Journal of Nonparametric Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gnst20

Robust covariance estimation with noisy high-frequency financial data

Jiandong Wang & Manying Bai

JOURNAL OF NONPARAMETRIC STATISTICS
https://doi.org/10.1080/10485252.2022.2075549

Robust covariance estimation with noisy high-frequency financial data

Jiandong Wang and Manying Bai

School of Economics and Management, Beihang University, Beijing, People's Republic of China

ABSTRACT

We propose consistent and efficient robust different time-scales estimators to mitigate the heavy-tail effect of high-frequency financial data. Our estimators are based on minimising the Huber loss function with a suitable threshold. We show these estimators are guaranteed to be robust to measurement noise of certain types and jumps. With only finite fourth moments of the observation log-price data, we develop the sub-Gaussian concentration of our estimators around the volatility. We conduct the simulation studies to show the finite sample performance of the proposed estimation methods. The simulation studies imply that our methods are also robust to financial data in the presence of jumps. Empirical studies demonstrate the practical relevance and advantages of our estimators.

ARTICLE HISTORY
Received 14 December 2020
Accepted 1 April 2022

KEYWORDS
Heavy-tail; Huber loss; different time-scales; concentration inequality

1. Introduction

Volatility plays a crucial role in modern finance both academics and practitioners. The wide availability of high-frequency financial data makes it possible to estimate precisely the integrated volatility of security prices. A consistent and efficient estimator of integrated volatility is the sum of squared log-return, which is called realized volatility (RV). And the RV estimator enjoys the optimal rate $n^{1/2}$. It is commonly agreed that high-frequency financial data is in the presence of microstructure noise, which would violate the desirable properties of this estimator. Parametric estimate the microstructure noise are investigated by Li, Xie, and Zheng (2016) and Clinet and Potiron (2018, 2021). Many researchers have developed well-performance methods to reduce the effect of microstructure noise. For example, Zhang, Mykland, and Aït-Sahalia (2005) proposed a nonparametric estimator, two-scale realized volatility (TSRV), which is the first consistent estimation with the additive microstructure noise, despite a relative slower convergence rate $n^{1/6}$. To achieve the optimal convergence rate $n^{1/4}$, which was shown by Gloter and Jacod (2001), Zhang (2006) subsequently extend the TSRV to develop the multi-scale realized volatility (MSRV). Other than the post average of TSRV and MSRV, Christensen, Kinnebrock, and Podolskij (2010) and Jacod, Li, Mykland, Podolskij, and Vetter (2009) consider the Pre-averaging realized volatility estimator (PRV). This estimator also performs well and enjoys

CONTACT Jiandong Wang jiandongdong4388@163.com

the optimal convergence rate. Other methods include realized kernel realized volatility (KRV) (Barndorff-Nielsen, Hansen, Lunde, and Shephard 2008), wavelet estimator (Fan and Wang 2007), a quasi-maximum likelihood estimator (QMLE, Xiu 2010).

Since one can estimate the single asset realized volatility, it is natural to extend to estimate the realized covariance of two or more assets. When considering the volatility estimation of multiple assets, Zhang (2011) based on TSRV estimator and proposed the two-scale realized covariance estimator (TSCV). Recently, Mykland, Zhang, and Chen (2019) derived a new algebraic property of two scales estimation in high frequency, called smoothed TSRV estimator. Aït-Sahalia, Fan, and Xiu (2010) extended the one-dimensional QMLE volatility estimator to a quasi-maximum likelihood covariance estimator. On the other hand, large-dimensional or high-dimensional matrix estimation needs to impose some structure assumptions such as sparsity or bandability. In the meantime, the factor model offers another useful approach to matrix estimation. Similar to the above usual matrix estimation, high-dimensional and high-frequency volatility matrix estimation methods can refer to the literature (Wang and Zou 2010; Tao, Wang, and Zhou 2013; Kim, Wang, and Zou 2016; Kim, Kong, Li, and Wang 2018; Kong 2018; Pelger 2019; Shin, Kim, and Fan 2021).

Unlike low-frequency financial data, which is always homogeneously spaced, high-frequency financial data of two assets are rarely simultaneous, usually occur randomly and asynchronously. Thus estimating the realized covariance between multiple assets in the asynchronous case will cause some undesirable features in inference. Especially early studies of the correlation of stock price returns, using high-frequency asynchronous data, shows that tends to have a strong bias towards zero as the sampling interval shrinks. This effect was called the 'Epps effect' as well, which was documented by Epps (1979). Before applying the covariance estimator, one needs to synchronize the data. Mykland et al. (2019) developed an asynchronous method that obtains the synchronous data by taking the mean of the asynchronous trading volumes in some time scale like 1 min. Refresh Time Scheme was suggested by Barndorff-Nielsen et al. (2008) to synchronize the high-frequency asynchronous financial data. Fan, Li, and Ke (2010) used the 'pairwise refresh' and 'all-fresh' scheme to handle the asynchronous trading. The generalized synchronized method was investigated by Aït-Sahalia et al. (2010). Zhang (2011) provide the previous tick method to deal with the asynchronous data. In this paper, we extend the previous tick method and propose the generalized previous tick scheme.

Under the bound spot volatility and sub-Gaussian assumption microstructural noise, Tao et al. (2013) show that the realized volatility matrix has sub-Gaussian concentration with convergence rate $n^{-1/4}$. As well-document stylized features of financial data, financial returns usually exhibit heavy-tails. Unfortunately, current volatility estimation methods usually rely on light-tail assumptions and trade-offs the convergence rate and the order finite moments. Thus, one cannot apply these methods directly and this demands developing a robust estimation method that can obtain the sub-Gaussian concentration with only finite fourth moments. Shin et al. (2021) proposed a robust pre-average method to deal with the heterogeneous heavy-tailed distributions. Recently, Fan, Wang, and Zhong (2018) introduced the robust estimator under i.i.d. observations condition, by minimizing the Huber loss function, introduced by Huber (1964), with a diverging threshold to reduce biases, the proposed estimator relax the sub-Gaussian tail assumption with the only finite fourth moment. For more relative work, one can also refer to Fan, Wang, and Zhong (2019). Motivated by the above robust estimation method, Fan and

Kim (2018) apply the Huber loss function to the high-frequency volatility matrix to obtain the robust estimator. Zhang (2011) point out there still exists the effect of microstructure noise with two different assets. Unlike the pre-averaging method, in this paper, we propose another robust estimators under the time-scale framework. We lighten the heavy-tail influence of microstructure noise by minimizing the Huber loss function to the two time-scale volatility with suitable tuning parameters. Finally, we obtain the robust estimator by combining the different time-scales quadratic variations. The proposed estimators also share the sub-Gaussian concentration around the volatility with only finite fourth moments of the observed log returns. Due to the first robust estimator cannot attain the optimal rate, we generalized the robust two time-scale to a robust multi-time scale estimator. The paper is organized as follows. Section 2 gives the model setup. Sections 3 and 4 detail two time-scale volatility estimators and covariance respectively. Section 5 extends the robust two time-scale estimators. The simulations is conducted in Section 6. Finally, we conduct the empirical study in Section 7.

2. Model setup

In this section, we recapitulate the Itô process of log returns. Consider the price processes of two assets, Suppose both $\{X_t\}$ and $\{Y_t\}$ follow an Itô process, namely,

$$\mathrm{d}\tilde{X}(t) = \mu^x(t)\,\mathrm{d}t + \sigma^x(t)\,\mathrm{d}W_t^x, \tag{1}$$

$$\mathrm{d}\tilde{Y}(t) = \mu^y(t)\,\mathrm{d}t + \sigma^y(t)\,\mathrm{d}W_t^y, \tag{2}$$

where $\mu^x(t), \mu^y(t)$ and $\sigma^x(t), \sigma^y(t)$ are the drift term and the instantaneous variance respectively, and W_t^x and W_t^y are two standard Brownian motion with correlation $\rho(t)$. Furthermore, stochastic processes $\tilde{X}(t)$, $\tilde{Y}(t)$, $\mu^x(t)$, $\mu^y(t)$, $\sigma^x(t)$, $\sigma^y(t)$ are defined on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ with $\{\mathcal{F}_t\}$ satisfies the usual condition. What we interest is the integrated volatility in the fixed time interval $[0, T]$, that is

$$\langle \tilde{X}, \tilde{X} \rangle_T = \int_0^T (\sigma^x(t))^2\,\mathrm{d}t,$$

$$\langle \tilde{Y}, \tilde{Y} \rangle_T = \int_0^T (\sigma^y(t))^2\,\mathrm{d}t,$$

where $\langle \cdot, \cdot \rangle$ is the inner product and the integrated covariance

$$\langle \tilde{X}, \tilde{Y} \rangle_T = \int_0^T \sigma^x(t)\sigma^y(t)\,\mathrm{d}\langle W_t^x, W_t^y \rangle$$

It is well agreed that financial data usually contains noise. Researchers commonly call this noise as the microstructure noise in the high-frequency finance literature. Due to the microstructure noise, the log price processes cannot be observed directly. They are usually assumed that the observed log price processes $X(t_i)$ and $Y(t_j)$ contain the latent log-returns with additive microstructure noises, i.e.

$$X(t_i) = \tilde{X}(t_i) + \epsilon^x(t_i), \quad i = 1, \ldots, n^x \tag{3}$$

$$Y(t_j) = \tilde{Y}(t_j) + \epsilon^y(t_j), \quad j = 1, \ldots, n^y \tag{4}$$

where n^x, n^x are the observable sample of x and y, and the microstructure noise $\epsilon^x(t_i), \epsilon^y(t_j)$ are independent with mean zero and variance η. We assume the microstructure noise are independent of the latent log price processes.

3. Robust two-scale realized volatility

3.1. Review of TSRV

In this subsection, we give a brief review of the TSRV estimator. The Motivation of TSRV is based on subsampling, averaging and bias-correction. The main idea is that separate original time, $\mathcal{S} = \{t_1, \dots, t_n\}$. And the subsampling of the original denote $\mathcal{S}^{(l)}$, $l = 1, \dots, K$, where $K = o(n)$. The $\mathcal{S}^{(K)}$ means that one starts at the lth observation time, t_l, and take the second observation time after time-scale K with K an integer, such as 5 min, that is t_{l+K}. Therefore one can obtain these different time-scale square variations by computing sum of square of these different time-scale log-returns, denoted as $[X, X]_T^{(\text{sparse},l)}$, $l = 1, \dots, K$. Finally, taking average above sparse quadratic variations as follows

$$[X, X]_T^{(K)} = \frac{1}{K} \sum_{l=1}^{K} [X, X]_T^{(\text{sparse},l)} = \frac{1}{K} \sum_{i=1}^{n-K+1} (X(t_{i+K}) - X(t_i))^2.$$

we denote $[X, X]_T^{(K)}$ as low frequency part. Since $[X, X]_T^{(K)}$ remain bias, then one can subsample the original more high frequency, denoted as $\mathcal{S}^{(J)}$, $l = 1, \dots, J$. Similar to the computing process, we can obtain the second subsample quadratic variations $[X, X]_T^{(J)}$. Now one can combine above two different time-scales subsample quadratic variations to obtain the TSRV estimator:

$$\widehat{\langle X, X \rangle}_T = [X, X]_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J} [X, X]_T^{(J)}, \tag{5}$$

where $1 \leq J \ll K$ and $K = O(n^{2/3})$, $\bar{n}_K = \frac{n-K+1}{K}$, $\bar{n}_J$ is similar to the $\bar{n}_K$. The adjust sample bias for finite sample, we have

$$\widehat{\langle X, X \rangle}_T^{(tsrv)} = \left(1 - \frac{\bar{n}_K}{\bar{n}_J}\right)^{-1} \left([X, X]_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J} [X, X]_T^{(J)}\right). \tag{6}$$

Zhang et al. (2005) show that the convergence rate of this estimator is not optimal:

$$\widehat{\langle X, X \rangle}_T^{(tsrv)} - \langle \tilde{X}, \tilde{X} \rangle_T = O_p(n^{-1/6}). \tag{7}$$

Despite the strengths of this estimator, it does not take into account for the market feature of heavy-tails. Therefore the sub-Gaussian assumption of microstructure noise is violated. Unfortunately, without the assumption of sub-Gaussian distribution for microstructural noise, these realized volatility matrix cannot achieve the sub-Gaussian concentration.

