

# CONVERGENCE OF COVARIANCE AND SPECTRAL DENSITY ESTIMATES FOR HIGH-DIMENSIONAL LOCALLY STATIONARY PROCESSES

BY DANNA ZHANG<sup>1</sup> AND WEI BIAO WU<sup>2</sup>

<sup>1</sup>*Department of Mathematics, University of California, San Diego, daz076@ucsd.edu*

<sup>2</sup>*Department of Statistics, University of Chicago, wbwu@galton.uchicago.edu*

Covariances and spectral density functions play a fundamental role in the theory of time series. There is a well-developed asymptotic theory for their estimates for low-dimensional stationary processes. For high-dimensional nonstationary processes, however, many important problems on their asymptotic behaviors are still unanswered. This paper presents a systematic asymptotic theory for the estimates of time-varying second-order statistics for a general class of high-dimensional locally stationary processes. Using the framework of functional dependence measure, we derive convergence rates of the estimates which depend on the sample size  $T$ , the dimension  $p$ , the moment condition and the dependence of the underlying processes.

**1. Introduction.** During the past several decades, there has been a well-developed theory for stationary processes. However, the assumption of stationarity may not be valid in many applications. Nonstationary time series analysis has gained popularity in finance, signal processing, neuroscience, meteorology, seismology and many other areas.

As an important class of nonstationary processes, locally stationary processes have attracted considerable attention in the past few years. Different approaches for modelling locally stationary processes have been developed. For example, Dahlhaus (1997, 2000a) adopted a time-varying spectral representation; see also Priestley (1981, 1988). Mallat, Papanicolaou and Zhang (1998) considered processes whose covariance operators are time-varying convolutions. Another method of modelling nonstationarity is to approximate nonstationary processes by piecewise stationary processes; see Adak (1998) and Ombao, von Sachs and Guo (2005). Other notable work includes Nason, von Sachs and Kroisandt (2000), Moulines, Priouret and Roueff (2005) and more recently Zhou (2010) and Vogt (2012); see Dahlhaus (2012) for a comprehensive overview.

Parametric locally stationary processes with time-varying coefficients have been largely studied; see, for example, time-varying AR models (Subba Rao (1970), Dahlhaus (1997), Moulines, Priouret and Roueff (2005)), ARMA models (Grenier (1983), Dahlhaus and Polonik (2009)), ARCH and GARCH models (Dahlhaus and Subba Rao (2006), Dahlhaus and Subba Rao (2007), Hafner and Linton (2010), Fryzlewicz and Subba Rao (2011)). In this paper, we consider nonparametric locally stationary processes. Let  $(X_{t,T})_{t=1}^T$  be the observed sequence generated from the model

$$(1.1) \quad X_{t,T} = X_t(t/T) = G(t/T, \mathcal{F}_t) = (X_{t1,T}, \dots, X_{tp,T})^\top,$$

where  $\mathcal{F}_t = (\dots, \varepsilon_{t-1}, \varepsilon_t)$ ,  $\varepsilon_t$ ,  $t \in \mathbb{Z}$ , are i.i.d. random elements,  $G(\cdot, \cdot) = (g_1(\cdot, \cdot), \dots, g_p(\cdot, \cdot))^\top$  is a  $\mathbb{R}^p$ -valued measurable function such that  $X_t(u) = G(u, \mathcal{F}_t)$  is a well-defined

---

Received November 2018; revised November 2019.

*MSC2020 subject classifications.* Primary 62M10; secondary 62M15.

*Key words and phrases.* High-dimensional time series, locally stationary processes, second-order statistics, convergence rate, Hanson–Wright-type inequalities.

random vector and the uniform stochastic Lipschitz continuity holds: there exists some constant  $\mathcal{K} > 0$  for which

$$(1.2) \quad \max_{1 \leq j \leq p} \|g_j(u, \mathcal{F}_t) - g_j(v, \mathcal{F}_t)\| \leq \mathcal{K}|u - v| \quad \text{for all } u, v \in [0, 1],$$

where, for a random variable  $X$ , the  $L^2$  norm  $\|X\| = (\mathbb{E}X^2)^{1/2}$ . In the scalar case with  $p = 1$ , [Zhou \(2010\)](#) performed nonparametric specification tests for quantile curves under the framework of (1.1). If  $G(u, \cdot)$  does not depend on  $u$ , then (1.1) becomes  $X_t = G(\mathcal{F}_t)$ , which defines a large class of stationary processes. Under this framework, [Chen, Xu and Wu \(2013\)](#) quantified the convergence rates in covariance and precision matrix estimation. [Zhang and Wu \(2017\)](#) derived a Gaussian approximation result for the maximum of the sample mean vector of high-dimensional stationary processes. It is still an open problem on whether an asymptotic theory for the estimates of second-order characteristics including covariance and spectral density matrices can be developed for high-dimensional nonstationary processes via such a general data-generating mechanism.

Estimating second-order characteristics is of fundamental importance in many aspects of statistics. During the past decades, estimation of various cases of second-order statistics has been studied for dependent and nonstationary processes. For example, in finance, [Jacquier, Polson and Rossi \(2004\)](#) concerned multivariate stochastic volatility models parameterized by time-varying covariance matrices with fat tails and correlated errors. In environmental science, [Wikle and Hooten \(2010\)](#) proposed nonlinear spatiotemporal dynamic models to accommodate quadratic interactions between processes which are critical for many geophysical ([Kondrashov et al. \(2005\)](#), [Majda, Abramov and Grote \(2005\)](#)) and ecological ([Hooten and Wikle \(2008\)](#)) processes. In electroencephalographic (EEG) studies, [Prado, West and Krystal \(2001\)](#) considered dynamic regression models with time-varying lag-lead structure to analyze multichannel EEG recordings of scalp electrical potential activity, and [Park, Eckley and Om-bao \(2014\)](#) developed multivariate locally stationary wavelet processes to capture the time-evolving scale-specific cross-dependence between components of the nonstationary signals. In essence, researchers face a number of challenges in solving these real-world problems: (i) nonlinear dynamics of data generating systems, (ii) temporally dependent and nonstationary observations, (iii) non-Gaussian distributions and/or (iv) high-dimensional data.

Motivated by those real-world applications, we shall study properties of estimates of second-order characteristics of a general class of locally stationary processes which can be high-dimensional and non-Gaussian, and lay a theoretical foundation for estimation consistency. In Section 2, we shall introduce the framework of high-dimensional locally stationary processes and some concepts about functional dependence measures that are useful for establishing an asymptotic theory. Section 3 concerns the estimation of time-varying autocovariance matrix functions. Section 4 introduces the nonparametric estimation of time-varying spectral density and coherence matrices. In Section 5, we use the constrained  $\ell_1$  minimization approach to estimate the inverse of the spectral density matrix which can be used to identify the graphical structure for high-dimensional locally stationary processes. Section 6 provides Hanson–Wright-type inequalities for tail probabilities for nonstationary processes with finite polynomial moments. Proofs are given in the Supplementary Material ([Zhang and Wu \(2020\)](#)).

We now introduce some notation. For a random variable  $X$  and  $q \geq 1$ , we define  $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}$ . Denote  $\|X\| = \|X\|_2$  and the operator  $\mathbb{E}_0$  with  $\mathbb{E}_0(X) := X - \mathbb{E}X$ . Define the projection operator  $\mathcal{P}_t \cdot = \mathbb{E}(\cdot | \mathcal{F}_t) - \mathbb{E}(\cdot | \mathcal{F}_{t-1})$  where  $\mathcal{F}_t = (\dots, \varepsilon_{t-1}, \varepsilon_t)$ . For a vector  $v = (v_1, \dots, v_p)^\top$  and  $q \geq 1$ , we define  $|v|_q = (\sum_{j=1}^p |v_j|^q)^{1/q}$  and  $|v|_\infty = \max_j |v_j|$ . For a matrix  $A = (a_{ij})_{i,j=1}^p$ , define the element-wise  $\ell_\infty$  norm  $|A|_\infty = \max_{i,j} |a_{ij}|$  and the matrix  $\ell_1$  norm  $|A|_{\ell_1} = \max_j \sum_i |a_{ij}|$ . Write the  $p \times p$  identity matrix as  $I_p$ . For an interval

$\mathcal{I} \subset \mathbb{R}$ , denote by  $\mathcal{C}^i \mathcal{I}$ ,  $i \in \mathbb{N}$ , be the collection of functions that have  $i$ -th order continuous derivatives on  $\mathcal{I}$ . For two real numbers, set  $x \vee y = \max(x, y)$  and  $x \wedge y = \min(x, y)$ . We use  $C, C_1, C_2, \dots$  to denote positive constants whose values may differ from place to place. A constant with a symbolic subscript is used to emphasize the dependence of the value on the subscript. Throughout the paper, we use  $r, s, t$  to denote time indexes and use  $i, j$  to denote dimension indexes.

**2. High-dimensional locally stationary processes.** Consider the  $p$ -dimensional process  $(X_{t,T})$  generated from the model (1.1). For convenience of notation, we shall abbreviate  $X_{t,T}$  and  $(X_{t_1,T}, \dots, X_{t_p,T})^\top$  as  $X_t$  and  $(X_{t_1}, \dots, X_{t_p})^\top$ , respectively. The stochastic continuity condition (1.2) indicates that  $X_{tj}(u) = g_j(u, \mathcal{F}_t)$  changes smoothly in  $u$ . One has local stationarity in the sense that, for a fixed  $u$ , the nonstationary process  $(X_{tj})$  for  $t$  over the window  $T(u - b) \leq t \leq T(u + b)$  with a small  $b$  can be approximated by the stationary process  $X_{tj}(u) = g_j(u, \mathcal{F}_t)$  in view of

$$\|X_{tj} - X_{tj}(u)\| \leq \mathcal{K} \left| \frac{t}{T} - u \right|,$$

which converges to 0 if  $t/T - u \rightarrow 0$ . In the stationary case in which  $G(\cdot, \cdot)$  does not depend on  $u$ , one can let  $\mathcal{K} = 0$  in (1.2). With condition (1.2), the form (1.1) provides a convenient framework for studying locally stationary processes and covers a large range of nonstationary time series models. In the scalar case with  $p = 1$ , Wiener (1958) studied stationary processes that can be coded by using i.i.d. random variables  $\varepsilon_t$  via a possibly nonlinear function  $G$ ; see also Rosenblatt (1971), Priestley (1988), Tong (1990), Wu (2005), Tsay (2005)) for classes of processes of this form. The representation  $X_t = G(\mathcal{F}_t)$  also includes recursive model of the form  $X_t = G(X_{t-1}, \varepsilon_t)$ , which includes Markov chain models and nonlinear autoregressive models such as threshold autoregressive models, autoregressive models with conditional heteroscedasticity and exponential autoregressive models. By allowing the data-generating function  $G$  to change flexibly over time  $u$ , it extends a large number of existing stationary processes into their nonstationary counterparts in a natural way.

To develop an asymptotic theory for estimators of time-varying second-order characteristics, we need to introduce appropriate dependence measures. Assume that  $\max_{1 \leq j \leq p} \sup_{u \in [0,1]} \|g_j(u, \mathcal{F}_0)\|_q < \infty$  for some  $q \geq 1$ . Let  $\varepsilon'_s, \varepsilon_t, s, t \in \mathbb{Z}$ , be i.i.d. random variables. For  $t \geq 0$  and  $1 \leq j \leq p$ , we define the element-wise functional dependence measures

$$(2.1) \quad \delta_{t,q,j} = \sup_{u \in [0,1]} \|g_j(u, \mathcal{F}_t) - g_j(u, \mathcal{F}_{t,\{0\}})\|_q,$$

where  $\mathcal{F}_{t,\{l\}} = (\dots, \varepsilon_{l-1}, \varepsilon'_l, \varepsilon_{l+1}, \dots, \varepsilon_t)$  is a coupled version of  $\mathcal{F}_t$  with  $\varepsilon_l$  in  $\mathcal{F}_t$  replaced by  $\varepsilon'_l$ , and the uniform or  $L^\infty$  functional dependence measure

$$(2.2) \quad \omega_{t,q} = \sup_{u \in [0,1]} \|G(u, \mathcal{F}_t) - G(u, \mathcal{F}_{t,\{0\}})\|_\infty \|g_j\|_q.$$

Note that  $\mathcal{F}_{t,\{0\}} = \mathcal{F}_t$  if  $t < 0$ . Hence,  $\delta_{t,q,j} = 0$  and  $\omega_{t,q} = 0$  for  $t < 0$ . Wu (2005) introduced a functional dependence measure for stationary processes in which the data-generating mechanism  $g_j$  does not vary with time  $u$ . In our setting, the quantity  $\delta_{t,q,j}$  measures the dependence of  $g_j(u, \mathcal{F}_t)$  on the single input  $\varepsilon_0$  over  $u \in [0, 1]$ , which can be viewed as the uniform dependence measure with lag  $t$  for locally stationary processes.

