Journal of Econometrics 200 (2017) 79–103

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Model-free approaches to discern non-stationary microstructure noise and time-varying liquidity in high-frequency data☆

Richard Y. Chen [1,*], Per A. Mykland [2]

Department of Statistics, The University of Chicago, United States

ARTICLE INFO

Article history:
Received 19 December 2015
Received in revised form 8 January 2017
Accepted 3 May 2017
Available online 27 June 2017

JEL classification:
C12
C13
C14
C58

Keywords:
Microstructure
High-frequency tests
Statistical powers
Stable central limit theorems
Non-stationarity
Volatility
Liquidity

ABSTRACT

In this paper, we provide non-parametric statistical tools to test stationarity of microstructure noise in general hidden Itô semimartingales, and discuss how to measure liquidity risk using high-frequency financial data. In particular, we investigate the impact of non-stationary microstructure noise on some volatility estimators, and design three complementary tests by exploiting edge effects, information aggregation of local estimates and high-frequency asymptotic approximation. The asymptotic distributions of these tests are available under both stationary and non-stationary assumptions, thereby enable us to conservatively control type-I errors and meanwhile ensure the proposed tests enjoy the asymptotically optimal statistical power. Besides, it also enables us to empirically measure aggregate liquidity risks by these test statistics. As byproducts, functional dependence and endogenous microstructure noise are briefly discussed. Simulation with a realistic configuration corroborates our theoretical results, and our empirical study indicates the prevalence of non-stationary microstructure noise in New York Stock Exchange.

1. Introduction

The introduction of high-tech trading mechanisms into markets, for example, electronic communication networks (ECNs) and other electronic trading platforms, provides an opportunity for speculators and market makers to take advantage of speed in trading and market making, and this technological innovation also brings new regulatory challenges. The subsequent high-frequency trading results in a huge amount of high-frequency transaction and quotation data, which in particular opens two potential gates for research in theoretical and empirical asset pricing: one is estimation methodology using high-frequency data, since practitioners and researchers can get access to the big data and estimate variables of interest with greater accuracy; the other is a "frog eyes' view" on market microstructure, since low-latency data offers a valuable chance to investigate trading behaviors with a higher resolution than ever before.

Correspondingly, this paper's contributions to the literature are twofold: (i) one is stationarity test of microstructure noise, we study the estimation problem when using high-frequency data with non-stationary noises, and then test non-stationarity in microstructure noise via several complementary model-free approaches; (ii) the other one is on empirical market microstructure, since the microstructure noise can capture some information about market quality and liquidity, we estimate noise levels as measures of time-varying bid–ask spreads, risk aversions of market participants, etc., and detect short-term liquidity variations.

1.1. Literature review

The high-frequency finance practice motivates two clearly distinct and closely related researches:

One is more accurate estimation in financial econometrics, to name a few but not all, the estimation of integrated volatilities, quadratic covariances, the activities of jumps, the leverage effects, the volatility of volatility, the lead–lag effect. This stream of research started from Jacod (1994) and Jacod and Protter (1998)

☆ The authors thank the referees for valuable suggestions and insights which make significant improvements. The authors also benefited much from discussions with Yingying Li, Xinhua Zheng, Yoann Potiron, Markus Bibinger. This research was funded by National Science Foundation Grant DMS 14-07812. All comments are gratefully welcomed.

* Corresponding author.
E-mail addresses: yrchen@uchicago.edu (R.Y. Chen), mykland@pascal.uchicago.edu (P.A. Mykland).
1 Richard Y. Chen is a Ph.D. student in Statistics.
2 Per A. Mykland is the Robert M. Hutchins Distinguished Service Professor.

http://dx.doi.org/10.1016/j.jeconom.2017.05.015
0304-4076/

from the perspective of stochastic calculus, and Foster and Nelson (1996), Engle (2000), Zhang (2001), Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002) in the context of econometrics. Now, the high-frequency financial econometrics has already developed into a considerably influential research field with numerous prominent scholars and there are already monographs on this area: Jacod and Shiryaev (2003) and Jacod and Protter (2012) developed probabilistic tools for high-frequency financial data analysis, Aït-Sahalia and Jacod (2014) provided an excellent overview in econometric literature, Hautsch (2012) is a good account from a financial standpoint. There are also academic chapters concisely reviewing high-frequency financial econometrics: (Russell and Engle, 2010; Mykland and Zhang, 2012; Jacod, 2012).

The other one is the study of market microstructure. Low-latency data allows financial practitioners and researchers to look at the financial markets at a higher resolution level, for example, one can know the bid/ask dynamics within each second, one can also know the order flow through the limit order book. The market microstructure theory studies how the latent demand and latent supply of market participants are ultimately translated into prices by studying the specific market structure in detail. The cornerstone papers include Glosten and Milgrom (1985) and Kyle (1985), both of them are using (pseudo)[3] game-theoretical argument in information economics. More comprehensive books include O'Hara (1995) and Hasbrouck (2007). However, when looking closely at the transaction or quotation prices, one can find that the price is no longer an Itô semimartingale, not even random walk. For this reason, according to market microstructure theory (O'Hara, 2003), the semimartingale model in classical asset pricing theory (Harrison and Pliska, 1981; Delbaen and Schachermayer, 1994) is not a photographic depiction of the real prices of financial assets, yet it is still a fairly good approximation to asset prices when the sampling frequency is relatively low, and that is the reason the literature suggests using at most 5-min subsampling.

Some estimation methods for integrated volatility using noisy high-frequency financial data have already been well established: (i) Zhang et al. (2005) found the first consistent estimator (two-time scale realized volatility) using subsampling and averaging in the presence of i.i.d. market microstructure noise and Zhang (2006) gave a multi-scale version with the optimal rate of convergence $n^{\frac{1}{4}}$, Li and Mykland (2007) discussed the robustness of TSRV to noise assumptions in general, Kalnina and Linton (2008) generalized the TSRV to the model with endogenous and diurnal noise and put forward a modified version of TSRV which we shall use in this paper. Later, Aït-Sahalia et al. (2011) generalized the model to allowing correlated noises under stationary and strong-mixing conditions; (ii) Barndorff-Nielsen et al. (2008) provided a kernel-based estimator under the model in which the noise process is temporarily dependent and stationary and possibly linearly correlated with the latent Itô process, their inference is also robust to endogenous spacing; (iii) Jacod et al. (2009) designed a generalized version of the pre-averaging approach (Podolskij and Vetter, 2009), under a Markovian noise model which allows arbitrary fashion of noise but without correlation between noise and the latent process; (iv) Motivated by the likelihood method from Aït-Sahalia et al. (2005) and Xiu (2010) established quasi-maximum likelihood method (QMLE) in the estimation of integrated volatility; (v) Bibinger et al. (2014) developed the local generalized method of moments to estimate quadratic covariation using noisy high-frequency data.

Many estimators of integrated volatilities using high-frequency noisy data were developed under the assumption that the microstructure noise is stationary. However, the literature in empirical finance, such as Admati and Pfleiderer (1988), Hasbrouck (1993), Andersen and Bollerslev (1997) and Gouriéroux et al. (1999), has already shown in 1990s that markets exhibit systematic intra-day patterns. Therefore, allowing heteroskedasticity and non-stationary in microstructure noise in integrated volatility estimation is of particular importance in application. Particularly, Kalnina and Linton (2008) used a parametric model to describe the diurnal pattern in microstructure noise. Aït-Sahalia and Yu (2009) used the estimates of noise variance in high-frequency data to measure the market liquidity from June 1996 to December 2005. There is other related research in the literature, Awartani et al. (2009) studied the changes in microstructure noise due to sampling frequency, Bandi et al. (2013) derived the optimal sampling frequency in terms of finite-sample forecast mean squared error in linear forecast model with non-stationary microstructure noise.

1.2. Structure of this paper

Section 2 describes our model and assumptions; after showing non-stationarity effect on the two scale estimator, complementary statistical tests are designed to detect microstructure noise stationarity based on high-frequency asymptotics, the asymptotic distributions under both null and alternative hypotheses and their implications for testing are shown in Sections 3–5; Section 6 introduces an aggregate measure of liquidity risk and studies its estimation problem; relation between volatility and variance of microstructure noise, as well as endogenous microstructure noise are discussed in Section 7; Sections 8 and 9 contain our simulation and empirical analysis; Section 10 concludes. Some proofs are given in the Appendix.

2. The model and assumptions

2.1. Model setup

Firstly, we have a filtered probability space $\left(\Omega^{(0)}, \mathcal{F}^{(0)}, \left\{\mathcal{F}_t^{(0)}\right\}_{t\geq 0}, \mathbb{P}^{(0)}\right)$ on which a latent Itô semimartingale $\{X_t\}_{t\geq 0}$ is adapted, and can be described by

$$X_t = X_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s + J_t \tag{1}$$

where $\{b_t\}_{t\geq 0}$ is the drift, σ_t^2 is the spot volatility in financial terminology (for example, its dynamics can be described by Heston model (Heston, 1993)); $\{W_t\}_{t\geq 0}$ is a 1-dimensional Wiener process; J_t is a jump process which is described in Section 2.3.

Secondly, we have another filtered probability space $\left(\Omega^{(1)}, \mathcal{F}^{(1)}, \left\{\mathcal{F}_t^{(1)}\right\}_{t\geq 0}, \mathbb{P}^{(1)}\right)$ on which the observable process $\{Y_t\}_{t\geq 0}$ is adapted. Then, we can define the market microstructure noise process $\{e_t\}_{t\geq 0}$,[4] as the difference between the latent and observable processes:

$$e_t \equiv Y_t - X_t \tag{2}$$

besides we define

$$Z_t \equiv E_{\mathbb{P}^{(1)}}(Y_t|\mathcal{F}^{(0)}) = X_t + E_{\mathbb{P}^{(1)}}(e_t|\mathcal{F}^{(0)}) \tag{3}$$

[3] To say it "pseudo" because in the model considered in Kyle (1985), the market maker does not aim to maximize their utility, instead his or her objective is only to guarantee market clearing.

[4] Although the noise is immaterial outside the observation times, it does not harm to assume there exist such a noise process in continuous time.

we call $\{Z_t\}_{t\geq 0}$ the "*estimable latent process*" because we can indeed estimate it from the actual observations via, for example, pre-averaging (Podolskij and Vetter, 2009; Jacod et al., 2009, 2010; Mykland and Zhang, 2016). It is natural to assume the process $\{Z_t\}_{t\geq 0}$ is an Itô semimartingale, for example, if we assume $Z_t = f(X_t)$ for some $f(\cdot) \in \mathcal{C}^2(\mathbb{R})$ (Li and Mykland, 2007) then $\{Z_t\}_{t\geq 0}$ is an Itô semimartingale.[5] Based upon $\{Z_t\}_{t\geq 0}$, we can define a noise process $\{\epsilon_t\}_{t\geq 0}$ of another form, which is not necessarily the difference between the observable process $\{Y_t\}_{t\geq 0}$ and the latent process $\{X_t\}_{t\geq 0}$, instead defined theoretically via

$$\epsilon_t \equiv Y_t - Z_t = e_t - E_{\mathbb{P}^{(1)}}(e_t|\mathcal{F}^{(0)}) \tag{4}$$

we call $\{\epsilon_t\}_{t\geq 0}$ the "*distinguishable noise*", which can be disentangled from the estimable latent process $\{Z_t\}_{t\geq 0}$ (Bandi and Russell, 2006).

Thirdly, we have a Markov kernel to provide a connection between the processes $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$, namely $Q_t(\omega^{(0)}, \mathrm{d}y) : (\Omega^{(0)}, \mathcal{F}^{(0)}) \times (\mathbb{R}, \mathcal{B}(\mathbb{R})) \longmapsto [0, 1]$, i.e., conditional on the whole latent process X, there exists a probability measure on the space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Thus, all the relevant process, either latent or observable, can be defined on the extended filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$[6] where

$$\begin{cases} \Omega \equiv \Omega^{(0)} \times \Omega^{(1)}, \mathcal{F} \equiv \mathcal{F}^{(0)} \otimes \mathcal{F}^{(1)} \\ \mathcal{F}_t \equiv \bigcap_{s>t} \mathcal{F}_s^{(0)} \otimes \mathcal{F}_s^{(1)} \\ \mathbb{P}(\mathrm{d}\omega^{(0)}, \mathrm{d}\omega^{(1)}) \equiv \mathbb{P}^{(0)}(\mathrm{d}\omega^{(0)}) \cdot \otimes_{t\geq 0} Q_t(\omega^{(0)}, \mathrm{d}y_t(\omega^{(1)})). \end{cases} \tag{5}$$

Moreover, define

$$g_t(\omega^{(0)}) = \int_{\mathbb{R}} \left|y - Z_t(\omega^{(0)})\right|^2 Q_t(\omega^{(0)}, \mathrm{d}y)$$

i.e., $g_{t_i} = E(\epsilon_{t_i}^2|\mathcal{F}^{(0)})$. By this definition, $\{g_t\}_{t\geq 0}$ is also a stochastic process. Note that g_t could depend on more than one latent random variables, i.e., it is possible that $g_t(\omega^{(0)}) = g_t(X_t(\omega^{(0)}), Z_t(\omega^{(0)}), \sigma_t^2(\omega^{(0)}), \ldots)$ for each t. In Sections 4 and 6 regarding some behaviors in the presence of non-stationary microstructure noise, we pose specific restrictions on the process $\{g_t\}_{t\geq 0}$, and let it be an Itô diffusion, and use asymptotic properties to show asymptotically optimal power and measure liquidity in high-frequency data.

2.2. Observational notation

This subsection can be skipped at the first reading. Please be advised to go back to this subsection when encounter the observational notation in later sections.

Suppose we focus on a finite interval $[0, T]$ on which ultra-high frequency data is recorded. Define $\mathcal{G}$ to be the finest time grid whence all the observations were obtained. Suppose we have n observations after the reference starting point 0, then $\mathcal{G}$ can be written as

$$\mathcal{G} \equiv \{t_0 = 0, t_1, t_2, \ldots, t_n\}. \tag{6}$$

We sometimes do sparse sampling, typically start from the kth observation and take one sample from every K observations. Formally, we define sub-grids $\mathcal{G}^{(K,k)}$'s indexed by $k = 0, \ldots, K-1$ for each $K \in \mathbb{N}_+$:

$$\mathcal{G}^{(k)} = \mathcal{G}^{(K,k)} \equiv \left\{t_k, t_{k+K}, t_{k+2K}, \ldots, t_{k+(\lfloor n/K \rfloor - 1)K}\right\}, \quad \text{where } k = 0, 1, \ldots, K-1. \tag{7}$$

To analyze the edge effect[7] and the modified TSRV, we need more observational notation:

$$\left.\begin{array}{ll} \mathcal{G}'^{(k)} = \mathcal{G}'^{(K,k)} & \equiv \{\min \mathcal{G}^{(K,k)} + 1, \min \mathcal{G}^{(K,k)} + 2, \ldots, \max \mathcal{G}^{(K,k)}\} \\ \mathcal{G}''^{(k)} = \mathcal{G}''^{(K,k)} & \equiv \{\min \mathcal{G}^{(K,k)} + 1, \min \mathcal{G}^{(K,k)} + 2, \ldots, \max \mathcal{G}^{(K,k)} - 1\} \\ \mathcal{G}^{(\min)} = \mathcal{G}^{(K,\min)} & \equiv \{\min \mathcal{G}^{(K,1)}, \min \mathcal{G}^{(K,2)}, \ldots, \min \mathcal{G}^{(K,K)}\} \\ \mathcal{G}^{(\min)} = \mathcal{G}^{(K,\max)} & \equiv \{\max \mathcal{G}^{(K,1)}, \max \mathcal{G}^{(K,2)}, \ldots, \max \mathcal{G}^{(K,K)}\} \end{array}\right\} \tag{8}$$

thus, we have $|\mathcal{G}^{(\min)}| = |\mathcal{G}^{(\max)}| = K$ and the following relationships

$$\bigcup_{k=1}^{K} \mathcal{G}^{(k)} = \mathcal{G}^{(\min)} \bigcup \left(\bigcup_{k=1}^{K} \mathcal{G}'^{(k)}\right) = \mathcal{G}^{(\min)} \bigcup \left(\bigcup_{k=1}^{K} \mathcal{G}''^{(k)}\right) \bigcup \mathcal{G}^{(\max)}$$

sometimes, we will also denote by $\mathcal{H}_i$ the ith time point in a given grid $\mathcal{H}$, for example, $\mathcal{G}_i^{(\min)} = \min \mathcal{G}^{(K,i)} = \min \mathcal{G}^{(i)}$, $\mathcal{G}_i'^{(k)} = \mathcal{G}_i''^{(k)}$.[8]

In order to define some of our tests in Section 4, we need to introduce some shrinking moving windows and local sampling grids. Later, we will partition the fixed time interval $[0, \mathcal{T}]$ (in application, $\mathcal{T}$ could be 5 business days or some longer periods) into r_n (depends on n and $r_n \to \infty$) sub-intervals $(T_i, T_{i+1}]$'s, such that each $(T_{i-1}, T_i]$ contains K_n observations, i.e. $T_i = t_{iK_n}$, $0 = T_0 \leq T_1 \leq T_2 \leq \cdots \leq T_{r_n}$ and $r_n = \lfloor n/K_n \rfloor$. We also let $\mathcal{S}_i$ denote the shrinking sampling grid $\{t_{(i-1)K_n}, \ldots, t_{iK_n}\}$ over $[T_{i-1}, T_i]$, i.e., $|\mathcal{S}_i| = K_n$, $\mathcal{S}_i = \mathcal{G} \cap [T_{i-1}, T_i]$, $\bigcup_{i=1}^{r_n} \mathcal{S}_i = \{t_i \in \mathcal{G} : t_i \leq T_{r_n}\}$.

2.3. Assumptions

Beyond the model setup in Section 2.1, we have to make the following *identification assumption* in order to achieve identifiability and estimability:

$$\mathrm{d}Z_t \equiv \mathrm{d}X_t = b_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t + J_t \tag{9}$$

otherwise all the estimation methods will break down (Jacod et al., 2009). Note that under the identification assumption (9), we have $\{e_t\}_{t\geq 0}$ and $\{\epsilon_t\}_{t\geq 0}$ are identical, and there is no correlation between noise and the latent process.

As a sum-up, the following assumptions will be needed for various results:

Assumption 1 (*Diffusion Part of Itô Semimartingale*)**.** The underlying model is (1), $\{b_t\}_{t\geq 0}$, $\{\sigma_t\}_{t\geq 0}$ and $\{W_t\}_{t\geq 0}$ are adapted, $\{b_t\}_{t\geq 0}$ and $\{\sigma_t\}_{t\geq 0}$ are càdlàg processes and locally bounded.

Assumption 2 (*Jumps of Itô Semimartingale*)**.** $J_t = \int_0^t \int_{\mathbb{R}} x\mathbb{1}_{\{|x|\leq 1\}}(\mu - \nu)(\mathrm{d}s, \mathrm{d}x) + \int_0^t \int_{\mathbb{R}} x\mathbb{1}_{\{|x|>1\}}\mu(\mathrm{d}s, \mathrm{d}x)$ with μ being a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ and ν being the predictable compensator

[5] The definition (3) suggests the possibility of our inability to recover the latent process $\{X_t\}$ from the noisy observations $\{Y_t\}$, since Z_t does not necessarily equal to X_t. More strikingly, as later discussed, this allows correlation between the microstructure noise and the latent process.

[6] This model combines the features of the models in Li and Mykland (2007) and Jacod and Protter (2012) (or Jacod et al., 2009, 2010), and is endowed with an additional feature that ϵ_{t_i}'s are not defined as the differences between the observations Y_{t_i}'s and the latent values X_{t_i}'s but the differences between the observations Y_{t_i}'s and the values we can actually recover which are Z_{t_i}'s.

[7] The edge effect is a pervasive phenomenon in non-parametric high-frequency econometrics. Verbally stated, edge effect is "information phasing in and phasing out at the edges of time intervals", which is caused by inhomogeneous usage of data. Although undesirable, this feature is inevitable in estimation.

[8] The time grids defined in (7) and (8) depend on the tuning parameter K which should be more properly written as K_n, however, the dependence on n will be suppressed in the observational notation, for the sake of readability and notational ease. Nonetheless, it is important to keep this implicit dependence on n in mind.

of μ in the sense that $(\mu - \nu)((0,t],A)$ is a local martingale for $\forall t > 0, \forall A \in \mathcal{B}(\mathbb{R})$. One could write $\nu(\mathrm{d}t, \mathrm{d}x) = \mathrm{d}t \otimes \lambda(\mathrm{d}x)$ where λ is a σ-finite measure (for example, Lebesgue measure) on $\mathbb{R}$.

Assumption 3 (*Finite Jumps of Itô Semimartingale*)**.** On top of Assumption 2, assume $\exists$ a function Γ on $\mathbb{R}$ such that $\int_{\mathbb{R}} \Gamma(x)\lambda(\mathrm{d}x) < \infty$ where $\Gamma \geq 1$.

Assumption 4 (*Identifiable Hidden Itô Semimartingale*)**.** The underlying process is (1); and we have the identifiability assumption (9).

Assumption 5 (*Conditional Independence*)**.** Conditional on the latent variable(s), the observations Y_{t_i}'s at different times are independent, i.e., $Y_{t_i} \perp\!\!\!\perp Y_{t_j}$ for $i \neq j$. This assumption simplifies the proof substantially.[9]

Assumption 6 (*Locally Boundedness of Microstructure Effect*)**.** $\forall l > 0$, and $\forall \alpha > 0$, $\exists M_{(\alpha,l)}$, such that $E\left(|\epsilon_{t_i}|^{\alpha}|\mathcal{F}^{(0)}\right) \leq M_{(\alpha,l)}$, when $X_{t_i} \in [-l, l], \sigma^2_{t_i} \in (0, l]$.

Assumption 7 (*Possibly Irregular Observational Grid with Shrinking Mesh*)**.** The sampling times can be irregular, but independent of the latent process. The Mesh of the grid $\mathcal{G}$ goes to zero, more specifically, $\max_{1\leq i\leq n} \Delta t_i = O_p\left(\frac{1}{n}\right)$.

Based on some of these assumptions, we provide results involving various modes of stochastic convergences. It is necessary to clarify our notation for these convergence modes: $\xrightarrow{\mathbb{P}}$ means convergence in probability, $\xrightarrow{\mathcal{L}}$ means convergence in law (convergence in distribution, weak convergence), $\xrightarrow{\mathcal{L}-s}$ means stable convergence in law.[10]

3. Testing stationarity/non-stationarity: the first test

In this article, we are considering testing the null hypothesis that the market microstructure noise is stationary:

$$H_0 : \{\epsilon_t\}_{t\geq 0} \text{ is stationary} \longleftrightarrow H_1 : \{\epsilon_t\}_{t\geq 0} \text{ is non-stationary}$$

and we concern the following questions:

- Could we find any *non-parametric test* to tell the stationarity of microstructure noise?
- Is any stationarity test valid in terms of controlling type-I error?
- Is it asymptotically optimal in that its statistical power is the largest in asymptotics?

3.1. Prelude: non-stationarity and its remedy in estimation

In this subsection, we divert our focus to the estimation of integrated volatility (or continuous quadratic variation in the terminology of stochastic calculus) using high-frequency data contaminated by (possibly non-stationary) market microstructure noise. Our first test statistic was inspired by this.

Two-time scale realized volatility estimator (TSRV) (Zhang et al., 2005) is the first consistent estimator of integrated volatility using noisy high frequency financial data. In this article, we define $[Y,Y]_{\mathcal{H}}$ as the realized variance of process $\{Y_t\}$ computed on a given sampling grid $\mathcal{H}$. The TSRV is defined as follows:

$$\widehat{\langle X,X\rangle}_T^{(TSRV,K_n)} \equiv [Y,Y]_T^{(avg,K_n)} - \frac{n-K_n+1}{nK_n}[Y,Y]_{\mathcal{G}} \tag{10}$$

where, according to the notation introduced in Section 2.2,

$$\begin{aligned}
[Y,Y]^{(avg,K_n)} &= \frac{1}{K_n}\sum_{k=0}^{K_n-1}[Y,Y]_{\mathcal{G}^{(k)}} \\
[Y,Y]_{\mathcal{G}} &= \sum_{i=1}^{n}(Y_{t_i} - Y_{t_{i-1}})^2 \\
[Y,Y]_{\mathcal{G}^{(k)}} &= \sum_{t_i\in\mathcal{G}'^{(k)}}(Y_{t_i} - Y_{t_{i-K}})^2, \text{ for } k = 0,\ldots,K-1.
\end{aligned}$$

The optimal choice for the tuning parameter is $K_n = O(n^{\frac{2}{3}})$,[11] which results in the best possible order of TSRV. In the identical fashion, we can define $[\epsilon,\epsilon]_{\mathcal{G}}$, $[\epsilon,\epsilon]_T^{(avg,K_n)}$ and $[Z,Z]_T^{(avg,K_n)}$.

The intuition behind the design of $\widehat{\langle X,X\rangle}_T^{(TSRV,K_n)}$ is sub-sampling and averaging: each $[Y,Y]_{\mathcal{G}^{(k)}}$ is computed on a sparser grid hence mitigate the microstructure effect, hence their average $[Y,Y]_T^{(avg,K_n)}$ should be more closer to $\langle X,X\rangle_T$; the second term $[Y,Y]_{\mathcal{G}}$ is a good proxy to the noise variance, hence it is to offset the bias due to the noise in $[Y,Y]_T^{(avg,K_n)}$.

