*The Annals of Statistics*
2019, Vol. 47, No. 3, 1663–1687
https://doi.org/10.1214/18-AOS1731


# GAUSSIAN APPROXIMATION OF MAXIMA OF WIENER FUNCTIONALS AND ITS APPLICATION TO HIGH-FREQUENCY DATA[1]

BY YUTA KOIKE

*University of Tokyo and JST CREST*

This paper establishes an upper bound for the Kolmogorov distance between the maximum of a high-dimensional vector of smooth Wiener functionals and the maximum of a Gaussian random vector. As a special case, we show that the maximum of multiple Wiener–Itô integrals with common orders is well approximated by its Gaussian analog in terms of the Kolmogorov distance if their covariance matrices are close to each other and the maximum of the fourth cumulants of the multiple Wiener–Itô integrals is close to zero. This may be viewed as a new kind of fourth moment phenomenon, which has attracted considerable attention in the recent studies of probability. This type of Gaussian approximation result has many potential applications to statistics. To illustrate this point, we present two statistical applications in high-frequency financial econometrics: One is the hypothesis testing problem for the absence of lead-lag effects and the other is the construction of uniform confidence bands for spot volatility.

**1. Introduction.** This study is originally motivated by the problem of testing whether there exists a (possibly) time-lagged correlation between two Brownian motions based on their high-frequency observation data. Roughly speaking, the setting considered here is described as follows. We discretely observe the following two continuous-time processes on the interval $[0, T]$:

$$X_t^1 = x_0^1 + \sigma_1 B_t^1, \qquad X_t^2 = x_0^2 + \sigma_2 B_{t-\vartheta}^2, \qquad t \in [0, T], \tag{1.1}$$

where $x_0^1, x_0^2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $B_t = (B_t^1, B_t^2)$ $(t \in \mathbb{R})$ is a bivariate two-sided Brownian motion with correlation $\rho \in (-1, 1)$ and $\vartheta \in \mathbb{R}$. For each $\nu = 1, 2$, the process $X^\nu$ is observed at the time points $0 \le t_0^\nu < t_1^\nu < \cdots < t_{n_\nu}^\nu \le T$, hence the observation times are possibly nonsynchronous. Based on the observation data $(X^1_{t_i^1})_{i=0}^{n_1}$ and $(X^2_{t_j^2})_{j=0}^{n_2}$, we aim at solving the following statistical hypothesis testing problem:

$$H_0 : \rho = 0 \quad \text{vs} \quad H_1 : \rho \neq 0. \tag{1.2}$$

Received September 2017; revised March 2018.

[1]Supported by JST CREST and JSPS KAKENHI Grant Numbers JP16K17105, JP17H01100.

*MSC2010 subject classifications.* 60G15, 60H07, 62E17, 62G20.

*Key words and phrases.* Bootstrap, fourth moment phenomenon, Malliavin calculus, maximum, Stein's method, uniform confidence bands.

Model (1.1) was introduced in Hoffmann, Rosenbaum and Yoshida [24] (as a more general one) to model lead-lag effects in high-frequency financial data (see also [47]). In [24], the problem of estimating the time-lag parameter $\vartheta$ is considered. To estimate $\vartheta$, Hoffmann, Rosenbaum and Yoshida [24] have introduced the following contrast function:

$$U_n(\theta) = \sum_{i,j} (X^1_{t^1_i} - X^1_{t^1_{i-1}})(X^2_{t^2_j} - X^2_{t^2_{j-1}}) 1_{\{(t^1_{i-1}, t^1_i] \cap (t^2_{j-1} - \theta, t^2_j - \theta] \neq \varnothing\}}.$$

$U_n(\theta)$ could be considered as the (sample) cross-covariance function between the returns of $X^1$ and $X^2$ at the lag $\theta$ computed by Hayashi and Yoshida's [23] method. Hoffmann, Rosenbaum and Yoshida [24] have shown that

$$\widehat{\vartheta}_n = \arg\max_{\theta \in \mathcal{G}_n} |U_n(\theta)|$$

is a consistent estimator for $\vartheta$ under some regularity conditions while one appropriately takes the finite set $\mathcal{G}_n$ as long as $\rho \neq 0$. The condition $\rho \neq 0$ is necessary because it is clearly impossible to identify the parameter $\vartheta$ if $\rho = 0$. Therefore, unless we can believe $\rho \neq 0$ due to some external information, we need to reject the null hypothesis in the above testing problem before we carry out estimation of $\vartheta$. A natural approach to solve testing problem (1.2) is to reject the null hypothesis if the value of $\max_{\theta \in \mathcal{G}_n} |U_n(\theta)|$ is *too* large. To implement this idea precisely, we need to derive or approximate the distribution of $\max_{\theta \in \mathcal{G}_n} |U_n(\theta)|$ under the null hypothesis $H_0$. One main purpose of this paper is to give an answer to this problem. More generally, we consider the problem of approximating the distributions of maximum-type statistics appearing in high-frequency financial econometrics. Indeed, we encounter such statistics in many problems of high-frequency financial econometrics, for example, construction of uniform confidence bands for spot volatility and other time-varying characteristics, family-wise error rate control for testing at many time points (cf. [3, 16]), change point analysis of volatility (cf. [5]), testing the absence of jumps (cf. [33, 46]) and so on.

From a mathematical point of view, this paper is built on two recent studies developed in different areas. The first one is the seminal work of Chernozhukov, Chetverikov and Kato [9, 11, 13, 14] which we call the *Chernozhukov–Chetverikov–Kato theory*, or the *CCK theory* for short. One main conclusion from the CCK theory is a bound for the Kolmogorov distance between the distributions of the maximum of a (high-dimensional) random vector and that of a Gaussian vector, which has an apparent connection to our purpose. However, their result is not directly applicable to our problem because their target random vector is a sum of independent random vectors [9, 11, 14] or Gaussian [13, 14]. In fact, one of our main target random vectors, $(U_n(\theta))_{\theta \in \mathcal{G}_n}$, is a sum of dependent random vectors even under the null hypothesis where the dependence is caused by the nonsynchronicity of the observation times. Although there are several extensions of the CCK theory to a sum of dependent random vectors (see, e.g., [7, 8, 12, 50, 51]),

it still seems difficult to apply such a result to our problem because the nonsynchronicity causes a quite complex, "nonstationary," dependence structure. In this aspect, this paper aims at extending the CCK theory suitably to our purpose, and our results indeed generalize several results of [13]. In particular, our results do not require that the target random vector should be written as a sum of random vectors and give a simpler bound than those of the preceding studies listed above.

It turns out that in the CCK theory the independence/Gaussianity assumption on the target vector is crucial for the application of *Stein's method*.[2] In other words, we can naturally extend the CCK theory to a case without independence as long as Stein's method is effectively applicable. This viewpoint leads us to using another important theory for this work, *Malliavin calculus*, in our problem. In fact, starting from the seminal work of Nourdin and Peccati [38], the recent studies show that "Stein's method and Malliavin calculus fit together admirably well" (page 3 of Nourdin [37]). This paper shows that this statement continues to hold true in the application to the CCK theory. Our application of Malliavin calculus is based on a multivariate extension of the ideas from [38], which is established in [43] (see also [39]). We refer to the monograph [40] for more information about this subject.

After developing the main Gaussian approximation results, we turn to the original problem of statistical applications in high-frequency data. In this paper, we demonstrate two applications: One is testing the absence of lead-lag effects and the other is constructing uniform confidence bands for spot volatility. We have already explained the background of the former problem in the above, so we briefly discuss the latter one. Estimation of spot volatility is one of major topics in high-frequency financial econometrics (see Chapter 8 of [1] and references therein). There are quite a few articles concerning construction of *pointwise* confidence bands for spot volatility; see, for example, [2, 31, 34, 36]. In contrast, only a few results are available on the behavior of *uniform* errors in spot volatility estimation: Kristensen [31] and [27] give uniform convergence rates for kernel-type spot volatility estimators, while Fan and Wang [20] consider a Gumbel approximation for the distribution of uniform errors of kernel-type spot volatility estimators. Besides, Sabel [48] implements multiscale inference for spot volatility via KMT construction. This paper contributes this relatively undeveloped areas by providing a new approach to construct uniform confidence bands for spot volatility in the spirit of the CCK theory: Construction of uniform confidence bands is a typical application of the CCK theory; cf. [10, 28, 29].

In the first application, the Gaussian approximation itself is still statistically infeasible because the covariance structure of the objective statistics is unknown. For this reason we also develop a wild (or multiplier) bootstrap procedure to approximate the quantiles of the error distribution of the test statistic, which is the

[2]The independence assumption also plays a role in deriving maximal moment inequalities, but this issue may be considered separately.

approach taken in the CCK theory. The Gaussian approximation result serves as validating such a bootstrap procedure.

The remainder of this paper is organized as follows. Section 2 presents the main Gaussian approximation results obtained in this study. In Section 3, we derive Gaussian approximation results for maxima of random symmetric quadratic forms as an application of the main results. We present two statistical applications of our results in high-frequency financial econometrics in Section 4. We especially propose a testing procedure for (1.2). The finite sample performance of this testing procedure is illustrated in Section 5. We put most technical parts of the paper in the Supplementary Material [30]: Appendix A collects the preliminary definitions and results used in Appendix B, which contains proofs of all the results presented in the main text of the paper.

**Notation.** Throughout the paper, $\mathfrak{C} = (\mathfrak{C}(i,j))_{1\le i,j\le d}$ denotes a $d\times d$ nonnegative definite symmetric matrix, and $Z = (Z_1,\dots,Z_d)^\top$ denotes a $d$-dimensional centered Gaussian random vector with covariance matrix $\mathfrak{C}$. For a vector $x = (x_1,\dots,x_d)^\top \in \mathbb{R}^d$, we write $x_\vee = \max_{1\le j\le d} x_j$. For any $\varepsilon > 0$ and any subset $A$ of $\mathbb{R}$, we write $A^\varepsilon = \{x\in\mathbb{R} : |x-y|\le\varepsilon \text{ for some } y\in A\}$. For a real-valued function $f$ defined on an interval $I\subset\mathbb{R}$ and $\eta>0$, we write $\|f\|_\infty = \sup\{|f(x)| : x\in I\}$ and $w(f;\eta) = \sup\{|f(s)-f(t)| : s,t\in I, |s-t|\le\eta\}$. For a random variable $\xi$ and $p\ge 1$, we write $\|\xi\|_p = \{E[|\xi|^p]\}^{1/p}$. For a matrix $A$, we denote by $\|A\|_{\mathrm{sp}}$ and $\|A\|_F$ its spectral norm and Frobenius norm, respectively.

