Kolmogorov Dependence Theory

Huitong Qiu [*], Fang Han [†], Han Liu [‡], and Brian Caffo [§]

April 3, 2021

Abstract

Serial dependence is a general phenomenon in time series data, and has motivated the development of many dependence conditions. However, the commonality between these dependence conditions is relatively less known. In this paper, we propose a new dependence measure, named Kolmogorov dependence measure, and develop the Kolmogorov dependence condition under this measure. We show that the Kolmogorov dependence condition unifies a number of widely used dependence conditions by serving as a common necessary condition. To demonstrate the applicability of the Kolmogorov dependence condition, we derive the rates of convergence for a class of large quantile-based scatter matrix estimators under this condition. This manifests the usefulness of the results devised in this paper since asymptotic analysis of quantile statistics under dependence is known to be challenging in high dimensions.

Keywords: weak dependence; scatter matrix estimation; quantile statistics

1 INTRODUCTION

Dependent data arises from a wide range of applications (Fan et al., 2014). For example, in finance, the series of asset returns commonly exhibit short-term or long-term memory (Andersen, 2009); in functional magnetic resonance imaging (fMRI), the images from consecutive scans are serially dependent (Purdon and Weisskoff, 1998; Woolrich et al., 2001); in geophysics, measurements made in geographical sites over time usually exhibit temporal dependence (Majda and Wang, 2006).

The prevalence of serial dependence has motivated the development of various dependence conditions. These conditions can be categorized into structural conditions and non-structural conditions. The former are based on specific models for the data generating mechanism. Examples of structural conditions include vector autoregressive (VAR) models and physical dependence conditions. We provide a brief review of these conditions and their applications in high dimensions.

- **VAR models**: The VAR models specify that the observed random vector depends linearly on its previous realizations. Under this model, Loh and Wainwright (2012) investigated sparse linear regression; Han and Liu (2013c) and Qiu et al. (2015b) proposed to estimate

[*]Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA; e-mail: `hqiu7@jhu.edu`
[†]Department of Statistics, University of Washington, Seattle,WA 98195, USA; e-mail: `fanghan@uw.edu`
[‡]Department of Statistics, Northwestern University, Evanston, IL 60208, USA; e-mail: `hanliu@northwestern.edu`
[§]Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA; e-mail: `bcaffo@jhu.edu`

the transition matrix via Dantzig-selector-type approaches; Wang et al. (2013) studied the performance of sparse principal component analysis; Qiu et al. (2016) considered estimating time-varying graphical models.

- **Physical dependence**: For stationary causal processes in the form of $\{X_t = g(\{\boldsymbol{\epsilon}_j\}_{j\le t})\}_{t\in\mathbb{Z}}$, the physical dependence condition (Wu, 2005) assumes that the difference between $\boldsymbol{X}_t = g(\{\boldsymbol{\epsilon}_j\}_{j\le t})$ and $X_t' = g(\{\boldsymbol{\epsilon}_0', \boldsymbol{\epsilon}_j : j \le t, j \ne 0\})$ decays to 0 as t goes to infinity. Here $\{\boldsymbol{\epsilon}_0', \boldsymbol{\epsilon}_j : j \in \mathbb{Z}\}$ is a sequence of independent and identically distributed random vectors, and g is a measurable function[1]. The difference between X_t and X_t' quantifies the dependence of X_t on $\boldsymbol{\epsilon}_0$. Under this condition, Xiao and Wu (2012) derived rates of convergence for banding and thresholding estimators of the autocovariance matrix for stationary time series; Chen et al. (2013) studied the estimation of covariance and inverse covariance matrices for stationary and locally stationary time series.

Despite the wide applications of the structural dependence conditions, some inconvenience exists in that the dependence measure relies on a model while the "true" generating mechanism is usually unknown[2]. In contrast, non-structural dependence conditions rely on model-free dependence measures. For a time series $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$, these dependence measures quantify the degree of dependence between the "past", $\{\boldsymbol{X}_t\}_{t\le 0}$, and the "future", $\{\boldsymbol{X}_t\}_{t\ge n}$. Examples of non-structural dependence conditions include, among many, the mixing conditions and the weak dependence conditions, as illustrated below with a focus on the applications in high dimensions.

- **Mixing conditions**: The mixing conditions are built on various mixing coefficients, which quantify the dependence strength between the σ-fields generated by $\{\boldsymbol{X}_t\}_{t\le 0}$ and $\{\boldsymbol{X}_t\}_{t\ge n}$. The mixing conditions specify that the mixing coefficients decay to 0 as n goes to infinity. Assuming exponentially decaying α-mixing coefficients, Fan et al. (2012) studied the asymptotic behavior of the sample covariance matrix; Fan et al. (2011) and Fan et al. (2013) considered covariance matrix estimation under factor models with factors observed and unobserved, respectively. Based on these covariance matrix estimators, Fan et al. (2015) derived limiting distributions for portfolio risk estimators; Bai and Liao (2016, 2013) derived limiting distributions for the estimated factors and factor loadings. Besides the α-mixing conditions, Pan and Yao (2008) and Lam et al. (2011) exploited the ϕ- and ψ-mixing conditions in estimating factors and factor loadings. Under ϕ-mixing conditions, Han and Liu (2013a) studied principal component analysis, Qiu et al. (2015a) studied quantile-based scatter matrices, and Fan et al. (2016) studied risk inference. Recently, Han (2018), Shen et al. (2020), and Han and Li (2019), among many others, also established relevant exponential inequalities.

- **Weak dependence**: The weak dependence conditions rely on a dependence measure quantified by the covariance between smooth functions of $\{\boldsymbol{X}_t\}_{t\le 0}$ and $\{\boldsymbol{X}_t\}_{t\ge n}$, and require that the covariance goes to 0 as n goes to infinity (Doukhan and Louhichi, 1999). Under the weak

[1] $X_t = g(\{\boldsymbol{\epsilon}_j\}_{j\le t})$ is interpreted as a physical system with $\{\boldsymbol{\epsilon}_j\}_{j\le t}$ as the inputs and X_t as the output.

[2] We note that the data generating mechanisms themselves can be fairly general. For example, linear processes are special cases of stationary causal processes with $g(\{\boldsymbol{\epsilon}_j\}_{j\le t}) = \sum_{k=0}^{\infty} \boldsymbol{\Phi}_k \boldsymbol{\epsilon}_{t-k}$, where $\boldsymbol{\Phi}_0 = \mathbf{I}_d$ and $\boldsymbol{\Phi}_k \in \mathbb{R}^{d\times d}$. Wold's decomposition theorem (Wold, 1938) states that any process where the only deterministic term is the mean term can be represented as a linear process.

dependence conditions, Kallabis and Neumann (2006) and Doukhan and Neumann (2007) derived various probability and moment inequalities; Fan et al. (2012) studied the sample covariance matrix; Sancetta (2008) considered shrinkage estimators of covariance matrices.

The mixing conditions have been criticized for being difficult to verify (Doukhan and Louhichi, 1999 and more recently Han and Wu, 2019). The difficulty is mainly due to the σ-fields involved in the definitions of the mixing coefficients. In comparison, the weak dependence conditions are easier to verify in many scenarios. However, the covariance-based dependence measure only considers smooth transformations of the data. These conditions are not directly applicable to many other scenarios, such as the analysis of many quantile-based statistics, where non-smooth transformations are involved.

The aforementioned dependence conditions are based on distinct measures of dependence, and the commonality between them is unclear.

In this paper, we propose a new dependence measure named the Kolmogorov dependence measure. This dependence measure is naturally formulated using the Kolmogorov distance. Specifically, for two sequences of random variables, we quantify their dependence by the Kolmogorov distance (a.k.a. the Kolmogorov metric, referred to as the celebrated Kolmogorov–Smirnov statistic) between their joint distribution and the product of their marginal distributions. Using this dependence measure, we develop the Kolmogorov dependence condition for multivariate time series. We show that the Kolmogorov dependence condition unifies a class of VAR models, mixing conditions, physical dependence conditions, and covariance-based weak dependence conditions by serving as a common necessary condition.

The major challenge in building the connections between the Kolmogorov dependence condition and other conditions lies in the fundamental difference in the dependence measures. In particular, the Kolmogorov dependence measure is essentially the covariance between non-smooth transformations of the data, and hence standard techniques for analyzing smooth transformations are no longer applicable. To overcome this challenge, we devise a technique that approximates the discontinuous functions via smooth ones. These techniques enable the establishment of the Kolmogorov dependence condition under a wide variety of existing dependence conditions.

To further demonstrate the usefulness of the Kolmogorov dependence condition, we analyze a family of quantile-based scatter matrix estimators under high dimensional dependent data. It is shown that the Kolmogorov dependence measure facilitates an easy analysis of these estimators, and also captures the impact of serial dependence on estimation accuracy in a natural way.

Our contributions lie in three main aspects. First, we propose a novel dependence condition with a novel dependence measure. It unifies a number of widely used dependence conditions by serving as a common necessary condition. Secondly, under the Kolmogorov dependence condition, we derive the rates of convergence for a family of quantile-based scatter matrix estimators. This enables the analysis of quantile statistics under much weaker dependence assumptions than ϕ-mixing that was used in Qiu et al. (2015a), the only paper that gives comparable results. Lastly, we develop a set of techniques for analyzing time series using the Kolmogorov dependence condition. These techniques are of independent interest.

1.1 Organization

We organize the rest of this paper as follows. In Section 2, we propose the Kolmogorov dependence condition, and demonstrate its relations with other dependence conditions. In Section 3, we apply the Kolmogorov dependence condition to analyzing a family of quantile-based scatter matrix estimators. In Section 4, we provide a discussion of the main contributions of this paper. We gather the proofs of the theoretical results, additional technical results, and supporting lemmas in a supplement.

1.2 Notation

Let $\mathbf{v} = (v_1, \ldots, v_d)^\mathsf{T} \in \mathbb{R}^d$ be a d-dimensional vector, and $\mathbf{M} = [M_{jk}] \in \mathbb{R}^{d_1 \times d_2}$ be a $d_1 \times d_2$ matrix with M_{jk} as the (j,k)-th entry. For $0 < q < \infty$, we denote the ℓ_q norm of $\mathbf{v}$ as $\|\mathbf{v}\|_q := (\sum_{j=1}^d |v_j|^q)^{1/q}$ and the ℓ_∞ norm of $\mathbf{v}$ as $\|\mathbf{v}\|_\infty := \max_j |v_j|$. Let the matrix $\ell_{\max}$ norm of $\mathbf{M}$ be $\|\mathbf{M}\|_{\max} := \max_{jk} |M_{jk}|$, the matrix ℓ_∞ norm of $\mathbf{M}$ be $\|\mathbf{M}\|_\infty := \max_j \sum_{k=1}^d |M_{jk}|$, the Frobenius norm be $\|\mathbf{M}\|_F := \sqrt{\sum_{jk} M_{jk}^2}$, and the spectral norm be $\|\mathbf{M}\|_2 := \lambda_{\max}(\sqrt{\mathbf{M}^\mathsf{T}\mathbf{M}})$, i.e., the largest singular value of $\mathbf{M}$.

For a sequence of numbers $a_1, \ldots, a_d$, we denote $\operatorname{diag}(a_1, \ldots, a_d)$ to be a diagonal matrix with $a_1, \ldots, a_d$ on the diagonal. Similarly, for a sequence of matrices $\mathbf{A}_1, \ldots, \mathbf{A}_d$, we denote $\operatorname{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_d)$ to be a block diagonal matrix with $\mathbf{A}_1, \ldots, \mathbf{A}_d$ on the diagonal.

Denote $\mathbb{Z}$ to be the set of all integers, and $\mathbb{Z}^+$ to be the set of all positive integers. Let $\mathcal{S}, \mathcal{T} \subseteq \mathbb{Z}$ be two sets. We denote $|\mathcal{S}|$ as the cardinality of $\mathcal{S}$, and $d(\mathcal{S}, \mathcal{T}) := \inf\{|s - t| : s \in \mathcal{S}, t \in \mathcal{T}\}$ as the minimal distance between the elements in $\mathcal{S}$ and $\mathcal{T}$.

Throughout the paper, we use $C, C_1, C_2, \ldots$ to denote generic constants, though the actual values may vary at different occasions. We use $\mathbf{e}_j$ to denote the j-th column of the identity matrix, and define $\mathbb{S}_e := \{\mathbf{e}_j, \mathbf{e}_j + \mathbf{e}_k, \mathbf{e}_j - \mathbf{e}_k : j \neq k \in \{1, \ldots, d\}\}$.

For a sequence of random variables $\{X_n\}$ and a sequence of non-negative real numbers $\{a_n\}$, we write $X_n = O_P(a_n)$ if X_n is stochastically bounded by a_n. That is, for any $\epsilon > 0$, there exist finite numbers $M > 0$ and $N > 0$ such that $\mathbb{P}(|X_n| > M|a_n|) < \epsilon$ for all $n > N$.

2 KOLMOGOROV DEPENDENCE

The following is the definition of the Kolmogorov dependence measure.

Definition 2.1. Let $\{X_s\}_{s \in \mathcal{S}}$ and $\{Y_t\}_{t \in \mathcal{T}}$ be two sequences of random variables indexed by sets $\mathcal{S}, \mathcal{T} \subseteq \mathbb{Z}$. We define the Kolmogorov dependence measure between the two sequences by

$$\begin{aligned}
&\boldsymbol{\kappa}(\{X_s\}_{s \in \mathcal{S}}, \{Y_t\}_{t \in \mathcal{T}}) := \\
&\quad \sup_{u \in \mathbb{R}} \bigg| \mathbb{P}\Big(X_s \le u, Y_t \le u, \forall s \in \mathcal{S}, t \in \mathcal{T}\Big) - \\
&\quad \mathbb{P}\Big(X_s \le u, \forall s \in \mathcal{S}\Big) \mathbb{P}\Big(Y_t \le u, \forall t \in \mathcal{T}\Big) \bigg|.
\end{aligned}$$

If we define $F(u) := \mathbb{P}(X_s \le u, Y_t \le u, \forall s \in \mathcal{S}, t \in \mathcal{T})$ and $G(u) := \mathbb{P}(X_s \le u, \forall s \in \mathcal{S})\mathbb{P}(Y_t \le u, \forall t \in \mathcal{T})$, the Kolmogorov dependence measure between $\{X_s\}_{s \in \mathcal{S}}$ and $\{Y_t\}_{t \in \mathcal{T}}$ is the Kolmogorov distance between F and G: $\boldsymbol{\kappa}(\{X_s\}_{s \in \mathcal{S}}, \{Y_t\}_{t \in \mathcal{T}}) = \sup_{u \in \mathbb{R}} |F(u) - G(u)|$. It is hence also clear that Kolmogorov dependence measure, which corresponds to the Kolmogorov distance, is strictly weaker than the α-dependence measure, which will be introduced later and corresponds to the total variation distance.

Based on the Kolmogorov dependence measure, we next introduce the Kolmogorov dependence condition for stationary time series.

Definition 2.2 (Kolmogorov dependence condition)**.** A stationary univariate time series $\{X_t\}_{t\in\mathbb{Z}}$ is called (Ψ,ρ)-Kolmogorov dependent if there exist a function $\Psi:\mathbb{Z}^+\times\mathbb{Z}^+\to\mathbb{R}$ and a sequence $\{\rho(n)\}_{n\geq 0}$ decreasing to 0 at infinity such that for any finite non-empty sets $\mathcal{S},\mathcal{T}\subseteq\mathbb{Z}$ with $\max(\mathcal{S})\leq\min(\mathcal{T})$, we have

$$\boldsymbol{\kappa}(\{X_s\}_{s\in\mathcal{S}},\{X_t\}_{t\in\mathcal{T}})\leq\Psi(|\mathcal{S}|,|\mathcal{T}|)\rho\{d(\mathcal{S},\mathcal{T})\}.$$

A stationary multivariate time series $\{\boldsymbol{X}_t\in\mathbb{R}^d\}_{t\in\mathbb{Z}}$ is called $(\mathbb{S},\Psi,\rho)$-Kolmogorov dependent if there exits a set $\mathbb{S}\subseteq\mathbb{R}^d$ such that for any $\boldsymbol{a}\in\mathbb{S}$, the time series $\{\boldsymbol{a}^\top\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is (Ψ,ρ)-Kolmogorov dependent.

Kolmogorov dependence condition falls in the category of non-structural conditions. Here $d(\mathcal{S},\mathcal{T})$ represents the gap in time between the "past", $\{\boldsymbol{X}_t\}_{t\in\mathcal{S}}$, and the "future", $\{\boldsymbol{X}_t\}_{t\in\mathcal{T}}$. The sequence $\{\rho(n)\}_{n\geq 0}$ characterizes how fast the dependence strength between them, measured by $\boldsymbol{\kappa}$, decays as the time gap increases. We also note that $\mathbb{S}$ controls the stringency of the condition. If $\mathbb{S}_1\subseteq\mathbb{S}_2$, $(\mathbb{S}_2,\Psi,\rho)$-Kolmogorov dependence implies $(\mathbb{S}_1,\Psi,\rho)$-Kolmogorov dependence.

In the following theorems, we connect the Kolmogorov dependence condition to several dependence conditions frequently used in the literature, i.e., VAR models, mixing conditions, weak dependence, and physical dependence. In particular, we show that time series satisfying these dependence conditions are Kolmogorov dependent.