3.2. Robust two-time scale realized volatility

Financial assets have a heavy-tail impact, which necessitates the development of novel estimators to mitigate this effect. Recently, Fan and Kim (2018) applied Huber loss function and developed a robust high-dimension volatility matrix estimation with high-frequency financial data. Motivated by their M-estimator, we construct a new robust volatility estimator as follows.

First, consider the low-frequency part. The log price $X(t_{i+K}) - X(t_i)$ consists of the true log price $\tilde{X}(t_{i+K}) - \tilde{X}(t_i)$ and microstructure noise $\epsilon(t_{i+K}) - \epsilon(t_i)$, $i = 1, \ldots, n-K+1$. Their volatility also comes from the true volatility and the variance of microstructure noise 2η. $X(t_{i+K}) - X(t_i)$ is heavy-tailed due to the heavy tails of the microstructural noises, and so we employ the Huber loss to mitigate the influence of the heavy tails as follows

$$\hat{\theta}_K = \arg\min_{\theta_K} \sum_{i=1}^{n-K+1} \ell_{\alpha_K}((X(t_{i+K}) - X(t_i))^2/K - \theta_K), \tag{8}$$

where α_K is the truncation parameter and will be given in following theorem and ℓ_α is the Huber loss

$$\ell_\alpha(x) = \begin{cases} 2\alpha|x| - \alpha^2, & |x| > \alpha, \\ x^2, & |x| \le \alpha. \end{cases} \tag{9}$$

We call the $\hat{\theta}_K$ as the robust low frequency volatility estimator. This estimator is the equivalent to the solution of the following equation:

$$\sum_{i=1}^{n-K+1} \psi\{\alpha_K^{-1}((X(t_{i+K}) - X(t_i))^2/K - \theta_K)\} = 0,$$

where the influence function $\psi(x) = \text{sign}(x)\min(1, |x|)$. The truncation parameter α_K trades off between the biases and the heaviness of tails. If we token α_K very large, then the robust estimator would degenerate to the $[X, X]_T^{(K)}$ and bias is reduced.

The robust low-frequency part $\hat{\theta}_K$ is biased due to the microstructure noise. Therefore we need to reduce the influence caused by the microstructure noise. Similar to the idea of TSRV estimator, we consider the high-frequency part. To ease the heavy-tail impact, we apply the Huber loss again to the high-frequency part with truncation parameter α_J:

$$\hat{\theta}_J = \arg\min_{\theta_J} \sum_{i=1}^{n-J+1} \ell_{\alpha_J}((X(t_{i+J}) - X(t_i))^2/J - \theta_J), \tag{10}$$

For the asymptotics, we requires $K = O(n^{2/3})$. In the classical two scales, one can set $J = 1$. Finally, we obtain the robust two-scale volatility estimator by combining the robust low frequency part estimator and high frequency part estimator:

$$\widehat{\langle X, X\rangle}^{(rtsrv)} = \left(1 - \frac{\bar{n}_K}{\bar{n}_J}\right)^{-1} \left(\hat{\theta}_K - \frac{\bar{n}_K}{\bar{n}_J}\hat{\theta}_J\right), \tag{11}$$

We call this robust estimator (9) the robust two-time scale realized volatility (RTSRV) estimator. Unlike the jump robust estimator, proposed by Mykland and Zhang (2016), simple truncated with the sample variance, our estimator is also adapted with the sample size. And this will show that our method has a better outcome with the jump existing and we show it in the simulation study. To investigate the asymptotic properties of the RTSRV estimator, we need some assumptions which are supposed to hold throughout this paper.

Assumption 3.1:

A1. *The drift term in* (1) *is zero, i.e.* $\mu(t) = 0$.
A2. *Without confusion, denote* $\gamma(t)$ *as the spot variance/covariance of* X, Y *or both and microstructure noise are bounded, that is*

$$\max_{t\in[0,T]} \gamma(t) \le \nu_\gamma, \quad \max_{t\in[0,T]} \mathbb{E}(\epsilon^4) \le \nu_\epsilon^2,$$

where ν_γ *and* ν_ϵ *are positive constant.*

Remark 3.1: Assumption A1 is imposed just for simplicity and it is no technical difficulty to obtain the same conclusion without this assumption, due to the Itô formula, the drift term $\mu(t)$ are negligible in higher-order terms as well. Assumption A2 indicates that the observed log-price $X(t_i)$ has only finite fourth moments due to the heavy-tailed microstructure noise. Furthermore, one can also relax the boundedness condition of the instantaneous volatility process $X(t_i)$ such as the locally bounded process (see Aït-Sahalia and Xiu 2017) using the localization argument made in Section 4.4.1 of Jacod and Protter (2012).

We show the RTSRV estimator has the sub-Gaussian concentration in the following theorem.

Theorem 3.1: *If the log-price process follows* (1) *and* (3)*, and suppose* $K = O(n^{2/3})$, $J = O(1)$ *and* $K \log \delta^{-1}/n = o(1)$ *for* $\delta > 0$. *Under the assumption* A1 *and* A2*, take*

$$\alpha_J = \sqrt{\frac{(n-J+1)\bar{V}_J}{J \log \delta^{-1}}}, \quad \alpha_K = \sqrt{\frac{(n-K+1)\bar{V}_K}{K \log \delta^{-1}}},$$

where

$$\bar{V}_J = \frac{1}{J^2} \sum_{k=1}^{n-J+1} \mathrm{var}\{(X(t_{k+J}) - X(t_k))^2\}, \quad \bar{V}_K = \frac{1}{K^2} \sum_{k=1}^{n-K+1} \mathrm{var}\{(X(t_{k+K}) - X(t_k))^2\}.$$

Then for large n, we have:

$$P\left\{ |\hat{\theta}_K - \theta_K| \ge C\sqrt{\frac{(\nu_\gamma + 2\nu_\epsilon/K)^4 K \log \delta^{-1}}{(n-K+1)\bar{V}_K}} + \frac{2K}{n}\nu_\gamma \right\} \le 3\sqrt{\frac{(n-K+1)}{K}}\delta. \quad (12)$$

Furthermore, we have

$$P\left\{ |\hat{\theta}_J - \theta_J| \ge C\sqrt{\frac{(\nu_\gamma + 2\nu_\epsilon/K)^4 J \log \delta^{-1}}{(n-J+1)\bar{V}_J}} + \frac{2J}{n}\nu_\gamma \right\} \le 3\sqrt{\frac{(n-J+1)}{J}}\delta. \quad (13)$$

Therefore, we have

$$P\left\{|\widehat{\langle X,X\rangle}_T^{(rtsrv)} - \langle X^0,X^0\rangle_T| \geq C\sqrt{\frac{(\nu_\gamma + 2\nu_\epsilon/K)^4 K\log\delta^{-1}}{(n-K+1)\bar{V}_K}} + \frac{2K}{n}\nu_\gamma\right\}$$
$$\leq 6\sqrt{\frac{(n-K+1)}{K}}\delta. \tag{14}$$

Remark 3.2: One can estimate the asymptotic variance of $\hat{\theta}_K$ as follows:

$$\hat{\hat{\theta}}_K = \frac{1}{K(n-K+1)}\sum_{k=1}^{n-K+1}(X(t_{k+K}) - X(t_k))^2.$$

then

$$\hat{\hat{V}}_K = \frac{1}{(n-K+1)}\sum_{k=1}^{n-K+1}((X(t_{k+K}) - X(t_k))^2/K - \hat{\hat{\theta}}_K)^2.$$

Similar to above algebraic manipulations, one can obtain the asymptotic variance of $\hat{\theta}_J$

Remark 3.3: More generally, consider the jump diffusion:

$$\mathrm{d}\tilde{X}(t) = \mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_t + \mathrm{d}J(t), \quad t \in [0,T]$$

and if we let the jump term $J(t)$ satisfies following assumption:

A3 $\sum_{t\in[0,T]}|\Delta J(t)| = O_p(1)$

Then we have the same theoretical result of Theorem 3.1 as well as the following result Theorem 5.1. This conclusions are also valid to the following theories results for the multi processes if we consider independent jump and satisfies the assumption A.3.

4. Robust realized covariance estimation

4.1. synchronized time point

If the two log price processes are synchronous, that is, if they were observed at the same time, one can stop here. Unfortunately, unlike low-frequency financial data, which is typically evenly spaced, high-frequency financial data is seldom simultaneous and frequently occurs asynchronously. In most circumstances, high-frequency tractions for two assets that occur at different times are non-synchronous. If the realized covariance is computed directly using the entire sample, a large bias will result. As a result, the transaction data must be synchronized. To address this issue, a variety of approaches have been offered. We presented a new strategy for dealing with asynchronous financial data in this study.

We first define the predetermined tick time sequence as $\{\pi_1,\ldots,\pi_n\}$ in the fixed time interval $[0,T]$, where $0 = \pi_1 < \cdots < \pi_n = T$ and n is the sampling frequency. The real transaction time of assets are denoted as $\{t_{i,j}, i = 1,\ldots,p, j = 1,\ldots,n_i\}$ where p is the number the assets, n_i is the number of the ith asset trade times in the fixed time T. Thus our previous tick time for the ith asset jth synchronized time can be defined as:

$$\tau_{i,j} = \max\{t_{i,j} : t_{i,j} \leq \pi_i, j = 1, \ldots, n_i\}, \quad i = 1 \cdots, p. \tag{15}$$

The sequence $\{\tau_{i,j}\}, i = 1, \ldots, p, j = 1, \ldots n$ are the final synchronized time sequence of p assets. One can obtain the predetermined tick time by following:

$$\pi_j = \max_{0 \leq i \leq p}\{t_{i,j}, \} \tag{16}$$

where $\pi_1 = \max\{t_{1,1}, t_{2,1}, \ldots, t_{p,1}\}$. Now, we give the formal definition of the general previous tick time sequence:

Definition 4.1: A sequence of sample time $\{\tau_1, \ldots, \tau_n\}$ named General Previous Tick Time for a collection of p assets, if

(1) $0 \leq \tau_1 < \cdots < \tau_n = T$;
(2) at least on observation in each asset between consecutive τ_js;
(3) $\sup_i |\tau_{i,j} - \tau_{i,j-1}| \xrightarrow{p} 0$.

Remark 4.1: The general previous time tick scheme become the General Synchronized Scheme provided the assets are heavy trade. Furthermore, if one considers two assets, the general previous tick scheme degenerate to the Previous Tick method, which is proposed by Zhang (2011).

