Submitted to the Annals of Statistics

HYPOTHESIS TESTING FOR HIGH-DIMENSIONAL TIME SERIES VIA SELF-NORMALIZATION

By Runmin Wang*,†, and Xiaofeng Shao*,†

University of Illinois at Urbana-Champaign†

Self-normalization has attracted considerable attention in the recent literature of time series analysis, but its scope of applicability has been limited to low-/fixed-dimensional parameters for low-dimensional time series. In this article, we propose a new formulation of self-normalization for inference about the mean of high-dimensional stationary processes. Our original test statistic is a U-statistic with a trimming parameter to remove the bias caused by weak dependence. Under the framework of nonlinear causal processes, we show the asymptotic normality of our U-statistic with the convergence rate dependent upon the order of the Frobenius norm of the long-run covariance matrix. The self-normalized test statistic is then constructed on the basis of recursive subsampled U-statistics and its limiting null distribution is shown to be a functional of time-changed Brownian motion, which differs from the pivotal limit used in the low-dimensional setting. An interesting phenomenon associated with self-normalization is that it works in the high-dimensional context even if the convergence rate of original test statistic is unknown. We also present applications to testing for bandedness of the covariance matrix and testing for white noise for high-dimensional stationary time series and compare the finite sample performance with existing methods in simulation studies. At the root of our theoretical arguments, we extend the martingale approximation to the high-dimensional setting, which could be of independent theoretical interest.

1. Introduction. In this paper, we study the problem of hypothesis testing for the mean vector of a p-dimensional stationary time series $\{Y_t\}_{t=1}^N$. Mean testing for independent and identically distributed (i.i.d. hereafter) data is a classical problem in multivariate analysis. When the dimension p is fixed as the sample size N grows, Hotelling's T^2 test is a classical one and it enjoys certain optimality properties under Gaussian assumptions [see Anderson (2003), Theorem 5.6.6 (pp. 196)]. There is a recent surge of interest in the high-dimensional setting, where p grows as the sample size $N \to \infty$, motivated by the collection of high-dimensional data from many areas such as biological science, finance and economics, and climate science among others. See Bai and Saranadasa (1996), Srivastava and Du (2008) , Srivastava (2009), Chen and Qin (2010), Lopes et al. (2011), Secchi et al. (2013), Cai et al. (2014), Gregory et al. (2015), Xu et al. (2016), Zhang (2017), Pini et al. (2018) and references cited in these papers. All of these works dealt with i.i.d. data, and the methods and theory developed may not be suitable when the high-dimensional data exhibits serial dependence. High-dimensional data with serial or temporal dependence occurs in many fields, such as large-dimensional panel data in economics, fMRI data collected over time in neuroscience, and spatio-temporal data analyzed in climate studies.

*Runmin Wang is Ph.D. Candidate and Xiaofeng Shao is Professor, Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820 (e-mail: rwang52@illinois.edu, xshao@illinois.edu).

MSC 2010 subject classifications: Primary 62H15, 60K35; Secondary 62G10, 62G20

Keywords and phrases: High-Dimensional Inference, Nonlinear Time Series, Pivotal, Self-Normalization, U-statistic

The focus of this article is on inference for the mean of a high-dimensional time series. When the dimension is low and fixed, several methods have been developed to perform hypothesis testing for the mean of a multivariate time series with weak dependence, e.g. normal approximation with consistent estimation of the long-run covariance matrix [Andrews (1991)], subsampling [Politis and Romano (1994)], moving block bootstrap [Künsch (1989), Liu and Singh (1992)] and variants, blockwise empirical likelihood [Kitamura (1997)] and the self-normalization method [Lobato (2001),Shao (2010)]. When the dimension is high and grows with respect to the sample size, little is known about the validity of the above-mentioned methods. It is worth noting that Jentsch and Politis (2015) showed the asymptotic validity of a multivariate version of the linear process bootstrap [McMurray and Politis (2010)] for inference about the mean when the dimension of a time series is allowed to increase with the sample size. However, the growth rate of p has to be slower than that of the sample size, which rules out the case $p > N$. Recently, Zhang and Wu (2017) considered the problem of approximating the maxima of sums of high-dimensional stationary time series by Gaussian vectors under the framework of functional dependence measure [Wu (2005)]. Their approach, which can be viewed as an extension of Chernozhukov et al. (2013) from the i.i.d. setting to the stationary time series setting, is applicable to tests about the mean of high-dimensional time series. Another related work along this line is Zhang and Cheng (2018), who obtained similar Gaussian approximation results as those presented in Zhang and Wu (2017) but under more stringent assumptions. Note that Zhang and Cheng (2018) used a blockwise multiplier bootstrap as an extension of multiplier bootstrap used in Chernozhukov et al. (2013) to accommodate weak serial dependence, whereas Zhang and Wu (2017) adopted direct estimation of the long-run covariance matrix, which also requires selecting a block size in its batched mean estimate. Both Zhang and Cheng (2018) and Zhang and Wu (2017) also extended their approaches to inference for other quantities beyond the mean, and their theory allows p to grow at either a polynomial or exponential rate as a function of N depending on the moment and dependence assumptions.

In this article, we propose to adopt a U-statistic based approach to the testing problem, extending the work of Chen and Qin (2010), who first proposed to use a U-statistic in a high-dimensional two-sample mean testing problem for independent data. Our U-statistic is however different from the one proposed for i.i.d. data in that we remove pairs of observations that are within m time points of each other, where m is a trimming parameter, to alleviate the bias caused by weak serial dependence. Under the framework of high-dimensional nonlinear causal processes, we show that our U-statistic is asymptotically normal under the null hypothesis. The norming sequence is dependent on the Frobenius norm of the long run covariance matrix (i.e., $\|\Gamma\|_F$), whose rate of divergence is not assumed to be known. To perform the test, one approach is to find a ratio-consistent estimator of $\|\Gamma\|_F$, say using $\|\widehat{\Gamma}\|_F$, so that

$$\frac{\|\widehat{\Gamma}\|_F}{\|\Gamma\|_F} \to 1 \text{ in probability,} \tag{1.1}$$

where $\widehat{\Gamma}$ is the usual lag window estimator. Such an estimator typically involves a bandwidth parameter, and its consistency has been shown in the low and fixed-dimensional context; see Andrews (1991), Newey and West (1987). In the high-dimensional context, Chen and Wu (2019) showed the so-called normalized Frobenious norm consistency, which implies the ratio consistency (1.1), in the context of trend testing. However, no discussion about the choice of the bandwidth parameter seems offered in Chen and Wu (2019) and their result is restricted to linear processes.

To circumvent the difficulty, we take an alternative approach, and our test is based on the idea of self-normalization (SN, hereafter). SN for the mean of a time series was first proposed by Lobato (2001); also see Kiefer et al. (2000) for a related development in the time series regression framework around the same time. Later SN was extended by Shao (2010) and coauthors to various inference problems in time series analysis; See Shao and Zhang (2010), Zhou and Shao (2013), Kim et al. (2015) and Zhang et al. (2011) among others. The basic idea of self-normalization in the time series context is that it uses an inconsistent variance estimator as the studentizer, and the resulting studentized test statistic can still be (asymptotically) pivotal and its limiting null distribution and critical values can be approximated by Monte-Carlo simulations. It has the appealing feature of requiring no tuning parameters for some problems or fewer tuning parameters compared to existing inference procedures, but all existing SN-based methods are limited to inference for a parameter with finite and fixed dimension; see Shao (2015) for a recent review. Here we make the first attempt to extend the idea of SN for inference in high-dimensional time series and for a parameter of high/growing dimension. To this end, we study the weak convergence of a recursive version of our full-sample based U-statistic. Under suitable assumptions, we show that the limiting process is a time-changed Brownian motion, which is different from the standard Brownian motion limit in the application of SN for low-dimensional weakly dependent time series. The limiting null distribution of our SN-based test statistic is still pivotal and its critical values are tabulated via simulations.

One appealing feature of our test statistic is its adaptiveness to the unknown order of $\|\Gamma\|_F$, which gets canceled out in the limit of our self-normalized test statistic. This seems to be discovered for the first time, as the convergence rate is typically known or needs to be estimated in the use of SN for a low-dimensional parameter; see Shao (2015). On the theory side, we extend the martingale approximation argument to the high-dimensional setting. In our result, the dimension p can grow at an exponential rate as a function of N under suitable moment and weak dependence assumptions on the processes. Compared to the maximum type tests proposed by Zhang and Wu (2017), Zhang and Cheng (2018), our test is of L_2 type and it targets dense and weak alternatives, whereas theirs are expected to be more powerful for strong and sparse alternatives. As two important applications, we apply our tests to testing for the bandedness of a covariance matrix and testing for white noise for high-dimensional time series. Finally, we mention a few recent works on inference for high-dimensional time series. Lam and Yao (2012) proposed a static factor model for high-dimensional time series and focused on estimating the number of factors; Basu and Michailidis (2015) investigated the theoretical properties of l_1-regularized estimates in the context of high-dimensional time series and introduced a measure of stability for stationary processes using their spectral properties that provides insight into the effect of dependence on the accuracy of the regularized estimates; Paul and Wang (2016) presented results related to asymptotic behavior of sample covariance and autocovariance matrices of high-dimensional time series using random matrix theory.

The rest of the article is organized as follows. Section 2 introduces the basic problem setting and the notations we use throughout the paper. Section 3 presents our self-normalized statistic as well as related asymptotic results. Section 4 introduces two extensions of the self-normalized statistic to bandedness and white noise testing and Section 5 presents all finite sample simulation results. Section 6 concludes. Finally all the technical details are included in the Appendix and Supplemental Material.

2. Problem Setting.

Assume that we have a p-dimensional stationary nonlinear time series

$$Y_t = \mu + g(\epsilon_t, \epsilon_{t-1}, \cdots)$$

for some measurable function g, where $\{\epsilon_t\}_{t=-\infty}^{\infty}$ are i.i.d. random elements in some measurable space. For the j-th element of Y_t, denoted as $Y_{t,j}$, assume

$$Y_{t,j} = \mu_j + g_j(\epsilon_t, \epsilon_{t-1}, \cdots),$$

where g_j is the jth component of the map g, and $\mu = (\mu_1, \cdots, \mu_p)^T$. We assume $\mathbb{E}[g(\epsilon_t, \epsilon_{t-1}, \cdots)] = 0$. Later we shall introduce suitable weak dependence assumptions under the above framework, which was initially proposed by Wu (2005), who advocated the use of physical dependence measure in asymptotic theory of time series analysis; see Wu (2011) for a review. Our weak dependence condition is characterized by a variant of the Geometric Moment Contraction [see Hsing and Wu (2004), Wu and Shao (2004), Wu and Min (2005)], which was found very useful for studying nonlinear time series and also verifiable for many linear and nonlinear time series models; see Shao and Wu (2007).

Throughout the paper, we let $\Sigma_0 = Var(Y_t)$ denote the marginal covariance matrix and $\Gamma := \sum_{k=-\infty}^{\infty} cov(Y_t, Y_{t+k})$ denote the long-run covariance matrix of Y_t. We define $\mathcal{F}_t = \sigma(\epsilon_t, \epsilon_{t-1}, \cdots, \epsilon_1, \epsilon_0, \epsilon_{-1}, \cdots)$ as the natural filtration generated by $\{\epsilon_t\}$, and define $\mathcal{F}_t' = \sigma(\epsilon_t, \epsilon_{t-1}, \cdots, \epsilon_1, \epsilon_0', \epsilon_{-1}', \cdots)$ where ϵ_t' is an i.i.d. copy of ϵ_t which is independent from $\{\epsilon_t\}_{t\in\mathbb{Z}}$. We use $\|\cdot\|_F$ to denote the Frobenius norm and $\|\cdot\|$ to denote the spectral norm for a matrix (vector). We let $\|\cdot\|_h$ be the L_h norm for random vectors. We define $\mathbb{E}_t(\cdot) := \mathbb{E}(\cdot|\mathcal{F}_t)$ and $\mathbb{E}_t'(\cdot) := \mathbb{E}(\cdot|\mathcal{F}_t')$. For any random element $X_t = X(\mathcal{F}_t)$ which is a function of $\mathcal{F}_t$, we define $X_t' = X(\mathcal{F}_t')$. All asymptotic results are under the regime $\min(N, p) \to \infty$.

Given a stretch of observations Y_t, $t = 1, \cdots, N$, from the above process, we are interested in testing the hypothesis that

$$\mathcal{H}_0 : \mu = \mu_0 \qquad \text{v.s.} \qquad \mathcal{H}_1 : \mu \neq \mu_0. \tag{2.1}$$

Without loss of generality, we let $\mu_0 = 0$. If $\mu_0 \neq 0$, we can apply our test to $\{Y_t - \mu_0\}_{t=1}^N$.

For this testing problem, a very natural test statistic is the distance between $\bar{Y}_N$ and 0, where $\bar{Y}_N = N^{-1}\sum_{t=1}^N Y_t$ is the sample mean. For example, if we use L_2 distance, then

$$\|\bar{Y}_N - 0\|_2^2 = \frac{1}{N^2}\sum_{t=1}^N\sum_{s=1}^N Y_t^T Y_s.$$

However the distribution for the above statistic is not easy to derive, in part because when t and s are close to each other, the correlation between Y_t and Y_s induces a 'bias' term (under the null) that needs to be eliminated by consistent estimation; see Ayyala et al. (2017). Since the auto-correlation can be viewed as a nuisance component for mean inference, we propose to avoid its direct estimation by removing the cross product between observations that are too close to each other in time. To this end, we consider the test statistic

$$T_n = \binom{n+1}{2}^{-1}\sum_{t=1}^n\sum_{s=1}^t Y_{t+m}^T Y_s, \tag{2.2}$$

where $n = N - m$ and $m < N$ is a trimming parameter which satisfies $1/m + m/N = o(1)$ as $\min(p, N) \to \infty$. See Chen and Wu (2019) for a similar trimming idea in testing for the form of the trend of multivariate time series. Let $\theta = \mu^T\mu \in \mathbb{R}^1$ be the scalar parameter of interest. Then $\mu = 0$ is equivalent to $\theta = 0$. Hence T_n can be viewed as a one sample U-statistic for time series; see Lee (1990). The trimming parameter controls the amount of bias since the bias $\mathbb{E}(T_n) - \theta$ depends on m and $tr(\Sigma_h), h = m, m+1, \cdots$, where $\Sigma_h = cov(Y_t, Y_{t+h})$. The larger values of m correspond to smaller bias, which is intimately related to the accuracy of size; the smaller values of m correspond to more pairs of observations used in the test, which can lead to more power. Section 5.1 offers numerical evidence and some discussion of the role of m in detail. It is worth noting that another commonly used distance is $\|\bar{Y}_N\|_\infty$, which has been studied recently in Zhang and Wu (2017) and Zhang and Cheng (2018). See Section 5.1 for some numerical comparison.

Throughout the paper, we use "$\xrightarrow{p}$" to denote convergence in probability and "$\xrightarrow{\mathcal{D}}$" to denote convergence in distribution. Let $\mathcal{D}[0,1]$ be the space of functions on $[0,1]$ which are right continuous and have left limits, endowed with the Skorokhod topology (Billingsley (2008)). Denote by "$\rightsquigarrow$" weak convergence in $\mathcal{D}[0,1]$. We use $A \lesssim B$ to represent that A is less than or equal to cB for some constant $c > 0$.