Theorem 2.1 (VAR model)**.** Let $\{\boldsymbol{X}_t\in\mathbb{R}^d\}_{t\in\mathbb{Z}}$ be a stationary time series satisfying the vector autoregressive model

$$\boldsymbol{X}_t=\mathbf{A}\boldsymbol{X}_{t-1}+\boldsymbol{\epsilon}_t,\ \text{ for any } t\in\mathbb{Z},$$

where $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random vectors. For $\mathbb{S}\subseteq\mathbb{R}^d$ such that $\sup_{\boldsymbol{x}\in\mathbb{S}}\|\boldsymbol{x}\|_2<\infty$, assume the following conditions hold:

1. $\|\mathbf{A}\|_2<1$.
2. $\mathbb{E}|\boldsymbol{a}^\top\mathbf{A}^\ell\boldsymbol{\epsilon}_1|\leq C\|\boldsymbol{a}\|_2\|\mathbf{A}\|_2^\ell$ for any $\boldsymbol{a}\in\mathbb{S}$, any $\ell\in\mathbb{Z}^+$, and some constant $C>0$.
3. For any $\boldsymbol{a}\in\mathbb{S}$, there exists a constant $H>0$ such that $\mathbb{P}(u\leq\boldsymbol{a}^T\boldsymbol{X}_t\leq u+v)\leq Hv$ for any $u\in\mathbb{R}$, $v>0$.

Then $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is $(\mathbb{S},\Psi,\rho)$-Kolmogorov dependent with

$$\Psi(u,v)\rho(n)=\Big\{4H+\frac{3C\sup_{\boldsymbol{x}\in\mathbb{S}}\|\boldsymbol{x}\|_2}{2(1-\|\mathbf{A}\|_2)}\Big\}(u+v)\|\mathbf{A}\|_2^{n/2}.$$

Remark 2.2. Condition 1 guarantees that $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is a stable process. Condition 3 is a smoothness condition on the distribution function. For Condition 2, when d is fixed, since $\mathbb{E}|\boldsymbol{a}^\top\mathbf{A}^\ell\boldsymbol{\epsilon}_1|\leq\|\boldsymbol{a}^\top\mathbf{A}^\ell\|_2\mathbb{E}\|\boldsymbol{\epsilon}_1\|_2\leq\|\boldsymbol{a}\|_2\|\mathbf{A}\|_2^\ell\mathbb{E}\|\boldsymbol{\epsilon}_1\|_2$, the condition is satisfied with $\mathbb{S}=\mathbb{R}^d$ provided that $\mathbb{E}\|\boldsymbol{\epsilon}_1\|_2<$

∞. When d may scale with the sample size, Condition 2 can be satisfied by assuming either Gaussian innovations, $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$, or certain sparsity structures on the transition matrix $\mathbf{A}$, depending on the structure of $\mathbb{S}$:

1. **Gaussian innovations**: Suppose that $\boldsymbol{\epsilon}_1 \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$ follows a Gaussian distribution with $\|\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\|_2 \le C$ for some constant C. By the properties of Gaussian distributions, we have
$$\boldsymbol{a}^\top \mathbf{A}^\ell \boldsymbol{\epsilon}_1 \sim N(0, \boldsymbol{a}^\top \mathbf{A}^\ell \boldsymbol{\Sigma}_\epsilon \mathbf{A}^{\ell^\top} \boldsymbol{a}).$$
Thus, we obtain
$$\begin{aligned}\mathbb{E}|\boldsymbol{a}^\top \mathbf{A}^\ell \boldsymbol{\epsilon}_1| &= \sqrt{\frac{2}{\pi}\boldsymbol{a}^\top \mathbf{A}^\ell \boldsymbol{\Sigma}_\epsilon \mathbf{A}^{\ell^\top} \boldsymbol{a}} \\ &\le \sqrt{\frac{2}{\pi}\|\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\|_2 \|\mathbf{A}^{\ell^\top}\boldsymbol{a}\|_2^2} \le \sqrt{\frac{2C}{\pi}}\|\boldsymbol{a}\|_2\|\mathbf{A}\|_2^\ell.\end{aligned}$$
Thus, Condition 2 is satisfied with $\mathbb{S} = \mathbb{R}^d$.

2. **Sparse transition matrix**: Consider $\mathbb{S} = \mathbb{S}_e$. In Section 3, we will show that $\mathbb{S}_e$ is what we need to obtain the rate of convergence for a class of scatter matrix estimators. Suppose that $\mathbf{A}$ is block diagonal: $\mathbf{A} = \operatorname{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_m)$, where $d_i := \dim(\mathbf{A}_i)$ is fixed for $i = 1, \ldots, m$ while m may scale with sample size. In other words, $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ consists of autoregressive blocks. For any $j \in \{1, \ldots, d\}$, let $i_0 = \min\{i : j \le d_1 + \cdots + d_i\}$ and $\boldsymbol{\epsilon}_1 = (\boldsymbol{\epsilon}_{11}, \ldots, \boldsymbol{\epsilon}_{1m})$ partitioned according to the dimensions of $(\mathbf{A}_1, \ldots, \mathbf{A}_m)$. We have $\mathbb{E}|\mathbf{e}_j^\top \mathbf{A}^\ell \boldsymbol{\epsilon}_1| \le \|\mathbf{A}_{i_0}\|_2^\ell \mathbb{E}\|\boldsymbol{\epsilon}_{1i_0}\|_2 \le \|\mathbf{A}\|_2^\ell \mathbb{E}\|\boldsymbol{\epsilon}_{1i_0}\|_2$. Thus, $\mathbb{E}|\boldsymbol{a}^\top \mathbf{A}^\ell \boldsymbol{\epsilon}_1| \le 2\|\mathbf{A}\|_2^\ell \mathbb{E}\|\boldsymbol{\epsilon}_{1i_0}\|_2$ for any $\boldsymbol{a} \in \mathbb{S}_e$. Therefore, Condition 2 is satisfied if $\mathbb{E}\|\boldsymbol{\epsilon}_{1i}\|_2 < \infty$ for $i = 1, \ldots, m$.

Next, we introduce the mixing conditions.

Definition 2.3 (Bradley (2005))**.** Let $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ be a stationary time series defined on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$. For $-\infty \le J \le L \le \infty$, define $\mathscr{F}_J^L := \sigma(\boldsymbol{X}_t : J \le t \le L, t \in \mathbb{Z})$ as the σ-field generated by $\{\boldsymbol{X}_t : J \le t \le L, t \in \mathbb{Z}\}$. Denote $\mathcal{L}^2(\mathscr{F}_j^L)$ to be the space of square-integrable, $\mathscr{F}_J^L$-measurable random variables. For any $n \ge 1$, define the following mixing coefficients:
$$\begin{aligned}
\boldsymbol{\alpha}(n) &:= \sup_{A\in\mathscr{F}_{-\infty}^0,\ B\in\mathscr{F}_n^\infty} \Big|\mathbb{P}(A\cap B) - P(A)P(B)\Big|;\\
\boldsymbol{\phi}(n) &:= \sup_{A\in\mathscr{F}_{-\infty}^0,\ B\in\mathscr{F}_n^\infty,\ \mathbb{P}(A)>0} \Big|\mathbb{P}(B\mid A) - \mathbb{P}(B)\Big|;\\
\boldsymbol{\psi}(n) &:= \sup_{A\in\mathscr{F}_{-\infty}^0,\ B\in\mathscr{F}_n^\infty,\ \mathbb{P}(A)>0,\ \mathbb{P}(B)>0} \Big|\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)\mathbb{P}(B)} - 1\Big|;\\
\boldsymbol{\varrho}(n) &:= \sup_{f\in\mathcal{L}(\mathscr{F}_{-\infty}^0),\ g\in\mathcal{L}(\mathscr{F}_n^\infty)} \Big|\operatorname{Corr}(f,g)\Big|;\\
\boldsymbol{\beta}(n) &:= \sup_{\substack{\{A_i\}_{i=1}^I \text{ partition } \Omega,\ A_i\in\mathscr{F}_{-\infty}^0\\ \{B_i\}_{i=1}^J \text{ partition } \Omega,\ B_i\in\mathscr{F}_n^\infty}} \\
&\qquad \frac{1}{2}\sum_{i=1}^I\sum_{j=1}^J \Big|\mathbb{P}(A_i\cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)\Big|.
\end{aligned}$$

The time series $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is α-, ϕ-, ψ-, ρ-, or β-mixing if and only if the corresponding mixing coefficient goes to 0 as n goes to infinity.

The mixing coefficients measure the dependence of two subsequences with time gap n. Analogous to the sequence $\rho(n)$ in the Kolmogorov dependence condition, the rate at which they converge to 0 characterizes the degree of dependence over the time series. If any of these mixing coefficients is 0 for all n, the series $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is independent. These mixing coefficients satisfy the following inequalities (Bradley, 2005):

$$\begin{cases} 2\boldsymbol{\alpha}(n) \le \boldsymbol{\beta}(n) \le \boldsymbol{\phi}(n) \le 1/2\boldsymbol{\psi}(n) \\ 4\boldsymbol{\alpha}(n) \le \boldsymbol{\varrho}(n) \le \boldsymbol{\psi}(n) \end{cases} \tag{2.1}$$

for any $n \in \mathbb{Z}^+$.

The connection between the Kolmogorov dependence measure and the mixing coefficients are straightforward. For any $\boldsymbol{a} \in \mathbb{R}^d$ and finite sets $\mathcal{S}, \mathcal{T} \subseteq \mathbb{Z}$ with $\max(\mathcal{S}) < \min(\mathcal{T})$, we have

$$\{\boldsymbol{a}^\top \boldsymbol{X}_t \le u : t \in \mathcal{S}\} \in \sigma\{\boldsymbol{X}_t : t \le \max(\mathcal{S})\},$$
$$\{\boldsymbol{a}^\top \boldsymbol{X}_t \le u : t \in \mathcal{T}\} \in \sigma\{\boldsymbol{X}_t : t \le \min(\mathcal{T})\}$$

for any $u \in \mathbb{R}$. Therefore, we have $\boldsymbol{\kappa}(\{\boldsymbol{a}^\top\boldsymbol{X}_t\}_{t\in\mathcal{S}}, \{\boldsymbol{a}^\top\boldsymbol{X}_t\}_{t\in\mathcal{T}}) \le \boldsymbol{\alpha}\{\min(\mathcal{T}) - \max(\mathcal{S})\}$. Using (2.1), we immediately have the following theorem.

Theorem 2.3 (Mixing)**.** α-, ϕ-, ψ-, ρ-, and β–mixing time series are $(\mathbb{R}^d, \Psi, \boldsymbol{\alpha})$-, $(\mathbb{R}^d, \Psi, \boldsymbol{\phi})$-, $(\mathbb{R}^d, \Psi, \boldsymbol{\psi})$-, $(\mathbb{R}^d, \Psi, \boldsymbol{\varrho})$-, $(\mathbb{R}^d, \Psi, \boldsymbol{\beta})$-Kolmogorov dependent ,respectively, for any function $\Psi \ge 1$.

Compared with the mixing conditions, the Kolmogorov dependence condition is easier to verify. For example, establishing the relation between VAR models and the α-mixing condition has proven to be difficult, mainly due to the σ-fields involved in the definition of the mixing coefficient (Chanda, 1974; Gorodetskii, 1978; Andrews, 1984; Pham and Tran, 1985). In comparison, the proof of Theorem 2.1 is natural and concise (Section A.1).

Next, we introduce the weak dependence measure of Doukhan and Louhichi (1999). For a function $g : (\mathbb{R}^d)^u \to \mathbb{R}$, define

$$\begin{aligned} \mathrm{Lip}(g) := \sup\Big\{ &\frac{|g(\mathbf{x}_1, \ldots, \mathbf{x}_u) - g(\mathbf{y}_1, \ldots, \mathbf{y}_u)|}{\|\mathbf{x}_1 - \mathbf{y}_1\| + \cdots + \|\mathbf{x}_u - \mathbf{y}_u\|} : \\ &(\mathbf{x}_1, \ldots, \mathbf{x}_u) \ne (\mathbf{y}_1, \ldots, \mathbf{y}_u)\Big\}, \end{aligned} \tag{2.2}$$

where $\|\cdot\|$ is a norm on $\mathbb{R}^d$. Denote $\Lambda := \{g : (\mathbb{R}^u)^d \to \mathbb{R} \text{ for some } u : \mathrm{Lip}(g) < \infty\}$ and $\Lambda^{(1)} := \{g \in \Lambda : \|g\|_\infty \le 1\}$, where $\|g\|_\infty := \sup_{\mathbf{x}} g(\mathbf{x})$.

Definition 2.4 (Doukhan and Louhichi (1999); Doukhan and Neumann (2007))**.** A stationary time series $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is $(\Lambda^{(1)}, \psi, \zeta)$-weakly dependent if and only if there exist a function $\psi : \mathbb{R}_+^2 \times \mathbb{N}^2 \to \mathbb{R}_+$ and a sequence $\{\zeta(n)\}_{n\ge 0}$ decreasing to 0 as n goes to infinity, such that for any $g_1, g_2 \in \Lambda^{(1)}$ with $g_1 : (\mathbb{R}^d)^u \to \mathbb{R}$, $g_2 : (\mathbb{R}^d)^v \to \mathbb{R}$, $u, v \in \mathbb{Z}^+$, and any u-tuple $(s_1, \ldots, s_u)$, v-tuple $(t_1, \ldots, t_v)$ with $s_1 \le \cdots \le s_u < t_1 \le \cdots \le t_v$, the following inequality is satisfied:

$$\begin{aligned} &\Big|\mathrm{Cov}\Big\{g_1(\boldsymbol{X}_{s_1}, \ldots, \boldsymbol{X}_{s_u}), g_2(\boldsymbol{X}_{t_1}, \ldots, \boldsymbol{X}_{t_v})\Big\}\Big| \\ &\le \psi(\mathrm{Lip}(g_1), \mathrm{Lip}(g_2), u, v)\zeta(t_1 - s_u). \end{aligned}$$

Important examples of $(\Lambda^{(1)}, \psi, \zeta)$-weakly dependent processes include θ-, η-, κ-, and λ-dependence. They correspond to specific choices of function ψ, as listed in Table 1.

Similar to the sequence $\rho(n)$ in Kolmogorov dependence condition, the sequence $\zeta(n)$ describes the degree of dependence over the $(\Lambda^{(1)}, \psi, \zeta)$-weakly dependent time series. The next theorem relates the weak dependence condition to Kolmogorov dependence.

Table 1: Important Examples of Weak Dependence (Definition 6 in Doukhan and Neumann, 2007).

θ-dependence:	$\psi(x, y, u, v) = vy$
η-dependence:	$\psi(x, y, u, v) = ux + vy$
κ-dependence:	$\psi(x, y, u, v) = uvxy$
λ-dependence:	$\psi(x, y, u, v) = ux + vy + uvxy$

Theorem 2.4 (Weak dependence)**.** Let $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ be a $(\Lambda^{(1)}, \psi, \zeta)$-weakly dependent time series, and $\|\cdot\|_*$ be the dual norm of $\|\cdot\|$ in (2.2). For $\mathbb{S} \subseteq \mathbb{R}^d$ such that $\sup_{\boldsymbol{x}\in\mathbb{S}} \|\boldsymbol{x}\|_* < \infty$, assume that for any $\boldsymbol{a} \in \mathbb{S}$, there exists a constant $H > 0$ such that $\mathbb{P}(u \leq \boldsymbol{a}^\top \boldsymbol{X}_t \leq u + v) \leq Hv$ for any $u \in \mathbb{R}$ and $v > 0$. Then the following statements hold:

1. If $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is θ- or η-dependent, then $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is $(\mathbb{S}, \Psi, \rho)$-Kolmogorov dependent with
$$\Psi(u, v)\rho(n) = \Big(4H + \frac{3}{4}\sup_{\boldsymbol{x}\in\mathbb{S}} \|\boldsymbol{x}\|_*\Big)(u + v)\sqrt{\zeta(n)}.$$
2. If $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is κ-dependent, then $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is $(\mathbb{S}, \Psi, \rho)$-Kolmogorov dependent with
$$\Psi(u, v)\rho(n) = \Big\{4H(u+v) + \frac{9}{16}\sup_{\boldsymbol{x}\in\mathbb{S}} \|\boldsymbol{x}\|_*^2 uv\Big\}\zeta(n)^{1/3}.$$
3. If $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is λ-dependent, then $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is $(\mathbb{S}, \Psi, \rho)$-Kolmogorov dependent with
$$\Psi(u, v)\rho(n) = \Big\{\Big(4H + \frac{3}{4}\sup_x \zeta(x)\sup_{\boldsymbol{x}\in\mathbb{S}} \|\boldsymbol{x}\|_*\Big)(u+v) + \frac{9}{16}\sup_{\boldsymbol{x}\in\mathbb{S}} \|\boldsymbol{x}\|_*^2 uv\Big\}\zeta(n)^{1/3}.$$

A special case of the mixing conditions and the weak dependence condition is m-dependence.

Definition 2.5. A stationary time series $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is m-dependent if and only if for any $t \in \mathbb{Z}$, $\{\boldsymbol{X}_s : s \leq t\}$ and $\{\boldsymbol{X}_s : s > t + m\}$ are independent.

If the time series $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is m-dependent, it is mixing with the mixing coefficients being 0 whenever $n > m$, and is $(\Lambda^{(1)}, \psi, \zeta)$-weakly dependent with $\zeta(n) = 0$ whenever $n > m$. In particular, we have the following corollary.

Corollary 1 (m-dependence)**.** An m-dependent time series $\{\boldsymbol{X}_t \in \mathbb{R}^d\}_{t\in\mathbb{Z}}$ is $(\mathbb{R}^d, \Psi, \rho)$-Kolmogorov dependent with $\rho(n) = I(n \leq m)$ and $\Psi = \sup_{\mathcal{S},\mathcal{T}\subseteq\mathbb{Z},\boldsymbol{a}\in\mathbb{R}^d} \boldsymbol{\kappa}(\{\boldsymbol{a}^\top\boldsymbol{X}_s\}_{s\in\mathcal{S}}, \{\boldsymbol{a}^\top\boldsymbol{X}_t\}_{t\in\mathcal{T}})$.

Lastly, we introduce the physical dependence measure proposed in Wu (2005)[3].

[3] We slightly generalize the definition in Wu (2005) to accommodate multivariate time series.

