# Multi-scale analysis of lead-lag relationships in high-frequency financial markets

Takaki Hayashi\*†‡

Yuta Koike §†¶

February 6, 2018

#### Abstract

We propose a novel estimation procedure for scale-by-scale lead-lag relationships of financial assets observed at high-frequency in a non-synchronous manner. The proposed estimation procedure does not require any interpolation processing of original datasets and is applicable to those with highest time resolution available. Consistency of the proposed estimators is shown under the continuous-time framework that has been developed in our previous work [16]. An empirical application shows promising results of the proposed approach.

*Keywords*: High-frequency data; Lead-lag relationship; Multi-scale analysis; Non-synchronous data; Stochastic volatility; Wavelet.

#### 1 Introduction

A financial market accommodates a diversified groups of participants. They have different sources of money, different time horizons and different risk attitudes, with different quality and quantity of information. In Müller *et al.* [27] it is argued that such differences are engraved in price formation at each of distinct time scales. They can cause a *multi-scale* structure embedded in the financial market.

This paper intends to study such a multi-scale structure of financial markets that can exist in a very short time period. In particular, we are to investigate *lead-lag relationships* between financial assets by the use of high-frequency data. Identification of lead-lag relationships among assets is fundamentally important both for theoretical and practical perspectives; the existence of such relationships may mean the inefficiency of financial markets for theorists but it may also provide opportunities for market participants to earn "excess" profits. So that so, it is quite natural that lead-lag analysis has been conducted in the finance literature for a long time. Since 90's as high-frequency data has become more and more accessible, lead-lag relationships with high-frequency data have been studied by such authors as [3, 8, 23, 31]. In the meantime, multi-scale

<sup>\*</sup>Keio University, Graduate School of Business Administration, 4-1-1 Hiyoshi, Yokohama 223-8526, Japan

<sup>&</sup>lt;sup>†</sup>Department of Business Administration, Graduate School of Social Sciences, Tokyo Metropolitan University, Marunouchi Eiraku Bldg. 18F, 1-4-1 Marunouchi, Chiyoda-ku, Tokyo 100-0005 Japan

<sup>&</sup>lt;sup>‡</sup>CREST, Japan Science and Technology Agency

<sup>§</sup> Graduate School of Mathematical Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914 Japan

The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

analysis with high-frequency financial data has been carried out; e.g., [2, 11, 15, 26, 32]. However, main interest of most of these articles is the estimation of volatilities of assets. There is little work that conducts multi-scale analysis of lead-lag relationships in the high-frequency domain; one exception is Hafner [12] which has examined multi-scale structures of the lead-lag relationships between the returns, durations and volumes of high-frequency transaction data of the IBM stock.

To our understanding, the main focus of those studies conducting multi-scale analysis is empirical application *per se*, not to develop a new estimation methodology. Their adopted approaches are theoretically based on "classical" discrete time series that appear to be more suitable for daily or lower frequency data with longer time horizons. On one hand, analysis of high-frequency financial data shall focus on a short time horizon, that is, one day or shorter. So, it is unclear whether one can reasonably apply such a "classical" method to high-frequency financial data without reservation. On the other hand, *continuous-time* modeling provides a convenient and powerful framework to analyze high-frequency data observed in a short horizon (cf. Aït-Sahalia and Jacod [1]).

With these in mind, in [16] the authors have developed a continuous-time framework that is designed specifically for multi-scale analysis of lead-lag relationships in high-frequency data. There, they introduce two Brownian motions  $B^1$  and  $B^2$  with a scale-by-scale correlation structure. More precisely, they have shown that, for any  $R_j \in [-1,1]$  and  $\theta_j \in \mathbb{R}$   $(j=0,1,\ldots)$ , there exists a bivariate Gaussian process  $B_t = (B_t^1, B_t^2)$   $(t \in \mathbb{R})$  with stationary increments such that

- (I) both  $B^1$  and  $B^2$  are two-sided Brownian motions,
- (II) the cross-spectral density of B is given by

$$f(\lambda) = \sum_{j=0}^{\infty} R_j e^{-\sqrt{-1}\theta_j \lambda} 1_{\Lambda_j}(\lambda), \qquad \lambda \in \mathbb{R},$$
(1)

where 
$$\Lambda_j = [-2^{j+1}\pi, -2^j\pi) \cup (2^j\pi, 2^{j+1}\pi]$$
 for every  $j \in \mathbb{Z}$ .

The frequency band  $\Lambda_j$  corresponds to the time scale between  $2^{-j}$  and  $2^{-j+1}$  in the time domain. Also, note that, if  $W_t = (W_t^1, W_t^2)$   $(t \in \mathbb{R})$  is a two-sided bivariate Brownian motion with correlation R, for  $\theta \in \mathbb{R}$  the process  $(W_t^1, W_{t-\theta}^2)$   $(t \in \mathbb{R})$  has the cross-spectral density  $Re^{-\sqrt{-1}\theta\lambda}$   $(\lambda \in \mathbb{R})$ . Therefore, we can consider that  $B^1$  and  $B^2$  have a lead-lag relationship with the time-lag  $\theta_j$  in the time scale between  $2^{-j}$  and  $2^{-j+1}$ . Hence, under this model we can understand the multi-scale structure of the lead-lag relationships by estimating the parameters  $\theta_j$  from observation data.

The main contribution of this paper is to develop a novel estimation procedure for the parameters  $\theta_j$  based on *non-synchronous* observations of (volatility-modulated versions of)  $B^1$  and  $B^2$ . Although a procedure has already been proposed by the current authors ([16]), it requires data interpolation in accordance with a regular grid with size equated to the finest time resolution, and hence is computationally formidable if we are analyzing a dataset with, for instance, one micro second time precision. Even in such a situation the proposed procedure in this paper is free from any interpolation processing of the original data and is applicable. A numerical experiment also shows that the new estimators seem to have better (finite sample) performance than the interpolation-based estimator when the sampling times are non-synchronous to a reasonable degree.

The rest of the paper is organized as follows. In Section 2 we present the theoretical setting considered in this paper in details. Our new estimation procedure is described in Section 3. We develop an asymptotic theory associated with the proposed estimators in Section 4. In Section 5 we assess finite sample performance of the proposed estimators by Monte Carlo experiments, and in Section 6 we apply our procedure to empirical datasets. Section 7 concludes the paper. All the proofs are collected in Section 8.

#### 2 Setting

Throughout the paper we focus on situations where the resolution of the timestamps of datasets is *very* small. We let the time resolution correspond to  $\tau_N := 2^{-N-1}$  for some  $N \in \mathbb{N}$ . We will develop an asymptotic theory in the high-frequency setting, i.e., when N tends to infinity, or the time resolution shrinks to zero.

As mentioned in the Introduction, our theoretical framework is based on a bivariate Gaussian process  $B_t = (B_t^1, B_t^2)$  ( $t \in \mathbb{R}$ ) with stationary increments satisfying properties (I)–(II). Since we are mainly interested in the lead-lag relationships at scales close to the finest observation resolution, it is convenient to "relabel" indices of the parameters  $R_j$  and  $\theta_j$  in (1) so that the finest resolution  $\tau_N$  corresponds to the level j=1 while we consider the asymptotic theory such that N tends to infinity. For this reason, as in [16] we replace property (II) with the following one: The cross-spectral density of B is given by

$$f_N(\lambda) = \sum_{j=1}^{N+1} R_j e^{-\sqrt{-1}\theta_j \lambda} 1_{\Lambda_{(j)}}(\lambda), \qquad \lambda \in \mathbb{R},$$
(2)

where (j) = N - j + 1. We also assume that  $\theta_j \in (-\delta, \delta)$  for every j with some  $\delta > 0$ .

Now, for each  $\nu=1,2$ , we consider the log price process  $X^{\nu}=(X^{\nu}_t)_{t\geq 0}$  of the  $\nu$ -th asset given by

$$X_t^{\nu} = X_0^{\nu} + \int_0^t \sigma_s^{\nu} dB_s^{\nu}, \qquad t \ge 0,$$
 (3)

where  $(\sigma_t^{\nu})_{t\geq 0}$  is a càdlàg process adapted to the filtration  $(\mathcal{F}_t^{\nu})$  such that the process  $(B_t^{\nu})$  is, respectively, a one-dimensional  $(\mathcal{F}_t^{\nu})$ -Brownian motion. We observe the process  $X^{\nu}$  on the interval  $[0,T+\delta]$  at the sampling times  $0 \leq t_0^{\nu} < t_1^{\nu} < \cdots < t_{n_{\nu}}^{\nu} \leq T+\delta$ . The sampling times  $(t_i^1)_{i=0}^{n_1}$  and  $(t_i^2)_{i=0}^{n_2}$  are random variables which are independent of  $(X^1,X^2)$  and implicitly depend on N such that

$$r_N := \max_{\nu=1,2} \max_{i=0,1,\dots,n_{\nu}+1} (t_i^{\nu} - t_{i-1}^{\nu}) \to^p 0$$

as  $N \to \infty$ , where we set  $t^{\nu}_{-1} := 0$  and  $t^{\nu}_{n_{\nu}+1} := T + \delta$  for each  $\nu = 1, 2$ .

**Remark 1.** Our model is generally not a semimartingale, hence it is generally not free of arbitrage in the absence of market frictions due to the well-known fundamental theorem of asset pricing (see e.g. [9]). However, if we take account of market frictions such as discrete trading or transaction costs, we can show that our model has no arbitrage; see [17] for details.

#### **3** Construction of the estimators

Our aim is to estimate the parameters  $\theta_j$  for each j based on discrete observation data  $(X^1_{t^1_i})_{i=0}^{n_1}$  and  $(X^2_{t^2_i})_{i=0}^{n_2}$ . We begin by introducing some notation. For each  $\nu=1,2$ , we associate the observation times

 $(t_i^{\nu})_{i=0}^{n_{\nu}}$  with the collection of intervals  $\Pi_N^{\nu} = \{(t_{i-1}^{\nu}, t_i^{\nu}] : i = 1, \dots, n_{\nu}\}$ . We will systematically employ the notation I (resp. J) for an element of  $\Pi_N^1$  (resp.  $\Pi_N^2$ ).

For an interval  $H \subset [0,\infty)$ , we set  $\overline{H} = \sup H$ ,  $\underline{H} = \inf H$ ,  $|H| = \overline{H} - \underline{H}$ . In addition, we set  $V(H) = V_{\overline{H}} - V_{\underline{H}}$  for a a stochastic process  $(V_t)_{t \geq 0}$ , and  $H_{\theta} = H + \theta$  for a real number  $\theta$ .

Now we explain how to construct our estimators. To explain the idea behind the construction, we focus on the case of  $\sigma_s^{\nu} \equiv 1$  for  $\nu = 1, 2$ . The parameter  $\theta_j$  is the unique maximizer of the scale-by-scale cross-covariance function  $\rho_{(j)}(\theta)$  between  $B^1$  and  $B^2$ , which is defined by

$$\rho_{(j)}(\theta) = E\left[\left(\int_{-\infty}^{\infty} \psi_{(j)}^{LP}(s-u)dB_s^1\right) \left(\int_{-\infty}^{\infty} \psi_{(j)}^{LP}(s-u-\theta)dB_s^2\right)\right], \qquad \theta \in \mathbb{R},$$

where  $\psi^{LP}_{(j)}(s)=2^{(j)/2}\psi^{LP}(2^{(j)}s)$  and  $\psi^{LP}$  denotes the Littlewood-Paley wavelet:

$$\psi^{LP}(s) = (\pi s)^{-1}(\sin(2\pi s) - \sin(\pi s))$$

(see Sections 2.2–2.3 of [16] for details). Motivated by this fact, we first construct a sensible *covariance* estimator  $\widehat{\rho}_{(j)}(\theta)$  for  $\rho_{(j)}(\theta)$ , and then construct the *lead-lag* estimator  $\widehat{\theta}_j$  for  $\theta_j$  as a maximizer of  $|\widehat{\rho}_{(j)}(\theta)|$  as in [21]. The idea behind the construction of the estimator  $\widehat{\rho}_{(j)}(\theta)$  is as follows. Let  $U^N(\theta)$  be the inverse Fourier transform of  $f_N(\lambda)$ . Then we have

$$\rho_{(j)}(\theta) = 2^{-\frac{(j)}{2}} (U^N * \psi_{(j)}^{LP})(\theta) = \int_{-\infty}^{\infty} U^N(\theta - s) \psi^{LP}(2^{(j)}s) ds$$

by the convolution theorem. This suggests us to consider the following estimator for  $\rho_{(i)}(\theta)$ :

$$\widehat{\rho}_{(j)}(\theta) := \widehat{\rho}_{(j)}^{N}(\theta) = \sum_{l=-L_{i}+1}^{L_{j}-1} \widehat{U}^{N}(\theta - l\tau_{N})\Psi_{j}(l),$$

where  $\widehat{U}^N(\theta)$  is an estimator for  $U^N(\theta)$  and  $\Psi_j(l)$  is an approximation of  $\psi^{LP}(2^{(j)}l\tau_N)$  (it turns out that the factor  $\tau_N$  corresponding to ds is unnecessary because  $2^{(j)}\tau_N=2^{-j}$  does not tend to 0 in our asymptotic setting), both of which are explicitly defined in the following. Since  $U^N(\theta)$  may be regarded as the "cross-covariance function between  $dB^1$  and  $dB^2$ ", we adopt the following estimator introduced in Hoffmann  $et\ al.$  [21] as  $\widehat{U}^N(\theta)$ :

$$\widehat{U}^{N}(\theta) = \begin{cases} \sum_{I \in \Pi_{N}^{1}, J \in \Pi_{N}^{2} : \overline{I} \leq T} X^{1}(I) X^{2}(J) K(I, J_{-\theta}) & \text{if } \theta \geq 0, \\ \sum_{I \in \Pi_{N}^{1}, J \in \Pi_{N}^{2} : \overline{J} \leq T} X^{1}(I) X^{2}(J) K(I_{\theta}, J) & \text{if } \theta < 0, \end{cases}$$

where we set  $K(I,J)=1_{\{I\cap J\neq\emptyset\}}$  for two intervals I and J. This  $\widehat{U}^N(\theta)$  can be regarded as the empirical cross-covariance estimator between the returns of  $X^1$  and  $X^2$  at the lag  $\theta$  computed by Hayashi and Yoshida [18]'s method to handle the non-synchronous sampling times. In the meantime, the Fourier inversion formula yields

$$\cdot \psi^{LP}(2^{(j)}l\tau_N) = 2^j \tau_N \int_{-\infty}^{\infty} e^{\sqrt{-1}l\tau_N \lambda} 1_{\Lambda_{(j)}}(\lambda) d\lambda = 2^j \int_{-\pi}^{\pi} e^{\sqrt{-1}l\lambda} 1_{\Lambda_{-j}}(\lambda) d\lambda,$$

hence the transfer function of  $(\psi^{LP}(2^{(j)}l\tau_N))_{l\in\mathbb{Z}}$  is  $2^j1_{\Lambda_{-j}}(\lambda)$ . In particular,  $\Psi_j(l)$  well approximates  $\psi^{LP}(2^{(j)}l\tau_N)$  if the transfer function of  $(\Psi_j(l))_{l=-L_j+1}^{L_j-1}$  well approximates  $2^j1_{\Lambda_{-j}}(\lambda)$ . We construct such a sequence  $(\Psi_j(l))_{l=-L_j+1}^{L_j-1}$  from Daubechies' wavelet filter as follows (we refer to Chapters 6–8 of [6], Section 4.8 of [30] and Section 3.4.5 of [34] for details about Daubechies' wavelet filter). Let  $(h_p)_{p=0}^{L-1}$  be Daubechies' wavelet filter of (even) length L whose power transfer function  $H_L(\lambda) = |\sum_{p=0}^{L-1} h_p e^{-\sqrt{-1}\lambda p}|^2$  is given by

$$H_L(\lambda) = 2\sin^L(\lambda/2) \sum_{p=0}^{L/2-1} {L/2-1+p \choose p} \cos^{2p}(\lambda/2), \qquad \lambda \in \mathbb{R}.$$

