Wavelet-based methods for high-frequency lead-lag analysis

Takaki Hayashi[*†‡]   Yuta Koike[†§‡]

February 15, 2017

Abstract

We propose a novel framework to investigate lead-lag relationships between two financial assets. Our framework bridges a gap between continuous-time modeling based on Brownian motion and the existing wavelet methods for lead-lag analysis based on discrete-time models and enables us to analyze the multi-scale structure of lead-lag effects. We also present a statistical methodology for the scale-by-scale analysis of lead-lag effects in the proposed framework and develop an asymptotic theory applicable to a situation including stochastic volatilities and irregular sampling. Finally, we report several numerical experiments to demonstrate how our framework works in practice.

Keywords: High-frequency data; Lead-lag effect; Multiscale modeling; Wavelet.

1 Introduction

A lead-lag effect is a phenomenon where one asset (called a "leader") is correlated with another asset (called a "lagger") at later times. Investigation of such a phenomenon has a long history in the economics literature; it is measured at various time scales mostly dailies or longer than daily, but relatively less for intra-daily data. It is not surprising that different lead-lag effects can be observed at different time scales in the financial markets, for there are a variety of participants in financial markets: They have different views for the economy and markets, different risk attitudes, with different sources of money and information, which makes them have different investment/trading horizons (cf. Müller *et al.* [34]).

The aim of this paper is to capture such multi-scale structures of lead-lag effects inherent in high-frequency financial data. The wavelet analysis provides a canonical framework to accomplish this purpose. The application of the wavelet analysis to finance has recently been expanded in various ways. We refer to the book Gençay *et al.* [21] for an introduction of such applications. The application of the wavelet analysis to investigating lead-lag relationships in the financial markets has also been investigated by many researchers *with middle or low-frequency data*; see [7, 11, 18, 19, 36] as well as Chapter 7 of [21] for example. On the other hand, there are few articles applying it to the investigation of lead-lag effects in the high-frequency domain. One exception is the paper by Hafner [24], which explores the lead-lag relation between the returns, durations and volumes of high-frequency transaction data of the IBM.

To our knowledge, all the existing studies on the wavelet analysis of lead-lag effects have their theoretical basis on discrete-time process modeling, which is mainly pioneered in Whitcher *et al.* [44, 45] and Serroukh and Walden [40, 41]. This is presumably because few results are available for statistical modeling of lead-lag effects in discretely observed continuous-time processes, even without taking their multi-scale structure into account. In the meantime, the

---
[*]Keio University, Graduate School of Business Administration, 4-1-1 Hiyoshi, Yokohama, Kanagawa 223-8526, Japan
[†]Tokyo Metropolitan University, Department of Business Administration, Graduate School of Social Sciences, Marunouchi Eiraku Bldg. 18F, 1-4-1 Marunouchi, Chiyoda-ku, Tokyo 100-0005 Japan
[‡]CREST, Japan Science and Technology Agency
[§]The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

modern, continuous-time finance theory is based on semimartingale processes, especially driven by Brownian motions (cf. Duffie [17]). Also, it has been getting recognized that modeling high-frequency financial data as discrete observations of continuous-time processes is quite effective for the statistical analysis (cf. Aït-Sahalia and Jacod [1]). From these perspectives it is desirable to develop a class of multivariate models to incorporate lead-lag relationships between coordinates for Brownian motion driven models, which is the primary goal of this study. We shall remark that some authors have recently developed statistical modeling of lead-lag effects in the continuous-time framework. Hoffmann *et al.* [25] have introduced a simple model to describe lead-lag effects between two continuous Itô semimartingales and developed a statistical estimation method for the lead-lag effects from (possibly asynchronous) discrete observation data. A similar model is studied in Robert and Rosenbaum [37] with different methodology. Empirical applications of Hoffmann *et al.* [25]'s methodology are found in [2, 6, 8, 27]. Apart from Brownian motion driven models, the Hawkes processes would be a credible candidate to describe lead-lag effects in the continuous-time framework; see Bacry *et al.* [3] and Da Fonseca and Zaatour [10] for example. However, none of them takes account of potential multi-scale structure of lead-lag effects. This paper attempts to fill in this gap.

Based on the proposed continuous-time model, we also aim at developing a statistical theory for the scale-by-scale analysis of lead-lag effects from discrete observation data. Our paper is virtually the first attempt to bridge the gap between the two distinct areas of research, namely wavelet analysis and continuous-time stochastic processes.

The paper is organized as follows. Section 2 presents two frameworks for scale-by-scale modeling of lead-lag effects between two Brownian motions. Section 3 develops an asymptotic theory for the framework presented in Section 2. We illustrate the numerical performance of the proposed approach on simulated data in Sections 4 and on real data in Section 5, respectively. Most of the proofs are given in Section 7.

Notation

For any function $f \in L^1(\mathbb{R})$, we denote by $\mathcal{F}f$ and $\mathcal{F}^{-1}f$ its Fourier transform and inverse Fourier transform, respectively. Specifically, they are defined by the following formulae:

$$(\mathcal{F}f)(\lambda) = \int_{-\infty}^{\infty} f(t)e^{-\sqrt{-1}\lambda t}dt, \qquad (\mathcal{F}^{-1}f)(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(\lambda)e^{\sqrt{-1}\lambda t}d\lambda.$$

The above definition can also be applied to a function $f \in L^2(\mathbb{R})$ by understanding that the convergences of the integrals take in $L^2(\mathbb{R})$. The Fourier inversion formula then reads $f = \mathcal{F}^{-1}(\mathcal{F}f)$ in $L^2(\mathbb{R})$ for all $f \in L^2(\mathbb{R})$. Also, the Parseval identity and the convolution theorem read

$$\int_{-\infty}^{\infty} (\mathcal{F}f)(\lambda)(\mathcal{F}g)(\lambda)d\lambda = 2\pi\int_{-\infty}^{\infty} f(t)g(t)dt$$

and

$$\mathcal{F}(f * g) = (\mathcal{F}f) \cdot (\mathcal{F}g) \qquad \text{in } L^2(\mathbb{R})$$

for all $f, g \in L^2(\mathbb{R})$.

2 Modeling scale-by-scale lead-lag effects in the continuous-time framework

In this section we propose two frameworks to introduce multiple lead-lag relationships between two Brownian motions B_t^1 and B_t^2 on a scale-by-scale basis. We also present sensible cross-covariance estimators constructed from discrete observations of the processes B_t^1 and B_t^2 on a fixed interval, which can be used for identifying lead-lag effects scale-by-scale. Specifically, we assume that $B_t = (B_t^1, B_t^2)$ is observed at the time points $i\tau_J$ for $i = 0, 1, \ldots, n$. Here, the unit time τ_J corresponds to the finest observable resolution, while n is the number of the unit times contained in the sampling

period. So, the interval $[0, n\tau_J]$ corresponds to the sampling period. We take $\tau_J = 2^{-J-1}$ to make the interpretation of wavelet analysis easier.

We denote by $(\Omega, \mathcal{F}, P)$ the probability space where B is defined.

2.1 Lévy-Ciesielski's construction revisited

We first revisit the classical Lévy-Ciesielski construction of Brownian motion (see [23] for a review on this topic). For each $\nu = 1, 2$, let ξ_0^ν be a standard normal variable. Also, let $(\xi_{jk}^\nu)_{j,k=0}^\infty$ be i.i.d. standard normal variables independent of ξ_0^ν. Then we define the process $B^\nu = (B_t^\nu)_{t \in [0,1]}$ by

$$B_t^\nu = \xi_0^\nu t + \sum_{j=0}^\infty \sum_{k=0}^{2^j - 1} \xi_{jk}^\nu \int_0^t \psi_{jk}(s) ds, \qquad \nu = 1, 2, \tag{1}$$

where ψ_{jk}'s are the Haar functions defined by $\psi_{jk}(s) = 2^{j/2} \psi(2^j s - k)$, where

$$\psi(s) = \begin{cases} 1 & 0 \le s < \frac{1}{2}, \\ -1 & \frac{1}{2} \le s < 1, \\ 0 & \text{otherwise}. \end{cases}$$

It is well-known that the infinite sum in the right side of (1) has the limit in $C[0,1]$ a.s. as the function of $t \in [0,1]$, and the limit process B^ν is a standard Brownian motion. The convergence is also valid in $L^2(P)$ for any $t \in [0,1]$.

Decomposition (1) naturally suggests that we could think that the process

$$B^\nu(j)_t = \sum_{k=0}^{2^j - 1} \xi_{jk}^\nu \int_0^t \psi_{jk}(s) ds, \qquad \nu = 1, 2; j = 0, 1, 2, \ldots,$$

as a "Brownian component at the level j". Here, the level j has the unit resolution $\tau_j = 2^{-j-1}$. This suggests that we could assess the lead-lag effect at the resolution τ_j by measuring that between $B^1(j)$ and $B^2(j)$ in some sense. A standard way to measure the lead-lag effect between two processes is assessing the cross-covariance function of their returns, provided that they are jointly stationary. However, this approach cannot be applied to the processes $B^1(j)$ and $B^2(j)$ directly because they are not of stationary increments. Instead, we propose assessing the cross-covariance function between $(\xi_{jk}^1)_{k=0}^\infty$ and $(\xi_{jk}^2)_{k=0}^\infty$, provided that their joint (weak) stationarity. Specifically, the objectives are given by

$$\rho_j^0(2 \cdot l\tau_j) = \text{Cov}\left(\xi_{jk}^1, \xi_{jk+l}^2\right), \qquad l = 0, \pm 1, \pm 2, \ldots. \tag{2}$$

Here, we write the cross-covariance between ξ_{jk}^1 and ξ_{jk+l}^2 with respect to the lag $2 \cdot l\tau_j$ instead of l alone to emphasize that their physical time difference is $2 \cdot l\tau_j$.

Since we can reproduce ξ_{jk}^ν's for $j \le J$ via the identity

$$\xi_{jk}^\nu = \int_0^1 \psi_{jk}(s) dB_s^\nu = 2^{\frac{j}{2}} \left(2B_{(2k+1)\tau_j}^\nu - B_{(2k+2)\tau_j}^\nu - B_{2k\tau_j}^\nu\right),$$

we can naturally use the following estimator for $\rho_j(2 \cdot l\tau_j)$:

$$\hat{\rho}_j^0(2 \cdot l\tau_j) = \frac{1}{M - l} \sum_{k=0}^{M-1-l} \left(\int_0^1 \psi_{jk}(s) dB_s^1\right) \left(\int_0^1 \psi_{jk+l}(s) dB_s^2\right),$$

where $M = \lfloor n/2^{J-j+1} \rfloor$. This class of estimators has an advantage in terms of computation, for the variables $\int_0^1 \psi_{jk}(s) dB_s^\nu$ are known as the *discrete wavelet transform* (DWT) of $(B_{(k+1)\tau_J}^\nu - B_{k\tau_J}^\nu)_{k=0}^{n-1}$ (ignoring the boundary variables) and fast

3

computation algorithms for them are known ("pyramid algorithm", cf. Section 7.3.1 of Mallat [32]). On the other hand, the main disadvantage of this approach is that we can only evaluate the cross-covariance function at a lag of the form $2 \cdot l\tau_j$ for some $l \in \mathbb{Z}$ and $j \leq J$, which causes undesirable restriction on lead-lag analyses. In the next subsection we propose another framework to deal with this issue.

2.2 Lévy-Ciesielski's construction based on dyadic wavelet transform

The major drawback of the first approach is that we cannot define the cross-covariance at all the discrete grid points generated for the finest observation resolution τ_J. To overcome this issue, we introduce an alternative decomposition to (1). For this purpose we reinterpret the Lévy-Ciesielski construction of Brownian motion as follows. Since $\phi := 1_{[0,1)}$ and ψ_{jk} ($j = 0, 1, \dots, k = 0, 1, \dots, 2^j - 1$) constitute an orthonormal basis of $L^2[0,1]$, we have the following expansion for any $f \in L^2[0,1]$:

$$f = \langle f, \phi \rangle \phi + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \langle f, \psi_{jk} \rangle \psi_{jk} \qquad \text{in } L^2[0,1], \tag{3}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of $L^2[0,1]$. This implies that

$$\int_0^1 f(s)dB_s^\nu = \langle f, \phi \rangle B_1^\nu + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \langle f, \psi_{jk} \rangle \int_0^1 \psi_{jk}(s)dB_s^\nu \tag{4}$$

for $\nu = 1, 2$. Substituting $f = 1_{(0,t]}$ in the above equation, we recover the Lévy-Ciesielski construction (1) with $\xi_0^\nu = B_1^\nu$ and $\xi_{jk}^\nu = \int_0^1 \psi_{jk}(s)dB_s^\nu$. This suggests that we could obtain a decomposition of Brownian motion analogous to the Lévy-Ciesielski construction once we have a "canonical" decomposition for any $f \in L^2[0,1]$ such as (3). Motivated by this idea, we consider alternative wavelet decomposition for functions which is suitable for our purpose.

We begin by recalling that decomposition (3) can be considered as a discretization of Calderón's reproducing identity for $f \in L^2(\mathbb{R})$ (cf. Sections 3.1–3.2 of [43]):

$$f(t) = \frac{1}{C_\psi} \int_0^\infty \left[\int_{-\infty}^\infty (W_a f)(b) \psi\left(\frac{t-b}{a}\right) db \right] \frac{1}{a^2} da \quad \text{in } L^2(\mathbb{R}), \tag{5}$$

where $*$ denotes convolution and we set

$$C_\psi = \int_0^\infty \frac{|(\mathcal{F}\psi)(\lambda)|^2}{\lambda} d\lambda$$

as well as we define the function $W_a f : \mathbb{R} \to \mathbb{R}$, which is called the *continuous wavelet transform* of f, by

$$W_a f(b) = a^{-\frac{1}{2}} \int_{-\infty}^\infty f(t) \psi\left(\frac{t-b}{a}\right) dt, \qquad b \in \mathbb{R}.$$

In fact, decomposition (3) is obtained by discretizing formula (5) in both the scale parameter a and the shift parameter b. Now, what is unsuitable for us in (4) is that we can only consider discretized time shifts of the forms $2 \cdot l\tau_j$ ($l \in \mathbb{Z}$). A natural solution of this issue is to only discretize the scale parameter a. This leads to the following expansion for $f \in L^2(\mathbb{R})$:

$$f = (f * \underline{\phi}) * \phi + \sum_{j=0}^{\infty} 2^j (f * \underline{\psi_j}) * \psi_j \qquad \text{in } L^2(\mathbb{R}). \tag{6}$$

Here, for any function g on $\mathbb{R}$ we define the functions $\underline{g}$ and g_j ($j \in \mathbb{Z}$) on $\mathbb{R}$ by setting $\underline{g}(t) = g(-t)$ and $g_j(t) = 2^{j/2} g(2^j t)$ for $t \in \mathbb{R}$. Note that $f * \underline{\psi_j} = W_{2^{-j}} f$. Decomposition (6) is indeed valid for any $f \in L^2(\mathbb{R})$ by Theorem 5.11 from [32], for we can deduce

$$|(\mathcal{F}\phi)(\lambda)|^2 + \sum_{j=-\infty}^{\infty} 2^j |(\mathcal{F}\psi_j)(\lambda)|^2 = 1$$

from the proof of Theorem 5.13 from [32]. The corresponding Lévy-Ciesielski type construction is given as follows:

4

Proposition 1. *Let $B = (B_t)_{t \in \mathbb{R}}$ be a two-sided Brownian motion. Suppose that real-valued functions $\tilde{\phi}, \tilde{\psi} \in L^2(\mathbb{R})$ satisfy*

$$|(\mathcal{F}\tilde{\phi})(\lambda)|^2 + \sum_{j=0}^{\infty} 2^j |(\mathcal{F}\tilde{\psi}_j)(\lambda)|^2 = 1 \tag{7}$$

for any $\lambda \in \mathbb{R}$. Then we have

$$B_t = \int_0^t \tilde{\xi} * \tilde{\phi}(s) ds + \sum_{j=0}^{\infty} 2^j \int_0^t \tilde{\xi}_j * \tilde{\psi}_j(s) ds \qquad in \ L^2(\mathbb{R}) \tag{8}$$

for any $t \in \mathbb{R}$, where

$$\tilde{\xi}(u) = \int_{-\infty}^{\infty} \tilde{\phi}(s-u) dB_s, \qquad \tilde{\xi}_j(u) = \int_{-\infty}^{\infty} \tilde{\psi}_j(s-u) dB_s.$$

A proof is given in Section 7.1. This decomposition suggests that we might think that the process $\tilde{\xi}_j(\cdot)$ as an alternative "Brownian component at the level j". However, unlike the original Lévy-Ciesielski construction, we do not generally have the independence of Brownian components across different levels, i.e. the processes $\tilde{\xi}_j(\cdot)$ and $\tilde{\xi}_{j'}(\cdot)$ are not independent for $j \neq j'$, especially when we adopt the Haar wavelets as $\tilde{\phi} = \phi$ and $\tilde{\psi} = \psi$. The lack of such independence makes it challenging to model/interpret lead-lag effects at different levels. We can avoid this issue by alternatively adopting band-limited wavelets (i.e. wavelets having compactly supported Fourier transforms). Specifically, we take the Littlewood-Paley wavelets as follows:[1]

$$\tilde{\phi}(s) = \phi^{LP}(s) := \frac{\sin(\pi s)}{\pi s}, \qquad \tilde{\psi}(s) = \psi^{LP}(s) := \phi^{LP}(2s) - \phi^{LP}(s).$$

We may regard the Littlewood-Paley wavelets as the "representative" band-limited wavelets because any band-limited function can be recovered from its discrete samples with interpolation based on the Littlewood-Paley scaling function ϕ^{LP}, according to the Shannon-Whittaker sampling theorem (cf. Theorem 3.2 of [32]). Now, since we have

$$(\mathcal{F}\phi^{LP})(\lambda) = 1_{[-\pi,\pi]}(\lambda), \qquad (\mathcal{F}\psi^{LP})(\lambda) = 1_{[-2\pi,-\pi)\cup(\pi,2\pi]}(\lambda),$$

condition (7) is satisfied. Moreover, for any $j, j' \geq 0$ and any $u, v \in \mathbb{R}$, the Parseval identity yields

$$E\left[\tilde{\xi}_j(u)\tilde{\xi}_{j'}(v)\right] = \int_{-\infty}^{\infty} \psi_j^{LP}(s-u)\psi_{j'}^{LP}(s-v) ds = \frac{1}{2^{\frac{j+j'}{2}+1}\pi} \int_{\Lambda_j \cap \Lambda_{j'}} e^{\sqrt{-1}(u-v)\lambda} d\lambda, \tag{9}$$

where $\Lambda_j = [-2^{j+1}\pi, -2^j\pi) \cup (2^j\pi, 2^{j+1}\pi]$ for $j \in \mathbb{Z}$. Since $\Lambda_j \cap \Lambda_{j'} = \emptyset$ if $j \neq j'$, we have the independence of Brownian components across different levels because they are Gaussian.