Definition 2.6 (Wu (2005))**.** Let $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$ be i.i.d. random vectors, and $\{\boldsymbol{\epsilon}'_t\}_{t\in\mathbb{Z}}$ be an i.i.d. copy of $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$. For a set $I \subseteq \mathbb{Z}$, let $\boldsymbol{\epsilon}_{t,I} := \boldsymbol{\epsilon}'_t$ if $t \in I$ and $\boldsymbol{\epsilon}_{t,I} := \boldsymbol{\epsilon}_t$ if $t \notin I$. Let $\mathcal{F}_t := \{\ldots, \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_t\}$ be a shift process, and $\mathcal{F}_{t,I} := \{\ldots, \boldsymbol{\epsilon}_{t-1,I}, \boldsymbol{\epsilon}_{t,I}\}$ be a coupled version of $\mathcal{F}_t$, where $\boldsymbol{\epsilon}_t$ is replaced by $\boldsymbol{\epsilon}'_t$ if $t \in I$. Let $\boldsymbol{g}$ be a measurable function to $\mathbb{R}^d$. We define the physical dependence measure to be

$$\delta_p(I, t, \boldsymbol{g}) := \|\boldsymbol{g}(\mathcal{F}_t) - \boldsymbol{g}(\mathcal{F}_{t,I})\|_p,$$

where $p \geq 1$ is a constant, and $\|\boldsymbol{g}(\mathcal{F}_t) - \boldsymbol{g}(\mathcal{F}_{t,I})\|_p := \{\mathbb{E}\|\boldsymbol{g}(\mathcal{F}_t) - \boldsymbol{g}(\mathcal{F}_{t,I})\|^p\}^{1/p}$ for a norm $\|\cdot\|$ on $\mathbb{R}^d$.

The process $\{\boldsymbol{X}_t = \boldsymbol{g}(\mathcal{F}_t)\}_{t\in\mathbb{Z}}$ is stationary, and is causal or non-anticipative in the sense that $\boldsymbol{X}_t$ does not depend on future innovations $\{\boldsymbol{\epsilon}_s : s > t\}$. $\mathcal{F}_t$ and $\boldsymbol{X}_t$ can be regarded as the inputs and output of a physical system $\boldsymbol{g}$. $\delta_p(I, t, \boldsymbol{g})$ quantifies the dependence of $\boldsymbol{g}(\mathcal{F}_t)$ on $\{\boldsymbol{\epsilon}_t : t \in I\}$. The next theorem connects physical dependence measure to Kolmogorov dependence.

Theorem 2.5 (Physical dependence)**.** Following the notations in Definition 2.6, let $\boldsymbol{X}_t = \boldsymbol{g}(\mathcal{F}_t)$, $\|\cdot\|_*$ be the dual norm of $\|\cdot\|$ on $\mathbb{R}^d$, and $\mathbb{S} \subseteq \mathbb{R}^d$ such that $\sup_{\boldsymbol{x}\in\mathbb{S}} \|\boldsymbol{x}\|_* < \infty$. Assume the following conditions hold:

1. For any $\boldsymbol{a} \in \mathbb{S}$, there exists a constant $H > 0$ such that $\mathbb{P}(u \leq \boldsymbol{a}^\top \boldsymbol{X}_t \leq u + v) \leq Hv$ holds for any $u \in \mathbb{R}$ and $v > 0$.
2. $\lim_{t\to\infty} \delta_p(I, t, \boldsymbol{g}) = 0$ for $I = \{0, -1, -2, \ldots\}$.

Then $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is $(\mathbb{S}, \Psi, \rho)$-Kolmogorov dependent with

$$\Psi(u, v)\rho(n) = \Big(4H + \frac{3}{2}\sup_{\boldsymbol{x}\in\mathbb{S}} \|\boldsymbol{x}\|_*\Big)(u + v)\sqrt{\delta_p(I, n, \boldsymbol{g})}.$$

For brevity, we denote $\delta_p(t, \boldsymbol{g}) := \delta_p(\{0\}, t, \boldsymbol{g})$. Although Theorem 2.5 only considers $I = \{0, -1, -2, \ldots\}$, the following theorem relates $\delta_p(I, t, \boldsymbol{g})$ to $\delta_p(t, \boldsymbol{g})$.

Theorem 2.6 (Wu (2005))**.** Let $p > 1$, $C_p := 18p^{3/2}(p-1)^{-1/2}$ if $1 < p < 2$, $C_p := \sqrt{2p}$ if $p \geq 2$. We have

$$\delta_p^{p'}(I, n, \boldsymbol{g}) \leq 2^{p'} C_p^{p'} \sum_{i\in I} \delta_p(n - i, \boldsymbol{g})^{p'},$$

where $p' := \min(p, 2)$ and $I \subseteq \mathbb{Z}$.

Theorem 2.6 shows that the physical dependence of $\boldsymbol{g}(\mathcal{F}_t)$ on $\{\boldsymbol{\epsilon}_t : t \in I\}$ can be upper bounded by the physical dependence of $\boldsymbol{g}(\mathcal{F}_t)$ on each $\boldsymbol{\epsilon}_t$ for $t \in I$ individually. In fact, a number of physical dependence conditions in the literature are given under $\delta_p(t, \boldsymbol{g})$ alone. For example, Xiao and Wu (2012) and Chen et al. (2013) considered the short-range dependence condition: $\sum_{n=m}^{\infty} \delta_p(m, \boldsymbol{g}) < \infty$ with ℓ_∞ norm. According to Theorem 2.6, Condition 2 in Theorem 2.5 is satisfied by the short-range dependence condition.

Although VAR models, Doukhan's weak dependence, and physical dependence appear to be of dramatically different forms, the derivations of their relations to Kolmogorov dependence are based on the same technique. They commonly rely on a smooth approximation of the indicator function, which connects the Kolmogorove dependence measure with the covariance between the "past" and the "future". We refer to Section A in the supplementary material for the detailed proof.

3 QUANTILE-BASED SCATTER MATRICES

In this section, we demonstrate the usefulness of the Kolmogorov dependence theory by deriving the rate of convergence for a family of quantile-based scatter matrices under high dimensional time series.

Let $Z \in \mathbb{R}$ be a random variable and $q \in [0,1]$ be a constant. We define the q-quantile of Z as

$$Q(Z;q) := \inf\{z : \mathbb{P}(Z \le z) \ge q\}.$$

$Q(Z;q)$ is unique if there exists a unique z such that $\mathbb{P}(Z \le z) = q$. If F is the distribution function of Z, we also use $Q(F;q)$ exchangeably with $Q(Z;q)$. Correspondingly, we define the empirical q-quantile of a sample, $\{z_t\}_{t=1}^T$, as

$$\widehat{Q}(\{z_t\}_{t=1}^T;q) := z^{(k)}, \text{ where } k = \min\Big\{t : \frac{t}{T} \ge q\Big\}. \tag{3.1}$$

Here $z^{(1)} \le z^{(2)} \le \cdots \le z^{(T)}$ are the order statistics of $z_1, \ldots, z_T$. Built on quantiles, the median absolute deviation (MAD) (Hampel, 1974) provides a robust measure of scale. The population and sample MADs are defined as[4]

$$\sigma^{\mathrm{M}}(Z;q) := Q\Big(\Big\{\Big|Z - Q\Big(Z;\frac{1}{2}\Big)\Big|\Big\};q\Big),$$
$$\widehat{\sigma}^{\mathrm{M}}(\{z_t\}_{t=1}^T;q) := \widehat{Q}\Big(\Big\{\Big|z_t - \widehat{Q}\Big(\{z_s\}_{s=1}^T;\frac{1}{2}\Big)\Big|\Big\}_{t=1}^T;q\Big).$$

In the rest of the paper, we suppress the parameter q and write $\sigma^{\mathrm{M}}(Z)$ and $\widehat{\sigma}^{\mathrm{M}}(\{z_t\}_{t=1}^T)$ for notational brevity. Let $\{\boldsymbol{X}_t\}_{t=1}^T$ be a stationary sequence of random vectors, where $\boldsymbol{X}_t = (X_{t1}, \ldots, X_{td})^\mathsf{T}$. As a generalization of MAD to the multivariate scenario, the population and sample MAD scatter matrices can be defined as $\mathbf{R} := [\mathbf{R}_{jk}]$ and $\widehat{\mathbf{R}} := [\widehat{\mathbf{R}}_{jk}]$, where the entries of $\mathbf{R}$ and $\widehat{\mathbf{R}}$ are given by

$$\begin{aligned}
\mathbf{R}_{jj} &= \sigma^{\mathrm{M}}(X_{1j})^2, \quad \widehat{\mathbf{R}}_{jj} = \widehat{\sigma}^{\mathrm{M}}(\{X_{tj}\}_{t=1}^T)^2,\\
\mathbf{R}_{jk} &= \frac{1}{4}\Big[\sigma^{\mathrm{M}}(X_{1j}+X_{1k})^2 - \sigma^{\mathrm{M}}(X_{1j}-X_{1k})^2\Big] \text{ for } j \ne k,\\
\widehat{\mathbf{R}}_{jk} &= \frac{1}{4}\Big[\widehat{\sigma}^{\mathrm{M}}(\{X_{tj}+X_{tk}\}_{t=1}^T)^2 - \widehat{\sigma}^{\mathrm{M}}(\{X_{tj}-X_{tk}\}_{t=1}^T)^2\Big]\\
&\quad \text{for } j \ne k.
\end{aligned}$$

In Han et al. (2014), $\mathbf{R}$ and $\widehat{\mathbf{R}}$ have been studied under independent data for elliptical distributions and beyond, which have been deeply studied in Liu et al. (2012), Han and Liu (2012), Han and Liu (2013b), Han and Liu (2014), Zhou et al. (2019), Han and Liu (2017), Han and Liu (2018), among many others. However, their properties under dependent data is unknown. Now, we investigate the consistency of $\widehat{\mathbf{R}}$ under Kolmogorov dependence. We begin by introducing two conditions on the sample $\{\boldsymbol{X}_t\}_{t=1}^T$.

Condition 3.1. $\{\boldsymbol{X}_t\}_{t=1}^T$ is a sample from a $(\mathbb{S}, \Psi, \rho)$-Kolmogorov dependent time series $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ such that

1. $\mathbb{S} = \mathbb{S}_e := \{\mathbf{e}_j, \mathbf{e}_j + \mathbf{e}_k, \mathbf{e}_j - \mathbf{e}_k : j \ne k \in \{1, \ldots, d\}\}$;

[4] In Hampel (1974), q was set to 1/2 to achieve the best possible 50% breakdown point (i.e., the maximum proportion of outliers that the estimate can safely tolerate) and the most sharply bounded influence function (Hampel et al., 1986).

2. Ψ is one of the following functions:

   (a) $\Psi(u,v) = 2v$,

   (b) $\Psi(u,v) = \beta(u+v) + (1-\beta)uv$, for some $\beta \in [0,1]$;

3. $\{\rho(n)\}_{n\geq 0}$ satisfies

$$\sum_{n=0}^{\infty}(n+1)^k\rho(n) \leq L_1 L^k (k!)^a, \text{ for any } k \in \mathbb{Z}^+, \tag{3.2}$$

   where $L_1 > 0$ and $a \geq 0$ are constants and L may scale with (T,d) such that

$$L = L(T,d) \leq \sqrt{\frac{L_1 T}{2^{7a+11}(\log d)^{2a+3}}}. \tag{3.3}$$

Equation (3.2) specifies the desired rate of decay in $\rho(n)$. The upper bound in (3.2) is adaptive to the sample size T and dimension d, in the sense that L is allowed to scale with (T,d) by the rate $\sqrt{T}/(\log d)^{a+3/2}$. Intuitively, larger sample size provides more information, which in turn allows for stronger dependence among the sample. On the other hand, larger dimension of the data entails weaker dependence. Overall, d is allowed to scale in the rate $\exp\{T^{1/(2a+3)}\}$ without collapsing L to 0.

The next condition is about the identifiability of the quantiles.

Condition 3.2. $\boldsymbol{X}_1$ is absolutely continuous and for any $\boldsymbol{a} \in \mathbb{S}_e$, we have

$$\begin{cases} \inf\limits_{|x-Q(F_{\boldsymbol{a}};q)|<\kappa} \dfrac{d}{dx}F_{\boldsymbol{a}}(x) \geq \eta, \\ \inf\limits_{|x-Q(\bar{F}_{\boldsymbol{a}};q)|<\kappa} \dfrac{d}{dx}\bar{F}_{\boldsymbol{a}}(x) \geq \eta \end{cases} \tag{3.4}$$

for some constants $\kappa, \eta > 0$, where $F_{\boldsymbol{a}}$ and $\bar{F}_{\boldsymbol{a}}$ are the distribution functions of $\boldsymbol{a}^\top\boldsymbol{X}_1$ and $|\boldsymbol{a}^\top\boldsymbol{X}_1 - Q(\boldsymbol{a}^\top\boldsymbol{X}_1;q)|$, respectively.

Condition 3.2 guarantees the identifiability of the quantiles of the distribution functions. This condition is standard in the literature on quantile statistics (Han et al., 2014; Belloni and Chernozhukov, 2011; Wang et al., 2012). With Conditions 3.1 and 3.2, we can obtain the rate of convergence for $\widehat{\mathbf{R}}$.

Theorem 3.3. Let $\{\boldsymbol{X}_t\}_{t=1}^T$ be a sample from a stationary Kolmogorov dependence time series, $\{\boldsymbol{X}_t \in \mathbb{R}^d\}_{t\in\mathbb{Z}}$, such that Conditions 3.1 and 3.2 hold. Suppose $(\log d)^{2a+3}/T \to 0$ as (T,d) go to infinity. Then, for (T,d) large enough and any $\alpha \in (0,1)$, with probability no smaller than $1-24\alpha^2$, we have

$$\begin{aligned} &\|\widehat{\mathbf{R}} - \mathbf{R}\|_{\max} \leq \\ &\quad \max\Big\{\frac{8}{\eta^2}\Big\{16\sqrt{\frac{2^a L_1(\log d + \log(1/\alpha))}{T}} + \frac{1}{T}\Big\}^2, \\ &\quad \frac{8\sigma^{\mathrm{M}}_{\max}}{\eta}\Big\{16\sqrt{\frac{2^a L_1(\log d + \log(1/\alpha))}{T}} + \frac{1}{T}\Big\}\Big\}, \end{aligned} \tag{3.5}$$

where $\sigma^{\mathrm{M}}_{\max} := \max\{\sigma^{\mathrm{M}}(\boldsymbol{a}^\top\boldsymbol{X}_t) : \boldsymbol{a} \in \mathbb{S}_e\}$, a and L_1 are defined in (3.2), and η is defined in (3.4).

The implications of Theorem 3.3 are as follows:

1. In Condition 3.1, the decaying rate of $\rho(n)$ is controlled by a and L_1, which capture the strength of serial dependence. Theorem 3.3 quantifies that these parameters affect the accuracy of the estimator $\widehat{\mathbf{R}}$ by modifying the upper bound of its estimation error.

2. When a, L_1, η, and $\sigma^{\mathrm{M}}_{\max}$ are fixed, the rate of convergence for $\widehat{\mathbf{R}}$ reduces to $O_P(\sqrt{\log d/T})$. Han et al. (2014) derived similar rates of convergence for $\widehat{\mathbf{R}}$ under independent data, and showed that the rate leads to optimal rates of convergence for various covariance estimators induced from $\widehat{\mathbf{R}}$.

Theorem 3.3 establishes the rate of convergence for $\widehat{\mathbf{R}}$ using the language of the Kolmogorov dependence. The proof of the theorem relies on the fact that both the Kolmogorov dependence measure and the quantile function are closely related to non-smooth transformations of the data. We refer to Lemma C.1 for technical details.

As is established by Theorems 2.1, 2.3, 2.4, and 2.5, the Kolmogorov dependence condition is a necessary condition for VAR models, mixing conditions, various weak dependence conditions, and physical dependence conditions. Therefore, Theorem 3.3 immediately implies rates of convergence for $\widehat{\mathbf{R}}$ under these other dependence conditions. Below we present the results for VAR models and α-mixing conditions as two examples.

Corollary 2 (VAR model)**.** Let $\{\boldsymbol{X}_t \in \mathbb{R}^d\}_{t=1}^T$ be a sample from a VAR process satisfying the conditions in Theorem 2.1 with $\mathbb{S} = \mathbb{S}_e$. Assume that Condition 3.2 holds and $(\log d)^{2a+3}/T \to 0$ as (T, d) go to infinity. Then we have

$$\|\widehat{\mathbf{R}} - \mathbf{R}\|_{\max} = O_P\Big[\frac{\sigma^{\mathrm{M}}_{\max}}{\eta}\Big\{\sqrt{\Big(4H + \frac{3C}{1-\|\mathbf{A}\|_2}\Big)\frac{1}{1-\sqrt{\|\mathbf{A}\|_2}}\frac{\log d}{T}} + \frac{1}{T}\Big\}\Big].$$

Here $\mathbf{A}$ is the transition matrix of the VAR process, $\sigma^{\mathrm{M}}_{\max} = \max\{\sigma^{\mathrm{M}}(\boldsymbol{a}^\top \boldsymbol{X}_t) : \boldsymbol{a} \in \mathbb{S}_e\}$, η is defined in Condition 3.2, and H, C are constants defined in Theorem 2.1.

Corollary 2 shows that for VAR models, the effect of serial dependence on the consistency of $\widehat{\mathbf{R}}$ essentially reduces to the spectral norm of the transition matrix. Similar findings have been noted for Pearson covariance matrix estimation in a number of literature (Loh and Wainwright, 2012; Han and Liu, 2013c; Qiu et al., 2016).

