{aligned}$$

Similar to the previous proof, we obtain

$$\begin{aligned}
(A1)&=\mathbb{E}\left[\frac{1}{bHT^2}\frac{\sum g^2(l/H)}{g_2}\sum_{i=1}^{N_T}\sum_{j=1}^{N_T}K\left(\frac{t_j-t_i}{bT}\right)\sigma^2\left(\frac{t_j}{T}\right)I_{\{t_j>t_i\}}\cdot\sigma^2\left(\frac{t_i}{T}\right)\right]\\
&\quad+\ \mathbb{E}\left[\frac{1}{bHT^2}\frac{\sum g^2(l/H)}{g_2}\sum_{i=1}^{N_T}\sum_{j=1}^{N_T}K\left(\frac{t_j-t_i}{bT}\right)\sigma^2\left(\frac{t_j}{T}\right)I_{\{t_j<t_i\}}\cdot\sigma^2\left(\frac{t_i}{T}\right)\right]\\
&=\frac{1}{bHT^2}\frac{\sum g^2(l/H)}{g_2}\int_0^T\int_0^T K\left(\frac{s-t}{bT}\right)\sigma^2\left(\frac{s}{T}\right)\sigma^2\left(\frac{t}{T}\right)\lambda\left(\frac{s}{T}\right)\lambda\left(\frac{t}{T}\right)I_{\{s\neq t\}}\,ds\,dt,
\end{aligned}$$

and

$$(A2)=\frac{1}{bHT}\frac{\omega^2\sum h^2(l/H)}{g_2}\int_0^T\int_0^T K\left(\frac{s-t}{bT}\right)\lambda\left(\frac{s}{T}\right)\lambda\left(\frac{t}{T}\right)I_{\{s\neq t\}}\,ds\,dt.$$

Further,

$$
\begin{aligned}
(B) &= \mathbb{E}\left[\sum_{i=1}^{N_T}\frac{1}{2bHT}\frac{\sum h^2(l/H)}{g_2}\sum_{j=1}^{N_T}K\left(\frac{t_j-t_i}{bT}\right)\mathbb{E}\left[\left(Y_{t_j,T}-Y_{t_{j-1},T}\right)^2 \;\middle|\; N_{\cdot,T}\right] I_{\{t_j\neq t_i\}}\cdot\sigma^2\left(\frac{t_i}{T}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{2bHT}\frac{\sum h^2(l/H)}{g_2}\sum_{i=1}^{N_T}\sum_{j=1}^{N_T}K\left(\frac{t_j-t_i}{bT}\right)\sigma^2\left(\frac{t_j}{T}\right)\frac{1}{T}I_{\{t_j\neq t_i\}}\cdot\sigma^2\left(\frac{t_i}{T}\right)\right] \\
&\quad + \mathbb{E}\left[\frac{1}{2bHT}\frac{\sum h^2(l/H)}{g_2}\sum_{i=1}^{N_T}\sum_{j=1}^{N_T}K\left(\frac{t_j-t_i}{bT}\right)2\omega^2 I_{\{t_j\neq t_i\}}\cdot\sigma^2\left(\frac{t_i}{T}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{2bHT^2}\frac{\sum h^2(l/H)}{g_2}\int_0^T\int_0^T K\left(\frac{s-t}{bT}\right)\sigma^2\left(\frac{s}{T}\right)\sigma^2\left(\frac{t}{T}\right)\lambda\left(\frac{s}{T}\right)\lambda\left(\frac{t}{T}\right)I_{\{s\neq t\}}\,ds\,dt\right] \\
&\quad + \mathbb{E}\left[\frac{\omega^2}{bHT}\frac{\sum h^2(l/H)}{g_2}\int_0^T\int_0^T K\left(\frac{s-t}{bT}\right)\sigma^2\left(\frac{t}{T}\right)\lambda\left(\frac{s}{T}\right)\lambda\left(\frac{t}{T}\right)I_{\{s\neq t\}}\,ds\,dt\right].
\end{aligned}
$$

Likewise, we can derive the second part

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^1\widehat{\sigma}^2_{clock,pavg}(u)\sigma^2(u)\lambda(u)du\right] \\
&= \mathbb{E}\left[\int_0^1\frac{1}{bH}\frac{1}{g_2}\sum_{j=1}^{N_T}K\left(\frac{t_j-t}{bT}\right)\left(\overline{\triangle Y}_{t_j,T}\right)^2\sigma^2\left(\frac{t}{T}\right)\lambda\left(\frac{t}{T}\right)\frac{dt}{T}\right] \\
&\quad - \mathbb{E}\left[\int_0^1\frac{1}{2bH}\frac{\sum h^2(l/H)}{g_2}\sum_{j=1}^{N_T}K\left(\frac{t_j-t}{bT}\right)\left(Y_{t_j,T}-Y_{t_{j-1},T}\right)^2\sigma^2\left(\frac{t}{T}\right)\lambda\left(\frac{t}{T}\right)\frac{dt}{T}\right] \\
&=: (C)-(D).
\end{aligned}
$$

