Efficient Estimation of Integrated Volatility Incorporating Trading Information *

Yingying Li† Shangyu Xie‡ Xinghua Zheng§

May 10, 2016

Abstract

We consider a setting where market microstructure noise is a parametric function of trading information, possibly with a remaining noise component. Assuming that the remaining noise is $O_p(1/\sqrt{n})$, allowing irregular times and jumps, we show that we can estimate the parameters at rate n, and propose a volatility estimator which enjoys $\sqrt{n}$ convergence rate. Simulation studies show that our method performs well even with model misspecification and rounding. Empirical studies demonstrate the practical relevance and advantages of our method. Furthermore, we find that a simple model can account for a high percentage of the total variation in microstructure noise.

Key Words: High frequency data, integrated volatility, market microstructure noise, realized volatility, efficiency.

JEL Classification: C14, C13, D40,

*Yingying Li's research is supported in part by RGC grants GRF 602710, GRF 606811 and GRF 16502014 of the HKSAR; Shagnyu Xie's research is supported by National Natural Science Foundation of China (71203025) and the State Key Program of National Natural Science Foundation of China (71331006), and the Fundamental Research Funds for the Central Universities in UIBE (16YQ05); Xinghua Zheng's research is supported in part by RGC grants GRF 602710, GRF 606811 and GRF 16305315 of the HKSAR.

†Hong Kong University of Science and Technology, HKSAR, China. yyli@ust.hk

‡RCAF and School of Banking and Finance, University of International Business and Economics, Beijing, China. xshyu@amss.ac.cn

§Hong Kong University of Science and Technology, HKSAR, China. xhzheng@ust.hk

1. Introduction

High-frequency data has attracted tremendous attention in recent years. In the vast literature of high frequency data studies, a central focus is to estimate volatilities consistently and efficiently. A major challenge arises from the presence of market microstructure noise, which is an integral part of the financial market.

A widely used assumption about market microstructure noise in the volatility estimation literature is that they are independent and identically distributed (i.i.d.) and additive to the log-price process. More specifically, over a time interval of interest $[0, T]$, one observes at times $0 = t_0 < t_1 < \ldots < t_n = T$,

$$Y_{t_k} = X_{t_k} + \varepsilon_{t_k}, \quad k = 0, 1, \cdots, n, \tag{1}$$

where X_{t_k} and ε_{t_k} denote the latent log-price and market microstructure noise at the observation time t_k respectively, and ε_{t_k}'s are i.i.d. and independent of X. Consistent estimators of the integrated volatility under this setting include the two-scales realized volatility (TSRV, Zhang et al. (2005)), multi-scale realized volatility (MSRV, Zhang (2006)), realized kernels (RK, Barndorff-Nielsen et al. (2008)), pre-averaging estimator (PAV, Jacod et al. (2009) and Podolskij and Vetter (2009)), and quasi-maximum likelihood estimator (QMLE, Xiu (2010)). The optimal rate of convergence is $n^{1/4}$ (Gloter and Jacod (2001)). MSRV, RK, PAV and QMLE are all rate-optimal.

On the other hand, studies on market microstructure noise can be traced back to the 1980s; see, Black (1986), Madhavan (2000), O'Hara (1995), Stoll (2003), and Hasbrouck (2007), among many others. An example of a simple model for microstructure noise is the "implicit measure of the effective bid-ask spread" as in Roll (1984):

$$\varepsilon_{t_k} = \alpha I_{b/s}(t_k), \tag{2}$$

where $I_{b/s}(t_k)$ denotes the trade type, indicating if the trade is buyer-initiated $(+1)$ or seller-initiated (-1); and the coefficient α can be interpreted as one-half of the effective bid-ask spread. Roll's model was extended in Glosten and Harris (1988) by incorporating the trading volume:

$$\varepsilon_{t_k} = I_{b/s}(t_k)\big(\alpha + \beta V_{t_k}\big), \tag{3}$$

where V_{t_k} denotes the trading volume at time t_k. Almgren and Chriss (2000) consider an optimal execution problem and they model the market impact as a function of trade type and trading rate. A variant in the spirit of (3) is then the following

$$\varepsilon_{t_k} = I_{b/s}(t_k)\big(\alpha + \beta V_{t_k}/\Delta t_k\big), \tag{4}$$

where $\Delta t_k := t_k - t_{k-1}$ denotes the duration between two consecutive transactions. A pioneer paper in high-frequency volatility estimation literature Aït-Sahalia et al. (2005) also models the market microstructure noise in a parametric way (but without covariates), and shows that even with model misspecification, such parametric modeling enables one to estimate the volatility at the optimal rate $n^{1/4}$ under Model (1).

Rich information is available in high-frequency data. For example, in trade data, in addition to transaction prices, trading volumes are also reported. Furthermore, quotes data are also publicly available, which contain even richer information. Individuals or institutions can also have additional trading information. This motivates us to consider taking advantage of the rich information available in the market and study the setting where the noise term in (1) can be further modeled using available trading information through a parametric function such as (2), (3) or (4). The function can be either linear or nonlinear. We show that in this case, even with irregular observation times and jumps, the parameters in the noise model can be estimated with high precision (with convergence rate n instead of $\sqrt{n}$ as in usual parametric estimations), and consequently the "latent log-prices" can be estimated highly accurately. This allows us to further obtain an efficient volatility estimator, based on the estimated log-prices. We call this estimator "estimated-price realized volatility" (ERV). We show that the proposed ERV, which is based on noisy observations, provides $\sqrt{n}$ rate of convergence and the same asymptotic properties as realized volatility (RV) based on latent log-prices.

Given the complexity of market microstructure noise, we further consider the setting where market microstructure noise admits an extra noise component. Under the assumption that the extra noise component is $O_p(1/\sqrt{n})$, we propose another volatility estimator ERV_{ext} which still enjoys $\sqrt{n}$ rate of convergence. Numerically, we demonstrate that ERV_{ext} (and E-QMLE, another estimator that we propose without establishing its asymptotic properties) performs well even in the situations where there are rounding errors and model misspecification on the parametric model. More importantly, extensive empirical studies demonstrate the relevance of our method and the advantages of our estimator. An interesting additional empirical finding is that, for various stocks examined, a simple model for market microstructure noise, which incorporates only trade type and trading rate, can account for around 70%-80% of the total variation in noise. Our analysis also provides a useful framework for studying the market microstructure.

An independent and concurrent research, Chaker (2013), shares the same spirit as this paper. There are however quite a few major differences. In this paper, the models for market microstructure noise are allowed to be nonlinear[1]; the observation times are allowed to be irregularly spaced, in fact the observation times can even be endogenous as what is considered in Li et al. (2014) and Li et al. (2013); and jumps are allowed in the latent price process. Furthermore, our small additional noise assumption leads to rather different estimators and asymptotic properties. Some earlier works along this line include Hansen and Lunde (2006) and Engle and Sun (2007). Hansen and Lunde (2006) consider in Section 6 of their paper how to estimate the efficient prices from bid and ask quotes and transaction prices, based on a vector autoregressive model. Engle and Sun (2007) use GARCH model for the efficient price process and a two-component ARMA model for the noise. Theoretical properties about the related estimators have not been discussed.

The rest of this paper is organized as follows. Section 2 presents our proposed ERV estimator and its extensions, together with their asymptotic properties. Sections 3 and 4 are devoted to simulation studies and empirical studies, respectively. Section 5 concludes and discusses related issues. Proofs are given in the Appendix.

2. Estimated-price Realized Volatility

2.1. When noise can be completely modeled

We assume that the latent log-price process has the following representation:

$$dX_t = \mu_t \, dt + \sigma_t \, dW_t + dJ_t, \quad t \in [0, T], \tag{5}$$

where W_t is a Brownian motion, μ_t and σ_t are adapted locally bounded random processes, and J_t is a pure jump process, all defined on a common filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. The quantity of interest is the quadratic variation (QV) $\int_0^T \sigma_t^2 dt + \sum_{t\leq T}(\Delta J_t)^2$ with $\Delta J_t := J_t - J_{t-}$, or more often, the continuous part of QV, commonly referred to as the integrated volatility $\text{IV} := \int_0^T \sigma_t^2 dt$. Without loss of generality, we set $T = 1$.

[1] Nonlinear models are relevant in practice. For example, Keim and Madhavan (1996) show that the price impact of block trades is a concave function of order size.

Following Li et al. (2014) we shall allow the observation times to be endogenous and adopt some of the notation therein. Denote the observation times at stage n by

$$0 = t_{n,0} < t_{n,1} < \ldots < t_{n,N} \leq 1. \tag{6}$$

Here n is a latent number that characterizes the observation frequency, and $N = N(n)$, which may be random, stands for the actual number of observations before time 1. See Section 3 of Li et al. (2014) for various examples in this regard. In the exogenous case when observation times are either deterministic or random but independent of the price process, without loss of generality, we can and will take $n = N$. More generally, we will establish a feasible asymptotic theory in terms of N under the assumption that N/n has a (possibly random) probability limit F. Let us mention that in the endogenous setting, while in general n may not be uniquely defined[2], the feasible asymptotic theory will be independent of n (see also Remark 1 in Li et al. (2014) or the discussion following Assumption (O) in Jacod et al. (2014)). For notational ease, when there is no confusion we shall write $t_{n,1}, t_{n,k}$ as t_1, t_k etc.

In this subsection we consider the setting where the market microstructure noise can be completely modeled by trading information, through a parametric function g:

$$Y_{t_k} = X_{t_k} + g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0), \tag{7}$$

where Y_{t_k} is the observed log-prices at time t_k, $\mathbf{Z}_{t_k}$ is the information set which can include, but not limited to, trade type, trading volume, and bid-ask bounds; and $\boldsymbol{\theta}_0$ is a (finite-dimensional) parameter. The aforementioned Models (2), (3), and (4) are all examples of g. We shall also consider some other forms of g in the numerical studies in Section 3. In our theoretical analysis, we allow the function g to be of any parametric form $g(\mathbf{Z}; \boldsymbol{\theta})$ (satisfying certain mild conditions to be specified later).

We first discuss how to estimate the parameter $\boldsymbol{\theta}_0$. Denote $\Delta X_{t_k} = X_{t_k} - X_{t_{k-1}}$, $\Delta Y_{t_k} = Y_{t_k} - Y_{t_{k-1}}$, and $\Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}) = g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}) - g(\mathbf{Z}_{t_{k-1}}; \boldsymbol{\theta})$. One then has

$$\Delta Y_{t_k} = \Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0) + \Delta X_{t_k}. \tag{8}$$

[2]To see this, similar to Examples 4 - 6 in Li et al. (2014), define the observation times $t_{n,i}$ to be successive hitting times: $t_{n,i+1} := \inf\{t > t_{n,i} : |X_t - X_{t_{n,i}}| \geq Z_{i+1}/\sqrt{n}\}$, where Z_i's are random variables which may or may not be i.i.d. Such a definition suggests that n is a natural characterization of the observation frequency. However, if another person takes $\widetilde{Z}_{i+1} = \sqrt{2} Z_{i+1}$, then $t_{n,i+1}$ can be equivalently defined as $\widetilde{t}_{2n,i+1} := \inf\{t > \widetilde{t}_{2n,i} : |X_t - X_{\widetilde{t}_{2n,i}}| \geq \widetilde{Z}_{i+1}/\sqrt{2n}\}$. The latter definition suggests $2n$ as another characterization of the observation frequency.

To estimate the parameter $\boldsymbol{\theta}_0$, we first change our viewpoint by regarding $\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0)$ as the "signal" and ΔX_{t_k} as the "noise". Observe that in the simplest case when the drift $\mu_t \equiv 0$, spot volatility $\sigma_t \equiv \sigma$ and the observation times $t_k = k/n$ for $k = 0, 1, \ldots, n$, we have that the "noise" $\Delta X_{t_k} \sim N(0, \sigma^2/n)$. Therefore if further ΔX_{t_k} is independent of $\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0)$, then the maximum-likelihood estimator (MLE) $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ is given by

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} Q_N(Y, \mathbf{Z}, \boldsymbol{\theta}), \quad \text{where} \quad Q_N(Y, \mathbf{Z}, \boldsymbol{\theta}) = \frac{1}{2}\sum_{k=1}^{N} \left(\Delta Y_{t_k} - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta})\right)^2. \tag{9}$$

When g is a linear function such as in (2), (3), and (4), Eq. (8) becomes a linear regression, and $\widehat{\boldsymbol{\theta}}$ is explicitly given by

$$\widehat{\boldsymbol{\theta}}_L = \left(\sum_{k=1}^{N} \Delta\mathbf{Z}_{t_k}^T \Delta\mathbf{Z}_{t_k}\right)^{-1} \cdot \left(\sum_{k=1}^{N} \Delta\mathbf{Z}_{t_k}^T \Delta Y_{t_k}\right). \tag{10}$$

We propose to estimate $\boldsymbol{\theta}_0$ defined in (9) for the general setup as well. We will show that the estimator $\widehat{\boldsymbol{\theta}}$ has a convergence rate of n.

2.1.1. The estimated-price realized volatility

It has long been established that realized volatility (RV)

$$\mathrm{RV} := \sum_{k=1}^{N} \Delta X_{t_k}^2 \tag{11}$$

is an efficient estimator of the quadratic variation. In practice, we observe Y_{t_k} instead of X_{t_k}; however by using the aforementioned model for microstructure noise (7) and the estimator $\widehat{\boldsymbol{\theta}}$, we can estimate the latent log-prices by

$$\widehat{X}_{t_k} := Y_{t_k} - g(\mathbf{Z}_{t_k};\widehat{\boldsymbol{\theta}}), \tag{12}$$

which leads to our proposed estimator, the estimated-price realized volatility (ERV)

$$\mathrm{ERV} = \sum_{k=1}^{N} \Delta\widehat{X}_{t_k}^2, \quad \text{where} \quad \Delta\widehat{X}_{t_k} = \widehat{X}_{t_k} - \widehat{X}_{t_{k-1}}. \tag{13}$$

To deal with jumps, following Mancini (2009) and Aït-Sahalia and Jacod (2009), for a suitable exponent $\xi > 0$ to be specified below, we define the thresholded ERV as follows:

$$\mathrm{ERV}_{threshold} = \sum_{k=1}^{N} \Delta\widehat{X}_{t_k}^2 \cdot \mathbf{1}_{|\Delta\widehat{X}_{t_k}| \le (\Delta t_k)^{\xi}}. \tag{14}$$

Alternatively, one can also use bi- and multi-power variation (Barndorff-Nielsen and Shephard (2004), Barndorff-Nielsen et al. (2006), Barndorff-Nielsen et al. (2006)). A slight difference is that thresholded ERV is more convenient to use in the irregular time setting.

We next state our assumptions, which can be categorized into two sets: Assumption A is about X, g, and $\mathbf{Z}_{t_k}$, and Assumption B is about observation times $\{t_k : k = 0, 1, \ldots, N\}$.

We will use the following notation: $\xrightarrow{p}$ stands for convergence in probability, $\xrightarrow{\mathcal{L}}$ represents convergence in law, $Y_n = o_p(f(n))$ means that $Y_n/f(n) \xrightarrow{p} 0$, and $Y_n = O_p(f(n))$ means that the sequence $|Y_n|/f(n)$ is tight. With a slight abuse of notation, we use $|\cdot|$ to denote the Euclidean norm in all dimensions.