Corollary 3 (α-mixing)**.** Let $\{\boldsymbol{X}_t \in \mathbb{R}^d\}_{t=1}^T$ be a sample from an α-mixing time series satisfying

$$\boldsymbol{\alpha}(n) \le C_1 \exp\Big(-C_2 n^r\Big), \tag{3.6}$$

where C_1, C_2 and r are positive constants. Assume that Condition 3.2 holds. Then we have

$$\|\widehat{\mathbf{R}}-\mathbf{R}\|_{\max} = O_P\Big\{\frac{\sigma^{\mathrm{M}}_{\max}}{\eta}\Big(\sqrt{\frac{C_1}{1-e^{-C_2}}\frac{\log d}{T}}+\frac{1}{T}\Big)\Big\}, \text{ if } r\geq 1; \tag{3.7}$$

$$\|\widehat{\mathbf{R}}-\mathbf{R}\|_{\max} = O_P\Big(\frac{\sigma^{\mathrm{M}}_{\max}}{\eta}\Big[\sqrt{2^{1/r}C_1\Big\{1+\sqrt{\frac{2\pi e}{r}}\Big(\frac{1}{2C_2 r}\Big)^{1/r}\Big\}\frac{\log d}{T}}+\frac{1}{T}\Big]\Big), \text{ if } 0<r<1. \tag{3.8}$$

Here $\sigma^{\mathrm{M}}_{\max}=\max\{\sigma^{\mathrm{M}}(\boldsymbol{a}^{\top}\boldsymbol{X}_t):\boldsymbol{a}\in\mathbb{S}_e\}$ and η is defined in Condition 3.2.

Condition (3.6) has been widely exploited in modeling dependence in financial time series. See, for example, Fan et al. (2011), Fan et al. (2012), Fan et al. (2013), Fan et al. (2015), Bai and Liao (2016), and Bai and Liao (2013).

In addition to VAR models and mixing conditions, rates of convergence under the weak dependence conditions and the physical dependence conditions can be easily obtained using Theorems 2.4, 2.5, and 3.3. For conciseness, we omit further discussions herein. Prior to this work, comparable rates were only derived under ϕ-mixing conditions for quantile statistics under high dimensions (Qiu et al., 2015a).

4 DISCUSSION

In this section, we discuss the uniqueness and the generality of the Kolmogorov dependence condition. The Kolmogorov dependence condition is closely related to the Doukhan's weak dependence conditions (Doukhan and Louhichi, 1999; Kallabis and Neumann, 2006; Doukhan and Neumann, 2007, 2008) developed for concentration inequalities. However, these conditions are not directly applicable to analyzing quantile-based statistics, since they are not invariant to non-smooth transformations of the data. In comparison, the Kolmogorov dependence condition is readily adapted to the non-smooth structure of quantile statistics. The Kolmogorov dependence condition also resembles the α-mixing conditions (Dedecker and Prieur, 2004; Kontorovich et al., 2008; Merlevède et al., 2009, 2011) in terms of the dependence measures. The key difference is that the dependence measure in α-mixing requires bounding the total variation distance, which makes the α-mixing conditions difficult to verify. In comparison, the Kolmogorov dependence condition relaxes the requirement for σ-fields, and therefore can be relatively easily verified under many popular dependence conditions, including the α-mixing conditions themselves.

The Kolmogorov dependence condition provides us a fairly general understanding of dependence. It serves as a common necessary condition unifying a number of dependence conditions. Thus, the theoretical results obtained under the Kolmogorov dependence condition shed light on the properties of other dependence conditions as well. As an example, we established the rate of convergence for the MAD scatter matrix estimators under the Kolmogorov dependence condition, and demonstrated that the corresponding rates under other dependence conditions can be immediately obtained as well.

A PROOF OF MAIN RESULTS

In this section, we present the proofs of the main theorems.

Lemma A.1. Let $\{Y_t\}_{t\in Z}$ be a stationary univariate time series. Assume that there exists a constant $H > 0$ such that $\mathbb{P}(u \le Y_t \le u+v) \le Hv$ for any $u \in \mathbb{R}$ and $v > 0$. Given any $u \in \mathbb{R}$ and $\epsilon > 0$, let $h(x) := I(x \le u)$ and $h_\epsilon(x)$ be a smoothed version of h:

$$h_\epsilon(x) := \begin{cases} h(x), \text{ if } x < u-\epsilon \text{ or } x > u+\epsilon; \\ \dfrac{1}{4\epsilon^3}\{x^3 - 3ux^2 + 3(u^2-\epsilon^2)x - u^3 + 3u\epsilon^2 + \\ \quad 2\epsilon^3\}, \text{ if } u-\epsilon \le x \le u+\epsilon. \end{cases} \tag{A.1}$$

Then we have

$$\begin{aligned}&\Big|\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S}\cup\mathcal{T}\Big) - \\ &\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S}\Big)\mathbb{P}\Big(Y_{t'} \le u, \forall\, t' \in \mathcal{T}\Big)\Big| \\ &\le 4H(|\mathcal{S}|+|\mathcal{T}|)\epsilon + \Big|\mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big\}\Big|\end{aligned}$$

for any finite non-empty sets $\mathcal{S}, \mathcal{T} \subseteq \mathbb{Z}$.

Proof. $h_\epsilon(x)$ is continuous with first order derivative

$$\frac{d}{dx}h_\epsilon(x) = \begin{cases} 0, \text{ if } x < u-\epsilon \text{ or } x > u+\epsilon; \\ \dfrac{3}{4\epsilon^2}\{(x-u)^2 - \epsilon^2\}, \text{ if } u-\epsilon \le x \le u+\epsilon. \end{cases}$$

Thus, $h_\epsilon(x)$ is Lipschitz continuous with $\mathrm{Lip}(h_\epsilon) = \sup_x |dh_\epsilon(x)/dx| = 3/(4\epsilon)$. By the definition of covariance and the triangle inequality, we have

$$\begin{aligned}&\Big|\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S}\cup\mathcal{T}\Big) - \\ &\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S}\Big)\mathbb{P}\Big(Y_{t'} \le u, \forall\, t' \in \mathcal{T}\Big)\Big| \\ =&\Big|\mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h(Y_t), \prod_{t\in\mathcal{T}} h(Y_t)\Big\}\Big| \\ \le&\underbrace{\Big|\mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h(Y_t), \prod_{t\in\mathcal{T}} h(Y_t)\Big\} - \mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big\}\Big|}_{A} \\ &+ \Big|\mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big\}\Big|.\end{aligned} \tag{A.2}$$

We now derive the upper bound for A. By the triangle inequality, we have

$$\begin{aligned}A \le&\Big|\mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h(Y_t), \prod_{t\in\mathcal{T}} h(Y_t) - \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big\}\Big| + \\ &\Big|\mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h(Y_t) - \prod_{t\in\mathcal{S}} h_\epsilon(Y_t), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big\}\Big|.\end{aligned} \tag{A.3}$$

For two random variables X and Y with $|X| \leq 1$, we have

$$\begin{aligned} |\operatorname{Cov}(X,Y)| =& |\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y| \\ \leq& \mathbb{E}|X||Y| + \mathbb{E}|X|\mathbb{E}|Y| \leq 2\mathbb{E}|Y|. \end{aligned} \tag{A.4}$$

Now, setting $X = \prod_{t\in\mathcal{S}} h(Y_t)$ and $Y = \prod_{t\in\mathcal{T}} h(Y_t) - \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)$, we have

$$\begin{aligned} &\Big|\operatorname{Cov}\Big\{\prod_{t\in\mathcal{S}} h(Y_t), \prod_{t\in\mathcal{T}} h(Y_t) - \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big\}\Big| \\ &\leq 2\mathbb{E}\Big|\prod_{t\in\mathcal{T}} h(Y_t) - \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big|. \end{aligned}$$

Setting $X = \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)$ and $Y = \prod_{t\in\mathcal{S}} h(Y_t) - \prod_{t\in\mathcal{S}} h_\epsilon(Y_t)$ in (A.4), we have

$$\begin{aligned} &\Big|\operatorname{Cov}\Big\{\prod_{t\in\mathcal{S}} h(Y_t) - \prod_{t\in\mathcal{S}} h_\epsilon(Y_t), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big\}\Big| \\ &\leq 2\mathbb{E}\Big|\prod_{t\in\mathcal{S}} h(Y_t) - \prod_{t\in\mathcal{S}} h_\epsilon(Y_t)\Big|. \end{aligned}$$

Plugging the above two inequalities into (A.3), we have

$$\begin{aligned} A \leq& 2\mathbb{E}\Big|\prod_{t\in\mathcal{T}} h(Y_t) - \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big| + 2\mathbb{E}\Big|\prod_{t\in\mathcal{S}} h(Y_t) - \prod_{t\in\mathcal{S}} h_\epsilon(Y_t)\Big| \\ \leq& 2(|\mathcal{S}| + |\mathcal{T}|)\mathbb{E}\Big|h(Y_t) - h_\epsilon(Y_t)\Big|. \end{aligned}$$

The last inequality is due to the fact that

$$\Big|\prod_{t=1}^{m} a_t - \prod_{t=1}^{m} b_t\Big| \leq \sum_{t=1}^{m} |a_t - b_t| \tag{A.5}$$

for $0 \leq a_t, b_t \leq 1$. Noting that $|h(Y_t) - h_\epsilon(Y_t)| \leq 1$ and $h(Y_t) - h_\epsilon(Y_t)$ is non-zero only when $u - \epsilon \leq Y_t \leq u + \epsilon$, we have

$$\begin{aligned} A \leq& 2(|\mathcal{S}| + |\mathcal{T}|)\mathbb{P}(u - \epsilon \leq Y_t \leq u + \epsilon) \\ \leq& 4H(|\mathcal{S}| + |\mathcal{T}|)\epsilon. \end{aligned} \tag{A.6}$$

Plugging (A.6) into (A.2) completes the proof. □

A.1 Proof of Theorem 2.1

Proof. Applying Lemma A.1 with $Y_t = \boldsymbol{a}^\top \boldsymbol{X}_t$, we have

$$\begin{aligned} &\Big|\mathbb{P}\Big(Y_t \leq u, \forall\, t \in \mathcal{S} \cup \mathcal{T}\Big) - \\ &\mathbb{P}\Big(Y_t \leq u, \forall\, t \in \mathcal{S}\Big)\mathbb{P}\Big(Y_{t'} \leq u, \forall\, t' \in \mathcal{T}\Big)\Big| \\ &\leq 4H(|\mathcal{S}| + |\mathcal{T}|)\epsilon + \underbrace{\Big|\operatorname{Cov}\Big\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big\}\Big|}_{B}, \end{aligned} \tag{A.7}$$

where h_ϵ is defined in (A.1). Now we derive an upper bound of B. Since $\|\mathbf{A}\|_2 < 1$, the process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ has moving average representation $\boldsymbol{X}_t = \sum_{\ell=0}^{\infty} \mathbf{A}^\ell \boldsymbol{\epsilon}_{t-\ell}$. Define $\boldsymbol{X}_t^{[n]}$ to be a finite order moving average process: $\boldsymbol{X}_t^{[n]} := \sum_{\ell=0}^{n-1} \mathbf{A}^\ell \boldsymbol{\epsilon}_{t-\ell}$ where $n = d(\mathcal{S}, \mathcal{T})$. Define $Y_t^{[n]} := \boldsymbol{a}^\top \boldsymbol{X}_t^{[n]}$. Using $Y_t^{[n]}$, we can upper bound B in (A.2) by

$$
\begin{aligned}
B \leq & \underbrace{\Big|\mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t) - \prod_{t\in\mathcal{S}} h_\epsilon(Y_t^{[n]}), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big\}\Big|}_{B_1} + \\
& \underbrace{\Big|\mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t^{[n]}), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t) - \prod_{t\in\mathcal{T}} h_\epsilon(Y_t^{[n]})\Big\}\Big|}_{B_2} + \\
& \underbrace{\Big|\mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t^{[n]}), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t^{[n]})\Big\}\Big|}_{B_3}.
\end{aligned}
$$

Note that $\{Y_t^{[n]} : t \in \mathcal{S}\}$ only depends on $\{\boldsymbol{\epsilon}_t : \min(\mathcal{S}) - n < t \leq \max(\mathcal{S})\}$ and $\{Y_t^{[n]} : t \in \mathcal{T}\}$ only depends on $\{\boldsymbol{\epsilon}_t : \min(\mathcal{T}) - n < t \leq \max(\mathcal{T})\}$. Since $n = d(\mathcal{S}, \mathcal{T}) = \min(\mathcal{T}) - \max(\mathcal{S})$, we have that $\prod_{t\in\mathcal{S}} h_\epsilon(Y_t^{[n]})$ and $\prod_{t\in\mathcal{T}} h_\epsilon(Y_t^{[n]})$ are independent. Thus, we have $B_3 = 0$. Regarding B_1, using (A.4) and (A.5), we have

$$
\begin{aligned}
B_1 \leq & 2\mathbb{E}\Big|\prod_{t\in\mathcal{S}} h_\epsilon(Y_t) - \prod_{t\in\mathcal{S}} h_\epsilon(Y_t^{[n]})\Big| \\
\leq & 2|\mathcal{S}|\mathbb{E}|h_\epsilon(Y_t) - h_\epsilon(Y_t^{[n]})| \\
\leq & 2|\mathcal{S}|\mathrm{Lip}(h_\epsilon)\mathbb{E}|Y_t - Y_t^{[n]}|.
\end{aligned} \tag{A.8}
$$

Plugging in $\mathrm{Lip}(h_\epsilon) = 3/(4\epsilon)$, $Y_t = \boldsymbol{a}^\top \boldsymbol{X}_t = \sum_{\ell=0}^{\infty} \boldsymbol{a}^\top \mathbf{A}^\ell \boldsymbol{\epsilon}_{t-\ell}$ and $Y_t^{[n]} = \boldsymbol{a}^\top \boldsymbol{X}_t^{[n]} = \sum_{\ell=0}^{n-1} \boldsymbol{a}^\top \mathbf{A}^\ell \boldsymbol{\epsilon}_{t-\ell}$, we obtain

$$
\begin{aligned}
B_1 \leq & \frac{3}{2\epsilon}|\mathcal{S}|\mathbb{E}\Big|\sum_{\ell=n}^{\infty} \boldsymbol{a}^\top \mathbf{A}^\ell \boldsymbol{\epsilon}_{t-\ell}\Big| \\
\leq & \frac{3}{2\epsilon}|\mathcal{S}|\sum_{\ell=n}^{\infty} \mathbb{E}\Big|\boldsymbol{a}^\top \mathbf{A}^\ell \boldsymbol{\epsilon}_{t-\ell}\Big| \leq \frac{3C|\mathcal{S}|\|\boldsymbol{a}\|_2\|\mathbf{A}\|_2^n}{2\epsilon(1 - \|\mathbf{A}\|_2)}.
\end{aligned}
$$

The last inequality is due to Conditions 1 and 2 on the VAR process. Applying similar arguments to B_2, we have $B_2 \leq 3C|\mathcal{T}|\|\boldsymbol{a}\|_2\|\mathbf{A}\|_2^n/\{2\epsilon(1 - \|\mathbf{A}\|_2)\}$. Thus, we have

$$
B \leq B_1 + B_2 \leq \frac{3C(|\mathcal{S}| + |\mathcal{T}|)\|\boldsymbol{a}\|_2\|\mathbf{A}\|_2^n}{2\epsilon(1 - \|\mathbf{A}\|_2)}. \tag{A.9}
$$

Plugging (A.9) into (A.7), we have

$$
\begin{aligned}
& \Big|\mathbb{P}\Big(Y_t \leq u, \forall\, t \in \mathcal{S} \cup \mathcal{T}\Big) - \\
& \mathbb{P}\Big(Y_t \leq u, \forall\, t \in \mathcal{S}\Big)\mathbb{P}\Big(Y_{t'} \leq u, \forall\, t' \in \mathcal{T}\Big)\Big| \\
& \leq (|\mathcal{S}| + |\mathcal{T}|)\Big\{4H\epsilon + \frac{3C\|\boldsymbol{a}\|_2\|\mathbf{A}\|_2^n}{2\epsilon(1 - \|\mathbf{A}\|_2)}\Big\}.
\end{aligned}
$$

Now setting $\epsilon = \|\mathbf{A}\|_2^{n/2}$ and taking supremum over $u \in \mathbb{R}$, we have

$$
\begin{aligned}
&\boldsymbol{\kappa}(\{Y_t\}_{t\in\mathcal{S}}, \{Y_t\}_{t\in\mathcal{T}}) \\
&\le \Big\{4H + \frac{3C\|\boldsymbol{a}\|_2}{2(1-\|\mathbf{A}\|_2)}\Big\}(|\mathcal{S}| + |\mathcal{T}|)\|\mathbf{A}\|_2^{n/2}.
\end{aligned}
$$

Since $\|\mathbf{A}\|_2 < 1$, $\{Y_t\}_{t\in\mathbb{Z}}$ is Kolmogorov dependent. This completes the proof. □

A.2 Proof of Theorem 2.4

Proof. Applying Lemma A.1 with $Y_t = \boldsymbol{a}^\top \boldsymbol{X}_t$, again we have

$$
\begin{aligned}
&\Big|\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S}\cup\mathcal{T}\Big) - \\
&\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S}\Big)\mathbb{P}\Big(Y_{t'} \le u, \forall\, t' \in \mathcal{T}\Big)\Big| \\
&\le 4H(|\mathcal{S}| + |\mathcal{T}|)\epsilon + \underbrace{\Big|\mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big\}\Big|}_{B},
\end{aligned} \tag{A.10}
$$

where h_ϵ be defined in (A.1). It remains to derive an upper bound for B. Since $\{\boldsymbol{X}_t\}_{t\in\boldsymbol{Z}}$ is $(\Lambda^{(1)}, \psi, \zeta)$-weakly dependent, we have

$$
\begin{aligned}
B =& \Big|\mathrm{Cov}\Big\{g(\{\boldsymbol{X}_t : t\in\mathcal{S}\}), g(\{\boldsymbol{X}_t : t\in\mathcal{T}\})\Big\}\Big| \\
\le& \psi(\mathrm{Lip}(g), \mathrm{Lip}(g), |\mathcal{S}|, |\mathcal{T}|)\zeta\{d(\mathcal{S}, \mathcal{T})\},
\end{aligned} \tag{A.11}
$$

where $g(\mathbf{x}_t : t\in\mathcal{S}) := \prod_{t\in\mathcal{S}} h_\epsilon(\boldsymbol{a}^\top \mathbf{x}_t)$. Since $\mathrm{Lip}(h_\epsilon) = 3/(4\epsilon)$, using (A.5), we have that for any $\{\mathbf{x}_t, \mathbf{y}_t : t\in\mathcal{S}\}$, we have

$$
\begin{aligned}
\Big|\prod_{t\in\mathcal{S}} h_\epsilon(\boldsymbol{a}^\top\mathbf{x}_t) - \prod_{t\in\mathcal{S}} h_\epsilon(\boldsymbol{a}^\top\mathbf{y}_t)\Big| \le& \frac{3}{4\epsilon}\sum_{t\in\mathcal{S}} |\boldsymbol{a}^\top(\mathbf{x}_t - \mathbf{y}_t)| \\
\le& \frac{3\|\boldsymbol{a}\|_*}{4\epsilon}\sum_{t\in\mathcal{S}} \|\mathbf{x}_t - \mathbf{y}_t\|,
\end{aligned}
$$

