Statistica Sinica Preprint No: SS-2022-0147

Title	Mean Tests For High-dimensional Time Series
Manuscript ID	SS-2022-0147
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0147
Complete List of Authors	Shuyi Zhang, Song Chen and Yumou Qiu
Corresponding Authors	Song Chen
E-mails	csx@gsm.pku.edu.cn

doi:10.5705/ss.202022.0147

Statistica Sinica

MEAN TESTS FOR HIGH-DIMENSIONAL TIME SERIES

Shuyi Zhang*, Song Xi Chen† and Yumou Qiu†

*East China Normal University and †Peking University

Abstract: This study considers testing for two-sample mean differences in high-dimensional temporally dependent data, which we then extend to the one-sample situation. To eliminate the bias caused by the temporal dependence in the time series observations, we propose a band-excluded U-statistic (BEU-statistic) to estimate the squared Euclidean distance between the two means that excludes cross-products of data vectors of temporally close time points. We derive the asymptotic normality of the BEU-statistic for the high-dimensional setting with "spatial" (column-wise) and temporal dependence. We also develop an estimator built on the kernel-smoothed cross-time covariances to estimate the variance of the BEU-statistic, facilitating a test procedure based on the standardized BEU-statistic. The proposed test is nonparametric and adaptive to a wide range of dependence and dimensionality, and has attractive power properties relative to those of a self-normalized test. A numerical simulation and a real-data analysis on the return and volatility of S&P 500 stocks before and after the 2008 financial crisis demonstrate the performance and utility of the proposed test.

Key words and phrases: High dimensionality, long-run variance estimation, L_2-type test, spatial and temporal dependence, U-statistics.

Joint Corresponding Authors: Song Xi Chen, Guanghua School of Management and Center for Statistical Science, Peking University, Beijing 100871, China, e-mail: csx@gsm.pku.edu.cn; Yumou Qiu, School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing 100871, China, e-mail: qiuyumou@math.pku.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

1. Introduction

High-dimensional data characterized by simultaneous measurements of many variables are common in social, economic, and environmental studies, especially in spatial econometrics (Arbia, 2016) and financial econometrics (Fan et al., 2020). The data are usually temporally dependent, and there is dependence between the high-dimensional components at each cross-section of time. Examples include high-frequency financial data on asset returns (Fan et al., 2011; Liu and Chen, 2020), economic panel data (Stock and Watson, 2002) with a large number of recorded variables, and large-scale spatio-temporal data from atmospheric environmental and climate change studies (Xu et al., 2020). Thus, we require inferences for mean vectors with both high dimensionality and temporal dependence in order to evaluate the treatment effects in such studies (Fan et al., 2015).

This study aims to provide an effective testing procedure for detecting differences in the means of two groups under various treatments, where the data exhibit temporal dependence and high dimensionality. The conventional Hotelling's test, published when the author became an economic Professor at Columbia University, and designed for independent and identically distributed (i.i.d.) data with a fixed dimension, cannot be used to evaluate a high-dimensional treatment effect (Bai and Saranadasa, 1996). Two-sample mean tests for high-dimensional data have been proposed, largely for i.i.d. data, including the L_2-type tests of Bai and Saranadasa (1996) based on a bias-corrected Euclidean statistic, and Chen and Qin (2010) formulated with U-statistics. These tests avoid using the sample covariance, owing to its adverse effects in high-dimensional settings. See also Chen et al. (2011), Feng et al. (2015), and Wang et al. (2015) for other formulations of L_2-type tests. Another test procedure is the L_∞ (maximum)-test, which takes the maximum standardized difference between all dimensions of the two sample means (Chernozhukov et al., 2013; Cai et al.,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

2014; Chang et al., 2017). A third type is the L_2 thresholding test (Zhong et al., 2013; Qiu et al., 2018; Chen et al., 2023), which improves on the higher criticism tests (Donoho and Jin, 2015; Hall and Jin, 2010). These tests apply a thresholding procedure in the marginal differences of two-sample means to exclude non-signal-bearing dimensions, and thus enhance the signal-to-noise ratio for better power under the sparse and faint signal setting; see also Huang et al. (2021) for high-dimensional mean tests.

Compared with studies on independent data, few works test for high-dimensional means in temporally dependent data, which are common in economics, mainly because of the difficulties in dealing with the temporal dependence, while accounting for the column-wise dependence between the high-dimensional components. Chernozhukov et al. (2019) extended the Gaussian approximation results for the maximum statistics to weakly dependent data under the β-mixing conditions. Using this result, they constructed an L_∞-test using a kernel-based multiplier bootstrap procedure; see Chang et al. (2018) and Qiu and Zhou (2022) for global and multiple testing procedures for high-dimensional precision and partial correlation matrices. However, the maximum test is less advantageous for detecting weak signals. For L_2-type tests, Ayyala et al. (2017) extended the procedure of Bai and Saranadasa (1996) to m-dependent Gaussian data, under the moderate dimensionality where p and n are of the same order. Wang and Shao (2020) considered one-sample testing for a high-dimensional mean using a U-statistic formulation under physical dependence with a geometric decaying rate. Instead of estimating the variance of the statistic, Wang and Shao (2020) construct the test using self-normalization.

In this study, we consider testing for two-sample means in high-dimensional weakly dependent time series data, without the Gaussian assumption, while allowing for exponential growth of the dimension. The L_2-type U-statistics originally proposed by Chen and Qin (2010) for independent data are no longer unbiased for $||\mu_1 - \mu_2||^2$, the squared Euclidean distance between two population means μ_1 and μ_2. Thus, for temporally dependent data, we construct a band-excluded U-statistic (BEU-statistic) for the two-sample setting that removes pairs of temporally close observations. We derive the asymptotic normality of the proposed test statistic under general weak column-wise dependence and temporal dependence, where the dimension can be much larger than the sample size. We also develop a kernel-smoothing method over the cross-time long-run covariances to estimate the variance of the test statistic, and propose a testing procedure with data-driven tuning parameter selection for the exclusion and smoothing bandwidths. We establish the theoretical properties of the proposed test under the null and alternative hypotheses, showing its proper asymptotic size control and power for dense and weak signals. The power of the proposed test is analyzed under both local and fixed alternatives. Then, we extend the test formulation to the one-sample setting, which is shown to be more powerful than the self-normalized (SN) test of Wang and Shao (2020) in both theoretical results and numerical simulations. Simulation studies are used to evaluate the performance and confirm the theoretical properties. Under the capital asset pricing model, we apply the proposed method to compare the adjusted returns of S&P 500 stocks by market index, and their specific volatility, before and after the 2008 financial crisis. Our results indicate that the crisis did not lead to a significant difference in the adjusted returns, but it did increase the volatility.

The remainder of the paper is organized as follows. Section 2 outlines the assumptions on the data distribution and the temporal dependence. The U-statistic formulation is introduced in Section 3, with the theoretical result on its asymptotic normality. Section 4 constructs the variance estimator of the proposed statistic and shows its ratio consistency. Section 5 provides the implementation and data adaptive tuning parameter selection for the proposed test. Section 6 analyzes the power of the proposed test and compares it with that of the SN test. Sections 7 and 8 report results from simulation studies and a real-data

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

analysis, respectively, on S&P 500 stock returns. All technical proofs are relegated to the Supplementary Material.

2. Preliminaries

Suppose we observe p-dimensional stationary time series $\{\boldsymbol{X}_{i,t}\}_{t=1}^{n_i}$ from two populations for $i = 1$ and 2, where $\boldsymbol{X}_{i,t} = (X_{i,t,1}, X_{i,t,2}, \ldots, X_{i,t,p})^{\mathrm{T}}$, and n_1 and n_2 are the respective sample sizes. We assume mutual independence between the two samples, and allow temporal dependence within each sample. Let $\boldsymbol{\mu}_i = (\mu_{i,1}, \ldots, \mu_{i,p})^{\mathrm{T}}$ and $\boldsymbol{\Sigma}_{i,0}$ be the mean and covariance matrix, respectively, of $\boldsymbol{X}_{i,1}$. Define the cross-covariance matrices $\boldsymbol{\Sigma}_{i,k} = \mathrm{Cov}(\boldsymbol{X}_{i,t+k}, \boldsymbol{X}_{i,t}) = (\sigma_{i,k,j_1j_2})_{p\times p}$, for $k = -(n_i - 1), \ldots, n_i - 1$, and $\boldsymbol{\Sigma}_{i,0}$ is the marginal covariance. Let $\boldsymbol{\Sigma}_{i,\infty} = \sum_{k=-\infty}^{+\infty} \boldsymbol{\Sigma}_{i,k} = (\sigma_{i,\infty,j_1j_2})_{p\times p}$ be the long-run covariance matrix of $\boldsymbol{X}_{i,t}$, provided $\{\sigma_{i,k,j_1j_2}\}$ are summable over k for all j_1, j_2.

Our aim is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{vs.} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \tag{2.1}$$

These are global hypotheses for two-sample means, which are studied extensively under independent data (Donoho and Jin, 2004; Chen and Qin, 2010; Feng et al., 2015; Wang et al., 2015). However, except for the work of Ayyala et al. (2017) for m-dependent data, the two-sample mean test for temporally dependent high-dimensional observations has not been studied sufficiently in the literature.

We make the following assumptions in the analysis.

Assumption 1. (i) $n_1/(n_1 + n_2) \to \kappa_0 \in (0, 1)$ as n_1, $n_2 \to \infty$. (ii) For a positive integer q and a constant $\Delta > 0$, $\max_{1\le i\le 2, 1\le j\le p} \mathrm{E}^{1/q}(|X_{i,t,j}|^q) \le \Delta$.

Assumption 2. Each $\boldsymbol{X}_{i,t}$ is generated from a linear innovation model such that $\boldsymbol{X}_{i,t} = \boldsymbol{\Gamma}_i \boldsymbol{Z}_{i,t} + \boldsymbol{\mu}_i$, for $i = 1, 2$, where $\boldsymbol{\Gamma}_i$ is a $p \times r$ matrix with $r \ge p$ such that $\boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_i^{\mathrm{T}} = \boldsymbol{\Sigma}_{i,\infty}$,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

and $\boldsymbol{Z}_{i,t} = (Z_{i,t,1}, Z_{i,t,2}, \ldots, Z_{i,t,r})^{\mathrm{T}}$ is the innovation random vector with $\mathrm{E}\big(\boldsymbol{Z}_{i,t}\big) = \mathbf{0}$. For each j, $\{Z_{i,t,j}\}_{t=1}^{n_i}$ is a second-order stationary time series with unit long-run variance, and $\max_{1\le j\le r} \mathrm{E}(Z_{i,t,j}^8) \le \Delta_z$ for a positive constant Δ_z and $i = 1, 2$. Furthermore, Z_{i,t_1,j_1} and Z_{i,t_2,j_2} are uncorrelated for any t_1 and t_2 if $j_1 \neq j_2$. For any sequences of time points $\{t_{11}, \ldots, t_{1a_1}\}, \ldots, \{t_{l1}, \ldots, t_{la_l}\}$ with $\sum_{k=1}^{l} a_k \le 8$ and distinct $j_1, \ldots, j_l$,

$$\mathrm{E}\left\{\prod_{k=1}^{l}\left(Z_{i,t_{k1},j_k} \ldots Z_{i,t_{ka_k},j_k}\right)\right\} = \prod_{k=1}^{l} \mathrm{E}\big(Z_{i,t_{k1},j_k} \cdots Z_{i,t_{ka_k},j_k}\big). \tag{2.2}$$

Assumption 1 (i) is a conventional assumption in two-sample problems, and (ii) is needed for the Davydov's inequality (Davydov, 1968) to control the temporal correlation between $\boldsymbol{X}_{i,t_1}$ and $\boldsymbol{X}_{i,t_2}$ under mixing conditions. Assumption 2 extends the linear innovation models of Bai and Saranadasa (1996) and Cui et al. (2020) for i.i.d. data to temporally dependent data. Assumption 2 uses a linear process model for data generation, with $\{Z_{i,t}\}_{t=1}^{n_i}$ as the innovation process. The linear process is widely used in time series analysis (Brockwell and Davis, 1991). Although a linear generation of multivariate data is considered in Bai and Saranadasa (1996) and other works for the independent setting, in contrast to these works, the innovation process here is temporarily dependent. For each time t, the innovation vector $\boldsymbol{Z}_{i,t}$ is assumed to be nearly independent to allow wider forms of innovation distributions. We could just assume $\boldsymbol{Z}_{i,t}$ has an independent column vector. However, the theoretical derivation can be made without the full independence, and assuming the weaker equation (2.2) is sufficient. Examples of (2.2) for non-independent cases can be found for non-Gaussian distributed data.