Under suitable moment and weak dependence assumptions on Y_t, we can show that

$$(n+1)T_n/(\sqrt{2}\|\Gamma\|_F) \xrightarrow{\mathcal{D}} N(0,1)$$

under the null; see Corollary 3.8. This motivates us to define the process

$$T_n(r) := \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} Y_{t+m}^T Y_s, \; r \in [0,1]$$

and study its process convergence in $D[0,1]$. Under the null hypothesis where $\mu_0 = 0$, consider the decomposition

$$Y_t = D_t - \xi_t,$$

where $D_t := \sum_{k=0}^{\infty} [\mathbb{E}_t(Y_{t+k}) - \mathbb{E}_{t-1}(Y_{t+k})]$ and $\xi_t := \widetilde{D}_t - \widetilde{D}_{t-1}$, where $\widetilde{D}_t := \sum_{k=1}^{\infty} \mathbb{E}_t(Y_{t+k})$.

By simple calculation we can show that $(D_t, \mathcal{F}_t)$ is a martingale difference sequence. Martingale approximation for the partial sums of a stationary process has been investigated by Gordin (1969), Hall and Heyde (2014), Wu and Woodroofe (2004), Wu (2007), among others. All these works are done in a low-/fixed-dimensional setting. By contrast, we shall show that it still works for our U-statistic and in the high-dimensional setting. Based on the above decomposition, we write

$$\begin{aligned} T_n(r) &= \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} Y_{t+m}^T Y_s \\ &= \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} (D_{t+m} - \xi_{t+m})^T (D_s - \xi_s) \\ &= \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} D_{t+m}^T D_s - \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} \xi_{t+m}^T D_s - \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} D_{t+m}^T \xi_s + \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} \xi_{t+m}^T \xi_s \\ &= S_n(r) - R_{1,n}(r) - R_{2,n}(r) + R_{3,n}(r), \end{aligned}$$

where $S_n(r) = \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} D_{t+m}^T D_s$, $R_{1,n}(r) = \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} \xi_{t+m}^T D_s$, $R_{2,n}(r) = \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} D_{t+m}^T \xi_s$, and $R_{3,n}(r) = \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} \xi_{t+m}^T \xi_s$. Note that $T_n(r) = 0$ if $r < 1/n$.

REMARK 2.1. It is worth mentioning that a straightforward extension of the SN idea in Lobato (2001) does not really work in the setting $p > N$. To elaborate the idea, we shall briefly review the SN method in Lobato (2001). Let $\mathcal{B}(r)$, $r \in [0,1]$ be the standard Brownian motion and $\mathcal{B}_q(r), r \in [0,1]$ be a q-dimensional vector of independent Brownian motions. Define

$$U_q = \mathcal{B}_q(1)^T J_q^{-1} \mathcal{B}_q(1), \text{ where } J_q = \int_0^1 [\mathcal{B}_q(r) - r\mathcal{B}_q(1)][\mathcal{B}_q(r) - r\mathcal{B}_q(1)]^T dr.$$

The critical values for $U_q, q = 1, \cdots, 20$ have been tabulated by Lobato (2001). For $Y_t \in \mathbb{R}^p$, let $D_N^2 = N^{-2} \sum_{t=1}^{N} \{\sum_{j=1}^{t} (Y_j - \bar{Y}_N)\}\{\sum_{j=1}^{t} (Y_j - \bar{Y}_N)\}^T$ be the $p \times p$ normalization matrix. If p is small and fixed, then under the null and suitable assumptions, we have $N(\bar{Y}_N-\mu_0)^T(D_N^2)^{-1}(\bar{Y}_N-\mu_0)\xrightarrow{\mathcal{D}} U_p$, as $N\to\infty$. The key ingredient is to replace the consistent estimator of Γ, as used in the traditional approach, with the inconsistent estimator D_N^2. Since the normalization factor D_N^2 is proportional to Γ, the nuisance parameter Γ is canceled out in the limiting distribution of the resulting statistic. It is not hard to see that the SN approach is not feasible when $p>N$, since D_N^2 is not invertible in this case. Even when $p<N$, both empirical and theoretical studies suggest that the approximation error grows with the dimension p [Sun (2014)]. So the use of this form of self-normalization can result in a big size distortion when p is comparable to N.

3. Technical Assumptions and Theoretical Results. To facilitate our methodological and theoretical development, we shall introduce some technical assumptions. We first extend the GMC (Geometric Moment Contraction) condition in Hsing and Wu (2004) and Wu and Shao (2004) to the high-dimensional setting.

DEFINITION 3.1. *Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a $p\times d$ matrix-valued stationary process with $Y_t=h(\mathcal{F}_t)$ for some h. It has the Uniform Geometric Moment Contraction (UGMC(k)) property if there exists some positive number k such that*

$$\sup_{i=1,\cdots,p,j=1,\cdots,d}\mathbb{E}[|Y_{0,i,j}|^k]<C<\infty$$

and

$$\sup_{i=1,\cdots,p,j=1,\cdots,d}\mathbb{E}(|Y_{t,i,j}-Y'_{t,i,j}|^k)\le C\rho^t,\ \ t\ge 1$$

for some $0<\rho<1$ and a positive constant C that do not depend on p or d. For vector-valued stationary process, the same definition can be applied by letting $d=1$.

REMARK 3.2. *Define $\mathcal{F}_{t,\{k\}}=\sigma(\epsilon_t,...,\epsilon_{k+1},\epsilon'_k,\epsilon_{k-1},...)$, and it is easy to see that $\mathcal{F}'_t=\mathcal{F}_{t,\{0,-1,...\}}$. Let $Y_{t,\{k\}}=g(\mathcal{F}_{t,\{k\}})$. In Zhang and Wu (2017) and Zhang and Cheng (2018), they defined the functional dependence measure for each component process as*

$$\theta_{t,q,j}=\|Y_{t,j}-Y_{t,j,\{0\}}\|_q=\|Y_{t,j}-g_j(\mathcal{F}_{t,\{0\}})\|_q,$$

and let $\Theta_{m,q,j}=\sum_{t=m}^{\infty}\theta_{t,q,j}$. Throughout these two papers, they imposed conditions on $\Theta_{m,q,j}$. Specifically, Zhang and Cheng (2018) considered a special case where $\max_j\Theta_{m,q,j}\le C\rho^m$ with $\rho\in(0,1)$ and some constant C. Under this condition,

$$\begin{aligned}\|Y_{t,j}-Y'_{t,j}\|_q&=\|Y_{t,j}-Y_{t,j,\{0\}}+Y_{t,j,\{0\}}-Y_{t,j,\{0,-1\}}+Y_{t,j,\{0,-1\}}-Y_{t,j,\{0,-1,-2\}}+\cdots\|_q\\ &\le\|Y_{t,j}-Y_{t,j,\{0\}}\|_q+\sum_{l=0}^{\infty}\|Y_{t,j,\{0,...,-l\}}-Y_{t,j,\{0,...,-l,-(l+1)\}}\|_q\le C\rho^t\left(\sum_{l=0}^{\infty}\rho^l\right)\le\left(\frac{C}{1-\rho}\right)\rho^t,\end{aligned}$$

and $\max_j\|Y_{t,j}-Y'_{t,j}\|_q\le\left(\frac{C}{1-\rho}\right)\rho^t$, which is just the definition of UGMC(q) defined above. Conversely, if we assume UGMC(q), then

$$\begin{aligned}\|Y_{t,j}-Y_{t,j,\{0\}}\|_q&\le\|Y_{t,j}-Y'_{t,j}\|_q+\|Y'_{t,j}-Y_{t,j,\{0\}}\|_q=\|Y_{t,j}-Y'_{t,j}\|_q+\|Y_{t+1,j}-Y'_{t+1,j}\|_q\\ &\le C\rho^t+C\rho^{t+1}=C(1+\rho)\rho^t,\end{aligned}$$

which means $\max_j\Theta_{m,q,j}\le C(1+\rho)\rho^m$. Hence our UGMC assumption is equivalent to that in Zhang and Cheng (2018).

In Zhang and Wu (2017), they defined a so-called "dependence adjusted norm" by letting

$$\|Y_{\cdot j}\|_{q,\alpha} = \sup_{m\geq 0}(m+1)^{\alpha}\Theta_{m,q,j}$$

which is equivalent to the classical L_q norm for i.i.d. data. Further they defined

$$\Psi_{q,\alpha} = \max_{1\leq j\leq q}\|Y_{\cdot j}\|_{q,\alpha} \qquad \text{and} \qquad \Upsilon_{q,\alpha} = \left(\sum_{j=1}^{p}\|Y_{\cdot j}\|_{q,\alpha}^{q}\right)^{1/q},$$

and imposed assumptions on these two quantities. Their weak dependence conditions are in general weaker than ours in the sense that no (uniform) moment assumptions are required for each component series and also algebraic decay of $\Theta_{m,q,j}$ for each j is allowed.

ASSUMPTION 3.3. *Assume that $\{Y_t\}_{t\in\mathbb{Z}}$ are $\mathbb{R}^p$-valued stationary time series with $\mathbb{E}(Y_t) = 0$ and they satisfy*

A.1 $\sup_{1\leq j\leq p}\sum_{k=0}^{\infty}\|\mathbb{E}_0[Y_{k,j}]\|_8 < C$.
A.2 $\{Y_t\}$ *is UGMC(8).*
A.3 $\sum_{h=0}^{\infty}\|\Sigma_h\| = o(\|\Gamma\|_F)$.
A.4 $p^4\rho^m = o(\|\Gamma\|_F^4)$ *and* $1/m + m/N = o(1)$.
A.5 *For* $h = 2,3,4$, $\sum_{j_1,\cdots,j_h=1}^{p}|cum(D_{t_1,j_1},\cdots,D_{t_l,j_l},\widetilde{D}_{t_{l+1},j_{l+1}},\cdots,\widetilde{D}_{t_h,j_h})| = O(\|\Gamma\|_F^h)$, *for any* $l = 0,\cdots,h$.
A.6 $\sum_{j_1,\cdots,j_4=1}^{p}|cov(\Delta_{t_1,j_1,j_2},\Delta_{t_2,j_3,j_4})| = O(\|\Gamma\|_F^4)$ *for any* t_1,t_2, *where* $\Delta_t = \mathbb{E}(D_{t+1}D_{t+1}^T|\mathcal{F}_t)$ *and* $\Delta_{t,i,j}$ *is the* (i,j)*th element of* Δ_t.

Throughout the paper, we use $cum(A_1,\cdots,A_d)$ to denote the joint cumulant of d random variables $A_1,\cdots,A_d$; see page 19 of Brillinger (2001) for a formal definition of the dth order joint cumulant.

REMARK 3.4. Assumption A.1 indicates that $\sup_j\|Y_{0,j}\|_8$, $\sup_j\|D_{0,j}\|_8$ and $\sup_j\|\widetilde{D}_{0,j}\|_8$ are all bounded. To see this, $\mathbb{E}_0[Y_{0,j}] = Y_{0,j}$ and $\|D_{0,j}\|_8 \leq \|Y_{0,j}\|_8 + \|\widetilde{D}_{0,j}\|_8 + \|\widetilde{D}_{-1,j}\|_8$. Moreover,

$$\|\widetilde{D}_{0,j}\|_8 \leq \sum_{k=1}^{\infty}\|\mathbb{E}_0[Y_{k,j}]\|_8 < C.$$

Assumption A.1 can be shown to be implied by Assumption A.2 but since Assumption A.1 was used explicitly at several places, we put it up for the ease of reference. Assumption A.2 can be verified for stationary ARMA processes. The geometric decay rate associated with UGMC condition can actually be relaxed to a polynomial rate, but at the expense of more complicated details. Assumption A.3 effectively restricts the growth rate of $\sum_{h=0}^{\infty}\|\Sigma_h\|$ relative to $\|\Gamma\|_F$, which can be verified for ARMA processes as well; see Section 3.2. Assumption A.4 is the only constraint on the trimming parameter m, and it implies that the bias after trimming is asymptotically negligible. Assumption A.5 can be verified under some mild conditions. See Section 3.2 for the verification for linear process. Furthermore, we have the following result.

PROPOSITION 3.5. *Assumption A.5 can be satisfied if either one of the following is true:*

1.

$$|cum(D_{t_1,j_1},\cdots,D_{t_l,j_l},\widetilde{D}_{t_{l+1},j_{l+1}},\cdots,\widetilde{D}_{t_h,j_h})| \leq C\rho^{\max_k j_k - \min_k j_k} \tag{3.1}$$

for any $t_1 \leq \cdots \leq t_h$, $l = 0,\cdots,h$, *and that all diagonal elements of* Γ *are greater than some positive constant* c_0.

2. The conditional expectations of component processes are q-dependent, i.e., $\mathbb{E}_{t_0}(g_i(\epsilon_{t_1},\epsilon_{t_1-1},\cdots))$ *is independent of* $\mathbb{E}_{s_0}(g_j(\epsilon_{s_1},\epsilon_{s_1-1},\cdots))$ *for any* $t_1 \geq t_0$, $s_1 \geq s_0$, *and* $|i-j| \geq q$, *where* q *is a positive fixed integer which is independent of* n *and* p.

THEOREM 3.6. *Under Assumptions 3.3, we have*

$$\frac{\sqrt{2}}{n\|\Gamma\|_F} S_n(r) \rightsquigarrow \mathcal{B}(r^2) \qquad \text{in } D[0,1].$$

THEOREM 3.7. *Under Assumptions 3.3, we have*

$$\max_{r\in[0,1]} \left| \frac{\sqrt{2}}{n\|\Gamma\|_F} R_{i,n}(r) \right| \xrightarrow{p} 0$$

for $i = 1,2,3$.

Theorems 3.6 and 3.7 suggest that the leading term $S_n(r)$ dominates in $T_n(r)$ and that the remainder terms $R_{i,n}(r), i = 1,2,3$ are asymptotically negligible. Thus the martingale approximation still works in our high-dimensional setting and for our U-statistic.

COROLLARY 3.8. *Under Assumption 3.3, we have* $\frac{\sqrt{2}}{n\|\Gamma\|_F} T_n(r) \rightsquigarrow \mathcal{B}(r^2)$ *in* $D[0,1]$.

We introduce our self-normalizer as

$$W_n^2 = \frac{1}{n}\sum_{k=1}^{n}\left(T_n(k/n) - \frac{k(k+1)}{n(n+1)}T_n(1)\right)^2. \tag{3.2}$$

Then we define our self-normalized test statistic $T_{SN,n}$ as

$$T_{SN,n} := \frac{T_n(1)^2}{W_n^2}. \tag{3.3}$$

THEOREM 3.9. *Under* $\mathcal{H}_0$ *and Assumptions 3.3, we have*

$$T_{SN,n} \xrightarrow{\mathcal{D}} \mathcal{K} := \frac{\mathcal{B}(1)^2}{\int_0^1 (\mathcal{B}(u^2) - u^2\mathcal{B}(1))^2 du}. \tag{3.4}$$

Compared to the use of self-normalization in the low-dimensional setting [Lobato (2001), Shao (2010)], there are some interesting differences we want to highlight. Firstly, due to the use of U-statistics, the limit of the process $T_n(r)$ (after some standardization) is a time-changed Brownian motion and it differs from the Bronwian motion limit for the partial sum process in Lobato (2001) and Shao (2010). Secondly, the null limit of the self-normalized test statistic $\mathcal{K}$ differs from that used in the low-dimensional case. Since it is still pivotal, we can obtain the simulated quantiles for $\mathcal{K}$, as presented in the table below. Thirdly, we had to introduce a trimming parameter m to eliminate the need to estimate autocovariances, which is not needed in the low-dimensional case. Such trimming serves as a bias reduction tool, and it seems necessary to preserve the main feature of self-normalization.