Now, analogously to the first approach, we measure the lead-lag effect at the level j by assessing the cross-covariance function between the processes $\tilde{\xi}_j^1(\cdot)$ and $\tilde{\xi}_j^2(\cdot)$. Here, we note that $\tilde{\xi}_j^1$ and $\tilde{\xi}_j^2$ are jointly stationary if the process $B_t = (B_t^1, B_t^2)$ is of stationary increments, i.e. $E\left[\left(B_{t_1+h}^1 - B_{s_1+h}^1\right)\left(B_{t_2+h}^2 - B_{s_2+h}^2\right)\right] = E\left[\left(B_{t_1}^1 - B_{s_1}^1\right)\left(B_{t_2}^2 - B_{s_2}^2\right)\right]$ for any $t_1, t_2, s_1, s_2, h \in \mathbb{R}$. As shown in the next subsection, the latter assumption is not restrictive. Consequently, we propose assessing the cross-covariance function:

$$\rho_j(\theta) = \text{Cov}\left[\tilde{\xi}_j^1(u), \tilde{\xi}_j^2(u+\theta)\right], \qquad \theta \in \mathbb{R},$$

provided their joint (weak) stationarity.

---

[1]See page 115 of Daubechies [13].

Remark 1. The "coefficients" $(f * \underline{\psi_j})_{j=0}^{\infty}$ in decomposition (6) are called the *(translation-invariant) dyadic wavelet transform* of f in the wavelet literature (cf. Section 5.2 of [32]). In conjunction with this terminology, the processes in (7) might be called the dyadic wavelet transform of dB^{ν}. Applications of dyadic wavelet transform based decomposition (6) are found in pattern recognition and denoising with translation-invariant thresholding estimators; see Chapter 6 and Section 11.3.1 of [32].

2.3 Model specification via cross-spectrum

For ease of interpretation, it is desirable that $|\rho_j(\theta)|$ has the unique maximum at some $\theta_j \in \mathbb{R}$ for each j. So we presuppose such a situation and introduce the following specification into our model. We first note that (9) implies that $\tilde{\xi}_j^{\nu}$ has the spectral density $2^{-j}1_{\Lambda_j}$, hence its spectrum is concentrated on Λ_j. We also remark that if $W_t = (W_t^1, W_t^2)$ ($t \in \mathbb{R}$) is a bivariate two-sided Brownian motion with correlation parameter R, then for any $\theta \in \mathbb{R}$ the process $(W_t^1, W_{t-\theta}^2)$ ($t \in \mathbb{R}$) is of stationary increments and has the cross-spectral density $Re^{-\sqrt{-1}\theta\lambda}$, $\lambda \in \mathbb{R}$. Motivated by these facts, let us suppose that the bivariate process $B_t = (B_t^1, B_t^2)$ is of stationary increments and its cross-spectral density is of the form $R_j e^{-\sqrt{-1}\theta_j \lambda}$ with $R_j \in [-1, 1]$ and $\theta_j \in \mathbb{R}$ for $\lambda \in \Lambda_j$, $j = 0, 1, \ldots$. Namely, the function $f : \mathbb{R} \to \mathbb{C}$, which is defined by

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

satisfies

$$E\left[\left(\int_{-\infty}^{\infty} u(s)dB_s^1\right)\overline{\left(\int_{-\infty}^{\infty} v(s)dB_s^2\right)}\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty}(\mathcal{F}u)(\lambda)\overline{(\mathcal{F}v)(\lambda)}f(\lambda)d\lambda \tag{10}$$

for any $u, v \in L^2(\mathbb{R})$. Therefore, noting that

$$(\mathcal{F}^{-1}1_{\Lambda_0})(s) = \psi^{LP}(s) \tag{11}$$

for $s \in \mathbb{R}$, in this situation we have for each j

$$\rho_j(\theta) = \frac{R_j}{2^{j+1}\pi}\int_{\Lambda_j} e^{\sqrt{-1}(\theta-\theta_j)\lambda}d\lambda = R_j\psi^{LP}(2^j(\theta-\theta_j)) = R_j\frac{\sin[2^j\pi(\theta-\theta_j)]}{2^j\pi(\theta-\theta_j)}(2\cos[2^j\pi(\theta-\theta_j)] - 1)$$

by the Fourier inversion formula, hence $|\rho_j(\theta)|$ has the unique maximum $|R_j|$ at $\theta = \theta_j$ as long as $R_j \neq 0$.

Now the question of matter is whether we can construct a bivariate process $B_t = (B_t^1, B_t^2)$ having the pre-described properties such that both B^1 and B^2 are respectively one-dimensional standard Brownian motions. The following proposition gives an affirmative answer:

Proposition 2. *Suppose that a measurable function $f : \mathbb{R} \to \mathbb{C}$ satisfies*

$$\|f\|_{\infty} \leq 1 \tag{12}$$

and

$$\overline{f(\lambda)} = f(-\lambda) \qquad \text{for almost all } \lambda \in \mathbb{R}. \tag{13}$$

Then there is 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) *f is the cross-spectral density of B, i.e. f satisfies* (10) *for any $t, s \in \mathbb{R}$.*

Conversely, if a bivariate process $B_t = (B_t^1, B_t^2)$ ($t \in \mathbb{R}$) with stationary increments satisfies condition (i), there is a measurable function $f : \mathbb{R} \to \mathbb{C}$ satisfying (12)–(13) *and condition (ii).*

We prove this result in Section 7.2.

In the next section we will consider an asymptotic theory when the unit length τ_J shrinks to zero, or equivalently, when J tends to infinity, which is a standard approach for theoretical analyses of statistics for high-frequency data (cf. Aït-Sahalia and Jacod [1]). In such a situation it is convenient to relabel indices of the parameters R_j and θ_j so that the finest resolution τ_J corresponds to the level $j = 1$. This suggests that we should model the cross-spectral density of $B_t = (B_t^1, B_t^2)$ as

$$f_J(\lambda) = \sum_{j=0}^{J} R_{(j)} e^{-\sqrt{-1}\theta_{(j)}\lambda} 1_{\Lambda_j}(\lambda) = \sum_{j=1}^{J+1} R_j e^{-\sqrt{-1}\theta_j\lambda} 1_{\Lambda_{(j)}}(\lambda), \tag{14}$$

where we set $(j) = J - j + 1$. Here, we omit all the components finer than τ_J from the model because they are not identifiable.

2.4 Construction of estimators

If continuous-time observation data of the process B_t on the whole real line were available, we could reproduce $\tilde{\xi}^\nu_{(j)}(s)$'s via

$$\tilde{\xi}^\nu_{(j)}(s) = \int_{-\infty}^{\infty} \psi^{LP}_{(j)}(s - u) dB_u^\nu, \qquad j = 1, \ldots, J + 1$$

for $\nu = 1, 2$ and construct estimators for $\rho_{(j)}(\theta)$ as in Section 2.1. Since we only have discrete observation data of B_t on $[0, n\tau_J]$, one natural way is to approximate the above integral by discretization. This is however problematic because discretization of ψ^{LP} is unstable due to oscillation and ψ^{LP} is not compactly supported. For these reasons we adopt another approach that using Daubechies' compactly supported wavelets. Daubechies' wavelets generate finite-length filters whose gain functions well approximate those of the Littlewood-Paley wavelets (cf. Lai [30]).

Specifically, we denote by $(h_p)_{p=0}^{L-1}$ Daubechies' wavelet filter of length L. Its squared gain function $H_L(\lambda) = |\sum_{p=0}^{L-1} h_p e^{-\sqrt{-1}\lambda p}|^2$ is given by

$$H_L(\lambda) = 2\sin^L(\lambda/2) \sum_{p=0}^{L/2-1} \binom{L/2 - 1 + p}{p} \cos^{2p}(\lambda/2)$$

(note that L is always an even integer). The associated scaling filter $(g_p)_{p=0}^{L-1}$ is determined by the quadrature mirror relationship[2]:

$$g_p = (-1)^{p+1} h_{L-p-1}, \qquad p = 0, 1, \ldots, L - 1.$$

Hence its squared gain function $G_L(\lambda) = |\sum_{p=0}^{L-1} g_p e^{-\sqrt{-1}\lambda p}|^2$ satisfies $G_L(\lambda) = H_L(\lambda - \pi)$. See Chapters 6–8 of [13] and Section 3.4.5 of [43] for more details.

Using the filters $(g_p)_{p=0}^{L-1}$ and $(h_p)_{p=0}^{L-1}$ as low-pass and high-pass filters, we can construct scale-by-scale band-pass filters. We denote by $(h_{j,p})_{p=0}^{L_j-1}$ the level j wavelet filters associated with the filters $(g_p)_{p=0}^{L-1}$ and $(h_p)_{p=0}^{L-1}$ for every j, where $L_j = (2^j - 1)(L - 1) + 1$. See Section 2 of Percival and Mofjeld [35] for the precise definition. For our analysis the form of its squared gain function $H_{j,L}(\lambda) = |\sum_{p=0}^{L-1} h_{j,p} e^{-\sqrt{-1}\lambda p}|^2$ is important, which is given by

$$H_{j,L}(\lambda) = H_L(2^{j-1}\lambda) \prod_{i=0}^{j-2} G_L(2^i\lambda), \qquad \lambda \in \mathbb{R}. \tag{15}$$

---

[2]We follow Gençay *et al.* [21] in using the notation that (h_p) denotes the wavelet filter and (g_p) denotes the scaling filter. Note that many authors adopt the reverse usage: (h_p) denotes the scaling filter and (g_p) denotes the wavelet filter. See [13, 32, 43] for instance.

We also remark that the filter $(h_{j,p})_{p=0}^{L_j-1}$ has unit energy

$$\sum_{p=0}^{L_j-1} h_{j,p}^2 = 1. \tag{16}$$

Now we approximates $\tilde{\xi}_{(j)}^{\nu}(\cdot)$ by

$$\mathcal{W}_{jk}^{\nu} = \sum_{p=0}^{L_j-1} h_{j,p} \left(B_{(k+1-p)\tau_J}^{\nu} - B_{(k-p)\tau_J}^{\nu} \right), \qquad k = L_j - 1, \dots, n-1 \tag{17}$$

for each $\nu = 1, 2$. We remark that the transformation of $(B_{(k+1)\tau_J}^{\nu} - B_{k\tau_J}^{\nu})_{k=0}^{n-1}$ in (17) is the so-called *maximal overlap discrete wavelet transform* (MODWT) up to multiplication of constants, and that the resulting $\mathcal{W}_{jk}^{\nu}$'s are the corresponding wavelet coefficients. See Section 3 of [35] and Section 4.5 of [21] for details. Then we define the cross-covariance estimators at the level j by

$$\widehat{\rho}_{(j)}(l\tau_J) = \begin{cases} \frac{\tau_J^{-1}}{n-l-L_j+1} \sum_{k=L_j-1}^{n-l-1} \mathcal{W}_{jk}^1 \mathcal{W}_{jk+l}^2 & \text{if } l \geq 0 \\ \frac{\tau_J^{-1}}{n+l-L_j+1} \sum_{k=L_j-1}^{n+l-1} \mathcal{W}_{jk-l}^1 \mathcal{W}_{jk}^2 & \text{otherwise.} \end{cases} \tag{18}$$

In the next section we will show that these are asymptotically sensible cross-covariance estimators while both J and L tend to infinity with appropriate rates.

Remark 2. The above estimators are formally the same as the *wavelet cross-covariance estimators* used in discrete-time modeling framework up to multiplication of constants (cf. Section 7.4 of [21]). Therefore, the results presented in the next section ensure the validity of using such estimators in the continuous-time modeling framework proposed in this paper.

3 Asymptotic theory

This section presents an asymptotic theory for the estimators constructed in the previous section. We also generalize the setting therein. First we assume that $n\tau_J \to T$ as $J \to \infty$ for some $T \in (0, \infty)$. This means that we observe data on a *fixed* interval (e.g. one day). Next, $B_t = (B_t^1, B_t^2)$ ($t \in \mathbb{R}$) denotes a bivariate Gaussian process with stationary increments such that both B^1 and B^2 are respectively one-dimensional two-sided standard Brownian motions and its cross-spectral density is given by (14). We denote by $(\Omega, \mathcal{F}, P)$ the probability space where the process B is defined. Now, we assume that for each $\nu = 1, 2$ we have a filtration $(\mathcal{F}_t^{\nu})_{t\geq 0}$ such that the process $(B_t^{\nu})_{t\geq 0}$ is an $(\mathcal{F}_t^{\nu})$-Brownian motion. Then, the ν-th log-price process $X^{\nu} = (X_t^{\nu})_{t\geq 0}$ is given by

$$X_t^{\nu} = X_0^{\nu} + \int_0^t \sigma_t^{\nu} dB_t^{\nu}, \qquad t \geq 0, \tag{19}$$

where σ_t^{ν} is an $(\mathcal{F}_t^{\nu})$-adapted càdlàg process (hence the above stochastic integral is well-defined in the usual sense).

