

Gaussian approximation for high dimensional vector under physical dependence

XIANYANG ZHANG¹ and GUANG CHENG²

¹*Department of Statistics, Texas A&M University, College Station, TX 77843, USA.
E-mail: zhangxiany@stat.tamu.edu*

²*Department of Statistics, Purdue University, West Lafayette, IN 47906, USA.
E-mail: chengg@purdue.edu*

We develop a Gaussian approximation result for the maximum of a sum of weakly dependent vectors, where the data dimension is allowed to be exponentially larger than sample size. Our result is established under the physical/functional dependence framework. This work can be viewed as a substantive extension of Chernozhukov et al. (*Ann. Statist.* **41** (2013) 2786–2819) to time series based on a variant of Stein’s method developed therein.

Keywords: Gaussian approximation; high dimensionality; physical dependence measure; Slepian interpolation; Stein’s method; time series

1. Introduction

Let $\{\epsilon_i\}_{i \in \mathbb{Z}}$ be independent and identically distributed (i.i.d.) random elements. Consider a p -dimensional random vector with the following causal representation:

$$x_i := (x_{i1}, \dots, x_{ip})' = \mathcal{G}_i(\dots, \epsilon_{i-1}, \epsilon_i), \quad (1)$$

where $\mathcal{G}_i = (\mathcal{G}_{i1}, \dots, \mathcal{G}_{ip})'$ is a measurable function such that x_i is well defined. Let $y_i = (y_{i1}, \dots, y_{ip})'$ be a Gaussian sequence which is independent of x_i and preserves the autocovariance structure of x_i .¹ Suppose $\mathbb{E}x_i = \mathbb{E}y_i = 0$. The major goal of this paper is to quantify the Kolmogorov distance between T_X and T_Y :

$$\rho_n := \sup_{t \in \mathbb{R}} |P(T_X \leq t) - P(T_Y \leq t)|, \quad (2)$$

where $T_X = \max_{1 \leq j \leq p} X_j$, $T_Y = \max_{1 \leq j \leq p} Y_j$, and

$$X = (X_1, \dots, X_p)' = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i, \quad Y = (Y_1, \dots, Y_p)' = \frac{1}{\sqrt{n}} \sum_{i=1}^n y_i.$$

Here, n is the sample size and p is allowed to be exponentially larger than n . Throughout this paper, $\{x_i\}$ is not necessarily assumed to be stationary (as \mathcal{G}_i is allowed to change with i).

¹This assumption can be relaxed, see more details in Remark 2.1.

The distribution of T_X is of great interest in high dimensional statistical inference such as model selection, simultaneous inference, and multiple testing [6–8,25,32]. When p increases slowly with n , the convergence of ρ_n to zero follows from the multivariate central limit theorem with growing dimension, see, for example, [5,16,23]. When $p = O(\exp(n^c))$ for some $c > 0$, Chernozhukov et al. [12] recently showed that ρ_n decays to zero at a polynomial rate if $\{x_i\}$ is an *independent* sequence. This result provides an astounding improvement over the previous results in [5] by allowing the data dimension to diverge exponentially fast. In this paper, we shall establish a similar high dimensional Gaussian approximation result in the more general setup where x_i admits the causal representation (1). It is worth pointing out that our results require non-trivial modifications of the technical tools developed in [12] in order to overcome the difficulties arising from the dependence across data vectors. In particular, we develop some new techniques in dealing with high dimensional dependent data such as the use of dependency graph, leave one-block-out argument, self-normalization and M -dependent approximation, which are also of interest in their own right.

To quantify the strength of dependence for time series, we adapt the physical dependence measure in [27,30] for low dimensional time series to the high dimensional setting. Specifically, we allow the structure of the physical system or filter $\mathcal{G}_i = \mathcal{G}_{i,n}$ to change with sample size, that is, we are dealing with triangular array. Compared to the classical mixing type conditions which involve complicated manipulation of taking the supremum over two sigma algebras, the framework of physical dependence (or its variants) is known to be very general and easy to verify for both linear and nonlinear data-generating mechanisms. One example given in [27] is a simple AR(1) process $X_i = (X_{i-1} + \epsilon_i)/2$, where ϵ_i are i.i.d. Bernoulli random variables with success probability $1/2$. The process X_i is not strong mixing [1], while it can be conveniently studied under the framework of physical dependence [27] as it admits the causal representation $X_i = \sum_{j=0}^{+\infty} 2^{-(j+1)} \epsilon_{i-j}$. We also remark that the physical dependence measure and mixing type conditions do not nest each other. Our results thus complement [11] which established a Gaussian approximation result for β -mixing time series around the same time when this manuscript was under preparation. While our work is being carried out, we note an arXiv work [31] which establishes the Gaussian approximation theory for *stationary* high-dimensional time series under different physical dependence assumptions.

Finally, we point out that although high dimensional statistics has witnessed unprecedented development, statistical inference for high dimensional time series remains largely untouched so far. The Gaussian approximation theory developed in this paper represents an initial step along this direction. In particular, it provides a theoretical framework in studying high dimensional bootstrap that works even when the autocovariance structure of $\{x_i\}$ is unknown. Also see [10, 14,17,18] for some other recent studies on high dimensional time series.

The rest of the article is organized as follows. Section 2.1 establishes a general result in the framework of dependency graph, which leads to delicate bounds in Section 2.2 on the Kolmogorov distance for weakly dependent time series under physical dependence. Some concrete examples such as non-stationary linear models and GARCH models are studied in Section 2.3, while Section 3 presents some numerical results. All the proofs are gathered in the Appendix.

Let $|\cdot| := |\cdot|_q$ be the Euclidean norm of \mathbb{R}^q . Denote by $C^k(\mathbb{R})$ the class of k times continuously differentiable functions from \mathbb{R} to itself, and denote by $C_b^k(\mathbb{R})$ a sub-class of $C^k(\mathbb{R})$ such that $\sup_{z \in \mathbb{R}} |\partial^j f(z)/\partial z^j| < \infty$ for $j = 0, 1, \dots, k$. For a sequence of random variables $\{z_i\}_{i=1}^n$, define

$\bar{\mathbb{E}}[z_i] = \sum_{i=1}^n \mathbb{E}z_i/n$. For a random variable z , let $\|z\|_q = (\mathbb{E}|z|^q)^{1/q}$. Write $a \lesssim b$ if a is smaller than or equal to b up to a universal positive constant. For two sequences a_n and b_n , denote by $a_n \asymp b_n$, if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. For $a, b \in \mathbb{R}$, let $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For two matrices \mathbf{A} and \mathbf{B} , denote by $\mathbf{A} \otimes \mathbf{B}$ their Kronecker product.

2. Gaussian approximation theory

2.1. Dependency graph

In this subsection, we introduce a generic framework in modeling the dependence among a sequence of p -dimensional (*not necessarily identically distributed*) random vectors $\{x_i\}_{i=1}^n$. We call it as dependency graph $G_n = (V_n, E_n)$, where $V_n = \{1, 2, \dots, n\}$ is a set of vertices and E_n is the corresponding set of undirected edges. For any two disjoint subsets of vertices $S, T \subseteq V_n$, if there is no edge from any vertex in S to any vertex in T , the collections $\{x_i\}_{i \in S}$ and $\{x_i\}_{i \in T}$ are independent. Let $D_{\max, n} = \max_{1 \leq i \leq n} \sum_{j=1}^n \mathbf{I}\{\{i, j\} \in E_n\}$ be the maximum degree of G_n and denote $D_n = 1 + D_{\max, n}$. Throughout the paper, we allow D_n to grow with the sample size n . For example, if an array $\{x_{i,n}\}_{i=1}^n$ is a $M := M_n$ dependent sequence (i.e., $x_{i,n}$ and $x_{j,n}$ are independent if $|i - j| > M$), then we have $D_n = 2M + 1$. Within this framework, the dependence structure of the underlying sequence is directly associated with the graph G_n , which allows a more general characterization for various forms of dependence such as temporal dependence, spatial dependence and dependence driven by other variables. In the low-dimensional setting, the dependency graph has been used for studying central limit theorem of dependent data when D_n is not too large; see [2–4, 21]. To further illustrate this concept, we provide the following example.

Example 2.1 (U -statistics). Let $\{\varepsilon_i\}_{i=1}^n$ be n i.i.d. random variables. Given a symmetric function $h(\cdot, \dots, \cdot)$ on \mathbb{R}^{m_0} , the degenerate U -statistic is defined as $\binom{n}{m_0}^{-1} \sum h(\varepsilon_{i_1}, \dots, \varepsilon_{i_{m_0}})$, where the sum extends over all $\binom{n}{m_0}$ subsets of indices from $\{1, 2, \dots, n\}$. Let $x_i = h(\varepsilon_{i_1}, \dots, \varepsilon_{i_{m_0}})$ with $i = \{i_1, \dots, i_{m_0}\}$. The dependence of x_i can be characterized through the corresponding dependence graph. Specifically, $\{i, j\} \in E$ with $i = \{i_1, \dots, i_{m_0}\}$ and $j = \{j_1, \dots, j_{m_0}\}$ if and only if $i \cap j \neq \emptyset$.

Recall that $T_X = \max_{1 \leq j \leq p} X_j$ and $T_Y = \max_{1 \leq j \leq p} Y_j$. The problem of comparing distributions of maxima is nontrivial since the maximum function $z = (z_1, \dots, z_p)' \rightarrow \max_{1 \leq j \leq p} z_j$ is non-differentiable. To overcome this difficulty, we consider a smooth approximation of the maximum function,

$$F_\beta(z) := \beta^{-1} \log \left(\sum_{j=1}^p \exp(\beta z_j) \right), \quad z = (z_1, \dots, z_p)',$$

where $\beta > 0$ is the smoothing parameter that controls the level of approximation. Simple algebra yields that (see [9]),

$$0 \leq F_\beta(z) - \max_{1 \leq j \leq p} z_j \leq \beta^{-1} \log p. \quad (3)$$

To handle unbounded random variables, we employ the truncation argument which is slightly different from the one used in [12]. Denote the truncated variables $\tilde{x}_{ij} = (x_{ij} \wedge M_x) \vee (-M_x) - \mathbb{E}[(x_{ij} \wedge M_x) \vee (-M_x)]$ and $\tilde{y}_{ij} = (y_{ij} \wedge M_y) \vee (-M_y) - \mathbb{E}[(y_{ij} \wedge M_y) \vee (-M_y)]$ for some $M_x, M_y > 0$. Note that the map $x \rightarrow (x \wedge M_x) \vee (-M_x)$ is Lipschitz continuous which facilitates our derivations in Section 2.2. Let $\tilde{x}_i = (\tilde{x}_{i1}, \dots, \tilde{x}_{ip})'$ and $\tilde{y}_i = (\tilde{y}_{i1}, \dots, \tilde{y}_{ip})'$. For $1 \leq i \leq n$, let $N_i = \{j : \{i, j\} \in E_n\}$ be the set of neighbors of i , and $\tilde{N}_i = \{i\} \cup N_i$. Let $\phi(M_x)$ be a constant depending on the threshold parameter M_x such that

$$\max_{1 \leq j, k \leq p} \frac{1}{n} \sum_{i=1}^n \left| \sum_{l \in \tilde{N}_i} (\mathbb{E} x_{ij} x_{lk} - \mathbb{E} \tilde{x}_{ij} \tilde{x}_{lk}) \right| \leq \phi(M_x).$$

Analogous quantity $\phi(M_y)$ can be defined for $\{y_i\}$. Set $\phi(M_x, M_y) = \phi(M_x) + \phi(M_y)$. Define

$$m_{x,k} = \left(\bar{\mathbb{E}} \max_{1 \leq j \leq p} |x_{ij}|^k \right)^{1/k}, \quad m_{y,k} = \left(\bar{\mathbb{E}} \max_{1 \leq j \leq p} |y_{ij}|^k \right)^{1/k},$$

$$\bar{m}_{x,k} = \max_{1 \leq j \leq p} \left(\bar{\mathbb{E}} |x_{ij}|^k \right)^{1/k}, \quad \bar{m}_{y,k} = \max_{1 \leq j \leq p} \left(\bar{\mathbb{E}} |y_{ij}|^k \right)^{1/k}.$$

Note that $\bar{m}_{x,k} \leq m_{x,k}$ and $\bar{m}_{y,k} \leq m_{y,k}$. Further define an indicator function,

$$\mathcal{I} := \mathcal{I}_\Delta = \mathbf{1} \left\{ \max_{1 \leq j \leq p} |X_j - \tilde{X}_j| \leq \Delta, \max_{1 \leq j \leq p} |Y_j - \tilde{Y}_j| \leq \Delta \right\},$$

where $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_p)' = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{x}_i$ and $\tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_p)' = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{y}_i$.

We next approximate the indicator function $I\{\cdot \leq t\}$ by a suitable smooth function $g(\cdot) \in C_b^3(\mathbb{R})$, and thus set $m(\cdot) = g \circ F_\beta(\cdot)$. In particular, g can be chosen as the convolution between $I\{\cdot \leq t\}$ and a Gaussian density. As an intermediate step, we derive in Proposition 2.1 a non-asymptotic upper bound for the quantity $|\mathbb{E}[m(X) - m(Y)]|$ using the Slepian interpolation [24]. The proof of Proposition 2.1 generalizes that of Theorem 2.1 in [12] by modifying Stein's leave-one-out argument [26] to the leave-one-block-out argument that captures the local dependence of the data.

Proposition 2.1. Assume that $2\sqrt{5}\beta D_n^2 M_{xy}/\sqrt{n} \leq 1$ with $M_{xy} = \max\{M_x, M_y\}$. Then we have for any $\Delta > 0$,

$$\begin{aligned} & |\mathbb{E}[m(X) - m(Y)]| \\ & \lesssim (G_2 + G_1\beta)\phi(M_x, M_y) + (G_3 + G_2\beta + G_1\beta^2) \frac{D_n^2}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3) \\ & \quad + (G_3 + G_2\beta + G_1\beta^2) \frac{D_n^3}{\sqrt{n}} (m_{x,3}^3 + m_{y,3}^3) + G_1\Delta + G_0\mathbb{E}[1 - \mathcal{I}], \end{aligned} \tag{4}$$

where $G_k = \sup_{z \in \mathbb{R}} |\partial^k g(z)/\partial z^k|$ for $k \geq 0$. In addition, if $2\sqrt{5}\beta D_n^3 M_{xy}/\sqrt{n} \leq 1$, we can replace $m_{x,3}^3 + m_{y,3}^3$ by $\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3$ in (4).

When specializing to a M -dependent sequence, we obtain the following result.

Corollary 2.1. *When $\{x_i\}$ is a M -dependent sequence, under the assumption that $2\sqrt{5}\beta(6M+1)M_{xy}/\sqrt{n} \leq 1$, we have*

$$\begin{aligned} |\mathbb{E}[m(X) - m(Y)]| &\lesssim (G_3 + G_2\beta + G_1\beta^2) \frac{(2M+1)^2}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3) \\ &\quad + (G_2 + G_1\beta)\phi(M_x, M_y) + G_1\Delta + G_0\mathbb{E}[1 - \mathcal{I}]. \end{aligned} \quad (5)$$

We remark that the upper bound in (5) can be further simplified using the self-normalization technique (see Lemma A.1) and certain arguments under weak dependence assumption.

Considering the approximation properties of F_β and g , Proposition 2.1 leads to an upper bound on the Kolmogorov distance ρ_n defined in (2). In particular, we obtain an explicit upper bound of ρ_n for the M -dependence sequence based on Corollary 2.1. Such a result is viewed as an intermediate one, and thus deferred to Section A.2.

Remark 2.1. In view of the proof of Proposition 2.1 (see, e.g., (31)), the assumption that $\{y_i\}$ preserves the autocovariance structure of $\{x_i\}$ can be weakened by assuming that for all i ,

$$\sum_{k \in \tilde{N}_i} \mathbb{E}x_i x'_k = \sum_{k \in \tilde{N}_i} \mathbb{E}y_i y'_k.$$

Thus $\{y_i\}$ is allowed to be a sequence of *independent* (mean-zero) p -dimensional Gaussian random variables such that $\text{cov}(y_i) = \sum_{k \in \tilde{N}_i} \mathbb{E}x_i x'_k$ provided that $\sum_{k \in \tilde{N}_i} \mathbb{E}x_i x'_k$ is positive-definite. In fact, when $\{x_i\}$ is M -dependent and stationary, we can construct $\{y_i\}$ as i.i.d. Gaussian sequence with the covariance $\sum_{j=i-M}^{i+M} \mathbb{E}x_i x'_j$. In this case, we need to replace $\phi(M_x, M_y)$ in (5) by $\phi(M_x, M_y) + \max_{1 \leq j, k \leq p} \sum_{l=1}^M l |\mathbb{E}x_{ij} x_{i+l, k}| / n$ due to the edge effect.