Equipped with the dependence measures in (2.1) and (2.2), we define in the following the dependence adjusted norms (d.a.n.):

$$(2.3) \quad \|X_{\cdot j}\|_{q,\alpha} = \sup_{m \geq 0} (m + 1)^\alpha \Delta_{m,q,j}, \alpha \geq 0 \quad \text{where } \Delta_{m,q,j} = \sum_{t=m}^{\infty} \delta_{t,q,j},$$

$$(2.4) \quad \| |X_{\cdot}|_{\infty} \|_{q,\alpha} = \sup_{m \geq 0} (m+1)^\alpha \Omega_{m,q}, \alpha \geq 0 \quad \text{where } \Omega_{m,q} = \sum_{t=m}^{\infty} \omega_{t,q}.$$

We use  $\alpha$  to depict the decay rate of the cumulative (tail) dependence measure  $\Delta_{m,q,j} = \sum_{t=m}^{\infty} \delta_{t,q,j}$  by noting that  $\Delta_{m,q,j} \leq \|X_{\cdot j}\|_{q,\alpha} (m+1)^{-\alpha}$  for all  $m \in \mathbb{N}$ . In this sense, it quantifies the strength of temporal dependence: larger  $\alpha$  implies faster decay of tail dependence measures, and thus weaker temporal dependence. We can interpret the quantity  $\|X_{\cdot j}\|_{q,\alpha}$  as the  $q$ th moment by taking dependence into account. Elementary calculations show if  $X_{tj}, t \in \mathbb{Z}$ , are i.i.d., then  $\|X_{ij}\|_q \leq \|X_{\cdot j}\|_{q,\alpha} \leq 2\|X_{ij}\|_q$ , suggesting that the dependence adjusted norm is equivalent to the classical  $L^q$  norm. Due to temporal dependence, it may happen that  $\max_t \|X_{tj}\|_q < \infty$  while  $\|X_{\cdot j}\|_{q,\alpha} = \infty$ . For example, if  $\delta_{t,q,j} \asymp t^{-\beta}, \beta > 1$ , then  $\|X_{\cdot j}\|_{q,\alpha} = \infty$  if  $\alpha > \beta - 1$  and  $\|X_{\cdot j}\|_{q,\alpha} < \infty$  if  $\alpha \leq \beta - 1$ .

The d.a.n.  $\|X_{\cdot j}\|_{q,\alpha}$  accounts for temporal dependence for the component process  $(X_{tj})_{t \in \mathbb{Z}}$ . To adjust for dimensionality, we further define respectively the overall and the uniform dependence adjusted norms

$$(2.5) \quad \Theta_{q,\alpha} = \left( \sum_{j=1}^p \|X_{\cdot j}\|_{q,\alpha}^{q/2} \right)^{2/q}, \quad \Phi_{q,\alpha} = \max_{1 \leq j \leq p} \|X_{\cdot j}\|_{q,\alpha}.$$

The quantities  $\| |X_{\cdot}|_{\infty} \|_{q,\alpha}, \Theta_{q,\alpha}$  and  $\Phi_{q,\alpha}$  provide a concise and natural measure of dependence which can effectively account for high dimensionality and temporal dependence. They will be imposed in our theorems. It can be easily seen that  $\Phi_{q,\alpha} \leq \| |X_{\cdot}|_{\infty} \|_{q,\alpha} \leq \Theta_{q,\alpha}$ . They may be unbounded functions in terms of the dimension  $p$ .

EXAMPLE 2.1 (Time-varying nonlinear vector autoregressive model). Let  $\varepsilon_t$  be i.i.d. and consider the  $p$ -dimensional process  $X_{t,T}^\circ, t = 1, \dots, T$ , which is generated from the time-varying recursive model

$$(2.6) \quad X_{t,T}^\circ = R(t/T, X_{t-1,T}^\circ, \varepsilon_t),$$

where  $\sup_{0 \leq u \leq 1} \| |R(u, x_0, \varepsilon_0)|_{\infty} \|_q < \infty$  for some  $q \geq 2$  and  $x_0$  and it satisfies

$$(2.7) \quad \chi := \sup_{u \in [0,1]} \sup_{x \neq y} \frac{\| |R(u, x, \varepsilon_0) - R(u, y, \varepsilon_0)|_{\infty} \|_q}{|x - y|_{\infty}} < 1.$$

The tvVAR(1) model  $X_{t,T}^\circ = A(t/T)X_{t-1,T}^\circ + \varepsilon_t$  for some transition matrix  $A(\cdot)$  is a special case of (2.6). It also includes other time-varying parametric models such as tvVAR(1) and tvTAR(1); see, for example, [Dahlhaus and Subba Rao \(2006\)](#) and [Zhou and Wu \(2009\)](#) for low-dimensional processes. We shall show that the process defined by the recursion (2.6) can be well approximated by our (1.1). For fixed  $u \in [0, 1]$ , the stationary approximation in this case is given by

$$(2.8) \quad X_t(u) = R(u, X_{t-1}(u), \varepsilon_t).$$

By the arguments of Theorem 2 in [Wu and Shao \(2004\)](#), for any  $u \in [0, 1]$ , (2.8) admits a unique stationary solution and iterations of (2.8) lead to  $X_t(u) = G(u, \mathcal{F}_t)$ . Assume  $\mathcal{M} := \sup_{u \in [0,1]} \| |K(X_t(u))| \|_q < \infty$  where

$$(2.9) \quad K(x) = \sup_{u \neq v} \frac{\| |R(u, x, \varepsilon_0) - R(v, x, \varepsilon_0)|_{\infty} \|_q}{|u - v|}.$$

By generalizing Lemma 4.5 in [Dahlhaus, Richter and Wu \(2019\)](#) to the vector case, we can obtain  $\| |X_t(u) - X_t(v)|_{\infty} \|_q \leq \mathcal{M}|u - v|/(1 - \chi)$  and

$$(2.10) \quad \sup_{t=1, \dots, T} \| |X_{t,T}^\circ - X_t(t/T)|_{\infty} \|_q \leq \mathcal{M} \frac{\chi}{(1 - \chi)^2} \cdot T^{-1}.$$

Note that the approximation error in (2.10) cannot be avoided; see equation (49) in Dahlhaus (2012) for the tvAR(1) case. For the more general version of (2.6),  $X_{t,T}^\circ = R(t/T, X_{t-1,T}^\circ, X_{t-2,T}^\circ, \dots, X_{t-d,T}^\circ, \varepsilon_t)$ ,  $d \geq 2$ , we can compute the approximation error similarly, as stated in Lemma A.1.

EXAMPLE 2.2 (Time-varying vector linear processes). Let  $\varepsilon_{tj}$ ,  $t, j \in \mathbb{Z}$ , be i.i.d. random variables with mean 0, variance 1 and finite  $q$ th moment  $\mu_q := \mathbb{E}(|\varepsilon_{tj}|^q) < \infty$ ,  $q > 2$ . Let  $A_m(u) = (a_{m,ij}(u))_{i,j=1}^p$  be  $p \times p$  matrices with real entries such that  $a_{m,ij}(u) \in C^1[0, 1]$ ,  $m \geq 0$ ,  $1 \leq i, j \leq p$ , and  $\sup_{u \in [0,1]} \sum_{m=0}^\infty \text{tr}[A_m(u)A_m(u)^\top] < \infty$ . Write  $\varepsilon_t = (\varepsilon_{t1}, \dots, \varepsilon_{tp})^\top$ . By Kolmogorov’s three series theorem, the  $p$ -dimensional linear process

$$(2.11) \quad X_t(u) = \sum_{m=0}^\infty A_m(u)\varepsilon_{t-m}$$

is well defined and the assumptions on  $A_m(u)$  ensure the local stationarity of the process  $X_t(t/T)$ . Let  $A_{m,j \cdot}(u)$  and  $A_{m \cdot j}(u)$  be the  $j$ th row and  $j$ th column of  $A_m(u)$ . By Lemma D.3, the element-wise and  $L^\infty$  functional dependence measures can be computed by

$$\begin{aligned} \delta_{t,q,j} &= \sup_{u \in [0,1]} \|A_{t,j \cdot}(u)\varepsilon_0\|_q \leq C_q \sup_{u \in [0,1]} |A_{t,j \cdot}(u)|_2 \mu_q^{1/q}, \\ \omega_{t,q} &= \sup_{u \in [0,1]} \| |A_t(u)\varepsilon_0|_\infty \|_q \\ &\leq C_q (1 \vee \log p)^{1/2} \sup_{u \in [0,1]} \left( \sum_{j=1}^p |A_{t \cdot j}(u)|_\infty^2 \right)^{1/2} p^{1/q} \mu_q^{1/q}, \end{aligned}$$

since  $\| |\varepsilon_0|_\infty \|_q \leq p^{1/q} \mu_q^{1/q}$ . If there exist  $c > 1$  and  $K_1, K_2 > 0$ , such that for all  $t \geq 0$  and  $1 \leq j \leq p$ ,  $\sup_{u \in [0,1]} |A_{t,j \cdot}(u)|_2 \leq K(t+1)^{-c}$  and  $\sup_{u \in [0,1]} (\sum_{j=1}^p |A_{t \cdot j}(u)|_\infty^2)^{1/2} \leq K_2(t+1)^{-c}$  hold, then with  $\alpha = c - 1$ , we have

$$\begin{aligned} \Theta_{q,\alpha} &\leq C_{q,\alpha} K_1 p^{2/q} \mu_q^{1/q}, & \Phi_{q,\alpha} &\leq C_{q,\alpha} K_1 \mu_q^{1/q}, \\ \| |X_\cdot|_\infty \|_{q,\alpha} &\leq C'_{q,\alpha} K_2 (1 \vee \log p)^{1/2} p^{1/q} \mu_q^{1/q}, \end{aligned}$$

where the constants  $C_{q,\alpha}, C'_{q,\alpha}$  both depend on  $q$  and  $\alpha$  only. It also applies to the process with a Lipschitz continuous transform of  $X_t(u)$  in (2.11):  $Y_t(u) = (Y_{t1}(u), \dots, Y_{tp}(u))^\top$ , where  $Y_{tj}(u) = g_j(X_{tj}(u))$  and  $g_j(\cdot)$  are Lipschitz continuous with uniformly bounded Lipschitz constants.

**3. Estimation of autocovariance matrix functions.** Autocovariances play an important role in almost every aspect of time series analysis. For zero-mean stationary processes  $X_t = G(\mathcal{F}_t)$  where the function  $G(u, \cdot)$  in (1.1) does not depend on  $u$ , we shall estimate the autocovariance matrices  $\Gamma_l = \mathbb{E}(X_0 X_l^\top)$  based on the observations  $X_1, \dots, X_T$  by

$$(3.1) \quad \hat{\Gamma}_l = \frac{1}{T} \sum_{t=l+1}^T X_{t-l} X_t^\top \quad \text{for } l \geq 0,$$

and  $\hat{\Gamma}_l = \hat{\Gamma}_{-l}^\top$  for  $l < 0$ . For the locally stationary process (1.1) of mean zero, the time-varying autocovariance matrix with lag  $l$  is defined by

$$(3.2) \quad \Gamma_l(u) = \mathbb{E}(X_0(u) X_l(u)^\top) \quad \text{where } X_t(u) = G(u, \mathcal{F}_t).$$

For fixed  $u \in (0, 1)$ , by local stationarity,  $X_{Tt} \approx g_j(u, \mathcal{F}_t)$  for  $t$  close to  $Tu$ . Thus we can use observations  $X_t$  with  $t$  close to  $Tu$  to construct the estimator  $\hat{\Gamma}_l(u)$ . Specifically, let  $b_T$  be the bandwidth

$$T_1(u) = \lfloor Tu \rfloor - \lfloor Tb_T \rfloor + 1, \quad T_2(u) = \lfloor Tu \rfloor + \lfloor Tb_T \rfloor$$

and  $M = 2\lfloor Tb_T \rfloor$  be the window width. For  $u \in [b_T, 1 - b_T]$  and  $0 \leq l < M$ , a natural estimator of  $\Gamma_l(u)$  from the sample  $\{X_t, t = T_1(u), \dots, T_2(u)\}$  is

$$(3.3) \quad \hat{\Gamma}_l(u) = \frac{1}{M} \sum_{r=l+T_1(u)}^{T_2(u)} X_{r-l} X_r^\top,$$

and  $\hat{\Gamma}_l(u) = \hat{\Gamma}_{-l}^\top(u)$  for  $l < 0$ .