To approximate the theoretical quantiles of $\mathcal{K}$, ($\mathcal{K}_\alpha$ denotes the α quantile of $\mathcal{K}$), we use a sequence of i.i.d. standard normal random variables with length 10^6 to approximate one realization of the standard Brownian motion path. We construct 10^6 Monte-Carlo replicates for this path and then the empirical quantiles for $\mathcal{K}$ are summarized in the following table.

α	0.8	0.9	0.95	0.99	0.995
$\mathcal{K}_\alpha$	18.19	34.15	54.70	118.49	153.94

TABLE 1

Upper quantiles of the distribution $\mathcal{K}$ simulated based on 10^6 Monte Carlo replications

REMARK 3.10. Chen and Qin (2010) first proposed to use a U-statistic in a high-dimensional two sample mean testing problem for independent data and they used normal approximation and a direct ratio-consistent variance estimate; see page 814 for the expression of the variance estimate and their Theorem 2 for the ratio-consistency statement. In comparison, our U-statistic is different from theirs in that (1) we are using a one sample U-statistic; (2) we have to introduce a trimming parameter m to remove pairs of observations that are within m lags to avoid direct estimation of the bias caused by the temporal correlation. Our U-statistic is tailored for weakly dependent time series, see Lee (1990); (3) the nuisance parameter associated with our test statistic is $\|\Gamma\|_F$, for which a ratio-consistent estimator still involves a bandwidth parameter (see Chen and Wu (2019)), whereas the nuisance parameter for the statistic in Chen and Qin (2010) is $\|\Sigma\|_F$ which can be consistently estimated without any tuning parameter. Our self-normalizer is not a consistent estimator of but proportional to $\|\Gamma\|_F$, and the resulting self-normalized test statistic has a pivotal limit under the null.

Another appealing and distinctive feature of the SN-based test in the high-dimensional setting is that the use of self-normalization in the low-dimensional context requires the knowledge of the convergence rate to a certain stochastic process, say standard Brownian motion. However in the high-dimensional setting we present here, we do not know the exact diverging rate of $\|\Gamma\|_F$, but within the self-normalization procedure, this nuisance parameter can be canceled out from both the numerator and the denominator. In other words, the applicability of the SN method is considerably broadened.

3.1. *Limit Theory under a Local Alternative.* Under the alternative, $\mathbb{E}[Y_t] = \mu \neq 0$. Then $T_n(r)$ can be decomposed as

$$\begin{aligned} T_n(r) &= \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} Y_{t+m}^T Y_s = \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} (Y_{t+m} - \mu + \mu)^T (Y_s - \mu + \mu) \\ &= \sum_{t=1}^{\lfloor nr \rfloor} \sum_{s=1}^{t} (Y_{t+m} - \mu)^T (Y_s - \mu) + \binom{\lfloor nr \rfloor + 1}{2} \|\mu\|_2^2 \\ &\quad + \sum_{t=1}^{\lfloor nr \rfloor} t (Y_{t+m} - \mu)^T \mu + \sum_{s=1}^{\lfloor nr \rfloor} (\lfloor nr \rfloor - s + 1)(Y_s - \mu)^T \mu. \end{aligned}$$

THEOREM 3.11. *Under Assumptions 3.3 and the alternative hypothesis $\mathbb{E}[Y_t] = \mu \neq 0$, we have*

1. *If* $\|\mu\| = o(n^{-1/2}\|\Gamma\|_F^{1/2})$*, then*

$$T_{SN,n} \xrightarrow{\mathcal{D}} \frac{\mathcal{B}(1)^2}{\int_0^1 (\mathcal{B}(u^2) - u^2\mathcal{B}(1))^2 du}$$

and $P(T_{SN,n} \geq \mathcal{K}_\alpha) \to \alpha$*. Thus the SN-based test has trivial power asymptotically.*

2. *If* $\sqrt{n}\|\mu\|\|\Gamma\|_F^{-1/2} \to c$*, where* $c \in (0, \infty)$*, then*

$$T_{SN,n} \xrightarrow{\mathcal{D}} \frac{(\mathcal{B}(1) + c^2/\sqrt{2})^2}{\int_0^1 (\mathcal{B}(u^2) - u^2\mathcal{B}(1))^2 du}$$

and $P(T_{SN,n} \geq \mathcal{K}_\alpha) \to \beta \in (\alpha, 1)$*. Thus our test has nontrivial power asymptotically.*

3. *If* $\sqrt{n}\|\mu\|\|\Gamma\|_F^{-1/2} \to \infty$*, then* $T_{SN,n} \xrightarrow{p} \infty$ *and* $P(T_{SN,n} \geq \mathcal{K}_\alpha) \to 1$*. Thus the limiting power is 1.*

REMARK 3.12. Theorem 3.11 suggests that the local neighborhood around the null for which there is a nontrivial power is characterized by $\|\mu\| = cn^{-1/2}\|\Gamma\|_F^{1/2}$. In the special case when $\Gamma = I_p$, $\mu = \delta(1, \cdots, 1)^T$ where $\delta = Cn^{-1/2}p^{-1/4}$, for some $C \neq 0$, existing methods which are designed to test against sparse alternatives fail to detect such dense and faint alternatives; see Cai et al. (2014). By contrast, $T_{SN,n}$ is able to achieve nontrivial power.

3.2. *Linear Processes.* A direct application of the main theorem is to the case of linear processes. Consider the data generating process

$$Y_t = \mu + \sum_{k=0}^{\infty} c_k \epsilon_{t-k}, \tag{3.5}$$

where ϵ_t are i.i.d. p-dimensional innovations with mean 0 and c_k are $p \times p$ coefficient matrices. Applying the martingale approximation, we can obtain by simple calculation that $D_t = C(1)\epsilon_t$ where $C(1) = \sum_{k=0}^{\infty} c_k$ and

$$\widetilde{D}_t = \sum_{j=0}^{\infty} \left(\sum_{k=j+1}^{\infty} c_k \right) \epsilon_{t-j}.$$

This is exactly the well-known Beveridge Nelson (BN) decomposition described in Phillips and Solo (1992). In this case, the long-run covariance matrix is $\Gamma = C(1)\Sigma_\epsilon C(1)^T$, where $\Sigma_\epsilon = Var(\epsilon_0)$.

ASSUMPTION 3.13. *Assume that* $\{Y_t\}$ *is generated from (3.5) with* $\mu = 0$ *and that*

B.1 $\sup_{1\leq j\leq p} \|\epsilon_{t,j}\|_8 < C$.
B.2 $\sum_{k=m}^{\infty} k\|c_k\| < C\rho^m$ *for some positive constant* C *and* $0 < \rho < 1$.
B.3 $\sum_{h=0}^{\infty} \|\Sigma_h\| = o(\|\Gamma\|_F)$.
B.4 $p^4\rho^m = o(\|\Gamma\|_F^4)$ *and* $1/m + m/N = o(1)$.
B.5 For $h = 2, 3, 4$, $\sum_{j_1,\cdots,j_h=1}^{p} |cum(D_{t_1,j_1}, \cdots, D_{t_l,j_l}, \widetilde{D}_{t_{l+1},j_{l+1}}, \cdots, \widetilde{D}_{t_h,j_h})| = O(\|\Gamma\|_F^h)$*, for any* $l = 0, \cdots, h$.

COROLLARY 3.14. *Under Assumptions 3.13, we have*

$$\frac{\sqrt{2}}{n\|\Gamma\|_F} T_n(r) \rightsquigarrow \mathcal{B}(r^2) \qquad in\ D[0,1].$$

The assumptions B.3 and B.5 can be verified for many weakly dependent time series models. In the following proposition, we shall present some more primitive assumptions for the vector AR(1) model, i.e., $Y_t = AY_{t-1} + \epsilon_t$, for $t \in \mathbb{Z}$. For simplicity, we assume A to be symmetric and ϵ_t to be i.i.d. p-dimensional random vectors with mean zero and covariance matrix Σ_ϵ.

PROPOSITION 3.15. *Assume that Y_t are generated from a $VAR(1)$ model satisfying*

1. $|\Gamma_{i,i}| > c_0 > 0$ *for some positive constant* c_0 *and all* $i = 1, \cdots, p$*;*
2. $\|\Sigma_\epsilon\| = o(\|\Gamma\|_F)$*,*
3. $\limsup_{p\to\infty} \|A\| < c_1 < 1$ *for some positive constant* c_1*.*

Then B.3 can be verified. Furthermore, if we substitute condition 3 with

4. $\limsup_{p\to\infty} \|A\|_1 < c_1 < 1$ *for some positive constant* c_1*,*

and in addition assume

5. $\sum_{k_1,\cdots,k_h=1}^{p} |cum(\epsilon_{0,k_1}, \cdots, \epsilon_{0,k_h})| = O(\|\Gamma\|_F^h)$*, for* $h = 2, 3, 4$*,*

then B.5 holds.

REMARK 3.16. Under the conditions in Proposition 3.15, it is easy to see that the order of $\|\Gamma\|_F$ is between $\sqrt{p}$ and p. When $\|\Gamma\|_F$ is of order p, theoretically we do not have any explicit restriction on the growth rate of p as a function of n (or N). In this case, the condition B.4 holds as long as the trimming parameter m grows to infinity but slower than N. When $\|\Gamma\|_F$ is of order $\sqrt{p}$, we can allow the order of p to be e^{n^β}, for any $\beta \in (0, 1)$, by choosing m to be of order n^γ, where $\gamma \in (\beta, 1)$. Condition 5 in Proposition 3.15 basically restricts the coordinate dependence of the innovation sequence ϵ_t. If the components of ϵ_t satisfy certain m-dependence or geometric moment contraction condition [see Wu and Shao (2004)], then $\sum_{k_1,\cdots,k_h=1}^{p} |cum(\epsilon_{0,k_1}, \cdots, \epsilon_{0,k_h})| = O(p)$, so condition 5 is satisfied.

4. Applications.

The SN-based test can be extended to test the bandedness of the covariance matrix of high-dimensional time series (HDTS). Assume that we have a stationary s-dimensional time series $(X_t)_{t\in\mathbb{Z}}$ with $\mathbb{E}(X_t) = 0$ for notational simplicity (we can apply our method to demeaned data in practice). For high-dimensional temporally-independent data, the covariance matrix $\Sigma = Cov(X_t, X_t) = (\gamma_{jk})_{j,k=1,\cdots,s}$ is an important measure of the dependence among components of X_t, and for time series, it measures the contemporaneous component-wise dependence. In this section, we slightly abuse the notation and use Σ to denote the covariance matrix of X_t, Σ_h to denote the autocovariance matrix of X_t at lag h. Qiu and Chen (2012) first developed a test for bandedness of Σ, motivated by promising results regarding banding and tapering the sample covariance in estimating Σ; see Bickel and Levina (2008), Cai et al. (2010) among others. Specifically, for a given bandwidth L, they test

$$H_{L,0} : \Sigma = B_L(\Sigma), \text{ versus } H_{L,1} : \Sigma \neq B_L(\Sigma)$$

where $B_L(\Sigma) = (\gamma_{jk}\mathbf{I}\{|j-k| \leq L\})_{s\times s}$ is a banded version of Σ with bandwidth L. Note that diagonal matrices are the simplest among banded matrices, and testing for Σ being diagonal (or the so-called sphericity hypothesis in classical multivariate analysis) in the high-dimensional setting has been considered in Ledoit and Wolf (2002), Jiang (2004), Schott (2005), Chen et al. (2010) and Cai and Jiang (2011), among others. All of the above works are for independent data, and they seem no longer applicable to HDTS due to temporal dependence. As a practical motivation, we note that in the analysis of fMRI functional connectivity for brain networks in the format of multivariate time series, Σ has been used to characterize functional connectivity; see Hutchison et al. (2013). As pointed out by a referee, Liu et al. (2018) studied the sphericity hypothesis testing in the context of high-dimensional time series. However, their test statistic seems infeasible as they assumed certain unknown quantities in their test to be known and did not offer any consistent estimates for these unknown quantities.

To test $H_{L,0}$, we let $X_t = (X_{t1}, \cdots, X_{ts})^T$ and $\gamma_{jk} = Cov(X_{tj}, X_{tk})$ for $j,k = 1,\cdots,s$. Further let $\mathcal{I} = \{(j,k) : |j-k| > L, j > k\}$. Then the null hypothesis $H_{L,0}$ is equivalent to $0 = \gamma = (\gamma_{jk})_{(j,k)\in\mathcal{I}} \in \mathbb{R}^{P_L}$, where $P_L = (s-L)(s-L-1)/2$. Let $Z_{t,jk} = X_{tj}X_{tk}$, and $Y_t = (Z_{t,jk})_{(j,k)\in\mathcal{I}} \in \mathbb{R}^{P_L\times 1}$. Then we can formulate this as a testing-many-means problem based on the transformed observations $(Y_t)_{t=1}^N$.

In addition, we can also apply the SN-based test to testing the white noise hypothesis for HDTS. Testing for white noise is an important problem in time series analysis and it is indispensable in diagnostic checking for linear time series modeling. There is a huge literature for univariate and low-dimensional vector time series; see Li (2004) for a review of the literature of univariate time series and Hosking (1980), Li and McLeod (1981) and Lütkepohl (2005), among others for the diagnostic checking methods for vector time series. The literature on white noise testing for high-dimensional time series is quite recent. Chang et al. (2017) proposed to use maximum of absolute autocorrelations and cross correlations of component series as a test statistic and its null distribution is approximated by Gaussian approximation [Chernozhukov et al. (2013)]. Li et al. (2018) used the sum of squares of the eigenvalues in the symmetrized sample autocovariance matrix at a certain lag, and the limiting null distribution is derived using tools from random matrix theory. Specifically, they both test $H_{0,d} : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_d = 0$, where d is a fixed and pre-specified lag and $\Sigma_h = Cov(X_t, X_{t-h}) = (\gamma_{h,jk})_{j,k=1,\cdots,s}$, where $\gamma_{h,jk} = Cov(X_{tj}, X_{(t-h)k}) = \mathbb{E}(Z_{t,h,jk})$ and $Z_{t,h,jk} = X_{tj}X_{(t-h)k}$. Let $Y_{t,h} = (Z_{t,h,jk})_{j,k=1,\cdots,p} = (Z_{t,h,11}, Z_{t,h,12}, \cdots, Z_{t,h,1s}, Z_{t,h,21}, Z_{t,h,22}, \cdots, Z_{t,h,2s}, \cdots, Z_{t,h,s1}, Z_{t,h,s2}, \cdots, Z_{t,ss})^T$, which is an $s^2\times 1$ vector. Then $\Sigma_h = 0$ is equivalent to $\mathbb{E}(Y_{t,h}) = 0$, and $H_{0,d}$ can be tested using our SN-based method and the new data sequence $(Y_t)_{t=d+1}^N$, where $Y_t = (Y_{t,1}^T, \cdots, Y_{t,d}^T)^T$.

5. Simulation Results.

In this section, we investigate the finite sample performance of SN-based methods for mean testing, bandedness testing of covariance matrices, and white noise testing in subsections 5.1, 5.2 and 5.3, respectively.