We assume that the processes X^1 and X^2 are observed at time stamps of the form $k\tau_J$ for some $k = 0, 1, \dots, n$. Unlike the previous section, we allow that observation data at some time points are missing. For each $\nu = 1, 2$ we denote by $\delta_{\nu,k}$ the indicator variable which is unity when X^{ν} is not observed at the time $k\tau_J$ and zero otherwise. So the observation data for X^{ν} are given by $\{X_{k\tau_J}^{\nu} : \delta_{\nu,k} = 0, k = 0, 1, \dots, n\}$. For simplicity we assume that the initial value X_0^{ν} is observed, i.e. $\delta_{\nu,0} \equiv 0$. To construct our estimators we create (pseudo) complete observation data at all the points $k\tau_J$'s by use of the previous-tick interpolation scheme. Then we construct our estimators based on this observation data as in the previous section. To derive the mathematical expression of the estimators constructed so, it is convenient to introduce Lo and MacKinlay [31]'s notation. For each $\nu = 1, 2$, setting

$$\chi_{\nu,k}(0) = 1 - \delta_{\nu,k+1}, \qquad \chi_{\nu,k}(\alpha) = (1 - \delta_{\nu,k+1}) \prod_{l=1}^{\alpha} \delta_{\nu,k+1-l}, \qquad \alpha = 1, \dots, k$$

for $k = 0, 1, \ldots, n-1$, we can write the pseudo observed returns for X^ν based on the interpolated data as

$$\Delta_k^o X^\nu = \sum_{\alpha=0}^{k} \chi_{\nu,k}(\alpha) \Delta_{k-\alpha} X^\nu, \qquad k = 0, 1, \ldots, n-1,$$

where $\Delta_k X^\nu = X^\nu_{(k+1)\tau_J} - X^\nu_{k\tau_J}$. Therefore, we have $\mathcal{W}^\nu_{jk} = \sum_{p=0}^{L_j-1} h_{j,p} \Delta^o_{k-p} X^\nu$ for $j = 0, 1, \ldots$ and $k = L_j - 1, \ldots, n-1$. The estimator $\widehat{\rho}_{(j)}(l\tau_J)$ is then constructed by (18). Here, for the construction of asymptotic results in a general form we additionally consider a functional version of $\widehat{\rho}_{(j)}(l\tau_J)$. We define the process $(\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t)_{t\geq 0}$ by

$$\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t = \begin{cases} \frac{\tau_J^{-1}}{n-l-L_j+1} \sum_{k=L_j-1}^{\lfloor \tau_J^{-1} t \rfloor - l} \mathcal{W}^1_{jk} \mathcal{W}^2_{jk+l} & \text{if } l \geq 0, \\ \frac{\tau_J^{-1}}{n+l-L_j+1} \sum_{k=L_j-1}^{\lfloor \tau_J^{-1} t \rfloor + l} \mathcal{W}^1_{jk-l} \mathcal{W}^2_{jk} & \text{otherwise} \end{cases}$$

for $t \geq 0$.[3] Since we have $\widehat{\rho}_{(j)}(l\tau_J) = \widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_{(n-1)\tau_J}$, it is enough to investigate the asymptotic properties of $(\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_j)_t)_{t\geq 0}$.

Regarding the mechanism of missing observations, we focus on the following simple situation as in [31] (known as *missing completely at random*):

(i) The observation for X^ν can be missing at each $i\tau_J$ with probability π_ν for $\nu = 1, 2$,

(ii) missing observations occur independently.

Under this situation, for each $\nu = 1, 2$ $(\delta_{\nu,k})_{k=1}^\infty$ is a sequence of i.i.d. Bernoulli variables with provabilities π_ν and $1 - \pi_\nu$ of taking variables 1 and 0. Also, $(\delta_{1,k})_{k=1}^\infty$ and $(\delta_{2,k})_{k=1}^\infty$ are mutually independent.

To avoid boundary issues, we assume that the true lag parameters $\theta_j \in (-\delta, \delta)$ for all j with some constant $\delta \in (0, T)$, and evaluate the cross-covariance function on the finite grid set $\mathcal{G}_J = \{l\tau_J : l \in \mathbb{Z}, |l\tau_J| < \delta\}$.

For processes $(Y_t^J)_{t\geq 0}$ $(J = 1, 2, \ldots)$ and a process $(Y_t)_{t\geq 0}$, we write $Y_t^J \xrightarrow{ucp} Y_t$ to express $\sup_{0\leq s\leq t} |Y_s^J - Y_s| \to^p 0$ as $J \to \infty$ for any $t > 0$.

Theorem 1. *Let $j \in \mathbb{N}$ be fixed. Suppose that $L \to \infty$ and $L^2 \tau_J \to 0$ as $J \to \infty$. Let (ϑ_J) be a sequence of real numbers such that $\vartheta_J \in \mathcal{G}_J$ for every J.*

(a) *If $L^{-\frac{1}{2}} \tau_J^{-1}(\vartheta_J - \theta_j) \to \infty$ as $J \to \infty$, then $\widehat{\boldsymbol{\rho}}_{(j)}(\vartheta_J)_t \xrightarrow{ucp} 0$ as $J \to \infty$.*

(b) *If $\tau_J^{-1}(\vartheta_J - \theta_j) \to b$ as $J \to \infty$ for some $b \in \mathbb{R}$, then $\widehat{\boldsymbol{\rho}}_{(j)}(\vartheta_J)_t \xrightarrow{ucp} \Sigma_t(\theta_j) R_j \int_{\Lambda_{-j}} D(\lambda) \Pi(\lambda) e^{\sqrt{-1}b\lambda} d\lambda$ as $J \to \infty$, where*

$$D(\lambda) = \frac{1}{2\pi} \left| \frac{e^{-\sqrt{-1}\lambda} - 1}{\lambda} \right|^2, \qquad \Pi(\lambda) = \frac{(1 - \pi_1)(1 - \pi_2)}{(1 - \pi_1 e^{\sqrt{-1}\lambda})(1 - \pi_2 e^{-\sqrt{-1}\lambda})}$$

and

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

The proof of Theorem 1 is given in Section 7.3.

Remark 3. The first part of Theorem 1 claims that our cross-covariance estimator $\widehat{\rho}_{(j)}(\theta)$ is close to zero if θ is sufficiently far from the true lag parameter θ_j. The second part of the theorem claims that our cross-covariance estimator $\widehat{\rho}_{(j)}(\theta)$ tends to the quantity R_j multiplied by some (random) constant. This constant consists of four sources: $\Sigma_t(\theta_j)$ comes from the presence of volatility. $D(\lambda)$ represents a discretization error caused by dX^ν being replaced by $\Delta_k X^\nu$'s. $\Pi(\lambda)$ is caused by previous-tick interpolation. $e^{\sqrt{-1}b\lambda}$ comes from the bias due to the discrepancy between θ and θ_j.

---

[3] We set $\sum_{k=p}^q \equiv 0$ if $p > q$ by convention.

Theorem 1 suggests that we could estimate the lag parameter θ_j by maximizing $|\widehat{\rho}_{(j)}(\theta)|$ over the finite grid $\mathcal{G}_J$ as in Hoffmann *et al.* [25]. Consequently, we choose the random variable $\widehat{\theta}_j \in \mathcal{G}_J$ such that

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

as an estimator for θ_j.

Theorem 2. *Let $j \in \mathbb{N}$ be fixed. Suppose that $L \to \infty$ and $L^2 \tau_J^\kappa \to 0$ as $J \to \infty$ for some $\kappa \in (0, 1)$. Suppose also that both σ^1 and σ^2 almost surely have γ-Hölder continuous paths for some $\gamma > 0$. Then, if a sequence v_J of positive numbers satisfies $L^{-\frac{1}{2}} \tau_J^{-1} v_J \to \infty$ as $J \to \infty$, then $v_J^{-1}(\widehat{\theta}_j - \theta_j) \to^p 0$ as $J \to \infty$, 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 $J \to \infty$.*

We prove Theorem 2 in Section 7.4.

Remark 4. The Hölder continuity assumption on the paths of the volatilities σ^1 and σ^2 is satisfied by a number of stochastic volatility models. In fact, it is satisfied if σ^ν is a continuous Itô semimartingale (with respect to the filtration $(\mathcal{F}_t^\nu)$) for every $\nu = 1, 2$. It is worth mentioning that the assumption is also satisfied by the so-called rough fractional stochastic volatility models introduced in Gatheral *et al.* [20] which have pointed out the practicality of such models.

Remark 5. Theorem 2 is a counterpart of Theorem 1 from Hoffmann *et al.* [25] in our framework. Our results are applicable for separating multiple lead-lag effects on a scale-by-scale basis. On the other hand, the convergence rate is slower than the estimator of Hoffmann *et al.* [25] by the factor $\sqrt{L}$, which might be regarded as a cost to separate multiple lead-lag effects.

4 Simulation study

In this section we implement a Monte Carlo experiment to evaluate finite sample performance of our scale-by-scale lead-lag parameter estimators $\widehat{\theta}_j$'s defined in the previous section. We set $J = 13$ and $n = 15,000$. As the search grid $\mathcal{G}_J$, we use $\mathcal{G}_J = \{l\tau_J : l \in \mathbb{Z}, |l| \leq 60\}$.

We simulate model (19) with constant volatilities $\sigma^\nu \equiv 1$ for $\nu = 1, 2$. The parameters for the spectral density (14) are chosen as in Table 1. To simulate the process B at the time points $k\tau_J$, $k = 0, 1, \ldots, n$, it is enough to generate bivariate variables $\Delta_k B$, $k = 0, 1, \ldots, n$. Since they are stationary and Gaussian, we can use the multivariate version of the circulant embedding method from [9] once we compute the cross-covariance function $E[\Delta_k B^1 \Delta_{k+l} B^2]$, $l \in \mathbb{Z}$. It can be computed by using (11) and the Fourier inversion formula as

$$E\left[\Delta_k B^1 \Delta_{k+l} B^2\right] = \sum_{j=0}^{J} R_{(j)} \int_0^{\tau_J} \int_0^{\tau_J} \psi^{LP}(2^j(u - v + l\tau_J - \theta_{(j)})) du dv, \quad l \in \mathbb{R},$$

hence we approximate it by

$$\tau_J^2 \sum_{j=0}^{J} R_{(j)} \psi^{LP}(2^j(l\tau_J - \theta_{(j)})).$$

Regarding the parameters π^1 and π^2 which controls the probabilities of missing observations, we consider two situations where $\pi_1 = \pi_2 = 0$ and $\pi_1 = \pi_2 = 0.5$. In the first situation no missing observation occurs, so the observation times are equidistant and synchronous. In the meantime, in the second situation the observation times are non-equidistant and asynchronous.

As the wavelet filter $(h_p)_{p=0}^{L-1}$ to construct our estimator, we examine the following three choices of Daubechies' wavelets:

Table 1: Parameters for the spectral density (14)

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
θ_j/τ_J	-1	-1	-2	-2	-3	-5	-7	-10	0

Haar Haar wavelets ($L = 2$),

LA(8) Least asymmetric wavelet with length $L = 8$,

LA(20) Least asymmetric wavelet with length $L = 20$.

See Chapter 8 of [13] for details on the least asymmetric wavelets.

We run 1,000 Monte Carlo iterations for each experiment. We report the sample median and median absolute deviation (MAD) of the estimates for each experiment in Tables 2. As the table reveals, at the finest resolution levels $j = 1, 2$ all the estimates are very precise, while at coarser resolution levels $j \geq 3$ the LA(8) and LA(20) based estimators tend to perform better than the Haar based estimators. This is not surprising because the consistency of our estimator is ensured in the asymptotics as L tends to infinity. The LA(20) based estimator shows an excellent performance at moderate resolution levels $j \leq 6$ even in the presence of missing observations. At the coarsest resolution levels $j = 7, 8$, the precisions of all the estimators fall. This would be due to the following reason: According to the proofs of Theorems 1–2, the convergence rate of our estimator $\widehat{\theta}_j$ is proportional to the square root of $L_j = (2^j - 1)(L - 1) + 1$, hence it is declined as j increases.

Table 2: Simulation results

j	1	2	3	4	5	6	7	8
True	-1	-1	-2	-2	-3	-5	-7	-10
				$\pi_1 = \pi_2 = 1$				
Haar	-1	-1	-1	-1	-1	-1	-2	-2
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(1)
LA(8)	-1	-1	-2	-2	-3	-4	-6	-8
	(0)	(0)	(0)	(0)	(0)	(1)	(3)	(10)
LA(20)	-1	-1	-2	-2	-3	-5	-6	-9
	(0)	(0)	(0)	(0)	(0)	(1)	(4)	(15)
				$\pi_1 = \pi_2 = 0.5$				
Haar	-1	-1	-2	-2	-2	-2	-2	-2
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
LA(8)	-1	-1	-2	-2	-3	-4	-6	-8
	(0)	(0)	(0)	(0)	(0)	(1)	(3)	(10)
LA(20)	-1	-1	-2	-2	-3	-5	-6	-9
	(0)	(0)	(0)	(0)	(0)	(1)	(4)	(15)

This table reports the median and the median absolute deviation (in parentheses) of the estimates.

5 Empirical illustration

In this section we apply our new method to evaluate lead-lag effects on a scale-by-scale basis in real financial data. Specifically, we analyze the lead-lag relationships between the log-prices of the S&P 500 index and the E-mini S&P500 futures in April 2016, containing 21 trading days. We have obtained our data set from the Bloomberg. The price data of the S&P 500 index are recorded with the second-by-second basis from 9:30 am EDT to 16:00 EDT each trading day. We use the transaction data of the E-mini S&P500 futures recorded between 9:30 am EDT and 16:00 EDT each trading day with the accuracy in the timestamp values being one second.

Before presenting the result, we remark that there are many researches examining the lead-lag effect between the S&P 500 index and index futures: See [15, 29, 42] for example. These studies have reported that the futures lead the index.

We regard the log-price process of the S&P 500 index as X^1 and the log-price process of the E-mini S&P 500 futures as X^2. We set

$$\mathcal{G}_J = \{-300s, -299s, \ldots, -1s, 0s, 1s, \ldots, 299s, 300s\}$$

as the search grid. We use LA(20) (Daubechies' least asymmetric wavelet filter with length 20) as the wavelet filter $(h_p)_{p=0}^{L-1}$.

Table 3 reports the estimated values of $\widehat{\theta}_j$ for $j = 1, \ldots, 8$ on April 1, 2016. We also depict the function $\widehat{\rho}_{(j)}(\theta)$ evaluated for $\theta \in \{-30s, -29s, \ldots, -1s, 0s, 1s, \ldots, 29s, 30s\}$ in Figure 1. The table shows that all the estimated lags are negative, which indicates that X^2 (futures) lead X^1 (index). This is consistent with the preceding studies. In Figure 2 we depict the time series of $\widehat{\theta}_j$'s evaluated every trading day. We find that the estimated values of $\widehat{\theta}_j$'s are quite stable in this period, especially at finer resolutions $j \leq 6$. This suggests that there might be a stable multi-scale structure in lead-lag effects between the S&P 500 index and the E-mini S&P500 futures.

Table 3: Estimated values of $\widehat{\theta}_j$ for April 1, 2016 (in seconds)

j	1	2	3	4	5	6	7	8
$\widehat{\theta}_j$	-1	-1	-2	-2	-3	-5	-7	-10

6 Summary

In this paper, we have proposed a new framework to model potential multi-scale structures of lead-lag relationships between financial assets. Our framework has accommodated traditional wavelet methods for analyzing lead-lag relationships to continuous-time modeling by establishing an explicit connection between wavelet and the Lévy-Ciesielski construction of Brownian motion from a statistical viewpoint. We have also developed a statistical methodology to estimate lead-lag times on a scale-by-scale basis in the proposed framework. An associated asymptotic theory has been shown in order to ensure the validity of our methodology. To complement the theory, we have conducted a simulation study which demonstrates the finite sample performance of our asymptotic theory. We have reported an empirical application as well to illustrate how the proposed framework performs in practice.

A drawback of the presented estimation method is that it requires us to interpolate data onto the grid with the finest observable resolution. In some cases this is computational challenging. For example, if the finest observable resolution is one micro-second, then the method requires us to store one million observations per one second. A solution to this issue is currently under investigation and will be presented in future work.

Figure 1: The function $\widehat{\rho}_{(j)}(\theta)$ for April 1, 2016
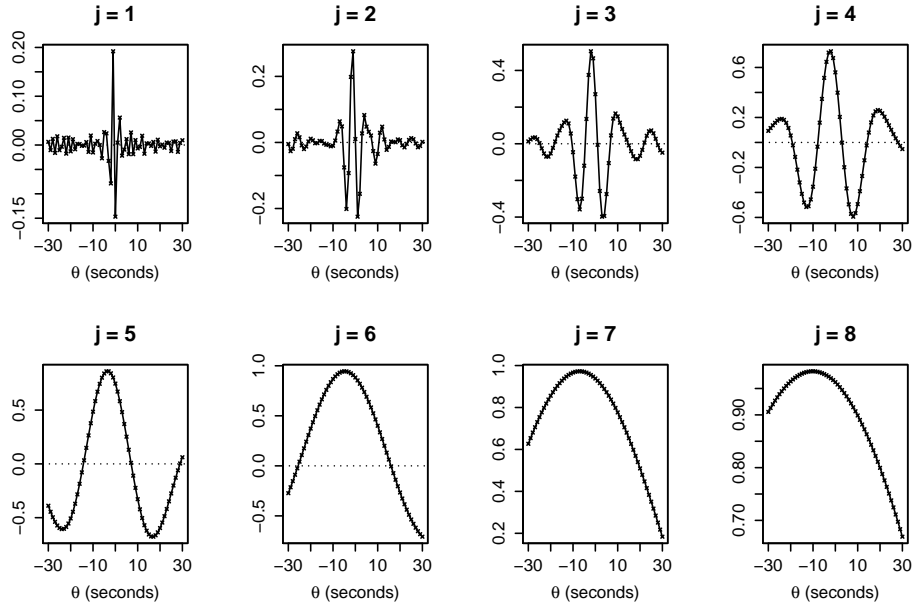
The values are divided by $\frac{\tau_J^{-1}}{n-L_j+1}\sqrt{(\sum_{k=L_j-1}^{n-1}(\mathcal{W}_{jk}^1)^2)(\sum_{k=L_j-1}^{n-1}(\mathcal{W}_{jk}^2)^2)}$ for each j.

