Estimation and Testing for Dependence of Market Microstructure Noise

Masato Ubukata and Kosuke Oya

Graduate School of Economics, Osaka University

October 17, 2008

Abstract

Market microstructure noise, which is induced by market frictions such as the bid-ask bounce and the discreteness of the price change, generates deviations of observed price from frictionless equilibrium price. Estimation of the noise dependence is of direct economic interest because it sheds light on market microstructure effects such that the bid-ask bounce and the clustering of order flow are related to negative and positive autocorrelations of the noises, respectively. This paper proposes a test statistic for the dependence and cross and auto covariance estimators of the bivariate noise processes, and derives their asymptotic distributions. The asymptotic distributions provide another test statistics for statistical significance for the cross and auto covariances. Monte Carlo simulation shows the covariance estimators and test statistics provide good performance in a finite sample. An empirical illustration confirms the proposed statistics and estimators capture various dependence patterns in market microstructure noise.

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

JEL classification: C12; D49

1 Introduction

Market microstructure noise is induced by some market frictions: the bid-ask bounce, the discreteness of price change and asymmetric information of traders, etc. A study about each of these market frictions have been pioneered in market microstructure literature. Roll (1984) derives a simple measure of estimating the bid-ask spread based on negative autocovariance of observed return. Other covariance spread model is established by Stoll (1989) and George, Kaul and Nimalendran (1991). A related literature includes inferences about the decomposition of the bid-ask spread using an indicator function driven by the direction of trade as in Glosten and Harris (1988), Huang and Stoll (1997) and Madhavan, Richardson and Roomans (1997). Harris (1990) studies the rounding effects by the discreteness of transaction price. The asymmetric information of traders causes adverse selection components within bid-ask spread in both market maker system and limit order market (see, e.g., Glosten and Milgrom 1985, Easley and O'Hara 1987, Copeland and Galai 1983, Foucault 1999 and Glosten 1994).

A recent microstructure noise analysis has provided a way to measure how much impact the noise causes to transaction price because it deviates from frictionless equilibrium price by microstructure effects. Bandi and Russell (2006) propose an estimator to evaluate full-information cost which is a difference between the transaction price and the unobserved price reflecting all private and public information about the asset. Medahhi (2002) and Hansen and Lunde (2006) study the variance of noise and correlation between the noise and an instanteneous volatility of semimartingale price process which appear in the market with no frictions.

Many empirical analyses exhibit autocorrelations and read-lag relationship of cross-correlation in observed intraday returns and these correlations could be induced by the dependence of microstructure noises. This paper studies a degree of the dependence of the microstructure noise. The main contributions and results of the study are the followings: we propose a test statistic for the dependence and cross and auto covariance estimators of bivariate noise processes and derive their asymptotic distributions. Furthermore, we provide test statistics for significance of the variance and covariance of the noise. We confirm that the proposed cross and auto covariance estimators and test statistics have a good performance in finite samples through Monte Carlo simulation. As an empirical illustration, the statistics are applied to high-frequency asset prices on the Osaka Securities Exchange and we find various dependence patterns in market microstructure noise.

The measurement of the noise dependence enables bias correction for the integrated variance and covariance estimation, which is important for option pricing, the measurement of value-at-risk, and portfolio evaluation. We must be careful to the noise which contaminates high-frequency prices. Realized variance and covariance are not necessarily the best approaches.

In the literature on integrated variance estimation with noise, Zhou (1996) proposes a kernel-based estimator, and Zhang, Mykland and Aït-Sahalia (2005) suggest two scales of realized variance as a linear combination of realized variances at two frequencies, with Zhang (2006) extending this estimator to multiple scales. Although these studies are conducted under an i.i.d. noise assumption, market microstructure noise possibly has time dependence. Under the dependent noise assumption, Aït-Sahalia, Mykland and Zhang (2006) modify the two- and multiple-scale realized variances, and Hansen and Lunde (2005, 2006) and Barndorff-Nielsen, Hansen, Lunde and Shephard (2008) develop the kernel-based estimator. Bandi and Russell (2008) show the optimal frequency based on the minimization of the mean squared error. For integrated covariance estimation with 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, it is important that the adequacy of these estimators such as their unbiasedness, consistency and efficiency depend on the dependence structure of the noise process. Therefore, we should know whether the market microstructure noise is time dependent and how the degrees of the dependence are for the estimation of integrated variance and covariance.

Voev and Lunde (2007) propose a test statistic for the cross-sectional dependence of the noises in order to determine the kernel bandwidth of their estimator. The main differences from Voev and Lunde (2007) and our study are a choice of the intervals used for the cross-covariance estimation of the noises and an evaluation of a variance of the cross-covariance estimator of the noises. First, Voev and Lunde's (2007) cross-covariance estimator is based on subsampled interval whose length is large enough in order to guarantee the unbiasedness of the estimator. If the interval is too wide, the cross-covariance estimator may have a large variance. To take care of it, we propose a cross-covariance estimator based on the interval with proper length where the dependence of the noise disappears. The length of the interval is determined by a testing procedure proposed in this paper. Second, although Voev and Lunde (2007) show the unbiasedness of the kernel-based cumulative covariance estimator under dependent noise, their t-statistic uses an approximated variance of the cross-covariance estimator under i.i.d. noise. As suggested in their paper, it is natural that this approximation leads to t-statistics that are somewhat larger than they should be. On the other hand, the test statistic derived in this paper does not require the i.i.d. noise approximation by using subsampling methods.

The paper itself proceeds as follows. In section 2, the transaction price model and properties of market microstructure noise are presented under a framework of high-frequency financial analysis. We propose the test statistic for the cross-sectional dependence of noises in section 3. We provide the cross and auto covariance estimators of the bivariate noise processes, their asymptotic distributions, and the test statistics for their significance in section 4. Section 5 includes a simulation experiment and an empirical illustration. We conclude the paper with an appendix that provides proofs for the several lemmas and theorems.

2 Price process and market microstructure noise

We assume logarithmic equilibrium price processes of two assets, $\{P_1^*(t)\}$ and $\{P_2^*(t)\}$, which follow two-dimensional Itô process without 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. We assume that the initial value of the price $P_l^*(0)$ is a constant and $\sigma_l(t) > 0$ is a bounded progressively measurable function. $\{P_1^*(t)\}$ and $\{P_2^*(t)\}$ correspond to the logarithmic prices which appear in the market without trading imperfections, frictions and informational effects.

We make some notation to represent intraday returns and an irregularly nonsynchronous trading. $r_{1,i} := P_1(t_i) - P_1(t_{i-1})$ 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, where t_i and s_j are the end times of the *i*-th and *j*-th intervals. The different notations of the transaction times for the two assets are due to the nonsynchronous trading. A simple way to model high-frequency transaction price is to use hidden semimartingale processes, as named by Mykland and Zhang (2005). In this framework, the logarithmic transaction price P_l is observed with market microstructure noise as follows:

$$P_1(t_i) = P_1^*(t_i) + \eta(t_i), \quad P_2(s_j) = P_2^*(s_j) + \delta(s_j), \tag{2}$$

where $P_l^*(t)$ is the logarithmic equilibrium price described in (1). $\eta(t_i)$ and $\delta(s_j)$ are the market microstructure noises in asset 1 and 2 which have many sources, including the presence of bid ask spreads, the discreteness of prices change and the difference in trade sizes, etc. We assume the market microstructure noises have the following properties.

Assumption 1. *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 mean.*

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

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

m *is a finite positive number.* $\Gamma(\ell)$ *is a covariance matrix with finite elements.*

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

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

To avoid complication of the subscript for (1b), $\gamma_{\eta\delta}(\ell)$ and $\gamma_{\delta\eta}(\ell)$ are rewritten as $\gamma(\ell)$ and $\gamma(-\ell)$ because of $\gamma_{\eta\delta}(\ell) = \mathrm{E}[\eta(t)\delta(t-\ell)]$ and $\gamma_{\delta\eta}(\ell) = \mathrm{E}[\eta(t-\ell)\delta(t)] = \mathrm{E}[\eta(t)\delta(t+\ell)]$. The auto and cross correlation coefficients of the two noises are defined as $\rho_\eta(\ell)$, $\rho_\delta(\ell)$ and $\rho(\ell)$. For (1d), even if P_l^* and $\boldsymbol{u}(t)$ are correlated, the dependence between the noises generally dominates the dependence between the equilibrium price and noise as the number of high-frequency observations increases. Furthermore, Hansen and Lunde (2006) suggest that the independence assumption between the equilibrium price and noise does not statistically damage the analysis of asset prices with high trading intensities.

Our interest of the study is to estimate the covariance matrix in Assumption (1b) and characterize the asymptotic and finite sample properties of the auto and cross covariance estimators of bivariate noise processes. A test statistic for the cross-sectional dependence of the noises using the subsampling method is introduced in section 3. We provide the auto and cross covariance estimators of the bivariate noise processes, their asymptotic distributions, and test statistics for their significance in section 4.

3 Cross-sectional dependence of noises

Microstructure noise as well as equilibrium price are unobservable. We introduce a simple identification approach to measure cross-sectional dependence of the noises and propose a test statistic to detect a distance where the dependence of the noises disappears.

3.1 How to measure cross-sectional dependence of noises

First we define expectation and variance conditional on the stochastic arrival times. Denote the 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 $\mathrm{V_{IJ}}[\,\cdot\,]$. We cannot identify the covariation of the equilibrium price processes that have martingale properties and the covariance of the market microstructure noises in the pair of overlapping intervals because the covariation of the equilibrium price processes is not zero. On the other hand, in the pair of the nonoverlapping intervals $\{I^i \cap J^j = \emptyset\}$, we have:

$$\mathrm{E_{IJ}}\Big[\sum_{i,j} r_{1,i} r_{2,j} 1_{\{I^i \cap J^j]=\emptyset\}}\Big] = \mathrm{E_{IJ}}\Big[\sum_{i,j} e_{\eta,i} e_{\delta,j} 1_{\{I^i \cap J^j=\emptyset\}}\Big], \tag{3}$$

where $e_{\eta,i} := \eta(t_i) - \eta(t_{i-1})$ and $e_{\delta,j} := \delta(s_j) - \delta(s_{j-1})$. The product of returns on the nonoverlapping intervals is used to identify the covariance of the noises as in Voev and Lunde (2007).

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 the case of $s_{j-1} - t_i > 0$, the difference of the nonoverlapping intervals is denoted by $\ell = -(s_{j-1} - t_i)$. In Figure 1, the top panel (a) and the lower panel (b) illustrate the former and 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, for the other cases, we have only to replace the corresponding definition of ℓ. It is noted that $\ell = t_{i-1} - s_j > 0$ implies $\{I^i \cap J^j = \emptyset\}$. We define the product of returns on the i-th and j-th intervals satisfying $t_{i-1} - s_j = \ell > 0$ as follows:

$$Z_{\ell,ij} = r_{1,i} r_{2,j}, \quad \text{for all}, i, j, \text{ such that } t_{i-1} - s_j = \ell. \tag{4}$$

The conditional expectation of $Z_{\ell,ij}$ is:

$$\begin{aligned} \mathrm{E_{IJ}}\big[Z_{\ell,ij}\big] &= \mathrm{E_{IJ}}\big[\eta(t_i)\delta(s_j)\big] - \mathrm{E_{IJ}}\big[\eta(t_i)\delta(s_{j-1})\big] - \mathrm{E_{IJ}}\big[\eta(t_{i-1})\delta(s_j)\big] + \mathrm{E_{IJ}}\big[\eta(t_{i-1})\delta(s_{j-1})\big] \\ &= \gamma(\ell + \Delta t_i) - \gamma(\ell + \Delta t_i + \Delta s_j) - \gamma(\ell) + \gamma(\ell + \Delta s_j), \end{aligned} \tag{5}$$

where $\Delta t_i := t_i - t_{i-1}$ and $\Delta s_j := s_j - s_{j-1}$. For all ℓ taking more than a large enough L such that $\gamma(L) = 0$, we obtain $\gamma(\ell) = 0$ and $\mathrm{E_{IJ}}\big[Z_{\ell,ij}\big] = 0$ 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 $\mathrm{E_{IJ}}\big[Z_{L,ij}\big] = \mathrm{E_{IJ}}\big[Z_{L-1,ij}\big] = \mathrm{E_{IJ}}\big[Z_{L-2,ij}\big] = \cdots = \mathrm{E_{IJ}}\big[Z_{L-s^*+1,ij}\big] = 0$ and $\mathrm{E_{IJ}}\big[Z_{L-s^*,ij}\big] \neq 0$ because of $\gamma(L - s^*) \neq 0$. Denote $\ell^* = L - s^*$.

Then we have $\gamma(\ell^*+1) = 0$ and $\gamma(\ell^*) \neq 0$ and conclude that the distance ℓ^* is the threshold value of dependence of the noises. We note that whether $\mathrm{E_{IJ}}\left[Z_{\ell,ij}\right] = 0$ does not necessarily imply whether $\gamma(\ell) = 0$. This is because the sum of all cross-covariances in (5) possibly takes a value of zero, even where $\gamma(\ell) \neq 0$. To avoid this situation, we apply the method of determination for ℓ^* described above. Thus, a test statistic for determining the threshold value can be constructed by using a sample mean of $Z_{\ell,ij}$, which satisfies the nonoverlapping intervals with the distance ℓ.

3.2 Test statistic for cross-sectional dependence of noises

From Assumption (1b), the dependence of the two noises disappears when the noises are sufficiently separated. In this subsection, we propose a test statistic to detect the threshold value as in the previous subsection. For the construction of the test statistic, we define a sequence that arranges $Z_{\ell,ij}$ in ascending order of index i as $\{Z_{\ell,k}\}_{k=1}^{N_\ell}$. 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 2 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 the sample mean of $Z_{\ell,k}$; that is $\bar{Z}_{\ell,N_\ell} := \frac{1}{N_\ell}\sum_{k=1}^{N_\ell} Z_{\ell,k}$, as the estimator of $\mathrm{E_{IJ}}[Z_{\ell,k}]$. Let $f_{\ell,N_\ell} := \left(\bar{Z}_{\ell,N_\ell} - \mathrm{E_{IJ}}\left[\bar{Z}_{\ell,N_\ell}\right]\right)N_\ell^{1/2}$ be the theoretical standardization for $\bar{Z}_{\ell,N_\ell}$. We make the following assumption for the asymptotic variance of f_{ℓ,N_ℓ} which is given by $\lim_{N_\ell\to\infty} \mathrm{E_{IJ}}\left[\left(f_{\ell,N_\ell}\right)^2\right] = \sigma_{\ell,f}^2$:

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

Next we consider the asymptotic normality of the estimator of $\mathrm{E_{IJ}}[Z_{\ell,k}]$. 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 of the estimator of $\mathrm{E_{IJ}}[Z_{\ell,k}]$.

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

$$\frac{f_{\ell,N_\ell}}{\sigma_{\ell,f}} \xrightarrow{a} N(0,1). \tag{6}$$

The proof is given in the Appendix.

Although the asymptotic variance $\sigma^2_{\ell,f}$ is unknown, we can construct a consistent estimator of $\sigma^2_{\ell,f}$ by applying a subsampling method first proposed by Carlstein (1986). Carlstein (1986) considers 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 h-th subseries using $\{Z_{\ell,k}\}_{k=1}^{N_\ell}$ as follows:

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

where $[\,\cdot\,]$ denotes the integer part of a real number, M_ℓ is a number of observations within each subseries $\{Z^h_\ell\}$ and K_ℓ is a total number of subseries. The superscript h of $\{Z^h_\ell\}$ represents that the subseries take the sample $Z_{\ell,k}$ with $k = hM_\ell + 1, \cdots, (h+1)M_\ell$. The variance estimator is given by:

$$\hat{\sigma}^2_{\ell,f} = \frac{M_\ell}{K_\ell} \sum_{h=0}^{K_\ell-1} \left(\bar{Z}^h_{\ell,M_\ell} - \frac{1}{K_\ell} \sum_{h=0}^{K_\ell-1} \bar{Z}^h_{\ell,M_\ell}\right)^2, \tag{7}$$

where $\bar{Z}^h_{\ell,M_\ell}$ is a sample mean of subseries $\{Z^h_\ell\}$. 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 $M_\ell/N_\ell \to 0$ as $N_\ell \to \infty$. Then we have:*

$$\hat{\sigma}^2_{\ell,f} \to_{L_2} \sigma^2_{\ell,f} \quad \textit{as} \;\; N_\ell \to \infty. \tag{8}$$

The proof is described in the Appendix.

The purpose of this subsection is to construct a test statistic to detect the threshold value of the dependence of the noises. 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:*

$$\tau(\ell) := \frac{\sqrt{N_\ell}\bar{Z}_{\ell,N_\ell}}{\hat{\sigma}_{\ell,f}} \xrightarrow{a} N(0,1) \tag{9}$$

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 (7) because the high-frequency transaction data yield a large number of N_ℓ. However, it is difficult to determine the optimal numbers of M_ℓ and K_ℓ that minimize the mean squared error of $\hat{\sigma}^2_{\ell,f}$ because we do not know the covariance structure of the noises. For fixed N_ℓ, an increase in M_ℓ (i.e. a decrease in K_ℓ) reduces the bias but increases the 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 the noise process. We investigate the influence of the numbers of M_ℓ and K_ℓ selected under $M_\ell = cN_\ell^{1/3}$ with some c through Monte Carlo simulation in section 5.

Next, we summarize the procedure to identify the threshold value where the cross-covariance of noises becomes zero. We first test the null hypothesis $\mathrm{E}_{\mathrm{IJ}}[Z_{L,k}] = 0$ for all k with a large value L using the test statistic (9). 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, we then test whether $\mathrm{E}_{\mathrm{IJ}}[Z_{L-1,k}]$ is zero. If $\mathrm{E}_{\mathrm{IJ}}[Z_{L-1,k}] = 0$ is not rejected, we proceed to judge the statistical significance of $\mathrm{E}_{\mathrm{IJ}}[Z_{L-2,k}]$. We continue to test sequentially until the null is 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 the critical value of the test statistic (9).

4 Consistent covariance estimators of noises

Once the threshold value ℓ^* as described in the section 3 has been selected, we find that the cross covariance of bivariate noise processes $\gamma(\ell)$ is zero when $\ell \geq \ell^*$. However, $\gamma(\ell)$ is still unknown for $\ell < \ell^*$. In subsection 4.1, we derive the cross-covariance estimator of bivariate noise processes and its asymptotic distribution, and propose a test statistic for statistical significance of the cross covariance. The related statistics for autocovariance of univariate noise process are established in subsection 4.2.