The associated scaling filter  $(g_p)_{p=0}^{L-1}$  is defined via the quadrature mirror relationship as  $g_p=(-1)^{p+1}h_{L-p-1}$ ,  $p=0,1,\ldots,L-1$ , hence its power transfer function  $G_L(\lambda)=|\sum_{p=0}^{L-1}g_pe^{-\sqrt{-1}\lambda p}|^2$  satisfies  $G_L(\lambda)=H_L(\lambda-\pi)$ . Then, for every j we construct the associated level j wavelet filter  $(h_{j,p})_{p=0}^{L_j-1}$  recursively by  $h_{1,p}=h_p$  for  $p=0,1,\ldots,L_1-1$  and  $h_{j,p}=\sum_{q=0}^{L_{j-1}-1}g_{p-2q}h_{j-1,q}$  for  $p=0,1,\ldots,L_j-1$ , where  $L_j=(2^j-1)(L-1)+1$  and  $g_p=0$  for  $p\notin\{0,1,\ldots,L-1\}$ . Now we define the sequence  $(\Psi_j(l))_{l=-L_j+1}^{L_j-1}$  by

$$\Psi_j(l) = \sum_{p=0}^{L_j - 1 - |l|} h_{j,p} h_{j,p+|l|}, \qquad l = 0, \pm 1, \dots, \pm (L_j - 1).$$

These quantities are identical to the *autocorrelation wavelets* from Nason *et al.* [28] (see Definition 3 from [28]). The transfer function  $H_{j,L}(\lambda) = \sum_{l=-L_j+1}^{L_j-1} \Psi_j(l) e^{-\sqrt{-1}l\lambda}$  of  $(\Psi_j(l))_{l=-L_j+1}^{L_j-1}$  is given by

$$H_{j,L}(\lambda) = H_L(2^{j-1}\lambda) \prod_{i=0}^{j-2} G_L(2^i\lambda), \qquad \lambda \in \mathbb{R}$$

(see Eq.(28) from [28]). Therefore,  $H_{j,L}(\lambda)$  well approximates  $2^j 1_{\Lambda_{-j}}(\lambda)$  as  $L \to \infty$  by Theorem 1 from Lai [24]<sup>2</sup> and thus  $\Psi_j(l)$  may be used an approximation of  $\psi^{LP}(2^{(j)}l\tau_N)$ . Finally, for every  $j \in \mathbb{N}$  we define the estimator  $\widehat{\theta}_j := \widehat{\theta}_j^N$  for  $\theta_j$  as a solution of the following equation:

$$\left|\widehat{\rho}_{(j)}(\widehat{\theta}_j)\right| = \max_{\theta \in \mathcal{G}^N} \left|\widehat{\rho}_{(j)}(\theta)\right|.$$

Here, we maximize the function  $\widehat{\rho}_{(j)}(\theta)$  regarding  $\theta$  over the finite grid

$$\mathcal{G}^N = \{l\tau_N : l \in \mathbb{Z}, |l| \le \Gamma_N\}$$

with some positive integer  $\Gamma_N$  as in [21].

**Remark 2.** Given the length L of Daubechies' wavelet filter, we still have several options of  $(h_p)_{p=0}^{L-1}$  such as the *external phase wavelet* and the *least asymmetric wavelet* (cf. Section 4.8 of [30]). However, all of them have the same power transfer function  $H_L(\lambda)$  by definition, hence  $(\Psi_j(l))_{l=-L_j+1}^{L_j-1}$  only depends on the length L of Daubechies' wavelet filters.

<sup>&</sup>lt;sup>1</sup>We use the notation that  $(h_p)$  denotes the wavelet filter and  $(g_p)$  denotes the scaling filter following [30]. Note that the reverse notation is often used in the literature.

<sup>&</sup>lt;sup>2</sup>Note that Lai [24] defines Daubechies' wavelet filter of length L as a filter whose power transfer function is given by  $H_L(\lambda)/2$ .

### 4 Asymptotic theory

For a function  $f \in L^1(\mathbb{R})$ , we denote by  $\mathcal{F}f$  the Fourier transform of f:

$$(\mathcal{F}f)(\lambda) = \int_{-\infty}^{\infty} f(t)e^{-\sqrt{-1}\lambda t}dt, \qquad \lambda \in \mathbb{R}.$$

We impose the following conditions to derive our asymptotic results.

**Assumption 1.** For every  $\nu=1,2$ , the paths of  $\sigma^{\nu}$  are almost surely  $\gamma$ -Hölder continuous for some  $\gamma>0$ 

**Assumption 2.** (i)  $r_N = O_p(\tau_N^{\xi})$  as  $N \to \infty$  for any  $\xi \in (0,1)$ .

(ii) There are constants  $\alpha > 1$ ,  $\beta \in (0,1)$ , Q > 1 and an absolutely continuous real-valued function D on  $[-\pi,\pi]$  such that

$$\tau_m \sum_{k=0}^{\lceil T\tau_m^{-1} \rceil - 1} \int_{-\pi}^{\pi} E\left[ \left| D_k^N(\lambda, \theta_N) - D(\lambda) \right|^Q \right] d\lambda = O(\tau_N^{\alpha})$$

as  $N \to \infty$  for any sequence  $(\theta_N)$  of real numbers satisfying  $\theta_N \in \mathcal{G}^N$  for every N, where  $m = \lceil \beta N \rceil$ ,

$$D_k^N(\lambda, \theta) = \begin{cases} \frac{1}{2\pi\tau_m \tau_N} \sum_{I,J: \underline{I} \in I_m(k)} (\mathcal{F}1_I)(\lambda/\tau_N)(\mathcal{F}1_{J-\theta})(-\lambda/\tau_N)K(I, J_{-\theta}) & \text{if } \theta \ge 0, \\ \frac{1}{2\pi\tau_m \tau_N} \sum_{I,J: \underline{J} \in I_m(k)} (\mathcal{F}1_{I_{\theta}})(\lambda/\tau_N)(\mathcal{F}1_J)(-\lambda/\tau_N)K(I_{\theta}, J) & \text{if } \theta < 0 \end{cases}$$

and  $I_m(k) = [kT\tau_m, (k+1)T\tau_m)$ . Moreover,  $D(\lambda) > 0$  for almost all  $\lambda \in [-\pi, \pi]$  and  $D' \in L^\infty(-\pi, \pi)$ . (iii) There is a constant a > 0 such that  $\inf_{I \in \Pi^1_N} |I| \ge a\tau_N$  and  $\inf_{J \in \Pi^2_N} |J| \ge a\tau_N$  for every N.

The simplest situation where Assumption 2 is satisfied is the equidistant and synchronous sampling case such that  $t_i^1 = t_i^2 = i\tau_N$  for every i. In this case one can easily see that

$$D_k^N(\lambda, \theta) = \frac{1}{2\pi} \left| \frac{e^{-\sqrt{-1}\lambda} - 1}{\lambda} \right|^2$$

for any  $\theta \in \mathcal{G}^N$ , hence Assumption 2 is satisfied with  $D(\lambda)$  being the quantity in the right side of the above equation. Another example is Lo and MacKinlay [25]'s sampling scheme as described by the following proposition:

**Proposition 1.** Suppose that, for each  $\nu=1,2$ , the observation times  $(t_i^{\nu})_{i=0}^{n_{\nu}}$  are randomly chosen from  $\{i\tau_N: i=0,1,\ldots,\lfloor (T+\delta)\tau_N^{-1}\rfloor\}$  using Bernoulli trials with success probability  $1-\pi_{\nu}$ . Then, Assumption 2 is satisfied with

$$D(\lambda) = \frac{1}{\pi \lambda^2} \Re \left[ \frac{(1 - \pi_1)(1 - \pi_2)(1 - e^{-\sqrt{-1}\lambda})}{(1 - \pi_1 e^{-\sqrt{-1}\lambda})(1 - \pi_2 e^{-\sqrt{-1}\lambda})} \right]$$
$$= \frac{1 - \cos \lambda}{\pi \lambda^2} \frac{(1 - \pi_1)(1 - \pi_2)(1 + \pi_1 + \pi_2 - \pi_1 \pi_2 (2\cos \lambda + 1))}{|1 - \pi_1 e^{-\sqrt{-1}\lambda}|^2 |1 - \pi_2 e^{-\sqrt{-1}\lambda}|^2}.$$

Now we state asymptotic results. The first result concerns the asymptotic behavior of the estimators  $\widehat{\rho}_{(j)}(\theta)$  and can be considered as a counterpart of Propositions 3–4 from [21]:

**Theorem 1.** Let j be a positive integer. Suppose that  $L \to \infty$  and  $L\tau_N^{\kappa} \to 0$  as  $N \to \infty$  for any  $\kappa > 0$ . Suppose also that  $(\Gamma_N + L_j)\tau_N < \delta$  for every N. Under Assumptions 1–2, the following statements hold true:

(a) If a sequence  $v_N>0$  satisfies  $L^{-\frac{1}{2}}\tau_N^{-1}v_N\to\infty$  as  $N\to\infty$ , then

$$\max_{\theta \in \mathcal{G}^N: |\theta - \theta_j| \ge v_N} |\widehat{\rho}_{(j)}(\theta)| \to^p 0$$

as  $N \to \infty$ .

(b) Let  $(\vartheta_N)$  be a sequence of real numbers such that  $\vartheta_N \in \mathcal{G}^N$  and  $\tau_N^{-1}(\vartheta_N - \theta_j) \to b$  as  $N \to \infty$  for some  $b \in \mathbb{R}$ . Then

$$\widehat{\rho}_{(j)}(\vartheta_N) \to^p 2^j \Sigma_T(\theta_j) R_j \int_{\Lambda_{-i}} D(\lambda) \cos(b\lambda) d\lambda$$

as  $N \to \infty$ , where

$$\Sigma_T(\theta) = \begin{cases} \int_0^T \sigma_s^1 \sigma_{s+\theta}^2 ds & \text{if } \theta \ge 0, \\ \int_0^T \sigma_{s-\theta}^1 \sigma_s^2 ds & \text{otherwise.} \end{cases}$$

The next theorem concerns consistency of the estimators  $\widehat{\theta}_j$  and can be considered as a counterpart of Theorem 1 from [21]:

**Theorem 2.** Let j be a positive integer. Suppose that  $L \to \infty$  and  $L\tau_N^{\kappa} \to 0$  as  $N \to \infty$  for any  $\kappa > 0$ . Suppose also that  $(\Gamma_N + L_j)\tau_N < \delta$  for every N, and  $\Gamma_N\tau_N \to \delta$  as  $N \to \infty$ . Under Assumptions 1–2, if a sequence  $v_N > 0$  satisfies  $L^{-\frac{1}{2}}\tau_N^{-1}v_N \to \infty$  as  $N \to \infty$ , then

$$v_N^{-1}(\widehat{\theta}_j - \theta_j) \to^p 0$$

as  $N\to\infty$ , provided that  $R_j\neq 0$  and  $\Sigma_T(\theta_j)\neq 0$  a.s. In particular, we have  $\widehat{\theta}_j\to^p\theta_j$  as  $N\to\infty$ .

**Remark 3.** Theorem 2 shows that the proposed estimators  $\hat{\theta}_j$  enjoy a similar asymptotic property to that of the estimator proposed in our previous work [16]. As stated in the Introduction, the estimators have a computational advantage when applying it to high-frequency data with fine timestamps such as one milli- or micro-second. In the next section we see that they can possess another advantage in terms of finite sample performance, especially when observations are non-synchronous.

## 5 Simulation study

In this section we assess the finite sample accuracy of the proposed estimators  $\hat{\theta}_j$  by a Monte Carlo study. We set N=14,  $T=n\tau_N$  with n=30,000.