Finally, we enumerate the notation from Malliavin calculus which are necessary to state our main results. We refer to [25, 40, 44] for a detailed description of Malliavin calculus. Also, see Section A.1 of Appendix A for a concise overview of the notions from Malliavin calculus used in this paper.

- Throughout the paper, $H$ denotes a real separable Hilbert space. The inner product and the norm of $H$ are denoted by $\langle\cdot,\cdot\rangle_H$ and $\|\cdot\|_H$, respectively.
- We assume that an isonormal Gaussian process $W = (W(h))_{h\in H}$ over $H$ defined on a probability space $(\Omega,\mathcal{F},P)$ is given. We denote by $L^2(W)$ the space $L^2(\Omega,\sigma(W),P)$ for short.
- For a nonnegative integer $q$, $H^{\otimes q}$ and $H^{\odot q}$ denote the $q$th tensor power and $q$th symmetric tensor power, respectively.
- For an element $f\in H^{\odot q}$, we denote by $I_q(f)$ the $q$th multiple Wiener–Itô integral of $f$.
- For any real number $p\ge 1$ and any integer $k\ge 1$, $\mathbb{D}_{k,p}$ denotes the stochastic Sobolev space of random variables which are $k$ times differentiable in the Malliavin sense and the derivatives up to order $k$ have finite moments of order $p$. If $F\in\mathbb{D}_{k,p}$, we denote by $D^kF$ the $k$th Malliavin derivative of $F$. We write $DF$ instead of $D^1F$ for short.
- $L$ denotes the Ornstein–Uhlenbeck operator. Also, the pseudo inverse of $L$ is denoted by $L^{-1}$.

**2. Main results.** Throughout this section, $F = (F_1, \dots, F_d)^\top$ denotes a $d$-dimensional random vector such that $F_j \in \mathbb{D}_{1,2}$ and $E[F_j] = 0$ for all $j = 1, \dots, d$. For each $\beta > 0$, we define the function $\Phi_\beta : \mathbb{R}^d \to \mathbb{R}$ by

$$\Phi_\beta(x) = \beta^{-1} \log\left( \sum_{j=1}^d e^{\beta x_j} \right) \qquad \bigl(x = (x_1, \dots, x_d)^\top \in \mathbb{R}^d\bigr).$$

Equation (1) from [13] states that

$$0 \le \Phi_\beta(x) - x_\vee \le \beta^{-1} \log d \tag{2.1}$$

for any $x \in \mathbb{R}^d$.

We first give a generalization of Theorem 1 from [13] as follows.

PROPOSITION 2.1. *Let $g : \mathbb{R} \to \mathbb{R}$ be a $C^2$ function with bounded first and second derivatives. Then, for any $\beta > 0$ we have*

$$\bigl|E\bigl[g(\Phi_\beta(F))\bigr] - E\bigl[g(\Phi_\beta(Z))\bigr]\bigr| \le \bigl(\|g''\|_\infty/2 + \beta\|g'\|_\infty\bigr)\Delta,$$

*where*

$$\Delta = E\Bigl[\max_{1 \le i,j \le d} \bigl|\mathfrak{C}(i,j) - \langle DF_i, -DL^{-1}F_j\rangle_H\bigr|\Bigr].$$

*In particular, it holds that*

$$\bigl|E\bigl[g(F_\vee)\bigr] - E\bigl[g(Z_\vee)\bigr]\bigr| \le \bigl(\|g''\|_\infty/2 + \beta\|g'\|_\infty\bigr)\Delta + 2\beta^{-1}\|g'\|_\infty \log d.$$

REMARK 2.1. We can indeed derive Theorem 1 of [13] from Proposition 2.1 in the following way. Suppose that the law of $F$ is the $d$-dimensional normal distribution with mean 0 and covariance matrix $\Sigma = (\Sigma(i,j))_{1 \le i,j \le d}$. Without loss of generality, we may assume that $F$ is expressed as $F = \Sigma^{1/2}G$ with $G$ being a $d$-dimensional standard Gaussian vector. Then we can define the isonormal Gaussian process $W$ over $H := \mathbb{R}^d$ by $W(h) = h^\top G$, $h \in H$ (cf. Example 2.1.3 of [40]), and we have $F_i = \sum_{j=1}^d \gamma_{ij} W(e_j)$ for every $i = 1, \dots, d$, where $\gamma_{ij}$ denotes the $(i,j)$th component of the matrix $\Sigma^{1/2}$ and $(e_1, \dots, e_d)$ denotes the canonical basis of $\mathbb{R}^d$. In this case, it holds that

$$\langle DF_i, -DL^{-1}F_j\rangle_H = \sum_{k,l=1}^d \gamma_{ik}\gamma_{jl}\langle e_k, e_l\rangle_H = \sum_{k=1}^d \gamma_{ik}\gamma_{jk} = \Sigma(i,j),$$

hence we obtain the conclusion of Theorem 1 from [13].

Proposition 2.1 and some elementary approximation arguments lead the following useful lemma.

LEMMA 2.1. *There is a universal constant $C > 0$ such that*

$$P(F_\vee \in A) \le P\big(Z_\vee \in A^{5\varepsilon}\big) + C\varepsilon^{-2}(\log d)\Delta$$

*for any Borel set $A$ of $\mathbb{R}$ and any $\varepsilon > 0$.*

REMARK 2.2. Lemma 2.1 is useful when we derive a Gaussian approximation for the supremum of statistics indexed by an infinite set (see Proposition 4.3 and its proof). In fact, Lemma 2.1 can be considered as a counterpart of Theorem 3.1 from [14], which is used to derive their Gaussian approximation results for suprema of empirical processes. An advantage of Lemma 2.1 over Theorem 3.1 from [14] is that the second term of the estimate is proportional to $\varepsilon^{-2}$ in Lemma 2.1, while it is proportional to $\varepsilon^{-3}$ in Theorem 3.1 from [14]. This difference generally leads a weaker condition and a better convergence rate in Gaussian approximation; see Remark 4.8 for details.

Combining Lemma 2.1 with several technical tools developed in the CCK theory, we obtain the following main result of this paper, which can be considered as a generalization of Theorem 2 from [13].

THEOREM 2.1. (a) *Suppose that $d \ge 2$ and there are constants $\underline{\sigma}, \overline{\sigma} > 0$ such that $\underline{\sigma}^2 \le \mathfrak{C}(j,j) \le \overline{\sigma}^2$ for all $j = 1, \dots, d$. Set $a_d = E[\max_{1\le j\le d}(Z_j/\sqrt{\mathfrak{C}(j,j)})]$. Then*

$$\sup_{x\in\mathbb{R}} \big|P(F_\vee \le x) - P(Z_\vee \le x)\big| \le C\Delta^{1/3}\big\{1 \vee a_d^2 \vee \log(1/\Delta)\big\}^{1/3}(\log d)^{1/3}, \tag{2.2}$$

*where $C > 0$ depends only on $\underline{\sigma}$ and $\overline{\sigma}$ (the right-hand side is understood to be 0 if $\Delta = 0$).*

(b) *Suppose that $d \ge 2$ and there is a constant $b > 0$ such that $\mathfrak{C}(j,j) \ge b$ for all $j = 1, \dots, d$. Then*

$$\sup_{x\in\mathbb{R}} \big|P(F_\vee \le x) - P(Z_\vee \le x)\big| \le C'\Delta^{1/3}(\log d)^{2/3}, \tag{2.3}$$

*where $C' > 0$ depends only on $b$.*

Since we have $\max_{1\le j\le d}|x_j| = \max\{\max_{1\le j\le d} x_j, \max_{1\le j\le d}(-x_j)\}$ for any real numbers $x_1, \dots, x_d$, we obtain the following result as a direct consequence of Theorem 2.1.

COROLLARY 2.1. *Under the assumptions of Theorem 2.1(b), we have*

$$\sup_{x\in\mathbb{R}} \Big|P\Big(\max_{1\le j\le d}|F_j| \le x\Big) - P\Big(\max_{1\le j\le d}|Z_j| \le x\Big)\Big| \le C'\Delta^{1/3}(\log d)^{2/3},$$

*where $C' > 0$ depends only on $b$.*

In order to make Theorem 2.1 and Corollary 2.1 useful, we need a reasonable bound for the quantity $\Delta$. When the random vector $F$ consists of multiple Wiener–Itô integrals with common order, we have the following useful bound for $\Delta$.

LEMMA 2.2. *Let $q \geq 2$ be an integer and suppose that $F_j = I_q(f_j)$ for some $f_j \in H^{\odot q}$ for $j = 1, \dots, d$. Then we have*

$$\Delta \leq \max_{1 \leq i,j \leq d} \bigl|\mathfrak{C}(i,j) - E[F_i F_j]\bigr| + C_q \log^{q-1}\bigl(2d^2 - 1 + e^{q-2}\bigr) \max_{1 \leq k \leq d} \sqrt{E\bigl[F_k^4\bigr] - 3E\bigl[F_k^2\bigr]^2},$$

*where $C_q > 0$ depends only on $q$.*

REMARK 2.3. Lemma 2.2 implies that, in order to bound the Kolmogorov distance between $F_\vee$ and $Z_\vee$, we only need to control the convergence rate of the covariance matrix of $F$ to that of $Z$ and the fourth cumulants of the components of $F$, as long as $F$ consists of multiple Wiener–Itô integrals with common order. This can be considered as a type of *fourth moment phenomenon*, which was first discovered by [45] while they derived central limit theorems for sequences of multiple Wiener–Itô integrals. For more information about the fourth moment phenomenon, we refer to [40] and references therein.