4.1 Cross-covariance estimator of noises

We consider the estimation of the cross-covariance $\gamma(\ell)$ in this subsection. For construction of an unbiased estimator, we only have to remedy the nonoverlapping intervals so that all cross-covariances in (5) except $\gamma(\ell)$ become zero by using the threshold value determined through the test statistic (9). We use m^+ and m^- for the threshold values in the cases of $\ell > 0$ and $\ell < 0$ instead of ℓ^*. 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 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 (4), is defined as follows:

$$Z_{\ell,ij}^{(\pm)} = \begin{cases} \bar{r}_{1,i}^{(+)}\, \underline{r}_{2,j}^{(+)} 1_{\{t_{i-1}-s_j=\ell\}} & \text{if} \quad \ell > 0 \\ \bar{r}_{1,i}^{(+)}\, \underline{r}_{2,j}^{(+)} 1_{\{t_{i-1}-s_j=0\}} + \underline{r}_{1,i}^{(-)}\, \bar{r}_{2,j}^{(-)} 1_{\{s_{j-1}-t_i=0\}} & \text{if} \quad \ell = 0 \\ \underline{r}_{1,i}^{(-)}\, \bar{r}_{2,j}^{(-)} 1_{\{s_{j-1}-t_i=-\ell\}} & \text{if} \quad \ell < 0 \end{cases} \tag{10}$$

The top panel (a), the middle panel (b) and the lower panel (c) in Figure 3 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^+ are zero. The nonoverlapping intervals described in (b) and (c) are given by the same idea as in (a). The conditional expectation of $Z^{(\pm)}_{\ell,ij}$ where the indicator function takes one is $\mathrm{E}_{\mathrm{IJ}}\big[Z^{(\pm)}_{\ell,ij}\big] = -\gamma(\ell)$. We select $Z^{(\pm)}_{\ell,ij}$ for all i, j such that the indicator function takes a value of one and define a sequence that arranges the selected $Z^{(\pm)}_{\ell,ij}$ in ascending order of index i as $\{Z^{(\pm)}_{\ell,k}\}_{k=1}^{N_\ell}$. Then the cross-covariance estimator is given by:

$$\hat{\gamma}(\ell) = -\frac{1}{N_\ell}\sum_{k=1}^{N_\ell} Z^{(\pm)}_{\ell,k}. \tag{11}$$

The following theorem states the asymptotic normality of the cross-covariance estimator $\hat{\gamma}(\ell)$.

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

$$N_\ell^{1/2}(\hat{\gamma}(\ell) - \gamma(\ell)) \xrightarrow{a} N(0, \omega_\ell^2), \tag{12}$$

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 the Appendix. Then the subsampling variance estimator is given by:

$$\hat{\omega}_\ell^2 \;=\; \frac{M_\ell}{K_\ell}\sum_{h=0}^{K_\ell - 1}\Big(\bar{Z}^{(\pm),h}_{\ell,M_\ell} - \frac{1}{K_\ell}\sum_{h=0}^{K_\ell - 1}\bar{Z}^{(\pm),h}_{\ell,M_\ell}\Big)^2, \tag{13}$$

where $\bar{Z}^{(\pm),h}_{\ell,M_\ell}$ is a sample mean of subseries $\{Z^{(\pm),h}_\ell\}$ and the h-th subseries is:

$$\{Z^{(\pm),h}_\ell\} := \big(Z^{(\pm)}_{\ell,hM_\ell+1}, Z^{(\pm)}_{\ell,hM_\ell+2}, \cdots, Z^{(\pm)}_{\ell,(h+1)M_\ell}\big),\;\; 0 \le h \le K_\ell - 1,\;\; K_\ell = \big[N_\ell/M_\ell\big].$$

The asymptotic distribution of $\hat{\gamma}(\ell)$ established in Theorem 2 and the subsampling variance estimator provide 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.*

As N_ℓ goes to infinity s.t. $M_\ell \to \infty$ and $M_\ell / N_\ell \to 0$, we have:

$$\tau^*(\ell) := \frac{\sqrt{N_\ell}\,\hat{\gamma}(\ell)}{\hat{\omega}_\ell} \xrightarrow{a} N(0,1) \tag{14}$$

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

It is noted for the integrated covariance estimation that the bias of cumulative covariance estimator by Hayashi and Yoshida (2005) is virtually zero when the test statistic (14) does not reject the null $\gamma(\ell) = 0$ for all ℓ. Even if the cumulative covariance estimator is unbiased, the noise has a strong impact on its variance. An estimator applying subsampling method to the cumulative covariance estimator by Hayashi and Yoshida (2005) is preferable for such case. It is the same as the subsampling estimator by Voev and Lunde (2007).

4.2 Autocovariance estimator of noise

In the previous subsection, we have proposed the cross-covariance estimator of the noises and have derived the test statistic for the significance of the cross-covariance. This framework is applicable for estimation of the autocovariance of a univariate noise process. We briefly describe a consistent autocovariance estimator of the noise in this subsection. We start with the construction of the test statistic to find the threshold value of the dependence of univariate noise process. Define the product of the returns on the i-th and j-th intervals for asset 1 satisfying $\ell = t_{j-1} - t_i \geq 0$ as follows:

$$Z_{1,\ell,ij} = r_{1,i} r_{1,j}, \qquad \text{for all } i, j, \text{ such that } t_{j-1} - t_i = \ell \geq 0. \tag{15}$$

Let $\{Z_{1,\ell,k}\}_{k=1}^{N_{1,\ell}}$ be a sequence that arranges $Z_{1,\ell,ij}$ in ascending order of index i. $N_{1,\ell}$ is the 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, which implies $\gamma_\eta(\ell) = 0$, and the alternative hypothesis consists of all possible deviations from the null.

The test statistic for this hypothesis is given by:

$$\tau_\eta(\ell) := \frac{\sqrt{N_{1,\ell}}\ \bar{Z}_{1,\ell,N_{1,\ell}}}{\hat{\sigma}_{1,\ell,f}}, \tag{16}$$

$\hat{\sigma}^2_{1,\ell,f}$ is a subsampling estimator of $\sigma^2_{1,\ell,f} = \lim_{N_{1,\ell}\to\infty} \mathrm{E}_{\mathrm{IJ}}\Big[\big(f_{1,\ell,N_{1,\ell}}\big)^2\Big]$, where $f_{1,\ell,N_{1,\ell}} := \big(\bar{Z}_{1,\ell,N_{1,\ell}} - \mathrm{E}_{\mathrm{IJ}}\big[\bar{Z}_{1,\ell,N_{1,\ell}}\big]\big) N^{1/2}_{1,\ell}$ and $\bar{Z}_{1,\ell,N_{1,\ell}}$ is a sample mean of $\{Z_{1,\ell,k}\}$.

The subsampling variance estimator is denoted by:

$$\hat{\sigma}^2_{1,\ell} = \frac{M_{1,\ell}}{K_{1,\ell}} \sum_{h=0}^{K_{1,\ell-1}} \Big(\bar{Z}^h_{1,\ell,M_{1,\ell}} - \frac{1}{K_{1,\ell}} \sum_{h=0}^{K_{1,\ell-1}} \bar{Z}^h_{1,\ell,M_{1,\ell}}\Big)^2, \tag{17}$$

where $\bar{Z}^h_{1,\ell,M_{1,\ell}}$ is sample mean of the h-th subseries $\{Z^h_{1,\ell}\}$:

$$\{Z^h_{1,\ell}\} := \big(Z_{1,\ell,hM_{1,\ell}+1}, Z_{1,\ell,hM_{1,\ell}+2}, \cdots, Z_{1,\ell,(h+1)M_{1,\ell}}\big),\ \ 0 \le h \le K_{1,\ell-1},\ \ K_{1,\ell} = \big[N_{1,\ell}/M_{1,\ell}\big].$$

Under the null we have $\tau_\eta(\ell) \xrightarrow{a} N(0,1)$ as $N_{1,\ell}$ goes to infinity s.t. $M_{1,\ell} \to \infty$ and $M_{1,\ell}/N_{1,\ell} \to 0$. 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 (16) enables us to identify the threshold value m_1.

To derive the autocovariance estimator of the noise, we construct $Z^{(\pm)}_{1,\ell,ij}$ for all i, j satisfying $t_{j-1} - t_i = \ell\ \ge 0$ using the selected threshold value m_1 as follows:

$$\begin{aligned} Z^{(\pm)}_{1,\ell,ij} &= \underline{r}^{(-)}_{1,i}\bar{r}^{(+)}_{1,j} = \Big(P(t_i) - P\big(\underline{t}^{(-)}_{i-1}\big)\Big)\Big(P\big(\bar{t}^{(+)}_j\big) - P(t_{j-1})\Big), \\ &\quad \text{for all } i, j, \text{ such that } t_{j-1} - t_i = \ell\ \ge 0, \end{aligned} \tag{18}$$

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

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

where $\omega^2_{\eta,\ell} = \lim_{N_{1,\ell}\to\infty} \mathrm{E}_{\mathrm{IJ}}\Big[\big\{N_{1,\ell}^{1/2}(\hat{\gamma}_\eta(\ell) - \gamma_\eta(\ell))\big\}^2\Big]$. The subsampling variance estimator of $\omega^2_{\eta,\ell}$ is given by:

$$\hat{\omega}^2_{\eta,\ell} = \frac{M_{1,\ell}}{K_{1,\ell}} \sum_{h=0}^{K_{1,\ell-1}} \Big(\bar{Z}^{(\pm),h}_{1,\ell,M_{1,\ell}} - \frac{1}{K_{1,\ell}} \sum_{h=0}^{K_{1,\ell-1}} \bar{Z}^{(\pm),h}_{1,\ell,M_{1,\ell}}\Big)^2, \tag{20}$$

where $\bar{Z}^{(\pm),h}_{1,\ell,M_\ell}$ is a sample means of the h-th subseries $\{Z^{(\pm),h}_{1,\ell}\}$:

$$\{Z^{(\pm),h}_{1,\ell}\} := \big(Z^{(\pm)}_{1,\ell,hM_{1,\ell}+1}, Z^{(\pm)}_{1,\ell,hM_{1,\ell}+2}, \cdots, Z^{(\pm)}_{1,\ell,(h+1)M_{1,\ell}}\big), \;\; 0 \le h \le K_{1,\ell-1}, \;\; K_{1,\ell} = \big[N_{1,\ell}/M_{1,\ell}\big].$$

The test statistic to find whether the autocovariance of the noise is zero 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 $\gamma_\eta(0) = 0$ and $\gamma_\eta(0) > 0$. As $N_{1,0}$ goes to infinity s.t. $M_{1,0} \to \infty$ and $M_{1,0}/N_{1,0} \to 0$, the one-sided test statistic for the variance of the noise is asymptotically distributed as chi-square with one degree of freedom:

$$\tau^*_\eta(0) = \Big(\frac{\sqrt{N_{1,0}}\;\hat{\gamma}_\eta(0)}{\hat{\omega}_{\eta,0}}\Big)^2 \xrightarrow{a} \chi^2(1) \tag{21}$$

  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 $\gamma_\eta(\ell) \neq 0$. As $N_{1,\ell}$ goes to infinity s.t. $M_{1,\ell} \to \infty$ and $M_{1,\ell}/N_{1,\ell} \to 0$, the test statistic for the significance of the *autocovariance of the noise is:*

$$\tau_\eta^*(\ell) := \frac{\sqrt{N_{1,\ell}}\, \hat{\gamma}_\eta(\ell)}{\hat{\omega}_{\eta,\ell}} \xrightarrow{a} N(0,1) \tag{22}$$

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

5 Monte Carlo simulation and empirical illustration

In this section, we investigate finite sample properties of auto and cross covariance estimators and the test statistics proposed in this paper through Monte Carlo simulation, and apply them to the high-frequency transaction prices of several stocks on the Osaka Securities Exchange as an empirical illustration.

5.1 Finite sample properties of covariance estimators of noises

We employ the data generation process introduced in Voev and Lunde (2007). The equilibrium price processes P_1^* and P_2^* follow the stochastic differential equations:

$$\begin{aligned} dP_l^*(t) &= \sigma_l(t)\Big[\sqrt{1-\lambda_l^2}\, dW_l^{(A)}(t) \;+\; \lambda_l\, dW_l^{(B)}(t)\Big], \\ d\sigma_l^2(t) &= \kappa_l(\theta_l - \sigma_l^2(t))\, dt \;+\; \omega_l \sigma_l^2(t)\, dW_l^{(B)}(t), \quad l = 1, 2, \end{aligned} \tag{23}$$

where $W_l^{(\cdot)}$ is a standard Brownian motion and $\sigma_l^2(t)$ follows the generalized autoregressive conditional heteroskedasticity (GARCH) diffusion process. $W_1^{(A)}$ and $W_2^{(A)}$ are correlated; that is, $d\langle W_1^{(A)}, W_2^{(A)}\rangle_t = \rho^*(t)dt$. We use the anti-Fisher transformation to generate stochastic correlation $\rho^*(t)$:

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

where $x(t)$ follows the GARCH diffusion process. We take $(\lambda_1, \lambda_2) = (0.5, 0.5)$, $(\kappa_1, \kappa_2, \kappa_3) = (0.006, 0.037, 0.051)$, $(\theta_1, \theta_2, \theta_3) = (2.719, 2.527, 0.200)$ and $(\omega_1, \omega_2, \omega_3) = (0.0382, 0.216, 0.130)$. The parameters κ_l, θ_l and ω_l, for $l = 1, 2$ are obtained by Drost and Werker's (1996) relationship between the continuous time and discrete time parameters which are the daily GARCH(1,1) estimates based on Intel Corporation and Microsoft Corporation returns. The trading time per day is set as 6.5 hours like in the NYSE and NASDAQ.

5.1.1 Bivariate moving average noise

In this simulation, we consider regular and synchronous sampling. First we employ bivariate moving average noise processes which are truncated at some finite lag. The noise $\eta(t)$ and $\delta(t)$ are generated by the following bivariate MA(2) model every a transaction time:

$$\begin{aligned}\eta(t) &= \epsilon_\eta(t) - 0.5\,\epsilon_\eta(t-1) + 0.4\,\epsilon_\eta(t-2) - 0.1\,\epsilon_\delta(t-1) + 0.15\,\epsilon_\delta(t-2),\\ \delta(t) &= \epsilon_\delta(t) - 0.4\,\epsilon_\eta(t-1) + 0.15\,\epsilon_\eta(t-2) - 0.1\,\epsilon_\delta(t-1) + 0.3\,\epsilon_\delta(t-2).\end{aligned}$$

To set variance of the noise we use a noise-to-signal ratio defined as the variance of the noise divided by the integrated variance. Hansen and Lunde (2006) report that noise-to-signal ratios take a range from 0.0002 to 0.006 using transaction data for 30 stocks on the NYSE and NASDAQ. We then set the variance of noise such that the noise-to-signal ratios of assets 1 and 2 average 0.005 and 0.002, respectively.

The transaction time interval is set to be 10, 15 and 30 seconds, respectively. Two transaction price series, which are the sum of the equilibrium price and noise, are sampled synchronously at equidistant times. In the regular and synchronous sampling, $\ell = 1$ represents one tick time whose length is equal to each sampling interval (10, 15 and 30 seconds). On each sampling interval, the sample size of N_ℓ ($N_{2,\ell}$) for $\ell > 0$ per day is about 2300, 1550 and 770, respectively. On the other hand, the irregular and non-synchronous sampling is applied in section 5.1.2.

We use the test statistic (9) and (16) to find the threshold value where the cross and auto covariances of the noises become zero, and estimate the cross and auto covariances of the bivariate noise processes by (11) and (19). We compare the performance of the following covariance estimators.

- Cross and auto covariance estimators constructed by the sample mean of $Z_{\ell,k}$ and $Z_{2,\ell,k}$; that is, $\tilde{\gamma}(\ell) := -\frac{1}{N_\ell}\sum_{k=1}^{N_\ell} Z_{\ell,k}$ and $\tilde{\gamma}_\delta(\ell) := -\frac{1}{N_{2,\ell}}\sum_{k=1}^{N_{2,\ell}} Z_{2,\ell,k}$. For example, the conditional expectation of $Z_{\ell,k}$ is equal to $\gamma(\ell+\Delta t_i) - \gamma(\ell+\Delta t_i+\Delta s_j) - \gamma(\ell) + \gamma(\ell+\Delta s_j)$ as represented in (5). Therefore, we expect the cross-covariance estimator $\tilde{\gamma}(\ell)$ with ℓ less than the threshold value is biased and the autocovariance estimator $\tilde{\gamma}_\delta(\ell)$ is also biased by the same reason.

- The proposed cross and auto covariance estimators $\hat{\gamma}(\ell)$ and $\hat{\gamma}_\delta(\ell)$ in (11) and (19). These estimators are unbiased if the threshold values are correctly selected. For the variance estimation in the test statistic (9) and (16), we use the asymptotic formula $M_\ell = cN_\ell^{1/3}$ and $M_{2,\ell} = cN_{2,\ell}^{1/3}$ with some c to determine M_ℓ (the number of observations within each subseries) and K_ℓ (the total number of subseries). Denote the cross and auto covariance estimators as $\hat{\gamma}(\ell)_{[c]}$ and $\hat{\gamma}(\ell)_{[c]}$. We set $c = 4, 2, 1$ and 0.5 to investigate the influence of M_ℓ and K_ℓ. The corresponding numbers of M_ℓ ($M_{2,\ell}$) and K_ℓ ($K_{2,\ell}$) are: (c, M_ℓ, K_ℓ) =$(4, 52, 44)$, $(2, 26, 88)$, $(1, 13, 176)$, $(0.5, 7, 330)$ with sample size = 2300, (c, M_ℓ, K_ℓ) =$(4, 46, 33)$, $(2, 23, 67)$, $(1, 12, 129)$, $(0.5, 6, 258)$ with sample size = 1550 and (c, M_ℓ, K_ℓ) =$(4, 36, 21)$, $(2, 18, 42)$, $(1, 9, 85)$, $(0.5, 5, 154)$ with sample size = 770, respectively.

- The Voev and Lunde's (2007) cross-covariance estimator:

$$\ddot{\gamma}(\ell) = \frac{1}{S_1 S_2}\sum_{b_1=1}^{S_1}\sum_{b_2=1}^{S_2}\ddot{\gamma}(\ell)_b\ ,$$

  where S_1 and S_2 are numbers of subgrids for asset 1 and 2. Voev and Lunde's (2007) cross-covariance estimator $\ddot{\gamma}(\ell)$ is based on subsampling method. They do not conduct a selection of the threshold value before the cross-covariance estimation. Instead of the selection, S_1 and S_2 are taken as the values where length of subsampled interval is large enough in order to eliminate a bias caused by three covariances $\gamma(\ell+\Delta t_i)$, $\gamma(\ell+\Delta t_i+\Delta s_j)$ and $\gamma(\Delta s_j)$. For the b_l-th subgrid ($b_l = 1, \ldots, S_l$, for $l = 1, 2$), define the subgrid returns $r_{1,i^b} := P_1(t_{i^b}) -$

$P_1(t_{i^b-1})$ with $\{t_{i^b}\}_{i^b=1}^{N_{1,\ell}^b}$ and $r_{2,j^b} := P_2(s_{j^b}) - P_2(s_{j^b-1})$ with $\{s_{j^b}\}_{j^b=1}^{N_{2,\ell}^b}$. The shorthand subscript b is used in place of b_1 and b_2. Then $\ddot{\gamma}(\ell)_b$ is denoted by:

$$\ddot{\gamma}(\ell)_b = \begin{cases} -\frac{1}{N_{\ell,b}} \sum_{i^b=1}^{N_{1,\ell}^b} \sum_{j^b=1}^{N_{2,\ell}^b} r_{1,i^b} r_{2,j^b} 1_{\{t_{i^b-1} - s_{j^b} = \ell\}} & \text{for} \quad \ell > 0, \\ -\frac{1}{N_{\ell,b}} \sum_{i^b=1}^{N_{1,\ell}^b} \sum_{j^b=1}^{N_{2,\ell}^b} r_{1,i^b} r_{2,j^b} 1_{\{t_{i^b-1} - s_{j^b} = 0\} \cup \{s_{j^b-1} - t_{i^b} = 0\}} & \text{for} \quad \ell = 0, \\ -\frac{1}{N_{\ell,b}} \sum_{i^b=1}^{N_{1,\ell}^b} \sum_{j^b=1}^{N_{2,\ell}^b} r_{1,i^b} r_{2,j^b} 1_{\{s_{j^b-1} - t_{i^b} = -\ell\}} & \text{for} \quad \ell < 0, \end{cases} \tag{24}$$

with

$$N_{\ell,b} = \begin{cases} \sum_{i^b=1}^{N_{1,\ell}^b} \sum_{j^b=1}^{N_{2,\ell}^b} 1_{\{t_{i^b-1} - s_{j^b} = \ell\}} & \text{for} \quad \ell > 0, \\ \sum_{i^b=1}^{N_{1,\ell}^b} \sum_{j^b=1}^{N_{2,\ell}^b} 1_{\{t_{i^b-1} - s_{j^b} = 0\} \cup \{s_{j^b-1} - t_{i^b} = 0\}} & \text{for} \quad \ell = 0, \\ \sum_{i^b=1}^{N_{1,\ell}^b} \sum_{j^b=1}^{N_{2,\ell}^b} 1_{\{s_{j^b-1} - t_{i^b} = -\ell\}} & \text{for} \quad \ell < 0. \end{cases} \tag{25}$$

On the other hand, we propose a cross-covariance estimator based on the interval with proper length where the dependence of the noise disappears. The length of the interval is determined through the test statistic (9) and (16) proposed in this paper.