We simulate model (3) with the following two scenarios of the volatility processes:

**Scenario 1** Constant volatilities.  $\sigma^{\nu} \equiv 1$  for  $\nu = 1, 2$ .

Scenario 2 Stochastic volatilities with a leverage effect. The Heston model is adopted to generate the volatility process  $\sigma_t^{\nu}$  for each  $\nu=1,2$ : The process  $v_t^{\nu}=(\sigma_t^{\nu})^2$  is the solution of the following stochastic differential equation:

$$dv_t^{\nu} = \kappa(\eta - v_t^{\nu})dt + \xi \sqrt{v_t^{\nu}} (\rho dB_t^{\nu} + \sqrt{1 - \rho^2} dW_t^{\nu}),$$

where  $W^{\nu}$  is a standard Wiener process and the initial value  $v_0^{\nu}$  is randomly drawn from the stationary distribution of the process  $v_t^{\nu}$  in each iteration, i.e.  $v_0^{\nu} \sim \operatorname{Gamma}(2\kappa\eta/\xi^2, 2\kappa/\xi^2)$ . We assume that the processes B,  $W^1$  and  $W^2$  are mutually independent. The parameters  $\kappa$ ,  $\eta$ ,  $\xi$  and  $\rho$  are chosen as in [4]:  $\kappa = 5$ ,  $\eta = 0.04$ ,  $\xi = 0.5$  and  $\rho = -0.5$ .

The parameters for the spectral density (2) are chosen as in Table 1. Simulation of the paths of the process B is performed in the same way as in [16].

Table 1: Parameters for the spectral density (2)

$\overline{j}$	1	2	3	4	5	6	7	8	9–14
$R_j$	0.3	0.5	0.7	0.5	0.5	0.5	0.5	0.5	0
$\theta_j/ au_N$	-1	-1	-2	-2	-3	-5	-7	-10	0

We use the Lo-MacKinlay sampling scheme presented in Section 4 to generate the sampling times  $(t_i^1)_{i=0}^{n_1}$ and  $(t_i^2)_{i=0}^{n_2}$ . We fix  $\pi_1$  as  $\pi_1=1/4$  and vary  $\pi_2$  as  $\pi_2\in\{1/4,1/2,3/4\}$ . Recall that  $\pi_\nu$  is the probability of occurrence of observation missing for the  $\nu$ -th asset  $X^{\nu}$ . The larger value  $\pi_2$  takes the less frequently  $X^{\nu}$  is observed, hence the degree of non-synchronicity becomes higher. We use L=20 as the length of Daubechies' wavelet filter and set  $\mathcal{G}^N = \{l\tau_N : l \in \mathbb{Z}, |l| \leq 100\}$ . For comparison we also compute the estimators for  $\theta_i$  proposed in [16], each of which is defined as a maximizer of the corresponding so-called wavelet cross-covariance estimator based on data synchronized by interpolation (we refer to it as "WCCF"). Here, the computation of the wavelet cross-covariance estimators requires the specification of Daubechies' wavelet and we use the least asymmetric wavelet with length 20. We run 1,000 Monte Carlo iterations with each of three experimental conditions in each scenario. Table 2 reports the sample median and the median absolute deviation (MAD) of the estimates for each experiment in Scenario 1. We see from Table 2 that both estimators accurately estimate the true values in the case of  $\pi_2 = 1/4$  at the levels  $j \leq 7$ . It is theoretically natural that the accuracy of the estimators declines as j increases because the contrast function  $|\hat{\rho}_{(i)}(\theta)|$ ,  $\theta \in \mathcal{G}_N$  gets flatter as  $L_j = (2^j - 1)(L - 1) + 1$  increases. In the cases of  $\pi_2 = 1/2$  and  $\pi_2 = 3/4$ , the WCCF estimators are apparently biased at the levels  $j \geq 3$ , while the estimators  $\hat{\theta}_j$  still keep the good precision. Hence our new estimators can handle high-frequency data with rather a high degree of nonsynchronicity. Table 3 shows simulation results in Scenario 2. As the table reveals, the presence of a time variation and a leverage effect in the volatilities does not affect the performance of the proposed estimators, which is aligned with the obtained asymptotic theory.

Table 2: Simulation results in Scenario 1

j	1	2	3	4	5	6	7	8
True	-1	-1	-2	-2	-3	-5	-7	-10
				$\pi_2 =$	1/4			
$\widehat{\theta}_j$	-1(0)	-1(0)	-2(0)	-2(0)	-3(0)	-5(1)	-7(3)	-9(9)
WCCF	-1(0)	-1(0)	-2(0)	-2(0)	-3(0)	-5(1)	-7(3)	-9(9)
				$\pi_2 =$	1/2			
$\widehat{ heta}_j$	-1(0)	-1(0)	-2(0)	-2(0)	-3(0)	-5(1)	-7(3)	-9(9)
WCCF	-1(0)	-1(0)	-1(0)	-1(0)	-2(0)	-4(1)	-6(3)	-8(9)
	$\pi_2 = 3/4$							
$\widehat{\theta}_j$	-1(0)	-1(0)	-2(0)	-2(0)	-3(0)	-5(1)	-7(3)	-9(9)
WCCF	-1(0)	-1(0)	-1(0)	0 (0)	0(0)	-2(1)	-4(3)	-7(9)

This table reports the median and the median absolute deviation (in parentheses) of the estimates in Scenario 1 (divided by  $\tau_N$ ).

Table 3: Simulation results in Scenario 2

j	1	2	3	4	5	6	7	8
True	-1	-1	-2	-2	-3	-5	-7	-10
				$\pi_2 =$	1/4			
$\widehat{ heta}_j$	-1(0)	-1(0)	-2(0)	-2(0)	-3(0)	-5(1)	-7(3)	-9(9)
WCCF	-1(0)	-1(0)	-2(0)	-2(0)	-3(0)	-5(1)	-7(3)	-9(9)
				$\pi_2 =$	1/2			
$\widehat{ heta}_j$	-1(0)	-1(0)	-2(0)	-2(0)	-3(0)	-5(1)	-7(3)	-9(9)
WCCF	-1(0)	-1(0)	-1(0)	-1(0)	-2(0)	-4(1)	-6(3)	-8(9)
	$\pi_2 = 3/4$							
$\widehat{ heta}_j$	-1(0)	-1(0)	-2(0)	-2(0)	-3(0)	-5(1)	-7(3)	-9(9)
WCCF	-1 (0)	-1(0)	-1(0)	0(0)	0(0)	-2(1)	-4(3)	-6 (9)

This table reports the median and the median absolute deviation (in parentheses) of the estimates in Scenario 2 (divided by  $\tau_N$ ).

## 6 Empirical application

In this section we apply the proposed method to actual high-frequency data in the U.S. stock market. We investigate the lead-lag relationships between transactions of a single asset traded concurrently at multiple exchanges.<sup>3</sup> We select four assets; Apple (AAPL), Cisco Systems (CSCO), Intel (INTC) and Microsoft (MSFT). They are among stocks that are listed on the NASDAQ exchange and are at the same time constituents of the 30 Dow Jones Industrial Average (DJIA) stocks as of August 2016. The exchanges chosen for the current analysis are the NASDAQ, NYSE Arca and BATS. We use trade data recorded between 9:30 and 16:00. The source is the Daily TAQ database, whose time resolution is one micro-second. The sample period is the whole month of August 2016, consisting of 21 trading days. We report in Table 4 the average daily numbers of trades of each asset on each exchange.

We use L=20 as the length of Daubechies' wavelet filter and set

$$\mathcal{G}^N = \{-2\text{ms}, -1.999\text{ms}, \dots, 1.999\text{ms}, 2\text{ms}\}$$

as the search grid.

For comparison we also compute the following two estimators for lead-lag times.

• Hoffmann-Rosenbaum-Yoshida (HRY) estimator [21]: This estimator is defined as a maximizer of  $\widehat{U}^N(\theta)$  over the grid  $\theta \in \mathcal{G}^N$ :

$$\widehat{\theta}^{HRY} = \arg\max_{\theta \in \mathcal{G}^N} |\widehat{U}^N(\theta)|.$$

• Dobrev-Schaumburg (DS) estimator [10]: This estimator is constructed as follows. For each  $\nu=1,2$  and each  $t\geq 0$ , we set  $I^{\nu}_t=1$  if  $t\in\{t^{\nu}_i:i=0,1,\ldots,n_{\nu}\}$  and  $I^{\nu}_t=0$  otherwise. Then we define

$$A(\theta) = \frac{1}{\min\{n_1, n_2\}} \sum_{k=1}^{\infty} I_{k\tau_N}^1 I_{k\tau_N + \theta}^2$$

for each  $\theta \in \mathbb{R}$ . Now, the DS estimator  $\widehat{\theta}^{DS}$  is defined as a maximizer of  $A(\theta)$  over the grid  $\mathcal{G}^N$ :

$$\widehat{\theta}^{DS} = \arg\max_{\theta \in \mathcal{G}^N} A(\theta).$$

Figures 1–3 are daily time series plots of the estimates  $\widehat{\theta}^{HRY}$ ,  $\widehat{\theta}_j$  ( $1 \leq j \leq 10$ ) and  $\widehat{\theta}^{DS}$  for INTC, evaluated every trading day. We can see that the estimates of  $\widehat{\theta}_j$  are substantially stable at the level j=10, which corresponds to the time scale between 1.024ms and 2.048ms. The estimates of  $\widehat{\theta}^{HRY}$  are scattered, exhibiting no regular patterns. The estimates of  $\widehat{\theta}^{DS}$  are mostly stable, suggesting the presence of consistent lead-lag relationships in trading activity across the three markets.

Tables 5–7 report the sample medians and the sample median absolute deviations (in parentheses) of the estimates  $\hat{\theta}^{HRY}$ ,  $\hat{\theta}_{i}$  ( $1 \leq j \leq 10$ ) and  $\hat{\theta}^{DS}$  over the whole sample period for the three pairs of the exchanges.

<sup>&</sup>lt;sup>3</sup>This is closely related to the issue of identifying the particular exchange at which *price discovery* of the assets actually takes place. Such an issue is one of the fundamental problems in financial econometrics and has been widely studied in the literature; see e.g. [13, 14, 29, 33].

We can find that the estimators  $\hat{\theta}_j$  tend to give more stable estimates than  $\hat{\theta}^{HRY}$ . In particular, the estimates of the  $\hat{\theta}_j$  at the levels j=9,10 are substantially stable in many cases.

These findings perhaps suggest that there are some market participants who are very active in trading on all the three markets for all of these four assets at the same time, operating with the time scale between 0.512ms and 2.048ms. Their aggregate activities could cause systematic lead-lag relationships which are reported here. We imagine that these participants are "machines" operated by *high-frequency trading* firms.

Interestingly, we further find that for the pairs NASDAQ-NYSE Arca and NYSE Arca-BATS the estimates of the  $\widehat{\theta}_j$  at the levels j=9,10 are relatively closer to those of  $\widehat{\theta}^{DS}$ . Be aware that the estimators  $\widehat{\theta}_j$  measure the lead-lag times between the asset *returns*, while the estimator  $\widehat{\theta}^{DS}$  measures the lead-lag times between the transaction *times* of the assets; they are not the identical quantities. Nevertheless, if the above mentioned aggressive market participants indeed cause the lead-lag relationships then the two lead-lag estimators can behave similarly.

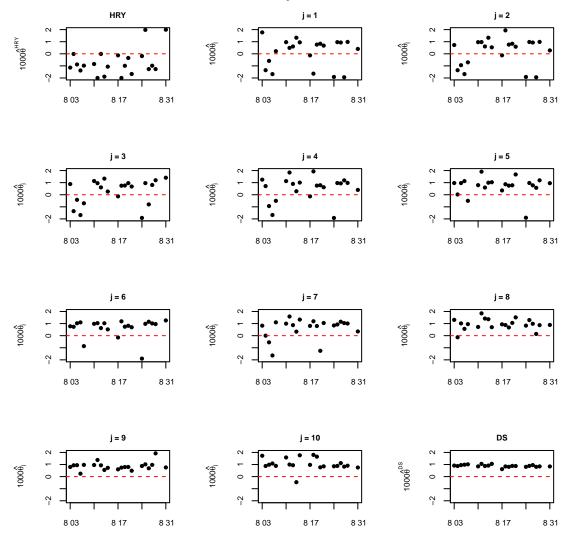
In contrast, the estimates of  $\widehat{\theta}^{DS}$  for the pair NASDAQ-BATS seem to capture a "deterministic," or "mechanical," lead-lag relationship. In order to focus on "stochastic" lead-lag relationships captured by the DS estimator, we re-compute  $\widehat{\theta}^{DS}$  for this pair using the *restricted* grid by removing the lags over -0.1ms and 0.1ms from the original,  $\mathcal{G}^N = \{-2\text{ms}, -1.999\text{ms}, \dots, -0.101\text{ms}, 0.101\text{ms}, \dots, 1.999\text{ms}, 2\text{ms}\}$ . The results are reported in Table 8. We can see that the DS estimates in this case get rather closer to those of  $\widehat{\theta}_j$  with j=10, thus the above remark seems also valid for the pair NASDAQ-BATS as well.

Overall, the estimates of our estimators  $\hat{\theta}_j$  as well as the DS estimator  $\hat{\theta}^{DS}$  indicate the following lead-lag relationships between the three exchanges: The NASDAQ exchange is the fastest, the NYSE Arca exchange is the slowest, and the BATS exchange is in the middle. Since all the four assets considered in this study are listed on the NASDAQ exchange, our finding that the NASDAQ is the fastest may not be surprising; the primary listing exchange typically dominates the *price discovery* process (cf. [13, 14, 29, 33]).