The TSRV was originally designed under the setting where microstructure noises are stationary; however, under non-stationary microstructure noises, TSRV has a bias term produced by edge effect due to the following lemma:

Lemma 1. *Under the Assumptions 1 and 3–7, we have*

$$\begin{aligned}
&[Y,Y]_T^{(avg,K_n)} - [Z,Z]_T^{(avg,K_n)} \\
&= \underbrace{\frac{2}{K_n}\sum_{k=1}^{K}\sum_{t_i\in\mathcal{G}''^{(k)}} g_{t_i} + \frac{1}{K_n}\sum_{t_i\in\mathcal{G}^{(\min)}} g_{t_i} + \frac{1}{K_n}\sum_{t_i\in\mathcal{G}^{(\max)}} g_{t_i}}_{\text{bias in } [Y,Y]_T^{(avg,K)} \text{ due to non-stationary noise}} + o_p(1).
\end{aligned} \tag{11}$$

From Lemma 1, we can see the noise in each time point does not contribute equally to the bias in the averaged realized variance $[Y,Y]_T^{(avg,K_n)}$. In the first and last K_n sample points, the conditional second moments of noises are multiplied by the factor $\frac{1}{K_n}$; in contrast, the conditional second moments of noises in the middle of the sample points are multiplied by the factor $\frac{2}{K_n}$. However, the noise correction term $[Y,Y]_{\mathcal{G}}$ in (10) acts as if g_{t_i}'s all have the same contribution to the noise component in $[Z,Z]_T^{(avg,K_n)}$. The modification to the TSRV and the first two tests are motivated by the inhomogeneity of utilization of information at the two edges of the time interval $[0,T]$.

To the best of our knowledge, Kalnina and Linton (2008) is the first study which considered the edge effect in TSRV due to the non-stationary microstructure noise, and they redefined the TSRV by $[Y,Y]_T^{(avg,K_n)} - \frac{n-K_n+1}{nK_n}[Y,Y]_T^{\{n\}}$ where

$$[Y,Y]_T^{\{n\}} = \frac{1}{2}\left(\sum_{i=1}^{n-K_n}(Y_{t_{i+1}} - Y_{t_i})^2 + \sum_{i=K_n}^{n-1}(Y_{t_{i+1}} - Y_{t_i})^2\right)$$

[9] An interpretation of this assumption is that the market microstructure effects occurred at different times are independent if the market participants know the latent efficient prices.

[10] The concept "stable convergence in law" may appear unfamiliar for some readers, please refer to Mykland and Zhang (2012) or chapter 2 in Jacod and Protter (2012) for definition.

[11] A caveat in application is to choose K_n such that $n - \lfloor n/K_n \rfloor K$ is sufficiently small, in order to reduce the edge effect.

under a parametric model which incorporates the diurnal and endogenous measurement error.[12] In the following, we used this design to attack the non-stationarity problem under the general hidden Itô semimartingale model given in Section 2.

In this paper, we call the new TSRV consisting of the modified version of realized variance in Kalnina and Linton (2008) as "sample-weighted TSRV", which is defined as

$$\widehat{\langle X,X\rangle}_T^{(WTSRV,K_n)} = [Y,Y]_T^{(avg,K_n)} - \frac{1}{K_n}[Y,Y]_T^{\{n\}}.$$

The sample-weighted TSRV enjoys the following asymptotic property under the general model in Section 2:

Theorem 1. *Suppose there are n observations in the finite time interval* $[0,T]$. *When we take* $K_n = cn^{2/3}$ *for some constant c, under the Assumptions* 1 and 3–7, *we have*

$$n^{1/6}\left(\widehat{\langle X,X\rangle}_T^{(WTSRV,K_n)} - [Z,Z]_T\right) \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(0, \frac{8}{Tc^2}\int_0^T g_t^2 dt + \frac{4cT}{3}\int_0^T \sigma_t^4\, dt\right). \tag{12}$$

The theorem tells us the sample-weighted TSRV in non-stationary noise setting enjoys the same asymptotic property as those of traditional TSRV in stationary noise setting (Zhang et al., 2005; Li and Mykland, 2007; Aït-Sahalia et al., 2011), in that the asymptotic distribution as well as the convergence rate remains unchanged; in other words, the asymptotic property of the sample-weighted TSRV is invariant with respect to non-stationary market microstructure noise.

3.2. The first test $N(Y,K_n)_T^n$

Assuming H_0 is true, both of the asymptotic distributions of the original TSRV and the sample-weighted TSRV are mixed normals. So, the asymptotic distribution of the difference between two different versions (after proper scaling) is also a mixed normal. Therefore, we can test the null H_0 based on the asymptotic behavior of their difference $D(Y,K_n)_T^n = \sqrt{K_n}\left(\widehat{\langle X,X\rangle}_T^{(WTSRV,K_n)} - \widehat{\langle X,X\rangle}_T^{(TSRV,K_n)}\right)$, note that

$$D(Y,K_n)_T^n = \frac{n-2(K_n-1)}{2nK_n^{1/2}}[Y,Y]_{\mathcal{G}^{(\min)}} + \frac{n-2(K_n-1)}{2nK_n^{1/2}}[Y,Y]_{\mathcal{G}^{(\max)}} - \frac{K_n-1}{nK_n^{1/2}}[Y,Y]_{\mathcal{G}/(\mathcal{G}^{(\min)}\cup\mathcal{G}^{(\max)})}. \tag{13}$$

The first test statistic $N(Y,K_n)_T^n$ is designed as follows:

$$N(Y,K_n)_T^n \equiv \begin{cases} \dfrac{D(Y,K_n)_T^n}{\sqrt{\frac{1}{n}[Y;4]_{\mathcal{G}} - \frac{3}{2n^2}[Y,Y]_{\mathcal{G}}^2}}, & \text{if } [Y;4]_{\mathcal{G}} - \dfrac{3}{2n}[Y,Y]_{\mathcal{G}}^2 \neq 0 \\ 0, & \text{if } [Y;4]_{\mathcal{G}} - \dfrac{3}{2n}[Y,Y]_{\mathcal{G}}^2 = 0 \end{cases} \tag{14}$$

where $[Y;4]_{\mathcal{G}} = \sum_{i=1}^n (Y_{t_i} - Y_{t_{i-1}})^4$ is the sample quarticity based on the observation Y_{t_i}'s.

Our first test statistic has the following asymptotic property:

Theorem 2. *If the noise process is stationary, under the Assumptions* 1, 2 and 5–7, *as long as* $K_n \to \infty$ *but* $K_n = o(n)$,

$$N(Y,K_n)_T^n \xrightarrow{\mathcal{L}} N(0,1). \tag{15}$$

We use this result to test the stationarity of the market microstructure noise in Section 9.2 (Fig. 9).

The denominator of the test statistic (15), namely $\frac{1}{n}[Y;4]_{\mathcal{G}} - \frac{3}{2n^2}([Y,Y]_{\mathcal{G}})^2$ is actually an estimator of $2E(\epsilon^4|\mathcal{F}^{(0)})$. This is formally introduced in (17), which is not only used in the first test statistic but also used in the second test statistic in Section 4.2. It is interesting in its own right, hence we here give the result:

Lemma 2. *If we define the process* $h_t(\omega^{(0)}) \equiv E(\epsilon_t^4|\mathcal{F}^{(0)})(\omega^{(0)})$, *then under the Assumptions* 1, 2 and 5–7, *we have*

$$\frac{1}{n}[Y;4]_{\mathcal{G}} = \frac{2}{T}\int_0^T h_t\, dt + \frac{6}{T}\int_0^T g_t^2\, dt + O_p\left(\frac{1}{\sqrt{n}}\right). \tag{16}$$

Remark 1. Based on Lemma 2, if the noise is stationary, $\frac{1}{2n}[Y,Y,Y,Y]_{\mathcal{G}}|\mathcal{F}^{(0)} \xrightarrow{\mathbb{P}} E(\epsilon^4|\mathcal{F}^{(0)}) + 3\left(E(\epsilon^2|\mathcal{F}^{(0)})\right)^2$, so a natural estimate of $E(\epsilon^4|\mathcal{F}^{(0)})$ is

$$\widehat{E(\epsilon^4|\mathcal{F}^{(0)})} = \frac{1}{2n}[Y;4]_{\mathcal{G}} - \frac{3}{4n^2}[Y,Y]_{\mathcal{G}}^2. \tag{17}$$

Remark 2. We now investigate the behavior of our first test statistic when microstructure noise is non-stationary. Since $\frac{1}{2n}[Y,Y]_{\mathcal{G}} = \frac{1}{T}\int_0^T g_t\, dt + o_p(1)$, hence

$$\widehat{E(\epsilon^4|\mathcal{F}^{(0)})} \xrightarrow{\mathbb{P}} D_T \equiv \begin{cases} E(\epsilon^4|\mathcal{F}^{(0)}) & \{\epsilon_t\}_{t\geq 0} \text{ is stationary} \\ \dfrac{1}{T}\displaystyle\int_0^T h_t\, dt + \dfrac{3}{T}\int_0^T g_t^2\, dt \\ \quad - \frac{3}{T^2}\left(\int_0^T g_t\, dt\right)^2 & \{\epsilon_t\}_{t\geq 0} \text{ is non-stationary.} \end{cases} \tag{18}$$

Since we assume local boundedness of noise variance, D_T is finite almost surely, regardless of noise stationarity. From the proof in Appendix A.3, we know

$$N(Y,K_n)_T^n = \sqrt{K_n} \times \frac{\overline{g}^{(\text{start})} + \overline{g}^{(\text{end})} - 2\overline{g}^{(\text{middle})}}{D_T} + O_p(1) \tag{19}$$

where

$$\begin{aligned} \overline{g}^{(\text{start})} &= \frac{1}{K_n}\sum_{t_i\in\mathcal{G}^{(\max)}} g_{t_i} \\ \overline{g}^{(\text{end})} &= \frac{1}{K_n}\sum_{t_i\in\mathcal{G}^{(\max)}} g_{t_i} \\ \overline{g}^{(\text{middle})} &= \frac{1}{n+1-2K_n}\sum_{k=1}^{K_n}\sum_{t_i\in\mathcal{G}_k''} g_{t_i}. \end{aligned} \tag{20}$$

Since $K_n = O\left(n^{2/3}\right)$ in our setup, $N(Y,K_n)_T$ explodes when the noise is not stationary. Thus, the type-II error of this test is asymptotically negligible.

Followed from Theorem 2 and Remark 2, we have

Corollary 1. *Assume* $\{g_t\}_{t\geq 0}$ *and* $\{h_t\}_{t\geq 0}$ *are càdlàg processes*[13] *on* $[0,T]$ *with* 0, *T being continuity points almost surly, additional we*

[12] The model upon which Kalnina and Linton (2008) was based is

$$\begin{aligned} dX_t &= \mu_t\, dt + \sigma_t\, dW_t \\ Y_{t_i} &= X_{t_i} + \epsilon_{t_i} \\ \epsilon_{t_i} &= u_{t_i} + v_{t_i} \\ u_{t_i} &= \delta\gamma_n(W_{t_i} - W_{t_{i-1}}) \\ v_{t_i} &= m(t_i) + n^{-\frac{\alpha}{2}}\omega(t_i)e_{t_i}, \alpha \in [0,1/2) \end{aligned}$$

where $e \perp\!\!\!\perp X$, i.i.d., $E(e) = 0$.

[13] The term "càdlàg" (French acronym of "continue droite, limite gauche") describes the property of a function that is everywhere right-continuous and has left limits everywhere, for example, a Brownian motion (sample path are continuous almost surly), Lévy processes (countably many jump discontinuities).

have the Assumptions 1, 2 and 5–7, and let $K_n \to \infty$, $K_n/n \to 0$. When the noise process is non-stationary,

$$N(Y, K_n)_T^n - \sqrt{K_n} \times \frac{g_0 + g_T - \frac{1}{T}\int_0^T g_t \mathrm{d}t}{\frac{1}{T}\int_0^T h_t\,\mathrm{d}t + \frac{3}{T}\int_0^T g_t^2\,\mathrm{d}t - \frac{3}{T^2}\left(\int_0^T g_t\,\mathrm{d}t\right)^2} \xrightarrow{\mathcal{L}} N(0,1) \tag{21}$$

on the event that $\int_0^T h_t\,\mathrm{d}t + 3\int_0^T g_t^2\,\mathrm{d}t - \frac{3}{T}\left(\int_0^T g_t\,\mathrm{d}t\right)^2 \neq 0$.

Remark 3. The test statistic $N(Y, K_n)_T^n$ can disclose potential non-stationarity in the market microstructure noise via two edges of the mesh $\mathcal{G}^{(\min)}$, $\mathcal{G}^{(\max)}$ and the middle of the mesh $\mathcal{G}/(\mathcal{G}^{(\min)} \cup \mathcal{G}^{(\max)})$. We can show there are, in latter subsections, schemes which are able not only to reflect the heterogeneity between two edges and the middle, but also to capture almost all of the information about the non-stationarity in the data, however, inevitably with more computational cost. We will discuss these schemes in Section 4.

4. The second and third tests

4.1. The general idea

The second and third tests are designed as an attempt to effectively utilize all the information relevant to noise stationarity contained in the data, in contrast to the first test $N(Y, K_n)^n$ (see Remark 3). The basic idea of the second and third tests is to conduct local tests on sub-intervals and then combine evidences from all the local tests.

To straighten the idea, recall the observational notation in Section 2.2, we partition the fixed time interval $[0, \mathcal{T}]$ into r_n subintervals $(T_i, T_{i+1}]$'s, such that each $(T_{i-1}, T_i]$ contains K_n observations. Similar to the definition of the first test statistic, but instead of the whole interval $[0, \mathcal{T}]$, the second test uses local test statistic defined on a moving window of the form $(T_{i-1}, T_{i-1+s_n}] \subset [0, \mathcal{T}]$ with a suitable window length s_n (in terms of the number of subintervals $(T_i, T_{i+1}]$'s):

$$D(Y, K_n, s_n)_i^n \equiv \frac{(s_n - 2)K_n + 2}{2 s_n K_n^{3/2}}\left([Y, Y]_{\mathcal{G}_i} + [Y, Y]_{\mathcal{G}_{i+s_n}}\right) - \frac{K_n - 1}{s_n K_n^{3/2}}[Y, Y]_{\cup_{k=2}^{s_n-1}\mathcal{G}_{i+k}} \tag{22}$$

Then, we use the overlapping window to calculate the quantity $U(Y, K_n, u)_{\mathcal{T}}^n$, which depends on the process Y, the stage of statistical experiment n, tuning parameter K_n and $s_n < \left\lfloor \frac{n}{K_n} \right\rfloor$, and a number $u > 0$:

$$U(Y, K_n, s_n, u)_{\mathcal{T}}^n \equiv \frac{1}{\lfloor n/K_n \rfloor - s_n + 1} \sum_{i=1}^{\lfloor n/K_n \rfloor - s_n + 1} \left|D(Y, K_n, s_n)_i^n\right|^u. \tag{23}$$

Similarly, we also define a quantity based on non-overlapping windows:

$$U'(Y, K_n, s_n, u)_{\mathcal{T}}^n \equiv \frac{1}{\lfloor n/(s_n K_n) \rfloor} \sum_{i=1}^{\lfloor n/(s_n K_n) \rfloor} \left|D(Y, K_n)_{(i-1)s_n+1}\right|^u \tag{24}$$

4.2. The second test $V(Y, K_n, s_n, 2)_{\mathcal{T}}^n$

We designed our second test statistic by

$$V(Y, K_n, s_n, 2)_{\mathcal{T}}^n \equiv \begin{cases} \sqrt{n/K_n}\left(U(Y, K_n, s_n, 2)_{\mathcal{T}}^n - \frac{1}{n}[Y; 4]_{\mathcal{G}} - \frac{3}{2n^2}[Y, Y]_{\mathcal{G}}^2\right)/\widehat{\eta}, & \text{if } \widehat{\eta} \neq 0 \\ 0, & \text{if } \widehat{\eta} = 0 \end{cases} \tag{25}$$

where

$$\widehat{\eta}^2 = \frac{6}{n^2}[Y; 4]_{\mathcal{G}}^2 - \frac{21}{n^3}[Y; 4]_{\mathcal{G}} \cdot [Y, Y]_{\mathcal{G}}^2 + \frac{39}{2n^4}[Y, Y]_{\mathcal{G}}^4. \tag{26}$$

We have the following result in regard to the asymptotic property of $V(Y, K_n, 2)_{\mathcal{T}}^n$:

Theorem 3 ($V(Y, K_n, s_n, 2)_{\mathcal{T}}^n$ *under the Null*)**.** *Under the Assumptions 1, 2 and 5–7, assume the noise process is stationary, and choose the tuning parameters such that $K_n/n \to 0$, $K_n/n^{1/3} \to \infty$, $s_n \to \infty$, s_n. Then, the test statistic $V(Y, K_n, 2)_{\mathcal{T}}^n$ has the following asymptotic property:*

$$V(Y, K_n, s_n, 2)_{\mathcal{T}}^n \xrightarrow{\mathcal{L}} N(0, 1) \tag{27}$$

on the event that $E(\epsilon^4|\mathcal{F})^2 - E(\epsilon^4|\mathcal{F})E(\epsilon^2|\mathcal{F})^2 + E(\epsilon^2|\mathcal{F})^4 \neq 0$.

We use this result to test the stationarity of the market microstructure noise in Section 9.2 (Fig. 10).

We can also define another quantity $V'(Y, K_n, s_n, 2)_{\mathcal{T}}^n$ based on non-overlapping intervals

$$V'(Y, K_n, s_n, 2)_{\mathcal{T}}^n \equiv \begin{cases} \sqrt{n/(s_n K_n)}\left(U'(Y, K_n, s_n, 2)_{\mathcal{T}}^n - \frac{1}{n}[Y; 4]_{\mathcal{G}} - \frac{3}{2n^2}[Y, Y]_{\mathcal{G}}^2\right)/\widehat{\eta}, & \text{if } \widehat{\eta} \neq 0 \\ 0, & \text{if } \widehat{\eta} = 0 \end{cases}$$

Following from Theorem 3, the asymptotic property of $V'(Y, K_n, s_n, 2)_{\mathcal{T}}^n$ can be derived.

Corollary 2. *Under the same conditions as in Theorem 3, assume the noise is stationary:*

$$V'(Y, K_n, s_n, 2)_{\mathcal{T}}^n \xrightarrow{\mathcal{L}} N(0, 1) \tag{28}$$

on the event that $E(\epsilon^4|\mathcal{F})^2 - E(\epsilon^4|\mathcal{F})E(\epsilon^2|\mathcal{F})^2 + E(\epsilon^2|\mathcal{F})^4 \neq 0$.

Remark 4. It is a little bit surprising when we compare Corollary 2 with Theorem 3, since the limiting mixed normals of $U(Y, K_n, s_n, 2)_{\mathcal{T}}^n$ and $U'(Y, K_n, s_n, 2)_{\mathcal{T}}^n$ have the same asymptotic variance which can be consistently estimated by $\widehat{\eta}$, though the convergence rate of the former is lower. However, the results only demonstrate the limiting behaviors. $V'(Y, K_n, s_n, 2)_{\mathcal{T}}^n$ required less computation, while $V(Y, K_n, s_n, 2)_{\mathcal{T}}^n$ is more accurate in terms of asymptotic approximation because of its higher rate of convergence.

4.3. The third test $\overline{V}(Y, K_n, 2)^n$

There is a moderate edge effect in the second test statistic (25) (coming from the first $s_n K_n$ and the last $s_n K_n$ observations). Motivated by Remark 3 regarding the first test statistic (14), we can design a complementary test statistic $\overline{V}(Y, K_n, 2)^n$ (defined by (30)) with the same asymptotic property with $V(Y, K_n, s_n, 2)^n$ when the noise is stationary, yet has a smaller edge effect in finite sample. However, we should keep $V(Y, K_n, s_n, 2)^n$ in our toolbox — although $\overline{V}(Y, K_n, 2)^n$ offers better approximation when noise is stationary, we will see $V(Y, K_n, s_n, 2)^n$ has more statistical power as indicated in Fig. 1.

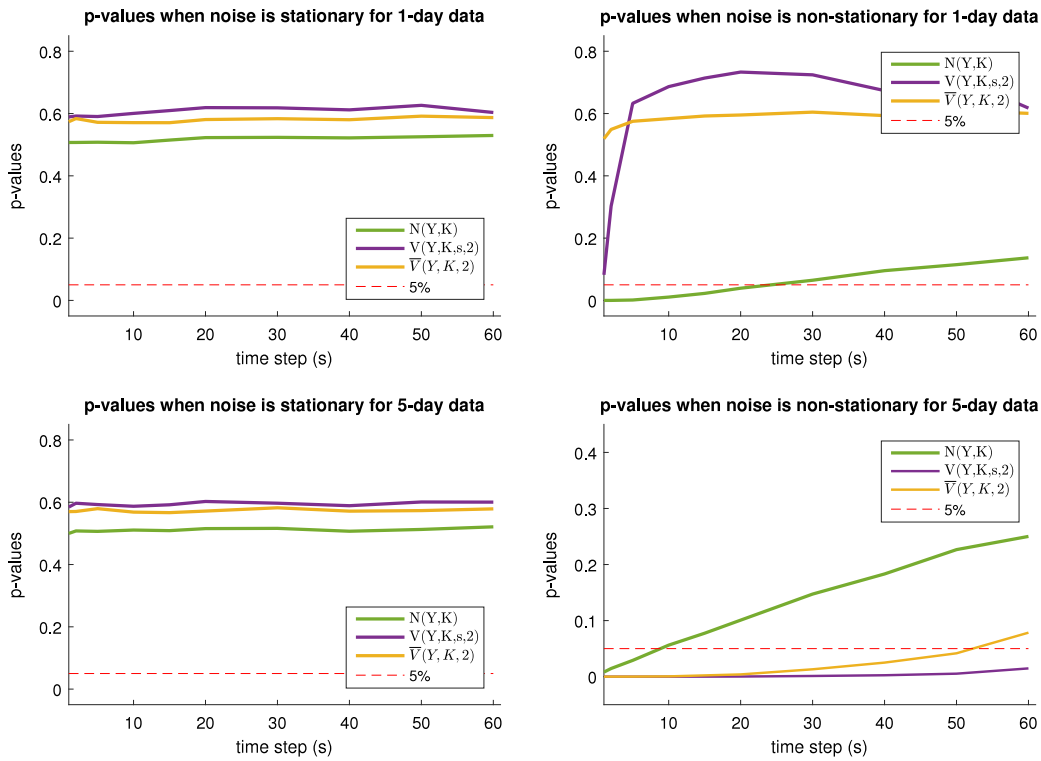


Fig. 1. Averaged p-values of the 3 tests proposed.
These plots show some properties of the tests we proposed:
1. For type-I error, all the tests can control their type-I errors pretty well regardless of the time span of the data, in that $N(Y,K)^n$ can very accurately control its type-I error in finite samples, $V(Y,K,s,2)^n$ and $\overline{V}(Y,K,2)^n$ are more conservative in the sense that their effective type-I error is smaller than specified;
2. For type-II error or statistical power, only $N(Y,K)^n$ performs satisfactorily on 1-day data, meanwhile, $V(Y,K,s,2)$ and $\overline{V}(Y,K,2)^n$ are much better when applied to multi-day data in terms of their larger statistical powers and faster convergence rates. Consistent with Corollary 3, $V(Y,K_n,s_n,2)^n$ has a better statistical power in finite sample.

The key component of the third test statistic is

$$\overline{U}(Y,K_n,2)^n_{\mathcal{T}} \equiv \frac{1}{4n}\sum_{i=1}^{\lfloor n/K_n\rfloor-1}\left|[Y,Y]_{\mathcal{S}_{i+1}}-[Y,Y]_{\mathcal{S}_i}\right|^2 \tag{29}$$

where each $\mathcal{S}_i$ denotes the shrinking sampling grid $\{t_{(i-1)K_n},\ldots,t_{iK_n}\}$ over $[T_{i-1},T_i]$ (recall the observational notation in Section 2.2), and $[Y,Y]_{\mathcal{S}_i}$ is the realized variance of process Y on the grid $\mathcal{S}_i$. Our third test statistic is defined as

$$\overline{V}(Y,K_n,2)^n_{\mathcal{T}} \equiv \begin{cases}\sqrt{n/K_n}\left(\overline{U}(Y,K_n,2)^n_{\mathcal{T}}-\frac{1}{n}[Y;4]_{\mathcal{G}}-\frac{3}{2n^2}[Y,Y]^2_{\mathcal{G}}\right)/\widehat{\eta}, & \text{if } \widehat{\eta}\neq 0\\ 0, & \text{if } \widehat{\eta}=0\end{cases} \tag{30}$$

where $\widehat{\eta}$ was defined in (26).

Theorem 4 ($\overline{V}(Y,K_n,2)^n_{\mathcal{T}}$ *under the Null*)**.** *Under the Assumptions 1, 2 and 5–7, assume the noise process is stationary, suppose $K_n\to\infty$, $K_n/n\to 0$ then the test statistic $\overline{V}(Y,K_n,2)^n_{\mathcal{T}}$ has the following asymptotic property:*

$$\overline{V}(Y,K_n,2)^n_{\mathcal{T}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \tag{31}$$

on the event that $E(\epsilon^4|\mathcal{F})^2-E(\epsilon^4|\mathcal{F})E(\epsilon^2|\mathcal{F})^2+E(\epsilon^2|\mathcal{F})^4\neq 0$.

4.4. Optimal statistical powers

How $V(Y,K_n,s_n,2)^n$ and $\overline{V}(Y,K_n,2)^n$ behave when the noise is non-stationary determine their statistical powers. If the test statistics tend to be large when the microstructure noise is non-stationary, they can easily detect non-stationarity.

The behaviors of $U(Y,K_n,s_n,2)^n$ and $\overline{U}(Y,K_n,2)^n$ largely indicate the behaviors of $V(Y,K_n,s_n,2)^n$ and $\overline{V}(Y,K_n,2)^n$. We investigate in this subsection the asymptotic behaviors of $U(Y,K_n,s_n,2)^n$ and $\overline{U}(Y,K_n,2)^n$ when microstructure noise is non-stationary under a slightly strengthened setting, we need 2 more assumptions on top of those assumptions in Section 2.3:

Assumption 8 (*Regular Sampling*)**.** The sample grid is equi-distant over the interval $[0,\mathcal{T}]$.