We also introduce Voev and Lunde's (2007) test statistic $\tau^*(\ell)_{VL} = \ddot{\gamma}(\ell)/\sqrt{\mathrm{V}[\ddot{\gamma}(\ell)]}$ to test the significance of cross-covariance of the noises. Their t-statistic uses an approximated variance of the cross-covariance estimator under i.i.d. noise. The approximated variance is given by:

$$\mathrm{V}[\ddot{\gamma}(\ell)] \approx \begin{cases} \frac{1}{(S_1 S_2)^2} \sum_{b=1}^{S_1 S_2} \frac{1}{N_{\ell,b}} (v_x + 5\sigma_\eta^2 \sigma_\delta^2) & \text{for} \quad \ell = 0, \\ \frac{1}{(S_1 S_2)^2} \sum_{b=1}^{S_1 S_2} \frac{1}{N_{\ell,b}} (v_x + 4\sigma_\eta^2 \sigma_\delta^2) & \text{for} \quad \ell \neq 0, \end{cases} \tag{26}$$

where $v_x := \sigma_{x\ sec,1}^2 \sigma_{x\ sec,2}^2 + 2\sigma_\eta^2 \sigma_{x\ sec,1}^2 + 2\sigma_\delta^2 \sigma_{x\ sec,2}^2$, $\sigma_{x\ sec,l}^2 = \frac{\int_0^1 \sigma_1^2(s) ds}{23400/x}$ is the x-second average integrated variance of each asset l. As described in Voev and Lunde (2007), the noise variances σ_η^2 and σ_δ^2 are estimated by subsampling realized variance with one-minute returns and dividing by twice the number of returns. And the integrated variance is estimated by a realized kernel-based estimator of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008). On the other hand, the test statistic derived in this paper is constructed by the subsampling variance estimator which does not require the i.i.d. noise approximation.

The proposed cross and auto covariance estimators (11) and (19) require the threshold value of noise dependence in advance. First we investigate an accuracy of the threshold value selected by the test statistic (9) and (16). The bivariate MA(2) noise process has the true threshold values m^+, m^- and m_2 being 2 tick times ($|\ \ell\ |= 2$). As long as the selected threshold value is not less than 2 ticks, the proposed covariance estimators $\hat{\gamma}(\ell)$ and $\hat{\gamma}_\delta(\ell)$ are unbiased. Figure 4 plots the probability of miss selection that the estimated threshold values $\hat{m}^+$, $\hat{m}^-$ and $\hat{m}_2$ are less than 2 ticks, where the number of repetitions is one thousand. Although the probability of miss selection with $c = 0.5$ is the lowest of all across the sampling interval, the differences of the probability between $c = 0.5$ and the others are small. And the probabilities of miss selection approach to zero as the time interval becomes shorter (the sample size increases). The influence caused by the miss selection of threshold values are reported in Tables 1 and 2. In case of the sample size = 770 (sampling interval = 30 seconds) in Table 1, we find that $\hat{\gamma}(\ell)_{[c]}$ and $\hat{\gamma}_\delta(\ell)_{[c]}$ with any c have a somewhat bias by the miss selection of the threshold value. However, the bias of the estimators with the higher trading intensity like 10 seconds becomes close to zero because the probability of the miss selection of the threshold value is quite low. In case of the low trading intensity, we use the sample over multiple days (for example 3 days) to obtain enough sample size. The fourth panel in Table 1 shows the biases of $\hat{\gamma}(\ell)_{[c]}$ and $\hat{\gamma}_\delta(\ell)_{[c]}$ using sample over 3 days with the sampling interval being 30 seconds are smaller than those in case of sample over 1 day (the third panel in Table 1). The behaviors of the proposed cross and auto covariance estimators are much improved by the use of sample over multiple days even in the low trading intensity.

$\tilde{\gamma}(\ell)$ and $\tilde{\gamma}_\delta(\ell)$ with $\ell = 0, 1$ are biased as mentioned above. The bias of Voev and Lunde's (2007) cross-covariance estimator $\ddot{\gamma}(\ell)$ is virtually zero for all cases because it is constructed by the subsampling returns with the intervals being wide enough. However, RMSE of $\ddot{\gamma}(\ell)$ is larger than $\hat{\gamma}(\ell)_{[c]}$ with any c and RMSE ratio of $\ddot{\gamma}(\ell)$ to $\hat{\gamma}(\ell)_{[1]}$ is more than 1 as summarized in Table 2. The large RMSE of $\ddot{\gamma}(\ell)$ comes from a large standard deviation induced by the longer intervals for construction of $\ddot{\gamma}(\ell)$ relative to that for $\hat{\gamma}(\ell)_{[c]}$.

The large standard deviation of the covariance estimator also makes an impact on behavior of the cross and auto correlations of the noises. We compare the cross-correlations computed by the proposed variance and covariance estimators; that is $\hat{\rho}(\ell) = \hat{\gamma}(\ell)_{[1]}/\sqrt{\hat{\gamma}_\eta(0)_{[1]}\hat{\gamma}_\delta(0)_{[1]}}$, and the method introduced in Voev and Lunde (2007). Voev and Lunde's (2007) cross-correlation is computed by $\ddot{\rho}(\ell) = \ddot{\gamma}(\ell)/(\hat{\sigma}_\eta\, \hat{\sigma}_\delta)$ where $\hat{\sigma}_\eta^2$ and $\hat{\sigma}_\delta^2$ are estimates of the noise variances by subsampling realized variance with one-minute returns and dividing by twice the number of returns. Figures 5 and 6 show empirical distributions of the cross-correlations in cases of the true cross-correlations $\rho(0) = 0.87$ and $\rho(2) = 0.26$. We find that the cross-correlation estimate $\hat{\rho}(\ell)$ have a higher degree of concentration around the true correlation value than that of $\ddot{\rho}(\ell)$ in all sampling intervals.

Table 3 summarizes the size and power of test statistics for significance of cross and auto covariance estimators as proposed in (14), (21), (22) and Voev and Lunde (2007). The size of the proposed test statistic takes around the nominal size, but the size of Voev and Lunde' (2007) test statistic $\tau^*(\ell)_{VL}$ is large. Figure 7 shows asymptotic distribution (standard normal) and empirical distributions of $\hat{\gamma}(\ell)_{[1]}$ and $\ddot{\gamma}(\ell)$ which are standardized by $\hat{\omega}_\ell$ and $\sqrt{\hat{V}[\ddot{\gamma}(\ell)]}$. Although the standardized empirical distribution of $\hat{\gamma}(\ell)_{[1]}$ is close to the standard normal distribution, that of $\ddot{\gamma}(\ell)$ is heavily tailed. The proposed test statistic also has a larger power as the sample size increases.

Next we summarize the influence of M_ℓ and K_ℓ selected at fixed N_ℓ under $M_\ell = cN_\ell^{1/3}$ on the variance estimation. Although the value of c should be determined in accordance with the dependence of the noises, we do not know about the dependence in advance. Even in this simulation setting, it is not easy to select the optimal value of c. In general, when noise processes have more dependence, the value of c should be large. The behavior of the covariance estimators and the test statistics would be also influenced by the value of c. However, we find that the performance of the proposed statistics with any c does not change so much as we have seen in Tables 1, 2 and 3.

So far, we have assumed that the equilibrium price follows an Itô process without drift as in (1). Now we consider the equilibrium price follows an Itô process with drift. The product of the noises dominates that of the drift term as the length of the interval shrinks to zero. However, it is not clear whether the influence of the drift is negligible because the estimators and test statistics proposed in this paper are based on the expanded intervals in (10) and (18). In this simulation, we add a drift term to the stochastic differential equations:

$$dP_l^*(t) = \mu_l dt + \sigma_l(t)\Big[\sqrt{1-\lambda_l^2}dW_l^{(A)}(t) + \lambda_l dW_l^{(B)}(t)\Big], \quad l = 1,2, \tag{27}$$

where μ_l is a constant and (μ_1, μ_2) is set as $(0.1, 0.1)$, $(0.5, 0.5)$ and $(1, 1)$. We estimate $\hat{\gamma}(\ell)_{[1]}$ and $\hat{\gamma}_\delta(\ell)_{[1]}$ under the above settings. Table 4 represents the bias and RMSE ratio of them to $\hat{\gamma}(\ell)_{[1]}$ and $\hat{\gamma}_\delta(\ell)_{[1]}$, which are estimated using the data generation process without drift as in (22). The biases of the estimators are quite small and the RMSE ratios take values of around one, even in the case of rapid trend acceleration where $(\mu_1, \mu_2) = (1, 1)$. We find that the cross and auto covariance estimators of the noises are not substantially affected by the presence of the nonzero drift term.

5.1.2 Bivariate autoregressive noise

We have investigated the behaviors of covariance estimators of moving average noise processes and the test statistics for finding the threshold values (9), (16) and the significance (14), (21), (22) under the regular and synchronous sampling. In the next experiment, we consider the irregular and non-synchronous sampling where the average observed time interval for asset 1 and 2 are 10 and 5 seconds because this sampling scheme is more realistic for the actual trading system. And we employ bivariate autoregressive noise processes whose autocorrelation decays gradually to zero, but is not truncated at some finite lag. Although Assumption (1b) is approximately valid, it is important to investigate the finite sample properties of the covariance estimators of autoregressive noises. We generate the noises $\eta(t)$ and $\delta(t)$ by the following bivariate AR(1) model every one second,

$$\begin{aligned}\eta(t) &= 0.6\,\eta(t-1) + 0.2\,\delta(t-1) + \epsilon_\eta(t),\\ \delta(t) &= 0.2\,\eta(t-1) + 0.6\,\delta(t-1) + \epsilon_\delta(t).\end{aligned}$$

After the summing of equilibrium price and noise at the same period, the observed price is randomly selected by Poisson sampling with mean durations for asset 1 and 2 being 10 and 5 seconds.

In the irregular and non-synchronous sampling, $\ell = 1$ represents one second. With these simulation settings, we obtain, on average, 900 realizations of sample size N_ℓ with $\ell = 0$. On the other hand, the average sample size of N_ℓ with $| \ell |> 0$ is 450. Then the corresponding numbers of M_ℓ and K_ℓ with some c are equal to (c, M_ℓ, K_ℓ) =$(4, 30, 15)$, $(2, 15, 30)$, $(1, 8, 45)$ and $(0.5, 4, 110)$.

Tables 5 and 6 summarize the bias and RMSE of the proposed cross and auto covariance estimators $\hat{\gamma}(\ell)_{[c]}$, $\hat{\gamma}_\delta(\ell)_{[c]}$ and the Voev and Lunde's (2007) cross-covariance estimator $\ddot{\gamma}(\ell)$. For the cross-covariance estimators, the biases of $\hat{\gamma}(\ell)_{[c]}$ and $\ddot{\gamma}(\ell)$ are virtually zero, and $\hat{\gamma}(\ell)_{[c]}$ has a smaller RMSE than $\ddot{\gamma}(\ell)$ for all ℓ. The smaller RMSE of $\hat{\gamma}(\ell)_{[c]}$ comes from the shorter intervals for construction of $\hat{\gamma}(\ell)_{[c]}$ relative to that for $\ddot{\gamma}(\ell)$. For the autocovariance estimators, the biases are close enough to zero and the RMSEs are the almost the same value for all c. [1]

Figure 8 shows the empirical distributions of cross-correlation and autocorrelaiton estimates $\hat{\rho}(\ell) = \hat{\gamma}(\ell)_{[1]}/\sqrt{\hat{\gamma}_\eta(0)_{[1]}\hat{\gamma}_\delta(0)_{[1]}}$ and $\hat{\rho}_\delta(\ell) = \hat{\gamma}_\delta(\ell)_{[1]}/\hat{\gamma}_\delta(0)_{[1]}$. We find that the correlation estimates take around the true correlation. Figures 9 and 10 show the standardized distribution of the cross and auto covariance estimators $\hat{\gamma}(\ell)_{[c]}$ and $\hat{\gamma}_\delta(\ell)_{[c]}$ with each value of c. The average sample size within each subseries M_ℓ is 4 when c is equal to 0.5. A few standardized empirical distributions have heavy tail caused by small M_ℓ and approach to the standard normal as c increases. In the empirical application, we should select the value of c not to be M_ℓ too small. However, most histograms could be approximated by the standard normal distribution and the influences by the different values of c are very small among the proposed statistics. Therefore, these simulation results show that the proposed covariance estimator and test statistics have a good performance and are robust for selection of c even in case of autoregressive noises.

5.2 Empirical illustration

We have established the covariance estimators (11) and (19) of bivariate noise processes and the test statistics (9) and (16) for finding the threshold values and the significance (14), (21), (22) in sections 3 and 4. In this subsection we provide the details of implementing them in practice. We

[1]It is noted that the RMSE of $\hat{\gamma}_\delta(\ell)$ for $\ell = 1, 2$ takes somewhat larger values than that of $\hat{\gamma}_\delta(\ell)$ for the other ℓ. This is caused by smaller $N_{2,\ell}$ for $\ell = 1, 2$; that is, a sample size for asset 2. In this experiment, $N_{2,1}$ and $N_{2,2}$ are about $1/6$ and $1/2$ of $N_{2,\ell}$ for the other ℓ because the average observed time interval of asset 2 is set as 5 seconds.

illustrate a scheme of whole procedure for the estimation and testing for dependence of microstructure noise as follows:

- **Univariate noise process**
  - **Threshold value step:** The threshold value m_1 of univariate noise dependence is statistically selected through the test statistic (16).
  - **Estimation step:** Estimate the variance and the autocovariance $\gamma_\eta(\ell)$ of the noise with $\ell \geq 0$ using (19).
  - **Test significance step:** Test the null hypotheses $\gamma_\eta(0) = 0$ by (21) and $\gamma_\eta(\ell) = 0$ by (22).
- **Bivariate noise processes**
  - **Threshold value step:** The threshold values of cross-sectional noise dependence m^+ and m^- are statistically selected through the test statistic (9).
  - **Estimation step:** Estimate the cross-covariance $\gamma(\ell)$ of the noises using (11).
  - **Test significance step:** Test the null hypothesis $\gamma(\ell) = 0$ by (14).

We apply these statistics to high-frequency transaction prices of four stocks on the Osaka Securities Exchange: OMRON Corporation (OC) and Murata Manufacturing Co., Ltd. (MM) from manufacturing industry, Ono Pharmaceutical Co., Ltd. (OP) and Santen Pharmaceutical Co., Ltd. (SP) from medical industry. The trading day on the Osaka Securities Exchange is divided into morning (9:00-11:00 local time) and afternoon (12:30-15:10) sessions. As illustrated in Andersen, Bollerslev and Cai (2000), transaction prices at the opening and closing are more volatile than any other transaction prices. This is due to the call auction to determine the opening and closing prices. In this empirical illustration, we use the transaction prices during from 9:05 to 15:05 to estimate the cross and auto covariances of the noises.

For univariate noise process, 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 (16) for the identification of m_1 as described in section 4.2. When the total number of $\{Z_{1,\ell,k}\}$ for each ℓ is not large enough for the inference, we only have to use samples in multiple days to obtain a large sample size. In the case of high-frequency data for D days, we define $n_\ell^{(d)}$ as the total number of $\{Z_{1,\ell,k}\}$ on the d-th day where $d = \{1, \cdots, D\}$. We set $D = 245$ where the sample period is from October 2, 2006 to September 28, 2007. We obtain $\{Z_{1,\ell,1}, \cdots, Z_{1,\ell,n_\ell^{(1)}}\}$ on the first day and $\{Z_{1,\ell,n^{(1)}+1}, \cdots, Z_{1,\ell,n_\ell^{(1)}+n_\ell^{(2)}}\}$ on the second day. The total number of $\{Z_{1,\ell,k}\}$ in the sample period is $N_{1,\ell} = \sum_{d=1}^{D} n_\ell^{(d)}$. We set $L = 70$ seconds. If the null $\mathrm{E}_{\mathrm{IJ}}[Z_{1,70,k}] = 0$ is not rejected, we test the null $\mathrm{E}_{\mathrm{IJ}}[Z_{1,69,k}] = 0$. We continue to test sequentially until the null is rejected. Finally, we regard the distance where the null is rejected the first time as the estimator of threshold value m_1.

Next we estimate the variance of noise $\gamma_\eta(0)$ by (19) and test the null hypothesis $\gamma_\eta(0) = 0$ using (21). If the test statistic (21) 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 (22). The significance level is set at 0.05. Table 7 (a) shows the variance estimate of the noise, the test statistic (21) for its significance and the noise-to-signal ratio which is the ratio between the variance estimate of the noise and the integrated variance estimate using a realized kernel-based estimator of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008). We confirm that the variance of market microstructure noise in each asset is significantly larger than zero. Table 7 (b) shows the test statistic (22) for the significance of the autocovariance. The dependence of the noise in OC, MM, OP and SP disappear at around 60 seconds. Figure 11 plots four autocorrelation functions $\hat{\rho}_\eta(\ell) = \hat{\gamma}_\eta(\ell)/\hat{\gamma}_\eta(0)$. We can see significantly negative and positive autocorrelations of microstructure noise in OC, MM and SP. The negative autocorrelations at small lags imply that there exists the opposite orders within small period that buying (selling) at present follows selling (buying) at small lags. On the other hand, the positive autocorrelations at higher lags would be induced by the clustering of order flow which occurs in case of buying or selling pressure. These features are confirmed in Bandi and Russell (2006).