Figure 2: The time series of the estimates $\hat{\theta}_j$ in April, 2016 ($1 \le j \le 8$)
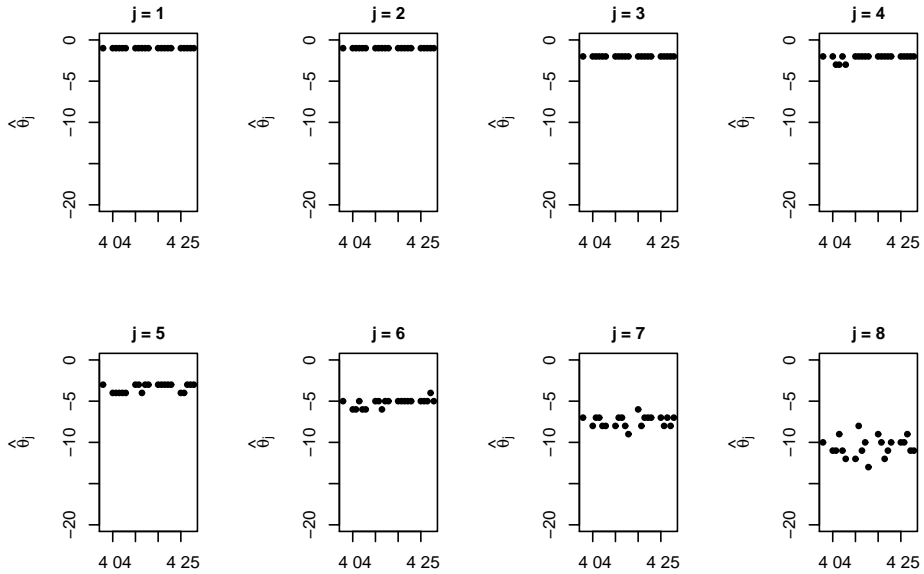
The horizontal axis is labeled by days. The vertical axis is in seconds.

7 Proofs

7.1 Proof of Proposition 1

By Theorem 5.11 from [32] we have $1_{[0,t]} = (1_{[0,t]} * \tilde{\phi}) * \tilde{\phi} + \sum_{j=0}^{\infty} 2^j (1_{[0,t]} * \underline{\tilde{\psi}_j}) * \tilde{\psi}_j$ in $L^2(\mathbb{R})$. Theorem 5.11 from [32] also implies that $(1_{[0,t]} * \underline{\tilde{\phi}}) * \tilde{\phi}, (1_{[0,t]} * \underline{\tilde{\psi}_j}) * \tilde{\psi}_j \in L^2(\mathbb{R})$. Therefore, we obtain

$$B_t = \int_{-\infty}^{\infty} (1_{[0,t]} * \underline{\tilde{\phi}}) * \tilde{\phi}(s) dB_s + \sum_{j=0}^{\infty} 2^j \int_{-\infty}^{\infty} (1_{[0,t]} * \underline{\tilde{\psi}_j}) * \tilde{\psi}_j(s) dB_s$$

in $L^2(\mathbb{R})$. Now, we show that

$$\int_{-\infty}^{\infty} (1_{[0,t]} * \underline{g}) * g(s) dB_s = \int_{-\infty}^{\infty} du (1_{[0,t]} * \underline{g})(u) \int_{-\infty}^{\infty} g(s - u) dB_s \qquad (20)$$

for $g = \tilde{\phi}, \tilde{\psi}_0, \tilde{\psi}_1, \ldots$. We have

$$\int_{-T}^{T} \left| \int_{|u|>A} (1_{[0,t]} * \underline{g})(u) g(s - u) du \right|^2 ds \leq 2T \|g\|_{L^2(\mathbb{R})} \int_{|u|>A} (1_{[0,t]} * \underline{g})(u)^2 du$$

by the Schwarz inequality. Since $(1_{[0,t]} * \underline{g})(u) \in L^2(\mathbb{R})$ by the Young inequality, we conclude that

$$\lim_{A \to \infty} \int_{-T}^{T} \left| \int_{|u|>A} (1_{[0,t]} * \underline{g})(u) g(s - u) du \right|^2 ds = 0.$$

Therefore, we have

$$\int_{-\infty}^{\infty} (1_{[0,t]} * \underline{g}) * g(s) dB_s = \lim_{T \to \infty} \lim_{A \to \infty} \int_{-T}^{T} dB_s \int_{-A}^{A} (1_{[0,t]} * \underline{g})(u) g(s - u) du$$

in $L^2(P)$. Applying stochastic Fubini's theorem (cf. Theorem 5.26 from [33]), we obtain

$$\int_{-\infty}^{\infty} (1_{[0,t]} * \underline{g}) * g(s) dB_s = \lim_{T \to \infty} \lim_{A \to \infty} \int_{-A}^{A} du (1_{[0,t]} * \underline{g})(u) \int_{-T}^{T} g(s - u) dB_s$$

in $L^2(P)$. Now, by an analogous argument to the above we deduce (20). Since we have

$$\int_{-\infty}^{\infty} du (1_{[0,t]} * \underline{g})(u) \int_{-\infty}^{\infty} g(s - u) dB_s = \int_0^t dv \int_{-\infty}^{\infty} g(v - u) \left(\int_{-\infty}^{\infty} g(s - u) dB_s \right) du$$

by Fubini's theorem, we obtain the desired result. $\qquad \square$

7.2 Proof of Proposition 2

We use some concepts on Schwartz's generalized functions and refer to Chapters 6–7 of [39] for details about them.

Let us denote by $\mathfrak{S}$ the set of all (complex-valued) rapidly decreasing functions on $\mathbb{R}$. Define the function $\mathbf{f} : \mathfrak{S} \to \mathbb{C}$ by $\mathbf{f}(u) = \int_{-\infty}^{\infty} (\mathcal{F}^{-1}u)(\lambda) f(\lambda) d\lambda$ for $u \in \mathfrak{S}$, which can be defined thanks to (12). $\mathbf{f}$ is obviously a tempered generalized function on $\mathbb{R}$. Moreover, if $u \in \mathfrak{S}$ is real-valued, then $\mathbf{f}(u) \in \mathbb{R}$. In fact, we have

$$\overline{\mathbf{f}(u)} = \int_{-\infty}^{\infty} \overline{(\mathcal{F}^{-1}u)(\lambda) f(\lambda)} d\lambda = \int_{-\infty}^{\infty} (\mathcal{F}^{-1}u)(-\lambda) f(-\lambda) d\lambda = \mathbf{f}(u)$$

by (13). Now, for any $u \in \mathfrak{S}$ we have $\mathcal{F}(\mathbf{f} * u) = (\mathcal{F}u)(\mathcal{F}\mathbf{f}) = (\mathcal{F}u) f$ in $\mathfrak{S}$, hence $\mathcal{F}(\mathbf{f} * u) \in L^2(\mathbb{R})$. Therefore, $\mathbf{f} * u \in L^2(\mathbb{R})$ and $\|\mathbf{f} * u\|_{L^2(\mathbb{R})} = \|(\mathcal{F}u) f\|_{L^2(\mathbb{R})} \leq 2\pi \|u\|_{L^2(\mathbb{R})}$ by the Parseval identity and (12). Hence, there is a (unique) continuous function $\mathbf{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ such that $\mathbf{F}(u) = \mathbf{f} * u$ for any $u \in \mathfrak{S}$. By continuity $\mathbf{F}(u)$ is real-valued as long

as $u \in L^2(\mathbb{R})$ is real-valued. Similarly, we define the function $\mathbf{g} : \mathfrak{S} \to \mathbb{C}$ by $\mathbf{g}(u) = \int_{-\infty}^{\infty} (\mathcal{F}^{-1}u)(\lambda)\sqrt{1 - |f(\lambda)|^2}d\lambda$ for $u \in \mathfrak{S}$. Then, by an analogous argument to the above there is a (unique) continuous function $\mathbf{G} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ such that $\mathbf{G}(u) = \mathbf{g} * u$ for any $u \in \mathfrak{S}$ and that $\mathbf{G}(u)$ is real-valued as long as $u \in L^2(\mathbb{R})$.

Now let $(W_t^1)_{t \in \mathbb{R}}$ and $(W_t^2)_{t \in \mathbb{R}}$ be two independent two-sided standard Brownian motions. Then we define the processes $(B_t^1)_{t \in \mathbb{R}}$ and $(B_t^2)_{t \in \mathbb{R}}$ by $B_t^1 = W_t^1$ and

$$
B_t^2 = \begin{cases} \int_{-\infty}^{\infty} \mathbf{F}(1_{(0,t]})(s)dW_s^1 + \int_{-\infty}^{\infty} \mathbf{G}(1_{(0,t]})(s)dW_s^2 & \text{if } t \geq 0, \\ -\int_{-\infty}^{\infty} \mathbf{F}(1_{(t,0]})(s)dW_s^1 - \int_{-\infty}^{\infty} \mathbf{G}(1_{(t,0]})(s)dW_s^2 & \text{otherwise.} \end{cases}
$$

B^2 is obviously real-valued and continuous. Moreover, noting that $\mathcal{F}\mathbf{F} = (\mathcal{F}u)f$ and $\mathcal{F}\mathbf{G} = (\mathcal{F}u)\sqrt{1 - |f|^2}$ in $L^2(\mathbb{R})$ for any $u \in L^2(\mathbb{R})$, we can easily check that B^2 is a two-sided standard Brownian motion due to the Parseval identity. Hence B^1 and B^2 satisfy condition (i). Condition (ii) also follows from the Parseval identity. This especially implies that the bivariate process $B_t = (B_t^1, B_t^2)$ is of stationary increments. B_t is evidently Gaussian by construction, hence we complete the proof of the first part of the proposition.

Conversely, suppose that a bivariate process $B_t = (B_t^1, B_t^2)$ ($t \in \mathbb{R}$) with stationary increments satisfies condition (i). Then, by the spectral representation theorem for the structural function of a process with stationary increments (see e.g. Theorem 4 from Chapter I, Section 4 of [22]) there is a function $F : \mathbb{R} \to \mathbb{C}$ of bounded variation such that

$$
E\left[\left(\int_{-\infty}^{\infty} u(s)dB_s^1\right)\overline{\left(\int_{-\infty}^{\infty} v(s)dB_s^2\right)}\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty} (\mathcal{F}u)(\lambda)\overline{(\mathcal{F}v)(\lambda)}dF(\lambda) \tag{21}
$$

for any $u, v \in L^2(\mathbb{R})$ and

$$
|F(\lambda_1) - F(\lambda_2)| \leq |\lambda_1 - \lambda_2| \tag{22}
$$

for any $\lambda_1, \lambda_2 \in \mathbb{R}$. (22) implies that F is absolutely continuous, so F is differentiable almost everywhere on $\mathbb{R}$ and $f := F' \in L^1_{\text{loc}}(\mathbb{R})$ by Theorem 7.20 from [38]. This f satisfies (10) due to (21). Moreover, (12) follows from (22) and Theorem 1.40 from [38]. Finally, by (10) we have

$$
\int_{-\infty}^{\infty} |(\mathcal{F}u)(\lambda)|^2 \overline{f(\lambda)}d\lambda = 2\pi E\left[\overline{\left(\int_{-\infty}^{\infty} u(s)dB_s^1\right)}\left(\int_{-\infty}^{\infty} u(s)dB_s^2\right)\right]
$$
$$
= 2\pi E\left[\left(\int_{-\infty}^{\infty} \overline{u(s)}dB_s^1\right)\overline{\left(\int_{-\infty}^{\infty} \overline{u(s)}dB_s^2\right)}\right] = \int_{-\infty}^{\infty} |(\mathcal{F}\overline{u})(\lambda)|^2 f(\lambda)d\lambda
$$
$$
= \int_{-\infty}^{\infty} |(\mathcal{F}u)(-\lambda)|^2 f(\lambda)d\lambda = \int_{-\infty}^{\infty} |(\mathcal{F}u)(\lambda)|^2 f(-\lambda)d\lambda
$$

for any $u \in L^2(\mathbb{R})$. Therefore, for any bounded Borel set $A \subset \mathbb{R}$, we have $\int_A \overline{f(\lambda)}d\lambda = \int_A f(-\lambda)d\lambda$ by taking $u = \mathcal{F}^{-1}1_A$. Consequently, f satisfies (13) due to Theorem 1.40 from [38]. $\qquad\square$

7.3 Proof of Theorem 1

Throughout the discussions, for sequences (x_J) and (y_J), $x_J \lesssim y_J$ means that there exists a constant $K \in [0, \infty)$ such that $x_J \leq Ky_J$ for large J. Also, for $r > 0$ $\|\cdot\|_r$ denotes the L^r-norm of random variables, i.e. $\|\xi\|_r = (E[|\xi|^r])^{1/r}$ for a random variable ξ.

First we note that, without loss of generality, the volatility processes σ^1 and σ^2 may be assumed to be bounded by a standard localization argument presented in e.g. Section 3 of [5]. In fact, for each $K > 0$ and each $\nu = 1, 2$, let us define $S_K^\nu = \inf\{t : |\sigma_t^\nu| > K\}$. We have $|\sigma_s^\nu| \leq K$ as long as $s < S_K^\nu$. Now define the process $\sigma_s^{\nu,(K)}$ by

$\sigma_s^{\nu,(K)} = \sigma_s^{\nu} 1_{\{s < S_K^{\nu}\}} + \sigma_{S_K^{\nu}-}^{\nu} 1_{\{s \geq S_K^{\nu}\}}$ for $s \geq 0$. $\sigma_s^{\nu,(K)}$ is obviously càdlàg and $(\mathcal{F}_t^{\nu})$-adapted because S_K^{ν} is an $(\mathcal{F}_t^{\nu})$-stopping time. So we can define the process $X_t^{\nu,(K)}$ by $X_t^{\nu,(K)} = \int_0^t \sigma_s^{\nu,(K)} dB_s^{\nu}$ for $t \geq 0$. We associate $\widehat{\rho}_{(j)}^{(K)}(\theta)$ with $X^{1,(K)}$ and $X^{2,(K)}$. Now since we have $\{\widehat{\rho}_{(j)}^{(K)}(\theta) \neq \widehat{\rho}_{(j)}(\theta))\} \subset \{S_K^1 \wedge S_K^2 < t+1\}$ and

$$\limsup_{K \to \infty} P\left(S_K^1 \wedge S_K^2 < t+1\right) \leq \limsup_{K \to \infty} P\left(\max_{\nu=1,2} \sup_{0 \leq s \leq t+1} |\sigma_s^{\nu}| > K\right) = 0$$

because both σ^1 and σ^2 are càdlàg, the results of Theorem 1 hold true once they hold true for $\widehat{\rho}_{(j)}^{(K)}(\theta)$. Consequently, in the following we assume that both σ^1 and σ^2 are bounded.

We use the notation $\mathcal{L}_J = \{l \in \mathbb{Z} : l\tau_J \in \mathcal{G}_J\}$ and $\mathcal{L}_J^+ = \{l \in \mathbb{Z}_+ : l\tau_J \in \mathcal{G}_J\}$. We start by proving the C-tightness of the target quantity.

Lemma 1. *Under the assumptions of Theorem 1, the process $(\widehat{\boldsymbol{\rho}}_{(j)}(\vartheta_J)_t)_{t \geq 0}$ is C-tight as $J \to \infty$.*

Proof. Since $\vartheta_J \in \mathcal{G}_J$, it can be written as $\vartheta_J = l\tau_J$ for some $l = l_J \in \mathcal{L}_J$. For the simplicity of presentation we assume that $l_J \in \mathcal{L}_J^+$ for all J (this assumption can easily be removed).

According to Proposition 3.26 from Chapter VI of [28], it suffices to prove

$$\lim_{A \to \infty} \limsup_{J \to \infty} P\left(\sup_{0 \leq t \leq m} \left|\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t\right| > A\right) = 0, \tag{23}$$

$$\lim_{\eta \to 0} \limsup_{J \to \infty} P\left(w_m(\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J), \eta) > \varepsilon\right) = 0 \tag{24}$$

for any $m \in \mathbb{N}$ and any $\varepsilon > 0$. Here for a function $g : [0, \infty) \to \mathbb{R}$ $w_m(g, \eta)$ denote the modulus of continuity of g on $[0, m]$, i.e. $w_m(g, \eta) = \sup\{|g(s) - s(t)|; s, t \in [0, m], |s - t| \leq \eta\}$.