We shall study the maximum deviation over the range  $0 \leq l < m$  with  $m \leq \lfloor M^\beta \rfloor$  for some  $0 \leq \beta < 1$ , that is,

$$(3.4) \quad \psi_T^* := \max_{0 \leq l < m} \sup_{u \in [b_T, 1-b_T]} |\hat{\Gamma}_l(u) - \Gamma_l(u)|_\infty,$$

or in the stationary case

$$\tilde{\psi}_T^* := \max_{0 \leq l < m} |\hat{\Gamma}_l - \Gamma_l|_\infty.$$

For univariate stationary processes with  $p = 1$ , uniform convergence of autocovariance estimates is closely related to the estimation of orders of ARMA processes or linear systems in general. The pioneering works in this direction were given by E. J. Hannan and his collaborators; see, for example, Hannan (1974) and An, Chen and Hannan (1982). Readers can find a summary of those works and references in Section 5.3 of Hannan and Deistler (1988). Giurcanu and Spokoiny (2004) obtained an upper bound of  $\max_{0 \leq l < m} |\hat{\Gamma}_l - \Gamma_l|$  for Gaussian stationary processes and also extended to the locally stationary case (cf. Propositions 2.3 and 3.5 therein). More recently, Xiao and Wu (2014) considered maximum deviations for sample autocovariances of univariate stationary processes.

Since the process  $X_t$  can be nonlinear, nonstationary, non-Gaussian and high-dimensional, it can be quite involved to derive an upper bound for  $\psi_T^*$  or  $\tilde{\psi}_T^*$ . Theorem 3.1 below provides a nonasymptotic bound of the stochastic part for locally stationary processes

$$(3.5) \quad \psi_T := \max_{0 \leq l < m} \sup_{u \in [b_T, 1-b_T]} |\hat{\Gamma}_l(u) - \mathbb{E}\hat{\Gamma}_l(u)|_\infty$$

with the existence of finite  $q$ th moment of the underlying process, while Theorem 3.2 concerns Gaussian processes. In our setting, with the framework of functional dependence measures, it turns out that we can have a close form of the upper bound in the form of (3.6) or (3.7). The convergence rate depends in a subtle way on the temporal dependence characterized by  $\alpha$  [cf. (2.3) and (2.4)], the dependence adjusted norms  $\|X\|_\infty$ ,  $\|X\|_q$ ,  $\Theta_{q,\alpha}$  and  $\Phi_{q,\alpha}$ , the sample size  $T$  and the dimension  $p$ . We present the results for the stationary case in Proposition 3.3.

**THEOREM 3.1.** *Assume that  $\mathbb{E}(X_t) = 0$  and  $\Theta_{q,\alpha} < \infty$  for some  $q > 4$  and  $\alpha > 0$ . Let  $b_T$  be the bandwidth and  $M = 2\lfloor Tb_T \rfloor$ . Assume  $M \leq T$  and  $m \leq \lfloor M^\beta \rfloor$  for some  $0 \leq \beta < 1$ . Let  $\ell = 1 \vee \log p$ . Then there exist an absolute constant  $C$ , constant  $C_\alpha$  only depending on  $\alpha$  and constant  $C_{q,\alpha}$  only depending on  $q$  and  $\alpha$  such that for any  $x > 0$ ,*

$$(3.6) \quad \mathbb{P}(\psi_T \geq x) \leq \frac{C_{q,\alpha} T H_{M,m}(\ell \|X\|_\infty \|X\|_q \wedge \Theta_{q,\alpha})^q}{(Mx)^{q/2}} + CTp^2 \exp\left(-\frac{Mx^2}{C_\alpha \Phi_{4,\alpha}^4}\right),$$

where  $H_{M,m} = m^{q/4}(\log M)^{q+1} + M^{q/4-\alpha q/2-1}m$  for  $\alpha < 1/2 - 2/q$ , and  $H_{M,m} = m^{q/4}(\log M)^{q+1}$  for  $\alpha > 1/2 - 2/q$ .

Despite the complicated nature of our problem which involves temporal dependencies, cross-sectional dependencies and possibly nonstationarity, one essentially only needs to deal with quantities  $\|X \cdot\|_{q,\alpha}$ ,  $\Theta_{q,\alpha}$  and  $\Phi_{4,\alpha}$  in our nonasymptotic bound in Theorem 3.1. They concisely quantify measure of dependence which can naturally account for high dimensionality. They are also used in other theorems in the following sections.

For the term  $\ell\|X \cdot\|_{q,\alpha} \wedge \Theta_{q,\alpha}$  in (3.6), consider the case in which each component process is balanced with a similar order of d.a.n., that is, there exist constants  $C_1, C_2 > 0$  such that  $C_1 \leq \|X \cdot\|_{q,\alpha} \leq C_2$  for all  $j$ . Then  $\Theta_{q,\alpha} \asymp p^{2/q}$ . Since  $\|X \cdot\|_{q,\alpha} \leq (\sum_{j=1}^p \|X \cdot\|_{q,\alpha}^q)^{1/q} \asymp p^{1/q}$ , the order of  $\ell\|X \cdot\|_{q,\alpha}$  is smaller than that of  $\Theta_{q,\alpha}$ . Then the term  $\ell\|X \cdot\|_{q,\alpha} \wedge \Theta_{q,\alpha} \asymp \ell p^{1/q}$ .

**THEOREM 3.2.** *Let  $(X_t)$  be a Gaussian process of form (1.1), which satisfies  $\mathbb{E}(X_t) = 0$  and  $\Phi_{2,0} < \infty$ . Let  $b_T$  be the bandwidth and  $M = 2\lfloor T b_T \rfloor$ . Assume  $M \leq T$  and  $m \leq \lfloor M^\beta \rfloor$  for some  $0 \leq \beta < 1$ . Then there exists an absolute constant  $C > 0$  such that for any  $x > 0$ ,*

$$(3.7) \quad \mathbb{P}(\psi_T > x) \leq 2Tm p^2 \exp\left[-C \min\left(\frac{Mx^2}{\Phi_{2,0}^4}, \frac{Mx}{\Phi_{2,0}^2}\right)\right].$$

**PROPOSITION 3.3.** *Consider the zero-mean stationary process  $X_t = G(\mathcal{F}_t)$ . Let  $\hat{\Gamma}_l$  be the autocovariance matrix estimator given in (3.1). Define*

$$(3.8) \quad \tilde{\psi}_T = \max_{0 \leq l < m} |\hat{\Gamma}_l - \mathbb{E}\hat{\Gamma}_l|_\infty.$$

(i) *Under the assumptions of Theorem 3.1, we have*

$$(3.9) \quad \mathbb{P}(\tilde{\psi}_T \geq x) \leq \frac{C_{q,\alpha} T H_{T,m}^* (\ell\|X \cdot\|_{q,\alpha} \wedge \Theta_{q,\alpha})^q}{(Tx)^{q/2}} + Cmp^2 \exp\left(-\frac{Tx^2}{C_\alpha \Phi_{4,\alpha}^4}\right),$$

where  $H_{T,m}^* = m^{q/4}$  for  $\alpha > 1/2 - 2/q$  and  $H_{T,m}^* = m^{q/4} + T^{q/4-\alpha q/2-1}m$  for  $\alpha < 1/2 - 2/q$ . (ii) *Under the assumptions of Theorem 3.2, we have*

$$(3.10) \quad \mathbb{P}(\tilde{\psi}_T > x) \leq 2mp^2 \exp\left[-C \min\left(\frac{Tx^2}{\Phi_{2,0}^4}, \frac{Tx}{\Phi_{2,0}^2}\right)\right].$$

For stationary processes, the bounds of  $\tilde{\psi}_T$  in Proposition 3.3 can be useful for nonlinear spectra estimation (cf. Paparoditis and Politis (2012)). For one-dimensional linear processes, Jirak (2011) proved the Gumbel convergence of  $\max_{0 \leq l < m} |\hat{\Gamma}_l - \mathbb{E}\hat{\Gamma}_l|$  for  $m$  growing at most logarithmic speed. And Xiao and Wu (2014) considered general stationary processes within our framework and relaxed the growth speed to be  $m = O(T^\beta)$  for some  $0 \leq \beta < 1$ . Our result (3.9) allows the same wide range and the same sharp bound when  $p = 1$  as the latter, that is,

$$\max_{0 \leq l < m} |\hat{\Gamma}_l - \mathbb{E}\hat{\Gamma}_l| = O_{\mathbb{P}}\left(\sqrt{\frac{\log T}{T}}\right)$$

with  $\beta < \min(1 - 4/q, \alpha q/2)$  as the requirement.

We now conduct a detailed discussion about how the different factors take effect on the convergence rate.

*Effect of local stationarity.* The stationary case admits a sharper bound than the locally stationary case, as it does not involve the maximum over the time  $u$ . For example, in comparison with (3.6) by letting  $M = T$ , (3.9) excludes the additional  $(\log T)^{q+1}$  in the polynomial term and has a slightly sharper exponential term by multiplying with  $mp^2$  instead of  $Tp^2$ .

*Effect of moment condition.* Theorem 3.1 is a Nagaev-type inequality and it indicates two types of bounds for the tail probability: polynomial tail and sub-Gaussian type tail, which respectively induce the two orders below

$$(3.11) \quad \mathcal{H}_1 = \frac{(TH_{M,m})^{2/q}}{M} (\ell \| |X \cdot|_{\infty} \|_{q,\alpha} \wedge \Theta_{q,\alpha})^2, \quad \mathcal{H}_2 = \sqrt{\frac{\log(pT)}{M}} \Phi_{4,\alpha}^2.$$

We have  $\psi_T = O_{\mathbb{P}}(\mathcal{H}_1 + \mathcal{H}_2)$ . For large (resp., small)  $x$ , the polynomial (resp., the sub-Gaussian) tail dominates. As a comparison, Theorem 3.2 admits an exponential bound for Gaussian processes and it implies  $\psi_T = O_{\mathbb{P}}(\mathcal{H}_3)$  where

$$(3.12) \quad \mathcal{H}_3 = \sqrt{\frac{\log(pT)}{M}} \Phi_{2,0}^2 \vee \frac{\log(pT)}{M} \Phi_{2,0}^2.$$

Looking into the rates in two cases, if  $\log(pT) \ll M$ ,  $\Phi_{4,\alpha} \asymp 1$  and  $\Phi_{2,0} \asymp 1$ , it holds that  $\mathcal{H}_2 \asymp \mathcal{H}_3$ , thus  $\mathcal{H}_1$  is the additional term characterized by the moment order  $q$  if each component process only has finite  $q$ th moment rather than the Gaussianity.

*Effect of dependence.* If  $X_i$  are i.i.d., then  $\delta_{t,q,j} = 0$  and  $\omega_{t,q} = 0$  for all  $t \geq 1$ ,  $\delta_{0,q,j} = \|X_{0j} - X'_{0j}\|_q$  and  $\omega_{0,q} = \| |X_0 - X'_0|_{\infty} \|_q$ , where  $X_0$  and  $X'_0$  are i.i.d. The quantities  $\| |X \cdot|_{\infty} \|_{q,\alpha}$ ,  $\Theta_{q,\alpha}$  and  $\Phi_{q,\alpha}$  in Proposition 3.3 thus reduce to  $\omega_{0,q}$ ,  $(\sum_{j=1}^p \delta_{0,q,j}^{q/2})^{2/q}$  and  $\max_{j \leq p} \delta_{0,q,j}$ , respectively. To account for temporal dependence, we need to use the dependence adjusted norms  $\| |X \cdot|_{\infty} \|_{q,\alpha}$ ,  $\Theta_{q,\alpha}$  and  $\Phi_{q,\alpha}$ , which are generally larger than the ones under independence.

**COROLLARY 3.4.** *Let  $\psi_T^*$  be the maximum deviation defined in (3.4). Let condition (1.2) be satisfied. Recall  $\mathcal{H}_1, \mathcal{H}_2$  defined in (3.11) and  $\mathcal{H}_3$  defined in (3.12). (i) Under the assumptions of Theorem 3.1, we have*

$$(3.13) \quad \psi_T^* = O_{\mathbb{P}}(\mathcal{H}_1 + \mathcal{H}_2 + \Delta_{\psi}),$$

where

$$\Delta_{\psi} = \frac{\mathcal{K}M}{T} \Phi_{2,0} + \frac{1 + m^{-\alpha+1}}{M} \Phi_{2,0} \Phi_{2,\alpha}.$$

(ii) Under the assumptions of Theorem 3.2,

$$(3.14) \quad \psi_T^* = O_{\mathbb{P}}(\mathcal{H}_3 + \Delta_{\psi}).$$

For the bias  $\max_{0 \leq l < m} |\mathbb{E} \hat{\Gamma}_l - \Gamma_l|_{\infty}$  in the stationary case, the first term in  $\Delta_{\psi}$  should disappear in view of  $\mathcal{K} = 0$ . Consequently, it follows that  $\max_{0 \leq l < m} |\mathbb{E} \hat{\Gamma}_l - \Gamma_l|_{\infty} = O((1 + m^{-\alpha+1}) \Phi_{2,0} \Phi_{2,\alpha} / T)$ .