4.2. Review of two time scale covariance estimator

Since the two-time scale realized volatility estimating approach is practical for a single log price process, it inspires us to apply it to two or more log price processes. Zhang (2011) introduced the two-time scale covariance (TSCV) estimation as an extension of the TSRV estimator. The following is a quick overview of the TSCV estimator:

Consider two log price processes $X(\tau_{1,j})$ and $Y(\tau_{2,j})$, synchronized by above synchronous method and follow (1)–(4) with the synchronous time point, denoting n as the total number of paired log price sample. Then the full sample square variation of these two log price processes is

$$[X, Y]_T = \sum_{j=1}^{n}(X(\tau_{1,j+1}) - X(\tau_{1,j}))(Y(\tau_{2,j+1}) - Y(\tau_{2,j})).$$

It's worth noting that $[X, Y]_T$ is often skewed toward zero if $X(\tau_{1,j+1}) - X(\tau_{1,j})$ and $Y(\tau_{2,j+1}) - Y(\tau_{2,j})$ are asynchronous. The well-known Epps effect (Epps 1979) is responsible for this prejudice. Zhang (2011) came up with an analytic formulation for this bias. Zhang has proposed a bias-free alternative approach that eliminates both the Epps effect and the impact of microstructure noise at the same time. It is defined in the same way as in the univariate example.

$$\widehat{\langle X, Y\rangle}_T^{(tscv)} = C_n\left([X, Y]_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}[X, Y]_T^{(J)}\right), \tag{17}$$

where $[X, Y]_T^{(K)}$ is the average lag K realized covariance:

$$[X, Y]_T^{(K)} = \frac{1}{K}\sum_{j=1}^{n-K+1}(X(\tau_{1,j+K}) - X(\tau_{1,j}))(Y(\tau_{2,j+K}) - Y(\tau_{2,j})),$$

the definition of $[X, Y]_T^{(J)}$ is similar to the $[X, Y]_T^{(K)}$ and where $\bar{n}_K = (n - K + 1)/K, \bar{n}_J = (n - J + 1)/J, C_n = 1 + O_p(n^{-1/6})$. One can choose $C_n = n/((K - J)\bar{n}_K)$.

4.3. Robust two scale time realized covariance estimation

To reduce the heavy-tail effect, we apply the Huber loss function to both the $[X, Y]_T^{(K)}$ and $[X, Y]_T^{(J)}$. The whole procedure can be presented in the following steps:

Step 1: obtain the low-frequency part robust estimator:

$$\hat{\Gamma}_K = \arg\min_{\Gamma_K}\sum_{j=1}^{n-K+1}\ell_{\tilde{\alpha}_K}(\tilde{Q}_K(t_j)/K - \Gamma_K), \tag{18}$$

where $\tilde{Q}_K(t_j) = (X(\tau_{1,j+K}) - X(\tau_{1,j}))(Y(\tau_{2,j+K}) - Y(\tau_{2,j}))$.
Step 2: minimize the Huber loss for low-frequency yield:

$$\hat{\Gamma}_J = \arg\min_{\Gamma_J}\sum_{j=1}^{n-J+1}\ell_{\tilde{\alpha}_J}(\tilde{Q}_J(t_j)/J - \Gamma_J), \tag{19}$$

where $\tilde{Q}_J(t_j) = (X(\tau_{1,j+J}) - X(\tau_{1,j}))(Y(\tau_{2,j+J}) - Y(\tau_{2,j}))$. J is constant and satisfies $J = O(1)$.
Step 3: combine above two steps to obtain the final estimator:

$$\widehat{\langle X, Y\rangle}^{(rtscv)} := C_n\left(\hat{\Gamma}_K - \frac{\bar{n}_K}{\bar{n}_J}\hat{\Gamma}_J\right), \tag{20}$$

where C_n is a constant that can be used to tun for small sample bias. And it satisfies $C_n = 1 + O_p(n^{-1/6})$.

Then we have following theorem

Theorem 4.1: *If two log-price processes follow* (1)–(2) *and* (3)–(4), *and suppose* $K = O(n^{2/3})$, $J = O(1)$ *and* $K\log\delta^{-1}/n = o(1)$ *for* $\delta > 0$. *Under the assumptions A1 and A2,* *take*

$$\tilde{\alpha}_J = \sqrt{\frac{(n-J+1)V_J}{J\log\delta^{-1}}}, \quad \tilde{\alpha}_K = \sqrt{\frac{(n-K+1)V_K}{K\log\delta^{-1}}},$$

where

$$V_J = \frac{1}{J^2}\sum_{k=1}^{n-J+1} \mathrm{Cov}\{X(\tau_{1,k+J}) - X(\tau_{1,k})), (Y(\tau_{2,k+J})) - Y(\tau_{2,k})))\},$$

$$V_K = \frac{1}{K^2}\sum_{k=1}^{n-K+1} \mathrm{Cov}\{(X(\tau_{1,k+K}) - X(\tau_{1,k})), (Y(\tau_{2,k+K})) - Y(\tau_{2,k})))\}$$

Then for large n, we have:

$$\mathbb{P}\left\{|\hat{\Gamma}_K - \Gamma_K| \geq C\sqrt{\frac{(\nu_\gamma + 2\nu_\epsilon/K)^4 K\log\delta^{-1}}{(n-K+1)V_K}} + \frac{2K}{n}\nu_\gamma\right\} \leq 3\sqrt{\frac{n-K+1}{K}}\delta. \tag{21}$$

and

$$\mathbb{P}\left\{|\hat{\Gamma}_J - \Gamma_J| \geq C\sqrt{\frac{(\nu_\gamma + 2\nu_\epsilon/J)^4 J\log\delta^{-1}}{(n-J+1)V_J}} + \frac{2J}{n}\nu_\gamma\right\} \leq 3\sqrt{\frac{n-J+1}{J}}\delta. \tag{22}$$

Therefore, we have

$$\mathbb{P}\left\{|\widehat{\langle X, Y\rangle}_T^{(rtscv)} - \langle\tilde{X}, \tilde{Y}\rangle_T| \geq C\sqrt{\frac{(\nu_\gamma + 2\nu_\epsilon/K)^4 K\log\delta^{-1}}{(n-K+1)V_K}} + \frac{2K}{n}\nu_\gamma\right\}.$$
$$\leq 6\sqrt{\frac{n-K+1}{K}}\delta. \tag{23}$$

Remark 4.2: Similar computation in Remark 3.2, we can obtain the asymptotic covariance of the $\hat{\Gamma}_K, \hat{\Gamma}_J$. This theorem shows that the robust volatility estimator $\widehat{\langle X, Y\rangle}_T^{(rtscv)}$ has the sub-Gaussian concentration with the convergence rate is $n^{-1/6}$.

5. Improvement of robust two-time scale estimators

5.1. Review of the MSRV estimator

Although TSRV is the first consistent and asymptotic (mixed) normal estimator of the quadratic variation $\langle X, X\rangle_T$ and even it is robust to the serially dependent noise, it does not achieve the optimal rate. At the cost of higher complexity and computation, it is possible to generalize two-time scale to multiple-time scale, by averaging not two-time scales but on multi-time scales. Simple review of MSRV is given as follows:

$$\widehat{\langle X, X\rangle}_T^{\mathrm{MSRV}} = \sum_{i=1}^{M} a_i[X, X]_T^{K_i} + \frac{1}{n}[X, X]_T, \tag{24}$$

where this estimator require the choice of K_i, without loss of generality we choose $K_i = 1, \ldots, M$, $M = O(n^{\frac{1}{2}})$. The first part in the right side of equation (24) is called slow time scales, the second term is called the fast time scale. Zhang (2006) shown that for suitably selected $a_i's$, $\widehat{\langle X, X\rangle}_T^{\mathrm{MSRV}}$ converges to the true $\langle X, X\rangle_T$ at a rate of $n^{-1/4}$ under the i.i.d. noise assumption. In that paper, Zhang suggests choosing a_i as follows:

$$a_i = \frac{i}{M}h\left(\frac{i}{M}\right) - \frac{1}{2M^2}\frac{i}{M}h'\left(\frac{i}{M}\right),$$

where h is a continuously differentiable real-value function with derivative h', and satisfying the following two conditions:

$$\int_0^1 xh(x)\,\mathrm{d}x = 1 \quad \text{and} \quad \int_0^1 h(x)\,\mathrm{d}x = 0.$$

5.2. Robust MSRV estimator

In this subsection, we extend the MSRV estimator to robust MSRV to ease the heavy-tail and the jump. For the low-time scales, we obtain the robust estimators as follows:

$$\hat{\theta}_{K_i} = \arg\min_{\theta_{K_i}} \sum_{i=1}^{n-K_i+1} \ell_{\alpha_{K_i}}((X(t_{i+K_i}) - X(t_i))^2/K_i - \theta_{K_i}), \quad K_i = 1, \ldots, M. \tag{25}$$

Consider the fast time-scale robust estimator, we need to minimize the following huber loss

$$\hat{\eta} = \arg\min_{\eta} \sum_{i=1}^{n} \ell_{\alpha_\eta}((X(t_{i+1}) - X(t_i))^2/2 - \eta). \tag{26}$$

where the tuning parameters α_{K_i}, $K_i = 1, \ldots, M$ and α_η are given in following theorem. Then we have the final robust MSRV estimator:

$$\widehat{\langle X, X\rangle}_T^{\mathrm{MSRV}} = \sum_{i=1}^{M} a_i\hat{\theta}_{K_i} + \frac{1}{n}\hat{\eta}. \tag{27}$$

Theorem 5.1: *If the log-price process follows* (1) *and* (3), *and suppose* $M = O(n^{1/2})$, *and* $M\log\delta^{-1}/n = o(1)$ *for* $\delta > 0$. *Under the assumptions* A1 *and* A2, *take*

$$\alpha_{K_i} = \sqrt{\frac{(n-K_i+1)\bar{V}_{K_i}}{K_i\log\delta^{-1}}}, \quad K_i = 1, \ldots, M.$$

where $\bar{V}_{K_i} = \frac{1}{K_i^2}\sum_{k=1}^{n-K_i+1}\mathrm{var}\{(X(t_{k+K_i}) - X(t_k))^2\}$. *The definition* $\hat{\theta}_{K_i}$ *is given in* (25). *Then for large n, we have:*

$$\mathbb{P}\left\{|\hat{\theta}_{K_i} - \theta_{K_i}| \geq C\sqrt{\frac{(\nu_\gamma + 2\nu_\epsilon/K_i)^4 K_i\log\delta^{-1}}{(n-K_i+1)\bar{V}_{K_i}}} + \frac{2K_i}{n}\nu_\gamma\right\} \leq 3\sqrt{\frac{(n-K_i+1)}{K_i}}\delta. \tag{28}$$

And if we take

$$\alpha_\eta = \sqrt{\frac{\bar{V}_{\epsilon^*}}{2\log\delta^{-1}}},$$

where $\bar{V}_{\epsilon^*} = \frac{1}{4n}\sum_{i=1}^{n}\mathrm{Var}[X(t_{i+1}) - X(t_i)]$. *Then we have for large n,*

$$\mathbb{P}\left\{|\hat{\eta} - \eta| \geq C\sqrt{\frac{\nu_\epsilon^4 \log\delta^{-1}}{n\bar{V}_{\epsilon^*}}} + \frac{1}{2n}\nu_\gamma\right\} \leq 2\sqrt{n}\delta. \tag{29}$$

Therefore, we have for large n,

$$\mathbb{P}\left\{|\widehat{\langle X, X\rangle}_T^{(\mathrm{RMSRV})} - \langle\tilde{X},\tilde{X}\rangle_T| \geq C\sqrt{\frac{(\nu_\gamma + 2\nu_\epsilon/M)^4 M\log\delta^{-1}}{(n-M+1)\bar{V}_M}} + \frac{2M}{n}\nu_\gamma\right\}$$
$$\leq 4\sqrt{\frac{(n-M+1)}{M}}\delta. \tag{30}$$

5.3. *Robust MSCV estimator*

In this subsection, we extend the MSRV estimator to robust MSRvV to ease the heavy-tail and the jump. We adopt the synchronous method of asynchronous data. For the low-time scales, we obtain the robust estimators as follows:

$$\hat{\Gamma}_{K_i} = \arg\min_{\Gamma_{K_i}} \sum_{i=1}^{n-K_i+1} \ell_{\tilde{\alpha}_{K_i}}((X(\tau_{1,i+K_i}) - X(\tau_{1,i}))(Y(\tau_{2,i+K_i}) - Y(\tau_{2,i}))/K_i - \Gamma_{K_i}), \tag{31}$$

where $K_i = 1,\ldots,M$. The tuning parameters $\tilde{\alpha}_{K_i}, K_i = 1,\ldots,M$ are given in following theorem. Then we have the final robust MSCV estimator:

$$\widehat{\langle X, Y\rangle}_T^{\mathrm{MSCV}} = \sum_{i=1}^{M} a_i\hat{\Gamma}_{K_i}. \tag{32}$$