It is often involved to compute the variables $L^{-1}F_j$ in the case that $F_j$'s are general Wiener functionals. It would be worth mentioning that we can avoid this issue if the variables $F_j$ are twice differentiable in the Malliavin sense and satisfy a suitable moment condition. To state such a result precisely, we make some definitions: For an $H \otimes H$-valued random variable $G$, we denote by $\|G\|_{\mathrm{op}}$ the operator norm of the (random) operator $H \ni h \mapsto \langle h, G\rangle_H \in H$. Also, we say that a random variable $Y$ is *sub-Gaussian relative to the scale $a > 0$* if $E[e^{\lambda Y}] \leq \exp(\lambda^2 a^2/2)$ for all $\lambda \in \mathbb{R}$.

LEMMA 2.3. *If $F_1, \dots, F_d \in \mathbb{D}_{2,4p}$ for a positive integer $p$, we have*

$$\Delta \leq \max_{1 \leq i,j \leq d} \bigl|\mathfrak{C}(i,j) - E[F_i F_j]\bigr| + d^{1/p}\sqrt{2p-1} \cdot \frac{3}{2}\Bigl(\max_{1 \leq i \leq d} \bigl\|\|D^2 F_i\|_{op}\bigr\|_{4p}\Bigr)\Bigl(\max_{1 \leq j \leq d} \bigl\|\|DF_j\|_H\bigr\|_{4p}\Bigr).$$

*Moreover, if there is a constant $a > 0$ such that both the variables $\|D^2 F_i\|_{op}$ and $\|DF_i\|_H$ are sub-Gaussian relative to the scale $a$ for all $i = 1, \dots, d$, we have*

$$\Delta \leq \max_{1 \leq i,j \leq d} \bigl|\mathfrak{C}(i,j) - E[F_i F_j]\bigr| + Ca^2 \log^{3/2}\bigl(2d^2 - 1 + \sqrt{e}\bigr),$$

*where $C > 0$ is a universal constant.*

REMARK 2.4. The above result (combined with Theorem 2.1) can be viewed as an analogy of the so-called *second-order Poincaré inequalities* proved in Nourdin, Peccati and Reinert [41]. Indeed, its proof is based on the lemmas proved there.

**3. Gaussian approximation of maxima of random symmetric quadratic forms.** In this section, we focus on the problem of approximating the distribution of maxima of symmetric quadratic forms. The next result can be easily derived from the results in the previous section.

THEOREM 3.1. *For each $n \in \mathbb{N}$, let $\boldsymbol{\xi}_n$ be an $N_n$-dimensional centered Gaussian vector with covariance matrix $\Sigma_n = (\Sigma_n(k,l))_{1\le k,l\le N_n}$ and $d_n \ge 2$ be an integer. Also, for each $k = 1, \dots, d_n$, let $A_{n,k}$ be an $N_n \times N_n$ symmetric matrix and $Z_n = (Z_{n,1}, \dots, Z_{n,d_n})^\top$ be a $d_n$-dimensional centered Gaussian vector with covariance matrix $\mathfrak{C}_n = (\mathfrak{C}_n(k,l))_{1\le k,l\le d_n}$. Set $F_{n,k} := \boldsymbol{\xi}_n^\top A_{n,k}\boldsymbol{\xi}_n - E[\boldsymbol{\xi}_n^\top A_{n,k}\boldsymbol{\xi}_n]$ and suppose that the following conditions are satisfied*:

(i) *There is a constant $b > 0$ such that $\mathfrak{C}_n(k,k) \ge b$ for every $n$ and every $k = 1, \dots, d_n$.*

(ii) $\max_{1\le k\le d_n}(E[F_{n,k}^4] - 3E[F_{n,k}^2]^2)\log^6 d_n \to 0$ *as* $n \to \infty$.

(iii) $\max_{1\le k,l\le d_n} |\mathfrak{C}_n(k,l) - E[F_{n,k}F_{n,l}]| \log^2 d_n \to 0$ *as* $n \to \infty$.

*Then we have*

$$\sup_{x\in\mathbb{R}} \Big| P\Big(\max_{1\le k\le d_n} F_{n,k} \le x\Big) - P\Big(\max_{1\le k\le d_n} Z_{n,k} \le x\Big)\Big| \to 0 \tag{3.1}$$

*and*

$$\sup_{x\in\mathbb{R}} \Big| P\Big(\max_{1\le k\le d_n} |F_{n,k}| \le x\Big) - P\Big(\max_{1\le k\le d_n} |Z_{n,k}| \le x\Big)\Big| \to 0$$

*as* $n \to \infty$.

REMARK 3.1. Since any symmetric Gaussian quadratic form can be written as a linear combination of independent $\chi^2$ random variables via eigenvalue decomposition (see, e.g., Section 3.2.1 of [17]), the readers may be wondering about whether it is possible to apply the original CCK theory to derive a similar result to Theorem 3.1 using eigenvalue decomposition. This is however impossible in general because the matrices $\Sigma_n^{1/2}A_{n,1}\Sigma_n^{1/2}, \dots, \Sigma_n^{1/2}A_{n,d_n}\Sigma_n^{1/2}$ are not necessarily simultaneously diagonalizable by an orthogonal matrix, which may induce an additional cross-sectional dependence after orthogonal transformation. To see this, suppose that $\Sigma_n$ is identity for simplicity. Then the aforementioned eigenvalue decomposition argument reads as follows: For each $k = 1, \dots, d_n$, we take an $N_n \times N_n$ real orthogonal matrix $U_{n,k}$ such that $U_{n,k}A_{n,k}U_{n,k}^\top$ is diagonal, and set $\boldsymbol{\varepsilon}_{n,k} = U_{n,k}\boldsymbol{\xi}_n$. Then the components of $\boldsymbol{\varepsilon}_{n,k}$ are independent and $F_{n,k}$ can be written as a linear combination of the squared components of $\boldsymbol{\varepsilon}_{n,k}$. However, for $k \ne l$,

the covariance matrix of $\boldsymbol{\varepsilon}_{n,k}$ and $\boldsymbol{\varepsilon}_{n,l}$ is given by $U_{n,k}U_{n,l}^{\top}$, which is generally not diagonal; for example, we have

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\left(\frac{1}{\sqrt{5}}\begin{pmatrix}2 & 1\\ 1 & -2\end{pmatrix}\right)^{\top} = \frac{1}{\sqrt{10}}\begin{pmatrix}3 & -1\\ 1 & 3\end{pmatrix}.$$

REMARK 3.2. Even if the matrices $\Sigma_n^{1/2}A_{n,1}\Sigma_n^{1/2}, \dots, \Sigma_n^{1/2}A_{n,d_n}\Sigma_n^{1/2}$ are simultaneously diagonalizable, there is gain to use Theorem 3.1 instead of the original CCK theory. To see this, suppose that each $F_{n,k}$ can be written as

$$F_{n,k} = \sum_{i=1}^{N_n} \lambda_{n,k}(i)\bigl(\eta_i^2 - 1\bigr),$$

where $\lambda_{n,k}(1), \dots, \lambda_{n,k}(N_n) \in \mathbb{R}$ and $(\eta_i)_{i=1}^{\infty}$ is a sequence of i.i.d. standard normal variables. In this case, if we assume that there are constants $\overline{b}, \underline{b} > 0$ such that

$$\underline{b} \le \sum_{i=1}^{N_n} \lambda_{n,k}(i)^2 \le \overline{b}$$

for all $n \in \mathbb{N}$ and $k = 1, \dots, d_n$ and that the matrix $\mathfrak{C}_n$ is equal to the covariance matrix of the variables $F_{n,1}, \dots, F_{n,d_n}$, Proposition 2.1 of [15] yields the convergence (3.1), provided that $B_n^2 \log^7(d_n N_n) = o(N_n)$ as $n \to \infty$, where

$$B_n = \sqrt{N_n} \max_{1\le k\le d_n} \max_{1\le i\le N_n} \bigl|\lambda_{n,k}(i)\bigr|.$$

Since we have

$$\max_{1\le k\le d_n} \bigl(E\bigl[F_{n,k}^4\bigr] - 3E\bigl[F_{n,k}^2\bigr]^2\bigr) \le \frac{B_n^2}{N_n}\overline{b}, \tag{3.2}$$

the convergence (3.1) is indeed implied by $B_n^2 \log^6 d_n = o(N_n)$ according to Theorem 3.1. In addition, the inequality (3.2) can be not tight. A cheap example is the case that

$$\lambda_{n,k}(i) = \begin{cases} 1/N_n^{1/4} & \text{if } i = k,\\ 1/\sqrt{N_n} & \text{otherwise.}\end{cases}$$

In this case, we have $B_n^2/N_n = 1/\sqrt{N_n}$, while it holds that

$$\max_{1\le k\le d_n} \bigl(E\bigl[F_{n,k}^4\bigr] - 3E\bigl[F_{n,k}^2\bigr]^2\bigr) = O\bigl(N_n^{-1}\bigr).$$

See also Remark 4.8 for another advantage of using our results instead of the original CCK theory.

REMARK 3.3 (Discussion on the fourth moment condition (ii)).

(i) In Theorem 3.1, the number $N_n$ does not necessarily diverge to get the convergence $E[F_{n,k}^4] - 3E[F_{n,k}^2]^2 \to 0$. This is because the variance of $\boldsymbol{\xi}_n$ is allowed to diverge in the setting of the theorem. To see this, suppose that $N_n = 1$, $\boldsymbol{\xi}_n$ is a centered Gaussian variable with variance $n$, and $A_{n,k} = 1/\sqrt{2n}$. In this case, we have $F_{n,k} = (\boldsymbol{\xi}_n^2 - n)/\sqrt{2n}$ and thus $E[F_{n,k}^2] = 1$ and $E[F_{n,k}^4] - 3E[F_{n,k}^2]^2 = 12/n \to 0$.