Let $\boldsymbol{\Sigma}_{i,k}^{z} = \mathrm{Cov}(\boldsymbol{Z}_{i,t+k}, \boldsymbol{Z}_{i,t})$ be the cross-time covariance for any integer k, and $\boldsymbol{\Sigma}_{i,\infty}^{z} = \sum_{k=-\infty}^{\infty} \boldsymbol{\Sigma}_{i,k}^{z}$ be the long-run covariance of $\boldsymbol{Z}_{i,t}$. Under Assumption 2, $\boldsymbol{\Sigma}_{i,k}^{z}$ is diagonal, satisfying $\boldsymbol{\Sigma}_{i,-k}^{z} = \boldsymbol{\Sigma}_{i,k}^{z}$ and $\boldsymbol{\Sigma}_{i,\infty}^{z} = \boldsymbol{I}_r$, where $\boldsymbol{I}_r$ is the $r \times r$ identity matrix. Moreover, $\boldsymbol{\Sigma}_{i,k} = \boldsymbol{\Gamma}_i \boldsymbol{\Sigma}_{i,k}^{z} \boldsymbol{\Gamma}_i^{\mathrm{T}}$ and $\boldsymbol{\Sigma}_{i,\infty} = \boldsymbol{\Gamma}_i \boldsymbol{\Sigma}_{i,\infty}^{z} \boldsymbol{\Gamma}_i^{\mathrm{T}} = \boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_i^{\mathrm{T}}$, for $i = 1, 2$. The condition of a unit long-run

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

variance of $\{Z_{i,t,j}\}$ for each j is not essential, because we can apply rescaling on $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Sigma}^z_{i,\infty}$ simultaneously to make it so. Note that the so-called column-wise dependence between the components of $\boldsymbol{X}_{i,t}$ is mostly induced by the matrices $\boldsymbol{\Gamma}_i$, whereas the temporal dependence of $\boldsymbol{X}_{i,t}$ results from the temporal dependence of the univariate innovations $\{Z_{i,t,j}\}$ over time for all $j = 1, \ldots, r$. If the elements of $\boldsymbol{\Sigma}_{i,\infty}$ are bounded and the diagonal values of $\boldsymbol{\Sigma}^z_{i,k}$ decrease to zero uniformly as the time lag k increases, this leads to all elements in $\boldsymbol{\Sigma}_{i,k}$ decaying to zero uniformly.

In spatial and temporal statistics, separability is a common assumption on covariances. The covariance structure of $\boldsymbol{X}_{i,t}$ implied from Assumption 2 includes the separable covariances as a special case. To see this, let $\mathbb{X}_i = (\boldsymbol{X}^{\mathrm{T}}_{i,1}, \boldsymbol{X}^{\mathrm{T}}_{i,2}, \ldots, \boldsymbol{X}^{\mathrm{T}}_{i,n_i})^{\mathrm{T}}$ be the vectorization of the data over all time points. Correspondingly, let $\mathbb{Z}_i = (\boldsymbol{Z}^{\mathrm{T}}_{i,1}, \boldsymbol{Z}^{\mathrm{T}}_{i,2}, \ldots, \boldsymbol{Z}^{\mathrm{T}}_{i,n_i})^{\mathrm{T}}$ and $\mathbb{G}_i = \mathrm{diag}(\boldsymbol{\Gamma}_i, \ldots, \boldsymbol{\Gamma}_i) = \boldsymbol{I}_{n_i} \otimes \boldsymbol{\Gamma}_i$, where $\otimes$ denotes the Kronecker product. It can be shown that $\mathrm{Var}(\mathbb{X}_i) = \mathbb{G}_i \mathrm{Var}(\mathbb{Z}_i) \mathbb{G}^{\mathrm{T}}_i$. Let $\boldsymbol{\Sigma}^z_{i,k} = \mathrm{diag}\{\sigma^z_{i,k,1}, \ldots, \sigma^z_{i,k,r}\}$, with diagonal elements $\{\sigma^z_{i,k,l}\}^r_{l=1}$. If $\sigma^z_{i,k,1} = \ldots = \sigma^z_{i,k,r} = \sigma^z_{i,k}$, for all k, we have $\mathrm{Var}(\mathbb{Z}) = \boldsymbol{C}_i \otimes \boldsymbol{I}_r$, where $\boldsymbol{C}_i = (\sigma^z_{i,k_1-k_2})_{n_i \times n_i}$. This implies that $\mathrm{Var}(\mathbb{X}_i) = \boldsymbol{C}_i \otimes \boldsymbol{\Gamma}_i \boldsymbol{\Gamma}^{\mathrm{T}}_i$, where $\boldsymbol{C}_i$ and $\boldsymbol{\Gamma}_i \boldsymbol{\Gamma}^{\mathrm{T}}_i$ are the temporal and spatial dependence, respectively, of $\mathbb{X}_i$. Therefore, if the diagonal elements of $\boldsymbol{\Sigma}^z_{i,k}$ are identical for each k, meaning all the univariate innovation time series $\{Z_{i,t,j}\}$ have the same cross-time covariances, $\mathbb{X}_i$ has a separable covariance matrix.

We assume the temporal dependence in the innovation time series $\{\boldsymbol{Z}_{i,t}\}$ is β-mixing, with the mixing coefficient

$$\beta^z_i(k) = \sup_t \mathrm{E}\Big\{ \sup_{B \in \mathcal{F}^{\infty}_{i,t+k}} \big|\mathbb{P}(B|\mathcal{F}^t_{i,-\infty}) - \mathbb{P}(B)\big| \Big\},$$

where $\mathcal{F}^t_{i,-\infty} = \sigma(\boldsymbol{Z}_{i,s}, s \le t)$ and $\mathcal{F}^{\infty}_{i,t+k} = \sigma(\boldsymbol{Z}_{i,s}, s \ge t+k)$ are the σ-fields generated by $\{\boldsymbol{Z}_{i,s}\}_{s \le t}$ and $\{\boldsymbol{Z}_{i,s}\}_{s \ge t+k}$, respectively. The β-mixing condition is as follows.

Assumption 3. For $a, c > 0$, the β-mixing coefficients of the innovation process $\{\boldsymbol{Z}_{i,t}\}$

satisfy $\max\{\beta_1^z(k), \beta_2^z(k)\} \le c\exp\{-ak\}$, for all positive integer k.

Let $\beta_i^x(k)$ be the β-mixing coefficient of $\{\boldsymbol{X}_{i,t}\}$. Because $\{\boldsymbol{X}_{i,t}\}$ are generated linearly by $\{\boldsymbol{Z}_{i,t}\}$, $\beta_i^x(k) \le \beta_i^z(k)$. Thus, Assumption 3 implies $\max\{\beta_1^x(k), \beta_2^x(k)\} \le c\exp\{-ak\}$. This condition is needed by the coupling method to derive the asymptotic distribution of the test statistic under time-dependent data. Similar conditions are made for high-dimensional inference in Chang et al. (2018), Chernozhukov et al. (2019), and Wong et al. (2020). Note that the exponential decay can be relaxed to a polynomial decay, with more involved technical derivations. We discuss the polynomial decay further after presenting the main theorems.

Under the fixed-dimension scenario, the β-mixing condition is a mild assumption in the time series literature. It is known that causal ARMA processes with continuous innovation distributions, stationary Markov chains with certain conditions, and stationary GARCH models with finite second moments and continuous innovation distributions all satisfy the β-mixing condition; see Doukhan (1994) and Bradley (2005). Under the high-dimensional scenario, where the dimension increases with the sample size, the β-mixing condition is more restrictive. Theorem 3.2 in Han and Wu (2023) provides a lower bound $\tilde{\beta}(k) \ge 1 - 2\exp(-\tilde{c}_1 p \tilde{\tau}^{2k})$ on the β-mixing coefficient for a high-dimensional stationary vector AR model $\tilde{Z}_{t,j} = \tilde{\tau}\tilde{Z}_{t,j} + \tilde{\epsilon}_{t,j}$, for $j = 1, \ldots, p$, with a common coefficient $\tilde{\tau}$ and i.i.d. innovation noise $\{\tilde{\epsilon}_{t,j}\}$, where $\tilde{c}_1$ is a positive constant. It implies that $\inf \lim_n \tilde{\beta}(n) > 0$ if $\lim_n \log p\tilde{\tau}^{2n} \ge \log 2/\tilde{c}_1$. However, under the condition of $b \ge c_1(\log n + \log p)$, for some c_1, as assumed in Proposition 1, this lower bound becomes trivial if $c_1 \ge -(2\log\kappa)^{-1}$, which can be achieved by choosing a sufficiently large c_1. Hence, the β-mixing condition in Assumption 3 is not violated. At the same time, the asymptotic results of the proposed tests should still be valid, under certain conditions on the τ-measure of dependence (Dedecker

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

and Prieur, 2005), following a similar investigation in Qiu and Zhou (2022), although the theoretical proof is more involved.

Assumption 4. For any $i, i_1, i_2 \in \{1, 2\}$ and $\boldsymbol{\mu}_{i_1}$, $\boldsymbol{\mu}_{i_2}$ such that $\boldsymbol{\mu}_{i_1}^{\mathrm{T}} \boldsymbol{\Sigma}_{i,\infty} \boldsymbol{\mu}_{i_2} \neq \mathbf{0}$, there exists a $C_0 > 0$ such that

$$\max\left\{ \limsup_{p\to+\infty} \sum_{k_1,k_2=-\infty}^{\infty} \frac{|\mathrm{tr}(\boldsymbol{\Sigma}_{i,k_1}\boldsymbol{\Sigma}_{i,k_2})|}{\mathrm{tr}(\boldsymbol{\Sigma}_{i,\infty}^2)}, \limsup_{p\to+\infty} \sum_{k=-\infty}^{+\infty} \frac{|\boldsymbol{\mu}_{i_1}^{\mathrm{T}}\boldsymbol{\Sigma}_{i,k}\boldsymbol{\mu}_{i_2}|}{|\boldsymbol{\mu}_{i_1}^{\mathrm{T}}\boldsymbol{\Sigma}_{i,\infty}\boldsymbol{\mu}_{i_2}|} \right\} \leq C_0.$$

Assumption 5. For two positive constants η and C_1, $\min(\lambda_{1,\min}, \lambda_{2,\min}) \geq C_1 p^{-\eta}$, where $\lambda_{i,\min}$ is the minimum eigenvalue of $\boldsymbol{\Sigma}_{i,\infty}$, for $i = 1, 2$.

