Discussion Papers In Economics And Business

A Test for Dependence and Covariance Estimator of Market Microstructure Noise

Masato Ubukata

Kosuke Oya

Discussion Paper 07-03-Rev.2

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

A Test for Dependence and Covariance Estimator of Market Microstructure Noise

Masato Ubukata

Kosuke Oya

Discussion Paper 07-03-Rev.2

March 2008

この研究は「大学院経済学研究科・経済学部記念事業」
基金より援助を受けた、記して感謝する。

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

A Test for Dependence and Covariance Estimator of Market Microstructure Noise *

Masato Ubukata†

Kosuke Oya‡

Abstract

There are many approaches for estimating an integrated variance and covariance in the presence of market microstructure noise. It is important to know a dependence of noise to construct the integrated variance and covariance estimators. We study a time dependence of bivariate noise processes in this paper. We propose a test statistic for the dependence of the noises and an autocovariance estimator of the noises and derive its asymptotic distribution. The asymptotic distribution of the autocovariance estimator provides us to another test statistic which is for significance of the autocovariances and for detection whether the noise exists or not. We obtain good performances of the test statistics and autocovariance estimator of the noises in a finite sample through Monte Carlo simulation. In empirical illustration, we confirm that the proposed statistics and estimators capture various dependence patterns of the market microstructure noises.

Keywords: test statistic; market microstructure noise; time-dependence; nonsynchronous observations; high frequency data.

JEL classification: C12; D49

---

*Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Special Purposes, 90254135, 2007 and for Scientific Research (A), 18203013, 2007.

†Graduate School of Economics, Osaka University, 560-0043, Machikaneyama, Toyonaka, Osaka, Japan. E-Mail: ubukata@econ.osaka-u.ac.jp

‡Graduate School of Economics, Osaka University, 560-0043, Machikaneyama, Toyonaka, Osaka, Japan. E-Mail: oya@econ.osaka-u.ac.jp

1 Introduction

Estimating covariance matrix of two diffusion processes are important for option pricing, the measurement of value-at-risk, and portfolio evaluation. High frequency financial prices have been used for estimation of integrated variance and covariance. However, we must be careful to the high frequency prices being contaminated with some noise. This noise, known as market microstructure noise, has many sources, including the presence of bid-ask spread and the discreteness of the prices. The realized volatility and covariance are not necessarily the best approach to estimate integrated variance and covariance in the presence of noise. For the literature on the integrated variance estimation with noise, Zhou (1996) proposes the kernel-based estimator. Zhang, Mykland and Aït-Sahalia (2005) propose two scales realized volatility which is the linear combination of realized volatilities at two frequencies. Zhang (2006) extend their estimator to multiple scales. Bandi and Russell (2006) show the optimal frequency based on the minimization of mean squared error. Although these studies are conducted under i.i.d. noise assumption, the market microstructure noise possibly has a time dependence. Under the dependent noise assumption, Aït-Sahalia, Mykland and Zhang (2006) modify the two scales and multiple scales realized volatilities. Hansen and Lunde (2005, 2006) and Barndorff-Nielsen, Hansen, Lunde and Shephard (2006) develop the kernel-based estimator. For the integrated covariance estimation with the dependent noise, Voev and Lunde (2007) show that Hayashi and Yoshida's (2005) cumulative covariance estimator is biased in the presence of cross-correlated noises and propose modified cumulative covariance estimators based on kernel and subsampling methods.

However, we need to know a characteristic of the noises to guarantee the adequacy of these estimators described above because the unbiasedness, consistency and efficiency of them depend on some assumptions for the noises. Whether the market microstructure noise is a time dependent or not and the amount of its dependence could be one of interesting objects on high frequency financial analysis. In this paper, we propose a test statistic to detect the dependence of bivariate market microstructure noise processes and an autocovariance estimator of the noises. Voev and Lunde (2007) propose a test statistic for the cross-sectional dependence of the noises in order to determine kernel bandwidth for their estimator. The main difference from Voev and Lunde's (2007) t-statistic is an evaluation of the variance of the cross-covariance estimator of the noises. Althogh Voev and Lunde (2007) show the unbiasedness of the kernel-based cumulative covariance estimator under the dependent noises, their t-statistic uses an approximated variance of the cross-covariance estimator under the i.i.d. noises and its approximation requires to estimate an integrated variance and a variance of the noise in advance. As denoted in their paper, it is natural that their assumption leads to somewhat larger t-statistic than it should be. On the other hand, we show that the test statistic can be constructed without such i.i.d. noise approximation by using subsampling method and provide an autocovariance estimator of the bivariate noise processes and its asymptotic distribution. Furthermore, we propose a test for significance of the autocovariance, including a test for variance of the noise. We confirm good performances of the proposed test statistic and autocovariance estimator in finite sample through Monte Carlo simulations. For an empirical illustration, we apply these statistics to high frequency asset prices in Osaka Securities Exchange and find that the market microstructure noises in some assets are significantly correlated.

The paper itself proceeds as follows. In section 2 we describe problem settings and background on high frequency financial analysis. We propose the test statistic for cross-sectional dependence of the noises in section 3. In section 4 we provide the autocovariance estimator of the bivariate noise processes, its asymptotic distribution and the test statistic for its significance. Section 5 includes a simulation experiment and an empirical illustration. In section 6 we conclude the paper with appendix that provide proofs about several lemmas and theorems.

2 Problem settings and background

Consider the logarithmic price processes of two assets, $\{P_1^*(t)\}$ and $\{P_2^*(t)\}$ which follow the one-dimensional Itô process with no drift on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$,

$$dP_l^*(t) = \sigma_l(t) dW_l(t), \ \ l = 1, 2, \ \ t \in [0, T], \tag{1}$$

$$d\langle W_1, W_2\rangle_t = \rho^*(t)dt, \quad \rho^*(t) \in (-1, 1),$$

where W_1 and W_2 are standard Brownian motions. The initial value of the price $P_l^*(0)$ is a constant. $\sigma_l(t) > 0$ of $P_l^*(t)$ is a bounded measurable function. The integrated variance and covariation of P_1^* and P_2^* within time horizon $[0, T]$ is given by

$$IV_{l,T} = \int_0^T \sigma_l^2(u)du, \qquad IC_T = \int_0^T \sigma_1(u)\sigma_2(u)\rho^*(u)du, \quad l = 1, 2. \tag{2}$$

We make some notation to represent intraday returns and irregularly nonsynchronous trading system. $r_{1,i} := P_1(t_i) - P_1(t_{i-1})$ is the *i*-th observed intraday return of an asset 1 and $r_{2,j} := P_2(s_j) - P_2(s_{j-1})$ is the *j*-th one of an asset 2. Let t_i and s_j be the end times of the *i*-th and *j*-th intervals of the asset 1 and 2. The different notations for the transaction times in two assets are due to the nonsynchronous trading.

Before showing the test statistic for the dependence and autocovariance estimator of noises, we first review Hayashi and Yoshida's (2005) cumulative covariance (CC) estimator for the nonsynchronously observed process. Figure 1 illustrates a case of $\{(t_{i-1}, t_i] \cap (s_{j-1}, s_j] \neq \emptyset\}$. Then the CC estimator is given by

$$CC = \sum_{i,j} r_{1,i} r_{2,j} 1_{\{(t_{i-1},t_i] \cap (s_{j-1},s_j] \neq \emptyset\}}. \tag{3}$$

The product of $r_{1,i}$ and $r_{2,j}$ contributes to the estimation of the integrated covariance only if intervals $(t_{i-1}, t_i]$ and $(s_{j-1}, s_j]$ are overlapping. The indicator function in (3) enables the CC estimator to use a raw data without any data interpolations. When the observed returns $r_{1,i}$ and $r_{2,j}$ are returns from true price process $r_{1,i}^* = P_1^*(t_i) - P_1^*(t_{i-1})$ and $r_{2,j}^* = P_2^*(s_j) - P_2^*(s_{j-1})$, the CC estimator is unbiased. Its consistency and asymptotic normality are proved in Hayashi and Yoshida (2005, 2006).

However, the integrated volatility and covariance estimators with finer sampling do not converge to the true values in many empirical studies because the transaction prices in high frequency data usually is contaminated with market microstructure noise. It is related to imperfections in the trading system established in the market microstructure literature. A simple way to model high frequency transaction data is to use hidden semi-martingale processes, as named by Mykland and Zhang (2005). In this framework the logarithmic price P_l is observed with market microstructure noises,

$$P_1(t_i) = P_1^*(t_i) + \eta(t_i), \quad P_2(s_j) = P_2^*(s_j) + \delta(s_j), \tag{4}$$

where $P_l^*(t)$ is the true logarithmic price described in (1) which appears in the market with no trading imperfections, frictions, or informational effects. $\eta(t_i)$ and $\delta(s_j)$ are the market microstructure noises in asset 1 and 2, respectively. We assume the properties of the market microstructure noises as follows.

Assumption 1. *The properties of the market microstructure noise.*
Let a vector of market microstructure noise of asset 1 and 2 be $\boldsymbol{u}(t) = (\eta(t)\ \delta(t))'$.

(1a) $\{\boldsymbol{u}(t)\}$ *is a sequence of random variables with zero means.*

(1b) *The bivariate noise processes are covariance stationary with autocovariance function which has finite dependence in the sense that*

$$\Gamma(\ell) \;=\; \mathrm{E}[\boldsymbol{u}(t)\boldsymbol{u}'(t-\ell)] = \begin{pmatrix} \gamma_\eta(\ell) & \gamma_{\eta\delta}(\ell) \\ \gamma_{\delta\eta}(\ell) & \gamma_\delta(\ell) \end{pmatrix} = \mathbf{0}, \quad for\ all \ \mid \ell \mid > m.$$

m *is a finite positive number.* $\Gamma(\ell)$ *is a decreasing function of* $\mid \ell \mid$.

(1c) *There exists some positive number* $\beta > 1$ *satisfies* $\mathrm{E}\left|\boldsymbol{u}(t)\boldsymbol{u}'(s)\right|^{4\beta} < \infty$ *for all* t, s.

(1d) *The noise process is independent with true price process.* $P_l^* \perp\!\!\!\perp \boldsymbol{u}(t),\ l = 1, 2.$

To avoid complication of the subscript for (1b) we rewrite $\gamma_{\eta\delta}(\ell)$ and $\gamma_{\delta\eta}(\ell)$ as $\gamma(\ell)$ and $\gamma(-\ell)$ because $\gamma_{\eta\delta}(\ell) = \mathrm{E}[\eta(t)\delta(t-\ell)]$ and $\gamma_{\delta\eta}(\ell) = \mathrm{E}[\eta(t-\ell)\delta(t)] = \mathrm{E}[\eta(t)\delta(t+\ell)]$. We define the autocorrelation coefficients and the cross-correlation coefficient of two noises as $\rho_\eta(\ell)$, $\rho_\delta(\ell)$ and $\rho(\ell)$. For (1d), even if P_l^* and $\boldsymbol{u}(t)$ are correlated, the dependence of the noises generally dominates the dependence between the true price and noise as the number of high frequency observations increases. Furthermore, Hansen and Lunde (2006) suggest that the independence assumption between the true price and noise does not cause the damage statistically for analysis of asset prices in more trading intensities. We denote conditional expectation and variance given intervals $I^i := (t_{i-1}, t_i]$ and $J^j := (s_{j-1}, s_j]$ for all i, j as $\mathrm{E_{IJ}}[\,\cdot\,]$ and $\mathrm{V_{IJ}}[\,\cdot\,]$. Then the conditional expectation of the CC estimator is

$$\mathrm{E_{IJ}}\Big[\sum_{i,j} r_{1,i}r_{2,j}1_{\{I^i\cap J^j]\neq\emptyset\}}\Big] = \mathrm{E_{IJ}}\Big[\sum_{i,j} r^*_{1,i}r^*_{2,j}1_{\{I^i\cap J^j\neq\emptyset\}}\Big] + \mathrm{E_{IJ}}\Big[\sum_{i,j} e_{\eta,i}e_{\delta,j}1_{\{I^i\cap J^j\neq\emptyset\}}\Big],$$

where $e_{\eta,i} := \eta(t_i) - \eta(t_{i-1})$ and $e_{\delta,j} := \delta(s_j) - \delta(s_{j-1})$. The first term converges to $\int_0^T \sigma_1(u)\sigma_2(u)\rho^*(u)du$ as shown in Hayashi and Yoshida (2005). The second term representing the bias of the CC estimator is further decomposed as

$$\begin{aligned} & \mathrm{E_{IJ}}\Big[\sum_{i,j} e_{\eta,i}e_{\delta,j}1_{\{I^i\cap J^j\neq\emptyset\}}\Big] \\ = \;& \sum_{i,j}\Big(\gamma(t_i - s_j) - \gamma(t_i - s_{j-1}) - \gamma(t_{i-1} - s_j) + \gamma(t_{i-1} - s_{j-1})\Big)1_{\{I^i\cap J^j\neq\emptyset\}}. \end{aligned}$$

It is obvious that the CC estimator is biased when the market microstructure noises are cross-correlated. Voev and Lunde (2007) propose the bias-corrected CC estimator based on kernels. They show that it is unbiased when cross-sectional dependence of the noises is correctly specified. For their estimator, we need to identify the cross-sectional dependence of the noises by using t-statistic proposed in Voev and Lunde (2007). In their t-statistic, they use an approximated variance of cross-covariance estimator which is derived under the i.i.d. and no cross-correlated noises assumptions. Therefore, the variance of cross-covariance estimator is undervalued and it leads to larger t-statistic than it should be. Further their t-statistic includes unknown parameters such that the integrated variance and the variance of the noises we have to estimate. To get rid of the difficulties, we propose alternative test statistic for the cross-sectional dependence of the noises using subsampling method in section 3 and we provide the autocovariance estimator of the bivariate noise processes, its asymptotic distribution and a test statistic for its significance in section 4.

3 The test statistic for cross-sectional dependence of noises

In this section, we propose a test statistic for cross-sectional dependence of noises. Although we cannot identify the covariation of the true price processes which have martingale property and the covariance of the market microstructure noises in the pair of the overlapping intervals, we have

$$\mathrm{E_{IJ}}\Big[\sum_{i,j} r_{1,i} r_{2,j} 1_{\{I^i \cap J^j] \neq \emptyset\}}\Big] = \mathrm{E_{IJ}}\Big[\sum_{i,j} e_{\eta,i} e_{\delta,j} 1_{\{I^i \cap J^j \neq \emptyset\}}\Big], \tag{5}$$

because of the covariation of the true price processes being zero in the pair of the nonoverlapping intervals $\{I^i \cap J^j = \emptyset\}$. We use the product of returns on the nonoverlapping intervals to identify the covariance of the noises if it exists, as well with Voev and Lunde (2007).