(ii) If $\sup_{n\in\mathbb{N}} \max_{1\le k\le d_n} E[F_{n,k}^2] < \infty$, a sufficient condition to prove the condition (ii) of Theorem 3.1 is $\max_{1\le k\le d_n} \|\Sigma_n^{1/2} A_{n,k} \Sigma_n^{1/2}\|_{\mathrm{sp}} \log^3 d_n \to 0$ as $n \to \infty$. This follows from the following inequality (see equation (11) of [17]):

$$E\big[F_{n,k}^4\big] - 3E\big[F_{n,k}^2\big]^2 = 48\,\mathrm{tr}\big[\big(\Sigma_n^{1/2} A_{n,k} \Sigma_n^{1/2}\big)^4\big] \le 24\big\|\Sigma_n^{1/2} A_{n,k} \Sigma_n^{1/2}\big\|_{\mathrm{sp}}^2 E\big[F_{n,k}^2\big]$$

(note that we always have $E[F_{n,k}^4] - 3E[F_{n,k}^2]^2 \ge 0$; see Remark 5.2.5 of [40]). In practice, it is often easier to check the condition on $\|\Sigma_n^{1/2} A_{n,k} \Sigma_n^{1/2}\|_{\mathrm{sp}}$ than to directly check the condition on $E[F_{n,k}^4] - 3E[F_{n,k}^2]^2$.

(iii) The condition $E[F_{n,k}^4] - 3E[F_{n,k}^2]^2 \to 0$ is necessary to approximate the distribution of the random variable $F_{n,k}$ by a Gaussian distribution if $\sup_{n\in\mathbb{N}} E[F_{n,k}^2] < \infty$ because there is a universal constant $c > 0$ such that

$$\sup_{n\in\mathbb{N}} E\big[F_{n,k}^8\big] \le c \sup_{n\in\mathbb{N}} E\big[F_{n,k}^2\big]^4 < \infty$$

(see, e.g., Theorem 5.10 of [25]), which implies the uniform integrability of the variables $F_{n,k}^2$ and $F_{n,k}^4$, $n = 1, 2, \ldots$. Actually, adopting an analogous discussion to the one from Chernozhukov, Chetverikov and Kato [15], we can easily generalize the conclusion of Theorem 3.1 to the convergence of the Kolmogorov distance between $F_n$ and $Z_n$ as follows:

$$\sup_{x_1,\ldots,x_{d_n}\in\mathbb{R}} \left| P\left(\bigcap_{k=1}^{d_n} \{F_{n,k} \le x_k\}\right) - P\left(\bigcap_{k=1}^{d_n} \{Z_{n,k} \le x_k\}\right)\right| \to 0.$$

Therefore, if $\sup_{n\in\mathbb{N}} \max_{1\le k\le d_n} E[F_{n,k}^2] < \infty$, the condition $\max_{1\le k\le d_n} (E[F_{n,k}^4] - 3E[F_{n,k}^2]^2) \log^6 d_n \to 0$ as $n \to \infty$ is indeed a necessary condition *when $d_n$ is fixed* (it is still unclear that this condition is necessary when $d_n \to \infty$ as $n \to \infty$, though).

In the next section, we will apply Theorem 3.1 to derive a Gaussian approximation of the null distribution of the test statistic for the absence of lead-lag effects. In order to implement the test in practice, we need to compute quantiles of the null distribution, but it is not easy to directly compute those of the derived Gaussian analog of the test statistic because its covariance structure contains unknown quantities for statisticians. For this reason, we will apply a wild bootstrap procedure to approximately compute quantiles of the null distribution. Theorem 3.1 is

still applicable for ensuring the validity of such a procedure as long as *Gaussian* wild bootstrapping is considered, but it turns out that a wild bootstrap procedure based on another distribution performs much better in finite samples. For this reason, we partially generalize Theorem 3.1 to a non-Gaussian case.

For every $n \in \mathbb{N}$, let $N_n \geq 1$ and $d_n \geq 2$ be integers and let $\Gamma_{n,k} = (\gamma_{n,k}(i, j))_{1 \leq i,j \leq N_n}$ be an $N_n \times N_n$ symmetric matrix for each $k = 1, \ldots, d_n$. We assume that $\gamma_{n,k}(i, i) = 0$ for all $i = 1, \ldots, N_n$, $k = 1, \ldots, d_n$ and $n \in \mathbb{N}$. Given a sequence $\xi = (\xi_i)_{i=1}^{\infty}$ of random variables, we set

$$Q_{n,k}(\xi) := \sum_{i,j=1}^{N_n} \gamma_{n,k}(i, j)\xi_i\xi_j, \qquad k = 1, \ldots, d_n$$

for every $n \in \mathbb{N}$.

Let $Y = (Y_i)_{i=1}^{\infty}$ be a sequence of independent variables such that $E[Y_i] = 0$ and $E[Y_i^2] = 1$ for every $i$. Also, let $G = (G_i)_{i=1}^{\infty}$ be a sequence of independent standard Gaussian variables. For every $i \in \mathbb{N}$, we define the random variables $(W_j^{(i)})_{j=1}^{\infty}$ by

$$W_j^{(i)} = \begin{cases} Y_j & \text{if } j \leq i, \\ G_j & \text{if } j > i. \end{cases}$$

THEOREM 3.2. *For each $n \in \mathbb{N}$, let $Z_n = (Z_{n,1}, \ldots, Z_{n,d_n})^{\top}$ be a $d_n$-dimensional centered Gaussian vector with covariance matrix $\mathfrak{C}_n = (\mathfrak{C}_n(k, l))_{1 \leq k,l \leq d_n}$, and set*

$$R_{n,1} = \sum_{i=1}^{N_n} E\left[\max_{1 \leq k \leq d_n} \left|\sum_{j=1}^{N_n} \gamma_{n,k}(i, j)W_j^{(i)}\right|^3\right]\left(E\left[|Y_i|^3\right] + E\left[|G_i|^3\right]\right),$$

$$R_{n,2} = \max_{1 \leq k,l \leq d_n} \left|\mathfrak{C}_n(k, l) - E\left[Q_{n,k}(G)Q_{n,l}(G)\right]\right|,$$

$$R_{n,3} = \max_{1 \leq k \leq d_n} \sqrt{E\left[Q_{n,k}(G)^4\right] - 3E\left[Q_{n,k}(G)^2\right]^2}.$$

*Suppose that there is a constant $b > 0$ such that $\mathfrak{C}_n(k, k) \geq b$ for every $n$ and every $k = 1, \ldots, d_n$. Then we have*

$$\sup_{x \in \mathbb{R}} \left|P\Big(\max_{1 \leq k \leq d_n} |Q_{n,k}(Y)| \leq x\Big) - P\Big(\max_{1 \leq k \leq d_n} |Z_{n,k}| \leq x\Big)\right| \to 0$$

*as $n \to \infty$, provided that $R_{n,1} \log^{\frac{7}{2}} d_n \vee R_{n,2} \log^2 d_n \vee R_{n,3} \log^3 d_n \to 0$.*

REMARK 3.4. The variables $W_j^{(i)}$ are related to the so-called *Lindeberg method*. In fact, our proof of Theorem 3.2 is based on the generalized Lindeberg method developed in [35, 42] (see also Chapter 11 of [40]).

REMARK 3.5. There is probably room for improvement in Theorem 3.2. In particular, the truncation arguments used in the CCK theory (based on Lemma A.6 of [9]) are apparently applicable to our case, which would significantly weaken the assumptions of Theorem 3.2. On the other hand, it is less obvious whether the other techniques used in the CCK theory (especially in Chernozhukov, Chetverikov and Kato [15]) are applicable to our case or not. Their excellent argument leads a very sharp bound, but it seems crucial in their argument that the statistics considered there is a linear function of independent random variables. More precisely, to apply their argument to our case, the independence between the variables $U_i$ and $V_i$ appearing in the proof of Theorem 3.2 seems necessary, but this is not the case (such a structure is necessary to get an analogous estimate to equation (30) of [15], e.g.). This issue is left to future research.

REMARK 3.6. Analogous quantities to $R_{n,2}$ and $R_{n,3}$ from Theorem 3.2 have already appeared in Theorem 3.1 and it is usually not difficult to bound them. On the other hand, as long as the third moments of $Y_i$'s are uniformly bounded, the quantity $R_{n,1}$ is bounded by the third moment of the maximum of a sum of (high-dimensional) independent random vectors, so we have many inequalities which can be used to bound it (see, e.g., Chapter 14 of [6]). Here, we give two examples of such inequalities. The first one only requires the uniform boundedness of the $p$th moments of $Y_i$'s for some $p \geq 3$, while the latter one is applicable when the variables $Y_i$ are sub-Gaussian.

LEMMA 3.1. (a) *Suppose that* $\sup_{i\in\mathbb{N}} \|Y_i\|_p < \infty$ *for some* $p \geq 3$. *Then*

$$\sum_{i=1}^{N_n} E\left[\max_{1\leq k\leq d_n}\left|\sum_{j=1}^{N_n}\gamma_{n,k}(i,j)W_j^{(i)}\right|^3\right] \leq 2d_n^{3/p}(p-1)^{3/2}\sup_{i\in\mathbb{N}}\|Y_i\|_p^3\sum_{i=1}^{N_n}\max_{1\leq k\leq d_n}\left(\sum_{j=1}^{N_n}\gamma_{n,k}(i,j)^2\right)^{3/2}$$

*for every* $n$.

(b) *Suppose that there is a constant* $a > 0$ *such that* $Y_i$ *is sub-Gaussian relative to the scale* $a$ *for all* $i = 1, 2, \ldots$. *Then*

$$\sum_{i=1}^{N_n} E\left[\max_{1\leq k\leq d_n}\left|\sum_{j=1}^{N_n}\gamma_{n,k}(i,j)W_j^{(i)}\right|^3\right] \leq 5^{3/2}a^3\log^{3/2}(2d_n-1+\sqrt{e})\sum_{i=1}^{N_n}\max_{1\leq k\leq d_n}\left(\sum_{j=1}^{N_n}\gamma_{n,k}(i,j)^2\right)^{3/2}$$

*for every* $n$.