First we show that there is a constant C such that

$$\left\|\mathcal{W}_{jk}^{\nu}\right\|_4 \leq C\sqrt{\tau_J} \tag{25}$$

for any j, k and $\nu = 1, 2$. Since we can rewrite $\mathcal{W}_{jk}^{\nu}$ as $\mathcal{W}_{jk}^{\nu} = \sum_{\alpha=0}^k \chi_{\nu,k-p}(\alpha) \sum_{p=0}^{(L_j-1) \wedge (k-\alpha)} h_{j,p} \Delta_{k-p-\alpha} X^{\nu}$, the Minkowski and the Burkholder-Davis-Gundy inequalities as well as (16) yield

$$\left\|\mathcal{W}_{jk}^{\nu}\right\|_4 \leq \sum_{\alpha=0}^k \|\chi_{\nu,k-p}(\alpha)\|_4 \left\|\sum_{p=0}^{(L_j-1) \wedge (k-\alpha)} h_{j,p} \Delta_{k-p-\alpha} X^{\nu}\right\|_4 \lesssim \sum_{\alpha=0}^k \pi_{\nu}^{\alpha/4} \left\{\left(\sum_{p=0}^{L_j-1} h_{j,p}^2\right)^2 \cdot \tau_J^2 K^4\right\}^{1/4} \lesssim \sqrt{\tau_J},$$

hence (25) holds true.

Next, (25) and the Schwarz inequality imply that $E\left[\sup_{0 \leq t \leq m} \left|\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t\right|\right] \lesssim \sum_{k=L_j-1}^{\lfloor \tau_J^{-1} m \rfloor - l} E\left[\left|\mathcal{W}_{jk}^1 \mathcal{W}_{jk+l}^2\right|\right] \lesssim m$, hence (23) holds true by the Markov inequality.

Finally, for $0 \leq s \leq t \leq m$, the Schwarz inequality yields

$$\left|\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t - \widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_s\right| \lesssim (\lfloor \tau_J^{-1} t \rfloor - \lfloor \tau_J^{-1} s \rfloor) \sum_{k=L_j-1}^{\lfloor \tau_J^{-1} m \rfloor - l} \left|\mathcal{W}_{jk}^1 \mathcal{W}_{jk+l}^2\right|^2,$$

hence we have $P\left(w_m\left(\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J), \eta\right) > \varepsilon\right) \lesssim \varepsilon^{-1}(\eta + \tau_J)m$ by the Markov and Schwarz inequalities as well as (25). This yields (24). This completes the proof. $\square$

Next we assess the errors induced by interpolation.

16

Lemma 2. *Under the assumptions of Theorem 1, we have* $\sup_{0 \le t \le t_0} \max_{l \in \mathcal{L}_J} \left\| \widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t - E\left[\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t | X\right] \right\|_r =$ $O\left(\sqrt{\tau_J}L_j\right)$ *as* $J \to \infty$ *for any* $t_0 > 0$, *any* $K > 0$ *and any even integer* r.

Proof. By symmetry it is enough to prove $\sup_{0 \le t \le t_0} \max_{l \in \mathcal{L}_J^+} \left\| \widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t - E\left[\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t | X\right] \right\|_r = O\left(\sqrt{\tau_J}L_j\right)$. Set $\mathcal{X}_{k,l,p,q}(\alpha,\beta) = \chi_{1,k-p}(\alpha)\chi_{2,k+l-q}(\beta) - E\left[\chi_{1,k-p}(\alpha)\chi_{2,k+l-q}(\beta)\right]$. For any $l \in \mathcal{L}_J^+$, we can rewrite $\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t - E\left[\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t | X\right]$ as

$$
\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t - E\left[\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t | X\right] = \frac{\tau_J^{-1}}{n-l-L_j+1} \sum_{p,q=0}^{L_j-1} \sum_{\alpha=0}^{\lfloor \tau_J^{-1}t \rfloor - l - p} \sum_{\beta=0}^{\lfloor \tau_J^{-1}t \rfloor - q} h_{j,p}h_{j,q}
$$
$$
\times \sum_{k=(L_j-1)\vee(\alpha+p)\vee(\beta+q-l)}^{\lfloor \tau_J^{-1}t \rfloor - l} \mathcal{X}_{k,l,p,q}(\alpha,\beta)\Delta_{k-p-\alpha}X^1 \Delta_{k+l-q-\beta}X^2.
$$

Therefore, the Minkowski inequality yields

$$
\left\| \widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t - E\left[\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t | X\right] \right\|_r
$$
$$
\lesssim \sum_{p,q=0}^{L_j-1} \sum_{\alpha,\beta=0}^{\infty} |h_{j,p}h_{j,q}| \left\| \sum_{k=(L_j-1)\vee(\alpha+p)\vee(\beta+q-l)}^{\lfloor \tau_J^{-1}t \rfloor - l} \mathcal{X}_{k,l,p,q}(\alpha,\beta)\Delta_{k-p-\alpha}X^1 \Delta_{k+l-q-\beta}X^2 \right\|_r.
$$

Now, by construction $\mathcal{X}_{k,l,p,q}(\alpha,\beta)$ and $\mathcal{X}_{k',l,p,q}(\alpha,\beta)$ are independent if $|k-k'| > \alpha \vee \beta$, and we have $E\left[|\mathcal{X}_{k,l,p,q}(\alpha,\beta)|^A\right] \lesssim \pi_1^{\alpha}\pi_2^{\beta}$ by construction and $E[|\Delta_{k-p-\alpha}X^1|^A] + E[|\Delta_{k+l-q-\beta}X^2|^A] \lesssim \tau_J^{A/2}$ by the Burkholder-Davis-Gundy inequality for any $A > 1$. Therefore, by Theorem 2 from [16] we obtain

$$
\sup_{0 \le t \le t_0} \max_{l \in \mathcal{L}_J^+} \left\| \sum_{k=(L_j-1)\vee(\alpha+p)\vee(\beta+q-l)}^{\lfloor \tau_J^{-1}t \rfloor - l} \mathcal{X}_{k,l,p,q}(\alpha,\beta)\Delta_{k-p-\alpha}X^1 \Delta_{k+l-q-\beta}X^2 \right\|_r^r
$$
$$
\lesssim \tau_J^r \left\{ \left(\tau_J^{-1}(\alpha \vee \beta + 2)\pi_1^{2\alpha}\pi_2^{2\beta}\right)^{r/2} + \tau_J^{-1}(\alpha \vee \beta + 2)^{r-1}\pi_1^{\alpha}\pi_2^{\beta} \right\},
$$

hence we conclude that

$$
\sup_{0 \le t \le t_0} \max_{l \in \mathcal{L}_J^+} \left\| \widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t - E\left[\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t | X\right] \right\|_r \lesssim \sqrt{\tau_J} \sum_{p,q=0}^{L_j-1} |h_{j,p}h_{j,q}| \le \sqrt{\tau_J}L_j,
$$

where we use the inequality

$$
\sum_{p=0}^{L_j-1} |h_{j,p}| \le \sqrt{L_j \sum_{p=0}^{L_j-1} h_{j,p}^2} = \sqrt{L_j}. \tag{26}
$$

Thus we complete the proof. $\qquad\square$

Now we investigate the asymptotic behavior of $E\left[\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t | X\right]$. For $\nu = 1, 2$ and $k \ge L_j - 1$, we define the variables Z_k^{ν} by $Z_k^{\nu} = E\left[\mathcal{W}_{jk}^{\nu} | X\right] = (1 - \pi_{\nu}) \sum_{p=0}^{L_j-1} \sum_{\alpha=0}^{k-p} h_{j,p}\pi_{\nu}^{\alpha}\Delta_{k-p-\alpha}X^{\nu}$. Thanks to the independence between δ_k^1's and δ_k^2's, we have

$$
E\left[\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t | X\right] = \frac{\tau_J^{-1}}{n-l-L_j+1} \sum_{k=L_j-1}^{\lfloor \tau_J^{-1}t \rfloor - l} Z_k^{\nu} Z_{k+l}^{\nu} \tag{27}
$$

17

for $l \geq 0$. To analyze the asymptotic behavior of this quantity, we introduce the "de-volatilized" version of Z_k^ν as follows:

$$\zeta_k^\nu = (1 - \pi_\nu) \sum_{p=0}^{L_j-1} \sum_{\alpha=0}^{k-p} h_{j,p} \pi_\nu^\alpha \Delta_{k-p-\alpha} B^\nu.$$

Since ζ_k^ν's are centered Gaussian variables, their distribution is completely determined by their covariance structure. To investigate their covariance structure we introduce the following auxiliary quantity for each $\theta \in \mathbb{R}$:

$$\bar{\rho}_j(\theta) = \int_{-\infty}^{\infty} D(\lambda) H_{j,L}(\lambda) \Pi(\lambda) e^{\sqrt{-1}\tau_J^{-1}\theta\lambda} f_J(\tau_J^{-1}\lambda) d\lambda.$$

In the following N denotes a positive integer such that $\tau_J^w N \to c$ for some $w \in (0, \frac{1}{2})$ and $c \in (0, \infty)$.

Lemma 3. (a) *We have* $\left| E\left[\zeta_k^\nu \zeta_{k'}^\nu \right] \right| \leq \tau_J (1 - \pi_\nu)^2 \sum_{\alpha,\beta=0}^{\infty} \pi_\nu^{\alpha+\beta} \sum_{p=0}^{L_j-1} |h_{j,p}| 1_{\{|k'-k-\beta+\alpha|<L_j\}}$ *for any* $\nu = 1, 2$ *and any* $k, k' \in \mathbb{Z}_+$.

(b) *We have* $\max_{k,k',l \in \mathbb{Z}_+ : k, k' \geq N+L_j} \left| E\left[\zeta_k^1 \zeta_{k'+l}^2 \right] - \tau_J \bar{\rho}_j((k'+l-k)\tau_J) \right| = O(\tau_J L_j^2 (\pi_1^N + \pi_2^N))$ *as* $J \to \infty$.

Proof. First, claim (a) follows from the following identity

$$E\left[\zeta_k^\nu \zeta_{k'}^\nu \right] = \tau_J (1 - \pi_\nu)^2 \sum_{\alpha=0}^{k} \sum_{\beta=0}^{k'} \sum_{\substack{p=0 \\ 0 \leq k'-\beta-(k-p-\alpha) \leq L_j-1}}^{(L_j-1)\wedge(k-\alpha)} h_{j,p} h_{j,k'-\beta-(k-p-\alpha)} \pi_\nu^{\alpha+\beta}$$

and the inequality $|h_{j,p}| \leq 1$.

Next, using the identity

$$E\left[\Delta_k B^1 \Delta_{k+l} B^2 \right] = \tau_J \int_{-\infty}^{\infty} D(\lambda) e^{\sqrt{-1}l} f_J(\tau_J^{-1}\lambda) d\lambda, \tag{28}$$

we have

$$E\left[\zeta_k^1 \zeta_{k'+l}^2 \right]$$

$$= (1 - \pi_1)(1 - \pi_2) \sum_{p,q=0}^{L_j-1} \sum_{\alpha=0}^{k-p} \sum_{\beta=0}^{k'+l-q} h_{j,p} h_{j,q} \pi_1^\alpha \pi_2^\beta E\left[\Delta_{k-p-\alpha} B^1 \Delta_{k'+l-q-\beta} B^2 \right]$$

$$= \tau_J (1 - \pi_1)(1 - \pi_2) \sum_{p,q=0}^{L_j-1} \sum_{\alpha=0}^{k-p} \sum_{\beta=0}^{k'+l-q} h_{j,p} h_{j,q} \pi_1^\alpha \pi_2^\beta \int_{-\infty}^{\infty} D(\lambda) e^{\sqrt{-1}(k'+l-q-\beta-k+p+\alpha)\lambda} f_J(\tau_J^{-1}\lambda) d\lambda.$$

Then, noting that the identity $\Pi(\lambda) = (1 - \pi_1)(1 - \pi_2) \sum_{\alpha,\beta=0}^{\infty} \pi_1^\alpha \pi_2^\beta e^{\sqrt{-1}\lambda(\alpha-\beta)}$, we obtain

$$E\left[\zeta_k^1 \zeta_{k'+l}^2 \right] = \tau_J \int_{-\infty}^{\infty} D(\lambda) H_{j,L}(\lambda) \Pi(\lambda) e^{\sqrt{-1}(k'+l-k)\lambda} f_J(\tau_J^{-1}\lambda) d\lambda + O(\tau_J L^2 (\pi_1^N + \pi_2^N))$$

$$= \tau_J \bar{\rho}_j((k'+l-k)\tau_J) + O(\tau_J L^2 (\pi_1^N + \pi_2^N))$$

uniformly in $k, k' \geq N + L_j$ and $l \in \mathbb{Z}_+$, where we use (26). Hence claim (b) also holds true. $\qquad\square$

From now on we investigate the asymptotic behavior of $\bar{\rho}_j(\theta)$. We need the following auxiliary result.

Lemma 4. *It holds that*

$$\sup_{\lambda \in [0,\pi]} \left| \frac{d}{d\lambda} H_{j,L}(\lambda) \right| \leq (2^j - 1) c_L, \text{ where } c_L = \binom{L-2}{L/2-1} \frac{L-1}{2^{L-1}}.$$

Moreover, $c_L = O(\sqrt{L})$ *as* $L \to \infty$.

Proof. From the proof of Theorem 3 from [30] we have

$$\frac{d}{d\lambda}G_L(\lambda) = \cos^{L-2}\left(\frac{\lambda}{2}\right)\binom{L-2}{L/2-1}(L-1)\sin^{L-2}\left(\frac{\lambda}{2}\right)\frac{\sin\lambda}{2} = \binom{L-2}{L/2-1}\frac{L-1}{2^{L-1}}\sin^{L-1}(\lambda),$$

where we use the double angle formula for the sine function. Since we have $H_L(\lambda) = G_L(\lambda + \pi)$, the first part of the lemma follows from (15) and the Leibniz rule.

The latter part is a consequence of Stirling's formula. $\quad\square$

We can rewrite $\bar{\rho}_j(\theta)$ as

$$\bar{\rho}_j(\theta) = \sum_{i=1}^{J+1} R_i \int_{\Lambda_{-i}} D(\lambda)H_{j,L}(\lambda)\Pi(\lambda)e^{\sqrt{-1}\tau_J^{-1}(\theta-\theta_i)\lambda}d\lambda =: \sum_{i=1}^{J+1} \bar{\rho}_{j,(i)}(\theta).$$

Lemma 5. *There is a constant A such that*

$$\left|\int_{\pi/2^i}^{\pi/2^{i-1}} D(\lambda)H_{j,L}(\lambda)\Pi(\lambda)e^{\sqrt{-1}a\lambda}d\lambda\right| \leq \frac{A}{|a|}\left\{1 + (2^j - 1)c_L\right\}$$

for any positive integers i, j, any even positive integer L and any non-zero real number a.

Proof. Integration by parts yields

$$\int_{\pi/2^i}^{\pi/2^{i-1}} D(\lambda)H_{j,L}(\lambda)\Pi(\lambda)e^{\sqrt{-1}a\lambda}d\lambda = \frac{1}{\sqrt{-1}a}\left[D(\lambda)H_{j,L}(\lambda)\Pi(\lambda)e^{\sqrt{-1}a\lambda}\right]_{\pi/2^i}^{\pi/2^{i-1}}$$

$$- \frac{1}{\sqrt{-1}a}\int_{\pi/2^i}^{\pi/2^{i-1}} \frac{d}{d\lambda}\left\{D(\lambda)H_{j,L}(\lambda)\Pi(\lambda)\right\}e^{\sqrt{-1}a\lambda}d\lambda.$$

We can easily see that $\sup_{\lambda\in\mathbb{R}}\left|\frac{d}{d\lambda}\left\{D(\lambda)\Pi(\lambda)\right\}\right| < \infty$, hence the desired result follows from Lemma 4. $\quad\square$

Lemma 6. *Under the assumptions of Theorem 1(a), we have $\bar{\rho}_j(\vartheta_J) \to 0$ as $J \to \infty$.*

Proof. Since we have $|\bar{\rho}_{j,(i)}(\vartheta_J)| \leq \pi(\tau_{i-1} - \tau_i)$ for every i, it is enough to prove

$$\bar{\rho}_{j,(i)}(\vartheta_J) \to 0 \tag{29}$$

as $J \to \infty$ for any fixed i due to the dominated convergence theorem. By Theorem 1 from [30], we have

$$\lim_{L\to\infty} H_{j,L}(\lambda) = \begin{cases} 1 & \text{if } \lambda \in (\frac{\pi}{2^j}, \frac{\pi}{2^{j-1}}), \\ 0 & \text{if } \lambda \in [0, \frac{\pi}{2^j}) \cup (\frac{\pi}{2^{j-1}}, \pi]. \end{cases} \tag{30}$$

Therefore, (29) holds true if $i \neq j$ due to the bounded convergence theorem. On the other hand, if $i = j$, Lemma 5 yields $\left|\bar{\rho}_{j,(i)}(\vartheta_J)\right| \lesssim \frac{\tau_J c_L}{\vartheta_J - \theta_j}$, hence we obtain the desired result by assumption and $c_L = O(\sqrt{L})$. $\quad\square$

Lemma 7. *Under the assumptions of Theorem 1(b), we have $\bar{\rho}_j(l\tau_J) \to R_j \int_{\Lambda_{-j}} D(\lambda)\Pi(\lambda)e^{\sqrt{-1}b\lambda}d\lambda$ as $J \to \infty$.*

Proof. From the above argument it suffices to prove

$$R_j \int_{\Lambda_{-j}} D(\lambda)H_{j,L}(\lambda)\Pi(\lambda)e^{\sqrt{-1}\tau_J^{-1}(l\tau_J-\theta_j)\lambda}d\lambda \to R_j \int_{\Lambda_{-j}} D(\lambda)\Pi(\lambda)e^{\sqrt{-1}b\lambda}d\lambda,$$

which follows from (30) and the bounded convergence theorem. $\quad\square$

Proof of Theorem 1. Since $\vartheta_J \in \mathcal{G}_J$, it can be written as $\vartheta_J = l\tau_J$ for some $l = l_J \in \mathcal{L}_J$. For the simplicity of presentation we assume that $l_J \in \mathcal{L}_J^+$ for all J.