For bivariate noise processes, we use the test statistic (9) for the identification of m^+ and m^-, estimate the cross-covariance by (11), and judge the significance of the cross-covariance through the test statistic (14). The results are reported in Table 8. In both pairs between the same industries and between different industries, there exist cross-covariances significantly different from zero within $-33 \le \ell \le 66$ for OC-MM, $-57 \le \ell \le 47$ for OC-OP, $-62 \le \ell \le 0$ for OC-SP, $-48 \le \ell \le 68$ for MM-OP, $-34 \le \ell \le 68$ for MM-SP, and $-59 \le \ell \le 42$ for OP-SP, respectively. Figure 12 plots the six cross-correlation functions $\hat{\rho}(\ell) = \hat{\gamma}(\ell)/\sqrt{\hat{\gamma}_\eta(0)\hat{\gamma}_\delta(0)}$. These plots show that the market microstructure noises among some assets display asymmetric and different cross-sectional dependence patterns. Especially the asymmetric cross-correlation in OC and SP captures their lead-lag relationship. Although we do not go into the details, we find that the proposed test statistics and cross and auto covariance estimators of the bivariate noise processes provide valuable insights for the analysis of market microstructure.

6 Concluding remarks

Market microstructure noise occurs in the market with trading imperfections, frictions and informational effects. In this paper, we propose unbiased and consistent estimators of cross and auto covariances of the noises and derive the asymptotic distributions of them. Through Monte Carlo simulation, we find that the proposed estimator has small mean squared error. For the cross-covariance estimation, our estimator has smaller mean squared error than Voev and Lunde's (2007). The larger mean squared error of Voev and Lunde's comes from a large standard deviation induced by a too wide time interval for construction of their cross-covariance estimator. Our estimator is based on the interval with proper length such that the dependence of the noise disappears. The length is determined by testing procedure proposed in this paper. Furthermore, we also propose the test statistics for the significance of the cross and auto covariances of the noises. We confirm that our test statistics have good empirical size and power properties.

The empirical illustration confirms that the proposed statistics enable to capture various noise dependence patterns in several assets. The statistical analysis of market microstructure noise would provide some evidence on the influence of market regularity and the trading mechanism on asset pricing in financial markets. For that reason, the proposed method will shed more light on market microstructure analysis.

Appendix

The proof of Lemma 1.

Let the start 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 theorems for finite dependence, as in 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 more general dependence cases like the mixing sequence. Assumption 2 implies that the conditional variance of a standardized sample mean of $\{Z_{\ell,k'+1}, \ldots, Z_{\ell,k'+n}\}$ for any k' approaches 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 (C1) for the application of their central limit theorem.

(C1) $\mathrm{E_{IJ}}|Z_{\ell,k}|^{2\beta} < \infty$, 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:

$$\begin{aligned} Z_{\ell,k} &= \big(P_1(\overline{A_k}) - P_1(\underline{A_k})\big)\big(P_2(\overline{B_k}) - P_2(\underline{B_k})\big) = \int_{A_k} \sigma_1(u)dW_1(u) \int_{B_k} \sigma_2(u)dW_2(u) \\ &+ \int_{A_k} \sigma_1(u)dW_1(u)\Delta\delta(B_k) + \int_{B_k} \sigma_2(u)dW_2(u)\Delta\eta(A_k) + \Delta\eta(A_k)\Delta\delta(B_k). \end{aligned} \tag{28}$$

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:

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

Because A_k and B_k are nonoverlapping, the high-order absolute moment of the first term in (28):

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

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

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

where $\parallel\delta(\cdot)\parallel_{2\beta}=\big(\mathrm{E_{IJ}}\big|\delta(\cdot)\big|^{2\beta}\big)^{\frac{1}{2\beta}}$. From (1d) in Assumption 1, the high-order absolute moment of the second term in (28) has:

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

For the third term of (28), $\mathrm{E_{IJ}}\Big|\int_{B_k}\sigma_2(u)dW_2(u)\Delta\eta(A_k)\Big|^{2\beta}<\infty$. From (1c) in Assumption 1, the high-order absolute moment of the fourth term of (28) has:

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

Finally, we have

$$\begin{aligned}\mathrm{E_{IJ}}\big|Z_{\ell,k}\big|^{2\beta}\ &\leq\Big(\parallel\int_{A_k}\sigma_1(u)dW_1(u)\int_{B_k}\sigma_2(u)dW_2(u)\parallel_{2\beta}+\parallel\int_{A_k}\sigma_1(u)dW_1(u)\Delta\delta(B_k)\parallel_{2\beta}\\&\quad+\parallel\int_{B_k}\sigma_2(u)dW_2(u)\Delta\eta(A_k)\parallel_{2\beta}+\parallel\Delta\eta(A_k)\Delta\delta(B_k)\parallel_{2\beta}\Big)^{2\beta}<\infty.\end{aligned}\tag{29}$$

Condition (C1) holds. We then 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 (7), we apply L_2-convergence of the subsampling estimator given by Lemma 4.6.1 in Politis, Romano and Wolf (1999). Because the strong mixing condition holds from (1b) in Assumption 1, it suffices 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.

The conditional expectation of $\hat{\gamma}(\ell)$ is:

$$\mathrm{E}_{\mathrm{IJ}}[\hat{\gamma}(\ell)]=-\frac{1}{N_\ell}\sum_{k=1}^{N_\ell}\mathrm{E}_{\mathrm{IJ}}[Z_{\ell,k}^{(\pm)}]=\gamma(\ell). \tag{30}$$

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

$$\begin{aligned}\mathrm{V}_{\mathrm{IJ}}[\hat{\gamma}(\ell)]&=\mathrm{V}_{\mathrm{IJ}}\Big[-\frac{1}{N_\ell}\sum_{k=1}^{N_\ell}Z_{\ell,k}^{(\pm)}\Big]=\frac{1}{N_\ell^2}\sum_{k=1}^{N_\ell}\mathrm{V}_{\mathrm{IJ}}\big[Z_{\ell,k}^{(\pm)}\big]+\frac{2}{N_\ell^2}\sum_{k=1}^{N_\ell}\sum_{j=1}^{N_\ell-k}\mathrm{Cov}_{\mathrm{IJ}}\big[Z_{\ell,k}^{(\pm)},Z_{\ell,k+j}^{(\pm)}\big]\\&\le\frac{1}{N_\ell^2}\sum_{k=1}^{N_\ell}\max_k\Big\{\mathrm{V}_{\mathrm{IJ}}\big[Z_{\ell,k}^{(\pm)}\big]\Big\}+\frac{2}{N_\ell^2}\sum_{k=1}^{N_\ell}\sum_{j=1}^{\tilde{m}_k}\max_k\Big\{\Big|\mathrm{Cov}_{\mathrm{IJ}}\big[Z_{\ell,k}^{(\pm)},Z_{\ell,k+j}^{(\pm)}\big]\Big|\Big\}=O\Big(\frac{1}{N_\ell}\Big),\end{aligned} \tag{31}$$

where $\tilde{m}_k$ is defined as $\max_j\{\text{Cov}_{\text{IJ}}[Z^{(\pm)}_{\ell,k}, Z^{(\pm)}_{\ell,k+j}] \neq 0, 0 < j \leq N_\ell - k\}$. Because of the finite dependence of $\{Z^{(\pm)}_{\ell,k+j}\}_{k=1}^{N_\ell}$, $\tilde{m}_k$ is finite. This implies $\text{V}_{\text{IJ}}[\hat{\gamma}(\ell)] \to 0$ as N_ℓ goes to infinity and the consistency of $\hat{\gamma}(\ell)$ holds. Let the asymptotic variance of $\hat{\gamma}(\ell)$ be $\omega_\ell^2 = \lim_{N_\ell\to\infty} \text{E}_{\text{IJ}}\Big[\big\{N_\ell^{1/2}(\hat{\gamma}(\ell) - \gamma(\ell))\big\}^2\Big]$. We find that the asymptotic normality of $\hat{\gamma}(\ell)$ can be proved by a similar argument to the proof of Lemma 1. The difference between $\{Z^{(\pm)}_{\ell,k}\}_{k=1}^{N_\ell}$ and $\{Z_{\ell,k}\}_{k=1}^{N_\ell}$ is the amount of dependence. A sequence of $\{Z^{(\pm)}_{\ell,k}\}_{k=1}^{N_\ell}$ has more dependence than $\{Z_{\ell,k}\}_{k=1}^{N_\ell}$ because $Z^{(\pm)}_{\ell,k}$ 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, $Z^{(\pm)}_{\ell,k}$ has finite dependence from (1b) and (1d) in Assumption 1. $\square$

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. We also thank Center for the Study of Finance and Insurance, Osaka University and Osaka Securities Exchange for providing the high-frequency data used in this paper.

Funding

Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Special Purposes, 90254135, 2007, for Scientific Research (A), 18203013, 2007 and for Young Scientists (Start-up), 20830047, 2008.

References

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

Andersen, Torben G., Tim Bollerslev and Jun Cai. (2000). ”Intraday and interday volatility in the Japanese stock market.” *Journal of International Financial Markets*, 10(2), 107-130.

Bandi, Federico M., and Jeffrey R. Russell. (2006). ”Full-information transaction costs.” Working paper, Graduate School of Business, University of Chicago.

Bandi, Federico M., and Jeffrey R. Russell. (2008). ”Microstructure Noise, Realized Variance, and Optimal Sampling.” *Review of Economic Studies*, 75(2), 339-369.

Barndorff-Nielsen, Ole E., Peter R. Hansen, Asger Lunde and Neil Shephard. (2008). ”Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise.” *Econometrica*, forthcoming.

Carlstein, Edward. (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.

Copeland, Thomas E., and Dan. Galai. (1983). ”Information Effects on the Bid-Ask Spread.” *Journal of Finance* 38(5), 1457-1469.

Drost, Feike C., and Bas J.M. Werker. (1996). ”Closing the GARCH gap: Continuous time GARCH modeling.” *Journal of Econometrics* 74(1), 31-57.

Easley, David, and Maureen O’Hara. (1987). ”Price, trade size, and information in securities markets.” *Journal of Financial Economics* 19(1), 69-90.

Foucault, Thierry. (1999). ”Order flow composition and trading costs in a dynamic limit order market.” *Journal of Financial Markets* 2(2), 99-134.

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

George, Thomas J, Gautam Kaul and M. Nimalendran. (1991). ”Estimation of the Bid-Ask Spread and its Components: A New Approach.” *The Review of Financial Studies* 4(4), 623-656.

Glosten, Lawrence R. (1994). ”Is the Electronic Open Limit Order Book Inevitable?.” *Journal of*

Finance 49(4), 1127-1161.

Glosten, Lawrence R., and Lawrence E. Harris. (1988). ”Estimating the components of the bid-ask spread.” *Journal of Financial Economics* 21(1), 123-142. Glosten, Lawrence R., and Paul R. Milgrom. (1985). ”Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders.” *Journal of Financial Economics* 14(1), 71-100.

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

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

Harris, Lawrence. (1990). ”Estimation of Stock Price Variances and Serial Covariances from Discrete Observations.” *The Journal of Financial and Quantitative Analysis* 25, 291-306.

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

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

Huang, Roger D., and Hans R. Stoll. (1997). ”The Components of the Bid-Ask Spread: A General Approach.” *The Review of Financial Studies* 10(4), 995-1034.

Madhavan, Ananth, Matthew Richardson and Mark Roomans. (1997). ”Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks.” *The Review of Financial Studies* 10(4), 1035-1064.

Meddahi, Nour. (2002). ”A Theoretical Comparison Between Integrated And Realized Volatility.” *Journal of Applied Econometrics* 17(5), 479-508.

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

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

Politis, Dimitris N., Joseph P. Romano, and Michael Wolf. (1997). ”Subsampling.” Springer Series in Statistics.

Roll, Richard. (1984). ”A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market.” *The Journal of Finance* 39(4), 1127-1139.

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

Stoll, Hans R. (1989). ”Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests.” *The Journal of Finance* 44(1), 115-134.

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

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

Zhang, Lan, Per A. Mykland and Yacine 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, Bin. (1996). ”High-Frequency Data and Volatility in Foreign-Exchange Rates.” *Journal of Business and Economic Statistics*, 14, 45-52.

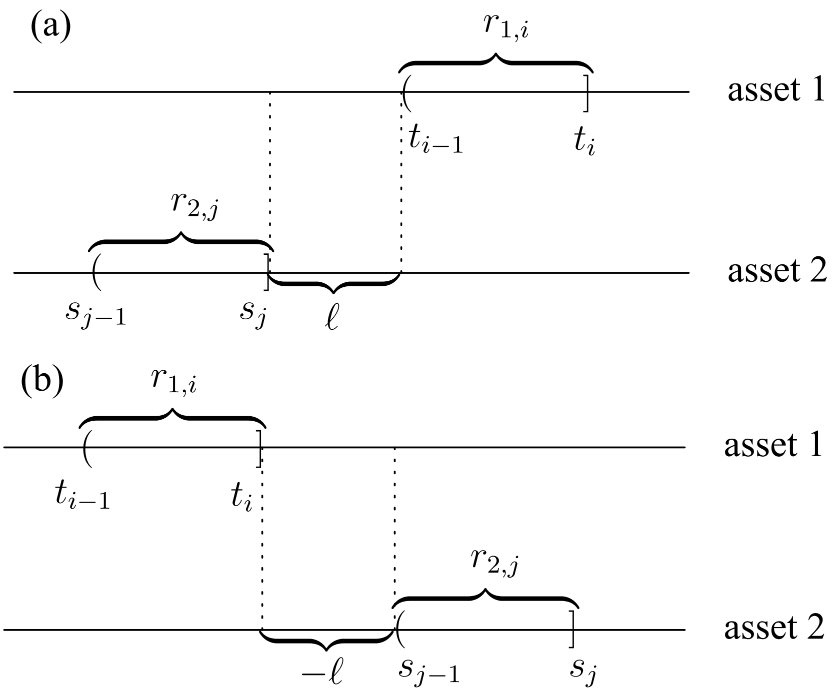

Figure 1: Pairs of returns on nonoverlapping intervals with (a) $\ell > 0$ and (b) $\ell < 0$.

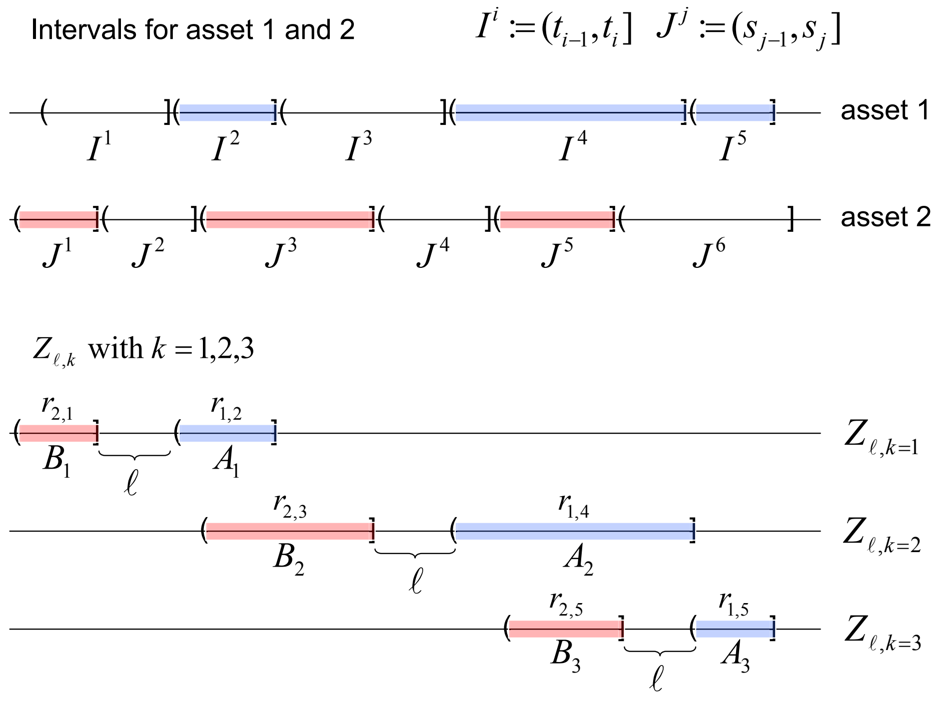

Figure 2: An example of each pair of the intervals (A_k, B_k) for $k = 1, 2, 3$.

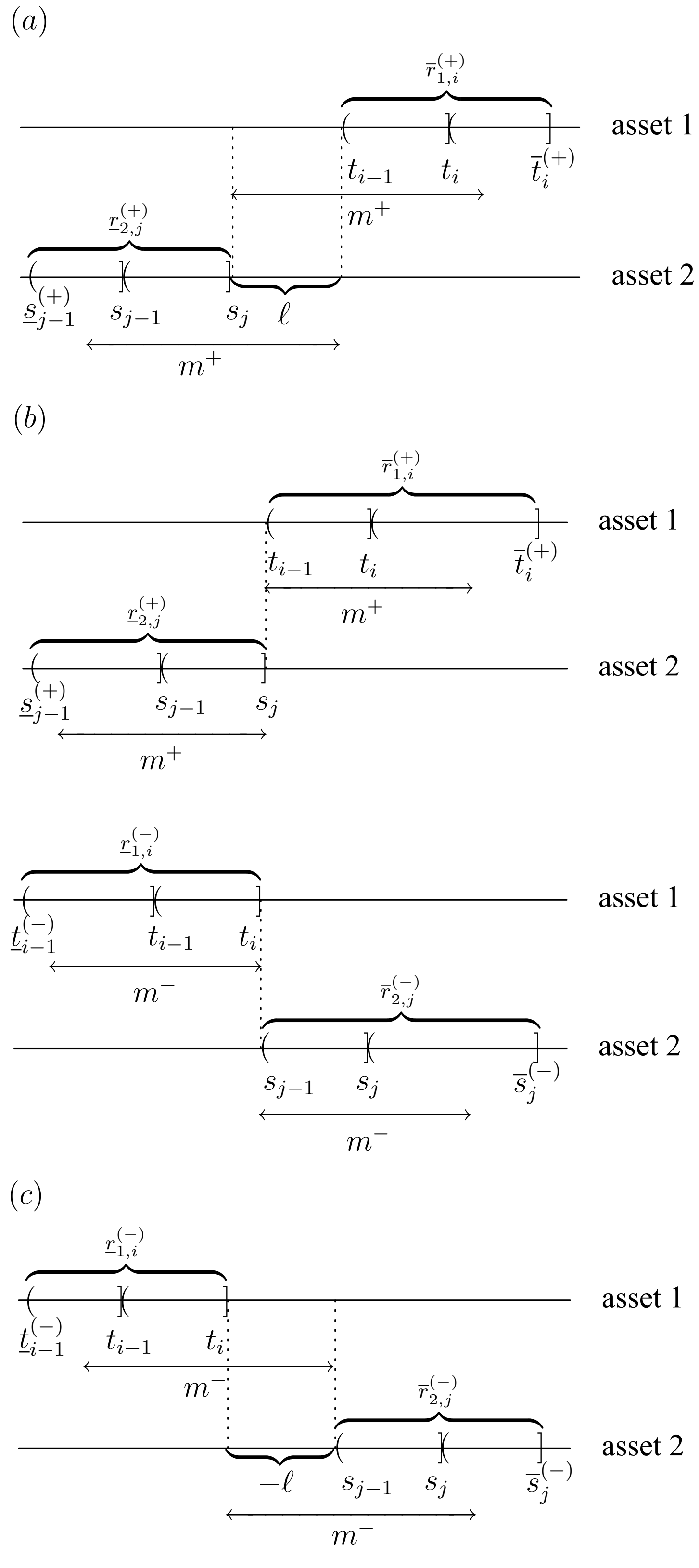

Figure 3: A pair of modified intervals in cases of (a) $\ell > 0$, (b) $\ell = 0$ and (c) $\ell < 0$.