By direct calculations we can show that (C) and (D) are equal to (A) and (B), respectively (except null sets), and therefore we have derived the equality (A.7.). Finally, we can estimate $\sigma^2(u_i)$ in (A.7.) by another leave-one-out estimate $\widehat{\sigma}^{2(i)}_{pavg}(u_i)$ with pre-choices of the segment length N and block size H. We conclude that, $\frac{1}{T}\sum_{i=1}^{N_T}\widehat{\sigma}^{2(i)}_{clock,pavg}(u_i)\widehat{\sigma}^{2(i)}_{pavg}(u_i)$ is an asymptotic unbiased estimator for $\int_0^1\widehat{\sigma}^2_{clock,pavg}(u)\sigma^2_{clock}(u)du$. □

Proof of limit distribution in (5.9)**.** The verification of the asymptotic normality of $\widehat{\lambda}(u_o)$ in the stochastic intensity model runs as before; see section 4.2. The intensity estimator is given by

$$
\begin{aligned}
\widehat{\lambda}(u_o) &= \frac{1}{\mathfrak{b}T}\sum_{i=1}^{N_T}\mathfrak{K}\left(\frac{t_i-u_oT}{\mathfrak{b}T}\right) \\
&= \frac{1}{\mathfrak{b}T}\int_{u_oT-\mathfrak{b}T}^{u_oT+\mathfrak{b}T}\mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right)dN_{t,T},
\end{aligned}
$$

since $\mathfrak{K}(x)$ assigns zero for x outside $[-1,1]$ (thus, only the data inside $[u_oT-\mathfrak{b}T, u_oT+\mathfrak{b}T]$ contributes to the estimate). We divide the interval $(u_oT-\mathfrak{b}T, u_oT+\mathfrak{b}T]$ into M_T subintervals of fixed length $\triangle$, and we assume that $\lfloor 2\mathfrak{b}T/\triangle \rfloor = M_T$ being an integer. By setting $\triangle_{i,T} := u_oT - \mathfrak{b}T + i\triangle$, we show that

$$\sqrt{\mathfrak{b}T}\left(\widehat{\lambda}(u_o) - \frac{1}{\mathfrak{b}T}\int_{u_oT-\mathfrak{b}T}^{u_oT+\mathfrak{b}T} \mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right)\lambda\left(\frac{t}{T}\right)\, dt\right) = \sum_{i=1}^{M_T}\int_{\triangle_{i-1,T}}^{\triangle_{i,T}} \frac{1}{\sqrt{\mathfrak{b}T}}\mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right) dM_{t,T}$$
$$=: \sum_{i=1}^{M_T} A_{i,T}$$

converges in distribution to a normal random variable. Recall that $M_{t,T} := N_{t,T} - \int_0^t \lambda(t/T)dt$ is a squared integrable martingale w.r.t. its natural filtration. Let a discrete filtration $\mathcal{G}_{i,T} := \sigma\left(N_{s,T}, \lambda_{s,T} : s \leq \triangle_{i,T}\right)$, in which $\mathcal{G}_{i,T}$ contains all information about the point process up to the time $u_oT - \mathfrak{b}T + i\triangle$. In particular

$$\lambda_{t_o-\mathfrak{b}T,T} = \lambda\left(\frac{t_o-\mathfrak{b}T}{T}\right) = \lambda(u_o-\mathfrak{b}) \text{ is } \mathcal{G}_{0,T}\text{-measurable}$$

(which plays an important role later). We see that the triangular array $A_{i,T}$ is a martingale difference sequence w.r.t. $\mathcal{G}_{i,T}$, satisfying conditions given in the stable version of central limit theorem for martingale difference sequences in Hall and Heyde [38, Corollary 3.1], i.e.