Assumptions 4 and 5 are mild technical conditions for deriving the asymptotic distribution of the proposed test statistic. Note that $\mathrm{tr}(\boldsymbol{\Sigma}_{i,\infty}^2) = \sum_{k_1,k_2} \mathrm{tr}(\boldsymbol{\Sigma}_{i,k_1}\boldsymbol{\Sigma}_{i,k_2})$. Assumption 4 requires $\{\mathrm{tr}(\boldsymbol{\Sigma}_{i,k_1}\boldsymbol{\Sigma}_{i,k_2})/\mathrm{tr}(\boldsymbol{\Sigma}_{i,\infty}^2)\}$ to be summable, which is analogous to the absolute summable condition on cross-time covariances of univariate time series. Similar conditions are made in Wang and Shao (2020) for one-sample testing. Assumption 5 puts a lower bound on the minimum eigenvalue of $\boldsymbol{\Sigma}_{i,\infty}$, which is allowed to diminish to zero.

3. BEU-statistic

We consider the L_2-type statistics estimating $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, the overall difference between the two population means. Let $\bar{\boldsymbol{X}}_i = \sum_{t=1}^{n_i} \boldsymbol{X}_{i,t}/n_i = (\bar{X}_{i,1}, \ldots, \bar{X}_{i,p})^{\mathrm{T}}$ be the sample means for $i = 1, 2$. Under Assumptions 2 and 3, it can be shown that $\bar{X}_{i,j}$ is asymptotic normal with mean $\mu_{i,j}$ and variance $\sigma_{i,\infty,jj}/n_i$, for all $j = 1, \ldots, p$. The leading-order term of $\mathrm{E}(\bar{X}_{i,j}^2)$ is $\mu_{i,j}^2 + \sigma_{i,\infty,jj}/n_i$, where the bias $\sigma_{i,\infty,jj}/n_i$ accumulates in the mean of the L_2-statistic $\sum_{j=1}^{p} \bar{X}_{i,j}^2$, and diverges when the dimension p is much larger than the sample size.

To reduce the bias induced by the temporal dependence, we construct

$$U_{i,j}(b) = \frac{1}{n_i(b)} \sum_{|t_1 - t_2| \geq b} X_{i,t_1,j} X_{i,t_2,j} \tag{3.1}$$

as an estimator for $\mu_{i,j}^2$, for $i = 1, 2$ and $j = 1, \ldots, p$. Here, b is a positive tuning parameter that defines a temporal exclusion band of width b to exclude products $X_{i,t_1}^{\mathrm{T}} X_{i,t_2}$ in t_1 and t_2 that are less than b apart in the above statistic, and $n_i(b) = (n_i - b)(n_i - b + 1)$ is the number of terms in the summation of (3.1). We take $b \to \infty$ as $n_i \to \infty$.

Let $V_j(b) = U_{1,j}(b) + U_{2,j}(b) - 2\bar{X}_{1,j}\bar{X}_{2,j}$ be the estimator of $(\mu_{1,j} - \mu_{2,j})^2$, for $j = 1, \ldots, p$. Summing $V_j(b)$ over j, we propose the BEU-statistic

$$T(b) = \frac{1}{n_1(b)} \sum_{|t_1 - t_2| \geq b} \boldsymbol{X}_{1,t_1}^{\mathrm{T}} \boldsymbol{X}_{1,t_2} + \frac{1}{n_2(b)} \sum_{|t_1 - t_2| \geq b} \boldsymbol{X}_{2,t_1}^{\mathrm{T}} \boldsymbol{X}_{2,t_2} - \frac{2}{n_1 n_2} \sum_{t_1=1}^{n_1} \sum_{t_2=1}^{n_2} \boldsymbol{X}_{1,t_1}^{\mathrm{T}} \boldsymbol{X}_{2,t_2} \tag{3.2}$$

as an estimator for $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2$.

Note that $T(0) = (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)^{\mathrm{T}}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)$ is the L_2-statistic used in Bai and Saranadasa (1996), and $T(1)$ is the U-statistic proposed by Chen and Qin (2010) for independent observations. The exclusion band of $|t_1 - t_2| \geq b$ removes pairs of observations $\boldsymbol{X}_{i,t_1}$ and $\boldsymbol{X}_{i,t_2}$ in (3.2) that are more strongly correlated. This effectively mitigates the bias of $T(b)$ induced by the temporal dependence. Bias reduction is the key to constructing L_2-type statistics for high-dimensional data, because the accumulation of the bias from each component decreases the asymptotic performance of the L_2-statistics if p is much larger than n (Feng et al., 2015).

Let $T_1(b) = n_1(b)^{-1} \sum_{|t_1 - t_2| \geq b} \boldsymbol{X}_{1,t_1}^{\mathrm{T}} \boldsymbol{X}_{1,t_2}$ be the first term on the right-hand side of (3.2). Then, $T_1(b)$ can be used to test the one-sample hypothesis $H_0 : \boldsymbol{\mu}_1 = \mathbf{0}$ vs. $H_1 : \boldsymbol{\mu}_1 \neq \mathbf{0}$, and a location shift enables us to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_{10}$ vs. $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_{10}$, for a known $\boldsymbol{\mu}_{10}$. Note that $T_1(b)$ is the statistic considered in Wang and Shao (2020) for testing $H_0 : \boldsymbol{\mu}_1 = \mathbf{0}$ under geometric regularized physical dependence. Instead of estimating the variance of $T_1(b)$, they use a self-normalized technique to formulate the testing procedure. We present

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

a power comparison between the proposed test and the SN test in Section 6 for the one-sample case.

Our plan is to derive and estimate the variance of $T(b)$, and to construct a test for the hypotheses (2.1) based on a standardized version of $T(b)$. For a positive integer k, let

$$\boldsymbol{M}_k = \kappa_0^{-1}\boldsymbol{\Sigma}_{1,k} + (1-\kappa_0)^{-1}\boldsymbol{\Sigma}_{2,k} \text{ and } \boldsymbol{M}_\infty = \sum_{k=-\infty}^{+\infty} \boldsymbol{M}_k \tag{3.3}$$

be the weighted lag-k cross-covariance and the weighted long-run covariance, respectively, of the two-sample time series. The following proposition provides the mean and variance of the BEU-statistic $T(b)$, which shows that $T(b)$ is asymptotically unbiased for $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2$.

Proposition 1. *Under Assumption 1 with $q > 4$ and Assumptions 2–5, if $\log p = o(n)$ and the exclusion bandwidth satisfies $b = o(n)$ and $b \geq c_1(\log n + \log p)$, for a positive constant c_1, we have as $n, p \to \infty$,*

$$\mathrm{E}\{T(b)\} = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2 + \sum_{i=1}^{2} \frac{2}{n_i(b)} \sum_{k=b}^{n_i-1} (n_i - k)\mathrm{tr}(\boldsymbol{\Sigma}_{i,k}) \text{ and}$$

$$\mathrm{Var}\{T(b)\} = \big\{2n^{-2}\mathrm{tr}(\boldsymbol{M}_\infty^2) + 4n^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}}\boldsymbol{M}_\infty(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\big\}\{1 + o(1)\}.$$

In Proposition 1, we allow an exponential growth rate of p relative to n. The proposition shows that the bias of $T(b)$ is asymptotically equal to $2n^{-1}\sum_{k\geq b}\mathrm{tr}(\boldsymbol{M}_k)$ by noting that $n_i(b) = (n_i - b)(n_i - b + 1)$, which is determined by the auto-covariance of $\{\boldsymbol{X}_{1,t}\}$ and $\{\boldsymbol{X}_{2,t}\}$ with time-lag larger than b. This bias term diminishes to zero at a polynomial rate of p and n if $b \geq c_1(\log n + \log p)$, for sufficiently large c_1, under Assumption 3.

Corollary 1. *Under the conditions of Proposition 1 and the null hypothesis of (2.1),*

$$\mathrm{Var}\{T(b)\} = \left\{\frac{2}{n_1^2}\mathrm{tr}(\boldsymbol{\Sigma}_{1,\infty}^2) + \frac{2}{n_2^2}\mathrm{tr}(\boldsymbol{\Sigma}_{2,\infty}^2) + \frac{4}{n_1 n_2}\mathrm{tr}(\boldsymbol{\Sigma}_{1,\infty}\boldsymbol{\Sigma}_{2,\infty})\right\}\{1 + o(1)\}. \tag{3.4}$$

Corollary 1 provides the variance of the BEU-statistic under the null hypothesis. An estimator of this variance is constructed in Section 4, and is used to formulate the proposed

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

testing procedure for hypotheses (2.1). For the one-sample problem, from the proof of Proposition 1, it can be shown that the leading variance of $T_1(b)$ is $2n_1^{-2}\mathrm{tr}(\mathbf{\Sigma}_{1,\infty}^2)$.

To derive the asymptotic normality of the BEU-statistic $T(b)$, we use the coupling method for time series and the martingale central limit theorem (Hall and Heyde, 1980) for the U-statistics. For both samples, we partition the time points $\{1,\ldots,n_i\}$ into a sequence of large segments of length a_1, followed by small segments of length a_2, where $a_2 = o(a_1)$. Let $d_i = \lfloor n_i/(a_1+a_2)\rfloor$ be the total number of large and small segments, for $i = 1, 2$, where $\lfloor\cdot\rfloor$ denotes the floor function. Let $\bar{\boldsymbol{X}}_{i,m}$ be the average of $\boldsymbol{X}_{i,t}$ over the mth-largest segment, for $m = 1,\ldots,d_i$ and $i = 1, 2$. By the coupling method, $\bar{\boldsymbol{X}}_{i,m_1}$ and $\bar{\boldsymbol{X}}_{i,m_2}$ can be regarded as independent, because they are separated by at least one small block. Therefore, the averages $\{\bar{\boldsymbol{X}}_{i,m}\}_{m=1}^{d_i}$ over the large blocks can be regarded as independent, and the martingale central limit theorem for independent observations can be applied to show the asymptotic normality of $T(b)$ under temporally dependent data. Detailed technical derivations are provided in the proof of Theorem 1 in the Supplementary Material.

To obtain the limiting distribution, we impose a condition on the trace of the long-run covariance $\mathbf{\Sigma}_{i,\infty}$, which is used to bound the higher moments of the data. A similar condition is made on $\mathbf{\Sigma}_{i,0}$ for independent data in Feng et al. (2015) and Wang et al. (2015).

Assumption 6. $\mathrm{tr}\big(\mathbf{\Sigma}_{i_1,\infty}\mathbf{\Sigma}_{i_2,\infty}\mathbf{\Sigma}_{i_3,\infty}\mathbf{\Sigma}_{i_4,\infty}\big) = o\big[\mathrm{tr}^2\{(\mathbf{\Sigma}_{1,\infty}+\mathbf{\Sigma}_{2,\infty})^2\}\big]$, for $i_1, i_2, i_3, i_4 = 1, 2$.

Let $\lambda_{i,\min}$ and $\lambda_{i,\max}$ be the minimum and maximum eigenvalues, respectively, of $\mathbf{\Sigma}_{i,\infty}$. Assumption 6 is valid if all the eigenvalues of $\mathbf{\Sigma}_{i,\infty}$ are bounded from zero and infinity. If $\lambda_{i,\max}$ are bounded away from infinity and $\lambda_{i,\min} = O(p^{\eta})$, one needs $\eta > -1/4$ to ensure Assumption 6. On the other hand, if $\lambda_{i,\min}$ are bounded from zero and $\lambda_{i,\max} = O(p^{\xi})$,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

Assumption 6 is valid if $\xi < 1/4$. More generally, if the eigenvalues are diverging such that $\lambda_{i,\min} = \gamma_{i,1} p^{\eta}$ and $\lambda_{i,\max} = \gamma_{i,2} p^{\xi}$ for some positive constants $\gamma_{i,1}$ and $\gamma_{i,2}$, then

$$\frac{\operatorname{tr}\left(\boldsymbol{\Sigma}_{i_1,\infty}\boldsymbol{\Sigma}_{i_2,\infty}\boldsymbol{\Sigma}_{i_3,\infty}\boldsymbol{\Sigma}_{i_4,\infty}\right)}{\operatorname{tr}^2\left\{(\boldsymbol{\Sigma}_{1,\infty}+\boldsymbol{\Sigma}_{2,\infty})^2\right\}} \le \frac{\gamma_{i_1,2}\gamma_{i_2,2}\gamma_{i_3,2}\gamma_{i_4,2}p^{4(\xi-\eta)-1}}{(\gamma_{1,1}+\gamma_{2,1})^4} \to 0 \quad \text{as } p \to \infty$$

for $i_1, i_2, i_3, i_4 \in \{1, 2\}$ and if $\xi - \eta < 1/4$.