Using the above lemma, we obtain a useful criterion to check the conditions appearing in Theorem 3.2 in terms of the so-called *influence indices*: Given a symmetric matrix $\Gamma = (\gamma(i,j))_{1\le i,j\le N}$, the *influence* of the variable $i$ of $\Gamma$ is defined by

$$\mathrm{Inf}_i(\Gamma) = \sum_{j=1}^{N} \gamma(i,j)^2$$

for $i = 1, \dots, N$. The influence indices play an important role in studies of the central limit theorem for random quadratic forms (and homogeneous sums more generally); see [21, 35, 42], for example.

COROLLARY 3.1. *Suppose that there is a constant $a > 0$ such that $Y_i$ is sub-Gaussian relative to the scale $a$ for all $i = 1, 2, \dots$. Then the convergences $R_{n,1}\log^{\frac{7}{2}} d_n \to 0$ and $R_{n,3}\log^3 d_n \to 0$ are implied by the following condition*:

$$\begin{aligned}&(\log d_n)^6 \max_{1\le k\le d_n} \mathrm{tr}\big(\Gamma_{n,k}^4\big)\\ &\quad + (\log d_n)^5 \Big(\max_{1\le i\le N_n} \sqrt{\Lambda_{n,i}}\Big) \sum_{i=1}^{N_n} \Lambda_{n,i} \to 0 \qquad \text{as } n\to\infty,\end{aligned} \tag{3.3}$$

*where*

$$\Lambda_{n,i} = \max_{1\le k\le d_n} \mathrm{Inf}_i(\Gamma_{n,k}), \qquad i = 1, \dots, N_n.$$

REMARK 3.7 (Implication of condition (3.3)). Let us consider the case that there is a symmetric matrix $\bar\Gamma_n = (\bar\gamma_n(i,j))_{1\le i,j\le N_n}$ such that $\mathrm{Inf}_i(\bar\Gamma_n) = \Lambda_{n,i}$ for all $i = 1, \dots, N_n$. Namely, the influence indices of the matrices $\Gamma_{n,1}, \dots, \Gamma_{n,d_n}$ are dominated by that of the matrix $\bar\Gamma_n$. In this case, condition (3.3) reads as

$$(\log d_n)^6 \max_{1\le k\le d_n} \mathrm{tr}\big(\Gamma_{n,k}^4\big) + (\log d_n)^{10} \|\bar\Gamma_n\|_F^4 \max_{1\le i\le N_n} \mathrm{Inf}_i(\bar\Gamma_n) \to 0 \qquad \text{as } n\to\infty.$$

The quantity $\|\bar\Gamma_n\|_F^2$ is the variance of the quadratic form

$$\bar Q_n(Y) = \sum_{i,j=1}^{N_n} \bar\gamma_n(i,j) Y_i Y_j.$$

Therefore, it would be natural to assume $\sup_{n\in\mathbb{N}} \|\bar\Gamma_n\|_F^2 < \infty$. Moreover, in many cases it is reasonable to expect $\max_{1\le i\le d_n} \mathrm{Inf}_i(\bar\Gamma_n) = O(N_n^{-1})$ because we have by definition

$$\max_{1\le i\le d_n} \mathrm{Inf}_i(\bar\Gamma_n) \le \|\bar\Gamma_n\|_{\mathrm{sp}}^2.$$

According to [21], $\|\bar{\Gamma}_n\|_{\mathrm{sp}}$ gives an optimal convergence rate for the Kolmogorov distance between

$$\bar{Q}_n(Y)/\sqrt{\mathrm{Var}[\bar{Q}_n(Y)]}$$

and a standard Gaussian variable, hence it seems reasonable to expect $\|\bar{\Gamma}_n\|_{\mathrm{sp}} = O(N_n^{-1/2})$ in view of the standard Berry–Esseen inequality. Moreover, since $\mathrm{tr}(\Gamma_{n,k}^4) \le \|\Gamma_{n,k}\|_{sp}^2 \|\Gamma_{n,k}\|_F^2$, we might expect $\max_{1\le k\le d_n} \mathrm{tr}(\Gamma_{n,k}^4) = O(N_n^{-1})$ due to a similar reason. Consequently, the condition (3.3) is typically satisfied when $\log^{10} d_n = o(N_n)$ as $n \to \infty$.

A typical example satisfying the above condition is the situations where $\Gamma_{n,k}$'s correspond to sample auto-covariances:

$$\gamma_{n,k}(i,j) = \begin{cases} 1/\sqrt{N_n} & \text{if } |j-i| = k, \\ 0 & \text{otherwise.} \end{cases}$$

In this case, the quantities $\sum_{j=1}^{N_n} \gamma_{n,k}(i,j)^2$ does not depend on $k$, so we can take $\bar{\Gamma}_n = \Gamma_{n,1}$ for example.

## 4. Application to high-frequency data.

4.1. *Testing the absence of lead-lag effects.* We turn to the problem of testing the absence of lead-lag effects which is mentioned at the beginning of the Introduction. Here, we consider a more general setting than the one described in the Introduction by allowing (deterministic) time-varying volatilities as well as the presence of multiple lead-lag times under the alternative.

Let $\rho_1, \dots, \rho_M$ be real numbers satisfying the condition $\sum_{m=1}^M |\rho_m| < 1$. Also, let $\theta_1, \dots, \theta_M$ be mutually different numbers. Then, by Proposition 2 from [22] there is a bivariate Gaussian process $B_t = (B_t^1, B_t^2)$ $(t \in \mathbb{R})$ with stationary increments such that both $B^1$ and $B^2$ are standard Brownian motions as well as $B^1$ and $B^2$ have the cross spectral density given by

$$\mathfrak{s}(\lambda) = \sum_{m=1}^M \rho_m e^{-\sqrt{-1}\theta_m\lambda}, \qquad \lambda \in \mathbb{R}.$$

This means that we have

$$E\left[\left(\int_{-\infty}^{\infty} f(t)\,dB_t^1\right)\left(\int_{-\infty}^{\infty} g(t)\,dB_t^2\right)\right] = \sum_{m=1}^M \rho_m \int_{-\infty}^{\infty} f(t)g(t+\theta_m)\,dt$$

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

For each $\nu = 1, 2$, we consider the process $X^\nu = (X_t^\nu)_{t\ge 0}$ given by

$$X_t^\nu = X_0^\nu + \int_0^t \sigma_\nu(s)\,dB_s^\nu, \qquad t \ge 0, \tag{4.1}$$

where $\sigma_\nu \in L^2(0,\infty)$ is nonnegative-valued and deterministic. We observe the process $X^\nu$ on the interval $[0,T]$ at the deterministic sampling times $0 \le t_0^\nu < t_1^\nu < \cdots < t_{n_\nu}^\nu \le T$, which implicitly depend on the parameter $n \in \mathbb{N}$ such that

$$r_n := \max_{\nu=1,2} \max_{i=0,1,\ldots,n_\nu+1} \bigl(t_i^\nu - t_{i-1}^\nu\bigr) \to 0$$

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

REMARK 4.1. It is not difficult to extend the following discussion to the case that the volatilities $\sigma_1, \sigma_2$ and the sampling times $(t_i^1)_{i=0}^{n_1}$, $(t_j^2)_{j=0}^{n_2}$ are random but independent of the process $B$, but we focus on the deterministic case for the simplicity of notation. Extension to a situation where the volatilities depend on $B$ is nontrivial because of the nonergodic nature of the problem (i.e., the asymptotic covariance matrix of the statistics $(U_n(\theta))_{\theta\in\mathcal{G}_n}$ defined below generally depends on $B$) and we leave it to future research.

Our aim is to construct a testing procedure for the following statistical hypothesis testing problem based on discrete observation data $(X^1_{t_i^1})_{i=0}^{n_1}$ and $(X^2_{t_i^2})_{i=0}^{n_2}$:

$$\begin{aligned} &H_0 : \rho_m = 0 \qquad \text{for all } m = 1, \ldots, M \quad \text{vs} \\ &H_1 : \rho_m \neq 0 \qquad \text{for some } m = 1, \ldots, M. \end{aligned} \tag{4.2}$$

We introduce some notation. For each $\nu = 1, 2$, we associate the observation times $(t_i^\nu)_{i=0}^{n_\nu}$ with the collection of intervals $\Pi_n^\nu = \{(t_{i-1}^\nu, t_i^\nu] : i = 1, \ldots, n_\nu\}$. We will systematically employ the notation $I$ (resp., $J$) for an element of $\Pi_n^1$ (resp., $\Pi_n^2$). For an interval $S \subset [0,\infty)$, we set $\overline{S} = \sup S$, $\underline{S} = \inf S$ and $|S| = \overline{S} - \underline{S}$. In addition, we set $V(S) = V_{\overline{S}} - V_{\underline{S}}$ for a a stochastic process $(V_t)_{t\ge0}$, and $S_\theta = S + \theta$ for a real number $\theta$. We define the Hoffmann–Rosenbaum–Yoshida cross-covariance estimator by

$$U_n(\theta) = \sum_{I\in\Pi_n^1, J\in\Pi_n^2} X^1(I)X^2(J)K(I, J_{-\theta}),$$

where we set $K(I,J) = 1_{\{I\cap J\neq\varnothing\}}$ for two intervals $I$ and $J$. Now our test statistic is given by

$$T_n = \sqrt{n} \max_{\theta\in\mathcal{G}_n} \bigl|U_n(\theta)\bigr|,$$

where $\mathcal{G}_n$ is a finite subset of $\mathbb{R}$.

To establish the asymptotic property of our test statistic $T_n$, we first investigate the asymptotic property of the following quantity:

$$F_n(\theta) = \sqrt{n}\bigl(U_n(\theta) - E\bigl[U_n(\theta)\bigr]\bigr).$$

We impose the following conditions:

[A1] $\sup_{t\in[0,T]}(\sigma_1(t)+\sigma_2(t))<\infty$.