This implies that the function g is Lipschitz with $\mathrm{Lip}(g) \le 3\|\boldsymbol{a}\|_*/(4\epsilon)$. Now, replacing ψ in (A.11) by the various specifications in Table 1, and combining the resulting inequality with (A.10), we get

$$
\begin{aligned}
&\Big|\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S} \cup \mathcal{T}\Big) - \\
&\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S}\Big)\mathbb{P}\Big(Y_{t'} \le u, \forall\, t' \in \mathcal{T}\Big)\Big| \\
\le &\begin{cases}
4H(|\mathcal{S}|+|\mathcal{T}|)\epsilon + \dfrac{3\|\boldsymbol{a}\|_*}{4\epsilon}|\mathcal{T}|\zeta\{d(\mathcal{S},\mathcal{T})\} \\
\quad \text{if } \{\boldsymbol{X}_t\}_{t\in\mathbb{Z}} \text{ is } \theta\text{-dependent;} \\
(|\mathcal{S}|+|\mathcal{T}|)\Big\{4H\epsilon + \dfrac{3\|\boldsymbol{a}\|_*}{4\epsilon}\zeta\{d(\mathcal{S},\mathcal{T})\}\Big\} \\
\quad \text{if } \{\boldsymbol{X}_t\}_{t\in\mathbb{Z}} \text{ is } \eta\text{-dependent;} \\
4H(|\mathcal{S}|+|\mathcal{T}|)\epsilon + \dfrac{9\|\boldsymbol{a}\|_*^2}{16\epsilon^2}|\mathcal{S}||\mathcal{T}|\zeta\{d(\mathcal{S},\mathcal{T})\} \\
\quad \text{if } \{\boldsymbol{X}_t\}_{t\in\mathbb{Z}} \text{ is } \kappa\text{-dependent;} \\
4H(|\mathcal{S}|+|\mathcal{T}|)\epsilon + \Big\{\dfrac{3\|\boldsymbol{a}\|_*}{4\epsilon}(|\mathcal{S}|+|\mathcal{T}|) + \dfrac{9\|\boldsymbol{a}\|_*^2}{16\epsilon^2} \\
\quad |\mathcal{S}||\mathcal{T}|\Big\}\zeta\{d(\mathcal{S},\mathcal{T})\} \text{ if } \{\boldsymbol{X}_t\}_{t\in\mathbb{Z}} \text{ is } \lambda\text{-dependent.}
\end{cases}
\end{aligned}
$$

Therefore, if $\{\boldsymbol{X}_t\}_{t\in\boldsymbol{Z}}$ is θ- or η-dependent, setting $\epsilon = \sqrt{\zeta\{d(\mathcal{S},\mathcal{T})\}}$ gives the desired result. If $\{\boldsymbol{X}_t\}_{t\in\boldsymbol{Z}}$ is κ- or λ-dependent, setting $\epsilon = \zeta\{d(\mathcal{S},\mathcal{T})\}^{1/3}$ gives the desired result. ☐

A.3 Proof of Theorem 2.5

Proof. Applying Lemma A.1 with $Y_t = \boldsymbol{a}^\top \boldsymbol{X}_t$, we have

$$
\begin{aligned}
&\Big|\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S} \cup \mathcal{T}\Big) - \\
&\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S}\Big)\mathbb{P}\Big(Y_{t'} \le u, \forall\, t' \in \mathcal{T}\Big)\Big| \\
&\le 4H(|\mathcal{S}|+|\mathcal{T}|)\epsilon + \underbrace{\Big|\mathrm{Cov}\Big\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\Big\}\Big|}_{B},
\end{aligned}
\tag{A.12}
$$

To derive an upper bound on B, let $\{\boldsymbol{\epsilon}'_t\}_{t\in\boldsymbol{Z}}$ and $\{\boldsymbol{\epsilon}''_t\}_{t\in\boldsymbol{Z}}$ be two i.i.d. copies of $\{\boldsymbol{\epsilon}_t\}_{t\in\boldsymbol{Z}}$. Let $n = d(\mathcal{S},\mathcal{T})$. Define $J(t,n) := \{t-n, t-n-1, t-n-2, \ldots\}$ and

$$
\begin{aligned}
\mathcal{G}_t &:= (\ldots, \boldsymbol{\epsilon}'_{t-n-1}, \boldsymbol{\epsilon}'_{t-n}, \boldsymbol{\epsilon}_{t-n+1}, \ldots, \boldsymbol{\epsilon}_t), \\
\mathcal{H}_t &:= (\ldots, \boldsymbol{\epsilon}''_{t-n-1}, \boldsymbol{\epsilon}''_{t-n}, \boldsymbol{\epsilon}_{t-n+1}, \ldots, \boldsymbol{\epsilon}_t).
\end{aligned}
$$

$\mathcal{G}_t$ and $\mathcal{H}_t$ are coupled versions of $\mathcal{F}_t$ with $\boldsymbol{\epsilon}_j$ replaced by $\boldsymbol{\epsilon}'_j$ and $\boldsymbol{\epsilon}''_j$ if $j \in J(t,n)$. Now define the process $\{\boldsymbol{X}_t^{[n]}\}_{t\in\boldsymbol{Z}}$ by

$$
\boldsymbol{X}_t^{[n]} := \begin{cases} \boldsymbol{g}(\mathcal{G}_t) \text{ if } t \in \mathcal{S}; \\ \boldsymbol{g}(\mathcal{H}_t) \text{ if } t \in \mathcal{T}. \end{cases}
$$

Again we upper bound B by

$$B \le \underbrace{\left|\mathrm{Cov}\left\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t) - \prod_{t\in\mathcal{S}} h_\epsilon(Y_t^{[n]}), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t)\right\}\right|}_{B_1} + \underbrace{\left|\mathrm{Cov}\left\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t^{[n]}), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t) - \prod_{t\in\mathcal{T}} h_\epsilon(Y_t^{[n]})\right\}\right|}_{B_2} + \underbrace{\left|\mathrm{Cov}\left\{\prod_{t\in\mathcal{S}} h_\epsilon(Y_t^{[n]}), \prod_{t\in\mathcal{T}} h_\epsilon(Y_t^{[n]})\right\}\right|}_{B_3}.$$

Note that by the definition of $\boldsymbol{X}_t^{[n]}$, $\{\boldsymbol{X}_t^{[n]} : t \in \mathcal{S}\}$ and $\{\boldsymbol{X}_t^{[n]} : t \in \mathcal{T}\}$ are independent. Thus, we still have $B_3 = 0$. Using the same technique as in (A.8), we have

$$\begin{aligned} B_1 \le & \frac{3}{2\epsilon}|\mathcal{S}|\mathbb{E}\left|Y_t - Y_t^{[n]}\right| \le \frac{3}{2\epsilon}|\mathcal{S}|\|\boldsymbol{a}\|_*\mathbb{E}\|\boldsymbol{X}_t - \boldsymbol{X}_t^{[n]}\| \\ \le & \frac{3}{2\epsilon}|\mathcal{S}|\|\boldsymbol{a}\|_*\delta_p(I,t,\boldsymbol{g}). \end{aligned}$$

The last equality is due to the fact that $\{\mathbb{E}\|\boldsymbol{X}_t - \boldsymbol{X}_t^{[n]}\|^p\}^{1/p}$ in an increasing function of p. Using similar arguments, we can also obtain $B_2 \le 3|\mathcal{T}|\|\boldsymbol{a}\|_*\delta_p(I,t,\boldsymbol{g})/(2\epsilon)$. Thus, we have

$$B \le B_1 + B_2 \le \frac{3}{2\epsilon}(|\mathcal{S}| + |\mathcal{T}|)\|\boldsymbol{a}\|_*\delta_p(I,t,\boldsymbol{g}).$$

Plugging the above inequality into (A.12) and setting $\epsilon = \sqrt{\delta_p(I,t,\boldsymbol{g})}$, we have

$$\begin{aligned} &\left|\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S}\cup\mathcal{T}\Big) - \right.\\ &\left.\mathbb{P}\Big(Y_t \le u, \forall\, t \in \mathcal{S}\Big)\mathbb{P}\Big(Y_{t'} \le u, \forall\, t' \in \mathcal{T}\Big)\right| \\ &\le \Big(4H + \frac{3}{2}\|\boldsymbol{a}\|_*\Big)(|\mathcal{S}| + |\mathcal{T}|)\sqrt{\delta_p(I,t,\boldsymbol{g})}. \end{aligned}$$

This completes the proof. □

A.4 Proof of Theorem 3.3

Proof. Equation (3.4) implies that

$$\begin{aligned} &F\Big\{F^{-1}\Big(\frac{1}{2}\Big) + \frac{x}{2}\Big\} - \frac{1}{2} \\ &= F\Big\{F^{-1}\Big(\frac{1}{2}\Big) + \frac{x}{2}\Big\} - F\Big\{F^{-1}\Big(\frac{1}{2}\Big)\Big\} \ge \frac{\eta x}{2}, \\ &\frac{1}{2} - F\Big\{F^{-1}\Big(\frac{1}{2}\Big) - \frac{x}{2}\Big\} \\ &= F\Big\{F^{-1}\Big(\frac{1}{2}\Big)\Big\} - F\Big\{F^{-1}\Big(\frac{1}{2}\Big) - \frac{x}{2}\Big\} \ge \frac{\eta x}{2}, \end{aligned}$$

for $0 < x/2 \le \kappa$ and $F \in \{F_{\boldsymbol{a}}, \bar{F}_{\boldsymbol{a}} : \boldsymbol{a} \in \mathbb{S}_e\}$. We allow D_2 in (C.2) to depend on T. Specifically, we define

$$D_{2,T} = 2\Big\{4L(T,d)\Big\}^{1/(a+2)}, \tag{A.13}$$

and correspondingly, let

$$\varphi_T(x) := \frac{Tx^2}{D_1 + D_{2,T}T^{(a+1)/(a+2)}x^{(2a+3)/(a+2)}} \tag{A.14}$$

for $x > 0$. It's easy to check that φ_T is non-decreasing on $(0, \infty)$ by investigating the derivative of $\log \varphi_T(x)$. Thus, using Lemma C.2, we have, for any $j \in \{1, \ldots, d\}$,

$$\begin{aligned}
&\mathbb{P}\Big\{\Big|\widehat{\sigma}^{\mathrm{M}}(\{X_{tj}\}_{t=1}^T) - \sigma^{\mathrm{M}}(X_j)\Big| > x\Big\} \\
&\le 3\exp\Big\{-\varphi_T\Big(\frac{\eta x}{2} - \frac{1}{T}\Big)\Big\} + 3\exp\Big\{-\varphi_T\Big(\frac{\eta x}{2}\Big)\Big\} \\
&\le 6\exp\Big\{-\varphi_T\Big(\frac{\eta x}{2} - \frac{1}{T}\Big)\Big\},
\end{aligned} \tag{A.15}$$

when $0 < x/2 < \kappa$ and $\eta x/2 > 1/T$. Now, by the definitions of $\widehat{\mathbf{R}}_{jj}$ and $\mathbf{R}_{jj}$, we have

$$\begin{aligned}
&\mathbb{P}\Big(|\widehat{\mathbf{R}}_{jj} - \mathbf{R}_{jj}| > x\Big) \\
&= \mathbb{P}\Big[\Big|\widehat{\sigma}^{\mathrm{M}}(\{X_{tj}\}_{t=1}^T)^2 - \sigma^{\mathrm{M}}(X_j)^2\Big| > x\Big] \\
&\le \mathbb{P}\Big[\Big\{\widehat{\sigma}^{\mathrm{M}}(\{X_{tj}\}_{t=1}^T) - \sigma^{\mathrm{M}}(X_j)\Big\}^2 + \\
&\qquad 2\Big|\sigma^{\mathrm{M}}(X_j)\Big\{\widehat{\sigma}^{\mathrm{M}}(\{X_{tj}\}_{t=1}^T) - \sigma^{\mathrm{M}}(X_j)\Big\}\Big| > x\Big] \\
&\le \mathbb{P}\Big\{\Big|\widehat{\sigma}^{\mathrm{M}}(\{X_{tj}\}_{t=1}^T) - \sigma^{\mathrm{M}}(X_j)\Big| > \sqrt{\frac{x}{2}}\Big\} + \\
&\qquad \mathbb{P}\Big\{\Big|\widehat{\sigma}^{\mathrm{M}}(\{X_{tj}\}_{t=1}^T) - \sigma^{\mathrm{M}}(X_j)\Big| > \frac{x}{4\sigma^{\mathrm{M}}(X_j)}\Big\}.
\end{aligned} \tag{A.16}$$

Applying (A.15), we have

$$\begin{aligned}
&\mathbb{P}\Big(|\widehat{\mathbf{R}}_{jj} - \mathbf{R}_{jj}| > x\Big) \\
&\le 6\exp\Big\{-\varphi_T\Big(\frac{\eta}{2}\sqrt{\frac{x}{2}} - \frac{1}{T}\Big)\Big\} + \\
&\qquad 6\exp\Big[-\varphi_T\Big\{\frac{\eta x}{8\sigma^{\mathrm{M}}(X_j)} - \frac{1}{T}\Big\}\Big] \\
&\le 12\max\Big\{\exp\Big\{-\varphi_T\Big(\frac{\eta}{2}\sqrt{\frac{x}{2}} - \frac{1}{T}\Big)\Big\}, \\
&\qquad \exp\Big\{-\varphi_T\Big(\frac{\eta x}{8\sigma^{\mathrm{M}}_{\max}} - \frac{1}{T}\Big)\Big\}\Big\}.
\end{aligned} \tag{A.17}$$

Next, we derive the concentration inequality about $\widehat{\mathbf{R}}_{jk}$ for $j \neq k$. Again, using Lemma C.2, we have, for $j \neq k$,

$$\begin{aligned}
&\mathbb{P}\Big\{\Big|\widehat{\sigma}^{\mathrm{M}}(\{X_{tj}+X_{tk}\}_{t=1}^{T})-\sigma^{\mathrm{M}}(X_j+X_k)\Big|>x\Big\} \\
&\leq 6\exp\Big\{-\varphi_T\Big(\frac{\eta x}{2}-\frac{1}{T}\Big)\Big\}, \qquad (A.18)\\
&\mathbb{P}\Big\{\Big|\widehat{\sigma}^{\mathrm{M}}(\{X_{tj}-X_{tk}\}_{t=1}^{T})-\sigma^{\mathrm{M}}(X_j-X_k)\Big|>x\Big\} \\
&\leq 6\exp\Big\{-\varphi_T\Big(\frac{\eta x}{2}-\frac{1}{T}\Big)\Big\}. \qquad (A.19)
\end{aligned}$$

By the definitions of $\widehat{\mathbf{R}}_{jk}$ and $\mathbf{R}_{jk}$, we have

$$\begin{aligned}
&\mathbb{P}\Big(|\widehat{\mathbf{R}}_{jk}-\mathbf{R}_{jk}|>x\Big) \\
=&\Big(\Big[\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}+X_{tk}\}_{t=1}^{T}\Big)^2-\sigma^{\mathrm{M}}(X_j+X_k)^2\Big]+ \\
&\Big[\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}-X_{tk}\}_{t=1}^{T}\Big)^2-\sigma^{\mathrm{M}}(X_j-X_k)^2\Big]>4x\Big) \\
\leq&\underbrace{\mathbb{P}\Big\{\Big|\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}+X_{tk}\}_{t=1}^{T}\Big)^2-\sigma^{\mathrm{M}}(X_j+X_k)^2\Big|>2x\Big\}}_{P_1}+ \\
&\underbrace{\mathbb{P}\Big\{\Big|\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}-X_{tk}\}_{t=1}^{T}\Big)^2-\sigma^{\mathrm{M}}(X_j-X_k)^2\Big|>2x\Big\}}_{P_2}. \qquad (A.20)
\end{aligned}$$

Using the same technique as in (A.16), we have

$$\begin{aligned}
P_1\leq&\mathbb{P}\Big[\Big\{\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}+X_{tk}\}_{t=1}^{T}\Big)-\sigma^{\mathrm{M}}(X_j+X_k)\Big\}^2+ \\
&2\Big|\sigma^{\mathrm{M}}(X_j+X_k)\Big\{\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}+X_{tk}\}_{t=1}^{T}\Big)- \\
&\sigma^{\mathrm{M}}(X_j+X_k)\Big\}\Big|>2x\Big] \\
\leq&\mathbb{P}\Big\{\Big|\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}+X_{tk}\}_{t=1}^{T}\Big)-\sigma^{\mathrm{M}}(X_j+X_k)\Big|>\sqrt{x}\Big\}+ \\
&\mathbb{P}\Big[\Big|\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}+X_{tk}\}_{t=1}^{T}\Big)-\sigma^{\mathrm{M}}(X_j+X_k)\Big|> \\
&\frac{x}{2\sigma^{\mathrm{M}}(X_j+X_k)}\Big], \qquad (A.21)
\end{aligned}$$

and similarly

$$
\begin{aligned}
P_2 \leq & \mathbb{P}\Big[\Big\{\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}-X_{tk}\}_{t=1}^T\Big)-\sigma^{\mathrm{M}}(X_j-X_k)\Big\}^2+\\
& \quad 2\Big|\sigma^{\mathrm{M}}(X_j-X_k)\Big\{\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}-X_{tk}\}_{t=1}^T\Big)-\\
& \quad \sigma^{\mathrm{M}}(X_j-X_k)\Big\}\Big|>2x\Big]\\
\leq & \mathbb{P}\Big\{\Big|\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}-X_{tk}\}_{t=1}^T\Big)-\sigma^{\mathrm{M}}(X_j-X_k)\Big|>\sqrt{x}\Big\}+\\
& \quad \mathbb{P}\Big[\Big|\widehat{\sigma}^{\mathrm{M}}\Big(\{X_{tj}-X_{tk}\}_{t=1}^T\Big)-\sigma^{\mathrm{M}}(X_j-X_k)\Big|\\
& \quad >\frac{x}{2\sigma^{\mathrm{M}}(X_j-X_k)}\Big].
\end{aligned}
\tag{A.22}
$$