Theorem 5.2: *If two log-price processes follow* (1)–(2) *and* (3)–(4), *and suppose* $M = O(n^{1/2})$, *and* $M\log\delta^{-1}/n = o(1)$ *for* $\delta > 0$. *Under the assumptions* A1 *and* A2, *take*

$$\alpha_{K_i} = \sqrt{\frac{(n-K_i+1)\bar{V}_{K_i}}{K_i\log\delta^{-1}}}, \quad K_i = 1,\ldots,M.$$

where $\bar{V}_{K_i} = \frac{1}{K_i^2}\sum_{k=1}^{n-K_i+1}\mathrm{Cov}\{X(\tau_{1,i+K_i}) - X(\tau_{1,i}), Y(\tau_{2,i+K_i}) - Y(\tau_{2,i})\}$. *The definition of* $\hat{\theta}_{K_i}$ *is given in* (31). *Then for large n, we have:*

$$\mathbb{P}\left\{|\hat{\Gamma}_{K_i} - \Gamma_{K_i}| \geq C\sqrt{\frac{(\nu_\gamma + 2\nu_\epsilon/K_i)^4 K_i\log\delta^{-1}}{(n-K_i+1)\bar{V}_{K_i}}} + \frac{2K_i}{n}\nu_\gamma\right\} \leq 3\sqrt{\frac{(n-K_i+1)}{K_i}}\delta. \tag{33}$$

Therefore, we have for large n,

$$\mathbb{P}\left\{|\widehat{\langle X, Y\rangle}_T^{(\mathrm{RMSCV})} - \langle\tilde{X},\tilde{Y}\rangle_T| \geq C\sqrt{\frac{(\nu_\gamma + 2\nu_\epsilon/M)^4 M\log\delta^{-1}}{(n-M+1)\bar{V}_M}} + \frac{2M}{n}\nu_\gamma\right\}$$

$$\leq 3\sqrt{\frac{(n-M+1)}{M}}\delta. \tag{34}$$

Remark 5.1: Combining (9) and (10), we can obtain the robust realized correlation of two assets (X, Y):

$$\widehat{\operatorname{Corr}(X,Y)}^{\text{RTSCV}} = \frac{\widehat{\langle X,Y\rangle}_T^{(rtscv)}}{\sqrt{\widehat{\langle X,X\rangle}_T^{(rtsrv)}}\sqrt{\widehat{\langle Y,Y\rangle}_T^{(rtsrv)}}}. \tag{35}$$

and

$$\widehat{\operatorname{Corr}(X,Y)}^{\text{RMSCV}} = \frac{\widehat{\langle X,Y\rangle}_T^{(rmscv)}}{\sqrt{\widehat{\langle X,X\rangle}_T^{(rmsrv)}}\sqrt{\widehat{\langle Y,Y\rangle}_T^{(rmsrv)}}}. \tag{36}$$

Remark 5.2: The normality of the above robust estimators is deferred in the latter work. Also one can consider the performance of the robust estimators when we consider the microstructure noise is serial dependence, that studied by Aït-Sahalia et al. (2011), Jacod, Li, and Zheng (2017), Da and Xiu (2021), and Li and Linton (2022, n.d.). The robust estimators, RMSRV and RMSCV, have a better performance overly at the cost of complexity and computation.

6. Simulation design

In this section, we conduct simulations for both the univariate price process and bivariate log price process and then apply these estimators to real data. For each estimator in the simulation, we compute the bias and RMSE. Denote by RV_i the integrated volatility of day i, as defined in (3), and let $\widehat{\text{RV}}_i$ be the estimator. The bias and root mean squared error is computed as follows:

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{\text{RV}}_i - \text{RV}_i), \quad \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{\text{RV}}_i - \text{RV}_i)^2}. \tag{37}$$

6.1. Univariate price process

The univariate price processes simulations are also based on a time interval of one day, that is $t \in [0,1]$. The log price processes were stochastic volatility model of Huang and Tauchen (2005):

$$\begin{aligned} X_t &= \tilde{X}_t + \epsilon_t, \quad \mathrm{d}\tilde{X}_t = \gamma\sigma_t\,\mathrm{d}W_t + \sqrt{1-\gamma^2}\sigma_t\,\mathrm{d}B_t, \\ \sigma_t &= \exp(\beta_0 + \beta_1 v_t), \quad \mathrm{d}v_t = \kappa v_t\,\mathrm{d}t + \mathrm{d}B_t. \end{aligned} \tag{38}$$

and the jump diffusion model:

$$\begin{aligned} X_t &= \tilde{X}_t + \epsilon_t, \quad \mathrm{d}\tilde{X}_t = \gamma\sigma_t\,\mathrm{d}W_t + \sqrt{1-\gamma^2}\sigma_t\,\mathrm{d}B_t + J_t\,\mathrm{d}N_t, \\ \sigma_t &= \exp(\beta_0 + \beta_1 v_t), \quad \mathrm{d}v_t = \kappa v_t\,\mathrm{d}t + \mathrm{d}B_t. \end{aligned} \tag{39}$$

where B_t, W_t are independent. The continuous part is simulated using the Euler scheme. The discontinuous part is $J_t \sim \exp(z)$, where $z \sim N(-5, 0.8)$, and N_t is a Poisson process independent Brownian motion with intensity λ. As for microstructure noise, we separate it from the Sub-Gaussian and heavy-tail cases. In the sub-Gaussian case, we adapt the noise following the norm distribution that is $\epsilon \sim N(0, 0.0005^2)$. Consider the heavy-tail case, we simulate the noise using t-distribution with the degree of freedom $df = 5$, that is $0.01t(5)$.

We assume that a day with consists of 4 hours of open trading, that is everyday sample path is 14,400. We simulate 500 days. The parameters $(\gamma, \beta_0, \beta_1, \kappa, \lambda)$ are set to $(-0.3, -5/16, 1/8, -1/40, 12)$. And the initial value of v_t for each day drawn from a norm distribution $N(0, -2/\kappa)$. Here, one may absorb the jump term and noise term as the big noise, that is big noise is equal to noise plus the jump. Here we can treat the heavy-tail of noise mainly cause by the jump. This is because the jump is dominate the microstructure noise.

In univariate case, we compute the Bi-power realized volatility estimator (BPRV, Barndorff-Nielsen and Shephard 2004), TSRV, MSRV, jump robust two scale realized volatility estimator (JTSRV), proposed by (Boudt and Zhang 2013). We select the scale $J = 1$, and $K = 300(\text{suggest}), 591(K = [n^{2/3}])$ and choose the tuning parameter $\delta = 1/1.01$. The numeric result shows in Table 1.

From Table 1, we found that these estimators have different performances. First, when the price is without jumps, traditional methods (BPRV and TSRV) have better performance than the truncation estimators (JTSRV and RTSRV). This indicates that our method cannot improve the estimation and this is due to the trading off bias and heaviness. And the performance of RTSRV estimator with heavy-tail noise has slightly better than the sub-Gaussian tail.

Table 1. Comparison of bias and RMSE of integrated variance estimators.

	Sub-Gaussian		Heavy-tail	
Estimator	Bias	RMSE	Bias	RMSE
No jump				
BPRV ($K = 300$)	−0.008	0.122	−0.008	0.123
TSRV	−0.041	0.255	−0.041	0.254
JTSRV	−0.103	0.544	−0.103	0.543
RTSRV	−0.279	0.978	−0.279	0.978
BPRV($K = 591$)	−0.075	0.477	−0.074	0.475
TSRV	−0.066	0.381	−0.067	0.382
JTSRV	−0.111	0.554	−0.110	0.555
RTSRV	−0.279	0.978	−0.279	0.978
Exist jump				
BPRV ($K = 300$)	1.268	1.276	1.268	1.276
TSRV	0.820	0.828	0.821	0.829
JTSRV	0.133	0.154	0.132	0.155
RTSRV	−0.063	0.093	−0.063	0.093
BPRV ($K = 591$)	2.131	2.136	2.132	2.137
TSRV	1.201	1.209	1.201	1.210
JTSRV	0.322	0.155	0.322	0.156
RPRV	0.054	0.088	0.054	0.080
RTSRV	−0.063	0.093	−0.063	0.093
RMSRV	−0.066	0.263	−0.066	0.263

When the price process contains a jump process, that is we consider the heavy-tail caused by the jump, our estimators have less bias and small RMSE, which shows that the RPRV, RTSRV, and RMSRV estimators outperform other methods. The RPRV has less bias but it has larger MSE than RMSRV. And the BPRV and TSRV estimators have large biases and RMSE compared with the no jump case and our estimator. Finally, the scale $K = 300$ with the suggested sample is better performance than the scale $K = 591$ with the convergence order. Although the MRSRV estimator has slightly large bias than the RTSRV as well as large RMSE. This indicates that the RMSRV have more cumulative error. Among other methods, RPRV has best performance in jump diffusion model.

6.2. Bivariate price process

The bivariate price process simulation are based on a time interval of one day, that is $t \in [0, 1]$. Like the univariate case, we still consider the stochastic volatility model and jump diffusion model. We simulate the continuous model as follows:

$$\begin{aligned}
X_t &= \tilde{X}_t + \epsilon_t^x, \quad \mathrm{d}\tilde{X}_t = \gamma_x \sigma_t^x \, \mathrm{d}B_t^x + \sqrt{1-\gamma_x^2}\sigma_t^x \, \mathrm{d}W_t, \\
Y_t &= \tilde{Y}_t + \epsilon_t^y, \quad \mathrm{d}\tilde{Y}_t = \gamma_y \sigma_t^y \, \mathrm{d}B_t^y + \sqrt{1-\gamma_y^2}\sigma_t^y \, \mathrm{d}W_t, \\
\sigma_t^x &= \exp(\beta_0 + \beta_1 v_t^x), \quad \mathrm{d}v_t^x = \kappa v_t^x \, \mathrm{d}t + \mathrm{d}B_t^x, \\
\sigma_t^y &= \exp(\beta_0 + \beta_1 v_t^y), \quad \mathrm{d}v_t^y = \kappa v_t^y \, \mathrm{d}t + \mathrm{d}B_t^y.
\end{aligned} \tag{40}$$

and the jump diffusion model:

$$\begin{aligned}
X_t &= \tilde{X}_t + \epsilon_t^x, \quad \mathrm{d}\tilde{X}_t = \gamma_x \sigma_t^x \, \mathrm{d}B_t^x + \sqrt{1-\gamma_x^2}\sigma_t^x \, \mathrm{d}W_t + J_t^x \, \mathrm{d}N_t^x, \\
Y_t &= \tilde{Y}_t + \epsilon_t^y, \quad d\tilde{Y}_t^y = \gamma_y \sigma_t^y \, \mathrm{d}B_t^y + \sqrt{1-\gamma_y^2}\sigma_t^y \, \mathrm{d}W_t + J_t^y \, \mathrm{d}N_t^y, \\
\sigma_t^x &= \exp(\beta_0 + \beta_1 v_t^x), \quad \mathrm{d}v_t^x = \kappa v_t^x \, \mathrm{d}t + \mathrm{d}B_t^x, \\
\sigma_t^y &= \exp(\beta_0 + \beta_1 v_t^y), \quad \mathrm{d}v_t^y = \kappa v_t^y \, \mathrm{d}t + \mathrm{d}B_t^y.
\end{aligned} \tag{41}$$

where B_t^x, B_t^y, W_t are independent Brownian motion. The continuous part are simulated by using the Euler scheme. The discontinuous part of X is $J_t^x \sim \exp(z)$, where $z \sim N(-5, 0.8)$, and N_t^x is a Poisson process independent Brownian motion with intensity λ_x. For the jump part of Y, $J_t^y \sim exp(s)$, where $s \sim N(-5, 1.2)$ and the Poisson process independent Brownian motion with intensity λ_y.