2.2. Weakly dependent time series under physical dependence

In this subsection, we shall develop Gaussian approximation theory for weakly dependent time series, which is the major interest of this paper. To this end, we need to introduce suitable dependence measure for high dimensional vector. We adapt the concept of physical dependence measure for non-stationary causal process in [30] to the high-dimensional setting for its broad applicability to linear and nonlinear processes as well as its theoretical convenience and elegance.

Recall that $x_i = \mathcal{G}_i(\mathcal{F}_i)$, where $\mathcal{F}_i = (\dots, \epsilon_{i-1}, \epsilon_i)$ and $\mathcal{G}_i = (\mathcal{G}_{i1}, \dots, \mathcal{G}_{ip})'$. To measure the strength of dependence, we let $\{\epsilon'_i\}$ be an i.i.d. copy of $\{\epsilon_i\}$. For $\|x_{ij}\|_q < \infty$, define

$$\theta_{k,j,q} = \sup_i \|\mathcal{G}_{ij}(\mathcal{F}_i) - \mathcal{G}_{ij}(\mathcal{F}_{i,i-k})\|_q, \quad \Theta_{k,j,q} = \sum_{l=k}^{+\infty} \theta_{l,j,q}, \quad (6)$$

where $\mathcal{F}_{i,k} = (\dots, \epsilon_{k-1}, \epsilon'_k, \epsilon_{k+1}, \dots, \epsilon_{i-1}, \epsilon_i)$ is a coupled version of \mathcal{F}_i . In the subsequent discussions, we assume that the dependence measure $\sup_{1 \leq j \leq p} \Theta_{k,j,q} < \infty$ for some $q > 0$. We

point out that the dimension of \mathcal{G}_i (i.e., p) is allowed to grow with n , which makes our setting different from the one in [30].

Before presenting our main result, we introduce the following assumptions which will be verified under specific models in Section 2.3. Let $h : [0, +\infty) \rightarrow [0, +\infty)$ be a strictly increasing convex function with $h(0) = 0$. Denote by $h^{-1}(\cdot)$ the inverse function of $h(\cdot)$. Let $l_n := l_n(p, \gamma) = \log(pn/\gamma) \vee 1$. Define $\sigma_{j,k} = \text{cov}(X_j, X_k) = \sum_{i,l=1}^n \text{cov}(x_{ij}, x_{lk})/n$.

Assumption 2.1. Assume that $\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} \mathbb{E}x_{ij}^4 < c_1$ for $c_1 > 0$ and there exists $\mathfrak{D}_n > 0$ such that one of the following two conditions holds

$$\max_{1 \leq i \leq n} \mathbb{E}h\left(\max_{1 \leq j \leq p} |x_{ij}|/\mathfrak{D}_n\right) \leq C_1, \quad (7)$$

$$\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} \mathbb{E} \exp(|x_{ij}|/\mathfrak{D}_n) \leq C_2, \quad (8)$$

for $C_1, C_2 > 0$.

Assumption 2.2. Assume there exist $M = M(n) > 0$ and $\gamma = \gamma(n) \in (0, 1)$ such that

$$n^{3/8} M^{-1/2} l_n^{-5/8} \geq C_3 \max\{\mathfrak{D}_n h^{-1}(n/\gamma), l_n^{1/2}\} \quad \text{under Condition (7),}$$

$$n^{3/8} M^{-1/2} l_n^{-5/8} \geq C_4 \max\{\mathfrak{D}_n l_n, l_n^{1/2}\} \quad \text{under Condition (8),}$$

for $C_3, C_4 > 0$, where \mathfrak{D}_n is given in Assumption 2.1.

Assumption 2.2 imposes constraints on the intermediate quantities M , l_n and γ so that the upper bound in (11) holds. These quantities are later on chosen to optimize the upper bound. We remark that the quantity M in Assumption 2.2 corresponds to an M -dependent sequence used in the proof of Theorem 2.1 for approximating the weakly dependent sequence $\{x_i\}$; see the end of this section. A larger value of M leads to a better approximation to the original data sequence, but also to an increasing upper bound given in Corollary 2.1. Hence, a proper choice of M is needed.

Assumption 2.3. Assume that

$$c_1 < \min_{1 \leq j \leq p} \sigma_{j,j} \leq \max_{1 \leq j \leq p} \sigma_{j,j} < c_2, \quad (9)$$

$$\sum_{j=1}^{+\infty} \max_{1 \leq k \leq p} j \theta_{j,k,3} < c_3, \quad (10)$$

for some constants $0 < c_1 < c_2$ and $c_3 > 0$.

Note (10) and the condition that $\max_{1 \leq j \leq p} \frac{1}{n} \sum_{i=1}^n \|x_{ij}\|_2^2 < c$ for some $c > 0$ imply the second part of (9), that is, $\max_{1 \leq j \leq p} \sigma_{j,j} < c_2$.

We are now in position to present the main result of this paper.

Theorem 2.1. Under Assumptions 2.1–2.3, we have for $q \geq 2$,

$$\begin{aligned} \rho_n \lesssim & n^{-1/8} M^{1/2} l_n^{7/8} + \gamma + (n^{1/8} M^{-1/2} l_n^{-3/8})^{q/(1+q)} \left(\sum_{j=1}^p \Theta_{M,j,q}^q \right)^{1/(1+q)} \\ & + \Xi_M^{1/3} (1 \vee \log(p/\Xi_M))^{2/3}, \end{aligned} \quad (11)$$

where $\Xi_M = \max_{1 \leq k \leq p} \sum_{j=M}^{+\infty} j \theta_{j,k,2}(x)$, and M and γ satisfy Assumption 2.2.

The key strategy in the proof of Theorem 2.1 is M -dependent approximation, which will be sketched in the end of this section.

Note that the conditions in Theorem 2.1 can be categorized into two types: tail restrictions and weak dependence assumptions. Assumptions 2.1 and 2.2 impose restrictions on the tail of $\{x_{ij}\}$ uniformly across j , which are needed even in the independence case [12]. Note that Assumption 2.1 is satisfied if $x_{ij} = \mathfrak{D}_n \zeta_{ij}$ with $\max_{i,j} \mathbb{E} \exp(|\zeta_{ij}|) \leq C_2$. Assumption 2.3 essentially requires weak dependence uniformly across all the components of $\{x_i\}$. We verify (10) for both linear and nonlinear time series models in the next section.

Under the assumption that $p \lesssim \exp(n^b)$ for $0 \leq b < 1/11$, we obtain that $\rho_n \lesssim n^{-(1-11b)/8}$ in Corollary 2.2 by optimizing the upper bound in (11) w.r.t. M and γ . In fact, the optimization is achieved when $\gamma \asymp n^{-(1-11b)/8}$ and $M = Cn^b$ for some large enough C .

Corollary 2.2. Assume that $\max_{1 \leq j \leq p} \Theta_{u,j,q} \lesssim \varrho^\mu$ for some $\varrho < 1$ and $q \geq 2$, and $p \lesssim \exp(n^b)$ for some $0 \leq b < 1/11$. Suppose one of the following two conditions holds

$$\max_{1 \leq i \leq n} \mathbb{E} \left(\max_{1 \leq j \leq p} |x_{ij}| / \mathfrak{D}_n \right)^4 \leq C_1, \quad \mathfrak{D}_n \lesssim n^{(3-25b)/32}, \quad (12)$$

$$\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} \mathbb{E} \exp(|x_{ij}| / \mathfrak{D}_n) \leq C_2, \quad \mathfrak{D}_n \lesssim n^{(3-17b)/8}, \quad (13)$$

for $C_1, C_2 > 0$. Under (9), we have $\rho_n \lesssim n^{-(1-11b)/8}$.

Remark 2.2. In general, we can assume that

$$\max_{1 \leq i \leq n} \mathbb{E} \left(\max_{1 \leq j \leq p} |x_{ij}| / \mathfrak{D}_n \right)^k \leq C_1, \quad \mathfrak{D}_n \lesssim n^{\frac{(3-9b)k - (9-11b)}{8k}}. \quad (14)$$

In this case, one can again choose $\gamma \asymp n^{-(1-11b)/8}$ and $M = cn^b$ to obtain the polynomial decay rate $n^{-(1-11b)/8}$.

In the end of this section, we discuss the M -approximation technique used in the proof of Theorem 2.1. Let $x_i^{(M)} = (x_{i1}^{(M)}, \dots, x_{ip}^{(M)})' = \mathbb{E}[x_i | \epsilon_{i-M}, \dots, \epsilon_i]$ be a M -dependent approximation sequence for $\{x_i\}$. Define $X^{(M)}$ in the same way as X by replacing x_i with $x_i^{(M)}$. Because

$|m(x) - m(y)| \leq 2G_0$ and $|m(x) - m(y)| \leq G_1 \max_{1 \leq j \leq p} |x_j - y_j|$ by the Lipschitz property of F_β (see e.g. [12]), we have

$$\begin{aligned} |\mathbb{E}[m(X) - m(X^{(M)})]| &\leq |\mathbb{E}[(m(X) - m(X^{(M)}))\mathcal{I}_M]| \\ &\quad + |\mathbb{E}[(m(X) - m(X^{(M)}))(1 - \mathcal{I}_M)]| \\ &\lesssim G_1 \Delta_M + G_0 \mathbb{E}[1 - \mathcal{I}_M], \end{aligned} \quad (15)$$

where $\mathcal{I}_M := \mathcal{I}_{\Delta_M, M} = \mathbf{1}\{\max_{1 \leq j \leq p} |X_j - X_j^{(M)}| \leq \Delta_M\}$ for some $\Delta_M > 0$ depending on M . Suppose $\max_{1 \leq j \leq p} \mathbb{E}\|x_{ij}\|_q < \infty$ for all i and some $q > 0$. By Lemma A.1 of [19], we have

$$(\mathbb{E}|X_j - X_j^{(M)}|^q)^{q'/q} \leq C_q n^{1-q'/2} \Theta_{M,j,q}^{q'},$$

where $q' = \min(2, q)$ and C_q is a positive constant depending on q (note that the results in Lemma A.1 of [19] are still valid for nonstationary process in view of their arguments). For any $q \geq 2$, we obtain

$$\begin{aligned} \mathbb{E}[1 - \mathcal{I}_M] &\leq \sum_{j=1}^p P(|X_j - X_j^{(M)}| \geq \Delta_M) \leq \sum_{j=1}^p \frac{1}{\Delta_M^q} \mathbb{E}|X_j - X_j^{(M)}|^q \\ &\leq \sum_{j=1}^p \frac{C_q^{q/2} \Theta_{M,j,q}^q}{\Delta_M^q} = \sum_{j=1}^p \frac{C_q^{q/2}}{\Delta_M^q} \left(\sum_{l=M}^{+\infty} \theta_{l,j,q} \right)^q. \end{aligned}$$

Optimizing the bound with respect to Δ_M in (15), we deduce that

$$|\mathbb{E}[m(X) - m(X^{(M)})]| \lesssim (G_0 G_1^q)^{1/(1+q)} \left(\sum_{j=1}^p \Theta_{M,j,q}^q \right)^{1/(1+q)}, \quad (16)$$

which along with (3) implies that

$$|\mathbb{E}[g(T_X) - g(T_{X^{(M)}})]| \lesssim (G_0 G_1^q)^{1/(1+q)} \left(\sum_{j=1}^p \Theta_{M,j,q}^q \right)^{1/(1+q)} + \beta^{-1} G_1 \log p,$$

with $T_{X^{(M)}} = \max_{1 \leq j \leq p} \sum_{i=1}^n x_{ij}^{(M)} / \sqrt{n}$.

Remark 2.3. Our results can also be combined with the notion of dependence adjusted norm recently proposed in Zhang and Wu [31]. We merely illustrate the idea here but do not intend to obtain the sharpest possible result. Define

$$\omega_{j,q} = \max_i \|\mathcal{G}_i(\mathcal{F}_i) - \mathcal{G}_i(\mathcal{F}_{i,j})\|_\infty \|_q, \quad (17)$$

and $\Omega_{M,q} = \sum_{j=M}^{+\infty} \omega_{j,q}$. Using the Burkholder type inequality in Theorem 4.1 of Pinelis [22] [also see Lemma C.5 of Zhang and Wu [31]], we have for any $x > 0$,

$$\mathbb{E}[1 - \mathcal{I}_M] = P(|X - X^{(M)}|_{\infty} \geq x) \lesssim \frac{(\log(p))^{q/2} \Omega_{M+1,q}^q}{x^q}.$$

Choosing Δ_M to optimize the bound, we obtain

$$\begin{aligned} |\mathbb{E}[m(X) - m(X^{(M)})]| &\lesssim G_1 \Delta_M + G_0 P(|X - X^{(M)}|_{\infty} \geq \Delta_M) \\ &\lesssim G_1 \Delta_M + G_0 \frac{(\log(p))^{q/2} \Omega_{M+1,q}^q}{\Delta_M^q} \\ &\lesssim (G_0 G_1^q)^{1/(q+1)} ((\log(p))^{1/2} \Omega_{M+1,q})^{q/(1+q)}. \end{aligned}$$

Combining with the arguments in the proof of Theorem 2.1, we have

$$\begin{aligned} \rho_n &\lesssim n^{-1/8} M^{1/2} l_n^{7/8} + \gamma + (n^{1/8} M^{-1/2} l_n^{-3/8})^{q/(1+q)} ((\log(p))^{1/2} \Omega_{M+1,q})^{q/(1+q)} \\ &\quad + \Xi_M^{1/3} (1 \vee \log(p/\Xi_M))^{2/3}. \end{aligned}$$

When $p \lesssim \exp(n^b)$, suppose $\Omega_{M+1,q} \asymp M^{-\alpha}$ with $\alpha > (1+b)/(1-7b)$. Then by picking $M \asymp n^c$ for some $\max\{(1+b)/(8\alpha+4), 2b/(\alpha-1)\} < c < (1-7b)/4$, we can still obtain the polynomial decay rate.

2.3. Example

To illustrate the applicability of our general theory, we verify assumptions in several commonly used time series models.

Example 2.2 (Nonstationary linear model). Consider a nonstationary linear model

$$x_i = \sum_{l=0}^{+\infty} \mathbf{A}_{i,l} \epsilon_{i-l}, \quad (18)$$

where $\{\mathbf{A}_{i,l}\}$ is a sequence of $p \times p$ matrices with $\mathbf{A}_{i,l} = (a_{i,l,jk})_{j,k=1}^p$, and $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{ip})$ is a sequence of i.i.d. p -dimensional random vectors with $\mathbb{E}\epsilon_i = 0$. In this case, \mathcal{G}_i is a linear function on the inputs $(\dots, \epsilon_{i-1}, \epsilon_i)$. It is easy to see that $\mathcal{G}_i(\mathcal{F}_i) - \mathcal{G}_i(\mathcal{F}_{i-1}) = \mathbf{A}_{i,l}(\epsilon_{i-l} - \epsilon'_{i-l})$ which implies $\theta_{l,j,2q} = \sup_i \|\sum_{k=1}^p a_{i,l,jk}(\epsilon_{i-l,k} - \epsilon'_{i-l,k})\|_{2q}$. Suppose $\max_{1 \leq j \leq p} \|\epsilon_{0j}\|_{2q} < \infty$ for some $q > 1$, and

$$\max_{1 \leq j \leq p} \sup_i \left(\sum_{k=1}^p a_{i,l,jk}^2 \right)^{1/2} \lesssim \varrho^l \quad \text{for some } \varrho < 1.$$

By Rosenthal's inequality, we further have

$$\theta_{l,j,2q}^{2q} \lesssim \sup_i \left\{ \sum_{k=1}^p a_{i,l,jk}^2 \mathbb{E}(\epsilon_{i-l,k} - \epsilon'_{i-l,k})^2 \right\}^q \lesssim \sup_i \left(\sum_{k=1}^p a_{i,l,jk}^2 \right)^q,$$

which induces that $\max_{1 \leq j \leq p} \Theta_{u,j,2q} \lesssim \varrho^u$. Further assume

$$\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} \mathbb{E} \exp \left(\frac{1}{c} \left| \sum_{l=0}^{\infty} \sum_{k=1}^p a_{i,l,jk} \epsilon_{i-l,k} \right| \right) \leq C_1, \quad (19)$$

and $\min_{1 \leq j \leq p} \frac{1}{n} \sum_{i,k=1}^n \text{cov}(x_{ij}, x_{kj}) > c'$ for $c, c', C_1 > 0$. By Jensen's inequality, (19) holds provided that $\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} \sum_{l=0}^{\infty} \sum_{k=1}^p |a_{i,l,jk}| < \infty$ and $\max_{1 \leq j \leq p} \mathbb{E} \exp(|\epsilon_{0j}|/c'') \leq C_2$ for some $c'', C_2 > 0$. Then by Corollary 2.2, we have $\rho_n \lesssim n^{-(1-11b)/8}$ for $p \lesssim \exp(n^b)$ with $b < 1/11$.