3.1 Details of the test statistic

From Assumption (1b), the cross-covariance of two noises disappears when the noises are separated enough. We propose a test statistic to detect the distance where the cross-covariances of the noises become zero in this subsection. For the nonoverlapping intervals $\{I^i \cap J^j = \emptyset\}$ and $t_{i-1} - s_j > 0$, the distance between the intervals is defined as $\ell = t_{i-1} - s_j$. In case of $s_{j-1} - t_i > 0$, the difference of the nonoverlapping intervals is denoted by $\ell = -(s_{j-1} - t_i)$. The top panel (a) and the lower panel (b) in Figure 2 illustrate the former and the latter cases, respectively. The nonoverlapping adjacent intervals such that $t_{i-1} - s_j = 0$ or $s_{j-1} - t_i = 0$ are used in the case of $\ell = 0$. In what follows we consider the case of $\ell > 0$ because we can construct the test statistic for the other cases by replacing the definition of ℓ. We define the product of returns on the i-th and j-th intervals using the indicator function which takes one if $\{I^i \cap J^j = \emptyset\}$ and $\ell = t_{i-1} - s_j > 0$ as follows,

$$Z_{\ell,ij} = r_{1,i} r_{2,j} 1_{\{\{I^i \cap J^j = \emptyset\} \,\cap\, \{t_{i-1} - s_j = \ell\}\}}, \text{ for all } i, j. \tag{6}$$

We take the conditional expectation of $Z_{\ell,ij}$ for any i, j such as the indicator function being one. For any i, j satisfying $\{I^i \cap J^j = \emptyset\}$ and $t_{i-1} - s_j = \ell$, which is a nonoverlapping interval with distance ℓ,

$$\begin{aligned} \mathrm{E}_{\mathrm{IJ}}\big[Z_{\ell,ij}\big] &= \mathrm{E}_{\mathrm{IJ}}\big[\eta(t_i)\delta(s_j)\big] - \mathrm{E}_{\mathrm{IJ}}\big[\eta(t_i)\delta(s_{j-1})\big] - \mathrm{E}_{\mathrm{IJ}}\big[\eta(t_{i-1})\delta(s_j)\big] + \mathrm{E}_{\mathrm{IJ}}\big[\eta(t_{i-1})\delta(s_{j-1})\big] \\ &= \gamma(\ell+\Delta t_i) - \gamma(\ell+\Delta t_i+\Delta s_j) - \gamma(\ell) + \gamma(\ell+\Delta s_j) \end{aligned} \tag{7}$$

For all ℓ taking more than a large enough L such that $\gamma(L) = 0$, we obtain $\gamma(\ell) = 0$ and $\mathrm{E}_{\mathrm{IJ}}\big[Z_{\ell,ij}\big] = 0$ from (1b) in Assumption 1. Now suppose $s^* := \min_s\{s \mid \gamma(L-s) \neq 0, s \geq 0\}$. This implies $\gamma(L) = \gamma(L-1) = \gamma(L-2) = \cdots = \gamma(L-s^*+1) = 0$ and $\gamma(L-s^*) \neq 0$, and $\mathrm{E}_{\mathrm{IJ}}\big[Z_{L,ij}\big] = \mathrm{E}_{\mathrm{IJ}}\big[Z_{L-1,ij}\big] = \mathrm{E}_{\mathrm{IJ}}\big[Z_{L-2,ij}\big] = \cdots = \mathrm{E}_{\mathrm{IJ}}\big[Z_{L-s^*+1,ij}\big] = 0$ and $\mathrm{E}_{\mathrm{IJ}}\big[Z_{L-s^*,ij}\big] \neq 0$. It is obvious that the source of $\mathrm{E}_{\mathrm{IJ}}\big[Z_{L-s^*,ij}\big] \neq 0$ is $\gamma(L-s^*) \neq 0$. Changing the point of view, we set a large enough L such that $\mathrm{E}_{\mathrm{IJ}}\big[Z_{L,ij}\big] = 0$ and find the distance ℓ^* such that $\ell^* = \max_\ell\{\mathrm{E}_{\mathrm{IJ}}\big[Z_{\ell,ij}\big] \neq 0, \ell \leq L\}$. Then we conclude that $\gamma(\ell^*+1) = 0$ and $\gamma(\ell^*) \neq 0$ and that the threshold value of the dependence becomes the distance ℓ^*. It is noted that whether $\mathrm{E}_{\mathrm{IJ}}\big[Z_{\ell,ij}\big] = 0$ or not is not necessarily implies whether $\gamma(\ell) = 0$ or not. This is because the sum of all cross-covariances in (7) incidentally takes zero even in case of $\gamma(\ell) \neq 0$. To avoid such situation we apply the method of determination for ℓ^* described above. The test statistic is constructed by using a sample mean of $Z_{\ell,ij}$ which satisfies nonoverlapping intervals with the distance ℓ to determine the threshold value.

For the construction of the test statistic, we define a sequence $\{Z_{\ell,k}\}_{k=1}^{N_\ell}$ as follows. First we select $Z_{\ell,ij}$ which satisfies $\{I^i \cap J^j = \emptyset\}$ and $\ell = t_{i-1} - s_j$. $\{Z_{\ell,k}\}_{k=1}^{N_\ell}$ is a sequence which arranges the selected $Z_{\ell,ij}$ in ascending order of index i. N_ℓ is the total number of the products of returns on the nonoverlapping intervals with the distance ℓ. We define the k-th pair of the selected intervals as A_k and B_k. Then $Z_{\ell,k}$ is defined as a product of returns on nonoverlapping intervals A_k and B_k. Figure 3 illustrates each pair of intervals (A_k, B_k), (A_{k+1}, B_{k+1}) and (A_{k+2}, B_{k+2}) for $k = 1$. We define $\bar{Z}_{\ell,N_\ell} := \frac{1}{N_\ell}\sum_{k=1}^{N_\ell} Z_{\ell,k}$ as a sample mean of $Z_{\ell,k}$ and make the following assumption for $Z_{\ell,k}$.

Assumption 2. $\mathrm{V}_{\mathrm{IJ}}\big[n^{-1/2}\sum_{k=k'+1}^{k'+n} Z_{\ell,k}\big] \to \sigma^2_{\ell,f}$, *uniformly in any* k', *as* $n \to \infty$. *This means: For any sequence* $\{n_{N_\ell}\}$ *that tends to infinity with* N_ℓ, $\sup_{k'}\big|\mathrm{V}_{\mathrm{IJ}}\big[n_{N_\ell}^{-1/2}\sum_{k=k'+1}^{k'+n_{N_\ell}} Z_{\ell,k}\big] - \sigma^2_{\ell,f}\big| \to 0$ *as* $N_\ell \to \infty$.

This assumption states that the conditional variance of a standardized sample mean of $\{Z_{\ell,k'+1}, \ldots, Z_{\ell,k'+n}\}$ is close to some limiting value as a sample size n goes to infinity. Let $f_{\ell,N_\ell} := \left(\bar{Z}_{\ell,N_\ell} - \mathrm{E}_{\mathrm{IJ}}\left[\bar{Z}_{\ell,N_\ell}\right]\right) N_\ell^{1/2}$ be the theoretical standardization for $\bar{Z}_{\ell,N_\ell}$. Then the asymptotic variance of f_{ℓ,N_ℓ} is given by $\lim_{N_\ell \to \infty} \mathrm{E}_{\mathrm{IJ}}\left[\left(f_{\ell,N_\ell}\right)^2\right] = \sigma^2_{\ell,f}$. It is noted that $\{Z_{\ell,k}\}_{k=1}^{N_\ell}$ is a sequence of dependent and heterogeneously distributed random scalars because the variance depends on the length of the irregularly observed interval and $\{Z_{\ell,k}\}$ is serially correlated. We obtain the following lemma for the asymptotic normality.

Lemma 1. *Suppose Assumptions 1 and 2 hold. As N_ℓ goes to infinity, we have*

$$\frac{f_{\ell,N_\ell}}{\sigma_{\ell,f}} \xrightarrow{a} N(0,1). \tag{8}$$

The proof is given in Appendix. Next we consider the estimation of $\sigma^2_{\ell,f}$. We construct a consistent estimator of $\sigma^2_{\ell,f}$ by applying a subsampling method which is first proposed by Carlstein (1986). Although Carlstein (1986) considers a variance estimation for a general statistic without specifying the dependence in a stationary sequence, Fukuchi (1999) and Politis, Romano and Wolf (1999) extend their results to heteroskedastic observations.

We define the subseries of $\{Z_{\ell,k}\}_{k=1}^{N_\ell}$ as follows.

$$\{Z_{\ell,N_\ell}^{hM_\ell}\} := \left(Z_{\ell,hM_\ell+1}, Z_{\ell,hM_\ell+2}, \cdots, Z_{\ell,(h+1)M_\ell}\right), \quad 0 \le h \le K_\ell - 1, \quad K_\ell = \left[N_\ell / M_\ell\right],$$

where $[\ \cdot\ ]$ denotes the integer part of a real number, M_ℓ is a number of element in subseries $\{Z_{\ell,M_\ell}^{hM_\ell}\}$ and K_ℓ is a total number of subseries. We assume that $M_\ell \to \infty$ and $M_\ell / N_\ell \to 0$ as $N_\ell \to \infty$. The variance estimator is given by

$$\hat{\sigma}^2_{\ell,f} = \frac{M_\ell}{K_\ell} \sum_{h=0}^{K_\ell - 1} \left(\bar{Z}_{\ell,M_\ell}^{hM_\ell} - \frac{1}{K_\ell} \sum_{h=0}^{K_\ell - 1} \bar{Z}_{\ell,M_\ell}^{hM_\ell}\right)^2, \tag{9}$$

where $\bar{Z}_{\ell,M_\ell}^{hM_\ell}$ is a sample mean of subseries $\{Z_{\ell,M_\ell}^{hM_\ell}\}$. We have the following lemma for the variance estimator $\hat{\sigma}^2_{\ell,f}$.

Lemma 2. *Suppose Assumptions 1 and 2 hold. Let M_ℓ be s.t. $M_\ell \to \infty$ and $M_\ell / N_\ell \to 0$. Then we have*

$$\hat{\sigma}^2_{\ell,f} \to_{L_2} \sigma^2_{\ell,f} \quad \textit{as} \;\; N_\ell \to \infty. \tag{10}$$

The proof is described in Appendix.

Our purpose in this section is to construct a test statistic to detect the distance where cross-covariance of the noises becomes zero. The test statistic is derived from the results of Lemmas 1 and 2. Let the null hypothesis be $\mathrm{E}_{\mathrm{IJ}}[Z_{\ell,k}] = 0$ for all k, for given ℓ. The alternative hypothesis consists of all possible deviations from the null. Then we have the following theorem for the test statistic.

Theorem 1. *Suppose Assumptions 1 and 2 hold. As N_ℓ goes to infinity, we have*

$$\tau(\ell) := \frac{\sqrt{N_\ell}\bar{Z}_{\ell,N_\ell}}{\hat{\sigma}_{\ell,f}} \xrightarrow{a} N(0,1) \tag{11}$$

under the null hypothesis. $\tau(\ell)$ diverges under the alternative.

The asymptotic normality of the test statistic follows directly from Lemmas 1 and 2.

The large numbers of M_ℓ and K_ℓ are available for the variance estimation in (9) because the high frequency transaction data yields a large number of N_ℓ. However, it is difficult to determine the optimal numbers of M_ℓ and K_ℓ which minimizes mean squared error of $\hat{\sigma}^2_{\ell,f}$ since we do not know the covariance structure of the noises. The increase of M_ℓ (i.e. the decrease of K_ℓ) in fixed N_ℓ reduces a bias but increases a variance of $\hat{\sigma}^2_{\ell,f}$. It is known that the optimal asymptotic rate of M_ℓ is proportional to $N_\ell^{1/3}$ for the subsampling variance estimation, that is, the asymptotic formula is $M_\ell = cN_\ell^{1/3}$ where c depends on underlying process. We investigate an influence by numbers of M_ℓ and K_ℓ selected under $M_\ell = cN_\ell^{1/3}$ with some c through Monte Carlo simulations in section 5.

Next we summarize a procedure to identify the distance where the cross-covariance of noises becomes zero. First we test the null hypothesis $\mathrm{E}_{\mathrm{IJ}}[Z_{L,k}] = 0$ for all k with a large value L using the test statistic (11). The null hypothesis $\mathrm{E}_{\mathrm{IJ}}[Z_{L,k}] = 0$ would not be rejected because the cross-covariance between sufficiently separated noises is zero from (1b) of Assumption 1. If $\mathrm{E}_{\mathrm{IJ}}[Z_{L,k}] = 0$ is not rejected as expected, secondly we will test whether $\mathrm{E}_{\mathrm{IJ}}[Z_{L-1,k}]$ is zero or not. If $\mathrm{E}_{\mathrm{IJ}}[Z_{L-1,k}] = 0$ is not reject, we proceed to judge the statistical significance for $\mathrm{E}_{\mathrm{IJ}}[Z_{L-2,k}]$. We continue to test sequentially until the null being rejected. Finally we regard the distance where the null is rejected the first time as $\ell^* = \max_\ell\{|\tau(\ell)| > c.v.\}$ where $c.v.$ is a critical value of the test statistic (11).