Since $l\tau_J \in (-\delta, \delta)$ for every $l = l_J$, all the subsequence of the sequence $(l\tau_J)_{J \geq 1}$ has converging subsequences. Therefore, without loss of generality we may assume that $l\tau_J \to \theta$ as $J \to \infty$ for some $\theta \in [-\delta, \delta]$. Note that $\theta = \theta_j$ for case (b) by assumption.

Set $\mathfrak{c} = 0$ for case (a) and $\mathfrak{c} = R_j \int_{\Lambda_{-j}} D(\lambda)\Pi(\lambda)e^{\sqrt{-1}b\lambda}d\lambda$ for case (b). We need to prove $\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t \xrightarrow{ucp} \Sigma_t(\theta)\mathfrak{c}$ as $J \to \infty$. By Lemma 1 it is enough to show that $\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t \to \Sigma_t(\theta)\mathfrak{c}$ as $J \to \infty$ for any fixed $t > 0$. We decompose $\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t$ as

$$
\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t = \left(\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t - E\left[\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t | X \right] \right)
$$

$$
+ \left(E\left[\widehat{\boldsymbol{\rho}}_{(j)}(l\tau_J)_t | X \right] - \frac{\tau_J^{-1}}{n - l - L_j + 1} \sum_{k=L_j-1}^{\lfloor t\tau_J^{-1} \rfloor - l} \sigma_{(k-L_j-N)+\tau_J}^1 \sigma_{\{(k-L_j-N)_++l\}\tau_J}^2 \zeta_k^1 \zeta_{k+l}^2 \right)
$$

$$
+ \frac{\tau_J^{-1}}{n - l - L_j + 1} \sum_{k=L_j-1}^{\lfloor t\tau_J^{-1} \rfloor - l} \sigma_{(k-L_j-N)+\tau_J}^1 \sigma_{\{(k-L_j-N)_++l\}\tau_J}^2 \zeta_k^1 \zeta_{k+l}^2
$$

$$
=: \mathbf{I}_J + \mathbf{II}_J + \mathbf{III}_J.
$$

First, since $\|\mathbf{I}_J\|_2 = O(\sqrt{\tau_J}L) = o(1)$ as $J \to \infty$ by Lemma 2, we obtain $\mathbf{I}_J \to^p 0$.

Next, noting (27), we decompose $\mathbf{II}_J$ as

$$
\mathbf{II}_J = \frac{\tau_J^{-1}}{n - l - L_j + 1} \sum_{k=L_j-1}^{\lfloor t\tau_J^{-1} \rfloor - l} \left\{ \left(Z_k^1 - \sigma_{(k-L_j-N)+\tau_J}^1 \zeta_k^1 \right) Z_{k+l}^2 \right.
$$

$$
\left. + \sigma_{(k-L_j-N)+\tau_J}^1 \left(Z_{k+l}^2 - \sigma_{\{(k-L_j-N)_++l\}\tau_J}^2 \zeta_{k+l}^2 \right) \right\}
$$

$$
= \mathbf{II}_J' + \mathbf{II}_J''.
$$

We have $|\mathbf{II}_J'| \lesssim \sum_{k=L_j-1}^{\lfloor t\tau_J^{-1} \rfloor - l} \left| Z_k^1 - \sigma_{(k-L_j-N)+\tau_J}^1 \zeta_k^1 \right| \left| Z_{k+l}^2 \right|$ Since we have

$$
Z_k^1 - \sigma_{(k-L_j-N)+\tau_J}^1 \zeta_k^1
$$

$$
= (1 - \pi_1) \left\{ \sum_{\alpha=0}^{k \wedge N} \pi_1^\alpha \sum_{p=0}^{(L_j-1) \wedge (k-\alpha)} h_{j,p} \int_{(k-p-\alpha)\tau_J}^{(k-p-\alpha+1)\tau_J} (\sigma_s^1 - \sigma_{(k-L_j-N)+\tau_J}^1)dB_s^1 \right.
$$

$$
\left. + \sum_{\alpha=N+1}^{k} \sum_{p=0}^{(L_j-1) \wedge (k-\alpha)} h_{j,p}\pi_1^\alpha \left(\Delta_{k-p-\alpha}X^1 - \sigma_{(k-L_j-N)+\tau_J}^1 \Delta_{k-p-\alpha}B^1 \right) \right\},
$$

it holds that

$$
\left\| Z_k^1 - \sigma_{i\tau_m}^1 \xi_k^1 \right\|_2 \lesssim \sqrt{\tau_J} \left\| \sup_{s \in [(k-L_j-N)+\tau_J, (k+1)\tau_J)} \left| \sigma_s^1 - \sigma_{(k-L_j-N)+\tau_J}^1 \right| \right\|_2 + \sqrt{\tau_J}\pi_1^N
$$

by the triangle inequality and (16). On the other hand, since we can rewrite Z_{k+l}^2 as

$$
Z_{k+l}^2 = (1 - \pi_2) \sum_{\beta=0}^{k+l} \pi_2^\beta \sum_{q=0}^{(L_j-1) \wedge (k+l-\beta)} h_{j,q}\Delta_{k+l-q-\beta}X^2,
$$

20

we have $\|Z_{k+l}^2\|_2 \lesssim \sqrt{\tau_J}$ by the triangle inequality, the boundedness of $\bar{\sigma}^2$ and (16). Hence we obtain

$$\|\mathbf{II}_J'\|_1 \lesssim \tau_J \sum_{k=L_j-1}^{\lfloor t\tau_J^{-1}\rfloor - l} \left\| \sup_{s\in[(k-L_j-N)_+\tau_J,(k+1)\tau_J)} \left| \sigma_s^1 - \sigma_{(k-L_j-N)_+\tau_J}^1 \right| \right\|_2 + \pi_1^N$$

by the Schwarz inequality. Since σ^1 is càdlàg and bounded, the bounded convergence theorem yields $\|\mathbf{II}_J'\|_1 \to 0$, hence $\mathbf{II}_J' \to^p 0$. By an analogous argument we can prove $\mathbf{II}_J'' \to^p 0$, hence we obtain $\mathbf{II}_J \to^p 0$.

Finally we prove $\mathbf{III}_J \to^p \Sigma_t(\theta)\mathfrak{c}$. It suffices to prove

$$\sup_{0\leq s\leq t} \left| \sum_{k=L_j-1}^{\lfloor s\tau_J^{-1}\rfloor - l} \sigma_{(k-L_j-N)_+\tau_J}^1 \sigma_{\{(k-L_j-N)_++l\}\tau_J}^2 \zeta_k^1 \zeta_{k+l}^2 - \Sigma_s(\theta)\mathfrak{c} \right| \to^p 0. \tag{31}$$

Define the process $A^J = (A_s^J)_{s\geq 0}$ by $A_s^J = \sum_{k=L_j-1}^{\lfloor s\tau_J^{-1}\rfloor - l} \left(\zeta_k^1 \zeta_{k+l}^2 - E\left[\zeta_k^1 \zeta_{k+l}^2\right] \right)$. A^J is obviously of (locally) bounded variation. Moreover, since it holds that $\|\zeta_k^\nu\|_2 \leq (1-\pi_\nu)\sum_{\alpha=0}^\infty \pi_\nu^\alpha \sqrt{\sum_{p=0}^{L_j-1} h_{j,p}^2 \tau_J} = \sqrt{\tau_J}$ by the triangular inequality and (16), we have $E\left[\sum_{k=L_j-1}^{\lfloor s\tau_J^{-1}\rfloor - l} |\zeta_k^1 \zeta_{k+l}^2 - E\left[\zeta_k^1 \zeta_{k+l}^2\right]|\right] = O(1)$ for any $s > 0$ by the Schwarz inequality. Hence A^J is P-UT by 6.6 from chapter VI of [28]. Moreover, the process $\left(\sigma_{(\lfloor s\tau_J^{-1}\rfloor - L_j-N)_+\tau_J}^1 \sigma_{\{(\lfloor s\tau_J^{-1}\rfloor - L_j-N)_++l\}\tau_J}^2\right)_{s\geq 0}$ converges in probability to the process $(\sigma_s^1\sigma_{s+\theta}^2)_{s\geq 0}$ for the Skorokhod topology by Proposition 6.37 from chapter VI of [28]. Therefore, according to Theorem 6.22 from chapter VI of [28], (31) follows once we show that

$$\sup_{0\leq s\leq t} \left| \sum_{k=L_j-1}^{\lfloor s\tau_J^{-1}\rfloor - l} \zeta_k^1 \zeta_{k+l}^2 - \mathfrak{c}(s-\theta)_+ \right| \to^p 0.$$

By Lemma 1 it is enough to prove $\sum_{k=L_j-1}^{\lfloor s\tau_J^{-1}\rfloor - l} \zeta_k^1 \zeta_{k+l}^2 \to^p \mathfrak{c}(s-\theta)_+$ for any fixed s. Moreover, by Lemmas 3 and 6–7 this follows from

$$\sum_{k=L_j-1}^{\lfloor s\tau_J^{-1}\rfloor - l} \left(\zeta_k^1 \zeta_{k+l}^2 - E\left[\zeta_k^1 \zeta_{k+l}^2\right] \right) \to^p 0. \tag{32}$$

To prove (32), let us define the random vector ζ by

$$\zeta = \left(\zeta_{L_j-1}^1, \cdots, \zeta_{\lfloor \tau_J^{-1}s\rfloor - l}^1, \zeta_{L_j-1+l}^2, \cdots, \zeta_{\lfloor \tau_J^{-1}s\rfloor}^2 \right)^\top. \tag{33}$$

Then we have $\sum_{k=L_j-1}^{\lfloor s\tau_J^{-1}\rfloor - l} \left(\zeta_k^1 \zeta_{k+l}^2 - E\left[\zeta_k^1 \zeta_{k+l}^2\right] \right) = \zeta^\top A_J \zeta - E\left[\zeta^\top A_J \zeta\right]$, where

$$A_J = \begin{pmatrix} 0 & E_{\lfloor \tau_J^{-1}s\rfloor - l - L_j + 2} \\ E_{\lfloor \tau_J^{-1}s\rfloor - l - L_j + 2} & 0 \end{pmatrix}.$$

Therefore, the proof of (32) is completed once we show that $\mathrm{Var}[\zeta^\top A_J \zeta] \to 0$ as $J \to \infty$. Since ζ is centered Gaussian, we have $\mathrm{Var}[\zeta^\top A_J \zeta] = 2\,\mathrm{tr}[(\Sigma_J A_J)^2]$ from the discussion in Section 3.2.1 of [12], where Σ_J denotes the covariance matrix of ζ. Since $\mathrm{tr}[(\Sigma_J A_J)^2] \leq \|\Sigma_J A_J\|_F^2 \leq \|\Sigma_J\|_F^2$ by Appendix II(ii)–(iii) from [14], it is enough to prove $\|\Sigma_J\|_F^2 \to 0$, where $\|\cdot\|_F$ denote the Frobenius norm of matrices.

First, by Lemma 3(a) and (26) we have $\sum_{k,k'=L_j-1}^{\lfloor s\tau_J^{-1}\rfloor - l} \left(\left|E\left[\zeta_k^1 \zeta_{k'}^1\right]\right|^2 + \left|E\left[\zeta_{k+l}^1 \zeta_{k'+l}^1\right]\right|^2 \right) = O\left(\tau_J L_j^2\right) = o(1)$. Next, by Lemma 3(b) we have $\sum_{k,k'=L_j-1}^{\lfloor s\tau_J^{-1}\rfloor - l} \left|E\left[\zeta_k^1 \zeta_{k'+l}^2\right]\right|^2 = \tau_J^2 \sum_{k,k'=L_j+N}^{\lfloor s\tau_J^{-1}\rfloor - l} \bar{\rho}_j((k'+l-k)\tau_J)^2 + o(1)$. Therefore, to complete

the proof of $\|\Sigma_J\|_F^2 \to 0$, it is enough to show that

$$\tau_J^2 \sum_{k,k'=L_j+N}^{\lfloor s\tau_J^{-1}\rfloor-l} \bar{\rho}_j((k'+l-k)\tau_J)^2 \to 0. \tag{34}$$

We rewrite the target quantity as

$$\tau_J^2 \sum_{k,k'=L_j+N}^{\lfloor s\tau_J^{-1}\rfloor-l} \bar{\rho}_j((k'+l-k)\tau_J)^2$$

$$= \tau_J^2 \sum_{i_1,i_2=1}^{J+1} \sum_{k,k'=L_j+N}^{\lfloor s\tau_J^{-1}\rfloor-l} \prod_{r=1}^{2} R_{i_r} \int_{\Lambda_{-i}} D(\lambda) H_{j,L}(\lambda) \Pi(\lambda) e^{\sqrt{-1}\lambda(k'+l-k-\theta_{i_r}\tau_J^{-1})} d\lambda$$

$$=: \sum_{i_1,i_2=1}^{J+1} \Xi_J(i_1,i_2).$$

Since we have $|\Xi_J(i_1,i_2)| \lesssim (\tau_{i_1-1}-\tau_{i_1})(\tau_{i_2-1}-\tau_{i_2})$, we obtain (34) by the dominated convergence theorem once we prove $\Xi_J(i_1,i_2) \to 0$ for any fixed i_1,i_2. By Lemma 5 we have

$$|\Xi_J(i_1,i_2)| \lesssim \tau_J + \tau_J^2 \sum_{\substack{k,k'=N+L_j \\ \theta_{i_1},\theta_{i_2}\neq(k'+l-k)\tau_J}}^{\lfloor \tau_J^{-1}t\rfloor-l} \prod_{r=1}^{2} \frac{c_L}{\left|k'+l-k-\theta_{i_r}\tau_J^{-1}\right|}$$

$$\lesssim \tau_J + \tau_J^2 c_L^2 \sum_{r=1}^{2} \sum_{\substack{k,k'=N+L_j \\ \theta_{i_r}\neq(k'+l-k)\tau_J}}^{\lfloor \tau_J^{-1}t\rfloor-l} \frac{1}{\left|k'+l-k-\theta_{i_r}\tau_J^{-1}\right|^2} \lesssim \tau_J + \tau_J c_L^2.$$

Since $c_L = O(\sqrt{L})$, we conclude that $\Xi_J(i_1,i_2) \to^p 0$.

Consequently, we complete the proof of the theorem. $\qquad\square$

7.4   Proof of Theorem 2

Similarly to Section 7.3, a localization procedure allows us to assume that both σ^1 and σ^2 are bounded as well as there is a constant $K > 0$ such that

$$|\sigma_t^1 - \sigma_s^1| + |\sigma_t^2 - \sigma_s^2| \leq K|t-s|^\gamma \tag{35}$$

for any $t, s \geq 0$.

The following result is a counterpart of Proposition 3 from Hoffmann *et al.* [25]:

Lemma 8. *Under the assumptions of Theorem 2, we have* $\max_{\theta\in\mathcal{G}_J:v_J^{-1}|\theta-\theta_j|>\varepsilon} \left|\widehat{\rho}_{(j)}(\theta)\right| \to^p 0$ *as* $J \to \infty$ *for any* $\varepsilon > 0$.