As for microstructure noise, we separate it from the sub-Gaussian and heavy-tail cases. In the sub-Gaussian case, we adapt the noise follow the norm distribution that is $\epsilon \sim N(0, 0.0005^2)$. We simulate the the heavy-tail noise from the t-distribution, i.e. $0.01t(df)$ with degree of freedom $df = 5$. We assume that a day with consist of 4 hours of open trading, that is every day sample path is 14,400. We simulate 500 days. The parameters $(\gamma_x, \gamma_y, \beta_0, \beta_1, \kappa, \lambda_x, \lambda_y)$ are set to $(-0.3, -0.3, -5/16, 1/8, -1/40, 12, 36)$. And the initial value of v_t^x and v_t^y for each day drawn from a norm distribution $N(0, -2/\kappa)$.

From Table 2, we found that the scale $K = 300$ has better performance than the scale $K = 591$. Compared with another method, without a jump, it can be seen that the performance of our method cannot improve the performance of other methods. This can be

Table 2. Comparison of bias and RMSE of integrated covariance estimators.

	Sub-Gaussian		Heavy-tail	
Estimator	Bias	RMSE	Bias	RMSE
No jump				
BPCV ($K = 300$)	−0.008	0.122	−0.008	0.123
TSCV	0.151	0.151	0.150	0.151
JTSCV	−0.012	0.116	−0.012	0.116
RTSCV	−0.067	0.263	−0.067	0.263
BPCV ($K = 591$)	−0.015	0.133	−0.015	0.133
TSCV	−0.066	0.381	−0.067	0.382
JTSCV	−0.111	0.554	−0.11	0.006
RTSCV	−0.067	0.263	−0.067	0.263
Exist jump				
BPCV ($K = 300$)	1.448	1.452	1.449	1.453
TSCV	0.933	0.936	0.933	0.936
JTSCV	0.998	0.996	0.998	0.996
RTSCV	−0.006	0.006	−0.006	0.006
BPCV ($K = 591$)	2.385	2.387	2.386	2.38
TSCV	1.335	1.339	1.335	1.339
JTSCV	0.322	0.155	0.322	0.156
RPRCV	0.113	0.114	0.113	0.114
RTSCV	−0.005	0.007	−0.005	0.007
RMSCV	−0.005	0.006	−0.005	0.006

explained by the trading off bias and heaviness and the mirocstructure noise in our simulation is very small it was domain by the true price. When the process exists jumps, our methods, RTSCV and RMSCV, outperforms another methods such as the RPRCV. This simulation also indicates that the BPCV and TSCV estimators are not robust with respect to jump.

The reason why RTSCV and RMSCV have a better performance than the RPRCV is that the RPRCV estimator of two different asserts without debias is not optimal as explained in Theorem 1 of Fan and Kim (2018). As pointed out by Zhang (2011), the variation $[\epsilon^X, \epsilon^Y]_T$ has order $[n^{2/3}\mathbb{E}(\epsilon^X)^2\mathbb{E}(\epsilon^Y)^2]^{1/2}$. Therefore one needs to ease the microstructure noise effect, our estimator considers this impact and removes this influence of microstructural noise.

7. Real data analysis

7.1. Realized volatility

In this subsection, we analyze high-frequency assets prices for two assets, IBM and MSFT. We use 1 min trading data over for 4 months. The sample period is from 16 July 2019, to 16 November 2019. Since two price processes are non-synchronous, we first apply the general previous tick method to synchronize two assets.

To check whether it exists jumps, we apply the jump test. There are many methods are proposed to test the existence of jump such as Jing, Kong, and Liu (2012), Jing, Liu, and Kong (2014), Kong, Liu, and Jing (2015) and Barndorff-Nielsen and Shephard (2006). Here we adopt the test method proposed by Barndorff-Nielsen and Shephard (2006). Test results are shown in Tables 3 and 4. Among these weeks of IBM and MSFT, these weeks are significant jump expect the third week in IBM and the 13th week in MSFT.

Table 3. Jump test for the IBM and MSFT.

Time	1st week	2st week	3rd week	4th week	5th week	6th week	7th week
IBM							
Test statistic	70.42	27.214	1.03	33.78	28.97	6.71	31.51
p-value	0	< 0.05	0.300	< 0.05	< 0.05	< 0.05	< 0.05
MSFT							
Test statistic	34.65	8.88	14.1	17.4	58.5	6.6	23.3
p-value	< 0.05	< 0.05	< 0.05	< 0.05	0	< 0.05	< 0.05

Table 4. Table 3 (continued).

Time	8th week	9th week	10th week	11th week	12th week	13th week	14th week
IBM							
Test statistic	202.8	26.37	17.14	6.86	217.05	10.9	2.83
p-value	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	0.005
MSFT							
Test statistic	43.9	14.1	38.2	17.2	80.6	0.08	6.93
p-value	0	< 0.05	0	< 0.05	< 0.05	0.93	< 0.05

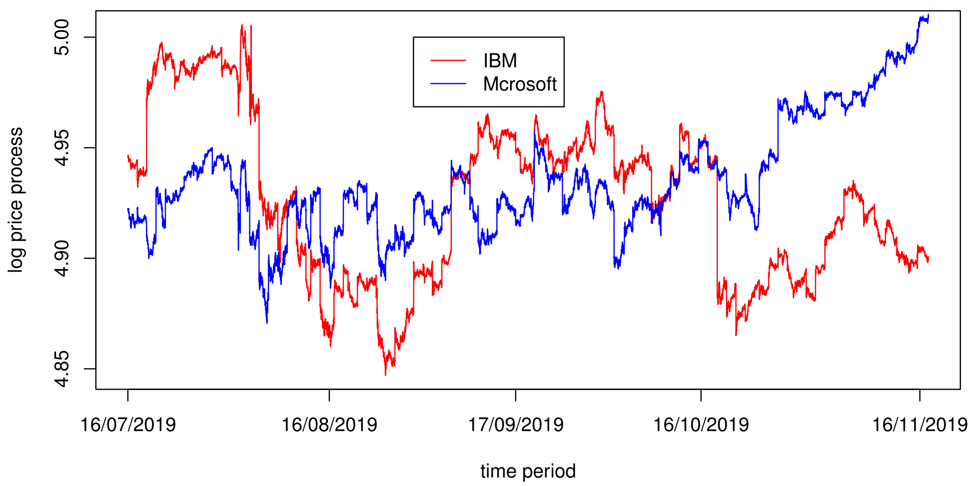

Figure 1. Log price process of IBM and Microsoft.

The log price processes of IBM and MSFT are plotted in Figure 1. Figure 1 indicates that two processes have jumps, which clarify the above test again. The realized volatility, covariance, and correlation of these two assets are shown in Figures 2–5.

As shown in the simulation results, RMSRV has less performance than the RTSRV and RPRV when considering the single price process. Figure 4 shows that the processes are negatively correlated in the first week, the RPRCV estimator always has a positive estimation that is unreasonable and the estimated result is out of this figure. We show the correlation estimated to result in Figure 5, the RTSRV have more flat than other and shows that these two processes have less correlated. It is worth noting that the RPRV estimator always has a positive correlation, which is violated to the process shown in Figure 1. Compared with the TSRV estimator, our RMSRV estimator is more robust.

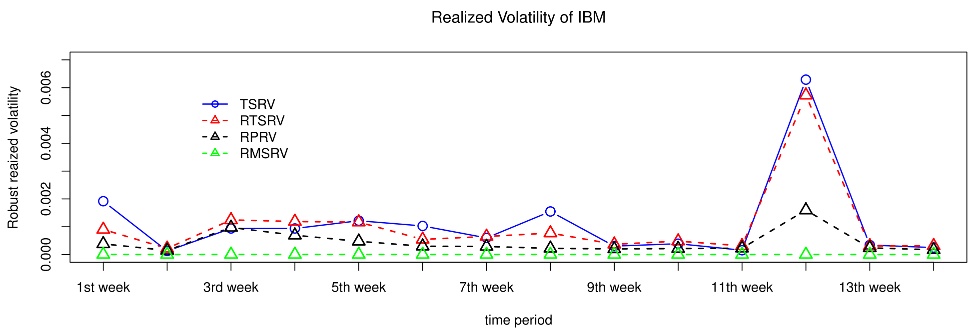

Figure 2. Realised volatility Estimation of IBM.

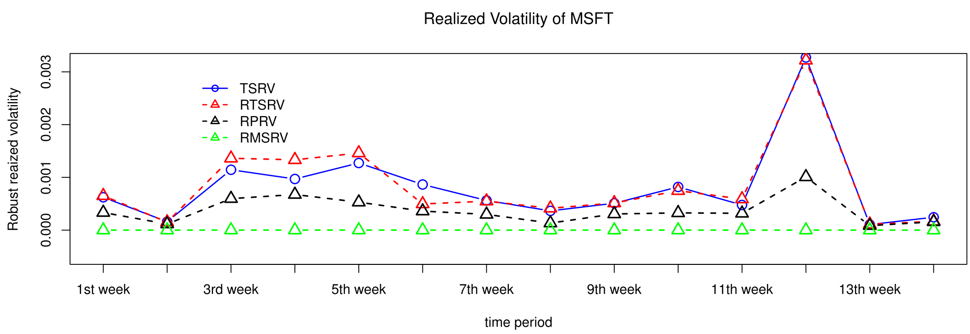

Figure 3. Realised volatility estimation of IBM.

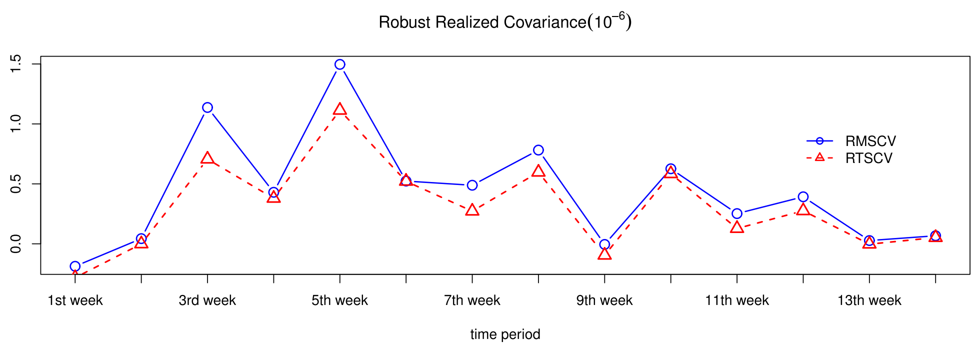

Figure 4. Covariance estimation of IBM and Microsoft.

7.2. Portfolio selection

In this subsection, we analyze 1 min trading data on a period (2 February 2015–32 July 2015) for ten stocks: Apple, Goldman, IBM, Intel, JPMorgan, Coca-Cola, 3M, MSFT, Nike, McDonald.