4 Consistent covariance estimators of noises

4.1 Cross-covariance estimator of noises

We consider an estimation of cross-covariance $\gamma(\ell)$ in this subsection. It is obvious that the sample mean of $Z_{\ell,k}$ does not give an unbiased estimator of $\gamma(\ell)$ for any ℓ being less than a threshold value. For construction of unbiased estimator, we have only to remedy the nonoverlapping intervals so as to all cross-covariances in (7) except $\gamma(\ell)$ become zero by using the threshold value determined through the test statistic (11). Suppose the bivariate noise processes have finite cross-sectional dependence in the sense that $\gamma(\ell) = 0$ for $\ell > m^+ > 0$ and for $-\ell > m^- > 0$. When ℓ is positive, we define $\bar{t}_i^{(+)}$ as the first transaction time of asset 1 which follows t_i subject to $\bar{t}_i^{(+)} - s_j > m^+$ and $\underline{s}_{j-1}^{(+)}$ as the last transaction time of asset 2 which is followed by s_{j-1} subject to $t_{i-1} - \underline{s}_{j-1}^{(+)} > m^+$. As ℓ is negative, we define $\bar{s}_j^{(-)}$ as the first transaction time of asset 2 which follows s_j subject to $\bar{s}_j^{(-)} - t_i > m^-$ and $\underline{t}_{i-1}^{(-)}$ as the last transaction times of asset 1 which is followed by t_{i-1} subject to $s_{j-1} - \underline{t}_{i-1}^{(-)} > m^-$. The returns on the intervals $(t_{i-1}, \bar{t}_i^{(+)}]$ and $(\underline{t}_{i-1}^{(-)}, t_i]$ are denoted by $\bar{r}_{1,i}^{(+)} := P_1\big(\bar{t}_i^{(+)}\big) - P_1(t_{i-1})$ and $\underline{r}_{1,i}^{(-)} := P_1(t_i) - P_1\big(\underline{t}_{i-1}^{(-)}\big)$. For the asset 2, the returns on the intervals $(s_{j-1}, \bar{s}_j^{(-)}]$ and $(\underline{s}_{j-1}^{(+)}, s_j]$ are denoted by $\bar{r}_{2,j}^{(-)} := P_2\big(\bar{s}_j^{(-)}\big) - P_2(s_{j-1})$ and $\underline{r}_{2,j}^{(+)} := P_2(s_j) - P_2\big(\underline{s}_{j-1}^{(+)}\big)$, respectively. Then $Z_{\ell,ij}^{(\pm)}$ which modifies $Z_{\ell,ij}$ in (6) is defined as

$$Z_{\ell,ij}^{(\pm)} = \begin{cases} \bar{r}_{1,i}^{(+)}\ \underline{r}_{2,j}^{(+)} 1_{\{t_{i-1}-s_j=\ell\}} & if \quad \ell > 0 \\ \bar{r}_{1,i}^{(+)}\ \underline{r}_{2,j}^{(+)} 1_{\{t_{i-1}-s_j=0\}} + \underline{r}_{1,i}^{(-)}\ \bar{r}_{2,j}^{(-)} 1_{\{s_{j-1}-t_i=0\}} & if \quad \ell = 0 \\ \underline{r}_{1,i}^{(-)}\ \bar{r}_{2,j}^{(-)} 1_{\{s_{j-1}-t_i=-\ell\}} & if \quad \ell < 0 \end{cases} \tag{12}$$

The top panel (a), the middle panel (b) and the lower panel (c) in Figure 4 illustrate each pair of intervals with $\ell > 0$, $\ell = 0$ and $\ell < 0$, respectively. For all $\ell > 0$, cross-covariances where the distances between the noises are further than m^+ take zero. The nonoverlapping intervals described in (b) and (c) are given by the same idea as (a). The conditional expectation of $Z^{(\pm)}_{\ell,ij}$ is $\mathrm{E}_{\mathrm{IJ}}\big[Z^{(\pm)}_{\ell,ij}\big] = -\gamma(\ell)$. We select $Z^{(\pm)}_{\ell,ij}$ for all i, j such as the indicator function taking one and define a sequence which arranges the selected $Z^{(\pm)}_{\ell,ij}$ in ascending order of index i as $\{Z^{(\pm)}_{\ell,k}\}_{k=1}^{N_\ell}$. Then the modified cross-covariance estimator is given by

$$\hat{\gamma}(\ell) = -\frac{1}{N_\ell}\sum_{k=1}^{N_\ell} Z^{(\pm)}_{\ell,k}. \tag{13}$$

We obtain the followings theorem for the modified cross-covariance estimator $\hat{\gamma}(\ell)$.

Theorem 2. *Suppose Assumptions 1 and 2 hold. Then we have*

$$N_\ell^{1/2}(\hat{\gamma}(\ell) - \gamma(\ell)) \xrightarrow{a} N(0, \omega_\ell^2), \tag{14}$$

where $\omega_\ell^2 = \lim_{N_\ell\to\infty} \mathrm{E}_{\mathrm{IJ}}\Big[\big\{N_\ell^{1/2}(\hat{\gamma}(\ell) - \gamma(\ell))\big\}^2\Big]$.

The proof is described in Appendix. The asymptotic distribution of $\hat{\gamma}(\ell)$ established in Theorem 2 provides us the test statistic for the null hypothesis $\gamma(\ell) = 0$ and the alternative $\gamma(\ell) \neq 0$.

Corollary 1. *The test for cross-covariance of market microstructure noises.*
Let a subsampling estimator of ω_ℓ^2 *be* $\hat{\omega}_\ell^2$. *As* N_ℓ *goes to infinity, we have*

$$\tau^*(\ell) := \frac{\sqrt{N_\ell}\,\hat{\gamma}(\ell)}{\hat{\omega}_\ell} \xrightarrow{a} N(0,1) \tag{15}$$

under the null hypothesis. $\tau^*(\ell)$ *diverges under the alternative.*

It is noted that the bias of CC estimator (3) is virtually zero when the test statistic (15) does not reject the null $\gamma(\ell) = 0$ for all ℓ. Even if the CC estimator is unbiased, the noise has a strong impact on its variance. Therefore, we should use the Voev and Lunde's (2007) subsampling cumulative covariance estimator without kernel method.

4.2 Autocovariance estimator of noise

We propose the cross-covariance estimator of the noises and derive the test statistic for significance of cross-covariance in the previous subsection. The framework is applicable for estimation of an autocovariance of univariate noise process. We will briefly describe a consistent autocovariance estimator of the noise in this subsection. The autocovariance estimator of the noise also requires the distance where the autocovariance becomes virtually zero as the cross-covariance estimator of the noises does. Then we start with the construction of the test statistic to identify its distance. we define the product of returns on the i-th and j-th intervals for asset 1 using the indicator function which takes one if $\ell = t_{j-1} - t_i \geq 0$ as follows,

$$Z_{1,\ell,ij} = r_{1,i} r_{1,j} 1_{\{t_{j-1} - t_i = \ell \geq 0\}}, \quad \text{for all } i, j. \tag{16}$$

Let $\{Z_{1,\ell,k}\}_{k=1}^{N_{1,\ell}}$ be a sequence which arranges $Z_{1,\ell,ij}$ satisfying $\ell = t_{j-1} - t_i \geq 0$ in ascending order of index i. $N_{1,\ell}$ is a total number of a sequence $\{Z_{1,\ell,k}\}$. The null hypothesis is $\mathrm{E}_{\mathrm{IJ}}[Z_{1,\ell,k}]$ $= 0$ for all k, for given ℓ and the alternative hypothesis consists of all possible deviations from the null. The test statistic for this hypothesis is given by

$$\tau_\eta(\ell) := \frac{\sqrt{N_{1,\ell}}\, \bar{Z}_{1,\ell,N_{1,\ell}}}{\hat{\sigma}_{1,\ell,f}}, \tag{17}$$

where $\bar{Z}_{1,\ell,N_{1,\ell}}$ is a sample mean of $\{Z_{1,\ell,k}\}$. $\hat{\sigma}^2_{1,\ell,f}$ is a subsampling estimator of $\sigma^2_{1,\ell,f} = \lim_{N_{1,\ell} \to \infty}$ $\mathrm{E}_{\mathrm{IJ}}\left[\left(f_{1,\ell,N_{1,\ell}}\right)^2\right]$, where $f_{1,\ell,N_{1,\ell}} := \left(\bar{Z}_{1,\ell,N_{1,\ell}} - \mathrm{E}_{\mathrm{IJ}}\left[\bar{Z}_{1,\ell,N_{1,\ell}}\right]\right) N_{1,\ell}^{1/2}$. Under the null we have $\tau_\eta(\ell) \xrightarrow{a}$ $N(0,1)$ as $N_{1,\ell}$ goes to infinity. We define the threshold value of the finite dependence of noise for asset 1 as m_1 in the sense that the autocovariance function $\gamma_\eta(\ell)$ for all $\ell > m_1$ is zero. The test statistic (17) enables to identify the threshold value m_1.

To derive the autocovariance estimator of noise, we construct $Z^{(\pm)}_{1,\ell,ij}$ for all i, j using identified threshold value m_1 as follows,

$$Z^{(\pm)}_{1,\ell,ij} = \underline{r}^{(-)}_{1,i} \overline{r}^{(+)}_{1,j} 1_{\{t_{j-1} - t_i = \ell \geq 0\}} = \Big(P(t_i) - P\big(\underline{t}^{(-)}_{i-1}\big)\Big)\Big(P\big(\overline{t}^{(+)}_j\big) - P(t_{j-1})\Big) 1_{\{t_{j-1} - t_i = \ell \geq 0\}}, \tag{18}$$

where $\bar{t}_j^{(+)}$ is the first transaction time which follows t_j subject to $t_j^{(+)} - t_i > m_1$ and $\underline{t}_{i-1}^{(-)}$ is the last transaction time which is followed by t_{i-1} subject to $t_{j-1} - t_{i-1}^{(-)} > m_1$. Then we have $\mathrm{E_{IJ}}[Z_{1,\ell,ij}^{(\pm)}] = \gamma_\eta(\ell)$ for all ℓ. We select $Z_{1,\ell,ij}^{(\pm)}$ such as the indicator function taking one and define a sequence which arranges the selected $Z_{1,\ell,ij}^{(\pm)}$ in ascending order of index i as $\{Z_{1,\ell,k}^{(\pm)}\}_{k=1}^{N_{1,\ell}}$. The autocovariance estimator of the noise and its asymptotic distribution are given by

$$\hat{\gamma}_\eta(\ell) = -\frac{1}{N_{1,\ell}} \sum_{k=1}^{N_{1,\ell}} Z_{1,\ell,k}^{(\pm)}, \quad N_{1,\ell}^{1/2}\big(\hat{\gamma}_\eta(\ell) - \gamma_\eta(\ell)\big) \xrightarrow{a} N\big(0, \omega_{\eta,\ell}^2\big), \tag{19}$$

where $\omega_{\eta,\ell}^2 = \lim_{N_{1,\ell}\to\infty} \mathrm{E_{IJ}}\left[\left\{N_{1,\ell}^{1/2}(\hat{\gamma}_\eta(\ell) - \gamma_\eta(\ell))\right\}^2\right]$. Once the asymptotic normality of $\hat{\gamma}_\eta(\ell)$ is established, a test statistic to detect whether the autocovariance of the noise is zero or not is defined as follows.

Corollary 2. *The test for autocovariance of market microstructure noise.*

- *Case of $\ell = 0$, that is, test for the variance of the noise.*

  Let the null hypothesis and the alternative be $\sigma_\eta^2 = 0$ and $\sigma_\eta^2 > 0$, and a subsampling estimator of $\omega_{\eta,0}^2$ be $\hat{\omega}_{\eta,0}^2$. As $N_{1,0}$ goes to infinity, the one-sided test statistic for the variance of the noise is

$$\tau_\eta^*(0) = \left(\frac{\sqrt{N_{1,0}}\ \hat{\gamma}_\eta(0)}{\hat{\omega}_{\eta,0}}\right)^2 \xrightarrow{a} \chi(1) \tag{20}$$

  under the null hypothesis. $\tau_\eta^*(0)$ diverges under the alternative.

- *Case of $\ell > 0$.*

  Let the null hypothesis and the alternative be $\gamma_\eta(\ell) = 0$ and $\gamma_\eta(\ell) \neq 0$, and a subsampling estimator of $\omega_{\eta,\ell}^2$ be $\hat{\omega}_{\eta,\ell}^2$. As $N_{1,\ell}$ goes to infinity, the test statistic for the significance of the autocovariance of the noise is

$$\tau_\eta^*(\ell) := \frac{\sqrt{N_{1,\ell}}\ \hat{\gamma}_\eta(\ell)}{\hat{\omega}_{\eta,\ell}} \xrightarrow{a} N(0,1) \tag{21}$$

  under the null hypothesis. $\tau_\eta^*(\ell)$ diverges under the alternative.

5 Monte Carlo simulation and Empirical illustration

5.1 Finite sample properties of autocovariance estimator of noises

We examine the bias and root mean squared error (RMSE) of the autocovariance estimator (13) and (19) of bivariate noise processes using threshold values identified by the test statistic (11) and (17) through Monte Carlo simulation. We use data generation process introduced in Voev and Lunde (2007). The true price processes P_1^* and P_2^* follows the stochastic differential models,

$$\begin{aligned} dP_l^*(t) &= \sigma_l(t)\Big[\sqrt{1-\lambda_l^2}dW_l^{(A)}(t) + \lambda_l dW_l^{(B)}(t)\Big], \\ d\sigma_l^2(t) &= \kappa_l(\theta_l - \sigma_l^2(t))dt + \omega_l\sigma_l^2(t)dW_l^{(B)}(t), \quad l = 1, 2, \end{aligned} \tag{22}$$

where $W_l^{(\cdot)}$ is a standard Brownian motion and $\sigma_l(t)$ follows the GARCH diffusion process. $W_1^{(A)}$ and $W_2^{(A)}$ are correlated, that is, $d\langle W_1^{(A)}, W_2^{(A)}\rangle_t = \rho^*(t)dt$. To generate stochastic correlation $\rho^*(t)$, it is represented by the anti-Fisher transformation,

$$\begin{aligned} \rho^*(t) &= \frac{\exp(2x(t)) - 1}{\exp(2x(t)) + 1}, \\ dx(t) &= \kappa_3(\theta_3 - x(t))dt + \omega_3 x(t)dW(t), \end{aligned}$$

where $x(t)$ follows the GARCH diffusion process. We take $(\lambda_1, \lambda_2) = (0.5, 0.5)$, $(\kappa_1, \kappa_2, \kappa_3) = (0.3, 0.2, 0.1)$, $(\theta_1, \theta_2, \theta_3) = (0.1, 0.1, 0.1)$ and $(\omega_1, \omega_2, \omega_3) = (0.2, 0.3, 0.1)$. The integrated variance and covariance of P_1^* and P_2^* within the time horizon $[0, T]$ is given by

$$IV_{l,T} = \int_0^T \sigma_l^2(u)du, \quad IC_T = \int_0^T \sigma_1(u)\sigma_2(u)\sqrt{(1-\lambda_1^2)(1-\lambda_2^2)}\rho^*(u)du, \quad l = 1, 2.$$

The subsequent simulation settings are the followings.