*Effect of the dimension  $p$ .* The terms  $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$  all involve the dimension  $p$ , where the former depends on  $p$  via the dependence norm  $\ell \| |X \cdot|_{\infty} \|_{q,\alpha} \wedge \Theta_{q,\alpha}$  and the latter two depend on  $p$  logarithmically. We further investigate how  $p$  takes effect analytically by examining the case where  $\beta = 0$ , and thus  $m = 1$ . We focus on the sample covariance matrix with lag  $l = 0$  only and consider the case with the existence of finite  $q$ th moments. Assume that  $\| |X \cdot|_{\infty} \|_{q,\alpha} \asymp$

1 and  $\|X_{\cdot}|_{\infty}\|_{q,\alpha} \asymp p^\tau$  for some  $\tau \geq 0$ . In the strongest cross-sectional dependence case with  $X_{tj} = a_j X_{t1}$  and  $C_1 \leq |a_j| \leq C_2$  for some constants  $C_1, C_2 > 0$ , we have  $\tau = 0$ . We can have

$$(3.15) \quad \mathcal{H}_1 \asymp \frac{T^{2/q} H_M^* \ell^2 p^{2\tau}}{M}, \quad \mathcal{H}_2 \asymp \sqrt{\frac{\log(pT)}{M}},$$

where  $H_M^* = (\log M)^{2+2/q}$  if  $\alpha > 1/2 - 2/q$  and  $H_M^* = M^{1/2-2/q-\alpha}$  for  $\alpha < 1/2 - 2/q$ . If we choose a relatively small window width, say  $M \leq T^{4/q}$ ,  $\mathcal{H}_1$  is always the dominant order. As a natural requirement of consistency, we need  $M/H_M^* \gg T^{2/q}$  and

$$(1 \vee \log p) p^\tau = o\left(\frac{M}{T^{2/q} H_M^*}\right).$$

As can be seen, if  $M/(T^{2/q} H_M^*) \asymp T^c$  for some  $0 < c < 1 - 2/q$ , we can allow ultra-high dimension  $p$  with  $\log p = o(T^c)$  for  $\tau = 0$  and polynomial increase with  $T$  which should satisfy  $p = o((T^c/\log T)^{1/\tau})$  for  $\tau > 0$ . Furthermore, a wider range of  $p$  is allowed if the temporal dependence is weaker in view of  $H_M^*$  which is nonincreasing with  $\alpha$ .

The larger the window width  $M$  we choose, the wider the range of  $p$  is allowed for consistency of the stochastic part. In view of Corollary 3.4, we need to balance the bias term  $\Delta_\psi$ . Below is a discussion on the choice of  $M$ .

*The choice of the window width  $M$ .* Regarding the choice of  $M$ , there is a trade-off between the deviation bound  $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$  and the bias order  $\Delta_\psi$ . Consider the case where the process only has finite  $q$ th moments,  $m = 1$  and  $\|X_{\cdot j}\|_{q,\alpha} \asymp 1$ , and recall the orders of  $\mathcal{H}_1$  and  $\mathcal{H}_2$  in (3.15). We examine the strong temporal dependence case  $\alpha < 1/2 - 2/q$ . To minimize  $\mathcal{H}_1 + \mathcal{H}_2 + \Delta_\psi$ ,  $M$  is chosen to be

$$M \asymp \max\{(T^{2/q} \ell^2 p^{2\tau} / \log(pT))^{1/(2/q+\alpha)}, (T^2 (\log(pT)))^{1/3}\},$$

which is nonincreasing in  $\alpha$ , nondecreasing in  $\tau$  and increasing with  $p$ . That is to say, a larger window width is required if the process has stronger temporal dependence and larger dimension. For Gaussian processes, we consider the case where  $\log(pT) \ll T$  and  $\Phi_{2,0} \asymp 1$ . To minimize  $\mathcal{H}_3 + \Delta_\psi$ , the optimal  $M$  satisfies

$$M \asymp (T^2 \log(pT))^{1/3}.$$

REMARK 1. Using the idea of local smoothing, one can consider the following weighted version of (3.3) with a kernel:

$$\hat{\Gamma}_l(u) = \frac{1}{M} \sum_{r=l+T_1(u)}^{T_2(u)} K\left(\frac{r - \lfloor Tu \rfloor}{M}\right) X_{r-l} X_r^\top,$$

where  $K(\cdot)$  is symmetric, nonnegative and differentiable with bounded derivatives on the support  $(-1/2, 1/2)$ , and  $K(0) = 1$ . A careful check of the proofs of Theorems 3.1 and 3.2 suggests that they still hold accordingly.

REMARK 2. Consider the time-varying recursive model (2.6) in Example 2.1. For ease of notation, we write  $X_s^\circ$  for  $X_{s,T}^\circ$ . Assume  $\mathbb{E}X_s^\circ = 0$  and let  $\Gamma_{s,t}^\circ = \mathbb{E}(X_s^\circ X_t^{\circ\top})$ . Since  $\Gamma_{s,t}^\circ \approx \Gamma_{s+r,t+r}^\circ$  for small  $r$ , we can estimate  $\Gamma_{s,t}^\circ$  by a similar form of (3.3)

$$\hat{\Gamma}_{s,t}^\circ = \frac{1}{M} \sum_{r=-\lfloor Tb_T \rfloor + 1}^{\lfloor Tb_T \rfloor} X_{s+r}^\circ X_{t+r}^{\circ\top},$$

where  $M = 2\lfloor Tb_T \rfloor \leq T$ . And we can draw a similar conclusion as Theorem 3.1 for  $\psi_T^\circ = \max_{|s-t| \leq m} |\hat{\Gamma}_{s,t}^\circ - \mathbb{E}\hat{\Gamma}_{s,t}^\circ|$ . Assume for convenience the starting point  $X_0^\circ = 0$  and  $\mathcal{U} := \sup_{0 \leq u \leq 1} \|R(u, 0, \varepsilon_0)\|_q < \infty$ . By (2.7), we have  $\|X_t^\circ - R(t/T, 0, \varepsilon_t)\|_p \leq \chi \|X_{t-1}^\circ\|_p$ , implying  $\|X_t^\circ\|_p \leq \mathcal{U} + \chi \|X_{t-1}^\circ\|_p$ , and hence  $\|X_t^\circ\|_p \leq \mathcal{U}/(1 - \chi)$  by recursion. We now compute the uniform functional dependence measure. Write  $X_t^\circ = R_{\varepsilon_t}^t \circ R_{\varepsilon_{t-1}}^{t-1} \circ \dots \circ R_{\varepsilon_1}^1(0) =: H_t(\mathcal{F}_t)$ ,  $t \geq 1$ , where the map  $R_\varepsilon^t(\cdot) = R(t/T, \cdot, \varepsilon)$  and  $H_t$  is a measurable function consisting of composites of  $R_\varepsilon^t$ . Recall (2.1) that  $\mathcal{F}_{t,\{l\}}$  is a coupled version of  $\mathcal{F}_t$  with  $\varepsilon_l$  in  $\mathcal{F}_t$  replaced by  $\varepsilon'_l$ . For  $k \geq 0$ , write  $X_{t,\{t-k\}}^\circ = H_t(\mathcal{F}_{t,\{t-k\}})$ , and  $L^\infty$  dependence measure for the process  $(X_t^\circ)$

$$\omega_{k,q}^\circ := \sup_t \|H_t(\mathcal{F}_t) - H_t(\mathcal{F}_{t,\{t-k\}})\|_q \leq 2\chi^k \|X_{t-k}^\circ\|_q \leq \frac{2\chi^k \mathcal{U}}{1 - \chi}$$

by recursion (2.7). Since  $\omega_{k,q}^\circ$  decays geometrically in  $k$ , we can simply let  $\alpha = 1$ . Then the uniform dependence adjusted norm  $\|X_\cdot\|_{q,\alpha} \leq c_\chi \mathcal{U}$  with  $c_\chi = 2 \max_{m \geq 0} (m + 1) \sum_{t=m}^\infty \chi^t / (1 - \chi)$ , and  $\Phi_{4,\alpha} \leq c_\chi \mathcal{U}$ . By Theorem 3.1, there exists constant  $C, C_\chi, C_{q,\chi}$  such that

$$\mathbb{P}(\psi_T^\circ \geq x) \leq \frac{C_{q,\chi} T m^{q/4} (\log M)^{q+1} \ell^q \mathcal{U}^q}{(Mx)^{q/2}} + CT p^2 \exp\left(-\frac{Mx^2}{C_\chi \mathcal{U}^4}\right).$$

**4. Spectral density and coherence matrix.** Spectral analysis is a fundamental tool to gain insights into the cyclical behavior of time series. The spectrum provides an adequate description of the frequency domain characteristics of stationary processes. Estimation of spectral density has been extensively studied in the univariate stationary case; see, for example, Anderson (1971), Priestley (1981), Rosenblatt (1985), among many others. Coherence, also known as the time series analogue in the frequency domain of the standard correlation coefficient, measures the linear relationship between a pair of time series as a function of frequency; see, for example, Brillinger (1975) and Brockwell and Davis (1991). Since nonstationary data with time-varying structural changes are increasingly common in diverse fields, time-varying spectrum and coherence become a popular tool to reveal the dynamics of the underlying mechanism. For example, in EEG data analysis, it has been widely used to measure brain functional connectivity; see Liu, Gaetz and Zhu (2010), Simpson, Bowman and Laurienti (2013), Lindquist et al. (2014) among others.

Various models and methods have been developed to estimate the time-varying spectra and coherences for nonstationary processes. Priestley and Tong (1973) concerned the cross-spectrum and coherence between oscillatory processes stemming from a time-varying spectral representation, which was later investigated by Dahlhaus (2000a) allowing for rigorous asymptotic considerations. Ombao et al. (2001) proposed a method based on the smooth localized complex exponentials to select the span which can be used to obtain the smoothed estimates of the time-varying spectra and coherence. Sanderson, Fryzlewicz and Jones (2010) and Park, Eckley and Ombao (2014) considered the problem of estimating time-evolving cross-dependence in a collection of locally stationary wavelet processes. Ombao and Van Bellegem (2008) developed a coherence estimation procedure using time-localized linear filtering. Many of the previous results require restrictive structural condition on the underlying processes such as linearity or Gaussianity.

Under the framework (1.1), the time-varying spectral density matrix function is defined as

$$F(u, \theta) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \Gamma_k(u) \exp(-ik\theta) \quad \text{where } \iota = \sqrt{-1}.$$

To estimate the time-varying spectral density matrix consistently, we use smoothing and consider the lag window estimate

$$(4.1) \quad \hat{F}(u, \theta) = \frac{1}{2\pi} \sum_{l=-m}^m K(l/m) \hat{\Gamma}_l(u) \exp(-l\theta),$$

where  $\hat{\Gamma}_l(u)$  is the estimate of the autocovariance matrix function with lag  $l$  defined in (3.3),  $m$  is the window width satisfying the natural conditions  $m \leq M^\beta$  for some  $0 < \beta < 1$  and  $K(\cdot)$  is a continuous symmetric nonnegative kernel function with the support  $[-1, 1]$  and  $K(0) = 1$ . In the special case of stationary processes  $X_t = G(\mathcal{F}_t)$ , the spectral density matrix  $F(\theta) = (2\pi)^{-1} \sum_{k \in \mathbb{Z}} \Gamma_k \exp(-ik\theta)$  and we can estimate  $F(\theta)$  by

$$(4.2) \quad \hat{F}(\theta) = \frac{1}{2\pi} \sum_{l=-m}^m K(l/m) \hat{\Gamma}_l \exp(-l\theta),$$

where  $\hat{\Gamma}_l$  are estimates of autocovariance matrices, given by (3.1).

Theorem 4.1 and Theorem 4.2 below provide nonasymptotic bounds for  $|\hat{F}(u, \theta) - \mathbb{E}\hat{F}(u, \theta)|_\infty$  uniformly over  $u$  and  $\theta$ , under the assumption of finite polynomial moments and Gaussianity, respectively, while Proposition 4.3 concerns the stationary case. Corollary 4.4 concerns the deviation of  $\hat{F}(u, \theta)$  and the true spectral density matrix  $F(u, \theta)$ .