We here study the behaviour of our realized covariance estimators for the return and risk of ten stocks' portfolios. First, we apply the jump test method to test whether the log price contains the presence of a jump. From Table 5, we can find that there exist jumps for Nike in February, Coca-cola in May, and Intel in June and July. We assume 10 assets stocks

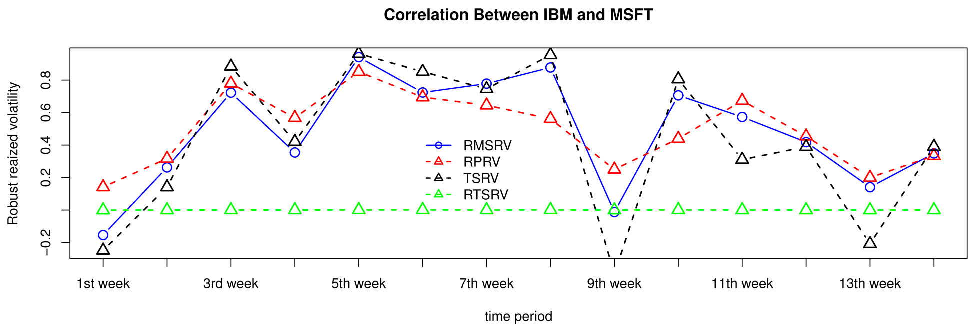

Figure 5. Correlation estimation of IBM and Microsoft.

Table 5. Jump test result of ten stocks.

Date	Feb.	Mar.	Apr.	May.	Jun.	Jul.
Apple	×	×	×	×	×	×
Goldman	×	×	×	×	×	×
IBM	×	×	×	×	×	×
Intel	×	×	×	×	✓	✓
JPMorgan	×	×	×	×	×	×
Coca-Cola	×	×	×	✓	×	×
3M	×	×	×	×	×	×
MSFT	×	×	×	×	×	×
Nike	✓	×	×	×	×	×
McDonald	×	×	×	×	×	×

in our portfolio and must determine the amount of money to invest in each. Let w_i denote the fraction of our budget invested in asset $i = 1, \dots, 10$. The returns $r \in R^{10}$ with known mean $\mathbb{E}[r] = \mu$ and covariance $\mathrm{Var}(r) = \Sigma$. Thus, given a portfolio $w \in R^{10}$, the overall return is $R = r^T w$.

We optimize the portfolio by the trade-off between the expected return $\mathbb{E}[R] = \mu^T w$ and associated risk, which we take as the return variance $Var(R) = w^T \Sigma w$. Initially, we consider only long portfolios, so our problem is

$$\begin{aligned} &\arg\max_{w} \quad \mu^T w - \gamma w^T \Sigma w \\ &\text{subject to} \quad \sum_{i=10}^{10} w_i = 1, \quad w \geq 0,\ i = 1, 2, \dots, 10. \end{aligned}$$

where the objective is the risk-adjusted return and $\gamma > 0$ is a risk aversion parameter. We set the risk aversion parameter $\gamma = 1$ for simplicity. We apply the same convex optimal method to compute the return and risk with different realized covariance estimators. We use the R package **CVXR** to apply above convex optimal method and further details of this compute process refer to the paper Fu, Narasimhan, and Boyd (2020).

Table 6 shows the estimation results of a portfolio under the five covariance estimation methods, namely, RV, PRCV, RPRCV, RTSCV, RMSCV. We don't report the return and risk under the TSCV and MSRCV here due to they have low returns compared with the above five covariance estimators.

Table 6. Return and risk estimations.

Date	Feb.	Mar.	Apr.	May.	Jun.	Jul.
Return	5.244	5.279	5.311	5.356	5.341	5.284
Risk						
RV	3.301×10^{-3}	3.301×10^{-3}	2.417×10^{-3}	1.647×10^{-3}	2.279×10^{-3}	3.148×10^{-3}
PRCV	6.619×10^{-3}	4.199×10^{-3}	3.249×10^{-3}	3.816×10^{-3}	4.716×10^{-3}	16.662×10^{-3}
RPRCV	1.328×10^{-3}	9.878×10^{-4}	7.573×10^{-4}	7.613×10^{-4}	1.123×10^{-3}	4.235×10^{-3}
RTCV	2.527×10^{-3}	1.424×10^{-3}	1.403×10^{-3}	1.358×10^{-3}	2.225×10^{-3}	8.293×10^{-3}
RMCV	5.951×10^{-7}	3.672×10^{-7}	2.793×10^{-7}	3.495×10^{-7}	4.303×10^{-7}	1.937×10^{-6}

We see in Table 6 that the base of the return on the above covariance estimators has the same returns. PRCV method has the highest risk compared with other methods. As for the robust estimators, the RMSCV has the lowest risk, RTSCV has nearly two times risk higher than the RPRCV method.

Acknowledgements

We are very grateful to the editor and two anonymous referees for their very valuable and constructive comments that led to significant improvement of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the National Natural Science Foundation of China [grant numbers 71571007, 71850007].

ORCID

Jiandong Wang http://orcid.org/0000-0002-1119-3851

References

Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010), 'High-Frequency Covariance Estimates With Noisy and Asynchronous Financial Data', *Journal of the American Statistical Association*, 105(492), 1504–1517.

Aït-Sahalia, Y., Mykland, P.A., and Zhang, L. (2011), 'Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise', *Journal of Econometrics*, 160, 160–175.

Aït-Sahalia, Y., and Xiu, D. (2017), 'Using Principal Component Analysis to Estimate a High Dimensional Factor Model with High-frequency Data', *Journal of Econometrics*, 201(2), 384–399.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., and Shephard, N. (2008), 'Multivariate Realised Kernels: Consistent Positive Semi-definite Estimators of the Covariation of Equity Prices with Noise and Non-synchronous Trading', *Journal of Econometrics*, 162(2), 149–169.

Barndorff-Nielsen, O.E., and Shephard, N. (2004), 'Power and Bipower Variation with Stochastic Volatility and Jumps', *Journal of Financial Econometrics*, 2(1), 1–37.

Barndorff-Nielsen, O.E., and Shephard, N. (2006), 'Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation', *Journal of Financial Econometrics*, 4(1), 1–30.

Boudt, K., and Zhang, J. (2013), 'Jump Robust Two Time Scale Covariance Estimation and Realized Volatility Budgets', *Quantitative Finance*, 15(6), 1041–1054.

Catoni, O. (2012), 'Challenging the Empirical Mean and Empirical Variance: A Deviation Study', *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(4), 1148–1185.

Christensen, K., Kinnebrock, S., and Podolskij, M. (2010), 'Pre-averaging Estimators of the Ex-post Covariance Matrix in Noisy Diffusion Models with Non-synchronous Data', *Journal of Econometrics*, 159(1), 116–133.

Clinet, S., and Potiron, Y. (2018), 'Efficient Asymptotic Variance Reduction when Estimating Volatility in High Frequency Data', *Journal of Econometrics*, 206(1), 103–142.

Clinet, S., and Potiron, Y. (2021), 'Disentangling Sources of High Frequency Market Microstructure Noise', *Journal of Business & Economic Statistics*, 39(1), 18–39.

Da, R., and Xiu, D. (2021), 'When Moving-Average Models Meet High-Frequency Data: Uniform Inference on Volatility', *Econometrica*, 89(6), 2787–2825.

Epps, T.W. (1979), 'Comovements in Stock Prices in the Very Short Run', *Journal of the American Statistical Association*, 74(366), 291–298.

Fan, J., and Kim, D. (2018), 'Robust High-Dimensional Volatility Matrix Estimation for High-Frequency Factor Model', *Journal of the American Statistical Association*, 113(523), 1268–1283.

Fan, J., Li, Y., and Ke, Y. (2010), 'Vast Volatility Matrix Estimation Using High-Frequency Data for Portfolio Selection', *Journal of the American Statistical Association*, 107(497), 412–428.

Fan, J., and Wang, Y. (2007), 'Multi-Scale Jump and Volatility Analysis for High-Frequency Financial Data', *Journal of the American Statistical Association*, 102(480), 1349–1362.

Fan, J., Wang, W., and Zhong, Y. (2018), 'An ℓ_∞ Eigenvector Perturbation Bound and Its Application to Robust Covariance Estimation', *Journal of Machine Learning Research*, 18(1), 1–42.

Fan, J., Wang, W., and Zhong, Y. (2019), 'Robust Covariance Estimation for Approximate Factor Models', *Journal of Econometrics*, 208, 5–22.

Fu, A., Narasimhan, B., and Boyd, S. (2020), 'CVXR: An R Package for Disciplined Convex Optimization', *Journal of Statistical Software*, 94(14), 1–34.

Gloter, A., and Jacod, J. (2001), 'Diffusions with Measurement Errors. I. Local Asymptotic Normality', *ESAIM: Probability and Statistics*, 5, 225–242.

Huang, X., and Tauchen, G. (2005), 'The Relative Contribution of Jumps to Total Price Variance', *Journal of Financial Econometrics*, 3(4), 456–499.

Huber, P.J. (1964), 'Robust Estimation of a Location Parameter', *The Annals of Mathematical Statistics*, 35(1), 73–101.

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., and Vetter, M. (2009), 'Microstructure Noise in the Continuous Case: The Pre-averaging Approach', *Stochastic Processes and their Applications*, 119(7), 2249–2276.

Jacod, J., Li, Y., and Zheng, X. (2017), 'Statistical Properties of Microstructure Noise', *Econometrica*, 85(4), 1133–1174.

Jacod, J., and Protter, P. (2012), *Discretization of Processes*, Berlin: Springer.

Jing, B.-Y., Kong, X.-B., and Liu, Z. (2012), 'Modeling High-frequency Financial Data by Pure Jump Processes', *The Annals of Statistics*, 40(2), 759–784.

Jing, B.-Y., Liu, Z., and Kong, X.-B. (2014), 'On the Estimation of Integrated Volatility with Jumps and Microstructure Noise', *Journal of Business & Economic Statistics*, 32(3), 457–467.

Kim, D., Kong, X.-B., Li, C.-X., and Wang, Y. (2018), 'Adaptive Thresholding for Large Volatility Matrix Estimation Based on High-frequency Financial Data', *Journal of Econometrics*, 203(1), 69–79.

Kim, D., Wang, Y., and Zou, J. (2016), 'Asymptotic Theory for Large Volatility Matrix Estimation Based on High-frequency Financial Data', *Stochastic Processes and their Applications*, 126(11), 3527–3577.

Kong, X.-B. (2018), 'On the Systematic and Idiosyncratic Volatility with Large Panel High-frequency Data', *The Annals of Statistics*, 46(3), 1077–1108.

Kong, X.-B., Liu, Z., and Jing, B.-Y. (2015), 'Testing for Pure-jump Processes for High-frequency Data', *The Annals of Statistics*, 43(2), 847–877.

Li, Y., Xie, S., and Zheng, X. (2016), 'Efficient Estimation of Integrated Volatility Incorporating Trading Information', *Journal of Econometrics*, 195(1), 33–50.

Li, Z.M., and Linton, O. (n.d.), 'A ReMeDI for Microstructure Noise,' Social Science Electronic Publishing.

Li, Z.M., and Linton, O. (2022), 'A ReMeDI for Microstructure Noise', *Econometrica*, 90(1), 367–389.

Mykland, P.A., and Zhang, L. (2016), 'Between Data Cleaning and Inference: Pre-averaging and Robust Estimators of the Efficient Price', *Journal of Econometrics*, 194(2), 242–262.

Mykland, P.A., Zhang, L., and Chen, D. (2019), 'The Algebra of Two Scales Estimation, and the S-TSRV: High Frequency Estimation that is Robust to Sampling Times', *Journal of Econometrics*, 208(1), 101–119.

Pelger, M. (2019), 'Large-dimensional Factor Modeling Based on High-frequency Observations', *Journal of Econometrics*, 208(1), 23–42. Special Issue on Financial Engineering and Risk Management.

Shin, M., Kim, D., and Fan, J. (2021), 'Adaptive Robust Large Volatility Matrix Estimation based on High-frequency Financial Data,' Preprint arXiv:2102.12752.

Tao, M., Wang, Y., and Zhou, H.H. (2013), 'Optimal Sparse Volatility Matrix Estimation for High-dimensional Itô Processes with Measurement Errors', *The Annals of Statistics*, 41, 1816–1864.