1. We set trading time per day as 6.5 hours like NYSE and NASDAQ, and the minimum time interval as one second.

2. We generate the noise process which follows a bivariate AR(1) process. To set a variance

of the noise we use a noise-to-signal ratio defined as the variance of the noise divided by integrated variance. Hansen and Lunde (2006) report the noise-to-signal ratios take a range from 0.0002 to 0.006 using transaction data for thirty equities in NYSE and NASDAQ. Then we set the variance of noise such that noise-to-signal ratios of asset 1 and 2 averagely take 0.005, 0.002, respectively. The average observed time intervals of asset 1 and 2 are chosen as 10 and 5 seconds from Poisson observation arrivals.

3. We identify the distance where the covariance of the noises becomes virtually zero through the test statistic (11) and (17), and estimate the autocovariance estimator of the bivariate noise processes (13) and (19).

On the cross-covariance of the noises, we evaluate performances of the following cross-covariance estimators.

- a cross-covariance estimator constructed by using $Z_{\ell,k}$, that is, $\tilde{\gamma}(\ell) := -\frac{1}{N_\ell}\sum_{k=1}^{N_\ell} Z_{\ell,k}$.
- the cross-covariance estimator $\hat{\gamma}(\ell)$ in (13). To determine M_ℓ (number of element in subseries) and K_ℓ (total number of subseries) on the variance estimation in the test statistic (11), we use the asymptotic formula $M_\ell = cN_\ell^{1/3}$. We denote the cross-covariance estimator as $\hat{\gamma}(\ell)_{[c]}$. We set $c = 4, 2, 1$ and 0.5 to investigate an influence by M_ℓ and K_ℓ.
- the cross-covariance estimator (13) through Voev and Lunde's (2007) t-statistic: $\hat{\gamma}(\ell)_{VL}$. We use their t-statistic to identify the distance where cross-covariance becomes virtually zero.

Table 1 summarizes the sample bias and RMSE of the cross-covariance estimators $\tilde{\gamma}(\ell), \hat{\gamma}(\ell)_{[1]}$ and $\hat{\gamma}(\ell)_{VL}$, where the number of repetition is one thousand. The second and third columns represent true cross-covariance $\gamma(\ell)$ and correlation coefficient $\rho(\ell)$ with each value of ℓ. $\tilde{\gamma}(\ell)$ is more biased as ℓ approaches to zero as expected. The biases of $\hat{\gamma}(\ell)_{[c=1]}$ and $\hat{\gamma}(\ell)_{VL}$ are virtually zero and $\hat{\gamma}(\ell)_{[1]}$ has the smaller RMSE than $\hat{\gamma}(\ell)_{VL}$ for all ℓ. The smaller RMSE of $\hat{\gamma}(\ell)_{[1]}$ comes from the shorter intervals for construction of $\hat{\gamma}(\ell)_{[1]}$ relative to that for $\hat{\gamma}(\ell)_{VL}$.

In the next simulation experiment, we investigate the influence of M_ℓ and K_ℓ selected at fixed N_ℓ under $M_\ell = cN_\ell^{1/3}$ on the variance estimation in the test statistic (11). Although we should determine the value of c in accordance with the dependence of the noises, we do not know about the dependence in advance. Even in our simulation settings, it is not easy to select the optimal value of c. In general, when the noise processes have more dependence, the value of c will be large. We conduct simulations under several values of c such that $c = 4, 2, 1$ and 0.5. In this simulation settings, we obtain averagely 450 and 900 realizations of N_ℓ with $\ell > 0$ and $\ell = 0$, respectively. Then the average numbers of M_ℓ and K_ℓ with some c is (c, M_ℓ, K_ℓ) $=(4, 30, 15)$, $(2, 15, 30)$, $(1, 8, 45)$ and $(0.5, 4, 110)$ in case of $\ell > 0$. Table 2 represents RMSE ratios of $\hat{\gamma}(\ell)_{[4]}$, $\hat{\gamma}(\ell)_{[2]}$, $\hat{\gamma}(\ell)_{[0.5]}$ and $\hat{\gamma}(\ell)_{VL}$ to $\hat{\gamma}(\ell)_{[1]}$. Figure 5 plots the RMSE ratios in Table 2. $\hat{\gamma}(\ell)_{VL}$ uniformly takes RMSE about 1.3 times as larger as the cross-covariance estimators with various c. The results show that the cross-covariance estimator is robust for selection of c.

Table 3 summarizes the bias and RMSE of the autocovariance estimator (19) of univariate noise through the test statistic (17) in asset 2. Figure 6 plots the bias and RMSE ratios of $\hat{\gamma}_\delta(\ell)_{[4]}$, $\hat{\gamma}_\delta(\ell)_{[2]}$ and $\hat{\gamma}_\delta(\ell)_{[0.5]}$ to $\hat{\gamma}_\delta(\ell)_{[1]}$. Although we will see some bias as c takes a smaller value, these biases are enough close to zero and RMSE is the almost same value for all c. It is noted that RMSE of $\gamma_\delta(\ell)$ for $\ell = 1, 2$ takes somewhat larger values than that of $\gamma_\delta(\ell)$ for the other ℓ as in Table 3. This is caused by smaller numbers of $N_{2,\ell}$ for $\ell = 1, 2$. In this experiment, $N_{2,1}$ and $N_{2,2}$ are about $1/6$, $1/2$ of $N_{2,\ell}$ for the other ℓ because the average observed time interval of asset 2 is set as 5 seconds.

So far, we have assumed that the true price follows Itô process with no drift as in (1). Now we consider the case with non-zero drift. The product of the noises dominates that of the drift term as the length of interval shrinks to zero. However, it is not clear about whether an influence of the drift is negligible or not since the estimators and test statistics proposed in this paper are based on the expanded intervals as in (12) and (18).

In this simulation, we add the drift term to the true price process,

$$dP_l^*(t) \quad = \quad \mu_l dt + \sigma_l(t)\Big[\sqrt{1-\lambda_l^2}\, dW_l^{(A)}(t) + \lambda_l dW_l^{(B)}(t)\Big], \;\; l = 1, 2, \qquad (23)$$

where μ_l is a constant and (μ_1, μ_2) is set as $(0.05, 0.05)$, $(0.1, 0.1)$ and $(0.5, 0.5)$. We estimate $\hat{\gamma}(\ell)_{[1]}$ and $\hat{\gamma}_\delta(\ell)_{[1]}$ under the above settings. Table 4 represents the biases and RMSE ratios of these estimates to $\hat{\gamma}(\ell)_{[1]}$ and $\hat{\gamma}_\delta(\ell)_{[1]}$ which are estimated using the data generation process with zero drift as in (22). The biases of the estimators are quite small and RMSE ratios take around one even in case of $(\mu_1, \mu_2) = (0.5, 0.5)$ which represents a rapid trend acceleration. We find that the autocovariance estimator of the noises are not much effected by the presence of the non-zero drift term.

5.2 Empirical illustration

We have established the test statistics and the autocovariance estimator of noises in section 3 and 4. In this subsection we apply these statistics to high-frequency transaction prices of three stocks in Osaka Securities Exchange; OMRON Corporation (OC), Murata Manufacturing Co., Ltd. (MM) and Nintendo Co., Ltd. (NI). We investigate whether the noise exists or not, estimate autocovariances of the noises and judge their significance if the noise exists.

First we test whether the noise exists or not using the test statistic (20) which requires the variance estimator (19). To identify m_1, we start with testing whether $\mathrm{E}_{\mathrm{IJ}}[Z_{1,L,k}] = 0$ or $\mathrm{E}_{\mathrm{IJ}}[Z_{1,L,k}] \neq 0$ for all k with a large value L through the test statistic (17). In this empirical illustration, we set $L = 60$. When the total number of $\{Z_{1,\ell,k}\}$ for each ℓ is not large enough for empirical analysis, we use multiple days to obtain a large sample size. In case of D days high-frequency data, we define $n_\ell^{(d)}$ as the total number of $\{Z_{1,\ell,k}\}$ in d-th day where $d = \{1, \cdots, D\}$. We set $D = 83$ for the illustration and the sample period is from March 1, 2007 until June 29, 2007 for 83 trading days. On the first day we obtain $\{Z_{1,60,1}, \cdots, Z_{1,60,n_{60}^{(1)}}\}$ and $\{Z_{1,60,n^{(1)}+1}, \cdots, Z_{1,60,n_{60}^{(1)}+n_{60}^{(2)}}\}$ is sampled on the second day. The total number of $\{Z_{1,60,k}\}$ in the sample period is $N_{1,60} = \sum_{d=1}^{D} n_{60}^{(d)}$. If the null $\mathrm{E}_{\mathrm{IJ}}[Z_{1,60,k}] = 0$ is not rejected, we will test whether $\mathrm{E}_{\mathrm{IJ}}[Z_{1,59,k}]$ is zero or not. We continue to test sequentially until the null being rejected. Finally we regard the distance where the null is rejected the first time as m_1. Then we estimate the variance of noise $\gamma_\eta(0)$ by (19) and test the null hypothesis $\sigma_\eta^2 = 0$ using (20). If the test statistic (20) rejects the null, we estimate the autocovariance of the noise $\gamma_\eta(\ell)$ by the estimator (19) and check its significance through the test statistic (21).

Table 5 (a) shows the test statistic (20) for the variance of noise at 5% significance level where the critical value is 3.84. We confirm that market microstructure noise in each asset exists because the variance of noise is significantly larger than zero. The variances of the noise in OC, MM and NI are estimated as 1.369×10^{-6}, 1.296×10^{-7} and 2.209×10^{-7}. Table 5 (b) shows the test statistic (21) for the significance of the autocovariance where the critical value is 1.96 at 5% significance level. The dependence of the noise differs for each asset, that is, the noise in OC is uncorrelated and the dependences in MM and NI disappear over 17 and 8 seconds. Figure 6 plots three autocorrelation functions $\hat{\rho}_\eta(\ell) = \hat{\gamma}_\eta(\ell)/\hat{\gamma}_\eta(0)$. We find that the microstructure noise in each asset except OC is negatively autocorrelated.

Next we estimate a cross-covariance of the noises for two assets. We use the test statistic (11) and the cross-covariance estimator (13) and judge the significance of the cross-covariance through the test statistic (15), similarly as the univariate noise case. The results are reported in Table 6. In OC-MM and MM-NI, the cross-covariances are significantly different from zero within $-8 \leq \ell \leq 15$ and $-3 \leq \ell \leq 7$ at 5% significance level. Figure 7 plots three cross-correlation functions $\hat{\rho}(\ell) = \hat{\gamma}(\ell)/\sqrt{\hat{\gamma}_\eta(0)\hat{\gamma}_\delta(0)}$. The negative and asymmetric cross-correlations are found in OC-MM and MM-NI. These empirical illustration shows that the market microstructure noise for each asset a various dependence patterns. In this way, the proposed test statistics and autocovariance estimator of bivariate noise processes gives us an insightful analysis of market microstructure noise.

6 Concluding Remarks

Numerous studies have proposed integrated variance and covariance estimators in the presence of market microstructure noise because realized variance, realized covariance and cumulative covariance estimators deteriorate with the noise. It is important to know a dependence of noise to construct a proper integrated variance and covariance estimator suited for the noise properties. In this paper we propose a test statistic to detect a distance where the autocovariance of bivariate noise processes becomes virtually zero for a consistent autocovariance estimator of the noises. Furthermore we show the asymptotic distribution of the autocovariance estimator and propose an another test statistic for its significance. We find that the test statistic and autocovariance estimator have good performance through Monte Carlo simulations. The empirical illustration confirms that the proposed statistics enable to capture various dependence patterns of the noises in several assets. The statistical analysis for the market microstructure noise will give us some evidence about the influences of the market regularity and trading mechanism to the asset pricing in the financial market. The proposed method will shed light on the market microstructure.

Appendix

The proof of Lemma 1.

Let the start times and end times of intervals A_k and B_k be $\underline{A_k}$, $\underline{B_k}$ and $\overline{A_k}$, $\overline{B_k}$, that is, $A_k = (\underline{A_k}, \overline{A_k}]$ and $B_k = (\underline{B_k}, \overline{B_k}]$. We consider the dependence between $Z_{\ell,k}$ and $Z_{\ell,k+h}$ for any h such that $\underline{B_{k+h}} - \overline{A_k} \geq 0$. It is obvious that $Z_{\ell,k}$ has finite dependence from (1b) and (1d) in Assumption 1. Although there are some central limit theorem for finite dependence such that Hoeffding and Robbins (1948) and Serfling (1968), we apply the results given by Theorem 3.1 in Politis, Romano and Wolf (1997) because our studies are applicable to the more general dependence cases such as the mixing sequence. Assumption 2 implies the conditional variance of a standardized sample mean of $\{Z_{\ell,k'+1}, \ldots, Z_{\ell,k'+n}\}$ for any k' approaches to the limiting value $\sigma^2_{\ell,f}$. The condition for the strong mixing coefficient in Theorem 3.1 of Politis, Romano and Wolf (1997) is satisfied from (1b) in Assumption 1. Therefore, it suffices to show the following condition for the application of their central limit theorem.

(C1) $\mathrm{E}_{\mathrm{IJ}}|Z_{\ell,k}|^{2\beta} < \infty$, for some $\beta > 1$.