Assumption 9 (*Noise Variance Process is Itô*)**.** $\{g_t\}_{t\geq 0}$ is an Itô diffusion (in time):

$$g_t=\int_0^t \mu_s^{(g)}\,\mathrm{d}s+\int_0^t \sigma_s^{(g)}\,\mathrm{d}B_s \tag{32}$$

where $\{\mu_t^{(g)}\}_{t\geq 0}$ is locally bounded, optional and càdlàg, $\{B_t\}_{t\geq 0}$ is a standard Brownian motion, $\{\sigma_t^{(g)}\}_{t\geq 0}$ is a locally bounded Itô diffusion.

As described in Section 2.2, we partition the whole time interval into r_n disjoint sub-intervals $(T_{i-1},T_i]$ for $i=1,2,\ldots,r_n$ such that in each sub-interval we have K_n observations, particularly we have $T_0=0$ and $\mathcal{T}-T_{r_n}=o(1)$, $r_nK_n/n\to 1$. Since Assumption 8 is assumed, we let $\Delta T=T_i-T_{i-1}$, $\forall i=1,2,\ldots,r_n$.

Theorem 5 ($U(Y,K_n,s_n,2)^n$ *under the Alternative*)**.** *Assume Assumptions 1 and 2 and 5–7 as well as Assumptions 8 and 9. Let $\frac{K_n}{n^{1/2}}\to\infty$,*

$s_n \to \infty$ and $\frac{s_n K_n}{n} \to 0$ but $\frac{s_n^{2/5} K_n}{n^{3/5}} \to \infty$. Then, we have

$$\sqrt{\frac{n}{K_n}}\left(\frac{n}{s_n K_n^2} U(Y, K_n, s_n, 2)_{\mathcal{T}}^n - E_n^{(1)} - E_n^{(2)} - E_n^{(3)}\right) \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(0, \frac{2\mathcal{T}}{9}\int_0^{\mathcal{T}} (\sigma_t^{(g)})^4\,\mathrm{d}t\right) \tag{33}$$

on the event that $\{\sigma_t^{(g)}\}_{t\in[0,\mathcal{T}]}$ is non-vanishing, where

$$E_n^{(1)} = \frac{(s_n-2)^2}{3s_n^2}\int_0^{\mathcal{T}} (\sigma_t^{(g)})^2\,\mathrm{d}t$$
$$E_n^{(2)} = -\frac{s_n K_n}{6n}\left[(\sigma_0^{(g)})^2 + (\sigma_{\mathcal{T}}^{(g)})^2\right]\mathcal{T}$$
$$E_n^{(3)} = \frac{2n}{s_n K_n^2 \mathcal{T}}\int_0^{\mathcal{T}} h_t\,\mathrm{d}t, \qquad h_t(\omega^{(0)}) \equiv E(\epsilon_t^4|\mathcal{F}^{(0)})(\omega^{(0)}).$$

Theorem 6 ($\overline{U}(Y, K_n, 2)^n$ *under the Alternative*)**.** *Assume Assumptions 1 and 2 and 5–7 as well as Assumptions 8 and 9. Let $\frac{K_n}{n^{1/2}} \to \infty$ and $\frac{K_n}{n^{2/3}} \to 0$. Then, we have*

$$\sqrt{\frac{n}{K_n}}\left(\frac{n}{K_n^2}\overline{U}(Y, K_n, 2)_{\mathcal{T}}^n - \overline{E}^{(1)} - \overline{E}_n^{(2)}\right) \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(0, \frac{2\mathcal{T}}{3}\int_0^{\mathcal{T}} (\sigma_t^{(g)})^4\,\mathrm{d}t\right) \tag{34}$$

on the event that $\{\sigma_t^{(g)}\}_{t\in[0,\mathcal{T}]}$ is non-vanishing, where

$$\overline{E}^{(1)} = \frac{2}{3}\int_0^{\mathcal{T}} (\sigma_t^{(g)})^2\,\mathrm{d}t$$
$$\overline{E}_n^{(2)} = \frac{2n}{K_n^2\mathcal{T}}\int_0^{\mathcal{T}} h_t\,\mathrm{d}t, \qquad h_t(\omega^{(0)}) \equiv E(\epsilon_t^4|\mathcal{F}^{(0)})(\omega^{(0)}).$$

Theorems 3 and 4 provide us the asymptotic distributions of $V(Y, K_n, s_n, 2)^n$ and $\overline{V}(Y, K_n, 2)^n$ under the stationarity hypothesis, which aid us to control the type-I error. On the other hand, Theorems 5 and 6 reveal asymptotic behaviors of $V(Y, K_n, s_n, 2)^n$ and $\overline{V}(Y, K_n, 2)^n$ under the alternative hypothesis by respectively analyzing $U(Y, K_n, s_n, 2)^n$ and $\overline{U}(Y, K_n, 2)^n$. Since the moments of noise are locally bounded, $\frac{1}{n}[Y;4]_{\mathcal{G}} - \frac{3}{2n^2}[Y,Y]_{\mathcal{G}}^2$ and $\widehat{\eta}$ are always finite. Following Theorems 5 and 6, we have the following corollary:

Corollary 3. *Assume Assumptions 1 and 2 and 5–7 as well as Assumptions 8 and 9. Adopt the choice of tuning parameters in Theorems 5 and 6, we have*

$$V(Y, K_n, s_n, 2)^n = O_p\left(s_n \cdot \frac{K_n}{n^{1/2}} \cdot K_n^{1/2}\right)$$
$$\overline{V}(Y, K_n, 2)^n = O_p\left(\frac{K_n}{n^{1/2}} \cdot K_n^{1/2}\right)$$

on the event $\widehat{\eta} \neq 0$. Besides, we have

$$N(Y, K_n)^n = O_p(K_n^{1/2})$$

on the event $[Y;4]_{\mathcal{G}} - \frac{3}{2n}[Y,Y]_{\mathcal{G}}^2 \neq 0$.

Recall the choices of tuning parameter, $K_n \to \infty$ for $N(Y, K_n)^n$, $K_n/n^{1/2} \to \infty$ for $V(Y, K_n, s_n, 2)^2$ and $\overline{V}(Y, K_n, 2)^2$, their asymptotic powers attain the optimal. As in finite samples, $\overline{V}(Y, K_n, 2)^n$ has more statistical power than $N(Y, K_n)^n$ by a factor of magnitude $K_n/n^{1/2}$; $V(Y, K_n, s_n, 2)^n$ is more powerful than $\overline{V}(Y, K_n, 2)^n$ by a factor of magnitude s_n.

5. A user's guide of stationarity tests

We currently have 3 complementary tests, namely $N(Y, K_n)^n$, $V(Y, K_n, s_n, 2)^n$ and $\overline{V}(Y, K_n, 2)^n$; each test has its own advantages as well as disadvantages. In this subsection, we are going to discuss their strength and weakness, and how to choose the optimal test for different circumstances.

(1) The first test $N(Y, K_n)^n$ divides the sample into 3 periods and compares the noise level in the middle with those on the edges. If we are interested in possible daily diurnal noise patterns, for example, let us test whether the noise level is higher in opening and closing trading hours, the best choice is to apply $N(Y, K)^n$ on 1-day high-frequency data. However, $N(Y, K_n)^n$ is not sensitive to local changes, for example, in case of a periodic change and the data sample covers several cycles, $N(Y, K_n)^n$ will likely misjudge the non-stationarity fact;
(2) The second test uses moving local windows each containing $s_n K_n$ observation, and compares noise levels in the edges and the middle of each local window; the third test also uses local windows but compares the noise level in one local widow with those in neighboring windows. Because they conduct test locally and aggregate local evidences, $V(Y, K_n, s_n, 2)^n$ and $\overline{V}(Y, K_n, 2)^n$ are more powerful in detecting local noise changes which $N(Y, K_n)^n$ could probably ignore. However, if the noise transition goes very smoothly but there is a systematic paradigm shift on a global scale, $V(Y, K_n, s_n, 2)^n$ and $\overline{V}(Y, K_n, 2)^n$ might lead to false stationarity conclusion;
(3) As said in Section 4.3, $V(Y, K_n, s_n, 2)^n$ has a smaller edge effect than $\overline{V}(Y, K_n, 2)^n$ hence $\overline{V}(Y, K_n, 2)^n$ is more a accurate test under the null hypothesis; whereas $V(Y, K_n, 2)^n$ enjoys larger statistical power (the lower right panel in Fig. 1). The intuition is that by construction the focus of $\overline{V}(Y, K_n, 2)^n$ is too local although it results in the smaller edge effect, which turns into its disadvantage when the noise is non-stationary.

As a simulation comparison, Fig. 1 shows averaged p-values computed from simulated 1-day/multi-day data with stationary/non-stationary noises. Fig. 2 shows their ROC curves. The simulation configuration is described in Section 8.1, and each p-values shown is the average of 3000 Monte Carlo samples.

As a summary, we list different considerations about the optimal choice of these tests in Table 1. We suggest some choices of the tuning parameters (K_n, s_n) to balance various errors in the high-frequency approximation. Table 2 shows the convergence rates and statistical powers of our tests under the suggested tuning parameters.

6. Measuring aggregate liquidity risks

6.1. A notion of "aggregate liquidity risk"

On one hand, $[g, g]$ as the quadratic variation of $\{g_t\}_{t\geq 0}$ over $(0, \mathcal{T}]$ is a reasonable measure of the "aggregate" variation of the process $\{g_t\}_{t\geq 0}$. On the other hand, microstructure noise variance g_{t_i} is a measure of market quality (Hasbrouck, 1993), or more specifically, market liquidity (Aït-Sahalia and Yu, 2009). Hence, it should not be utterly unreasonable to interpret $[g, g]_{\mathcal{T}}$ as "*aggregate liquidity risks*". In this section, we are going to define a notion of "*aggregate liquidity risks*" and provide an estimator with an associated CLT.

6.2. Estimating aggregated liquidity risk $[g, g]$

Recall Theorem 6 and note that $\frac{n}{K_n^2}\int_0^{\mathcal{T}} h_t\,\mathrm{d}t \to 0$, $\frac{3n}{2K_n^2}\overline{U}(Y, K_n, 2)_{\mathcal{T}}^n$ is a consistent estimator of $[g, g]_{\mathcal{T}}$, i.e.

$$\frac{3n}{2K_n^2}\overline{U}(Y, K_n, 2)_{\mathcal{T}}^n \xrightarrow{\mathbb{P}} [g, g]_{\mathcal{T}}. \tag{35}$$

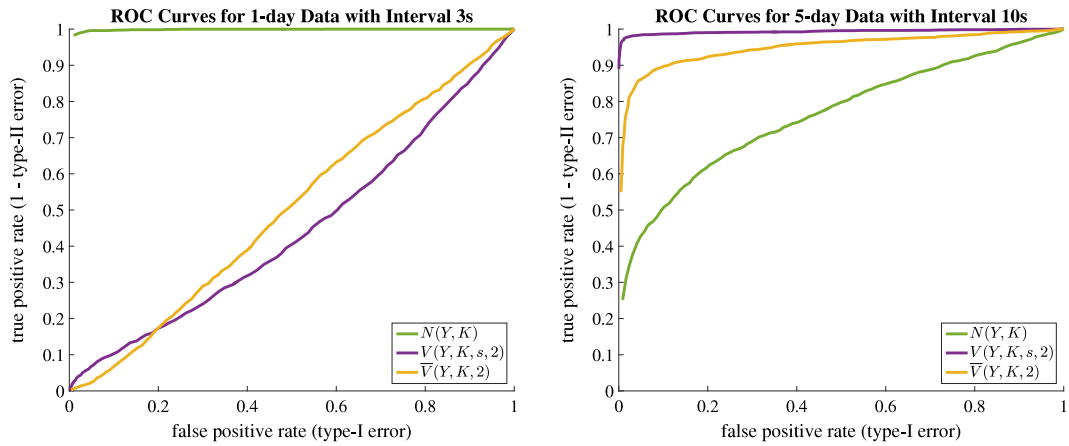


Fig. 2. ROC curves. ROC curves show how type-II changes as type-I error varies. Once again, the ROC curves indicates that when microstructure noise exhibits a daily diurnal pattern, $N(Y,K_n)^n$ is optimal for 1-day data, $V(Y,K_n,s_n,2)^n$ is optimal for multi-day data.

Table 1
Strengths and weakness of the tests.

Test statistics	$N(Y,K_n)^n_{\mathcal{T}}$	$V(Y,K_n,s_n,2)^n_{\mathcal{T}}$	$\overline{V}(Y,K_n,2)^n_{\mathcal{T}}$
Type-I error control	Most accurate	Least accurate	Moderately accurate
Strength in detection[a]	Global change	Local change	Local change (suboptimal)
Length requirement	1/multi-day data	Multi-day data	Multi-day data
Frequency requirement[b]	$\leq$20 s/$\leq$10 s	$\leq$60 s	$\leq$50 s
Computational cost	Relative small	Relative large	Relative large

[a] Evaluated in terms of statistical power.

[b] The minimal thresholds are expressed in terms of averaged time gap between consecutive observations. They are estimated from our simulation whose configuration is fairly realistic (Section 8.1).

Table 2
Rates of convergence and statistical powers of the tests.

	$N(Y,K_n)^n$	$V(Y,K_n,s_n,2)^n$	$\overline{V}(Y,K_n,2)^n$
K_n	$n^{1/2}$	$n^{2/3}$	$n^{7/12}$
s_n	–	$n^{1/6}$	–
Rates under H_0	$n^{1/4}$	$n^{1/6}$	$n^{5/24}$
Magnitudes under H_1	$n^{1/4}$	$n^{2/3}$	$n^{3/8}$

However, We can rewrite (34) as

$$\sqrt{\frac{n}{K_n}}\left(\frac{3n}{2K_n^2}\overline{U}(Y,K_n,2)^n_{\mathcal{T}}-[g,g]_{\mathcal{T}}\right)-\frac{3n^{3/2}}{K_n^{5/2}\mathcal{T}}\int_0^{\mathcal{T}}h_t\,\mathrm{d}t \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(0,\frac{3\mathcal{T}}{2}\int_0^{\mathcal{T}}(\sigma_t^{(g)})^4\,\mathrm{d}t\right)$$

depending the relation between the number of blocks and the number of observations within each block, we have different second-order properties. If we let $\frac{K_n}{n^{3/5}}\to\infty$, we have an unbiased central limit theorem for estimating $[g,g]$. Otherwise, in case $K_n \asymp n^{3/5}$ (or $K_n/n^{3/5}\to 0$), we have a CLT with a finite (or diverging) bias.

Corollary 4. [14] *Assume Assumptions 1 and 2 and 5–7 as well as Assumptions 8 and 9. Let $\frac{K_n}{n^{3/5}}\to\infty$, $\frac{K_n}{n^{2/3}}\to 0$, then we have*

$$\sqrt{\frac{n}{K_n}}\left(\frac{3n}{2K_n^2}\overline{U}(Y,K_n,2)^n_{\mathcal{T}}-[g,g]_{\mathcal{T}}\right) \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(0,\frac{3\mathcal{T}}{2}\int_0^{\mathcal{T}}(\sigma_t^{(g)})^4\,\mathrm{d}t\right). \tag{37}$$

Remark 5. Toward a better finite-sample performance, for example, to get a more accurate confidence interval for the aggregate liquidity risk, we suggest to use the estimate of

$$\left(\underbrace{\frac{3\mathcal{T}}{2}\int_0^{\mathcal{T}}(\sigma_t^{(g)})^4\mathrm{d}t}_{\text{due to discretization (non-vanishing)}}+\underbrace{\frac{54n^2}{K_n^4\mathcal{T}}\int_0^{\mathcal{T}}\left[h_t^2-h_tg_t^2+g_t^4\right]\mathrm{d}t}_{\text{due to noise (vanishing)}}\right)\times\frac{K_n}{n}$$

as the approximation to the finite-sample variance, in order to avoid the situation in which we underestimate the finite-sample variance and become overoptimistic about the accuracy of our estimate. The 95% confidence interval for our measure "*aggregated*

[14] Corollary 4 in some sense is an extension of the "integral-to-spot device" in Mykland and Zhang (2016): for a semimartingale $\{\theta_t\}$ on $[0,\mathcal{T}]$, let $\Theta_{(T_i,T_{i+q}]}=\int_{T_i}^{T_{i+q}}\theta_t\,\mathrm{d}t$, and $\mathrm{QV}_q(\Theta)=\frac{1}{q}\sum_{i=q}^{r_n-q}\left(\theta_{(T_i,T_{i+q}]}-\theta_{(T_{i-q},T_i]}\right)^2$, then under some regularity conditions (to guarantee standard stable convergence plus additional restriction on edge effects), as $q\to\infty$ and $q\Delta T\to 0$,

$$\frac{1}{(q\Delta T)^2}\mathrm{QV}_q(\Theta)\xrightarrow{\mathbb{P}}\frac{2}{3}[\theta,\theta]_{\mathcal{T}-}.$$

Define $G_i\equiv\int_{T_{i-1}}^{T_i}g_t\,\mathrm{d}t$ and $\widehat{G}_i\equiv\frac{\Delta T}{2K_n}[Y,Y]_{\mathcal{S}_i}$. Under some regularity conditions, according to the "integral-to-spot device" in Mykland and Zhang (2016)

$$\frac{3}{2(\Delta T)^2}\sum_{i=1}^{\lfloor n/K_n\rfloor}(G_i-G_{i-1})^2\xrightarrow{\mathbb{P}}[g,g]_{\mathcal{T}}.$$

However, we do not know the true values of G_i's in application, after swapping G_i for $\widehat{G}_i$,

$$\frac{3}{2(\Delta T)^2}\sum_{i=1}^{\lfloor n/K_n\rfloor}\left(\widehat{G}_i-\widehat{G}_{i-1}\right)^2\xrightarrow{\mathbb{P}}[g,g]_{\mathcal{T}-}+(\text{possibly additional terms}). \tag{36}$$

Note that $\frac{n}{K_n^2}\overline{U}(Y,K_n,2)^n_{\mathcal{T}}=\frac{1}{(\Delta T)^2}\sum_{i=1}^{\lfloor n/K_n\rfloor-1}\left(\widehat{G}_i-\widehat{G}_{i-1}\right)^2$, Corollary 4 reveals the possible additional terms and provides the central limit theorem associated with (36). Upon choosing K_n appropriately, the additional terms on the right side of (36) is zero and we have an unbiased central limit theorem.

liquidity risk" is

$$\left[\frac{3n}{2K_n^2}\overline{U}(Y,K_n,2)_{\mathcal{T}}^n - 1.96\times\widehat{\Gamma},\ \frac{3n}{2K_n^2}\overline{U}(Y,K_n,2)_{\mathcal{T}}^n + 1.96\times\widehat{\Gamma}\right] \tag{38}$$

where

$$\widehat{\Gamma}^2 = \frac{27}{128\,l_n^2K_n^4}\sum_{i=1}^{\lfloor n/K_n\rfloor - l_n}\left[\sum_{j=1}^{l_n}\left([Y,Y]_{\mathcal{S}_{i+j}} - [Y,Y]_{\mathcal{S}_{i+j-1}}\right)^2\right]^2$$
$$+\frac{27}{8\,K_n^2}\sum_{i=1}^{\lfloor n/K_n\rfloor}\left(\frac{4}{K_n^2}[Y;4]_{\mathcal{S}_i}^2 - \frac{14}{K_n^3}[Y;4]_{\mathcal{S}_i}[Y,Y]_{\mathcal{S}_i}^2 + \frac{13}{K_n^4}[Y,Y]_{\mathcal{S}_i}^4\right)$$

and $l_n \asymp \sqrt{n/K_n}$.

7. Noise functional dependency and model extension

The law of microstructure noise is represented via a Markov kernel $Q_t(\omega^{(0)}, \mathrm{d}y)$ for each time t, through which the second moment of the noise evolves according to a random function in time $g_t(\omega^{(0)}) = E(\epsilon_t^2|\mathcal{F}^{(0)})(\omega^{(0)}) = \int_{\mathbb{R}}|y - Z_t(\omega^{(0)})|^2\,Q_t(\omega^{(0)}, \mathrm{d}y)$ on the probability space $(\Omega^{(0)}, \mathcal{F}^{(0)}, \{\mathcal{F}_t^{(0)}\}_{t\geq 0}, \mathbb{P}^{(0)})$. The random function $g_t(\omega^{(0)})$ could depend on various latent variables, and more generally the form of $Q_t(\omega^{(0)}, \mathrm{d}\omega^{(1)})$ allows a wide range of correlation structures between the efficient price process and the microstructure noise. In this section, we discuss an elementary empirical evidence about the dependence of $g_t(\omega^{(0)})$ on $\sigma_t^2(\omega^{(0)})$ and the implication of the violation of our *identification assumption* i.e. Assumption 4.

7.1. Regression: market microstructure noises and spot volatilities

In this subsection, we go beyond the recognition that the second moment of microstructure noise is evolving over time, document the dependence of g_t on volatility σ_t^2. In doing so, we conduct time series linear regression of g_{t_i}'s on the latent variables $\sigma_{t_i}^2$'s.[15] We assume that the latent market microstructure noise variance and the latent volatility are correlated:

Assumption 10. With probability 1, we have $\forall t \geq 0$,

$$g_t = \beta\sigma_t^2 + \alpha + \zeta_t \tag{39}$$

where $\zeta_t \perp\!\!\!\perp \sigma_t^2$.

Since both g_t and σ_t^2 are unobservable, we need some preliminary estimates for both variables. Here, we use scaled sample-weighted TSRV and realized variance calculated from local samples to estimate spot volatilities $\sigma_{t_i}^2$'s and local noise levels, respectively, i.e., $\widehat{\sigma}_{\tau_{i-1}}^2 = \frac{1}{\tau_i-\tau_{i-1}}\widehat{\langle X,X\rangle}_{(\tau_{i-1},\tau_i]}^{(WTSRV)}$ and $\widehat{g}_{\tau_{i-1}} = \frac{1}{2|\mathcal{H}_i|}[Y,Y]_{\mathcal{H}_i}$, where $\{\tau_0, \tau_1, \ldots\} \subset \mathcal{G}$, $\mathcal{H}_i = \mathcal{G}\cap(\tau_{i-1},\tau_i]$, $|\mathcal{H}_i|$ is the cardinality of $\mathcal{H}_i$. Then, we can conduct linear regression on these pairs of volatility-noise estimates $(\widehat{g}_{\tau_i}, \widehat{\sigma}_{\tau_i}^2)$'s by ordinary least squares:

$$\widehat{g}_{\tau_i} = \widehat{\beta}_m\widehat{\sigma}_{\tau_i}^2 + \widehat{\alpha}_m + \eta_{\tau_i}^{(m)} \tag{40}$$

where m is the number of observation in the small time interval $(\tau_i, \tau_{i+1}]$, and $\eta_t^{(m)}$ denotes a component in the noise variance not captured by the volatility estimator $\widehat{\sigma}_{\tau_i}^2$. Besides, we use m in the subscripts of estimators $\hat{\alpha}_m$ and $\hat{\beta}_m$ to emphasize that the values of the estimators in (40) depend on the sample size n, and the distribution of $\eta_t^{(m)}$ also depends on m.

Lemma 3. *Suppose Assumptions 1 and 3–7 ae well as Assumption 10 hold, let* $\min|H_i| \to \infty$ *and* $\max|H_i|/n \to 0$, *then*

$$\widehat{\beta}_m \xrightarrow{\mathbb{P}} \beta$$
$$\widehat{\alpha}_m \xrightarrow{\mathbb{P}} \alpha.$$

By Lemma 3, if there is a linear relationship between the noise variance and the spot volatility, the regression (40) provides consistent estimates of linear coefficients. Fig. 3 shows the least square regression plots for high-frequency transaction data in April, 2013 of 6 stocks: International Business Machines (IBM), Goldman Sachs (GS), Johnson & Johnson (JNJ), Nike, Inc. (NKE), Chevron Corporation (CVX), McDonald's (MCD).

The time series regression and empirical analysis here are preliminary. One can investigate the statistical properties of this type of linear regression in more detail. Perhaps, there are non-linear relations. These issues will be addressed in our future research.

7.2. Model extension: endogenous noise

In our model, we allow arbitrary fashion of the noise process up to the time-varying Markov kernel $Q_t(\cdot,\cdot)$ plus the identification assumption (Assumption 4). As documented in Jacod et al. (2009), the identification assumption is restrictively strong. If one is interested in the stationarity of $\{\epsilon_t\}_{t\geq 0}$, our methods are valid regardless the identification assumption holds or not. However, if one is concerned about $\{e_t\}_{t\geq 0}$ our methods will break down when the identification assumption is violated. Nevertheless, this extension is indispensable for empirically compatible modeling and it allows endogenous microstructure noise (noise which is correlated with the efficient price (Hansen and Lunde, 2006)).

Note that in Section 2.1, conditioning on all latent variables, ϵ_t is a mean-zero random variable, i.e., $\int_{\mathbb{R}}(y - Z_t(\omega^{(0)}))\,Q_t(\omega^{(0)}, \mathrm{d}y) = 0$ since $\int_{\mathbb{R}} y\,Q_t(\omega^{(0)}, \mathrm{d}y) = Z_t(\omega^{(0)})$. However, the conditional mean of e_t is not necessarily 0 since $E(e_t|\mathcal{F}^{(0)}) = E(Y_t - X_t|\mathcal{F}^{(0)}) = Z_t - X_t$.

This observation enables us to, non-parametrically, introduce endogenous noise into our model. We can allow instantaneous/realized correlation between the latent process $\{X_t\}_{t\geq 0}$ and the noise process $\{e_t\}_{t\geq 0}$. Although $E(e_t|\mathcal{F}^{(0)})$ is not necessarily 0, we assume the unconditional mean $E_{\mathbb{P}}(e_t)$ is zero, then calculation shows

$$\begin{aligned}
\mathrm{Cov}(X_t, e_t) &= E_{\mathbb{P}^{(0)}}[X_tZ_t] - E_{\mathbb{P}^{(0)}}[X_t^2]\\
\mathrm{Cov}(Z_t, e_t) &= E_{\mathbb{P}^{(0)}}[Z_t^2] - E_{\mathbb{P}^{(0)}}[X_tZ_t]\\
\mathrm{Cov}(X_t, \epsilon_t) &= 0\\
\mathrm{Cov}(Z_t, \epsilon_t) &= 0.
\end{aligned}$$

Jacod et al. (2009) assumed $Z_t = X_t$, so there is no endogenous noise in their model. However, as long as $E_{\mathbb{P}^{(0)}}[X_tZ_t] \neq E_{\mathbb{P}^{(0)}}[X_t^2]$, there is correlation between the latent process $\{X\}_{t\geq 0}$ defined by (1) and the noise process $\{e_t\}_{t\geq 0}$ defined by (2).