Table 4: The average daily numbers of transactions

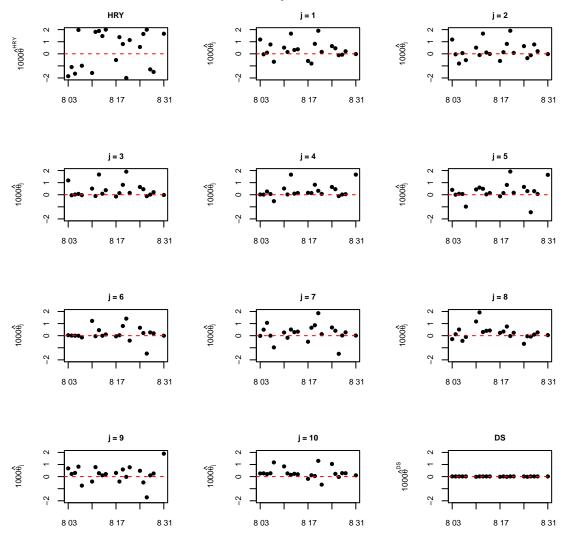
	AAPL	CSCO	INTC	MSFT
NASDAQ	84476	23710	27224	45072
NYSE Arca	58730	13722	18386	24984
BATS	50988	14959	18485	25150

Figure 1: Daily time series of the estimates  $\hat{\theta}^{HRY}$ ,  $\hat{\theta}_j$  ( $1 \leq j \leq 10$ ) and  $\hat{\theta}^{DS}$  for INTC: NASDAQ vs Arca



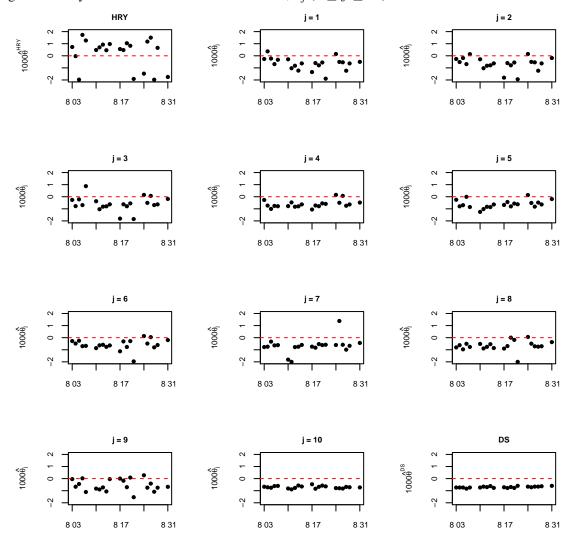
In this figure we depict daily time series of the estimates  $\widehat{\theta}^{HRY}$ ,  $\widehat{\theta}_j$  ( $1 \leq j \leq 10$ ) and  $\widehat{\theta}^{DS}$  for the INTC transaction prices between the NASDAQ and the NYSE Arca exchanges. The upper-left figure corresponds to  $\widehat{\theta}^{HRY}$ , while the lower-right figure corresponds  $\widehat{\theta}^{DS}$ . The remaining figures correspond to  $\widehat{\theta}_j$  for  $j=1,\ldots,10$ . The horizontal axis represents dates. The vertical axis is expressed in mili-seconds. The red dash line denotes 0ms. The positive value imply that the NASDAQ leads the NYSE Arca and vice versa.

Figure 2: Daily time series of the estimates  $\widehat{\theta}^{HRY}$ ,  $\widehat{\theta}_j$  ( $1 \leq j \leq 10$ ) and  $\widehat{\theta}^{DS}$  for INTC: NASDAQ vs BATS



In this figure we depict daily time series of the estimates  $\widehat{\theta}^{HRY}$ ,  $\widehat{\theta}_j$  ( $1 \leq j \leq 10$ ) and  $\widehat{\theta}^{DS}$  for the INTC transaction prices between the NASDAQ and the BATS exchanges. The upper-left figure corresponds to  $\widehat{\theta}^{HRY}$ , while the lower-right figure corresponds  $\widehat{\theta}^{DS}$ . The remaining figures correspond to  $\widehat{\theta}_j$  for  $j=1,\ldots,10$ . The horizontal axis represents dates. The vertical axis is expressed in mili-seconds. The red dash line denotes 0ms. The positive value imply that the NASDAQ leads the BATS and vice versa.

Figure 3: Daily time series of the estimates  $\widehat{\theta}^{HRY}$ ,  $\widehat{\theta}_j$  ( $1 \leq j \leq 10$ ) and  $\widehat{\theta}^{DS}$  for INTC: Arca vs BATS



In this figure we depict daily time series of the estimates  $\widehat{\theta}^{HRY}$ ,  $\widehat{\theta}_j$  ( $1 \leq j \leq 10$ ) and  $\widehat{\theta}^{DS}$  for the INTC transaction prices between the NASDAQ and the NYSE Arca exchanges. The upper-left figure corresponds to  $\widehat{\theta}^{HRY}$ , while the lower-right figure corresponds  $\widehat{\theta}^{DS}$ . The remaining figures correspond to  $\widehat{\theta}_j$  for  $j=1,\ldots,10$ . The horizontal axis represents dates. The vertical axis is expressed in mili-seconds. The red dash line denotes 0ms. The positive value imply that the NYSE Arca leads the BATS and vice versa.

Table 5: The medians and MADs of the estimates: NASDAQ vs Arca

	AAPL	CSCO	INTC	MSFT
HRY	0.041 (0.200)	-0.745 (0.933)	-0.986 (0.950)	-0.628 (0.636)
j = 1	0.536 (1.029)	0.539 (1.082)	0.605 (0.557)	0.940 (0.691)
j = 2	0.872 (0.469)	0.900 (0.683)	0.607 (0.549)	0.992 (0.710)
j = 3	0.906 (0.337)	0.880 (0.572)	0.748 (0.669)	0.929 (0.239)
j = 4	0.862 (0.603)	0.910 (0.412)	0.786 (0.575)	0.847 (0.565)
j = 5	0.926 (0.291)	0.842 (0.476)	0.868 (0.246)	0.793 (0.498)
j = 6	0.899 (0.394)	0.894 (0.461)	0.960 (0.286)	0.980 (0.222)
j = 7	0.960 (0.405)	0.903 (0.715)	0.871 (0.344)	0.878 (0.297)
j = 8	0.855 (0.409)	0.725 (0.614)	0.936 (0.351)	1.080 (0.302)
j = 9	0.874 (0.381)	1.043 (0.307)	0.797 (0.219)	0.987 (0.199)
j = 10	0.799 (0.654)	0.924 (0.215)	0.940 (0.170)	0.944 (0.154)
DS	0.839 (0.007)	0.885 (0.053)	0.875 (0.053)	0.839 (0.053)

This table reports the sample medians and the sample median absolute deviations (in parentheses) of the estimates  $\widehat{\theta}^{HRY}$  (HRY),  $\widehat{\theta}_j$  ( $j=1,\ldots,10$ ) and  $\widehat{\theta}^{DS}$  (DS) over the whole sample period when  $X^1$  is the transaction price process executed on the NASDAQ exchange and  $X^2$  is the one executed on the NYSE Arca exchange.

Table 6: The medians and MADs of the estimates: NASDAQ vs BATS

	AAPL	CSCO	INTC	MSFT
HRY	0.388 (0.645)	-0.615 (2.033)	0.806 (1.760)	-0.381 (2.159)
j = 1	0.123 (0.519)	0.134 (0.519)	0.210 (0.443)	0.016 (0.329)
j = 2	0.080 (0.322)	0.256 (0.789)	0.071 (0.639)	0.236 (0.331)
j = 3	0.034 (0.391)	0.244 (0.338)	0.131 (0.353)	0.059 (0.342)
j = 4	0.108 (0.203)	0.139 (0.470)	0.139 (0.188)	0.050 (0.270)
j = 5	0.173 (0.224)	0.150 (0.289)	0.166 (0.350)	0.088 (0.308)
j = 6	0.061 (0.222)	0.166 (0.279)	0.035 (0.234)	0.102 (0.294)
j = 7	0.075 (0.328)	0.085 (0.388)	0.274 (0.409)	0.066 (0.230)
j = 8	0.088 (0.295)	0.340 (0.403)	0.227 (0.385)	0.154 (0.322)
j = 9	0.167 (0.083)	0.239 (0.248)	0.256 (0.494)	0.249 (0.136)
j = 10	0.139 (0.085)	0.235 (0.142)	0.230 (0.130)	0.190 (0.188)
DS	0.015 (0.000)	0.015 (0.000)	0.015 (0.000)	0.015 (0.000)

This table reports the sample medians and the sample median absolute deviations (in parentheses) of the estimates  $\widehat{\theta}^{HRY}$  (HRY),  $\widehat{\theta}_j$  ( $j=1,\ldots,10$ ) and  $\widehat{\theta}^{DS}$  (DS) over the whole sample period when  $X^1$  is the transaction price process executed on the NASDAQ exchange and  $X^2$  is the one executed on the BATS exchange.

Table 7: The medians and MADs of the estimates: Area vs BATS

	AAPL	CSCO	INTC	MSFT
HRY	-0.104 (0.307)	0.771 (0.930)	0.657 (0.566)	1.109 (0.491)
j = 1	-0.747 (0.224)	-0.736 (0.771)	-0.602 (0.406)	-0.673 (0.715)
j = 2	-0.744 (0.154)	-0.740 (0.403)	-0.602 (0.329)	-0.620 (0.704)
j = 3	-0.746 (0.206)	-0.719 (0.350)	-0.627 (0.286)	-0.685 (0.715)
j = 4	-0.753 (0.212)	-0.718 (0.455)	-0.726 (0.151)	-0.778 (0.221)
j = 5	-0.788 (0.077)	-0.713 (0.363)	-0.640 (0.265)	-0.676 (0.383)
j = 6	-0.848 (0.252)	-0.638 (0.283)	-0.609 (0.283)	-0.697 (0.277)
j = 7	-0.733 (0.203)	-0.658 (0.400)	-0.641 (0.160)	-0.721 (0.292)
j = 8	-0.752 (0.289)	-0.678 (0.208)	-0.709 (0.270)	-0.711 (0.423)
j = 9	-0.740 (0.132)	-0.645 (0.199)	-0.679 (0.568)	-0.699 (0.165)
j = 10	-0.713 (0.173)	-0.753 (0.297)	-0.714 (0.105)	-0.787 (0.193)
DS	-0.710 (0.022)	-0.700 (0.099)	-0.715 (0.076)	-0.669 (0.053)

This table reports the sample medians and the sample median absolute deviations (in parentheses) of the estimates  $\widehat{\theta}^{HRY}$  (HRY),  $\widehat{\theta}_j$  ( $j=1,\ldots,10$ ) and  $\widehat{\theta}^{DS}$  (DS) over the whole sample period when  $X^1$  is the transaction price process executed on the NYSE Arca exchange and  $X^2$  is the one executed on the BATS exchange.

Table 8: The medians and MADs of the DS estimates (different grid): NASDAQ vs BATS

	AAPL	CSCO	INTC	MSFT
DS	0.139 (0.046)	0.247 (0.039)	0.247 (0.053)	0.170 (0.061)

This table reports the sample medians and the sample median absolute deviations (in parentheses) of the estimates  $\widehat{\theta}^{DS}$  (DS) with the search grid  $\mathcal{G}^N=\{-2\mathrm{ms},-1.999\mathrm{ms},\dots,-0.101\mathrm{ms},0.101\mathrm{ms},\dots,1.999\mathrm{ms},2\mathrm{ms}\}$  over the whole sample period when  $X^1$  is the transaction price process executed on the NASDAQ exchange and  $X^2$  is the one executed on the BATS exchange.

#### 7 Conclusion

In this paper we have proposed a new estimation method for multi-scale analysis of lead-lag relationships between two assets based on their high-frequency observation data when they are non-synchronously observed. The new method is based on the novel estimator for the scale-by-scale cross-covariance functions  $\rho_{(j)}(\theta)$  that are constructed via a variant of wavelet transform of the empirical cross-covariance function used in [21]. We have also developed an associated asymptotic theory to obtain consistency of the proposed estimators in the modeling framework proposed in our previous work [16]. Compared with the estimation method proposed in [16], which essentially adopts the same method as the traditional one in the wavelet literature, the newly proposed method is more appropriate in applications to high-frequency financial data from a computational point of view. The simulation study has shown that the proposed estimators are also substantially more suitable for non-synchronously observed data than the previous one, as intended. The empirical results have demonstrated that the new method can provide a deep insight into lead-lag relationships in the financial markets in the high-frequency domain.

#### 8 Proofs

Throughout the discussions, for sequences  $(x_N)$  and  $(y_N)$ ,  $x_N \lesssim y_N$  means that there exists a constant  $C \in [0, \infty)$  such that  $x_N \leq Cy_N$  for large N. Also, for  $r > 0 \| \cdot \|_r$  denotes the  $L^r$ -norm of random variables, i.e.  $\|Z\|_r = (E[|Z|^r])^{1/r}$  for a random variable Z.

## 8.1 Proof of Proposition 1

We begin by proving some lemmas. Let us set  $\tilde{\Pi}_N^{\nu} = \Pi_N^{\nu} \cup \{(0, t_0^{\nu}], (t_{n_1}^{\nu}, T + \delta]\}$  for  $\nu = 1, 2$  and  $\tilde{r}_N = (\sup_{I \in \tilde{\Pi}_N^1} |I|) \vee (\sup_{J \in \tilde{\Pi}_N^2} |J|)$ . We denote by  $P^{\Pi^1}$  (resp.  $E^{\Pi^1}$ ) the conditional probability (resp. conditional expectation) given  $(t_i^1)_{i=0}^{n_1}$ .

**Lemma 1.** 
$$P(\tau_N^{-1}\tilde{r}_N > x) \le C\tau_N^{-1}e^{-x/C}$$

This lemma can be shown in a similar manner to the proof of Lemma 4 from [5], so we omit the proof.

**Lemma 2.** Let  $\varpi \in (0,1)$  and set  $q = \lceil \varpi N \rceil$ . Suppose that  $\theta_N \ge 0$  for all N. For any M > 0, there is a constant  $C_M > 0$  such that

$$\left| E^{\Pi^1} \left[ \frac{1}{2\pi \tau_m \tau_N} \sum_{J \in \tilde{\Pi}_N^2} (\mathcal{F}1_I) (\lambda/\tau_N) (\mathcal{F}1_{J-\theta_N}) (-\lambda/\tau_N) K(I, J_{-\theta_N}) \right] - \frac{\tau_N}{\pi \tau_m \lambda^2} \Re \left[ \left( 1 - e^{-\sqrt{-1}\lambda \tau_N^{-1}|I|} \right) \frac{1 - \pi_2}{1 - \pi_2 e^{-\sqrt{-1}\lambda}} \right] \right| \le C_M \tau_N^M$$

for any  $N \in \mathbb{N}$ , any  $\lambda \in \mathbb{R}$  and any  $I \in \tilde{\Pi}_N^1$  such that  $T\tau_q \leq \underline{I} < \overline{I} \leq T(1-\tau_q)$ .