Wang, Y., and Zou, J. (2010), 'Vast Volatility Matrix Estimation for High-frequency Financial Data', *The Annals of Statistics*, 38(2), 101–119.

Xiu, D. (2010), 'Quasi-maximum Likelihood Estimation of Volatility with High Frequency Data', *Journal of Econometrics*, 159(1), 235–250.

Zhang, L. (2006), 'Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach', *Bernoulli*, 12(6), 1019–1043.

Zhang, L. (2011), 'Estimating Covariation: Epps Effect, Microstructure Noise', *Journal of Econometrics*, 160(1), 33–47.

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(472), 1394–1411.

Appendices

In this section, we give proofs of Theorems 3.1, 4.1, 5.1, and 5.2. We first prove Theorem 4.1 due to Theorem 3.1 is the special case of Theorem 4.1. Then we give the proof of the Theorems 5.1 and 5.2.

Appendix 1. Proof of Theorem 4.1

Proof: We start with the proof of (21) in Theorem 4.1. Without loss generality, we assume that $n = K(L+1) - 1$, for some $L \in N$. Without confusion, let synchronized sample time $\tau_{1,i}, \tau_{2,i}$ be t_k and denote the integrated covariance of two assets $\int_0^T \gamma_{ij}(t)\,\mathrm{d}t := \int_0^T \sigma^x(t)\sigma^y(t)\,\mathrm{d}\langle W_t^x, W_t^y\rangle$. Denote noise as η_{ij}. Define

$$
\begin{aligned}
f(\Gamma_K) &= \frac{\tilde{\alpha}_K}{n-K+1}\sum_{k=0}^{n-K}\psi(\tilde{\alpha}_K^{-1}(\tilde{Q}_K(t_k) - \Gamma_K)),\\
V_K(t_k) &= \mathbb{E}\{(\tilde{Q}_K(t_k) - A_K(t_k))^2 \mid \mathcal{F}_{t_k}\},\\
\bar{A}_K(t_k) &= \frac{1}{n-K+1}\sum_{k=0}^{n-K} A_K(t_k) = \frac{2}{K}\eta_{ij} + \frac{1}{K}\sum_{k=0}^{n-K}\mathbb{E}\left\{\int_{t_k}^{t_{k+K}}\gamma_{ij}(t)\,\mathrm{d}t \mid \mathcal{F}_{t_k}\right\},\\
A_K(t_k) &= \frac{1}{K}\left\{\mathbb{E}\left(\int_{t_k}^{t_{k+K}}\gamma_{ij}(t)\,\mathrm{d}t \mid \mathcal{F}_{t_k}\right) + 2\eta_{ij}\right\},\\
\bar{V}_K(t_k) &= \frac{1}{n-K+1}\sum_{k=0}^{n-K} V_K(t_k),
\end{aligned}
$$

where $\tilde{Q}_K(t_k) = (X(t_{k+K}) - X(t_k))(Y(t_{k+K}) - Y(t_k))/K$. Using the notations above, it is easy to show

$$\begin{aligned}
\tilde{Q}_K(t_k) - A_K(t_k) &= \frac{1}{K}((\tilde{X}(t_{k+K}) - \tilde{X}(t_k))(\tilde{Y}(t_{k+K}) - \tilde{Y}(t_k))) \\
&\quad - \frac{1}{K}\left\{\mathbb{E}\left[\int_{t_k}^{t_{k+K}} \gamma_{ij}(t)\,\mathrm{d}t \mid \mathcal{F}_{t_k}\right]\right\} \\
&\quad + \frac{1}{K}\{(\epsilon^x(t_{k+K}) - \epsilon^x(t_k))(\epsilon^y(t_{k+K}) - \epsilon^y(t_k)) - 2\eta_{ij}\} \\
&\quad + \frac{1}{K}((\tilde{X}(t_{k+K}) - \tilde{X}(t_k))(\epsilon^x(t_{k+K}) - \epsilon^x(t_k)) \\
&\quad + \frac{1}{K}(\tilde{Y}t_{k+K}) - \tilde{Y}(t_k))(\epsilon^y(t_{k+K}) - \epsilon^y(t_k)) \\
&= \nabla_1 + \nabla_2 + \nabla_3 + \nabla_4.
\end{aligned}$$

For the first term ∇_1, by Itô formula, we have

$$\begin{aligned}
\mathbb{E}[\nabla_1|\mathcal{F}_{t_k}] &= \frac{1}{K}\mathbb{E}\left\{(\tilde{X}(t_{k+K}) - \tilde{X}(t_k))(\tilde{Y}(t_{k+K}) - \tilde{Y}(t_k)) - \int_{t_k}^{t_{k+K}} \gamma_{ij}(t)\,\mathrm{d}t \mid \mathcal{F}_{t_k}\right\} \\
&= 0. \quad \text{a.s.}
\end{aligned}$$

For the second term ∇_2, by the mutual independence of the microstructure noise,

$$\mathbb{E}[\nabla_2|\mathcal{F}_{t_k}] = \mathbb{E}\left\{\frac{1}{K}[(\epsilon^x(t_{k+K}) - \epsilon^x(t_k))(\epsilon^y(t_{k+K}) - \epsilon^y(t_k)) - 2\eta_{ij}(t) \mid \mathcal{F}_{t_k}]\right\} = 0. \quad \text{a.s.}$$

For the last two terms, since $\tilde{X}, \tilde{Y}$ are independent of ϵ^x, ϵ^y,

$$\mathbb{E}[\nabla_3 + \nabla_4 \mid \mathcal{F}_{t_k}] = 0. \quad \text{a.s.}$$

Therefore

$$\mathbb{E}[\tilde{Q}_K(t_k) - A_K(t_k) \mid \mathcal{F}_{t_k}] = 0. \quad \text{a.s.} \tag{A1}$$

On the other hand, after some algebra calculations, we can also show that

$$V_K(t_k) = \mathbb{E}[(\tilde{Q}_K(t_k) - A_K(t_k))^2 \mid \mathcal{F}_{t_k}] \le C\left(\nu_\gamma^2 + 4\frac{n}{K}\nu_\gamma\nu_\epsilon + \frac{2n}{K^2}\nu_\epsilon^2\right). \quad \text{a.s.}$$

Similar to Proposition 2.2 of Catoni (2012), we define

$$B_-(\Gamma_K) = \frac{1}{n-K+1}\sum_{k=0}^{n-K}\left\{A_K(t_k) - \Gamma_K - \tilde{\alpha}_K^{-1}[V_K(t_k) + (A_K(t_k) - \Gamma_K)^2] - \frac{\log(1/\delta)}{L\tilde{\alpha}_K^{-1}}\right\},$$

$$B_+(\Gamma_K) = \frac{1}{n-K+1}\sum_{k=0}^{n-K}\left\{A_K(t_k) - \Gamma_K + \tilde{\alpha}_K^{-1}[V_K(t_k) + (A_K(t_k) - \Gamma_K)^2] + \frac{\log(1/\delta)}{L\tilde{\alpha}_K^{-1}}\right\},$$

Denote $b(t_k) = A_K(t_k) - \Gamma_K - \tilde{\alpha}_K^{-1}[V_K(t_k) + (A_K(t_k) - \Gamma_K)^2] - \frac{\log(1/\delta)}{L\tilde{\alpha}_K^{-1}}$, then $B_-(\Gamma_K) = \frac{1}{n-K+1}\sum_{k=0}^{n-K} b(t_k)$. Using the same argument of Proposition 2.1. in Catoni (2012), we have

$$\begin{aligned}
\mathbb{P}(f(\Gamma_K) < B_-(\Gamma_K)) &= \mathbb{P}(\exp\{\tilde{\alpha}_K^{-1}\bar{n}_K B_-(\Gamma_K) < \exp\{\tilde{\alpha}_K^{-1}\bar{n}_K f(\Gamma_K)\}\}) \\
&\le \mathbb{E}(1\{\exp\{\tilde{\alpha}_K^{-1}\bar{n}_K B_-(\Gamma_K) < \exp\{\tilde{\alpha}_K^{-1}\bar{n}_K f(\Gamma_K)\}\}) \\
&\le \mathbb{E}\left(\frac{\exp\{\tilde{\alpha}_K^{-1}\bar{n}_K f(\Gamma_K)\}}{\exp\{\tilde{\alpha}_K^{-1}\bar{n}_K B_-(\Gamma_K)\}}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left(\prod_{m=0}^{K-1}\prod_{k=0}^{L-1}\frac{\exp(K^{-1}\tilde{\alpha}_K^{-1}b(t_{kK+m}))}{\exp(K^{-1}\psi(\tilde{\alpha}_K^{-1}(\tilde{Q}(t_{kK+m})-\Gamma_K)))}\right)\\
&\le \prod_{m=0}^{K-1}\mathbb{E}\left(\prod_{k=0}^{L-2}\frac{\exp(K^{-1}\tilde{\alpha}_K^{-1}b(t_{kK+m}))}{\exp(K^{-1}\psi(\tilde{\alpha}_K^{-1}(\tilde{Q}(t_{kK+m})-\Gamma_K)))}\right)\\
&\quad\cdot\mathbb{E}\left(\frac{\exp(\tilde{\alpha}_K^{-1}b(t_{K(L-1)+m}))}{\exp(\psi(\tilde{\alpha}_K^{-1}(\tilde{Q}(t_{K(L-1)+m})-\Gamma_K)))}|\mathcal{F}_{t_{K(L-1)+m}}\right)^{1/K}\\
&\le \prod_{m=0}^{K-1}\mathbb{E}\left(\prod_{k=0}^{L-2}\frac{\exp(K^{-1}\tilde{\alpha}_K^{-1}b(t_{kK+m}))}{\exp(K^{-1}\psi(\tilde{\alpha}_K^{-1}(\tilde{Q}(t_{kK+m})-\Gamma_K)))}\right)\delta^{1/KL}\\
&\le \delta.
\end{aligned}
$$

where the third inequality is due to the Hölder's inequality and fourth inequality holds by following inequality (A2). Since $-\log(1-x-x^2)\le\psi(x)\le\log(1+x+x^2)$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\exp\{\tilde{\alpha}_K^{-1}b(t_k)\}}{\exp\{\psi(\tilde{\alpha}_K^{-1}(\tilde{Q}_K(t_k)-\Gamma_{K-})\}}\middle|\mathcal{F}_{t_k}\right]\\
&\quad\le \mathbb{E}\left[\frac{\exp\{\tilde{\alpha}_K^{-1}b(t_k)\}}{1-\tilde{\alpha}_K^{-1}(\tilde{Q}_K(t_k)-\Gamma_{K-})-\tilde{\alpha}_K^{-2}(\tilde{Q}_K(t_k)-\Gamma_{K-})^2}\middle|\mathcal{F}_{t_k}\right]\\
&\quad= \frac{\exp\{\tilde{\alpha}_K^{-1}b(t_k)\}}{1-\tilde{\alpha}_K^{-1}(A_K(t_k)-\Gamma_K)-\tilde{\alpha}_K^{-2}\{V_K(t_k)+A_K(t_k)-\Gamma_{K-}\}}\\
&\quad\le \frac{\exp\{\tilde{\alpha}_K^{-1}b(t_k)\}}{\exp\{\tilde{\alpha}_K^{-1}(A_K(t_k)-\Gamma_{K-})-\tilde{\alpha}_K^{-2}(V_K(t_k)+A_K(t_k)-\Gamma_{K-})^2\}}\\
&\quad\le \delta^{1/L}.\quad \text{a.s.}
\end{aligned}\tag{A2}
$$