(i) $\mathbb{E}\left[A_{i,T} \;\middle|\; \mathcal{G}_{i-1,T}\right] = 0,$

(ii) $\sum_{i=1}^{M_T}\mathbb{E}\left[A_{i,T}^2 \;\middle|\; \mathcal{G}_{i-1,T}\right] \xrightarrow{\mathbb{P}} \eta^2$, where η^2 is $\mathcal{G}_{0,T}$-measurable, therefore nested in every σ-algebra $\mathcal{G}_{i,T}$, and

(iii) $\sum_{i=1}^{M_T}\mathbb{E}\left[A_{i,T}^4 \;\middle|\; \mathcal{G}_{i-1,T}\right] \xrightarrow{\mathbb{P}} 0.$

Note that the nesting condition is substituted by the measurability of η^2 such that the theorem also holds, however, without stability. In fact, we will show that

$$\frac{1}{\eta}\sum_{i=1}^{M_T} A_{i,T} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

By construction, it is clear that (i) is fulfilled. To show (ii), we see

$$\sum_{i=1}^{M_T}\mathbb{E}\left[A_{i,T}^2 \;\middle|\; \mathcal{G}_{i-1,T}\right] = \sum_{i=1}^{M_T}\mathbb{E}\left[\left(\int_{\triangle_{i-1,T}}^{\triangle_{i,T}} \frac{1}{\sqrt{\mathfrak{b}T}}\mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right) dM_{t,T}\right)^2 \;\middle|\; \mathcal{G}_{i-1,T}\right]$$

$$
\begin{aligned}
&= \sum_{i=1}^{M_T} \mathbb{E}\left[\int_{\triangle_{i-1,T}}^{\triangle_{i,T}} \frac{1}{\mathfrak{b}T} \mathfrak{K}^2\left(\frac{t-u_oT}{\mathfrak{b}T}\right) \lambda\left(\frac{t}{T}\right) dt \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&= \lambda(u_o-\mathfrak{b}) \sum_{i=1}^{M_T} \mathbb{E}\left[\int_{\triangle_{i-1,T}}^{\triangle_{i,T}} \frac{1}{\mathfrak{b}T} \mathfrak{K}^2\left(\frac{t-u_oT}{\mathfrak{b}T}\right) dt \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&\quad + \sum_{i=1}^{M_T} \mathbb{E}\left[\int_{\triangle_{i-1,T}}^{\triangle_{i,T}} \frac{1}{\mathfrak{b}T} \mathfrak{K}^2\left(\frac{t-u_oT}{\mathfrak{b}T}\right) \left\{ \lambda\left(\frac{t}{T}\right) - \lambda(u_o-\mathfrak{b}) \right\} dt \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&= \lambda(u_o-\mathfrak{b}) \int \mathfrak{K}^2(x)dx \;+\; O(\mathfrak{b}^{\gamma'}),
\end{aligned}
$$

by the conditional isometry property. The statement (iii) is provided by

$$
\begin{aligned}
\sum_{i=1}^{M_T} \mathbb{E}\left[A_{i,T}^4 \;\middle|\; \mathcal{G}_{i-1,T} \right] &= \sum_{i=1}^{M_T} \mathbb{E}\left[\left(\int_{\triangle_{i-1,T}}^{\triangle_{i,T}} \frac{1}{\sqrt{\mathfrak{b}T}} \mathfrak{K}\left(\frac{t-u_oT}{\mathfrak{b}T}\right) dM_{t,T} \right)^4 \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&\le C \cdot \sum_{i=1}^{M_T} \mathbb{E}\left[\left(\int_{\triangle_{i-1,T}}^{\triangle_{i,T}} \frac{1}{\mathfrak{b}T} \mathfrak{K}^2\left(\frac{t-u_oT}{\mathfrak{b}T}\right) \lambda\left(\frac{t}{T}\right) dt \right)^2 \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&\le C \cdot \frac{1}{\mathfrak{b}^2T^2} \sum_{i=1}^{M_T} \triangle^2 = O\left(\frac{1}{\mathfrak{b}T}\right),
\end{aligned}
$$

by Burkholder-Davis-Gundy's inequality (see Jacod and Protter [48, p.39]). We remark that the derivation of the asymptotic bias is the same as in Theorem 4.7, therefore omitted. Thus, we can conclude that the normal limit distribution (5.9) holds. □

Discussion on asymptotic normality of volatility estimators

As was mentioned, it is much more difficult to prove central limit theorems for tick-time/clock-time volatility estimators in the general setting of stochastic volatility than to prove those for deterministic one. In the following, we will give only the outline of the proof of the limit distribution for $\widehat{\sigma}^2_{clock}(\cdot)$, and emphasize that the asymptotic variance is not the same as in the exogenous case, in that an additional variance term will appear in this general case (see the first term of (A.9.)). Similar to the last proof, we rewrite

$$
\begin{aligned}
\widehat{\sigma}^2_{clock}(u_o) &= \sum_{i=1}^{M_T} \frac{1}{b} \sum_{\triangle_{i-1,T} < t_j \le \triangle_{i,T}} K\left(\frac{t_j-u_oT}{bT}\right) \left(X_{t_j,T} - X_{t_{j-1},T}\right)^2 \\
&=: \sum_{i=1}^{M_T} \spadesuit_{i,T}.
\end{aligned}
$$