Proof. By symmetry and the triangular inequality, it suffices to prove the following equations:

$$\max_{l\in\mathcal{L}_J^+:v_J^{-1}|l\tau_J-\theta_j|>\varepsilon} \left|\widehat{\rho}_{(j)}(l\tau_J) - E\left[\widehat{\rho}_{(j)}(l\tau_J)|X\right]\right| \to^p 0, \tag{36}$$

$$\max_{l\in\mathcal{L}_J^+:v_J^{-1}|l\tau_J-\theta_j|>\varepsilon} \left|E\left[\widehat{\rho}_{(j)}(l\tau_J)|X\right]\right| \to^p 0. \tag{37}$$

First we show (36). Take $\varepsilon > 0$ arbitrarily. Then, for any even integer r the Markov inequality yields

$$P\left(\max_{l\in\mathcal{L}_J^+:v_J^{-1}|l\tau_J-\theta_j|>\varepsilon} \left|\widehat{\rho}_{(j)}(l\tau_J) - E\left[\widehat{\rho}_{(j)}(l\tau_J)|X\right]\right| > \varepsilon\right)$$

$$\leq \varepsilon^{-r} E\left[\max_{l\in\mathcal{L}_J^+:v_j^{-1}|l\tau_J-\theta_j|>\varepsilon} \left|\widehat{\rho}_{(j)}(l\tau_J)-E\left[\widehat{\rho}_{(j)}(l\tau_J)|X\right]\right|^r\right]$$

$$\leq \varepsilon^{-r}\sum_{l\in\mathcal{L}_J^+} E\left[\left|\widehat{\rho}_{(j)}(l\tau_J)-E\left[\widehat{\rho}_{(j)}(l\tau_J)|X\right]\right|^r\right] \leq \varepsilon^{-r}\#\mathcal{L}_J^+ \max_{l\in\mathcal{L}_J^+} E\left[\left|\widehat{\rho}_{(j)}(l\tau_J)-E\left[\widehat{\rho}_{(j)}(l\tau_J)|X\right]\right|^r\right].$$

We can take sufficiently large r such that $2/r < 1-\kappa$. Then we have

$$\#\mathcal{L}_J^+ \max_{l\in\mathcal{L}_J^+} E\left[\left|\widehat{\rho}_{(j)}(l\tau_J)-E\left[\widehat{\rho}_{(j)}(l\tau_J)|X\right]\right|^r\right] = O\left(\left(\tau_J^{1-2/r}L^2\right)^{r/2}\right) = o(1)$$

by Lemma 2. This yields the desired result.

Next we show (37). We adopt an analogous strategy to the one used in the proof of Theorems 3–4 from [4]. Set $m = \{(1-\kappa)\wedge\frac{1}{4}\}\frac{J}{2}$ and

$$I_m(i) = \{k\in\mathbb{Z}_+ : (k-L_j-N)\tau_J \in [i\tau_m, (i+1)\tau_m)\}$$
$$= \{2^{J-m}i+L_j+N, 2^{J-m}i+L_j+N+1, \ldots, 2^{J-m}(i+1)+L_j+N-1\}.$$

We decompose $E\left[\widehat{\rho}_{(j)}(\theta)|X\right]$ as

$$E\left[\widehat{\rho}_{(j)}(\theta)|X\right]$$
$$= \frac{\tau_J^{-1}}{n-l-L_j+1}\left\{\sum_{k=L_j-1}^{L_j+N-1} Z_k^1 Z_{k+l}^2 + \sum_{k=2^{J-m}M_J}^{n-1-l} Z_k^1 Z_{k+l}^2 + \sum_{i=0}^{M_J-1}\sum_{k\in I_m(i)}\left(Z_k^1-\sigma_{i\tau_m}^1\zeta_k^1\right)Z_{k+l}^2\right.$$

$$+ \sum_{i=0}^{M_J-1}\sum_{k\in I_m(i)}\sigma_{i\tau_m}^1\zeta_k^1\left(Z_{k+l}^2-\sigma_{i\tau_m+l\tau_J}^2\zeta_{k+l}^2\right) + \sum_{i=0}^{M_J-1}\sigma_{i\tau_m}^1\sigma_{i\tau_m+l\tau_J}^2\sum_{k\in I_m(i)}\left(\zeta_k^1\zeta_{k+l}^2-E\left[\zeta_k^1\zeta_{k+l}^2\right]\right)$$

$$\left.+ \sum_{i=0}^{M_J-1}\sigma_{i\tau_m}^1\sigma_{i\tau_m+l\tau_J}^2\sum_{k\in I_m(i)} E\left[\zeta_k^1\zeta_{k+l}^2\right]\right\}$$

$$=: \mathbb{I}_J(l)+\mathbb{II}_J(l)+\mathbb{III}_J(l)+\mathbb{IV}_J(l)+\mathbb{V}_J(l)+\mathbb{VI}_J(l),$$

where $M_J = \lfloor 2^{m-J}(n-l-L_j-N)\rfloor$.

First we prove $\max_{l\in\mathcal{L}_J^+}|\mathbb{I}_J(l)| \to^p 0$. Since we have $\|Z_k^\nu\|_4 \lesssim \sqrt{\tau_J}$ by the Minkowski and Burkholder-Davis-Gundy inequalities as well as (16), we obtain $\|\mathbb{I}_J(l)\|_2 \lesssim \sum_{k=L_j-1}^{L_j+N-1}\|Z_k^1\|_4\|Z_{k+l}^2\|_4 \lesssim N\tau_J$ by the triangle and Schwarz inequalities. Therefore, the Markov inequality yields $P\left(\max_{l\in\mathcal{L}_J^+}|\mathbb{I}_J(l)| > \varepsilon\right) \leq \varepsilon^{-1}\sum_{l\in\mathcal{L}_J^+}\|\mathbb{I}_J(l)\|_2^2 \lesssim N^2\tau_J$ for any $\varepsilon > 0$, hence $\max_{l\in\mathcal{L}_J^+}|\mathbb{I}_J(l)| \to^p 0$.

Noting that $L^2\tau_J \to 0$, we can prove $\max_{l\in\mathcal{L}_J^+}|\mathbb{II}_J(l)| \to^p 0$ in an analogous manner to the above.

Next we prove $\max_{l\in\mathcal{L}_J^+}|\mathbb{III}_J(l)| \to^p 0$. For any $k \in I_m(i)$ we have

$$Z_k^1-\sigma_{i\tau_m}^1\zeta_k^1 = (1-\pi_1)\left\{\sum_{\alpha=0}^{k\wedge N}\pi_1^\alpha\sum_{p=0}^{(L_j-1)\wedge(k-\alpha)} h_{j,p}\int_{(k-p-\alpha)\tau_J}^{(k-p-\alpha+1)\tau_J}(\sigma_s^1-\sigma_{i\tau_m}^1)dB_s^1\right.$$

$$\left.+ \sum_{\alpha=N+1}^{k}\sum_{p=0}^{(L_j-1)\wedge(k-\alpha)} h_{j,p}\pi_1^\alpha\left(\Delta_{k-p-\alpha}X^1-\sigma_{i\tau_m}^1\Delta_{k-p-\alpha}B^1\right)\right\},$$

hence it holds that

$$\left\|Z_k^1-\sigma_{i\tau_m}^1\zeta_k^1\right\|_r \lesssim \sqrt{\tau_J}\left\|\sup_{s\in[i\tau_m,(i+1)\tau_m+(L_j+N+1)\tau_J)}\left|\sigma_s^1-\sigma_{i\tau_m}^1\right|\right\|_r + \sqrt{\tau_J}\pi_1^N$$

$$\lesssim \sqrt{\tau_J}\left((\tau_m + (L_j + N + 1)\tau_J)^\gamma + \pi_1^N\right)$$

for any $r \geq 1$ by the Minkowski and Burkholder-Davis-Gundy inequalities as well as (35). Hence we obtain

$$P\left(\max_{l \in \mathcal{L}_J^+}|\mathbb{III}_J(l)| > \varepsilon\right) \lesssim \sum_{l \in \mathcal{L}_J^+}\left\{\sum_{i=0}^{M_J-1}\sum_{k \in I_m(i)}\left\|Z_k^1 - \sigma_{i\tau_m}^1\zeta_k^1\right\|_{2r}\left\|Z_{k+l}^2\right\|_{2r}\right\}^r$$

$$\lesssim \tau_J^{-1}\left\{M_J \cdot 2^{J-m} \cdot \tau_J\left((\tau_m + (L_j + N + 1)\tau_J)^\gamma + \pi_1^N\right)\right\}^r$$

$$= O\left(\tau_J^{-1}\left\{(\tau_m + (L_j + N + 1)\tau_J)^\gamma + \pi_1^N\right\}^r\right)$$

for any $\varepsilon > 0$. We can take large enough $r \geq 1$ such that $\tau_J^{-1}\left\{(\tau_m + (L_j + N + 1)\tau_J)^\gamma + \pi_1^N\right\}^r \to 0$, hence we obtain $\max_{l \in \mathcal{L}_J^+}|\mathbb{III}_J(l)| \to^p 0$.

We can prove $\max_{l \in \mathcal{L}_J^+}|\mathbb{IV}_J(l)| \to^p 0$ in an analogous manner.

Now we prove $\max_{l \in \mathcal{L}_J^+}|\mathbb{V}_J(l)| \to^p 0$. Let $\Sigma_{m,l}(i)$ be the covariance matrix of $((\zeta_k^1)_{k \in I_m(i)}, (\zeta_{k+l}^2)_{k \in I_m(i)})^\top$ for every i, and set $C_{m,l}(i) = \Sigma_{m,l}(i)^{1/2}A_m\Sigma_{m,l}(i)^{1/2}$, where

$$A_m = \begin{pmatrix} 0 & E_{2^{J-m}} \\ E_{2^{J-m}} & 0 \end{pmatrix}.$$

We first prove the following equations:

$$\begin{aligned}
\max_{l \in \mathcal{L}_J^+}\max_{0 \leq i \leq M_J-1}\|C_{m,l}(i)\|_{\mathrm{sp}} &= o\left(\tau_J L_j^{3/2}|\log \tau_J|^2\right), \\
\max_{l \in \mathcal{L}_J^+}\max_{0 \leq i \leq M_J-1}\|C_{m,l}(i)\|_F^2 &= o\left(\tau_J \tau_m(L_j|\log \tau_J|)^2\right),
\end{aligned} \tag{38}$$

where $\|\cdot\|_{\mathrm{sp}}$ denotes the spectral norm of matrices. For any $l \in \mathcal{L}_J^+$ we have

$$\|C_{m,l}(i)\|_F^2 \leq \|\Sigma_{m,l}(i)\|_F^2 = \sum_{k,k' \in I_m(i)}E\left[\zeta_k^1\zeta_{k'}^1\right]^2 + \sum_{k,k' \in I_m(i)}E\left[\zeta_{k+l}^2\zeta_{k'+l}^2\right]^2 + 2\sum_{k,k' \in I_m(i)}E\left[\zeta_k^1\zeta_{k'+l}^2\right]^2$$

by Appendix II(ii)–(iii) from [14], while we have

$$\|C_{m,l}(i)\|_{\mathrm{sp}} \leq \|\Sigma_{m,l}(i)\|_{\mathrm{sp}}\|A_m\|_{\mathrm{sp}}$$

$$\leq \max\left\{\max_{k \in I_m(i)}\sum_{k' \in I_m(i)}\left|E\left[\zeta_k^1\zeta_{k'}^1\right]\right|, \max_{k \in I_m(i)}\sum_{k' \in I_m(i)}\left|E\left[\zeta_{k+l}^2\zeta_{k'+l}^2\right]\right|\right\} + \max_{k \in I_m(i)}\sum_{k' \in I_m(i)}\left|E\left[\zeta_k^1\zeta_{k'+l}^2\right]\right|$$

by Corollary 4.5.11 and Theorem 5.6.9 from [26]. Now Lemma 3(a) implies that

$$\max_{l \in \mathcal{L}_J^+}\max_{0 \leq i \leq M_J-1}\left\{\sum_{k,k' \in I_m(i)}E\left[\zeta_k^1\zeta_{k'}^1\right]^2 + \sum_{k,k' \in I_m(i)}E\left[\zeta_{k+l}^2\zeta_{k'+l}^2\right]^2\right\} \lesssim \tau_J^2 \# I_m(i)L_j^2 = O\left(\tau_J\tau_m L_j^2\right),$$

and

$$\max_{l \in \mathcal{L}_J^+}\max_{0 \leq i \leq M_J-1}\max\left\{\max_{k \in I_m(i)}\sum_{k' \in I_m(i)}\left|E\left[\zeta_k^1\zeta_{k'}^1\right]\right|, \max_{k \in I_m(i)}\sum_{k' \in I_m(i)}\left|E\left[\zeta_{k+l}^2\zeta_{k'+l}^2\right]\right|\right\} \lesssim \tau_J L_j^{3/2}.$$

Moreover, by Lemma 3(b) we have

$$\sum_{k,k' \in I_m(i)}E\left[\zeta_k^1\zeta_{k'+l}^2\right]^2 = \tau_J^2\sum_{k,k' \in I_m(i)}\bar{\rho}_j((k' + l - k)\tau_J)^2 + o(\tau_J\tau_m L_j^2)$$

and

$$\max_{k} \sum_{k' \in I_m(i)} \left| E\left[\zeta_k^1 \zeta_{k'+l}^2\right] \right| = \max_{k} \tau_J \sum_{k' \in I_m(i)} \left| \bar{\rho}_j((k'+l-k)\tau_J) \right| + o(\tau_J L_j^{3/2})$$

uniformly in $l \in \mathcal{L}_J^+$ and $i = 0, 1, \ldots, M_J - 1$. Since it holds that

$$\sum_{k,k' \in I_m(i)} \bar{\rho}_j((k'+l-k)\tau_J)^2 = \sum_{k,k' \in I_m(0)} \bar{\rho}_j((k'+l-k)\tau_J)^2,$$

$$\max_{k \in I_m(i)} \sum_{k' \in I_m(i)} \left| \bar{\rho}_j((k'+l-k)\tau_J) \right| = \max_{k \in I_m(0)} \sum_{k' \in I_m(0)} \left| \bar{\rho}_j((k'+l-k)\tau_J) \right|,$$

the proof of (38) is completed once we show that

$$\tau_J^2 \max_{l \in \mathcal{L}_J^+} \sum_{k,k' \in I_m(0)} \bar{\rho}_j((k'+l-k)\tau_J)^2 = o(\tau_J \tau_m (L_j |\log \tau_J|)^2), \tag{39}$$

$$\tau_J \max_{l \in \mathcal{L}_J^+} \max_{k \in I_m(0)} \sum_{k' \in I_m(0)} \left| \bar{\rho}_j((k'+l-k)\tau_J) \right| = o(\tau_J L_j^{3/2} |\log \tau_J|^2). \tag{40}$$

By Lemma 5 we have

$$\tau_J^2 \max_{l \in \mathcal{L}_J^+} \sum_{k,k' \in I_m(0)} \bar{\rho}_j((k'+l-k)\tau_J)^2$$

$$\lesssim \tau_J \tau_m + \tau_J^2 \max_{l \in \mathcal{L}_J^+} \sum_{i_1,i_2=1}^{J+1} \sum_{\substack{k,k' \in I_m(0) \\ \theta_{i_1}, \theta_{i_2} \neq (k'+l-k)\tau_J}} \prod_{r=1}^{2} \frac{c_L}{\left| k'+l-k - \theta_{i_r} \tau_J^{-1} \right|}$$

$$\lesssim \tau_J \tau_m + \tau_J \tau_m (J+1)^2 c_L^2 = O(\tau_J \tau_m L |\log \tau_J|^2),$$

hence we obtain (39). Analogously we can prove (40). Hence we complete the proof of (38).