The following theorem states the asymptotic normality of the BEU-statistic $T(b)$.

Theorem 1. *Under the conditions of Proposition 1 and Assumption 6, we have*

$$\frac{T(b) - \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2}{\sqrt{\operatorname{Var}\{T(b)\}}} \xrightarrow{d} N(0,1) \quad as \quad n, p \to \infty.$$

Under Assumption 3, our proposed BEU-statistic can be used for ultrahigh-dimensional series. A weaker condition on the mixing coefficients, say a polynomial decay, will put restrictions on the dimension p of the series, leading to more involved technical derivations. The challenge is mainly due to the slower convergence rate of the cross-covariance induced by the Davydov's inequality under the polynomial decay condition, compared with that under Assumption 3 for the exponential decay. This makes terms such as $\sum_{|k_1|,|k_2|>K} |\operatorname{tr}(\boldsymbol{\Sigma}_{i,k_1}\boldsymbol{\Sigma}_{i,k_2})|$ in Lemma 2 in the Supplementary Material converge at a slower rate. In the polynomial decay case, it can be shown that $\sum_{|k_1|,|k_2|>K} |\operatorname{tr}(\boldsymbol{\Sigma}_{i,k_1}\boldsymbol{\Sigma}_{i,k_2})| = o\{\operatorname{tr}(\boldsymbol{\Sigma}_{i,\infty}^2)\}$, for K being of a polynomial order with respect to n under $p \le \bar{a} n^{\bar{b}}$, for some constants $\bar{a}, \bar{b} > 0$. Hence, for series with stronger dependence, which corresponds to the polynomial decay, our proposed BEU-statistic is still applicable, with more restrictions on the polynomial increase of p with respect to n.

The current L_2 proposal is for temporally dependent data. Another important choice of the test statistic is based on the thresholding procedure (Chen et al., 2019). L_2-type statistics such as the proposed BEU-statistic and the thresholding-type statistics target different signal regimes. The former are powerful for dense but weak signals, where the signal strength from each component can be much smaller than the order $n^{-1/2}$, whereas

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

the latter are powerful for sparse signals with strength of order at least $\{(\log p)/n\}^{1/2}$. To establish a thresholding test similar to that of Chen et al. (2019) for dependent data requires first establishing moderate deviation results, which are not yet available, after which, a similar thresholding test can be developed for dependent time series data.

Based on the asymptotic normality, we can construct a test for the null hypothesis in (2.1) by obtaining a ratio-consistent estimator for the null variance of $T(b)$. The latter task is the focus of the next section.

4. Variance Estimation

Under H_0 of (2.1), from (3.4), the leading-order null variance of $T(b)$ is determined by $\mathrm{tr}(\mathbf{\Sigma}_{i,\infty}^2)$, $\mathrm{tr}(\mathbf{\Sigma}_{2,\infty}^2)$, and $\mathrm{tr}(\mathbf{\Sigma}_{1,\infty}\mathbf{\Sigma}_{2,\infty})$. In order to formulate a test, those trace quantities need to be estimated, which amounts to estimating $\mathrm{tr}(\mathbf{\Sigma}_{i,k_1}\mathbf{\Sigma}_{i,k_2})$ and $\mathrm{tr}(\mathbf{\Sigma}_{1,k_1}\mathbf{\Sigma}_{2,k_2})$ in the expansions

$$\mathrm{tr}(\mathbf{\Sigma}_{i,\infty}^2) = \sum_{k_1,k_2=-\infty}^{\infty} \mathrm{tr}(\mathbf{\Sigma}_{i,k_1}\mathbf{\Sigma}_{i,k_2}) \ \text{ and } \ \mathrm{tr}(\mathbf{\Sigma}_{1,\infty}\mathbf{\Sigma}_{2,\infty}) = \sum_{k_1,k_2=-\infty}^{\infty} \mathrm{tr}(\mathbf{\Sigma}_{1,k_1}\mathbf{\Sigma}_{2,k_2}). \tag{4.5}$$

To estimate $\mathrm{tr}(\mathbf{\Sigma}_{i,k_1}\mathbf{\Sigma}_{i,k_2})$, we apply a similar band-exclusion technique to that used to construct $T(b)$ in (3.2). Let $|\mathcal{N}|$ denote the cardinality of a set $\mathcal{N}$. For $i = 1, 2$ and another positive exclusion bandwidth parameter $\tilde{b}$, let

$$\begin{aligned} G_{i,1}(k_1, k_2; \tilde{b}) &= \left|\mathcal{N}_{i,1}(k_1, k_2; \tilde{b})\right|^{-1} \sum_{(t_1,t_2)\in\mathcal{N}_{i,1}(k_1,k_2;\tilde{b})} \boldsymbol{X}_{i,t_2}^{\mathrm{T}}\boldsymbol{X}_{i,t_1}\boldsymbol{X}_{i,t_1-k_1}^{\mathrm{T}}\boldsymbol{X}_{i,t_2+k_2}, \\ G_{i,2}(k; \tilde{b}) &= \left|\mathcal{N}_{i,2}(k; \tilde{b})\right|^{-1} \sum_{(t_1,t_2,t_3)\in\mathcal{N}_{i,2}(k;\tilde{b})} \boldsymbol{X}_{i,t_2}^{\mathrm{T}}\boldsymbol{X}_{i,t_1}\boldsymbol{X}_{i,t_1-k}^{\mathrm{T}}\boldsymbol{X}_{i,t_3}, \\ G_{i,3}(\tilde{b}) &= \left|\mathcal{N}_{i,3}(\tilde{b})\right|^{-1} \sum_{(t_1,t_2,t_3,t_4)\in\mathcal{N}_{i,3}(\tilde{b})} \boldsymbol{X}_{i,t_1}^{\mathrm{T}}\boldsymbol{X}_{i,t_2}\boldsymbol{X}_{i,t_3}^{\mathrm{T}}\boldsymbol{X}_{i,t_4} \end{aligned} \tag{4.6}$$

be the estimators for $\mathrm{tr}\{\mathrm{E}(\boldsymbol{X}_{i,k_1+1}\boldsymbol{X}_{i,1}^{\mathrm{T}})\mathrm{E}(\boldsymbol{X}_{i,k_2+1}\boldsymbol{X}_{i,1}^{\mathrm{T}})\}$, $\boldsymbol{\mu}_i^{\mathrm{T}}\mathrm{E}(\boldsymbol{X}_{i,k+1}\boldsymbol{X}_{i,1}^{\mathrm{T}})\boldsymbol{\mu}_i$, and $(\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\mu}_i)^2$,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

respectively, where $|k_1|, |k_2|, |k| < \min(n_i, n_2)/2$, and

$$\mathcal{N}_{i,1}(k_1,k_2;\tilde{b}) = \big\{(t_1,t_2) : |t_1-t_2| \ge \tilde{b}+|k_1|+|k_2|, 1 \le t_1, t_1-k_1, t_2, t_2+k_2 \le n_i\big\},$$

$$\mathcal{N}_{i,2}(k;\tilde{b}) = \big\{(t_1,t_2,t_3) : \min_{1\le j_1<j_2\le 3} |t_{j_1}-t_{j_2}| \ge \tilde{b}+|k|, 1 \le t_1, t_1-k, t_2, t_3 \le n_i\big\} \text{ and}$$

$$\mathcal{N}_{i,3}(\tilde{b}) = \big\{(t_1,t_2,t_3,t_4) : \min_{1\le j_1<j_2\le 4} |t_{j_1}-t_{j_2}| \ge \tilde{b}, 1 \le t_1,t_2,t_3,t_4 \le n_i\big\}$$

are the index sets with certain time separation. These index sets are designed to ensure sufficient temporal distance to reduce the temporal dependence. For example, the set $\mathcal{N}_{i,1}(k_1,k_2;\tilde{b})$ makes $\boldsymbol{X}_{i,t_1}\boldsymbol{X}_{i,t_1-k_1}^{\mathrm{T}}$ and $\boldsymbol{X}_{i,t_2+k_2}\boldsymbol{X}_{i,t_2}^{\mathrm{T}}$ in $G_{i,1}(k_1,k_2;\tilde{b})$ at least $\tilde{b}$ apart. Let

$$\widehat{\mathrm{tr}}(\boldsymbol{\Sigma}_{i,k_1}\boldsymbol{\Sigma}_{i,k_2};\tilde{b}) = G_{i,1}(k_1,k_2;\tilde{b}) - G_{i,2}(k_1;\tilde{b}) - G_{i,2}(k_2;\tilde{b}) + G_{i,3}(\tilde{b}) \tag{4.7}$$

be estimators of $\mathrm{tr}(\boldsymbol{\Sigma}_{i,k_1}\boldsymbol{\Sigma}_{i,k_2})$, for $i = 1, 2$. Similar to the diminishing bias attained by $T(b)$, as shown in Proposition 1, the bias of $\widehat{\mathrm{tr}}(\boldsymbol{\Sigma}_{i,k_1}\boldsymbol{\Sigma}_{i,k_2};\tilde{b})$ diminishes to zero as $\tilde{b} \to \infty$. Specifically, under Assumption 3, it suffices to choose $\tilde{b}$ at the order of $\log p$.

Similar estimators can be constructed for $\mathrm{tr}(\boldsymbol{\Sigma}_{1,k_1}\boldsymbol{\Sigma}_{2,k_2})$. Because observations from different groups are independent, band exclusion is not needed between two samples. Thus, we construct estimators for $\mathrm{tr}\{\mathrm{E}(\boldsymbol{X}_{1,k_1+1}\boldsymbol{X}_{1,1}^{\mathrm{T}})\mathrm{E}(\boldsymbol{X}_{2,k_2+1}\boldsymbol{X}_{2,1}^{\mathrm{T}})\}$, $\boldsymbol{\mu}_1^{\mathrm{T}}\mathrm{E}(\boldsymbol{X}_{2,k+1}\boldsymbol{X}_{2,1}^{\mathrm{T}})\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2^{\mathrm{T}}\mathrm{E}(\boldsymbol{X}_{1,k+1}\boldsymbol{X}_{1,1}^{\mathrm{T}})\boldsymbol{\mu}_2$, and $\|\boldsymbol{\mu}_1\|_2^2\|\boldsymbol{\mu}_2\|_2^2$ as

$$G_a(k_1,k_2) = |\mathcal{N}_a(k_1,k_2)|^{-1} \sum_{(t_1,t_2)\in\mathcal{N}_a(k_1,k_2)} \boldsymbol{X}_{2,t_2}^{\mathrm{T}}\boldsymbol{X}_{1,t_1}\boldsymbol{X}_{1,t_1-k_1}^{\mathrm{T}}\boldsymbol{X}_{2,t_2+k_2},$$

$$G_{c,1}(k;\tilde{b}) = n_1(\tilde{b})^{-1}|\mathcal{N}_{c,2}(k)|^{-1} \sum_{t_1\in\mathcal{N}_{c,2}(k)} \sum_{|t_2-t_3|\ge\tilde{b}} \boldsymbol{X}_{1,t_2}^{\mathrm{T}}\boldsymbol{X}_{2,t_1}\boldsymbol{X}_{2,t_1-k}^{\mathrm{T}}\boldsymbol{X}_{1,t_3},$$