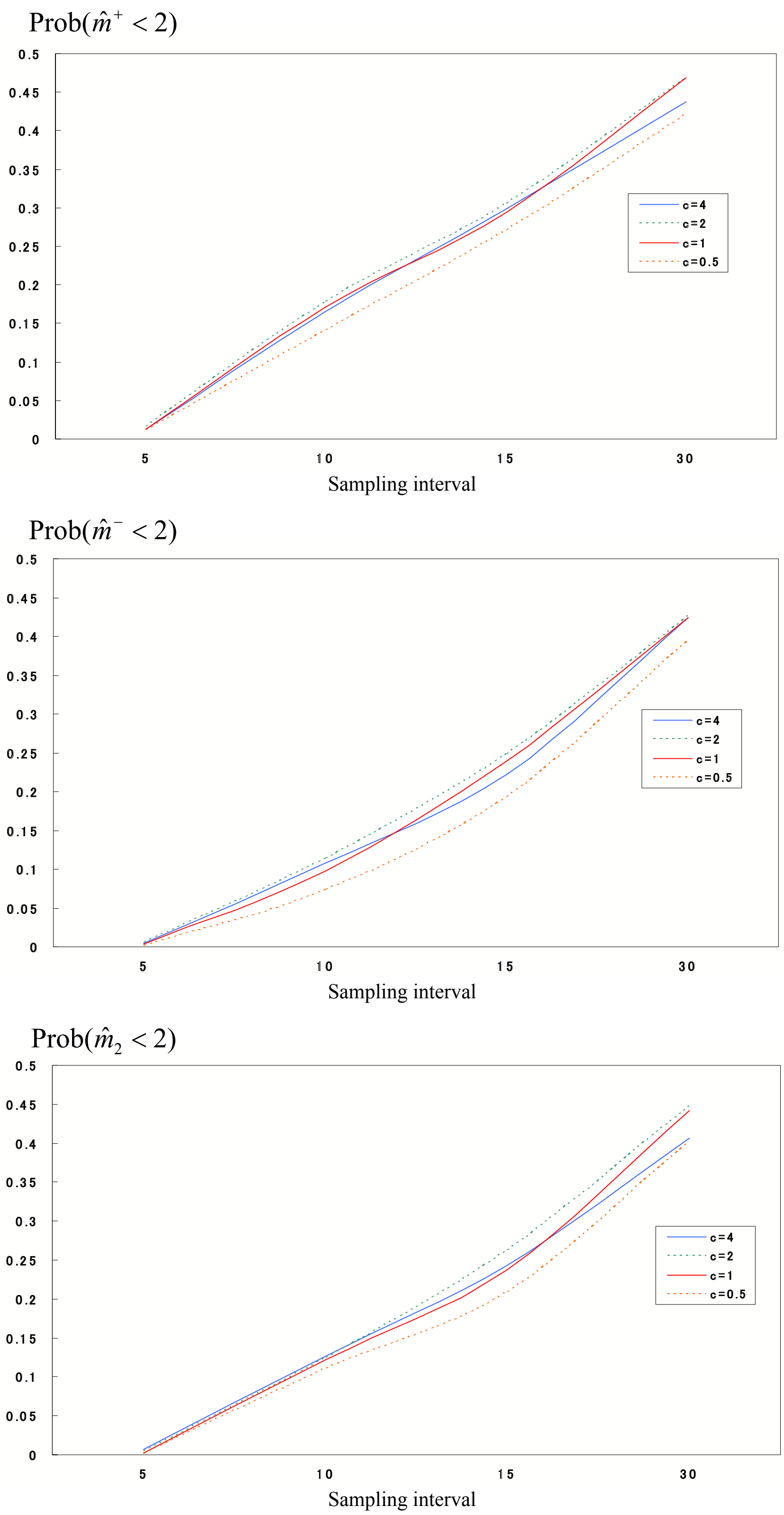

Figure 4: Probability of the estimated threshold value being less than 2.

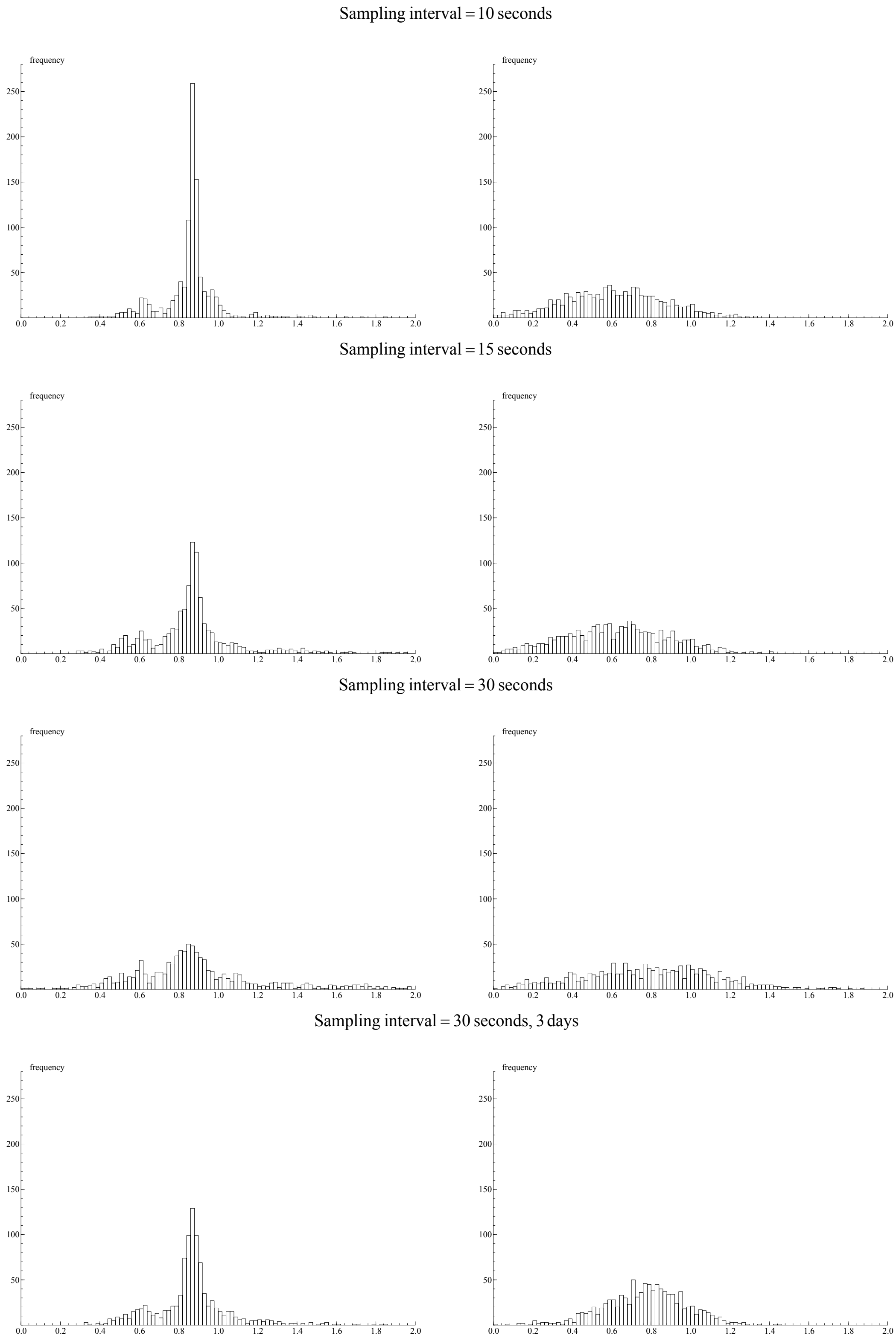

Figure 5: Empirical distributions of $\hat{\rho}(0)$ (left side) and $\ddot{\rho}(0)$ (right side), where true cross-correlation $\rho(0) = 0.87$.

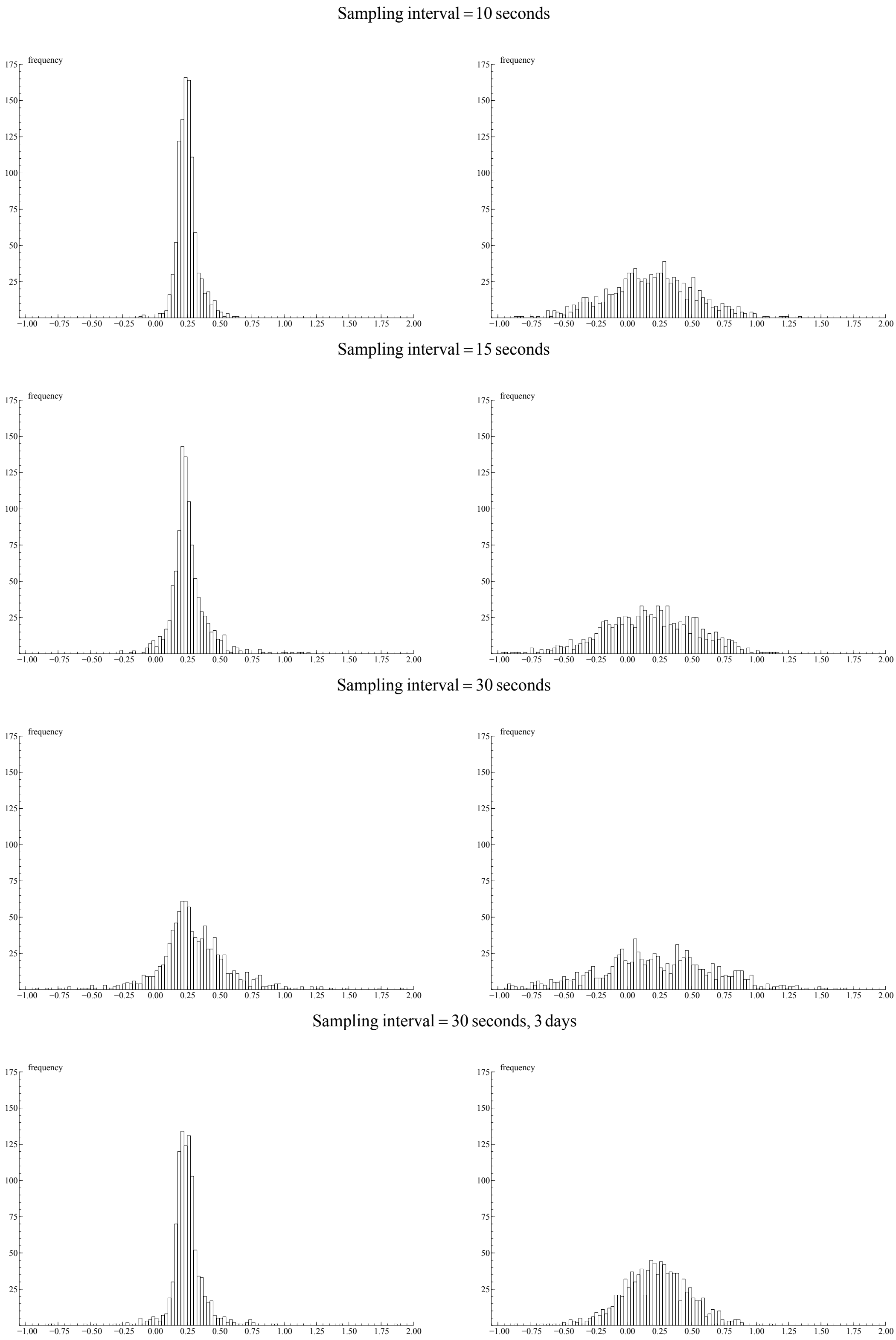

Figure 6: Empirical distributions of $\hat{\rho}(2)$ (left side) and $\ddot{\rho}(2)$ (right side), where true cross-correlation $\rho(2) = 0.26$.

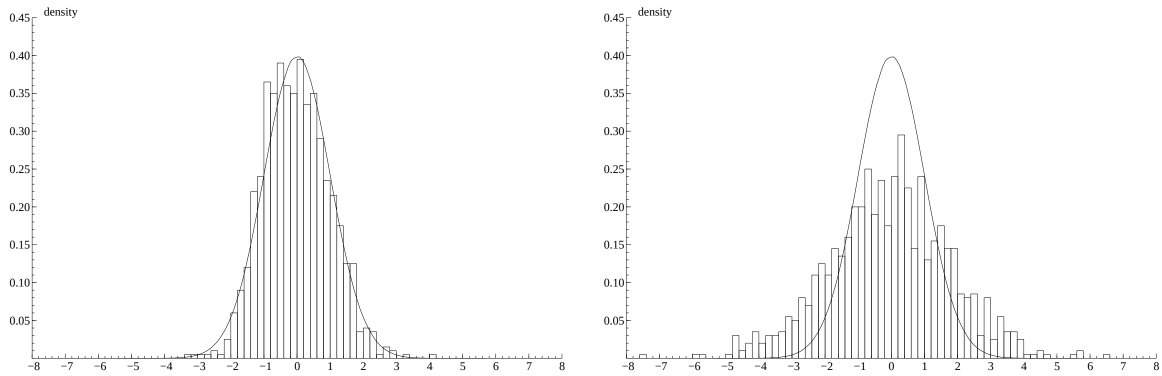

Figure 7: Standardized empirical distributions of $\hat{\gamma}(\ell)$ (left side) and $\ddot{\gamma}(\ell)$ (right side), and the standard normal distributions (solid line).

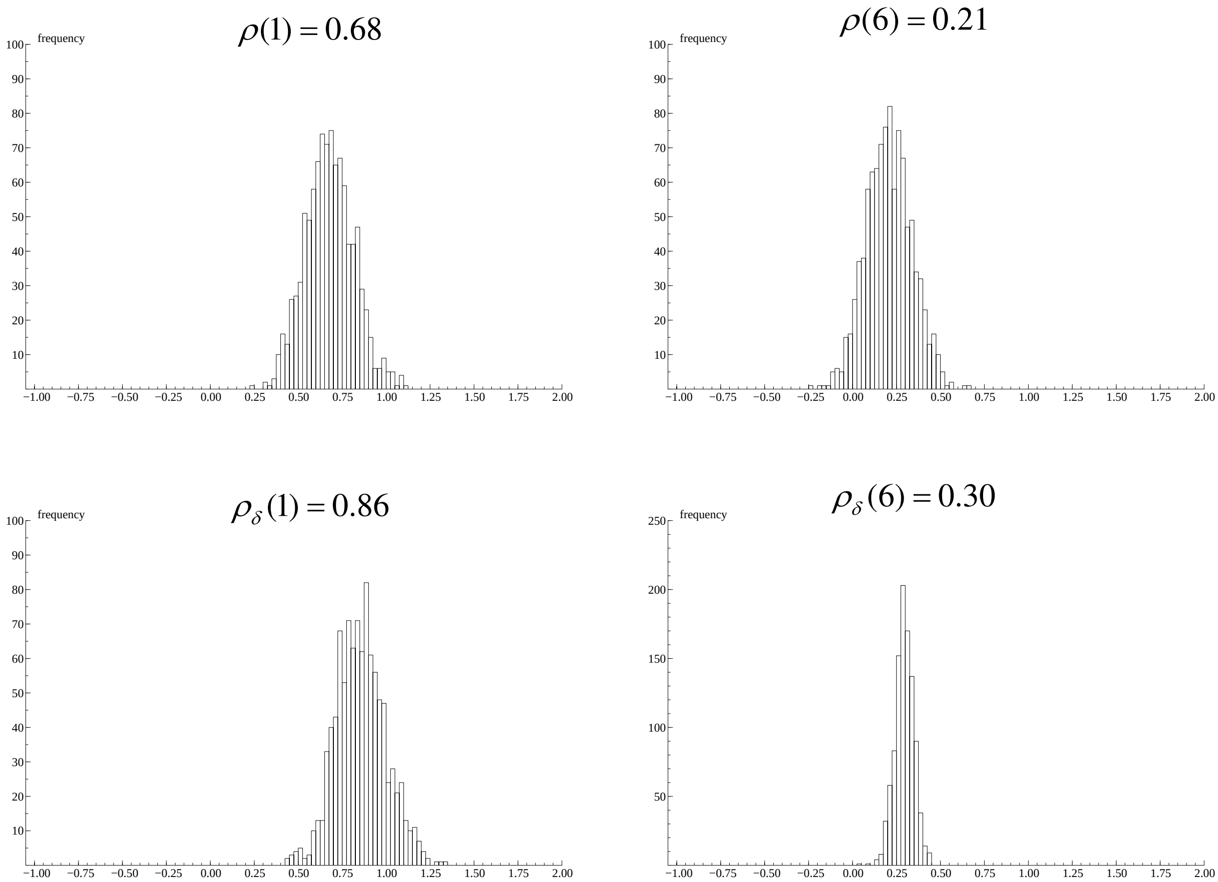

Figure 8: Empirical distributions of $\hat{\rho}(\ell)$ and $\hat{\rho}_\delta(\ell)$.

Note: In the simulation for autocovariance estimator of univariate noise process, the sample sizes of $N_{2,1}$ and $N_{2,2}$ are about $1/6$ and $1/2$ of $N_{2,\ell}$ for the other ℓ because the average observed time interval of asset 2 is set as 5 seconds. Empirical distributions of $\hat{\rho}_\delta(\ell)$ is from sample over 3 days to obtain a large sample size $N_{2,1}$.