5.1. *Mean Inference.*

In this subsection, we study the finite sample performance of the proposed method for mean inference. Consider the data generating process

$$Y_t - \mu = A(Y_{t-1} - \mu) + \epsilon_t,$$

which is a p-dimensional VAR(1) model. Here $\mu = \mathbb{E}[Y_t]$ and the innovation sequence $\{\epsilon_t\}$ are i.i.d. according to a multivariate normal distribution with mean 0 and covariance Σ_ϵ where $\Sigma_\epsilon^{1/2}$ is a tri-diagonal matrix with diagonal elements all equal to 1, and the first off-diagonal entries all equal to 0.5. We consider two sample sizes, $N \in \{120, 480\}$ and three dimensions, $p \in \{50, 100, 200\}$. For the coefficient matrix A, we simply let $A = \rho I_p$ and pick $\rho \in \{0.2, 0.8, -0.5\}$.

Under the null hypothesis, μ is simply a vector of zeroes. We let $\mu = 0.8\times(1/\sqrt{p}, 1/\sqrt{p}, \cdots, 1/\sqrt{p})^T$ under the alternative. We include four methods and ten statistics in the simulation: (1) Self-Normalized Statistic ($m = 5, 10, 20, 30$) (Denoted as "SN(5)", "SN(10)", "SN(20)", "SN(30)"); (2) the test proposed in Ayyala et al. (2017). Note that their test assumed the q-dependence for the data generating process but in practice we typically do not know the value of q. We shall set $q = 5$ and 10 here so the test is denoted as "AY(5)" and "AY(10)", respectively; (3) the approach proposed in Zhang and Cheng (2018) with the block size used in the block bootstrap $b_{ZC} = \{10, 20\}$, denoted as "ZC(10)" and "ZC(20)"; (4) the approach proposed in Zhang and Wu (2017) with the block size used in batched mean estimate $b_{ZW} = \{10, 20\}$, denoted as "ZW(10)" and "ZW(20)". Note that there seems no data driven formula for the block size used in Zhang and Cheng (2018) and Zhang and Wu (2017), and it is indeed an open problem on how to select the optimal block size in the high-dimensional setting. For the choice of our trimming parameter m, we shall let $m = 5, 10, 20, 30$ and leave the detailed discussion on its role later.

We set the nominal level as 5% and perform 1000 Monte Carlo simulations for $N = 120$ and $N = 480$. The computation for the test in Ayyala et al. (2017) is very expensive so only the result for the case $N = 120$ is shown here. The results are summarized in Tables 2 and 3. Under the null hypothesis, SN has an accurate size mostly when the dependence is weak (i.e., $\rho = 0.2$). When the dependence gets stronger (i.e., $\rho = 0.8$), there are some fairly large size distortion corresponding to $m = 5$, which is likely due to the bias incurred by using a small m, and the size corresponding to larger m (i.e., $m = 20, 30$) appears much better. When $\rho = -0.5$, there are slight conservativeness in the size of SN test, but most are quite close to nominal level, especially when $m = 10, 20, 30$ and $N = 480$. By contrast, ZC and ZW showed much more severe size distortion, especially in the (relatively) strong dependent case (i.e., $\rho = 0.8$). For both block sizes 10 and 20, ZW method appears to fail to provide a reasonable size in almost all cases, whereas ZC method seems to perform better when the block size is 20, although the size appears too liberal when $\rho = 0.8$ and too conservative when $\rho = 0.2$ and -0.5. Also we can observe the sensitivity of both ZW and ZC with respect to the block size, the choice of which seems to be an open problem in the high-dimensional setting. The test AY(5) exhibits huge size distortion when $\rho = 0.8$, which is presumably due to the fact that $q = 5$ is too small, whereas AY(10) shows much improvements although it is still quite oversized. In the case $\rho = 0.2$, there are some noticeable size distortions with AY(5) and again AY(10) exhibits more accurate size. Overall the size of SN-based test seems much more satisfactory and stable than ZC and ZW, and outperforms that of AY slightly.

As seen from Table 3, which presents the power, SN-based test exhibits highest power when $\rho = -0.5$, and the power in the case $\rho = 0.8$ is quite low. This can be explained by the fact that in the limit the power is a monotonic increasing function of $\sqrt{n}\|\mu\|\|\Gamma\|_F^{-1/2}$, which takes the largest value when $\rho = -0.5$ and admits the smallest value when $\rho = 0.8$. The powers for ZW and ZC are a bit hard to interpret due to the strong over-rejection under the null hypothesis. One can present size-adjusted power, but given the severely distorted size we decide not to pursue this. The tests by Ayyala et al. (2017) exhibit higher power than SN-based tests in almost all cases and the power gain appears quite moderate in some cases. This might suggest that if we can completely remove the bias caused by weak temporal dependence and choose the tuning parameter properly, the normal approximation can work reasonably well, outperforming the SN-based test in power. This is consistent with the "better size but less power" phenomenon observed for SN-based test as compared to normal approximation in the low-dimensional setting; see Shao (2010). It is also worth noting that for our SN-based test the powers corresponding to $m = 20, 30$ are a little lower than that for $m = 10$ in most cases, and the power for $m = 30$ is comparable or slightly less than that for $m = 20$. As we increase m, we expect less power as there are fewer pairs of observations used in the test.

Discussion on the role of m: The trimming parameter m plays an important role in balancing the tradeoff between size distortion and power loss. If m is too small, then the bias might not be negligible especially in the strongly dependent case and this could lead to a big size distortion. If m is too large, then the effective sample size, which is proportional to $N - m$, is less than the optimal level, which could result in power loss. In general, it would be desirable to come up with a data-driven formula for m that is adaptive to the magnitude of serial dependence, which may require a theoretical characterization of the leading term for the bias and the asymptotic power function. An empirical way of choosing m is to visualize the auto and cross correlations of the time series at hand, and choose a m such that majority of auto and cross-correlations are smaller than some threshold for lags beyond m. Our simulation experience suggests that the size and power performance are relatively stable over a certain range $[m_0, m_1]$, which might suggest that the optimal choice of m is not that critical, as long as we choose a m that is in the suitable interval. A careful study of this issue is beyond the scope of this paper and is left for future investigation.

ρ	N	p	SN(5)	SN(10)	SN(20)	SN(30)	AY(5)	AY(10)	ZC(10)	ZC(20)	ZW(10)	ZW(20)
$\rho = 0.2$	120	50	4.5	5.5	5.7	5.5	5.5	5.6	4.6	4.9	43.4	84.5
		100	5.3	5.7	5.9	5.7	5.0	4.1	3.1	2.7	59.2	96.9
		200	6.5	7.0	6.5	7.5	6.5	3.6	2.4	1.8	74.5	100.0
	480	50	4.9	4.9	5.6	5.3			5.2	3.9	13.2	20.7
		100	5.4	5.4	5.5	5.8			4.6	2.8	17.9	32.1
		200	5.5	5.4	5.5	5.8			4.5	2.8	17.9	32.1
$\rho = 0.8$	120	50	21.1	9.5	8.7	9.4	39.3	11.2	26.9	10.1	80.5	93.2
		100	25.7	11.9	9.8	9.3	52.9	11.8	28.6	7.5	94.5	99.4
		200	33.2	15.4	11.7	10.4	74.6	14.4	30.4	7.1	99.7	100.0
	480	50	14.4	5.0	5.1	5.9			33.1	10.6	51.4	36.7
		100	21.4	7.0	5.9	5.5			40.1	10.7	64.3	50.2
		200	23.8	10.8	4.6	4.4			48.1	13.7	76.4	62.8
$\rho = -0.5$	120	50	5.2	4.8	3.8	3.7	12.5	7.8	1.1	1.4	30.4	79.8
		100	5.2	3.4	3.2	3.4	12.5	4.5	0.8	1.1	41.5	95.5
		200	7.7	2.7	4.2	2.6	13.5	5.0	0.7	0.8	55.2	99.8
	480	50	7.1	5.4	5.0	5.5			1.8	2.3	6.5	13.4
		100	7.8	4.3	5.1	4.6			1.9	2.6	7.4	18.0
		200	10.0	5.1	5.0	5.4			1.1	1.2	6.1	22.4

TABLE 2

Empirical Rejection Rate (in percentage) for the mean testing ($\mathcal{H}_0$)

ρ	N	p	SN(5)	SN(10)	SN(20)	SN(30)	AY(5)	AY(10)	ZC(10)	ZC(20)	ZW(10)	ZW(20)
$\rho = 0.2$	120	50	30.6	30.9	27.2	26.1	64.3	61.0	20.1	16.3	76.0	95.8
		100	20.7	20.2	17.9	18.8	46.2	37.5	11.3	8.9	78.4	99.2
		200	16.5	16.3	14.3	13.5	28.4	22.0	5.3	3.9	85.1	99.9
	480	50	94.4	94.1	94.4	94.2			82.7	77.2	93.2	96.4
		100	87.5	87.6	87.1	87.0			54.5	45.5	78.6	87.3
		200	78.5	78.4	77.9	77.1			31.5	23.2	64.7	80.7
$\rho = 0.8$	120	50	23.0	10.3	8.9	9.3	44.4	14.0	29.8	12.0	84.7	93.8
		100	28.8	13.7	9.6	9.0	56.8	14.4	31.5	8.9	93.9	99.1
		200	33.8	17.0	11.9	10.5	76.8	15.6	30.9	7.9	99.4	100.0
	480	50	26.1	12.7	9.0	9.2			46.9	20.6	63.8	49.5
		100	29.8	12.8	8.0	8.0			47.6	17.8	72.0	56.4
		200	29.2	14.6	8.1	7.0			49.7	14.5	80.1	56.3
$\rho = -0.5$	120	50	85.4	86.3	86.5	81.0	99.5	98.9	44.4	46.0	95.2	99.9
		100	68.8	78.4	74.8	67.7	98.4	94.1	15.7	18.0	91.4	100.0
		200	41.0	61.1	57.5	50.5	95.0	81.0	5.4	5.3	89.0	100.0
	480	50	100.0	100.0	100.0	100.0			100.0	100.0	100.0	100.0
		100	100.0	100.0	100.0	100.0			97.3	97.5	99.8	100.0
		200	99.9	100.0	100.0	100.0			67.8	69.4	94.1	99.5

TABLE 3

Empirical Rejection Rate (in percentage) for the mean testing ($\mathcal{H}_1$)

5.2. *Testing the Bandedness of Covariance Matrix.* In this subsection we shall present the simulation result for testing the bandedness of a covariance matrix. We modify the model assumptions in Qiu and Chen (2012) by allowing temporal dependence. In particular, we generate p-dimensional X_t from the model

$$X_{t,i} = \sum_{l=0}^{k_0} \gamma_l Z_{t,i-l} + \delta X_{t-1,i},$$

where k_0 is the bandwidth of the covariance matrix, $\gamma_0 = 1$ for all cases and other coefficients γ_l will be specified later on. We let $\delta \in \{0, 0.4\}$ and sample $Z_{t,i}$ independently from $N(0,1)$. Notice that when $\delta = 0$, the observations X_t are i.i.d. We choose the sample size $N \in \{20, 50, 100\}$ and the dimension $p \in \{20, 60\}$.

We calculate the statistic proposed in Qiu and Chen (2012), denoted as T_{QC} and compare with our test statistic, denoted as T_{SN}. Note that T_{QC} requires $X_t's$ to be i.i.d. whereas T_{SN} does not, thus when $\delta \neq 0$, we should expect an impact on the size of T_{QC}. There is no tuning for T_{QC} and for T_{SN} we set the trimming parameter $m = 10$. Under the null hypothesis, we consider three cases for the bandwidth $k_0 \in \{0, 2, 5\}$. For $k_0 = 2$, we let $\gamma_1 = 0.5, \gamma_2 = 0.25$, and for $k_0 = 5$, we let $\gamma_1 = \cdots = \gamma_5 = 0.4$. To examine the power, we let $k_0 = 2, 5$ and test the null hypothesis that $\Sigma = B_{k_0-2}(\Sigma)$. Those coefficients are the same as those with the same k_0 in evaluating the size. We set the nominal level as 5% and run the experiment for 1000 times and record the empirical rejection rate.

The results under the null are summarized in Table 4. For the i.i.d. case, both methods provide reasonably accurate empirical sizes. It is worth noting that the T_{QC} under-rejects the null when both N and p are small for $\mathcal{H}_0 : \Sigma = B_5(\Sigma)$. When data are weakly dependent, we observe that T_{QC} fails for all cases with a huge size distortion. This is somewhat expected as T_{QC} strongly relies on the i.i.d. assumption. In comparison, T_{SN} still delivers very good size in most cases, except for some size distortion when the sample size N is too small, i.e., $N = 20$.

Table 5 shows the power under the alternative. For i.i.d. case T_{QC} has higher power than T_{SN} for all cases. The power gain of T_{QC} over T_{SN} seems to diminish as sample size increases from $N = 20$ to $N = 100$. For weakly dependent data, although T_{QC} has a very high empirical rejection rate in most cases, it should not be taken too seriously because of the huge size distortion under the null. Further, it is noted that the power for T_{SN} is only slightly lower than the i.i.d. case which indicates T_{SN} still works under the weakly dependent case. In summary, T_{SN} provides a robust alternative to T_{QC}, which is specifically designed for the i.i.d. data.

	i.i.d.						weakly dependent					
	T_{QC}			T_{SN}			T_{QC}			T_{SN}		
	N			N			N			N		
p	20	50	100	20	50	100	20	50	100	20	50	100
						$H_0 : \Sigma = B_0(\Sigma)$						
20	6.3	5.6	5.5	6.3	4.8	4.2	84.2	90.2	91.8	9.0	5.6	5.3
60	6.7	6.0	5.2	6.7	5.2	4.8	100.0	100.0	100.0	10.1	6.2	5.2
						$H_0 : \Sigma = B_2(\Sigma)$						
20	4.8	4.3	4.1	6.3	4.5	4.8	47.7	52.4	55.2	6.6	5.3	5.1
60	6.6	5.5	5.3	6.1	5.3	4.9	99.9	100.0	100.0	9.3	5.7	5.4
						$H_0 : \Sigma = B_5(\Sigma)$						
20	2.6	2.1	2.2	4.6	4.4	4.8	12.3	13.2	15.4	5.9	5.6	5.0
60	4.9	4.4	4.3	5.9	4.8	4.5	85.4	91.8	93.2	6.8	5.5	4.8

TABLE 4
Empirical Rejection Rate (in percentage) for Testing the Bandedness (under null)

5.3. *Testing for White Noise.* In this subsection we investigate the finite sample properties of our test for white noise. For the trimming parameter we fix $m = 10$. The nominal level is set as 5%,

	i.i.d.						weakly dependent					
	T_{QC}			T_{SN}			T_{QC}			T_{SN}		
	N			N			N			N		
p	20	50	100	20	50	100	20	50	100	20	50	100
	$H_0 : \Sigma = B_0(\Sigma)$ when $\Sigma = B_2(\Sigma)$											
20	97.7	100.0	100.0	24.0	97.4	100.0	100.0	100.0	100.0	18.6	92.6	99.9
60	98.8	100.0	100.0	26.7	99.3	100.0	100.0	100.0	100.0	20.4	96.1	100.0
	$H_0 : \Sigma = B_3(\Sigma)$ when $\Sigma = B_5(\Sigma)$											
20	19.7	76.3	99.9	6.0	43.1	88.7	52.0	94.1	100.0	6.7	30.0	76.0
60	29.2	84.5	100.0	5.7	40.9	93.2	98.7	100.0	100.0	6.4	26.6	76.5

TABLE 5

Empirical Rejection Rate (in percentage) for Testing the Bandedness (under alternative)

and we take $N \in \{75, 150, 300\}$ and $p \in \{50, 100\}$. For each experiment we have 1000 Monte-Carlo replicates.We compare our test statistic T_{SN} with the test statistic T_C developed in Chang et al. (2017) (with time series PCA), which targets the sparse alternative and has been implemented in the R package "HDtest".