An intuitive interpretation is that e_t encodes some relevant information about the processes defined on the latent probability space if it is correlated with the latent random variables X_t and Z_t. In contrast, ϵ_t is a pure noise and conveys no useful information about the latent processes, the correlation between ϵ_t and any latent random variable is zero. For this reason, we call e_t "*endogenous microstructure noise*", and call ϵ_t "*exogenous microstructure*

[15] Aït-Sahalia and Sağlam (2016) provide a theoretical underpinning for the negative correlation between volatility and liquidity, equivalently, positive correlation between σ_t^2 and g_t: a higher volatility indicates a higher risk that arbitrageurs might take advantage of market makers' previous orders to act against market makers; hence, it reduces the activities of liquidity provision.

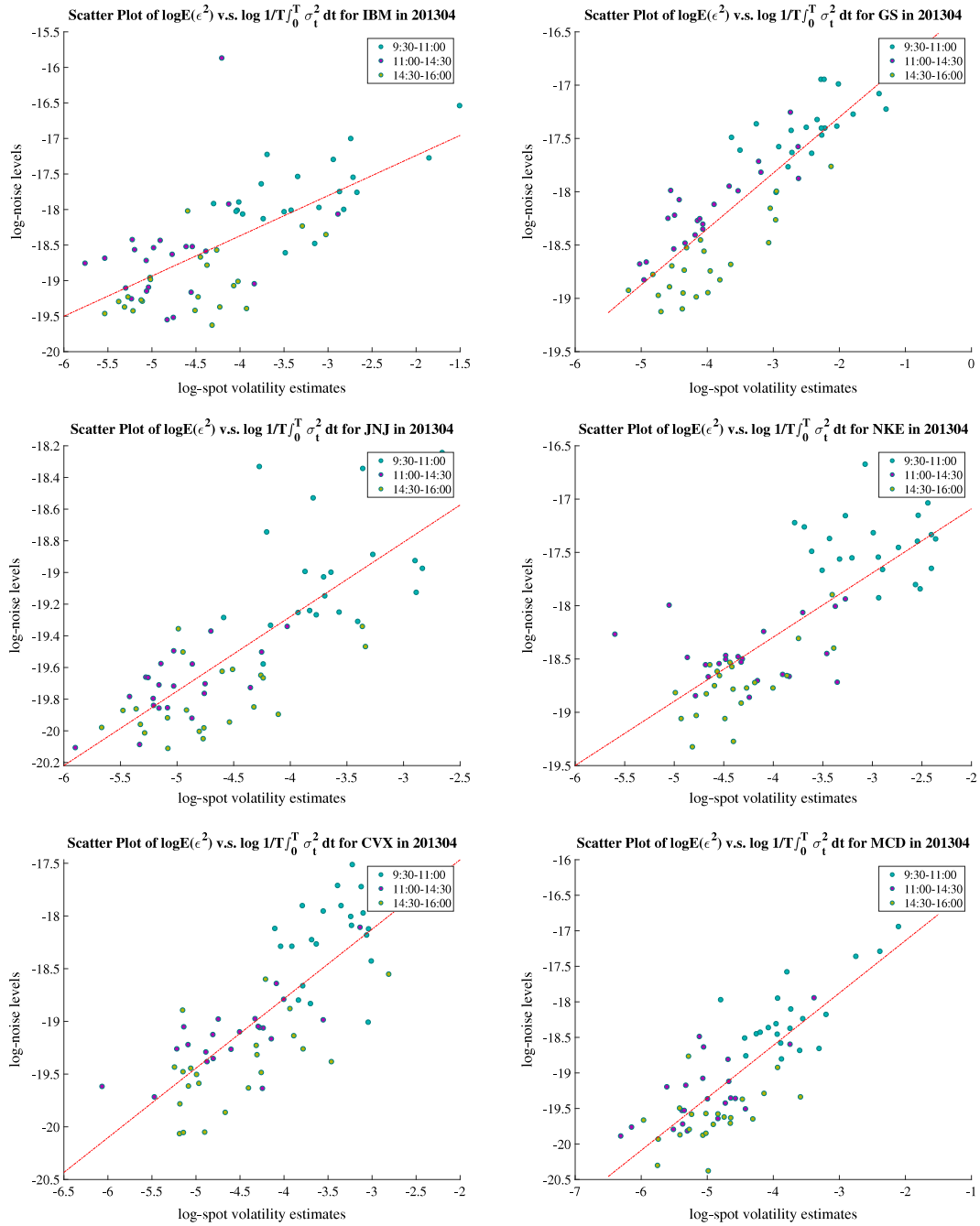
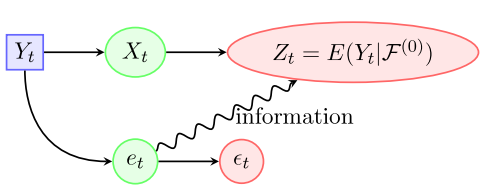


Fig. 3. Scatter plot of $\log(\widehat{g}_{\tau_i})$ against $\log(\widehat{\sigma}^2_{\tau_i})$, where each τ_i represents a particular period in each day. The red dotted lines are the fitted regression lines for IBM (IT), GS (finance), JNJ (medicine and pharmacy), NKE (manufacturing), CVX (energy), MCD (fast food), from left to right and top to bottom, respectively.

noise".



Remark 6. When one tries to estimate the integrated volatility, the quantity which is actually estimated is $\langle Z, Z\rangle_T$, not necessarily the usually desired target $\langle X, X\rangle_T$. This is discussed by Li and Mykland (2007). In contrast to Jacod et al. (2009), we do not assume $\int_{\mathbb{R}} y\, Q_t(\omega^{(0)}, dy) = X_t(\omega^{(0)})$. In other words, in the case where $Z_t \neq X_t$, the integrated volatility $\langle X, X\rangle_T$ is not identifiable; however, if we are satisfied with estimating $\langle Z, Z\rangle_T$, then we are able to introduce some conditional correlation between the efficient price and the microstructure noise.

One conceptual finding from the model extension is the informational content in microstructure noise $\{e_t\}_{t\geq 0}$ with respect to the efficient price in financial term (or latent process in statistical term) which is modeled as an Itô semimartingale. The interpretation comes from market microstructure theory (O'Hara, 1995, 2003). As in the classical asset pricing theory, we take the price as given and exogenous, and conduct trading and hedging strategies, portfolio allocation and risk management. But, the price discovery and price formation depend on the behaviors of market participants, no price will be produced without investment activities of various market participants. It is the balance between demand and

supply from investors, it is the psychology of people in the market, it is the synthesis of microscopic effects of beliefs and behaviors of market participants, that determine the prices. Thus, the efficient price should be an endogenous process in the financial market. It is one of striking difference between asset pricing and market microstructure theory: the classical asset pricing theory assumes frictionless and competitive market in which people do not have to worry about the price impact and liquidity constraint. While, in market microstructure theory, the modelers need to look inside the "black box" of the trading processes, and take market making, price discovery, liquidity formation, inventory control, asymmetric information into account.

Since we consider the price as endogenous, which, for example, affected by transaction costs (like bid–ask spread), inventory control, discrete adjustment of price, lagged incorporation of new information, insider trading and adverse selection brought by asymmetric information, lack of liquidity caused by one or several of the factors mentioned above, the Itô process is merely an approximation to the efficient price observed at high-frequency, at which market microstructure effects manifest itself to such extent that the accumulated noise swamps the integrated volatility of the latent Itô process and the variation in microstructure noise dominates the total variance.

Therefore, it is reasonable (even indispensable) to extend our model to allow endogenous microstructure noise, at least from the viewpoint of microstructure theory, and for sake of realistic modeling at low-latency and millisecond level. This topic is not the focus of this paper; in-depth discussion and treatment on endogenous microstructure noise will be addressed in our future research.

8. Simulation

8.1. Simulation scenario

The configuration of our simulation design is

$$Y_{t_i} = \left\lfloor \frac{X_{t_i} + \epsilon_{t_i}}{\alpha} \right\rfloor \alpha \tag{41}$$

$$\mathrm{d}X_t = \mu\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t + J_t^X\,\mathrm{d}N_t^X \tag{42}$$

$$\mathrm{d}\sigma_t^2 = \kappa(\bar{\sigma}^2 - \sigma_t^2)\,\mathrm{d}t + \delta\sigma_t\,\mathrm{d}B_t + \sigma_{t-}J_t^V\,\mathrm{d}N_t^V \tag{43}$$

where $E(\mathrm{d}W_t \cdot \mathrm{d}B_t) = \rho\,\mathrm{d}t$, N^X and N^V are Poisson processes $\perp\!\!\!\perp W, B$ with parameters λ_X and λ_V, respectively, the jump sizes satisfy $J^X \sim N(\theta_X, \nu_X)$ and $J_t^V = e^z$ with $Z \sim N(\theta_V, \nu_V)$. The stationary microstructure noise behaves as $\epsilon_{t_i}^{(s)} \overset{\text{i.i.d.}}{\sim} N(0, a_0^2)$, whereas the non-stationary microstructure noises are distributed as

$$\begin{aligned} \epsilon_{t_i}^{(ns)} &= \sqrt{\frac{60}{17}}\left[\left(\frac{i}{n} - 0.5\right)^2 + 0.2\right]^{\frac{1}{2}} \times e_{t_i} \\ e_{t_i} &= z_i + \sum_{j=1}^{M} \binom{u+j-1}{j} z_{i-j} \\ z_k &\overset{\text{i.i.d.}}{\sim} N\left(0, \omega^2\right), \quad \omega^2 = a_1\left(\frac{1}{n}\sum_{j=1}^{n}\sigma_{it_j}^4\right)^{\frac{1}{2}} \end{aligned} \tag{44}$$

where $u \in (-0.5, 0.5)$ and n is the number of high-frequency observations in 1 business day. In (44), the noise variance of $\{\epsilon_{t_i}^{(ns)}\}_i$ changes according to a U-curve, which means that the noise is of relatively higher levels around opening and closed hours. The U-curve is chosen such that the averaged noise variance within a day is ω^2. The noise conforms to the empirical feature that the variance of microstructure noise increases with the level of volatility (Bandi and Russell, 2006). The parameters are chosen so that they are consistent with Aït-Sahalia and Yu (2009):

X parameters	X_0	μ	ρ	λ_X	θ_X	ν_X
	$\ln(100)$	0.03	−0.6	6	0.0016	0.004
σ parameters	κ	$\bar{\sigma}^2$	δ	λ_V	θ_V	ν_V
	6	0.16	0.5	12	−5	0.8
Noise parameters	a_0	a_1	α	M	u	
	5×10^{-3}	1.54×10^{-4}	1×10^{-5}	10	0.3	

Furthermore, σ_0^2 is sampled from the stationary distribution of Cox–Ingersoll–Ross process (Cox et al., 1985), i.e., Gamma $\left(\frac{2\kappa\bar{\sigma}^2}{\delta^2}, \frac{\delta^2}{2\kappa}\right)$ so the unconditional mean of the volatility is $\bar{\sigma}^2$. a_1 is chosen such that $\mathrm{Var}(\epsilon^{(s)}) = \mathrm{Var}(\epsilon^{(ns)})$ in average. We also adopted a random sampling scheme according to an inhomogeneous Poisson process Poisson$(\lambda_t \times \Delta)$ where Δ is averaged sampling duration and the trading intensity evolves periodically $\lambda_t = 1 + 0.5 \times \cos(2\pi t/T)$ with T being the length of 1 business day.

8.2. Simulation results

In Fig. 4, 5 and 6, we show the simulation results of $N(Y, K_n)_T^n$, $V(Y, K_n, s_n, 2)_T^n$ and $\overline{V}(Y, K_n, 2)_T^n$ where T is taken to be 1 business day (left panel in each figure) and 5 business days (right panel in each figure). For each test and each time span, the simulation is conducted in 2 different circumstances: stationary noise (upper picture in each column), U-shape noise (44) (lower picture in each column). The plots show various empirical densities function of our proposed tests against the density of $N(0, 1)$. Each group of tests were computed from 3000 sample paths with averaged sampling interval 1 s.

9. Empirical studies

9.1. Empirical evidence of non-stationary microstructure noise

Fig. 7 shows daily variations of microstructure noise levels in 2008. Fig. 8 exhibits intra-day variations in microstructure noises of individual stocks in the first 4 months of 2013.

9.2. Empirical tests

In this subsection, we apply our tests onto high-frequency financial transaction data of stocks. We take several components in Dow Jones Industrial Average (DJIA30): Intel Corporation (INTC), International Business Machines Corporation (IBM), Goldman Sachs (GS), JPMorgan Chase (JPM), Exxon Mobil Corporation (XOM), and WalMart (WMT). We compute the test statistics and the p-values for some stocks during the 22 business days in April, which is shown in Table 3. Besides, in Fig. 9, we plot the whole trend of the test statistics $N(Y, K)_T^n$ during the period January 3, 2006 to December 31, 2013 as measures of liquidity.

10. Conclusion

In this paper, we mainly concern hypothesis testing of microstructure noise stationarity in a hidden Itô semimartingale model. The null hypothesis is that microstructure noise is stationary, and the alternative hypothesis is that microstructure noise is non-stationary with arbitrage dynamics up to a Markov kernel. Our tests work in fairly general settings where the latent Itô semimartingale may have jumps with any degree of activity, the microstructure allows white noise and rounding error, and the observation times can be irregularly spaced.

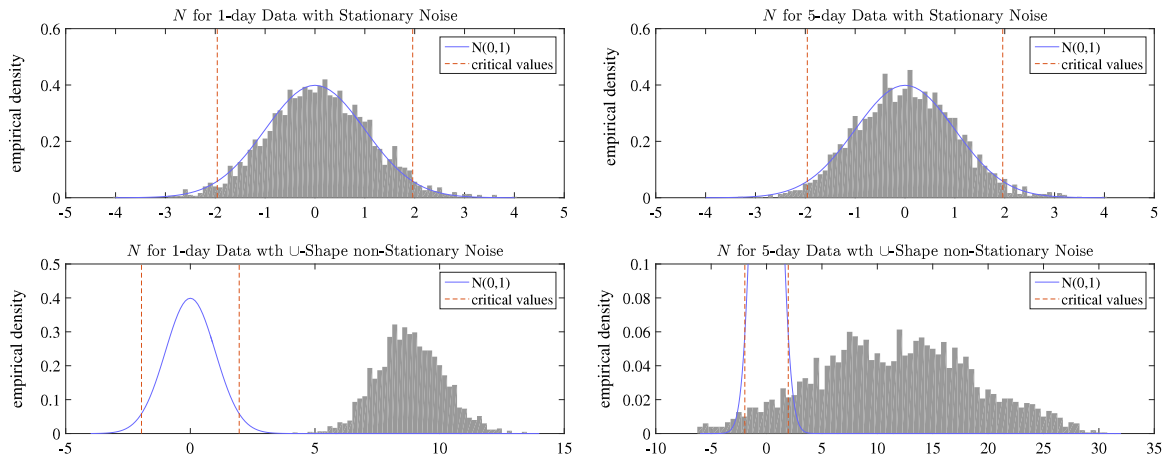


Fig. 4. Empirical density of $N(Y, K_n)^n$. These plots show the empirical densities of $N(Y, K_n)^n$ when it applies to 1-day/5-day data with stationary/non-stationary noises. Compared the simulation of other tests, we can see $N(Y, K_n)^n$ converges faster to $N(0, 1)$ when microstructure noise is stationary. On the other hand, if the microstructure noise is non-stationary and exhibits daily diurnal pattern, $N(Y, K_n)^n$ is the best for 1-day data.

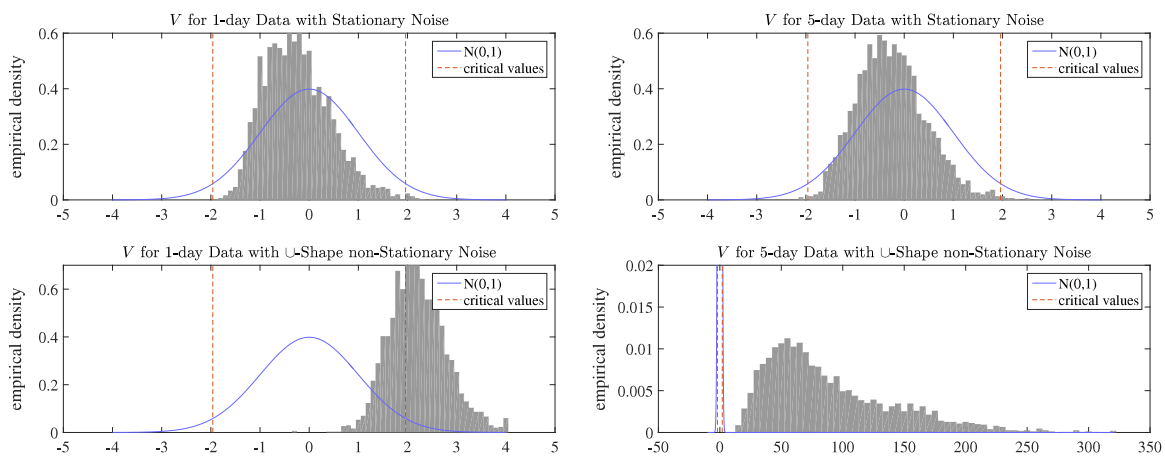


Fig. 5. Empirical density of $V(Y, K_n, s_n, 2)^n$. These plots show the empirical densities of $V(Y, K_n, s_n, 2)^n$ when it applies to 1-day/5-day data with stationary/non-stationary noises. Compared the simulation of other tests, we can see $V(Y, K_n, s_n, 2)^n$ is more conservative due to its relatively large edge effect when microstructure noise is stationary. On the other hand, if the microstructure noise is non-stationary and exhibits daily diurnal pattern, $V(Y, K_n, s_n, 2)^n$ is the best for multi-day data and enjoys the largest statistical power.

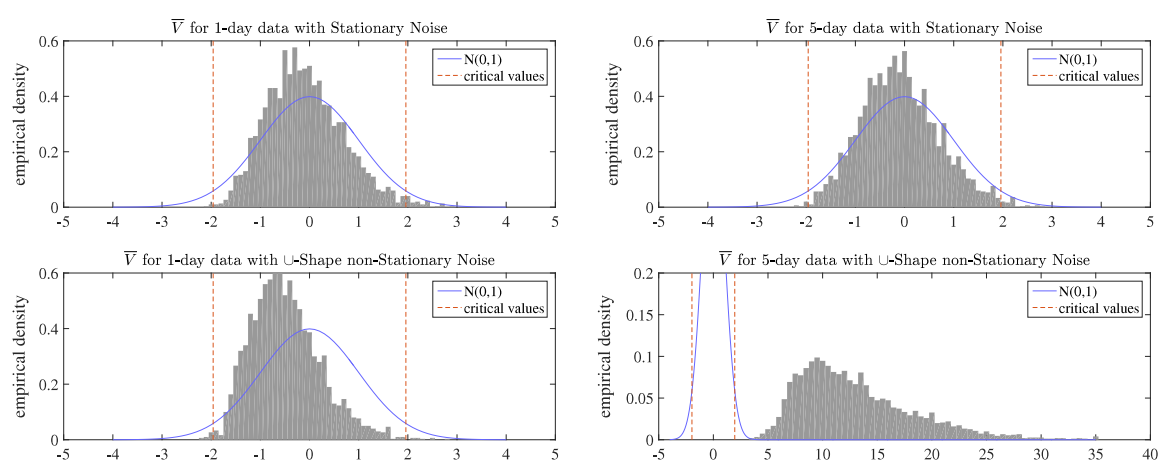


Fig. 6. Empirical density of $\overline{V}(Y, K_n, 2)^n$. These plots show the empirical densities of $\overline{V}(Y, K_n, 2)^n$ when it applies to 5-day/10-day data with stationary/non-stationary noises. Compared the simulation of other tests, we can see $\overline{V}(Y, K_n, 2)^n$ controls type-I error more accurately than $V(Y, K_n, s_n, 2)^n$ does when microstructure noise is stationary. On the other hand, if the microstructure noise is non-stationary and exhibits daily diurnal pattern, $N(Y, K_n)$ is better for multi-day data.

The first test is motivated by the behavior of the two-scaled estimator (TSRV) under contamination of non-stationary noise, whose negative impact can be eliminated by a modification of TSRV (Kalnina and Linton, 2008) under our general model. Based on the remedy for non-stationary microstructure noise, the first test $N(Y, K_n)^n$ is designed as a functional of volatility estimators, its

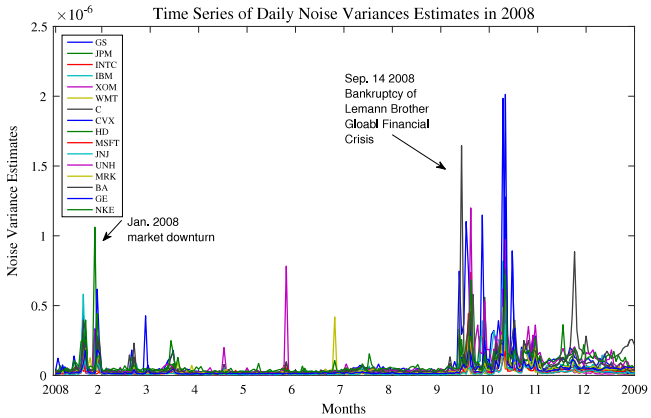


Fig. 7. Daily noise variance estimates in 2008, with a simple event-history analysis. During the turmoil of financial crisis, the market microstructure noise surged up, the quality of the market worsened strikingly.

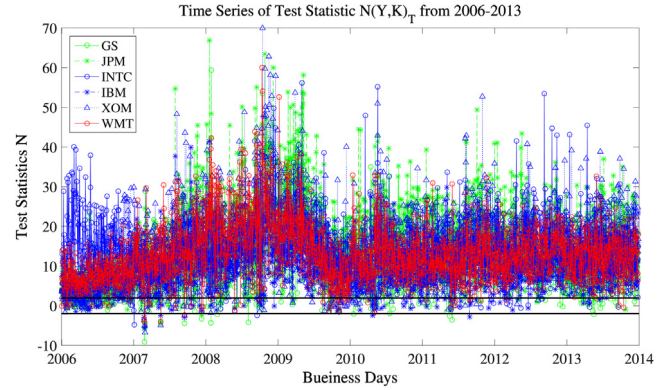


Fig. 9. The six time series of the first test statistic $N(Y, K_n)^n$ computed daily using intra-day ultrahigh frequency financial data. The black horizontal lines around zero are .025% and .975% quantiles of the standard normal distribution.

type-I error can be controlled by associated central limit theorem under the null hypothesis. We also demonstrated that $N(Y, K_n)^n$ explodes in high-frequency asymptotics when microstructure noise is non-stationary.

Besides, we have other complementary tests, namely $V(Y, K_n, s_n, 2)_T^n$ and $\overline{V}(Y, K_n, 2)^n$. They are defined as functionals of $N(Y, K_n)^n$'s and realized variances, respectively, which are computed in different local time windows. $V(Y, K_n, s_n, 2)^n$ and $\overline{V}(Y, K_n, 2)^n$ are asymptotically equivalent and share the same convergence rate under the null hypothesis. Asymptotic approximation to $\overline{V}(Y, K_n, 2)^n$ in finite sample is more accurate than that of $V(Y, K_n, s_n, 2)^n$ under the null hypothesis, however,

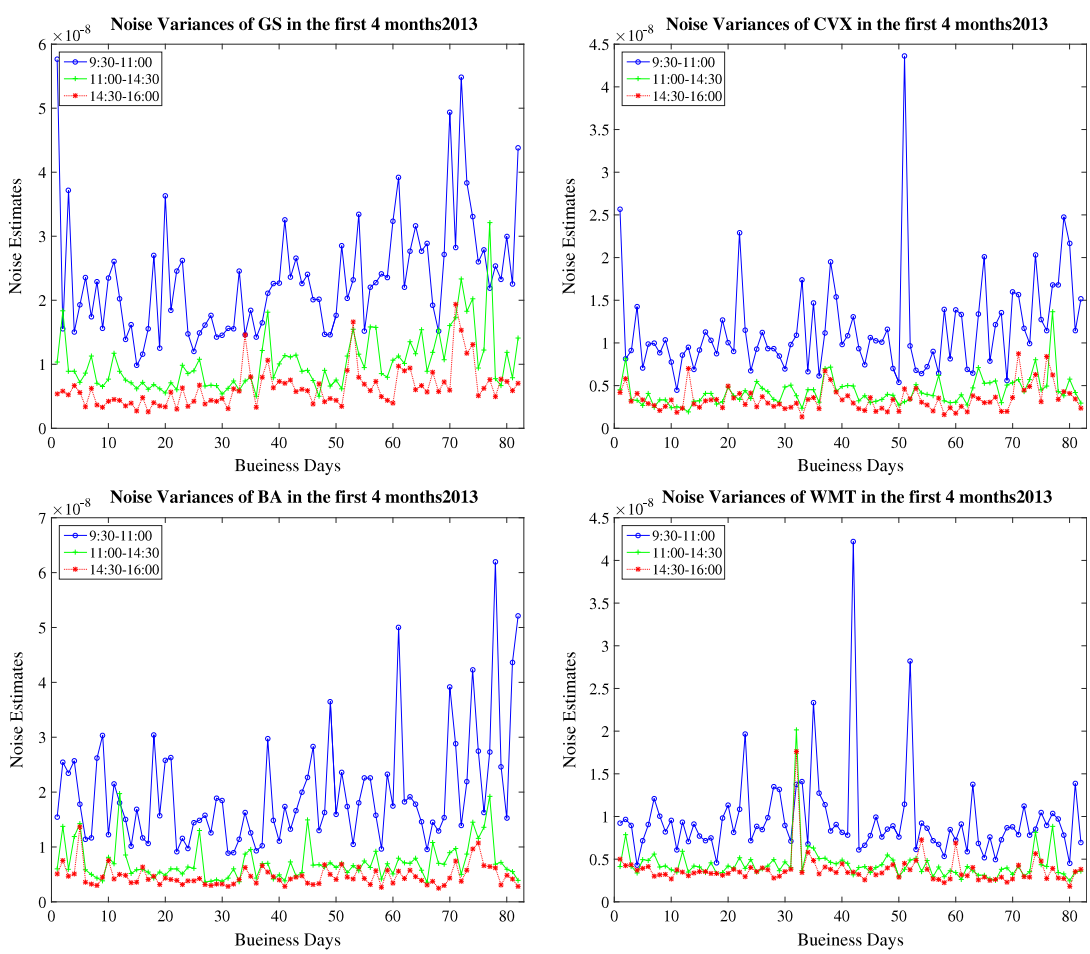


Fig. 8. Temporal dynamics of $\widehat{E(\epsilon_t^2)}$ in different segments of trading hours 9:30–16:00 EST. For example, the blue line is the time series plot of estimated noise level around morning across different business days in the first 4 months, 2013. From left to right and top to bottom, the companies are Goldman Sachs (GS), Chevron Corporation (CVX), Boeing (BA), Walmart (WMT).