Let $\Delta\eta(A_k) := \eta(\overline{A_k}) - \eta(\underline{A_k})$ and $\Delta\delta(B_k) := \delta(\overline{B_k}) - \delta(\underline{B_k})$ be the differences between the noises on each interval A_k and B_k. $Z_{\ell,k}$ is decomposed as

$$\begin{aligned} Z_{\ell,k} &= \big(P_1(\overline{A_k}) - P_1(\underline{A_k})\big)\big(P_2(\overline{B_k}) - P_2(\underline{B_k})\big) = \int_{A_k} \sigma_1(u)dW_1(u) \int_{B_k} \sigma_2(u)dW_2(u) \\ &+ \int_{A_k} \sigma_1(u)dW_1(u)\Delta\delta(B_k) + \int_{B_k} \sigma_2(u)dW_2(u)\Delta\eta(A_k) + \Delta\eta(A_k)\Delta\delta(B_k). \end{aligned} \quad (24)$$

We take $\sigma_1(t), \sigma_2(t) < C$ where C is a constant because $\sigma_1(t)$ and $\sigma_2(t)$ are bounded. On each interval A_k and B_k,

$$\mathrm{E}_{\mathrm{IJ}}\Big|\int_{A_k}\sigma_1(u)dW_1(u)\Big|^{2\beta} < \big|C\big|^{2\beta}\mathrm{E}_{\mathrm{IJ}}\Big|\int_{A_k}dW_1(u)\Big|^{2\beta} < \infty \text{ and } \mathrm{E}_{\mathrm{IJ}}\Big|\int_{B_k}\sigma_2(u)dW_2(u)\Big|^{2\beta} < \infty.$$

Since A_k and B_k are nonoverlapping, the high-order absolute moment of the first term in (24)

$$\mathrm{E}_{\mathrm{IJ}}\Big|\textstyle\int_{A_k}\sigma_1(u)dW_1(u)\int_{B_k}\sigma_2(u)dW_2(u)\Big|^{2\beta} = \mathrm{E}_{\mathrm{IJ}}\Big|\int_{A_k}\sigma_1(u)dW_1(u)\Big|^{2\beta}\mathrm{E}_{\mathrm{IJ}}\Big|\int_{B_k}\sigma_2(u)dW_2(u)\Big|^{2\beta}$$

is bounded. From (1c) in Assumption 1 and Minkowski's inequality,

$$\mathrm{E}_{\mathrm{IJ}}\big|\Delta\delta(B_k)\big|^{2\beta} \leq \big(\parallel\delta(\overline{B_k})\parallel_{2\beta} + \parallel\delta(\underline{B_k})\parallel_{2\beta}\big)^{2\beta} < \infty,$$

where $\parallel\delta(\cdot)\parallel_{2\beta} = \big(\mathrm{E}_{\mathrm{IJ}}\big|\delta(\cdot)\big|^{2\beta}\big)^{\frac{1}{2\beta}}$. From (1d) in Assumption 1 the high-order absolute moment of the second term in (24) has

$$\mathrm{E}_{\mathrm{IJ}}\Big|\int_{A_k}\sigma_1(u)dW_1(u)\Delta\delta(B_k)\Big|^{2\beta} = \mathrm{E}_{\mathrm{IJ}}\Big|\int_{A_k}\sigma_1(u)dW_1(u)\Big|^{2\beta}\mathrm{E}_{\mathrm{IJ}}\Big|\Delta\delta(B_k)\Big|^{2\beta} < \infty.$$

For the third term of (24), $\mathrm{E}_{\mathrm{IJ}}\Big|\int_{B_k}\sigma_2(u)dW_2(u)\Delta\eta(A_k)\Big|^{2\beta} < \infty$. From (1c) in Assumption 1 the high-order absolute moment of the fourth term of (24) has

$$\begin{aligned}
&\mathrm{E}_{\mathrm{IJ}}\big|\Delta\eta(A_k)\Delta\delta(B_k)\big|^{2\beta}\\
&= \mathrm{E}_{\mathrm{IJ}}\big|\eta(\overline{A_k})\delta(\overline{B_k}) - \eta(\overline{A_k})\delta(\underline{B_k}) - \eta(\underline{A_k})\delta(\overline{B_k}) + \eta(\underline{A_k})\delta(\underline{B_k})\big|^{2\beta}\\
&\leq \Big(\parallel\eta(\overline{A_k})\delta(\overline{B_k})\parallel_{2\beta} + \parallel\eta(\overline{A_k})\delta(\underline{B_k})\parallel_{2\beta} + \parallel\eta(\underline{A_k})\delta(\overline{B_k})\parallel_{2\beta} + \parallel\eta(\underline{A_k})\delta(\underline{B_k})\parallel_{2\beta}\Big)^{2\beta}\\
&< \infty.
\end{aligned}$$

Finally, we have

$$\mathrm{E}_{\mathrm{IJ}}\big|Z_{\ell,k}\big|^{2\beta} \leq \Big(\parallel\textstyle\int_{A_k}\sigma_1(u)dW_1(u)\int_{B_k}\sigma_2(u)dW_2(u)\parallel_{2\beta} + \parallel\int_{A_k}\sigma_1(u)dW_1(u)\Delta\delta(B_k)\parallel_{2\beta}$$

$$+ \| \int_{B_k} \sigma_2(u) dW_2(u) \Delta\eta(A_k) \|_{2\beta} + \| \Delta\eta(A_k)\Delta\delta(B_k) \|_{2\beta} \Big)^{2\beta} < \infty. \tag{25}$$

The condition (C1) holds. Then we obtain the asymptotic normality of f_{ℓ,N_ℓ} from the central limit result in Politis, Romano and Wolf (1997). □

The proof of Lemma 2.

To show the consistency of the variance estimator in (9) we apply L_2-convergence of the subsampling estimator given by Lemma 4.6.1 in Politis, Romano and Wolf (1999). Because strong mixing condition holds from (1b) in Assumption 1, we suffice to show the following conditions (C2) and (C3) for the application of Lemma 4.6.1.

(C2) $K_\ell^{-1} \sum_{h=0}^{K_\ell - 1} \mathrm{V}_{\mathrm{IJ}}\left[M_\ell^{1/2} \bar{Z}_{\ell,M_\ell}^{hM_\ell} \right] \to \sigma_{\ell,f}^2$ s.t. $M_\ell \to \infty$ and $M_\ell/N_\ell \to 0$ as $N_\ell \to \infty$.

(C3) $(f_{\ell,N_\ell})^4$ is uniformly integrable.

Denote $\mathrm{V}_{\mathrm{IJ}}\left[M_\ell^{1/2} \bar{Z}_{\ell,M_\ell}^{hM_\ell} \right]$ as $\sigma_{\ell,h}^2$. Then we have

$$\frac{1}{K_\ell} \sum_{h=0}^{K_\ell - 1} \sigma_{\ell,h}^2 - \sigma_{\ell,f}^2 \le \frac{1}{K_\ell} \sum_{h=0}^{K_\ell - 1} \left| \sigma_{\ell,h}^2 - \sigma_{\ell,f}^2 \right| = \sup_{0 \le h \le K_\ell - 1} \left| \sigma_{\ell,h}^2 - \sigma_{\ell,f}^2 \right| \to 0$$

s.t. $M_\ell \to \infty$ as $N_\ell \to \infty$ from Assumption 2. Thus (C2) holds.

For (C3), we can show $\mathrm{E}_{\mathrm{IJ}}\left| Z_{\ell,k} \right|^{4\beta} < \infty$ for some $\beta > 1$ from (1c) in Assumption 1 by the similar argument as the proof of (C1). Let the centered $Z_{\ell,k}$ be $Z_{\ell,k}^*$. It is obvious that $\mathrm{E}_{\mathrm{IJ}}\left| \sum_{k=1}^{N_\ell} Z_{\ell,k}^* \right|^{4\beta} = O\left(N_\ell^{2\beta}\right)$ because the sequence $\{Z_{\ell,k}^*\}$ is m-dependent with $\mathrm{E}_{\mathrm{IJ}}\left| Z_{\ell,k}^* \right|^{4\beta} < \infty$. Therefore, the order of $\mathrm{E}_{\mathrm{IJ}}\left| f_{\ell,N_\ell} \right|^{4\beta} = \mathrm{E}_{\mathrm{IJ}}\left| N_\ell^{-1/2} \sum_{k=1}^{N_\ell} Z_{\ell,k}^* \right|^{4\beta}$ becomes $O(1)$. Then (C3) holds because $\mathrm{E}_{\mathrm{IJ}}\left| f_{\ell,N_\ell} \right|^{4\beta} < \infty$ implies that $(f_{\ell,N_\ell})^4$ is uniformly integrable. Finally these results yield $\hat{\sigma}_{\ell,f}^2 \xrightarrow{a} \sigma_{\ell,f}^2$ as $N_\ell \to \infty$. □

The proof of Theorem 2.

It is to see that the conditional expectation of $\hat{\gamma}(\ell)$ is

$$\mathrm{E}_{\mathrm{IJ}}[\hat{\gamma}(\ell)] = -\frac{1}{N_\ell} \sum_{k=1}^{N_\ell} \mathrm{E}_{\mathrm{IJ}}[Z_{\ell,k}^{(\pm)}] = \gamma(\ell). \tag{26}$$

For the conditional variance of $\hat{\gamma}(\ell)$, we have

$$\begin{aligned}\mathrm{V}_{\mathrm{IJ}}[\hat{\gamma}(\ell)] &= \mathrm{V}_{\mathrm{IJ}}\Big[-\frac{1}{N_\ell}\sum_{k=1}^{N_\ell} Z_{\ell,k}^{(\pm)}\Big] = \frac{1}{N_\ell^2}\sum_{k=1}^{N_\ell}\mathrm{V}_{\mathrm{IJ}}\big[Z_{\ell,k}^{(\pm)}\big] + \frac{2}{N_\ell^2}\sum_{k=1}^{N_\ell}\sum_{j=1}^{N_\ell-k}\mathrm{Cov}_{\mathrm{IJ}}\big[Z_{\ell,k}^{(\pm)}, Z_{\ell,k+j}^{(\pm)}\big] \\ &\leq \frac{1}{N_\ell^2}\sum_{k=1}^{N_\ell}\max_k\Big\{\mathrm{V}_{\mathrm{IJ}}\big[Z_{\ell,k}^{(\pm)}\big]\Big\} + \frac{2}{N_\ell^2}\sum_{k=1}^{N_\ell}\sum_{j=1}^{\tilde{m}_k}\max_k\Big\{\Big|\mathrm{Cov}_{\mathrm{IJ}}\big[Z_{\ell,k}^{(\pm)}, Z_{\ell,k+j}^{(\pm)}\big]\Big|\Big\} = O\Big(\frac{1}{N_\ell}\Big), \quad (27)\end{aligned}$$

where $\tilde{m}_k$ is defined as $\max_j\{\mathrm{Cov}_{\mathrm{IJ}}[Z_{\ell,k}^{(\pm)}, Z_{\ell,k+j}^{(\pm)}] \neq 0, 0 < j \leq N_\ell - k\}$. Because of the finite dependence of $\{Z_{\ell,k+j}^{(\pm)}\}_{k=1}^{N_\ell}$, $\tilde{m}_k$ is finite. This implies $\mathrm{V}_{\mathrm{IJ}}[\hat{\gamma}(\ell)] \to 0$ as N_ℓ goes to infinity and the consistency of $\hat{\gamma}(\ell)$ holds. Let the asymptotic variance of $\hat{\gamma}(\ell)$ be $\omega_\ell^2 = \lim_{N_\ell\to\infty}\mathrm{E}_{\mathrm{IJ}}\Big[\big\{N_\ell^{1/2}(\hat{\gamma}(\ell) - \gamma(\ell))\big\}^2\Big]$. We find the asymptotic normality of $\hat{\gamma}(\ell)$ can be proved by the similar argument as the proof of Lemma 1. The difference point of $\{Z_{\ell,k}^{(\pm)}\}_{k=1}^{N_\ell}$ and $\{Z_{\ell,k}\}_{k=1}^{N_\ell}$ is the amount of dependence. A sequence of $\{Z_{\ell,k}^{(\pm)}\}_{k=1}^{N_\ell}$ has more dependence than $\{Z_{\ell,k}\}_{k=1}^{N_\ell}$ because $Z_{\ell,k}^{(\pm)}$ is constructed by the product of returns on the nonoverlapping intervals where the length of each interval is longer than those of A_k and B_k. However, $Z_{\ell,k}^{(\pm)}$ is not correlated with $Z_{\ell,k+h}^{(\pm)}$ for the distance of them being large enough. $\square$

Acknowledgements

The authors are grateful to the associate editor and two anonymous referees for helpful comments. Special thanks go to the associate editor for constructive suggestions on previous drafts of the article and we go through many revisions about section 3 and 4. We would like to thank Toshiaki Watanabe and participants of the 2007 Japanese Economic Association Annual Spring Meeting for valuable comments.

References

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2006), ”Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise,” Working Paper, Princeton University.

Bandi, F. M. and J. R. Russell (2006), ”Separating microstructure noise from volatility,” *Journal of Financial Economics*, 79(3), 655-692.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde and N. Shephard (2006), ”Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise,” Working Paper, University of Oxford, Department of Economics.

Carlstein, E. (1986), ”The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence,” *The Annals of Statistics* 14(3), 1171-1179.

Fukuchi, J. (1999), ”Subsampling and model selection in time series analysis,” *Biometrika* 86(3), 591-604.

Hansen, P. R. and A. Lunde (2005), ”A realized variance for the whole day based on intermittent high-frequency data,” *Journal of Financial Econometrics* 20(3), 525-554.

Hansen, P. R. and A. Lunde (2006), ”Realized Variance and Market Microstructure Noise.” *Journal of Business and Economic Statistics* 24(2), 127-218.

Hayashi, T. and N. Yoshida (2005), ”On Covariance Estimation of Nonsynchronously Observed Diffusion Processes,” *Bernoulli* 11(2), 359-379.

Hayashi, T. and N. Yoshida (2006), ”Estimating Correlations with nonsynchronous observations in Continuous Diffusion Models,” unpublished manuscript.

Hoeffding, W. and R. Herbert (1948), ”The central limit theorem for dependent random variables,” *Duke Math. J.* 15, 773-780.

Mykland, P. A. and L. Zhang (2005), ”Comment: A Selective Overview of Nonparametric Methods in Financial Econometrics,” *Statistical Science* 20(4), 347-350.

Politis, D. N., J. P. Romano, and M. Wolf (1997), ”Subsampling for heteroskedastic time series,” *Journal of Econometrics* 81, 281-317.

Politis, D. N., J. P. Romano, and M. Wolf (1997), ”Subsampling,” Springer Series in Statistics.

Serfling, R. J. (1968), ”Contributions To Central Limit Theory For Dependent Variables,” *The Annals of Mathematical Statistics* 39(4), 1158-1175.

Voev, V. and A. Lunde (2007), ”Integrated Covariance Estimation using High-frequency Data in the Presence of Noise,” *Journal of Financial Econometrics* 5(1) 68-104.