Assumption A:

(A.i) (μ_t) is locally bounded;

(A.ii) (σ_t) is locally bounded with $\inf_{t\in(0,1]} \sigma_t > 0$ almost surely;

(A.iii) (J_t) satisfies that $\sum_t |\Delta J_t| = O_p(1)$;

(A.iv) For all k, $\mathbf{Z}_{t_k}$ and ΔX_{t_k} are conditionally independent given $\mathcal{F}_{t_{k-1}}$;

(A.v) $\max_k |\mathbf{Z}_{t_k}| = O_p(1)$;

(A.vi) The parameter space Θ for $\boldsymbol{\theta}$ is a compact set in $\mathbb{R}^d$ for some $d \in \mathbb{N}$, and $g(\mathbf{z}; \boldsymbol{\theta})$ is twice continuously differentiable in $\boldsymbol{\theta}$ in a neighborhood $\mathcal{N}(\boldsymbol{\theta}_0) \subset \Theta$;

(A.vii) For all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{N}(\boldsymbol{\theta}_0)$, $|g(\mathbf{z}; \boldsymbol{\theta}_1) - g(\mathbf{z}; \boldsymbol{\theta}_2)| \le L_0(\mathbf{z})|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|$, where $L_0(\mathbf{z})$ is locally bounded;

(A.viii) For all $\boldsymbol{\theta} \in \mathcal{N}(\boldsymbol{\theta}_0)$, $|\frac{\partial g}{\partial \boldsymbol{\theta}}(\mathbf{z}; \boldsymbol{\theta})| \le f(\mathbf{z})$ where $f(\mathbf{z})$ is locally bounded;

(A.ix) $|\frac{\partial g}{\partial \boldsymbol{\theta}}(\mathbf{z}; \boldsymbol{\theta}_1) - \frac{\partial g}{\partial \boldsymbol{\theta}}(\mathbf{z}; \boldsymbol{\theta}_2)| \le L_1(\mathbf{z})|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|$, where $L_1(\mathbf{z})$ is locally bounded;

(A.x) For any $\varepsilon > 0$, almost surely,

$$\inf_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\ge\varepsilon} \sum_{k=1}^{N} |\Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}) - \Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0)|^2 \to \infty \quad \text{as} \quad n \to \infty;$$

(A.xi) $\left|\left|\left(\frac{1}{N}\sum_{k=1}^{N} \Delta \frac{\partial g}{\partial \boldsymbol{\theta}}(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0) \Delta \frac{\partial g}{\partial \boldsymbol{\theta}^T}(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0)\right)^{-1}\right|\right| = O_p(1)$, where for any square matrix A, $||A||$ stands for its spectral norm;

(A.xii) For any $i, j = 1, \ldots, d$, for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{N}(\boldsymbol{\theta}_0)$, $\left|\frac{\partial^2 g}{\partial\theta_i\partial\theta_j}(\mathbf{z}; \boldsymbol{\theta}_1) - \frac{\partial^2 g}{\partial\theta_i\partial\theta_j}(\mathbf{z}; \boldsymbol{\theta}_2)\right| \le L_2(\mathbf{z})|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|$, where $L_2(\mathbf{z})$ is locally bounded;

(A.xiii) For any $i, j = 1, \ldots, d$, for all $\boldsymbol{\theta} \in \mathcal{N}(\boldsymbol{\theta}_0)$, $\left|\frac{\partial^2 g}{\partial\theta_i\partial\theta_j}(\mathbf{z}; \boldsymbol{\theta})\right| \le L_3(\mathbf{z})$, where $L_3(\mathbf{z})$ is locally bounded.

Remark 1. *Assumption* (A.iv) *is analogous to the usual assumption in regression that the predictor and noise are independent. Under Models* (2)*,* (3) *and* (4)*, the assumption amounts* *to assume that the immediate next trading depends solely on the market information up to the latest transaction. We believe this is a reasonable assumption when transactions occur at high frequency, and particularly so in algorithmic trading where whether to buy or to sell as well as how much to trade are determined by past transactions.*

Remark 2. *Assumptions* (A.vii), (A.ix), *and* (A.xii) *are actually implied by Assumption* (A.vi)*; they are explicitly stated for ease of reference in the proofs. Assumption* (A.x) *is analogous to the identifiability condition in MLE, and Assumption* (A.xi) *corresponds to the invertibility condition of the Fisher information matrix.*

Assumption B:

(B.i) The number of observations N in (6) satisfies that $N/n \xrightarrow{p} F$ for some positive random variable F;

(B.ii) The observation times $\mathcal{G}_N := \{t_k : k = 0, 1, \ldots, N\}$ is nonrandom or independent of the process X;

(B.iii) Let $\Delta t_k = t_k - t_{k-1}$ for all $k = 1, 2, \ldots$. Then $\Delta(\mathcal{G}) := \max_k \Delta t_k$ satisfies that $\sqrt{N}\Delta(\mathcal{G}) = o_p(1)$ and $N^2 \sum_{k=1}^{N} (\Delta t_k)^3 = O_p(1)$;

(B.iv) The asymptotic quadratic variation of time (AQVT) H_t (see Mykland and Zhang (2006)) exists:

$$H_t = \lim_{n\to\infty} N \sum_{t_k \le t} (\Delta t_k)^2, \quad t \in [0, 1].$$

(B.v) The shortest inter-observational period $\delta(\mathcal{G}) := \min_k \Delta t_k$ satisfies that $N\sqrt{\delta(\mathcal{G})} \xrightarrow{p} \infty$.

Remark 3. *Assumption B is* not *required in establishing the convergence rate of the estimator* $\widehat{\boldsymbol{\theta}}$*, and Assumption* (B.i) alone *is sufficient in establishing the convergence rate of ERV/ERV$_{threshold}$. Assumptions* (B.ii)$\sim$(B.iv) *are used only in establishing a special form of the central limit theorem (CLT) for ERV/ERV$_{threshold}$, see Remark 6 below. Recall also that if Assumption*(B.ii) *holds, then one can take $n = N$ and hence Assumption* (B.i) *automatically holds with $F \equiv 1$.*

Note that in Assumption (B.iii) we write $\sqrt{N}\Delta(\mathcal{G}) = o_p(1)$ instead of $\Delta(\mathcal{G}) = o_p(1/\sqrt{N})$ because in our setting N is allowed to be a random variable. The same remark applies to $N^2 \sum_{k=1}^{N} (\Delta t_k)^3 = O_p(1)$ and other similar statements below.

We now state our first set of main results. The first result concerns the convergence rate of $\widehat{\boldsymbol{\theta}}$.

Theorem 1. *Under Assumption A, we have that* $N(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1)$.

Remark 4. *We do mean* $N(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1)$ *instead of* $\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1)$ *(which is what one has in usual parametric estimations). The reason for such a higher convergence rate, intuitively speaking, is that in regression* (8)*, in the simple case when observation times are equally spaced, the noise term* ΔX_{t_k} *is of order* $O_p(\sqrt{\Delta t_k}) = O_p(\sqrt{1/n})$*, which contributes an extra* $O_p(\sqrt{1/n})$*. In general, when observation times can be irregular, the extra* $O_p(\sqrt{1/n})$ *is because the sum of squared "noise"* $\sum_{k=1}^{N} \Delta X_{t_k}^2$ *is* $O_p(1)$ *instead of* $O_p(n)$*, which is the rate of growth in usual regression settings.*

Remark 5. *CLTs for* $\widehat{\boldsymbol{\theta}}$ *can be readily derived* if *one is willing to impose convergence properties on* $\{\mathbf{Z}_{t_k}\}$*. We do not pursue this because we are considering a general model and* $\{\mathbf{Z}_{t_k}\}$ *might contain information for which there is no consensus on how it should be modeled.*

To state our next result concerning the ERV/ERV$_{threshold}$ estimator, we first recall the definition of stable convergence.

Definition 1. *Let* U_n *be a sequence of* χ*-measurable variables,* $\mathcal{F}_1 \subseteq \chi$*. We say that* U_n *converges* $\mathcal{F}_1$*-stably (or stably) in law to* U *as* $n \to \infty$ *if* U *is measurable with respect to an extension of* χ*, so that for all* $A \in \mathcal{F}_1$ *and for any bounded continuous function* h*,* $E(\mathbf{1}_A h(U_n)) \to E(\mathbf{1}_A h(U))$ *as* $n \to \infty$*, where* $\mathbf{1}_{\cdot}$ *stands for the indicator function.*

Theorem 2. *Under Assumption A,*

(i) if Assumption (B.i) *holds, then as* $n \to \infty$*,*

$$\sqrt{N}(ERV - RV) = o_p(1); \tag{15}$$

(ii) in particular, if Assumptions (B.ii)$\sim$(B.iv) *hold and* $J_t \equiv 0$*, then as* $n \to \infty$*,*

$$\sqrt{N}\,(ERV - IV) \xrightarrow{\mathcal{L}} \Phi \times \left(2\int_0^1 \sigma_t^4 \; dH_t\right)^{1/2} \quad \textit{stably}, \tag{16}$$

where Φ *is a standard normal random variable independent of* $\mathcal{F}_1$*;*

(iii) more generally, if (J_t) *admits only finitely many jumps, then under Assumptions* (B.ii)$\sim$(B.v)*, for any* $\xi \in (0, 1/2)$*, as* $n \to \infty$*,*

$$\sqrt{N}\,(ERV_{threshold} - IV) \xrightarrow{\mathcal{L}} \Phi \times \left(2\int_0^1 \sigma_t^4 \; dH_t\right)^{1/2} \quad \textit{stably}. \tag{17}$$

Remark 6. *If the observation times are endogenous as what is considered in Li et al. (2014), then* $\sqrt{N}\,(ERV_{threshold} - IV)$ *will converge to the limit with* $t = 1$ *in Eq.(9) in Li et al. (2014) .*

Remark 7. *In practice, the threshold* Δt_k^{ξ} *in* (14) *can be chosen in a data-driven way, for example,* $\widetilde{\sigma} \cdot \Delta t_k^{\xi}$ *where* $\widetilde{\sigma}$ *is a rough estimate of the daily volatility, or as in Aït-Sahalia et al. (2013), chosen as* $4\widetilde{\sigma} \cdot \Delta t_k^{1/2}$.

Furthermore we have a feasible CLT for $\mathrm{ERV}_{threshold}$ as follows.

Proposition 1. *Under Assumptions A and B, if* (J_t) *admits only finitely many jumps, then for any* $\xi \in (0, 1/2)$*, as* $n \to \infty$*, the statistic*

$$N_1^n := \frac{ERV_{threshold} - IV}{\sqrt{\frac{2}{3}\sum_{k=1}^{N} |\Delta \widehat{X}_{t_k}|^4 \cdot \mathbf{1}_{|\Delta \widehat{X}_{t_k}| \le (\Delta t_k)^{\xi}}}} \xrightarrow{\mathcal{L}} \Phi \text{ stably.} \tag{18}$$

2.2. Extensions: when there is an extra noise component

What if model (7) is not sufficient? In other words, what if the noise admits another source of error which cannot be explained by past trading information? Here we consider the following extension to model (7):

$$Y_{t_k} = X_{t_k} + g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0) + \varepsilon_{t_k}, \tag{19}$$

where (ε_{t_k}) are i.i.d. with mean 0, standard deviation σ_n, and also independent of $\mathcal{F}_1$. We propose the following approach in such a setting:

(i). Estimate the parameter $\boldsymbol{\theta}_0$ using (9). Define the estimated log-prices $\widehat{X}_{t_k}$ just as in (12), which is now an estimate of $X_{t_k} + \varepsilon_{t_k}$.

(ii). Then apply existing estimators for the noisy setting (such as TSRV, MSRV, RK, PAV and QMLE) to $\widehat{X}_{t_k}$.

The advantages of this strategy, compared with directly applying existing estimators to the raw log-prices, are twofold:

(i). Modeling and removing the noise component caused by trading and then applying existing estimators to the estimated log-prices yields a higher efficiency;

(ii). The potential serial dependence as well as time variation in market microstructure noise (see, e.g., Hansen and Lunde (2006), Jacod et al. (2014)) may be caused by dependence and variation in trades. Filtering out the component caused by trading can help remove the dependence and idiosyncrasies in the noise, consequently make the remaining noise more like white noise[3].

On the other hand, under model (19), if one finds it reasonable to impose an additional assumption on (ε_{t_k}) that

$$n\sigma_n^2 := nE(\varepsilon_{t_k}^2) \to \sigma_\varepsilon^2 \in [0,\infty), \quad \text{and} \quad n^2E(\varepsilon_{t_k}^4) = O(1), \tag{20}$$

then a more convenient estimator of IV is the following

$$\begin{aligned} \mathrm{ERV}_{ext} = &\sum_{k=1}^{N} \left(\Delta\widehat{X}_{t_k}\right)^2 \cdot \mathbf{1}_{|\Delta\widehat{X}_{t_k}|\le(\Delta t_k+1/\sqrt{N})^\xi} \\ &+ 2\sum_{k=2}^{N} \Delta\widehat{X}_{t_k}\cdot\Delta\widehat{X}_{t_{k-1}}\mathbf{1}_{|\Delta\widehat{X}_{t_k}|\le(\Delta t_k+1/\sqrt{N})^\xi,\ |\Delta\widehat{X}_{t_{k-1}}|\le(\Delta t_{k-1}+1/\sqrt{N})^\xi}. \end{aligned} \tag{21}$$

Remark 8. *Condition (20) is similar in spirit to the small (rounding) error assumption in Delattre and Jacod (1997) where the authors assume that the rounding level α_n satisfies that $\sqrt{n}\alpha_n \to \beta \in [0,\infty)$. Similar assumptions appeared in Hansen et al. (2008), Hansen and Horel (2009) and Large (2011), among others. More generally, small noise assumption has appeared in, for example, Aït-Sahalia et al. (2005) (who consider in Section 9.2 a setting when the variance contributed by the noise to the observed return is of the same order as the latent return), Barndorff-Nielsen et al. (2008) (who consider in Section 4.7 a setting when the noise is of order $n^{-\alpha}$ for $\alpha \in [0,1/2)$, hence the (total) noise is small), Rosenbaum (2009) (who studies the case when the rounding level goes to zero), Li and Mykland (2007), Li and Mykland (2015) and Li et al. (2015) (who consider both rounding and noise). Our motivation for considering such a small additional noise assumption comes from the intuition that if we can model a big portion of noise by trading information via the function g, then it is reasonable to view what is left as a quantity of a smaller order. In fact, under model* (19)*, a useful concept is the following proportion*

$$\pi_{exp} := \frac{g_V}{g_V+\varepsilon_V}, \quad \textit{where} \quad g_V = \sum_{k=1}^{N}(g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0))^2 \quad \textit{and} \quad \varepsilon_V = \sum_{k=1}^{N}\varepsilon_{t_k}^2. \tag{22}$$

[3]This is desirable, despite that several volatility estimators have been designed to accommodate serially dependent and/or time-varying noise. Such estimators include the (multivariate) realized kernel estimator in Barndorff-Nielsen et al. (2011), the Markov chain estimator in Hansen and Horel (2009), and moving average-based estimator in Hansen et al. (2008), the extension of TSRV in Aït-Sahalia et al. (2011) and the extension of PAV in Jacod *et al.* (2015). These estimators are all more sophisticated than their counterparts when microstructure noise is white.

The two quantities, g_V and ε_V, can be interpreted as follows: g_V stands for the total variation in the explained noise component g, and ε_V is the total variation in the unexplained noise component (ε_{t_k}). The ratio π_{exp} is hence analogous to the coefficient of determination (R-squared) in least squares regression, and measures the proportion of variation that is explained by the model (g in our setting). The quantities in (22) *can be estimated by*

$$\widehat{\pi_{exp}} := \frac{\widehat{g_V}}{\widehat{g_V} + \widehat{\varepsilon_V}} \quad \textit{where} \quad \widehat{g_V} := \sum_{k=1}^{N} (g(\mathbf{Z}_{t_k}, \widehat{\boldsymbol{\theta}}))^2 \quad \textit{and} \quad \widehat{\varepsilon_V} := \frac{\sum_{k=1}^{N} (\Delta \widehat{X}_{t_k})^2 - ERV_{ext}}{2}. \tag{23}$$

In our empirical studies, for various stocks examined, we find that with a simple function g, the proportion of explained variation is around 70% – 80%, indicating that the variation left in (ε_{t_k}) only accounts for a small proportion of total variation in the noise.

Remark 9. *The extra $1/\sqrt{N}$ in the threshold $(\Delta t_k + 1/\sqrt{N})^\xi$ in* (21) *is used to account for the extra term $\Delta\varepsilon_{t_k} := \varepsilon_{t_k} - \varepsilon_{t_{k-1}}$ in $\Delta\widehat{X}_{t_k}$ which approximates $\Delta X_{t_k} + \Delta\varepsilon_{t_k}$. Under the assumption on the fourth moment in* (20), *we have*

$$P\big(|\Delta\varepsilon_{t_k}| \geq (1/\sqrt{n})^\xi\big) \leq n^{2\xi} E|\Delta\varepsilon_{t_k}|^4 = O(1/n^{2-2\xi}).$$

Therefore, recall that N/n converges in probability to a positive (random) variable F, we obtain $P\big(\max_{k=1,\ldots,N} |\Delta\varepsilon_{t_k}| \geq (1/\sqrt{n})^\xi\big) = o(1)$ if $\xi < 1/2$. Using again $N/n \xrightarrow{p} F > 0$ we see that

$$P\big(\max_k |\Delta\varepsilon_{t_k}| \geq (1/\sqrt{N})^\xi\big) = o(1) \quad \textit{if} \quad \xi < 1/2. \tag{24}$$

In practice, the threshold can again be chosen in a data-driven way. Under subgaussian tail assumption on (ε_{t_k}), the threshold can be chosen as $4\max\Big(\sqrt{\widehat{B}\Delta t_k}, \sqrt{\widehat{B}/N}\Big)$, where $\widehat{B} = \sum_{k=1}^{N} \Delta\widehat{X}_{t_k}^2$, which estimates $IV + 2F\sigma_\varepsilon^2 + \sum_t (\Delta J_t)^2$.