[A2] There are positive constants $\underline{v}, \overline{v}$ such that $\underline{v} \le V_n(\theta) \le \overline{v}$ for all $n \in \mathbb{N}$ and $\theta \in \mathcal{G}_n$, where

$$V_n(\theta) = n \sum_{I\in\Pi_n^1, J\in\Pi_n^2} \left(\int_I \sigma_1(t)^2\,dt\right)\left(\int_J \sigma_2(t)^2\,dt\right) K(I, J_{-\theta}).$$

[A3] $\Sigma(\theta_m) > 0$ for all $m = 1, \dots, M$, where

$$\Sigma(\theta) = \begin{cases} \displaystyle\int_0^{T-\theta} \sigma_1(t)\sigma_2(t+\theta)\,dt & \text{if } \theta \ge 0, \\ \displaystyle\int_0^{T+\theta} \sigma_1(t-\theta)\sigma_2(t)\,dt & \text{if } \theta < 0. \end{cases}$$

[A4] The grid $\mathcal{G}_n$ satisfies the following conditions:

(i) There is a constant $\gamma > 0$ such that $\#\mathcal{G}_n = O(n^\gamma)$ as $n \to \infty$.

(ii) There is a sequence $(\upsilon_n)_{n\in\mathbb{N}}$ of positive numbers such that

$$\{\theta_1, \dots, \theta_M\} \subset \bigcup_{\theta\in\mathcal{G}_n} [\theta - \upsilon_n, \theta + \upsilon_n]$$

and $\lim_{n\to\infty} \upsilon_n \min\{n_1, n_2\} = 0$.

REMARK 4.2. Assumption [A1] is standard in the literature and satisfied when $\sigma_1$ and $\sigma_2$ are càdlàg, for example. [A2] roughly says that the scaling factor $\sqrt{n}$ is appropriate [the quantity $V_n(\theta)$ is related to the variance of $U_n(\theta)$]. [A2] holds true, for example, when $0 < \inf_{t\in[0,T]} \sigma_\nu(t) \le \sup_{t\in[0,T]} \sigma_\nu(t) < \infty$ for every $\nu = 1, 2$, $n\sum_{I\in\Pi_n^1} |I|^2 + n\sum_{J\in\Pi_n^2} |J|^2 = O(1)$ as $n \to \infty$ and there is a constant $c > 0$ such that $n(|I| \wedge |J|) \ge c$ for every $n$ and all $I \in \Pi_n^1$, $J \in \Pi_n^2$. [A3] ensures that $\max_{1\le m\le M} |E[U_n(\theta_m)]|$ does not vanish under $H_1$. [A4] ensures that $\mathcal{G}_n$ is sufficiently fine to capture the cross-covariance at the lag $\theta_m$ for every $m$. Note that [A4] is also assumed in [24] (see Assumption B3 of [24]).

PROPOSITION 4.1. *For each $n \in \mathbb{N}$, let $(Z_n(\theta))_{\theta\in\mathcal{G}_n}$ be a family of centered Gaussian variables such that $E[Z_n(\theta)Z_n(\theta')] = E[F_n(\theta)F_n(\theta')]$ for all $\theta, \theta' \in \mathcal{G}_n$. Under assumptions* [A1]–[A2], *we have*

$$\sup_{x\in\mathbb{R}} \left| P\left(\max_{\theta\in\mathcal{G}_n} |F_n(\theta)| \le x\right) - P\left(\max_{\theta\in\mathcal{G}_n} |Z_n(\theta)| \le x\right)\right| \to 0$$

*as $n \to \infty$, provided that $nr_n^2 \log^6(\#\mathcal{G}_n) \to 0$.*

REMARK 4.3. It is impossible to apply the original CCK theory (at least naively) to prove Proposition 4.1 because we need to apply Theorem 3.1 to a situation where the matrices $\Sigma_n^{1/2} A_{n,1} \Sigma_n^{1/2}, \dots, \Sigma_n^{1/2} A_{n,d_n} \Sigma_n^{1/2}$ are not simultaneously

diagonalizable. In fact, if we consider the synchronous and equidistant sampling with the step size $1/n$, the matrices corresponding to $U_n(\pm 1/n)$ are of the form

$$\begin{pmatrix} O & A \\ A^\top & O \end{pmatrix},$$

where we take the matrix $A = (a_{ij})$ as

$$a_{ij} = \begin{cases} 1 & \text{if } j - i = \pm 1, \\ 0 & \text{otherwise.} \end{cases}$$

We can easily check that those matrices are not commutative unless the size of $A$ is less than or equal to 2.

The above proposition suggests that the null distribution of our test statistic $T_n$ could be approximated by that of $\max_{\theta\in\mathcal{G}_n} |Z_n(\theta)|$ for sufficiently large $n$. However, it is not easy to evaluate the distribution of $\max_{\theta\in\mathcal{G}_n} |Z_n(\theta)|$ directly, so we rely on a (wild) bootstrap procedure to construct critical regions for our test. The above Gaussian approximation result plays a role in validating the bootstrap procedure.

Let $(w_I^1)_{I\in\Pi_n^1}$ and $(w_J^2)_{J\in\Pi_n^2}$ be mutually independent sequence of i.i.d. random variables which are independent of the processes $X^1$ and $X^2$. Then we set

$$U_n^*(\theta) = \sum_{I\in\Pi_n^1, J\in\Pi_n^2} \big(w_I^1 X^1(I)\big)\big(w_J^2 X^2(J)\big) K(I, J_{-\theta}).$$

Given a significance level $\alpha$, we denote by $q_n^*(1-\alpha)$ the $100(1-\alpha)\%$ quantile of the bootstrapped test statistic $T_n^* = \sqrt{n}\max_{\theta\in\mathcal{G}_n} |U_n^*(\theta)|$, conditionally on $X^1$ and $X^2$:

$$q_n^*(1-\alpha) = \inf\big\{z \in \mathbb{R} : P\big(T_n^* \le z | \mathcal{F}^X\big) \ge 1-\alpha\big\},$$

where $\mathcal{F}^X$ is the $\sigma$-field generated by the processes $X^1$ and $X^2$.

REMARK 4.4. We generate the bootstrap observations under the null hypothesis $H_0$. This is a typical approach in the bootstrap test literature (see, e.g., [4]). Moreover, as discussed in Section 4 of [19] as well as Section 2 of [18], this approach often serves as refining the performance of the test.

PROPOSITION 4.2. *Suppose that* [A1]–[A4] *are satisfied. Suppose also that* $E[w_I^1] = E[w_J^2] = 0$, $E[(w_I^1)^2] = E[(w_J^2)^2] = 1$ *for all* $I, J$ *and there is a constant* $a > 0$ *such that both* $w_I^1$ *and* $w_J^2$ *are sub-Gaussian relative to the scale a for all* $I, J$. *Suppose further that* $r_n = O(n^{-3/4-\eta})$ *as* $n \to \infty$ *for some* $\eta > 0$. *Then the following statements hold true for all* $\alpha \in (0, 1)$:

(a) *Under* $H_0$, *we have* $P(T_n \ge q_n^*(1-\alpha)) \to \alpha$ *as* $n \to \infty$.

(b) *Under $H_1$, we have $P(T_n \geq q_n^*(1-\alpha)) \to 1$ as $n \to \infty$.*

By Proposition 4.2, given a significance level $\alpha \in (0,1)$, we obtain a consistent and asymptotically level $\alpha$ test for (4.2) by rejecting the null hypothesis if $T_n \geq q_n^*(1-\alpha)$. Of course, in the practical implementation we replace $q_n^*(1-\alpha)$ by a simulated one. For example, given observation data, we generate i.i.d. copies $T_n^*(1), \ldots, T_n^*(R)$ of $T_n^*$ (conditionally on the observation data) with some sufficiently large integer $R$. Then we replace the function $P(T_n^* \leq z|\mathcal{F}^X)$ of $z$ by its empirical counterpart $\frac{1}{R}\sum_{r=1}^R 1_{\{T_n^*(r)\leq z\}}$ and compute $q_n^*(1-\alpha)$ accordingly. Note that this is equivalent to computing the *bootstrap p-value* $\hat{p}^* = \frac{1}{R}\sum_{r=1}^R 1_{\{T_n^*(r)>T_n\}}$ and rejecting the null hypothesis if $\hat{p}^* \leq \alpha$.

REMARK 4.5. The proposed test is evidently invariant under multiplying a constant. In particular, the factor $\sqrt{n}$ can be dropped when we implement the test in practice.

REMARK 4.6 (Choice of the multiplier variables). Choice of the distribution of the multiplier variables $(w_I^1)_{I\in\Pi_n^1}$ and $(w_J^2)_{J\in\Pi_n^2}$ are important for the finite sample property of the test. In our situation, it turns out that choosing Rademacher variables induces a quite good finite sample performance of our testing procedure. Namely, the proposed test performs very well in finite samples when the distributions of $w_I^1$ and $w_J^2$ are chosen according to

$$P(w_I^1 = 1) = P(w_I^1 = -1) = P(w_J^2 = 1) = P(w_J^2 = -1) = \frac{1}{2}.$$

This is presumably because the above choice makes the *unconditional* distribution of the bootstrapped test statistics of $T_n^*$ coincide with the distribution of $T_n$. This can be shown in the same way as in the proof of Theorem 1 from [18]. For this reason, we recommend that we should use Rademacher variables as the multiplier variables for the above testing procedure (and we do so in the simulation study of Section 5).

4.2. *Uniform confidence bands for spot volatility.* To illustrate another potential application of our main results, we present an application of our result to constructing uniform confidence bands for spot volatility. This section is only for an illustration purpose, so we do neither pursue the generality of the theory nor discuss practical problems on implementation such as the choice of a bandwidth and a kernel function. We refer to Section 6 of [31] for a discussion on the latter issue.