To examine the size, we generate the data from the model $\epsilon_t = Az_t$, which is the same as the setting considered in Chang et al. (2017), where A is $p \times p$ and z_t are p-dimensional i.i.d. from $N(0, I_p)$. For different loadings,

$\mathcal{M}_1$: Let $S = (s_{kl})_{1\le k,l\le p}$ for $s_{kl} = 0.995^{|k-l|}$ and then let $A = S^{1/2}$,
$\mathcal{M}_2$: Let $A = (a_{kl})_{1\le k,l\le p}$ with the a_{kl} being independently generated from $U(-1, 1)$.

To evaluate the power, we let $k_0 = 12$ and generate the data from the model

$\mathcal{M}_3$: $\epsilon_t = A\epsilon_{t-1} + e_t$ where $\{e_t\}_{t\ge1}$ are i.i.d. p-dimensional random vectors with independent components from t_8 distribution. For the coefficient matrix A, we let $A_{i,j}$ from $U(-0.25, 0.25)$ independently for any $1 \le i, j \le k_0$ and $A_{i,j} = 0$ otherwise.

$\mathcal{M}_4$: $\epsilon_t = Az_t$, where $z_t = (z_{t,1}, \cdots, z_{t,p})^T$. For $1 \le k \le k_0$, $z_{1,k}, \cdots, z_{N,k}$ are $N(0, \Sigma)$, where Σ is $N \times N$ matrix with 1 as the main diagonal elements, $0.5|i-j|^{-0.6}$ as the (i, j)-th element for $1 \le |i-j| \le 7$, and 0 for other elements. For $k > k_0$, $z_{1,k}, \cdots, z_{N,k}$ are independent standard normal random variables. The coefficient matrix A is generated as follows: $a_{k,l} \sim U(-1, 1)$ with probability 1/3 and $a_{k,l} = 0$ with probability 2/3 independently for $1 \le k \ne l \le p$, and $a_{k,k} = 0.8$ for $1 \le k \le p$.

$\mathcal{M}_5$: $\epsilon_t = A\epsilon_{t-1} + e_t$ where $\{e_t\}_{t\ge1}$ are i.i.d. p-dimensional random vectors with independent components from t_8 distribution. For the coefficient matrix A, we let $A_{i,j} = 0.9^{|i-j|}$ for an AR(1) type structure, then we normalize A so that $\|A\| = 0.9$.

Note that the models $\mathcal{M}_1 - \mathcal{M}_4$ were used in Chang et al. (2017) whereas $\mathcal{M}_5$ is added to examine the behavior of our test in the case of dense alternative. Results are summarized in Table 6. Under the null hypothesis ($\mathcal{M}_1$ & $\mathcal{M}_2$), T_{SN} has an accurate and stable empirical rejection rate, comparing to the designed nominal level 5%. T_C tends to under-reject the null a lot, especially when N is small. We notice that the empirical rejection rate of T_C is not stable. For example, it over-rejects the null under $\mathcal{M}_2$ with $p = 100$ and $N = 300$ but the size appears quite accurate when $p = 100$ and $N = 150$. This phenomenon may be due to the bootstrap procedure used in their test which involves a choice of block size, and a sound data-driven choice seems difficult in the high-dimensional setting.

		$N=75$				$N=150$				$N=300$			
		$p=50$		$p=100$		$p=50$		$p=100$		$p=50$		$p=100$	
	lag d	T_{SN}	T_C	T_{SN}	T_C	T_{SN}	T_C	T_{SN}	T_C	T_{SN}	T_C	T_{SN}	T_C
$\mathcal{M}_1$	2	3.8	0.0	3.0	0.0	3.1	0.7	3.2	0.8	3.0	3.6	3.7	2.3
	4	3.8	0.0	2.2	0.0	3.1	0.7	3.9	0.2	3.4	4.1	3.7	2.9
	6	3.0	0.0	2.2	0.0	3.2	1.0	3.5	0.2	2.6	3.7	3.0	2.7
	8	1.8	0.0	3.4	0.0	3.9	0.7	3.4	0.2	3.2	3.2	2.7	2.8
	10	2.6	0.0	1.6	0.0	3.8	0.7	3.3	0.1	2.6	2.6	4.3	2.8
$\mathcal{M}_2$	2	4.8	0.1	4.1	0.4	4.5	1.2	6.0	5.9	4.8	4.2	6.0	12.3
	4	4.7	0.0	5.1	0.3	5.4	1.2	4.4	4.9	5.0	3.1	4.6	13.3
	6	4.9	0.1	4.1	0.2	5.4	0.8	5.1	4.6	4.0	3.6	5.6	14.6
	8	2.4	0.0	3.3	0.2	4.8	0.6	5.1	3.8	5.9	3.3	6.1	14.6
	10	1.8	0.0	1.2	0.1	4.3	0.5	5.2	2.8	5.2	3.6	4.6	14.0
$\mathcal{M}_3$	2	10.6	0.8	0.0	0.0	29.2	39.9	15.5	10.3	66.2	100	34.3	98.0
	4	5.4	0.0	2.4	0.0	6.1	27.7	6.0	6.0	6.9	99.8	5.1	94.8
	6	4.4	0.0	2.4	0.0	5.3	19.8	6.0	4.5	5.6	99.5	5.4	91.5
	8	3.4	0.0	2.4	0.0	4.7	15.2	5.3	3.5	5.5	99.0	5.6	88.8
	10	1.4	0.0	2.4	0.0	3.7	13.3	4.7	3.0	6.0	98.7	4.6	86.8
$\mathcal{M}_4$	2	11.2	1.6	5.3	0.4	22.2	14.0	9.1	7.1	57.9	63.1	20.0	30.1
	4	5.7	1.3	5.5	0.1	8.4	8.7	6.0	6.0	19.8	53.9	8.3	26.1
	6	4.9	0.6	4.4	0.1	6.6	6.6	5.5	5.0	10.2	48.4	6.7	24.1
	8	3.5	0.3	2.8	0.0	4.5	5.4	5.5	4.0	5.9	44.2	4.9	22.1
	10	2.0	0.2	2.0	0.0	4.5	4.2	4.7	3.2	5.1	41.2	4.8	19.6
$\mathcal{M}_5$	2	35.5	2.5	28.1	5.8	55.1	4.6	62.7	3.2	89.0	44.8	94.2	24.4
	4	16.7	1.7	12.9	4.8	17.4	3.1	13.8	2.0	34.6	33.9	34.0	18.9
	6	10.7	1.3	8.2	3.8	6.3	1.3	6.5	1.4	12.2	28.6	10.5	14.8
	8	6.5	0.7	7.1	4.3	6.1	0.6	4.3	0.9	6.5	23.1	5.7	11.9
	10	3.5	0.4	5.3	3.8	4.9	0.5	4.7	0.5	4.0	19.8	5.5	9.6

TABLE 6

Empirical Rejection Rate (in percentage) for White Noise Test

Under the alternative ($\mathcal{M}_3$), we can observe that both methods can have nontrivial power. For T_C the overall performance is good when $N = 150$, 300 and has better power than T_{SN} especially when d is large, but performs very poorly when $N = 75$. T_{SN} has a decent power only at the case $d = 2$ and trivial power in other cases. Similar results can be found for $\mathcal{M}_4$. The loss of power for the SN-based test at larger d can be explained by the fact that as d increases the alternative becomes more sparse, which has less impact on the power of T_C than on that of T_{SN}.

For $\mathcal{M}_5$, T_{SN} outperforms T_C in most cases. This is due to the highly dense alternative under the model $\mathcal{M}_5$. It is also clear that when d gets larger, the power of T_{SN} decreases for the same reason explained for models $\mathcal{M}_3$ and $\mathcal{M}_4$.

6. Conclusion. In this paper we propose a new formulation of self-normalization for the inference of mean in the high-dimensional time series setting. We use a one sample U-statistic with trimming to accommodate weak time series dependence, and show its asymptotic normality under the general nonlinear causal process framework. To avoid direct consistent estimation of the nuisance parameter, which is the Frobenious norm of long run covariance matrix, we apply the idea of self-normalization. Different from the low-dimensional case, the recursive U-statistic based on subsamples (upon suitable standarization) converges to a time-changed Brownian motion and the self-normalized test statistic has a different pivotal limit. More interestingly, the convergence rate of our original U-statistic, which depends on the diverging rate of $\|\Gamma\|_F$, is not required to be known. This phenomenon seems new, as the convergence rate is typically known [See Shao (2010), Shao (2015)] or needs to be estimated (see e.g., long memory time series setting in Shao (2011)) in the use of SN for low-dimensional time series. Simulation studies show that our SN-based test statistic has accurate size, and it is not overly sensitive to the trimming parameter involved, whereas the size of the maximum type tests in Zhang and Cheng (2018) and Zhang and Wu (2017) can critically depend on the block size.

To conclude, it is worth pointing out a few important future research directions. An obvious one is to come up with a good data-driven formula for m, the trimming parameter involved in our test. In addition, we assume stationarity throughout, while in practice the series may be heteroscedastic and exhibits time varying dependence. This may be accommodated by using the local stationary framework in Zhou (2013), the use of which seems to be limited to the low-dimensional setting. Also we do not usually have a priori knowledge on whether the alternative is sparse or dense. It would be interesting to develop an adaptive test in the high-dimensional time series setting. One possibility is to extend recent work of He et al. (2018) from i.i.d. to dependent data. We shall leave these topics for future research.

Acknowledgements. We thank the Editor Richard Samworth, the Associate Editor and the reviewers, whose comments led to substantial improvements in the paper. We are grateful to Deepak Nag Ayyala for sending us the code used in Ayyala et al. (2017). The research is partially supported by NSF-DMS 1807023.

SUPPLEMENTAL MATERIALS

Supplement to: Hypothesis Testing for High-Dimensional Time Series via Self-normalization contain technical proofs of most theoretical results stated in the paper, some auxiliary lemmas and additional simulation results.

References.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience; 3rd edition.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.

Ayyala, D. N., Park, J., and Roy, A. (2017). Mean vector testing for high-dimensional dependent observations. *Journal of Multivariate Analysis*, 153:136–155.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6(2):311–329.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567.

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227.

Billingsley, P. (2008). *Probability and Measure*. John Wiley & Sons.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*. Classics in Applied Mathematics, 36. SIAM: Society for Industrial and Applied Mathematics.

Cai, T., Liu, W., and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):349–372.

Cai, T. T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *The Annals of Statistics*, 39(3):1496–1525.

Cai, T. T., Zhang, C.-H., and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, 38(4):2118–2144.

Chang, J., Yao, Q., and Zhou, W. (2017). Testing for high-dimensional white noise using maximum cross-correlations. *Biometrika*, 104(1):111–127.

Chen, L. and Wu, W. B. (2019). Testing for trends in high-dimensional time series. *Journal of the American Statistical Association*, 526:869–881.

Chen, S. X. and Qin, Y. L. (2010). A two sample test for high dimensional data with application to gene-set testing. *The Annals of Statistics*, 38:808–835.

Chen, S. X., Zhang, L.-X., and Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41:2786–2819.

Gordin, M. I. (1969). The central limit theorem for stationary processes. In *Doklady Akademii Nauk*, volume 188, pages 739–741. Russian Academy of Sciences.

Gregory, K. B., Carroll, R. J., Baladandayuthapani, V., and Lahiri, S. N. (2015). A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association*, 110(510):837–849.

Hall, P. and Heyde, C. C. (2014). *Martingale Limit Theory and Its Application.* Academic press.

He, Y., Xu, G., Wu, C., and Pan, W. (2018). Asymptotically independent u-statistics in high dimensional testing. *https://arxiv.org/abs/1809.00411*.

Hosking, J. R. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, 75(371):602–608.

Hsing, T. and Wu, W. B. (2004). On weighted u statistics for stationary processes. *Annals of Probability*, 32:1600–1631.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Della Penna, S., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., et al. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80:360–378.

Jentsch, C. and Politis, D. N. (2015). Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *The Annals of Statistics*, 43(3):1117–1140.

Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *The Annals of Applied Probability*, 14(2):865–880.

Kiefer, N. M., Vogelsang, T. J., and Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, 68(3):695–714.

Kim, S., Zhao, Z., and Shao, X. (2015). Nonparametric functional central limit theorem for time series regression with application to self-normalized confidence interval. *Journal of Multivariate Analysis*, 133:277–290.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Annals of Statistics*, 25:2084–2102.

Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17:1217–1241.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *Annals of Statistics*, 40(2):694–726.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30(4):1081–1102.

Lee, A. J. (1990). *U-statistics, Theory and Practice.* CRC Press.

Li, W. K. (2004). *Diagnostic Checks in Time Series.* Chapman and Hall/CRC.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate arma time series models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43:231–239.

Li, Z., Lam, C., Yao, J., and Yao, Q. (2018). On testing for high-dimensional white noise. *Annals of Statistics, to appear. Available at arXiv preprint arXiv:1808.03545.*

Liu, R. Y. and Singh, K. (1992). Moving block jackknife and bootstrap capture weak dependence. *Exploring the Limits of Bootstrap , Ed. R. LePage and L. Billard*, pages 225–248.

Liu, Y., Tamura, Y., and Taniguchi, M. (2018). Asymptotic theory of test statistic for sphericity of high-dimensional time series. *Journal of Time Series Analysis*, 39(3):402–416.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, 96(455):1066–1076.

Lopes, M., Jacob, L., and Wainwright, M. J. (2011). A more powerful two-sample test in high dimensions using random projection. In *Advances in Neural Information Processing Systems*, pages 1206–1214.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer Science & Business Media.

McMurray, T, L. and Politis, D, N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, 31:471–482.

Newey, W. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

Paul, D. and Wang, L. (2016). Discussion of estimating structured high-dimensional covariance and precision matrices: optimal rates and adaptive estimation. *Electronic Journal of Statistics*, 10(1):74–80.

Phillips, P. C. and Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, 20(2):971–1001.

Pini, A., Stamm, A., and Vantini, S. (2018). Hotellings t^2 in separable hilbert spaces. *Journal of Multivariate Analysis*, 167:284–305.

Politis, D. and Romano, J. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics*, 22(4):2031–2050.

Qiu, Y. and Chen, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *The Annals of Statistics*, 40(3):1285–1314.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, 92(4):951–956.

Secchi, P., Stamm, A., and Vantini, S. (2013). Inference for the mean of large p small n data: A finite-sample high-dimensional generalization of hotellings theorem. *Electronic Journal of Statistics*, 7:2005–2031.

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):343–366.

Shao, X. (2011). A simple test of changes in mean in the presence of long-range dependence. *Journal of Time Series Analysis*, 32:598–606.

Shao, X. (2015). Self-normalization for time series: a review of recent developments. *Journal of the American Statistical Association*, 110(512):1797–1817.

Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Annals of Statistics*, 35(4):1773–1801.

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105(491):1228–1240.

Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis*, 100(3):518–532.

Srivastava, M. S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 99(3):386–402.

Sun, Y. (2014). Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. *Journal of Econometrics*, 178:659–677.

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

Wu, W. B. (2007). Strong invariance principles for dependent random variables. *The Annals of Probability*, 35(6):2294–2320.

Wu, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and its Interface*, 4(2):207–226.

Wu, W. B. and Min, W. (2005). On linear processes with dependent innovations. *Stochastic Processes and their Applications*, 115(6):939–958.

Wu, W. B. and Shao, X. (2004). Limit theorems for iterated random functions. *Journal of Applied Probability*, 41(2):425–436.