Similarly, let $B(t_k)=A_K(t_k)-\Gamma_K+\tilde{\alpha}_K^{-1}[V_K(t_k)+(A_K(t_k)-\Gamma_K)^2]+\frac{\log(1/\delta)}{L\tilde{\alpha}_K^{-1}}$. By the same argument we have $P\{f(\Gamma_K<B_-(\Gamma_K))\}\le\delta$. It is easy to show $|A(t_k)|\le\nu_\gamma+\frac{2}{K}\nu_\epsilon$. By (A1) and the bounded of $|A(t_k)|$, we have for large n,

$$
\tilde{\alpha}_K^{-2}\bar{V}_K+\tilde{\alpha}_K^{-2}\{\bar{A}_K^{(2)}-(\bar{A}_K)^2\}+\frac{\log\delta^{-1}}{L}<\frac{1}{4}.
$$

Let Γ_{K+} be the smallest solution of the quadratic equation $B_+(\Gamma_K)=0$, Γ_{K-} be the largest solution of the quadratic equation $B_-(\Gamma_K)=0$. First consider Γ_{K+}, i.e.

$$
\tilde{\alpha}_K^{-1}\Gamma_K^2-\Gamma_K(1+2\tilde{\alpha}_K^{-1}\bar{A}_K)+\bar{A}_K+\tilde{\alpha}_K^{-1}\bar{V}_{\tilde{\alpha},ij}+\tilde{\alpha}_K^{-1}\bar{A}_K^{(2)}+\frac{\log\delta^{-1}}{L\tilde{\alpha}_K^{-1}}=0.
$$

Solving above equation with respective to Γ_K, we obtain

$$
\begin{aligned}
\Gamma_{K+}&=\bar{A}_K+2\left(\tilde{\alpha}_K^{-1}\bar{V}_K+\tilde{\alpha}_K^{-1}\{\bar{A}_K^{(2)}-(\bar{A}_K)^2\}+\frac{\log\delta^{-1}}{L\tilde{\alpha}_K^{-1}}\right)\\
&\quad\times\left(1+\sqrt{1-4\left\{\tilde{\alpha}_K^{-2}\bar{V}_K+\tilde{\alpha}_K^{-2}\{\bar{A}_K^{(2)}-(\bar{A}_K)^2\}+\frac{\log\delta^{-1}}{L}\right\}}\right)^{-1}\\
&\le \bar{A}_K+2\left(\tilde{\alpha}_K^{-1}\bar{V}_K+\tilde{\alpha}_K^{-1}\{\bar{A}_K^{(2)}-(\bar{A}_K)^2\}+\frac{\log 1/\delta}{L\tilde{\alpha}_K^{-1}}\right).
\end{aligned}
$$

Similarly, we have

$$\Gamma_{K_-} \geq \bar{A}_K - 2\left(\tilde{\alpha}_K^{-1}\bar{V}_K + \tilde{\alpha}_K^{-1}\{\bar{A}_K^{(2)} - (\bar{A}_K)^2\} + \frac{\log 1/\delta}{L\tilde{\alpha}_K^{-1}}\right).$$

Above result cannot be directly used due to $\Gamma_{K_\pm}$ are random variables. Therefore we need to deal with the randomness of $\Gamma_{K_\pm}$. For large n, we have

$$-2\left(\nu_\gamma + \frac{2}{K}\nu_\epsilon\right) < \Gamma_{K_-} < \Gamma_{K_+} < 2\left(\nu_\gamma + \frac{2}{K}\nu_\epsilon\right) \quad \text{a.s.}$$

For simplicity, we assume $s = \sqrt{L}$ is an integer value. Define $\mathcal{G} = \{x_0, \ldots, x_s\}$ and $x_0 = -2(\nu_\gamma + \frac{2}{K}\nu_\epsilon)$, $x_s = 2(\nu_\gamma + \frac{2}{K}\nu_\epsilon)$ and $x_{i+1} - x_i = \frac{4}{s}(\nu_\gamma + \frac{2}{K}\nu_\epsilon)$ for $i = 0, \ldots, s-1$. Let

$$\begin{aligned}
E_{1,i} &= \left\{\Gamma_{K_-} - \frac{4}{s}\left(\nu_\gamma + \frac{2}{K}\nu_\epsilon\right) < x_i \leq \Gamma_{K_-}\right\},\\
E_{2,i} &= \{B_-(x_i) \leq f(x_i)\},\\
E_{3,i} &= \left\{\Gamma_{K_+} \leq x_i < \Gamma_{K_+} + \frac{4}{s}\left(\nu_\gamma + \frac{2}{K}\nu_\epsilon\right)\right\},\\
E_{4,i} &= \{f(x_i) \leq B_+(x_i)\}.
\end{aligned}$$

Then we have

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{i=0}^{s}\bigcup_{j=0}^{s}(E_{1,i} \cap E_{2,i} \cap E_{3,j} \cap E_{4,j})\right)\\
&= \mathbb{P}\left\{\left[\bigcup_{i=0}^{s}(E_{1,i} \cap E_{2,i})\right] \cap \left[\bigcup_{j=0}^{s}(E_{3,j} \cap E_{4,j})\right]\right\}\\
&= 1 - \mathbb{P}\left\{\left[\bigcup_{i=0}^{s}(E_{1,i} \cap E_{2,i})\right]^c \cup \left[\bigcup_{j=0}^{s}(E_{3,j} \cap E_{4,j})\right]^c\right\}\\
&\geq 1 - \mathbb{P}\left\{\left[\bigcup_{i=0}^{s}(E_{1,i} \cap E_{2,i})\right]^c\right\} - \mathbb{P}\left\{\left[\bigcup_{j=0}^{s}(E_{3,j} \cap E_{4,j})\right]^c\right\}\\
&\geq 1 - 2(s+1)\delta.
\end{aligned}$$

The last inequality is due to following inequality. Since $E_{1,i}$, $i = 10, \ldots, s$ are mutually exclusive, then we have

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=0}^{s}(E_{1,i} \cap E_{2,i})\right) &= \sum_{i=0}^{s}\mathbb{P}(E_{1,i} \cap E_{2,i})\\
&\geq s + 1 - \sum_{i=0}^{s}(\mathbb{P}(E_{1,i}^c) + \mathbb{P}(E_{2,i}^c)) \geq 1 - (s+1)\delta.
\end{aligned}$$

Similarly, we have $\mathbb{P}(\bigcup_{j=0}^{s}(E_{3,j} \cap E_{4,j})) \geq 1 - 2(s+1)\delta$. Thus with probability greater than $1 - 2(s+1)\delta$, there exist $X_L, X_U \in \mathcal{G}$ s.t.

$$\Gamma_{K_-} - \frac{4}{s}\left(\nu_\gamma + \frac{2}{K}\nu_\epsilon\right) < X_L \leq \Gamma_{K_-}, \quad B_-(X_L) \leq f(X_L),$$

$$B_+ \leq X_U < \Gamma_{K_+} + \frac{4}{s}\left(\nu_\gamma + \frac{2}{K}\nu_\epsilon\right), \quad f(X_U) \leq B_+(X_U).$$

For large n, we have

$$B_+(X_U) \le 0, \text{ a.s.} \quad 0 \le B_-(X_L), \text{ a.s.}$$

On the other hand, since $f(\Gamma_K)$ is non-increasing function with respect to Γ_K, and with probability greater than $1-2(s+1)\delta$

$$\Gamma_{K_-} - \frac{4}{s}\left(\nu_\gamma + \frac{2}{K}\nu_\epsilon\right) < X_L \le \hat{\Gamma}_K \le X_L < \Gamma_{K_+} + \frac{4}{s}(\nu_\gamma + \frac{2}{K}\nu_\epsilon),$$

this implies

$$|\hat{\Gamma}_K - \bar{A}_K| \le 2\left(\tilde{\alpha}_K^{-1}\bar{V}_K + \tilde{\alpha}_K^{-1}\{\bar{A}_K^{(2)} - (\bar{A}_K)^2\} + \frac{\log\delta^{-1}}{L\tilde{\alpha}_K^{-1}}\right) + \frac{4}{s}\left(\nu_\gamma + \frac{2}{K}\nu_\epsilon\right).$$

Now, we turn to investigate $\bar{A}_K$. Since $\sum_{k=0}^{n-K}\{E(\int_{t_k}^{t_{k+K}}\gamma_{ij}(t)\,\mathrm{d}t \mid \mathcal{F}_{t_k}) - \int_{t_k}^{t_{k+K}}\gamma_{ij}(t)\,\mathrm{d}t\}$ is a martingale and $\gamma_{ij}(t)$ is bounded by assumption. It is suffice to show

$$\mathbb{P}\left(\left|\sum_{k=0}^{n-K}\left(\left\{\mathbb{E}\int_{t_k}^{t_{k+K}}\gamma_{ij}(t)\,\mathrm{d}t \mid \mathcal{F}_{t_k}\right\} - \int_{t_k}^{t_{k+K}}\gamma_{ij}(t)\,\mathrm{d}t\right)\right| \ge C\sqrt{\frac{\log\delta^{-1}}{n}}\right) \le \delta.$$

Simple algebraic manipulations show:

$$|\bar{A}_K - \Gamma_K| \le \left|\frac{1}{K}\frac{1}{n-K+1}\sum_{k=0}^{n-K}\left(\mathbb{E}\left\{\int_{t_k}^{t_{k+K}}\gamma_{ij}(t)\,\mathrm{d}t \mid \mathcal{F}_{t_k}\right\} - \int_{t_k}^{t_{k+K}}\gamma_{ij}(t)\,\mathrm{d}t\right)\right| + \frac{2K}{n}\nu_\gamma,$$

These imply that with probability at least $1-2(s+2)\delta$

$$|\hat{\Gamma}_K - \Gamma_K| \le C\sqrt{\frac{(\nu_\gamma + \frac{2}{K}\nu_\epsilon)^4\log\delta^{-1}}{\bar{n}_K\bar{V}_K}} + \frac{2K}{n}\nu_\gamma.$$

Therefore, this establishes the desired result of (21). Next, we define

$$f(\Gamma_J) = \frac{\tilde{\alpha}_J}{n-K+1}\sum_{k=0}^{n-J}\psi(\tilde{\alpha}_J^{-1}(\tilde{Q}_J(t_k) - \Gamma_J)),$$

$$A_J(t_k) = \frac{1}{J}\left\{\mathbb{E}\left(\int_{t_k}^{t_{k+J}}\gamma_{ij}(t)\,\mathrm{d}t \mid \mathcal{F}_{t_k}\right) + 2\eta_{ij}\right\},$$

$$V_J(t_k) = \mathbb{E}\{(\tilde{Q}_J(t_k) - A_J(t_k))^2 \mid \mathcal{F}_{t_k}\},$$

$$\bar{A}_J(t_k) = \frac{1}{n-J+1}\sum_{k=0}^{n-J}A_J(t_k),$$

$$\bar{V}_J(t_k) = \frac{1}{n-J+1}\sum_{k=0}^{n-J}V_J(t_k).$$

where $\tilde{Q}_J(t_k) = \frac{1}{J}(X(t_{k+J}) - X(t_k))(Y(t_{k+J}) - Y(t_k))$. Similar to the proofs of (21), we can obtain the conclusion of (22) by the same argument in (21) with above notations. Finally, Combining (22) and (21) and after some simple algebraic manipulations, the final result of (23) can be proved. ■

Appendix 2. Proof of Theorem 3.1

Proof: Proof of Theorem 3.1 is similar to the proof of Theorem 4.1, the result follows only replace Y by X in Theorem 4.1. If we let $J = 1$ in the Theorem 3.1, we can obtain (29). Detail of this proof can refer to Fan and Kim (2018). ■

Appendix 3. Proof of Theorem 5.1

Proof: Following the same argument in Theorem 3.1 in every step, we can obtain the result of Theorem 5.1. ■

Appendix 4. Proof of Theorem 5.2

Proof: The result of Theorem 5.2 can be obtained by the same argument in the proof of Theorem 4.1. ■