Now for any $\varepsilon > 0$ we have

$$P\left(\max_{l \in \mathcal{L}_J^+} |\mathbb{V}_J(l)| > \varepsilon \right) \lesssim \sum_{l \in \mathcal{L}_J^+} \sum_{i=0}^{M_J-1} P\left(\left| \sum_{k \in I_m(i)} \left(\zeta_k^1 \zeta_{k+l}^2 - E\left[\zeta_k^1 \zeta_{k+l}^2\right] \right) \right| > \frac{\varepsilon}{M_J K^2} \right)$$

with some constant $K > 0$. The Markov inequality yields

$$P\left(\left| \sum_{k \in I_m(i)} \left(\zeta_k^1 \zeta_{k+l}^2 - E\left[\zeta_k^1 \zeta_{k+l}^2\right] \right) \right| > \frac{\varepsilon}{M_J K^2} \right)$$

$$\leq \exp\left(-\frac{\varepsilon u_J}{M_J K^2} \right) \sum_{\varsigma \in \{-1,1\}} E\left[\exp\left(\varsigma u_J \sum_{k \in I_m(i)} \left(\zeta_k^1 \zeta_{k+l}^2 - E\left[\zeta_k^1 \zeta_{k+l}^2\right] \right) \right) \right],$$

where $u_J = (\tau_J L_j^{3/2} |\log \tau_J|^2)^{-1} \wedge (\sqrt{\tau_J \tau_m} L_j |\log \tau_J|)^{-1}$. Now, from the discussion in Section 3.2.1 of [12], we have

$$\log E\left[\exp\left(\varsigma u_J \sum_{k \in I_m(i)} \left(\zeta_k^1 \zeta_{k+l}^2 - E\left[\zeta_k^1 \zeta_{k+l}^2\right] \right) \right) \right] = -\frac{1}{2} \log \det(E - 2\varsigma u_J C_{m,l}(i)) - \mathrm{tr}[\varsigma u_J C_{m,l}(i)]$$

for sufficiently large J uniformly in $l \in \mathcal{L}_J^+$ and $i = 0, 1, \ldots, M_J - 1$ due to the first equation of (38). Therefore, by Appendix II-(v) from [14] we obtain

$$\log E\left[\exp\left(\varsigma u_J \sum_{k \in I_m(i)} \left(\zeta_k^1 \zeta_{k+l}^2 - E\left[\zeta_k^1 \zeta_{k+l}^2\right] \right) \right) \right] \leq \frac{u_J^2}{2} \|C_{m,l}(i))\|_F^2 + \frac{u_J^3}{3} \frac{\|C_{m,l}(i))\|_{\mathrm{sp}} \|C_{m,l}(i))\|_F^2}{(1 - \|C_{m,l}(i))\|_{\mathrm{sp}})^3}$$

for sufficiently large J uniformly in $l \in \mathcal{L}_J^+$ and $i = 0, 1, \ldots, M_J - 1$ due to the first equation of (38). Consequently, (38) implies that

$$\sum_{l \in \mathcal{L}_J} \sum_{i=0}^{M_J - 1} P\left(\left|\sum_{k \in I_m(i)} (\zeta_k^1 \zeta_{k+l}^2 - E\left[\zeta_k^1 \zeta_{k+l}^2\right])\right| > \frac{\varepsilon}{M_J K^2}\right) \lesssim \tau_J^{-1} M_J \exp\left(-\frac{\varepsilon u_J}{M_J K^2}\right).$$

Since we have $u_J / M_J = (\tau_J \tau_m^{-1} L_j^{3/2} |\log \tau_J|^2)^{-1} \wedge (\tau_m^{-1/2} \sqrt{\tau_J} L_j |\log \tau_J|)^{-1}$, it holds that $\tau_J^a u_J / M_J \to \infty$ as $J \to \infty$ for some $a > 0$ because $m < \{(1 - \kappa) \wedge \frac{1}{4}\} J$. So we obtain $\limsup_{J \to \infty} P\left(\max_{l \in \mathcal{L}_J} |\mathbb{V}_J(l)| > \varepsilon\right) = 0$. This is the desired result.

Finally, by Lemmas 3(b) and 6 we have $\max_{l \in \mathcal{L}_J^+} |\mathbb{VII}_J(l)| = o_p(M_J \cdot \tau_J^{-1} \tau_m \cdot \tau_J)$. Since $M_J = O(\tau_m^{-1})$, we obtain $\max_{l \in \mathcal{L}_J^+} |\mathbb{VII}_J(l)| \to^p 0$. This completes the proof of the lemma. $\square$

Lemma 9. $\int_{\Lambda_{-j}} D(\lambda)\Pi(\lambda) e^{\sqrt{-1}b\lambda} d\lambda \neq 0$ for any $j \in \mathbb{N}$ and $b \in [-\frac{1}{2}, \frac{1}{2}]$.

Proof. Since we have $\int_{\Lambda_{-j}} D(\lambda)\Pi(\lambda) e^{\sqrt{-1}b\lambda} d\lambda = 2\int_{\pi/2^j}^{\pi/2^{j-1}} D(\lambda)\Re\left[\Pi(\lambda) e^{\sqrt{-1}b\lambda}\right] d\lambda$ and $D(\lambda) > 0$ for any $\lambda \in \mathbb{R}$, it is enough to prove $\Re\left[\Pi(\lambda) e^{\sqrt{-1}b\lambda}\right] > 0$ for any $\lambda \in (0, \pi)$. We have

$$\Re\left[\Pi(\lambda) e^{\sqrt{-1}b\lambda}\right] = \frac{(1 - \pi_1)(1 - \pi_2)}{|(1 - \pi_1 e^{\sqrt{-1}\lambda})(1 - \pi_2 e^{-\sqrt{-1}\lambda})|^2} \mathfrak{C},$$

where $\mathfrak{C} = (1 + \pi_1 \pi_2) \cos b\lambda - \pi_1 \cos(b - 1)\lambda - \pi_2 \cos(b + 1)\lambda$. Hence it suffices to prove $\mathfrak{C} > 0$. Since we have $\mathfrak{C} = (1 + \pi_1 \pi_2) \cos(-b\lambda) - \pi_1 \cos((-b + 1)\lambda) - \pi_2 \cos((-b - 1)\lambda)$, by symmetry we may assume $b \geq 0$.

First we note that $\cos b\lambda > 0$ because $b\lambda \in [0, \frac{\pi}{2})$. Next, we can rewrite $\mathfrak{C}$ as

$$\mathfrak{C} = (1 - \pi_1)(1 - \pi_2) \cos b\lambda + \pi_1(\cos(-b)\lambda - \cos(b - 1)\lambda) + \pi_2(\cos b\lambda - \cos(b + 1)\lambda).$$

Since $-\pi \leq (b - 1)\lambda \leq (-b)\lambda \leq 0$ due to $0 \leq b \leq \frac{1}{2}$, we have $\cos(-b)\lambda - \cos(b - 1)\lambda \geq 0$ because $\cos$ is increasing on $[-\pi, 0]$. Also, if $\lambda \geq \frac{\pi}{2}$, we have $\frac{\pi}{2} \leq (b + 1)\lambda \leq \frac{3}{2}\pi$, hence $\cos(b + 1)\lambda \leq 0$. So $\cos b\lambda - \cos(b + 1)\lambda \geq 0$. Otherwise, we have $0 \leq b\lambda \leq (b + 1)\lambda \leq \frac{3}{4}\pi$, hence we have $\cos b\lambda - \cos(b + 1)\lambda \geq 0$ because $\cos$ is decreasing on $[0, \pi]$. Consequently, we have $\mathfrak{C} \geq (1 - \pi_1)(1 - \pi_2) \cos b\lambda > 0$. This completes the proof. $\square$

Proof of Theorem 2. Suppose that there is a number $\varepsilon > 0$ such that $P(v_J^{-1}|\widehat{\theta}_j - \theta_j| > \varepsilon)$ does not converge to 0 as $J \to \infty$. Then there is a sequence $(J_m)_{m \geq 1}$ of positive integers such that $J_m \uparrow \infty$ as $m \to \infty$ and $P(v_{J_m}^{-1}|\widehat{\theta}_j - \theta_j| > \varepsilon) \to a$ as $m \to \infty$ for some $a > 0$. Moreover, for every m we can take an integer $l_m \in \mathcal{L}_{J_m}$ such that $|l_m \tau_{J_m} - \theta_j| \leq \tau_{J_m}/2$. In particular, the sequence $(\tau_{J_m}^{-1}(l_m \tau_{J_m} - \theta_j))_{m \geq 1}$ has a converging subsequence. Without loss of generality we may assume that $\tau_{J_m}^{-1}(l_m \tau_{J_m} - \theta_j) \to b$ as $m \to \infty$ for some $b \in [-\frac{1}{2}, \frac{1}{2}]$. Now since $\left|\widehat{\rho}_{(j)}(\widehat{\theta}_j)\right| > \max_{\theta \in \mathcal{G}_J : v_J^{-1}|\theta - \theta_j| > \varepsilon} \left|\widehat{\rho}_{(j)}(\theta)\right|$ implies that $v_J^{-1}|\widehat{\theta}_j - \theta_j| \leq \varepsilon$, we have

$$P\left(v_{J_m}^{-1}\left|\widehat{\theta}_j - \theta_j\right| > \varepsilon\right) \leq P\left(\left|\widehat{\rho}_{(j)}(\widehat{\theta}_j)\right| \leq \max_{\theta \in \mathcal{G}_{J_m} : v_{J_m}^{-1}|\theta - \theta_j| > \varepsilon} \left|\widehat{\rho}_{(j)}(\theta)\right|\right)$$

$$\leq P\left(\left|\widehat{\rho}_{(j)}(l_m \tau_{J_m})\right| \leq \max_{\theta \in \mathcal{G}_{J_m} : v_{J_m}^{-1}|\theta - \theta_j| > \varepsilon} \left|\widehat{\rho}_{(j)}(\theta)\right|\right)$$

$$\leq P\left(\left|\widehat{\rho}_{(j)}(l_m \tau_{J_m})\right| \leq \frac{|\mathfrak{r}|}{2}\right) + \leq P\left(\frac{|\mathfrak{r}|}{2} < \max_{\theta \in \mathcal{G}_{J_m} : v_{J_m}^{-1}|\theta - \theta_j| > \varepsilon} \left|\widehat{\rho}_{(j)}(\theta)\right|\right),$$

where $\mathfrak{r} = \Sigma_t(\theta_j) R_j \int_{\Lambda_{-j}} D(\lambda)\Pi(\lambda) e^{\sqrt{-1}b\lambda} d\lambda$. $\widehat{\rho}_{(j)}(l_m \tau_{J_m}) \to^p \mathfrak{r}$ as $m \to \infty$ by Lemma 7. Moreover, $\mathfrak{r} \neq 0$ because of Lemma 9 and assumption. Therefore, by Lemma 8 we obtain $\limsup_{m \to \infty} P\left(v_{J_m}^{-1}\left|\widehat{\theta}_j - \theta_j\right| > \varepsilon\right) = 0$. This contradicts $\lim_{m \to \infty} P(v_{J_m}^{-1}|\widehat{\theta}_j - \theta_j| > \varepsilon) = a > 0$. $\square$

Acknowledgments

References

[1] Aït-Sahalia, Y. and Jacod, J. (2014). *High-frequency financial econometrics*. Princeton University Press.

[2] Alsayed, H. and McGroarty, F. (2014). Ultra-high-frequency algorithmic arbitrage across international index futures. *J. Forecast.* **33**, 391–408.

[3] Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Process. Appl.* **123**, 2475–2499.

[4] Barndorff-Nielsen, O. E., Corcuera, J. M. and Podolskij, M. (2011). Multipower variation for Brownian semistationary processes. *Bernoulli* **17**, 1159–1194.

[5] Barndorff-Nielsen, O. E., Graversen, S. E. G., Jacod, J., Podolskij, M. and Shephard, N. (2006). A central limit theorem for realised power and bipower variations of continuous semimartingales. In Y. Kabanov, R. Liptser and J. Stoyanov, eds., *From stochastic calculus to mathematical finance*, the Shiryaev Festschrift. Springer Verlag, Berlin, pp. 33–69.

[6] Bollen, N. P., O'Neill, M. J. and Whaley, R. E. (2016). Tail wags dog: Intraday price discovery in VIX markets. *Journal of Futures Markets (forthcoming)* DOI: 10.1002/fut.21805.

[7] Cardinali, A. (2009). A generalized multiscale analysis of the predictive content of Eurodollar implied volatilities. *International Journal of Theoretical and Applied Finance* **12**, 1–18.

[8] Ceron, A., Curini, L. and Iacus, S. M. (2016). First- and second-level agenda setting in the Twittersphere: An application to the Italian political debate. *Journal of Information Technology & Politics* **13**, 159–174.

[9] Chan, G. and Wood, A. T. A. (1999). Simulation of stationary Gaussian vector fields. *Statist. Comput.* **9**, 265–268.

[10] Da Fonseca, J. and Zaatour, R. (2016). Correlation and lead-lag relationships in a Hawkes microstructure model. Working paper. Available at SSRN: http://ssrn.com/abstract=2506845.

[11] Dajčman, S. (2013). Interdependence between some major European stock markets — a wavelet lead/lag analysis. *Prague Economic Papers* **1**, 28–49.

[12] 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.

[13] Daubechies, I. (1992). *Ten lectures on wavelets*. SIAM.

[14] Davies, R. B. (1973). Asymptotic inference in stationary Gaussian time-series. *Adv. in Appl. Probab.* **5**, 469–497.

[15] de Jong, F. and Nijman, T. (1997). High frequency analysis of lead-lag relationships between financial markets. *Journal of Empirical Finance* **4**, 259–277.

[16] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84**, 313–342.

[17] Duffie, D. (1991). The theory of value in security markets. In W. Hildenbrand and H. Sonnenschein, eds., *Handbook of mathematical economics*, vol. IV, chap. 31. Elsevier Science Publishers B.V., pp. 1615–1682.

[18] Fernández-Macho, J. (2012). Wavelet multiple correlation and cross-correlation: A multiscale analysis of Eurozone stock markets. *Phys. A* **391**, 1097–1104.

[19] Gallegati, M. (2008). Wavelet analysis of stock returns and aggregate economic activity. *Comput. Statist. Data Anal.* **52**, 3061–3074.

[20] Gatheral, J., Jaisson, T. and Rosenbaum, M. (2014). Volatility is rough. Working paper, Available at arXiv: http://arxiv.org/abs/1410.3394.

[21] Gençay, R., Selçuk, F. and Whitcher, B. (2002). *An introduction to wavelets and other filtering methods in finance and economics*. Academic Press.

[22] Gikhman, I. I. and Skorokhod, A. V. (1969). *Introduction to the theory of random processes*. W. B. Saunders Company.

[23] Grebenkov, D. S., Belyaev, D. and Jones, P. W. (2016). A multiscale guide to Brownian motion. *J. Phys. A* **49**, 1–31.

[24] Hafner, C. M. (2012). Cross-correlating wavelet coefficients with applications to high-frequency financial time series. *J. Appl. Stat.* **39**, 1363–1379.

[25] Hoffmann, M., Rosenbaum, M. and Yoshida, N. (2013). Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli* **19**, 426–461.

[26] Horn, R. A. and Johnson, C. R. (2013). *Matrix analysis*. Cambridge University Press, 2nd edn.

[27] Huth, N. and Abergel, F. (2014). High frequency lead/lag relationships — empirical facts. *Journal of Empirical Finance* **26**, 41–58.

[28] Jacod, J. and Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*. Springer, 2nd edn.

[29] Kawaller, I. G., Koch, P. D. and Koch, T. W. (1987). The temporal price relationship between S&P 500 futures and the S&P 500 index. *Journal of Finance* **42**, 1309–1329.

[30] Lai, M.-J. (1995). On the digital filter associated with Daubechies' wavelets. *IEEE Trans. Signal Process.* **43**, 2203–2205.

[31] Lo, A. W. and MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *J. Econometrics* **45**, 181–211.

[32] Mallat, S. (2009). *A wavelet tour of signal processing*. Elsevier Inc., 3rd edn.

[33] Medvegyev, P. (2007). *Stochastic integration theory*. Oxford University Press.

[34] 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.

[35] Percival, D. B. and Mofjeld, H. O. (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *J. Amer. Statist. Assoc.* **92**, 868–880.

[36] Ranta, M. (2010). *Wavelet multiresolution analysis of financial time series*. Ph.D. thesis, University of Vaasa.

[37] Robert, C. Y. and Rosenbaum, M. (2010). On the limiting spectral distribution of the covariance matrices of time-lagged processes. *J. Multivariate Anal.* **101**, 2434–2451.

[38] Rudin, W. (1987). *Real and complex analysis*. McGraw-Hill, Inc., 3rd edn.

[39] Rudin, W. (1991). *Functional analysis*. McGraw-Hill, Inc., 2nd edn.

[40] Serroukh, A. and Walden, A. (2000). Wavelet scale analysis of bivariate time series i: motivation and estimation. *J. Nonparametr. Stat.* **13**, 1–36.

[41] Serroukh, A. and Walden, A. (2000). Wavelet scale analysis of bivariate time series ii: statistical properties for linear processes. *J. Nonparametr. Stat.* **13**, 37–56.

[42] Stoll, H. R. and Whaley, R. E. (1990). The dynamics of stock index and stock index futures returns. *Journal of Financial and Quantitative Analysis* **25**, 441–468.

[43] Vidakovic, B. (1999). *Statistical modeling by wavelets*. John Wiley & Sons, Inc.

[44] Whitcher, B., Guttorp, P. and Percival, D. B. (1999). Mathematical background for wavelet estimators of cross-covariance and cross-correlation. Technical report 038, NRCSE.

[45] Whitcher, B., Guttorp, P. and Percival, D. B. (2000). Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research* **105**, 14941–14962.