When jump does not exist, for all n large enough, ERV_{ext} reduces to

$$\mathrm{ERV}_{ext,\ no\ jump} := \sum_{k=1}^{N} \big(\Delta\widehat{X}_{t_k}\big)^2 + 2\sum_{k=2}^{N} \Delta\widehat{X}_{t_k} \cdot \Delta\widehat{X}_{t_{k-1}}, \tag{25}$$

which equals $\mathrm{RV}_{ext,\ no\ jump} + o_p(1)$, where

$$\mathrm{RV}_{ext,\ no\ jump} := \sum_{k=1}^{N} \big(\Delta X_{t_k} + \Delta\varepsilon_{t_k}\big)^2 + 2\sum_{k=2}^{N} \big(\Delta X_{t_k} + \Delta\varepsilon_{t_k}\big)\big(\Delta X_{t_{k-1}} + \Delta\varepsilon_{t_{k-1}}\big), \tag{26}$$

see (75) in the proof of Theorem 3. Observe that ERV (as defined in (13)) converges to $\mathrm{IV}+2F\sigma_\varepsilon^2$ (see (78) below), hence it is no longer a consistent estimator of IV unless $\sigma_\varepsilon = 0$.

We impose the following additional assumptions on the observation times.

(B.vi) $N \sum_k \Delta t_{k-1} \Delta t_k = O_p(1)$;

(B.vii) The *neighboring-variation of time (NQVT)* Q_t exists:

$$Q_t = \lim_{n\to\infty} N \sum_{t_k \le t} \Delta t_{k-1} \Delta t_k, \quad t \in [0,1].$$

Remark 10. *Certainly either Assumption* (B.iv) *or* (B.vii) *implies* (B.vi)*. Note however that the two limits, H_t and Q_t, in Assumptions* (B.iv) *and* (B.vii) *could be different. A simple example is as follows: for each n, define $t_k = t_{n,k}$ so that $\Delta t_k = 1/(2n)$ when k is odd, and $\Delta t_k = 3/(2n)$ when k is even. In this case, $H_t = 5/2 \cdot t$ while $Q_t = 3/4 \cdot t$. Another example is when $(t_k = t_{n,k})$ are successive arrival times of a Poisson process with rate n. In this case, $H_t = 2t$ while $Q_t = t$.*

As to the CLT for the estimator ERV_{ext}, we will further need the volatility itself to be an Itô semimartingale. More precisely, we assume

(C.i) The process $\nu_t := \sigma_t^2$ is an Itô semimartingale satisfying Assumption (H-2) in Aït-Sahalia and Jacod (2014).

Theorem 3. *Under model* (19)*, under Assumption A and* (20)*,*

(i) if Assumption (B.i) *holds, then as $n \to \infty$,*

$$N(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1); \tag{27}$$

(ii) under Assumptions (B.i), (B.v), (B.vi) *and if (J_t) admits only finitely many jumps, then for any $\xi \in (0, 1/2)$, as $n \to \infty$,*

$$\sqrt{N}\,(ERV_{ext} - IV) = O_p(1); \tag{28}$$

(iii) under Assumptions (B.ii)∼(B.v), (B.vii) *and* (C.i)*, if (J_t) admits only finitely many jumps, then for any $\xi \in (0, 1/2)$, stably in law,*

$$\sqrt{N}\,(ERV_{ext} - IV) \xrightarrow{\mathcal{L}} \Phi \times \left(2\int_0^1 \sigma_t^4 \, dH_t + 4\int_0^1 \sigma_t^4 \, dQ_t + 8\sigma_\varepsilon^2 \; IV + 8\sigma_\varepsilon^4\right)^{1/2}, \tag{29}$$

where Φ is a standard normal random variable independent of $\mathcal{F}_1$.

As a special case of (29), in the equidistant case when $t_k = k/n$, if (J_t) admits only finitely many jumps, then, stably in law,

$$\sqrt{N}\,(\mathrm{ERV}_{ext} - \mathrm{IV}) \xrightarrow{\mathcal{L}} \Phi \times \left(6\int_0^1 \sigma_t^4 \, dt + 8\sigma_\varepsilon^2 \int_0^1 \sigma_t^2 \, dt + 8\sigma_\varepsilon^4\right)^{1/2}. \tag{30}$$

Finally we give a feasible CLT. Assume additionally that

$$n^2E(\varepsilon_{t_k}^4)\to\kappa\in[0,\infty),\quad \text{and}\quad n^3E(\varepsilon_{t_k}^6)=O(1). \tag{31}$$

We introduce the following estimator of the variance in (29):

$$\begin{aligned}\widehat{\mathrm{AVar}}_\varepsilon :=&\frac{2N}{3}\left(\sum_{k=2}^{N}(\widehat{X}_{t_k}-\widehat{X}_{t_{k-2}})^4\cdot \mathbf{1}_{|\Delta\widehat{X}_{t_k}|\le(\Delta t_k+1/\sqrt{N})^\xi,\ |\Delta\widehat{X}_{t_{k-1}}|\le(\Delta t_{k-1}+1/\sqrt{N})^\xi}\right.\\ &\left.-\sum_{k=1}^{N}(\Delta\widehat{X}_{t_k})^4\cdot\mathbf{1}_{|\Delta\widehat{X}_{t_k}|\le(\Delta t_k+1/\sqrt{N})^\xi}\right)\\ &+2\left(\sum_{k=1}^{N}(\Delta\widehat{X}_{t_k})^2\cdot\mathbf{1}_{|\Delta\widehat{X}_{t_k}|\le(\Delta t_k+1/\sqrt{N})^\xi}-\mathrm{ERV}_{ext}\right)^2.\end{aligned} \tag{32}$$

Proposition 2. *Under the assumptions of Theorem 3(iii), assuming further that* (31) *holds, then for any* $\xi\in(0,1/2)$*, as* $n\to\infty$,

$$N_{1,\varepsilon}^n:=\frac{\sqrt{N}(ERV_{ext}-IV)}{\sqrt{\widehat{AVar}_\varepsilon}}\xrightarrow{\mathcal{L}}\Phi \text{ stably.} \tag{33}$$

The significance of the extended estimator ERV_{ext} is that it is robust to small deviations from model (7). In addition, it is also seen in numerical studies that ERV_{ext} performs well even with misspecification of the noise model g. Loosely speaking, if under model (7) or model (19), the function g is misspecified, then the "residual" after fitting g will be absorbed into the ε term in (19). To avoid the risk of model misspecification, in the empirical studies below we shall adopt this estimator ERV_{ext} instead of our original estimator ERV or $\mathrm{ERV}_{threshold}$. Accordingly we only demonstrate the performance of ERV_{ext} in the simulation studies.

2.2.1. Discussion about the estimator ERV_{ext}

The estimator ERV_{ext} also fits the framework of QMLE studied in Aït-Sahalia et al. (2005) and Xiu (2010). To see this, consider the simplest setting where the latent process (X_t) satisfies that $dX_t=\sigma dW_t$, the additional noise $\varepsilon_{t_k}\sim_{i.i.d.}N(0,\sigma_\varepsilon^2/n)$, and the observation times $t_k=k/n$ for $k=0,1,\ldots,n$. In this case, $V_k:=\sqrt{n}\cdot\left(\Delta X_{t_k}+\Delta\varepsilon_{t_k}\right)$ can be written as an MA(1) process:

$$V_k=\delta_k+\eta\delta_{k-1},\quad \delta_k\sim_{i.i.d.}N(0,\gamma^2),$$

where γ^2 and η satisfy that

$$\begin{aligned}\gamma^2(1+\eta^2)&=\mathrm{Var}(V_k)=\sigma^2+2\sigma_\varepsilon^2\\ \gamma^2\eta&=\mathrm{Cov}(V_k,V_{k-1})=-\sigma_\varepsilon^2,\end{aligned}$$

see equations (8), (9) and (10) in Aït-Sahalia et al. (2005). To estimate the parameters, note that by ergodicity, we have

$$\frac{\sum_{k=1}^{n} V_k^2}{n} = \sum_{k=1}^{n} \left(\Delta X_{t_k} + \Delta \varepsilon_{t_k}\right)^2 \xrightarrow{p} \sigma^2 + 2\sigma_\varepsilon^2, \quad \text{and}$$
$$\frac{\sum_{k=2}^{n} V_k V_{k-1}}{n} = \sum_{k=2}^{n} \left(\Delta X_{t_k} + \Delta \varepsilon_{t_k}\right)\left(\Delta X_{t_{k-1}} + \Delta \varepsilon_{t_{k-1}}\right) \xrightarrow{p} -\sigma_\varepsilon^2.$$

Therefore, $\mathrm{RV}_{ext,\ no\ jump}$ defined in (26) converges to σ^2, and since $\mathrm{ERV}_{ext} = \mathrm{RV}_{ext,\ no\ jump} + o_p(1)$, so does ERV_{ext}.

Such a connection also suggests the following alternative strategy:

(i). Estimate the parameter $\boldsymbol{\theta}_0$ using (9). Define the estimated log-prices $\widehat{X}_{t_k}$ just as in (12), which is now an estimate of $X_{t_k} + \varepsilon_{t_k}$.

(ii). Then apply QMLE [4] to $\widehat{X}_{t_k}$.

We conjecture that this alternative estimator (call it, say, E-QMLE with "E" for estimated price) has the same convergence rate of $\sqrt{n}$ under condition (20). Under some situations, E-QMLE may yield a smaller asymptotic variance than ERV_{ext}, at the price of a higher computational cost. Rigorous treatment of this proposal is beyond the scope of this paper[5]. We do try this estimator in our simulation studies, and find that indeed sometimes it can yield a smaller RMSE than ERV_{ext}.

3. Simulation Studies

3.1. The feasible CLT, when required conditions are met

We conduct simulation studies to examine the performance of our proposed estimators, $\widehat{\boldsymbol{\theta}}$ and ERV_{ext}. We directly consider a challenging situation which involves irregular sampling times, jumps, and additional noise.

To motivate our simulation design, we first recall two concepts introduced in Barndorff-Nielsen et al. (2008):

$$\xi^2 = \frac{\omega^2}{\sqrt{\int_0^1 \sigma_t^4\, dt}} \quad \text{and} \quad \rho = \frac{\int_0^1 \sigma_t^2\, dt}{\sqrt{\int_0^1 \sigma_t^4\, dt}},$$

[4] One common advantage of ERV_{ext} and QMLE is that they are both free of tuning parameters.

[5] In a private communication, we learned from Dacheng Xiu that he is working on QMLE assuming the small noise assumption (20). This future result can facilitate a complete analysis of E-QMLE.

where ω^2 stands for the variance of noise, see p.1492 therein. ξ^2 can be regarded as the noise-to-signal ratio, and ρ measures the heteroskedasticity. As is explained in Barndorff-Nielsen et al. (2008), $\rho \leq 1$, and $\rho = 1$ corresponds to the constant volatility case.

We first discuss the model for the latent log-price process X. In our empirical studies, we found that ρ ranges from 0.3 to 0.7. In the simulation studies below, we choose a simple model for the volatility process to feature $\rho \approx 0.5$ and a U-shaped pattern:

$$\nu_t := \sigma_t^2 = \begin{cases} 15\zeta, & \text{if } t < 0.05 \text{ or } t \geq 0.95, \\ \zeta, & \text{otherwise,} \end{cases} \tag{34}$$

where $\zeta > 0$ is any fixed constant. Under such a setting, the ratio between the maximum and minimum (spot) volatilities is $\sqrt{15} \approx 3.9$. One can of course perturb the function above to make it continuous or even stochastic while still have a similar ρ. As to ζ which determines the volatility level, we take it to be 0.000125 so that the daily integrated volatility is 0.0003, which is a typical scale of the stocks analyzed in empirical studies.

With such a ν_t, we define our latent log-price process X to be

$$dX_t = (\mu - \nu_t/2)dt + \sigma_t dB_t + J_t \, dN_t, \tag{35}$$

where N_t is a Poisson process with intensity λ, and J_t denotes the jump size which is assumed to be independent of everything else. We assume that J_t follows a normal distribution with mean zero and variance σ_J^2. We set the parameters as $\mu = 0.0002$, $\lambda = 0.02$ and $\sigma_J = 0.015$ (the parameter setting in Section 7.1.3 of Aït-Sahalia et al. (2013) is taken for reference, where annualized values were used). We further set X_0 as $\log(30)$.

Next, we discuss the model for microstructure noise. We assume part of the noise can be modeled through a parametric function g:

$$Y_{t_k} = X_{t_k} + g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0) + \varepsilon_{t_k}, \quad \text{where } \varepsilon_{t_k} \sim_{i.i.d.} N(0, \sigma_n^2). \tag{36}$$

We consider the following three forms for the function g:

$$g_1(V_{t_k}, I_{b/s}(t_k); \alpha, \beta) = I_{b/s}(t_k)\big(\alpha + \beta V_{t_k}/\Delta t_k\big); \tag{37}$$

$$g_2(V_{t_k}, I_{b/s}(t_k); \alpha, \beta^+, \beta^-) = \begin{cases} \alpha + \beta^+ V_{t_k}/\Delta t_k, & \text{if buyer-initiated,} \\ -\alpha - \beta^- V_{t_k}/\Delta t_k, & \text{if seller-initiated;} \end{cases} \tag{38}$$

$$g_3(V_{t_k}, I_{b/s}(t_k); \beta, \gamma) = I_{b/s}(t_k) \log(\gamma + \beta V_{t_k}/\Delta t_k). \tag{39}$$

Model g_1 is the same as (4). Model g_2 allows for asymmetric impact of buys and sells. Model g_3 is nonlinear in trading rates, in particular, it is concave for buys and convex for sells, which is consistent with the findings in Keim and Madhavan (1996). When the trading rate is low, g_3 is approximately the same as g_1, whereas when the trading rate is high, $|g_3|$ grows more slowly than linearly.

The trade type process $\{I_{b/s}(t_k)\}$ is simulated as a Bernoulli process (± 1 valued) with a success probability of $p = 1/2$ and with an autocorrelation of 0.3. Here the correlation is incorporated to mimic the positive autocorrelation in trade types that are found in various empirical studies. With the correlated trade types, the g component and consequently the total microstructure noise becomes serially dependent.

As to the trading volume, inspired by Białkowski et al. (2008) and Kim and Murphy (2011)[6], we simulate the trading volume process $\{V_{t_k}\}$ by rounding $\{|V^*_{t_k}|\}$ up to hundreds, where $\{V^*_{t_k}\}$ are generated as

$$V^*_{t_k} = H_{t_k} + e_{t_k}, \tag{40}$$

where $H_{t_k}, k = 0, 1 \ldots,$ are independent Gamma random variables with mean $\int_{t_{k-1}}^{t_k} h_t dt$ and standard deviation $5,000$, in which $h_t = (0.025(t - 0.5)^2 + 0.05) \times 10^8$ features a "U-shaped pattern"[7]; and $\{e_{t_k}\}$ follows an ARMA(1,1) model

$$e_{t_k} = \phi_1 e_{t_{k-1}} + \psi_1 u_{t_{k-1}},$$

where $\phi_1 = 0.5, \psi_1 = 0.5$, and $\{u_{t_k}\}$ consists of i.i.d. normal random variables with mean 0 and standard deviation 100. The resulting trading volumes have means approximately 400 and standard deviations approximately 5,000.

Regarding the observation times $\{t_k\}$, we consider the irregular time setting generated by

Poisson arrival time: t_k's are the arrival times of a Poisson process with rate $23,400$.

We now discuss the choice for the parameters σ_n, α, β etc. in equations (36) – (39). We take $\sigma_n = 0.00012$, $\alpha = 1.875 \times 10^{-4}$, $\beta = 0.75 \times 10^{-12}$, $\beta^+ = 0.75 \times 10^{-12}$, $\beta^- = 0.30 \times$

[6]Białkowski et al. (2008) presented a decomposition for modeling intraday volume: one part reflects volume changes caused by market evolution, the other part describes the stock specific volume pattern. The dynamic of the specific volume part is depicted using autoregressive moving average (ARMA) and self-exciting threshold autoregressive (SETAR) models. Kim and Murphy (2011) observed that in 2009, the average size of an individual trade for the S&P 500 ETF (SPY) is 400 shares with a standard deviation of 5,100 shares.

[7]Biais et al. (1995), Gourieroux et al. (1999), and Białkowski et al. (2008), among others, observed that a U-shaped pattern is present in intraday volume.

10^{-12} and $\gamma = 1 + \alpha = 1.0001875$. Under these choices, the realized values of g_1, g_2, and g_3 have means approximately zero and standard deviations approximately $0.00024, 0.000226$ and 0.00024 respectively. These numbers are typical values of what we found in empirical studies.