Wu, W. B. and Woodroofe, M. (2004). Martingale approximations for sums of stationary processes. *The Annals of Probability*, 32(2):1674–1690.

Xu, G., Lin, L., Wei, P., and Pan, W. (2016). An adaptive two-sample test for high-dimensional means. *Biometrika*, 103(3):609–624.

Zhang, D. and Wu, W. B. (2017). Gaussian approximation for high dimensional time series. *The Annals of Statistics*, 45(5):1895–1919.

Zhang, X. (2017). Testing high dimensional mean under sparsity. *Preprint, available at "http://www.stat.tamu.edu/ zhangxiany/LR-HD2017.pdf"*.

Zhang, X. and Cheng, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli*, 24:2640–2675, see https://arxiv.org/abs/1406.1037 for an early arxiv version.

Zhang, X., Shao, X., Hayhoe, K., and Wuebbles, D. J. (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electronic Journal of Statistics*, 5:1765–1796.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association*, 108:726–740.

Zhou, Z. and Shao, X. (2013). Inference for linear models with dependent errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(2):323–343.

7. Appendix: Proof of Main Theorems.

7.1. *Proof of Theorem 3.6.* To show the process convergence, we need to prove the following two facts. Under Assumptions 3.3,

1. For any $r_1, \cdots, r_k \in [0,1]$,

$$\left(\frac{\sqrt{2}}{n\|\Gamma\|_F} S_n(r_1), \cdots, \frac{\sqrt{2}}{n\|\Gamma\|_F} S_n(r_k)\right) \xrightarrow{\mathcal{D}} (\mathcal{B}(r_1^2), \cdots, \mathcal{B}(r_k^2)). \tag{7.1}$$

2. The process $\frac{\sqrt{2}}{n\|\Gamma\|_F} S_n(r)$ is tight. It suffices to show for all $0 \le a \le b \le 1$ and $n \ge n_0$ for some $n_0 > 0$,

$$\mathbb{E}\left[\left(\frac{\sqrt{2}}{n\|\Gamma\|_F} S_n(b) - \frac{\sqrt{2}}{n\|\Gamma\|_F} S_n(a)\right)^4\right] \le C\left((\lfloor nb \rfloor - \lfloor na \rfloor)/n\right)^2 \tag{7.2}$$

   according to Lemma 9.8 in the supplement material.

7.1.1. *Proof of* (7.1). For simplicity we only prove the case when $k = 2$, since for a general $k \ge 2$, the result can be proved by similar arguments. By the Cramer-Wold device, it is equivalent to show for any $\alpha_1, \alpha_2 \in \mathbb{R}$,

$$\frac{\sqrt{2}}{n\|\Gamma\|_F}(\alpha_1 S_n(r_1) + \alpha_2 S_n(r_2)) \xrightarrow{\mathcal{D}} \alpha_1 \mathcal{B}(r_1^2) + \alpha_2 \mathcal{B}(r_2^2).$$

WLOG, we assume $r_1 \le r_2$. By simple calculation we can see that

$$\begin{aligned}
\frac{\sqrt{2}}{n\|\Gamma\|_F}(\alpha_1 S_n(r_1) + \alpha_2 S_n(r_2)) &= \frac{\sqrt{2}}{n\|\Gamma\|_F}\left(\alpha_1 \sum_{t=1}^{\lfloor nr_1 \rfloor} \sum_{s=1}^{t} D_{t+m}^T D_s + \alpha_2 \sum_{t=1}^{\lfloor nr_2 \rfloor} \sum_{s=1}^{t} D_{t+m}^T D_s\right) \\
&= \frac{\sqrt{2}}{n\|\Gamma\|_F}\left((\alpha_1 + \alpha_2) \sum_{t=1}^{\lfloor nr_1 \rfloor} \sum_{s=1}^{t} D_{t+m}^T D_s + \alpha_2 \sum_{t=\lfloor nr_1 \rfloor + 1}^{\lfloor nr_2 \rfloor} \sum_{s=1}^{t} D_{t+m}^T D_s\right) \\
&= \sum_{t=1}^{\lfloor nr_2 \rfloor} \eta_{t+m},
\end{aligned}$$

where

$$\eta_{t+m} = \begin{cases} \frac{(\alpha_1+\alpha_2)\sqrt{2}}{n\|\Gamma\|_F} D_{t+m}^T \sum_{s=1}^t D_s, & \text{for any } 1 \le t \le \lfloor nr_1 \rfloor, \\ \frac{\alpha_2\sqrt{2}}{n\|\Gamma\|_F} D_{t+m}^T \sum_{s=1}^t D_s, & \text{for any } \lfloor nr_1 \rfloor + 1 \le t \le \lfloor nr_2 \rfloor. \end{cases}$$

It can be easily verified that for any fixed n, $\{\eta_{t+m}\}_{t=1}^n$ is a martingale difference sequence with respect to $\mathcal{F}_{t+m}$. Direct application of Theorem 35.12 in Billingsley (2008) (Martingale CLT) indicates that we need to show the following:

1. $\forall \epsilon \ge 0, \sum_{t=1}^{\lfloor nr_2 \rfloor} \mathbb{E}[\eta_{t+m}^2 \mathbf{1}\{|\eta_{t+m}| > \epsilon\} | \mathcal{F}_{t+m-1}] \xrightarrow{p} 0$,

2. $V_n = \sum_{t=1}^{\lfloor nr_2 \rfloor} \mathbb{E}[\eta_{t+m}^2 | \mathcal{F}_{t+m-1}] \xrightarrow{p} \sigma^2 = (\alpha_1^2 r_1^2 + \alpha_2^2 r_2^2 + 2\alpha_1\alpha_2 r_1^2)$.

To show 1, it suffices to show

$$\sum_{t=1}^{\lfloor nr_2\rfloor}\mathbb{E}[\eta_{t+m}^4]\to 0. \tag{7.3}$$

To show 2, we can simplify V_n as

$$\begin{aligned}
V_n =& \sum_{t=1}^{\lfloor nr_2\rfloor}\mathbb{E}[\eta_{t+m}^2|\mathcal{F}_{t+m-1}]\\
=& \sum_{t=1}^{\lfloor nr_1\rfloor}\mathbb{E}[\frac{2(\alpha_1+\alpha_2)^2}{n^2\|\Gamma\|_F^2}(D_{t+m}^T\sum_{s=1}^t D_s)^2|\mathcal{F}_{t+m-1}] + \sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor}\mathbb{E}[\frac{2\alpha_2^2}{n^2\|\Gamma\|_F^2}(D_{t+m}^T\sum_{s=1}^t D_s)^2|\mathcal{F}_{t+m-1}]\\
=& \frac{2(\alpha_1^2+2\alpha_1\alpha_2)}{n^2\|\Gamma\|_F^2}\sum_{t=1}^{\lfloor nr_1\rfloor}\sum_{s_1=1}^t\sum_{s_2=1}^t D_{s_1}^T\mathbb{E}[D_{t+m}D_{t+m}^T|\mathcal{F}_{t+m-1}]D_{s_2}\\
&+\frac{2\alpha_2^2}{n^2\|\Gamma\|_F^2}\sum_{t=1}^{\lfloor nr_2\rfloor}\sum_{s_1=1}^t\sum_{s_2=1}^t D_{s_1}^T\mathbb{E}[D_{t+m}D_{t+m}^T|\mathcal{F}_{t+m-1}]D_{s_2}.
\end{aligned}$$

This implies that we only need to show

$$\frac{2}{n^2\|\Gamma\|_F^2}\sum_{t=1}^{n}\sum_{s_1=1}^t\sum_{s_2=1}^t D_{s_1}^T\mathbb{E}[D_{t+m}D_{t+m}^T|\mathcal{F}_{t+m-1}]D_{s_2}\xrightarrow{p}1. \tag{7.4}$$

Proof of (7.3) Note that

$$\begin{aligned}
\sum_{t=1}^{\lfloor nr_2\rfloor}\mathbb{E}[\eta_{t+m}^4] =& \frac{4(\alpha_1+\alpha_2)^4}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{\lfloor nr_1\rfloor}\mathbb{E}\left[\left(D_{t+m}^T\sum_{s=1}^t D_s\right)^4\right]+\frac{4\alpha_2^4}{n^4\|\Gamma\|_F^4}\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor}\mathbb{E}\left[\left(D_{t+m}^T\sum_{s=1}^t D_s\right)^4\right]\\
\le& \frac{4((\alpha_1+\alpha_2)^4\vee\alpha_2^4)}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\mathbb{E}\left[\left(D_{t+m}^T\sum_{s=1}^t D_s\right)^4\right].
\end{aligned}$$

For the summation,

$$\begin{aligned}
&\frac{1}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\mathbb{E}\left[\left(D_{t+m}^T\sum_{s=1}^t D_s\right)^4\right]\\
\lesssim&\frac{1}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\sum_{s_1\le\cdots\le s_4=1}^{t}\sum_{j_1,\cdots,j_4=1}^{p}\mathbb{E}[D_{t+m,j_1}D_{t+m,j_2}D_{t+m,j_3}D_{t+m,j_4}D_{s_1,j_1}D_{s_2,j_2}D_{s_3,j_3}D_{s_4,j_4}].
\end{aligned}$$

By strict stationarity, we have

$$\begin{aligned}
&\mathbb{E}[D_{t+m,j_1}D_{t+m,j_2}D_{t+m,j_3}D_{t+m,j_4}D_{s_1,j_1}D_{s_2,j_2}D_{s_3,j_3}D_{s_4,j_4}]\\
=&\mathbb{E}[D_{t+m-s_4,j_1}D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4}D_{s_1-s_4,j_1}D_{s_2-s_4,j_2}D_{s_3-s_4,j_3}D_{0,j_4}]\\
=&\mathbb{E}[D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D'_{t+m-s_4,j_3}D'_{t+m-s_4,j_4}D_{s_1-s_4,j_1}D_{s_2-s_4,j_2}D_{s_3-s_4,j_3}D_{0,j_4}]\\
&+\mathbb{E}[(D_{t+m-s_4,j_1}D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4}-D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D'_{t+m-s_4,j_3}D'_{t+m-s_4,j_4})\\
&\quad D_{s_1-s_4,j_1}D_{s_2-s_4,j_2}D_{s_3-s_4,j_3}D_{0,j_4}]\\
=&L_1+L_2.
\end{aligned}$$

For L_1, since D_{t+m-s_4} is independent of D_s for any $s \le 0$,

$$L_1 = \mathbb{E}[D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D'_{t+m-s_4,j_3}D'_{t+m-s_4,j_4}]\mathbb{E}[D_{s_1-s_4,j_1}D_{s_2-s_4,j_2}D_{s_3-s_4,j_3}D_{0,j_4}].$$

For L_2, since Y_t is UGMC(8), $\|D_{t,j} - D'_{t,j}\|_8 \le 2C\rho^t$ for any $j = 1, \cdots, p$. By Hölder's inequality we have

$$\begin{aligned}L_2 \le& \|D_{t+m-s_4,j_1}D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4} - D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D'_{t+m-s_4,j_3}D'_{t+m-s_4,j_4}\|_2\\ &\|D_{s_1-s_4,j_1}D_{s_2-s_4,j_2}D_{s_3-s_4,j_3}D_{0,j_4}\|_2.\end{aligned}$$

To utilize the property of UGMC, we manipulate the first term as

$$\begin{aligned}&\|D_{t+m-s_4,j_1}D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4} - D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D'_{t+m-s_4,j_3}D'_{t+m-s_4,j_4}\|_2\\ \le&\|D_{t+m-s_4,j_1}D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4} - D'_{t+m-s_4,j_1}D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4}\|_2\\ &+\|D'_{t+m-s_4,j_1}D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4} - D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4}\|_2\\ &+\|D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4} - D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D'_{t+m-s_4,j_3}D_{t+m-s_4,j_4}\|_2\\ &+\|D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D'_{t+m-s_4,j_3}D_{t+m-s_4,j_4} - D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D'_{t+m-s_4,j_3}D'_{t+m-s_4,j_4}\|_2,\end{aligned}$$

where the first term in the above expression satisfies

$$\begin{aligned}&\|D_{t+m-s_4,j_1}D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4} - D'_{t+m-s_4,j_1}D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4}\|_2\\ =&\|(D_{t+m-s_4,j_1} - D'_{t+m-s_4,j_1})D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4}\|_2\\ \le&\|D_{t+m-s_4,j_1} - D'_{t+m-s_4,j_1}\|_8\|D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4}\|_{8/3}\\ \le&2C\sup_j\|D_{0,j}\|_8^3\rho^{t-s_4+m},\end{aligned}$$

the same bound of which holds for the other three terms in the summation. Hence

$$\begin{aligned}L_2 \le&\|D_{t+m-s_4,j_1}D_{t+m-s_4,j_2}D_{t+m-s_4,j_3}D_{t+m-s_4,j_4} - D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D'_{t+m-s_4,j_3}D'_{t+m-s_4,j_4}\|_2\\ &\|D_{s_1-s_4,j_1}D_{s_2-s_4,j_2}D_{s_3-s_4,j_3}D_{0,j_4}\|_2\\ \le&8C\sup_j\|D_{0,j}\|_8^3\rho^{t-s_4+m}\|D_{s_1-s_4,j_1}D_{s_2-s_4,j_2}D_{s_3-s_4,j_3}D_{0,j_4}\|_2 \le 8C\sup_j\|D_{0,j}\|_8^7\rho^{t-s_4+m}.\end{aligned}$$

Moreover, by definition of joint cumulants we have

$$\begin{aligned}\mathbb{E}[D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D'_{t+m-s_4,j_3}D'_{t+m-s_4,j_4}] =& \Gamma_{j_1,j_2}\Gamma_{j_3,j_4} + \Gamma_{j_1,j_3}\Gamma_{j_2,j_4} + \Gamma_{j_1,j_4}\Gamma_{j_2,j_3}\\ &+ cum(D_{0,j_1}, D_{0,j_2}, D_{0,j_3}, D_{0,j_4}),\end{aligned}$$

which indicates that, under Assumption A.5,

$$\sum_{j_1,j_2,j_3,j_4=1}^{p}\mathbb{E}[D'_{t+m-s_4,j_1}D'_{t+m-s_4,j_2}D'_{t+m-s_4,j_3}D'_{t+m-s_4,j_4}] = O(\|\Gamma\|_F^4)$$

and

$$\begin{aligned}&|\mathbb{E}[D_{s_1-s_4,j_1}D_{s_2-s_4,j_2}D_{s_3-s_4,j_3}D_{0,j_4}]|\\ =&|\Gamma_{j_1,j_2}\Gamma_{j_3,j_4}\mathbf{1}\{s_1 = s_2\}\mathbf{1}\{s_3 = s_4\} + \Gamma_{j_1,j_3}\Gamma_{j_2,j_4}\mathbf{1}\{s_1 = s_3\}\mathbf{1}\{s_2 = s_4\}\\ &+ \Gamma_{j_1,j_4}\Gamma_{j_2,j_3}\mathbf{1}\{s_1 = s_4\}\mathbf{1}\{s_2 = s_3\} + cum(D_{s_1-s_4,j_1}, D_{s_2-s_4,j_2}, D_{s_3-s_4,j_3}, D_{0,j_4})|\\ \le&|\Gamma_{j_1,j_2}\Gamma_{j_3,j_4}\mathbf{1}\{s_1 = s_2\}\mathbf{1}\{s_3 = s_4\}| + |\Gamma_{j_1,j_3}\Gamma_{j_2,j_4}\mathbf{1}\{s_1 = s_3\}\mathbf{1}\{s_2 = s_4\}|\\ &+ |\Gamma_{j_1,j_4}\Gamma_{j_2,j_3}\mathbf{1}\{s_1 = s_4\}\mathbf{1}\{s_2 = s_3\}| + 2C\rho^{s_4-s_1}.\end{aligned}$$