First, we want to show that

$$
\sqrt{bT} \left\{ \sum_{i=1}^{M_T} \spadesuit_{i,T} - \mathbb{E}\left[\spadesuit_{i,T} \;\middle|\; \mathcal{G}_{i-1,T} \right] \right\} \;=:\; \sum_{i=1}^{M_T} B_{i,T} \;\xrightarrow{\mathcal{D}}\; \mathcal{N}(0,1),
$$

where now $\mathcal{G}_{i,T}$ contains, roughly speaking, all information up to time $u_oT - bT + i\triangle$. Due to the central limit theorem given by Hall and Heyde [38, Corollary 3.1], it is sufficient to show that

i) $B_{i,T} \in \mathcal{G}_{i-1,T}$,

ii) $\mathbb{E}\left[B_{i,T} \;\middle|\; \mathcal{G}_{i-1,T}\right] = 0$,

iii) $\sum_{i=1}^{M_T} \mathbb{E}\left[B_{i,T}^2 \;\middle|\; \mathcal{G}_{i-1,T}\right] \xrightarrow{\mathbb{P}} \kappa^2$, where κ^2 is $\mathcal{G}_{0,T}$-measurable, and

iv) $\sum_{i=1}^{M_T} \mathbb{E}\left[B_{i,T}^4 \;\middle|\; \mathcal{G}_{i-1,T}\right] \xrightarrow{\mathbb{P}} 0$.

By construction, it is obvious that i) and ii) hold. It is noteworthy that the distinction between the asymptotic variances of endogenous and exogenous cases lies in the conditional variance term iii), which will be displayed below. In what follows, $(*)$ means that the equality holds in asymptotic sense. Similarly, we adopt the notations (corresponding to each interval that is being considered):

$$t_{i_l}^{(k)} = \inf\{t_j : t_j > \triangle_{k-1,T}\} \qquad \text{and} \qquad t_{i_u}^{(k)} := \sup\{t_j : t_j \leq \triangle_{k,T}\}.$$

We will write shortly t_{i_l} and t_{i_u} for $t_{i_l}^{(k)}$ and $t_{i_u}^{(k)}$ if it is clear from the context, on which interval they are considered. We begin by looking at

$$\begin{aligned}
\mathbb{E}\left[\spadesuit_{i,T} \;\middle|\; \mathcal{G}_{i-1,T}\right] &= \frac{1}{b}\mathbb{E}\left[\sum_{\triangle_{i-1,T}<t_j\leq\triangle_{i,T}} K\left(\frac{t_j - u_oT}{bT}\right)\left(X_{t_j,T} - X_{t_{j-1},T}\right)^2 \;\middle|\; \mathcal{G}_{i-1,T}\right] \\
&= \frac{1}{b}\mathbb{E}\left[\sum_{\triangle_{i-1,T}<t_j<t_{i_u}} K\left(\frac{t_j - u_oT}{bT}\right)\left(X_{t_j,T} - X_{t_{j-1},T}\right)^2 \;\middle|\; \mathcal{G}_{i-1,T}\right] \\
&\quad + \frac{1}{b}\mathbb{E}\left[\mathbb{E}\left[K\left(\frac{t_{i_u} - u_oT}{bT}\right)\left(X_{t_{i_u},T} - X_{t_{i_u-1},T}\right)^2 \;\middle|\; \mathcal{F}_{t_{i_u-},T}\right] \;\middle|\; \mathcal{G}_{i-1,T}\right] \\
&= \frac{1}{b}\mathbb{E}\left[\sum_{\triangle_{i-1,T}<t_j<t_{i_u}} K\left(\frac{t_j - u_oT}{bT}\right)\left(X_{t_j,T} - X_{t_{j-1},T}\right)^2 \;\middle|\; \mathcal{G}_{i-1,T}\right] \\
&\quad + \frac{1}{b}\mathbb{E}\left[K\left(\frac{t_{i_u} - u_oT}{bT}\right)\sigma^2\left(\frac{t_{i_u}}{T}\right)\frac{1}{T} \;\middle|\; \mathcal{G}_{i-1,T}\right] \\
&= \cdots \\
&= \frac{1}{bT}\mathbb{E}\left[\sum_{\triangle_{i-1,T}<t_j\leq\triangle_{i,T}} K\left(\frac{t_j - u_oT}{bT}\right)\sigma^2\left(\frac{t_j}{T}\right) \;\middle|\; \mathcal{G}_{i-1,T}\right] \\
&= \frac{1}{bT}\mathbb{E}\left[\int_{\triangle_{i-1,T}}^{\triangle_{i,T}} K\left(\frac{t - u_oT}{bT}\right)\sigma^2\left(\frac{t}{T}\right)\lambda\left(\frac{t}{T}\right)dt \;\middle|\; \mathcal{G}_{i-1,T}\right].
\end{aligned}$$