**THEOREM 4.1.** *Assume that  $\mathbb{E}(X_t) = 0$  and  $\Theta_{q,\alpha} < \infty$  for some  $q > 4$  and  $\alpha > 0$ . Let  $b_T$  be the bandwidth and  $M = 2\lfloor Tb_T \rfloor$ . Assume  $M \leq T$  and  $m \leq \lfloor M^\beta \rfloor$  for some  $0 < \beta < 1$ . Let  $\ell = 1 \vee \log p$ . Let*

$$(4.3) \quad \varphi_T = \sup_{u \in [b_T, 1-b_T]} \max_{\theta} |\hat{F}(u, \theta) - \mathbb{E}\hat{F}(u, \theta)|_\infty.$$

Then for any  $x > 0$ , we have

$$(4.4) \quad \begin{aligned} \mathbb{P}(\varphi_T \geq x) &\leq C_{q,\alpha} (Mx)^{-q/2} T m R_{M,m} (\ell^{5/4} \|X_\cdot\|_{q,\alpha} \wedge \Theta_{q,\alpha})^q \\ &\quad + CTp^2 \exp\left(-\frac{Mx^2}{C_\alpha \Phi_{4,\alpha}^4 m}\right), \end{aligned}$$

where  $R_{M,m} = m^{q/2-1} (\log M)^{q+1}$  for  $\alpha > 1/2 - 2/q$ , and  $R_{M,m} = m^{q/2-1} \cdot (\log M)^{q+1} + M^{q/4-1-\alpha q/2} m^{q/4}$  for  $\alpha < 1/2 - 2/q$ .

**THEOREM 4.2.** *Let  $(X_t)$  be a Gaussian process of the form (1.1), which satisfies  $\mathbb{E}(X_t) = 0$  and  $\Phi_{2,0} < \infty$ . Let  $b_T$  be the bandwidth and  $M = 2\lfloor Tb_T \rfloor$ . Assume  $M \leq T$  and  $m \leq \lfloor M^\beta \rfloor$  for some  $0 < \beta < 1$ . Then there exist universal constants  $C_1, C_2 > 0$  such that for any  $x > 0$ ,*

$$(4.5) \quad \mathbb{P}(\varphi_T > x) \leq C_1 T m p^2 \exp\left[-C_2 \min\left(\frac{Mx^2}{m\Phi_{2,0}^4}, \frac{Mx}{m\Phi_{2,0}^2}\right)\right].$$

**PROPOSITION 4.3.** *For zero-mean stationary processes  $X_t = G(\mathcal{F}_t)$ , define*

$$\tilde{\varphi}_T = \max_{\theta} |\hat{F}(\theta) - \mathbb{E}\hat{F}(\theta)|_\infty.$$

(i) *Under the assumptions of Theorem 4.1, we have*

$$(4.6) \quad \begin{aligned} \mathbb{P}(\tilde{\varphi}_T \geq x) &\leq C_{q,\alpha} (Tx)^{-q/2} T m R_{T,m}^* (\ell^{5/4} \|X_\cdot\|_{q,\alpha} \wedge \Theta_{q,\alpha})^q \\ &\quad + C m p^2 \exp\left(-\frac{Tx^2}{C_\alpha \Phi_{4,\alpha}^4 m}\right), \end{aligned}$$

where  $R_{T,m}^* = m^{q/2-1}$  for  $\alpha > 1/2 - 2/q$ , and  $R_{T,m}^* = m^{q/2-1} + T^{q/4-1-\alpha q/2} m^{q/4}$  for  $\alpha < 1/2 - 2/q$ . (ii) For Gaussian stationary processes, it becomes

$$(4.7) \quad \mathbb{P}(\tilde{\varphi}_T > x) \leq C_1 m p^2 \exp\left[-C_2 \min\left(\frac{T x^2}{m \Phi_{2,0}^4}, \frac{T x}{m \Phi_{2,0}^2}\right)\right].$$

As expected, the tail probabilities for  $\tilde{\varphi}_T$  for stationary processes are sharper than the ones in Theorems 4.1 and 4.2, as  $\tilde{\varphi}_T$  does not concern the supremum over the time index. The proof of Proposition 4.3 is simpler and the argument is given in the Supplementary Material (Zhang and Wu (2020)).

REMARK 3. The long-run covariance matrix (a.k.a. asymptotic covariance matrix) can be determined by the spectral density matrix at the zero frequency. The estimation of the long-run covariance matrix is an important problem in statistical inference for time series and has been extensively studied in the low-dimensional stationary case; see Newey and West (1987), Politis, Romano and Wolf (1999), Bühlmann (2002), Lahiri (2003), Alexopoulos and Goldsman (2004). For locally stationary processes, we can estimate the time-varying long-run covariance matrix  $\Sigma(u) = \sum_{k=-\infty}^{\infty} \Gamma_k(u)$  by the idea of smoothing similarly as (4.1). Then nonasymptotic results similar to Theorem 4.1 and Theorem 4.2 can be established in high dimensions without extra difficulty. The convergence rate of the long-run covariance matrix estimator is sharper than that of  $\varphi_T$  given in (4.4) or (4.5) since there is no need to account for the supremum over  $\theta$ ; see Corollary B.1 in the Supplementary Material for details (Zhang and Wu (2020)).

COROLLARY 4.4. Define

$$\Delta_\varphi = \sup_{u \in [b_T, 1-b_T]} \max_{\theta} |\mathbb{E} \hat{F}(u, \theta) - F(u, \theta)|_\infty,$$

$$\varphi_T^* = \sup_{u \in [b_T, 1-b_T]} \max_{\theta} |\hat{F}(u, \theta) - F(u, \theta)|_\infty.$$

Under condition (1.2), it follows that  $\Delta_\varphi \leq \mathcal{V}_{m,M,T} + \mathcal{W}_m$ , where

$$\mathcal{V}_{m,M,T} = \frac{2\mathcal{K}M\sqrt{m}}{\pi T} \Phi_{2,0} + \pi^{-1} \left(m^{-\alpha} + \frac{r(m)}{M}\right) \Phi_{2,0} \Phi_{2,\alpha},$$

$$\mathcal{W}_m = \pi^{-1} \sup_u \sum_{l=1}^m (1 - K(l/m)) |\Gamma_l(u)|_\infty,$$

and  $r(m) = 1$  if  $\alpha > 1$ ,  $r(m) = \log m$  if  $\alpha = 1$  and  $r(m) = m^{1-\alpha}$  if  $\alpha < 1$ . Consequently,  $\varphi_T^* = O_{\mathbb{P}}(\mathcal{R}_1 + \mathcal{R}_2 + \Delta_\varphi)$  under the assumptions of Theorem 4.1 and  $\varphi_T^* = O_{\mathbb{P}}(\mathcal{R}_3 + \Delta_\varphi)$  under the assumptions of Theorem 4.2 with

$$\mathcal{R}_1 = \frac{(TmR_{M,m})^{2/q}}{M} (\ell^{5/4} \| |X \cdot|_\infty \|_{q,\alpha} \wedge \Theta_{q,\alpha})^2, \quad \mathcal{R}_2 = \sqrt{\frac{m \log(pT)}{M}} \Phi_{4,\alpha}^2,$$

$$\mathcal{R}_3 = \sqrt{\frac{m \log(pT)}{M}} \Phi_{2,0}^2 \vee \frac{m \log(pT)}{M} \Phi_{2,0}^2.$$

The term  $\mathcal{W}_m$  depends on the kernel function. Its order is determined by the smoothness of  $K(\cdot)$  at zero. In particular, this term vanishes if  $K(\cdot)$  is the rectangular kernel. In general, flat-top kernels which take value 1 at a neighbor of 0 have been employed to render a bias reduced estimator of spectral density; see for example, Politis and Romano (1995, 1999) and

Politis (2011). If  $1 - K(x) = O(|x|^a)$  at  $x = 0$  for some  $a > 0$  and  $\sup_u |\Gamma_l(u)|_\infty = O(l^{-b})$  for some  $b > 1$ , then  $\mathcal{W}_m = O(m^{-a} + m^{1-b})$ .

For the bias  $\Delta_{\hat{\varphi}} = \max_\theta |\mathbb{E}\hat{F}(\theta) - F(\theta)|_\infty$  in the stationary case, since  $\mathcal{K} = 0$ , we can obtain  $\Delta_{\hat{\varphi}} \leq \tilde{\mathcal{V}}_{m,T} + \tilde{\mathcal{W}}_m$  where

$$\tilde{\mathcal{V}}_{m,T} = \pi^{-1} \left( m^{-\alpha} + \frac{r(m)}{T} \right) \Phi_{2,0} \Phi_{2,\alpha},$$

$$\tilde{\mathcal{W}}_m = \pi^{-1} \sup_u \sum_{l=1}^m (1 - K(l/m)) |\Gamma_l|_\infty,$$

with  $r(m)$  defined the same as in Corollary 4.4.

A similar discussion as Section 3 can be made to concern the effects of local stationarity, moment condition, dependence strength and dimension on the convergence rates. The details are omitted here.

Next, we shall discuss how our results can be used.

*Regularized estimation of sparse spectral density matrices.* In the above Theorems 4.1 and 4.2, sparseness conditions are not imposed for spectral density matrix estimation. Sun et al. (2018) investigated regularized estimation of high-dimensional spectral density matrices for stationary Gaussian processes by imposing weak sparsity on the spectral density matrix in the sense that it falls within a small  $\ell^d$  ball in  $\mathbb{C}^{p \times p}$  for some  $0 \leq d < 1$ . In particular, they proposed hard thresholding of averaged periodograms to estimate the spectral density matrix and established nonasymptotic bounds for the concentration of the estimator around its expectation using spectral norm and Frobenius norm. The idea of thresholding has been widely used in high-dimensional covariance matrix estimation; see, for example, Bickel and Levina (2008) for i.i.d. vectors and Chen, Xu and Wu (2013) for time series. Applying the thresholding procedure to the lag-window estimate  $\hat{F}(u, \theta)$  in (4.1), we can introduce the regularized estimate in our regime:

$$(4.8) \quad T_\tau(\hat{F}(u, \theta)) = (\hat{F}_{ij}(u, \theta) \mathbf{1}\{|\hat{F}_{ij}(u, \theta)| \geq \tau\})_{i,j=1}^p,$$

where  $\tau > 0$  is a tuning parameter and  $T_\tau(\cdot)$  is a thresholding operator. Assume that  $F(u, \theta)$  has weak sparsity, that is, for some  $0 \leq d < 1$ ,  $\sup_{u \in [0,1]} \max_\theta \max_i \sum_{j=1}^p |F_{ij}(u, \theta)|^d \leq R_p$ . Recall the definition of  $\varphi_T^*$  in Corollary 4.4. We can adopt similar techniques in Bickel and Levina (2008) to obtain that under the event  $\varphi_T^* \leq \tau/2$ ,

$$\sup_{u \in [b_T, 1-b_T]} \max_\theta \|T_\tau(\hat{F}(u, \theta)) - F(u, \theta)\| \leq 7\tau^{1-d} R_p,$$

$$\sup_{u \in [b_T, 1-b_T]} \max_\theta \|T_\tau(\hat{F}(u, \theta)) - F(u, \theta)\|_F \leq 13\tau^{1-d} R_p.$$

Hence, the nonasymptotic bound in  $L^\infty$  norm can be used to derive the uniform convergence in spectral norm and Frobenius norm for thresholded estimates in the sparse case. In comparison with Sun et al. (2018), we allow more general processes which can be non-Gaussian and nonstationary, and we can provide a uniform bound by taking supreme over the frequency while they established a pointwise result at each single frequency. We shall comment that for the stationary Gaussian case considered in Sun et al. (2018) we can obtain the same result as Proposition 3.6 in that paper.