Thus,

$$\begin{aligned}
&\frac{1}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\sum_{s_1\le\cdots\le s_4=1}^{t}\sum_{j_1,\cdots,j_4=1}^{p}|L_1| \\
\le&\frac{1}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\sum_{s_1\le\cdots\le s_4=1}^{t}\sum_{j_1,\cdots,j_4=1}^{p}|(\Gamma_{j_1,j_2}\Gamma_{j_3,j_4}+\Gamma_{j_1,j_3}\Gamma_{j_2,j_4}+\Gamma_{j_1,j_4}\Gamma_{j_2,j_3} \\
&+cum(D_{0,j_1},D_{0,j_2},D_{0,j_3},D_{0,j_4}))||(\Gamma_{j_1,j_2}\Gamma_{j_3,j_4}\mathbf{1}\{s_1=s_2\}\mathbf{1}\{s_3=s_4\} \\
&+\Gamma_{j_1,j_3}\Gamma_{j_2,j_4}\mathbf{1}\{s_1=s_3\}\mathbf{1}\{s_2=s_4\}+\Gamma_{j_1,j_4}\Gamma_{j_2,j_3}\mathbf{1}\{s_1=s_4\}\mathbf{1}\{s_2=s_3\}+2C\rho^{s_4-s_1})| \\
\le&\frac{3}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\sum_{1\le s_1\le s_3\le t}\left(\sup_{j_1,j_2=1,\cdots,p}|\Gamma_{j_1,j_2}|\right)^2 O(\|\Gamma\|_F^4)+\frac{1}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\sum_{1\le s_1\le\cdots\le s_4\le t}2C\rho^{s_4-s_1}O(\|\Gamma\|_F^4) \\
\le&O(n^{-1})+\frac{2C}{n^4}\sum_{t=1}^{n}\sum_{s_1=1}^{t}\sum_{l=0}^{t}l^2\rho^l O(1)\le O(n^{-1})+O(n^{-2})=O(n^{-1})\to 0
\end{aligned}$$

by Assumption A.5. Note that we have used the fact that $\sup_{j_1,j_2=1,\cdots,p}|\Gamma_{j_1,j_2}|=O(1)$ under A.1 in Assumption 3.3. In addition,

$$\begin{aligned}
&\frac{1}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\sum_{s_1\le\cdots\le s_4=1}^{t}\sum_{j_1,\cdots,j_4=1}^{p}|L_2| \\
\le&\frac{1}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\sum_{s_1\le\cdots\le s_4=1}^{t}\sum_{j_1,\cdots,j_4=1}^{p}8C\sup_j\|D_{0,j}\|_8^7\rho^{t-s_4+m} \\
\lesssim&\frac{p^4}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\sum_{s_4=1}^{t}s_4^3\rho^{t-s_4+m}\le\frac{p^4}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\sum_{l=0}^{t-1}t^3\rho^{l+m}\le\frac{p^4\rho^m}{\|\Gamma\|_F^4}=o(1)
\end{aligned}$$

by Assumption A.4. Together with previous results we have

$$\frac{1}{n^4\|\Gamma\|_F^4}\sum_{t=1}^{n}\mathbb{E}\left[(D_{t+m}^T\sum_{s=1}^{t}D_s)^4\right]\to 0.$$

The proof is complete.

Proof of (7.4) To simplify notations, we let $\Delta_{t+m-1}=\mathbb{E}[D_{t+m}D_{t+m}^T|\mathcal{F}_{t+m-1}]$, and $\Delta_{t+m-1,i,j}$ is the (i,j) component of Δ_{t+m-1}.

Note that

$$\begin{aligned}
&\frac{2}{n^2\|\Gamma\|_F^2}\sum_{t=1}^{n}\sum_{s_1=1}^{t}\sum_{s_2=1}^{t}D_{s_1}^T\Delta_{t+m-1}D_{s_2} \\
=&\frac{2}{n^2\|\Gamma\|_F^2}\sum_{t=1}^{n}\sum_{s_1=1}^{t}\sum_{s_2=1}^{t}D_{s_1}^T\Gamma D_{s_2}+\frac{2}{n^2\|\Gamma\|_F^2}\sum_{t=1}^{n}\sum_{s_1=1}^{t}\sum_{s_2=1}^{t}D_{s_1}^T(\Delta_{t+m-1}-\Gamma)D_{s_2} \\
=&L_3+L_4.
\end{aligned}$$

By simple calculation, $\mathbb{E}[L_3] \to 1$. Moreover,

$$\begin{aligned}
\mathbb{E}[L_3^2] &= \frac{4}{n^4\|\Gamma\|_F^4} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1,s_2=1}^{t_1}\sum_{s_3,s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p} \mathbb{E}[D_{s_1,j_1}D_{s_2,j_2}D_{s_3,j_3}D_{s_4,j_4}]\Gamma_{j_1,j_2}\Gamma_{j_3,j_4} \\
&= \frac{4}{n^4\|\Gamma\|_F^4} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1,s_2=1}^{t_1}\sum_{s_3,s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p} cum[D_{s_1,j_1},D_{s_2,j_2},D_{s_3,j_3},D_{s_4,j_4}]\Gamma_{j_1,j_2}\Gamma_{j_3,j_4} \\
&+ \frac{4}{n^4\|\Gamma\|_F^4} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1,s_2=1}^{t_1}\sum_{s_3,s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p} \mathbf{1}\{s_1=s_2\}\mathbf{1}\{s_3=s_4\}\Gamma_{j_1,j_2}^2\Gamma_{j_3,j_4}^2 \\
&+ \frac{4}{n^4\|\Gamma\|_F^4} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1,s_2=1}^{t_1}\sum_{s_3,s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p} \mathbf{1}\{s_1=s_3\}\mathbf{1}\{s_2=s_4\}\Gamma_{j_1,j_3}\Gamma_{j_2,j_4}\Gamma_{j_1,j_2}\Gamma_{j_3,j_4} \\
&+ \frac{4}{n^4\|\Gamma\|_F^4} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1,s_2=1}^{t_1}\sum_{s_3,s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p} \mathbf{1}\{s_1=s_4\}\mathbf{1}\{s_2=s_3\}\Gamma_{j_1,j_4}\Gamma_{j_2,j_3}\Gamma_{j_1,j_2}\Gamma_{j_3,j_4} \\
&= K_{3,1}+K_{3,2}+K_{3,3}+K_{3,4},
\end{aligned}$$

since $\mathbb{E}[D_{s,i}D_{t,j}]=0$ if $s\neq t$, for any $i,j=1,\cdots,p$, by property of martingale difference sequence.

For the first term,

$$\begin{aligned}
|K_{3,1}| &\lesssim \frac{4}{n^4\|\Gamma\|_F^4} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1\le\cdots\le s_4=1}^{n}\sum_{j_1,\cdots,j_4=1}^{p} |cum[D_{s_1,j_1},D_{s_2,j_2},D_{s_3,j_3},D_{s_4,j_4}]||\Gamma_{j_1,j_2}||\Gamma_{j_3,j_4}| \\
&\le \frac{4}{n^4\|\Gamma\|_F^4} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1\le\cdots\le s_4=1}^{n} \rho^{s_4-s_1}\left(\sum_{j_1,j_2=1}^{p}|\Gamma_{j_1,j_2}|\right)^2 \\
&\le \frac{4}{n^4\|\Gamma\|_F^4} n^2 \sum_{s_1=1}^{n}\sum_{s_4-s_1=0}^{n-1}(s_4-s_1)^2\rho^{s_4-s_1}O(\|\Gamma\|_F^4) = O(n^{-1}) \to 0.
\end{aligned}$$

For the second one,

$$K_{3,2} = \frac{4}{n^4\|\Gamma\|_F^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{t_1}\sum_{s_3=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p}\Gamma_{j_1,j_2}^2\Gamma_{j_3,j_4}^2 \to 1.$$

For the third term,

$$\begin{aligned}
|K_{3,3}| &= \left|\frac{4}{n^4\|\Gamma\|_F^4} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1,s_2=1}^{t_1}\sum_{s_3,s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p} \mathbf{1}\{s_1=s_3\}\mathbf{1}\{s_2=s_4\}\Gamma_{j_1,j_3}\Gamma_{j_2,j_4}\Gamma_{j_1,j_2}\Gamma_{j_3,j_4}\right| \\
&= \frac{4}{n^4\|\Gamma\|_F^4} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1,s_2=1}^{t_1}\sum_{s_3,s_4=1}^{t_2} \mathbf{1}\{s_1=s_3\}\mathbf{1}\{s_2=s_4\}\sum_{j_1,j_2=1}^{p}[(\Gamma^2)_{j_1,j_2}]^2 \\
&\le \frac{4}{n^4\|\Gamma\|_F^4} \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1,s_2=1}^{n}\sum_{s_3,s_4=1}^{n} \mathbf{1}\{s_1=s_3\}\mathbf{1}\{s_2=s_4\}\|\Gamma^2\|_F^2 \\
&\le \frac{4}{n^4\|\Gamma\|_F^4} n^4\|\Gamma\|_F^2\|\Gamma\|^2 = O\left(\frac{\|\Gamma\|^2}{\|\Gamma\|_F^2}\right) \to 0,
\end{aligned}$$

where the second from the last line is due to Lemma 9.1 in the supplement material. And by similar arguments we have $|K_{3,4}| \to 0$. Combining these results, we have $\mathbb{E}[(L_3-1)^2] \to 0$, which implies $L_3 \xrightarrow{p} 1$ by Chebyshev's inequality.

For L_4, it suffices to show $L_4 \xrightarrow{p} 0$, which is implied by $\mathbb{E}[L_4^2] \to 0$. To this end, we note that

$$\begin{aligned}\mathbb{E}[L_4^2] =& \frac{4}{n^4\|\Gamma\|_F^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{t_1}\sum_{s_2=1}^{t_1}\sum_{s_3=1}^{t_2}\sum_{s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p}\mathbb{E}[D_{s_1,j_1}D_{s_2,j_2}D_{s_3,j_3}D_{s_4,j_4}\\ &(\Delta_{t_1+m-1,j_1,j_2}-\Gamma_{j_1,j_2})(\Delta_{t_2+m-1,j_3,j_4}-\Gamma_{j_3,j_4})].\end{aligned}$$

Here we need to deal with an expectation of a product of six random variables. By definition of joint cumulants, it can be decomposed as a summation of products of joint cumulants. It is tedious to list all cases since the derivation for those cases are very similar. Hence only some representative cases will be shown here.

1. $cum(D_{s_1,j_1}, D_{s_2,j_2}, D_{s_3,j_3}, D_{s_4,j_4}, \Delta_{t_1+m-1,j_1,j_2}, \Delta_{t_2+m-1,j_3,j_4})$

$$\begin{aligned}&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{t_1}\sum_{s_2=1}^{t_1}\sum_{s_3=1}^{t_2}\sum_{s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p}\\ &|cum(D_{s_1,j_1}, D_{s_2,j_2}, D_{s_3,j_3}, D_{s_4,j_4}, \Delta_{t_1+m-1,j_1,j_2}, \Delta_{t_2+m-1,j_3,j_4})|\\ \lesssim&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{1\le s_1\le s_2\le t_1}\sum_{1\le s_3\le s_4\le t_2}\sum_{j_1,\cdots,j_4=1}^{p}C\rho^{t_1-s_1+m-1}\\ \le&\frac{4Cn^3p^4}{n^4\|\Gamma\|_F^4}\sum_{t_1=1}^{n}\sum_{l=0}^{t_1-1}l\rho^{l+m-1}=O\left(\frac{p^4\rho^m}{\|\Gamma\|_F^4}\right)\to 0.\end{aligned}$$

2. $cum(D_{s_1,j_1}, D_{s_2,j_2}, D_{s_3,j_3}, D_{s_4,j_4})cum(\Delta_{t_1+m-1,j_1,j_2}, \Delta_{t_2+m-1,j_3,j_4})$

$$\begin{aligned}&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{t_1}\sum_{s_2=1}^{t_1}\sum_{s_3=1}^{t_2}\sum_{s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p}\\ &|cum(D_{s_1,j_1}, D_{s_2,j_2}, D_{s_3,j_3}, D_{s_4,j_4})||cum(\Delta_{t_1+m-1,j_1,j_2}, \Delta_{t_2+m-1,j_3,j_4})|\\ \lesssim&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{1\le t_1\le t_2\le n}\sum_{1\le s_1\le s_2\le s_3\le s_4\le t_2}C\rho^{s_4-s_1}\sum_{j_1,\cdots,j_4=1}^{p}|cum(\Delta_{t_1+m-1,j_1,j_2}, \Delta_{t_2+m-1,j_3,j_4})|\\ \le&\frac{4Cn^2}{n^4\|\Gamma\|_F^4}\left(\sum_{s_1=1}^{n}\sum_{l=1}^{n}l^2\rho^l\right)O(\|\Gamma\|_F^4)=O(n^{-1})\to 0,\end{aligned}$$

where the second line from the last is due to Assumption A.6.