Example 2.3 (Random coefficient autoregressive process). Let \mathbf{A}_i be a $p \times p$ random matrix and B_i be a $p \times 1$ random vector. Define a random coefficient autoregressive process as

$$x_i = \mathbf{A}_i x_{i-1} + B_i, \quad (20)$$

where (\mathbf{A}_i, B_i) are i.i.d. which ensures that $\{x_i\}$ is stationary. It can be shown that x_i has a causal representation $x_i = \mathcal{G}(\dots, \epsilon_{i-1}, \epsilon_i)$ for $\epsilon_i = (\mathbf{A}_i, B_i)$. Note that Bilinear and GARCH models fall within the framework of (20). We assume that \mathbf{A}_i is block diagonal,² that is,

$$\mathbf{A}_i = \begin{pmatrix} \mathbf{A}_{i1} & & & \\ & \mathbf{A}_{i2} & & \\ & & \ddots & \\ & & & \mathbf{A}_{iB} \end{pmatrix}, \quad (21)$$

where $\mathbf{A}_{i1}, \dots, \mathbf{A}_{iB}$ are $D \times D$ random matrices with $D \times B = p$. For a $p \times p$ matrix \mathbf{A} , denote by $\lambda^2(\mathbf{A})$ the eigenvalue of $\mathbf{A}'\mathbf{A}$. Let $x_i^* = \mathcal{G}(\mathcal{F}_{i,i-l})$ such that $x_i^* = \mathbf{A}_i x_{i-1}^* + B_i$. Suppose $x_i = (z'_{i1}, \dots, z'_{iB})'$ and $x_i^* = (z'^*_{i1}, \dots, z'^*_{iB})'$ according to the partition in (21), where $z_{ik}, z_{ik}^* \in \mathbb{R}^D$. Then we have

$$\begin{aligned} |z_{ik} - z_{ik}^*| &= |\mathbf{A}_{ik} \mathbf{A}_{i-1,k} \cdots \mathbf{A}_{i-l+1,k} (z_{i-l,k} - z_{i-l,k}^*)| \\ &\leq \prod_{j=0}^{l-1} \lambda(\mathbf{A}_{i-j}) |z_{i-l,k} - z_{i-l,k}^*|. \end{aligned}$$

²The block structure only needs to hold up to an unknown permutation of the components of x_i .

For any j belonging to the k th block,

$$\begin{aligned}\theta_{l,j,q} &= \|x_{ij} - x_{ij}^*\|_q \leq \|z_{ik} - z_{ik}^*\|_q \\ &\leq \prod_{j=0}^{l-1} \|\lambda(\mathbf{A}_{i-j})\|_q \|z_{i-l,k} - z_{i-l,k}^*\|_q \\ &\leq 2 \|\lambda(\mathbf{A}_0)\|_q^l \|z_{0k}\|_q.\end{aligned}\quad (22)$$

Suppose that

$$\|\lambda(\mathbf{A}_0)\|_q^l \max_{1 \leq k \leq B} \|z_{0k}\|_q \lesssim \varrho^l, \quad \varrho < 1. \quad (23)$$

Using the representation $x_i = \sum_{k=0}^{\infty} \mathbf{A}_i \mathbf{A}_{i-1} \cdots \mathbf{A}_{i-k+1} B_{i-k}$, it can be verified that (23) holds if $\|\lambda(\mathbf{A}_0)\|_q \leq \varrho$ and $\max_{1 \leq k \leq B} \|B_{0k}\|_q < c$ for some $c > 0$, where $B_0 = (B'_{01}, \dots, B'_{0B})'$ with $B_{0j} \in \mathbb{R}^D$. By (22) and (23), we have $\max_{1 \leq j \leq p} \Theta_{u,j,q} \lesssim \varrho^u$.

Remark 2.4. Motivated by Example 2.3, let $x_i = (z'_{i1}, \dots, z'_{iB})'$ with $z_{ij} \in \mathbb{R}^D$ for $1 \leq j \leq B$. Consider the blockwise model

$$z_{ij} = \tilde{\mathcal{G}}_{ij}(\dots, \epsilon_{i-1,j}, \epsilon_{ij}), \quad 1 \leq j \leq B, \quad (24)$$

where $\epsilon_i = (\epsilon'_{i1}, \dots, \epsilon'_{iB})'$ is a sequence of i.i.d. random vectors. In particular, when $D = 1$, we have the following model,

$$x_{ij} = \tilde{\mathcal{G}}_{ij}(\dots, \epsilon_{i-1,j}, \epsilon_{ij}), \quad 1 \leq j \leq p. \quad (25)$$

Because x_{ij} only depends on $\{\epsilon_{ij}\}$, we shall call (25) the componentwise model. Although the time series model is defined in a componentwise fashion, the components of x_i are dependent through the sequence $\{\epsilon_i\}$. For componentwise model, the analysis in the univariate case (see [28, 30]) can be applied separately for each component, and Conditions (8)–(10) can be translated into suitable restrictions on $\tilde{\mathcal{G}}_{ij}$ and the tail behavior of $\{\epsilon_{ij}\}$ uniformly across the index j .

Remark 2.5. The block assumption in Example 2.3 can be replaced by (26) below. For a matrix $\mathbf{A} = (a_{ij})_{i,j=1}^p$, define $\|\mathbf{A}\|_{\infty} = \max_{1 \leq i \leq p} \sum_{j=1}^p |a_{ij}|$. Using the fact that $|\mathbf{A}a|_{\infty} \leq \|\mathbf{A}\|_{\infty} |a|_{\infty}$, we have

$$\begin{aligned}|x_i - x_i^*|_{\infty} &\leq |\mathbf{A}_i \mathbf{A}_{i-1} \cdots \mathbf{A}_{i-l+1} (x_{i-l} - x_{i-l}^*)|_{\infty} \\ &\leq \prod_{k=i-l+1}^i \|\mathbf{A}_k\|_{\infty} |x_{i-l} - x_{i-l}^*|_{\infty}.\end{aligned}$$

Hence, we obtain for any $1 \leq j \leq p$

$$\begin{aligned}\theta_{l,j,q} &= \|x_{ij} - x_{ij}^*\|_q \leq \|x_i - x_i^*\|_\infty \|e_j\|_q \\ &\leq \prod_{k=i-l+1}^i \|\mathbf{A}_k\|_\infty \|e_j\|_q \|x_{i-l} - x_{i-l}^*\|_\infty \\ &\leq 2 \|\mathbf{A}_0\|_\infty^l \|x_{i-1} - x_{i-1}^*\|_\infty.\end{aligned}$$

An alternative assumption is given by

$$\|\mathbf{A}_0\|_\infty^l \|x_0 - x_0^*\|_\infty \lesssim \varrho^l, \quad \varrho < 1. \quad (26)$$

Here we impose constraint on the coefficient matrix \mathbf{A}_0 in terms of the $\|\cdot\|_\infty$ norm. This assumption is weaker than those in Example 2.3 as we drop the block assumption but it can be stronger when $\lambda(\mathbf{A}) \leq \|\mathbf{A}\|_\infty$.

Example 2.4 (Nonlinear Markov chain). Consider a nonlinear Markov chain defined by an iterated random function $H_i(\cdot, \epsilon_i)$,

$$x_i = H_i(x_{i-1}, \epsilon_i).$$

Here ϵ_i 's are i.i.d. innovations, and $H_i(\cdot, \cdot)$ is an \mathbb{R}^p -valued and jointly measurable function, which satisfies the following two conditions: (i) there exists some x_0 such that $\vartheta := \sup_i \|H_i(x_0, \epsilon_0)\|_q < \infty$ for $q > 2$, and (ii)

$$\vartheta \sup_i \|L_i\|_q^l \leq \varrho^l < 1, \quad L_i = \sup_{x \neq x'} \frac{|H_i(x, \epsilon_0) - H_i(x', \epsilon_0)|}{|x - x'|}. \quad (27)$$

Then it can be shown that $\max_{1 \leq j \leq p} \Theta_{l,j,q}(x) = O(\varrho^l)$ (see the derivations in [29]). In fact, (27) is a relatively strong assumption as ϑ generally grows with p , and the Lipschitz constant $\sup_i \|L_i\|_q$ can also be large when p is large. Assume a block structure on H_i (as in Remark 2.4): $H_i = (H'_{i1}, \dots, H'_{iB})'$ with $H_{ij} \in \mathbb{R}^D$. Then, we have

$$\tilde{x}_{ij} = H_{ij}(\tilde{x}_{i-1,j}, \tilde{\epsilon}_{ij}), \quad 1 \leq j \leq B,$$

where $x_i = (\tilde{x}'_{i1}, \dots, \tilde{x}'_{iB})'$ and $\epsilon_i = (\tilde{\epsilon}'_{i1}, \dots, \tilde{\epsilon}'_{iB})'$ with $\tilde{x}_{ij}, \tilde{\epsilon}_{ij} \in \mathbb{R}^D$. Under the above block structure, (27) can be weakened by replacing ϑ with $\max_{1 \leq j \leq B} \sup_i \|H_{ij}(x_0, \epsilon_0)\|_q$. Under tail conditions (12) or (13), and that $\min_{1 \leq j \leq p} \frac{1}{n} \sum_{i,k=1}^n \text{cov}(x_{ij}, x_{kj}) > c$ for $c > 0$, Corollary 2.2 can be applied, which suggests that the Kolmogorov distance decreases to zero at some polynomial rate.

In the high-dimensional setting, certain characteristics of time series models (such as the structures of the coefficient matrices) are allowed to vary with the dimension p . Regularity conditions

are thus required to account for such high dimensionality. These conditions are usually case-by-case and their suitability depends on the problem of interest. One set of assumptions may be replaced by others which concern a different aspect of the time series models. Here we focus on three concrete examples and discuss some sufficient conditions for our theory to hold. It is of interest to consider a broader class of time series models. Again we expect certain regularity conditions to hold besides those commonly assumed in the low dimensional setting.

3. Numerical studies

In this section, some numerical experiments are conducted to verify the Gaussian approximation phenomenon predicted by our general theory. We consider the following three linear models and one nonlinear model, where the designs are mainly motivated by the examples in [20].

1. VAR(2): $x_i = \mathbf{A}_1 x_{i-1} + \mathbf{A}_2 x_{i-2} + \epsilon_i$, where $\mathbf{A}_i = \mathbf{I}_{p/3} \otimes \tilde{\mathbf{A}}_i$ with $\mathbf{I}_{p/3}$ being the $p/3 \times p/3$ identity matrix and

$$\tilde{\mathbf{A}}_1 = \begin{pmatrix} 0.7 & 0.1 & 0.0 \\ 0.0 & 0.4 & 0.1 \\ 0.9 & 0.0 & 0.8 \end{pmatrix}, \quad \tilde{\mathbf{A}}_2 = \begin{pmatrix} -0.2 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.1 \\ 0.0 & 0.0 & 0.0 \end{pmatrix}.$$

2. VARMA(2, 1): $x_i = \mathbf{A}_1 x_{i-1} + \mathbf{A}_2 x_{i-2} + \epsilon_i + \mathbf{B}_1 \epsilon_{i-1}$, where $\mathbf{A}_i = \mathbf{I}_{p/2} \otimes \tilde{\mathbf{A}}_i$ and $\mathbf{B}_1 = \mathbf{I}_{p/2} \otimes \tilde{\mathbf{B}}_1$ with

$$\tilde{\mathbf{A}}_1 = \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix}, \quad \tilde{\mathbf{A}}_2 = \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix}, \quad \tilde{\mathbf{B}}_1 = \begin{pmatrix} 0.6 & 0.2 \\ 0.0 & 0.3 \end{pmatrix}.$$

3. Time-varying VAR(1): $x_i = \mathbf{A}_i x_{i-1} + \epsilon_i$, where $\mathbf{A}_i = \sin(2\pi i/n) \tilde{\mathbf{A}}$. Here $\tilde{\mathbf{A}}$ is symmetric and its entries are i.i.d. realizations from the Bernoulli distribution with success probability 0.25. We rescale $\tilde{\mathbf{A}}$ such that its largest eigenvalue is equal to 0.5.
4. BEKK-ARCH(1): $x_i = \Sigma_{i|i-1}^{1/2} \epsilon_i$, where $\epsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \mathbf{I}_p)$ or $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{ip})'$ with $\epsilon_{ij}/\sqrt{3} + 1$ being i.i.d. uniform random variables on $[0, 2]$, and $\Sigma_{i|i-1} = \mathbf{B} + \mathbf{A} x_i x_i' \mathbf{A}'$. Here $\mathbf{B} = \mathbf{I}_{p/2} \otimes \tilde{\mathbf{B}}$ and $\mathbf{A} = \mathbf{I}_{p/2} \otimes \tilde{\mathbf{A}}$ with

$$\tilde{\mathbf{A}} = \begin{pmatrix} 0.4 & 0 \\ 0 & 0.3 \end{pmatrix}, \quad \tilde{\mathbf{B}} = \begin{pmatrix} 0.8 & 0.5 \\ 0.5 & 0.7 \end{pmatrix}.$$

For models (1)–(3), we consider the following data generating processes for the errors. In cases (a)–(d) below, $\epsilon_i = \tilde{\mathbf{\Gamma}}^{1/2} \varepsilon_i$ where $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{ip})'$ with $\varepsilon_{ij}/\sqrt{3} + 1$ being i.i.d. uniform random variables on $[0, 2]$. We consider four covariance structures (a) AR(1): $\tilde{\mathbf{\Gamma}} = (\tilde{\gamma}_{ij})$ for $\tilde{\gamma}_{ij} = 0.25^{|i-j|}$; (b) Block diagonal: $\tilde{\mathbf{\Gamma}} = \mathbf{I}_{p/2} \otimes \mathbf{C}$ for $\mathbf{C} = (c_{ij})_{i,j=1}^2$, where $c_{11} = c_{22} = 1$ and $c_{12} = c_{21} = 0.8$; (c) Banded: $\tilde{\mathbf{\Gamma}} = (\tilde{\gamma}_{ij})$, where $\tilde{\gamma}_{ij} = 1$ for $i = j$, $\tilde{\gamma}_{ij} = 0.4$ for $|i - j| = 1$,

$\tilde{\gamma}_{ij} = 0.2$ for $|i - j| = 2, 3$, $\tilde{\gamma}_{ij} = 0.1$ for $|i - j| = 4$, and $\tilde{\gamma}_{ij} = 0$ otherwise; (d) Exchangeable: $\tilde{\Gamma} = (\tilde{\gamma}_{ij})$, where $\tilde{\gamma}_{ij} = 1$ for $i = j$ and $\tilde{\gamma}_{ij} = 0.25$ for $i \neq j$. In cases (e) and (f) below, $\epsilon_i = r_{ij}e_i$, where r_{ij} 's are fixed i.i.d. realizations generated from the uniform distribution on $[0, 1]$ and $\{e_i\}$ is a sequence of i.i.d. univariate random variables. For the distribution of e_i , we consider (e) $e_i = (v_i - 5)/\sqrt{5}$ with v_i being a Gamma distribution with shape parameter 5 and scale parameter 1; (f) $e_i = v'_i/\sqrt{2}$ with v'_i being a t distribution with degrees of freedom 4.

In all cases, we generate a Gaussian sequence $\{y_i\}$ which preserves the autocovariance structure of the non-Gaussian sequence $\{x_i\}$. We consider $n = 100$ and $p = 120, 240, 480, 960$. The results are obtained based on 10 000 Monte Carlo replications. Figures 1–3 show the P-P plots comparing the distributions of T_X and T_Y in linear models (1)–(3). Moreover, we present in Table 1 the probability $P(T_X \leq Q_{T_Y}(\alpha))$ with $\alpha = 90\%, 95\%, 97.5\%$ and 99% , where $Q_{T_Y}(\alpha)$ denotes the α th quantile of T_Y . The results suggest that the Gaussian approximation is quite accurate in all the linear cases considered here. Figure 4 and Table 2 present the results for BEKK-ARCH(1) model. The approximation is again accurate in the nonlinear case. It is also worth pointing out that the Gaussian approximation is in general very precise for the tail of T_X , which is most relevant in statistical inference. Overall, the numerical results clearly demonstrate the practical relevance of the Gaussian approximation theory.

Appendix

Define the generic constants C and C' that are independent of n and p . For a set \mathcal{A} , denote by $|\mathcal{A}|$ its cardinality.