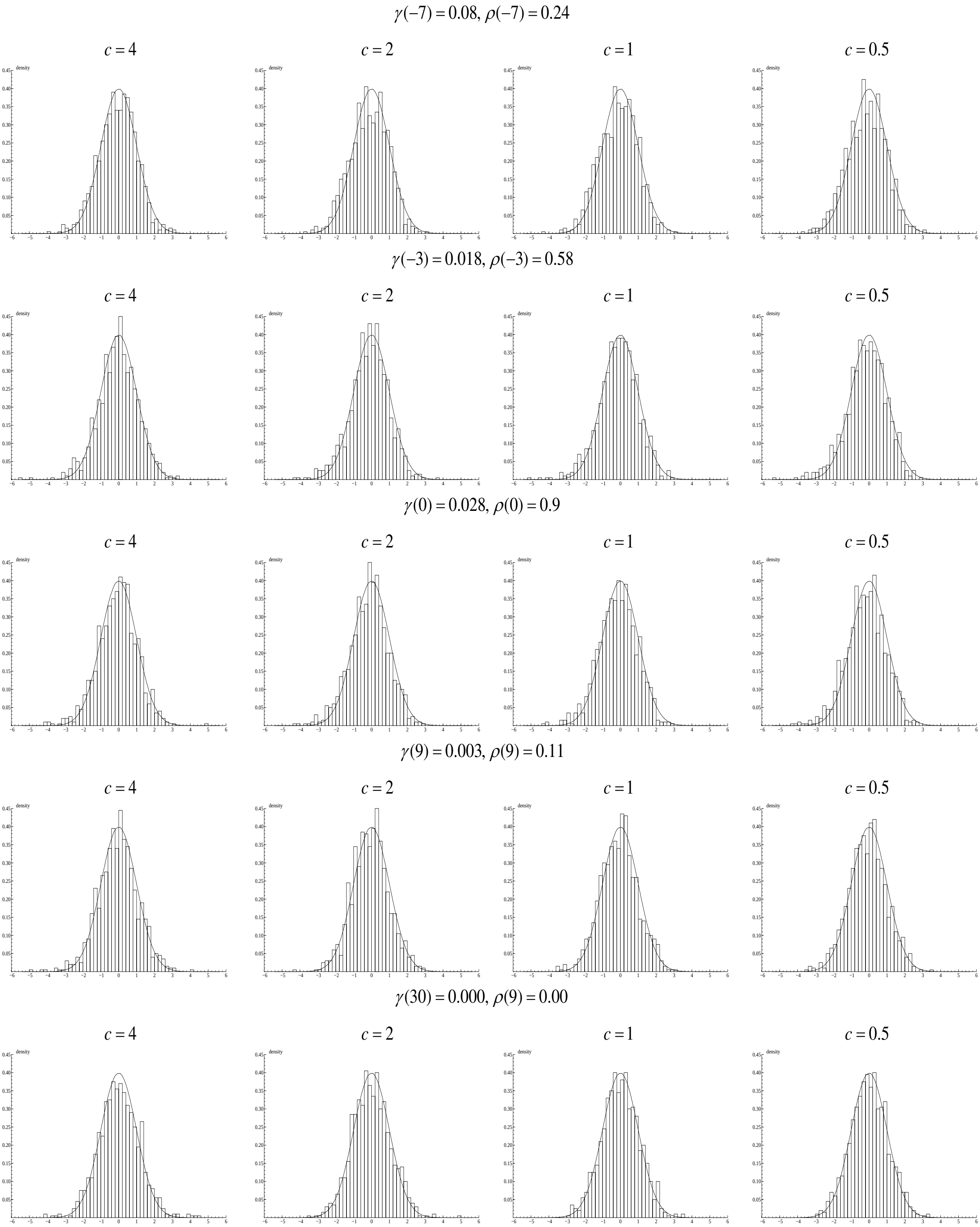

Figure 9: Standardized empirical distributions of the cross-covariance estimator of noises (histogram) and the standard normal distributions (solid line).

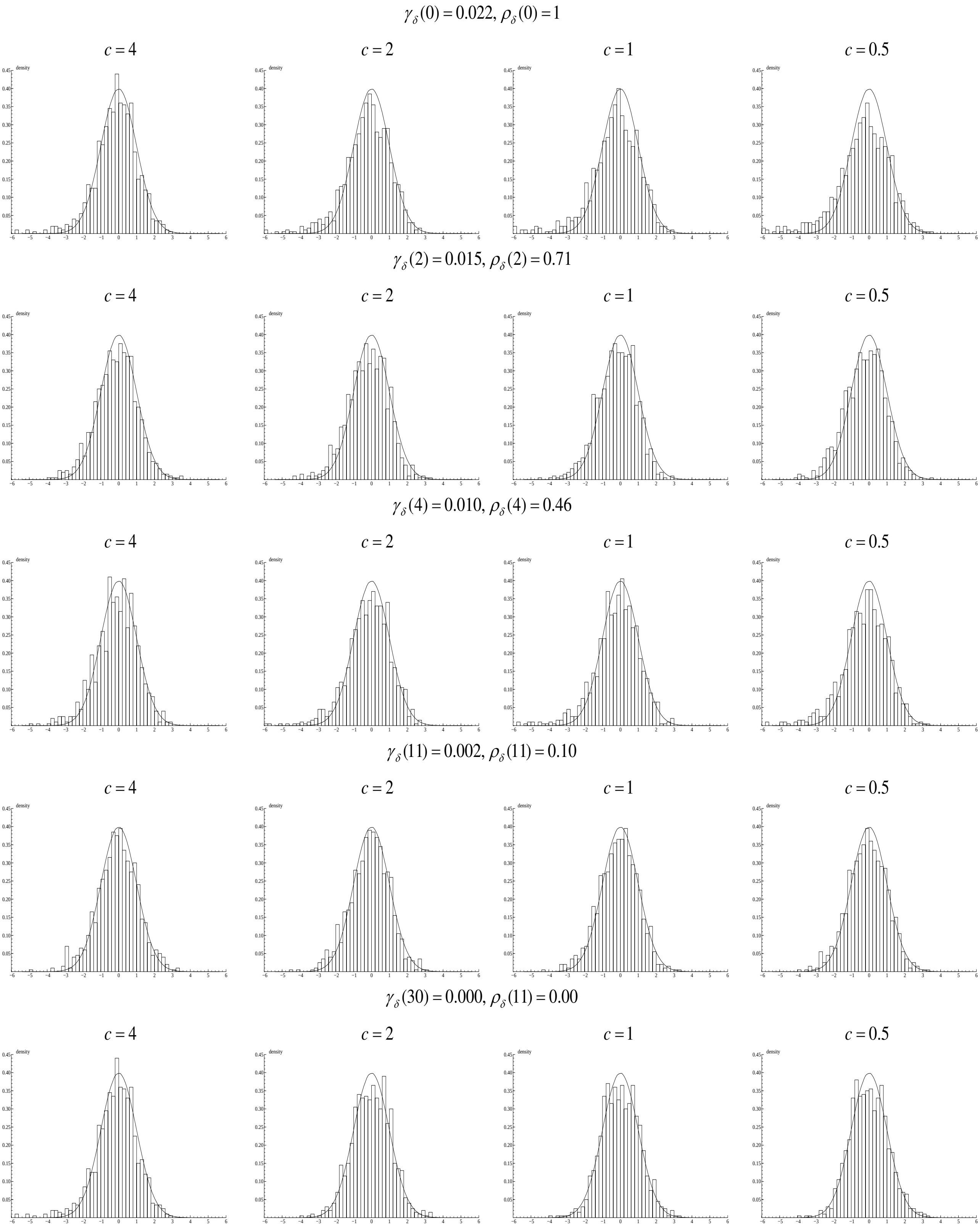

Figure 10: Standardized empirical distributions of the autocovariance estimator of noises (histogram) and the standard normal distributions (solid line).

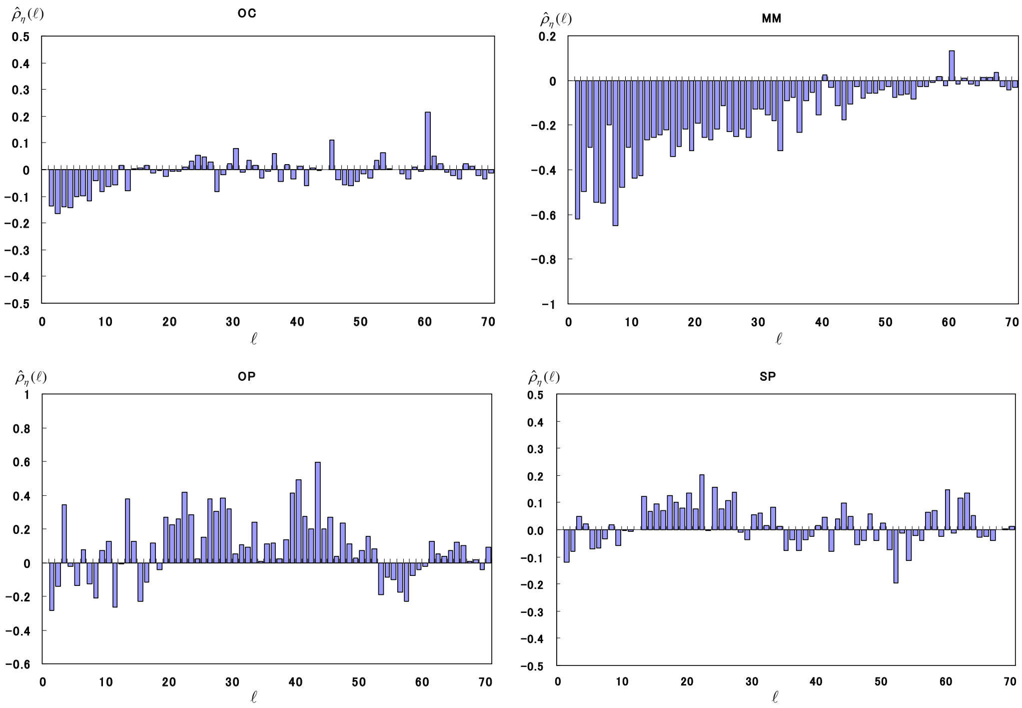

Figure 11: Autocorrelation functions of the noise in OC, MM, OP and SP.

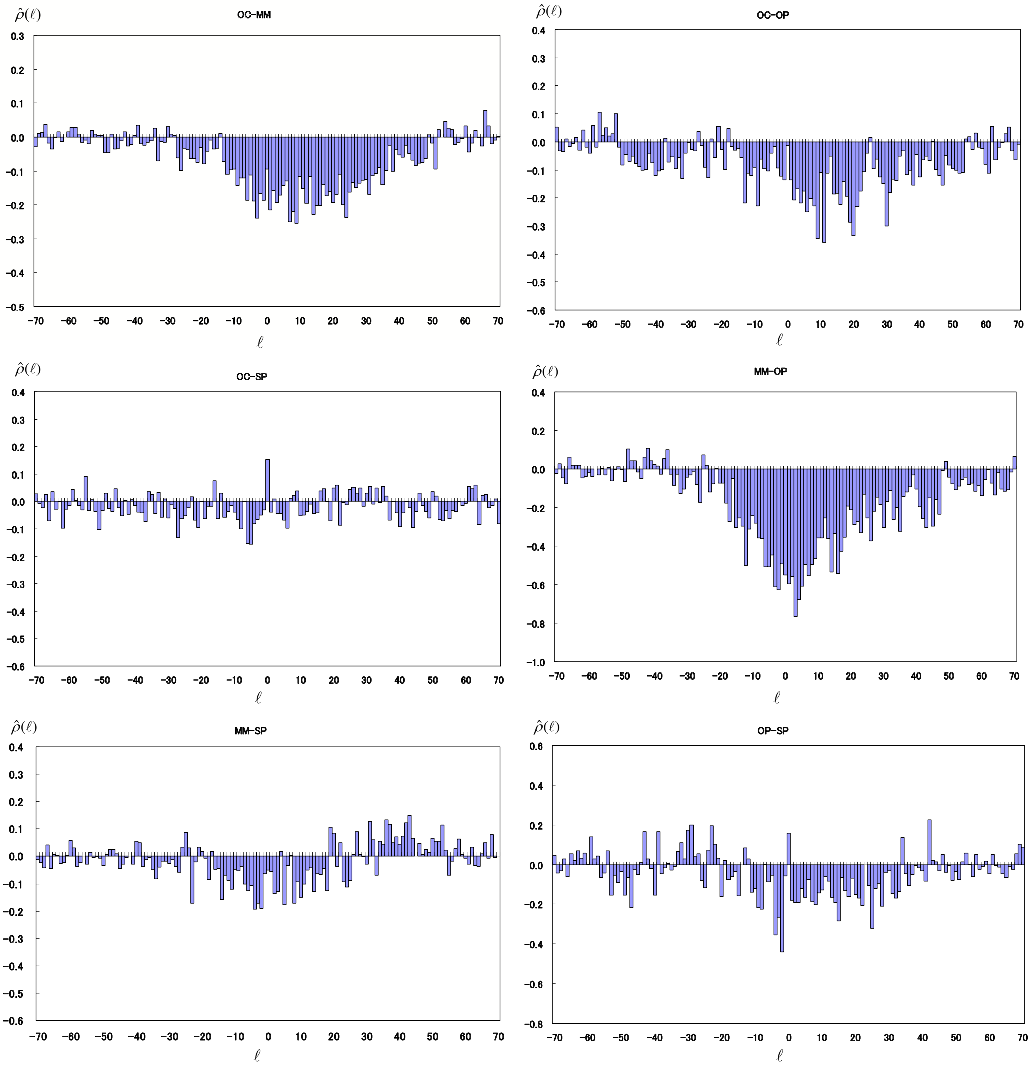

Figure 12: Cross-correlation functions of the noises among the four assets.

Table 1: Bias of cross and auto covariance estimators of bivariate MA noise processes.

	true covariances				$c=4$		$c=2$		$c=1$		$c=0.5$		
ℓ	$\gamma(\ell)$	$\gamma_\delta(\ell)$	$\tilde{\gamma}(\ell)$	$\tilde{\gamma}_\delta(\ell)$	$\hat{\gamma}(\ell)_{[4]}$	$\hat{\gamma}_\delta(\ell)_{[4]}$	$\hat{\gamma}(\ell)_{[2]}$	$\hat{\gamma}_\delta(\ell)_{[2]}$	$\hat{\gamma}(\ell)_{[1]}$	$\hat{\gamma}_\delta(\ell)_{[1]}$	$\hat{\gamma}(\ell)_{[0.5]}$	$\hat{\gamma}_\delta(\ell)_{[0.5]}$	$\ddot{\gamma}(\ell)$
sample size: 2300, sampling interval: 10 seconds													
-4	0		-0.0001		-0.0002		-0.0002		-0.0002		-0.0001		-0.0008
-3	0		0.0001		0.0003		0.0003		0.0003		0.0002		-0.0008
-2	0.0088		0.0000		-0.0006		-0.0006		-0.0006		-0.0005		-0.0007
-1	-0.0180		-0.0177		-0.0012		-0.0013		-0.0010		-0.0007		-0.0007
0	0.0269	0.0217	0.0415	0.0290	-0.0019	-0.0014	-0.0020	-0.0013	-0.0018	-0.0013	-0.0014	-0.0011	-0.0004
1	-0.0151	-0.0115	-0.0160	-0.0121	-0.0018	-0.0011	-0.0020	-0.0010	-0.0019	-0.0010	-0.0016	-0.0009	0.0000
2	0.0080	0.0061	0.0001	0.0000	-0.0007	-0.0004	-0.0008	-0.0004	-0.0007	-0.0004	-0.0006	-0.0004	0.0001
3	0	0	-0.0001	0.0000	0.0003	0.0002	0.0004	0.0002	0.0003	0.0002	0.0003	0.0002	0.0000
4	0	0	0.0000	0.0000	-0.0001	-0.0001	-0.0002	-0.0001	-0.0002	-0.0001	-0.0001	-0.0001	0.0001
sample size: 1550, sampling interval: 15 seconds													
-4	0		-0.0001		-0.0003		-0.0003		-0.0003		-0.0003		0.0001
-3	0		0.0001		0.0006		0.0007		0.0007		0.0006		-0.0001
-2	0.0088		0.0000		-0.0011		-0.0012		-0.0012		-0.0010		0.0001
-1	-0.0180		-0.0177		-0.0028		-0.0032		-0.0031		-0.0024		0.0000
0	0.0269	0.0217	0.0415	0.0291	-0.0037	-0.0028	-0.0040	-0.0029	-0.0037	-0.0025	-0.0032	-0.0022	0.0000
1	-0.0151	-0.0115	-0.0159	-0.0121	-0.0034	-0.0024	-0.0037	-0.0025	-0.0034	-0.0022	-0.0032	-0.0019	-0.0001
2	0.0080	0.0061	-0.0001	-0.0001	-0.0012	-0.0009	-0.0012	-0.0010	-0.0013	-0.0009	-0.0012	-0.0009	-0.0002
3	0	0	0.0001	0.0001	0.0009	0.0002	0.0008	0.0003	0.0009	0.0004	0.0008	0.0004	-0.0002
4	0	0	0.0000	-0.0001	-0.0004	-0.0004	-0.0003	-0.0003	-0.0003	-0.0003	-0.0004	-0.0003	-0.0002
sample size: 770, sampling interval: 30 seconds													
-4	0		-0.0001		-0.0002		-0.0003		-0.0002		-0.0003		-0.0004
-3	0		0.0000		0.0010		0.0010		0.0009		0.0009		-0.0003
-2	0.0088		0.0001		-0.0011		-0.0012		-0.0011		-0.0012		-0.0005
-1	-0.0180		-0.0179		-0.0059		-0.0059		-0.0058		-0.0055		-0.0007
0	0.0269	0.0217	0.0417	0.0290	-0.0062	-0.0046	-0.0064	-0.0051	-0.0063	-0.0048	-0.0058	-0.0043	-0.0005
1	-0.0151	-0.0115	-0.0160	-0.0121	-0.0053	-0.0040	-0.0058	-0.0046	-0.0059	-0.0044	-0.0052	-0.0039	-0.0005
2	0.0080	0.0061	-0.0001	0.0000	-0.0014	-0.0011	-0.0015	-0.0013	-0.0015	-0.0010	-0.0014	-0.0009	-0.0006
3	0	0	0.0002	0.0001	0.0007	0.0004	0.0009	0.0003	0.0009	0.0005	0.0008	0.0005	-0.0004
4	0	0	-0.0002	-0.0001	-0.0002	-0.0004	-0.0003	-0.0005	-0.0004	-0.0003	-0.0004	-0.0002	-0.0007
sample size: 2300, sampling interval: 30 seconds in 3 days													
-4	0		-0.0001		-0.0002		-0.0002		-0.0002		-0.0002		0.0001
-3	0		0.0001		0.0003		0.0004		0.0004		0.0003		0.0002
-2	0.0088		0.0000		-0.0007		-0.0008		-0.0007		-0.0007		0.0002
-1	-0.0180		-0.0177		-0.0014		-0.0015		-0.0014		-0.0010		0.0002
0	0.0269	0.0217	0.0415	0.0290	-0.0022	-0.0015	-0.0022	-0.0015	-0.0021	-0.0016	-0.0019	-0.0016	-0.0001
1	-0.0151	-0.0115	-0.0160	-0.0121	-0.0022	-0.0012	-0.0021	-0.0012	-0.0021	-0.0013	-0.0020	-0.0013	-0.0004
2	0.0080	0.0061	0.0001	0.0000	-0.0008	-0.0005	-0.0009	-0.0005	-0.0009	-0.0005	-0.0008	-0.0005	-0.0004
3	0	0	-0.0001	0.0000	0.0003	0.0003	0.0004	0.0002	0.0004	0.0002	0.0002	0.0001	-0.0004
4	0	0	0.0000	0.0000	-0.0002	-0.0001	-0.0002	-0.0001	-0.0003	-0.0001	-0.0002	-0.0002	-0.0004

Note: The market microstructure noises are generated by a bivariate MA(2) process. $\ell = 1$ represents one tick whose length is equal to each sampling interval (5, 10, 15 and 30 seconds). The second and third columns represent true cross and auto covariances $\gamma(\ell)$, $\gamma_\delta(\ell)$.