Applying (A.18) and (A.19) to the above two inequalities and noting that $\sigma^{\mathrm{M}}(X_j+X_k)\leq\sigma^{\mathrm{M}}_{\max}$, $\sigma^{\mathrm{M}}(X_j-X_k)\leq\sigma^{\mathrm{M}}_{\max}$, we obtain

$$
\begin{aligned}
P_1 \leq & 6\exp\Big\{-\varphi_T\Big(\frac{\eta\sqrt{x}}{2}-\frac{1}{T}\Big)\Big\}+6\exp\Big\{-\varphi_T\Big(\frac{\eta x}{4\sigma^{\mathrm{M}}_{\max}}-\frac{1}{T}\Big)\Big\},\\
P_2 \leq & 6\exp\Big\{-\varphi_T\Big(\frac{\eta\sqrt{x}}{2}-\frac{1}{T}\Big)\Big\}+6\exp\Big\{-\varphi_T\Big(\frac{\eta x}{4\sigma^{\mathrm{M}}_{\max}}-\frac{1}{T}\Big)\Big\}.
\end{aligned}
$$

Plugging the above two inequalities into (A.20), we have

$$
\begin{aligned}
& \mathbb{P}\Big(|\widehat{\mathbf{R}}_{jk}-\mathbf{R}_{jk}|>x\Big)\\
& \leq 12\exp\Big\{-\varphi_T\Big(\frac{\eta\sqrt{x}}{2}-\frac{1}{T}\Big)\Big\}+12\exp\Big\{-\varphi_T\Big(\frac{\eta x}{8\sigma^{\mathrm{M}}_{\max}}-\frac{1}{T}\Big)\Big\}\\
& \leq 24\max\Big\{\exp\Big\{-\varphi_T\Big(\frac{\eta\sqrt{x}}{2}-\frac{1}{T}\Big)\Big\},\\
& \qquad \exp\Big\{-\varphi_T\Big(\frac{\eta x}{8\sigma^{\mathrm{M}}_{\max}}-\frac{1}{T}\Big)\Big\}\Big\}.
\end{aligned}
\tag{A.23}
$$

Combining (A.17) and (A.23), we have

$$
\begin{aligned}
& \mathbb{P}\Big(\Big\|\widehat{\mathbf{R}}-\mathbf{R}\Big\|_{\max}>x\Big)\\
& \leq 24\max\Big\{\exp\Big\{2\log d-\varphi_T\Big(\frac{\eta}{2}\sqrt{\frac{x}{2}}-\frac{1}{T}\Big)\Big\},\\
& \qquad \exp\Big\{2\log d-\varphi_T\Big(\frac{\eta x}{8\sigma^{\mathrm{M}}_{\max}}-\frac{1}{T}\Big)\Big\}\Big\}.
\end{aligned}
\tag{A.24}
$$

Next, we simplify the above concentration bound using the special structure of function φ_T. Let

$$
\begin{aligned}
b_1(x) &:= \exp\Big\{2\log d-\varphi_T\Big(\frac{\eta}{2}\sqrt{\frac{x}{2}}-\frac{1}{T}\Big)\Big\},\\
b_2(x) &:= \exp\Big\{2\log d-\varphi_T\Big(\frac{\eta x}{8\sigma^{\mathrm{M}}_{\max}}-\frac{1}{T}\Big)\Big\}.
\end{aligned}
$$

We discuss the form of the concentration bound in two scenarios:

(i) If $b_1(x) \geq b_2(x)$, we focus on $b_1(x)$. We remind that by the definition of function φ_T, we have

$$\varphi_T\Big(\frac{\eta}{2}\sqrt{\frac{x}{2}} - \frac{1}{T}\Big) = \frac{T\Big(\frac{\eta}{2}\sqrt{\frac{x}{2}} - \frac{1}{T}\Big)^2}{D_1 + D_{2,T}T^{(a+1)/(a+2)}\Big(\frac{\eta}{2}\sqrt{\frac{x}{2}} - \frac{1}{T}\Big)^{(2a+3)/(a+2)}},$$

where D_1 and $D_{2,T}$ are defined in (C.3) and (A.13). To simplify the denominator on the right-hand side of the above equation, we require that

$$D_1 \geq D_{2,T}T^{(a+1)/(a+2)}\Big(\frac{\eta}{2}\sqrt{\frac{x}{2}} - \frac{1}{T}\Big)^{(2a+3)/(a+2)}. \tag{A.25}$$

Then we have $\varphi_T\{\eta\sqrt{x}/(2\sqrt{2}) - 1/T\} \geq T\{\eta\sqrt{x}/(2\sqrt{2}) - 1/T\}^2/(2D_1)$. By the definition of $b_1(x)$, we have

$$b_1(x) \leq \exp\Big\{2\log d - \frac{T}{2D_1}\Big(\frac{\eta}{2}\sqrt{\frac{x}{2}} - \frac{1}{T}\Big)^2\Big\}.$$

Setting $\exp\Big[2\log d - T\Big\{\eta\sqrt{x}/(2\sqrt{2}) - 1/T\Big\}^2/(2D_1)\Big] = \alpha^2$ for some $\alpha \in (0,1)$, we obtain

$$\begin{aligned} x &= \frac{8}{\eta^2}\Big\{\sqrt{\frac{4D_1(\log d - \log\alpha)}{T}} + \frac{1}{T}\Big\}^2 \\ &:= x_1(T,d). \end{aligned} \tag{A.26}$$

Under (A.26), for $d > 1/\alpha$, we have $\eta\sqrt{x}/(2\sqrt{2}) - 1/T \leq \sqrt{8D_1\log d/T}$. Thus, (A.25) holds if we require

$$D_1 \geq D_{2,T}T^{(a+1)/(a+2)}(8D_1\log d/T)^{(a+3/2)/(a+2)}.$$

Plugging the definitions of D_1 and $D_{2,T}$ into the above inequality, it follows that (A.25) holds when we have

$$L(T,d) \leq \sqrt{\frac{L_1 T}{2^{7a+11}(\log d)^{2a+3}}}. \tag{A.27}$$

(ii) If $b_1(x) < b_2(x)$, we follow a similar argument as in (i) and require that

$$D_1 \geq D_{2,T}T^{(a+1)/(a+2)}\Big(\frac{\eta x}{8\sigma^{\mathrm{M}}_{\max}} - \frac{1}{T}\Big)^{(2a+3)/(a+2)}. \tag{A.28}$$

This leads to $\varphi_T\{\eta x/(8\sigma^{\mathrm{M}}_{\max}) - 1/T\} \geq T\{\eta x/(8\sigma^{\mathrm{M}}_{\max}) - 1/T\}^2/(2D_1)$. By the definition of $b_2(x)$, we have

$$b_2(x) \leq \exp\Big\{2\log d - \frac{T}{2D_1}\Big(\frac{\eta x}{8\sigma^{\mathrm{M}}_{\max}} - \frac{1}{T}\Big)^2\Big\}.$$

Setting $\exp\Big[2\log d - T\Big\{\eta x/(8\sigma^{\mathrm{M}}_{\max}) - 1/T\Big\}^2/(2D_1)\Big] = \alpha^2$, we obtain

$$\begin{aligned} x &= \frac{8\sigma^{\mathrm{M}}_{\max}}{\eta}\Big\{\sqrt{\frac{4D_1(\log d - \log\alpha)}{T}} + \frac{1}{T}\Big\} \\ &:= x_2(T, d). \end{aligned} \tag{A.29}$$

Under (A.29), we have $\eta x/(8\sigma^{\mathrm{M}}_{\max}) - 1/T \le \sqrt{8D_1\log d/T}$ if $d > 1/\alpha$. Thus, (A.28) holds if we again require

$$D_1 \ge D_{2,T}T^{(a+1)/(a+2)}(8D_1\log d/T)^{(a+3/2)/(a+2)}.$$

Now, using the definitions of D_1 and $D_{2,T}$, we obtain that (A.28) is also guaranteed by (A.27).

Now we summarize the discussion above and derive the final rate of convergence. In (A.24), we set $x = \max\Big\{x_1(T, d), x_2(T, d)\Big\}$ and require that (A.27) holds. When $x_1(T, d) \ge x_2(T, d)$, we have $x = x_1(T, d)$. Thus, together with (A.27), we have $b_1(x) \le \alpha^2$. Since $b_2(x)$ is non-increasing in x, we have $b_2\{x_1(T, d)\} \le b_2\{x_2(T, d)\} \le \alpha^2$. The last inequality is ensured by (A.27). Thus, we obtain

$$\begin{aligned} \mathbb{P}\Big(\|\widehat{\mathbf{R}} - \mathbf{R}\|_{\max} > x\Big) &\le 24\max\Big\{b_1(x), b_2(x)\Big\} \\ &\le 24\alpha^2. \end{aligned} \tag{A.30}$$

On the other hand, when $x_1(T, d) < x_2(T, d)$, we have $x = x_2(T, d)$. Thus, together with (A.27), we have $b_2(x) \le \alpha^2$. Since $b_1(x)$ is non-increasing in x, we have $b_1\{x_2(T, d)\} \le b_1\{x_1(T, d)\} \le \alpha^2$, where the last inequality is ensured by (A.27). Thus, again, we can obtain (A.30). So, in either case, we have

$$\mathbb{P}\Big(\|\widehat{\mathbf{R}} - \mathbf{R}\|_{\max} > \max\Big\{x_1(T, d), x_2(T, d)\Big\}\Big) \le 24\alpha^2,$$

when T and d are large enough. This completes the proof of (3.5). □

A.5 Proof of Theorem 2

Proof. Using Theorem 3.3, to derive the rate of convergence for $\widehat{\mathbf{R}}$, we only need to verify Condition 3.1 for the VAR time series $\{\boldsymbol{X}_t\}_{t=1}^T$. Theorem 2.1 shows that $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is $(\mathbb{S}_e, \Psi, \rho)$-Kolmogorov dependent with $\Psi(u, v) = u + v$ and $\rho(n) = \{4H + 3C/(1 - \|\mathbf{A}\|_2)\}\|\mathbf{A}\|_2^{n/2}$. To verify (3.2), note that for any $k \ge 0$, we have

$$\begin{aligned} \sum_{n=0}^{\infty}(n+1)^k\|\mathbf{A}\|_2^{n/2} &\le \sum_{n=0}^{\infty}(n+1)\cdots(n+k)\|\mathbf{A}\|_2^{n/2} \\ &= \frac{d^k}{dx^k}\Big(\frac{1}{1-x}\Big)\Big|_{x=\sqrt{\|\mathbf{A}\|_2}} = \frac{k!}{(1-\sqrt{\|\mathbf{A}\|_2})^{k+1}}. \end{aligned}$$

Thus, (3.2) holds with

$$\begin{aligned} L_1 &= \Big(4H + \frac{3C}{1 - \|\mathbf{A}\|_2}\Big)\frac{1}{1 - \sqrt{\|\mathbf{A}\|_2}}, \\ L &= \frac{1}{1 - \sqrt{\|\mathbf{A}\|_2}}, \text{ and } a = 1. \end{aligned}$$

Plugging the above parameters into (3.5) leads to the desired result. $\square$

A.6 Proof of Theorem 3

Proof. Using Theorem 3.3, to derive the rate of convergence for $\widehat{\mathbf{R}}$, we only need to verify Condition 3.1 for the α-mixing time series $\{\boldsymbol{X}_t\}_{t=1}^T$. Theorem 2.3 shows that $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is $(\mathbb{S}_e,\Psi,\rho)$-Kolmogorov dependent with $\rho(n)=C_1\exp(-C_2n^r)$ and any of the Ψ functions required by Theorem 3.3.

When $r\ge 1$, for any $k\le 0$, we have

$$\begin{aligned}&\sum_{n=0}^{\infty}(n+1)^k\exp(-C_2n^r)\le\sum_{n=0}^{\infty}(n+1)^k\exp(-C_2n)\\&\le\sum_{n=0}^{\infty}(n+1)\cdots(n+k)\exp(-C_2n)\\&=\frac{d^k}{dx^k}\Big(\frac{1}{1-x}\Big)\Big|_{x=e^{-C_2}}=\frac{k!}{(1-e^{-C_2})^{k+1}}.\end{aligned}$$

Thus, (3.2) holds with $L_1=C_1/(1-c^{-C_2})$, $L=1/(1-c^{-C_2})$, and $a=1$. Plugging these parameters into (3.5) leads to (3.7).

When $0<r<1$ and $k=0$, we have

$$\begin{aligned}&\sum_{n=0}^{\infty}(n+1)^k\rho(n)=C_1\sum_{n=0}^{\infty}\exp(-C_2n^r)\\&\le C_1\Big\{1+\int_0^{\infty}\exp(-C_2u^r)du\Big\}=C_1\Big\{1+\frac{1}{rC_2^{1/r}}\Gamma\Big(\frac{1}{r}\Big)\Big\},\end{aligned}$$

where $\Gamma(x)=\int_0^\infty u^{x-1}e^{-u}du$ is the Gamma function. By Lemma C.3, we can upper bound $\Gamma(1/r)$ by

$$\Gamma\Big(\frac{1}{r}\Big)\le\sqrt{2\pi}\Big(\frac{1/r-1/2}{e}\Big)^{1/r-1/2}\le\sqrt{2\pi}\Big(\frac{1}{er}\Big)^{1/r-1/2}.$$

Thus, we have

$$\sum_{n=0}^{\infty}(n+1)^k\rho(n)\le C_1\Big\{1+\sqrt{\frac{2\pi e}{r}}\Big(\frac{1}{eC_2r}\Big)^{1/r}\Big\}. \tag{A.31}$$

Now we consider $0<r<1$ and $k\ge 1$. Denote $f(u):=(u+1)^k\exp(-C_2u^r)$ and $g(u):=(u+2)^k\exp(-C_2u^r)$. For any $n\in\mathbb{Z}^+$, we have $f(n+1)\le g(u)$ for all $u\in[n,n+1]$, and thus, $f(n+1)\le\int_n^{n+1}g(u)du$. Therefore, we have

$$\begin{aligned}&\sum_{n=0}^{\infty}(n+1)^k\rho(n)=C_1\Big\{1+2^ke^{-C_2}+\sum_{n=1}^{\infty}f(n+1)\Big\}\\&\le C_1\Big\{1+2^ke^{-C_2}+\int_1^{\infty}g(u)du\Big\} \qquad (A.32)\end{aligned}$$

Next we derive an upper bound for $\int_1^\infty g(u)du$. We have

$$
\begin{aligned}
\int_1^\infty g(u)du &\le 3^k \int_0^\infty u^k \exp(-C_2 u^r) du \\
&= \frac{3^k}{r C_2^{(k+1)/r}} \Gamma\Big(\frac{k+1}{r}\Big).
\end{aligned} \tag{A.33}
$$

Using Lemma C.3, we can upper bound $\Gamma\{(k+1)/r\}$ by

$$
\begin{aligned}
&\Gamma\Big(\frac{k+1}{r}\Big) \le \sqrt{2\pi}\Big\{\frac{(k+1)/r - 1/2}{e}\Big\}^{(k+1)/r-1/2} \\
&\le \sqrt{2\pi}\Big(\frac{k+1}{er}\Big)^{(k+1)/r-1/2} \\
&= \sqrt{2\pi}\Big(\frac{1}{r}\Big)^{(k+1)/r-1/2}\Big(\frac{e}{k+1}\Big)^{1/(2r)+1/2}\Big[\Big(\frac{k+1}{e}\Big)^{k+3/2}\Big]^{1/r} \\
&\le \sqrt{2\pi}\Big(\frac{1}{r}\Big)^{(k+1)/r-1/2}\Big(\frac{e}{2}\Big)^{1/(2r)+1/2}\Big[\Big(\frac{k+3/2}{e}\Big)^{k+3/2}\Big]^{1/r}.
\end{aligned} \tag{A.34}
$$

Using Lemma C.3 again, we can upper bound the last term in the product by

$$
\Big[\Big(\frac{k+3/2}{e}\Big)^{k+3/2}\Big]^{1/r} \le \Big\{\frac{(k+1)!}{\sqrt{2e}}\Big\}^{1/r} \le \Big(\frac{2^k k!}{\sqrt{2e}}\Big)^{1/r}. \tag{A.35}
$$

Putting together (A.32), (A.34), and (A.35), we obtain

$$
\int_1^\infty g(u)du \le \sqrt{\frac{\pi e}{r}}\Big(\frac{1}{2C_2 r}\Big)^{1/r}\Big\{3\Big(\frac{2}{C_2 r}\Big)^{1/r}\Big\}^k (k!)^{1/r}.
$$

Plugging the above equation into (A.32), we obtain

$$
\begin{aligned}
&\sum_{n=0}^\infty (n+1)^k \rho(n) \\
&\le C_1\Big[1 + 2^k e^{-C_2} + \sqrt{\frac{\pi e}{r}}\Big(\frac{1}{2C_2 r}\Big)^{1/r}\Big\{3\Big(\frac{2}{C_2 r}\Big)^{1/r}\Big\}^k (k!)^{1/r}\Big] \\
&\le C_1\Big\{1 + \sqrt{\frac{\pi e}{r}}\Big(\frac{1}{2C_2 r}\Big)^{1/r}\Big\}\max\Big\{3, 3\Big(\frac{2}{C_2 r}\Big)^{1/r}\Big\}^k (k!)^{1/r},
\end{aligned}
$$

where the last equation uses $1 + 2^k e^{-C_2} \le 3^k$ for $k \ge 1$. Now, combining the above equation with (A.31), we obtain $\sum_{n=0}^\infty (n+1)^k \rho(n) \le L_1 L^k (k!)^a$ with

$$
\begin{aligned}
L_1 &= C_1\Big\{1 + \sqrt{\frac{2\pi e}{r}}\Big(\frac{1}{2C_2 r}\Big)^{1/r}\Big\}, \\
L &= \max\Big\{3, 3\Big(\frac{2}{C_2 r}\Big)^{1/r}\Big\}, \text{ and } a = \frac{1}{r}.
\end{aligned}
$$

Plugging these parameters into (3.5) leads to (3.8). This completes the proof. □

B CONCENTRATION INEQUALITIES UNDER WEAK DEPENDENCE

In this section, we develop a concentration inequality for sums of weakly dependent random variables. We first reformulate Theorem 1 in Doukhan and Neumann (2007).