3. $cum(D_{s_1,j_1}, D_{s_2,j_2}, \Delta_{t_2+m-1,j_3,j_4})cum(D_{s_3,j_3}, D_{s_4,j_4}, \Delta_{t_1+m-1,j_1,j_2})$

$$
\begin{aligned}
&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{t_1}\sum_{s_2=1}^{t_1}\sum_{s_3=1}^{t_2}\sum_{s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p}\\
&|cum(D_{s_1,j_1}, D_{s_2,j_2}, \Delta_{t_2+m-1,j_3,j_4})cum(D_{s_3,j_3}, D_{s_4,j_4}, \Delta_{t_1+m-1,j_1,j_2})|\\
\lesssim&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{1\le t_1\le t_2\le n}^{n}\sum_{1\le s_1\le s_2\le t_1}\sum_{s_3=1}^{t_2}\sum_{s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p}\\
&|cum(D_{s_1,j_1}, D_{s_2,j_2}, \Delta_{t_2+m-1,j_3,j_4})||cum(D_{s_3,j_3}, D_{s_4,j_4}, \Delta_{t_1+m-1,j_1,j_2})|\\
\le&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{1\le t_1\le t_2\le n}^{n}\sum_{1\le s_1\le s_2\le t_1}\sum_{s_3=1}^{t_2}\sum_{s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p}\\
&C\rho^{t_2+m-1-s_1}(|\mathbb{E}(D_{s_3,j_3}D_{s_4,j_4}\Delta_{t_1+m-1,j_1,j_2})|+|\mathbb{E}(D_{s_3,j_3}D_{s_4,j_4})\mathbb{E}[\Delta_{t_1+m-1,j_1,j_2}]|)\\
\le&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{1\le t_1\le t_2\le n}^{n}\sum_{1\le s_1\le s_2\le t_1}\sum_{s_3=1}^{t_2}\sum_{s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p}\\
&C\rho^{t_2+m-1-s_1}(\|D_{s_3,j_3}\|_4\|D_{s_4,j_4}\|_4\|\Delta_{t_1+m-1,j_1,j_2}\|_2+\|D_{s_3,j_3}\|_2\|D_{s_4,j_4}\|_2|\Gamma_{j_1,j_2}|)\\
\le&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{1\le t_1\le t_2\le n}^{n}\sum_{1\le s_1\le s_2\le t_1}\sum_{s_3=1}^{t_2}\sum_{s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p}C\rho^{t_2+m-1-s_1}\left(\sup_{j=1,\cdots,p}\|D_{0,j}\|_4\right)^4\\
=&O\left(\frac{\rho^m p^4}{\|\Gamma\|_F^4}\right)\to 0.
\end{aligned}
$$

4. $cum(D_{s_1,j_1}, D_{s_2,j_2})cum(D_{s_3,j_3}, D_{s_4,j_4})cum(\Delta_{t_1+m-1,j_1,j_2}, \Delta_{t_2+m-1,j_3,j_4})$

$$
\begin{aligned}
&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{t_1}\sum_{s_2=1}^{t_1}\sum_{s_3=1}^{t_2}\sum_{s_4=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p}\\
&|cum(D_{s_1,j_1}, D_{s_2,j_2})||cum(D_{s_3,j_3}, D_{s_4,j_4})||cum(\Delta_{t_1+m-1,j_1,j_2}, \Delta_{t_2+m-1,j_3,j_4})|\\
\lesssim&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{1\le t_1\le t_2\le n}\sum_{s_1=1}^{t_1}\sum_{s_3=1}^{t_2}\sum_{j_1,\cdots,j_4=1}^{p}|\Gamma_{j_1,j_2}||\Gamma_{j_3,j_4}|C\rho^{t_2-t_1}\\
\le&\frac{4Cn^3}{n^4\|\Gamma\|_F^4}\left(\sum_{l=1}^{\infty}\rho^l\right)O(\|\Gamma\|_F^4)=O(n^{-1})\to 0.
\end{aligned}
$$

By similar arguments we can show the summations of cumulants for other partitions are vanished. This implies $L_4 \xrightarrow{p} 0$. Together with previous arguments, the proof of (7.4) is complete. □

7.1.2. *Proof of* (7.2). By the definition and Burkholder's inequality (Theorem 2.10, Hall and Heyde (2014)) the left hand side can be simplified as

$$\mathbb{E}\left[\left(\frac{\sqrt{2}}{n\|\Gamma\|_F}(S_n(b)-S_n(a))\right)^4\right] = \mathbb{E}\left[\left(\frac{\sqrt{2}}{n\|\Gamma\|_F}\sum_{t=\lfloor na\rfloor+1}^{\lfloor nb\rfloor}\sum_{s=1}^{t} D_{t+m}^T D_s\right)^4\right]$$

$$\lesssim \frac{4}{n^4\|\Gamma\|_F^4}\mathbb{E}\left[\left(\sum_{t=\lfloor na\rfloor+1}^{\lfloor nb\rfloor}\left(D_{t+m}^T\sum_{s=1}^{t} D_s\right)^2\right)^2\right]$$

$$\le \frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le t_1}\sum_{1\le s_3\le s_4\le t_2}\sum_{j_1,\cdots,j_4=1}^{p}$$

$$\mathbb{E}[D_{t_1+m,j_1}D_{t_1+m,j_2}D_{t_2+m,j_3}D_{t_2+m,j_4}D_{s_1,j_1}D_{s_2,j_2}D_{s_3,j_3}D_{s_4,j_4}].$$

We only need to consider the case that $\lfloor nb\rfloor-\lfloor na\rfloor\ge 1$, since otherwise it is trivially satisfied. Here we mainly deal with the expectation of 8 random variables. By similar argument in the proof of (7.4), we only consider some representative cases for the joint cumulants when we decompose the expectation. For simplicity, C is a generic constant which vary from line by line.

1. $|cum(D_{t_1+m,j_1},D_{t_1+m,j_2},D_{t_2+m,j_3},D_{t_2+m,j_4},D_{s_1,j_1},D_{s_2,j_2},D_{s_3,j_3},D_{s_4,j_4})|$

$$\frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le t_1}\sum_{1\le s_3\le s_4\le t_2}\sum_{j_1,\cdots,j_4=1}^{p}$$

$$|cum(D_{t_1+m,j_1},D_{t_1+m,j_2},D_{t_2+m,j_3},D_{t_2+m,j_4},D_{s_1,j_1},D_{s_2,j_2},D_{s_3,j_3},D_{s_4,j_4})|$$

$$\le \frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le t_1}\sum_{1\le s_3\le s_4\le t_2}\sum_{j_1,\cdots,j_4=1}^{p} C\rho^{t_2+m-s_1}$$

$$\le \frac{Cp^4}{n^4\|\Gamma\|_F^4}(\lfloor nb\rfloor-\lfloor na\rfloor)^2 n^2\sum_{l=0}^{\infty} l\rho^{l+m} \le \frac{Cp^4\rho^m}{\|\Gamma\|_F^4}\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^2}{n^2}\le C\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^2}{n^2}$$

2. $cum(D_{t_2+m,j_3},D_{t_2+m,j_4})cum(D_{t_1+m,j_1},D_{t_1+m,j_2},D_{s_1,j_1},D_{s_2,j_2},D_{s_3,j_3},D_{s_4,j_4})$

$$\frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le t_1}\sum_{1\le s_3\le s_4\le t_2}\sum_{j_1,\cdots,j_4=1}^{p}$$

$$|cum(D_{t_2+m,j_3},D_{t_2+m,j_4})cum(D_{t_1+m,j_1},D_{t_1+m,j_2},D_{s_1,j_1},D_{s_2,j_2},D_{s_3,j_3},D_{s_4,j_4})|$$

$$\lesssim \frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le s_3\le s_4\le t_1}\sum_{j_1,\cdots,j_4=1}^{p} C\rho^{t_1+m-s_1}\left(\sup_{j_1,j_2=1,\cdots,p}|\Gamma_{j_1,j_2}|\right)$$

$$+\frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le s_3\le t_1}\sum_{\lfloor na\rfloor+1\le s_4\le\lfloor nb\rfloor}\sum_{j_1,\cdots,j_4=1}^{p} C\rho^{t_1+m-s_1}\left(\sup_{j_1,j_2=1,\cdots,p}|\Gamma_{j_1,j_2}|\right)$$

$$+\frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le t_1}\sum_{\lfloor na\rfloor+1\le s_3\le s_4\le\lfloor nb\rfloor}\sum_{j_1,\cdots,j_4=1}^{p} C\rho^{t_1+m-s_1}\left(\sup_{j_1,j_2=1,\cdots,p}|\Gamma_{j_1,j_2}|\right),$$

which is further bounded by

$$\begin{aligned}&\frac{Cp^4}{n^4\|\Gamma\|_F^4}(\lfloor nb\rfloor-\lfloor na\rfloor)^2\left(\sum_{l=0}^{n} l^3\rho^{l+m}\right)+\frac{Cp^4}{n^4\|\Gamma\|_F^4}(\lfloor nb\rfloor-\lfloor na\rfloor)^3\left(\sum_{l=0}^{n} l^2\rho^{l+m}\right)\\&+\frac{Cp^4}{n^4\|\Gamma\|_F^4}(\lfloor nb\rfloor-\lfloor na\rfloor)^4\left(\sum_{l=0}^{n} l\rho^{l+m}\right)\\&\le\frac{Cp^4\rho^m}{n^2\|\Gamma\|_F^4}\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^2}{n^2}+\frac{Cp^4\rho^m}{n\|\Gamma\|_F^4}\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^3}{n^3}+\frac{Cp^4\rho^m}{\|\Gamma\|_F^4}\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^4}{n^4}\\&\le C\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^2}{n^2},\end{aligned}$$

where the last line is due to Assumption A.4, and we only need to consider the case that $\lfloor nb\rfloor\ge\lfloor na\rfloor+1$.

3. $cum(D_{t_2+m,j_3},D_{t_2+m,j_4},D_{s_1,j_1})cum(D_{t_1+m,j_1},D_{t_1+m,j_2},D_{s_2,j_2},D_{s_3,j_3},D_{s_4,j_4})$

$$\begin{aligned}&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le t_1}\sum_{1\le s_3\le s_4\le t_2}\sum_{j_1,\cdots,j_4=1}^{p}\\&|cum(D_{t_2+m,j_3},D_{t_2+m,j_4},D_{s_1,j_1})cum(D_{t_1+m,j_1},D_{t_1+m,j_2},D_{s_2,j_2},D_{s_3,j_3},D_{s_4,j_4})|\\&\lesssim\frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le s_3\le s_4\le t_1}\sum_{j_1,\cdots,j_4=1}^{p}C\rho^{t_2+m-s_1}\rho^{t_1+m-s_2}\\&+\frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le s_3\le t_1}\sum_{\lfloor na\rfloor+1\le s_4\le\lfloor nb\rfloor}\sum_{j_1,\cdots,j_4=1}^{p}C\rho^{t_2+m-s_1}\rho^{t_1+m-s_2}\\&+\frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le t_1}\sum_{\lfloor na\rfloor+1\le s_3\le s_4\le\lfloor nb\rfloor}\sum_{j_1,\cdots,j_4=1}^{p}C\rho^{t_2+m-s_1}\rho^{t_1+m-s_2}\\&\le\frac{C\rho^{2m}p^4}{\|\Gamma\|_F^4}\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^2}{n^4}+\frac{C\rho^{2m}p^4}{\|\Gamma\|_F^4}\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^3}{n^4}+\frac{C\rho^{2m}p^4}{\|\Gamma\|_F^4}\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^4}{n^4}\le C\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^2}{n^2}\end{aligned}$$

4. $cum(D_{t_1+m,j_1},D_{t_1+m,j_2},D_{t_2+m,j_3},D_{t_2+m,j_4})cum(D_{s_1,j_1},D_{s_2,j_2},D_{s_3,j_3},D_{s_4,j_4})$

$$\begin{aligned}&\frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{1\le s_1\le s_2\le t_1}\sum_{1\le s_3\le s_4\le t_2}\sum_{j_1,\cdots,j_4=1}^{p}\\&|cum(D_{t_1+m,j_1},D_{t_1+m,j_2},D_{t_2+m,j_3},D_{t_2+m,j_4})cum(D_{s_1,j_1},D_{s_2,j_2},D_{s_3,j_3},D_{s_4,j_4})|\\&\lesssim\frac{4}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\sum_{j_1,\cdots,j_4=1}^{p}|cum(D_{t_1+m,j_1},D_{t_1+m,j_2},D_{t_2+m,j_3},D_{t_2+m,j_4})|\\&\left(\sum_{1\le s_1\le s_2\le s_3\le s_4\le n}|cum(D_{s_1,j_1},D_{s_2,j_2},D_{s_3,j_3},D_{s_4,j_4})|\right)\\&\le\frac{4C}{n^4\|\Gamma\|_F^4}\sum_{\lfloor na\rfloor+1\le t_1\le t_2\le\lfloor nb\rfloor}\left(\sum_{s_1=1}^{n}\sum_{l=0}^{n}l^2\rho^l\right)O(\|\Gamma\|_F^4)\le C\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^2}{n^3}\le C\frac{(\lfloor nb\rfloor-\lfloor na\rfloor)^2}{n^2}.\end{aligned}$$

5. $cum(D_{t_1+m,j_2}, D_{t_2+m,j_3}, D_{t_2+m,j_4}) cum(D_{s_1,j_1}, D_{s_2,j_2}, D_{s_3,j_3}) cum(D_{t_1+m,j_1}, D_{s_4,j_4})$
   Notice that $cum(D_{t_1+m,j_1}, D_{s_4,j_4}) \neq 0$ only if $t_1 + m = s_4$. Since $s_4 \leq t_2$, this implies that $t_2 - t_1 \geq m$. Then

$$\begin{aligned}
&\frac{4}{n^4\|\Gamma\|_F^4} \sum_{\lfloor na\rfloor+1\leq t_1\leq t_2\leq\lfloor nb\rfloor} \sum_{1\leq s_1\leq s_2\leq t_1} \sum_{1\leq s_3\leq s_4\leq t_2} \sum_{j_1,\cdots,j_4=1}^{p} \\
&|cum(D_{t_1+m,j_2}, D_{t_2+m,j_3}, D_{t_2+m,j_4}) cum(D_{s_1,j_1}, D_{s_2,j_2}, D_{s_3,j_3}) cum(D_{t_1+m,j_1}, D_{s_4,j_4})| \\
\lesssim &\frac{4}{n^4\|\Gamma\|_F^4} \sum_{t_1=\lfloor na\rfloor+1}^{\lfloor nb\rfloor-m} \sum_{t_2=t_1+m}^{\lfloor nb\rfloor} \sum_{1\leq s_1\leq s_2\leq s_3\leq n} \sum_{j_1,\cdots,j_4=1}^{p} \\
&|cum(D_{t_1+m,j_2}, D_{t_2+m,j_3}, D_{t_2+m,j_4}) cum(D_{s_1,j_1}, D_{s_2,j_2}, D_{s_3,j_3}) \Gamma_{j_1,j_4}| \\
\leq &\frac{Cp^2\rho^m}{n^4\|\Gamma\|_F^2} (\lfloor nb\rfloor - \lfloor na\rfloor)^2 \sum_{s_1=1}^{n} \sum_{l=0}^{\infty} l\rho^l \leq C\left(\frac{\lfloor nb\rfloor - \lfloor na\rfloor}{n}\right)^2
\end{aligned}$$

6. $cum(D_{t_1+m,j_1}, D_{t_1+m,j_2}) cum(D_{t_2+m,j_3}, D_{t_2+m,j_4}) cum(D_{s_1,j_1}, D_{s_2,j_2}) cum(D_{s_3,j_3}, D_{s_4,j_4})$

$$\begin{aligned}
&\frac{4}{n^4\|\Gamma\|_F^4} \sum_{\lfloor na\rfloor+1\leq t_1\leq t_2\leq\lfloor nb\rfloor} \sum_{1\leq s_1\leq s_2\leq t_1} \sum_{1\leq s_3\leq s_4\leq t_2} \sum_{j_1,\cdots,j_4=1}^{p} \\
&|cum(D_{t_1+m,j_1}, D_{t_1+m,j_2}) cum(D_{t_2+m,j_3}, D_{t_2+m,j_4}) cum(D_{s_1,j_1}, D_{s_2,j_2}) cum(D_{s_3,j_3}, D_{s_4,j_4})| \\
= &\frac{4}{n^4\|\Gamma\|_F^4} \sum_{\lfloor na\rfloor+1\leq t_1\leq t_2\leq\lfloor nb\rfloor} \sum_{s_1=1}^{t_1} \sum_{s_3=1}^{t_2} \sum_{j_1,\cdots,j_4=1}^{p} \Gamma_{j_1,j_2}^2 \Gamma_{j_3,j_4}^2 \\
\leq &\frac{4(\lfloor nb\rfloor - \lfloor na\rfloor)^2}{n^2}
\end{aligned}$$

Other partitions can be proved by similar arguments. Combining the results we have

$$\mathbb{E}\left[\left(\frac{\sqrt{2}}{n\|\Gamma\|_F} S_n(b) - \frac{\sqrt{2}}{n\|\Gamma\|_F} S_n(a)\right)^4\right] \leq C\frac{(\lfloor nb\rfloor - \lfloor na\rfloor)^2}{n^2}$$

for every $n \geq n_0$ for some fixed $n_0 > 0$. □

R. Wang
Department of Statistics
University of Illinois at Urbana-Champaign
725 South Wright Street
Champaign, IL 61820
USA
E-mail: rwang52@illinois.edu

X. Shao
Department of Statistics
University of Illinois at Urbana-Champaign
725 South Wright Street
Champaign, IL 61820
USA
E-mail: xshao@illinois.edu