A.1. Proofs of the main results in Section 2.1

Proof of Proposition 2.1. We first prove (4). Define $Z(t) = \sum_{i=1}^n Z_i(t)$ with the Slepian interpolation $Z_i(t) = (\sqrt{t}\tilde{x}_i + \sqrt{1-t}\tilde{y}_i)/\sqrt{n}$ and $0 \leq t \leq 1$. Let $\Psi(t) = \mathbb{E}m(Z(t))$. Define $V^{(i)}(t) = \sum_{j \in \tilde{N}_i} Z_j(t)$ and $Z^{(i)}(t) = Z(t) - V^{(i)}(t)$. Write $\partial_j m(x) = \partial m(x)/\partial x_j$, $\partial_{jk} m(x) = \partial^2 m(x)/\partial x_j \partial x_k$ and $\partial_{jkl} m(x) = \partial^3 m(x)/\partial x_j \partial x_k \partial x_l$ for $j, k, l = 1, 2, \dots, p$, where $x = (x_1, x_2, \dots, x_p)'$. Note that

$$\begin{aligned} \mathbb{E}m(\tilde{X}) - \mathbb{E}m(\tilde{Y}) &= \Psi(1) - \Psi(0) = \int_0^1 \Psi'(t) dt \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^p \int_0^1 \mathbb{E}[\partial_j m(Z(t)) \dot{Z}_{ij}(t)] dt \\ &= \frac{1}{2} (I_1 + I_2 + I_3), \end{aligned} \quad (28)$$

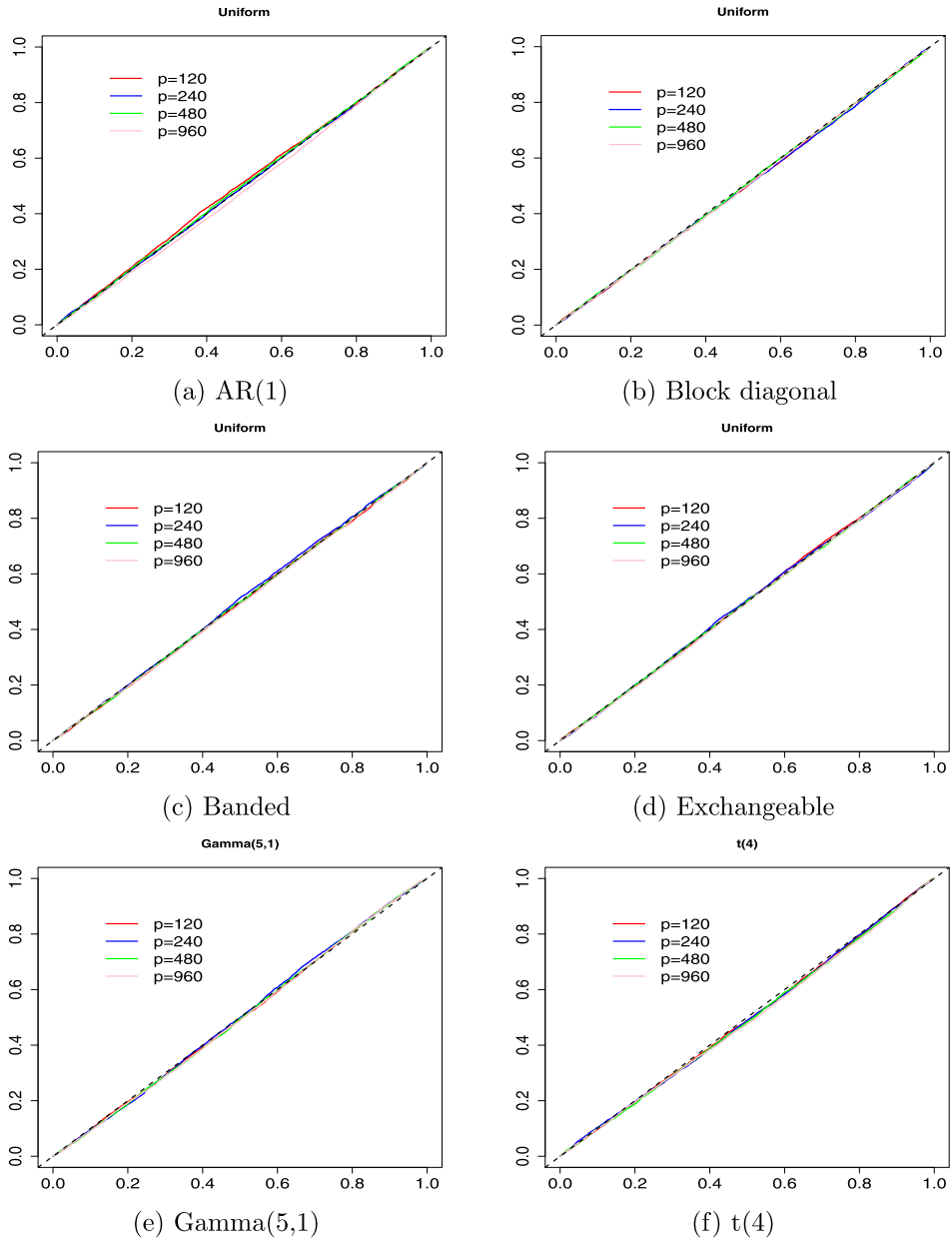


Figure 1. P-P plots comparing the distributions of T_X and T_Y , where the data are generated from the VAR(2) model.

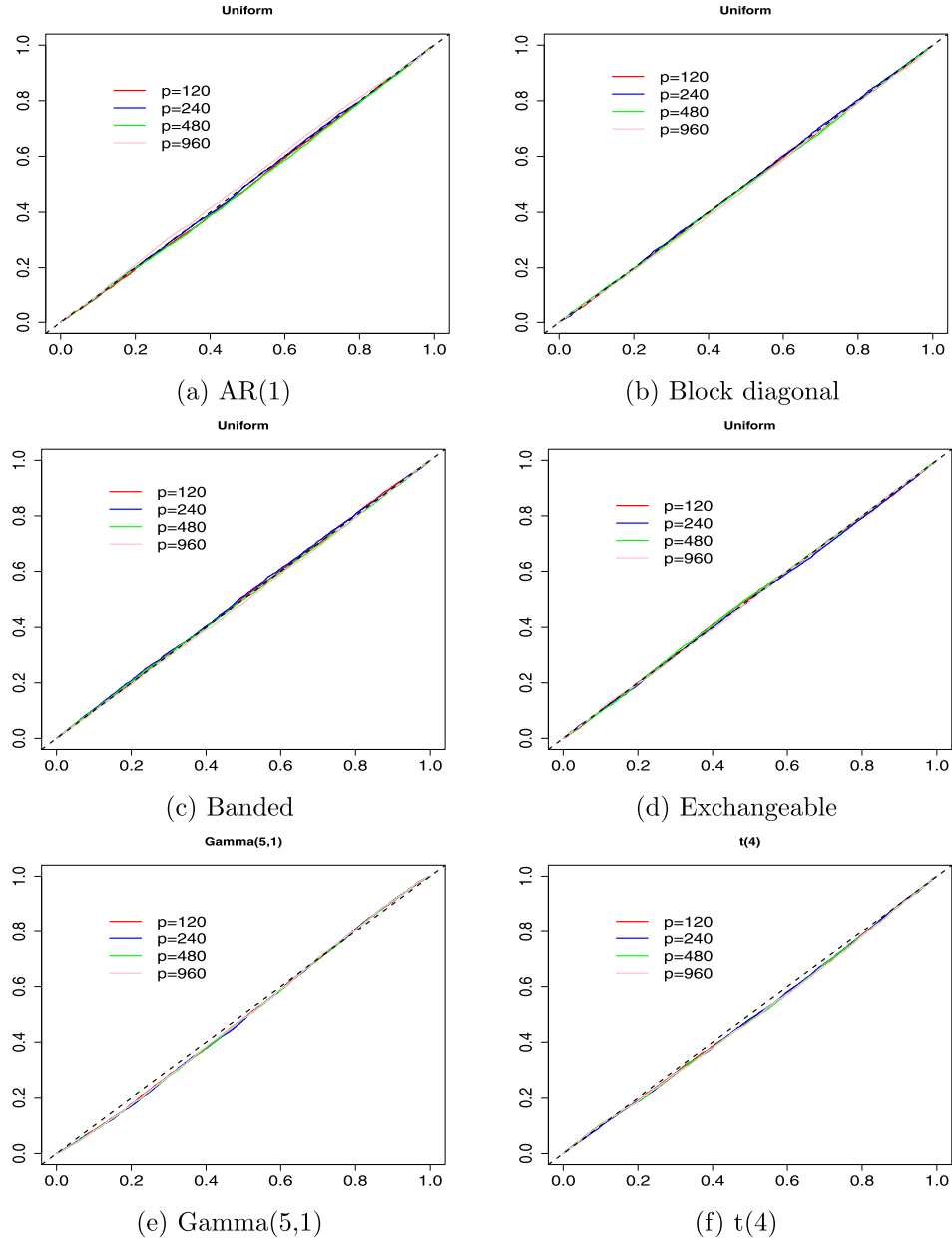


Figure 2. P-P plots comparing the distributions of T_X and T_Y , where the data are generated from the VARMA(2, 1) model.

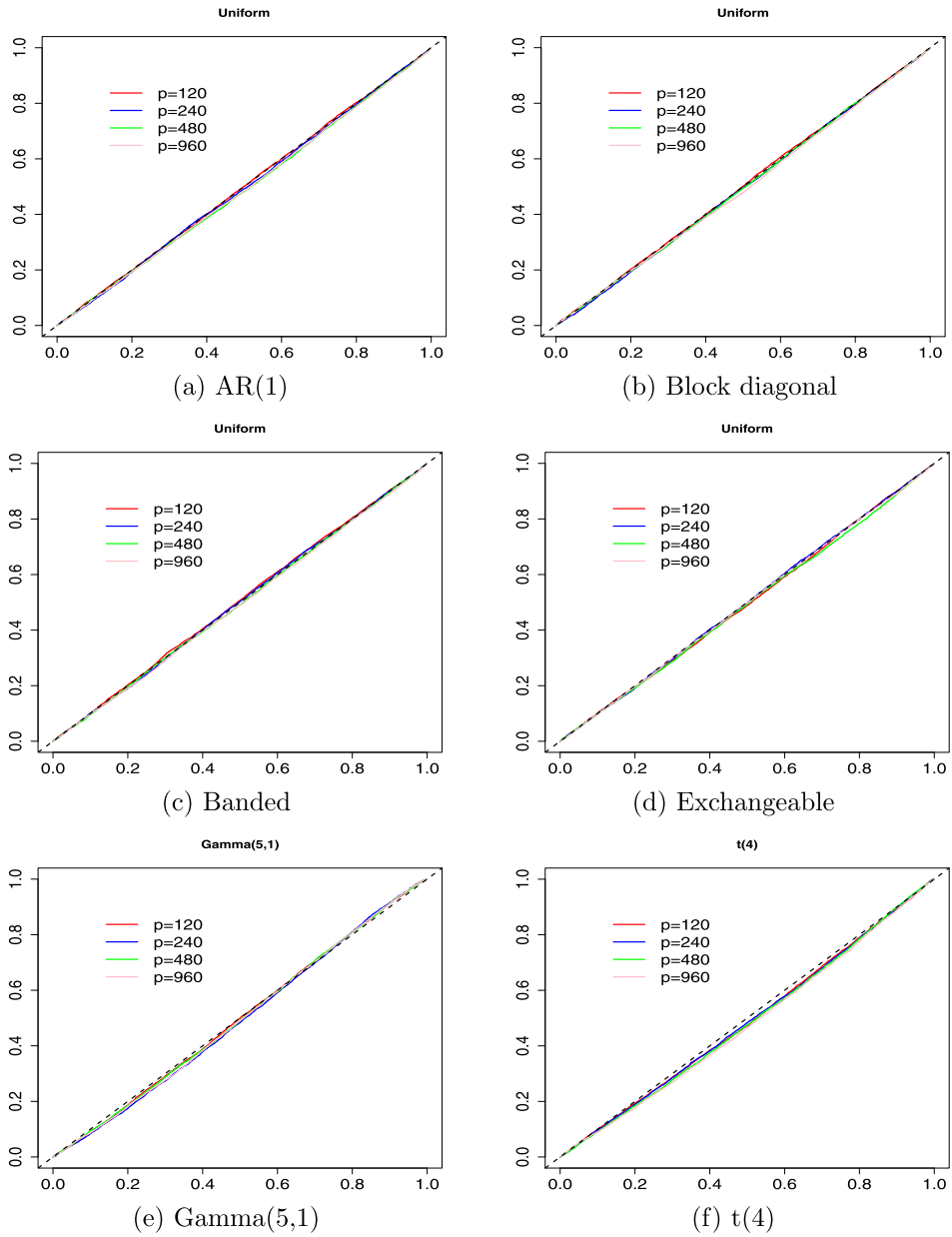


Figure 3. P-P plots comparing the distributions of T_X and T_Y , where the data are generated from the time-varying VAR(1) model.

Table 1. The simulated probability $P(T_X \leq Q_{T_Y}(\alpha))$, where $\alpha = 90\%, 95\%, 97.5\%, 99\%$, and $n = 100$. The results are obtained based on 10 000 Monte Carlo replications

	<i>p</i>	VAR(2)				VARMA(2, 1)				Time-varying VAR(1)			
		90%	95%	97.5%	99%	90%	95%	97.5%	99%	90%	95%	97.5%	99%
AR(1)	120	89.6	95.1	97.3	98.8	90.2	95.2	97.5	98.8	90.4	95.4	97.6	99.1
	240	90.2	95.1	97.7	99.1	90.1	95.0	97.5	98.9	89.9	95.3	97.8	99.1
	480	89.9	94.6	97.3	98.8	90.6	95.4	97.5	99.1	90.7	95.5	98.0	99.1
	960	90.4	95.2	97.7	99.1	89.9	95.3	97.9	99.2	90.9	95.8	98.1	99.5
Block diagonal	120	89.9	95.3	97.6	99.1	90.5	95.2	97.5	98.9	90.1	95.5	97.9	99.1
	240	90.4	94.8	97.3	99.0	89.8	94.8	97.3	99.0	90.5	95.4	97.9	99.2
	480	90.5	95.5	97.9	99.2	90.2	95.0	97.5	99.1	90.4	95.3	97.8	99.2
	960	90.0	94.9	97.6	99.3	90.7	96.0	98.2	99.3	90.6	95.3	97.9	99.2
Banded	120	90.6	95.5	97.6	99.1	89.3	94.7	97.4	98.9	89.7	95.3	97.9	99.3
	240	89.7	95.0	97.6	99.2	89.9	94.8	97.8	99.0	89.8	95.0	97.8	99.0
	480	90.1	95.3	97.6	99.2	90.3	95.1	97.5	99.1	90.2	95.3	97.9	99.2
	960	90.7	95.3	97.5	99.1	90.2	95.1	97.5	99.1	90.8	95.6	97.8	99.3
Exchangeable	120	90.2	95.3	97.6	99.0	90.5	95.3	97.7	99.2	90.0	94.9	97.5	99.0
	240	90.7	95.7	98.0	99.2	90.4	95.3	97.7	99.0	90.1	94.9	97.8	99.0
	480	90.1	95.0	97.6	99.0	89.9	95.0	97.3	99.0	90.9	95.3	97.8	99.2
	960	90.2	95.4	97.4	99.0	90.1	95.0	97.9	99.2	90.8	95.5	98.0	99.3
Gamma(5, 1)	120	89.0	94.1	96.8	98.7	89.0	93.9	96.7	98.4	88.9	94.0	96.4	98.2
	240	88.6	94.2	97.0	98.7	88.9	94.1	96.4	98.3	88.5	93.7	96.5	98.4
	480	88.9	93.8	96.7	98.6	88.9	94.0	96.5	98.5	88.8	93.9	96.7	98.3
	960	88.7	93.9	96.7	98.4	88.9	94.0	96.4	98.3	88.9	93.8	96.5	98.2
t(4)	120	90.0	94.7	97.1	98.6	90.2	95.1	97.5	98.8	90.7	95.3	97.3	98.6
	240	90.2	94.8	97.3	98.9	90.1	95.5	97.7	98.7	90.8	95.0	97.4	98.6
	480	90.9	95.1	97.3	98.8	90.3	95.0	97.5	98.9	90.0	94.7	97.3	98.8
	960	90.9	95.2	97.4	98.8	90.3	95.3	97.4	98.8	90.8	95.4	97.6	98.8

where $\dot{Z}_{ij}(t) = \{\tilde{x}_{ij}/\sqrt{t} - \tilde{y}_{ij}/\sqrt{1-t}\}/\sqrt{n}$, and

$$I_1 = \sum_{i=1}^n \sum_{j=1}^p \int_0^1 \mathbb{E}[\partial_j m(Z^{(i)}(t)) \dot{Z}_{ij}(t)] dt,$$
$$I_2 = \sum_{i=1}^n \sum_{k,j=1}^p \int_0^1 \mathbb{E}[\partial_k \partial_j m(Z^{(i)}(t)) \dot{Z}_{ij}(t) V_k^{(i)}(t)] dt, \tag{29}$$
$$I_3 = \sum_{i=1}^n \sum_{k,l,j=1}^p \int_0^1 \int_0^1 (1-\tau) \mathbb{E}[\partial_l \partial_k \partial_j m(Z^{(i)}(t) + \tau V^{(i)}(t)) \dot{Z}_{ij}(t) V_k^{(i)}(t) V_l^{(i)}(t)] d\tau dt.$$

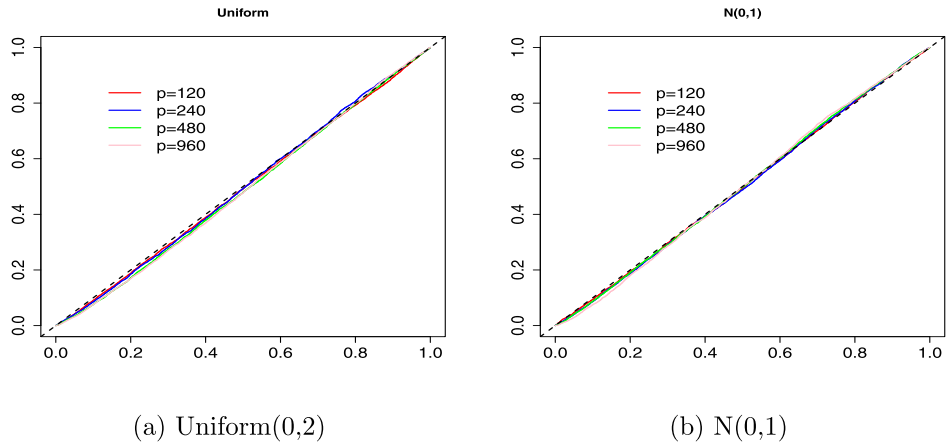


Figure 4. P-P plots comparing the distributions of T_X and T_Y , where the data are generated from the BEKK-ARCH(1) model.