Table 3
Test statistics $N(Y, K_n)_T$ for DJIA components (T is 1 business day).

Dates	IBM		XOM		INTC		GS		GE	
yyyy-mm-dd	$N(Y,K)_T$	p-value	$N(Y,K)_T$	p-value	$N(Y,K)_T$	p-value	$N(Y,K)_T$	p-value	$N(Y,K)_T$	p-value
2013-04-01	0.5942	0.2762	6.0114	9.1947e−10	17.3676	0	0.9125	0.1807	6.4765	4.6925e−11
2013-04-02	3.8894	5.0246e−05	16.7202	0	12.3133	0	8.6813	0	26.2744	0
2013-04-03	6.8579	3.4941e−12	11.1238	0	12.6015	0	9.9089	0	4.5688	2.4529e−06
2013-04-04	4.5851	2.2690e−06	11.7737	0	11.8105	0	7.4771	3.7970e−14	8.3468	0
2013-04-05	8.6943	0	19.6103	0	21.9399	0	13.0797	0	7.7996	3.1086e−15
2013-04-08	12.0086	0	10.2720	0	19.6533	0	12.0044	0	8.7725	0
2013-04-09	4.4107	5.1507e−06	4.2196	1.0152e−05	14.5840	0	3.9217	4.3971e−05	2.8118	0.0025
2013-04-10	10.7967	0	20.3985	0	12.4934	0	1.4729	0.0704	12.1430	0
2013-04-11	10.5358	0	8.4332	0	19.8102	0	5.5467	1.4557e−08	7.6796	7.9936e−15
2013-04-12	9.8741	0	18.8744	0	9.7960	0	10.4689	0	11.3813	0
2013-04-15	8.6767	0	37.0635	0	11.8791	0	5.0028	2.8247e−07	6.6791	1.2023e−11
2013-04-16	11.5517	0	25.8213	0	11.0252	0	5.5612	1.3384e−08	16.5744	0
2013-04-17	11.2338	0	4.2163	1.2419e−05	20.6048	0	5.5168	1.7261e−08	13.6559	0
2013-04-18	15.1748	0	14.7396	0	49.2313	0	3.1477	8.2284e−04	10.6347	0
2013-04-19	29.7852	0	18.3013	0	10.8806	0	9.7611	0	18.5074	0
2013-04-22	13.4899	0	7.0150	1.1479e−12	10.0430	0	7.5659	1.9207e−14	12.9960	0
2013-04-23	11.0911	0	0.9798	0.1636	1.5144	0.065	0.4083	0.3415	26.3066	0
2013-04-24	10.6420	0	26.4967	0	22.6824	0	9.6762	0	20.8122	0
2013-04-25	12.8092	0	13.4558	0	15.4190	0	9.4322	0	9.3956	0
2013-04-26	7.1480	4.4031e−13	14.8469	0	15.1904	0	3.0896	0.0010	6.0681	6.4723e−10
2013-04-29	3.4021	3.3438e−04	19.2697	0	0.8441	0.1993	9.3275	0	−00481	0.4808
2013-04-30	0.4047	0.3428	12.8344	0	−02676	0.3945	10.1050	0	7.4785	3.7637e−14

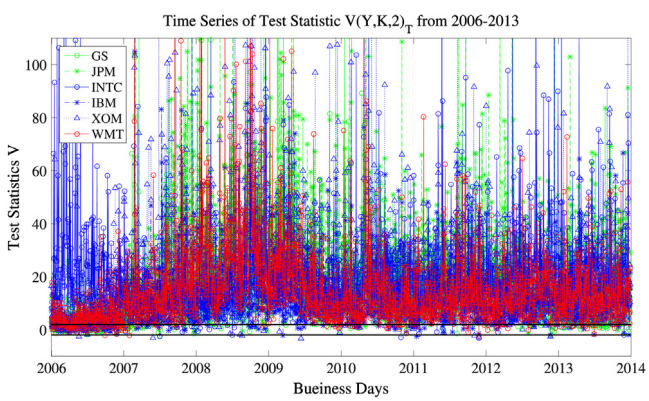


Fig. 10. The six time series of the second test statistic $V(Y, K_n, s_n, 2)^n$ computed daily using intra-day ultrahigh frequency financial data. The black horizontal lines around zero are .025% and .975% quantile of the standard normal distribution.

$V(Y, K_n, s_n, 2)^n$ has more advantage under the alternative hypothesis in that it has a larger statistical power. Compared to $N(Y, K_n)^n$ which is more suited for 1-day data, $V(Y, K_n, s_n, 2)^n_T$ and $\overline{V}(Y, K_n, 2)^n$ are more suited for multi-day data. How to choose these complementary tests are discussed in detail.

Since microstructure noise could be a measure of the market quality (market liquidity, market depth, etc.) (Hasbrouck, 1993; O'Hara, 2003; Aït-Sahalia and Yu, 2009), our test statistics can be measures of liquidity risk. Particularly, assuming microstructure noise variance evolves like an Itô diffusion, not only the asymptotic distributions of the test statistics under the alternative hypothesis are available, but also a notation of "aggregate liquidity risk" and a consistent estimator with an associated central limit theorem.

Some high-frequency financial data from NYSE are analyzed using the tests. As some DJIA components from 2006 to 2013 shows, variances of microstructure noise indeed changed both daily and intra-daily, which agrees with the empirical literature. Moreover, we find that the timing of the sudden increase in noise variance in Sep. 2008 coincided with the beginning of the global financial catastrophe triggered by the mortgage subprime crisis. The time series of our test statistics reveals a pattern which indicates increases in daily and weekly transaction costs during the financial turmoil.

Appendix

All the calculations are conditional on $\mathcal{F}^{(0)}$. Assuming Proposition 1 and Lemmas 4 and 5 which can also be found in Zhang et al. (2005) and Li and Mykland (2007):

Proposition 1. *Assume that $E(|A_n||\mathcal{F}^{(0)})$ is $O_p(1)$. Then, A_n is $O_p(1)$.*

Lemma 4. *Under the model (1), (3) and (4), and the assumptions in Section 2.3, we have*

$$[Y, Y]_{\mathcal{G}} = [\epsilon, \epsilon]_{\mathcal{G}} + O_p(1) \tag{45}$$

$$[Y, Y]_T^{(avg,K)} = [Z, Z]_T^{(avg,K)} + [\epsilon, \epsilon]_T^{(avg,K)} + O_p(1/\sqrt{K}). \tag{46}$$

Besides, define $\mathcal{G}_{K+1}^{(\min)}$ as the right immediate neighbor of $\max \mathcal{G}^{(\min)}$ in the full grid $\mathcal{G}$, and define $\mathcal{G}_0^{(\max)}$ as the left immediate neighbor of $\min \mathcal{G}^{(\max)}$ in the full grid $\mathcal{G}$.

In order to describe the edge effect and the behaviors of some test statistics below, we need to introduce some random variables:

$$\begin{aligned} M_T^{(1)} &\equiv \frac{1}{\sqrt{n}} \sum_{i=0}^{n} \left(\epsilon_{t_i}^2 - g_{t_i}\right) \\ M_T^{(2)} &\equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_{t_i}\epsilon_{t_{i-1}} \qquad M_T^{(3)} \equiv \frac{1}{\sqrt{n}} \sum_{k=0}^{K-1} \sum_{t_i \in \mathcal{G}'(k)} \epsilon_{t_i}\epsilon_{t_{i-K}} \end{aligned} \tag{47}$$

and denote $h_t(\omega^{(0)}) \equiv E(\epsilon_t^4|\mathcal{F}^{(0)})(\omega^{(0)})$. Note that $M_T^{(1)}$, $M_T^{(2)}$ and $M_T^{(3)}$ are the *end-points of martingales* with respect to filtration $\mathcal{F}_i = \sigma(\epsilon_{t_j}, j \le i; X_t, \forall t)$. By the argument from the Appendix A.2 in Aït-Sahalia et al. (2005), we have

Lemma 5. *$M_T^{(1)}$, $M_T^{(2)}$ and $M_T^{(3)}$ are asymptotically conditionally independent mixed normals, they have conditional variances $\frac{1}{T}\int_0^T h_t - g_t^2\,\mathrm{d}t$, $\frac{1}{T}\int_0^T g_t^2\,\mathrm{d}t$, $\frac{1}{T}\int_0^T g_t^2\,\mathrm{d}t$, respectively.*

A.1. Robustness to jumps in noise inference

In proving the testing theorems, namely, Theorems 2 and 3, Corollary 2 and Theorem 4, we can assume the $J_t = 0, \forall t > 0$ in (1)

without loss of generality, as long as the noise is uncorrelated with neither the continuous part nor the jump part. Under Assumptions 1 and 2, there are 3 components in the realized variance:

(1) finite quadratic variation of the discontinuous Itô semi-martingale $[X,X]_T = \langle X,X\rangle_T + \sum_{t\le T}|\Delta X_t|^2$, where $\Delta X_t = X_t - \lim_{s\nearrow t} X_s$ (a well-known result in stochastic calculus);
(2) variation due to noise, which is of order $O_p(n)$;
(3) asymptotically negligible terms, which are cross terms between noise, continuous martingale and jumps.

Under Assumption 5, by a similar argument to those in the proof of Lemma 1 in Li and Mykland (2007),[16] we have a result similar to Lemma 4:

$$[Y,Y]_{\mathcal{G}} = [\epsilon,\epsilon]_{\mathcal{G}} + \underbrace{2\sum_{i=1}^{n}(J_{t_i} - J_{t_{i-1}})(\epsilon_{t_i} - \epsilon_{t_{i-1}}) + [X,X]_T}_{O_p(1)} + O_p(1) \tag{48}$$

which suggests that normalized realized variance of the fastest time scale $\frac{1}{2n}[Y,Y]_{\mathcal{G}}$ is a consistent estimator of $E(\epsilon^2|\mathcal{F}^{(0)})$ provided the noise is stationary even if there exist jumps, i.e., Lemma 4 still holds. For this reason, the asymptotic distributions remain the same for the test statistics, even if jump is present.

A.2. Proof of Lemma 1

Proof. By our assumptions, we can write

$$\begin{aligned}[Y,Y]_T^{(avg,K)} &= [Z,Z]_T^{(avg,K)} + \frac{2\sqrt{n}}{K}\left(M_T^{(1)} - M_T^{(3)}\right)\\ &\quad + \frac{2}{K}\sum_{k=1}^{K}\sum_{t_i\in\mathcal{G}''(k)} g_{t_i} + \frac{1}{K}\sum_{t_i\in\mathcal{G}^{(\min)}} g_{t_i}\\ &\quad + \frac{1}{K}\sum_{t_i\in\mathcal{G}^{(\max)}} g_{t_i} + O_p\left(\frac{1}{\sqrt{K}}\right).\end{aligned}$$

By the conditional Lyapunov condition, and Lemma 5

$$M_T^{(1)} - M_T^{(3)} \xrightarrow{\mathbb{P}} \mathcal{MN}\left(0, \frac{1}{T}\int_0^T h_t\,dt\right)$$

thus (11) follows. □

A.3. Proof of Theorem 1

Proof. Asymptotically,[17] the new version of realized variance in Kalnina and Linton (2008) can be written as follows:

$$[Y,Y]_T^{\{n\}} = \frac{1}{2}[Y,Y]_{\mathcal{G}^{(\min)}} + [Y,Y]_{\bigcup_{k=1}^K \mathcal{G}''(k)} + \frac{1}{2}[Y,Y]_{\mathcal{G}^{(\min)}}.$$

Since for any grid $\mathcal{H}$, $[Y,Y]_{\mathcal{H}} = [Z,Z]_{\mathcal{H}} + 2[Z,\epsilon]_{\mathcal{H}} + [\epsilon,\epsilon]_{\mathcal{H}}$, we have

$$\begin{aligned}[Y,Y]_T^{\{n\}} &= [Z,\epsilon]_{\mathcal{G}^{(\min)}} + 2[Z,\epsilon]_{\bigcup_{k=1}^K \mathcal{G}''(k)} + [Z,\epsilon]_{\mathcal{G}^{(\max)}}\\ &\quad + \frac{1}{2}[\epsilon,\epsilon]_{\mathcal{G}^{(\min)}} + [\epsilon,\epsilon]_{\bigcup_{k=1}^K \mathcal{G}''(k)} + \frac{1}{2}[\epsilon,\epsilon]_{\mathcal{G}^{(\max)}} + O_p(1).\end{aligned}$$

Note that $[Z,\epsilon]_{\mathcal{G}^{(\min)}} + 2[Z,\epsilon]_{\bigcup_{k=1}^K \mathcal{G}''(k)} + [Z,\epsilon]_{\mathcal{G}^{(\max)}} \le 2[Z,\epsilon]_{\mathcal{G}}$. Define $\Delta Z_{t_i} = Z_{t_i} - Z_{t_{i-1}}$, then

$$\begin{aligned}&E\left(([Z,\epsilon]_{\mathcal{G}})^2 I_{\{\tau_l>T\}}|\mathcal{F}^{(0)}\right)\\ &= I_{\{\tau_l>T\}}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta Z_{t_i}\Delta Z_{t_j} E\left[\left(\epsilon_{t_i} - \epsilon_{t_{i-1}}\right)\left(\epsilon_{t_j} - \epsilon_{t_{j-1}}\right)|\mathcal{F}^{(0)}\right].\end{aligned}$$

By assumption, the noises are mutually independent conditioning on the whole path of latent process X, thus

$$E\left[(\epsilon_{t_i} - \epsilon_{t_{i-1}})(\epsilon_{t_j} - \epsilon_{t_{j-1}})|\mathcal{F}^{(0)}\right] = \begin{cases} -g_{t_{i\wedge j}}, & |i-j| = 1\\ g_{t_{i-1}} + g_{t_i}, & j = i\\ 0, & \text{otherwise.}\end{cases}$$

So, if $\tau_l > T$, we have

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta Z_{t_i}\Delta Z_{t_j} E\left[\left(\epsilon_{t_i} - \epsilon_{t_{i-1}}\right)\left(\epsilon_{t_j} - \epsilon_{t_{j-1}}\right)|\mathcal{F}^{(0)}\right]\\ &= \sum_{i=0}^{n-1}\left(\Delta Z_{t_{i+1}}\right)^2 g_{t_i} + \sum_{i=1}^{n}\left(\Delta Z_{t_i}\right)^2 g_{t_i}\\ &\quad - 2\sum_{i=1}^{n-1}\Delta Z_{t_i}\Delta Z_{t_{i+1}} g_{t_i} \le 4M_{(2,l)}\cdot[Z,Z]_{\mathcal{G}} = O_p(1)\end{aligned}$$

by Proposition 1 and $\mathbb{P}^{(0)}\{\tau_l > T\} \longrightarrow 1$ as $l \longrightarrow \infty$, we know $[Z,\epsilon]_{\mathcal{G}} = O_p(1)$. So, the following relation holds:

$$[Y,Y]_T^{\{n\}} = [\epsilon,\epsilon]_{\mathcal{G}} - \frac{1}{2}\left([\epsilon,\epsilon]_{\mathcal{G}^{(\min)}} + [\epsilon,\epsilon]_{\mathcal{G}^{(\max)}}\right) + O_p(1).$$

By our assumption, we have the following:

$$\begin{aligned}K[\epsilon,\epsilon]_T^{(avg,K)} &= 2\sqrt{n}\left(M_T^{(1)} - M_T^{(3)}\right) + 2\sum_{k=1}^{K}\sum_{t_i\in\mathcal{G}'(k)} g_{t_i}\\ &\quad + \sum_{t_i\in\mathcal{G}^{(\min)}} g_{t_i} - \sum_{t_i\in\mathcal{G}^{(\max)}} g_{t_i} + O_p(\sqrt{K})\end{aligned} \tag{49}$$

$$[\epsilon,\epsilon]_{\mathcal{G}} = 2\sqrt{n}\left(M_T^{(1)} - M_T^{(2)}\right) + 2\sum_{i=0}^{n} g_{t_i} + O_p(1). \tag{50}$$

Define the following quantities:

$$\begin{aligned}\underline{m}_T^{(1)} &\equiv \frac{1}{\sqrt{K}}\sum_{i=1}^{K}\left(\epsilon^2_{\mathcal{G}_i^{(\min)}} - g_{\mathcal{G}_i^{(\min)}}\right)\\ \underline{m}_T^{(2)} &\equiv \frac{1}{\sqrt{K}}\sum_{i=1}^{K}\epsilon_{\mathcal{G}_{i+1}^{(\min)}}\epsilon_{\mathcal{G}_i^{(\min)}}\\ \overline{m}_T^{(1)} &\equiv \frac{1}{\sqrt{K}}\sum_{i=1}^{K}\left(\epsilon^2_{\mathcal{G}_i^{(\max)}} - g_{\mathcal{G}_i^{(\max)}}\right)\\ \overline{m}_T^{(2)} &\equiv \frac{1}{\sqrt{K}}\sum_{i=1}^{K}\epsilon_{\mathcal{G}_i^{(\max)}}\epsilon_{\mathcal{G}_{i-1}^{(\max)}}.\end{aligned} \tag{51}$$

Similarly to (50),

$$\begin{aligned}[\epsilon,\epsilon]_{\mathcal{G}^{(\min)}} &= 2\sqrt{K}\left(\underline{m}_T^{(1)} - \underline{m}_T^{(2)}\right) + 2\sum_{t_i\in\mathcal{G}^{(\min)}} g_{t_i} + O_p(1)\\ [\epsilon,\epsilon]_{\mathcal{G}^{(\max)}} &= 2\sqrt{K}\left(\overline{m}_T^{(1)} - \overline{m}_T^{(2)}\right) + 2\sum_{t_i\in\mathcal{G}^{(\max)}} g_{t_i} + O_p(1).\end{aligned}$$

Combine Lemma 4 with these results, the difference between sample-weighted TSRV and the averaged realized variance of the theoretical process Z is

$$\widehat{\langle X,X\rangle}_T^{(WTSRV,K)} - [Z,Z]_T^{(avg,K)}$$

[16] Lemma 1 on p. 606 in Li and Mykland (2007).

[17] The caveat is $\mathcal{G}$ and $\mathcal{G}^{(\min)}\bigcup\left(\bigcup_{k=1}^K\mathcal{G}''(k)\right)\bigcup\mathcal{G}^{(\max)}$ might not equal, the difference is $\{t_{\lfloor n/K\rfloor\cdot K+1},\ldots,t_n\}$. Whereas, upon an appropriate choice of K, this difference is asymptotically negligible.

$$= \frac{2\sqrt{n}}{K}\left(M_T^{(2)} - M_T^{(3)}\right) + \frac{1}{\sqrt{K}}\left(\underline{m}_T^{(1)} - \underline{m}_T^{(2)} + \overline{m}_T^{(1)} - \overline{m}_T^{(2)}\right) + O_p\left(\frac{1}{\sqrt{K}}\right) \tag{52}$$

therefore

$$\frac{K}{\sqrt{n}}\left(\widehat{\langle X, X\rangle}_T^{(WTSRV,K)} - [Z,Z]_T^{(avg,K)}\right) = 2\left(M_T^{(2)} - M_T^{(3)}\right) + o_p(1) \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(0, \frac{8}{T}\int_0^T g_t^2\,\mathrm{d}t\right).$$

The remaining argument discussing the error term due to discretization $[Z,Z]_T^{(avg)} - \langle Z, Z\rangle_T$ to which an identical technique in Appendix A.3 in Zhang et al. (2005) applies, whence we get the claim of Theorem 1. □

A.4. Proof of Theorem 2

Proof. Recall the definition (51), by (52) and the asymptotic behavior of the original TSRV (Aït-Sahalia et al., 2005; Li and Mykland, 2007), under the null H_0, we have

$$\widehat{\langle X, X\rangle}_T^{(WTSRV,K)} - \widehat{\langle X, X\rangle}_T^{(TSRV,K)} = \frac{2(K-1)}{K\sqrt{n}}\left(M_T^{(2)} - M_T^{(1)}\right) + \frac{1}{\sqrt{K}}\left(\underline{m}_T^{(1)} - \underline{m}_T^{(2)} + \overline{m}_T^{(1)} - \overline{m}_T^{(2)}\right) + O_p\left(\frac{1}{K}\right)$$

i.e., when H_0 holds, we have $N(Y,K)_T^n = \left(\underline{m}_T^{(1)} - \underline{m}_T^{(2)} + \overline{m}_T^{(1)} - \overline{m}_T^{(2)}\right) + o_p(1)$.

To prove the limiting distribution is normal, again, we will use martingale limit central by exploiting the discrete predictable quadratic variations of those in (51).

$$\langle \underline{m}_T^{(1)}, \underline{m}_T^{(1)}\rangle_T | \mathcal{F}^{(0)} = \frac{1}{K}\sum_{i=0}^{K}\left[E\left(\epsilon^4_{\mathcal{G}_i^{(\min)}} | \mathcal{F}^{(1)}_{\mathcal{G}_{i-1}^{(\min)}}, \mathcal{F}^{(0)}\right) - g^2_{\mathcal{G}_i^{(\min)}}\right] = \frac{1}{K}\sum_{i=0}^{K}\left[h_{\mathcal{G}_i^{(\min)}} - g^2_{\mathcal{G}_i^{(\min)}}\right] \longrightarrow h_0 - g_0^2.$$

By the same calculation, we know $\langle \overline{m}_T^{(1)}, \overline{m}_T^{(1)}\rangle_T | \mathcal{F}^{(0)} = \frac{1}{K}\sum_{i=0}^{n}\left[h_{\mathcal{G}_i^{(\max)}} - g^2_{\mathcal{G}_i^{(\max)}}\right] \longrightarrow h_T - g_T^2$. Besides,

$$\begin{aligned}\langle \underline{m}^{(2)}, \underline{m}^{(2)}\rangle_T | \mathcal{F}^{(0)} &= \frac{1}{K}\sum_{i=1}^{K}\epsilon^2_{\mathcal{G}_i^{(\min)}} E\left(\epsilon^2_{\mathcal{G}_{i+1}^{(\min)}} | \mathcal{F}^{(1)}_{\mathcal{G}_i^{(\min)}}, \mathcal{F}^{(0)}\right)\\ &= \frac{1}{K}\sum_{i=1}^{K}\epsilon^2_{\mathcal{G}_i^{(\min)}} g_{\mathcal{G}_{i+1}^{(\min)}}\\ &= \frac{1}{K}\sum_{i=1}^{K}\left[\epsilon^2_{\mathcal{G}_i^{(\min)}} - g_{\mathcal{G}_i^{(\min)}}\right] g_{\mathcal{G}_{i+1}^{(\min)}}\\ &\quad + \frac{1}{K}\sum_{i=1}^{K} g_{\mathcal{G}_i^{(\min)}} g_{\mathcal{G}_{i+1}^{(\min)}}.\end{aligned}$$

Since $\mathcal{G}^{(\min)} \longrightarrow 0$ is a shrinking sub-grid, so $\frac{1}{K}\sum_{i=1}^{K} g_{\mathcal{G}_i^{(\min)}} g_{\mathcal{G}_{i+1}^{(\min)}} \longrightarrow g_0^2$. Besides,

$$\begin{aligned}&E\left(\left(\frac{1}{K}\sum_{i=1}^{K}(\epsilon^2_{\mathcal{G}_i^{(\min)}} - g_{\mathcal{G}_i^{(\min)}}) g_{\mathcal{G}_{i+1}^{(\min)}} \mathbf{1}_{\{\tau_l > T\}}\right)^2 | \mathcal{F}^{(0)}\right)\\ &= \frac{1}{K^2}\sum_{i=1}^{K} E\left(\left(\epsilon^2_{\mathcal{G}_i^{(\min)}} - g_{\mathcal{G}_i^{(\min)}}\right)^2 | \mathcal{F}^{(0)}\right) g^2_{\mathcal{G}_{i+1}^{(\min)}} \mathbf{1}_{\{\tau_l > T\}}\\ &\le \frac{1}{K^2}\sum_{i=1}^{K} M_{(4,l)} \cdot M^2_{(2,l)} = O_p\left(\frac{1}{K}\right)\end{aligned}$$

by Proposition 1 and the fact that $P(\tau_l > T) \xrightarrow{\mathbb{P}} 1$ as $l \longrightarrow \infty$, we know $\frac{1}{K}\sum_{i=1}^{K}(\epsilon^2_{\mathcal{G}_i^{(\min)}} - g_{\mathcal{G}_i^{(\min)}}) g_{\mathcal{G}_{i+1}^{(\min)}} \xrightarrow{\mathbb{P}} 0$, hence $\langle \underline{m}^{(2)}, \underline{m}^{(2)}\rangle_T \longrightarrow g_0^2$. Similarly, $\langle \overline{m}^{(2)}, \overline{m}^{(2)}\rangle_T \longrightarrow g_T^2$. By martingale central limit theorem, we know $\underline{m}_T^{(1)}, \overline{m}_T^{(1)}, \underline{m}_T^{(2)}, \overline{m}_T^{(2)}$ are asymptotically mixed normal. Because $\mathcal{G}^{(\min)}$ and $\mathcal{G}^{(\max)}$ are disjoint sets of observation times, so $\underline{m}_T^{(2)}$ and $\overline{m}_T^{(2)}$, or $\underline{m}_T^{(2)}$ and $\underline{m}_T^{(2)}$ are independent conditional on $\mathcal{F}^{(0)}$, so $\underline{m}_T^{(1)} + \overline{m}_T^{(1)} \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(0, h_0 - g_0^2 + h_T - g_T^2\right)$ and $\underline{m}_T^{(2)} + \overline{m}_T^{(2)} \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(0, g_0^2 + g_T^2\right)$.