Table 2: RMSE of cross and auto covariance estimators of bivariate MA noise processes.

	true covariances				$c=4$		$c=2$		$c=1$		$c=0.5$			RMSE ratio
ℓ	$\gamma(\ell)$	$\gamma_\delta(\ell)$	$\tilde{\gamma}(\ell)$	$\tilde{\gamma}_\delta(\ell)$	$\hat{\gamma}(\ell)_{[4]}$	$\hat{\gamma}_\delta(\ell)_{[4]}$	$\hat{\gamma}(\ell)_{[2]}$	$\hat{\gamma}_\delta(\ell)_{[2]}$	$\hat{\gamma}(\ell)_{[1]}$	$\hat{\gamma}_\delta(\ell)_{[1]}$	$\hat{\gamma}(\ell)_{[0.5]}$	$\hat{\gamma}_\delta(\ell)_{[0.5]}$	$\ddot{\gamma}(\ell)$	$\frac{\ddot{\gamma}(\ell)}{\hat{\gamma}(\ell)_{[1]}}$
sample size: 2300, sampling interval: 10 seconds														
-4	0		0.0031		0.0020		0.0021		0.0020		0.0020		0.0156	7.6173
-3	0		0.0032		0.0023		0.0023		0.0023		0.0022		0.0158	6.9754
-2	0.0088		0.0032		0.0025		0.0026		0.0025		0.0025		0.0156	6.1666
-1	-0.0180		0.0180		0.0054		0.0054		0.0051		0.0046		0.0158	3.0850
0	0.0269	0.0217	0.0416	0.0291	0.0057	0.0048	0.0059	0.0048	0.0056	0.0048	0.0052	0.0046	0.0121	2.1403
1	-0.0151	-0.0115	0.0163	0.0124	0.0059	0.0041	0.0061	0.0040	0.0060	0.0040	0.0055	0.0039	0.0152	2.5456
2	0.0080	0.0061	0.0032	0.0023	0.0025	0.0021	0.0025	0.0020	0.0025	0.0020	0.0024	0.0021	0.0152	6.1137
3	0	0	0.0032	0.0022	0.0024	0.0020	0.0024	0.0019	0.0024	0.0019	0.0023	0.0019	0.0152	6.4281
4	0	0	0.0032	0.0022	0.0021	0.0019	0.0021	0.0018	0.0021	0.0018	0.0020	0.0018	0.0153	7.4416
sample size: 1550, sampling interval: 15 seconds														
-4	0		0.0039		0.0031		0.0031		0.0031		0.0031		0.0165	5.3057
-3	0		0.0039		0.0034		0.0033		0.0033		0.0034		0.0162	4.8833
-2	0.0088		0.0039		0.0035		0.0037		0.0036		0.0036		0.0164	4.5473
-1	-0.0180		0.0182		0.0080		0.0082		0.0080		0.0074		0.0162	2.0107
0	0.0269	0.0217	0.0418	0.0293	0.0081	0.0071	0.0083	0.0072	0.0082	0.0069	0.0077	0.0068	0.0126	1.5406
1	-0.0151	-0.0115	0.0164	0.0124	0.0082	0.0062	0.0083	0.0062	0.0082	0.0060	0.0079	0.0058	0.0162	1.9887
2	0.0080	0.0061	0.0040	0.0028	0.0036	0.0032	0.0036	0.0032	0.0037	0.0032	0.0037	0.0032	0.0159	4.3586
3	0	0	0.0040	0.0028	0.0035	0.0033	0.0034	0.0031	0.0034	0.0031	0.0035	0.0032	0.0161	4.7868
4	0	0	0.0039	0.0028	0.0033	0.0031	0.0032	0.0030	0.0033	0.0030	0.0034	0.0030	0.0160	4.8777
sample size: 770, sampling interval: 30 seconds														
-4	0		0.0057		0.0071		0.0071		0.0071		0.0073		0.0165	2.3181
-3	0		0.0057		0.0073		0.0072		0.0073		0.0074		0.0163	2.2304
-2	0.0088		0.0057		0.0072		0.0072		0.0073		0.0075		0.0165	2.2569
-1	-0.0180		0.0189		0.0129		0.0128		0.0128		0.0124		0.0162	1.2608
0	0.0269	0.0217	0.0422	0.0295	0.0119	0.0128	0.0120	0.0126	0.0120	0.0127	0.0118	0.0125	0.0128	1.0720
1	-0.0151	-0.0115	0.0169	0.0128	0.0127	0.0115	0.0125	0.0112	0.0125	0.0113	0.0123	0.0111	0.0167	1.3340
2	0.0080	0.0061	0.0056	0.0041	0.0072	0.0082	0.0069	0.0077	0.0066	0.0078	0.0070	0.0079	0.0164	2.4911
3	0	0	0.0057	0.0041	0.0077	0.0085	0.0072	0.0080	0.0071	0.0081	0.0075	0.0082	0.0168	2.3509
4	0	0	0.0058	0.0041	0.0071	0.0081	0.0068	0.0078	0.0067	0.0078	0.0070	0.0078	0.0162	2.4104
sample size: 2300, sampling interval: 30 seconds in 3 days														
-4	0		0.0033		0.0039		0.0038		0.0039		0.0039		0.0095	2.4521
-3	0		0.0033		0.0040		0.0041		0.0041		0.0042		0.0095	2.3299
-2	0.0088		0.0033		0.0040		0.0040		0.0040		0.0041		0.0095	2.3448
-1	-0.0180		0.0180		0.0068		0.0068		0.0067		0.0064		0.0095	1.4298
0	0.0269	0.0217	0.0417	0.0292	0.0067	0.0067	0.0068	0.0065	0.0067	0.0066	0.0064	0.0067	0.0072	1.0714
1	-0.0151	-0.0115	0.0163	0.0124	0.0072	0.0057	0.0070	0.0056	0.0071	0.0056	0.0069	0.0057	0.0093	1.3175
2	0.0080	0.0061	0.0033	0.0024	0.0039	0.0041	0.0037	0.0038	0.0036	0.0039	0.0037	0.0040	0.0091	2.4818
3	0	0	0.0033	0.0023	0.0040	0.0042	0.0038	0.0040	0.0038	0.0040	0.0039	0.0042	0.0091	2.3827
4	0	0	0.0033	0.0024	0.0038	0.0042	0.0035	0.0039	0.0035	0.0039	0.0036	0.0041	0.0092	2.6264

Note: The market microstructure noises are generated by a bivariate MA(2) processes. $\ell = 1$ represents one tick time whose length is equal to each sampling interval (5, 10, 15 and 30 seconds). The second and third columns represent true cross and auto covariances $\gamma(\ell)$, $\gamma_\delta(\ell)$.

Table 3: Size and power of test statistics for significance in (14), (21) and (22).

	true covariances and correlations				$c=4$		$c=2$		$c=1$		$c=0.5$		
ℓ	$\gamma(\ell)$	$\rho(\ell)$	$\gamma_\delta(\ell)$	$\rho_\delta(\ell)$	$\hat\gamma(\ell)_{[4]}$	$\hat\gamma_\delta(\ell)_{[4]}$	$\hat\gamma(\ell)_{[2]}$	$\hat\gamma_\delta(\ell)_{[2]}$	$\hat\gamma(\ell)_{[1]}$	$\hat\gamma_\delta(\ell)_{[1]}$	$\hat\gamma(\ell)_{[0.5]}$	$\hat\gamma_\delta(\ell)_{[0.5]}$	$\ddot\gamma(\ell)$
sample size: 2300, sampling interval: 10 seconds													
-4	0.0000	0.00			0.045		0.042		0.045		0.055		0.560
-3	0.0000	0.00			0.055		0.058		0.058		0.073		0.577
-2	0.0088	0.29			0.843		0.838		0.854		0.891		0.580
-1	-0.0180	-0.58			1.000		1.000		1.000		1.000		0.724
0	0.0269	0.87	0.0217	1.00	1.000	0.998	1.000	0.998	1.000	0.999	1.000	1.000	0.932
1	-0.0151	-0.49	-0.0115	-0.53	0.991	0.990	0.992	0.987	0.993	0.990	0.996	0.990	0.698
2	0.0080	0.26	0.0061	0.28	0.77	0.775	0.756	0.793	0.779	0.801	0.819	0.819	0.596
3	0.0000	0.00	0.0000	0.00	0.061	0.058	0.058	0.047	0.056	0.053	0.071	0.068	0.546
4	0.0000	0.00	0.0000	0.00	0.047	0.050	0.042	0.047	0.048	0.051	0.056	0.072	0.526
sample size: 1550, sampling interval: 15 seconds													
-4	0.0000	0.00			0.037		0.036		0.030		0.054		0.484
-3	0.0000	0.00			0.052		0.042		0.050		0.076		0.460
-2	0.0088	0.29			0.637		0.624		0.646		0.697		0.549
-1	-0.0180	-0.58			0.973		0.981		0.983		0.987		0.687
0	0.0269	0.87	0.0217	1.00	1.000	0.973	1.000	0.974	1.000	0.983	1.000	0.988	0.918
1	-0.0151	-0.49	-0.0115	-0.53	0.944	0.909	0.952	0.930	0.961	0.921	0.970	0.931	0.628
2	0.0080	0.26	0.0061	0.28	0.54	0.552	0.554	0.557	0.565	0.597	0.590	0.613	0.519
3	0.0000	0.00	0.0000	0.00	0.046	0.046	0.042	0.044	0.051	0.054	0.071	0.082	0.484
4	0.0000	0.00	0.0000	0.00	0.051	0.047	0.040	0.049	0.046	0.053	0.058	0.062	0.486
sample size: 770, sampling interval: 30 seconds													
-4	0.0000	0.00			0.038		0.041		0.038		0.059		0.281
-3	0.0000	0.00			0.046		0.044		0.052		0.071		0.271
-2	0.0088	0.29			0.316		0.308		0.324		0.363		0.333
-1	-0.0180	-0.58			0.859		0.864		0.870		0.886		0.537
0	0.0269	0.87	0.0217	1.00	0.929	0.845	0.937	0.860	0.952	0.864	0.960	0.880	0.846
1	-0.0151	-0.49	-0.0115	-0.53	0.794	0.765	0.820	0.798	0.835	0.809	0.845	0.808	0.482
2	0.0080	0.26	0.0061	0.28	0.237	0.232	0.234	0.219	0.240	0.232	0.286	0.275	0.339
3	0.0000	0.00	0.0000	0.00	0.047	0.048	0.040	0.044	0.045	0.054	0.064	0.068	0.281
4	0.0000	0.00	0.0000	0.00	0.038	0.037	0.029	0.033	0.035	0.040	0.042	0.048	0.274
sample size: 2300, sampling interval: 30 seconds in 3 days													
-4	0.0000	0.00			0.048		0.053		0.055		0.067		0.386
-3	0.0000	0.00			0.067		0.071		0.067		0.080		0.383
-2	0.0088	0.29			0.72		0.715		0.726		0.753		0.584
-1	-0.0180	-0.58			0.939		0.945		0.951		0.956		0.845
0	0.0269	0.87	0.0217	1.00	0.989	0.939	0.996	0.944	0.997	0.951	0.997	0.955	0.986
1	-0.0151	-0.49	-0.0115	-0.53	0.913	0.851	0.923	0.876	0.926	0.882	0.941	0.887	0.800
2	0.0080	0.26	0.0061	0.28	0.638	0.635	0.648	0.657	0.643	0.658	0.677	0.658	0.501
3	0.0000	0.00	0.0000	0.00	0.067	0.060	0.060	0.062	0.060	0.064	0.074	0.075	0.354
4	0.0000	0.00	0.0000	0.00	0.039	0.058	0.035	0.055	0.040	0.053	0.046	0.068	0.362

Note: The market microstructure noises are generated by a bivariate MA(2) processes. $\ell = 1$ represents one tick time whose length is equal to each sampling interval (5, 10, 15 and 30 seconds). From second to fifth columns represent true cross and auto covariances $\gamma(\ell)$, $\gamma_\delta(\ell)$ and true cross and auto correlation coefficients $\rho(\ell)$, $\rho(\ell)$.

Table 4: The bias and RMSE ratio of $\hat{\gamma}(\ell)_{[1]}$ and $\hat{\gamma}_\delta(\ell)_{[1]}$ with drift parameter (μ_1, μ_2) to $\hat{\gamma}(\ell)_{[1]}$ and $\hat{\gamma}_\delta(\ell)_{[1]}$ without drift.

	no drift		(0.1,0,1)				(0.5,0.5)				(1,1)			
ℓ	bias × 100		bias × 100		RMSE ratio		bias × 100		RMSE ratio		bias × 100		RMSE ratio	
sample size: 2300, sampling interval: 10 seconds														
-4	-0.0155		-0.0156		1.0001		-0.0157		1.0004		-0.0161		1.0010	
-3	0.0321		0.0321		1.0001		0.0316		0.9993		0.0312		1.0000	
-2	-0.0594		-0.0594		1.0000		-0.0593		0.9989		-0.0597		0.9992	
-1	-0.1001		-0.1001		1.0000		-0.1008		0.9998		-0.1012		0.9999	
0	-0.1798	-0.1280	-0.1798	-0.1280	1.0000	1.0001	-0.1803	-0.1270	1.0004	0.9971	-0.1809	-0.1271	1.0006	0.9977
1	-0.1944	-0.1024	-0.1944	-0.1024	1.0000	1.0001	-0.1946	-0.1017	1.0001	0.9964	-0.1951	-0.1016	1.0002	0.9976
2	-0.0675	-0.0439	-0.0675	-0.0439	1.0000	1.0000	-0.0678	-0.0436	1.0001	1.0005	-0.0682	-0.0443	1.0003	0.9998
3	0.0347	0.0179	0.0346	0.0179	1.0000	1.0002	0.0344	0.0175	1.0001	1.0002	0.0340	0.0173	1.0003	1.0030
4	-0.0167	-0.0125	-0.0167	-0.0125	1.0001	1.0002	-0.0169	-0.0124	1.0006	1.0010	-0.0174	-0.0127	1.0014	1.0030
sample size: 1550, sampling interval: 15 seconds														
-4	-0.0333		-0.0333		0.9999		-0.0335		0.9995		-0.0343		0.9994	
-3	0.0704		0.0703		0.9998		0.0701		0.9988		0.0693		0.9978	
-2	-0.1163		-0.1163		1.0001		-0.1165		1.0005		-0.1173		1.0013	
-1	-0.3067		-0.3067		0.9999		-0.3070		0.9996		-0.3077		0.9991	
0	-0.3686	-0.2507	-0.3686	-0.2496	0.9999	0.9984	-0.3690	-0.2498	0.9998	0.9979	-0.3705	-0.2522	0.9998	0.9993
1	-0.3384	-0.2190	-0.3385	-0.2179	1.0000	0.9981	-0.3387	-0.2180	0.9998	0.9975	-0.3400	-0.2199	0.9997	0.9976
2	-0.1253	-0.0912	-0.1253	-0.0912	1.0000	0.9997	-0.1255	-0.0913	1.0003	0.9988	-0.1259	-0.0923	1.0008	0.9981
3	0.0890	0.0372	0.0890	0.0374	0.9999	0.9996	0.0888	0.0372	0.9998	0.9984	0.0876	0.0365	1.0000	0.9982
4	-0.0349	-0.0275	-0.0349	-0.0275	0.9998	0.9994	-0.0351	-0.0276	0.9992	0.9975	-0.0357	-0.0286	0.9985	0.9972
sample size: 770, sampling interval: 30 seconds														
-4	-0.0168		-0.0168		1.0006		-0.0179		1.0031		-0.0219		1.0074	
-3	0.0931		0.0931		1.0007		0.0923		1.0040		0.0891		1.0086	
-2	-0.1132		-0.1133		1.0007		-0.1136		1.0040		-0.1172		1.0083	
-1	-0.5770		-0.5770		1.0001		-0.5794		1.0005		-0.5848		1.0016	
0	-0.6288	-0.4794	-0.6288	-0.4782	1.0000	1.0002	-0.6304	-0.4805	1.0002	1.0026	-0.6351	-0.4845	1.0023	1.0061
1	-0.5895	-0.4351	-0.5894	-0.4339	0.9999	0.9999	-0.5898	-0.4357	0.9994	1.0021	-0.5918	-0.4382	0.9990	1.0054
2	-0.1517	-0.0997	-0.1516	-0.0998	0.9996	1.0011	-0.1519	-0.1016	0.9987	1.0059	-0.1529	-0.1047	0.9962	1.0140
3	0.0917	0.0489	0.0918	0.0488	0.9997	1.0012	0.0914	0.0470	0.9986	1.0063	0.0901	0.0442	0.9970	1.0148
4	-0.0378	-0.0272	-0.0377	-0.0275	0.9996	1.0011	-0.0381	-0.0293	0.9984	1.0060	-0.0380	-0.0326	0.9925	1.0140
sample size: 2300, sampling interval: 30 seconds in 3 days														
-4	-0.0244		-0.0245		1.0003		-0.0258		1.0024		-0.0292		1.0070	
-3	0.0359		0.0358		1.0003		0.0346		1.0022		0.0312		1.0057	
-2	-0.0719		-0.0720		0.9999		-0.0733		1.0000		-0.0767		1.0020	
-1	-0.1356		-0.1357		1.0002		-0.1370		1.0011		-0.1404		1.0026	
0	-0.2098	-0.1575	-0.2100	-0.1577	1.0002	1.0005	-0.2114	-0.1591	1.0015	1.0031	-0.2167	-0.1638	1.0062	1.0071
1	-0.2076	-0.1320	-0.2078	-0.1321	1.0001	1.0005	-0.2087	-0.1332	1.0004	1.0031	-0.2127	-0.1368	1.0037	1.0070
2	-0.0913	-0.0488	-0.0915	-0.0490	1.0003	1.0003	-0.0934	-0.0509	1.0025	1.0014	-0.0979	-0.0545	1.0079	1.0050
3	0.0365	0.0203	0.0363	0.0202	1.0004	1.0003	0.0354	0.0187	1.0009	1.0022	0.0305	0.0150	1.0077	1.0059
4	-0.0269	-0.0065	-0.0271	-0.0067	1.0006	1.0005	-0.0291	-0.0078	1.0030	1.0035	-0.0326	-0.0114	1.0090	1.0081

Note: The market microstructure noises are generated by a bivariate MA(2) processes. $\ell = 1$ represents one tick time whose length is equal to each sampling interval. In the simulation we use the stochastic differential equations with drift parameter $(\mu_1, \mu_2) = (0.1, 0.1)$, $(0.5, 0.5)$ and $(1, 1)$.