We run three sets of simulations for the situations where the g component in the noise follows Model g_1, g_2 and g_3 respectively, each with $2,500$ replications. The parameters are estimated using (10) for linear models g_1 and g_2; whereas for g_3, the parameters are estimated by minimizing the objective function (9), which is performed in R by using the function "optim".

In Table 1, `target` records the true parameter values ($(\alpha \times 10^4, \beta \times 10^{12})$ for g_1, $(\alpha \times 10^4, \beta^+ \times 10^{12}, \beta^- \times 10^{12})$ for g_2 and $(\gamma, \beta \times 10^{12})$ for g_3); `bias` gives average values for the difference between estimated values and target values; `RMSE` records the square root of the mean squared deviations. These statistics indicate for both parametric estimation and IV estimation, under all Models g_1, g_2 and g_3, our estimators yield small estimation errors.

	$g_1(V_{t_k}, I_{b/s}(t_k); \alpha, \beta)$			$g_3(V_{t_k}, I_{b/s}(t_k); \beta, \gamma)$		
	$\widehat{\alpha} \times 10^4$	$\widehat{\beta} \times 10^{12}$	ERV_{ext}	$\widehat{\gamma}$	$\widehat{\beta} \times 10^{12}$	ERV_{ext}
target	1.875	0.75	0.0003	1.0001875	0.75	0.0003
bias	4.80×10^{-4}	-1.29×10^{-3}	5.45×10^{-7}	-1.67×10^{-7}	4.00×10^{-4}	3.80×10^{-7}
RMSE	1.29×10^{-2}	6.92×10^{-3}	1.39×10^{-5}	6.90×10^{-7}	1.24×10^{-2}	1.46×10^{-5}

	$g_2(V_{t_k}, I_{b/s}(t_k); \alpha, \beta^+, \beta^-)$			
	$\widehat{\alpha} \times 10^4$	$\widehat{\beta^+} \times 10^{12}$	$\widehat{\beta^-} \times 10^{12}$	ERV_{ext}
target	1.875	0.75	0.30	0.0003
bias	4.18×10^{-4}	-1.67×10^{-3}	-5.33×10^{-4}	1.67×10^{-7}
RMSE	1.29×10^{-2}	9.79×10^{-3}	6.06×10^{-3}	1.38×10^{-5}

Table 1

Estimation results based on 2,500 replications for the Poisson arrival time case for model (36) *with X following* (35)*, g being one of the g_1, g_2 and g_3 in* (37)–(39)*. The RVs based on the raw observations have large biases of* 2.7×10^{-3}, 2.4×10^{-3} *and* 2.7×10^{-3}, *which are omitted in the table.*

Figure 1 displays the Q-Q plots and histograms for the statistic $N^n_{1,\varepsilon}$ defined in (33) for all three sets of simulations. The asymptotic standard normality of $N^n_{1,\varepsilon}$ is clearly supported.

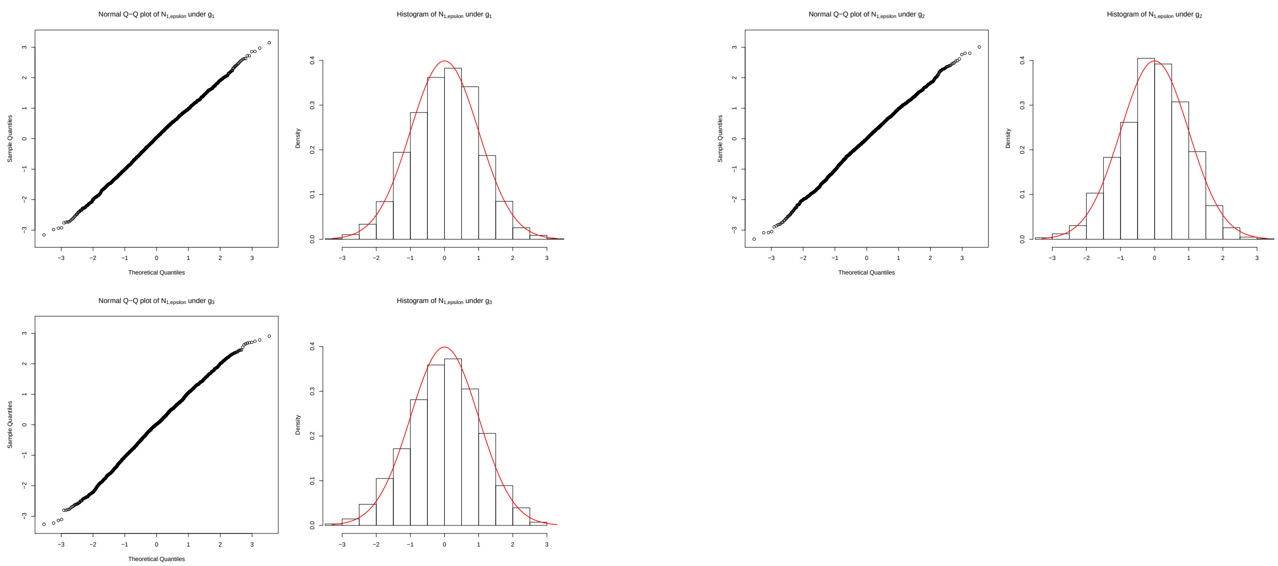

Fig. 1. *Normal Q-Q plot and histogram based on 2,500 replications of* $N_{1,\varepsilon}^n$ *under model* (36) *with* X *satisfying* (35), *and* g *being* g_1 *(upper-left),* g_2 *(upper-right) and* g_3 *(lower), for the Poisson arrival time case. In each of histogram, the red curve is the density plot of the standard normal distribution.*

3.2. Comparisons

3.2.1. Existing popular estimators to be compared with

In this subsection we compare our proposed estimators ERV_{ext} and E-QMLE with existing widely used volatility estimators MSRV, RK, PAV and QMLE. The estimator TSRV is not included in the comparison due to its suboptimal convergence rate. To make the comparisons fair, below we create a setting where microstructure noise is i.i.d. and observation times are equally spaced, so that the required assumptions for all these estimators are satisfied.

In the following, we briefly introduce the aforementioned estimators and how we compute them and their associated standard errors under Model (1). Let θ be a positive constant that can take different values for different estimators. Let $\widehat{Q}_1^N$ be defined as in (3.14) in Jacod et al. (2009) (see (47) below), which is a consistent estimator of the quarticity $\int_0^1 \sigma_t^4 \, dt$. Let $\widehat{E\varepsilon^2} = \sum_{k=1}^N (\Delta Y_{t_k})^2/(2N)$ and $\widehat{\mathrm{Var}(\varepsilon^2)} = \sum_{k=1}^N (\Delta Y_{t_k})^4/(2N) - 4(\widehat{E\varepsilon^2})^2$ be the estimators of the variances of the noise and squared noise, respectively (p.1402 in Zhang et al. (2005)).

(i) Let $M_N = [\theta\sqrt{N}]$. The MSRV (Zhang (2006)) is a weighted average over M_N different time scales:

$$\text{MSRV} = \sum_{i=1}^{M_N} \alpha_{N,i} \left\{ \frac{1}{K_i} \sum_{k=0}^{N-K_i} (Y_{t_{k+K_i}} - Y_{t_k})^2 \right\}, \tag{41}$$

where the weights $\alpha_{N,i}$ and K_i are given by (see Eq. (20) in Zhang (2006))

$$\alpha_{N,i} = 12 \frac{i}{M_N^2} \frac{i/M_N - 1/2 - 1/(2M_N)}{1 - 1/M_N^2}, \quad \text{and} \quad K_i = i.$$

The asymptotic standard deviation can be estimated by (Eq.(37) in Zhang (2006))

$$N^{-1/4} \left\{ 48\theta^{-3}(\widehat{E\varepsilon^2})^2 + \frac{52}{35}\theta\, \widehat{Q}_1^N + \frac{12}{5}\theta^{-1}(\widehat{\text{Var}(\varepsilon^2)}) + \frac{48}{5}\theta^{-1}(\widehat{E\varepsilon^2})\text{MSRV} \right\}^{1/2}. \tag{42}$$

The tuning parameter θ is chosen to be the value that minimizes (42) with MSRV estimated using $\theta = 1$ (call it MSRV_1), namely, if one lets $a = 48(\widehat{E\varepsilon^2})^2$, $b = \frac{52}{35}\widehat{Q}_1^N$, $c = \frac{12}{5}\widehat{\text{Var}(\varepsilon^2)} + \frac{48}{5}\widehat{E\varepsilon^2}\text{MSRV}_1$, then the optimal tuning parameter is chosen as $\sqrt{\frac{6a}{\sqrt{c^2+12ab}-c}}$.

(ii) Let $H = [\theta\sqrt{N}]$ and $K(x)$ ($x \in [0,1]$) be a kernel function. The RK (Barndorff-Nielsen et al. (2008)) is defined as

$$\text{RK} = \sum_{k=1}^{N}(\Delta Y_{t_k})^2 + \sum_{h=1}^{H} K\left(\frac{h-1}{H}\right) \left\{ \sum_{k=1}^{N} \Delta Y_{t_k}\Delta Y_{t_{k-h}} + \sum_{k=1}^{N} \Delta Y_{t_k}\Delta Y_{t_{k+h}} \right\}. \tag{43}$$

We choose $K(\cdot)$ to be the Parzen kernel, namely, $K(x) = 1 - 6x^2 + 6x^3$ when $0 \le x \le 1/2$ and $2(1-x)^3$ when $1/2 \le x \le 1$. The asymptotic standard deviation is estimated by (the first paragraph on p.1495 and Table II in Barndorff-Nielsen et al. (2008))

$$N^{-1/4}\sqrt{8.54(\widehat{E\varepsilon^2})^{1/2}(\widehat{Q}_1^N)^{3/4}}. \tag{44}$$

In our numerical studies, we use the estimated optimal $\theta = 4.77\sqrt{\widehat{E\varepsilon^2}}/(\widehat{Q}_1^N)^{1/4}$ (see Table II on p.1495, and p.1492 for the definition of ξ).

(iii) The PAV (Jacod et al. (2009)) with weight function $g(x) = x \wedge (1-x)$ for $x \in (0,1)$ is defined as follows: with $k_N = [\theta\sqrt{N}]$ being the window length over which the averaging takes place, let

$$\text{PAV} = \frac{12}{\theta\sqrt{N}} \sum_{k=0}^{N-k_N+1} (\overline{Y}_k^N)^2 - \frac{6}{\theta^2 N}\sum_{k=1}^{N}(\Delta Y_{t_k})^2, \tag{45}$$

where $\overline{Y}_k^N = (\sum_{i=k_N/2}^{k_N-1} Y_{t_{k+i}} - \sum_{i=0}^{k_N/2-1} Y_{t_{k+i}})/k_N$. The estimator of the asymptotic standard deviation is given by $N^{-1/4}\sqrt{\Gamma^N}$, where (Eq.(3.7) in Jacod et al. (2009))

$$\begin{aligned}\Gamma^N =&\frac{1812}{35\theta}\sum_{k=0}^{N-k_N+1}(\overline{Y}_k^N)^4 - \frac{2916}{35\theta^3 N}\sum_{k=0}^{N-2k_N+1}(\overline{Y}_k^N)^2\sum_{j=k+k_N}^{k+2k_N-1}(\Delta Y_{t_j})^2\\ &+\frac{939}{35\theta^3 N}\sum_{k=1}^{N-2}(\Delta Y_{t_k})^2(\Delta Y_{t_{k+2}})^2.\end{aligned} \tag{46}$$

An estimator of the quarticity $\int_0^1 \sigma_t^4\, dt$ is given by

$$\begin{aligned}\widehat{Q}_1^N =&\frac{48}{\theta^2}\sum_{k=0}^{N-k_N+1}(\overline{Y}_k^N)^4 - \frac{144}{\theta^4\ N}\sum_{k=0}^{N-2k_N+1}(\overline{Y}_k^N)^2\sum_{j=k+k_N}^{k+2k_N-1}(\Delta Y_{t_j})^2\\ &+\frac{36}{\theta^4\ N}\sum_{k=1}^{N-2}(\Delta Y_{t_k})^2(\Delta Y_{t_{k+2}})^2.\end{aligned} \tag{47}$$

The tuning parameter θ is taken to be $4.777\sqrt{\widehat{E\varepsilon^2}/\widehat{\sigma^2}}$ (see Remark 2 of Jacod et al. (2009)), where $\widehat{\sigma^2}$ is taken to be PAV with tuning parameter $\theta = 0.5$, as in Aït-Sahalia et al. (2013).

(iv) The QMLE (Xiu (2010)) is the estimated σ^2 obtained by maximizing the following quasi-likelihood function

$$l(a^2, \sigma^2) = -\frac{1}{2}\log\det(\Omega) - \frac{N}{2}\log(2\pi) - \frac{1}{2}\mathbf{Y}^T\Omega^{-1}\mathbf{Y}, \tag{48}$$

where

$$\Omega = \begin{pmatrix} \sigma^2\Delta t_1 + 2a^2 & -a^2 & 0 & \dots & 0 \\ -a^2 & \sigma^2\Delta t_2 + 2a^2 & -a^2 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots - a^2 & \sigma^2\Delta t_N + 2a^2 \end{pmatrix}.$$

The asymptotic standard deviation is estimated by (p.240 in Xiu (2010))

$$N^{-1/4}\left\{5(\widehat{E\varepsilon^2})^{1/2}\ \widehat{Q}_1^N/\mathrm{QMLE}^{1/2} + 3\ \mathrm{QMLE}^{3/2}(\widehat{E\varepsilon^2})^{1/2}\right\}^{1/2}. \tag{49}$$

3.2.2. The comparison

In this subsection, we create a setting where

(i). the latent log-price process is continuous;

(ii). the observation times are equally spaced; and

(iii). the market microstructure noises are i.i.d. and also independent with the latent log-price process.

Again, such a design is so that the required assumptions for all these estimators are satisfied.

More specifically, for the latent log-price we consider a similar model to (35) but without jumps:

$$dX_t = (\mu - \nu_t/2)dt + \sigma_t dB_t. \tag{50}$$

The parameter μ is chosen to be the same as in Section 3.1, namely, $\mu = 0.0002$. We again set $X_0 = \log(30)$.

As to the noise, we adopt (36) with g being g_1, namely,

$$Y_{t_k} = X_{t_k} + g_1(\mathbf{Z}_{t_k}; \theta_0) + \varepsilon_{t_k}. \tag{51}$$

The ε_{t_k}'s are taken to be i.i.d. normal with mean 0 and standard deviation 0.00012.

To make the whole noise i.i.d., we simulate the trade type process $\{I_{b/s}(t_k)\}$ as a Bernoulli process (± 1 valued) with a success probability of $p = 1/2$, and simulate the trading volumes $\{V_{t_k}\}$ as i.i.d. Gamma random variables with mean 400 and standard deviation 5,000.

Finally, regarding the observation times $\{t_k\}$, we consider the

<u>Equidistant time case</u>: $t_k = k/23,400$ for $k = 0, 1, \ldots, 23, 400$.

With the specifications above, the conditions required by the aforementioned volatility estimators are satisfied and these estimators can be readily applied. The standard deviation of $g_1(\mathbf{Z}_{t_k}; \theta_0)$ is about 0.00021, which leads to a proportion of explained variation π_{exp} as $0.00021^2/(0.00021^2 + 0.00012^2) \approx 75\%$, and the noise-to-signal ratio ξ^2 about $(0.00021^2 + 0.00012^2)/\sqrt{0.9\zeta^2 + 0.1 \times 15^2\zeta^2} \approx 0.0001$.

In the comparison we compute the estimated volatilities and confidence intervals based on the estimators MSRV, RK, PAV, QMLE, and our proposed estimators ERV_{ext} and E-QMLE. Table 2 reports the `bias`, `RMSE`, average width of confidence intervals (`AWidth`), and the coverage rate of the confidence intervals (`CR`, which is the percentage of times that the true integrated volatility is inside the confidence intervals), based on 2,500 replications. The first column `Intv` records the sampling interval. It ranges from one observation per second to one observation per 10 seconds. The advantages of ERV_{ext} and E-QMLE are clearly demonstrated:

(i). ERV_{ext} has smaller RMSE's, so does E-QMLE. The reduction in RMSE compared with the smallest RMSE achieved by MSRV, RK, PAV and QMLE ranges from 8% to 18% across different frequencies;

(ii). ERV_{ext} yields confidence intervals with coverage rates close to the nominal level.