*Application to the estimation of locally stationary generalized dynamic factor models.* Barigozzi et al. (2019) considered Time-varying Generalized Dynamic Factor Models (tvGDFM), extending the influential GDFM introduced in Forni et al. (2000) to locally stationary processes. To perform theoretical analysis of the estimation procedure for the tvGDFM, one needs to establish a moment bounds for uniform distance

$\sup_u \max_\theta |\hat{F}_{ij}(u, \theta) - F_{ij}(u, \theta)|$ . Our Theorem 4.1 and Corollary 4.4 can provide such a theoretical foundation. Using  $\mathbb{E}\varphi_T^2 = \int_0^\infty \mathbb{P}(\varphi_T^2 \geq v) dv$ , we obtain

$$(4.9) \quad \mathbb{E}|\varphi_T^*|^2 = \mathbb{E} \sup_u \max_\theta |\hat{F}(u, \theta) - F(u, \theta)|_\infty^2 \leq C_q (\mathcal{R}_1^2 + \mathcal{R}_2^2 + \Delta_\varphi^2),$$

where  $\mathcal{R}_1, \mathcal{R}_2$  and  $\Delta_\varphi$  have been defined in Corollary 4.4. Note that the term  $\mathcal{R}_1^2 + \mathcal{R}_2^2$  in (4.9) should be  $\mathcal{R}_3^2$  for Gaussian processes. For a more user-friendly bound, consider  $\alpha > 1/2 - 2/q$  and choose the rectangular kernel  $K(\cdot)$ . Let  $m \asymp \lfloor M^\beta \rfloor$  for some  $0 < \beta < 1$ . Assume that  $\|X_{\cdot j}\|_{q, \alpha} \asymp 1$  for all  $1 \leq j \leq p$ ,  $\mathcal{K} \asymp 1$  and  $\|X_{\cdot}\|_{q, \alpha} \asymp p^\tau, 0 \leq \tau \leq 2/q$ . Then (4.9) becomes

$$(4.10) \quad \mathbb{E}|\varphi_T^*|^2 \lesssim \frac{T^{4/q} (\log M)^{4+4/q} \min\{p^{4\tau} (1 \vee \log p)^5, p^{8/q}\}}{M^{2-2\beta}} + \frac{\log(pT)}{M^{1-\beta}} + \frac{M^{2+\beta}}{T^2},$$

and, by considering the cross spectral density for each pair of component processes in Theorem 4.1 and Corollary 4.4, we have

$$\max_{1 \leq i, j \leq p} \mathbb{E} \sup_u \max_\theta |\hat{F}_{ij}(u, \theta) - F_{ij}(u, \theta)|^2 \lesssim \frac{T^{4/q} (\log M)^{4+4/q}}{M^{2-2\beta}} + \frac{\log T}{M^{1-\beta}} + \frac{M^{2+\beta}}{T^2}.$$

*Estimation of coherence matrices.* In many applications, it is of interest to estimate the coherence matrix. In our framework, the time-varying coherence matrix is given by

$$\mathcal{C}(u, \theta) = \text{diag}[F(u, \theta)]^{-1/2} F(u, \theta) \text{diag}[F(u, \theta)]^{-1/2}.$$

We estimate the coherence matrix  $\mathcal{C}(u, \theta)$  by the plug-in estimator

$$(4.11) \quad \hat{\mathcal{C}}(u, \theta) = \text{diag}[\hat{F}(u, \theta)]^{-1/2} \hat{F}(u, \theta) \text{diag}[\hat{F}(u, \theta)]^{-1/2},$$

where  $\hat{F}(u, \theta)$  is the estimate of the spectral density matrix given by (4.1). We shall concern the bound for the maximum deviation

$$\rho_T = \sup_{u \in [b_T, 1-b_T]} \max_\theta |\hat{\mathcal{C}}(u, \theta) - \mathcal{C}(u, \theta)|_\infty.$$

Corollary 4.5 below gives a bound of  $\rho_T$  in terms of  $\varphi_T^*$ , by which the results for  $\varphi_T^*$  can be used to bound  $\rho_T$ .

**COROLLARY 4.5.** *Assume  $c_0 = \inf_u \min_\theta \min_{1 \leq j \leq p} F_{jj}(u, \theta) > 0$ . Then*

$$(4.12) \quad \rho_T \leq \frac{3\varphi_T^*}{c_0} + \frac{2\varphi_T^{*2}}{c_0^2},$$

where  $\varphi_T^* = \sup_u \max_\theta |\hat{F}(u, \theta) - F(u, \theta)|_\infty$  as defined in Corollary 4.4.

**5. Graphical model and inverse spectral density matrix.** The concept of graphical model for multivariate data has been extended to multivariate time series (e.g., Brillinger (1996), Dahlhaus (2000b), Timmer et al. (2000), Eichler (2012) among others). A vertex of the graph represents a component process and each edge indicates the partial correlation of the two corresponding components given others. Hence, for stationary Gaussian processes, this induced graph is a conditional independence graph in the frequency domain, the properties of which has been investigated largely (cf. Dahlhaus (2000b), Fried and Didelez (2003), Bach and Jordan (2004), etc.). For non-Gaussian processes, it is termed partial correlation

graph in Dahlhaus (2000b) using partial spectral coherence as a measure for the dependence between two marginal time series after removing the linear effects of some other components. Partial spectral coherence has been widely used in many real-world applications; see, for example, Gather, Imhoff and Fried (2002), Salvador et al. (2005), Eichler (2007), Medkour, Walden and Burgess (2009). Recently researchers study functional connectivities of brain networks in neuroscience based on inverse of spectral density matrices; see, for example, Baccalá and Sameshima (2001), Eichler, Dahlhaus and Sandkühler (2003), Blinowska (2011) and Lennartz et al. (2018). Baccalá and Sameshima (2001) proposed partial direct coherence, a normalized quantity for inverse of spectral density matrices which measures frequency domain direct causal relations. A zero value in the inverse of spectral density matrices suggests no partial direct coherence. For locally stationary processes, it is natural to study the time-varying functional connectivity based on the inverse spectral density matrix function. However, in the high-dimensional case where the dimension  $p$  can be even much larger than the sample size  $T$ , since the estimated spectral density matrix may not be invertible, classical methods developed under the low-dimensional setting are no longer applicable. In this section, we shall provide a solution to this challenging problem under the more general setting in which the process can be locally stationary and hence the inverse spectral density matrix varies with time.

For  $0 \leq u \leq 1$  and  $\theta$ , denote by  $\Omega^0(u, \theta) = F(u, \theta)^{-1}$ , the inverse of the spectral density matrix. We estimate the spectral density matrix by the lag window estimate [cf. (4.1)]. For simplicity, we consider the rectangular kernel, that is,  $K(x) = 1$  for  $|x| \leq 1$  and  $K(x) = 0$  otherwise. Then we use the constrained  $\ell_1$  minimization approach to estimate  $\Omega^0(u, \theta)$ . Let

$$(5.1) \quad \hat{\Omega}(u, \theta) = \arg \min |\Omega(u, \theta)|_{\ell_1} \quad \text{subject to } |\hat{F}(u, \theta)\Omega(u, \theta) - I_p|_{\infty} \leq \lambda,$$

where  $\lambda > 0$  is a tuning parameter. The constrained  $\ell_1$  minimization approach has been adopted in many applications; see Candès and Tao (2007), Bickel, Ritov and Tsybakov (2009), Cai, Liu and Luo (2011) among many others. The optimization program (5.1) can be decomposed into  $p$  parallel vector minimization sub-problems. Let  $e_i$  be the standard unit vector in  $\mathbb{R}^p$  with 1 in the  $i$ th coordinate and 0 in all others. For  $1 \leq i \leq p$ , let  $\hat{w}_i(u, \theta)$  be the solution to the following convex optimization problem:

$$(5.2) \quad \min |w|_1 \quad \text{subject to } |\hat{F}(u, \theta)w - e_i|_{\infty} \leq \lambda,$$

where  $w$  is a vector in  $\mathbb{R}^p$ . By a similar argument as Lemma 1 of Cai, Liu and Luo (2011), we can show that solving the optimization problem (5.1) is equivalent to solving the  $p$  optimization problems (5.2), that is,

$$(5.3) \quad \hat{\Omega}(u, \theta) = (\hat{w}_1(u, \theta), \dots, \hat{w}_p(u, \theta)).$$

We estimate  $\Omega^0(u, \theta)$  by

$$(5.4) \quad \tilde{\Omega}(u, \theta) = \frac{\hat{\Omega}(u, \theta) + \hat{\Omega}^\dagger(u, \theta)}{2},$$

where  $\dagger$  is the conjugate transpose of a matrix. Theorem 5.1 provides a nonasymptotic bound concerning the uniform convergence of  $\tilde{\Omega}(u, \theta)$ . To this end, we need to introduce quantity  $\kappa_0$  which characterizes the sparseness of  $\Omega^0(u, \theta)$ . Note that sparseness conditions are not needed for covariance and spectral density matrix estimates in Theorems 3.1 and 4.1.

**THEOREM 5.1.** *Define  $\kappa_0 = \sup_{0 \leq u \leq 1} \max_{\theta} |\Omega^0(u, \theta)|_{\ell_1}$  and*

$$\varrho_T = \sup_{u \in [b_T, 1-b_T]} \max_{\theta} |\tilde{\Omega}(u, \theta) - \Omega^0(u, \theta)|_{\infty}.$$

Recall Corollary 4.4 for  $\mathcal{V}_{m,M,T}$ . (i) Let the assumptions of Theorem 4.1 be satisfied. For any  $x \geq \mathcal{V}_{m,M,T} \vee (\lambda/\kappa_0)$ , we have

$$(5.5) \quad \mathbb{P}(Q_T \geq 7x\kappa_0^2) \leq \frac{C_{q,\alpha} T m R_{M,m} (\ell^{5/4} \|X\|_{\infty} \|q,\alpha \wedge \Theta_{q,\alpha}\|^q)}{(Mx)^{q/2}} + CTp^2 \exp\left(-\frac{Mx^2}{C_\alpha m \Phi_{4,\alpha}^4}\right).$$

(ii) Let the assumptions of Theorem 4.2 be satisfied. For any  $x \geq \mathcal{V}_{m,M,T} \vee (\lambda/\kappa_0)$ , we have

$$(5.6) \quad \mathbb{P}(Q_T \geq 7x\kappa_0^2) \leq C_1 T m p^2 \exp\left[-C_2 \min\left(\frac{Mx^2}{m \Phi_{2,0}^4}, \frac{Mx}{m \Phi_{2,0}^2}\right)\right].$$

REMARK 4. In the special case of stationary processes, denote the inverse spectral density matrix by  $\Omega^0(\theta) = F(\theta)^{-1}$ . As in (5.2) and (5.3), we can similarly consider the minimizer  $\hat{\Omega}(\theta) = (\hat{w}_1(\theta), \dots, \hat{w}_p(\theta))$  in which  $\hat{w}_i(\theta) \in \mathbb{R}^p$ ,  $1 \leq i \leq p$ , is the solution to the following convex optimization problem:

$$(5.7) \quad \min |w|_1 \quad \text{subject to } |\hat{F}(\theta)w - e_i|_\infty \leq \lambda,$$

where  $w$  is a vector in  $\mathbb{R}^p$  and  $\hat{F}(\theta)$  is given in (4.2). Let  $\tilde{\Omega}(\theta) = (\hat{\Omega}(\theta) + \hat{\Omega}^\dagger(\theta))/2$ . The tail probability bound in the right-hand side of (5.5) and (5.6) concerning  $\max_\theta |\tilde{\Omega}(\theta) - \Omega^0(\theta)|_\infty$  should be the same as that in (4.6) and (4.7) respectively. To prove it, we can follow all the arguments in the proof of Theorem 5.1 and then replace the last step by the corresponding result for stationary processes established in Proposition 4.3.

Fiecas et al. (2019) adopted the essentially same approach as (5.7) to estimate the inverse spectral density matrix for stationary processes. But they required more restrictive assumptions to establish the nonasymptotic bound. For one thing, they assumed geometric moment contraction, that is,  $\Delta_{m,2,j} = O(\lambda^m)$  for some  $0 < \lambda < 1$ , while we can deal with much stronger dependence with algebraic decay characterized by the parameter  $\alpha$ . For another, they required the existence of finite exponential moment for each component process while we can allow the mild condition with the existence of polynomial moment.

Note that the bound for the tail probability in Case (i) (resp., Case (ii)) of Theorem 5.1 is the same to Theorem 4.1 (resp., Theorem 4.2). We shall discuss the newly introduced parameter  $\kappa_0$ , which characterizes the sparseness of  $\Omega^0(u, \theta)$ . Consider the class of high-dimensional vector autoregressive models:  $X_t = AX_{t-1} + \varepsilon_t$ . Assume that  $\text{Cov}(\varepsilon_t) = I_p$  and the spectral radius of  $A = (a_{ij})_{i,j=1}^p$  is smaller than 1. By elementary calculation, we can obtain  $\Omega^0(\theta) = 2\pi(I_p + A^\top A - Ae^{-i\theta} - A^\top e^{i\theta})$ . Let  $L = \max(|A|_{\ell_1}, |A^\top|_{\ell_1})$  where  $|A|_{\ell_1} := \max_i \sum_j |a_{ij}|$ . Then  $\kappa_0 \leq 2\pi(1+L)^2$  and the sparseness of  $A$  ensures the sparseness of  $\Omega^0(\theta)$  in terms of the  $\ell_1$  norm.

Fiecas et al. (2019) also incorporated the  $\ell_1$  norm of  $\Omega^0(\theta)$  for high-dimensional stationary processes and additionally assumed the inverse spectral density matrix falls within a small  $\ell^d$  ball,  $0 \leq d < 1$ . Similarly, if we further assume  $\Omega^0(u, \theta)$  is weakly sparse within a small  $\ell^d$  ball, we can also work out the bounds in spectral norm and Frobenius norm accordingly; see Remark B.1 for detailed results.