Furthermore,

$$\begin{aligned}&\langle \underline{m}^{(1)}, \underline{m}^{(2)}\rangle_T | \mathcal{F}^{(0)}\\ &= \frac{1}{K}\sum_{i=1}^{K} E\left[\left(\epsilon^2_{\mathcal{G}_i^{(\min)}} - g_{\mathcal{G}_i^{(\min)}}\right)\cdot\left(\epsilon_{\mathcal{G}_i^{(\min)}}\epsilon_{\mathcal{G}_{i+1}^{(\min)}}\right) | \mathcal{F}^{(0)}, \mathcal{F}^{(1)}_{\mathcal{G}_i^{(\min)}}\right]\\ &= \frac{1}{K}\sum_{i=1}^{K}\left[\left(\epsilon^2_{\mathcal{G}_i^{(\min)}} - g_{\mathcal{G}_i^{(\min)}}\right)\epsilon_{\mathcal{G}_i^{(\min)}} \cdot E\left(\epsilon_{\mathcal{G}_{i+1}^{(\min)}} | \mathcal{F}^{(1)}_{\mathcal{G}_i^{(\min)}}\right) | \mathcal{F}^{(0)}\right]\\ &= 0\end{aligned}$$

and $\langle \overline{m}^{(1)}, \overline{m}^{(2)}\rangle_T | \mathcal{F}^{(0)} = \frac{1}{K}\sum_{i=1}^{K}\epsilon_{\mathcal{G}_{i-1}^{(\max)}} E\left(\epsilon^3_{\mathcal{G}_i^{(\max)}} | \mathcal{F}^{(0)}, \mathcal{F}^{(1)}_{\mathcal{G}_{i-1}^{(\max)}}\right)$.

$$\begin{aligned}&E\left(\left(\langle \overline{m}^{(1)}, \overline{m}^{(2)}\rangle_T\right)^2 \cdot I_{\{\tau_l > T\}} | \mathcal{F}^{(0)}\right)\\ &= \frac{1}{K^2}\sum_{i=1}^{K} g_{\mathcal{G}_{i-1}^{(\max)}} \cdot \left(E\left(E(\epsilon^3_{\mathcal{G}_i^{(\max)}} | \omega^{(0)}) | \mathcal{F}^{(1)}_{\mathcal{G}_{i-1}^{(\max)}}\right)\right)^2 \cdot I_{\{\tau_l > T\}}\\ &\le \frac{1}{K} \cdot M_{(2,l)} \cdot M^2_{(3,l)} \cdot I_{\{\tau_l > T\}}\end{aligned}$$

by Proposition 1 and the fact that $\mathbb{P}\{\tau_l > T\} \longrightarrow 1$ as $l \longrightarrow \infty$, we know $\langle \overline{m}^{(1)}, \overline{m}^{(2)}\rangle_T = O_p\left(\frac{1}{\sqrt{K}}\right) \xrightarrow{\mathbb{P}} 0$. Thus, $\underline{m}_T^{(1)}, \underline{m}_T^{(2)}, \overline{m}_T^{(1)}$ and $\overline{m}_T^{(2)}$ are asymptotically independent and mixed normals, and

$$\underline{m}_T^{(1)} - \underline{m}_T^{(2)} + \overline{m}_T^{(1)} - \overline{m}_T^{(2)} \xrightarrow{\mathcal{L}-s} \mathcal{MN}(0, h_0 + h_T). \quad \square$$

A.5. Proof of Lemma 2

Proof. Since $\{Z_t\}_{t\ge 0}$ is an Itô semimartingale

$$[Y;4]_{\mathcal{G}} = \sum_{i=1}^{n}(\epsilon_{t_i} - \epsilon_{t_{i-1}})^4 + 4\sum_{i=1}^{n}(\epsilon_{t_i} - \epsilon_{t_{i-1}})^3(Z_{t_i} - Z_{t_{i-1}}) + 4\sum_{i=1}^{n}(\epsilon_{t_i} - \epsilon_{t_{i-1}})(Z_{t_i} - Z_{t_{i-1}})^3 + O_p(1)$$

$$\begin{aligned}&E\left[\left(\sum_{i=1}^{n}(\epsilon_{t_i} - \epsilon_{t_{i-1}})^3(Z_{t_i} - Z_{t_{i-1}})\right)^2 \mathbf{1}_{\{\tau_l > T\}} | \mathcal{F}^{(0)}\right]\\ &= \sum_{i=1}^{n} E\left[(\epsilon_{t_i} - \epsilon_{t_{i-1}})^6(Z_{t_i} - Z_{t_{i-1}})^2 | \mathcal{F}^{(0)}\right]\mathbf{1}_{\{\tau_l > T\}}\\ &\le \sum_{i=1}^{n} E\left[(\epsilon_{t_i} - \epsilon_{t_{i-1}})^{12} | \mathcal{F}^{(0)}\right]^{1/2} \cdot E\left[(Z_{t_i} - Z_{t_{i-1}})^4 | \mathcal{F}^{(0)}\right]^{1/2}\mathbf{1}_{\{\tau_l > T\}}\end{aligned}$$

$$\leq (990)^{1/2}\left(\max_{2\leq k\leq 12} M_{(k,l)}\vee 1\right)\mathbf{1}_{\{\tau_l>T\}}$$
$$\times\sum_{i=1}^{n} E\left[(Z_{t_i}-Z_{t_{i-1}})^4|\mathcal{F}^{(0)}\right]^{1/2}=O_p(1)$$

$$E\left[\left(\sum_{i=1}^{n}(\epsilon_{t_i}-\epsilon_{t_{i-1}})(Z_{t_i}-Z_{t_{i-1}})^3\right)^2\mathbf{1}_{\{\tau_l>T\}}|\mathcal{F}^{(0)}\right]$$
$$=\sum_{i=1}^{n}E\left[(\epsilon_{t_i}-\epsilon_{t_{i-1}})^2(Z_{t_i}-Z_{t_{i-1}})^6|\mathcal{F}^{(0)}\right]\mathbf{1}_{\{\tau_l>T\}}$$
$$\leq\sum_{i=1}^{n}E\left[(\epsilon_{t_i}-\epsilon_{t_{i-1}})^4|\mathcal{F}^{(0)}\right]^{1/2}\cdot E\left[(Z_{t_i}-Z_{t_{i-1}})^{12}|\mathcal{F}^{(0)}\right]^{1/2}\mathbf{1}_{\{\tau_l>T\}}$$
$$\leq\sqrt{6}\left(\max_{2\leq k\leq 4} M_{(k,l)}\vee 1\right)\mathbf{1}_{\{\tau_l>T\}}\cdot\sum_{i=1}^{n}E\left[(Z_{t_i}-Z_{t_{i-1}})^{12}|\mathcal{F}^{(0)}\right]^{1/2}$$
$$=o_p(1).$$

Thus, $[Y;4]_{\mathcal{G}}=\sum_{i=1}^{n}(\epsilon_{t_i}-\epsilon_{t_{i-1}})^4+O_p(1)$. Note that

$$\sum_{i=1}^{n}(\epsilon_{t_i}-\epsilon_{t_{i-1}})^4=2\sum_{i=1}^{n}h_{t_i}+6\sum_{i=1}^{n}g_{t_{i-1}}g_{t_i}$$
$$+\sqrt{n}\left(2L_T^{(1)}+6L_T^{(2)}-4L_T^{(3)}-4L_T^{(4)}\right)+O_p(1) \tag{53}$$

where

$$L_T^{(1)}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\epsilon_{t_i}^4-h_{t_i}\right]$$
$$L_T^{(2)}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\epsilon_{t_{i-1}}^2\epsilon_{t_i}^2-E(\epsilon_{t_{i-1}}^2\epsilon_{t_i}^2|\mathcal{F}^{(0)})\right]$$
$$L_T^{(3)}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{t_{i-1}}^3\epsilon_{t_i}\qquad L_T^{(4)}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{t_{i-1}}\epsilon_{t_i}^3.$$

We can show that $L_T^{(1)}, L_T^{(2)}, L_T^{(3)}$ and $L_T^{(4)}$ are mixed normals. Observe that

$$\frac{1}{n}\sum_{i=1}^{n}h_{t_i}=\frac{1}{T}\sum_{i=1}^{n}h_{t_i}\frac{T}{n}\longrightarrow\frac{1}{T}\int_0^T h_t\,dt$$
$$\frac{1}{n}\sum_{i=1}^{n}g_{t_{i-1}}g_{t_i}=\frac{1}{T}\sum_{i=1}^{n}g_{t_{i-1}}g_{t_i}\frac{T}{n}\longrightarrow\frac{1}{T}\int_0^T g_t^2\,dt$$

then (16) follows. □

For each $i=1,2,\ldots,r_n$, we define $m_i\equiv m_i^{(1)}-m_i^{(2)}$ where

$$m_i^{(1)}\equiv\frac{1}{\sqrt{K_n}}\sum_{k=1}^{K_n}\left(\epsilon_{t_{(i-1)K_n+k}}^2-g_{t_{(i-1)K_n+k}}\right)$$
$$m_i^{(2)}\equiv\frac{1}{\sqrt{K_n}}\sum_{k=1}^{K_n}\epsilon_{t_{(i-1)K_n+k}}\epsilon_{t_{(i-1)K_n+k-1}}. \tag{54}$$

To prove Theorem 3, we need an additional lemma:

Lemma 6. *Assume the microstructure noise is stationary, and under the moment assumptions on the noise process $\{\epsilon_t\}_{t\geq 0}$, we have for each $i\in\{1,2,\ldots,r_n\}$,*

$$E(m_i^2|\mathcal{F}^{(0)})=E(\epsilon^4|\mathcal{F}^{(0)})$$
$$E(m_i^4|\mathcal{F}^{(0)})$$
$$=6\left[E(\epsilon^4|\mathcal{F}^{(0)})^2-E(\epsilon^4|\mathcal{F}^{(0)})E(\epsilon^2|\mathcal{F}^{(0)})^2+E(\epsilon^2|\mathcal{F}^{(0)})^4\right]$$
$$+O_p\left(\frac{1}{K}\right).$$

Proof. For the ease of notation, let us suppress the notation $K=K_n$, and denote $\epsilon_{(i-1)K_n+k}$ by $\xi_{i,k}$, and $g_{(i-1)K_n+k}$ by $g_{i,k}$ for each $i\in\{1,2,\ldots,r_n\}$ and $k\in\{0,1,2,\ldots,K\}$. Note that under our new notation

$$m_i^{(1)}\equiv\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\left(\xi_{i,k}^2-g_{i,k}\right)\quad m_i^{(2)}\equiv\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\xi_{i,k-1}\xi_{i,k}$$

and

$$E\left[\left(m_i^{(1)}\right)^2|\mathcal{F}^{(0)}\right]=\frac{1}{K}\sum_{k=1}^{K}E\left[\xi_{i,k}^4-g_{i,k}^2|\mathcal{F}^{(0)}\right]$$
$$=E(\epsilon^4|\mathcal{F}^{(0)})-E(\epsilon^2|\mathcal{F}^{(0)})^2$$
$$E\left[\left(m_i^{(2)}\right)^2|\mathcal{F}^{(0)}\right]=\frac{1}{K}\sum_{k=1}^{K}E\left[\xi_{i,k-1}^2\xi_{i,k}^2|\mathcal{F}^{(0)}\right]=E(\epsilon^2|\mathcal{F}^{(0)})^2$$
$$E\left[m_i^{(1)}m_i^{(2)}|\mathcal{F}^{(0)}\right]=\frac{1}{K}\sum_{k=1}^{K}\sum_{j=1}^{K}E\left[\left(\xi_{i,k}^2-g_{i,k}\right)\cdot\xi_{i,j-1}\cdot\xi_{i,j}|\mathcal{F}^{(0)}\right]=0$$

thus $E(m_i^2|\mathcal{F}^{(0)})=E(\epsilon^4|\mathcal{F}^{(0)})$. Note that

$$m_i^2=\left(m_i^{(1)}\right)^2+\left(m_i^{(2)}\right)^2-2m_i^{(1)}m_i^{(2)}$$
$$m_i^4=\left(m_i^{(1)}\right)^4-4\left(m_i^{(1)}\right)^3m_i^{(2)}+6\left(m_i^{(1)}\right)^2\left(m_i^{(2)}\right)^2$$
$$-4m_i^{(1)}\left(m_i^{(2)}\right)^3+\left(m_i^{(2)}\right)^4.$$

Some calculation yields

$$E\left[\left(m_i^{(1)}\right)^4|\mathcal{F}^{(0)}\right]$$
$$=\frac{1}{K^2}\left[\sum_{k=1}^{K}E\left(\xi_{i,k}^2-g_{i,k}\right)^4+6\sum_{k=1}^{K}\sum_{j\neq k}E\left(\xi_{i,k}^2-g_{i,k}\right)^2E\left(\xi_{i,j}^2-g_{i,j}\right)^2\right]$$
$$=6\left[E(\epsilon^4|\mathcal{F}^{(0)})-E(\epsilon^2|\mathcal{F}^{(0)})^2\right]^2+O_p\left(\frac{1}{K}\right)$$
$$E\left[\left(m_i^{(1)}\right)^3m_i^{(2)}|\mathcal{F}^{(0)}\right]$$
$$=\frac{3}{K^2}E\left[\left(\sum_{k=1}^{K}\sum_{j\neq k}(\xi_{i,k}^2-g_{i,k})^2(\xi_{i,j}^2-g_{i,j})\right)\cdot\left(\sum_{j=1}^{K}\xi_{i,j-1}\xi_{i,j}\right)|\mathcal{F}^{(0)}\right]$$
$$=\frac{6}{K^2}\sum_{k=2}^{K}E\left[(\xi_{i,k-1}^2-g_{i,k-1})^2\xi_{i,k-1}\cdot(\xi_{i,k}^2-g_{i,k})\xi_{i,k}|\mathcal{F}^{(0)}\right]=O_p\left(\frac{1}{K}\right)$$
$$E\left[\left(m_i^{(1)}\right)^2\left(m_i^{(2)}\right)^2|\mathcal{F}^{(0)}\right]$$
$$=\underbrace{\frac{1}{K^2}E\left[\sum_{k=1}^{K}\left(\xi_{i,k}^2-g_{i,k}\right)^2\cdot\left(\sum_{j=1}^{K}\xi_{i,j-1}\xi_{i,j}\right)^2|\mathcal{F}^{(0)}\right]}_{(m21)}$$
$$+\underbrace{\frac{1}{K^2}E\left[\sum_{k=1}^{K}\sum_{j\neq k}\left(\xi_{i,k}^2-g_{i,k}\right)\left(\xi_{i,j}^2-g_{i,j}\right)\cdot\left(\sum_{j=1}^{K}\xi_{i,j-1}\xi_{i,j}\right)^2|\mathcal{F}^{(0)}\right]}_{(m22)}.$$

Note that

$$(m21)=\frac{1}{K^2}E\left[\sum_{k=1}^{K}\sum_{j\neq k,k+1}\left(\xi_{i,k}^2-g_{i,k}\right)^2\cdot\xi_{i,j-1}^2\xi_{i,j}^2|\mathcal{F}^{(0)}\right]$$
$$+\frac{1}{K^2}E\left[\sum_{k=1}^{K}\left(\xi_{i,k}^2-g_{i,k}\right)^2\cdot\xi_{i,k-1}^2\xi_{i,k}^2|\mathcal{F}^{(0)}\right]$$

$$+\frac{1}{K^2}E\left[\sum_{k=1}^{K-1}\left(\xi_{i,k}^2-g_{i,k}\right)^2\cdot\xi_{i,k}^2\xi_{i,k+1}^2|\mathcal{F}^{(0)}\right]$$
$$=\left[E(\epsilon^4|\mathcal{F}^{(0)})-E(\epsilon^2|\mathcal{F}^{(0)})^2\right]E(\epsilon^2|\mathcal{F}^{(0)})^2+O_p\left(\frac{1}{K}\right)$$

and

$$\begin{aligned}(\text{m22}) &= \frac{1}{K^2}E\left[\left(\sum_{k=1}^{K}\sum_{j\neq k}\left(\xi_{i,k}^2-g_{i,k}\right)\left(\xi_{i,j}^2-g_{i,j}\right)\right)\right.\\ &\quad\left.\cdot\left(\sum_{j=1}^{K}\xi_{i,j-1}^2\xi_{i,j}^2+\sum_{j=1}^{K-1}\xi_{i,j-1}\xi_{i,j}^2\xi_{i,j+1}\right)|\mathcal{F}^{(0)}\right]\\ &= \frac{1}{K^2}E\left[\sum_{k=2}^{K}\left(\xi_{i,k-1}^4-\xi_{i,k-1}^2g_{i,k-1}\right)\left(\xi_{i,k}^4-\xi_{i,k}^2g_{i,k}\right)|\mathcal{F}^{(0)}\right]\\ &\quad+\frac{1}{K^2}E\left[\sum_{k=2}^{K-1}\left(\xi_{i,k-1}^3-\xi_{i,k-1}g_{i,k-1}\right)\right.\\ &\quad\left.\cdot\xi_{i,k}^2\cdot\left(\xi_{i,k+1}^3-\xi_{i,k+1}g_{i,k+1}\right)|\mathcal{F}^{(0)}\right]=O_p\left(\frac{1}{K}\right)\end{aligned}$$

so we have

$$E\left[\left(m_i^{(1)}\right)^2\left(m_i^{(2)}\right)^2|\mathcal{F}^{(0)}\right]=\left[E(\epsilon^4|\mathcal{F}^{(0)})-E(\epsilon^2|\mathcal{F}^{(0)})^2\right]E(\epsilon^2|\mathcal{F}^{(0)})^2+O_p\left(\frac{1}{K}\right)$$

$$\begin{aligned}E\left[m_i^{(1)}\left(m_i^{(2)}\right)^3|\mathcal{F}^{(0)}\right] &= \frac{3}{K^2}\sum_{k=2}^{K}E\left[\xi_{i,k-2}^2\xi_{i,k-1}^3\xi_{i,k}^3|\mathcal{F}^{(0)}\right]\\ &\quad+\frac{3}{K^2}\sum_{k=1}^{K-2}E\left[\xi_{i,k}^3\xi_{i,k+1}^3\xi_{i,k+2}^2|\mathcal{F}^{(0)}\right]\\ &\quad+\frac{1}{K^2}\sum_{k=1}^{K}E\left[\left(\xi_{i,k}^5-\xi_{i,k}^3g_{i,k}\right)\cdot\xi_{i,k-1}^3|\mathcal{F}^{(0)}\right]\\ &\quad+\frac{1}{K^2}\sum_{k=1}^{K-1}E\left[\left(\xi_{i,k}^5-\xi_{i,k}^3g_{i,k}\right)\cdot\xi_{i,k+1}^3|\mathcal{F}^{(0)}\right]=O_p\left(\frac{1}{K}\right)\end{aligned}$$

$$\begin{aligned}E\left[\left(m_i^{(2)}\right)^4|\mathcal{F}^{(0)}\right] &= \frac{1}{K^2}\sum_{k=1}^{K}E\left(\xi_{i,k-1}^4\xi_{i,k}^4|\mathcal{F}^{(0)}\right)\\ &\quad+\frac{6}{K^2}\sum_{k=1}^{K}\sum_{j\neq k}E\left(\xi_{i,k-1}^2\xi_{i,k}^2\xi_{i,j-1}^2\xi_{i,j}^2|\mathcal{F}^{(0)}\right)\\ &= 6E(\epsilon^2|\mathcal{F}^{(0)})^4+O_p\left(\frac{1}{K}\right).\end{aligned}$$

Thus, from the above calculation, we have

$$\begin{aligned}&E(m_i^4|\mathcal{F}^{(0)})\\ &= E\left[\left(m_i^{(1)}\right)^4|\mathcal{F}^{(0)}\right]+6E\left[\left(m_i^{(1)}\right)^2\left(m_i^{(2)}\right)^2|\mathcal{F}^{(0)}\right]\\ &\quad+E\left[\left(m_i^{(2)}\right)^4|\mathcal{F}^{(0)}\right]+O_p\left(\frac{1}{K}\right)\\ &= 6\left[E(\epsilon^4|\mathcal{F}^{(0)})^2-E(\epsilon^4|\mathcal{F}^{(0)})E(\epsilon^2|\mathcal{F}^{(0)})^2+E(\epsilon^2|\mathcal{F}^{(0)})^4\right]\\ &\quad+O_p\left(\frac{1}{K}\right).\quad\square\end{aligned}$$

A.6. Proof of Theorem 3

Proof. Under the assumption of this theorem, we know $g_t(\omega^{(0)})=E(\epsilon^2|\mathcal{F}^{(0)})(\omega^{(0)})$ has a constant value. By the proof of Theorem 2, we know $m_i\xrightarrow{\mathcal{L}-s}\mathcal{MN}\left(0,E(\epsilon^4|\mathcal{F}^{(0)})\right)$ under the null hypothesis. By continuous mapping theorem, $m_i^2\xrightarrow{\mathcal{L}-s}E(\epsilon^4|\mathcal{F}^{(0)})\cdot\chi_1^2$ where χ_1^2 denotes a centered Chi-square distribution with degree of freedom 1 and independent of $\mathcal{F}^{(1)}$. Note that

$$\begin{aligned}U(Y,K_n,s_n,2)_T^n &= \frac{1}{r_n-s_n+1}\\ &\quad\times\left[\sum_{i=1}^{r_n-s_n+1}m_i^2+\sum_{i=1}^{r_n-s_n+1}m_{i+s_n-1}^2+2\sum_{i=1}^{r_n-s_n+1}m_im_{i+s_n-1}\right]\\ &\quad+O_p\left(\frac{1}{K_n}\bigvee\frac{K_n}{n}\right)\end{aligned}$$

we can write

$$\begin{aligned}&\sqrt{r_n-s_n+1}\left(U(Y,K_n,2)_T^n-2E(\epsilon^4|\mathcal{F}^{(0)})\right)\\ &\quad=2H_T^{(1)}+2H_T^{(2)}+R_T+O_p\left(\frac{\sqrt{r_n}}{K_n}\bigvee\frac{1}{\sqrt{r_n}}\right)\end{aligned}\tag{55}$$

where

$$\begin{aligned}H_T^{(1)} &= \frac{1}{\sqrt{r_n-s_n+1}}\sum_{i=1}^{r_n-s_n+1}(m_i-E(m_i^2|\mathcal{F}^{(0)}))\\ H_T^{(2)} &= \frac{1}{\sqrt{r_n-s_n+1}}\sum_{i=1}^{r_n-s_n+1}m_im_{i-s_n+1}\\ R_T &= \frac{1}{\sqrt{r_n-s_n+1}}\\ &\quad\times\left[\sum_{i=r_n-s_n+2}^{r_n}(m_i^2-E(\epsilon^4|\mathcal{F}^{(0)}))-\sum_{i=1}^{s_n-1}(m_i^2-E(\epsilon^4|\mathcal{F}^{(0)}))\right].\end{aligned}$$

Furthermore, note that on the coarser filtered probability space $\left(\Omega^{(1)},\mathcal{F}^{(1)},\{\mathcal{F}_{t_{(i-1)K_n}}^{(1)}\}_{i\in\mathbb{N}},\mathbb{P}^{(1)}\right)$, $H_t^{(1)}$ and $H_t^{(2)}$ are two discrete martingales, and the increments of $H_t^{(1)}$ and $H_t^{(2)}$, namely

$$\begin{aligned}&\left\{\frac{1}{\sqrt{r_n-s_n+1}}\left(m_i^2-E(m_i^2|\omega^{(0)})\right)\right\}_{n\in\mathbb{N}^+,i\in\mathbb{N}^+\leq r_n-s_n+1}\\ &\quad\times\left\{\frac{1}{\sqrt{r_n-s_n+1}}\left(m_im_{i+s_n-1}\right)\right\}_{n\in\mathbb{N}^+,i\in\mathbb{N}^+\leq r_n-s_n+1}\end{aligned}$$

are two triangular sequences to which we can apply martingale central limit theorem. By the results of Lemma 6

$$\begin{aligned}\langle H^{(1)},H^{(1)}\rangle_T|\mathcal{F}^{(0)} &= \frac{1}{r_n-s_n+1}\\ &\quad\times\sum_{i=1}^{r_n-s_n+1}\left[E\left(m_i^4|\mathcal{F}_{t_{(i-1)K}}^{(1)}\right)-E\left(m_i^2|\mathcal{F}^{(0)}\cup\mathcal{F}_{t_{(i-1)K}}^{(1)}\right)^2\right]|\mathcal{F}^{(0)}\\ &= 5E(\epsilon^4|\mathcal{F}^{(0)})^2-6E(\epsilon^4|\mathcal{F}^{(0)})E(\epsilon^2|\mathcal{F}^{(0)})^2+6E(\epsilon^2|\mathcal{F}^{(0)})^4\end{aligned}$$

and

$$\begin{aligned}&\langle H^{(2)},H^{(2)}\rangle_T|\mathcal{F}^{(0)}\\ &\quad=\frac{1}{r_n-s_n+1}\sum_{i=1}^{r_n-s_n+1}E(m_i^2|\mathcal{F}^{(0)})\cdot E\left(m_{i+s_n-1}^2|\mathcal{F}^{(0)}\right)\\ &\quad+\frac{1}{r_n-s_n+1}\sum_{i=1}^{r_n-s_n+1}\left(m_i^2-E(m_i^2|\mathcal{F}^{(0)})\right)\cdot E\left(m_{i+s_n-1}^2|\mathcal{F}^{(0)}\right)\end{aligned}$$

since $P(\tau_l > T) \xrightarrow{\mathbb{P}^{(0)}} 1$ as $l \longrightarrow \infty$ and

$$E\left(\left(\frac{1}{r_n - s_n + 1}\sum_{i=1}^{r_n - s_n + 1}\left(m_i^2 - E(m_i^2|\mathcal{F}^{(0)})\right)\right.\right.$$
$$\left.\left.\cdot E\left(m_{i+s_n-1}^2|\mathcal{F}^{(0)}\right)\mathbf{1}_{\{\tau_l>T\}}\right)^2|\mathcal{F}^{(0)}\right)$$
$$= \frac{1}{(r_n - s_n + 1)^2}$$
$$\times \sum_{k=1}^{r_n - s_n + 1} Var\left(m_i^2 - E(m_i^2|\mathcal{F}^{(0)})\right)\left(E\left(m_{i+s_n-1}^2|\mathcal{F}^{(0)}\right)\right)^2\mathbf{1}_{\{\tau_l>T\}}$$
$$\le \frac{1}{(r_n - s_n + 1)^2}\sum_{k=1}^{r_n - s_n + 1} M_{(4,l)}\cdot M_{(2,l)}^2 = O_p\left(\frac{1}{r_n - s_n + 1}\right)$$

by Proposition 1, we know $\frac{1}{r_n - s_n + 1}\sum_{i=1}^{r_n - s_n + 1}\left(m_i^2 - E(m_i^2|\mathcal{F}^{(0)})\right)\cdot E\left(m_{i+s_n-1}^2|\mathcal{F}^{(0)}\right) \xrightarrow{\mathbb{P}} 0$, thus we have

$$\langle H^{(2)}, H^{(2)}\rangle_T|\mathcal{F}^{(0)} \xrightarrow{\mathbb{P}} E(\epsilon^4|\mathcal{F}^{(0)})^2.$$