Let us consider the stochastic process $X = (X_t)_{t\in[0,T]}$ which is defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ and of the form

$$X_t = X_0 + \int_0^t \sigma(s)\, dB_s, \qquad t \in [0,T],$$

where $B = (B_t)_{t\in[0,T]}$ is a standard $(\mathcal{F}_t)$-Brownian motion and $\sigma = (\sigma(t))_{t\in[0,T]}$ is a continuous $(\mathcal{F}_t)$-adapted process. The aim of this section is to construct uniform confidence bands for the *spot volatility* $\sigma^2$ based on the high-frequency observation data $\{X_{t_i}\}_{i=0}^n$, where $t_i = Ti/n$, $i = 0, 1, \dots, n$.

Specifically, we consider the following kernel-type estimator for $\sigma^2$ (cf. [20, 31]):

$$\widehat{\sigma}_n^2(t) := \sum_{i=1}^n K_h(t_{i-1} - t)(X_{t_i} - X_{t_{i-1}})^2, \qquad t \in [0, T],$$

where $h := h_n > 0$ is a bandwidth parameter, $K_h(x) = K(x/h)/h$ for $x \in \mathbb{R}$ and $K : \mathbb{R} \to \mathbb{R}$ is a kernel function. We derive a Gaussian approximation result for the supremum of the Studentization of $\widehat{\sigma}_n^2(t)$. Let us set

$$\mathfrak{s}_n(t) = \sqrt{\frac{2}{n^2}\sum_{i=1}^n K_h(t_{i-1} - t)^2}$$

for $t \in [0, T]$. In view of Theorem 3 from [31], $\sigma^2(t)\mathfrak{s}_n(t)$ can be seen as an approximation of the asymptotic standard error of $\widehat{\sigma}_n^2(t)$. We define the Gaussian analog of the Studentization of $\widehat{\sigma}_n^2(t)$ as follows. For each $n \in \mathbb{N}$, let $(z_i^n)_{i=1}^n$ be a sequence of i.i.d. centered Gaussian variables with variance $2/n^2$, and we set

$$Z_n(t) = \frac{1}{\mathfrak{s}_n(t)}\sum_{i=1}^n K_h(t_{i-1} - t)z_i^n, \qquad t \in [0, T].$$

We impose the following conditions:

[B1] $w(\sigma; \eta) = O_p(\eta^\gamma)$ as $\eta \to 0$ for some $\gamma \in (0, 1]$. Moreover, $\sigma^2(t) > 0$ for all $t \in [0, T]$ almost surely.

[B2] The kernel function $K : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous and compactly supported as well as satisfies $\int_{-\infty}^{\infty} K(t)\,dt = 1$.

We also impose the following strengthened version of assumption [B1] when deriving the convergence rate of Gaussian approximation:

[SB1] There is a constant $\Lambda > 0$ such that $\Lambda^{-1} \le |\sigma(t)| \le \Lambda$ and $w(\sigma; \eta) \le \Lambda\eta^\gamma$ for all $t \in [0, T]$ and $\eta \in (0, 1)$.

PROPOSITION 4.3. *Suppose that* [B1]–[B2] *are satisfied. Suppose also that the bandwidth parameter $h$ satisfies $nh^{1+2\gamma}\log n \to 0$ and $\log^6 n/nh \to 0$ as $n \to \infty$. Let $a_n$ be a sequence of positive numbers such that $a_n \to 0$ and $a_n/h \to \infty$ as $n \to \infty$. Then we have*

$$\sup_{x\in\mathbb{R}}\left|P\left(\sup_{t\in[a_n,T-a_n]}\left|\frac{\widehat{\sigma}_n^2(t) - \sigma^2(t)}{\sigma^2(t)\mathfrak{s}_n(t)}\right| \le x\right) - P\left(\sup_{t\in[a_n,T-a_n]}|Z_n(t)| \le x\right)\right| \to 0 \tag{4.3}$$

*as* $n \to \infty$. *Moreover, if we additionally assume* [SB1], *we have*

$$\sup_{x\in\mathbb{R}}\left|P\Big(\sup_{t\in[a_n,T-a_n]}\left|\frac{\widehat{\sigma}_n^2(t)-\sigma^2(t)}{\sigma^2(t)\mathfrak{s}_n(t)}\right|\le x\Big)-P\Big(\sup_{t\in[a_n,T-a_n]}|Z_n(t)|\le x\Big)\right| = O\big(\sqrt{nh^{1+2\gamma}\log n}\big)+O\Big(\frac{\log n}{(nh)^{\frac{1}{6}}}\Big) \tag{4.4}$$

*as* $n \to \infty$.

REMARK 4.7. We introduce the parameters $a_n$ in Proposition 4.3 to avoid boundary effects. See Section 4 of [31] for more details about this topic.

REMARK 4.8. Although we use Lemma 2.1 to prove Proposition 4.3 (see the proof of Lemma B.17), we can indeed use Theorem 3.1 of [14] instead to derive a similar result. However, the result requires a (slightly) stronger condition on the bandwidth $h$ and leads to a worse convergence rate. In fact, an inspection of the proof of Proposition 2.1 from [15] implies that we need to replace $\varepsilon^{-2}$ by $\varepsilon^{-3}$ in the inequality (B.9) if we use Theorem 3.1 of [14] instead of Lemma 2.1 to prove Lemma B.17. Then the optimal choice of $\varepsilon$ in the proof of Proposition 4.3 becomes $\varepsilon = (nh)^{-\frac{1}{8}}\log^{3/8} n$, which changes the order of the second term on the right-hand side of (4.4) to $O((nh)^{-\frac{1}{8}}\log^{7/8} n)$. Hence we need the condition $\log^7 n/nh \to 0$ as $n \to \infty$ to get the convergence (4.3).

In contrast to the previous subsection, the process $Z_n(t)$ does not contain any unknown parameter, so Proposition 4.3 is readily applicable to construction of uniform confidence bands for $\sigma^2$: Given a significance level $\alpha \in (0, 1)$, let $q_n(1-\alpha)$ be the $100(1-\alpha)\%$ quantile of the variable $\sup_{t\in[a_n,T-a_n]}|Z_n(t)|$ (which can be computed, e.g., by simulation). Then

$$\left[\frac{\widehat{\sigma}_n^2(t)}{1+\mathfrak{s}_n(t)\cdot q_n(1-\alpha)},\frac{\widehat{\sigma}_n^2(t)}{1-\mathfrak{s}_n(t)\cdot q_n(1-\alpha)}\right],\qquad t\in[a_n,T-a_n],$$

give asymptotically uniformly valid $100(1-\alpha)\%$ confidence bands for $\sigma^2(t)$, $t \in [a_n, T-a_n]$.

REMARK 4.9. The applications considered in this section concerns asymptotic settings where the terminal value $T$ of the sampling interval is fixed. Here, we briefly discuss applicability of our theory to asymptotic settings where the terminal value $T$ of the sampling interval tends to infinity. In such a setting, a typical problem which our theory seems to fit would be constructing uniform confidence bands for the coefficient functions of an ergodic diffusion process. Nonparametric estimation of the coefficient functions of a diffusion process from high-frequency data is extensively studied in the literature, but most studies focus only on point-wise

inference (except for Kanaya [26], where uniform convergence rates of kernel-based estimators have been derived; see also Söhl and Trabs [49] where the authors construct uniform confidence bands for the drift coefficient of a diffusion process based on low-frequency observation data), so it would be important to consider this problem. In such a problem, estimators typically have deterministic asymptotic covariance matrices, hence the issue indicated in Remark 4.1 does not arise. Unfortunately, however, we encounter another issue that it seems difficult (at least not straightforward) to get a reasonable estimate for the quantity $\Delta$ in this problem. This is perhaps because we do not take account of special properties of the underlying diffusion process (such as the Markov and mixing properties) when deriving our estimate. Therefore, this issue might be resolved by adopting the approach from Kusuoka and Yoshida [32] where Malliavin calculus is locally applied to the underlying process with taking account of the Markov and mixing properties. However, a rigorous treatment of this idea is rather demanding, so we leave it to future work.

**5. Numerical illustration.** In this section, we illustrate the finite sample performance of the testing procedure for the absence of lead-lag effects, which is proposed in Section 4.1.[3] The setting of our numerical experiments is basically adopted from Section 5 of [24]. Specifically, we simulate model (1.1) with $T = 1, \vartheta = 0.1, x_0^1 = x_0^2 = 0, \sigma_1 = \sigma_2 = 1$. We vary the correlation parameter as $\rho \in \{0, 0.25, 0.5, 0.75\}$ to examine the size and the power of the testing procedure. We consider both synchronous and nonsynchronous sampling scenarios. For the synchronous sampling scenario $t_i^1 = t_i^2 = ih_n$, $i = 0, 1, \dots, \lfloor Th_n^{-1} \rfloor$, we examine three kinds of sampling frequencies: $h_n \in \{10^{-3}, 3 \cdot 10^{-3}, 6 \cdot 10^{-3}\}$. Also, in these scenarios we set $\mathcal{G}_n = \{kh_n : k \in \mathbb{Z}, |kh_n| \leq 0.3\}$ as the search grid. On the other hand, for the nonsynchronous sampling scenario, we first simulate the processes on the equidistant times $i \cdot 10^{-3}$, $i = 0, 1, \dots, 1000$, then we randomly pick 300 sampling times for $X^1$. We do so for $X^2$ independently of the sampling for $X^1$. In this scenario, we set $\mathcal{G}_n = \{k \cdot 10^{-3} : k \in \mathbb{Z}, |k| \leq 300\}$ as the search grid. For the testing procedure, we use Rademacher variables as the multiplier variables and 999 bootstrap replications to construct the critical regions. We run 10,000 Monte Carlo iterations in each experiment.

Table 1 reports the rejection rates of the proposed test in each experiment. For the case $\rho = 0$, the numbers should be close to the corresponding significance levels, and this is true for all the experiments. Turning to the power performance, we find that in the low correlation case $\rho = 0.25$ the power of the test is rather weak except for the most frequent sampling scenario. This is reasonable in view of the simulation results reported in [24], which indicate that the contrast function $U_n(\theta)$ becomes flat in that case. For the moderate and the high correlation cases $\rho = 0.5$ and $\rho = 0.75$, the power of the test is satisfactory.

[3]The proposed testing procedure is implemented in the R package **yuima** as the function `llag.test` since version 1.7.2.