	MSRV				RK			
Intv	bias	RMSE	AWidth	CR	bias	RMSE	AWidth	CR
1	-0.089	0.168	0.551	88.3%	-0.001	0.151	0.544	91.9%
5	-0.204	0.321	0.834	76.7%	-0.003	0.253	0.825	89.3%
10	-0.250	0.415	1.023	72.4%	-0.006	0.349	1.016	84.1%
	PAV				QMLE			
Intv	bias	RMSE	AWidth	CR	bias	RMSE	AWidth	CR
1	0.043	0.190	0.602	89.6%	-0.004	0.150	0.574	93.7%
5	-0.109	0.328	0.957	82.3%	-0.010	0.259	0.858	90.3%
10	-0.060	0.465	1.250	79.8%	-0.015	0.344	1.050	85.4%
	ERV_{ext}				E-QMLE		Reduction in RMSE	
Intv	bias	RMSE	AWidth	CR	bias	RMSE	by ERV_{ext}	by E-QMLE
1	-0.007	0.127	0.506	95.5%	-0.005	0.123	15.3%	18.0%
5	-0.012	0.220	0.879	95.1%	-0.005	0.225	13.0%	11.1%
10	-0.026	0.315	1.202	94.3%	-0.007	0.315	8.4%	8.4%

Table 2

Comparison among different estimators based on 2,500 replications under the equidistant time setting for model (51) *with X following* (50)*. The true IV is 0.0003. All values (except* `Intv`*,* `CR` *and* `Reduction in RMSE`*) are reported after being multiplied by* 10^4*. The bottom right corner reports the reductions in RMSE by using ERV_{ext} or E-QMLE compared with the smallest RMSE achieved by MSRV, RK, PAV and QMLE.*

Remark 11. *Notice that the comparisons are made under an ideal setting where the assumptions for all the alternative estimators are satisfied. Our estimator ERV_{ext} is applicable to much more general situations, notably, microstructure noise can be serially dependent, can have diurnal features, and observation times can be irregular. See Section 3.1 for an illustration.*

3.3. When there is model misspecification, irregular observation times and rounding in both price and time

In this subsection, we demonstrate that ERV_{ext} still performs well even with some degrees of model misspecification. The setting considered in this subsection is the same as in Section 3.2.2 except that

(i). In (51) the function g_1 is changed to be g_3;

(ii). observations occur at irregular times, with the observation times rounded down to the previous second, and

(iii). the observed prices are rounded to cents.

The observations arrive as a Poisson process with rate 11,700. It is chosen such that at the one second sampling interval, the previous tick method yields a sample size of about 9,000, a typical size for the stocks in the empirical studies. The observation times are further rounded down to the previous second, to mimic the empirical data that we have access to which has transaction times recorded only accurate to seconds. All estimators are to be applied to data subsampled by the previous tick method.

To evaluate the performance of our estimators when model is misspecified, in the parametric estimation step we use the wrong model g_1 for the g component in the noise. Model g_1 involves the trading durations (Δt_k in (37)). In the data generating procedure we purposely rounded the observation times, which induces errors in the trading durations. The error is particularly severe when there are multiple observations in a same second, in which case the times of all the observations are recorded to have a same time stamp, which leads to zero trading durations. To reduce the errors, we adopt the following proxy: if there are m transactions recorded at a same time t, then the transaction times of the ith ($i = 1, \cdots, m$) transaction is set to be $t+i/m$ seconds. Taking the difference of successive transaction times yields the trading durations Δt_k.

The estimation results based on 2,500 replications are shown in Table 3. The percentage of explained variation π_{exp} and noise-to-signal ratios ξ^2 are estimated to be about 71.5% and 0.00011 respectively. In the table, `Intv` records the sampling interval (in units of seconds), and `Freq` records the average total number of observations when we sample every `Intv` seconds using the previous tick method. We only report the biases and RMSE's. The summaries about confidence intervals are omitted because with the violation of assumptions, the inference theories for all estimators become less reliable. From Table 3 we again see that ERV_{ext} and

E-QMLE yield smaller RMSE's, the reduction ranging from roughly 6% to 10% compared with the best among MSRV, RK, PAV and QMLE.

Let us add that if the rounding in time is less severe, for example, if one has access to millisecond data, then the performance of ERV_{ext} (and E-QMLE) can be further improved. In our unreported simulation studies, with the true time stamps, ERV_{ext} yields a reduction in RMSE of 13.0% at the one second frequency, which is more than twice the reduction reported in Table 3.

		MSRV		RK		PAV		QMLE	
Intv	Freq	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
1	9199	-0.186	0.269	-0.002	0.216	-0.008	0.253	-0.009	0.219
5	4294	-0.209	0.332	-0.010	0.267	-0.081	0.327	-0.015	0.271
10	2324	-0.244	0.412	-0.017	0.340	-0.096	0.450	-0.025	0.340
		ERV_{ext}		E-QMLE		Reduction in RMSE			
Intv	Freq	bias	RMSE	bias	RMSE	by ERV_{ext}		by E-QMLE	
1	9197	0.003	0.203	-0.007	0.195	6.0%		9.7%	
5	4294	-0.004	0.246	-0.013	0.248	7.9%		7.1%	
10	2323	-0.012	0.315	-0.019	0.320	7.4%		5.9%	

Table 3

Comparison among different estimators based on 2,500 replications for model (36) with X following (50) and $g = g_3$ (but misspecified as g_1 when applying ERV/E-QMLE), under the setting when observation times are irregular and there is rounding in both price and time. The true IV is 0.0003. All values (except `Intv`*,* `Freq` *and* `Reduction in RMSE`*) are reported after being multiplied by 10^4. The bottom right corner reports the reductions in RMSE by using ERV_{ext} or E-QMLE compared with the smallest RMSE achieved by MSRV, RK, PAV and QMLE.*

Figure 2 displays the Q-Q plot and histogram for the statistic $N^n_{1,\varepsilon}$ defined in (33) based on 1-second data. The asymptotic standard normality of $N^n_{1,\varepsilon}$ appears to still roughly hold.

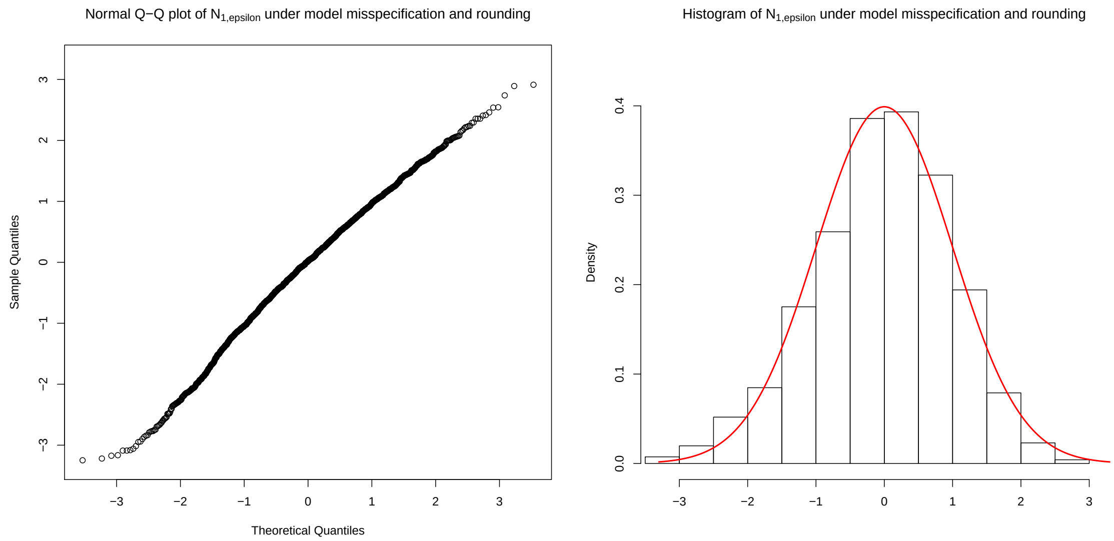

Fig. 2. *Normal Q-Q plot and histogram of $N^n_{1,\varepsilon}$ under model misspecification and rounding in both price and time. The red curve in the histogram is the density plot of the standard normal distribution.*

4. Empirical Studies

We analyze trade and quote (TAQ) data on a random day (June 1, 2012) for four stocks: Arch Coal Inc.(NYSE:ACI), Dell Inc.(NASDAQ:DELL), EMC Corp.(NYSE:EMC) and General Electric Co.(NYSE:GE).

Our method requires the information of trade type: whether a trade is buyer-initiated (a "buy") or seller-initiated (a "sell"). We obtain such information by applying the Lee Ready algorithm (program provided by WRDS) to the WRDS-derived Trades files (WCT datasets) which was generated from TAQ Trades and Quotes datasets. See
https://wrds-web.wharton.upenn.edu/wrds/research/applications/intraday/index.cfm
for detailed information. In the analysis below we remove the transactions with undefined trade type (i.e. LeeReady=0 in the generated dataset).

We assume model (19) and use g_1 in (37) to conduct the estimations. For each stock, we study three different sampling intervals: (`Intv`) 10s, 5s, and 1s. Same as in Section 3.3, all estimators are to be applied to data sampled by the previous tick method.

As we discussed in Section 3.3, Model g_1 involves another variable, the trading durations (Δt_k in (37)). The database that we have access to has transaction times recorded only accurate to seconds, hence the durations cannot be accurately calculated especially when there are multiple trades in single seconds. We adopt the same proxy as in Section 3.3, namely, if there are m transactions recorded with a same time stamp t, then the transaction times of the ith ($i = 1, \cdots, m$) transaction is set to be $t + i/m$ seconds. The trading durations Δt_k are then obtained by taking differences of successive transaction times.

Tables 4-7 show the estimation results of the six volatility estimators that we consider in Section 3, namely, MSRV, RK, PAV, QMLE, ERV_{ext} and E-QMLE. We adopt the non-truncated version of ERV_{ext} in (25), to be consistent with other estimators that do not remove jumps. The tuning parameters for MSRV, RK and PAV are chosen as in Section 3.2. We also record the estimates of RV.

We also report the estimates of the two variations, g_V and ε_V, and the ratio π_{exp} that we discussed in Remark 8. Recall that π_{exp} stands for the percentage of variation in the noise that is explained by the model g. Based on these quantities, one can estimate the standard deviations of g and ε_{t_k} by $\sqrt{\widehat{g_V}/N}$ and $\sqrt{\widehat{\varepsilon_V}/N}$ respectively, which are also reported in the tables.

In Tables 4-7, `Intv` records the sampling interval (in units of seconds), and `Freq` records the total number of observations when we sample every `Intv` seconds. The highest frequency we use is 1-second. We do not go into higher frequencies so that (1) the time information is reasonably accurate, and (2) when applying the alternative estimators, the required assumption of equidistant observation times will not be severely violated. Again, since there may be no observation in some sampling intervals, `Freq` may not equal 23,400/`Intv`.

Intv	Freq	RV	MSRV	RK	PAV	QMLE	ERV_{ext}	E-QMLE	$\widehat{g_V}$	$\widehat{\varepsilon_V}$	$\widehat{\pi_{exp}}$	$\widehat{sd(g)}$	$\widehat{sd(\varepsilon)}$
1	8364	84.890	18.012	18.273	18.407	18.717	18.589	18.940	31.405	8.052	79.6%	6.128	3.103
5	3810	48.918	19.743	20.746	19.562	20.245	20.243	20.199	14.753	1.874	88.7%	6.223	2.218
10	2214	37.628	18.735	20.168	20.691	20.271	19.783	19.545	8.661	1.648	84.0%	6.255	2.728

Table 4

Estimation results for stock ACI on June 1, 2012. All volatility estimates, $\widehat{g_V}$, $\widehat{\varepsilon_V}$, $\widehat{sd(g)}$ and $\widehat{sd(\varepsilon)}$ are reported after being multiplied by 10^4.

Intv	Freq	RV	MSRV	RK	PAV	QMLE	ERV_{ext}	E-QMLE	$\widehat{g_V}$	$\widehat{\varepsilon_V}$	$\widehat{\pi_{exp}}$	$\widehat{sd(g)}$	$\widehat{sd(\varepsilon)}$
1	8338	17.277	2.776	3.013	2.980	3.144	3.438	3.275	6.088	1.848	76.7%	2.702	1.489
5	3937	9.619	3.265	3.423	3.254	3.549	3.354	3.420	2.680	0.886	75.2%	2.609	1.500
10	2247	7.082	2.737	3.782	4.018	3.857	3.495	3.576	1.579	0.485	76.5%	2.651	1.469

Table 5

Estimation results for stock DELL on June 1, 2012. All volatility estimates, $\widehat{g_V}$, $\widehat{\varepsilon_V}$, $\widehat{sd(g)}$ *and* $\widehat{sd(\varepsilon)}$ *are reported after being multiplied by* 10^4.

Intv	Freq	RV	MSRV	RK	PAV	QMLE	ERV_{ext}	E-QMLE	$\widehat{g_V}$	$\widehat{\varepsilon_V}$	$\widehat{\pi_{exp}}$	$\widehat{sd(g)}$	$\widehat{sd(\varepsilon)}$
1	9654	7.050	1.573	1.737	1.809	1.715	1.783	1.803	2.287	0.773	74.7%	1.539	0.895
5	4091	3.992	1.762	1.854	1.832	1.889	1.818	1.818	0.876	0.340	72.0%	1.464	0.912
10	2279	3.079	1.428	1.918	2.017	1.917	1.875	1.872	0.519	0.130	79.9%	1.509	0.756

Table 6

Estimation results for stock EMC on June 1, 2012. All volatility estimates, $\widehat{g_V}$, $\widehat{\varepsilon_V}$, $\widehat{sd(g)}$ *and* $\widehat{sd(\varepsilon)}$ *are reported after being multiplied by* 10^4.

Intv	Freq	RV	MSRV	RK	PAV	QMLE	ERV_{ext}	E-QMLE	$\widehat{g_V}$	$\widehat{\varepsilon_V}$	$\widehat{\pi_{exp}}$	$\widehat{sd(g)}$	$\widehat{sd(\varepsilon)}$
1	13404	12.985	1.801	1.845	1.699	1.888	1.947	1.884	4.353	1.513	74.2%	1.802	1.062
5	4508	5.922	2.069	2.136	2.030	2.201	2.078	2.105	1.479	0.532	73.5%	1.811	1.087
10	2331	4.331	1.826	2.510	2.656	2.538	2.172	2.269	0.805	0.340	70.3%	1.859	1.207

Table 7

Estimation results for stock GE on June 1, 2012. All volatility estimates, $\widehat{g_V}$, $\widehat{\varepsilon_V}$, $\widehat{sd(g)}$ *and* $\widehat{sd(\varepsilon)}$ *are reported after being multiplied by* 10^4.

We see from Tables 4-7 that

(i). ERV_{ext} and E-QMLE are close to each other, both of which are close to the typical estimates based on the alternative noise-robust estimators.

(ii). Across different sampling intervals, ERV_{ext} and E-QMLE provide stable estimates.

- They do not exhibit "volatility signature plot" pattern (Andersen et al. (2000), the pattern that we see on RV, which gives larger estimated daily volatilities as the sampling interval goes smaller);

- ERV_{ext} and E-QMLE appear to give more stable estimates across different sampling intervals compared with the alternative estimators.

(iii). The ratio $\widehat{\pi_{exp}}$ is stable across different sampling intervals.

- For the stocks reported (and many other stocks we analyzed), the ratios are around 70%-80%, by using the simple model g_1. This suggests that microstructure noise can indeed be largely accounted for by trade type, trading volume/rate, etc.

The analyses also indicate that $\widehat{\pi_{exp}}$ provides a useful criterion to search for better models for the market microstructure, which will be of both theoretical and practical interest.

5. Conclusion and Discussion

This paper is among the first attempts to make efficient use of available information in the market for volatility estimation. Under a general parametric model for microstructure noise, in a setting where the observation times can be irregular and jumps are allowed, we show that we can estimate the parameters with rate n, which allows us to estimate the latent log-prices with high accuracy. Based on the estimated log-prices, we build "estimated-price realized volatility" (ERV), which provides $\sqrt{n}$ rate of convergence and the same asymptotic properties as realized volatility (RV) based on latent log-prices. To adapt to broader realistic situations, we further propose an extended version ERV_{ext} in the presence of an additional noise component. Under the assumption that the additional noise is "small" ($O_p(1/\sqrt{n})$ to be precise), ERV_{ext} also enjoys $\sqrt{n}$ rate of convergence.

The superior performance of our estimators, both for parameter estimation and volatility estimation, is demonstrated via simulation studies. ERV_{ext} is seen to perform well even with rounding in both price and time and model misspecifications on the parametric model. About the other estimator, E-QMLE, that we propose, although we do not establish its asymptotic properties in this article, numerical studies show that it also performs well in various settings.

Empirically, our estimators also perform favorably. An interesting additional finding is that a simple model for market microstructure noise, which incorporates only trade types and trading rates, can account for a high percentage of the total variation in the noise.