Lemma B.1. Suppose $\{X_t\}_{t=1}^T$ is a sequence of random variables with mean 0, defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\Psi : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{R}$ be one of the following functions:

(a) $\Psi(u, v) = 2v$;

(b) $\Psi(u, v) = \beta(u + v) + (1 - \beta)uv$, for some $\beta \in [0, 1]$.

Assume that there exist constants $M, L_1, L_2 > 0$, $a, b \geq 0$, and a non-increasing sequence of real coefficients $\{\rho(n)\}_{n\geq 0}$ such that for any u-tuple $(s_1, \ldots, s_u)$ and v-tuple $(t_1, \ldots, t_v)$ with $1 \leq s_1 \leq \cdots \leq s_u < t_1 \leq \cdots \leq t_v \leq T$, we have

$$\begin{aligned}&\Big|\mathrm{Cov}\Big(\prod_{i=1}^{u} X_{s_i}, \prod_{j=1}^{v} X_{t_j}\Big)\Big| \\ &\leq M^{u+v}\{(u + v)!\}^b \Psi(u, v)\rho(t_1 - s_u),\end{aligned} \tag{B.1}$$

where the sequence $\{\rho(n)\}_{n\geq 0}$ satisfies

$$\sum_{n=0}^{\infty}(n+1)^k \rho(n) \leq L_1 L_2^k (k!)^a, \text{ for any } k \in \mathbb{Z}^+. \tag{B.2}$$

Moreover, we require that the following moment condition holds:

$$\mathbb{E}|X_t|^k \leq (k!)^b M^k, \text{ for any } t = 1, \ldots, T, k \in \mathbb{Z}^+. \tag{B.3}$$

Then, for $S_T := \sum_{t=1}^T X_t$ and any $x > 0$, we have

$$\mathbb{P}(S_T \geq x) \leq \exp\Big\{-\frac{x^2}{C_1 T + C_2 x^{(2a+2b+3)/(a+b+2)}}\Big\},$$

where C_1 and C_2 are constants given by

$$C_1 = 2^{a+b+4}M^2 L_1 \text{ and } C_2 = 2(2ML_2)^{1/(a+b+2)}.$$

Proof. The proof follows that of Theorem 1 in Doukhan and Neumann (2007) with minor modifications, as listed below. We inherit the notations in Doukhan and Neumann (2007), and set $K = 1$.

Equation (30) in Doukhan and Neumann (2007) can be strengthened to

$$\mathbb{E}|Y_j| \leq 2^{k-j-1}\{(k - j + 1)!\}^b M^k \rho(t_{t+1} - t_t).$$

This leads to

$$|\overline{\mathbb{E}}(X_{t_1} \cdots X_{t_k})| \leq 2^{k-1}(k!)^b k^2 M^k \rho(t_{t+1} - t_t), \tag{B.4}$$

which corresponds to Lemma 13 in Doukhan and Neumann (2007). Using (B.4), we obtain that

$$
\begin{aligned}
&\Big|\Gamma(X_{t_1},\ldots,X_{t_k})\Big| \\
&\leq \sum_{\nu=1}^{k}\sum_{\bigcup_{p=1}^{\nu} I_p = I} N_\nu(I_1,\ldots,I_\nu)2^{k-\nu}(k!)^b M^k \min_{1\leq t<k}\rho(t_{t+1}-t_t) \\
&\leq 2^{k-1}M^k(k!)^b\{(k-1)!\}\min_{1\leq t<k}\rho(t_{t+1}-t_t).
\end{aligned}
$$

Thus, we have

$$
\Big|\Gamma_k(S_T)\Big| \leq n2^{k-1}M^k(k!)^{b+1}\sum_{s=0}^{T-1}(s+1)^{k-2}\rho(s). \tag{B.5}
$$

Equation (B.5) corresponds to Lemma 14 in Doukhan and Neumann (2007). The rest follows the same technique as in Doukhan and Neumann (2007). □

Equations (B.1) and (B.2) characterize the dependence structure of the sequence $\{X_t\}_{t=1}^T$. In detail, the covariance between the past, $\{X_{s_i}\}_{i=1}^u$, and the future, $\{X_{t_j}\}_{j=1}^v$, converges to 0 as the gap in time between them increases to infinity. (B.2) specifies the speed of the convergence. Equation (B.3) is a moment condition. In the next lemma, we further show that these conditions are location and scale invariant.

Lemma B.2. Let $\{X_t\}_{t=1}^T$ be a sequence of random variables satisfying (B.1)-(B.3). Let $\{\mu_t\}_{t=1}^T$ and $\{\gamma_t\}_{t=1}^T$ be uniformly bounded real sequences in the sense that $|\mu_t| \leq \mu$ and $0 < \gamma_t \leq \gamma$ for $t = 1,\ldots,T$, where μ and γ are constants. Let $\{Y_t\}_{t=1}^T$ be a location-scale transformed sequence defined as

$$
Y_t := \gamma_t(X_t + \mu_t),\ t = 1\ldots,T.
$$

Then (B.1)-(B.3) are satisfied by $\{Y_t\}_{t=1}^T$ with M replaced by $\gamma(M+\mu)$.

Proof. Equation (B.3) can be easily verified for $\{Y_t\}_{t=1}^T$:

$$
\begin{aligned}
\mathbb{E}|Y_t|^k = \mathbb{E}\Big|\gamma_t(X_t+\mu_t)\Big|^k &\leq \gamma^k\sum_{j=0}^{k}\mathbb{E}|X_t|^j|\mu_t|^{k-j} \\
&\leq (k!)^b\Big\{\gamma(M+\mu)\Big\}^k.
\end{aligned}
$$

The last inequality follows from (B.3). Next, we verify that $\{Y_t\}_{t=1}^T$ also satisfy (B.1) and (B.2). Let $\mathcal{S} := \{s_1,\ldots,s_u\}$, $\mathcal{T} := \{t_1,\ldots,t_v\}$, and $\mathcal{R} := \mathcal{S}\cup\mathcal{T}$. By the definition of $\{Y_t\}_{t=1}^T$, we have

$$
\begin{aligned}
\mathbb{E}\prod_{t\in\mathcal{R}} Y_t &= \prod_{t\in\mathcal{R}}\gamma_t\mathbb{E}\prod_{j\in\mathcal{R}}(X_j+\mu_j) \\
&= \prod_{t\in\mathcal{R}}\gamma_t\sum_{\mathcal{U}\subseteq\mathcal{R}}\prod_{j\in\mathcal{R}\setminus\mathcal{U}}\mu_j\mathbb{E}\prod_{k\in\mathcal{U}}X_k \\
&= \prod_{t\in\mathcal{R}}\gamma_t\sum_{\mathcal{U}\subseteq\mathcal{S},\mathcal{V}\subseteq\mathcal{T}}\prod_{j\in\mathcal{R}\setminus(\mathcal{U}\cup\mathcal{V})}\mu_j\mathbb{E}\prod_{k\in\mathcal{U}\cup\mathcal{V}}X_k.
\end{aligned} \tag{B.6}
$$

Applying the same technique to $\mathbb{E}\prod_{t\in\mathcal{S}} Y_t$ and $\mathbb{E}\prod_{j\in\mathcal{T}} Y_j$, we obtain

$$\begin{aligned}
&\mathbb{E}\prod_{t\in\mathcal{S}} Y_t \mathbb{E}\prod_{j\in\mathcal{T}} Y_j \\
&= \prod_{t\in\mathcal{R}} \gamma_t \Big(\sum_{\mathcal{U}\subseteq\mathcal{S}} \prod_{j\in\mathcal{S}\setminus\mathcal{U}} \mu_j \mathbb{E}\prod_{k\in\mathcal{U}} X_k \Big) \Big(\sum_{\mathcal{V}\subseteq\mathcal{T}} \prod_{j\in\mathcal{T}\setminus\mathcal{V}} \mu_j \mathbb{E}\prod_{k\in\mathcal{V}} X_k \Big) \\
&= \prod_{t\in\mathcal{R}} \gamma_t \sum_{\mathcal{U}\subseteq\mathcal{S},\mathcal{V}\subseteq\mathcal{T}} \prod_{j\in\mathcal{R}\setminus(\mathcal{U}\cup\mathcal{V})} \mu_j \mathbb{E}\prod_{k\in\mathcal{U}} X_k \mathbb{E}\prod_{\ell\in\mathcal{V}} X_l. \qquad \text{(B.7)}
\end{aligned}$$

By the definition of covariance, we have

$$\Big|\mathrm{Cov}(\prod_{t\in\mathcal{S}} Y_t, \prod_{t\in\mathcal{T}} Y_t)\Big| = \Big|\mathbb{E}\prod_{t\in\mathcal{R}} Y_t - \mathbb{E}\prod_{j\in\mathcal{S}} Y_j \mathbb{E}\prod_{k\in\mathcal{T}} Y_k\Big|.$$

Plugging (B.6) and (B.7) into the above equation, we have

$$\begin{aligned}
&\Big|\mathrm{Cov}(\prod_{t\in\mathcal{S}} Y_t, \prod_{t\in\mathcal{T}} Y_t)\Big| \\
&= \Big| \prod_{t\in\mathcal{R}} \gamma_t \Big\{ \sum_{\mathcal{U}\subseteq\mathcal{S},\mathcal{V}\subseteq\mathcal{T}} \prod_{j\in\mathcal{R}\setminus(\mathcal{U}\cup\mathcal{V})} \mu_j \\
&\qquad \Big(\mathbb{E}\prod_{k\in\mathcal{U}\cup\mathcal{V}} X_k - \mathbb{E}\prod_{\ell\in\mathcal{U}} X_l \mathbb{E}\prod_{m\in\mathcal{V}} X_m \Big) \Big\} \Big| \\
&\le \prod_{t\in\mathcal{R}} \gamma_t \Big\{ \sum_{\mathcal{U}\subseteq\mathcal{S},\mathcal{V}\subseteq\mathcal{T}} \prod_{j\in\mathcal{R}\setminus(\mathcal{U}\cup\mathcal{V})} \mu_j \\
&\qquad \Big| \mathbb{E}\prod_{k\in\mathcal{U}\cup\mathcal{V}} X_k - \mathbb{E}\prod_{\ell\in\mathcal{U}} X_l \mathbb{E}\prod_{m\in\mathcal{V}} X_m \Big| \Big\} \\
&= \prod_{t\in\mathcal{R}} \gamma_t \Big\{ \sum_{\mathcal{U}\subseteq\mathcal{S},\mathcal{V}\subseteq\mathcal{T}} \prod_{j\in\mathcal{R}\setminus(\mathcal{U}\cup\mathcal{V})} \mu_j \Big|\mathrm{Cov}(\prod_{k\in\mathcal{U}} X_k, \prod_{\ell\in\mathcal{V}} X_l)\Big| \Big\}. \qquad \text{(B.8)}
\end{aligned}$$

Now, (B.1) for $\{X_t\}_{t=1}^T$ implies that

$$\begin{aligned}
&\Big|\mathrm{Cov}(\prod_{t\in\mathcal{U}} X_t, \prod_{t\in\mathcal{V}} X_t)\Big| \\
&\le K^2 M^{|\mathcal{U}|+|\mathcal{V}|} \Big\{ (|\mathcal{U}|+|\mathcal{V}|)! \Big\}^b \Psi(|\mathcal{U}|,|\mathcal{V}|) \rho\Big\{ d(\mathcal{U},\mathcal{V}) \Big\} \\
&\le K^2 M^{|\mathcal{U}|+|\mathcal{V}|} \Big\{ (u+v)! \Big\}^b \Psi(u,v) \rho(t_1 - s_u), \qquad \text{(B.9)}
\end{aligned}$$

where the last inequality is due to $\mathcal{U} \subseteq \mathcal{S}$ and $\mathcal{V} \subseteq \mathcal{T}$. Plugging (B.9) into (B.8), we have

$$\begin{aligned}
&\Big|\mathrm{Cov}(\prod_{t\in\mathcal{S}} Y_t, \prod_{t\in\mathcal{T}} Y_t)\Big| \\
&\leq \prod_{t\in\mathcal{R}} \gamma_t \Big\{K^2\Big\{(u+v)!\Big\}^b \Psi(u,v)\rho(t_1-s_u) \\
&\qquad \sum_{\mathcal{U}\subseteq\mathcal{S},\mathcal{V}\subseteq\mathcal{T}} M^{|\mathcal{U}|+|\mathcal{V}|} \prod_{j\in\mathcal{R}\setminus(\mathcal{U}\cup\mathcal{V})} \mu_j\Big\} \\
&= \prod_{t\in\mathcal{R}} \gamma_t K^2\Big\{(u+v)!\Big\}^b \Psi(u,v)\rho(t_1-s_u) \\
&\qquad \Big(\sum_{\mathcal{W}\subseteq\mathcal{R}} M^{|\mathcal{W}|} \prod_{j\in\mathcal{R}\setminus\mathcal{W}} \mu_j\Big).
\end{aligned}$$

Noting that $\sum_{\mathcal{W}\subseteq\mathcal{R}} M^{|\mathcal{W}|} \prod_{j\in\mathcal{R}\setminus\mathcal{W}} \mu_j = \prod_{j\in\mathcal{R}}(M+\mu_j)$, we further obtain

$$\begin{aligned}
&\Big|\mathrm{Cov}(\prod_{t\in\mathcal{S}} Y_t, \prod_{t\in\mathcal{T}} Y_t)\Big| \\
&\leq K^2 \prod_{t\in\mathcal{R}} \gamma_t \prod_{j\in\mathcal{R}} (M+\mu_j)\Big\{(u+v)!\Big\}^b \Psi(u,v)\rho(t_1-s_u) \\
&\leq K^2\Big\{\gamma(M+\mu)\Big\}^{u+v}\Big\{(u+v)!\Big\}^b \Psi(u,v)\rho(t_1-s_u).
\end{aligned}$$

Thus, (B.1) and (B.2) are satisfied by $\{Y_t\}_{t=1}^T$ with M replaced by $\gamma(M+\mu)$. This completes the proof. □

Using Lemma B.2, we can remove the zero-mean requirement for $\{X_t\}_{t=1}^T$ in Lemma B.1. The next theorem summarizes Lemmas B.1 and B.2.

Theorem B.3. Let $\{X_t\}_{t=1}^T$ be a sequence of random variables satisfying (B.1)-(B.3). Suppose $\mathbb{E}X_t = \mu_t$ and $|\mu_t| \leq \mu$ for $t = 1, \dots, T$, where $\mu > 0$ is a constant. Let $S_T := \sum_{t=1}^T (X_t - \mu_t)$. Then, for any $x > 0$, we have

$$\begin{aligned}
&\mathbb{P}(S_T \geq x) \\
&\leq \exp\Big\{-\frac{x^2}{D_1 T + D_2 x^{(2a+2b+3)/(a+b+2)}}\Big\}.
\end{aligned} \tag{B.10}$$

Here D_1 and D_2 are constants defined by

$$\begin{aligned}
D_1 &= 2^{a+b+4}(M+\mu)^2 L_1, \\
D_2 &= 2\{2(M+\mu)L_2\}^{1/(a+b+2)},
\end{aligned}$$

where a, b, M, L_1, L_2 are constants defined in (B.1)-(B.3).

C SUPPORTING LEMMAS

Lemmas C.1 and C.2 are used in the proof of Theorem 3.3. They provide tail probabilities for related quantile-based statistics. Lemma C.3 is a Stirling-type bound on the Gamma function, and is used in the proof of Theorem 3.