Proof. Set

$$J_I = \bigcup_{J \in \tilde{\Pi}_N^2: K(I, J_{-\theta_N}) = 1} J_{-\theta_N}.$$

Then we have

$$\begin{split} &\frac{1}{2\pi\tau_{m}\tau_{N}}\sum_{J\in\tilde{\Pi}_{N}^{2}}(\mathcal{F}1_{I})(\lambda/\tau_{N})(\mathcal{F}1_{J_{-\theta}})(-\lambda/\tau_{N})K(I,J_{-\theta_{N}})\\ &=\frac{1}{2\pi\tau_{m}\tau_{N}}(\mathcal{F}1_{I})(\lambda/\tau_{N})\left\{(\mathcal{F}1_{[\underline{J_{I}},\underline{I}})(-\lambda/\tau_{N})+(\mathcal{F}1_{I})(-\lambda/\tau_{N})+(\mathcal{F}1_{[\overline{I},\overline{J_{I}})})(-\lambda/\tau_{N})\right\}\\ &=:\mathbf{I}+\mathbf{II}+\mathbf{III}. \end{split}$$

First we consider I. We can rewrite it as

$$\mathbf{I} = \frac{\tau_N}{2\pi\tau_m\lambda^2} \left( e^{-\sqrt{-1}\lambda\tau_N^{-1}|I|} - 1 \right) \left( 1 - e^{-\sqrt{-1}\lambda\tau_N^{-1}(\underline{I} - \underline{J}_I)} \right).$$

Conditionally on  $(t_i^1)_{i=0}^{n_1}$ ,  $\tau_N^{-1}(\underline{I}-\underline{J_I})$  follows the geometric distribution with success probability  $1-\pi_2$  truncated from above by  $\tau_N^{-1}\underline{I}$ . More precisely, we have

$$P^{\Pi^{1}}(\tau_{N}^{-1}(\underline{I}-\underline{J_{I}})=k) \begin{cases} \pi_{2}^{k}(1-\pi_{2}) & \text{if } 0 \leq k < \tau_{N}^{-1}\underline{I}, \\ \pi_{2}^{\tau_{N}^{-1}}\underline{I} & \text{if } k=\tau_{N}^{-1}\underline{I}. \end{cases}$$

Therefore, we obtain

$$\begin{split} E^{\Pi^{1}} \left[ 1 - e^{-\sqrt{-1}\lambda\tau_{N}^{-1}(\underline{I} - \underline{J}_{\underline{I}})} \right] \\ &= 1 - (1 - \pi_{2}) \sum_{k=0}^{\tau_{N}^{-1}\underline{I} - 1} \pi_{2}^{k} e^{-\sqrt{-1}\lambda k} - \pi_{2}^{\tau_{N}^{-1}\underline{I}} e^{-\sqrt{-1}\lambda\tau_{N}^{-1}\underline{I}} \\ &= 1 - \pi_{2}^{\tau_{N}^{-1}\underline{I}} - \frac{(1 - \pi_{2})(1 - \pi_{2}^{\tau_{N}^{-1}\underline{I}} e^{-\sqrt{-1}\lambda\tau_{N}^{-1}\underline{I}})}{1 - \pi_{2}e^{-\sqrt{-1}\lambda}} + \pi_{2}^{\tau_{N}^{-1}\underline{I}} (1 - e^{-\sqrt{-1}\lambda\tau_{N}^{-1}\underline{I}}) \\ &= \frac{\pi_{2}(1 - e^{\sqrt{-1}\lambda})}{1 - \pi_{2}e^{-\sqrt{-1}\lambda}} - \pi_{2}^{\tau_{N}^{-1}\underline{I}} \frac{1 - e^{-\sqrt{-1}\lambda\tau_{N}^{-1}\underline{I}} + \pi_{2}(1 - e^{-\sqrt{-1}\lambda})}{1 - \pi_{2}e^{-\sqrt{-1}\lambda}} + \pi_{2}^{\tau_{N}^{-1}\underline{I}} (1 - e^{-\sqrt{-1}\lambda\tau_{N}^{-1}\underline{I}}). \end{split}$$

Consequently, we have

$$\left| E^{\Pi^{1}}[\mathbf{I}] - \frac{\tau_{N}}{2\pi\tau_{m}\lambda^{2}} \left( e^{-\sqrt{-1}\lambda\tau_{N}^{-1}|I|} - 1 \right) \frac{\pi_{2}(1 - e^{-\sqrt{-1}\lambda})}{1 - \pi_{2}e^{-\sqrt{-1}\lambda}} \right| \lesssim \tau_{N}^{M}$$

uniformly in  $\lambda \in \mathbb{R}$  and  $I \in \tilde{\Pi}_N^1$  such that  $T\tau_q \leq \underline{I} < \overline{I} \leq T(1-\tau_q)$ . Here, we use the inequality  $|1-e^{-\sqrt{-1}x}| \leq |x|$  holding for all  $x \in \mathbb{R}$  and Lemma 1.

Next we consider III. We can rewrite it as

$$\mathbf{III} = \frac{\tau_N}{2\pi\tau_m\lambda^2} \left(1 - e^{\sqrt{-1}\lambda\tau_N^{-1}|I|}\right) \left(e^{\sqrt{-1}\lambda\tau_N^{-1}(\overline{J_I} - \overline{I})} - 1\right).$$

Now, an analogous argument to the above yields

$$\left| E^{\Pi^1}[\mathbf{III}] - \frac{\tau_N}{2\pi\tau_m\lambda^2} \left( 1 - e^{\sqrt{-1}\lambda\tau_N^{-1}|I|} \right) \frac{\pi_2(e^{\sqrt{-1}\lambda} - 1)}{1 - \pi_2e^{\sqrt{-1}\lambda}} \right| \lesssim \tau_N^M$$

uniformly in  $\lambda \in \mathbb{R}$  and  $I \in \tilde{\Pi}_N^1$  such that  $T\tau_q \leq \underline{I} < \overline{I} \leq T(1-\tau_q)$ . Hence we have

$$\left| E^{\Pi^1}[\mathbf{III}] - \overline{E^{\Pi^1}[\mathbf{I}]} \right| \lesssim \tau_N^M$$

uniformly for  $\lambda \in \mathbb{R}$  and  $I \in \tilde{\Pi}_N^1$  such that  $T\tau_q \leq \underline{I} < \overline{I} \leq T(1-\tau_q)$ 

Finally, we have

$$E^{\Pi_N^1}[\mathbf{II}] = \mathbf{II} = \frac{\tau_N}{\pi \tau_m \lambda^2} \left( 1 - \Re \left[ e^{-\sqrt{-1}\lambda \tau_N^{-1}|I|} \right] \right)$$

Therefore, a simple computation yields the desired result.

*Proof of Proposition 1.* Assumption 2(i) follows from Lemma 1. Assumption 2(iii) is evidently satisfied by definition.

Take a constant  $\beta \in (0,1)$  arbitrarily. We prove with this  $\beta$  that there are constants  $\alpha, Q > 1$  such that Assumption 2(ii) holds true. For simplicity of exposition, we assume  $\theta_N \geq 0$  for all N (this assumption can be easily removed).

Set

$$\tilde{D}_k^N(\lambda,\theta) = \frac{1}{2\pi\tau_m\tau_N} \sum_{I \in \tilde{\Pi}_N^1, J \in \tilde{\Pi}_N^2: \underline{I} \in I_m(k)} (\mathcal{F}1_I)(\lambda/\tau_N)(\mathcal{F}1_{J_{-\theta}})(-\lambda/\tau_N)K(I,J_{-\theta}).$$

It is obvious that

$$\tau_m \sum_{k=0}^{\lceil T\tau_m^{-1} \rceil - 1} \int_{-\pi}^{\pi} E\left[ \left| D_k^N(\lambda, \theta_N) - \tilde{D}_k^N(\lambda, \theta_N) \right|^p \right] d\lambda = O(\tau_N^p \tau_m^{p-1})$$

as  $N \to \infty$  for any p > 1. Therefore, it suffices to show that there are constants  $\alpha, Q > 1$  such that

$$\tau_m \sum_{k=0}^{\lceil T\tau_m^{-1} \rceil - 1} \int_{-\pi}^{\pi} E\left[ \left| \tilde{D}_k^N(\lambda, \theta_N) - D_k^N(\lambda, \theta_N) \right|^Q \right] d\lambda = O(\tau_N^{\alpha}) \tag{4}$$

as  $N \to \infty$ . For the proof we adopt a similar strategy to the proof of Proposition 6 from [5]. Let  $\varpi$  be a number such that  $\beta < \varpi < 1$  and set  $q = \lceil N\varpi \rceil$ . Let  $\mathcal E$  be the event on which the interval  $I_{q+2}(u)$  contains at least one point from  $\{t_i^1: i=0,1,\ldots,n_1\}$  and one point from  $\{t_i^2: i=0,1,\ldots,n_2\}$  for every  $u=0,\tau_m\tau_{q+2}^{-1},\ldots,T\tau_{q+2}^{-1}-1$ . We have

$$P(\mathcal{E}^c) \le \tau_{q+2}^{-1} (\pi_1^{\tau_N^{-1} \tau_{q+2}} + \pi_2^{\tau_N^{-1} \tau_{q+2}}). \tag{5}$$

In the following we denote by  $E^A$  the conditional expectation given an event A.

For  $u \in \mathbb{Z}_+$  and  $\lambda \in \mathbb{R}$ , we set

$$\eta_u^N(\lambda) = \frac{1}{2\pi\tau_m\tau_N} \sum_{I \in \tilde{\Pi}_N^1, J \in \tilde{\Pi}_N^2: \underline{I} \in I_q(u)} (\mathcal{F}1_I)(\lambda/\tau_N)(\mathcal{F}1_{J_{-\theta}})(-\lambda/\tau_N)K(I, J_{-\theta}).$$

Then, we decompose  $\tilde{D}_k^N(\lambda, \theta_N) - D(\lambda)$  as

$$\tilde{D}_k^N(\lambda, \theta_N) - D(\lambda) = \left\{ E^{\mathcal{E}} \left[ \sum_{u=k\tau_m \tau_q^{-1}}^{(k+1)\tau_m \tau_q^{-1} - 1} \eta_u^N(\lambda) \right] - D(\lambda) \right\}$$

$$\begin{split} &+\sum_{\substack{u=k\tau_m\tau_q^{-1}\\u\text{ is odd}}}^{(k+1)\tau_m\tau_q^{-1}-1}\left(\eta_u^N(\lambda)-E^{\mathcal{E}}[\eta_u^N(\lambda)]\right)+\sum_{\substack{u=k\tau_m\tau_q^{-1}\\u\text{ is even}}}^{(k+1)\tau_m\tau_q^{-1}-1}\left(\eta_u^N(\lambda)-E^{\mathcal{E}}[\eta_u^N(\lambda)]\right)\\ =:\mathbb{I}_k^N(\lambda)+\mathbb{II}_k^N(\lambda)+\mathbb{III}_k^N(\lambda). \end{split}$$

First we consider  $\mathbb{I}_k^N(\lambda)$ . Using the inequality  $|e^{\sqrt{-1}x}-1|\leq |x|$  holding for  $x\in\mathbb{R}$ , we have

$$|\eta_u^N(\lambda)| \lesssim \tau_m^{-1} \tau_N (\tau_N^{-1} r_N)^2 \# \{ I \in \tilde{\Pi}_N^1 : \underline{I} \in I_q(u) \}$$

uniformly in  $\lambda \in \mathbb{R}$  and  $u \in \mathbb{Z}_+$ . Therefore, noting that  $t_i^1 - t_{i-1}^1 \ge \tau_N$  for every  $i = 1, \dots, n_1 - 1$ , we obtain

$$|\eta_u^N(\lambda)| \lesssim \tau_m^{-1} \tau_q(\tau_N^{-1} r_N)^2 \lesssim \tau_N^{\varpi - \beta} (\tau_N^{-1} r_N)^2 \tag{6}$$

uniformly in  $\lambda \in \mathbb{R}$  and  $u \in \mathbb{Z}_+$ . Now, similarly to the proof of Eq.(36) form [5], we can prove

$$E[r_N^p] = O(\tau_N^p |\log \tau_N|^p) \tag{7}$$

for any p > 0. Hence we have

$$\left| E^{\mathcal{E}} \left[ \sum_{u=k\tau_{m}\tau_{q}^{-1}}^{(k+1)\tau_{m}\tau_{q}^{-1}-1} \eta_{u}^{N}(\lambda) \right] - E \left[ \sum_{u=k\tau_{m}\tau_{q}^{-1}}^{(k+1)\tau_{m}\tau_{q}^{-1}-1} \eta_{u}^{N}(\lambda) \right] \right| 
\lesssim \tau_{m}\tau_{q}^{-1} \left\{ E[\tau_{m}^{-1}\tau_{q}(\tau_{N}^{-1}r_{N})^{2}] P(\mathcal{E}^{c}) + E[\tau_{m}^{-1}\tau_{q}(\tau_{N}^{-1}r_{N})^{2}1_{\mathcal{E}^{c}}] \right\} 
\lesssim \tau_{N}^{2} \left\{ E[r_{N}^{2}] P(\mathcal{E}^{c}) + \sqrt{E[r_{N}^{4}]} \sqrt{P(\mathcal{E}^{c})} \right\} 
\lesssim |\log \tau_{N}|^{2} \tau_{m}\tau_{q+2}^{-1} \sqrt{\tau_{1}^{\tau_{N}^{-1}\tau_{q+2}} + \tau_{2}^{\tau_{N}^{-1}\tau_{q+2}}}$$