Of course the market information to be incorporated should by no means be restricted to trade type and trading rate, and the model for the noise should not be restricted to the ones that we illustrated. Users of our estimators are advised to further explore more sophisticated noise models and incorporate more available information and use the general framework provided in this paper.

The framework proposed in this paper and the concept of percentage of explained variation in the noise are useful in studies of market microstructure.

Appendix A. Proofs

First observe that, by using standard localization techniques (see, for example, Section 2.4.5 of Mykland and Zhang (2012)), for proving Theorems 1 - 3 and Propositions 1 - 2, we can replace the assumptions (A.i), (A.ii) and (A.v) with the following stronger assumptions:
(SA.i) $(|\mu_t|)$ is bounded;
(SA.ii) (σ_t) is bounded from both below and above;
(SA.iv) $\max_k |\mathbf{Z}_{t_k}|$ is bounded.
Similarly, as in Appendix (A.5) in Jacod and Protter (2012), Assumption (C.i) can be replaced with Assumption (SH-2) therein, which, in particular, implies that for any $p \geq 2$, for any finite stopping times $s \leq t$:

$$E\Big(|\nu_t - \nu_s|^p \mid \mathcal{F}_s\Big) \leq C\, E(t - s \mid \mathcal{F}_s), \tag{52}$$

see (A.67) therein. We always assume these strengthened assumptions below, mostly without special mention. Also in all the sequel, C, C_1 etc. denote generic constants whose values may change from line to line.

A.1. Proof of Theorem 1

Proof. We first prove consistency of $\widehat{\boldsymbol{\theta}}$:

$$\widehat{\boldsymbol{\theta}} \stackrel{p}{\longrightarrow} \boldsymbol{\theta}_0. \tag{53}$$

In the case when (σ_t) satisfies certain additional assumptions (e.g., (σ_t) itself is an Itô process), jump is not present, both $\{t_k\}$ and $\{\mathbf{Z}_{t_k}\}$ are independent of (X_t), and $\{\mathbf{Z}_{t_k}\}$ are mutually independent, one can work with the probability measure P_n^* in Mykland and Zhang (2009) and apply Theorem 2.4 in White (1980) to prove (53). In general, to establish (53), it suffices to show that for any $\varepsilon > 0$,

$$P\left(\inf_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\geq\varepsilon} (Q_N(Y,\mathbf{Z},\boldsymbol{\theta}) - Q_N(Y,\mathbf{Z},\boldsymbol{\theta}_0)) > 0\right) \to 1. \tag{54}$$

In fact,

$$\begin{aligned} &2\,(Q_N(Y,\mathbf{Z},\boldsymbol{\theta}) - Q_N(Y,\mathbf{Z},\boldsymbol{\theta}_0)) \\ =&\sum_{k=1}^N (\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}))^2 + 2\sum_{k=1}^N (\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}))\,\Delta X_{t_k}. \end{aligned} \tag{55}$$

Using Assumption (A.x) we only need to show that

$$\sum_{k=1}^{N} \left(\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta})\right) \Delta X_{t_k} = O_p(1). \tag{56}$$

In fact, by (5),

$$\begin{aligned}
&\sum_{k=1}^{N} \left(\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta})\right) \Delta X_{t_k} \\
=&\sum_{k=1}^{N} \left(\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta})\right) \int_{t_{k-1}}^{t_k} \mu_t \, dt \\
&+ \sum_{k=1}^{N} \left(\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta})\right) \int_{t_{k-1}}^{t_k} \sigma_t \, dW_t \\
&+ \sum_{k=1}^{N} \left(\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta})\right) \int_{t_{k-1}}^{t_k} dJ_t \\
:=& I + II + III.
\end{aligned}$$

Note that by Assumptions (A.vii) and (SA.iv),

$$|\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta})| \le \left(L_0(\mathbf{Z}_{t_k}) + L_0(\mathbf{Z}_{t_{k-1}})\right) |\boldsymbol{\theta}_0 - \boldsymbol{\theta}| \le C|\boldsymbol{\theta}_0 - \boldsymbol{\theta}|. \tag{57}$$

Hence, for term I, by Assumption (SA.i), we have

$$I \le \sum_{k=1}^{N} C|\boldsymbol{\theta}_0 - \boldsymbol{\theta}| \cdot C\Delta t_k \le C|\boldsymbol{\theta}_0 - \boldsymbol{\theta}|.$$

As to term II, by Assumptions (A.iv) and (SA.ii), it is a martingale, hence by the Burkholder-Davis-Gundy (BDG) inequality,

$$\begin{aligned}
E(II^2) \le& CE\left(\sum_{k=1}^{N} \left(\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta})\right)^2 \Delta t_k\right) \\
\le& C|\boldsymbol{\theta}_0 - \boldsymbol{\theta}|^2.
\end{aligned}$$

Finally, for term III, by (57) again we have

$$|III| \le C|\boldsymbol{\theta}_0 - \boldsymbol{\theta}| \cdot \sum_{t} |\Delta J_t|.$$

Combining the three estimates above and Assumption (A.iii) we see that (56) and consequently (53) hold.

Next we prove the stronger conclusion that $N(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1)$. By (53), with probability approaching one, $\widehat{\boldsymbol{\theta}} \in \mathcal{N}(\boldsymbol{\theta}_0) \subset \Theta$, under which case since $\widehat{\boldsymbol{\theta}}$ minimizes $Q_N(Y, \mathbf{Z}, \boldsymbol{\theta})$, we must have $\frac{\partial Q_N}{\partial \boldsymbol{\theta}}(Y, \mathbf{Z}, \widehat{\boldsymbol{\theta}}) = 0$. It follows from the mean-value theorem that for each $j = 1, \ldots, d$, there exists a $\boldsymbol{\theta}_j^*$ which lies on the line segment connecting $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$ such that

$$\frac{\partial Q_N}{\partial \theta_j}(Y, \mathbf{Z}, \boldsymbol{\theta}_0) + \frac{\partial^2 Q_N}{\partial \boldsymbol{\theta} \partial \theta_j}(Y, \mathbf{Z}, \boldsymbol{\theta}_j^*)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = 0.$$

Hence

$$N(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\frac{\partial Q_N}{\partial \boldsymbol{\theta}}(Y, \mathbf{Z}, \boldsymbol{\theta}_0) \left(\frac{1}{N} \frac{\partial^2 Q_N}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}(Y, \mathbf{Z}, \boldsymbol{\theta}^*) \right)^{-1}, \tag{58}$$

where

$$\frac{\partial^2 Q_N}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}(Y, \mathbf{Z}, \boldsymbol{\theta}^*) := \left(\frac{\partial^2 Q_N}{\partial \boldsymbol{\theta} \partial \theta_j}(Y, \mathbf{Z}, \boldsymbol{\theta}_j^*) \right)_{j=1,\ldots,d}.$$

Observe that the jth component of $\frac{\partial Q_N}{\partial \boldsymbol{\theta}}(Y, \mathbf{Z}, \boldsymbol{\theta}_0)$ is given by

$$\frac{\partial Q_N}{\partial \theta_j}(Y, \mathbf{Z}, \boldsymbol{\theta}_0) = -\sum_{k=1}^{N} \Delta \frac{\partial g}{\partial \theta_j}(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0) \Delta X_{t_k}. \tag{59}$$

By Assumption (A.viii) (and Assumption (SA.iv)) and using a similar argument to that for (56) we can show that

$$\frac{\partial Q_N}{\partial \boldsymbol{\theta}}(Y, \mathbf{Z}, \boldsymbol{\theta}_0) = O_p(1).$$

It remains to show that

$$\left(\frac{1}{N} \frac{\partial^2 Q_N}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}(Y, \mathbf{Z}, \boldsymbol{\theta}^*) \right)^{-1} = O_p(1). \tag{60}$$

Note that for any $\boldsymbol{\theta}$,

$$\begin{aligned} &\frac{1}{N} \frac{\partial^2 Q_N}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}(Y, \mathbf{Z}, \boldsymbol{\theta}) \\ =& \frac{1}{N} \sum_{k=1}^{N} \Delta \frac{\partial g}{\partial \boldsymbol{\theta}}(\mathbf{Z}_{t_k}; \boldsymbol{\theta}) \Delta \frac{\partial g}{\partial \boldsymbol{\theta}^T}(\mathbf{Z}_{t_k}; \boldsymbol{\theta}) - \frac{1}{N} \sum_{k=1}^{N} (\Delta Y_{t_k} - \Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta})) \Delta \frac{\partial^2 g}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}(\mathbf{Z}_{t_k}; \boldsymbol{\theta}). \end{aligned}$$

Hence the (i, j)th entry of $\frac{1}{N} \frac{\partial^2 Q_N}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}(Y, \mathbf{Z}, \boldsymbol{\theta}^*)$ equals

$$\begin{aligned} &\frac{1}{N} \frac{\partial^2 Q_N}{\partial \theta_i \partial \theta_j}(Y, \mathbf{Z}, \boldsymbol{\theta}_j^*) \\ =& \frac{1}{N} \sum_{k=1}^{N} \Delta \frac{\partial g}{\partial \theta_i}(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_j^*) \Delta \frac{\partial g}{\partial \theta_j}(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_j^*) - \frac{1}{N} \sum_{k=1}^{N} \left(\Delta Y_{t_k} - \Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_j^*)\right) \Delta \frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_j^*) \\ :=& I_{ij} - II_{ij}, \quad i, j = 1, \ldots, d. \end{aligned} \tag{61}$$

We have

$$I_{ij} = \frac{1}{N} \sum_{k=1}^{N} \Delta \frac{\partial g}{\partial \theta_i}(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0) \Delta \frac{\partial g}{\partial \theta_j}(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0) + I_{1,ij}, \tag{62}$$

where

$$
\begin{aligned}
|I_{1,ij}| =& \left| \frac{1}{N}\sum_{k=1}^{N} \Delta \frac{\partial g}{\partial \theta_i}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_j^*)\Delta \frac{\partial g}{\partial \theta_j}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_j^*) - \frac{1}{N}\sum_{k=1}^{N} \Delta \frac{\partial g}{\partial \theta_i}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0)\Delta \frac{\partial g}{\partial \theta_j}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0)) \right| \\
\leq& \frac{C}{N}\sum_{k=1}^{N}(L_1(\mathbf{Z}_{t_k}) + L_1(\mathbf{Z}_{t_{k-1}}))|\boldsymbol{\theta}_j^* - \boldsymbol{\theta}_0| \qquad (63)\\
\leq& \frac{C}{N}\sum_{k=1}^{N}|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| \xrightarrow{p} 0,
\end{aligned}
$$

where the inequalities hold thanks to Assumptions (A.ix), (A.viii) and (SA.iv) (and the simple fact that $|\boldsymbol{\theta}_j^* - \boldsymbol{\theta}_0| \leq |\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|$), and the last convergence is due to (53).

As to II_{ij}, we have

$$
\begin{aligned}
II_{ij} =& \frac{1}{N}\sum_{k=1}^{N} \Delta X_{t_k} \Delta \frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) + \frac{1}{N}\sum_{k=1}^{N}(\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_j^*))\Delta \frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_j^*) \\
&+ \frac{1}{N}\sum_{k=1}^{N} \Delta X_{t_k}\left(\Delta \frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_j^*) - \Delta \frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0)\right) \\
:=& II_{1,ij} + II_{2,ij} + II_{3,ij}.
\end{aligned} \qquad (64)
$$

For $II_{1,ij}$, by a similar argument to (56) and using Assumption (A.xiii), one can show that $II_{1,ij} = o_p(1)$. Moreover, by Assumptions (A.vii) and (A.xiii) and (53),

$$
|II_{2,ij}| \leq \frac{C}{N}\sum_{k=1}^{N}(L_0(\mathbf{Z}_{t_k}) + L_0(\mathbf{Z}_{t_{k-1}}))|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| \left| \Delta \frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_j^*) \right| \xrightarrow{p} 0.
$$

Finally, by the Cauchy-Schwarz inequality, Assumption (A.xii) and (53),

$$
\begin{aligned}
|II_{3,ij}| \leq& \frac{1}{N}\sqrt{\sum_{k=1}^{N} \Delta X_{t_k}^2 \cdot \sum_{k=1}^{N}\left(\Delta \frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_j^*) - \Delta \frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0)\right)^2} \\
\leq& \frac{1}{N}\sqrt{\sum_{k=1}^{N} \Delta X_{t_k}^2 \cdot \sum_{k=1}^{N} 2(L_2(\mathbf{Z}_{t_k})^2 + L_2(\mathbf{Z}_{t_{k-1}})^2)|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|^2} \xrightarrow{p} 0.
\end{aligned}
$$

Combining the estimates above we see that

$$
\frac{1}{N}\frac{\partial^2 Q_N}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}}(Y, \mathbf{Z}, \boldsymbol{\theta}^*) = \frac{1}{N}\sum_{k=1}^{N} \Delta \frac{\partial g}{\partial \boldsymbol{\theta}}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0)\Delta \frac{\partial g}{\partial \boldsymbol{\theta}^T}(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0) + o_p(1).
$$

and hence (60) holds by Assumption (A.xi). □

A.2. Proof of Theorem 2

Proof. Rewrite ERV as follows:

$$\begin{aligned}\mathrm{ERV} =& \sum_{k=1}^{N} \left(\Delta Y_{t_k} - \Delta g(\mathbf{Z}_{t_k}; \widehat{\boldsymbol{\theta}})\right)^2 = \sum_{k=1}^{N} \Big(\Delta X_{t_k} - (\Delta g(\mathbf{Z}_{t_k}; \widehat{\boldsymbol{\theta}}) - \Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0))\Big)^2 \\ =& \sum_{k=1}^{N} \Delta X_{t_k}^2 + \sum_{k=1}^{N} \Big(\Delta g(\mathbf{Z}_{t_k}; \widehat{\boldsymbol{\theta}}) - \Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0)\Big)^2 - 2\sum_{k=1}^{N} \Big(\Delta g(\mathbf{Z}_{t_k}; \widehat{\boldsymbol{\theta}}) - \Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0)\Big) \Delta X_{t_k} \\ :=& I + II - III.\end{aligned}$$

Note that term I is just the realized volatility RV. Under Assumptions (B.ii)∼(B.iv) and if further that $J_t \equiv 0$, we have,

$$\sqrt{N}\left(\mathrm{RV} - \int_0^1 \sigma_t^2 dt\right) \xrightarrow{\mathcal{L}} \Phi \times \left(2\int_0^1 \sigma_t^4 \, dH_t\right)^{1/2} \quad \text{stably}, \tag{65}$$

see, for example, Corollary 2.30 of Mykland and Zhang (2012). Therefore to prove parts (i) and (ii) of the theorem, it suffices to show that, under Assumptions A and (B.i), we have

$$\sqrt{N}(II - III) = o_p(1). \tag{66}$$

We first deal with term II. By Assumption (A.vii),

$$\Big|\Delta g(\mathbf{Z}_{t_k}; \widehat{\boldsymbol{\theta}}) - \Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0)\Big| \le \big(L_0(\mathbf{Z}_{t_k}) + L_0(\mathbf{Z}_{t_{k-1}})\big) \cdot |\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|.$$

Using Theorem 1 (and Assumption (SA.iv)) one can then easily verify that $N \cdot II = O_p(1)$, which certainly implies that $\sqrt{N} II = o_p(1)$.