**6. Hanson–Wright-type inequalities.** In this section, we shall provide Hanson–Wright-type tail probability inequalities for locally stationary processes. The celebrated Hanson–Wright inequality provided a concentration result for quadratic forms of sub-Gaussian i.i.d. random variables; see Hanson and Wright (1971), Wright (1973) and Rudelson and

Vershynin (2013). There has been a large literature concerning large/moderate deviations for quadratic forms of Gaussian processes; see, for example, Bercu, Gamboa and Rouault (1997), Bryc and Dembo (1997), Zani (2002), Kakizawa (2007) among others. Xiao and Wu (2012) obtained tail probability upper bounds for quadratic forms of stationary processes with finite polynomial moments. We aim to relax the (i) i.i.d., (ii) Gaussian/sub-Gaussian, (iii) one-dimensional or (iv) stationary assumptions which were imposed in previous works, and develop tail probability inequalities for quadratic forms for high-dimensional locally stationary processes.

Consider the quadratic form of the high-dimensional locally stationary process  $(X_t)$ :

$$Q_T = \sum_{1 \leq s \leq t \leq T} a_{s,t} X_s X_t^\top,$$

where the coefficients in our setting satisfy  $a_{s,t} = a_{t-s}$  ( $t \geq s$ ), which depends on the distance  $t - s$ . Moreover, we assume  $\sup_{s,t} |a_{s,t}| \leq 1$  and  $a_{s,t} = 0$  if  $t - s > B$ , where  $B \leq T$ . Theorem 6.1 and Theorem 6.3 provide tail probability inequalities for  $|Q_T - \mathbb{E}Q_T|_\infty$ . The former assumes the existence of finite polynomial moments and the latter assumes the Gaussianity.

**THEOREM 6.1.** *For the process (1.1), assume  $\mathbb{E}(X_t) = 0$  and, for some  $q > 4$  and  $\alpha > 0$ ,  $\|X_\cdot\|_\infty \|q, \alpha < \infty$ . Let  $\ell = 1 \vee \log p$  and let  $B \leq T$ . Then there exist constants  $C, C_\alpha, C_{q,\alpha} > 0$  such that for any  $x > 0$ ,*

$$\begin{aligned} \mathbb{P}(|Q_T - \mathbb{E}Q_T|_\infty \geq x) &\leq C_{q,\alpha} x^{-q/2} \ell^{5q/4} \|X_\cdot\|_\infty^q \|q, \alpha F_{T,B} \\ &\quad + Cp^2 \exp\left(-\frac{x^2}{C_\alpha \Phi_{4,\alpha}^4 TB}\right), \end{aligned}$$

where  $F_{T,B} = TB^{q/2-1}$  (resp.,  $TB^{q/2-1} + T^{q/4-\alpha q/2} B^{q/4}$ ) if  $\alpha > 1/2 - 2/q$  (resp.,  $\alpha < 1/2 - 2/q$ ).

Proposition 6.2 below concerns the special case of one-dimensional processes, by letting  $\ell = 1$  and replacing the  $L^\infty$  dependence adjusted norm  $\|X_\cdot\|_\infty \|q, \alpha$  in Theorem 6.1 by the component-wise dependence adjusted norms  $\|X_{\cdot i}\|_{q,\alpha}$  and  $\|X_{\cdot j}\|_{q,\alpha}$ .

**PROPOSITION 6.2.** *For  $1 \leq i, j \leq p$ , let  $Q_{T,ij} = \sum_{1 \leq s \leq t \leq T} a_{s,t} X_{si} X_{tj}$ . Under the assumptions of Theorem 6.1, we have*

$$\begin{aligned} \mathbb{P}(|Q_{T,ij} - \mathbb{E}Q_{T,ij}| \geq x) &\leq C_{q,\alpha} x^{-q/2} \|X_{\cdot i}\|_{q,\alpha}^{q/2} \|X_{\cdot j}\|_{q,\alpha}^{q/2} F_{T,B} \\ &\quad + C \exp\left(-\frac{x^2}{C_\alpha \Phi_{4,\alpha}^4 TB}\right), \end{aligned}$$

where  $F_{T,B}$  is defined the same as in Theorem 6.1.

**THEOREM 6.3.** *Let  $(X_t)$  be a Gaussian process of the form (1.1), which satisfies  $\mathbb{E}(X_t) = 0$  and  $\Phi_{2,0} < \infty$ . Let  $B \leq T$ . Then there exists a universal constant  $C > 0$  such that for any  $x > 0$ ,*

$$\mathbb{P}(|Q_T - \mathbb{E}Q_T|_\infty > x) \leq 2p^2 \exp\left[-C \min\left(\frac{x^2}{TB\Phi_{2,0}^4}, \frac{x}{B\Phi_{2,0}^2}\right)\right].$$

Theorems 6.4 and 6.6 below concern the following special case of  $Q_T$ :

$$(6.1) \quad L_T(B) := \sum_{B+1 \leq t \leq T} a_t X_{t-B} X_t^\top,$$

where  $\sup_t |a_t| \leq 1$  and  $0 \leq B < T$ . Proposition 6.5 applies to the one-dimensional case. They are useful to prove the results of the estimates of autocovariance matrices in Section 3.

**THEOREM 6.4.** *For the process (1.1), assume  $\mathbb{E}(X_t) = 0$  and, for some  $q > 4$  and  $\alpha > 0$ ,  $\|X_{\cdot}\|_{q,\alpha} < \infty$ . Let  $\ell = 1 \vee \log p$  and let  $B < T$ . Then there exists a constant  $C_{q,\alpha} > 0$  such that for any  $x > 0$ ,*

$$\begin{aligned} \mathbb{P}(|L_T(B) - \mathbb{E}L_T(B)|_\infty \geq x) &\leq C_{q,\alpha} x^{-q/2} \ell^q \|X_{\cdot}\|_{q,\alpha}^q D_{T,B} \\ &\quad + Cp^2 \exp\left(-\frac{x^2}{C_\alpha \Phi_{4,\alpha}^4 T}\right), \end{aligned}$$

where  $D_{T,B} = TB^{q/4-1}$  (resp.,  $TB^{q/4-1} + T^{q/4-\alpha q/2}$ ) if  $\alpha > 1/2 - 2/q$  (resp.,  $\alpha < 1/2 - 2/q$ ).

**PROPOSITION 6.5.** *Let  $L_{T,ij}(B) = \sum_{B+1 \leq t \leq T} a_t X_{(t-B)i} X_{tj}$ ,  $1 \leq i, j \leq p$ . Under the assumptions of Theorem 6.4, we have*

$$\begin{aligned} \mathbb{P}(|L_{T,ij}(B) - \mathbb{E}L_{T,ij}(B)| \geq x) &\leq C_{q,\alpha} x^{-q/2} \|X_{\cdot i}\|_{q,\alpha}^{q/2} \|X_{\cdot j}\|_{q,\alpha}^{q/2} D_{T,B} \\ &\quad + C \exp\left(-\frac{x^2}{C_\alpha \Phi_{4,\alpha}^4 T}\right), \end{aligned}$$

where  $D_{T,B}$  is defined the same as in Theorem 6.4.

**THEOREM 6.6.** *Let  $(X_t)$  be a Gaussian process of the form (1.1), which satisfies  $\mathbb{E}(X_t) = 0$  and  $\Phi_{2,0} < \infty$ . Then there exists a universal constant  $C > 0$  such that for any  $x > 0$ ,*

$$\mathbb{P}(|L_T(B) - \mathbb{E}L_T(B)|_\infty > x) \leq 2p^2 \exp\left[-C \min\left(\frac{x^2}{T\Phi_{2,0}^4}, \frac{x}{\Phi_{2,0}^2}\right)\right].$$

**7. Concluding remarks.** High-dimensional nonstationary processes arise in a wide range of disciplines. In this paper, we have made contributions towards a general theory for high-dimensional locally stationary processes that goes beyond the investigation of specific parametric models. We showed that many commonly seen parametric recursive models fit approximately within the framework of functional dependence measure, a convenient framework to depict the temporal dependence for high-dimensional processes. Equipped with functional dependence measure, the main tools we developed are tail probability inequalities for quadratic forms involving high-dimensional processes. We established a Nagaev-type bound on tail probability of quadratic forms with the existence of finite polynomial moments and a Hanson–Wright-type bound for Gaussian processes, based on which, we were able to estimate the autocovariance functions, spectral density matrix and inverse spectral density matrix. The convergence rate depends on the temporal dependence, the moment condition, the dimension and the sample size. To perform statistical inference of the estimates such as hypothesis testing and construction of simultaneous confidence bands, one needs to develop the more refined result in terms of asymptotic distributional theory. The latter is more challenging and we leave it as future work.

**Acknowledgements.** The authors thank the Editor, an Associate Editor and three anonymous referees for their many constructive comments. The authors also thank Marc Hallin, Matteo Barigozzi and Rainer von Sachs for their interest in our work.

## SUPPLEMENTARY MATERIAL

**Supplement to “Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes”** (DOI: [10.1214/20-AOS1954SUPP](https://doi.org/10.1214/20-AOS1954SUPP); .pdf). This supplemental file contains technical proofs of the results and theoretical justification of some conclusions in the main body.

## REFERENCES

- ADAK, S. (1998). Time-dependent spectral analysis of nonstationary time series. *J. Amer. Statist. Assoc.* **93** 1488–1501. MR1666643 <https://doi.org/10.2307/2670062>
- ALEXOPOULOS, C. and GOLDSMAN, D. (2004). To batch or not to batch? *ACM Trans. Model. Comput. Simul.* **14** 76–114.
- AN, H. Z., CHEN, Z. G. and HANNAN, E. J. (1982). Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.* **10** 926–936. MR0663443
- ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York. MR0283939
- BACCALÁ, L. A. and SAMEISHIMA, K. (2001). Partial directed coherence: A new concept in neural structure determination. *Biol. Cybernet.* **84** 463–474.
- BACH, F. R. and JORDAN, M. I. (2004). Learning graphical models for stationary time series. *IEEE Trans. Signal Process.* **52** 2189–2199. MR2085580 <https://doi.org/10.1109/TSP.2004.831032>
- BARIGOZZI, M., HALLIN, M., SOCCORSI, S. and VON SACHS, R. (2019). Time-varying general dynamic factor models and the measurement of financial connectedness. Available at SSRN 3329445.
- BERCU, B., GAMBOA, F. and ROUAULT, A. (1997). Large deviations for quadratic forms of stationary Gaussian processes. *Stochastic Process. Appl.* **71** 75–90. MR1480640 [https://doi.org/10.1016/S0304-4149\(97\)00071-9](https://doi.org/10.1016/S0304-4149(97)00071-9)
- BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008 <https://doi.org/10.1214/08-AOS600>
- BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 <https://doi.org/10.1214/08-AOS620>
- BLINOWSKA, K. J. (2011). Review of the methods of determination of directed connectivity from multichannel data. *Med. Biol. Eng. Comput.* **49** 521–529.
- BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory*. Holt, Rinehart, and Winston, New York. MR0443257
- BRILLINGER, D. R. (1996). Remarks concerning graphical models for time series and point processes. *Revista de Econometria* **16** 23.
- BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1093459 <https://doi.org/10.1007/978-1-4419-0320-4>
- BRYC, W. and DEMBO, A. (1997). Large deviations for quadratic functionals of Gaussian processes. *J. Theoret. Probab.* **10** 307–332. MR1455147 <https://doi.org/10.1023/A:1022656331883>
- BÜHLMANN, P. (2002). Bootstraps for time series. *Statist. Sci.* **17** 52–72. MR1910074 <https://doi.org/10.1214/ss/1023798998>
- CAI, T., LIU, W. and LUO, X. (2011). A constrained  $\ell_1$  minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973 <https://doi.org/10.1198/jasa.2011.tm10155>
- CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ . *Ann. Statist.* **35** 2313–2351. MR2382644 <https://doi.org/10.1214/009053606000001523>
- CHEN, X., XU, M. and WU, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Ann. Statist.* **41** 2994–3021. MR3161455 <https://doi.org/10.1214/13-AOS1182>
- DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916 <https://doi.org/10.1214/aos/1034276620>
- DAHLHAUS, R. (2000a). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762–1794. MR1835040 <https://doi.org/10.1214/aos/1015957480>
- DAHLHAUS, R. (2000b). Graphical interaction models for multivariate time series. *Metrika* **51** 157–172. MR1790930 <https://doi.org/10.1007/s001840000055>
- DAHLHAUS, R. (2012). Locally stationary processes. *Handb. Statist.* **30** 351–412.
- DAHLHAUS, R. and POLONIK, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli* **15** 1–39. MR2546797 <https://doi.org/10.3150/08-BEJ137>