uniformly in  $\lambda \in \mathbb{R}$  and  $k \in \mathbb{Z}_+$ . Moreover, Lemma 2 and (6)–(7) imply that

$$E\left[\sum_{u=k\tau_{m}\tau_{q}^{-1}-1}^{(k+1)\tau_{m}\tau_{q}^{-1}-1}\eta_{u}^{N}(\lambda)\right]$$

$$=\frac{\tau_{N}}{\pi\tau_{m}\lambda^{2}}\Re\left[E\left[\sum_{I\in\tilde{\Pi}_{N}^{1}:I\in I_{m}(k)}\left(1-e^{-\sqrt{-1}\lambda\tau_{N}^{-1}|I|}\right)\right]\frac{1-\pi_{2}}{1-\pi_{2}e^{-\sqrt{-1}\lambda}}\right]+O(\tau_{N}^{\varpi-\beta}|\log\tau_{N}|^{2})$$

uniformly in  $\lambda$  and k. Since  $(t_i^1 - t_{i-1}^1)_{i=1}^{n_1}$  is a sequence of i.i.d. variables whose distributions are the geometric distribution with success probability  $1 - \pi_1$ , the Wald identity yields

$$E\left[\sum_{u=k\tau_{m}\tau_{q}^{-1}-1}^{(k+1)\tau_{m}\tau_{q}^{-1}-1}\eta_{u}^{N}(\lambda)\right]$$

$$=\frac{\tau_{N}}{\pi\tau_{m}\lambda^{2}}\Re\left[E\left[\#\{I\in\tilde{\Pi}_{N}^{1}:\underline{I}\in I_{m}(k)\}\right]\frac{(1-\pi_{2})(1-e^{-\sqrt{-1}\lambda})}{(1-\pi_{1}e^{-\sqrt{-1}\lambda})(1-\pi_{2}e^{-\sqrt{-1}\lambda})}\right]+O(\tau_{N}^{\varpi-\beta}|\log\tau_{N}|^{2})$$

$$= D(\lambda) + O(\tau_N^{\varpi - \beta} |\log \tau_N|^2)$$

uniformly in  $\lambda$  and k. Consequently, we obtain

$$E\left[\left|\mathbb{I}_{k}^{N}(\lambda)\right|^{p}\right] = \left|\mathbb{I}_{k}^{N}(\lambda)\right|^{p} = O(\tau_{N}^{(\varpi-\beta)p}|\log \tau_{N}|^{2p})$$

uniformly in  $\lambda$  and k for any p > 1.

Next we consider  $\mathbb{II}_k^n(\lambda)$ . By construction  $(\eta_u^N(\lambda))_{u: \text{odd}}$  is independent conditionally to  $\mathcal{E}$ . Therefore, the Burkholder-Davis-Gundy (henceforth BDG) inequality and (6)–(7) yield

$$E^{\mathcal{E}}\left[\left|\mathbb{II}_{k}^{N}(\lambda)\right|^{p}\right] \lesssim (\tau_{m}\tau_{q}^{-1})^{p/2}E\left[\left|\tau_{m}^{-1}\tau_{q}(\tau_{N}^{-1}r_{N})^{2}\right|^{p}\right] \lesssim (\tau_{m}^{-1}\tau_{q})^{p/2}|\log \tau_{N}|^{2p}$$

uniformly in  $\lambda$  and k for any p > 1. Moreover, (5)–(7) imply that

$$E^{\mathcal{E}^c}\left[\left|\mathbb{II}_k^N(\lambda)\right|^p\right] = O((\tau_m^{-1}\tau_q)^{p/2}|\log\tau_N|^{2p})$$

uniformly in  $\lambda$  and k for any p > 1. Consequently, we obtain

$$E\left[\left|\mathbb{II}_{k}^{N}(\lambda)\right|^{p}\right] = O((\tau_{m}^{-1}\tau_{q})^{p/2}|\log \tau_{N}|^{2p})$$

uniformly in  $\lambda$  and k for any p > 1. An analogous argument yields

$$E\left[\left|\mathbb{III}_{k}^{N}(\lambda)\right|^{p}\right] = O((\tau_{m}^{-1}\tau_{q})^{p/2}|\log \tau_{N}|^{2p})$$

uniformly in  $\lambda$  and k for any p > 1.

After all, we have

$$\tau_m \sum_{k=0}^{\lceil T\tau_m^{-1} \rceil - 1} \int_{-\pi}^{\pi} E\left[ \left| \tilde{D}_k^N(\lambda, \theta_N) - D(\lambda) \right|^p \right] d\lambda = O(\tau_N^{(\varpi - \beta)p/2} |\log \tau_N|^{2p})$$

for any p>1. Now we take Q>1 so that  $(\varpi-\beta)Q>4$  and set  $\alpha=(\varpi-\beta)Q/4$ . Then (4) holds true.

#### 8.2 Proof of Theorem 1

First we remark that a standard localization procedure presented e.g. at the beginning of Section 7.3 of [16] allows us to assume that there is a constant K > 0 such that

$$|\sigma_t^1| + |\sigma_t^2| \le K, \qquad |\sigma_t^1 - \sigma_s^1| + |\sigma_t^2 - \sigma_s^2| \le K|t - s|^{\gamma}$$

for any  $t, s \ge 0$  throughout the proof.

Next we introduce some notation. For each  $k \in \mathbb{Z}_+$  and  $\theta \in (-\delta, \delta)$ , we set

$$\widehat{U}_k^N(\theta) = \begin{cases} \sum_{I,J:\underline{I} \in I_m(k)} X^1(I) X^2(J) K(I, J_{-\theta}) & \text{if } \theta \ge 0, \\ \sum_{I,J:\underline{J} \in I_m(k)} X^1(I) X^2(J) K(I_{\theta}, J) & \text{if } \theta < 0 \end{cases}$$

and

$$c_k^N(\theta) = \begin{cases} \sigma_{k\tau_m}^1 \sigma_{(k\tau_m + \theta - r_N)_+}^2 & \text{if } \theta \ge 0, \\ \sigma_{(k\tau_m - \theta - r_N)_+}^1 \sigma_{k\tau_m}^2 & \text{if } \theta < 0 \end{cases}$$

and

$$\widetilde{U}_{k}^{N}(\theta) = \begin{cases} \sum_{I,J:\underline{I} \in I_{m}(k)} B^{1}(I)B^{2}(J)K(I,J_{-\theta}) & \text{if } \theta \ge 0, \\ \sum_{I,J:J\in I_{m}(k)} B^{1}(I)B^{2}(J)K(I_{\theta},J) & \text{if } \theta < 0 \end{cases}$$

and

$$\overline{U}_k^N(\theta) = \begin{cases} \sum_{I,J:\underline{I} \in I_m(k)} E^{\Pi} \left[ B^1(I)B^2(J) \right] K(I,J_{-\theta}) & \text{if } \theta \ge 0, \\ \sum_{I,J:\underline{J} \in I_m(k)} E^{\Pi} \left[ B^1(I)B^2(J) \right] K(I_{\theta},J) & \text{if } \theta < 0. \end{cases}$$

In the following we denote by  $E^{\Pi}$  the conditional expectation given  $(t_i^1)_{i=0}^{n_1}$  and  $(t_j^2)_{j=0}^{n_2}$ .

**Lemma 3.** For any p > 1, there is a constant  $C_p > 0$  such that

$$E^{\Pi} \left[ \left| \widehat{U}_k^N(\theta) - c_k^N(\theta) \widetilde{U}_k^N(\theta) \right|^p \right] \le C_p \tau_m^{(1+\gamma)p}$$

for any  $N \in \mathbb{N}$ ,  $k \in \mathbb{Z}_+$  and  $\theta \in (-\delta, \delta)$ .

**Proof.** By symmetry it is enough to consider the case of  $\theta \geq 0$ .

First we apply the so-called reduction procedures used in [19, 20] to every realization of  $(I)_{I\in\Pi_N^1}$  and  $(J_{-\theta})_{J\in\Pi_N^2}$  (see also the proof of Lemma 2 from [5]). We define a new partition  $\tilde{\Pi}_N^1$  as follows:  $I\in\tilde{\Pi}_N^1$  if and only if either  $I\in\Pi_N^1$  and it has non-empty intersection with two distinct intervals from  $\Pi_N^2$  or there is  $J\in\Pi_N^2$  such that I is the union of all intervals from  $\Pi_N^1$  included in J. We also define a new partition  $\tilde{\Pi}_N^2$  as follows:  $J\in\tilde{\Pi}_N^2$  if and only if either  $J\in\Pi_N^2$  and  $J_{-\theta}$  has non-empty intersection with two distinct intervals from  $\Pi_N^1$  or there is  $I\in\Pi_N^1$  such that J is the union of all intervals from  $J'\in\Pi_N^2$  such that  $J'_{-\theta}$  is included in I. Due to bilinearity both  $\hat{U}_k^N(\theta)$  and  $\tilde{U}_k^N(\theta)$  are invariant under this procedure.  $r_N$  is also unchanged by this application because of its definition. Moreover, by construction we have

$$\max_{J\in \tilde{\Pi}_N^2} \sum_{I\in \tilde{\Pi}_N^1} K(I,J_{-\theta}) \leq 3, \qquad \max_{I\in \tilde{\Pi}_N^1} \sum_{J\in \tilde{\Pi}_N^2} K(I,J_{-\theta}) \leq 3.$$

Consequently, for the proof we may replace  $(\Pi_N^1, \Pi_N^2)$  by  $(\tilde{\Pi}_N^1, \tilde{\Pi}_N^2)$ . This allows us to assume that

$$\max_{J \in \Pi_N^2} \sum_{I \in \Pi_N^1} K(I, J_{-\theta}) \le 3, \qquad \max_{I \in \Pi_N^1} \sum_{J \in \Pi_N^2} K(I, J_{-\theta}) \le 3.$$
 (8)

throughout the proof without loss of generality.

We turn to the main body of the proof. We decompose the target quantity as

$$\begin{split} \widehat{U}_{k}^{N}(\theta) - c_{k}^{N}(\theta) \widetilde{U}_{k}^{N}(\theta) \\ &= \sum_{I,J:\underline{I} \in I_{m}(k)} \left\{ \int_{I} (\sigma_{s}^{1} - \sigma_{k\tau_{m}}^{1}) dB_{s}^{1} X^{2}(J) + \sigma_{k\tau_{m}}^{1} B^{1}(I) \int_{J} (\sigma_{s}^{2} - \sigma_{(k\tau_{m} + \theta - r_{N})_{+}}^{2}) dB_{s}^{2} \right\} K(I, J_{-\theta}) \\ &=: \mathbf{A}_{N} + \mathbf{B}_{N}. \end{split}$$

Let us consider  $A_N$ . The Minkovski, Schwarz and BDG inequalities yield

$$E^{\Pi} [|\mathbf{A}_{N}|^{p}] \leq \left\{ \sum_{I,J:\underline{I} \in I_{m}(k)} \left( E^{\Pi} \left[ \left| \int_{I} (\sigma_{s}^{1} - \sigma_{k\tau_{m}}^{1}) dB_{s}^{1} X^{2}(J) \right|^{p} \right] \right)^{1/p} K(I,J_{-\theta}) \right\}^{p}$$

$$\leq \left\{ \sum_{I,J:\underline{I} \in I_{m}(k)} \left( E^{\Pi} \left[ \left| \int_{I} (\sigma_{s}^{1} - \sigma_{k\tau_{m}}^{1}) dB_{s}^{1} \right|^{2p} \right] \right)^{1/2p} \left( E^{\Pi} \left[ \left| X^{2}(J) \right|^{2p} \right] \right)^{1/2p} K(I,J_{-\theta}) \right\}^{p}$$

$$\lesssim \left\{ \sum_{I,J:\underline{I} \in I_{m}(k)} \left\| \sup_{s \in I_{m}(k)} |\sigma_{s}^{1} - \sigma_{k\tau_{m}}^{1}| \left\| \int_{2p} \sqrt{|I||J|} K(I,J_{-\theta}) \right\}^{p}, \right\}$$

hence by assumption and (8) we obtain

$$E^{\Pi}\left[|\mathbf{A}_N|^p\right] \lesssim \tau_m^{p\gamma} \left\{ \sum_{I,J:\underline{I} \in I_m(k)} (|I| + |J|) K(I,J_{-\theta}) \right\}^p \lesssim \tau_m^{p(\gamma+1)}.$$

By symmetry we also have  $E^{\Pi}[|\mathbf{B}_N|^p] \lesssim \tau_m^{p(\gamma+1)}$ . This completes the proof.

Let us take a number  $\xi \in (\frac{\beta+4}{5},1)$  and set  $u_N = (\tau_N^{(5\xi-4+\beta)/2}|\log \tau_N|)^{-1}$ 

**Lemma 4.** There is a constant C such that

$$E^{\Pi} \left[ \exp \left( \varsigma u_N \left\{ \widetilde{U}_k^N(\theta) - \overline{U}_k^N(\theta) \right\} \right) \right] \le C$$

for any  $N \in \mathbb{N}$ ,  $k \in \mathbb{Z}_+$ ,  $\varsigma \in \{-1, 1\}$  and  $\theta \in (-\delta, \delta)$ .

**Proof.** Again, by symmetry it suffices to consider the case of  $\theta \ge 0$ . Moreover, as in the proof of Lemma 3, we may assume (8) without loss of generality.