It is shown that

$$bT \cdot \sum_{i=1}^{M_T} \left(\mathbb{E}\left[\spadesuit_{i,T} \;\middle|\; \mathcal{G}_{i-1,T} \right] \right)^2 \overset{(*)}{=} \sigma^4 (u_o - b) \lambda^2 (u_o - b) \frac{1}{bT} \sum_{i=1}^{M_T} \left(\int_{\triangle_{i-1,T}}^{\triangle_{i,T}} K\left(\frac{t - u_o T}{bT} \right) dt \right)^2 \tag{A.8.}$$

Next,

$$\begin{aligned}
& bT \sum_{i=1}^{M_T} \mathbb{E}\left[\spadesuit_{i,T}^2 \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&= \frac{bT}{b^2} \sum_{i=1}^{M_T} \mathbb{E}\left[\left(\sum_{\triangle_{i-1,T} < t_j \le \triangle_{i,T}} K\left(\frac{t_j - u_o T}{bT} \right) \left(X_{t_j,T} - X_{t_{j-1},T} \right)^2 \right)^2 \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&= \frac{bT}{b^2} \sum_{i=1}^{M_T} \mathbb{E}\left[\left(\sum_{\triangle_{i-1,T} < t_j < t_{i_u}} K\left(\frac{t_j - u_o T}{bT} \right) \left(X_{t_j,T} - X_{t_{j-1},T} \right)^2 \right)^2 \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&\quad + \frac{bT}{b^2} \sum_{i=1}^{M_T} \mathbb{E}\left[2(\cdots) \left(K\left(\frac{t_{i_u} - u_o T}{bT} \right) \left(X_{t_{i_u},T} - X_{t_{i_u-1},T} \right)^2 \right) \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&\quad + \frac{bT}{b^2} \sum_{i=1}^{M_T} \mathbb{E}\left[\left(K\left(\frac{t_{i_u} - u_o T}{bT} \right) \left(X_{t_{i_u},T} - X_{t_{i_u-1},T} \right)^2 \right)^2 \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&=: \; (I) + (I_1) + (I_2),
\end{aligned}$$

where

$$\begin{aligned}
(I_1) &= \frac{2bT}{b^2} \sum_{i=1}^{M_T} \mathbb{E}\left[(\cdots)\, \mathbb{E}\left[K\left(\frac{t_{i_u} - u_o T}{bT} \right) \left(X_{t_{i_u},T} - X_{t_{i_u-1},T} \right)^2 \;\middle|\; \mathcal{F}_{t_{i_u}-,T} \right] \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&= \frac{2bT}{b^2} \sum_{i=1}^{M_T} \mathbb{E}\left[(\cdots) \left(K\left(\frac{t_{i_u} - u_o T}{bT} \right) \sigma^2\left(\frac{t_{i_u}}{T} \right) \frac{1}{T} \right) \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&\overset{(*)}{=} \frac{2bT}{b^2} \sum_{i=1}^{M_T} \frac{\sigma^2(u_o - b)}{T} \mathbb{E}\left[\sum_{\triangle_{i-1,T} < t_j < t_{i_u}} K^2\left(\frac{t_j - u_o T}{bT} \right) \left(X_{t_j,T} - X_{t_{j-1},T} \right)^2 \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&= \cdots \\
&\overset{(*)}{=} \frac{2bT}{b^2} \sum_{i=1}^{M_T} \frac{\sigma^2(u_o - b)}{T} \frac{\sigma^2(u_o - b)\lambda(u_o - b)}{T} \mathbb{E}\left[\int_{\triangle_{i-1,T}}^{t_{i_u}} K^2\left(\frac{t - u_o T}{bT} \right) dt \;\middle|\; \mathcal{G}_{i-1,T} \right] \\
&= \frac{2\sigma^4(u_o - b)\lambda(u_o - b)}{bT} \sum_{i=1}^{M_T} \mathbb{E}\left[\int_{\triangle_{i-1,T}}^{t_{i_u}} K^2\left(\frac{t - u_o T}{bT} \right) dt \;\middle|\; \mathcal{G}_{i-1,T} \right]
\end{aligned}$$