It remains to show that $\sqrt{N} III = o_p(1)$. Note that since $\widehat{\boldsymbol{\theta}}$ depends on the whole process $((X_t), (\mathbf{Z}_{t_k}))$, term III is not a martingale even if $\mu_t = J_t \equiv 0$, and hence the BDG inequality is not applicable. To circumvent this difficulty, observe that since we have proved that $N(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1)$, by Assumption (B.i) that $N/n \xrightarrow{p} F > 0$, to show $\sqrt{N} III = o_p(1)$, it suffices to show that for any $K > 0$,

$$\sqrt{n} \sup_{|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \le K/n} \left|\sum_{k=1}^{N} (\Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}) - \Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0)) \Delta X_{t_k}\right| = o_p(1). \tag{67}$$

To show (67), first note that by Assumptions (A.vii) and (SA.iv), for any $\boldsymbol{\theta}_1, \ \boldsymbol{\theta}_2 \in \mathcal{N}(\boldsymbol{\theta}_0)$,

$$|\Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_1) - \Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_2| \le (L_0(\mathbf{Z}_{t_k}) + L_0(\mathbf{Z}_{t_{k-1}})) \cdot |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| \le C|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|.$$

Therefore, using further Assumption (SA.i) and (A.iii) we obtain that

$$\begin{aligned}
&\sqrt{n} \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\leq K/n} \left| \sum_{k=1}^{N} (\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0)) \cdot \int_{t_{k-1}}^{t_k} \mu_t \; dt \right| \\
\leq & \sqrt{n} \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\leq K/n} \left| \sum_{k=1}^{N} C|\boldsymbol{\theta}-\boldsymbol{\theta}_0| \; \Delta t_k \right| \\
\leq & C\sqrt{n} \cdot K/n \to 0,
\end{aligned}$$

and

$$\begin{aligned}
&\sqrt{n} \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\leq K/n} \left| \sum_{k=1}^{N} (\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0)) \cdot \int_{t_{k-1}}^{t_k} dJ_t \right| \\
\leq & \sqrt{n} \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\leq K/n} C|\boldsymbol{\theta}-\boldsymbol{\theta}_0| \cdot \sum_t |\Delta J_t| \\
\leq & C\sqrt{n} \cdot K/n \cdot \sum_t |\Delta J_t| \xrightarrow{p} 0.
\end{aligned}$$

It remains to show that

$$\sqrt{n} \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\leq K/n} \left| \sum_{k=1}^{N} (\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0)) \cdot \int_{t_{k-1}}^{t_k} \sigma_t \; dW_t \right| = o_p(1). \tag{68}$$

Define

$$F_N(\boldsymbol{\theta}) = \sum_{k=1}^{N} (\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_0)) \cdot \int_{t_{k-1}}^{t_k} \sigma_t \; dW_t.$$

Further define for any function $\phi : \mathcal{N}(\boldsymbol{\theta}_0) \to \mathbb{R}$, the modulus of continuity as follows

$$w(\phi, h) := \sup\{|\phi(\boldsymbol{\theta}_1) - \phi(\boldsymbol{\theta}_2)| : |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| \leq h\}, \quad \text{for any } h \geq 0.$$

Below we show that for all n such that $B(\boldsymbol{\theta}_0, K/n)(= \{\boldsymbol{\theta} : |\boldsymbol{\theta}-\boldsymbol{\theta}_0| \leq K/n\}) \subseteq \mathcal{N}(\boldsymbol{\theta}_0)$, for all $\ell \in \mathbb{N}$ large enough,

$$E(\sqrt{n} \; w(F_N, K/n))^{2\ell} = o(1), \tag{69}$$

which clearly implies (67). The proof of (69) is via a modification of the Kolmogorov-Centsov continuity theorem. More specifically, for any $\boldsymbol{\theta}_1, \; \boldsymbol{\theta}_2 \in \mathcal{N}(\boldsymbol{\theta}_0)$, by the BDG inequality (and Assumptions (SA.ii)), for any $\ell \in \mathbb{N}$ there exists $C_\ell < \infty$ such that

$$\begin{aligned}
E|F_N(\boldsymbol{\theta}_1) - F_N(\boldsymbol{\theta}_2)|^{2\ell} &\leq C_\ell E \left(\sum_{k=1}^{N} (\Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_1) - \Delta g(\mathbf{Z}_{t_k};\boldsymbol{\theta}_2))^2 \, \Delta t_k \right)^{\ell} \\
&\leq CE \left(\sum_{k=1}^{N} \left((L_0(\mathbf{Z}_{t_k}) + L_0(\mathbf{Z}_{t_{k-1}})) \cdot |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| \right)^2 \Delta t_k \right)^{\ell} \\
&\leq C|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|^{2\ell},
\end{aligned}$$

where in the second to the last inequality we used Assumption (A.vii) (and in the last inequality we used Assumption (SA.iv)). Following the argument in the proof of Corollary 14.9 in Kallenberg (1997) one obtains that for any $\ell > d/2$ and any $m \in \mathbb{N}$,

$$E(w(F_N, 2^{-m}))^{2\ell} \leq C 2^{-m(2\ell-d)}.$$

Taking m be such that $2^{-m} \geq K/n > 2^{-m-1}$ yields

$$E(\sqrt{n} w(F_N, K/n))^{2\ell} \leq C n^{\ell} \cdot (K/n)^{2\ell-d} = O(n^{-(\ell-d)}),$$

and therefore (69) holds for all $\ell > d$.

We now prove part (iii) of the theorem, which deals with the thresholded ERV. Note that

$$\Delta \widehat{X}_{t_k} = \Delta X_{t_k} + (\Delta g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0) - \Delta g(\mathbf{Z}_{t_k}; \widehat{\boldsymbol{\theta}})) := a_k + b_k. \tag{70}$$

By Assumption (A.vii) (and Assumption (SA.iv)), there exists $C > 0$ such that for all k, $|b_k| \leq C|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|$. Using Theorem 1 and Assumption (B.v) we then obtain that, with probability approaching one, for any $\xi \in (0, 1/2)$, $|b_k| \leq (\Delta t_k)^{\xi}/2$ for all k, under which case $|\Delta \widehat{X}_{t_k}| \geq (\Delta t_k)^{\xi}$ only if $|a_k| = |\Delta X_{t_k}| \geq (\Delta t_k)^{\xi}/2$. However, by Lévy's modulus of continuity theorem, under Assumption (B.iii), for all n large enough the latter could only happen when there is a jump during $[t_{k-1}, t_k]$. On the other hand, since (J_t) admits only finitely many jumps, for all n large enough, each interval $[t_{k-1}, t_k]$ would contain at most one jump, and if indeed there is a jump during $[t_{k-1}, t_k]$, then we will have that $|\Delta X_{t_k}| > 3/2 \cdot (\Delta t_k)^{\xi}$, which implies that $|\Delta \widehat{X}_{t_k}| > (\Delta t_k)^{\xi}$. Therefore we conclude that, with probability approaching one,

$$|\Delta \widehat{X}_{t_k}| \cdot \mathbf{1}_{|\Delta \widehat{X}_{t_k}| > (\Delta t_k)^{\xi}} = |\Delta X_{t_k} + b_k| \cdot \mathbf{1}_{\text{jump exists in } [t_{k-1}, t_k]} \quad \text{for all } k. \tag{71}$$

Using again the assumption that J_t admits only finitely many jumps and that $\sqrt{N}\Delta(\mathcal{G}) \xrightarrow{p} 0$ in assumption (B.iii) we see that

$$\begin{aligned} \text{ERV}_{\textit{threshold}} &= \sum_k |\Delta \widehat{X}_{t_k}|^2 \cdot \mathbf{1}_{|\Delta \widehat{X}_{t_k}| \leq (\Delta t_k)^{\xi}} \\ &= \sum_k (\Delta \widetilde{X}_{t_k} + b_k)^2 + o_p(1/\sqrt{n}), \end{aligned} \tag{72}$$

where $(\widetilde{X}_t)$ is the continuous part of (X_t) satisfying $d\widetilde{X}_t = \mu_t dt + \sigma_t \, dW_t$. The conclusion then follows from part (ii). □

A.3. Proof of Proposition 1

Proof. By a similar argument to (72), without loss of generality, we can assume that $J_t \equiv 0$. In this case, by Proposition 2 in Mykland and Zhang (2006), under Assumptions (B.ii)∼(B.iv), as $n \to \infty$, we have

$$\frac{N}{3} \sum_{k=1}^{N} |\Delta X_{t_k}|^4 \xrightarrow{p} \int_0^1 \sigma_t^4 \, dH_t. \tag{73}$$

To prove the proposition, it then suffices to show that

$$\frac{N}{3} \sum_{k=1}^{N} \left(|\Delta \widehat{X}_{t_k}|^4 - |\Delta X_{t_k}|^4 \right) \xrightarrow{p} 0. \tag{74}$$

Using the notation in the decomposition (70), we have that

$$\begin{aligned}&\sum_{k=1}^{N} \left(|\Delta \widehat{X}_{t_k}|^4 - |\Delta X_{t_k}|^4 \right) \\ =&4 \sum_{k=1}^{N} a_k^3 b_k + 6 \sum_{k=1}^{N} a_k^2 b_k^2 + 4 \sum_{k=1}^{N} a_k b_k^3 + \sum_{k=1}^{N} b_k^4.\end{aligned}$$

By Hölder's inequality, for $i = 1, 2, 3$,

$$\left| \sum_{k=1}^{N} a_k^i b_k^{4-i} \right| \leq \left(\sum_{k=1}^{N} a_k^4 \right)^{i/4} \left(\sum_{k=1}^{N} b_k^4 \right)^{(4-i)/4}.$$

By (73), if we can show that

$$N \sum_{k=1}^{N} b_k^4 = o_p(1),$$

then (74) will follow. The last claim is true thanks to Assumptions (A.vii) and (SA.iv) and Theorem 1. □

A.4. Proof of Theorem 3

Proof. Part (i) can be proved by slight modifications of the proof of Theorem 1. The key property used is that

$$\sum_k (\Delta \varepsilon_{t_k})^2 \leq 2 \sum_k \varepsilon_{t_k}^2 = O_p(1)$$

thanks to condition (20), where $\Delta \varepsilon_{t_k} = \varepsilon_{t_k} - \varepsilon_{t_{k-1}}$ for any k.

We now prove part (ii). Firstly, by a similar reasoning for (72) and using (24), we can assume without loss of generality that the jump does not exist. In this case, we can further assume that

(SA.i) $\mu_t \equiv 0$,

see, for example, Section 2.2 (p.1407-1409) of Mykland and Zhang (2009). Furthermore, using the conclusion of part (i) and by a similar argument for (66), we have that

$$\sqrt{N}\left(\mathrm{ERV}_{ext} - \mathrm{RV}_{ext}\right) = o_p(1), \tag{75}$$

where

$$\mathrm{RV}_{ext} = \sum_{k=1}^{N}(\Delta X_{t_k} + \Delta\varepsilon_{t_k})^2 + 2\sum_{k=2}^{N}(\Delta X_{t_k} + \Delta\varepsilon_{t_k})(\Delta X_{t_{k-1}} + \Delta\varepsilon_{t_{k-1}}).$$

Simple algebra and that $\sqrt{N}\Delta(\mathcal{G}) \xrightarrow{p} 0$ in assumption (B.iii) yields

$$\begin{aligned}
&\mathrm{RV}_{ext}\\
=&\sum_k \Delta X_{t_k}^2 + 2\sum_k \Delta X_{t_{k-1}}\Delta X_{t_k} - 2\sum_k \big((X_{t_k} - X_{t_{k-1}}) - (X_{t_{k-3}} - X_{t_{k-4}})\big)\varepsilon_{t_{k-2}}\\
&+ 2\sum_k(\varepsilon_{t_{k-1}} - \varepsilon_{t_{k-2}})\varepsilon_{t_k} + o_p(1/\sqrt{n})\\
:=& I + 2II - 2III + 2IV + o_p(1/\sqrt{n}),
\end{aligned}$$

where the $o_p(1/\sqrt{n})$ is due to the edge effects. Term I is just the realized volatility, for which we have the CLT (65) under Assumptions (B.ii) – (B.iv). As to term II, it is a martingale under Assumptions (SA.i) and (SA.ii), and we have that

$$E(II^2) = E\left(\sum_k \Delta X_{t_{k-1}}^2 \Delta X_{t_k}^2\right) \leq CE\left(\sum_k \Delta t_{k-1}\Delta t_k\right).$$

By Assumption (B.vi) we then get that $\sqrt{N}II = O_p(1)$. As to term III, we have

$$E(III^2|\mathcal{F}_1) = E\left(\sum_k \big((X_{t_k} - X_{t_{k-1}}) - (X_{t_{k-3}} - X_{t_{k-4}})\big)^2 \varepsilon_{t_{k-2}}^2 \middle| \mathcal{F}_1\right) \leq 2\sigma_n^2 \cdot \sum_k (X_{t_k} - X_{t_{k-1}})^2,$$

and hence $\sqrt{N}III = O_p(1)$ by Assumptions (B.i) and (20). Finally,

$$E(IV^2|\mathcal{F}_1) = E\left(\sum_k (\varepsilon_{t_{k-1}} - \varepsilon_{t_{k-2}})^2\varepsilon_{t_k}^2 |\mathcal{F}_1\right) = 2N\sigma_n^4,$$

hence, again by Assumptions (B.i) and (20), we have that $\sqrt{N}IV = O_p(1)$.

We now establish CLT under Assumptions (B.ii)∼(B.v) and (B.vii). In this case, we have $N = n$, and (65) holds. Furthermore, by the martingale CLT we have, stably in law,

$$\begin{aligned}
\sqrt{N}II &\xrightarrow{\mathcal{L}} \Phi_2 \times \left(\int_0^1 \sigma_t^4\, dQ_t\right)^{1/2},\\
\sqrt{N}III &\xrightarrow{\mathcal{L}} \Phi_3 \times \left(2\sigma_\varepsilon^2 \int_0^1 \sigma_t^2\, dt\right)^{1/2},\\
\sqrt{N}IV &\xrightarrow{\mathcal{L}} \Phi_4 \times \left(2\sigma_\varepsilon^4\right)^{1/2},
\end{aligned} \tag{76}$$

where $\Phi_i,\ i=2,3,4$, are standard normal random variables independent of $\mathcal{F}_1$. To show the desired convergence, it then remains to show that the $\Phi_i, i=2,3,4$ in (76) and the Φ in (65) are mutually independent.

We shall only show that Φ and Φ_1 are independent; the independence of other pairs can be shown similarly (and actually slightly more easily).

To prove that Φ and Φ_1 are independent, firstly we have

$$\sqrt{N}\left(I-\int_0^1 \sigma_t^2\ dt\right)=M_1,$$

where $(M_t)_{t\in[0,1]}$, a martingale, is the interpolated and rescaled error process defined by

$$dM_t = 2\sqrt{N}(X_t - X_{t_{k^*}})\ dX_t\ ,\quad M_0=0,$$

where k^* is the largest k such that $t_k\le t$ (see, for example, the proof of Proposition 2 (p. 1952) of Mykland and Zhang (2006) or the proof of Theorem 1 in Li et al. (2014)). Similarly, term II is the end value of a martingale $(II_t)_{t\in[0,1]}$ defined as

$$dII_t=\sqrt{N}\Delta X_{t_{k^*-1}}\ dX_t,\quad II_0=0.$$

Recall that $\nu_t=\sigma_t^2$. It follows that the quadratic covariation of (M_t) and (II_t) equals

$$\langle M, II\rangle_t = 2N\left(\sum_{k\le k^*}\Delta X_{t_{k-1}}\int_{t_{k-1}}^{t_k}(X_t-X_{t_{k-1}})\nu_t\ dt + \Delta X_{t_{k^*-1}}\int_{t_{k^*}}^{t}(X_t-X_{t_{k^*}})\nu_t\ dt\right).$$

To show that Φ and Φ_2 are independent, it suffices to show that $\langle M, II\rangle_t\to 0$ for all $t\le 1$. For ease of exposition, we shall only deal with the major term

$$\begin{aligned}&2N\sum_{k\le k^*}\Delta X_{t_{k-1}}\int_{t_{k-1}}^{t_k}(X_t-X_{t_{k-1}})\nu_t\ dt\\ =&2N\sum_{k\le k^*}\Delta X_{t_{k-1}}\nu_{t_{k-1}}\int_{t_{k-1}}^{t_k}(X_t-X_{t_{k-1}})\ dt+2N\sum_{k\le k^*}\Delta X_{t_{k-1}}\int_{t_{k-1}}^{t_k}(X_t-X_{t_{k-1}})(\nu_t-\nu_{t_{k-1}})\ dt\\ :=&2(A+B).\end{aligned}$$

It is easy to show by conditioning that for any $k_1<k_2$,

$$E\left(\Delta X_{t_{k_1-1}}\nu_{t_{k_1-1}}\int_{t_{k_1-1}}^{t_{k_1}}(X_t-X_{t_{k_1-1}})\ dt\cdot\Delta X_{t_{k_2-1}}\nu_{t_{k_2-1}}\int_{t_{k_2-1}}^{t_{k_2}}(X_t-X_{t_{k_2-1}})\ dt\right)=0.$$

Therefore

$$
\begin{aligned}
E(A^2) &= N^2 \sum_{k\le k^*} E\left(\Delta X_{t_{k-1}}^2 \nu_{t_{k-1}}^2 \left(\int_{t_{k-1}}^{t_k} (X_t - X_{t_{k-1}})\ dt\right)^2\right) \\
&\le N^2 \sum_{k\le k^*} E\left(\Delta X_{t_{k-1}}^2 \nu_{t_{k-1}}^2 \Delta t_k \int_{t_{k-1}}^{t_k} (X_t - X_{t_{k-1}})^2\ dt\right) \\
&= N^2 \sum_{k\le k^*} E\left(\Delta X_{t_{k-1}}^2 \nu_{t_{k-1}}^2 \Delta t_k \int_{t_{k-1}}^{t_k} E\left((X_t - X_{t_{k-1}})^2 \mid \mathcal{F}_{t_{k-1}}\right)\ dt\right) \\
&\le CN^2 \sum_{k\le k^*} E\left(\Delta X_{t_{k-1}}^2 \nu_{t_{k-1}}^2 \Delta t_k \int_{t_{k-1}}^{t_k} (t - t_{k-1})\ dt\right) \\
&\le CN^2 \sum_{k\le k^*} \Delta t_{k-1} (\Delta t_k)^3,
\end{aligned}
$$

where in the first inequality we used the Cauchy-Schwarz inequality, and in the last two inequalities we used Assumptions (SA.i) and (SA.ii). By Assumption (B.iii), the last term is $o(1)$.