Let  $\Sigma_N$  be the covariance matrix of  $(B^1(I))_{I \in \Pi^1_N : \underline{I} \in I_m(k)}, B^2(J))_{J \in \Pi^2_N : \underline{J} \in \tilde{I}_m(k)})^{\top}$ , where  $\tilde{I}_m(k) = I_m(k) + \theta - r_N$ , and set  $C_N = \Sigma_N^{1/2} A_N \Sigma_N^{1/2}$ , where

$$A_N = \begin{pmatrix} 0 & K_N \\ K_N^\top & 0 \end{pmatrix}, \qquad K_N = (K(I, J_{-\theta})/2)_{(I,J) \in \Pi_N^1 \times \Pi_N^2 : \underline{I} \in I_m(k), \underline{J} \in \tilde{I}_m(k)}.$$

We first prove the following equations:

$$||C_N||_{\text{sp}} = O(\tau_N^{-2+3\xi} |\log \tau_N|^2), \qquad ||C_N||_F^2 = O(\tau_N^{-4+5\xi+\beta} |\log \tau_N|^2),$$
 (9)

where  $\|\cdot\|_{\rm sp}$  denotes the spectral norm of matrices. By Theorem 5.6.9 from [22] and (8), we have  $\|A_N\|_{\rm sp} \le \frac{3}{2}$ . Therefore, Appendix II(ii)–(iii) from [7] yield

$$\begin{split} \|C_N\|_F^2 &\leq \frac{9}{4} \|\Sigma_N\|_F^2 = \frac{9}{4} \sum_{I \in \Pi_N^1: \underline{I} \in I_m(k)} E^{\Pi} \left[ B^1(I)^2 \right]^2 + \frac{9}{4} \sum_{J \in \Pi_N^2: \underline{J} \in \tilde{I}_m(k)} E^{\Pi} \left[ B^2(J)^2 \right]^2 \\ &+ \frac{9}{2} \sum_{(I,J) \in \Pi_N^1 \times \Pi_N^2: \underline{I} \in I_m(k), \underline{J} \in \tilde{I}_m(k)} E^{\Pi} \left[ B^1(I) B^2(J) \right]^2, \end{split}$$

$$\lesssim r_N \tau_m + \sum_{(I,J) \in \Pi^1_N \times \Pi^2_N : \underline{I} \in I_m(k), \underline{J} \in \tilde{I}_m(k)} E^{\Pi} \left[ B^1(I) B^2(J) \right]^2$$

while Corollary 4.5.11 and Theorem 5.6.9 from [22] imply that

$$||C_{N}||_{\mathrm{sp}} \leq \frac{3}{2} ||\Sigma_{N}||_{\mathrm{sp}} \leq \frac{3}{2} \max \left\{ \max_{I \in \Pi_{N}^{1}: \underline{I} \in I_{m}(k)} E^{\Pi} \left[ B^{1}(I)^{2} \right], \max_{J \in \Pi_{N}^{2}: \underline{J} \in \tilde{I}_{m}(k)} E^{\Pi} \left[ B^{2}(J)^{2} \right] \right\}$$

$$+ \frac{3}{2} \max_{I \in \Pi_{N}^{1}: \underline{I} \in I_{m}(k)} \sum_{J \in \Pi_{N}^{2}: \underline{J} \in \tilde{I}_{m}(k)} |E^{\Pi} \left[ B^{1}(I)B^{2}(J) \right] |$$

$$\lesssim r_{N} + \max_{I \in \Pi_{N}^{1}: \underline{I} \in I_{m}(k)} \sum_{J \in \Pi_{N}^{2}: \underline{J} \in \tilde{I}_{m}(k)} |E^{\Pi} \left[ B^{1}(I)B^{2}(J) \right] |.$$

It holds that

$$\begin{split} E^{\Pi} \left[ B^{1}(I)B^{2}(J) \right] \\ &= \frac{1}{2\pi\tau_{N}} \int_{-\infty}^{\infty} (\mathcal{F}1_{I})(\tau_{N}^{-1}\lambda) \overline{(\mathcal{F}1_{J})(\tau_{N}^{-1}\lambda)} f_{N}(\tau_{N}^{-1}\lambda) d\lambda \\ &= \frac{1}{2\pi\tau_{N}} \sum_{i=1}^{N+1} R_{i} \int_{\Lambda_{-i}} (\mathcal{F}1_{I})(\tau_{N}^{-1}\lambda) \overline{(\mathcal{F}1_{J})(\tau_{N}^{-1}\lambda)} e^{-\sqrt{-1}\tau_{N}^{-1}\lambda\theta_{i}} d\lambda \\ &= \frac{1}{2\pi\tau_{N}} \sum_{i=1}^{N+1} R_{i} \int_{\Lambda_{-i}} \left( \int_{0}^{|I|} e^{-\sqrt{-1}\tau_{N}^{-1}\lambda s} ds \right) \left( \int_{0}^{|J|} e^{\sqrt{-1}\tau_{N}^{-1}\lambda s} ds \right) e^{\sqrt{-1}\tau_{N}^{-1}\lambda(\underline{J}-\underline{I}-\theta_{i})} d\lambda. \end{split}$$

Since we have

$$\begin{split} &\frac{d}{d\lambda}\left\{\left(\int_0^{|I|}e^{-\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\left(\int_0^{|J|}e^{\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\right\}\\ &=-\sqrt{-1}\tau_N^{-1}\left(\int_0^{|I|}se^{-\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\left(\int_0^{|J|}e^{\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\\ &+\sqrt{-1}\tau_N^{-1}\left(\int_0^{|I|}e^{-\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\left(\int_0^{|J|}se^{\sqrt{-1}\tau_N^{-1}\lambda s}ds\right), \end{split}$$

we obtain

$$\left| \frac{d}{d\lambda} \left\{ \left( \int_0^{|I|} e^{-\sqrt{-1}\tau_N^{-1} \lambda s} ds \right) \left( \int_0^{|J|} e^{\sqrt{-1}\tau_N^{-1} \lambda s} ds \right) \right\} \right| \leq \frac{\tau_N^{-1}}{2} (|I|^2 |J| + |I||J|^2).$$

Therefore, integration by parts implies that

$$\begin{split} &\left| \int_{\Lambda_{-i}} \left( \int_{0}^{|I|} e^{-\sqrt{-1}\tau_{N}^{-1}\lambda s} ds \right) \left( \int_{0}^{|J|} e^{\sqrt{-1}\tau_{N}^{-1}\lambda s} ds \right) e^{\sqrt{-1}\tau_{N}^{-1}\lambda (\underline{J} - \underline{I} - \theta_{i})} d\lambda \right| \\ & \leq \left\{ |I||J| + \frac{\tau_{N}^{-1}}{2} (|I|^{2}|J| + |I||J|^{2}) \right\} \frac{2^{-i+1}\pi}{\tau_{N}^{-1}|\underline{J} - \underline{I} - \theta_{i}|} \end{split}$$

as long as  $\underline{J} - \underline{I} \neq \theta_i$ . Hence we obtain

$$\begin{split} \sum_{J \in \Pi_N^2 : \underline{J} \in \tilde{I}_m(k)} \left| E^{\Pi} \left[ B^1(I) B^2(J) \right] \right| &\lesssim \tau_N^{-1} r_N^2 + \tau_N^{-2} r_N^3 \sum_{i=1}^{N+1} \sum_{\substack{J \in \Pi_N^2 : \underline{J} \in \tilde{I}_m(k) \\ \underline{J} - \underline{I} - \theta_i \neq 0}} \frac{1}{\tau_N^{-1} |\underline{J} - \underline{I} - \theta_i|} \\ &= O(\tau_N^{-2 + 3\xi} |\log \tau_N|^2) \end{split}$$

by Assumption 2(i) and (iii). Consequently, we obtain the first equation of (9). Moreover, it holds that

$$\begin{split} & \sum_{(I,J): \underline{I} \in I_m(k), \underline{J} \in \tilde{I}_m(k)} \left| E^{\Pi} \left[ B^1(I) B^2(J) \right] \right|^2 \\ & \leq \frac{N+1}{2\pi\tau_N} \sum_{i=1}^{N+1} \sum_{(I,J): \underline{I} \in I_m(k), \underline{J} \in \tilde{I}_m(k)} \left| \int_{\Lambda_{-i}} \left( \int_0^{|I|} e^{-\sqrt{-1}\tau_N^{-1}\lambda s} ds \right) \left( \int_0^{|J|} e^{\sqrt{-1}\tau_N^{-1}\lambda s} ds \right) e^{\sqrt{-1}\tau_N^{-1}\lambda (\underline{J} - \underline{I} - \theta_i)} d\lambda \right|^2 \\ & \lesssim N\tau_N^{-2} r_N^3 \tau_m + N\tau_N^{-2} \sum_{i=1}^{N+1} \sum_{\substack{(I,J): \underline{I} \in I_m(k), \underline{J} \in \tilde{I}_m(k) \\ \underline{J} - \underline{I} - \theta_i \neq 0}} \left\{ |I| |J| + \frac{\tau_N^{-1}}{2} (|I|^2 |J| + |I| |J|^2) \right\}^2 \frac{1}{(\tau_N^{-1} |\underline{J} - \underline{I} - \theta_i|)^2} \\ & \lesssim N\tau_N^{-2} r_N^3 \tau_m + N^2 \tau_N^{-4} r_N^5 \tau_m = O(\tau_N^{-4 + 5\xi + \beta} |\log \tau_N|^2), \end{split}$$

hence we obtain the second equation of (9).

Now, noting that  $-2 + 3\xi - (5\xi - 4 + \beta)/2 = (\xi - \beta)/2 > 2(1 - \beta)/5 > 0$ , from the discussion in Section 3.2.1 of [5], we have

$$\log E^{\Pi} \left[ \exp \left( \varsigma u_N \left\{ \widetilde{U}_k^N(\theta) - \overline{U}_k^N(\theta) \right\} \right) \right] = -\frac{1}{2} \log \det(E - 2\varsigma u_N C_N) - \operatorname{tr}[\varsigma u_N C_N]$$

for sufficiently large N due to the first equation of (9). Therefore, by Appendix II-(v) from [7] we obtain

$$\log E^{\Pi} \left[ \exp \left( \varsigma u_N \left\{ \widetilde{U}_k^N(\theta) - \overline{U}_k^N(\theta) \right\} \right) \right] \le \frac{u_N^2}{2} \|C_N\|_F^2 + \frac{|u_N|^3}{3} \frac{\|C_N\|_{\text{sp}} \|C_N\|_F^2}{(1 - \|C_N\|_{\text{sp}})^3}$$

for sufficiently large N due to the first equation of (9). Consequently, (9) yields the desires result.

**Lemma 5.** We have 
$$\left|\overline{U}_k^N(\theta)\right| \leq 6(\tau_m + r_N)$$
 for any  $N \in \mathbb{N}$ ,  $k \in \mathbb{Z}_+$  and  $\theta \in (-\delta, \delta)$ .

**Proof.** Similarly to the above proofs, we may assume  $\theta \ge 0$  and that (8) holds true without loss of generality. Then we have

$$\left| \overline{U}_{k}^{N}(\theta) \right| = \left| \sum_{I,J:\underline{I}\in I_{m}(k),\underline{J}\in \tilde{I}_{m}(k)} E^{\Pi} \left[ B^{1}(I)B^{2}(J) \right] K(I,J_{-\theta}) \right|$$

$$\leq \sum_{I,J:\underline{I}\in I_{m}(k),\underline{J}\in \tilde{I}_{m}(k)} \left\{ E^{\Pi} \left[ B^{1}(I)^{2} \right] + E \left[ B^{2}(J)^{2} \right] \right\} K(I,J_{-\theta})$$

$$\leq 3 \left\{ \sum_{I:\underline{I}\in I_{m}(k)} |I| + \sum_{I:\underline{J}\in \tilde{I}_{m}(k)} |J| \right\} \leq 3 \cdot 2(\tau_{m} + \bar{\tau}_{N}).$$

This completes the proof.

Lemma 6. We have

$$\max_{\theta \in \mathcal{G}^N} \left| \widehat{\rho}_{(j)}(\theta) - \tau_m \sum_{k=0}^{M_N - 1} c_k^N(\theta) \int_{-\pi}^{\pi} D(\lambda) H_{j,L}(\lambda) e^{\sqrt{-1}\lambda \theta \tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda \right| \to^p 0$$

as  $N \in \mathbb{N}$ .

**Proof**. We decompose the target quantity as

$$\begin{split} \widehat{\rho}_{(j)}(\theta) - \tau_m \sum_{k=0}^{M_N-1} c_k^N(\theta) \int_{-\pi}^{\pi} D(\lambda) H_{j,L}(\lambda) e^{\sqrt{-1}\lambda\theta\tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda \\ &= \left( \widehat{\rho}_{(j)}(\theta) - \sum_{l=-L_j-1}^{L_j-1} \Psi_j(l) \sum_{k=0}^{M_N-1} c_k^N(\theta - l\tau_N) \widetilde{U}_k^N(\theta - l\tau_N) \right) \\ &+ \sum_{l=-L_j+1}^{L_j-1} \Psi_j(l) \sum_{k=0}^{M_N-1} c_k^N(\theta - l\tau_N) \left\{ \widetilde{U}_k^N(\theta - l\tau_N) - \overline{U}_k^N(\theta - l\tau_N) \right\} \\ &+ \sum_{l=-L_j+1}^{L_j-1} \Psi_j(l) \sum_{k=0}^{M_N-1} \left\{ c_k^N(\theta - l\tau_N) - c_k^N(\theta) \right\} \overline{U}_k^N(\theta - l\tau_N) \\ &+ \sum_{l=-L_j+1}^{M_N-1} c_k^N(\theta) \left( \sum_{l=-L_j+1}^{L_j-1} \Psi_j(l) \overline{U}_k^N(\theta - l\tau_N) - \tau_m \int_{-\pi}^{\pi} D(\lambda) H_{j,L}(\lambda) e^{\sqrt{-1}\lambda\theta\tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda \right) \\ &=: \mathbf{I}_N(\theta) + \mathbf{II}_N(\theta) + \mathbf{III}_N(\theta) + \mathbf{IV}_N(\theta). \end{split}$$

First, since we have  $\sum_{p=0}^{L_j-1} h_{j,p}^2 = 1$ , it holds that  $|\Psi_j(l)| \leq 1$  for every l by the Schwarz inequality. Therefore, we have

$$|\mathbf{I}_N(\theta)| \le \sum_{l=-L_j-1}^{L_j-1} \sum_{k=0}^{M_N-1} \left| \widehat{U}_k^N(\theta - l\tau_N) - c_k^N(\theta - l\tau_N) \widetilde{U}_k^N(\theta - l\tau_N) \right|,$$

for any  $\varepsilon > 0$  and p > 1 we obtain

$$P\left(\max_{\theta \in \mathcal{G}^{N}} |\mathbf{I}_{N}(\theta)| > \varepsilon\right)$$

$$\leq \left(\frac{(2L_{j} - 1)M_{N}}{\varepsilon}\right)^{p} \sum_{\theta \in \mathcal{G}^{N}} \sum_{l=-L_{j}-1}^{L_{j}-1} \sum_{k=0}^{M_{N}-1} E\left[\left|\widehat{U}_{k}^{N}(\theta - l\tau_{N}) - c_{k}^{N}(\theta - l\tau_{N})\widetilde{U}_{k}^{N}(\theta - l\tau_{N})\right|^{p}\right]$$

$$\lesssim \tau_{N}^{-1} L M_{N} \cdot (L M_{N})^{p} \tau_{m}^{(\gamma+1)p} = O(\tau_{N}^{-1} L^{1+p} \tau_{m}^{p\gamma-1})$$

by the Markov inequality and Lemma 3. Since we can take the number p large enough such that  $\tau_N^{-1}L^{1+p}\tau_m^{p\gamma-1}\to 0$  as  $N\to\infty$  by assumption, we obtain  $\max_{\theta\in\mathcal{G}^N}|\mathbf{I}_N(\theta)|\to^p 0$ .