- DAHLHAUS, R., RICHTER, S. and WU, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044. MR3920364 <https://doi.org/10.3150/17-bej1011>
- DAHLHAUS, R. and SUBBA RAO, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34** 1075–1114. MR2278352 <https://doi.org/10.1214/009053606000000227>
- DAHLHAUS, R. and SUBBA RAO, S. (2007). A recursive online algorithm for the estimation of time-varying ARCH parameters. *Bernoulli* **13** 389–422. MR2331257 <https://doi.org/10.3150/07-BEJ5009>
- EICHLER, M. (2007). Granger causality and path diagrams for multivariate time series. *J. Econometrics* **137** 334–353. MR2354948 <https://doi.org/10.1016/j.jeconom.2005.06.032>
- EICHLER, M. (2012). Graphical modelling of multivariate time series. *Probab. Theory Related Fields* **153** 233–268. MR2925574 <https://doi.org/10.1007/s00440-011-0345-8>
- EICHLER, M., DAHLHAUS, R. and SANDKÜHLER, J. (2003). Partial correlation analysis for the identification of synaptic connections. *Biol. Cybernet.* **89** 289–302. <https://doi.org/10.1007/s00422-003-0400-3>
- FIECAS, M., LENG, C., LIU, W. and YU, Y. (2019). Spectral analysis of high-dimensional time series. *Electron. J. Stat.* **13** 4079–4101. MR4017528 <https://doi.org/10.1214/19-EJS1621>
- FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Rev. Econ. Stat.* **82** 540–554.
- FRIED, R. and DIDELEZ, V. (2003). Decomposability and selection of graphical models for multivariate time series. *Biometrika* **90** 251–267. MR1986645 <https://doi.org/10.1093/biomet/90.2.251>
- FRYZLEWICZ, P. and SUBBA RAO, S. (2011). Mixing properties of ARCH and time-varying ARCH processes. *Bernoulli* **17** 320–346. MR2797994 <https://doi.org/10.3150/10-BEJ270>
- GATHER, U., IMHOFF, M. and FRIED, R. (2002). Graphical models for multivariate time series from intensive care monitoring. *Stat. Med.* **21** 2685–2701.
- GIURCANU, M. and SPOKOINY, V. (2004). Confidence estimation of the covariance function of stationary and locally stationary processes. *Statist. Decisions* **22** 283–300. MR2158265 <https://doi.org/10.1524/std.22.4.283.64315>
- GRENIER, Y. (1983). Time-dependent ARMA modeling of nonstationary signals. *IEEE Trans. Acoust. Speech Signal Process.* **31** 899–911.
- HAFNER, C. M. and LINTON, O. (2010). Efficient estimation of a multivariate multiplicative volatility model. *J. Econometrics* **159** 55–73. MR2720844 <https://doi.org/10.1016/j.jeconom.2010.04.007>
- HANNAN, E. J. (1974). The uniform convergence of autocovariances. *Ann. Statist.* **2** 803–806. MR0365961
- HANNAN, E. J. and DEISTLER, M. (1988). *The Statistical Theory of Linear Systems*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Wiley, New York. MR0940698
- HANSON, D. L. and WRIGHT, F. T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.* **42** 1079–1083. MR0279864 <https://doi.org/10.1214/aoms/1177693335>
- HOOTEN, M. B. and WIKLE, C. K. (2008). A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian collared-dove. *Environ. Ecol. Stat.* **15** 59–70. MR2412679 <https://doi.org/10.1007/s10651-007-0040-1>
- JACQUIER, E., POLSON, N. G. and ROSSI, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *J. Econometrics* **122** 185–212. MR2083256 <https://doi.org/10.1016/j.jeconom.2003.09.001>
- JIRAK, M. (2011). On the maximum of covariance estimators. *J. Multivariate Anal.* **102** 1032–1046. MR2793874 <https://doi.org/10.1016/j.jmva.2011.02.003>
- KAKIZAWA, Y. (2007). Moderate deviations for quadratic forms in Gaussian stationary processes. *J. Multivariate Anal.* **98** 992–1017. MR2325456 <https://doi.org/10.1016/j.jmva.2006.07.004>
- KONDRASHOV, D., KRAVTSOV, S., ROBERTSON, A. W. and GHIL, M. (2005). A hierarchy of data-based ENSO models. *J. Climate* **18** 4425–4444.
- LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer Series in Statistics. Springer, New York. MR2001447 <https://doi.org/10.1007/978-1-4757-3803-2>
- LENNARTZ, C., SCHIEFER, J., ROTTER, S., HENNIG, J. and LEVAN, P. (2018). Sparse estimation of resting-state effective connectivity from fMRI cross-spectra. *Front. Neurosci.* **12** 287. <https://doi.org/10.3389/fnins.2018.00287>
- LINDQUIST, M. A., XU, Y., NEBEL, M. B. and CAFFO, B. S. (2014). Evaluating dynamic bivariate correlations in resting-state fMRI: A comparison study and a new approach. *NeuroImage* **101** 531–546.
- LIU, C., GAETZ, W. and ZHU, H. (2010). Estimation of time-varying coherence and its application in understanding brain functional connectivity. *EURASIP Journal on Advances in Signal Processing* **2010**.
- MAJDA, A. J., ABRAMOV, R. V. and GROTE, M. J. (2005). *Information Theory and Stochastics for Multiscale Nonlinear Systems*. CRM Monograph Series **25**. Amer. Math. Soc., Providence, RI. MR2166171
- MALLAT, S., PAPANICOLAOU, G. and ZHANG, Z. (1998). Adaptive covariance estimation of locally stationary processes. *Ann. Statist.* **26** 1–47. MR1611808 <https://doi.org/10.1214/aos/1030563977>

- MEDKOUR, T., WALDEN, A. T. and BURGESS, A. (2009). Graphical modelling for brain connectivity via partial coherence. *J. Neurosci. Methods* **180** 374–383.
- MOULINES, E., PRIOURET, P. and ROUEFF, F. (2005). On recursive estimation for time varying autoregressive processes. *Ann. Statist.* **33** 2610–2654. MR2253097 <https://doi.org/10.1214/009053605000000624>
- NASON, G. P., VON SACHS, R. and KROISANDT, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 271–292. MR1749539 <https://doi.org/10.1111/1467-9868.00231>
- NEWAY, W. K. and WEST, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708. MR0890864 <https://doi.org/10.2307/1913610>
- OMBAO, H. and VAN BELLEGEM, S. (2008). Evolutionary coherence of nonstationary signals. *IEEE Trans. Signal Process.* **56** 2259–2266. MR2516630 <https://doi.org/10.1109/TSP.2007.914341>
- OMBAO, H., VON SACHS, R. and GUO, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Amer. Statist. Assoc.* **100** 519–531. MR2160556 <https://doi.org/10.1198/016214504000001448>
- OMBAO, H. C., RAZ, J. A., VON SACHS, R. and MALOW, B. A. (2001). Automatic statistical analysis of bivariate nonstationary time series. *J. Amer. Statist. Assoc.* **96** 543–560. MR1946424 <https://doi.org/10.1198/016214501753168244>
- PAPARODITIS, E. and POLITIS, D. N. (2012). Nonlinear spectral density estimation: Thresholding the correlogram. *J. Time Series Anal.* **33** 386–397. MR2915091 <https://doi.org/10.1111/j.1467-9892.2011.00771.x>
- PARK, T., ECKLEY, I. A. and OMBAO, H. C. (2014). Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Trans. Signal Process.* **62** 5240–5250. MR3268108 <https://doi.org/10.1109/TSP.2014.2343937>
- POLITIS, D. N. (2011). Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory* **27** 703–744. MR2822363 <https://doi.org/10.1017/S0266466610000484>
- POLITIS, D. N. and ROMANO, J. P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Series Anal.* **16** 67–103. MR1323618 <https://doi.org/10.1111/j.1467-9892.1995.tb00223.x>
- POLITIS, D. N. and ROMANO, J. P. (1999). Multivariate density estimation with general flat-top kernels of infinite order. *J. Multivariate Anal.* **68** 1–25. MR1668848 <https://doi.org/10.1006/jmva.1998.1774>
- POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling. Springer Series in Statistics.* Springer, New York. MR1707286 <https://doi.org/10.1007/978-1-4612-1554-7>
- PRADO, R., WEST, M. and KRISTAL, A. D. (2001). Multichannel electroencephalographic analyses via dynamic regression models with time-varying lag–lead structure. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **50** 95–109.
- PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series. Vol. 1.* Academic Press, London. MR0628735
- PRIESTLEY, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis.* Academic Press, London. MR0991969
- PRIESTLEY, M. B. and TONG, H. (1973). On the analysis of bivariate non-stationary processes. *J. Roy. Statist. Soc. Ser. B* **35** 153–166, 179–188. MR0336946
- ROSENBLATT, M. (1971). *Markov Processes. Structure and Asymptotic Behavior.* Springer, New York. MR0329037
- ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields.* Birkhäuser, Inc., Boston, MA. MR0885090 <https://doi.org/10.1007/978-1-4612-5156-9>
- RUDELSON, M. and VERSHYNIN, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82, 9. MR3125258 <https://doi.org/10.1214/ECP.v18-2865>
- SALVADOR, R., SUCKLING, J., SCHWARZBAUER, C. and BULLMORE, E. (2005). Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **360** 937–946.
- SANDERSON, J., FRYZLEWICZ, P. and JONES, M. W. (2010). Estimating linear dependence between non-stationary time series using the locally stationary wavelet model. *Biometrika* **97** 435–446. MR2650749 <https://doi.org/10.1093/biomet/asq007>
- SIMPSON, S. L., BOWMAN, F. D. and LAURIENTI, P. J. (2013). Analyzing complex functional brain networks: Fusing statistics and network science to understand the brain. *Stat. Surv.* **7** 1–36. MR3161730 <https://doi.org/10.1214/13-SS103>
- SUBBA RAO, T. (1970). The fitting of non-stationary time-series models with time-dependent parameters. *J. Roy. Statist. Soc. Ser. B* **32** 312–322. MR0269065
- SUN, Y., LI, Y., KUCEYESKI, A. and BASU, S. (2018). Large spectral density matrix estimation by thresholding. Preprint. Available at arXiv:1812.00532.
- TIMMER, J., LAUK, M., HÄUSSLER, S., RADT, V., KÖSTER, B., HELLWIG, B., GUSCHLBAUER, B., LÜCKING, C. H., EICHLER, M. et al. (2000). Cross-spectral analysis of tremor time series. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **10** 2595–2610.

- TONG, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford Statistical Science Series **6**. The Clarendon Press, Oxford University Press, New York. MR1079320
- TSAY, R. S. (2005). *Analysis of Financial Time Series*, 2nd ed. Wiley Series in Probability and Statistics. Wiley Interscience, Hoboken, NJ. MR2162112 <https://doi.org/10.1002/0471746193>
- VOGT, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40** 2601–2633. MR3097614 <https://doi.org/10.1214/12-AOS1043>
- WIENER, N. (1958). *Nonlinear Problems in Random Theory*. Technology Press Research Monographs. Wiley, New York. MR0100912
- WIKLE, C. K. and HOOTEN, M. B. (2010). A general science-based framework for dynamical spatio-temporal models. *TEST* **19** 417–451. MR2745992 <https://doi.org/10.1007/s11749-010-0209-z>
- WRIGHT, F. T. (1973). A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Ann. Probab.* **1** 1068–1070. MR0353419 <https://doi.org/10.1214/aop/1176996815>
- WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 <https://doi.org/10.1073/pnas.0506715102>
- WU, W. B. and SHAO, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436. MR2052582 <https://doi.org/10.1239/jap/1082999076>
- XIAO, H. and WU, W. B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.* **40** 466–493. MR3014314 <https://doi.org/10.1214/11-AOS967>
- XIAO, H. and WU, W. B. (2014). Portmanteau test and simultaneous inference for serial covariances. *Statist. Sinica* **24** 577–599. MR3235390
- ZANI, M. (2002). Large deviations for quadratic forms of locally stationary processes. *J. Multivariate Anal.* **81** 205–228. MR1906377 <https://doi.org/10.1006/jmva.2001.2003>
- ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156 <https://doi.org/10.1214/16-AOS1512>
- ZHANG, D. and WU, W. B. (2021). Supplement to “Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes.” <https://doi.org/10.1214/20-AOS1954SUPP>.
- ZHOU, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *Ann. Statist.* **38** 2187–2217. MR2676887 <https://doi.org/10.1214/09-AOS769>
- ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444 <https://doi.org/10.1214/08-AOS636>