Table 5: Bias of cross and auto covariance estimators of bivariate AR noise processes.

	true covariances				$c=4$		$c=2$		$c=1$		$c=0.5$		
ℓ	$\gamma(\ell)$	$\rho(\ell)$	$\gamma_\delta(\ell)$	$\rho_\delta(\ell)$	$\hat{\gamma}(\ell)_{[4]}$	$\hat{\gamma}_\delta(\ell)_{[4]}$	$\hat{\gamma}(\ell)_{[2]}$	$\hat{\gamma}_\delta(\ell)_{[2]}$	$\hat{\gamma}(\ell)_{[1]}$	$\hat{\gamma}_\delta(\ell)_{[1]}$	$\hat{\gamma}(\ell)_{[0.5]}$	$\hat{\gamma}_\delta(\ell)_{[0.5]}$	$\ddot{\gamma}(\ell)$
-30	0.0000	0.00			0.0000		-0.0001		-0.0001		-0.0001		-0.0010
-25	0.0001	0.00			0.0001		0.0001		0.0000		0.0000		-0.0008
-20	0.0004	0.01			0.0000		-0.0001		-0.0003		-0.0001		-0.0008
-15	0.0013	0.04			0.0000		-0.0001		0.0000		-0.0001		-0.0005
-10	0.0039	0.12			-0.0003		-0.0004		-0.0004		-0.0004		-0.0005
-9	0.0048	0.16			0.0000		-0.0001		-0.0002		-0.0003		-0.0011
-8	0.0060	0.19			-0.0003		-0.0004		-0.0004		-0.0005		-0.0007
-7	0.0075	0.24			-0.0001		-0.0003		-0.0003		-0.0004		-0.0007
-6	0.0094	0.30			0.0001		-0.0001		-0.0002		-0.0001		-0.0006
-5	0.0117	0.38			-0.0001		-0.0002		-0.0002		-0.0002		-0.0004
-4	0.0145	0.47			-0.0002		-0.0003		-0.0003		-0.0004		-0.0001
-3	0.0179	0.58			0.0000		-0.0003		-0.0002		-0.0003		-0.0010
-2	0.0217	0.70			-0.0001		-0.0004		-0.0005		-0.0005		-0.0004
-1	0.0255	0.83			-0.0001		-0.0002		-0.0003		-0.0003		-0.0007
0	0.0278	0.90	0.0217	1.00	0.0000	-0.0002	-0.0002	-0.0003	-0.0003	-0.0003	-0.0004	-0.0003	-0.0003
1	0.0210	0.68	0.0186	0.86	-0.0002	0.0001	-0.0003	0.0000	-0.0004	-0.0001	-0.0004	0.0000	-0.0005
2	0.0163	0.53	0.0154	0.71	-0.0001	-0.0003	-0.0002	-0.0005	-0.0002	-0.0004	-0.0002	-0.0006	-0.0001
3	0.0129	0.42	0.0125	0.58	0.0000	-0.0001	0.0000	-0.0002	0.0000	-0.0002	-0.0002	-0.0003	-0.0005
4	0.0102	0.33	0.0101	0.46	-0.0002	-0.0001	-0.0002	-0.0002	-0.0002	-0.0002	-0.0003	-0.0003	-0.0004
5	0.0081	0.26	0.0081	0.37	0.0001	-0.0001	0.0000	-0.0001	0.0000	-0.0002	0.0000	-0.0002	0.0002
6	0.0065	0.21	0.0065	0.30	-0.0001	-0.0001	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0001
7	0.0052	0.17	0.0052	0.24	-0.0003	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0003	-0.0005
8	0.0042	0.13	0.0042	0.19	-0.0003	0.0000	-0.0002	0.0000	-0.0002	0.0000	-0.0001	0.0000	-0.0004
9	0.0033	0.11	0.0033	0.15	0.0000	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0005
10	0.0027	0.09	0.0027	0.12	-0.0002	-0.0001	-0.0001	-0.0002	-0.0001	-0.0002	-0.0002	-0.0002	-0.0006
15	0.0009	0.03	0.0009	0.04	-0.0003	0.0000	-0.0002	-0.0001	-0.0001	0.0000	-0.0002	0.0000	-0.0002
20	0.0003	0.01	0.0003	0.01	0.0001	0.0000	0.0001	0.0000	0.0002	0.0000	0.0001	0.0000	-0.0003
25	0.0001	0.00	0.0001	0.00	0.0002	0.0000	0.0001	0.0000	0.0002	0.0000	0.0001	0.0000	-0.0004
30	0.0000	0.00	0.0000	0.00	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0003

Note: The market microstructure noises are generated by a bivariate AR(1) processes. From second to fifth columns represent true cross and auto covariances $\gamma(\ell)$, $\gamma_\delta(\ell)$ and true cross and auto correlation coefficients $\rho(\ell)$, $\rho(\ell)$.

Table 6: RMSE of cross and auto covariance estimators of bivariate AR noise processes.

	true covariances				$c=4$		$c=2$		$c=1$		$c=0.5$			RMSE ratio
ℓ	$\gamma(\ell)$	$\rho(\ell)$	$\gamma_\delta(\ell)$	$\rho_\delta(\ell)$	$\hat\gamma(\ell)_{[4]}$	$\hat\gamma_\delta(\ell)_{[4]}$	$\hat\gamma(\ell)_{[2]}$	$\hat\gamma_\delta(\ell)_{[2]}$	$\hat\gamma(\ell)_{[1]}$	$\hat\gamma_\delta(\ell)_{[1]}$	$\hat\gamma(\ell)_{[0.5]}$	$\hat\gamma_\delta(\ell)_{[0.5]}$	$\ddot\gamma(\ell)$	$\frac{\ddot\gamma(\ell)}{\hat\gamma(\ell)_{[1]}}$
-30	0.0000	0.00			0.0034		0.0034		0.0033		0.0033		0.0185	5.6022
-25	0.0001	0.00			0.0036		0.0035		0.0035		0.0034		0.0188	5.3962
-20	0.0004	0.01			0.0037		0.0037		0.0037		0.0035		0.0185	5.0531
-15	0.0013	0.04			0.0038		0.0038		0.0038		0.0038		0.0181	4.7665
-10	0.0039	0.12			0.0039		0.0038		0.0038		0.0039		0.0184	4.8153
-9	0.0048	0.16			0.0039		0.0039		0.0039		0.0038		0.0182	4.6998
-8	0.0060	0.19			0.0039		0.0040		0.0040		0.0039		0.0185	4.6768
-7	0.0075	0.24			0.0039		0.0040		0.0039		0.0040		0.0189	4.7858
-6	0.0094	0.30			0.0042		0.0042		0.0042		0.0041		0.0183	4.3739
-5	0.0117	0.38			0.0040		0.0039		0.0040		0.0040		0.0179	4.4678
-4	0.0145	0.47			0.0042		0.0043		0.0043		0.0043		0.0190	4.3749
-3	0.0179	0.58			0.0042		0.0041		0.0042		0.0041		0.0188	4.4923
-2	0.0217	0.70			0.0041		0.0041		0.0042		0.0043		0.0187	4.4703
-1	0.0255	0.83			0.0044		0.0046		0.0045		0.0046		0.0185	4.1387
0	0.0278	0.90	0.0217	1.00	0.0036	0.0018	0.0036	0.0019	0.0035	0.0018	0.0036	0.0019	0.0135	3.8068
1	0.0210	0.68	0.0186	0.86	0.0043	0.0059	0.0044	0.0059	0.0043	0.0057	0.0044	0.0057	0.0180	4.1603
2	0.0163	0.53	0.0154	0.71	0.0043	0.0037	0.0043	0.0038	0.0041	0.0036	0.0041	0.0037	0.0182	4.4145
3	0.0129	0.42	0.0125	0.58	0.0040	0.0029	0.0041	0.0029	0.0040	0.0029	0.0041	0.0029	0.0188	4.6493
4	0.0102	0.33	0.0101	0.46	0.0040	0.0026	0.0039	0.0027	0.0039	0.0026	0.0040	0.0026	0.0184	4.6939
5	0.0081	0.26	0.0081	0.37	0.0041	0.0025	0.0041	0.0025	0.0041	0.0025	0.0040	0.0025	0.0183	4.4672
6	0.0065	0.21	0.0065	0.30	0.0041	0.0023	0.0039	0.0024	0.0040	0.0023	0.0039	0.0023	0.0182	4.5523
7	0.0052	0.17	0.0052	0.24	0.0041	0.0024	0.0041	0.0024	0.0041	0.0023	0.0041	0.0023	0.0182	4.4815
8	0.0042	0.13	0.0042	0.19	0.0040	0.0024	0.0040	0.0024	0.0039	0.0024	0.0039	0.0024	0.0176	4.5172
9	0.0033	0.11	0.0033	0.15	0.0039	0.0024	0.0038	0.0024	0.0039	0.0024	0.0039	0.0024	0.0180	4.6349
10	0.0027	0.09	0.0027	0.12	0.0039	0.0024	0.0040	0.0024	0.0039	0.0023	0.0038	0.0023	0.0183	4.7445
15	0.0009	0.03	0.0009	0.04	0.0038	0.0022	0.0037	0.0023	0.0036	0.0022	0.0036	0.0022	0.0184	5.1160
20	0.0003	0.01	0.0003	0.01	0.0037	0.0021	0.0037	0.0020	0.0036	0.0020	0.0036	0.0020	0.0180	5.0089
25	0.0001	0.00	0.0001	0.00	0.0035	0.0020	0.0034	0.0020	0.0033	0.0020	0.0033	0.0019	0.0184	5.5534
30	0.0000	0.00	0.0000	0.00	0.0034	0.0019	0.0034	0.0018	0.0033	0.0018	0.0033	0.0018	0.0174	5.2596

Note: The market microstructure noises are generated by a bivariate AR(1) processes. From second to fifth columns represent true cross and auto covariances $\gamma(\ell)$, $\gamma_\delta(\ell)$ and true cross and auto correlation coefficient $\rho(\ell)$, $\rho(\ell)$.

Table 7: Autocovariance of univariate noise process.

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

Industry	Manufacturers		Medicals	
	OC	MM	OP	SP
noise variance	1.17×10^{-6}	1.37×10^{-7}	2.54×10^{-7}	1.04×10^{-6}
test statistics	2240.84^{*}	159.09^{*}	142.27^{*}	1251.67^{*}
NSR	0.0049	0.0009	0.0014	0.0042

(b) The test statistics for the autocovariance.

ℓ	OC		MM		OP		SP	
1	-3.22 *	(25132)	-5.00 *	(43753)	-1.19	(14838)	-1.99 *	(12163)
2	-3.11 *	(15821)	-3.55 *	(31658)	-0.46	(8542)	-0.95	(6936)
3	-2.71 *	(14909)	-1.52	(29671)	1.07	(7227)	0.30	(5813)
4	-2.58 *	(14696)	-3.66 *	(28661)	-0.08	(6415)	0.24	(5677)
5	-1.92	(12958)	-3.18 *	(27087)	-0.35	(5680)	-0.83	(4855)
6	-2.04 *	(12599)	-0.99	(27243)	0.21	(5567)	-0.72	(4786)
7	-2.00 *	(11752)	-4.02 *	(25902)	-0.45	(5130)	-0.37	(4385)
8	-0.68	(11599)	-2.60 *	(25476)	-0.63	(5125)	0.25	(4390)
9	-1.52	(10923)	-1.63	(24362)	0.22	(4855)	-0.71	(3956)
10	-1.36	(11018)	-2.88 *	(23916)	0.51	(4879)	-0.03	(3960)
20	-0.14	(9123)	-1.54	(21664)	0.72	(3643)	1.46	(3223)
22	0.17	(8870)	-2.10 *	(21306)	1.46	(3432)	2.51 *	(3032)
24	1.16	(8726)	-0.65	(21264)	0.13	(3476)	2.60 *	(3048)
30	1.74	(9672)	-1.07	(21414)	0.23	(3730)	0.76	(3581)
40	0.33	(8284)	0.23	(20047)	1.81	(3072)	0.22	(2690)
43	-0.04	(8148)	-2.15 *	(19972)	3.05 *	(3077)	0.63	(2749)
50	-0.45	(8211)	-0.50	(19996)	0.56	(2983)	0.39	(2775)
52	1.00	(8101)	-1.10	(19868)	0.63	(2879)	-3.14 *	(2666)
54	0.16	(8093)	-1.83	(19886)	-0.70	(2927)	-2.00 *	(2798)
55	-0.02	(7985)	-0.54	(19890)	-0.85	(2903)	-0.40	(2582)
60	5.79 *	(9709)	4.89 *	(21240)	-0.43	(3521)	3.19 *	(3545)
61	1.71	(8705)	-0.59	(20239)	1.96 *	(3273)	-0.20	(3018)
62	0.82	(8497)	0.31	(19754)	0.68	(3009)	2.52 *	(2823)
63	-0.36	(8336)	-0.53	(19989)	0.55	(2911)	2.81 *	(2843)
70	-0.45	(7989)	-1.06	(19547)	1.24	(2850)	0.22	(2598)

Note: In the top table (a), the test statistic for the variance of noise is given by (21). The critical value at 5% significance level is equal to 3.84. NSR is an abbreviation for noise-to-signal ratio which is the ratio between the variance estimate of the noise and the integrated variance estimate using a realized kernel-based estimator of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008). The bottom table (b) shows the test statistic (22) for the significance of the autocovariance of noise with $\ell > 0$. The critical value of the test statistic (22) is ± 1.96 at 5% significance level. Superscript * denotes significance at the 5% levels. The numbers in parentheses represent a sample size $N_{1,\ell}$.

Table 8: Test statistics for the cross-covariance of bivariate noise processes.

ℓ	OC-MM		OC-OP		OC-SP		MM-OP		MM-SP		OP-SP	
-70	-0.91	(10615)	1.30	(4217)	0.74	(3952)	-0.55	(6527)	-0.36	(6114)	0.89	(2366)
-62	-0.50	(10801)	1.07	(4245)	-2.72 *	(4105)	-1.06	(6500)	-0.56	(6124)	0.53	(2360)
-59	1.01	(10987)	1.28	(4207)	1.13	(4043)	-0.87	(6467)	0.80	(6006)	2.48 *	(2456)
-57	0.19	(11071)	2.25 *	(4092)	-0.30	(3939)	-0.72	(6528)	-0.70	(6161)	0.82	(2478)
-55	-0.30	(10966)	0.90	(4143)	2.18 *	(3863)	-0.68	(6355)	-0.80	(6107)	-0.52	(2417)
-48	-1.60	(10905)	-1.26	(4171)	-0.57	(3825)	2.51 *	(6500)	0.74	(6283)	-0.68	(2455)
-34	0.93	(10832)	-1.32	(4205)	-0.88	(3889)	-0.96	(6572)	-2.22 *	(6213)	-0.07	(2442)
-33	-2.68 *	(10826)	-0.79	(4317)	0.67	(3874)	-0.28	(6496)	-1.06	(6166)	0.60	(2386)
-30	0.92	(10959)	-0.05	(4225)	-1.06	(3959)	-0.44	(6562)	-0.59	(6306)	1.47	(2506)
-25	-0.95	(10917)	-1.25	(4251)	-0.88	(4051)	0.80	(6464)	1.56	(6191)	-1.06	(2442)
-17	-0.13	(10862)	-0.19	(4290)	-0.29	(4069)	-2.36 *	(6634)	0.31	(6283)	-0.38	(2448)
-10	-1.71	(11441)	-1.16	(4398)	-0.77	(4016)	-1.74	(6868)	-0.72	(6431)	-0.91	(2547)
-9	-2.62 *	(11197)	-2.47 *	(4335)	-1.12	(4225)	-1.98 *	(6863)	-0.79	(6549)	-1.49	(2649)
-8	-1.85	(11214)	-0.87	(4506)	-1.74	(4096)	-2.44 *	(7025)	-0.51	(6474)	-1.52	(2668)
-7	-2.05 *	(11295)	-1.09	(4528)	-0.03	(4108)	-1.97 *	(6993)	-1.62	(6556)	0.02	(2663)
-6	-3.19 *	(11457)	-1.35	(4575)	-2.25 *	(4082)	-2.86 *	(7158)	-1.82	(6564)	-0.57	(2616)
-5	-1.45	(11411)	-0.47	(4627)	-2.53 *	(4129)	-3.88 *	(7121)	-1.45	(6512)	-0.34	(2706)
-4	-2.72 *	(11358)	-0.19	(4703)	-1.34	(4200)	-2.90 *	(7129)	-2.90 *	(6659)	-2.23 *	(2824)
-3	-4.29 *	(11577)	-1.01	(4776)	-1.13	(4404)	-2.97 *	(7300)	-2.17 *	(6800)	-1.87	(2894)
-2	-1.90	(11671)	-1.25	(4808)	-0.88	(4474)	-2.96 *	(7410)	-2.96 *	(6739)	-2.53 *	(2980)
-1	-2.52 *	(11942)	-1.49	(5019)	-0.53	(4657)	-2.46 *	(7824)	-0.78	(6948)	-0.41	(3414)
0	-2.32 *	(27263)	-0.21	(12576)	3.58 *	(11042)	-4.51 *	(17583)	-0.81	(15319)	2.03 *	(9256)
1	-3.37 *	(12401)	-1.39	(5239)	-0.67	(4751)	-2.98 *	(7648)	-0.67	(6822)	-1.54	(3257)
2	-2.64 *	(12046)	-1.94	(4832)	0.07	(4461)	-3.56 *	(7391)	-1.54	(6787)	-1.54	(2922)
3	-3.19 *	(11628)	-1.67	(4712)	-0.68	(4437)	-4.81 *	(7149)	-1.33	(6681)	-1.32	(2807)
4	-2.99 *	(11473)	-2.03 *	(4790)	-0.78	(4245)	-4.39 *	(7141)	0.24	(6637)	-1.12	(2853)
5	-2.17 *	(11454)	-1.61	(4619)	-1.37	(4235)	-3.23 *	(7197)	-2.05 *	(6659)	-1.64	(2773)
6	-1.78	(11371)	-2.08 *	(4496)	-1.78	(4073)	-3.30 *	(7154)	-0.34	(6568)	-0.73	(2647)
7	-3.87 *	(11308)	-1.96	(4554)	0.24	(4026)	-3.24 *	(7081)	0.03	(6547)	-1.62	(2613)
8	-3.54 *	(11486)	-2.27 *	(4602)	0.38	(4141)	-3.16 *	(6987)	-2.14 *	(6554)	-1.91	(2572)
9	-4.08 *	(11307)	-3.57 *	(4541)	0.72	(4089)	-2.93 *	(6975)	-1.02	(6233)	-1.42	(2640)
10	-1.87	(11247)	-1.13	(4399)	-0.97	(4205)	-2.32 *	(6770)	-1.75	(6362)	-1.31	(2551)
15	-2.98 *	(11265)	-2.17 *	(4373)	-0.73	(4168)	-2.45 *	(6874)	-0.85	(6347)	-3.34 *	(2459)
20	-2.83 *	(11053)	-2.97 *	(4368)	0.89	(4049)	-1.35	(6556)	1.19	(6300)	-1.53	(2419)
30	-1.98 *	(11014)	-3.06 *	(4281)	0.53	(4024)	-2.18 *	(6700)	-0.44	(6251)	-0.30	(2382)
42	-0.48	(10801)	-0.68	(4206)	-0.02	(3961)	-2.16 *	(6531)	1.91	(6247)	4.18 *	(2345)
43	-1.04	(10696)	-0.99	(4187)	-0.42	(3981)	-2.77 *	(6462)	2.28 *	(6156)	0.44	(2369)
47	-1.55	(10989)	-2.25 *	(4116)	-0.29	(3897)	-1.93	(6547)	0.11	(6064)	-0.66	(2336)
64	-0.05	(10916)	-0.47	(4225)	-1.83	(3913)	-2.39 *	(6398)	-0.91	(6073)	-0.65	(2337)
66	2.86 *	(10848)	0.61	(4067)	0.63	(3996)	-1.86	(6341)	1.19	(6108)	-0.10	(2424)
68	-0.67	(10927)	-0.72	(4127)	-0.39	(4011)	-2.16 *	(6560)	2.04 *	(6092)	1.04	(2372)
70	0.04	(10733)	-0.24	(4146)	-1.25	(3922)	1.73	(6445)	-0.02	(6140)	1.63	(2381)

Note: The test statistic for the significance of the cross-covariance of the bivariate noise processes is given by (14). The critical value of the test statistics (14) is ± 1.96 at 5% significance level. Superscript * denote significance at the 5% levels. The numbers in parentheses represent a sample size N_ℓ.