and

$$(I_2) = \frac{bT}{b^2}\sum_{i=1}^{M_T}\mathbb{E}\left[\left(K\left(\frac{t_{i_u}-u_oT}{bT}\right)\left(X_{t_{i_u},T}-X_{t_{i_u-1},T}\right)^2\right)^2 \;\Big|\; \mathcal{G}_{i-1,T}\right]$$
$$= \frac{bT}{b^2}\sum_{i=1}^{M_T}\mathbb{E}\left[K^2\left(\frac{t_{i_u}-u_oT}{bT}\right)\sigma^4\left(\frac{t_{i_u}}{T}\right)\frac{1}{T^2}\mu_4 \;\Big|\; \mathcal{G}_{i-1,T}\right],$$

where we have assumed that $\mathbb{E}\left[U_i^4 \;\Big|\; \mathcal{F}_{t_i-,T}\right] = \mu_4 < \infty$. The first term (I) can be split up again into the next step that $(I) = (II) + (II_1) + (II_2)$, and repeatedly $(II) = (III) + (III_1) + (III_2)$ and so on. Thus

$$\begin{aligned}
&bT\sum_{i=1}^{M_T}\mathbb{E}\left[\spadesuit_{i,T}^2 \;\Big|\; \mathcal{G}_{i-1,T}\right]\\
&= \frac{2\sigma^4(u_o-b)\lambda(u_o-b)}{bT}\sum_{i=1}^{M_T}\mathbb{E}\left[\int_{\triangle_{i-1,T}}^{t_{i_u}}K^2\left(\frac{t-u_oT}{bT}\right)dt+\cdots+\int_{\triangle_{i-1,T}}^{t_{i_l}}K^2\left(\frac{t-u_oT}{bT}\right)dt \;\Big|\; \mathcal{G}_{i-1,T}\right]\\
&\quad+\ \frac{\mu_4}{bT}\sum_{i=1}^{M_T}\mathbb{E}\left[K^2\left(\frac{t_{i_u}-u_oT}{bT}\right)\sigma^4\left(\frac{t_{i_u}}{T}\right)+\cdots+K^2\left(\frac{t_{i_l}-u_oT}{bT}\right)\sigma^4\left(\frac{t_{i_l}}{T}\right) \;\Big|\; \mathcal{G}_{i-1,T}\right]\\
&= \frac{2\sigma^4(u_o-b)\lambda(u_o-b)}{bT}\sum_{i=1}^{M_T}\mathbb{E}\left[\int_{\triangle_{i-1,T}}^{\triangle_{i,T}}\int_{\triangle_{i-1,T}}^{t'}K^2\left(\frac{t-u_oT}{bT}\right)dt\ dN_{t',T} \;\Big|\; \mathcal{G}_{i-1,T}\right]\\
&\quad+\ \frac{\mu_4}{bT}\sum_{i=1}^{M_T}\mathbb{E}\left[\int_{\triangle_{i-1,T}}^{\triangle_{i,T}}K^2\left(\frac{t-u_oT}{bT}\right)\sigma^4\left(\frac{t}{T}\right)dN_{t,T} \;\Big|\; \mathcal{G}_{i-1,T}\right]\\
&= \frac{2\sigma^4(u_o-b)\lambda(u_o-b)}{bT}\sum_{i=1}^{M_T}\mathbb{E}\left[\int_{\triangle_{i-1,T}}^{\triangle_{i,T}}\int_{\triangle_{i-1,T}}^{t'}K^2\left(\frac{t-u_oT}{bT}\right)\lambda\left(\frac{t'}{T}\right)dt\ dt' \;\Big|\; \mathcal{G}_{i-1,T}\right]\\
&\quad+\ \frac{\mu_4}{bT}\sum_{i=1}^{M_T}\mathbb{E}\left[\int_{\triangle_{i-1,T}}^{\triangle_{i,T}}K^2\left(\frac{t-u_oT}{bT}\right)\sigma^4\left(\frac{t}{T}\right)\lambda\left(\frac{t}{T}\right)dt \;\Big|\; \mathcal{G}_{i-1,T}\right]\\
&\overset{(*)}{=} 2\sigma^4(u_o-b)\lambda^2(u_o-b)\cdot\frac{1}{bT}\sum_{i=1}^{M_T}\int_{\triangle_{i-1,T}}^{\triangle_{i,T}}\int_{\triangle_{i-1,T}}^{t'}K^2\left(\frac{t-u_oT}{bT}\right)dt\ dt'\\
&\quad+\mu_4\sigma^4(u_o-b)\lambda(u_o-b)\cdot\frac{1}{bT}\sum_{i=1}^{M_T}\int_{\triangle_{i-1,T}}^{\triangle_{i,T}}K^2\left(\frac{t-u_oT}{bT}\right)dt. \qquad \text{(A.9.)}
\end{aligned}$$

With (A.8.) and (A.9.), we can proceed with the condition iii), since

$$\sum_{i=1}^{M_T}\mathbb{E}\left[B_{i,T}^2 \;\Big|\; \mathcal{G}_{i-1,T}\right] = bT\cdot\sum_{i=1}^{M_T}\mathbb{E}\left[\spadesuit_{i,T}^2 \;\Big|\; \mathcal{G}_{i-1,T}\right] - \left(\mathbb{E}\left[\spadesuit_{i,T} \;\Big|\; \mathcal{G}_{i-1,T}\right]\right)^2.$$

□

Bibliography

[1] Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6, 701-726.