Using the fact that $Z^{(i)}(t)$ and $\dot{Z}_{ij}(t)$ are independent, and $\mathbb{E}\dot{Z}_{ij}(t) = 0$, we have $I_1 = 0$. To bound the second term, define the expanded neighborhood around N_i ,

$$\mathcal{N}_i = \{j : \{j, k\} \in E_n \text{ for some } k \in N_i\},$$

and $\mathcal{Z}^{(i)}(t) = Z(t) - \sum_{l \in \mathcal{N}_i \cup \tilde{N}_i} Z_l(t) = Z^{(i)}(t) - \mathcal{V}^{(i)}(t)$, where $\mathcal{V}^{(i)}(t) = \sum_{l \in \mathcal{N}_i \setminus \tilde{N}_i} Z_l(t)$ with $\mathcal{N}_i \setminus \tilde{N}_i = \{k \in \mathcal{N}_i : k \notin \tilde{N}_i\}$. By Taylor expansion, we have

$$\begin{aligned} I_2 &= \sum_{i=1}^n \sum_{k,j=1}^p \int_0^1 \mathbb{E}[\partial_k \partial_j m(\mathcal{Z}^{(i)}(t)) \dot{Z}_{ij}(t) V_k^{(i)}(t)] dt \\ &\quad + \sum_{i=1}^n \sum_{k,j,l=1}^p \int_0^1 \int_0^1 \mathbb{E}[\partial_k \partial_j \partial_l m(\mathcal{Z}^{(i)}(t) + \tau \mathcal{V}^{(i)}(t)) \dot{Z}_{ij}(t) V_k^{(i)}(t) \mathcal{V}_l^{(i)}(t)] dt d\tau \end{aligned}$$

Table 2. The simulated probability $P(T_X \leq Q_{T_Y}(\alpha))$, where $\alpha = 90\%, 95\%, 97.5\%, 99\%$, and $n = 100$. The results are obtained based on 10 000 Monte Carlo replications

p	BEKK-ARCH(1), Uniform(0, 2)				BEKK-ARCH(1), $N(0, 1)$			
	90%	95%	97.5%	99%	90%	95%	97.5%	99%
120	90.7	95.2	97.5	99.0	89.6	94.6	97.3	98.9
240	89.5	94.8	97.3	99.0	89.2	94.2	97.0	98.7
480	90.0	94.9	97.5	99.1	89.1	94.4	96.8	98.8
960	89.4	94.6	97.2	99.0	89.3	94.4	97.2	98.6

$$\begin{aligned}
&= \sum_{i=1}^n \sum_{k,j=1}^p \int_0^1 \mathbb{E}[\partial_k \partial_j m(\mathcal{Z}^{(i)}(t))] \mathbb{E}[\dot{Z}_{ij}(t) V_k^{(i)}(t)] dt \\
&\quad + \sum_{i=1}^n \sum_{k,j,l=1}^p \int_0^1 \int_0^1 \mathbb{E}[\partial_k \partial_j \partial_l m(\mathcal{Z}^{(i)}(t) + \tau \mathcal{V}^{(i)}(t)) \dot{Z}_{ij}(t) V_k^{(i)}(t) \mathcal{V}_l^{(i)}(t)] dt d\tau \\
&= I_{21} + I_{22},
\end{aligned}$$

where we have used the fact that $\dot{Z}_{ij}(t) V_k^{(i)}(t)$ and $\mathcal{Z}^{(i)}(t)$ are independent.

Let $M_{xy} = \max\{M_x, M_y\}$. By the assumption that $2\sqrt{5}\beta D_n^2 M_{xy}/\sqrt{n} \leq 1$,

$$\begin{aligned}
\max_{1 \leq j \leq p} \left| \sum_{l \in \mathcal{N}_i \cup \tilde{\mathcal{N}}_i} Z_{lj}(t) \right| &\leq \max_{1 \leq j \leq p} \sum_{l \in \mathcal{N}_i \cup \tilde{\mathcal{N}}_i} |Z_{lj}(t)| \leq D_n^2 \sup_{t \in [0,1]} (2\sqrt{t} + \sqrt{1-t}) M_{xy}/\sqrt{n} \\
&\leq \sqrt{5} D_n^2 M_{xy}/\sqrt{n} \leq \beta^{-1}/2 \leq \beta^{-1},
\end{aligned}$$

where the second inequality comes from the facts that $|\tilde{x}_{ij}| \leq 2M_{xy}$, $|\tilde{y}_{ij}| \leq M_{xy}$ and $|\mathcal{N}_i \cup \tilde{\mathcal{N}}_i| \leq D_n^2$. By Lemma A.5 in [12], we have for every $1 \leq j, k, l \leq p$,

$$|\partial_j \partial_k m(z)| \leq U_{jk}(z), \quad |\partial_j \partial_k \partial_l m(z)| \leq U_{jkl}(z),$$

where $U_{jk}(z)$ and $U_{jkl}(z)$ satisfy that

$$\sum_{j,k=1}^p U_{jk}(z) \leq (G_2 + 2G_1\beta), \quad \sum_{j,k,l=1}^p U_{jkl}(z) \leq (G_3 + 6G_2\beta + 6G_1\beta^2),$$

with $G_k = \sup_{z \in \mathbb{R}} |\partial^k g(z)/\partial z^k|$ for $k \geq 0$. Along with Lemma A.6 in [12], we obtain

$$\begin{aligned}
|I_{21}| &\leq \sum_{i=1}^n \sum_{k,j=1}^p \int_0^1 \mathbb{E}[U_{jk}(\mathcal{Z}^{(i)}(t))] |\mathbb{E}[\dot{Z}_{ij}(t) V_k^{(i)}(t)]| dt \\
&\lesssim \sum_{i=1}^n \sum_{k,j=1}^p \int_0^1 \mathbb{E}[U_{jk}(Z(t))] |\mathbb{E}[\dot{Z}_{ij}(t) V_k^{(i)}(t)]| dt \\
&\lesssim (G_2 + G_1\beta) \int_0^1 \max_{1 \leq j,k \leq p} \sum_{i=1}^n |\mathbb{E}[\dot{Z}_{ij}(t) V_k^{(i)}(t)]| dt.
\end{aligned}$$

Since $2\sqrt{5}\beta D_n^2 M_{xy}/\sqrt{n} \leq 1$, we have

$$\begin{aligned}
|I_{22}| &\leq \sum_{i=1}^n \sum_{k,j,l=1}^p \int_0^1 \int_0^1 \mathbb{E}[|\partial_k \partial_j \partial_l m(\mathcal{Z}^{(i)}(t) + \tau \mathcal{V}^{(i)}(t))| \cdot |\dot{Z}_{ij}(t) V_k^{(i)}(t) \mathcal{V}_l^{(i)}(t)|] dt d\tau \\
&\leq \sum_{i=1}^n \sum_{k,j,l=1}^p \int_0^1 \int_0^1 \mathbb{E}[U_{kjl}(\mathcal{Z}^{(i)}(t) + \tau \mathcal{V}^{(i)}(t)) |\dot{Z}_{ij}(t) V_k^{(i)}(t) \mathcal{V}_l^{(i)}(t)|] dt d\tau
\end{aligned}$$

$$\begin{aligned}
&\lesssim \sum_{i=1}^n \sum_{k,j,l=1}^p \int_0^1 \mathbb{E}[U_{kjl}(Z(t)) |\dot{Z}_{ij}(t) V_k^{(i)}(t) \mathcal{V}_l^{(i)}(t)|] dt d\tau \\
&\leq \int_0^1 \mathbb{E} \left[\sum_{k,j,l=1}^p U_{kjl}(Z(t)) \max_{1 \leq k,j,l \leq p} \sum_{i=1}^n |\dot{Z}_{ij}(t) V_k^{(i)}(t) \mathcal{V}_l^{(i)}(t)| \right] dt d\tau \\
&\lesssim (G_3 + G_2\beta + G_1\beta^2) \int_0^1 \mathbb{E} \max_{1 \leq k,j,l \leq p} \sum_{i=1}^n |\dot{Z}_{ij}(t) V_k^{(i)}(t) \mathcal{V}_l^{(i)}(t)| dt d\tau.
\end{aligned} \tag{30}$$

To bound the integration on (30), we let $w(t) = 1/(\sqrt{t} \wedge \sqrt{1-t})$ and note that

$$\begin{aligned}
&\int_0^1 \mathbb{E} \max_{1 \leq k,j,l \leq p} \sum_{i=1}^n |\dot{Z}_{ij}(t) V_k^{(i)}(t) \mathcal{V}_l^{(i)}(t)| dt \\
&\leq \int_0^1 \mathbb{E} \max_{1 \leq k,j,l \leq p} \left(\sum_{i=1}^n |\dot{Z}_{ij}(t)|^3 \right)^{1/3} \left(\sum_{i=1}^n |V_k^{(i)}(t)|^3 \right)^{1/3} \left(\sum_{i=1}^n |\mathcal{V}_l^{(i)}(t)|^3 \right)^{1/3} dt \\
&\leq \int_0^1 w(t) \left(\mathbb{E} \max_{1 \leq j \leq p} \sum_{i=1}^n |\dot{Z}_{ij}(t)/w(t)|^3 \mathbb{E} \max_{1 \leq k \leq p} \sum_{i=1}^n |V_k^{(i)}(t)|^3 \mathbb{E} \max_{1 \leq l \leq p} \sum_{i=1}^n |\mathcal{V}_l^{(i)}(t)|^3 \right)^{1/3} dt.
\end{aligned}$$

As for I_{21} , by the assumption that $\mathbb{E}y_{ij}y_{lk} = \mathbb{E}x_{ij}x_{lk}$ (in fact, we only need to require that $\sum_{k \in \tilde{N}_i} \mathbb{E}x_i x'_k = \sum_{k \in \tilde{N}_i} \mathbb{E}y_i y'_k$ for all i), we have

$$\begin{aligned}
&\max_{1 \leq j,k \leq p} \sum_{i=1}^n |\mathbb{E}[\dot{Z}_{ij}(t) V_k^{(i)}(t)]| \\
&= \max_{1 \leq j,k \leq p} \frac{1}{n} \sum_{i=1}^n \left| \sum_{l \in \tilde{N}_i} (\mathbb{E}\tilde{x}_{ij}\tilde{x}_{lk} - \mathbb{E}\tilde{y}_{ij}\tilde{y}_{lk}) \right| \\
&= \max_{1 \leq j,k \leq p} \frac{1}{n} \sum_{i=1}^n \left| \sum_{l \in \tilde{N}_i} (\mathbb{E}\tilde{x}_{ij}\tilde{x}_{lk} - \mathbb{E}x_{ij}x_{lk}) + \sum_{l \in \tilde{N}_i} (\mathbb{E}y_{ij}y_{lk} - \mathbb{E}\tilde{y}_{ij}\tilde{y}_{lk}) \right| \\
&\leq \max_{1 \leq j,k \leq p} \frac{1}{n} \sum_{i=1}^n \left| \sum_{l \in \tilde{N}_i} \{ \mathbb{E}y_{lk}(y_{ij} - \tilde{y}_{ij}) + \mathbb{E}\tilde{y}_{ij}(y_{lk} - \tilde{y}_{lk}) \} \right| \\
&\quad + \max_{1 \leq j,k \leq p} \frac{1}{n} \sum_{i=1}^n \left| \sum_{l \in \tilde{N}_i} \{ \mathbb{E}x_{lk}(x_{ij} - \tilde{x}_{ij}) + \mathbb{E}\tilde{x}_{ij}(x_{lk} - \tilde{x}_{lk}) \} \right| \\
&\leq \phi(M_x, M_y).
\end{aligned} \tag{31}$$

Using similar arguments as above, we have $|I_3| \lesssim (G_3 + G_2\beta + G_1\beta^2)I_{31}$ with

$$I_{31} \leq \int_0^1 w(t) \left(\mathbb{E} \max_{1 \leq j \leq p} \sum_{i=1}^n |\dot{Z}_{ij}(t)/w(t)|^3 \mathbb{E} \max_{1 \leq k \leq p} \sum_{i=1}^n |V_k^{(i)}(t)|^3 \mathbb{E} \max_{1 \leq l \leq p} \sum_{i=1}^n |V_l^{(i)}(t)|^3 \right)^{1/3} dt.$$

We first consider the term $\mathbb{E} \max_{1 \leq j \leq p} \sum_{i=1}^n |\dot{Z}_{ij}(t)/w(t)|^3$. Using the fact that $|\dot{Z}_{ij}(t)/w(t)| \leq (|\tilde{x}_{ij}| + |\tilde{y}_{ij}|)/\sqrt{n}$, we get

$$\begin{aligned} \mathbb{E} \max_{1 \leq j \leq p} \sum_{i=1}^n |\dot{Z}_{ij}(t)/w(t)|^3 &\lesssim \frac{1}{n^{3/2}} \mathbb{E} \max_{1 \leq j \leq p} \sum_{i=1}^n (|\tilde{x}_{ij}|^3 + |\tilde{y}_{ij}|^3) \\ &\lesssim \frac{1}{\sqrt{n}} (m_{x,3}^3 + m_{y,3}^3). \end{aligned}$$

On the other hand, notice that

$$\begin{aligned} \mathbb{E} \max_{1 \leq k \leq p} \sum_{i=1}^n |V_k^{(i)}(t)|^3 &\leq D_n^2 \mathbb{E} \max_{1 \leq k \leq p} \sum_{i=1}^n \sum_{j \in \tilde{N}_i} |Z_{jk}(t)|^3 \\ &\lesssim \frac{D_n^2}{n^{3/2}} \mathbb{E} \max_{1 \leq k \leq p} \sum_{i=1}^n \sum_{j \in \tilde{N}_i} (|\tilde{x}_{jk}|^3 + |\tilde{y}_{jk}|^3) \\ &\lesssim \frac{D_n^3}{\sqrt{n}} (m_{x,3}^3 + m_{y,3}^3). \end{aligned}$$

Similarly, we have

$$\begin{aligned} \mathbb{E} \max_{1 \leq l \leq p} \sum_{i=1}^n |V_l^{(i)}(t)|^3 &\leq D_n^4 \mathbb{E} \max_{1 \leq l \leq p} \sum_{i=1}^n \sum_{j \in \mathcal{N}_i} |Z_{jl}(t)|^3 \\ &\leq \frac{D_n^4}{n^{3/2}} \mathbb{E} \max_{1 \leq l \leq p} \sum_{i=1}^n \sum_{j \in \mathcal{N}_i} (|\tilde{x}_{jl}|^3 + |\tilde{y}_{jl}|^3) \\ &\lesssim \frac{D_n^6}{\sqrt{n}} (m_{x,3}^3 + m_{y,3}^3). \end{aligned}$$

Note that $\int_0^1 w(t) dt \lesssim 1$. Summarizing the above results, we have

$$\begin{aligned} I_2 &\lesssim (G_2 + G_1\beta)\phi(M_x, M_y) + (G_3 + G_2\beta + G_1\beta^2) \frac{D_n^3}{\sqrt{n}} (m_{x,3}^3 + m_{y,3}^3), \\ I_3 &\lesssim (G_3 + G_2\beta + G_1\beta^2) \frac{D_n^2}{\sqrt{n}} (m_{x,3}^3 + m_{y,3}^3). \end{aligned}$$