$$G_{c,2}(k;\tilde{b}) = n_2(\tilde{b})^{-1}|\mathcal{N}_{c,1}(k)|^{-1} \sum_{t_1\in\mathcal{N}_{c,1}(k)} \sum_{|t_2-t_3|\ge\tilde{b}} \boldsymbol{X}_{2,t_2}^{\mathrm{T}}\boldsymbol{X}_{1,t_1}\boldsymbol{X}_{1,t_1-k}^{\mathrm{T}}\boldsymbol{X}_{2,t_3} \quad \text{and}$$

$$G_d(\tilde{b}) = n_1(\tilde{b})^{-1}n_2(\tilde{b})^{-1} \sum_{|t_1-t_3|\ge\tilde{b}} \sum_{|t_2-t_4|\ge\tilde{b}} \boldsymbol{X}_{1,t_1}^{\mathrm{T}}\boldsymbol{X}_{2,t_2}\boldsymbol{X}_{1,t_3}^{\mathrm{T}}\boldsymbol{X}_{2,t_4},$$

respectively, where $n_i(\tilde{b}) = (n_i-\tilde{b})(n_i-\tilde{b}+1)$, $\mathcal{N}_a(k_1,k_2) = \big\{(t_1,t_2) : 1 \le t_1, t_1-k_1 \le n_1, 1 \le t_2, t_2+k_2 \le n_2\big\}$, and $\mathcal{N}_{c,i}(k) = \big\{t : 1 \le t, t-k \le n_i\big\}$, for $i = 1, 2$. Then, based on these statistics, we construct the estimator for $\mathrm{tr}(\mathbf{\Sigma}_{1,k_1}\mathbf{\Sigma}_{2,k_2})$ as

$$\widehat{\mathrm{tr}}(\mathbf{\Sigma}_{1,k_1}\mathbf{\Sigma}_{2,k_2};\tilde{b}) = G_a(k_1,k_2) - G_{c,1}(k_2;\tilde{b}) - G_{c,2}(k_1;\tilde{b}) + G_d(\tilde{b}). \tag{4.8}$$

Because the elements in $\mathbf{\Sigma}_{i,k}$ decay to zero as $|k|$ increases under Assumption 3, and according to (4.5), we consider a weighted sum of $\widehat{\mathrm{tr}}(\mathbf{\Sigma}_{i,k_1}\mathbf{\Sigma}_{i,k_2};\tilde{b})$ and $\widehat{\mathrm{tr}}(\mathbf{\Sigma}_{1,k_1}\mathbf{\Sigma}_{2,k_2};\tilde{b})$ to estimate $\mathrm{tr}(\mathbf{\Sigma}_{i,\infty}^2)$ and $\mathrm{tr}(\mathbf{\Sigma}_{1,\infty}\mathbf{\Sigma}_{2,\infty})$. The weights are determined by a kernel function such that a larger (smaller) weight is allocated for terms with smaller (larger) $|k_1|$ and $|k_2|$. This idea is connected to the kernel-type estimator for fixed-dimensional long-run covariances of Andrews (1991), and the smoothing of periodograms method for estimating the spectral density at a zero frequency for fixed-dimensional time series (Priestley, 1981).

Let $\mathcal{K}(\cdot)$ be a symmetric function on $\mathbb{R}$ that is continuous at zero and satisfies $\mathcal{K}(0) = 1$, $\sup_{u\in\mathbb{R}} |\mathcal{K}(u)| \le 1$, and $\int_{-\infty}^{\infty} |\mathcal{K}(u)|du < \infty$. We propose the following kernel-smoothing estimators:

$$\begin{aligned}\widehat{\mathrm{tr}}(\mathbf{\Sigma}_{i,\infty}^2;\tilde{b},s_0) &= \sum_{k_1=-n_i+1}^{n_i-1}\sum_{k_2=-n_i+1}^{n_i-1} \mathcal{K}(k_1/s_0)\mathcal{K}(k_2/s_0)\widehat{\mathrm{tr}}(\mathbf{\Sigma}_{i,k_1}\mathbf{\Sigma}_{i,k_2};\tilde{b}) \quad \text{and} \\ \widehat{\mathrm{tr}}(\mathbf{\Sigma}_{1,\infty}\mathbf{\Sigma}_{2,\infty};\tilde{b},s_0) &= \sum_{k_1=-n_1+1}^{n_1-1}\sum_{k_2=-n_2+1}^{n_2-1} \mathcal{K}(k_1/s_0)\mathcal{K}(k_2/s_0)\widehat{\mathrm{tr}}(\mathbf{\Sigma}_{1,k_1}\mathbf{\Sigma}_{2,k_2};\tilde{b})\end{aligned} \tag{4.9}$$

for $\mathrm{tr}(\mathbf{\Sigma}_{i,\infty}^2)$ and $\mathrm{tr}(\mathbf{\Sigma}_{1,\infty}\mathbf{\Sigma}_{2,\infty})$, respectively, where s_0 is a smoothing bandwidth diverging to ∞ as $n,p \to \infty$. According to the expression of the null variance in (3.4), we propose the smoothed band-exclusion statistic (SBE-statistic)

$$V_n(\tilde{b},s_0) = \frac{2}{n_1^2}\widehat{\mathrm{tr}}(\mathbf{\Sigma}_{1,\infty}^2;\tilde{b},s_0) + \frac{2}{n_2^2}\widehat{\mathrm{tr}}(\mathbf{\Sigma}_{2,\infty}^2;\tilde{b},s_0) + \frac{4}{n_1n_2}\widehat{\mathrm{tr}}(\mathbf{\Sigma}_{1,\infty}\mathbf{\Sigma}_{2,\infty};\tilde{b},s_0) \tag{4.10}$$

for estimating $\mathrm{Var}\{T(b)\}$ under H_0.

Andrews (1991) studied the kernel weighted estimator $\sum_k \mathcal{K}(k/s_0)\widehat{\mathbf{\Sigma}}_{i,k}$ of the long-run covariance $\mathbf{\Sigma}_{i,\infty}$ for various kernels under the fixed-dimension case, where $\widehat{\mathbf{\Sigma}}_{i,k}$ is the sample cross-time covariances, and showed that the quadratic spectral (QS) kernel

$$\mathcal{K}_{QS}(u) = \frac{25}{12\pi^2u^2}\left\{\frac{\sin(6\pi u/5)}{6\pi u/5} - \cos(6\pi u/5)\right\}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

is optimal for the long-run covariance estimation, in the sense of minimizing the asymptotic truncated mean squared error. We use the QS kernel in the numerical implementation, and a data-driven procedure to select the smoothing bandwidth s_0 in the next section. Simulation results reported in Section 7 show that the BEU-statistic $T(b)$ with the smoothed band-exclusion variance estimator $V_n(\tilde{b}, s_0)$ and the QS kernel performs well in high-dimensional scenarios. Note that there are other estimation methods for long-run covariances under fixed-dimensional settings, including the moving block bootstraps (Lahiri, 2003; Nordman and Lahiri, 2005). Using these other methods to estimate the variances of the L_2-type statistics for high-dimensional time series is worth further investigation.

To show the ratio consistence of the SBE variance estimator, we impose the following mild technical condition on the eigenvalue of the innovation loading matrix $\mathbf{\Gamma}_i$ in Assumption 2.

Assumption 7. Let $\boldsymbol{B}_i = \mathbf{\Gamma}_i^{\mathrm{T}}\mathbf{\Gamma}_i = (b_{i,j_1j_2})_{r\times r}$, $\tilde{\boldsymbol{B}}_i = (|b_{i,j_1j_2}|)_{r\times r}$, and $\lambda_{\max}(\tilde{\boldsymbol{B}}_i)$ be the maximum eigenvalue of $\tilde{\boldsymbol{B}}_i$, for $i = 1, 2$. There exist two positive constants, ψ and C_2, such that $\max\{\lambda_{\max}(\tilde{\boldsymbol{B}}_1), \lambda_{\max}(\tilde{\boldsymbol{B}}_2)\} \le C_2 p^{\psi}$.

Note that $\lambda_{\max}(\boldsymbol{B}) = \lambda_{\max}(\mathbf{\Sigma}_{i,\infty})$. This assumption describes the maximum eigenvalue of the absolute matrix of $\boldsymbol{B}$, which is allowed to diverge to infinity at a polynomial rate of p. The following theorem shows the ratio consistency of the proposed SBE variance estimator.

Theorem 2. *Assume the exclusion bandwidth $\tilde{b}$ and the smoothing bandwidth s_0 satisfy $\tilde{b} = o(n^{1/5})$ and $\tilde{b} \ge c_2(\log n + \log p + s_0)$, for a positive constant c_2. Under Assumption 1 with $q > 8$, Assumptions 2–7, and the null hypothesis of (2.1),*

$$\frac{V_n(\tilde{b}, s_0)}{\mathrm{Var}\{T(b)\}} \to 1 \quad \textit{in probability as } n, p \to \infty.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

From Theorems 1 and 2, the proposed BEU test rejects the null hypothesis in (2.1) if

$$T(b) > z_\alpha V_n^{1/2}(\tilde{b}, s_0), \tag{4.11}$$

where z_α is the upper α quantile of $N(0,1)$. Note that the requirements on the moment q and the exclusion bandwidth $\tilde{b}$ in Theorem 2 are more restrictive than those in Theorem 1. This is because establishing the consistency of the variance estimator needs to control higher order moments than those needed to derive the properties of the BEU-statistic $T(b)$.

As discussed in the second paragraph after (3.2), the statistic

$$T_1(b) = n_1(b)^{-1} \sum_{|t_1-t_2|\geq b} \boldsymbol{X}_{1,t_1}^{\mathrm{T}} \boldsymbol{X}_{1,t_2}$$

can be used to test the one-sample hypotheses $H_0 : \boldsymbol{\mu}_1 = \mathbf{0}$ vs. $H_a : \boldsymbol{\mu}_1 \neq \mathbf{0}$. Following the same derivation as Proposition 1, it can be shown that $\mathrm{Var}\{T_1(b)\} = 2n_1^{-2}\mathrm{tr}(\boldsymbol{\Sigma}_{1,\infty}^2)$, which can be estimated by $2n_1^{-2}\widehat{\mathrm{tr}}(\boldsymbol{\Sigma}_{1,\infty}^2; \tilde{b}, s_0)$ from (4.9). Therefore, similarly to the two-sample test in (4.11), the one-sample BEU test rejects the null hypothesis $\boldsymbol{\mu}_1 = \mathbf{0}$ if

$$T_1(b) > z_\alpha n_1^{-1}\{2\widehat{\mathrm{tr}}(\boldsymbol{\Sigma}_{1,\infty}^2; \tilde{b}, s_0)\}^{1/2}. \tag{4.12}$$

For this one-sample hypothesis, we compare the power of the BEU test with that of the self-normalized test of Wang and Shao (2020) in Sections 6 and 7.

5. Computation and Tuning Parameter Selection

In this section, we discuss computation and implementation aspects of the proposed BEU test, and propose a data-driven procedure to select the tuning parameters b, $\tilde{b}$, and s_0. When calculating the test statistic, matrix operations should be used wherever possible to improve the computation efficiency. Recall that $\boldsymbol{X}_i = (\boldsymbol{X}_{i,1}, \ldots, \boldsymbol{X}_{i,n_i})^{\mathrm{T}}$ is the $n_i \times p$ data matrix for the ith sample. Let $\boldsymbol{W}_i(b) = (w_{i,t_1t_2})_{n_i\times n_i}$ be an indicator matrix, with $w_{i,t_1t_2} = 1$ if $|t_1 - t_2| \geq b$, and zero otherwise. Let $\circ$ denote the Hadamard product of two matrices with the same dimensions. Then, the summation $\sum_{|t_1-t_2|\geq b} \boldsymbol{X}_{i,t_1}^{\mathrm{T}} \boldsymbol{X}_{i,t_2}$ in the BEU-statistic

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

$T(b)$ in (3.2) can be computed by summing over all elements in $(\boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}}) \circ \boldsymbol{W}_i(b)$.