Lemma C.1. Let $\{X_t\}_{t\in\mathbb{Z}}$ be a (Ψ,ρ)-Kolmogorov dependent time series, and $\{X_t\}_{t=1}^T$ be a sequence of observations. Assume that the sequence $\{\rho(n)\}_{n\geq 0}$ satisfies

$$\sum_{n=0}^{\infty}(n+1)^k\rho(n) \leq L_1 L_2^k (k!)^a,\ \forall\ k\geq 0, \tag{C.1}$$

for some constants $L_1, L_2 > 0$ and $a \geq 0$. Then, for any $x > 0$ and $q \in (0,1)$, we have

$$\begin{aligned}
&\mathbb{P}(|\widehat{Q}(\{X_t\};q) - Q(X_1;q)| \geq x)\\
&\leq \exp\Big(-\varphi\Big[F\Big\{F^{-1}(q)+x\Big\} - q - \frac{1}{T}\Big]\Big) + \\
&\quad \exp\Big(-\varphi\Big[q - F\Big\{F^{-1}(q)-x\Big\}\Big]\Big),
\end{aligned}$$

whenever we have $F\{F^{-1}(q)+x\} > q + 1/T$. Here the function φ is defined as

$$\varphi(x) := \frac{Tx^2}{D_1 + D_2 T^{(a+1)/(a+2)} x^{(2a+3)/(a+2)}},\ \ \text{for}\ \ x > 0, \tag{C.2}$$

where D_1 and D_2 are constants given by

$$D_1 = 2^{a+6}L_1, \tag{C.3}$$

$$D_2 = 2(4L_2)^{1/(a+2)}. \tag{C.4}$$

Proof. Let F_T be the empirical distribution function of the sample $\{X_t\}_{t=1}^T$ and denote $F_T^{-1}(q) = \widehat{Q}(\{X_t\}_{t=1}^T;q)$. By the definition of $\widehat{Q}(\cdot;\cdot)$ in (3.1), we have, for any $\epsilon\in[0,1]$,

$$\epsilon \leq F_T\{F_T^{-1}(\epsilon)\} \leq \epsilon + \frac{1}{T}. \tag{C.5}$$

By definition, we have

$$\begin{aligned}
&\mathbb{P}\Big\{\widehat{Q}(\{X_t\};q) - Q(X;q) \geq x\Big\} = \mathbb{P}\Big\{F_T^{-1}(q) - F^{-1}(q) \geq x\Big\}\\
&\leq\ \mathbb{P}\Big[F_T\{F_T^{-1}(q)\} \geq F_T\{F^{-1}(q)+x\}\Big],
\end{aligned}$$

where the last inequality is because F_T is non-decreasing. By (C.5), we have

$$\begin{aligned}
&\mathbb{P}\Big\{\widehat{Q}(\{X_t\}_{t=1}^T;q) - Q(X;q) \geq x\Big\}\\
&\leq \mathbb{P}\Big[q + \frac{1}{T} \geq F_T\{x + F^{-1}(q)\}\Big].
\end{aligned}$$

By the definition of F_T, we further have

$$\begin{aligned}
&\mathbb{P}\Big\{\widehat{Q}(\{X_t\}_{t=1}^T;q) - Q(X;q) \geq x\Big\}\\
&\leq \mathbb{P}\Big[\sum_{t=1}^T I\{X_t \leq F^{-1}(q)+x\} \leq Tq + 1\Big]\\
&= \mathbb{P}\Big(\sum_{t=1}^T\Big[-I\{X_t \leq F^{-1}(q)+x\} + F\{F^{-1}(q)+x\}\Big]\\
&\qquad \geq T\Big[F\{F^{-1}(q)+x\} - q - \frac{1}{T}\Big]\Big).
\end{aligned}$$

Since $\{X_t\}_{t\in\mathbb{Z}}$ is (Ψ,ρ)-Kolmogorov dependent, we have

$$\begin{aligned}&\operatorname{Cov}\Big[\prod_{t\in\mathcal{S}} I\Big\{X_t \le F^{-1}(q)+x\Big\}, \prod_{t\in\mathcal{T}} I\Big\{X_t \le F^{-1}(q)+x\Big\}\Big]\\ &\le \Psi(|\mathcal{S}|,|\mathcal{T}|)\rho\Big\{d(\mathcal{S},\mathcal{T})\Big\},\end{aligned}$$

for any $\mathcal{S},\mathcal{T}\subseteq\{1,\ldots,T\}$ with $\max(\mathcal{S})\le\min(\mathcal{T})$. Thus, by Theorem B.3, we have

$$\begin{aligned}&\mathbb{P}\Big\{\widehat{Q}(\{X_t\}_{t=1}^T;q) - Q(X;q) \ge x\Big\}\\ &\le \exp\Big(-\varphi\Big[F\{F^{-1}(q)+x\} - q - \frac{1}{T}\Big]\Big),\end{aligned} \tag{C.6}$$

where function φ is defined in (C.2). On the other hand, we have

$$\begin{aligned}&\mathbb{P}\Big\{\widehat{Q}(\{X_t\}_{t=1}^T;q) - Q(X;q) \le -x\Big\}\\ &= \mathbb{P}\Big\{F_T^{-1}(q) - F^{-1}(q) \le -x\Big\}\\ &\le \mathbb{P}\Big[F_T\{F_T^{-1}(q)\} \le F_T\{F^{-1}(q)-x\}\Big].\end{aligned}$$

Using (C.5) again, we have

$$\begin{aligned}&\mathbb{P}\Big\{\widehat{Q}(\{X_t\}_{t=1}^T;q) - Q(X;q) \le -x\Big\}\\ &\le \mathbb{P}\Big[q \le F_T\{F^{-1}(q)-x\}\Big]\\ &= \mathbb{P}\Big(\sum_{t=1}^T \Big[I\{X_t \le F^{-1}(q)-x\} - F\{F^{-1}(q)-x\}\Big]\\ &\qquad \ge T\Big[q - F\{F^{-1}(q)-x\}\Big]\Big).\end{aligned}$$

Thus, by Theorem B.3, we have

$$\begin{aligned}&\mathbb{P}\Big\{\widehat{Q}(\{X_t\}_{t=1}^T;q) - Q(X;q) \le -x\Big\}\\ &\le \exp\Big(-\varphi\Big[q - F\{F^{-1}(q)-x\}\Big]\Big),\end{aligned} \tag{C.7}$$

where function φ is defined in (C.2). Combining (C.6) and (C.7) completes the proof. $\square$

Lemma C.2. Under the assumptions in Lemma C.1, we have, for any $x>0$,

$$\begin{aligned}&\mathbb{P}\Big(|\widehat{\sigma}^{\mathrm{M}}(\{X_t\}_{t=1}^T) - \sigma^{\mathrm{M}}(X)| > x\Big)\\ &\le 2\exp\Big(-\varphi\Big[F\Big\{F^{-1}(q)+\frac{x}{2}\Big\} - q - \frac{1}{T}\Big]\Big) +\\ &\quad 2\exp\Big(-\varphi\Big[q - F\Big\{F^{-1}(q)-\frac{x}{2}\Big\}\Big]\Big) +\\ &\quad \exp\Big(-\varphi\Big[\bar{F}\Big\{\bar{F}^{-1}(q)+\frac{x}{2}\Big\} - q - \frac{1}{T}\Big]\Big) +\\ &\quad \exp\Big(-\varphi\Big[q - \bar{F}\Big\{\bar{F}^{-1}(q)-\frac{x}{2}\Big\}\Big]\Big).\end{aligned}$$

whenever $F\{F^{-1}(q)+x/2\}-q>1/T$ and $\bar{F}\{\bar{F}^{-1}(q)+x/2\}-q>1/T$. Here φ is defined in (C.2).

Proof. We denote $\widehat{m} := \widehat{Q}(\{X_t\}_{t=1}^T; q)$ and $m := Q(X; q)$ to be the sample and population q quantiles. By the definition of $\widehat{\sigma}^{\mathrm{M}}(\cdot)$, we have

$$
\begin{aligned}
&\mathbb{P}\Big\{\widehat{\sigma}^{\mathrm{M}}\Big(\{X_t\}_{t=1}^T\Big) - \sigma^{\mathrm{M}}(X) > x\Big\}\\
&= \mathbb{P}\Big\{\widehat{Q}\Big(\Big\{|X_t - \widehat{m}|\Big\}_{t=1}^T; q\Big) - Q\Big(|X - m|; q\Big) > x\Big\}\\
&\le \mathbb{P}\Big\{\widehat{Q}\Big(\Big\{|X_t - m|\Big\}_{t=1}^T; q\Big) + |\widehat{m} - m| - Q\Big(|X - m|; q\Big) > x\Big\}\\
&\le \mathbb{P}\Big\{\widehat{Q}\Big(\Big\{|X_t - m|\Big\}_{t=1}^T; q\Big) - Q\Big(|X - m|; q\Big) > \frac{x}{2}\Big\} +\\
&\qquad \mathbb{P}\Big(|\widehat{m} - m| > \frac{x}{2}\Big).
\end{aligned}
\tag{C.8}
$$

On the other hand, using the same technique, we have

$$
\begin{aligned}
&\mathbb{P}\Big\{\widehat{\sigma}^{\mathrm{M}}\Big(\{X_t\}_{t=1}^T\Big) - \sigma^{\mathrm{M}}(X) < -x\Big\}\\
&= \mathbb{P}\Big\{\widehat{Q}\Big(\Big\{|X_t - \widehat{m}|\Big\}_{t=1}^T; q\Big) - Q\Big(|X - m|; q\Big) < -x\Big\}\\
&\le \mathbb{P}\Big\{\widehat{Q}\Big(\Big\{|X_t - m|\Big\}_{t=1}^T; q\Big) - |\widehat{m} - m| - Q\Big(|X - m|; q\Big) < -x\Big\}\\
&\le \mathbb{P}\Big\{\widehat{Q}\Big(\Big\{|X_t - m|\Big\}_{t=1}^T; q\Big) - Q\Big(|X - m|; q\Big) < -\frac{x}{2}\Big\} +\\
&\qquad \mathbb{P}\Big(|\widehat{m} - m| > \frac{x}{2}\Big).
\end{aligned}
\tag{C.9}
$$

Combining (C.8) and (C.9), we have

$$
\begin{aligned}
&\mathbb{P}\Big\{|\widehat{\sigma}^{\mathrm{M}}\Big(\{X_t\}_{t=1}^T\Big) - \sigma^{\mathrm{M}}(X)| > x\Big\}\\
&\le \mathbb{P}\Big\{\Big|\widehat{Q}\Big(\Big\{|X_t - m|\Big\}_{t=1}^T; q\Big) - Q\Big(|X - m|; q\Big)\Big| > \frac{x}{2}\Big\} +\\
&\qquad 2\mathbb{P}\Big(|\widehat{m} - m| > \frac{x}{2}\Big).
\end{aligned}
\tag{C.10}
$$

Using Lemma C.1, we have

$$
\begin{aligned}
&\mathbb{P}\Big\{\Big|\widehat{Q}\Big(\Big\{|X_t - m|\Big\}_{t=1}^T; q\Big) - Q\Big(|X - m|; q\Big)\Big| > \frac{x}{2}\Big\}\\
&\le \exp\Big(-\varphi\Big[\bar{F}\Big\{\bar{F}^{-1}(q) + \frac{x}{2}\Big\} - q - \frac{1}{T}\Big]\Big) +\\
&\qquad \exp\Big(-\varphi\Big[q - \bar{F}\Big\{\bar{F}^{-1}(q) - \frac{x}{2}\Big\}\Big]\Big),
\end{aligned}
\tag{C.11}
$$

$$
\begin{aligned}
&\mathbb{P}\Big(|\widehat{m} - m| > \frac{x}{2}\Big)\\
&\le \exp\Big(-\varphi\Big[F\Big\{F^{-1}(q) + \frac{x}{2}\Big\}\Big] - q - \frac{1}{T}\Big) +\\
&\qquad \exp\Big(-\varphi\Big[q - F\Big\{F^{-1}(q) - \frac{x}{2}\Big\}\Big]\Big),
\end{aligned}
\tag{C.12}
$$

whenever $F\{F^{-1}(q) + x/2\} - q > 1/T$ and $\bar{F}\{\bar{F}^{-1}(q) + x/2\} - q > 1/T$. Combining (C.10), (C.11), and (C.12) leads to the desired result. □

Lemma C.3 (Batir (2008)). For any $x > 0$, the following inequalities hold:

$$\sqrt{2e}\Big(\frac{x+1/2}{e}\Big)^{x+1/2} \leq \Gamma(x+1) \leq \sqrt{2\pi}\Big(\frac{x+1/2}{e}\Big)^{x+1/2},$$

where $\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du$ is the Gamma function.

References

Andersen, T. G. (2009). *Handbook of Financial Time Series*. Springer.

Andrews, D. W. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability*, 21(4):930–934.

Bai, J. and Liao, Y. (2013). Statistical inferences using large estimated covariances for panel data and factor models. *arXiv preprint arXiv:1307.2662*.

Bai, J. and Liao, Y. (2016). Efficient estimation of approximate factor models via penalized maximum likelihood. *Journal of econometrics*, 191(1):1–18.

Batir, N. (2008). Inequalities for the Gamma function. *Archiv der Mathematik*, 91(6):554–563.

Belloni, A. and Chernozhukov, V. (2011). ℓ_1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, 39(1):82–130.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2(2):107–144.

Chanda, K. C. (1974). Strong mixing properties of linear stochastic processes. *Journal of Applied Probability*, 11(2):401–408.

Chen, X., Xu, M., and Wu, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, 41(6):2994–3021.

Dedecker, J. and Prieur, C. (2004). Coupling for τ-dependent sequences and applications. *Journal of Theoretical Probability*, 17(4):861–885.

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, 84(2):313–342.

Doukhan, P. and Neumann, M. H. (2007). Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Processes and their Applications*, 117(7):878–903.

Doukhan, P. and Neumann, M. H. (2008). The notion of ψ-weak dependence and its applications to bootstrapping time series. *Probability Surveys*, 5:146–168.

Fan, J., Han, F., and Liu, H. (2014). Challenges of big data analysis. *National Science Review*, 1(2):293–314.

Fan, J., Han, F., Liu, H., and Vickers, B. (2016). Robust inference of risks of large portfolios. *Journal of Econometrics*, 194(2):298–308.

Fan, J., Liao, Y., and Mincheva, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics*, 39(6):3320–3356.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680.

Fan, J., Liao, Y., and Shi, X. (2015). Risks of large portfolios. *Journal of Econometrics*, 186(2):367–387.

Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592–606.

Gorodetskii, V. (1978). On the strong mixing property for linear sequences. *Theory of Probability and Its Applications*, 22(2):411–413.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley.

Han, F. (2018). An exponential inequality for u-statistics under mixing conditions. *Journal of Theoretical Probability*, 31(1):556–578.

Han, F. and Li, Y. (2019). Moment bounds for large autocovariance matrices under dependence. *Journal of Theoretical Probability*, pages 1–48.

Han, F. and Liu, H. (2012). Transelliptical component analysis. In *Advances in Neural Information Processing Systems*, pages 359–367.

Han, F. and Liu, H. (2013a). Principal component analysis on non-Gaussian dependent data. In *Proceedings of the 30th International Conference on Machine Learning*, pages 240–248.

Han, F. and Liu, H. (2013b). Robust sparse principal component regression under the high dimensional elliptical model. In *Advances in Neural Information Processing Systems*, pages 1941–1949.

Han, F. and Liu, H. (2013c). Transition matrix estimation in high dimensional time series. In *Proceedings of the 30th International Conference on Machine Learning*, pages 172–180.

Han, F. and Liu, H. (2014). Scale-invariant sparse PCA on high-dimensional meta-elliptical data. *Journal of the American Statistical Association*, 109(505):275–287.

Han, F. and Liu, H. (2017). Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. *Bernoulli: official journal of the Bernoulli Society for Mathematical Statistics and Probability*, 23(1):23.

Han, F. and Liu, H. (2018). ECA: High-dimensional elliptical component analysis in non-Gaussian distributions. *Journal of the American Statistical Association*, 113(521):252–268.

Han, F., Lu, J., and Liu, H. (2014). Robust scatter matrix estimation for high dimensional distributions with heavy tails. Technical report, Johns Hopkins University.

Han, F. and Wu, W. (2019). Probability inequalities for high dimensional time series under a triangular array framework. *arXiv preprint arXiv:1907.06577*.

Kallabis, R. S. and Neumann, M. H. (2006). An exponential inequality under weak dependence. *Bernoulli*, 12(2):333–350.

Kontorovich, L. A., Ramanan, K., et al. (2008). Concentration inequalities for dependent random variables via the martingale method. *The Annals of Probability*, 36(6):2126–2158.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.

Liu, H., Han, F., and Zhang, C.-h. (2012). Transelliptical graphical models. In *Advances in neural information processing systems*, pages 800–808.

Loh, P.-L. and Wainwright, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics*, 40(3):1637–1664.

Majda, A. and Wang, X. (2006). *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*. Cambridge University Press.

Merlevède, F., Peligrad, M., and Rio, E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High dimensional probability V: the Luminy volume*, pages 273–292. Institute of Mathematical Statistics.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3-4):435–474.

Pan, J. and Yao, Q. (2008). Modeling multiple time series via common factors. *Biometrika*, 95(2):365–379.

Pham, T. D. and Tran, L. T. (1985). Some mixing properties of time series models. *Stochastic Processes and their Applications*, 19(2):297–303.

Purdon, P. L. and Weisskoff, R. M. (1998). Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Human Brain Mapping*, 6(4):239–249.

Qiu, H., Han, F., Liu, H., and Caffo, B. (2015a). Robust portfolio optimization. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, pages 46–54.

Qiu, H., Han, F., Liu, H., and Caffo, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(2):487–504.

Qiu, H., Xu, S., Han, F., Liu, H., and Caffo, B. (2015b). Robust estimation of transition matrices in high dimensional heavy-tailed vector autoregressive processes. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 1843–1851.

Sancetta, A. (2008). Sample covariance shrinkage for high dimensional dependent data. *Journal of Multivariate Analysis*, 99(5):949–967.

Shen, Y., Han, F., and Witten, D. (2020). Exponential inequalities for dependent V-statistics via random Fourier features. *Electronic Journal of Probability*, 25:1–18.

Wang, L., Wu, Y., and Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, 107(497):214–222.

Wang, Z., Han, F., and Liu, H. (2013). Sparse principal component analysis for high dimensional multivariate time series. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics*, pages 48–56.

Wold, H. (1938). A study in the analysis of stationary time series. *Journal of the Royal Statistical Society*, 102(2):295–298.

Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage*, 14(6):1370–1386.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40):14150–14154.

Xiao, H. and Wu, W. B. (2012). Covariance matrix estimation for stationary time series. *The Annals of Statistics*, 40(1):466–493.

Zhou, C., Han, F., Zhang, X., and Liu, H. (2019). An extreme-value approach for testing the equality of large U-statistic based correlation matrices. *Bernoulli*, 25(2):1472–1503.