Alternatively, we can bound I_3 in the following way. By Lemmas A.5 and A.6 in [12], we have

$$\begin{aligned}
 |I_3| &= \sum_{i=1}^n \sum_{k,l,j=1}^p \int_0^1 \int_0^1 (1-\tau) \mathbb{E}[\partial_l \partial_k \partial_j m(Z^{(i)}(t) + \tau V^{(i)}(t)) \dot{Z}_{ij}(t) V_k^{(i)}(t) V_l^{(i)}(t)] dt d\tau \\
 &\lesssim \sum_{i=1}^n \sum_{k,l,j=1}^p \int_0^1 \mathbb{E}[U_{kjl}(Z^{(i)}(t))] \mathbb{E}|\dot{Z}_{ij}(t) V_k^{(i)}(t) V_l^{(i)}(t)| dt \\
 &\lesssim \sum_{i=1}^n \sum_{k,l,j=1}^p \int_0^1 \mathbb{E}[U_{kjl}(Z(t))] \mathbb{E}|\dot{Z}_{ij}(t) V_k^{(i)}(t) V_l^{(i)}(t)| dt \\
 &\leq n(G_3 + G_2\beta + G_1\beta^2) \\
 &\quad \times \int_0^1 w(t) \max_{1 \leq j,k,l \leq p} (\bar{\mathbb{E}}|\dot{Z}_{ij}(t)/w(t)|^3)^{1/3} (\bar{\mathbb{E}}|V_k^{(i)}(t)|^3)^{1/3} (\bar{\mathbb{E}}|V_l^{(i)}(t)|^3)^{1/3} dt.
 \end{aligned}$$

Notice that

$$\max_{1 \leq j \leq p} \bar{\mathbb{E}}|\dot{Z}_{ij}(t)/w(t)|^3 \leq \frac{1}{n^{3/2}} \max_{1 \leq j \leq p} \bar{\mathbb{E}}(|\tilde{x}_{ij}| + |\tilde{y}_{ij}|)^3 \lesssim \frac{1}{n^{3/2}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3).$$

It is not hard to see that

$$\max_{1 \leq k \leq p} \bar{\mathbb{E}}|V_k^{(i)}(t)|^3 \leq D_n^2 \max_{1 \leq k \leq p} \bar{\mathbb{E}} \sum_{j \in \tilde{N}_i} |Z_{jk}(t)|^3 \lesssim \frac{D_n^3}{n^{3/2}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3).$$

Thus, we derive that

$$I_3 \lesssim (G_3 + G_2\beta + G_1\beta^2) \frac{D_n^2}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3).$$

Therefore, we obtain

$$\begin{aligned}
 &|\mathbb{E}[m(\tilde{X}) - m(\tilde{Y})]| \\
 &\lesssim (G_2 + G_1\beta) \phi(M_x, M_y) + (G_3 + G_2\beta + G_1\beta^2) \frac{D_n^3}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3) \\
 &\quad + (G_3 + G_2\beta + G_1\beta^2) \frac{D_n^2}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3).
 \end{aligned} \tag{32}$$

Using the above arguments, we can show that

$$I_{22} \lesssim (G_3 + G_2\beta + G_1\beta^2) \frac{D_n^3}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3), \tag{33}$$

provided that $2\sqrt{5}\beta D_n^3 M_{xy}/\sqrt{n} \leq 1$. This proves the last statement of Proposition 2.1.

Note that $|m(x) - m(y)| \leq 2G_0$ and $|m(x) - m(y)| \leq G_1 \max_{1 \leq j \leq p} |x_j - y_j|$ with $x = (x_1, \dots, x_p)'$ and $y = (y_1, \dots, y_p)'$. So

$$\begin{aligned} |\mathbb{E}[m(X) - m(\tilde{X})]| &\leq |\mathbb{E}[(m(X) - m(\tilde{X}))\mathcal{I}]| + |\mathbb{E}[(m(X) - m(\tilde{X}))(1 - \mathcal{I})]| \\ &\lesssim G_1 \Delta + G_0 \mathbb{E}[1 - \mathcal{I}], \\ |\mathbb{E}[m(Y) - m(\tilde{Y})]| &\lesssim G_1 \Delta + G_0 \mathbb{E}[1 - \mathcal{I}]. \end{aligned} \quad (34)$$

Therefore, (4) follows by combining (32), (33) and (34). \square

Proof of Corollary 2.1. Notice that $D_n = 2M + 1$, $|\tilde{N}_i| \leq 2M + 1$ and $|\mathcal{N}_i \cup \tilde{N}_i| \leq 4M + 1$. Define the $\mathfrak{N}_i = \{j : \{j, k\} \in E_n \text{ for some } k \in \mathcal{N}_i\}$. Then $|\mathfrak{N}_i \cup \mathcal{N}_i \cup \tilde{N}_i| \leq 6M + 1$. Following the arguments in the proof of Proposition 2.1, we can show that

$$\max_{1 \leq l \leq p} \mathbb{E} |\mathcal{V}_l^{(i)}(t)|^3 \lesssim \frac{D_n^3}{n^{3/2}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3),$$

which implies that

$$I_{22} \lesssim (G_3 + G_2\beta + G_1\beta^2) \frac{D_n^2}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3).$$

The conclusion follows from the proof of Proposition 2.1. \square

A.2. Some results for M -dependent time series

This subsection is devoted to the analysis of M -dependent time series, which fits in the framework of dependency graph. Here, we allow M to grow slowly with the sample size n . Let $n = (N + M)r$, where $N \geq M$ and $N, M, r \rightarrow +\infty$ as $n \rightarrow +\infty$. Define the block sums

$$A_{ij} = \sum_{l=iN+(i-1)M-N+1}^{iN+(i-1)M} x_{lj}, \quad B_{ij} = \sum_{l=i(N+M)-M+1}^{i(N+M)} x_{lj}. \quad (35)$$

It is not hard to see that $\{A_{ij}\}_{i=1}^r$ and $\{B_{ij}\}_{i=1}^r$ with $1 \leq j \leq p$ are two sequences of independent random variables. Let $V_{nj} = \sqrt{V_{1nj}^2 + V_{2nj}^2}$ with $V_{1nj}^2 = \sum_{i=1}^r A_{ij}^2$ and $V_{2nj}^2 = \sum_{i=1}^r B_{ij}^2$. By generalizing Theorem 2.16 of de la Peña et al. [15], we obtain the following lemma, which is particularly useful in controlling the last two terms in (5).

Lemma A.1. Suppose $\{x_i\}$ is a p -dimensional M -dependent sequence. Assume that there exist $a_j, b_j > 0$ such that

$$P\left(\sum_{i=1}^n x_{ij} > a_j\right) \leq 1/4, \quad P(V_{nj}^2 > b_j^2) \leq 1/4.$$

Then we have

$$P\left(\left|\sum_{i=1}^n x_{ij}\right| \geq x(a_j + b_j + V_{nj})\right) \leq 8 \exp(-x^2/8), \quad (36)$$

for any $1 \leq j \leq p$. In particular, we can choose $a_j^2 = 2b_j^2 = 2\mathbb{E}V_{nj}^2$.

Proof. We only need to prove the result for $x > 1$ as the inequality holds trivially for $x < 1$. Suppose that the distributions of A_i and B_i are both symmetric, then we have

$$\begin{aligned} P\left(\sum_{i=1}^n x_{ij} > xV_{nj}\right) &\leq P\left(\sum_{i=1}^r (A_{ij} + B_{ij}) > xV_{nj}\right) \\ &\leq P\left(\sum_{i=1}^r A_{ij} > xV_{nj}/2\right) + P\left(\sum_{i=1}^r B_{ij} > xV_{nj}/2\right) \\ &\leq P\left(\sum_{i=1}^r A_{ij} > xV_{1nj}/2\right) + P\left(\sum_{i=1}^r B_{ij} > xV_{2nj}/2\right) \\ &\leq 2 \exp(-x^2/8), \end{aligned}$$

where we have used Theorem 2.15 in [15].

Let $\{\xi_{ij}\}_{i=1}^n$ be an independent copy of $\{x_{ij}\}_{i=1}^n$ in the sense that $\{\xi_{ij}\}_{i=1}^n$ have the same joint distribution as that for $\{x_{ij}\}_{i=1}^n$, and define V'_{nj} (A'_{ij} and B'_{ij}) in the same way as V_{nj} (A_{ij} and B_{ij}) by replacing $\{x_{ij}\}_{i=1}^n$ with $\{\xi_{ij}\}_{i=1}^n$. Following the arguments in the proof of Theorem 2.16 in [15], we deduce that for $x > 1$,

$$\begin{aligned} &\left\{ \sum_{i=1}^n x_{ij} > x(a_j + b_j + V_{nj}), \sum_{i=1}^n \xi_{ij} \leq a_j, V'_{nj} \leq b_j \right\} \\ &\subset \left\{ \sum_{i=1}^n (x_{ij} - \xi_{ij}) \geq x(a_j + b_j + V_{nj}) - a_j, V'_{nj} \leq b_j \right\} \\ &\subset \left\{ \sum_{i=1}^n (x_{ij} - \xi_{ij}) \geq x(a_j + b_j + V_{nj}^* - V'_{nj}) - a_j, V'_{nj} \leq b_j \right\} \\ &\subset \left\{ \sum_{i=1}^n (x_{ij} - \xi_{ij}) \geq xV_{nj}^* \right\}, \end{aligned}$$

where we have used the fact that

$$V_{nj}^* \equiv \sqrt{\sum_{l=1}^r (A_{lj} - A'_{lj})^2 + \sum_{l=1}^r (B_{lj} - B'_{lj})^2} \leq V_{nj} + V'_{nj}.$$

We note that $A_{lj} - A'_{lj}$ and $B_{lj} - B'_{lj}$ are symmetric, and

$$P\left(\sum_{i=1}^n \xi_{ij} \leq a_j, V'_{nj} \leq b_j\right) \geq 1/2.$$

Thus, we obtain

$$\begin{aligned} & P\left(\sum_{i=1}^n x_{ij} \geq x(a_j + b_j + V_{nj})\right) \\ &= \frac{P(\sum_{i=1}^n x_{ij} \geq x(a_j + b_j + V_{nj}), \sum_{i=1}^n \xi_{ij} \leq a_j, V'_{nj} \leq b_j)}{P(\sum_{i=1}^n \xi_{ij} \leq a_j, V'_{nj} \leq b_j)} \\ &\leq 2P\left(\sum_{i=1}^n x_{ij} \geq x(a_j + b_j + V_{nj}), \sum_{i=1}^n \xi_{ij} \leq a_j, V'_{nj} \leq b_j\right) \\ &\leq 2P\left(\sum_{i=1}^n (x_{ij} - \xi_{ij}) \geq xV_{nj}^*\right) \\ &\leq 4\exp(-x^2/8). \end{aligned}$$

Hence, we get

$$P\left(\left|\sum_{i=1}^n x_{ij}\right| \geq x(a_j + b_j + V_{nj})\right) \leq 8\exp(-x^2/8).$$

In particular, we can choose $b_j^2 = 4\mathbb{E}V_{nj}^2$ and $a_j^2 = 2b_j^2 = 8\mathbb{E}V_{nj}^2$ because

$$4\mathbb{E}\left(\sum_{i=1}^n x_{ij}\right)^2 \leq 8\mathbb{E}\left(\sum_{j=1}^r A_j\right)^2 + 8\mathbb{E}\left(\sum_{j=1}^r B_j\right)^2 = 8\mathbb{E}V_{nj}^2.$$

□

Let $\varphi(M_x) := \varphi_{N,M}(M_x)$ be the smallest finite constant which satisfies that uniformly for i and j ,

$$E(A_{ij} - \check{A}_{ij})^2 \leq N\varphi^2(M_x), \quad E(B_{ij} - \check{B}_{ij})^2 \leq M\varphi^2(M_x), \quad (37)$$

where \check{A}_{ij} and \check{B}_{ij} are the truncated versions of A_{ij} and B_{ij} defined as follows:

$$\begin{aligned} \check{A}_{ij} &= \sum_{l=iN+(i-1)M-N+1}^{iN+(i-1)M} (x_{lj} \wedge M_x) \vee (-M_x), \\ \check{B}_{ij} &= \sum_{l=i(N+M)-M+1}^{i(N+M)} (x_{lj} \wedge M_x) \vee (-M_x). \end{aligned}$$

Similarly, we can define the quantity $\varphi(M_y)$ for the Gaussian sequence $\{y_i\}$. Set $\varphi(M_x, M_y) = \varphi(M_x) \vee \varphi(M_y)$. Further let $u_x(\gamma)$ and $u_y(\gamma)$ be the smallest quantities such that

$$P\left(\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} |x_{ij}| \leq u_x(\gamma)\right) \geq 1 - \gamma, \quad P\left(\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} |y_{ij}| \leq u_y(\gamma)\right) \geq 1 - \gamma. \quad (38)$$

Building on the above results, we are ready to derive an upper bound for ρ_n . Consider a “smooth” indicator function $g_0 \in C_b^3(\mathbb{R}) : \mathbb{R} \rightarrow [0, 1]$ such that $g_0(s) = 1$ for $s \leq 0$ and $g_0(s) = 0$ for $s \geq 1$. Fix any $t \in \mathbb{R}$ and define $g(s) = g_0(\psi(s - t - e_\beta))$ with $e_\beta = \beta^{-1} \log p$. For this function g , $G_0 = 1$, $G_1 \lesssim \psi$, $G_2 \lesssim \psi^2$ and $G_3 \lesssim \psi^3$. Here, ψ is a smoothing parameter we will choose carefully in the proof. Lemma A.1 and Corollary 2.1 imply the following result.

Proposition A.1. *Consider a M -dependent time series $\{x_i\}$ and its Gaussian counterpart $\{y_i\}$. Suppose $2\sqrt{5}\beta(6M+1)M_{xy}/\sqrt{n} \leq 1$ with $M_{xy} = \max\{M_x, M_y\}$, and $M_x > u_x(\gamma)$ and $M_y > u_y(\gamma)$ for some $\gamma \in (0, 1)$. Further suppose that there exists constants $0 < c_1 < c_2$ such that $c_1 < \min_{1 \leq j \leq p} \sigma_{j,j} \leq \max_{1 \leq j \leq p} \sigma_{j,j} < c_2$ uniformly holds for all large enough M and p , where $\sigma_{j,k} = \text{cov}(X_j, X_k)$. Then for any $\psi > 0$,*

$$\begin{aligned} \rho_n &= \sup_{t \in \mathbb{R}} |P(T_X \leq t) - P(T_Y \leq t)| \\ &\lesssim (\psi^2 + \psi\beta)\phi(M_x, M_y) + (\psi^3 + \psi^2\beta + \psi\beta^2) \frac{(2M+1)^2}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3) \\ &\quad + \psi\varphi(M_x, M_y)\sqrt{\log(p/\gamma)} + \gamma + (\beta^{-1} \log(p) + \psi^{-1})\sqrt{1 \vee \log(p\psi)}. \end{aligned}$$

Proof. Note that

$$\begin{aligned} \mathbb{E}[1 - \mathcal{I}] &\leq P\left(\max_{1 \leq j \leq p} |X_j - \tilde{X}_j| > \Delta\right) + P\left(\max_{1 \leq j \leq p} |Y_j - \tilde{Y}_j| > \Delta\right) \\ &\leq \sum_{j=1}^p \{P(|X_j - \tilde{X}_j| > \Delta) + P(|Y_j - \tilde{Y}_j| > \Delta)\}. \end{aligned}$$

Let

$$\begin{aligned} \Lambda_j &\equiv (2 + 2\sqrt{2}) \sqrt{\sum_{i=1}^r \mathbb{E}(A_{ij} - \tilde{A}_{ij})^2/n + \sum_{j=1}^r \mathbb{E}(B_{ij} - \tilde{B}_{ij})^2/n} \\ &\quad + \sqrt{\sum_{i=1}^r (A_{ij} - \tilde{A}_{ij})^2/n + \sum_{i=1}^r (B_{ij} - \tilde{B}_{ij})^2/n} = \Lambda_{1j} + \Lambda_{2j}, \end{aligned}$$

where

$$\tilde{A}_{ij} = \sum_{l=(i-1)(N+M)+1}^{iN+(i-1)M} \tilde{x}_{lj}, \quad \tilde{B}_{ij} = \sum_{l=iN+(i-1)M+1}^{i(N+M)} \tilde{x}_{lj}.$$

Applying Lemma A.1 and using the union bound, we have with probability at least $1 - 8\gamma$,

$$|X_j - \tilde{X}_j| \leq \Lambda_j \sqrt{8 \log(p/\gamma)}, \quad 1 \leq j \leq p.$$

By the assumption,

$$P\left(\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} |x_{ij}| \leq M_x\right) \geq 1 - \gamma, \quad P\left(\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} |y_{ij}| \leq M_y\right) \geq 1 - \gamma.$$

Therefore with probability at least $1 - \gamma$,

$$\begin{aligned} \Lambda_j &\leq (2 + 2\sqrt{2}) \sqrt{\sum_{i=1}^r \mathbb{E}(A_{ij} - \check{A}_{ij})^2/n + \sum_{i=1}^r \mathbb{E}(B_{ij} - \check{B}_{ij})^2/n} \\ &\quad + \sqrt{\sum_{i=1}^r (\mathbb{E}\check{A}_{ij})^2/n + \sum_{i=1}^r (\mathbb{E}\check{B}_{ij})^2/n}, \\ &\leq (3 + 2\sqrt{2})\varphi(M_x)\sqrt{Nr/n + Mr/n} \lesssim \varphi(M_x), \end{aligned}$$

where we have used the fact that $\mathbb{E}A_{ij} = \mathbb{E}B_{ij} = 0$ and the Cauchy-Schwarz inequality. The same argument applies to the Gaussian sequence $\{y_i\}$.