To estimate $\mathrm{tr}(\boldsymbol{\Sigma}_{i,k_1}\boldsymbol{\Sigma}_{i,k_2})$, the estimators in (4.6) can be computed on the centered data $\{\boldsymbol{X}_{i,t} - \bar{\boldsymbol{X}}_i\}$, so that $G_{i,2}(k;\tilde{b})$ and $G_{i,3}(\tilde{b})$ become smaller order terms that are negligible in the construction of the variance estimator. If the computing resource is a constraint, we can only compute $G_{i,1}(k_1, k_2;\tilde{b})$ on the centered data in the estimator $\widehat{\mathrm{tr}}(\boldsymbol{\Sigma}_{i,k_1}\boldsymbol{\Sigma}_{i,k_2};\tilde{b})$ in (4.7). Note that $G_{i,2}(k;\tilde{b})$ and $G_{i,3}(\tilde{b})$ require computation complexity of order n^3 and n^4, respectively. Centering the data can greatly reduce the computational burden. Similar arguments apply when estimating $\mathrm{tr}(\boldsymbol{\Sigma}_{1,k_1}\boldsymbol{\Sigma}_{2,k_2})$ in (4.8).

Note that $G_{i,1}(k_1, k_2;\tilde{b})$ in (4.6) can be computed using matrix operations as well. Let $\boldsymbol{A}[c_1 : c_2,]$ denote the sub-matrix of $\boldsymbol{A}$ with the c_1th row to the c_2th row. For any integer $-n_i < k < n_i$, let $\boldsymbol{Z}_i(k)$ be a row-shifted matrix of $\boldsymbol{X}_i$ in the following way. If $k = 0$, there is no shift and $\boldsymbol{Z}_i(0) = \boldsymbol{X}_i$; if $k < 0$, the first $|k|$ rows of $\boldsymbol{Z}_i(k)$ are zero and $\boldsymbol{Z}_i(k)[|k| + 1 : n_i,] = \boldsymbol{X}_i[1 : n_i - |k|,]$; if $k > 0$, the last k rows of $\boldsymbol{Z}_i(k)$ are zero and $\boldsymbol{Z}_i(k)[1 : n_i - k,] = \boldsymbol{X}_i[k + 1 : n_i,]$. Then, the summation of $\boldsymbol{X}_{i,t_2}^{\mathrm{T}}\boldsymbol{X}_{i,t_1}\boldsymbol{X}_{i,t_1-k_1}^{\mathrm{T}}\boldsymbol{X}_{i,t_2+k_2}$ over $(t_1, t_2) \in \mathcal{N}_{i,1}(k_1, k_2;\tilde{b})$ in (4.6) can be computed by simply summing over all the elements in $(\boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}}) \circ (\boldsymbol{Z}_i(-k_1)\boldsymbol{Z}_i(k_2)^{\mathrm{T}}) \circ \boldsymbol{W}_i(\tilde{b} + |k_1| + |k_2|)$. A similar algorithm can be applied for the statistic $G_a(k_1, k_2)$ in (4.8).

The tuning parameters b, $\tilde{b}$, and s_0 required in the proposed BEU test are chose adaptively based on the time course data. In particular, the exclusion bandwidths b and $\tilde{b}$ used in the BEU-statistic $T(b)$ and its variance estimator may be determined by using the sample autocorrelation function (ACF). Specifically, for each dimension j, we calculate the sample ACF of the univariate time series $\{X_{i,t,j}\}_{t=1}^{n_i}$, denoted as $\mathrm{AC}_{i,j}(k)$. Let $\mathrm{AC}_i(k) = \max_{1\le j\le p} |\mathrm{AC}_{i,j}(k)|$ be the maximal absolute sample ACF at time lag k, and let

$\mathbb{Z}^+$ be the set of all positive integers. Let

$$b_i = \min\{k \in \mathbb{Z}^+ : \mathrm{AC}_i(k) < \mathrm{mAC}_i\}$$

be the first time lag such that $\mathrm{AC}_i(k)$ is smaller than a data-driven threshold mAC_i. Here, $\mathrm{mAC}_i = \mathrm{Median}\{\mathrm{AC}_i(k) : n/10 \le k \le n/4\}$ is the median of the maximal absolute sample ACF with large time lags, such that the time dependence between observations is fairly weak. We choose $b = \max\{b_1, b_2\}$ and set $\tilde{b} = b$. The optimal bandwidth using the QS kernel to estimate the long-run covariances is derived in Andrews (1991). We choose the estimated optimal bandwidth based on the data-driven procedure introduced in (6.2) and (6.4) of Andrews (1991) as the smoothing bandwidth s_0 in $V_n(\tilde{b}, s_0)$. The results of our simulation in Section 6 show that the proposed BEU test with adaptively chosen exclusion and smoothing bandwiths works well, with accurate size and good power.

6. Power Analysis

Theorem 1 allows us to discuss the power properties of the proposed test. We consider two forms of alternative hypotheses for $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$. The first one is

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}} \boldsymbol{M}_\infty (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o\big\{n^{-1}\mathrm{tr}(\boldsymbol{M}_\infty^2)\big\} \text{ as } n, p \to \infty, \tag{6.1}$$

which describes the so-called local alternative. The second is

$$n^{-1}\mathrm{tr}(\boldsymbol{M}_\infty^2) = o\big\{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}} \boldsymbol{M}_\infty (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\big\} \text{ as } n, p \to \infty, \tag{6.2}$$

which may be viewed as the fixed alternative, because it allows stronger signals. Note that $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}} \boldsymbol{M}_\infty (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ is a weighted distance between $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and measures the strength of signals in terms of identifying $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. The local alternative (6.1) represents a weak signal case, so that this weighted distance is at a smaller order of $n^{-1}\mathrm{tr}(\boldsymbol{M}_\infty^2)$. The fixed alternative (6.2) implies that the weighted distance between $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ is at a larger order than $n^{-1}\mathrm{tr}(\boldsymbol{M}_\infty^2)$, which is a reverse of the local alternative condition in (6.1).

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

Let $\beta(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \mathrm{P}\{T(b) > z_\alpha V_n^{1/2}(\tilde{b}, s_0) | \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2\}$ be the power of the proposed test. From Theorem 1, we have $\mathrm{Var}\{T(b)\} = 2n^{-2}\mathrm{tr}(\boldsymbol{M}_\infty^2)\{1 + o(1)\}$ under the local alternative (6.1), and $\mathrm{Var}\{T(b)\} = 4n^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}}\boldsymbol{M}_\infty(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\{1 + o(1)\}$ under the fixed alternative (6.2). Let $\Phi(\cdot)$ be the standard normal distribution function and $\lambda_{\max}$ be the largest eigenvalue of $\boldsymbol{M}_\infty$. The following two theorems describe the power of the test under the two forms of alternatives.

Theorem 3. *Under the conditions of Theorems 1 and 2 and the local alternative (6.1), the power function of the proposed test is*

$$\beta(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \Phi\left\{ - z_\alpha + \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2}{\sqrt{2n^{-2}\mathrm{tr}(\boldsymbol{M}_\infty^2)}}\right\}\{1 + o(1)\}, \tag{6.3}$$

and $\beta(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \to \Phi(-z_\alpha + d^2/\sqrt{2})$ *if* $n\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2 \mathrm{tr}^{-1/2}(\boldsymbol{M}_\infty^2) \to d^2 \in [0, +\infty)$.

Theorem 4. *Under the conditions of Theorems 1 and 2 and the fixed alternative (6.2), the power function of the proposed test is*

$$\beta(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \Phi\left\{ \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2}{\sqrt{4n^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}}\boldsymbol{M}_\infty(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}}\right\}\{1 + o(1)\}. \tag{6.4}$$

In Theorems 3 and 4, note that $n\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2 \mathrm{tr}^{-1/2}(\boldsymbol{M}_\infty^2)$ and $\sqrt{n}\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2\{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}}\boldsymbol{M}_\infty(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\}^{-1/2}$ are the signal-to-noise ratios of the proposed test for the two-sample hypotheses (2.1) under the local and fixed alternatives, respectively, for weak and strong signals. From Theorems 3 and 4, the power of the proposed test is bounded from below by

$$\beta(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \geq \Phi\left(- z_\alpha + \frac{n\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2}{\sqrt{2p}\lambda_{\max}}\right) \quad \text{and} \quad \beta(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \geq \Phi\left(\frac{\sqrt{n}\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2}{2\sqrt{\lambda_{\max}}}\right)$$

under the local and fixed alternatives, respectively. Let $\tilde{p}$ be the number of nonzero $\mu_{1,j} - \mu_{2,j}$, for $j = 1, \ldots, p$. If $|\mu_{1,j} - \mu_{2,j}| = \delta$ for all nonzero $\mu_{1,j} - \mu_{2,j}$, then $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2 = \sqrt{\tilde{p}}\delta$. In this case, the lower bounds of the power function are

$$\beta(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \geq \Phi\left(- z_\alpha + \frac{n\tilde{p}\delta^2}{\sqrt{2p}\lambda_{\max}}\right) \quad \text{and} \quad \beta(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \geq \Phi\left(\frac{\delta}{2}\sqrt{\frac{n\tilde{p}}{\lambda_{\max}}}\right)$$

under the local and fixed alternatives, respectively. If the nonzero components of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

are dense, so that $\tilde{p}$ is of the same order as p and $\lambda_{\max}$ is bounded, the proposed test can detect the difference δ as weak at the order $n^{-1/2}p^{-1/4}$ under the local alternative.

Let $\beta_{\text{prop}}(d) = \Phi(-z_\alpha + d^2/\sqrt{2})$. Theorem 3 shows that $\beta_{\text{prop}}(d)$ is the limiting power of the proposed test under the local alternative of the two-sample hypotheses specified in (6.1), where $d^2 = n\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2 \text{tr}^{-1/2}(\boldsymbol{M}_\infty^2)$.

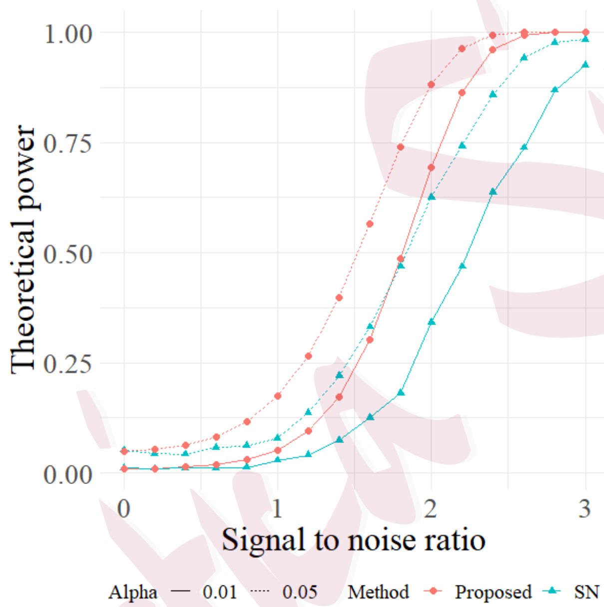


Figure 1: Theoretical power curves of the proposed test $\beta_{\text{prop}}(d_1) = \Phi(-z_\alpha + d_1^2/\sqrt{2})$ (red curve) and the self-normalized test $\beta_{\text{SN}}(d_1)$ of Wang and Shao (2020) (blue curve, labeled as SN) for the one-sample hypothesis $H_0 : \boldsymbol{\mu}_1 = 0$ vs. $\boldsymbol{\mu}_1 \neq 0$ under the local alternative.