Zhang, L. (2006), ”Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach,” *Bernoulli* 12(6), 1019-1043.

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

Zhou, B. (1996), ”High-Frequency Data and Volatility in Foreign-Exchange Rates,” *Journal of Business and Economic Statistics*, 14, 45-52.

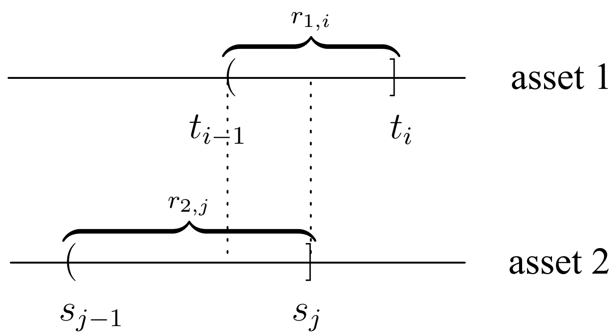

Figure 1: A case of $r_{1,i}$ and $r_{2,j}$ be overlapping.

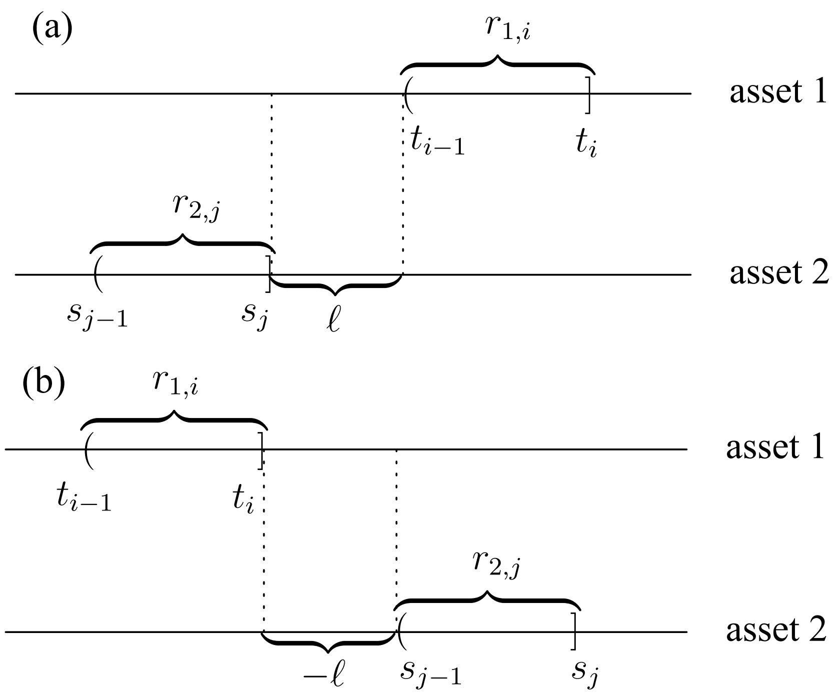

Figure 2: Pairs of returns on nonoverlapping intervals with (a) $\ell > 0$ and (b) $\ell < 0$.

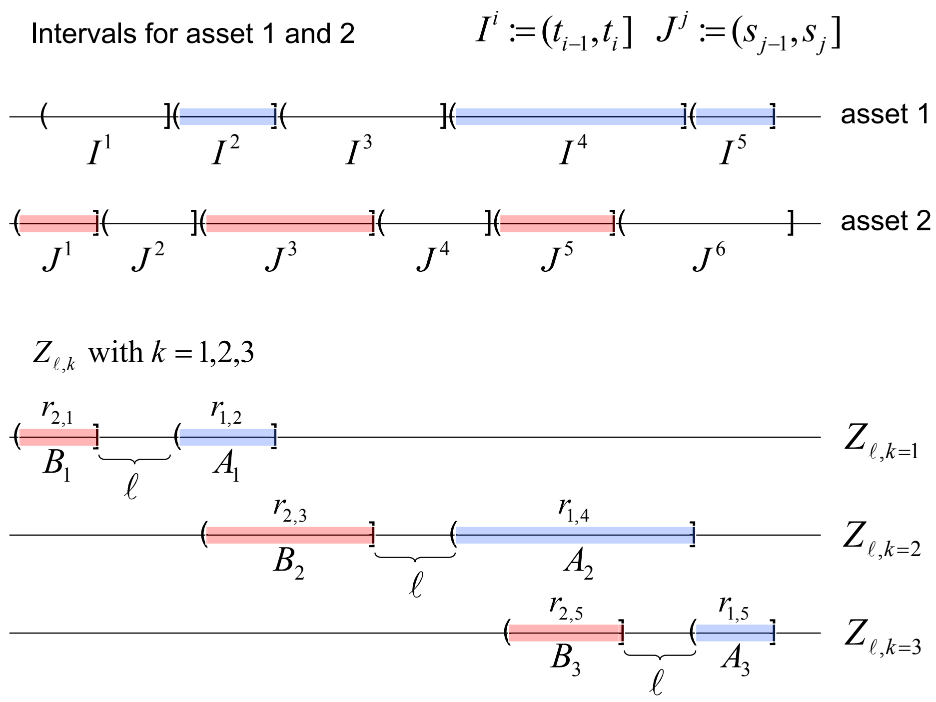

Figure 3: An example of each pair of the intervals (A_k, B_k) for $k = 1, 2, 3$.

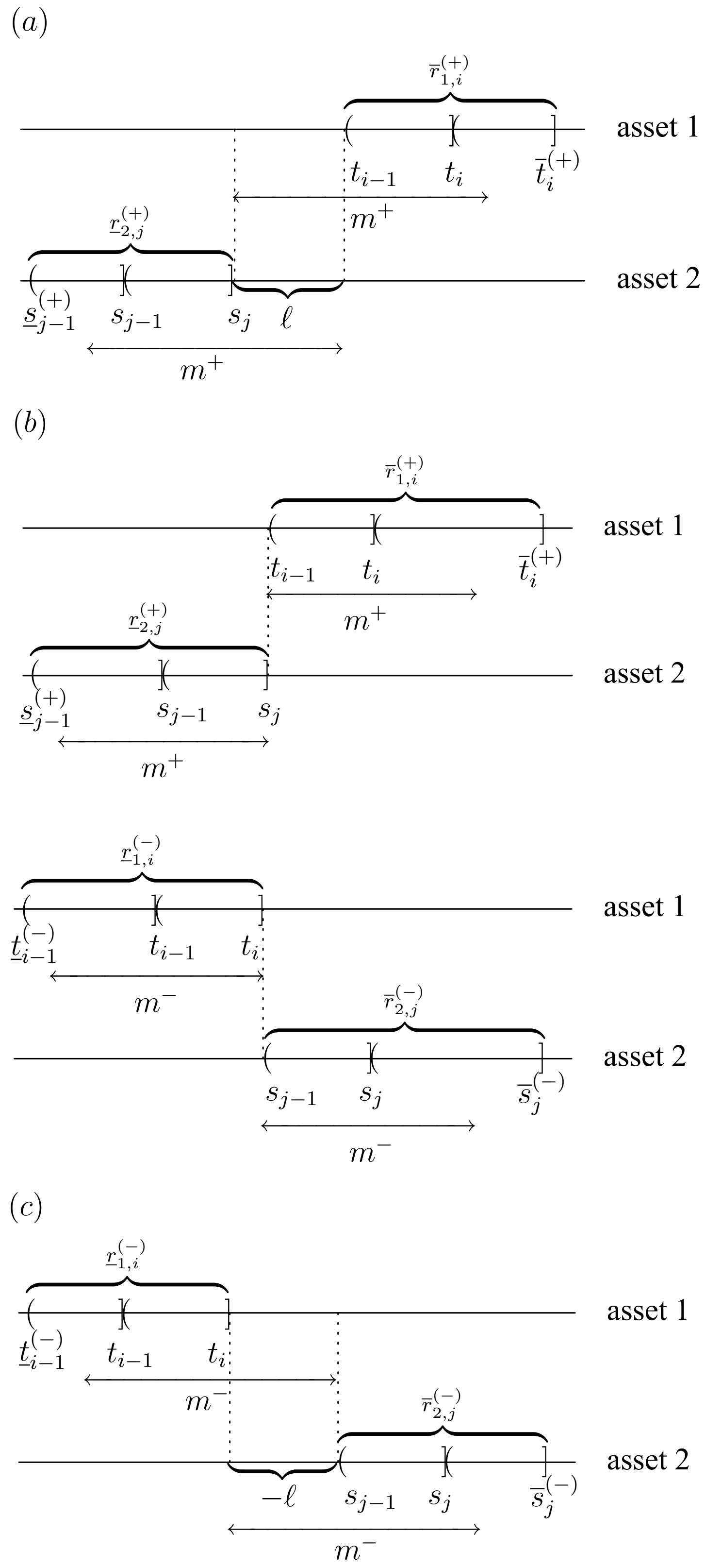

Figure 4: A pair of modified intervals in cases of (a) $\ell > 0$, (b) $\ell = 0$ and (c) $\ell < 0$.

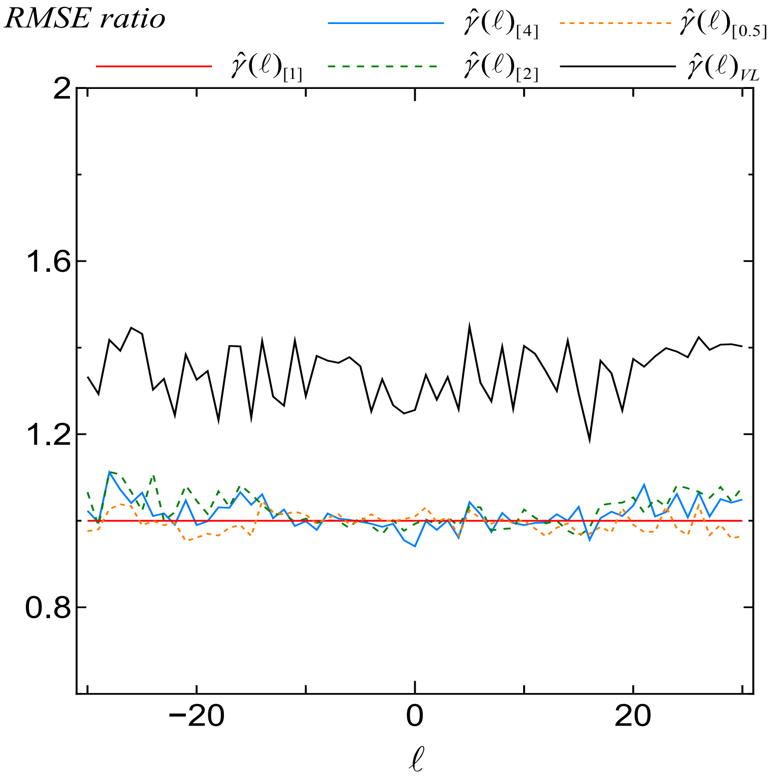

Figure 5: RMSE ratio plots of $\hat{\gamma}(\ell)_{[4]}$, $\hat{\gamma}(\ell)_{[2]}$, $\hat{\gamma}(\ell)_{[0.5]}$ and $\hat{\gamma}(\ell)_{VL}$ standardized by RMSE of $\hat{\gamma}(\ell)_{[1]}$.

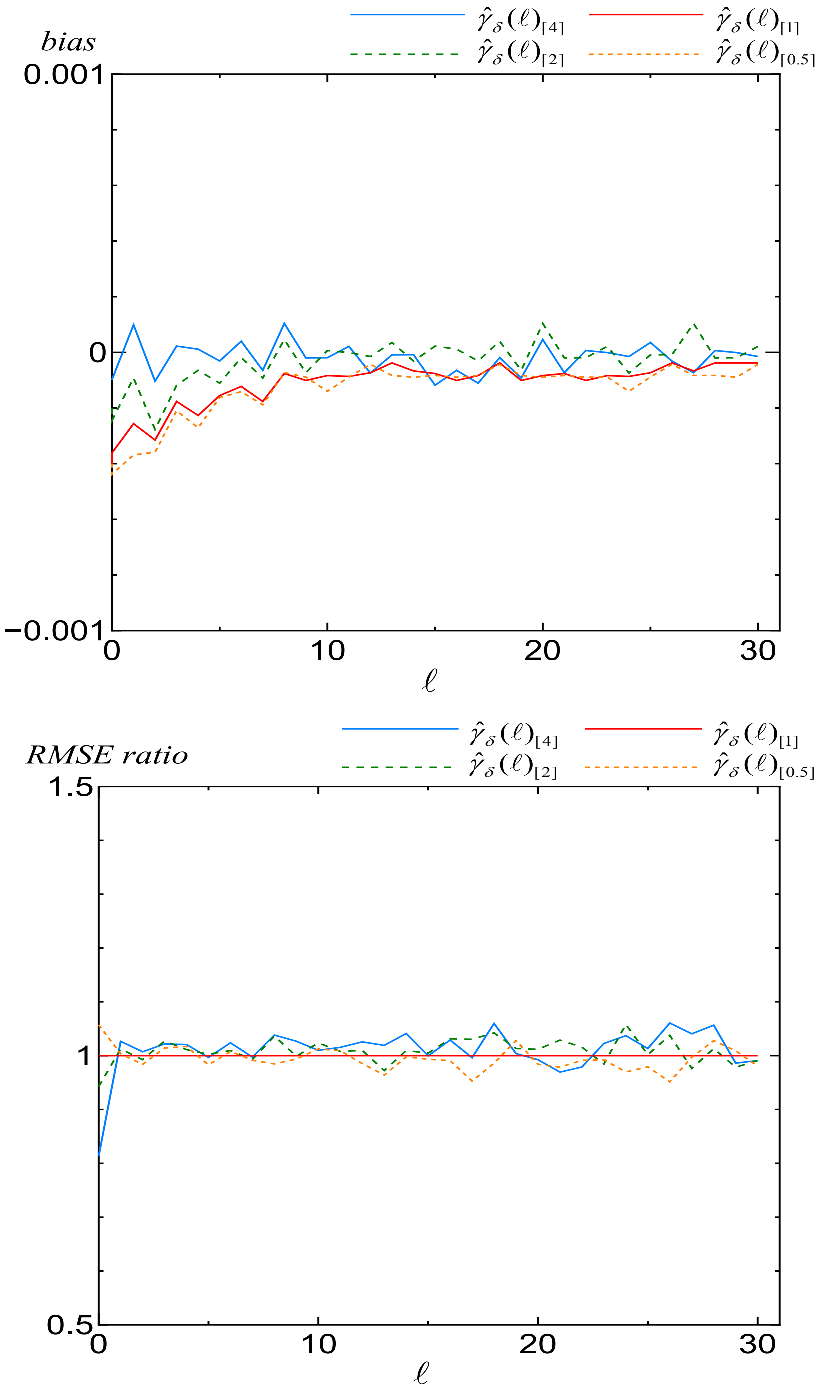

Figure 6: The bias and the RMSE ratio plots of $\hat{\gamma}_\delta(\ell)_{[4]}$, $\hat{\gamma}_\delta(\ell)_{[2]}$ and $\hat{\gamma}_\delta(\ell)_{[0.5]}$ to $\hat{\gamma}_\delta(\ell)_{[1]}$.