TABLE 1
*Rejection rates of the proposed test*

| | $\rho = 0$ | $\rho = 0.25$ | $\rho = 0.50$ | $\rho = 0.75$ |
|---|---|---|---|---|
| | | Synchronous sampling scenario | | |
| $h_n = 10^{-3}$ | | | | |
| $\alpha = 0.01$ | 0.011 | 1.000 | 1.000 | 1.000 |
| $\alpha = 0.05$ | 0.050 | 1.000 | 1.000 | 1.000 |
| $\alpha = 0.10$ | 0.100 | 1.000 | 1.000 | 1.000 |
| $h_n = 3 \cdot 10^{-3}$ | | | | |
| $\alpha = 0.01$ | 0.010 | 0.139 | 0.977 | 1.000 |
| $\alpha = 0.05$ | 0.051 | 0.281 | 0.993 | 1.000 |
| $\alpha = 0.10$ | 0.101 | 0.382 | 0.997 | 1.000 |
| $h_n = 6 \cdot 10^{-3}$ | | | | |
| $\alpha = 0.01$ | 0.011 | 0.041 | 0.634 | 0.997 |
| $\alpha = 0.05$ | 0.050 | 0.131 | 0.802 | 1.000 |
| $\alpha = 0.10$ | 0.099 | 0.214 | 0.867 | 1.000 |
| | | Nonsynchronous sampling scenario | | |
| $\alpha = 0.01$ | 0.010 | 0.056 | 0.753 | 1.000 |
| $\alpha = 0.05$ | 0.051 | 0.152 | 0.879 | 1.000 |
| $\alpha = 0.10$ | 0.099 | 0.235 | 0.919 | 1.000 |

*Note*. $\alpha$ denotes the significance level.

**Acknowledgments.** The author thanks Kengo Kato and Daisuke Kurisu for their helpful comments. The author is also grateful to the Editor and two anonymous referees for their constructive comments which have significantly improved the original version of the paper.

## SUPPLEMENTARY MATERIAL

**Supplement to "Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data"** (DOI: 10.1214/18-AOS1731SUPP; .pdf). This supplement file contains the technical materials of the paper and consists of two appendices. Appendix A collects the preliminary definitions and results used in Appendix B, which contains proofs of all the results presented in the main text of the paper.

## REFERENCES

[1] AÏT-SAHALIA, Y. and JACOD, J. (2014). *High-Frequency Financial Econometrics*. Princeton Univ. Press, Princeton, NJ.

[2] ALVAREZ, A., PANLOUP, F., PONTIER, M. and SAVY, N. (2012). Estimation of the instantaneous volatility. *Stat. Inference Stoch. Process.* **15** 27–59. MR2892587

[3] ANDERSEN, T. G., BOLLERSLEV, T. and DOBREV, D. (2007). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and

i.i.d. noise: Theory and testable distributional implications. *J. Econometrics* **138** 125–180. MR2380695

[4] BENNEDSEN, M., HOUNYO, U., LUNDE, A. and PAKKANEN, M. S. (2016). The local fractional bootstrap. Working paper. Available at https://arxiv.org/abs/1605.00868.

[5] BIBINGER, M., JIRAK, M. and VETTER, M. (2017). Nonparametric change-point analysis of volatility. *Ann. Statist.* **45** 1542–1578. MR3670188

[6] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data*: *Methods, Theory and Applications*. Springer, Heidelberg. MR2807761

[7] CHANG, J., QIU, Y., YAO, Q. and ZOU, T. (2017). On the statistical inference for large precision matrices with dependent data. Working paper. Available at https://arxiv.org/abs/1603.06663.

[8] CHEN, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.* **46** 642–678. MR3782380

[9] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448

[10] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818. MR3262468

[11] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461

[12] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Testing many moment inequalities. Working paper. Available at https://arxiv.org/abs/1312.7614v4.

[13] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040

[14] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Process. Appl.* **126** 3632–3651. MR3565470

[15] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963

[16] CHRISTENSEN, K., OOMEN, R. and RENÒ, R. (2016). The drift burst hypothesis. Available at https://ssrn.com/abstract=2842535.

[17] 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. MR2841074

[18] DAVIDSON, R. and FLACHAIRE, E. (2008). The wild bootstrap, tamed at last. *J. Econometrics* **146** 162–169. MR2459651

[19] DAVIDSON, R. and MACKINNON, J. G. (1999). The size distortion of bootstrap tests. *Econometric Theory* **15** 361–376. MR1704228

[20] FAN, J. and WANG, Y. (2008). Spot volatility estimation for high-frequency data. *Stat. Interface* **1** 279–288. MR2476744

[21] GÖTZE, F. and TIKHOMIROV, A. N. (1999). Asymptotic distribution of quadratic forms. *Ann. Probab.* **27** 1072–1098. MR1699003

[22] HAYASHI, T. and KOIKE, Y. (2016). Wavelet-based methods for high-frequency lead-lag analysis. Working paper. Available at https://arxiv.org/abs/1612.01232.

[23] HAYASHI, T. and YOSHIDA, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **11** 359–379. MR2132731

[24] HOFFMANN, M., ROSENBAUM, M. and YOSHIDA, N. (2013). Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli* **19** 426–461. MR3037160

[25] JANSON, S. (1997). *Gaussian Hilbert Spaces*. *Cambridge Tracts in Mathematics* **129**. Cambridge Univ. Press, Cambridge. MR1474726

[26] KANAYA, S. (2017). Uniform convergence rates of kernel-based nonparametric estimators for continuous time diffusion processes: A damping function approach. *Econometric Theory* **33** 874–914. MR3667040

[27] KANAYA, S. and KRISTENSEN, D. (2016). Estimation of stochastic volatility models by nonparametric filtering. *Econometric Theory* **32** 861–916. MR3530455

[28] KATO, K. and KURISU, D. (2017). Bootstrap confidence bands for spectral estimation of Lévy densities under high-frequency observations. Working paper. Available at https://arxiv.org/abs/1705.00586.

[29] KATO, K. and SASAKI, Y. (2016). Uniform confidence bands in deconvolution with unknown error distribution. Working paper. Available at https://arxiv.org/abs/1608.02251.

[30] KOIKE, Y. (2019). Supplement to "Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data." DOI:10.1214/18-AOS1731SUPP.

[31] KRISTENSEN, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory* **26** 60–93. MR2587103

[32] KUSUOKA, S. and YOSHIDA, N. (2000). Malliavin calculus, geometric mixing, and expansion of diffusion functionals. *Probab. Theory Related Fields* **116** 457–484. MR1757596

[33] LEE, S. S. and MYKLAND, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Rev. Financ. Stud.* **21** 2535–2563.

[34] MANCINO, M. E. and RECCHIONI, M. C. (2015). Fourier spot volatility estimator: Asymptotic normality and efficiency with liquid and illiquid high-frequency data. *PLoS ONE* **10** 1–33.

[35] MOSSEL, E., O'DONNELL, R. and OLESZKIEWICZ, K. (2010). Noise stability of functions with low influences: Invariance and optimality. *Ann. of Math.* (2) **171** 295–341. MR2630040

[36] MYKLAND, P. A. and ZHANG, L. (2008). Inference for volatility-type objects and implications for hedging. *Stat. Interface* **1** 255–278. MR2476743

[37] NOURDIN, I. (2013). Lectures on Gaussian approximations with Malliavin calculus. In *Séminaire de Probabilités XLV* (C. Donati-Martin, A. Lejay and A. Rouault, eds.). *Lecture Notes in Math.* **2078** 3–89. Springer, Cham. MR3185909

[38] NOURDIN, I. and PECCATI, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. MR2520122

[39] NOURDIN, I. and PECCATI, G. (2010). Stein's method meets Malliavin calculus: A short survey with new estimates. In *Recent Development in Stochastic Dynamics and Stochastic Analysis* (J. Duan, S. Luo and C. Wang, eds.). *Interdiscip. Math. Sci.* **8** 207–236. World Sci. Publ., Hackensack, NJ. MR2807823

[40] NOURDIN, I. and PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus*: *From Stein's Method to Universality*. *Cambridge Tracts in Mathematics* **192**. Cambridge Univ. Press, Cambridge. MR2962301

[41] NOURDIN, I., PECCATI, G. and REINERT, G. (2009). Second order Poincaré inequalities and CLTs on Wiener space. *J. Funct. Anal.* **257** 593–609. MR2527030

[42] NOURDIN, I., PECCATI, G. and REINERT, G. (2010). Invariance principles for homogeneous sums: Universality of Gaussian Wiener chaos. *Ann. Probab.* **38** 1947–1985. MR2722791

[43] NOURDIN, I., PECCATI, G. and RÉVEILLAC, A. (2010). Multivariate normal approximation using Stein's method and Malliavin calculus. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 45–58. MR2641769

[44] NUALART, D. (2006). *The Malliavin Calculus and Related Topics*, 2nd ed. Springer, Berlin. MR2200233

[45] NUALART, D. and PECCATI, G. (2005). Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* **33** 177–193. MR2118863

[46] PALMES, C. and WOERNER, J. H. C. (2016). A mathematical analysis of the Gumbel test for jumps in stochastic volatility models. *Stoch. Anal. Appl.* **34** 852–881. MR3529199

[47] 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. MR2719873

[48] SABEL, T. (2014). Simultaneous confidence statements about the diffusion coefficient of an Itô-process with application to spot volatility estimation. Ph.D. thesis, Georg-August Univ. School of Science.

[49] SÖHL, J. and TRABS, M. (2016). Adaptive confidence bands for Markov chains and diffusions: Estimating the invariant measure and the drift. *ESAIM Probab. Stat.* **20** 432–462. MR3581829

[50] ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156

[51] ZHANG, X. and CHENG, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675. MR3779697

GRADUATE SCHOOL OF MATHEMATICAL SCIENCES
UNIVERSITY OF TOKYO
3-8-1 KOMABA, MEGURO-KU
TOKYO 153-8914
JAPAN
E-MAIL: kyuta@ms.u-tokyo.ac.jp