When testing the one-sample hypotheses $H_0 : \boldsymbol{\mu}_1 = 0$, the proposed test based on (4.12) has the same power function as $\beta_{\text{prop}}(d_1)$ under the local alternative, where $d_1^2 = n_1\|\boldsymbol{\mu}_1\|_2^2 \text{tr}^{-1/2}(\boldsymbol{\Sigma}_{1,\infty}^2)$ is the signal-to-noise ratio of testing $\boldsymbol{\mu}_1 = \mathbf{0}$. Theorem 3.11 in Wang and Shao (2020) shows that

$$\beta_{\text{SN}}(d_1) = \mathbb{P}\left[\frac{\{\mathcal{B}(1) + d_1^2/\sqrt{2}\}^2}{\int_0^1 \{\mathcal{B}(u^2) - u^2\mathcal{B}(1)\}^2 du} \geq z_{1,\alpha}\right]$$

is the asymptotic power function of the SN test under the local alternative, where $\mathcal{B}(u)$ denotes the standard Brownian motion for $u \in [0, 1]$, and $z_{1,\alpha}$ is the upper α-quantile of

$\mathcal{B}(1)^2[\int_0^1\{\mathcal{B}(u^2) - u^2\mathcal{B}(1)\}^2 du]^{-1}$. As shown in Figure 1, given the same signal-to-noise ratio d_1, the power of the one-sample BEU test given in (4.12) is higher than that of the SN test, such that $\beta_{\text{prop}}(d_1) \geq \beta_{\text{SN}}(d_1)$. Specifically, Figure 1 plots the two power functions against the signal-to-noise ratio under $\alpha = 0.01$ and 0.05, which shows the superiority of the proposed test.

7. Numerical Studies

This section reports results from simulation experiments designed to evaluate the empirical size and power of the proposed test for the two-sample hypotheses (2.1). For comparison purposes, the test of Chen and Qin (2010) (CQ) for independent data and the test of Ayyala et al. (2017) (APR) for m-dependent data considered in the two-sample case. We also compare our proposed test with the SN test of Wang and Shao (2020) under the one-sample scenario.

First, we consider the two-sample case, where we use the moving average (MA) model and the auto-regressive (AR) model to generate temporally dependent data:

- MA model: $\boldsymbol{X}_{1,t} = \boldsymbol{\epsilon}_{1,t} + \rho_{\text{time}}\boldsymbol{\epsilon}_{1,t-1}$ and $\boldsymbol{X}_{2,t} = \boldsymbol{\mu}_2 + \boldsymbol{\epsilon}_{2,t} + \rho_{\text{time}}\boldsymbol{\epsilon}_{2,t-1}$;
- AR model: $\boldsymbol{X}_{1,t} = \rho_{\text{time}}\boldsymbol{X}_{1,t-1} + (1-\rho_{\text{time}}^2)^{1/2}\boldsymbol{\epsilon}_{1,t}$ and $\boldsymbol{X}_{2,t} = \boldsymbol{\mu}_2 + \rho_{\text{time}}\boldsymbol{X}_{2,t-1} + (1-\rho_{\text{time}}^2)^{1/2}\boldsymbol{\epsilon}_{2,t}$;

where $\{\boldsymbol{\epsilon}_{i,t}\}_{t=1}^{n_i}$ are i.i.d. p-dimensional random vectors from $N(0, \boldsymbol{\Sigma}_\epsilon)$, and ρ_{time} is the temporal dependence parameter that characterizes the strength of the temporal dependence. We set ρ_{time} as 0.1, 0.3, and 0.5 in the simulation. The spatial dependence is given by $\boldsymbol{\Sigma}_\epsilon = (\sigma_{\epsilon,j_1j_2})$, where $\sigma_{\epsilon,j_1j_2} = 0.7^{|j_1-j_2|}$. By default, $\boldsymbol{\mu}_1 = 0$. Under the alternative hypotheses, we consider different combinations of signal strength and sparsity for $\boldsymbol{\mu}_2$, where the first δ_0 proportion of the components in $\boldsymbol{\mu}_2$ are set as r_0, and the rest are set to zero. We

Table 1: Empirical sizes of the proposed test, the APR test (Ayyala et al., 2017), and the CQ test (Chen and Qin, 2010) for the two-sample hypotheses under the temporal dependence parameter $\rho_{\text{time}} = 0.1, 0.3, 0.5$, $n_0 = 100, 150$, $p = 100, 400$, and the MA and AR models.

Method	(n_0, p)	ρ_{time} under MA			ρ_{time} under AR		
		0.1	0.3	0.5	0.1	0.3	0.5
Proposed	(100,100)	0.069	0.045	0.065	0.061	0.074	0.076
	(150,100)	0.064	0.062	0.073	0.081	0.078	0.077
	(100,400)	0.079	0.056	0.070	0.075	0.053	0.059
	(150,400)	0.067	0.055	0.071	0.073	0.075	0.050
APR	(100,100)	0.057	0.051	0.066	0.068	0.077	0.266
	(150,100)	0.068	0.055	0.069	0.078	0.102	0.221
	(100,400)	0.053	0.046	0.050	0.055	0.066	0.453
	(150,400)	0.047	0.043	0.061	0.062	0.113	0.449
CQ	(100,100)	0.191	0.603	0.802	0.245	0.844	1.000
	(150,100)	0.204	0.662	0.847	0.285	0.858	0.999
	(100,400)	0.473	0.978	1.000	0.513	1.000	1.000
	(150,400)	0.486	0.971	0.997	0.546	1.000	1.000

choose $\delta_0 = 0.2$ and 0.3, and r_0 takes values from a sequence ranging from 0.05 to 0.3, with increments of 0.05. Here, δ_0 and r_0 denote the signal sparsity and strength, respectively. We set $n_1 = n_2 = n_0 = 100$ and 150, $p = 100$ and 400, and the significance level to 0.05. All simulations are repeated 1000 times under each setting. The MA model satisfies the m-dependence assumption required by the APR method, whereas the AR model is not m-dependent. The time dependence lag parameter m in the APR test is chosen as two, as suggested by Ayyala et al. (2017), based on their simulation studies.

Table 1 reports the empirical sizes of the proposed, APR, and CQ tests under MA and AR models with different time dependence parameters ρ_{time}. The CQ test is designed for independent data. We include it to gain empirical information about the consequences of ignoring temporal dependence in two-sample tests. Table 1 shows that the proposed test controls the size for testing the hypotheses (2.1) around the nominal level for all cases

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

considered. It is not unexpected that the CQ test, designed for independent samples, cannot control the size with severe size distortion as ρ_{time} increases. Thus, the consequence of ignoring the time dependence is severe. The APR test controls the size under the MA model, because the MA model describes an m-dependent series with $m = 1$, which meets the assumptions of the APR test (Ayyala et al., 2017). However, for the AR model, the APR test does not control the size around 0.05, especially when the temporal dependence parameter ρ_{time} increases to 0.5, with the size reaching over 0.4 for $p = 400$.

Figures 2 and 3 report the power of the proposed and APR tests, respectively. We empirically adjusted the critical values for the proposed and APR tests based on their simulated distributions under the null hypothesis so that they have the same empirical size of 0.05, for fairer power evaluation. Figures 2 and 3 suggest that the proposed and APR methods have comparable power under all combinations of signal proportion and strength. This is because both tests are constructed from sum-of-square statistics, which have a similar power profile in terms of signal detection. Note that the power of the APR method is slightly higher than that of the proposed test under a couple of settings. This may be because it uses more observations than the BEU-statistic $T(b)$ does, with a larger b, as selected by the proposed algorithm. Similar results are observed in the simulation studies of Ayyala et al. (2017), where the power of the APR test decreases with an increase of its time lag tuning parameter. The main problem with the APR test is that it cannot control the size for general temporal dependence, which limits its general applicability. In contrast, the proposed test can be used with proper control on the size, and exhibits reasonable power.

To investigate the performance of the proposed test for other distributions, we consider

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

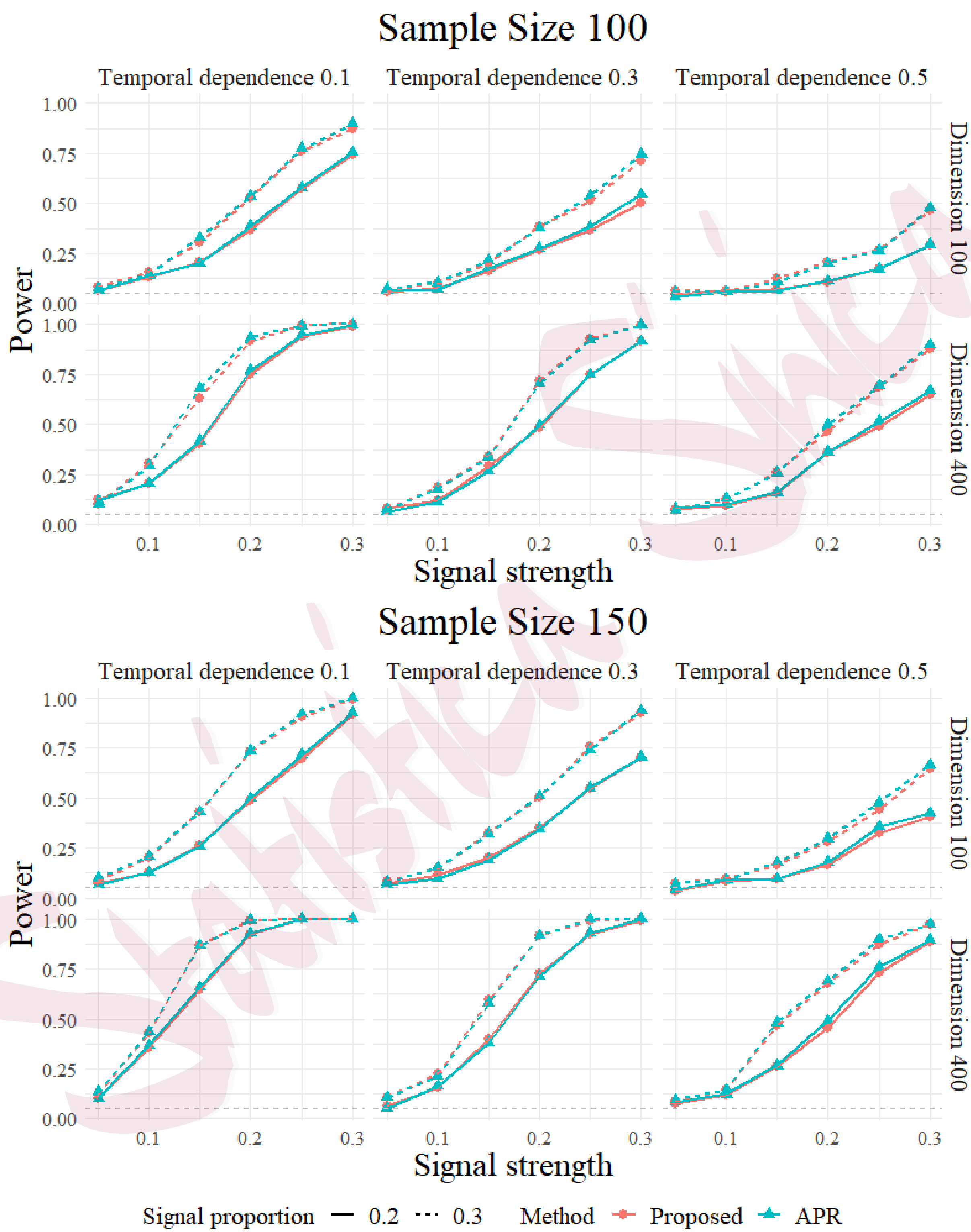


Figure 2: Empirical powers of the proposed test (red) and the APR test (blue) with respect to the signal strength r_0 (horizontal axis) for the two-sample hypotheses under the AR model, the sample size $n_0 = 100, 150$, the dimension $p = 100, 400$, three levels of temporal dependence ρ_{time}, and two values of signal proportions.