Summarizing the above results and along with (5), we deduce that

$$\begin{aligned} &|\mathbb{E}[m(X) - m(Y)]| \\ &\lesssim (G_2 + G_1\beta)\phi(M_x, M_y) + (G_3 + G_2\beta + G_1\beta^2) \frac{(2M+1)^2}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3) \\ &\quad + G_1\varphi(M_x, M_y)\sqrt{8 \log(p/\gamma)} + G_0\gamma, \end{aligned} \quad (39)$$

which also implies that

$$\begin{aligned} &|\mathbb{E}[g(T_X) - g(T_Y)]| \\ &\lesssim (G_2 + G_1\beta)\phi(M_x, M_y) + (G_3 + G_2\beta + G_1\beta^2) \frac{(2M+1)^2}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3) \\ &\quad + G_1\varphi(M_x, M_y)\sqrt{8 \log(p/\gamma)} + G_0\gamma + \beta^{-1}G_1 \log p, \end{aligned} \quad (40)$$

for M -dependent sequence, provided that $2\sqrt{5}\beta(6M+1)M_{xy}/\sqrt{n} < 1$. Consider a “smooth” indicator function $g_0 \in C^3(\mathbb{R}) : \mathbb{R} \rightarrow [0, 1]$ such that $g_0(s) = 1$ for $s \leq 0$ and $g_0(s) = 0$ for $s \geq 1$. Fix any $t \in \mathbb{R}$ and define $g(s) = g_0(\psi(s - t - e_\beta))$ with $e_\beta = \beta^{-1} \log p$. The conclusion follows from the proof of Corollary F.1 in [12] and Lemma 2.1 in [13] regarding the anti-concentration property for Gaussian distribution. We omit the details to conserve the space. \square

A.3. Proofs of the main results in Section 2.2

Proof of Theorem 2.1. For clarity, we present the proof in the following five steps.

Step 1: Construct the M -dependent sequence as

$$x_i := x_i^{(M)} = \mathbb{E}[\mathcal{G}(\dots, \epsilon_{i-1}, \epsilon_i) | \epsilon_{i-M}, \epsilon_{i-M+1}, \dots, \epsilon_i].$$

By construction, x_{1j} and $x_{(1+l)k}^{(l-1)}$ are independent for any $1 \leq j, k \leq p$. The triangle inequality and (16) imply that

$$|\mathbb{E}[m(X) - m(Y^{(M)})]| \lesssim |\mathbb{E}[m(X^{(M)}) - m(Y^{(M)})]| + (G_0 G_1^q)^{1/(1+q)} \left(\sum_{j=1}^p \Theta_{M,j,q}^q \right)^{1/(1+q)},$$

where $X^{(M)} = \sum_{i=1}^n x_i^{(M)} / \sqrt{n}$ and $Y^{(M)} = \sum_{i=1}^n y_i^{(M)} / \sqrt{n}$ with $y_i^{(M)}$ being the M -dependent approximation for $\{y_i\}$. By (39)

$$\begin{aligned} & |\mathbb{E}[m(X) - m(Y^{(M)})]| \\ & \lesssim (G_2 + G_1 \beta) \phi^{(M)}(M_x, M_y) + (G_3 + G_2 \beta + G_1 \beta^2) \frac{(2M+1)^2}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3) \\ & \quad + G_1 \varphi^{(M)}(M_x, M_y) \sqrt{8 \log(p/\gamma)} + G_0 \gamma + (G_0 G_1^q)^{1/(1+q)} \left(\sum_{j=1}^p \Theta_{M,j,q}^q \right)^{1/(1+q)}, \end{aligned}$$

where $\phi^{(M)}(M_x, M_y)$ and $\varphi^{(M)}(M_x, M_y)$ are defined based on $\{x_i^{(M)}\}$ and $\{y_i^{(M)}\}$. Following the arguments in the proof of Proposition A.1, we have

$$\begin{aligned} \rho_n & \lesssim (\psi^2 + \psi \beta) \phi^{(M)}(M_x, M_y) + (\psi^3 + \psi^2 \beta + \psi \beta^2) \frac{(2M+1)^2}{\sqrt{n}} (\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3) \\ & \quad + \psi \varphi^{(M)}(M_x, M_y) \sqrt{\log(p/\gamma)} + \gamma + (\beta^{-1} \log(p) + \psi^{-1}) \sqrt{1 \vee \log(p\psi)} \\ & \quad + \psi^{q/(1+q)} \left(\sum_{j=1}^p \Theta_{M,j,q}^q \right)^{1/(1+q)}, \end{aligned} \quad (41)$$

where φ , M , M_x , M_y and β will be chosen properly.

Step 2: Next we quantify $\phi^{(M)}(M_x)$ and $\varphi^{(M)}(M_x)$. To this end, define the projection operator

$$\mathcal{P}_j x_{ik} = \mathbb{E}[x_{ik} | \epsilon_{i-j}, \dots, \epsilon_i] - \mathbb{E}[x_{ik} | \epsilon_{i-j+1}, \dots, \epsilon_i].$$

Note that

$$\mathcal{P}_j x_{ik} = \mathbb{E}[\mathcal{G}_{ik}(\dots, \epsilon_{i-1}, \epsilon_i) - \mathcal{G}_{ik}(\dots, \epsilon'_{i-j}, \epsilon_{i-j+1}, \dots, \epsilon_{i-1}, \epsilon_i) | \epsilon_{i-j}, \dots, \epsilon_i].$$

Jensen's inequality yields that $\|\mathcal{P}_j x_{ik}\|_q \leq \theta_{j,k,q}(x)$. Let $\check{x}_{ij} = x_{ij} - \tilde{x}_{ij}$ and $\chi_{ij} = (x_{ij} \wedge M_x) \vee (-M_x)$. Based on $\{x_i^{(M)}\}$, we define the variables $A_{ij}^{(M)}, \tilde{A}_{ij}^{(M)}, \check{A}_{ij}^{(M)}, B_{ij}^{(M)}, \tilde{B}_{ij}^{(M)}$ in a similar way as before. Similarly, we can define $\check{x}_{ij}^{(l)}$ and $\chi_{ij}^{(l)}$ based on $x_{ij}^{(l)}$. For $M \geq l$, we note that $x_{ik}^{(M)} - x_{ik}^{(l)} = \sum_{j=l+1}^M \mathcal{P}_j x_{ik}$. Because $x_{ij}^{(M)}$ and $x_{(i+l)k}^{(l-1)}$ are independent for any $1 \leq j, k \leq p$ and $\mathbb{E}x_{ij} = \mathbb{E}\check{x}_{ij} = 0$, we obtain for $l > 0$,

$$\begin{aligned} |\mathbb{E}\check{x}_{ij}^{(M)} x_{(i+l)k}^{(M)}| &= |\mathbb{E}\check{x}_{ij}^{(M)} (x_{(i+l)k}^{(M)} - x_{(i+l)k}^{(l-1)})| \leq \|\check{x}_{ij}^{(M)}\|_2 \|x_{(i+l)k}^{(M)} - x_{(i+l)k}^{(l-1)}\|_2 \\ &\leq \|x_{ij}\|_4^2 \|x_{(i+l)k}^{(M)} - x_{(i+l)k}^{(l-1)}\|_2 / M_x \lesssim \sum_{j=l}^M \theta_{j,k,2} / M_x, \end{aligned}$$

where we have used the fact that $\|\check{x}_{ij}^{(M)}\|_2^2 \leq \mathbb{E}(x_{ij}^{(M)} - \chi_{ij}^{(M)})^2 \mathbf{I}\{|x_{ij}^{(M)}| > M_x\} \leq \mathbb{E}x_{ij}^4 / M_x^2$. Using the fact that the map $x \rightarrow (x \wedge M_x) \vee (-M_x)$ is Lipschitz continuous, we deduce that

$$\begin{aligned} |\mathbb{E}\check{x}_{ij}^{(M)} \check{x}_{(i+l)k}^{(M)}| &= |\mathbb{E}\check{x}_{ij}^{(M)} \{x_{(i+l)k}^{(M)} - x_{(i+l)k}^{(l-1)} - (\chi_{(i+l)k}^{(M)} - \chi_{(i+l)k}^{(l-1)})\}| \\ &\quad \times \mathbf{I}\{|x_{(i+l)k}^{(M)}| > M_x \text{ or } |x_{(i+l)k}^{(l-1)}| > M_x\}| \\ &\lesssim (\mathbb{E}|x_{ij}|^3)^{1/3} (\mathbb{E}|x_{(i+l)k}^{(M)} - x_{(i+l)k}^{(l-1)}|^3 + \mathbb{E}|\chi_{(i+l)k}^{(M)} - \chi_{(i+l)k}^{(l-1)}|^3)^{1/3} \\ &\quad \times \{P(|x_{(i+l)k}^{(M)}| > M_x) + P(|x_{(i+l)k}^{(l-1)}| > M_x)\}^{1/3} \\ &\lesssim \|x_{ij}\|_3 \|x_{(i+l)k}^{(M)} - x_{(i+l)k}^{(l-1)}\|_3 \|x_{(i+l)k}\|_3 / M_x \lesssim \sum_{j=l}^M \theta_{j,k,3} / M_x. \end{aligned}$$

Note $|\mathbb{E}\check{x}_{ij}^{(M)} x_{ik}^{(M)}| \leq \|x_{ij}\|_4^2 \|x_{ik}\|_2 / M_x$. It is not hard to show that the above result holds if $x_{ij}^{(M)}$ (or $x_{(i+l)k}^{(M)}$) is replaced by its $\tilde{x}_{ij}^{(M)}$ (or $\tilde{x}_{(i+l)k}^{(M)}$). Therefore, we have

$$\begin{aligned} &\max_{1 \leq j, k \leq p} \frac{1}{n} \sum_{i=1}^n \left| \sum_{l=(i-M) \vee 1}^{(i+M) \wedge n} (\mathbb{E}x_{ij}^{(M)} x_{lk}^{(M)} - \mathbb{E}\tilde{x}_{ij}^{(M)} \tilde{x}_{lk}^{(M)}) \right| \\ &\leq \max_{1 \leq j, k \leq p} \frac{1}{n} \sum_{i=1}^n \left| \sum_{l=(i-M) \vee 1}^{(i+M) \wedge n} (\mathbb{E}x_{ij}^{(M)} \check{x}_{lk}^{(M)} + \mathbb{E}\check{x}_{ij}^{(M)} \tilde{x}_{lk}^{(M)}) \right| \\ &\lesssim \max_{1 \leq k \leq p} \sum_{l=1}^M \sum_{j=l}^M \theta_{j,k,3} / M_x \lesssim 1 / M_x. \end{aligned}$$

Next, we consider $\varphi^{(M)}(M_x)$. Similar argument implies that for $l > 0$,

$$\begin{aligned} |\mathbb{E}\check{x}_{ik}^{(M)} \check{x}_{(i+l)k}^{(M)}| &= |\mathbb{E}\check{x}_{ik}^{(M)} \{\check{x}_{(i+l)k}^{(M)} - \check{x}_{(i+l)k}^{(l-1)} - \mathbb{E}(\chi_{(i+l)k}^{(M)} - \chi_{(i+l)k}^{(l-1)})\}| \\ &\quad \times \mathbf{I}\{|x_{(i+l)k}^{(M)}| > M_x \text{ or } |x_{(i+l)k}^{(l-1)}| > M_x\}| \end{aligned}$$

$$\begin{aligned}
&\lesssim (\mathbb{E}|\check{x}_{ik}^{(M)}|^2)^{1/2} (\mathbb{E}|x_{(i+l)k}^{(M)} - x_{(i+l)k}^{(l-1)}|^3 + \mathbb{E}|\chi_{(i+l)k}^{(M)} - \chi_{(i+l)k}^{(l-1)}|^3)^{1/3} \\
&\quad \times \{P(|x_{(i+l)k}^{(M)}| > M_x) + P(|x_{(i+l)k}^{(l-1)}| > M_x)\}^{1/6} \\
&\lesssim (\mathbb{E}|x_{ik}|^4/M_x^2)^{1/2} \|x_{(i+l)k}^{(M)} - x_{(i+l)k}^{(l-1)}\|_3 \\
&\quad \times (\mathbb{E}|x_{(i+l)k}^{(M)}|^4/M_x^4 + \mathbb{E}|x_{(i+l)k}^{(l-1)}|^4/M_x^4)^{1/6} \\
&\lesssim \|x_{ik}\|_4^2 \|x_{(i+l)k}\|_4^{2/3} \|x_{(i+l)k}^{(M)} - x_{(i+l)k}^{(l-1)}\|_3 / M_x^{5/3} \lesssim \sum_{j=l}^M \theta_{j,k,3} / M_x^{5/3}.
\end{aligned} \tag{42}$$

Note $|\mathbb{E}\check{x}_{ik}^{(M)}\check{x}_{ik}^{(M)}| \lesssim 1/M_x^2$. Thus, we obtain

$$\mathbb{E}(A_{ij}^{(M)} - \tilde{A}_{ij}^{(M)})^2 / N \lesssim \sum_{l=1}^N \sum_{j=l}^M \theta_{j,k,3} / M_x^{5/3} + 1/M_x^2 \lesssim 1/M_x^{5/3}.$$

Similarly $\mathbb{E}(B_{ij}^{(M)} - \tilde{B}_{ij}^{(M)})^2 / M \lesssim 1/M_x^{5/3}$. Notice that

$$\begin{aligned}
&\mathbb{E}(A_{ij}^{(M)} - \check{A}_{ij}^{(M)})^2 / N \\
&= \mathbb{E}(A_{ij}^{(M)} - \tilde{A}_{ij}^{(M)})^2 / N + (\mathbb{E}\check{A}_{ij}^{(M)})^2 / N \\
&\lesssim 1/M_x^{5/3} + \left\{ \sum_{l=iN+(i-1)M-N+1}^{iN+(i-1)M} \mathbb{E}(\chi_{lj}^{(M)} - x_{lj}^{(M)}) \mathbf{I}\{|x_{lj}^{(M)}| > M_x\} \right\}^2 / N \\
&\lesssim 1/M_x^{5/3} + N \left(\max_{ij} \mathbb{E}|x_{ij}|^4 / M_x^3 \right)^2
\end{aligned}$$

We can choose $\varphi^{(M)}(M_x) = C'(1/M_x^{5/6} + \sqrt{N}/M_x^3)$ for some constant $C' > 0$.