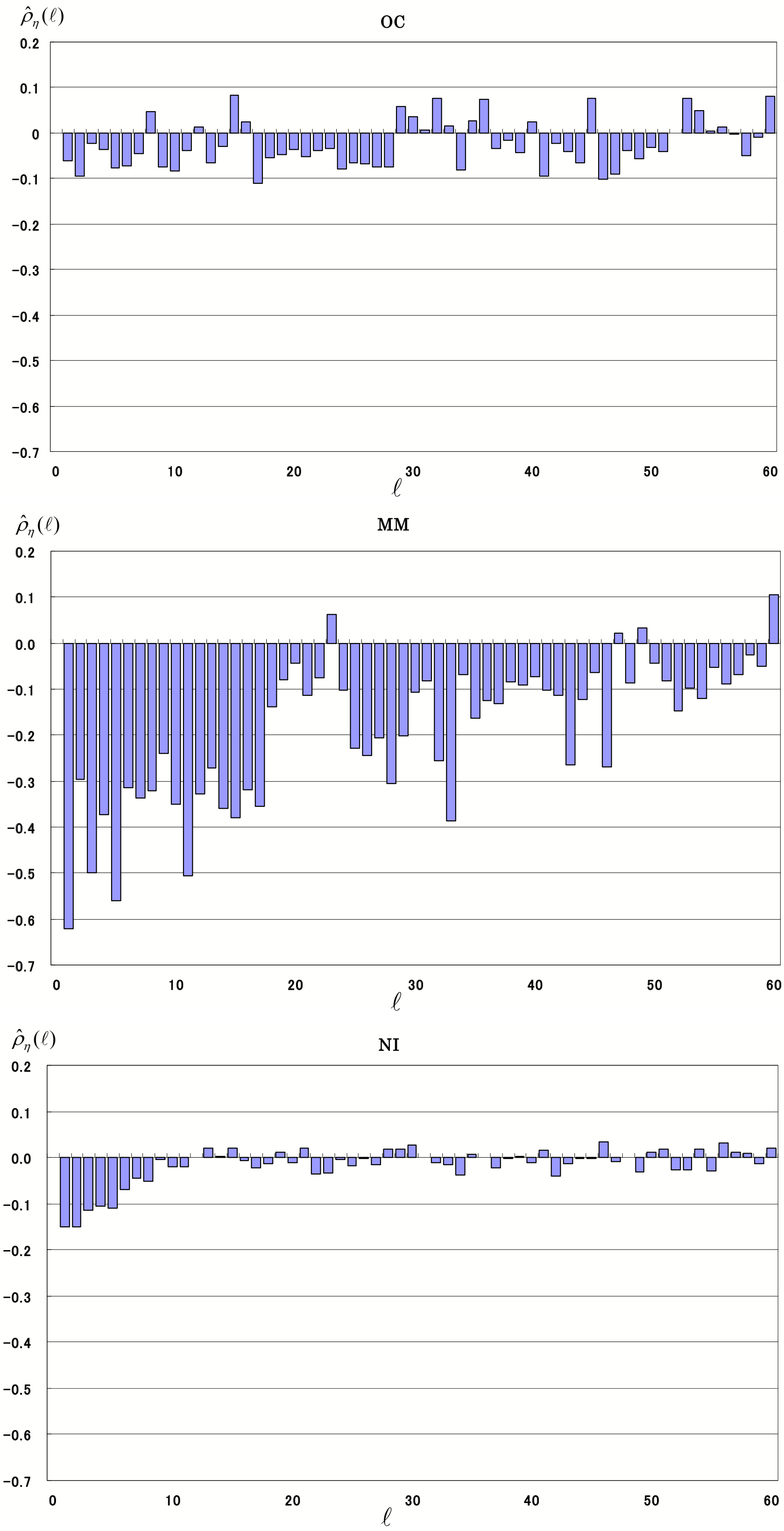

Figure 7: Autocorrelation functions of the noise in OC, MM and NI.

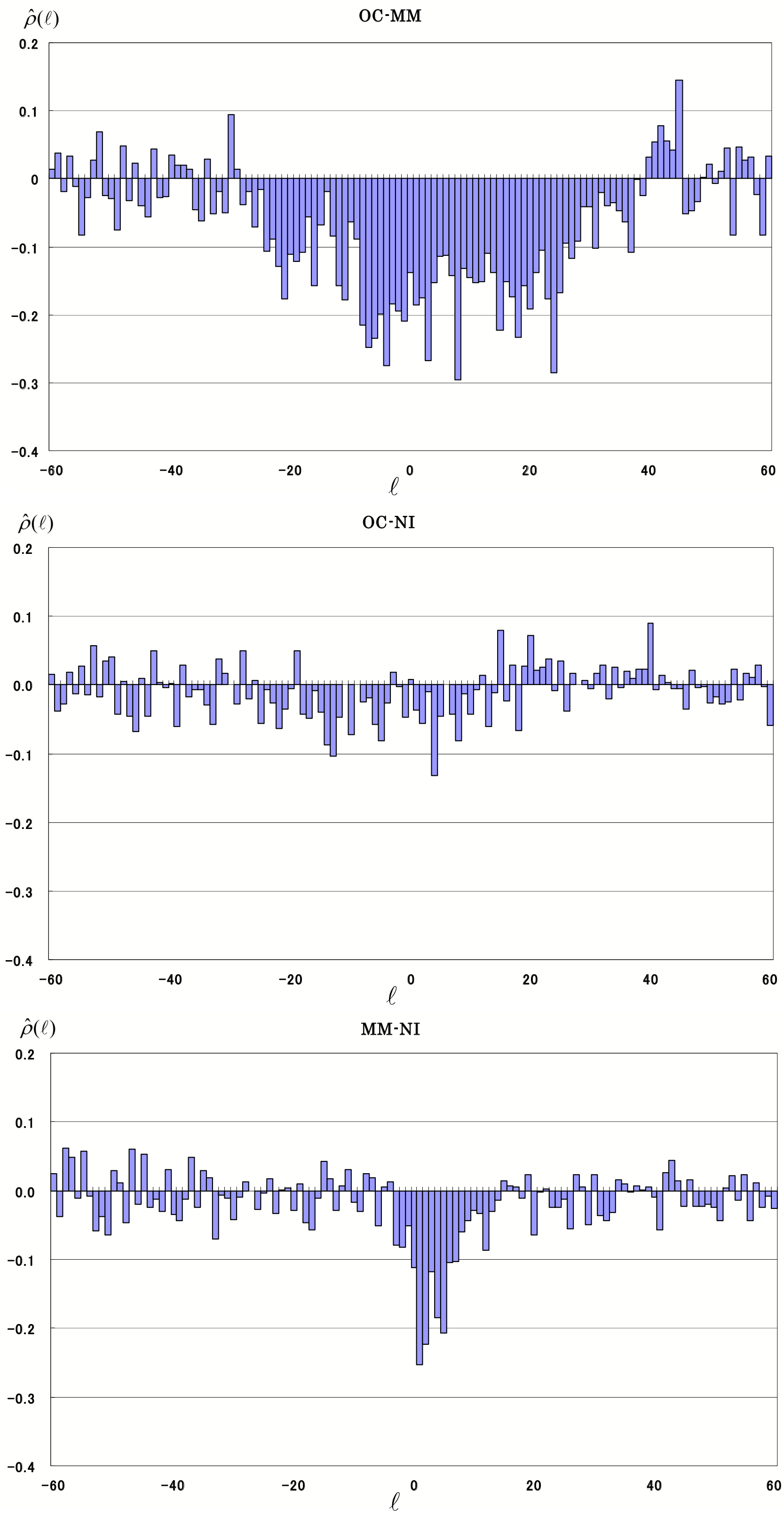

Figure 8: Cross-correlation functions of the noises in OC-MM, OC-NI and MM-NI.

Table 1: The bias and RMSE of cross-covariance estimators $\tilde{\gamma}(\ell)$, $\hat{\gamma}(\ell)_{[1]}$ and $\hat{\gamma}(\ell)_{VL}$.

ℓ	$\gamma(\ell)$	$\rho(\ell)$	$\tilde{\gamma}(\ell)$		$\hat{\gamma}(\ell)_{[1]}$		$\hat{\gamma}(\ell)_{VL}$	
			bias	RMSE	bias	RMSE	bias	RMSE
-30	0.0000	0.0000	0.0000	0.0022	0.0000	0.0025	-0.0001	0.0035
-25	0.0001	0.0000	0.0000	0.0022	0.0002	0.0025	0.0002	0.0036
-20	0.0004	0.0100	-0.0001	0.0022	0.0000	0.0028	-0.0002	0.0038
-15	0.0012	0.0400	-0.0003	0.0022	0.0000	0.0029	0.0002	0.0036
-14	0.0015	0.0500	-0.0007	0.0024	0.0001	0.0026	0.0001	0.0037
-13	0.0019	0.0600	-0.0009	0.0024	0.0001	0.0029	-0.0002	0.0037
-12	0.0024	0.0700	-0.0011	0.0025	0.0000	0.0028	0.0002	0.0036
-11	0.0029	0.0900	-0.0014	0.0027	-0.0001	0.0028	0.0000	0.0039
-10	0.0037	0.1200	-0.0017	0.0028	-0.0001	0.0029	0.0001	0.0038
-9	0.0046	0.1400	-0.0020	0.0030	0.0000	0.0029	0.0001	0.0040
-8	0.0057	0.1800	-0.0026	0.0034	-0.0001	0.0028	0.0000	0.0038
-7	0.0072	0.2300	-0.0032	0.0040	0.0001	0.0028	0.0000	0.0039
-6	0.0090	0.2800	-0.0040	0.0045	0.0000	0.0029	0.0001	0.0040
-5	0.0112	0.3500	-0.0051	0.0056	-0.0001	0.0030	0.0004	0.0040
-4	0.0139	0.4400	-0.0063	0.0067	-0.0001	0.0030	0.0001	0.0038
-3	0.0172	0.5400	-0.0079	0.0082	-0.0001	0.0029	0.0001	0.0038
-2	0.0210	0.6600	-0.0098	0.0100	-0.0003	0.0031	0.0001	0.0039
-1	0.0251	0.7900	-0.0122	0.0124	-0.0004	0.0032	-0.0001	0.0040
0	0.0286	0.9000	-0.0138	0.0139	-0.0004	0.0029	0.0000	0.0037
1	0.0222	0.7000	-0.0098	0.0101	-0.0003	0.0030	0.0001	0.0040
2	0.0175	0.5500	-0.0077	0.0081	-0.0001	0.0030	0.0001	0.0038
3	0.0139	0.4400	-0.0062	0.0066	-0.0001	0.0029	0.0002	0.0039
4	0.0111	0.3500	-0.0051	0.0055	-0.0002	0.0030	0.0001	0.0038
5	0.0089	0.2800	-0.0040	0.0046	-0.0001	0.0028	-0.0001	0.0040
6	0.0071	0.2200	-0.0032	0.0040	-0.0001	0.0029	0.0002	0.0038
7	0.0057	0.1800	-0.0024	0.0033	-0.0002	0.0029	-0.0001	0.0037
8	0.0045	0.1400	-0.0020	0.0030	-0.0001	0.0028	-0.0001	0.0039
9	0.0036	0.1100	-0.0017	0.0029	-0.0001	0.0029	0.0001	0.0037
10	0.0029	0.0900	-0.0014	0.0027	0.0000	0.0028	0.0001	0.0040
11	0.0023	0.0700	-0.0010	0.0024	0.0001	0.0028	-0.0004	0.0039
12	0.0018	0.0600	-0.0008	0.0023	-0.0001	0.0028	0.0000	0.0037
13	0.0015	0.0500	-0.0006	0.0023	0.0001	0.0028	0.0001	0.0036
14	0.0012	0.0400	-0.0005	0.0024	0.0001	0.0027	0.0001	0.0039
15	0.0010	0.0300	-0.0004	0.0022	0.0000	0.0028	0.0001	0.0037
20	0.0003	0.0100	0.0001	0.0022	0.0002	0.0027	0.0000	0.0037
25	0.0001	0.0000	-0.0002	0.0022	0.0002	0.0028	-0.0002	0.0038
30	0.0000	0.0000	0.0000	0.0022	0.0002	0.0025	-0.0001	0.0036

Note: The market microstructure noises are generated by a bivariate AR(1) process. $\gamma(\ell)$ and $\rho(\ell)$ are true cross-covariance and cross-correlation. $\tilde{\gamma}(\ell)$ is the estimator constructed by using $Z_{\ell,k}$. $\hat{\gamma}(\ell)_{[1]}$ is the cross-covariance estimator (13) using $Z^{(\pm)}_{\ell,k}$ with $c = 1$ on the variance estimation in the test statistic (11). $\hat{\gamma}(\ell)_{VL}$ is the estimator through Voev and Lunde's (2007) t-statistic.

Table 2: The RMSE ratios of $\hat{\gamma}(\ell)_{[4]}$, $\hat{\gamma}(\ell)_{[2]}$, $\hat{\gamma}(\ell)_{[0.5]}$ and $\hat{\gamma}(\ell)_{VL}$ to $\hat{\gamma}(\ell)_{[1]}$.