Besides,

$$\langle H^{(1)}, H^{(2)}\rangle_T|\mathcal{F}^{(0)} = \frac{1}{r_n - s_n + 1}\sum_{i=1}^{r_n - s_n + 1}\left(m_i^3 - m_i E(m_i^2|\mathcal{F}^{(0)})\right)$$
$$\cdot E\left(m_{i+s_n-1}|\mathcal{F}^{(0)}\right) = 0.$$

Therefore, we have the following joint asymptotic distribution for $H_T^{(1)}$ and $H_T^{(2)}$:

$$\begin{pmatrix} H_T^{(1)} \\ H_T^{(2)} \end{pmatrix} \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \zeta^2 & 0 \\ 0 & E(\epsilon^4|\mathcal{F}^{(0)})^2 \end{pmatrix}\right) \tag{56}$$

where $\zeta^2 = 5E(\epsilon^4|\mathcal{F}^{(0)})^2 - 6E(\epsilon^4|\mathcal{F}^{(0)})E(\epsilon^2|\mathcal{F}^{(0)})^2 + 6E(\epsilon^2|\mathcal{F}^{(0)})^4$. Lastly, note that $R_T = o_p(1)$, this is because $P(\tau_l > T) \xrightarrow{\mathbb{P}^{(0)}} 1$ as $l \to \infty$, and

$$E(R_T^2\mathbf{1}_{\{\tau_l>T\}}|\mathcal{F}^{(0)}) = \frac{1}{r_n - s_n + 1}\left[\sum_{i=1}^{s_n-1}E(m_i^4|\mathcal{F}^{(0)})\right.$$
$$\left. + \sum_{i=r_n-s_n+2}^{r_n}E(m_i^4|\mathcal{F}^{(0)}) - 2(s_n - 1)E(\epsilon^4|\mathcal{F}^{(0)})^2\right]$$
$$= O_p\left(\frac{s_n}{r_n - s_n + 1}\right) = o_p(1).$$

Plug in these results into (55), we can get

$$\sqrt{r_n - s_n + 1}\left(U(Y, K_n, s_n, 2)_T^n - 2E(\epsilon^4|\mathcal{F}^{(0)})\right)$$
$$= 2\left(H_T^{(1)} + H_T^{(2)}\right) + o_p(1) \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(0, \eta^2\right)$$

where $\eta^2 = 24[E(\epsilon^4|\mathcal{F}^{(0)})^2 - E(\epsilon^4|\mathcal{F}^{(0)})E(\epsilon^2|\mathcal{F}^{(0)})^2 + E(\epsilon^2|\mathcal{F}^{(0)})^4]$.

According to Remark 1 about the consistent estimator of $E(\epsilon^4|\mathcal{F}^{(0)})$, $\widehat{\eta}^2 - \eta^2 = O_p(1/\sqrt{n})$ when the noise is stationary due to (16), as well as $\frac{1}{2n}[Y, Y]_{\mathcal{G}} - E(\epsilon^2|\mathcal{F}^{(0)}) = O_p(1/\sqrt{n})$, plus the stable convergence for $U(Y, K_n, s_n, 2)_{\mathcal{T}}^n$, (27) follows. □

A.7. Proof of Theorems 5 and 6

In this proof, we write K and r without the subscript n in order to avoid clustered notation. We give the proof for Theorem 6 first, and dictate how to modify the proof to prove Theorem 5.

A.7.1. The law of large number: the limit quantity

Under the assumption of Theorem 6, and from Lemma 4, we have

$$\frac{1}{2K}[Y, Y]_{\mathcal{S}_i} = \frac{1}{2K}[\epsilon, \epsilon]_{\mathcal{S}_i} + O_p\left(\frac{1}{K}\right)$$
$$= \frac{1}{K}\sum_{t_j\in(T_{i-1},T_i]} g_{t_j} + \frac{1}{\sqrt{K}}\left(m_i^{(1)} - m_i^{(2)}\right) + O_p\left(\frac{1}{K}\right)$$

where $m_i^{(1)}$ and $m_i^{(2)}$ are defined in (54) which are asymptotically mixing normal. Thus,

$$\frac{1}{2K}[Y, Y]_{\mathcal{S}_{i+1}} - \frac{1}{2K}[Y, Y]_{\mathcal{S}_i}$$
$$= \underbrace{\frac{1}{K}\sum_{j=1}^{K}\sum_{l=1}^{K}\left(g_{t_{(i-1)K+j+l}} - g_{t_{(i-1)K+j+l-1}}\right)}_{(\mathrm{A})}$$
$$+ \frac{1}{\sqrt{K}}(m_{i+1} - m_i) + O_p\left(\frac{1}{K}\right)$$

notice that:

$$(\mathrm{A})^2 = \frac{1}{K^2}\left[\sum_{j=1}^{K}(j-1)\Delta g_{t_{(i-1)K+j}} + \sum_{j=K+1}^{2K}(K-(j-1))\Delta g_{t_{(i-1)K+j}}\right]^2$$
$$= \sum_{j=1}^{K}\frac{(j-1)^2}{K^2}(\Delta g_{t_{(i-1)K+j}})^2$$
$$+ \sum_{j=K+1}^{2K}\frac{(K-(j-1))^2}{K^2}(\Delta g_{t_{(i-1)K+j}})^2 + (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III})$$

where

$$(\mathrm{I}) = \sum_{j=1}^{K}\sum_{l\neq j}\frac{(j-1)(l-1)}{K^2}\Delta g_{t_{(i-1)K+j}}\Delta g_{t_{(i-1)K+l}}$$
$$(\mathrm{II}) = \sum_{j=1}^{K}\sum_{l\neq j}\frac{(K-(j-1))(K-(l-1))}{K^2}\Delta g_{t_{iK+j}}\Delta g_{t_{iK+l}} \tag{57}$$
$$(\mathrm{III}) = \sum_{j=1}^{K}\sum_{l=1}^{K}\frac{(j-1)(K-(l-1))}{K^2}\Delta g_{t_{(i-1)K+j}}\Delta g_{t_{iK+l}}$$

are mean-0 martingales. By standard localization procedure, we can strengthen the condition by assuming $\sigma_t^{(g)} \le \sigma_+^{(g)}, \forall t \in [0, \mathcal{T}]$; therefore,

$$E[(\mathrm{I})^2] \le \frac{T^2(\sigma_+^{(g)})^4}{n^2}\sum_{j=1}^{K}\sum_{l=1}^{K}\left[\frac{(j-1)(l-1)}{K^2}\right]^2$$
$$= \frac{T^2(\sigma_+^{(g)})^4}{n^2}\cdot\sum_{j=1}^{K}\frac{(j-1)^2}{K^2}\cdot\sum_{j=1}^{K}\frac{(l-1)^2}{K^2} = O_p\left(\frac{K^2}{n^2}\right)$$

by Chebyshev inequality, $(\mathrm{I}) = O_p\left(\frac{K}{n}\right)$. Similarly, $(\mathrm{II}), (\mathrm{III}) = O_p\left(\frac{K}{n}\right)$. Furthermore, we can know $(\mathrm{A}) = O_p\left(\sqrt{\frac{K}{n}}\right)$. Thus,

$$\sum_{i=1}^{r-1}\left(\frac{1}{2K}[Y, Y]_{\mathcal{S}_{i+1}} - \frac{1}{2K}[Y, Y]_{\mathcal{S}_i}\right)^2$$
$$= \sum_{j=1}^{K}\frac{(j-1)^2}{K^2}(\Delta g_{t_j})^2 + \sum_{j=1}^{K}\frac{(K-(j-1))^2}{K^2}(\Delta g_{t_{(r-1)K+j}})^2$$
$$+ \sum_{i=2}^{r-2}\sum_{j=1}^{K}\frac{(j-1)^2 + (K-(j-1))^2}{K^2}(\Delta g_{t_{(i-1)K+j}})^2$$

$$+\underbrace{\sum_{i=1}^{r-1}[(\mathrm{I})+(\mathrm{II})+(\mathrm{III})]}_{O_p\left(\frac{\sqrt{r}K}{n}\right)=O_p\left(\frac{1}{\sqrt{r}}\right)}+\underbrace{O_p\left(\frac{r}{K}\right)}_{\text{error due to noises}}$$

note that the error due to noises (of the stochastic order $O_p\left(\frac{r}{K}\right)$) approximately equals to $\frac{2r}{\mathcal{T}K}\int_0^{\mathcal{T}} h_t\,\mathrm{d}t$ by the proof of Theorem 3. Hence,

$$\begin{aligned}
&\sum_{i=1}^{r-1}\left(\frac{1}{2K}[Y,Y]_{\mathcal{S}_{i+1}}-\frac{1}{2K}[Y,Y]_{\mathcal{S}_i}\right)^2\\
&\quad-\frac{2}{3}\sum_{j=1}^{n}(\Delta g_{t_j})^2-\frac{2r}{\mathcal{T}K}\int_0^{\mathcal{T}} h_t\,\mathrm{d}t\\
&=\sum_{i=1}^{r-2}\sum_{j=1}^{K}\left[\frac{1}{3}-2\frac{K-(j-1)}{K}\frac{j-1}{K}\right](\Delta g_{t_{iK+j}})^2\\
&\quad+\sum_{j=1}^{K}\left[\frac{(j-1)^2}{K^2}-\frac{2}{3}\right](\Delta g_{t_j})^2\\
&\quad+\sum_{j=1}^{K}\left[\frac{(K-(j-1))^2}{K^2}-\frac{2}{3}\right](\Delta g_{t_{(r-1)K+j}})^2+(\mathrm{E1})+(\mathrm{E2})
\end{aligned}$$

where

$$(\mathrm{E1})=\sum_{i=1}^{r-1}[(\mathrm{I})+(\mathrm{II})+(\mathrm{III})]=O_p\left(\frac{1}{\sqrt{r}}\right)$$

end points of martingale in $\{g_t\}_{t\in[0,\mathcal{T}]}$

$$(\mathrm{E2})=\frac{2r}{K}\left(\frac{1}{r}\sum_{i=1}^{r-1}m_i^2-\frac{1}{\mathcal{T}}\int_0^{\mathcal{T}} h_t\,\mathrm{d}t\right)=O_p\left(\frac{\sqrt{r}}{K}\right)$$

error due to noise.

The order of (E1) will be analyzed later; (E2) comes from negligible remaining of microstructure noise and its order is obtained from the proof of Theorem 3:

$$\frac{K}{r}\cdot\sqrt{r}(\mathrm{E2})\xrightarrow{\mathcal{L}-s}\mathcal{MN}\left(0,\frac{24}{\mathcal{T}}\int_0^{\mathcal{T}}\left[h_t^2-h_tg_t^2+g_t^4\right]\mathrm{d}t\right).$$

Moreover,

$$\begin{aligned}
&\sum_{i=1}^{r-2}\sum_{j=1}^{K}\left[\frac{1}{3}-2\frac{K-(j-1)}{K}\frac{j-1}{K}\right](\Delta g_{t_{iK+j}})^2\\
&\quad+\sum_{j=1}^{K}\left[\frac{(j-1)^2}{K^2}-\frac{2}{3}\right](\Delta g_{t_j})^2\\
&\quad+\sum_{j=1}^{K}\left[\frac{(K-(j-1))^2}{K^2}-\frac{2}{3}\right](\Delta g_{t_{(r-1)K+j}})^2\\
&=O_p\left(\frac{K}{n}\right)=O_p\left(\frac{1}{r}\right).
\end{aligned}$$

Because $\sum_{j=1}^{n}(\Delta g_{t_j})^2-\langle g,g\rangle_{\mathcal{T}}=O_p\left(\frac{1}{\sqrt{n}}\right)$,[18] so

$$\begin{aligned}
&\frac{r}{K}U(Y,K,2)_{\mathcal{T}}^2-\frac{2}{3}\langle g,g\rangle_{\mathcal{T}}-\frac{2r}{\mathcal{T}K}\int_0^{\mathcal{T}} h_t\,\mathrm{d}t\\
&=(\mathrm{E1})+(\mathrm{E2})+O_p\left(\frac{1}{r}\right).
\end{aligned}$$

[18] The proofs can be found in Jacod and Protter (1998) and Mykland and Zhang (2006).

A.7.2. Decomposition of the discretization error process

Followed from (57), if we define the following two quantities:

$$\begin{aligned}
N_{\mathcal{T}}^{(1)}&\equiv 2\sqrt{r}\sum_{i=0}^{r-1}\sum_{j=2}^{K}\Delta g_{t_{iK+j}}\\
&\quad\cdot\left[\sum_{l=1}^{j-1}\left(1+2\frac{j-1}{K}\frac{l-1}{K}-\frac{j-1}{K}-\frac{l-1}{K}\right)\Delta g_{t_{iK+l}}\right] \qquad (58)\\
N_{\mathcal{T}}^{(2)}&\equiv\sqrt{r}\sum_{i=1}^{r-1}\sum_{j=1}^{K}\left(1-\frac{j-1}{K}\right)\Delta g_{t_{iK+j}}\cdot\left(\sum_{l=1}^{K}\frac{l-1}{K}\Delta g_{t_{(i-1)K+l}}\right) \qquad (59)
\end{aligned}$$

then we have

$$(\mathrm{E1})=\frac{1}{\sqrt{r}}N_{\mathcal{T}}^{(1)}+\frac{1}{\sqrt{r}}N_{\mathcal{T}}^{(2)}+\underbrace{O_p\left(\frac{K}{n}\right)}_{\text{the edge in (E1)}} \qquad (60)$$

furthermore, by (58)

$$\begin{aligned}
\langle N^{(1)},N^{(2)}\rangle_{\mathcal{T}}&=2r\sum_{i=1}^{r-1}\sum_{j=2}^{K}\left(1-\frac{j-1}{K}\right)\Delta\langle g,g\rangle_{t_{iK+j}}\\
&\quad\times\sum_{l=1}^{K}\frac{l-1}{K}\Delta g_{t_{(i-1)K+l}}\\
&\quad\times\sum_{l=1}^{j-1}\left(1+2\frac{j-1}{K}\frac{l-1}{K}-\frac{j-1}{K}-\frac{l-1}{K}\right)\Delta g_{t_{iK+l}}
\end{aligned}$$

$$\begin{aligned}
E\left[\langle N^{(1)},N^{(2)}\rangle_{\mathcal{T}}^2\right]&=4r^2\sum_{i=1}^{r-1}\sum_{j=2}^{K}\left(1-\frac{j-1}{K}\right)^2\left(\Delta\langle g,g\rangle_{t_{iK+j}}\right)^2\\
&\quad\times\left(\sum_{l=1}^{K}\frac{(l-1)^2}{K^2}E\left[(\Delta g_{t_{(i-1)K+l}})^2\right]\right)\\
&\quad\times\left(\sum_{l=1}^{j-1}\left(\frac{(j-1)(l-1)}{K^2}+\frac{(K-(j-1))(K-(l-1))}{K^2}\right)^2\right.\\
&\quad\left.\times E\left[(\Delta g_{t_{iK+l}})^2\right]\right)
\end{aligned}$$

thus we can know $E\left[\langle N^{(1)},N^{(2)}\rangle_{\mathcal{T}}^2\right]=O_p\left(\frac{r^3K^3}{n^4}\right)=O_p\left(\frac{1}{n}\right)$, So

$$\begin{aligned}
&\langle N^{(1)}+N^{(2)},N^{(1)}+N^{(2)}\rangle_{\mathcal{T}}\\
&\quad=\langle N^{(1)},N^{(1)}\rangle_{\mathcal{T}}+\langle N^{(2)},N^{(2)}\rangle_{\mathcal{T}}+O_p\left(\frac{1}{\sqrt{n}}\right). \qquad (61)
\end{aligned}$$

A.7.3. Calculating $\langle n^{(1)},N^{(1)}\rangle_{\mathcal{T}}$

By (58),

$$\begin{aligned}
\langle N^{(1)},N^{(1)}\rangle_{\mathcal{T}}&=4r\sum_{i=0}^{r-1}\sum_{j=2}^{K}\Delta\langle g,g\rangle_{t_{iK+j}}\\
&\quad\times\left[\sum_{l=1}^{j-1}\left(\frac{(j-1)(l-1)}{K^2}+\frac{(K-(j-1))(K-(l-1))}{K^2}\right)\Delta g_{t_{iK+l}}\right]^2\\
&=(\mathrm{A1})+(\mathrm{A2})
\end{aligned}$$

where

$$(\mathrm{A1})=4r\sum_{i=0}^{r-1}\sum_{j=2}^{K}\Delta\langle g,g\rangle_{t_{iK+j}}$$

$$\times \left[\sum_{l=1}^{j-1}\left(\frac{2(j-1)-K}{K^2}(l-1)+\frac{K-(j-1)}{K}\right)^2(\Delta g_{t_{iK+l}})^2\right]$$
$$= 4r\sum_{i=0}^{r-1}\sum_{j=2}^{K}\left(\sigma^{(g)}_{t_{iK+j}}\right)^4\Delta_n^2$$
$$\times \sum_{l=1}^{j-1}\left[\frac{2(j-1)-K}{K^2}(l-1)+\frac{K-(j-1)}{K}\right]^2+O_p\left(\frac{1}{rK^{1/2}}\right)$$

the error term appears because $\sigma^{(g)}$ is an Itô process, and the error due to the local-consistency approximation for $\sigma^{(g)}$ is of a smaller order than $4r\sum_{i=0}^{r-1}\sum_{j=2}^{K}\sqrt{j}\Delta_n^3\sum_{l=1}^{j-1}\left[\frac{2(j-1)-K}{K^2}(l-1)+\frac{K-(j-1)}{K}\right]^2 = O_p\left(\frac{1}{rK^{1/2}}\right)$, besides

$$(\text{A2}) = 8r\sum_{i=0}^{r-1}\sum_{j=3}^{K}\Delta\langle g,g\rangle_{t_{iK+j}}\cdot\phi_j$$

where

$$\phi_j = \sum_{l=2}^{j-1}\sum_{k=1}^{l-1}\left[\frac{2(j-1)-K}{K^2}(l-1)+\frac{K-(j-1)}{K}\right]$$
$$\cdot\left[\frac{2(j-1)-K}{K^2}(k-1)+\frac{K-(j-1)}{K}\right]\Delta g_{t_{iK+l}}\Delta g_{t_{iK+k}}$$

by Burkholder–Davis–Gundy inequality, $\exists C_1\in\mathbb{R}^+$ such that

$$\|\phi_j\|_2^2 \le C_1^2\|\langle\phi_j,\phi_j\rangle\|_1$$
$$\le \frac{\mathcal{T}^2C_1^2\left(\sigma^{(g)}_+\right)^2}{n^2}\sum_{l=2}^{j-1}\left[\frac{2(j-1)-K}{K^2}(l-1)+\frac{K-(j-1)}{K}\right]^2$$
$$\times\sum_{k=1}^{l-1}\left[\frac{2(j-1)-K}{K^2}(k-1)+\frac{K-(j-1)}{K}\right]^2$$

thus, $\|\phi_j\|_2 = O_p\left(\frac{j^3}{K^2n}\right)$ and $\sum_{j=3}^{K}\|\phi_j\|_2^2 \le \sum_{j=3}^{K}C_1^2\|\langle\phi_j,\phi_j\rangle\|_1 = O_p\left(\frac{K^3}{n^2}\right)$. Define $(\text{A2})' \equiv 8r\sum_{i=0}^{r-1}\sum_{j=3}^{K}\left(\sigma^{(g)}_{t_{iK}}\right)^2\Delta_n\phi_j$, and apply Burkholder–Davis–Gundy inequality again, but on $(\text{A2})'$, we get $\|(\text{A2})'\|_2^2 \le 64r^2C_2^2\sum_{i=0}^{r-1}\sum_{j=3}^{K}\left(\sigma^{(g)}_+\right)^4\Delta_n^2\times\|\langle\phi_j,\phi_j\rangle\|_1 = O_p\left(\frac{r^3K^3}{n^4}\right) = O_p\left(\frac{1}{n}\right)$, so

$$(\text{A2})' = O_p\left(\frac{1}{\sqrt{n}}\right) \tag{62}$$

by Cauchy–Schwarz inequality,

$$\|(\text{A2})\text{–}(\text{A2})'\|_1$$
$$\le 8r^2\Delta(\mathcal{G})\sum_{j=3}^{K}\left\|\sup_{|t-s|\le K\Delta(\mathcal{G})}\left[\left(\sigma^{(g)}_t\right)^2-\left(\sigma^{(g)}_s\right)^2\right]\right\|_2\cdot\|\phi_j\|_2$$
$$\le 8r^2K(\Delta(\mathcal{G}))^{\frac{3}{2}}\left(\sum_{j=3}^{K}\|\phi_j\|_2^2\right)^{\frac{1}{2}} = O_p\left(\frac{1}{\sqrt{r}}\right) \tag{63}$$

from (62) and (63), we can know $(\text{A2}) = o_p(1)$, and more importantly,

$$\langle N^{(1)},N^{(1)}\rangle_T = 4r\sum_{i=0}^{r-1}\sum_{j=2}^{K}\left(\sigma^{(g)}_{t_{iK+j}}\right)^4\Delta_n^2$$
$$\times\underbrace{\sum_{l=1}^{j-1}\left[\frac{2(j-1)-K}{K^2}(l-1)+\frac{K-(j-1)}{K}\right]^2}_{(1)}+o_p(1)$$

notice that $(1) = \frac{4}{3}\frac{j^5}{K^4}-\frac{10}{3}\frac{j^4}{K^3}+\frac{13}{3}\frac{j^3}{K^2}-3\frac{j^2}{K}+j+O(1)$, so

$$\langle N^{(1)},N^{(1)}\rangle_T$$
$$= 4r\sum_{i=0}^{r-1}\sum_{j=2}^{K}(1)\cdot\left[\left(\sigma^{(g)}_{t_{iK}}\right)^2+O_p\left(\sqrt{K\Delta_n}\right)\right]^2\Delta_n^2+o_p(1)$$
$$= 4r\sum_{i=0}^{r-1}\left[\left(\sum_{j=2}^{K}(1)\right)\times\frac{\Delta_n}{K}\left(\sigma^{(g)}_{t_{iK}}\right)^4K\Delta_n\right]+o_p(1)$$

by Faulhaber's formula, wee know

$$\sum_{j=2}^{K}(1) = \left(\frac{4}{3}\times\frac{1}{6}-\frac{10}{3}\times\frac{1}{5}+\frac{13}{3}\times\frac{1}{4}-3\times\frac{1}{3}+\frac{1}{2}\right)K^2$$
$$+O(K) = \frac{5}{36}K^2+O(K)$$

so

$$\langle N^{(1)},N^{(1)}\rangle_T = T\sum_{i=1}^{r-1}\left[\frac{5}{9}+O_p\left(\frac{1}{K}\right)\right]\left(\sigma^{(g)}_{t_{iK}}\right)^4K\Delta_n+o_p(1)$$
$$\longrightarrow \frac{5T}{9}\int_0^T\left(\sigma^{(g)}_t\right)^4\mathrm{d}t.$$

A.7.4. Calculating $\langle n^{(2)},N^{(2)}\rangle_T$

By (58),

$$\langle N^{(2)},N^{(2)}\rangle_T = r\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{(K-(j-1))^2}{K^2}\Delta\langle g,g\rangle_{t_{iK+j}}$$
$$\times\left(\sum_{l=1}^{K}\frac{(l-1)}{K}\Delta g_{t_{(i-1)K+l}}\right)^2$$
$$= (\text{B1})+(\text{B2})$$

where

$$(\text{B1}) = r\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{(K-(j-1))^2}{K^2}\Delta\langle g,g\rangle_{t_{iK+j}}$$
$$\times\sum_{l=1}^{K}\frac{(l-1)^2}{K^2}(\Delta g_{t_{(i-1)K+l}})^2$$
$$= r\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{(K-(j-1))^2}{K^2}\left(\sigma^{(g)}_{t_{iK+j}}\right)^4\Delta_n^2$$
$$\times\sum_{l=1}^{K}\frac{(l-1)^2}{K^2}+O_p\left(\frac{1}{rK^{1/2}}\right)$$

the error term just above comes from the local-constancy approximation on $(\sigma^{(g)})^2$, it is of the stochastic order of $O_p(r\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{(K-(j-1))^2}{K^2}\sqrt{K}\Delta_n^3\times\sum_{l=1}^{K}\frac{(l-1)^2}{K^2}) = O_p\left(\frac{1}{rK^{1/2}}\right)$. Besides,

$$(\text{B2}) = 2r\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{(K-(j-1))^2}{K^2}\Delta\langle g,g\rangle_{t_{iK+j}}\psi_i$$

where

$$\psi_i = \sum_{l=2}^{K}\sum_{k=1}^{l-1}\frac{l-1}{K}\frac{k-1}{K}\Delta g_{t_{(i-1)K+l}}\Delta g_{t_{(i-1)K+k}}.$$