[2] Aït-Sahalia, Y., and Jacod, J. (2014). *High-frequency financial econometrics*. Princeton University Press.

[3] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1), 43-76.

[4] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). Modeling and forecasting realized volatility (No. w8160). *National Bureau of Economic Research.*

[5] Ané, T., and Geman, H. (2000). Order flow, transaction clock, and normality of asset returns. *The Journal of Finance*, 55(5), 2259-2284.

[6] Bandi, F. M., and Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies*, 75(2), 339-369.

[7] Barndorff-Nielsen, O. E., and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 253-280.

[8] Barndorff-Nielsen, O. E., and Shephard, N. (2003). Realized power variation and stochastic volatility models. *Bernoulli*, 9(2), 243-265.

[9] Barndorff-Nielsen, O. E., and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1), 1-37.

[10] Barndorff-Nielsen, O. E., and Shephard, N. (2006). Power variation and time change. *Theory of Probability and Its Applications*, 50(1), 1-15.

[11] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6), 1481-1536.

[12] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, 12(3), C1-C32.

[13] Barndorff-Nielsen, O. E., Shephard, N., and Winkel, M. (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Processes and their Applications*, 116(5), 796-806.

[14] Belomestny, D. (2011). Statistical inference for time-changed Lévy processes via composite characteristic function estimation. *The Annals of Statistics*, 2205-2242.

[15] Bibinger, M., and Reiß, M. (2014). Spectral estimation of covolatility from noisy observations using local weights. *Scandinavian Journal of Statistics*, 41(1), 23-50.

[16] Brémaud, P. (1981). *Point processes and queues*. New York: Springer.

[17] Brockwell, P., Chadraa, E., and Lindner, A. (2006). Continuous-time GARCH processes. *The Annals of Applied Probability*, 16(2), 790-826.

[18] Brooks, M. M., and Marron, J. S. (1991). Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Processes and their Application*s, 38(1), 157-165.

[19] Carr, P., Geman, H., Madan, D. B., and Yor, M. (2003). Stochastic volatility for Lévy processes. *Mathematical Finance*, 13(3), 345-382.

[20] Carr, P., and Wu, L. (2004). Time-changed Lévy processes and option pricing. *Journal of Financial economics*, 71(1), 113-141.

[21] Cheng, M. Y. (1997). A bandwidth selector for local linear density estimators. *The Annals of Statistics*, 25(3), 1001-1013.

[22] Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: Journal of the Econometric Society*, 135-155.

[23] Cont, R., and Tankov, P. (2004). *Financial modelling with jump processes*. Chapman & Hall/CRC.

[24] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25(1), 1-37.

[25] Dahlhaus, R., and Neddermeyer, J. C. (2013). Online Spot Volatility-Estimation and Decomposition with Nonlinear Market Microstructure Noise Models. *Journal of Financial Econometrics*, 12(1), 174-212.

[26] Delattre, S., and Jacod, J. (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli*, 1-28.

[27] Delbaen, F., and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300(1), 463-520.

[28] Durrett, R. (2010). *Probability: theory and examples*. Cambridge university press.

[29] Fan, J., Gasser, T., Gijbels, I., Brockmann, M., and Engel, J. (1993). *Local polynomial fitting: a standard for nonparametric regression.*

[30] Fan, J., and Gijbels, I. (1996). *Local polynomial modelling and its applications.* Chapman & Hall/CRC.

[31] Fan, J., Jiang, J., Zhang, C., and Zhou, Z. (2003). Time-dependent diffusion models for term structure dynamics. *Statistica Sinica*, 13(4), 965-992.

[32] Fan, J., and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, 102(480), 1349-1362.

[33] Fan, J., and Wang, Y. (2008). Spot volatility estimation for high-frequency data. *Statistics and its Interface*, 1, 279-288.

[34] Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H.E. (2003) A theory of power-law distri- butions in financial market fluctuations. *Nature*, 423, 267-270.

[35] Geman, H., Madan, D. B., and Yor, M. (2001). Time changes for Lévy processes. *Mathematical Finance*, 11(1), 79-96.