As to term B, we have

$$
\begin{aligned}
E|B| \le& N \sum_{k\le k^*} E\Big(|\Delta X_{t_{k-1}}| \cdot \int_{t_{k-1}}^{t_k} E\big(\ |(X_t - X_{t_{k-1}}) \cdot (\nu_t - \nu_{t_{k-1}})|\big|\ \mathcal{F}_{t_{k-1}}\big)\ dt\Big) \\
\le& N \sum_{k\le k^*} E\Big(|\Delta X_{t_{k-1}}| \cdot \int_{t_{k-1}}^{t_k} \sqrt{E((X_t - X_{t_{k-1}})^2 | \mathcal{F}_{t_{k-1}}) \cdot E((\nu_t - \nu_{t_{k-1}})^2 | \mathcal{F}_{t_{k-1}})}\ dt\Big) \\
\le& CN \sum_{k\le k^*} E\Big(|\Delta X_{t_{k-1}}| \cdot \int_{t_{k-1}}^{t_k} (t - t_{k-1})\ dt\Big) \\
\le& CN \sum_{k\le k^*} \sqrt{\Delta t_{k-1}} (\Delta t_k)^2,
\end{aligned}
$$

where in the last two inequalities we used Assumptions (SA.i), (SA.ii) and (52). Finally, note that

$$
N \sum_{k\le k^*} \sqrt{\Delta t_{k-1}} (\Delta t_k)^2 \le \sqrt{\sum_{k\le k^*} \Delta t_{k-1} \cdot N^2 \sum_{k\le k^*} (\Delta t_k)^4},
$$

which is $o(1)$ again by Assumption (B.iii). □

A.5. Proof of Proposition 2

Proof. The proof is decomposed into several steps, each step towards estimating certain term in the asymptotic variance in (29). By similar arguments as at the beginning of the proof of

Theorem 3, we can assume that the jump does not exist, (SA.i) holds, and furthermore the conclusion would follow if we can show that

$$\begin{aligned}&\frac{2N}{3}\left(\sum_{k=2}^{N}(X_{t_k}-X_{t_{k-2}}+\varepsilon_{t_k}-\varepsilon_{t_{k-2}})^4-\sum_{k=1}^{N}(\Delta X_{t_k}+\Delta\varepsilon_{t_k})^4\right)\\&+2\left(\sum_{k=1}^{N}(\Delta X_{t_k}+\Delta\varepsilon_{t_k})^2-\mathrm{ERV}_{ext}\right)^2\\&\xrightarrow{p}2\int_0^1\sigma_t^4\,dH_t+4\int_0^1\sigma_t^4\,dQ_t+8\sigma_\varepsilon^2\,\mathrm{IV}+8\sigma_\varepsilon^4.\end{aligned}\tag{77}$$

<u>Step 1</u>: Estimate IV: by Theorem 3, $\mathrm{ERV}_{ext}\xrightarrow{p}\mathrm{IV}$;

<u>Step 2</u>: Estimate $\sigma_\varepsilon^2(=\lim_{n\to\infty}n\sigma_n^2)$. Consider

$$\begin{aligned}B:=\sum_{k=1}^{N}(\Delta X_{t_k}+\Delta\varepsilon_{t_k})^2=&\sum_{k=1}^{N}(\Delta X_{t_k})^2+2\sum_{k=1}^{N}\Delta X_{t_k}\cdot\Delta\varepsilon_{t_k}+\sum_{k=1}^{N}(\Delta\varepsilon_{t_k})^2\\:=&\mathrm{RV}+B.2+B.3.\end{aligned}$$

We have that RV $\xrightarrow{p}$ IV, and furthermore by using (20) it is easy to show that $B.2\xrightarrow{p}0$ and $B.3\xrightarrow{p}2\lim_n N\sigma_n^2=2\sigma_\varepsilon^2$ (recall that under Assumption (B.ii) we can and we do take $n=N$). To sum up we get that

$$B\xrightarrow{p}\mathrm{IV}+2\sigma_\varepsilon^2,\tag{78}$$

therefore,

$$\frac{B-\mathrm{ERV}_{ext}}{2}\xrightarrow{p}\sigma_\varepsilon^2.\tag{79}$$

<u>Step 3</u>: Estimate the quarticity. we have

$$\begin{aligned}&C:=N\sum_{k=1}^{N}(\Delta X_{t_k}+\Delta\varepsilon_{t_k})^4\\=&N\sum_{k=1}^{N}(\Delta X_{t_k})^4+N\sum_{k=1}^{N}(\Delta\varepsilon_{t_k})^4+6N\sum_{k=1}^{N}(\Delta X_{t_k})^2(\Delta\varepsilon_{t_k})^2\\&+4N\sum_{k=1}^{N}(\Delta X_{t_k})^3\cdot(\Delta\varepsilon_{t_k})+4N\sum_{k=1}^{N}(\Delta X_{t_k})\cdot(\Delta\varepsilon_{t_k})^3\\:=&C.1+C.2+6\ C.3+4\ C.4+4\ C.5.\end{aligned}$$

By (73), $C.1\xrightarrow{p}3\int_0^1\sigma_t^3\,dH_t$. Furthermore, by (20) and (31) it is straightforward to show that $C.2\xrightarrow{p}2\kappa+6\sigma_\varepsilon^4$. As to $C.3$, by (20) one can show that $C.3\xrightarrow{p}2\sigma_\varepsilon^2$ IV. Finally, by computing the second moment and using Assumption (B.iii) and (31) one can show that both $C.4$ and $C.5$ are $o_p(1)$. To sum up, we have

$$C \xrightarrow{p} 3\int_0^1 \sigma_t^3 \; dH_t + 2\kappa + 6\sigma_\varepsilon^4 + 12\sigma_\varepsilon^2 \text{ IV}. \tag{80}$$

Step 4: Estimate the "interlaced quarticity" $\int_0^1 \sigma_t^4 \; dQ_t$. Consider

$$\begin{aligned} D :=& N\sum_{k=2}^{N}(X_{t_k} - X_{t_{k-2}} + \varepsilon_{t_k} - \varepsilon_{t_{k-2}})^4 \\ =& N\sum_{k=2}^{N}(X_{t_k} - X_{t_{k-2}})^4 + N\sum_{k=2}^{N}(\varepsilon_{t_k} - \varepsilon_{t_{k-2)}})^4 + 6\; N\sum_{k=2}^{N}(X_{t_k} - X_{t_{k-2}})^2(\varepsilon_{t_k} - \varepsilon_{t_{k-2}})^2 \\ &+ \text{remainder} \\ :=& D.1 + D.2 + 6\; D.3 + o_p(1), \end{aligned}$$

where the remainder term is $o_p(1)$ by the same reasoning as in the previous step. Furthermore, by similar proofs to that for (73) and using (B.vii) we have

$$\begin{aligned} D.1 &= N\sum_{k=2}^{N} 3\sigma_{t_{k-2}}^4 \cdot (\Delta t_k + \Delta t_{k-1})^2 + o_p(1) \\ &= 3N\sum_{k=1}^{N} \sigma_{t_{k-2}}^4((\Delta t_k)^2 + (\Delta t_{k-1})^2) + 6N\sum_{k=2}^{N} \sigma_{t_{k-2}}^4(\Delta t_k)(\Delta t_{k-1}) + o_p(1) \\ &\xrightarrow{p} 6\int_0^1 \sigma_t^4 \; dH_t + 6\int_0^1 \sigma_t^4 \; dQ_t. \end{aligned}$$

As to $D.2$ and $D.3$, just as in the previous step we have that $D.2 \xrightarrow{p} 2\kappa + 6\sigma_\varepsilon^4$, and $D.3 \xrightarrow{p} 4\sigma_\varepsilon^2$ IV. To sum up we obtain that

$$D \xrightarrow{p} 6\int_0^1 \sigma_t^4 \; dH_t + 6\int_0^1 \sigma_t^4 \; dQ_t + 2\kappa + 6\sigma_\varepsilon^4 + 24\sigma_\varepsilon^2 \text{ IV}. \tag{81}$$

Combining this with (80) we get

$$\frac{D - C}{3} \xrightarrow{p} \int_0^1 \sigma_t^4 \; dH_t + 2\int_0^1 \sigma_t^4 \; dQ_t + 4\sigma_\varepsilon^2 \text{ IV}. \tag{82}$$

Step 5: Combining the aforementioned convergences we get the desired convergence (77).

□

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2000), "Great realizations," *Risk*, 13, 105–108.

Almgren, R. and Chriss, N. (2000), "Optimal execution of portfolio transactions," *Journal of Risk*, 3, 5–39.

Aït-Sahalia, Y., Fan, J., and Li, Y. (2013), "The leverage effect puzzle: Disentangling sources of bias at high frequency," *Journal of Financial Economics*, 109, 224–249.

Aït-Sahalia, Y., Jacod, J., 2009. Testing for jumps in a discretely observed process. Annals of Statistics 37, 184–222.

Aït-Sahalia, Y. and Jacod, J. (2014), "High Frequency Financial Econometrics," *Princeton University Press.*

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005). How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. *Review of Financial Studies*, **18** 351–416.

— (2011), Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise. *Journal of Econometrics* **160**, 190-203.

Barndorff-Nielsen, O. E. and Shephard, N. (2004), "Power and bipower variation with stochastic volatility and jumps (with discussion)," *Journal of Financial Econometrics*, 2, 1-48.

Barndorff-Nielsen, O.E., Graversen, S.E., J. Jacod, and Shephard, N. (2006), "Limit theorems for bipower variation in financial econometrics," *Econometric Theory*, 22, 677–719.

Barndorff-Nielsen, O.E., Graversen, S.E., J. Jacod, Podolskij, M. and Shephard, N. (2006), "A central limit theorem for realised power and bipower variations of continuous semimartingales," in *From Stochastic Calculus to Mathematical Finance, The Shiryaev Festschrift*, Yu. Kabanov and R. Liptser and J. Stoyanov, eds. (Springer-Verlag, Berlin), p. 33-69.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008), "Designing realized kernels to measure the ex post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

— (2009), “Realized kernels in practice: Trades and quotes,” *The Econometrics Journal*, 12, C1–C32.

Barndorff-Nielsen, O. E., Hansen, P.R., Lunde, A. and Shephard N. (2011). “Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading.” *Journal of Econometrics*, 162, 149–169.

Biais, B., Hillion, P., and Spatt, C. (1995), “An empirical analysis of the limit order book and the order flow in the Paris Bourse,” *Journal of Finance*, 50, 1655–1689.

Białkowski, J., Darolles, S., and Le Fol, G. (2008), “Improving vwap strategies: A dynamic volume approach,” *Journal of Banking & Finance*, 32, 1709–1722.

Black, F. (1986), “Noise,” *Journal of Finance*, 41, 529–543.

Chaker, S. (2013), “Volatility and Liquidity Costs,” *Working Paper.*

Delattre, S. and Jacod, J. (1997), “A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors,” *Bernoulli*, 3, 1–28.

Ellis, K., Michaely, R., and O’Hara, M. (2000), “The accuracy of trade classification rules: evidence from Nasdaq,” *Journal of Financial and Quantitative Analysis*, 35, 529–551.

Engle, R. F. and Sun, Z. (2007), “When is Noise Not Noise - A Microstructure Estimate of Realized Volatility,” *NYU Working Paper No. FIN-07-047. Available at SSRN: http://ssrn.com/abstract=1293630.*

Glosten, L. R. and Harris, L. E. (1988), “Estimating the components of the bid/ask spread,” *Journal of financial Economics*, 21, 123–142.

Gloter, A. and Jacod, J. (2001), “Diffusions with measurement errors. ii - optimal estimators,” *ESAIM*, 5, 243–260.

Gourieroux, C., Jasiak, J., and Le Fol, G. (1999), “Intra-day market activity,” *Journal of Financial Markets*, 2, 193–226.

Hansen, P. R. and Horel, G. (2009), “Quadratic variation by Markov chains,” *working paper.*

Hansen, P. R.,Large, J., and Lunde, A. (2008), “Moving average-based estimators of integrated variance,” *Econometric Reviews*,27, 79C111.

Hansen, P. R. and Lunde, A. (2006), "Realized variance and market microstructure noise," *Journal of Business and Economic Statistics*, 24, 127–218, with comments and a rejoinder by the authors.

Hasbrouck, J. (2007), *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*, Oxford.

Heston, S. L. (1993), "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of financial studies*, 6, 327–343.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009), "Microstructure noise in the continuous case: the pre-averaging approach," *Stochastic Processes and their Applications*, 119, 2249–2276.

Jacod J., Li, Y. and Zheng, X. (2014). Statistical Properties of Microstructure Noise: II. Available at SSRN: *http://ssrn.com/abstract=2212119*.

— (2015). Estimating the Integrated Volatility When Microstructure Noise is Dependent and Observation Times are Irregular. Available at SSRN: *http://ssrn.com/abstract=2659615*

Jacod, J. and Protter, P. (2012), *Discretization of processes*, Stochastic Modelling and Applied Probability, Heidelberg: Springer.

Kallenberg, O. (1997), *Foundations of modern probability*, Probability and its Applications (New York), New York: Springer-Verlag.

Keim, D. B. and Madhavan, A. (1996), "The Upstairs Market for Large-Block Transactions: Analysis and Measurement of Price Effects," *Review of Financial Studies*, 9, 1–36.

Kim, S. and Murphy, D. (2011), "The impact of high-frequency trading on stock market liquidity measures," Tech. rep., Working Paper, Kellogg School of Management at Northwestern University.

Large, J. (2011), "Estimating quadratic variation when quoted prices change by a constant increment," *Journal of Econometrics*, 160, 2–11.

Lee, C. and Mark J. R. (1991), "Inferring trade direction from intraday data," *The Journal of Finance*, 46, 733–746.

Li, Y., and Mykland, P.A. (2007). Are volatility Estimators Robust with Respect to Modeling Assumptions? *Bernoulli,* **13** 601-622.

— (2015). Rounding Errors and Volatility Estimation. *Journal of Financial Econometrics,* 13(2), 2015, 478 - 504.

Li, Y., Mykland, P. A., Renault, E., Zhang, L., and Zheng, X. (2014), "Realized volatility when sampling times are possibly endogenous," *Econometric Theory*, 30, 580–605.

Li, Y., Zhang, Z., and Li, Y. (2015), "A Unified Approach to Volatility Estimation in the Presence of Both Rounding and Random Market Microstructure Noise," *working paper.*

Li, Y., Zhang, Z., and Zheng, X. (2013), "Volatility inference in the presence of both endogenous time and microstructure noise," *Stochastic Processes and their Applications*, 123, 2696–2727.

Madhavan, A. (2000), "Market microstructure: A survey," *Journal of Financial Markets*, 3, 205 – 258.

Madhavan, C. (2009), "Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps," *Scandinavian Journal of Statistics. Theory and Applications*, 36, 270–296.

Mykland, P. A. and Zhang, L. (2006), "ANOVA for diffusions and Itô processes," *Annals of Statistics*, 34, 1931–1963.

— (2009), "Inference for continuous semimartingales observed at high frequency," *Econometrica*, 77, 1403–1445.

— (2012), "The econometrics of high frequency data," *Statistical methods for stochastic differential equations*, 109–190.

O'Hara, M. (1995), *Market Microstructure Theory*, Oxford: Blackwell.

Podolskij, M. and Vetter, M. (2009), "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli*, 15, 634–658.

Roll, R. (1984), "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39, 1127–1139.

Rosenbaum, M. (2009). Integrated Volatility and Round Off Error. *Bernoulli,* **15** 687–720.

Stoll, H. (2003), *Market microstructure, Handbook of Economics and Finance*, Elsevier Science B.V.

White, H. (1980), "Nonlinear regression on cross-section data," *Econometrica: Journal of the Econometric Society*, 721–746.

Xiu, D. (2010), "Quasi-maximum likelihood estimation of volatility with high frequency data," *Journal of Econometrics*, 159, 235–250.

Zhang, L. (2006), "Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach," *Bernoulli*, 12, 1019–1043.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005), "A tale of two time scales: determining integrated volatility with noisy high-frequency data," *Journal of the American Statistical Association*, 100, 1394–1411.