Apply Burkholder–Davis–Gundy on ψ_i, since $(\psi_i)_t \equiv \sum_{l=2}^{K}\sum_{k=1}^{l-1} \frac{l-1}{K}\frac{k-1}{K}\Delta g_{t_{(i-1)K+k}} \int_{t_{(i-1)K+l-1}}^{t_{(i-1)K+l}\wedge t} \mathrm{d}g_t$ is a continuous martingale by assumption of the Theorem 6,

$$\begin{aligned}\|\psi_i\|_2^2 &\le D_1^2\|\langle\psi_i,\psi_i\rangle\|_1 = D_1^2 E\sum_{l=2}^{K}\frac{(l-1)^2}{K^2}\Delta\langle g,g\rangle_{t_{(i-1)K+l}} \\ &\quad\times\left(\sum_{k=1}^{l-1}\frac{k-1}{K}\Delta g_{t_{(i-1)K+k}}\right)^2 \\ &\le D_1^2\left(\sigma_+^{(g)}\right)^4(\Delta(\mathcal{G}))^2\times\sum_{l=2}^{K}\frac{(l-1)^2}{K^2}\sum_{k=1}^{l-1}\frac{(k-1)^2}{K^2} \\ &= O_p\left(\frac{K^2}{n^2}\right)\end{aligned}$$

so $\|\psi_i\|_2^2 \le D_1\|\langle\psi_i,\psi_i\rangle\|_1 = O_p\left(\frac{1}{r^2}\right)$. Define $(\mathrm{B2})' \equiv 2r\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{(K-(j-1))^2}{K^2}\left(\sigma_{t_{(i-1)K}}^{(g)}\right)^2\Delta_n\psi_i$, apply Burkholder–Davis–Gundy inequality again on $(\mathrm{B2})'$,

$$\begin{aligned}\|(\mathrm{B2})'\|_2^2 &\le 4r^2D_2^2\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{(K-(j-1))^4}{K^4}\left(\sigma_t^{(g)}\right)^4\Delta_n^2\times\|\langle\psi_i,\psi_i\rangle\|_1 \\ &= O_p\left(\frac{r^2}{n^2}\right)\times\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{(K-(j-1))^4}{K^4}\times O_p\left(\frac{1}{r^2}\right) \\ &= O_p\left(\frac{1}{n}\right)\end{aligned}$$

therefore,

$$(\mathrm{B2})' = O_p\left(\frac{1}{\sqrt{n}}\right). \tag{64}$$

By Cauchy–Schwarz inequality,

$$\begin{aligned}\|(\mathrm{B2})-(\mathrm{B2})'\|_1 &\le 2r\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{(K-(j-1))^2}{K^2} \\ &\quad\times\left\|\Delta\langle g,g\rangle_{t_{iK+j}}-\left(\sigma_{t_{(i-1)K}}^{(g)}\right)^2\Delta_n\right\|_2\cdot\|\psi_i\|_2 \\ &\le 2r^2K\Delta(\mathcal{G})\cdot\left\|\sup_{|t-s|\le 2K\Delta(\mathcal{G})}\left[\left(\sigma_t^{(g)}\right)^2-\left(\sigma_s^{(g)}\right)^2\right]\right\|_2 \\ &\quad\cdot\sup_i\|\psi_i\|_2 = O_p\left(\frac{1}{\sqrt{r}}\right)\end{aligned} \tag{65}$$

combine (64) and (65), we can get $(\mathrm{B2}) = o_p(1)$. More importantly,

$$\begin{aligned}\langle N^{(2)},N^{(2)}\rangle_T &= r\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{(K-(j-1))^2}{K^2}\left[\left(\sigma_{t_{iK}}^{(g)}\right)^4+O_p(\sqrt{K\Delta_n})\right] \\ &\quad\times\left(\frac{K}{3}+O_p(1)\right)\times\Delta_n^2+o_p(1) \\ &= r\sum_{i=1}^{r-1}\frac{K^2}{9}\left(\sigma_{t_{iK}}^{(g)}\right)^4\Delta_n^2+o_p(1)\longrightarrow\frac{T}{9}\int_0^T\left(\sigma_t^{(g)}\right)^4\mathrm{d}t.\end{aligned}$$

A.7.5. Proof of the stable convergence

Based on Appendices A.7.2–A.7.4,

$$\begin{aligned}\langle\sqrt{r}(\mathrm{E1}),\sqrt{r}(\mathrm{E1})\rangle &= \langle N^{(1)},N^{(1)}\rangle_T+\langle N^{(2)},N^{(2)}\rangle_T+o_p(1) \\ &= \frac{2T}{3}\int_0^T\left(\sigma_t^{(g)}\right)^4\mathrm{d}t+o_p(1).\end{aligned} \tag{66}$$

Following the similar method as that in the proof of Theorem 3, we know

$$\begin{aligned}\langle\sqrt{r}(\mathrm{E2}),\sqrt{r}(\mathrm{E2})\rangle_T &= \frac{24r^2}{TK^2}\int_0^T\left[h_t^2-h_tg_t^2+g_t^4\right]\mathrm{d}t+o_p\left(\frac{r^2}{K^2}\right) \\ &= O_p\left(\frac{r^2}{K^2}\right).\end{aligned} \tag{67}$$

We need a technical condition on the filtration $\{\mathcal{F}_t\}_{t\ge0}$ to which all the relevant processes are adapted:

Assumption 11 (*Condition on the Filtration*)**.** There are Brownian motions $W^{(1)}, W^{(2)}, \ldots, W^{(p)}$ that generate the filtration $\{\mathcal{F}_t\}_{t\ge0}$.

Consider the normalized error process,

$$\begin{aligned}\sqrt{r}(\mathrm{E}) &= \sqrt{r}(\mathrm{E1})+\sqrt{r}(\mathrm{E2}) \\ &= N_T^{(1)}+N_T^{(2)}+\underbrace{O_p\left(\frac{1}{\sqrt{r}}\right)}_{\text{the edge in } N_T^{(1)}+N_T^{(2)}}+\underbrace{\sqrt{r}(\mathrm{E2})}_{O_p\left(\frac{r}{K}\right)} \\ &= 2\sqrt{r}\sum_{i=0}^{r-1}\sum_{j=2}^{K}\Delta g_{t_{iK+j}} \\ &\quad\times\sum_{l=1}^{j-1}\left(1+2\frac{j-1}{K}\frac{l-1}{K}-\frac{j-1}{K}-\frac{l-1}{K}\right)\Delta g_{t_{iK+l}} \\ &\quad+\sqrt{r}\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{K-(j-1)}{K}\Delta g_{t_{iK+j}}\times\sum_{l=1}^{K}\frac{l-1}{K}\Delta g_{t_{(i-1)K+l}} \\ &\quad+O_p\left(\frac{1}{\sqrt{r}}\vee\frac{r}{K}\right).\end{aligned}$$

Define

$$\begin{aligned}N_t^n &= 2\sqrt{r}\sum_{i=0}^{r-1}\sum_{j=2}^{K}\left[\sum_{l=1}^{j-1}\left(1+2\frac{j-1}{K}\frac{l-1}{K}-\frac{j-1}{K}-\frac{l-1}{K}\right)\Delta g_{t_{iK+l}}\right] \\ &\quad\cdot\Delta g_{t_{iK+j}\wedge t} \\ &\quad+\sqrt{r}\sum_{i=1}^{r-1}\sum_{j=1}^{K}\left(\sum_{l=1}^{K}\frac{l-1}{K}\Delta g_{t_{(i-1)K+l}}\right)\cdot\frac{K-(j-1)}{K}\Delta g_{t_{iK+j}\wedge t}\end{aligned}$$

then $\sqrt{r}(\mathrm{E})_t = N_T^n + o_p(1)$. Suppose $t_{iK+j-1} = \max\{t_k, k = 0, 1, \ldots, n, t_k \le t\}$, then

$$\begin{aligned}&\mathrm{d}\langle N^n,W^{(i)}\rangle_t \\ &= 2\sqrt{r}\left[\sum_{l=1}^{j-1}\left(1+2\frac{j-1}{K}\frac{l-1}{K}-\frac{j-1}{K}-\frac{l-1}{K}\right)\Delta g_{t_{iK+l}}\right] \\ &\quad\times\mathrm{d}\langle g,W^{(i)}\rangle_t \\ &\quad+\sqrt{r}\left[\sum_{l=1}^{K}\frac{l-1}{K}\Delta g_{t_{(i-1)K+l}}\cdot\frac{K-(j-1)}{K}\right]\mathrm{d}\langle g,W^{(i)}\rangle_t\end{aligned}$$

for $i = 1, 2, \ldots, p$, by Kunita–Watanabe inequality,

$$\begin{aligned}&\left|\langle g,W^{(i)}\rangle_{t+h}-\langle g,W^{(i)}\rangle_t\right| \le \sqrt{\langle g,g\rangle_{t+h}-\langle g,g\rangle_t} \\ &\quad\cdot\sqrt{\langle W^{(i)},W^{(i)}\rangle_{t+h}-\langle W^{(i)},W^{(i)}\rangle_t} \le \sigma_+^{(g)}h\end{aligned}$$

so $\Delta\langle g,W^{(i)}\rangle_{t_k} \le \sigma_+^{(g)}\Delta(\mathcal{G})$. We have

$$\begin{aligned}&\langle N^n,W^{(i)}\rangle_T \\ &= \underbrace{2\sqrt{r}\sum_{i=0}^{r-1}\sum_{j=2}^{K}\left[\sum_{l=1}^{j-1}\frac{(j-1)(l-1)+(K-(j-1))(K-(l-1))}{K^2}\Delta g_{t_{iK+l}}\right]\Delta\langle g,W^{(i)}\rangle_{t_{iK+j}}}_{(\mathrm{NW1})}\end{aligned}$$

$$+\underbrace{\sqrt{r}\sum_{i=1}^{r-1}\sum_{j=1}^{K}\left[\frac{K-(j-1)}{K}\sum_{l=1}^{K}\frac{l-1}{K}\Delta g_{t_{(i-1)K+l}}\right]\Delta\langle g, W^{(i)}\rangle_{t_{iK+j}}}_{\text{(NW2)}}$$

note that

$$E(\text{NW1})^2 \le \frac{4\left(\sigma_+^{(g)}\right)^4 r}{n^3} \times \sum_{i=0}^{r-1}\sum_{j=2}^{K}\sum_{l=1}^{j-1}\left[\frac{2(j-1)-K}{K^2}(l-1)+\frac{K-(j-1)}{K}\right]^2 = O_p\left(\frac{r^2K^2}{n^3}\right) = O_p\left(\frac{1}{n}\right)$$

$$E(\text{NW2})^2 \le \frac{\left(\sigma_+^{(g)}\right)^4 r}{n^3}\sum_{i=1}^{r-1}\sum_{j=1}^{K}\frac{(K-(j-1))^2}{K^2} \times \sum_{l=1}^{K}\frac{(l-1)^2}{K^2} = O_p\left(\frac{r^2K^2}{n^3}\right) = O_p\left(\frac{1}{n}\right)$$

the first equality in the first line follows the calculation of $\langle N^{(1)}, N^{(1)}\rangle_T$ in Appendix A.7.3.

Hence, $\langle N^n, W^{(i)}\rangle_T = O_p\left(\frac{1}{\sqrt{n}}\right)$, combine the result for $\langle N^{(1)}, N^{(1)}\rangle_T$ and $\langle N^{(2)}, N^{(2)}\rangle_T$, Theorem 6 follows.

A.7.6. The sketchy proof for Theorem 5

The first step in proving Theorem 5 is to write the error term

$$E_{\mathcal{T}}^n \equiv \sqrt{\frac{n}{K}}\left(\frac{n}{sK^2}U(Y,K,s,2)_{\mathcal{T}}^n - E_n^{(1)} - E_n^{(2)} - E_n^{(3)}\right)$$

as a summation of independent items (as (60) did). One intermediary (key) step is to write

$$E_{\mathcal{T}}^n = A_{\mathcal{T}}^n + B_{\mathcal{T}}^n + (\text{I})_{\mathcal{T}}^n + (\text{II})_{\mathcal{T}}^n + (\text{III})_{\mathcal{T}}^n + (\text{IV})_{\mathcal{T}}^n + o_p(1)$$

where

$$A_{\mathcal{T}}^n = \frac{1}{s}\left(\frac{n^{1/2}}{K}\right)^3\left(\frac{1}{r-s+1}\sum_{i=1}^{r-s+1}(m_i+m_{i+s})^2 - \frac{2}{\mathcal{T}}\int_0^{\mathcal{T}} h_t\,\mathrm{d}t\right)$$

$$\begin{aligned}B_{\mathcal{T}}^n = &\sum_{i=1}^{r-s+1}\sum_{l=1}^{K}\frac{(l-1)^2}{K^2}(\Delta g_{t_{(i-1)K+l}})^2 \\ &+ \sum_{i=1}^{r-s+1}\sum_{l=1}^{K}\frac{(K-(l-1))^2}{K^2}(\Delta g_{t_{(i+s-2)K+l}})^2 \\ &+ \sum_{i=1}^{r-s+1}\sum_{j=1}^{s-2}\sum_{l=1}^{K}\left[2\frac{j-1}{s-2}+\left(2\frac{l-1}{(s-2)K}-1\right)\right]^2 \\ &\times(\Delta g_{t_{(i+j-1)K+l}})^2 - E_n^{(1)} - E_n^{(2)}\end{aligned}$$

besides, $(\text{I})_{\mathcal{T}}^n = \sum_{i=1}^{r-s+1}(\text{I})_i^n$, $(\text{II})_{\mathcal{T}}^n = \sum_{i=1}^{r-s+1}(\text{II})_i^n$, $(\text{III})_{\mathcal{T}}^n = \sum_{i=1}^{r-s+1}(\text{III})_i^n$, and $(\text{IV})_{\mathcal{T}}^n = \sum_{i=1}^{r-s+1}(\text{IV})_i^n$ where

$$\begin{aligned}(\text{I})_i^n = &\frac{2r^{1/2}}{s}\sum_{l=2}^{K}\frac{l-1}{K}\Delta g_{t_{(i-1)K+l}} \times \sum_{h=1}^{l-1}\frac{h-1}{K}\Delta g_{t_{(i-1)K+h}} \\ &+ \frac{2r^{1/2}}{s}\sum_{l=2}^{K}\frac{K-(l-1)}{K}\Delta g_{t_{(i+s-2)K+l}} \\ &\times \sum_{h=1}^{l-1}\frac{K-(h-1)}{K}\Delta g_{t_{(i+s-2)K+h}} \\ &- \frac{r^{1/2}}{s}\sum_{l=1}^{K}\frac{K-(l-1)}{K}\Delta g_{t_{(i+s-2)K+l}} \\ &\times \sum_{l=1}^{K}\frac{l-1}{K}\Delta g_{t_{(i-1)K+l}}\end{aligned}$$

$$\begin{aligned}(\text{II})_i^n = &\frac{r^{1/2}}{s}\sum_{l=1}^{K}\frac{K-(l-1)}{K}\Delta g_{t_{(i+s-2)K+l}} \\ &\times \sum_{j=1}^{s-2}\sum_{l=1}^{K}\left(\frac{2(l-1)}{(s-2)K}+\frac{2(j-1)}{s-2}-1\right)\Delta g_{t_{(i+j-1)K+l}} \\ &- \frac{r^{1/2}}{s}\sum_{j=1}^{s-2}\sum_{l=1}^{K}\left(\frac{2(l-1)}{(s-2)K}+\frac{2(j-1)}{s-2}-1\right) \\ &\times \Delta g_{t_{(i+j-1)K+l}} \times \sum_{l=1}^{K}\frac{l-1}{K}\Delta g_{t_{(i-1)K+l}}\end{aligned}$$

$$\begin{aligned}(\text{III})_i^n = &\frac{2r^{1/2}}{s}\sum_{j=1}^{s-2}\sum_{l=2}^{K}\left(\frac{2(l-1)}{(s-2)K}+\frac{2(j-1)}{s-2}-1\right)\Delta g_{t_{(i+j-1)K+l}} \\ &\times \sum_{h=1}^{l-1}\left(\frac{2(h-1)}{(s-2)K}+\frac{2(j-1)}{s-2}-1\right)\Delta g_{t_{(i+j-1)K+h}}\end{aligned}$$

$$\begin{aligned}(\text{IV})_i^n = &\frac{2r^{1/2}}{s}\sum_{j=2}^{s-2}\sum_{l=1}^{K}\left(\frac{2(l-1)}{(s-2)K}+\frac{2(j-1)}{s-2}-1\right)\Delta g_{t_{(i+j-1)K+l}} \\ &\times \sum_{k=1}^{j-1}\sum_{l=1}^{K}\left(\frac{2(l-1)}{(s-2)K}+\frac{2(k-1)}{s-2}-1\right)\Delta g_{t_{(i+k-1)K+l}}.\end{aligned}$$

The next task is to show the joint asymptotics of $A_{\mathcal{T}}^n$, $B_{\mathcal{T}}^n$, $(\text{I})_{\mathcal{T}}^n$, $(\text{II})_{\mathcal{T}}^n$, $(\text{III})_{\mathcal{T}}^n$ and $(\text{IV})_{\mathcal{T}}^n$ (which is a similar task of Appendices A.7.3 and A.7.4). By Theorem 3, $A_{\mathcal{T}}^n = O_p\left(\frac{1}{s}\cdot\frac{n}{K^2}\right)$. Some calculation (omitted here) yields $B_{\mathcal{T}}^n = O_p(s/r)$, $(\text{I})_{\mathcal{T}}^n = O_p(1/s)$, $(\text{II})_{\mathcal{T}}^n = O_p(1/s^{1/2})$, $(\text{III})_{\mathcal{T}}^n = O_p(1/s^{1/2})$, and

$$(\text{IV})_{\mathcal{T}}^n \xrightarrow{\mathcal{L}-s} \mathcal{MN}\left(0, \frac{2\mathcal{T}}{9}\int_0^{\mathcal{T}}(\sigma_t^{(g)})^4\,\mathrm{d}t\right)$$

hence Theorem 5 follows. □

A.8. Proof of Lemma 3

Proof. Followed from (39), $\widehat{g}_t + (g_t - \widehat{g}_t) = \widehat{\beta}_m\widehat{\sigma}_t^2 + \widehat{\alpha}_m + (\beta\sigma_t^2 - \widehat{\beta}_m\widehat{\sigma}_t^2) + (\alpha - \widehat{\alpha}_m) + \zeta_t$, so we have

$$\begin{aligned}\eta_t^{(m)} &= \widehat{g}_t - \widehat{\beta}_m\widehat{\sigma}_t^2 - \widehat{\alpha}_m \\ &= (\widehat{g}_t - g_t) + \beta(\sigma_t^2 - \widehat{\sigma}_t^2) + \beta\widehat{\sigma}_t^2 \\ &\quad - \widehat{\beta}_m\widehat{\sigma}_t^2 + (\alpha - \widehat{\alpha}_m) + \zeta_t.\end{aligned} \tag{68}$$

Note that $\widehat{\sigma}_t^2 \xrightarrow{\mathbb{P}} \sigma_t^2$ and $\widehat{g}_t \xrightarrow{\mathbb{P}} g_t$, thus plugging (68) into (40), we get

$$\widehat{g}_t = \beta\widehat{\sigma}_t^2 + \alpha + \zeta_t + o_p(1)$$

so the estimates obtained from linear regression on the pairs $(\widehat{\sigma}_t^2, \widehat{g}_t)$'s are consistent, i.e., $\widehat{\beta}_n$ converges to β and $\widehat{\alpha}_n$ converges to α in the in-fill asymptotic setting, provided (39) holds. □

References

Admati, A.R., Pfleiderer, P., 1988. A theory of intraday patterns: volume and price variability. Rev. Financ. Stud. 1 (1), 3–40.

Aït-Sahalia, Y., Jacod, J., 2014. High-Frequency Financial Econometrics. Princeton University Press, Princeton, New Jersey.

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Rev. Financ. Stud. 18 (2), 351–416.

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2011. Ultra high frequency volatility estimation with dependent microstructure noise. J. Econometrics 160, 160–175.
Aït-Sahalia, Y., Sağlam, M., 2016. High Frequency Market Making. Working Paper.
Aït-Sahalia, Y., Yu, J., 2009. High Frequency market microstructure noise estimates and liquidity measures. Annu. Appl. Stat. 3 (1), 422–457.
Andersen, T., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. J. Empir. (4), 115–158.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. J. Amer. Statist. Assoc. 96 (453), 42–55.
Awartani, B., Corradi, V., Distaso, W., 2009. Assessing market microstructure effects via realized volatility measures with an application to the Dow Jones Industrial Average stocks. J. Bus. Econom. Statist. 27 (2), 251–265.
Bandi, F.M., Russell, J.R., 2006. Separating microstructure noise from volatility. J. Financ. Econ. 79, 655–692.
Bandi, F., Russell, J., Yang, C., 2013. Realized volatility forecasting in the presence of time-varying noise. J. Bus. Econom. Statist. 31 (3), 331–345.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., 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.
Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. J. R. Stat. Soc. Ser. B Stat. Methodol. B (64), 253–280.
Bibinger, M., Hautsch, N., Malec, P., Reiß, M., 2014. Estimating the quadratic covariation matrix from noisy observations: local method of moments and efficiency. Ann. Statist. 42 (4), 1312–1346.
Cox, J.C., Ingersoll Jr., Jonathan E., Ross, S.A., 1985. A Theory of the term structure of interest rates. Econometrica 53 (2), 385–407.
Delbaen, F., Schachermayer, W., 1994. A general version of the fundamental theorem of asset pricing. Math. Ann. (300), 463–520.
Engle, R.F., 2000. The econometrics of ultra-high-frequency data. Econometrica 68 (1), 1–22.
Foster, D., Nelson, D., 1996. Continuous record asymptotics for rolling sample variance estimators. Econometrica 64 (64), 139–174.
Glosten, L., Milgrom, P., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. J. Financ. Econ. 14, 71–100.
Gouriéroux, C., Jasiak, J., Fol, G.L., 1999. Intra-day market activity. J. Financ. Mark. (2), 193–226.
Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. J. Bus. Econom. Statist. 24 (2), 127–161.
Harrison, M., Pliska, S., 1981. Martingales and stochastic integrals in the theory of continuous trading. Stochastic Process. Appl. (11), 215–260.
Hasbrouck, J., 1993. Assessing the quality of a security market: a new approach to transaction-cost measurement. Rev. Financ. Stud. 6 (1), 191–212.
Hasbrouck, J., 2007. Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading. Oxford University Press.
Hautsch, N., 2012. Econometrics of Financial High-Frequency Data. Springer-Verlag, Berlin, Heidelberg.
Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. 6 (2), 327–343.
Jacod, J., 1994. Limit of Random Mesaures Associated with the Increments of a Browninan Semimartingale. Tech. Rep. Université De Paris II.
Jacod, J., 2012. Statistics and high frequency data. In: Statisitical Methods for Stochastic Differential Equations. In: Monographs on Statistics & Applied Probability (Book 124), Chapman & Hall/CRC, pp. 191–310 (Chapter 3).
Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Process. Appl. 119, 2249–2276.
Jacod, J., Podolskij, M., Vetter, M., 2010. Limit theorems for moving averages of discretized processes plus noise. Ann. Statist. 38 (3), 1478–1545.
Jacod, J., Protter, P., 1998. Asymptotic error distributions for the Euler method for stochastic differencial equations. Ann. Probab. 26 (26), 267–307.
Jacod, J., Protter, P., 2012. Discretization of Processes. Springer-Verlag, Berlin, Heidelberg.
Jacod, J., Shiryaev, A., 2003. Limit Theorems fo Stochastic Processes, second ed. Springer-Verlag, Berlin, Heidelberg.
Kalnina, I., Linton, O., 2008. Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. J. Econometrics 147, 47–59.
Kyle, A., 1985. Continuous auctions and insider trading. Econometrica 53 (6), 1315–1335.
Li, Y., Mykland, P.A., 2007. Are volatility estimators robust with respect to modeling assumption? Bernoulli 13 (3), 601–622.
Mykland, P.A., Zhang, L., 2006. ANOVA for diffusions and Itô processes. Ann. Statist. 34 (4), 1931–1963.
Mykland, P.A., Zhang, L., 2012. The econometrics of high frequency data. In: Statisitical Methods for Stochastic Differential Equations. In: Monographs on Statistics & Applied Probability (Book 124), Chapman and Hall/CRC, pp. 109–190 (Chapter 2).
Mykland, P.A., Zhang, L., 2016. Assessment of Uncertainty in High Frequency Data: The Observed Asymptotic Variance. Working Paper.
O'Hara, M., 1995. Market Microstructure Theory.
O'Hara, M., 2003. Liquidity and price discovery. J. Finance LVIII (4), 1335–1354.
Podolskij, M., Vetter, M., 2009. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Bernoulli 15 (3), 634–658.
Russell, J.R., Engle, R.F., 2010. Analysis of high-frequency data. In: Handbook of Financial Econometrics. Volume 1 — Tools and Techniques. In: Handbooks in Finance, North-Holland, pp. 383–426.
Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. J. Econometrics 159, 235–250.
Zhang, L., 2001. From Martingales to ANOVA: Implied and Realized Volatility. (Ph.D. thesis) The University of Chicago.
Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli 12 (6), 1019–1043.
Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale two time scales: determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100 (472), 1394–1411.