Step 3: We consider the quantities associated with $\{y_i^{(M)}\}$. First, note that

$$\begin{aligned}
|\mathbb{E}\check{y}_{ij}^{(M)} y_{(i+l)k}^{(M)}| &\leq |\mathbb{E}[(y_{ij}^{(M)} - M_y) y_{(i+l)k}^{(M)} \mathbf{I}\{y_{ij}^{(M)} > M_y\}]| \\
&\quad + |\mathbb{E}[(y_{ij}^{(M)} + M_y) y_{(i+l)k}^{(M)} \mathbf{I}\{y_{ij}^{(M)} < -M_y\}]|.
\end{aligned} \tag{43}$$

Because $y_{(i+l)k}^{(M)}$ is Gaussian conditional on $y_{ij}^{(M)}$ ($y_{(i+l)k}^{(M)}$ and $y_{ij}^{(M)}$ are jointly Gaussian), we have

$$\mathbb{E}[y_{(i+l)k}^{(M)} | y_{ij}^{(M)}] = \frac{\mathbb{E}[x_{ij}^{(M)} x_{(i+l)k}^{(M)}]}{\mathbb{E}[|x_{ij}^{(M)}|^2]} y_{ij}^{(M)},$$

which implies that

$$\begin{aligned}
 & |\mathbb{E}[(y_{ij}^{(M)} - M_y) y_{(i+l)k}^{(M)} \mathbf{I}\{y_{ij}^{(M)} > M_y\}]| \\
 &= |\mathbb{E}[(y_{ij}^{(M)} - M_y) \mathbf{I}\{y_{ij}^{(M)} > M_y\} \mathbb{E}[y_{(i+l)k}^{(M)} | y_{ij}^{(M)}]]| \\
 &= \frac{|\mathbb{E}[x_{ij}^{(M)} x_{(i+l)k}^{(M)}]|}{\mathbb{E}[|x_{ij}^{(M)}|^2]} |\mathbb{E}[(y_{ij}^{(M)} - M_y) y_{ij}^{(M)} \mathbf{I}\{y_{ij}^{(M)} > M_y\}]| \\
 &\leq \frac{|\mathbb{E}[x_{ij}^{(M)} x_{(i+l)k}^{(M)}]| \mathbb{E}[|y_{ij}^{(M)}|^4]}{\mathbb{E}[|x_{ij}^{(M)}|^2] M_y^2} \\
 &\leq \frac{3 |\mathbb{E}[x_{ij}^{(M)} x_{(i+l)k}^{(M)}]| \mathbb{E}[|x_{ij}^{(M)}|^2]}{M_y^2}.
 \end{aligned}$$

Due to the Gaussian tail, the degree of M_y can be made arbitrarily larger but the current choice suffices for our analysis. The same argument applies to the second term on the RHS of (43), which leads to

$$|\mathbb{E}[\check{y}_{ij}^{(M)} y_{(i+l)k}^{(M)}]| \lesssim |\mathbb{E}[x_{ij}^{(M)} x_{(i+l)k}^{(M)}]| / M_y^2 \lesssim \sum_{j=l}^M \theta_{j,k,2} / M_y^2. \quad (44)$$

To deal with $\mathbb{E}[\check{y}_{ij}^{(M)} \check{y}_{(i+l)k}^{(M)}]$, we first state a result. For $x \sim N(\mu, \sigma^2)$, it can be shown that

$$\begin{aligned}
 |\mathbb{E}[(x \wedge M) \vee (-M)]| &\leq |\mathbb{E}[(x - \mu) \mathbf{I}\{|x| \leq M\}]| + |\mu| \\
 &\quad + |MP(x > M) - MP(x < -M)| \\
 &\lesssim |\mu| + M|\mu|/\sigma.
 \end{aligned}$$

Using this fact and the Gaussian assumption, we have

$$\begin{aligned}
 \mathbb{E}[\check{y}_{ij}^{(M)} \check{y}_{(i+l)k}^{(M)}] &= \mathbb{E}[\check{y}_{ij}^{(M)} \mathbb{E}[\check{y}_{(i+l)k}^{(M)} | y_{ij}^{(M)}]] \\
 &\lesssim \frac{|\mathbb{E}[x_{ij}^{(M)} x_{(i+l)k}^{(M)}]| \mathbb{E}[|y_{ij}^{(M)}|^5]}{\mathbb{E}[|x_{ij}^{(M)}|^2] M_y^2} \lesssim \frac{\sum_{j=l}^M \theta_{j,k,2}}{M_y^2}.
 \end{aligned} \quad (45)$$

Therefore, using the same argument as that for x_i , we can set $\phi^{(M)}(M_y) = C/M_y^2$. By (44) and (45), we get

$$\mathbb{E}[\check{y}_{ij}^{(M)} \check{y}_{(i+l)k}^{(M)}] \lesssim \frac{\sum_{j=l}^M \theta_{j,k,2}}{M_y^2}.$$

Similar arguments as before show that $\varphi^{(M)}(M_y) = C'/M_y^2$ for some $C' > 0$.

Step 4: By the assumption that $\max_{i,j} \|x_{ij}\|_4 < \infty$ and the fact that $\|y_{ij}^{(M)}\|_2 = \|x_{ij}^{(M)}\|_2$, we have $E|x_{ij}^{(M)}|^3 \leq (E|x_{ij}|^4)^{3/4}$, and $E|y_{ij}^{(M)}|^3 \lesssim E|x_{ij}|^3 < \infty$, which implies that $\bar{m}_{x,3}^3 + \bar{m}_{y,3}^3 < \infty$. Thus, we ignore the constants and set

$$\psi \asymp n^{1/8} M^{-1/2} l_n^{-3/8}, \quad M_x = M_y = u \asymp n^{3/8} M^{-1/2} l_n^{-5/8}.$$

Let $2\sqrt{5}\beta(6M+1)M_{xy}/\sqrt{n} = 1$, that is $\beta \asymp \sqrt{n}/(uM)$. Under the assumption that $n^{7/4} \times M^{-1} l_n^{-9/4} \geq C_3 M$, it is straightforward to check the following:

$$(\psi^2 + \psi\beta)\phi(M_x, M_y) \lesssim \psi^2/u + \psi\sqrt{n}/(u^2 M) \lesssim n^{-1/8} M^{1/2} l_n^{7/8}, \quad (46)$$

$$(\psi^3 + \psi^2\beta + \psi\beta^2) \frac{(2M+1)^2}{\sqrt{n}} \lesssim \frac{\psi^3 M^2}{\sqrt{n}} + \frac{\psi^2 M}{u} + \frac{\psi\sqrt{n}}{u^2} \lesssim n^{-1/8} M^{1/2} l_n^{7/8}, \quad (47)$$

$$\psi\varphi(M_x, M_y)\sigma_j\sqrt{8\log(p/\gamma)} \lesssim \frac{\psi l_n^{1/2}}{u^{5/6}} + \frac{\psi l_n^{1/2}\sqrt{N}}{u^3} \lesssim n^{-1/8} M^{1/2} l_n^{7/8}, \quad (48)$$

$$(\beta^{-1}\log(p) + \psi^{-1})\sqrt{1 \vee \log(p\psi)} \lesssim \frac{l_n^{3/2}Mu}{\sqrt{n}} + \psi^{-1}l_n^{1/2} \lesssim n^{-1/8} M^{1/2} l_n^{7/8}. \quad (49)$$

By Lemma A.2, we have $c_1/2 < \min_{1 \leq j \leq p} \sigma_{j,j}^{(M)} \leq \max_{1 \leq j \leq p} \sigma_{j,j}^{(M)} < 2c_2$ for large enough M , where $\sigma_{j,k}^{(M)} = \text{cov}(X_j^{(M)}, X_k^{(M)})$. It remains to verify that the selected u satisfying that

$$P\left(\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} |x_{ij}| \leq u\right) \geq 1 - \gamma, \quad P\left(\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} |y_{ij}^{(M)}| \leq u\right) \geq 1 - \gamma. \quad (50)$$

We first consider Condition (7). Using the convexity of h , we have

$$\begin{aligned} \mathbb{E}h\left(\max_{1 \leq j \leq p} |x_{ij}^{(M)}|/\mathcal{D}_n\right) &\leq \mathbb{E}h\left(\mathbb{E}\left[\max_{1 \leq j \leq p} |x_{ij}|/\mathcal{D}_n \mid \epsilon_{i-M}, \dots, \epsilon_i\right]\right) \\ &\leq \mathbb{E}h\left(\max_{1 \leq j \leq p} |x_{ij}|/\mathcal{D}_n\right) \leq C_1. \end{aligned}$$

By the fact that $\max_{1 \leq i \leq n} \mathbb{E}h(\max_{1 \leq j \leq p} |x_{ij}^{(M)}|/\mathcal{D}_n) \leq C_1$ and the arguments in the proof of Lemma 2.2 in [12], we have $u_x(\gamma) \lesssim \max\{\mathcal{D}_n h^{-1}(n/\gamma), l_n^{1/2}\}$ and $u_y(\gamma) \lesssim l_n^{1/2}$. Because $n^{3/8} M^{-1/2} l_n^{-5/8} \geq C \max\{\mathcal{D}_n h^{-1}(n/\gamma), l_n^{1/2}\}$, we can always choose $u = O(n^{3/8} M^{-1/2} l_n^{-5/8})$ such that (50) is fulfilled. We can prove a similar result under Condition (8). Therefore by (41), (46), (47), (48) and (49), we get

$$\begin{aligned} &\sup_{t \in \mathbb{R}} |P(T_X \leq t) - P(T_{Y^{(M)}} \leq t)| \\ &\lesssim n^{-1/8} M^{1/2} l_n^{7/8} + \gamma + (n^{1/8} M^{-1/2} l_n^{-3/8})^{q/(1+q)} \left(\sum_{j=1}^p \Theta_{M,j,q}^q\right)^{1/(1+q)}. \end{aligned} \quad (51)$$

Step 5: Let $\tilde{\Delta} = \max_{1 \leq j, l \leq p} |\text{cov}(X_j, X_l) - \text{cov}(X_j^{(M)}, X_l^{(M)})|$. With similar arguments in the proof of Lemma A.2, we have $\tilde{\Delta} \lesssim \max_{1 \leq j \leq p} \sum_{l=M+1}^{+\infty} l \theta_{l,j,2}(x) = \Xi_M$. Thus by Theorem 2 in [13], we have

$$\sup_{t \in \mathbb{R}} |P(T_Y \leq t) - P(T_{Y^{(M)}} \leq t)| \lesssim \Xi_M^{1/3} (1 \vee \log(p/\Xi_M))^{2/3}, \quad (52)$$

where we have used the fact that $f(x) = x^{1/3} (1 \vee \log(p/x))^{2/3}$ is monotonic increasing when $\log(p/x) > 2$. The conclusion follows by combining (51) and (52). \square

Lemma A.2. *Consider the M -dependent approximation sequence $\{x_i^{(M)}\}$. Suppose that $c_1 < \min_{1 \leq j \leq p} \sigma_{j,j} \leq \max_{1 \leq j \leq p} \sigma_{j,j} < c_2$, $\max_{i,j} \|x_{ij}\|_4 \leq c_3$, and $\sum_{j=1}^{+\infty} \max_{1 \leq k \leq p} j \theta_{j,k,2}(x) < \infty$. Then we have $c_1/2 < \min_{1 \leq j \leq p} \sigma_{j,j}^{(M)} \leq \max_{1 \leq j \leq p} \sigma_{j,j}^{(M)} < 2c_2$ for large enough M , where $\sigma_{j,k}^{(M)} = \text{cov}(X_j^{(M)}, X_k^{(M)})$.*

Proof. We claim that as $M \rightarrow +\infty$,

$$\max_{1 \leq j \leq p} n^{-1} \sum_{i,k=1}^n |\mathbb{E} x_{ij}^{(M)} x_{kj}^{(M)} - \mathbb{E} x_{ij} x_{kj}| \rightarrow 0, \quad (53)$$

which implies that $\max_{1 \leq j \leq p} |\sigma_{j,j}^{(M)} - \sigma_{j,j}| \rightarrow 0$. The conclusion thus follows from the assumption that $c_1 < \min_{1 \leq j \leq p} \sigma_{j,j} \leq \max_{1 \leq j \leq p} \sigma_{j,j} < c_2$. To show (53), we note that

$$\begin{aligned} |\mathbb{E} x_{ij}^{(M)} x_{kj}^{(M)} - \mathbb{E} x_{ij} x_{kj}| &\leq \|x_{ij}^{(M)} - x_{ij}\|_2 \|x_{kj}\|_2 + \|x_{kj}^{(M)} - x_{kj}\|_2 \|x_{ij}\|_2 \\ &\lesssim \sum_{l=M+1}^{+\infty} (\|\mathcal{P}_l x_{ij}\|_2 + \|\mathcal{P}_l x_{kj}\|_2) \lesssim \sum_{l=M+1}^{+\infty} \theta_{l,j,2}, \end{aligned}$$

and for $h > M$,

$$|\mathbb{E} x_{ij} x_{(i+h)j}| \leq |\mathbb{E} x_{ij} (x_{(i+h)j} - x_{(i+h)j}^{(h-1)})| \leq \|x_{ij}\|_2 \|x_{(i+h)j} - x_{(i+h)j}^{(h-1)}\|_2 \lesssim \sum_{l=h}^{+\infty} \theta_{l,j,2}.$$

Thus, we have

$$\begin{aligned} &\max_{1 \leq j \leq p} n^{-1} \sum_{i,k=1}^n |\mathbb{E} x_{ij}^{(M)} x_{kj}^{(M)} - \mathbb{E} x_{ij} x_{kj}| \\ &\leq \max_{1 \leq j \leq p} n^{-1} \sum_{|i-k| \leq M} |\mathbb{E} x_{ij}^{(M)} x_{kj}^{(M)} - \mathbb{E} x_{ij} x_{kj}| \\ &\quad + \max_{1 \leq j \leq p} n^{-1} \sum_{|i-k| > M} |\mathbb{E} x_{ij} x_{kj}| \end{aligned}$$

$$\begin{aligned}
&\lesssim \max_{1 \leq j \leq p} M \sum_{l=M+1}^{+\infty} \theta_{l,j,2} + \max_{1 \leq j \leq p} \sum_{h=M+1}^{n-1} \sum_{l=h}^{+\infty} \theta_{l,j,2} \\
&\lesssim \max_{1 \leq j \leq p} \sum_{l=M+1}^{+\infty} l \theta_{l,j,2} \leq \sum_{l=M+1}^{+\infty} \max_{1 \leq j \leq p} l \theta_{l,j,2},
\end{aligned}$$

which implies that $\max_{1 \leq j \leq p} n^{-1} \sum_{i,k=1}^n |\mathbb{E} x_{ij}^{(M)} x_{kj}^{(M)} - \mathbb{E} x_{ij} x_{kj}| \rightarrow 0$ as $M \rightarrow +\infty$. \square

Acknowledgements

Research was sponsored by NSF CAREER Award DMS-1151692, DMS-1418042, DMS-1607320, Simons Fellowship in Mathematics, Office of Naval Research (ONR N00014-15-1-2331) and a grant from Indiana Clinical and Translational Sciences Institute.

References

- [1] Andrews, D.W.K. (1984). Nonstrong mixing autoregressive processes. *J. Appl. Probab.* **21** 930–934. [MR0766830](#)
- [2] Avram, F. and Bertsimas, D. (1993). On central limit theorems in geometrical probability. *Ann. Appl. Probab.* **3** 1033–1046. [MR1241033](#)
- [3] Baldi, P. and Rinott, Y. (1989). Asymptotic normality of some graph-related statistics. *J. Appl. Probab.* **26** 171–175. [MR0981262](#)
- [4] Baldi, P. and Rinott, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17** 1646–1650. [MR1048950](#)
- [5] Bentkus, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. [MR1965117](#)
- [6] Cai, T.T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. [MR2850210](#)
- [7] Cai, T.T., Liu, W. and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. [MR3164870](#)
- [8] Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Statist.* **35** 2313–2351. [MR2382644](#)
- [9] Chatterjee, S. (2005). An error bound in the Sudakov–Fernique inequality. Preprint. Available at [arXiv:math/0510424](#).
- [10] 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](#)
- [11] Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Testing many moment inequalities. Preprint. Available at [arXiv:1312.7614](#).
- [12] 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.
- [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] Cho, H. and Fryzlewicz, P. (2014). Multiple change-point detection for high-dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 475–507.
- [15] de la Peña, V.H., Lai, T.L. and Shao, Q.-M. (2009). *Self-Normalized Processes: Limit Theory and Statistical Applications. Probability and Its Applications (New York)*. Berlin: Springer. [MR2488094](#)
- [16] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739.
- [17] Han, F. and Liu, H. (2014). A direct estimation of high dimensional stationary vector autoregressions. Preprint. Available at [arXiv:1307.0293](#).
- [18] Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. [MR2933663](#)
- [19] Liu, W. and Lin, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Process. Appl.* **119** 249–280.
- [20] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer. [MR2172368](#)
- [21] Petrovskayam, B. and Leontovicha, M. (1982). The central limit theorem for a sequence of random variables with a slowly growing number of dependencies. *Theory Probab. Appl.* **27** 815–825.
- [22] Pinelis, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** 1679–1706. [MR1331198](#)
- [23] Portnoy, S. (1986). On the central limit theorem in \mathbb{R}^p when $p \rightarrow \infty$. *Probab. Theory Related Fields* **73** 571–583.
- [24] Röllin, A. (2013). Stein’s method in high dimensions with applications. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 529–549. [MR3088380](#)
- [25] Romano, J. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc.* **100** 94–108.
- [26] Stein, C. (1986). *Approximate Computation of Expectations. Institute of Mathematical Statistics Lecture Notes – Monograph Series 7*. Hayward, CA: IMS. [MR0882007](#)
- [27] Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. [MR2172215](#)
- [28] Wu, W.B. (2011). Asymptotic theory for stationary processes. *Stat. Interface* **4** 207–226.
- [29] Wu, W.B. and Shao, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436. [MR2052582](#)
- [30] Wu, W.B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21** 1397–1413. [MR2827528](#)
- [31] Zhang, D. and Wu, W.B. (2016). Gaussian approximation for high dimensional time series. *Ann. Statist.* To appear.
- [32] Zhang, X. and Cheng, G. (2016). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.* To appear.

Received January 2016 and revised September 2016