Next, for any  $\varepsilon > 0$  we have

$$P\left(\max_{\theta \in \mathcal{G}^N} |\mathbf{II}_N(\theta)| > \varepsilon\right)$$

$$\leq \sum_{\theta \in G^N} \sum_{l=-L_i+1}^{L_j-1} \sum_{k=0}^{M_N-1} P\left( \left| \widetilde{U}_k^N(\theta - l\tau_N) - \overline{U}_k^N(\theta - l\tau_N) \right| > \frac{\varepsilon}{KLM_N} \right)$$

with some constant K > 0. Therefore, the Markov inequality and Lemma 4 yield

$$P\left(\max_{\theta\in\mathcal{G}^N}|\mathbf{II}_N(\theta)|>\varepsilon\right)\lesssim \tau_N^{-1}LM_N\exp\left(-\frac{\varepsilon u_N}{KLM_N}\right).$$

Since  $\tau_N^c u_N/M_N \to \infty$  as  $N \to \infty$  for some c > 0, by assumption we conclude  $\max_{\theta \in \mathcal{G}^N} |\mathbf{II}_N(\theta)| \to^p 0$ . Now we prove  $\max_{\theta \in \mathcal{G}^N} |\mathbf{III}_N(\theta)| \to^p 0$ . Since  $\sum_{l=-L_j+1}^{L_j-1} |\Psi_j(l)| \le 2L_j - 1$ , we have

$$\max_{\theta \in \mathcal{G}^N} |\mathbf{III}_N(\theta)|$$

$$\leq \left( (2L_j - 1) \max_{\theta \in \mathcal{G}^N} \max_{l \in \mathbb{Z}: |l| < L_j} \max_{k = 0, 1, \dots, M_N - 1} \left| c_k^N(\theta - l\tau_N) - c_k^N(\theta) \right| \right) \sum_{k = 0}^{M_N - 1} \max_{\theta \in \mathcal{G}^N} \left| \overline{U}_k^N(\theta) \right|.$$

By the Hölder continuity of  $\sigma^1, \sigma^2$  and the assumption on L, we have

$$(2L_j - 1) \max_{\theta \in \mathcal{G}^N} \max_{l \in \mathbb{Z}: |l| < L_j} \max_{k=0,1,\dots,M_N - 1} \left| c_k^N(\theta - l\tau_N) - c_k^N(\theta) \right| \to^p 0$$

as  $N \to \infty$ , while Lemma 5 yields  $\sum_{k=0}^{M_N-1} \max_{\theta \in \mathcal{G}^N} \left| \overline{U}_k^N(\theta) \right| = O_p(1)$ . Hence we obtain the desired result.

Finally we prove  $\max_{\theta \in \mathcal{G}^N} |\mathbf{IV}_N(\theta)| \to^p 0$ . Noting that

$$\int_{-\pi}^{\pi} D(\lambda) H_{j,L}(\lambda) e^{\sqrt{-1}\lambda\theta\tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda = \sum_{l=-L_j-1}^{L_j-1} \Psi_j(l) \int_{-\pi}^{\pi} D(\lambda) e^{\sqrt{-1}\lambda(\theta-l\tau_N)\tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda,$$

for any  $\varepsilon > 0$  we have

$$P\left(\max_{\theta \in \mathcal{G}^{N}} |\mathbf{IV}_{N}(\theta)| > \varepsilon\right)$$

$$\leq \sum_{\theta \in \mathcal{G}^{N}} \sum_{l=-L_{j}+1}^{L_{j}-1} P\left(\tau_{m} \sum_{k=0}^{M_{N}-1} \left|\tau_{m}^{-1} \overline{U}_{k}^{N}(\theta) - \int_{-\pi}^{\pi} D(\lambda) e^{\sqrt{-1}\lambda(\theta - l\tau_{N})\tau_{N}^{-1}} f_{N}(\lambda/\tau_{N}) d\lambda\right| > \frac{\varepsilon}{KL}\right)$$

with some constant K > 0. Since we have

$$\overline{U}_k^N(\theta - l\tau_N) = \tau_m \int_{-\pi}^{\pi} D_k^N(\lambda, \theta - l\tau_N) e^{\sqrt{-1}\lambda(\theta - l\tau_N)\tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda,$$

it holds that

$$E\left[\left\{\tau_{m}\sum_{k=0}^{M_{N}-1}\left|\tau_{m}^{-1}\overline{U}_{k}^{N}(\theta)-\int_{-\pi}^{\pi}D(\lambda)e^{\sqrt{-1}\lambda(\theta-l\tau_{N})}f_{N}(\lambda/\tau_{N})d\lambda\right|\right\}^{Q}\right]$$

$$\leq (2\pi)^{Q-1}\tau_{m}\sum_{k=0}^{M_{N}-1}\int_{-\pi}^{\pi}E\left[\left|D_{k}^{N}(\lambda,\theta-l\tau_{N})-D(\lambda)\right|^{Q}\right]d\lambda$$

by the Jensen inequality. Therefore, by the Markov inequality we obtain

$$P\left(\max_{\theta \in \mathcal{G}^N} |\mathbf{IV}_N(\theta)| > \varepsilon\right) \lesssim \tau_N^{-1} L^{1+Q} \tau_m \max_{\theta \in \mathcal{G}^N} \sum_{k=0}^{M_N - 1} \int_{-\pi}^{\pi} E\left[\left|D_k^N(\lambda, \theta) - D(\lambda)\right|^Q\right] d\lambda.$$

Consequently, Assumption 2 and the assumption on L imply the desired result. This completes the proof.  $\Box$ 

**Proof of Theorem 1**. (a) From Lemma 6 it is enough to prove

$$\max_{\theta \in \mathcal{G}^N : |\theta - \theta_j| \ge v_N} \left| \tau_m \sum_{k=0}^{M_N - 1} c_k^N(\theta) \int_{-\pi}^{\pi} D(\lambda) H_{j,L}(\lambda) e^{\sqrt{-1}\lambda \theta \tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda \right| \to^p 0$$

as  $N \to \infty$ . The above equation follows once we show the following statements: If  $\vartheta_N \in \mathcal{G}^N$  ( $N = 1, 2, \ldots$ ) satisfy  $|\vartheta_N - \theta_j| \ge v_N$  for every N, then

$$\tau_m \sum_{k=0}^{M_N-1} c_k^N(\vartheta_N) \int_{-\pi}^{\pi} D(\lambda) H_{j,L}(\lambda) e^{\sqrt{-1}\lambda\vartheta_N \tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda \to^p 0$$

as  $N \to \infty$ . This can be shown in an analogous manner to the proof of Lemma 6 from [16].

(b) From Lemma 6 it suffices to prove

$$\tau_m \sum_{k=0}^{M_N-1} c_k^N(\theta) \int_{-\pi}^{\pi} D(\lambda) H_{j,L}(\lambda) e^{\sqrt{-1}\lambda\theta\tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda \to^p 2^j R_j \int_{\Lambda_{-j}} D(\lambda) \cos(b\lambda) d\lambda$$

as  $N \to \infty$ , which can be shown in an analogous manner to the proof of Lemma 7 from [16].

## 8.3 Proof of Theorem 2

Noting that  $\int_{\Lambda_{-j}} D(\lambda) \cos(b\lambda) d\lambda > 0$  for any  $b \in [-\frac{1}{2}, \frac{1}{2}]$  by assumption, the theorem can be shown in an analogous manner to the proof of Theorem 2 from [16] (using Theorem 1 instead of Lemmas 7–8 from [16]).

## Acknowledgments

Takaki Hayashi's research was supported by JSPS Grant-in-Aid for Scientific Research (C) Grant Number JP16K03601. Yuta Koike's research was supported by JST, CREST and JSPS Grant-in-Aid for Young Scientists (B) Grant Number JP16K17105.

#### References

- [1] Aït-Sahalia, Y. and Jacod, J. (2014). High-frequency financial econometrics. Princeton University Press.
- [2] Barunik, J. and Vacha, L. (2015). Realized wavelet-based estimation of integrated variance and jumps in the presence of noise. *Quant. Finance* **15**, 1347–1364.
- [3] Chan, K. (1992). A further analysis of the lead–lag relationship between the cash market and stock index futures market. *Review of Financial Studies* **5**, 123–152.
- [4] Christensen, K., Podolskij, M., Thamrongrat, N. and Veliyev, B. (2017). Inference from high-frequency data: A subsampling approach. *J. Econometrics* **197**, 245–272.

- [5] Dalalyan, A. and Yoshida, N. (2011). Second-order asymptotic expansion for a non-synchronous covariation estimator. *Ann. Inst. Henri Poincaré Probab. Stat.* **47**, 748–789.
- [6] Daubechies, I. (1992). Ten lectures on wavelets. SIAM.
- [7] Davies, R. B. (1973). Asymptotic inference in stationary Gaussian time-series. Adv. in Appl. Probab. 5, 469–497.
- [8] de Jong, F. and Nijman, T. (1997). High frequency analysis of lead-lag relationships between financial markets. *Journal of Empirical Finance* **4**, 259–277.
- [9] Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann.* **300**, 463–520.
- [10] Dobrev, D. and Schaumburg, E. (2016). High-frequency cross-market trading: Model free measurement and applications. Working paper.
- [11] Gençay, R., Gradojevic, N., Selçuk, F. and Whitcher, B. (2010). Asymmetry of information flow between volatilities across time scales. *Quant. Finance* **10**, 895–915.
- [12] Hafner, C. M. (2012). Cross-correlating wavelet coefficients with applications to high-frequency financial time series. *J. Appl. Stat.* **39**, 1363–1379.
- [13] Harris, F. H. d., McInish, T. H. and Wood, R. A. (2002). Security price adjustment across exchanges: an investigation of common factor components for Dow stocks. *Journal of Financial Markets* **5**, 277–308.
- [14] Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *Journal of Finance* **50**, 1175–1199.
- [15] Hasbrouck, J. (2015). High frequency quoting: Short-term volatility in bids and offers. Working paper, available at SSRN: http://ssrn.com/abstract=2237499.
- [16] Hayashi, T. and Koike, Y. (2016). Wavelet-based methods for high-frequency lead-lag analysis. Working paper. Available at arXiv: https://arxiv.org/abs/1612.01232.
- [17] Hayashi, T. and Koike, Y. (2017). No arbitrage and lead-lag relationships. Working paper. Available at arXiv: https://arxiv.org/abs/1712.09854.
- [18] Hayashi, T. and Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11, 359–379.
- [19] Hayashi, T. and Yoshida, N. (2008). Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *Ann. Inst. Statist. Math.* **60**, 357–396.
- [20] Hayashi, T. and Yoshida, N. (2011). Nonsynchronous covariation process and limit theorems. *Stochastic Process*. *Appl.* **121**, 2416–2454.
- [21] Hoffmann, M., Rosenbaum, M. and Yoshida, N. (2013). Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli* **19**, 426–461.
- [22] Horn, R. A. and Johnson, C. R. (2013). *Matrix analysis*. Cambridge University Press, 2nd edn.
- [23] Huth, N. and Abergel, F. (2014). High frequency lead/lag relationships empirical facts. *Journal of Empirical Finance* **26**, 41–58.
- [24] Lai, M.-J. (1995). On the digital filter associated with Daubechies' wavelets. *IEEE Trans. Signal Process.* **43**, 2203–2205.
- [25] Lo, A. W. and MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *J. Econometrics* **45**, 181–211.

- [26] Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V. and von Weizsäcker, J. E. (1997). Volatilities of different time resolutions analyzing the dynamics of market components. *Journal of Empirical Finance* **4**, 213–239.
- [27] Müller, U. A., Dacorogna, M. M., Davé, R. D., Pictet, O. V., Olsen, R. B. and Ward, J. R. (1993). Fractals and intrinsic time a challenge to econometricians. Tech. Rep. UAM.1993-08-16, Olsen and Associates.
- [28] Nason, G. P., von Sachs, R. and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62**, 271–292.
- [29] Ozturk, S. R., van der Wel, M. and van Dijk, D. (2017). Intraday price discovery in fragmented markets. *Journal of Financial Markets* **32**, 28–48.
- [30] Percival, D. B. and Walden, A. T. (2000). Wavelet methods for time series analysis. Cambridge University Press.
- [31] Renò, R. (2003). A closer look at the Epps effect. Int. J. Theor. Appl. Finance 6, 87–102.
- [32] Subbotin, A. (2008). A multi-horizon scale for volatility. Unpublished paper. Available at HAL: https://halshs.archives-ouvertes.fr/halshs-00261514.
- [33] Subrahmanyam, A. (1997). Multi-market trading and the informativeness of stock trades: An empirical intraday analysis. *Journal of Economics and Business* **49**, 515–531.
- [34] Vidakovic, B. (1999). Statistical modeling by wavelets. John Wiley & Sons, Inc.