[36] Gloter, A., and Jacod, J. (2001). Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM: Probability and Statistics*, 5, 225-242.

[37] Griffin, J. E., and Oomen, R. C. (2008). Sampling returns for realized variance calculations: tick time or transaction time?. *Econometric Reviews*, 27(1-3), 230-253.

[38] Hall, P., and Heyde, C. C. (1980). *Martingale limit theory and its application.* New York: Academic press.

[39] Hansen, P. R., and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2), 127-161.

[40] Härdle, W. (1991). *Smoothing Techniques: With Implementation in S.* Springer Science and Business Media.

[41] Härdle, W., Hall, P., and Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum?. *Journal of the American Statistical Association*, 83(401), 86-95.

[42] Härdle, W., and Linton, O. (1994). Applied nonparametric methods. *Handbook of econometrics*, 4, 2295-2339.

[43] Hart, J. D. (1996). Some automated methods of smoothing time-dependent data. *Journal of Nonparametric Statistics*, 6(2-3), 115-142.

[44] Hasbrouck, J. (2006). *Empirical market microstructure: The institutions, economics, and econometrics of securities trading.* Oxford University Press.

[45] Hull, J., and White, A. (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2), 281-300.

[46] Hurd, T. R. (2009). Credit risk modeling using time-changed Brownian motion. *International journal of theoretical and applied finance*, 12(08), 1213-1230.

[47] Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and their Applications*, 119(7), 2249-2276.

[48] Jacod, J., and Protter, P. E. (2012). *Discretization of processes*. Berlin Heidelberg: Springer-Verlag.

[49] Jacod, J., and Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*. Berlin: Springer-Verlag.

[50] Jones, C. M., Kaul, G., and Lipson, M. L. (1994). Transactions, volume, and volatility. *Review of Financial Studies*, 7(4), 631-651.

[51] Kalnina, I. (2011). Subsampling high frequency data. *Journal of Econometrics*, 161(2), 262-283.

[52] Kalnina, I., and Linton, O. (2008). Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *Journal of Econometrics*, 147(1), 47-59.

[53] Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: a kernel-based approach. *Econometric Theory*, 26(01), 60-93.

[54] Kuo, H.-H. (2006). *Introduction to Stochastic Integration*. Springer.

[55] Malliavin, P., and Mancino, M. E. (2009). A Fourier transform method for nonparametric estimation of multivariate volatility. *The Annals of Statistics*, 37(4), 1983-2010.

[56] Mancino, M. E., and Sanfelici, S. (2008). Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. *Computational Statistics & data analysis*, 52(6), 2966-2989.

[57] Monroe, I. (1978). Processes that can be embedded in Brownian motion. *The Annals of Probability*, 42-56.

[58] Munk, A., and Schmidt-Hieber, J. (2010). Lower bounds for volatility estimation in microstructure noise models. *In Borrowing Strength: Theory Powering Applications -A Festschrift for Lawrence D. Brown (pp. 43-55)*. Institute of Mathematical Statistics.

[59] Plerou, V., Gopikrishnan, P., Gabaix, X., A Nunes Amaral, L., and Stanley, H.E. (2001) Price fluctuations, market activity and trading volume. *Quantitative Finance*, 1, 262-269.

[60] Podolskij, M., and Vetter, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3), 634-658.

[61] Podolskij, M., and Vetter, M. (2010). Understanding limit theorems for semimartingales: a short survey. *Statistica Neerlandica*, 64(3), 329-351.

[62] Protter, P. E. (2013). *Stochastic integration and differential equations*. Vol. 21. Springer.

[63] Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics*, 453-466.

[64] Reiß, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *The Annals of Statistics*, 39(2), 772-802.

[65] Robinson, P. M. (1989). *Nonparametric estimation of time-varying parameters* (pp. 253-264). Berlin Heidelberg: Springer.

[66] Rosenbaum, M. (2009). Integrated volatility and round-off error. *Bernoulli*, 15(3), 687-720.

[67] Schmidt-Hieber, A. J. (2011). *Nonparametric methods in spot volatility estimation* (Doctoral dissertation, Niedersächsische Staats-und Universitätsbibliothek Göttingen).

[68] Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). CRC press.

[69] Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12(6), 1019-1043.

[70] Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales. *Journal of the American Statistical Association*, 100(472).

[71] Zhang, S., and Karunamuni, R. J. (1998). On kernel density estimation near endpoints. *Journal of Statistical Planning and Inference*, 70(2), 301-316.

[72] Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1), 45-52.

[73] Zumbach, G. O., Corsi, F., and Trapletti, A. (2002). Efficient estimation of volatility using high frequency data. Available at *SSRN 306002*.