ℓ	$\hat{\gamma}(\ell)_{[4]}$	$\hat{\gamma}(\ell)_{[2]}$	$\hat{\gamma}(\ell)_{[0.5]}$	$\hat{\gamma}(\ell)_{VL}$
-30	1.023	1.066	0.976	1.363
-25	1.065	1.023	0.989	1.432
-20	0.990	1.046	0.961	1.326
-15	1.037	1.061	0.965	1.239
-14	1.061	1.037	1.046	1.415
-13	1.006	1.015	1.020	1.287
-12	1.026	1.016	1.015	1.266
-11	0.988	0.997	1.021	1.416
-10	0.999	1.005	1.014	1.288
-9	0.979	0.995	0.993	1.381
-8	1.017	1.001	1.007	1.370
-7	1.005	1.000	1.015	1.365
-6	1.002	0.983	0.992	1.378
-5	0.998	1.010	0.999	1.357
-4	0.993	0.988	1.015	1.253
-3	0.986	0.969	1.001	1.327
-2	0.993	1.001	0.999	1.267
-1	0.955	0.977	1.004	1.248
0	0.941	0.993	1.010	1.256
1	0.999	1.003	1.031	1.337
2	0.979	0.991	0.997	1.280
3	1.001	1.008	1.008	1.332
4	0.961	0.988	0.966	1.258
5	1.043	1.033	1.024	1.448
6	1.015	1.031	1.006	1.319
7	0.974	0.978	0.993	1.276
8	1.018	0.981	1.004	1.402
9	0.995	0.983	0.995	1.259
10	0.990	1.026	1.001	1.404
11	0.995	1.007	0.982	1.386
12	0.996	0.994	0.964	1.345
13	1.015	1.000	0.984	1.300
14	1.000	0.977	0.994	1.416
15	1.032	0.964	0.970	1.294
20	1.035	1.054	0.991	1.374
25	1.008	1.075	0.967	1.378
30	1.049	1.078	0.965	1.403

Note: $\hat{\gamma}(\ell)_{[4]}$, $\hat{\gamma}(\ell)_{[2]}$, $\hat{\gamma}(\ell)_{[1]}$ and $\hat{\gamma}(\ell)_{[0.5]}$ are the cross-covariance estimator (13) with $c = 4, 2, 1, 0.5$ on the variance estimation in the test statistic (11). $\hat{\gamma}(\ell)_{VL}$ is the estimator through Voev and Lunde's (2007) t-statistic.

Table 3: The bias and RMSE of autocovariance estimators $\hat{\gamma}_\delta(\ell)_{[4]}$, $\hat{\gamma}_\delta(\ell)_{[2]}$, $\hat{\gamma}_\delta(\ell)_{[1]}$ and $\hat{\gamma}_\delta(\ell)_{[0.5]}$.

ℓ	$\gamma_\delta(\ell)$	$\rho_\delta(\ell)$	$\hat{\gamma}_\delta(\ell)_{[4]}$		$\hat{\gamma}_\delta(\ell)_{[2]}$		$\hat{\gamma}_\delta(\ell)_{[1]}$		$\hat{\gamma}_\delta(\ell)_{[0.5]}$	
			bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
0	0.0253	1.00	-0.0001	0.0015	-0.0002	0.0018	-0.0004	0.0019	-0.0004	0.0020
1	0.0209	0.83	0.0001	0.0056	-0.0001	0.0055	-0.0003	0.0054	-0.0004	0.0055
2	0.0170	0.67	-0.0001	0.0035	-0.0003	0.0034	-0.0003	0.0034	-0.0004	0.0034
3	0.0137	0.54	0.0000	0.0027	-0.0001	0.0028	-0.0002	0.0027	-0.0002	0.0027
4	0.0110	0.44	0.0000	0.0024	-0.0001	0.0023	-0.0002	0.0023	-0.0003	0.0023
5	0.0088	0.35	0.0000	0.0021	-0.0001	0.0021	-0.0002	0.0021	-0.0002	0.0021
6	0.0071	0.28	0.0000	0.0022	0.0000	0.0021	-0.0001	0.0021	-0.0001	0.0021
7	0.0057	0.22	-0.0001	0.0022	-0.0001	0.0022	-0.0002	0.0022	-0.0002	0.0022
8	0.0045	0.18	0.0001	0.0022	0.0000	0.0022	-0.0001	0.0021	-0.0001	0.0021
9	0.0036	0.14	0.0000	0.0022	-0.0001	0.0021	-0.0001	0.0022	-0.0001	0.0021
10	0.0029	0.11	0.0000	0.0022	0.0000	0.0022	-0.0001	0.0022	-0.0001	0.0022
11	0.0023	0.09	0.0000	0.0022	0.0000	0.0022	-0.0001	0.0021	-0.0001	0.0022
12	0.0019	0.07	-0.0001	0.0021	0.0000	0.0021	-0.0001	0.0021	0.0000	0.0020
13	0.0015	0.06	0.0000	0.0020	0.0000	0.0019	0.0000	0.0019	-0.0001	0.0019
14	0.0012	0.05	0.0000	0.0019	0.0000	0.0018	-0.0001	0.0018	-0.0001	0.0018
15	0.0010	0.04	-0.0001	0.0019	0.0000	0.0019	-0.0001	0.0019	-0.0001	0.0019
16	0.0008	0.03	-0.0001	0.0018	0.0000	0.0018	-0.0001	0.0017	-0.0001	0.0017
17	0.0006	0.02	-0.0001	0.0017	0.0000	0.0018	-0.0001	0.0018	-0.0001	0.0017
18	0.0005	0.02	0.0000	0.0018	0.0000	0.0018	0.0000	0.0017	0.0000	0.0017
19	0.0004	0.02	-0.0001	0.0018	-0.0001	0.0018	-0.0001	0.0018	-0.0001	0.0018
20	0.0003	0.01	0.0000	0.0017	0.0001	0.0017	-0.0001	0.0017	-0.0001	0.0017
25	0.0001	0.00	0.0000	0.0015	0.0000	0.0015	-0.0001	0.0015	-0.0001	0.0014
30	0.0000	0.00	0.0000	0.0014	0.0000	0.0014	0.0000	0.0014	0.0000	0.0014

Note: The market microstructure noises are generated by a bivariate AR(1) process. $\hat{\gamma}_\delta(\ell)_{[4]}$, $\hat{\gamma}_\delta(\ell)_{[2]}$, $\hat{\gamma}_\delta(\ell)_{[1]}$ and $\hat{\gamma}_\delta(\ell)_{[0.5]}$ are the estimator (19) with $c = 4, 2, 1, 0.5$ in asset 2.

Table 4: The results of autocovariance estimators. Case of true price process with non-zero drift.
(a) The bias and RMSE ratio of $\hat{\gamma}(\ell)_{[1]}$ with drift parameter (μ_1, μ_2) to $\hat{\gamma}(\ell)_{[1]}$ without drift

ℓ	no drift	(μ_1, μ_2) $(0.05, 0.05)$		$(0.1, 0.1)$		$(0.5, 0.5)$	
	bias×100	bias×100	RMSE ratio	bias×100	RMSE ratio	bias×100	RMSE ratio
-30	0.0010	0.0010	1.0001	0.0010	1.0001	0.0012	1.0006
-25	0.0209	0.0210	0.9999	0.0210	0.9999	0.0216	0.9993
-20	-0.0017	-0.0017	1.0000	-0.0018	1.0001	-0.0022	1.0004
-15	0.0001	0.0001	0.9999	0.0001	0.9997	0.0004	0.9988
-10	-0.0115	-0.0115	1.0000	-0.0116	1.0001	-0.0116	1.0004
-5	-0.0059	-0.0059	0.9999	-0.0060	0.9998	-0.0065	0.9988
-4	-0.0069	-0.0069	1.0001	-0.0069	1.0002	-0.0073	1.0012
-3	-0.0120	-0.0120	0.9999	-0.0120	0.9999	-0.0125	0.9995
-2	-0.0335	-0.0335	1.0001	-0.0335	1.0002	-0.0337	1.0011
-1	-0.0353	-0.0353	1.0000	-0.0353	0.9999	-0.0354	0.9998
0	-0.0396	-0.0397	1.0000	-0.0397	1.0000	-0.0397	1.0002
1	-0.0297	-0.0297	1.0000	-0.0298	0.9999	-0.0304	0.9997
2	-0.0095	-0.0095	0.9999	-0.0095	0.9998	-0.0099	0.9990
3	-0.0076	-0.0076	1.0000	-0.0076	1.0001	-0.0080	1.0004
4	-0.0219	-0.0219	1.0000	-0.0219	0.9999	-0.0219	0.9996
5	-0.0053	-0.0053	1.0000	-0.0054	0.9999	-0.0058	0.9996
10	0.0025	0.0025	1.0000	0.0025	1.0000	0.0029	0.9999
15	0.0009	0.0009	1.0000	0.0009	0.9999	0.0009	0.9997
20	0.0236	0.0236	1.0001	0.0236	1.0003	0.0239	1.0014
25	0.0228	0.0228	1.0000	0.0229	1.0000	0.0233	0.9999
30	0.0178	0.0178	1.0000	0.0179	0.9999	0.0184	0.9996

(b) The bias and RMSE ratio of $\hat{\gamma}_\delta(\ell)_{[1]}$ with drift parameter (μ_1, μ_2) to $\hat{\gamma}_\delta(\ell)_{[1]}$ without drift.

ℓ	no drift	(μ_1, μ_2) $(0.05, 0.05)$		$(0.1, 0.1)$		$(0.5, 0.5)$	
	bias×100	bias×100	RMSE ratio	bias×100	RMSE ratio	bias×100	RMSE ratio
0	-0.0361	-0.0361	1.0000	-0.0361	1.0000	-0.0365	0.9998
1	-0.0256	-0.0256	1.0000	-0.0257	1.0001	-0.0262	1.0004
2	-0.0315	-0.0315	1.0000	-0.0315	1.0000	-0.0318	1.0000
3	-0.0176	-0.0177	0.9998	-0.0177	0.9997	-0.0183	0.9984
4	-0.0227	-0.0227	1.0000	-0.0227	1.0000	-0.0230	1.0004
5	-0.0155	-0.0155	0.9999	-0.0155	0.9998	-0.0158	0.9989
10	-0.0083	-0.0083	0.9999	-0.0083	0.9997	-0.0085	0.9989
15	-0.0076	-0.0076	1.0000	-0.0076	1.0001	-0.0080	1.0005
20	-0.0083	-0.0083	1.0000	-0.0083	0.9999	-0.0085	0.9997
25	-0.0073	-0.0073	1.0001	-0.0074	1.0001	-0.0078	1.0007
30	-0.0038	-0.0038	1.0001	-0.0038	1.0002	-0.0039	1.0011

Note: In the simulation we use the stochastic differential model with non-zero drift in (23). We set (μ_1, μ_2) as $(0.05, 0.05)$, $(0.1, 0.1)$ and $(0.5, 0.5)$.

Table 5: Autocovariance of univariate noise process.

(a) The variance estimates and the test statistics for the variance.

	OC	MM	NI
estimates	1.369×10^6	1.296×10^7	2.209×10^7
test statistics	1411.64^*	185.70^*	3412.69^*

(b) The test statistics for the autocovariance.

ℓ	OC	MM	NI
1	-0.91	-3.98*	-4.34*
2	-0.95	-2.05*	-4.91*
3	-0.24	-3.04*	-3.82*
4	-0.44	-2.26*	-3.81*
5	-1.07	-3.72*	-4.61*
6	-1.04	-2.18*	-3.03*
7	-0.73	-2.19*	-1.97*
8	0.64	-2.12*	-2.50*
9	-0.95	-1.59	-0.18
10	-1.18	-2.16*	-1.06
11	-0.49	-3.19*	-0.91
12	0.17	-2.37*	0.03
13	-0.98	-1.66	0.96
14	-0.46	-2.40*	0.03
15	1.32	-2.67*	0.95
16	0.40	-2.57*	-0.27
17	-1.71	-2.63*	-1.07
18	-1.01	-0.97	-0.58
19	-0.71	-0.62	0.64
20	-0.66	-0.38	-0.62
30	0.66	-0.77	1.28
40	0.41	-0.54	-0.52
50	-0.43	-0.42	0.53
60	1.69	1.59	0.92

Note: In the top table (a), the test statistic for the variance of noise is given by (20). The critical value at 5% significance level is 3.84. The bottom table (b) shows the test statistic (21) for the autocovariance of noise with $\ell > 0$. The critical value of the test statistic (21) is 1.96 at 5% significance level. Superscript * denotes significance at the 5% levels.

Table 6: Test statistics for a cross-covariance of bivariate noise processes.

ℓ	OC-MM	OC-NI	MM-NI
-60	0.30	0.51	0.67
-50	-0.64	1.15	0.91
-40	0.51	0.05	-1.01
-30	1.85	0.03	-1.29
-20	-1.74	-0.11	-0.95
-15	-0.98	-0.68	1.30
-10	-0.89	-1.35	-0.66
-9	-1.31	0.02	-0.96
-8	-3.29^{*}	-0.36	0.72
-7	-3.37^{*}	-0.34	0.56
-6	-3.23^{*}	-0.99	-1.49
-5	-2.87^{*}	-1.56	0.18
-4	-3.33^{*}	-0.42	0.37
-3	-2.17^{*}	0.31	-2.48^{*}
-2	-2.17^{*}	0.01	-2.24^{*}
-1	-2.27^{*}	-0.83	-1.27
0	-2.07^{*}	0.20	-3.41^{*}
1	-2.43^{*}	-0.63	-4.66^{*}
2	-2.14^{*}	-0.86	-4.02^{*}
3	-2.40^{*}	-0.15	-2.25^{*}
4	-1.81	-1.66	-3.94^{*}
5	-1.10	-0.71	-4.35^{*}
6	-1.05	0.06	-2.33^{*}
7	-1.34	-0.77	-2.48^{*}
8	-2.83^{*}	-1.27	-1.51
9	-1.51	-0.21	-1.03
10	-1.40	-0.66	-0.82
15	-1.99^{*}	1.25	0.40
20	-1.61	1.25	-1.85
30	-0.53	-0.17	0.74
40	0.54	1.79	-0.31
50	0.39	-0.86	-0.73
60	0.69	-1.84	-0.79

Note: The test statistic for the cross-covariance of the bivariate noise processes is given by (15). The critical value of the test statistics (15) is 1.96 at 5% significance level. Superscript * denote significance at the 5% levels.