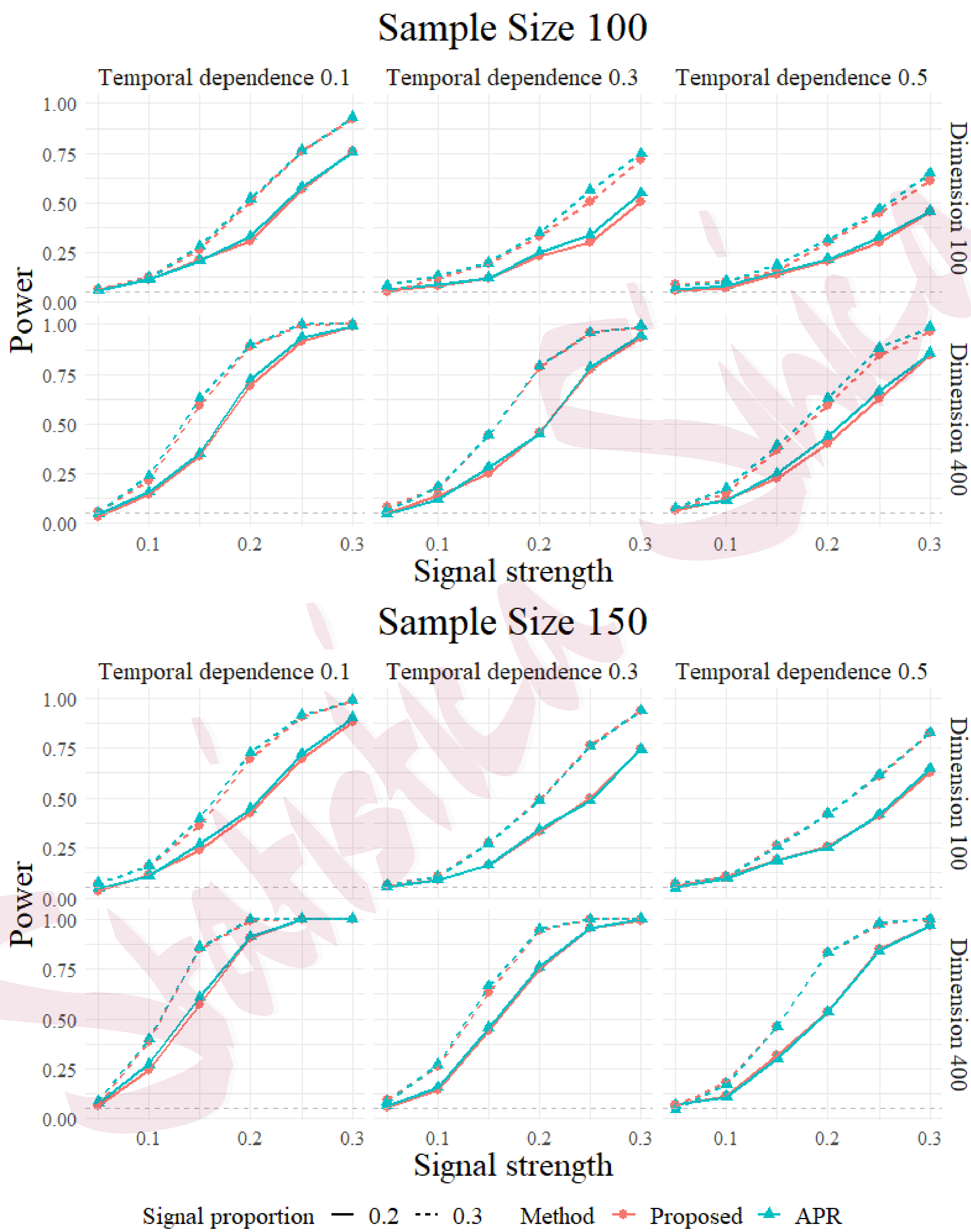


Figure 3: Empirical powers of the proposed test (red) and the APR test (blue) with respect to the signal strength r_0 (horizontal axis) for the two-sample hypotheses under the MA model, the sample size $n_0 = 100, 150$, the dimension $p = 100, 400$, three levels of temporal dependence ρ_{time}, and two values of signal proportions.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

the AR model

$$\boldsymbol{X}_{1,t} = \rho_{\text{time}}\boldsymbol{X}_{1,t-1} + (1-\rho_{\text{time}}^2)^{1/2}\boldsymbol{\epsilon}_{1,t} \quad \text{and} \quad \boldsymbol{X}_{2,t} = \boldsymbol{\mu}_2 + \rho_{\text{time}}\boldsymbol{X}_{2,t-1} + (1-\rho_{\text{time}}^2)^{1/2}\boldsymbol{\epsilon}_{2,t},$$

where the following two distributions are assigned to the i.i.d. errors $\{\boldsymbol{\epsilon}_{i,t}\}_{t=1}^{n_i}$:

- Multivariate t-distribution: $\boldsymbol{\epsilon}_{i,t} = \boldsymbol{e}_{i,t}/\sqrt{\chi^2_{i,t}(6)/6}$, where $\{\boldsymbol{e}_{i,t}\}_{t=1}^{n_i}$ are i.i.d. p-dimensional random vectors from $N(\mathbf{0}, \boldsymbol{\Sigma}_e)$, with $\boldsymbol{\Sigma}_e = (0.5^{|j_1-j_2|})$, $\{\chi^2_{i,t}(6)\}_{t=1}^{n_i}$ are i.i.d. random variables from a chi-squared distribution with six degrees of freedom, and $\{\boldsymbol{e}_{i,t}\}_{t=1}^{n_i}$ and $\{\chi^2_{i,t}(6)\}_{t=1}^{n_i}$ are mutually independent;

- Gamma distribution: $\epsilon_{i,t,j} = e_{i,t,j} + \gamma_i e_{i,t,j-1}$, where the i.i.d. $\{e_{i,t,j}\}_{j=0}^{p}$ follow a centralized Gamma$(1,1)$ distribution and $\gamma_1 = \gamma_2 = 0.5$.

Here, we choose $\rho_{\text{time}} = 0.1, 0.3, 0.5$ and $(n_0, p) = (100, 400)$. The settings of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the same as those in the AR model with the normally distributed error.

Table 2 and Figure 4 show the empirical size and power of the proposed test, the APR test, and the CQ test, with the error terms following a multivariate t-distribution and a gamma distribution, respectively. Compared with the case of normally distributed errors, here, the sizes of the APR test are nearly zero in all cases, whereas our proposed test exhibits reasonable sizes around the nominal level of 5%, in contrast to the APR test and the CQ test. The critical values used to compute powers are adjusted according to the distribution of the test statistic under the null hypothesis. Figure 4 shows that the power of the APR test is quite sensitive to the error distribution. When the errors follow a multivariate t-distribution or a gamma distribution, the power of the APR test remains small and flat as the signal strength increases and the sparsity level decreases. Under all settings of temporal dependence, the proposed test exhibits better performance, with higher power than that of the APR test, a result that becomes more pronounced when the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

sparsity level decreases. Here, the superiority of the proposed test is more evident than in the case with normally distributed errors.

Table 2: Empirical sizes of the proposed test, the APR test (Ayyala et al., 2017), and the CQ test (Chen and Qin, 2010) for the two-sample hypotheses under the temporal dependence parameter $\rho_{\text{time}} = 0.1, 0.3, 0.5$, $(n_0, p) = (100, 400)$, and the error term following a multivariate t-distribution and a gamma distribution.

Method	ρ_{time} under t distribution			ρ_{time} under Gamma distribution		
	0.1	0.3	0.5	0.1	0.3	0.5
Proposed	0.052	0.046	0.030	0.072	0.054	0.036
APR	0.000	0.000	0.000	0.000	0.000	0.000
CQ	0.732	1.000	1.000	0.804	1.000	1.000

Next, we compare the proposed test with the SN test of Wang and Shao (2020) in the one-sample testing problem. We use one time series generated from the MA model and the AR model in the two-sample setting as the observed data. Table 3 reports the empirical sizes of the proposed and the SN tests under the MA and AR models. The results show that both tests control their sizes under the model settings for the experiment sample sizes and dimensions. Figures 5 and 6 report the empirical power of each of the two tests. To make the power comparison fair, we perform the same adjustment on the critical values of the tests as in the two-sample simulation to make the two tests have the same empirical size of 0.05.

The results show that our proposed test has considerably higher power than that of the SN test in all cases. Although the SN test is able to control its size around the nominal level, it suffers some power loss by avoiding estimating the long-run covariance matrix of the test statistic. This is consistent with the results of the theoretical power comparison of the two tests in Section 6. Our testing procedure is based on a novel kernel-smoothing estimator for the variance of the L_2-type BEU-statistic under high-dimensional time series data. Compared with the SN approach, the advantage in terms of power is a main contribution

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

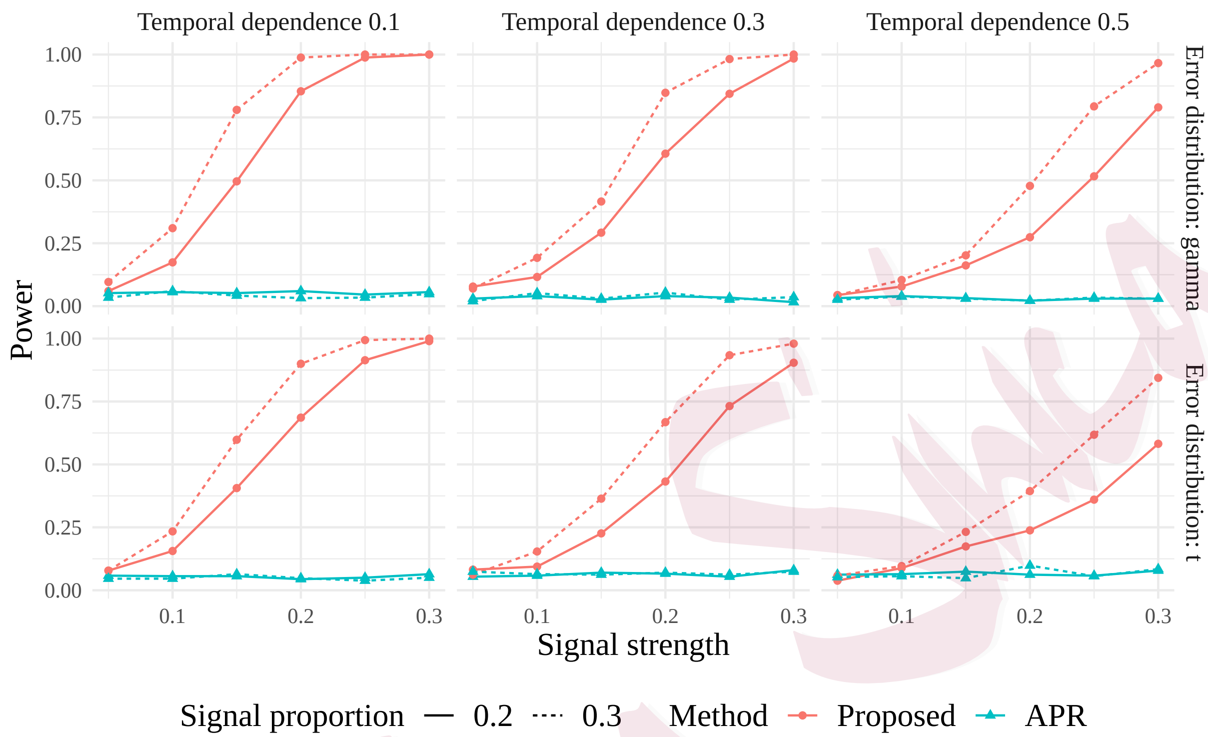


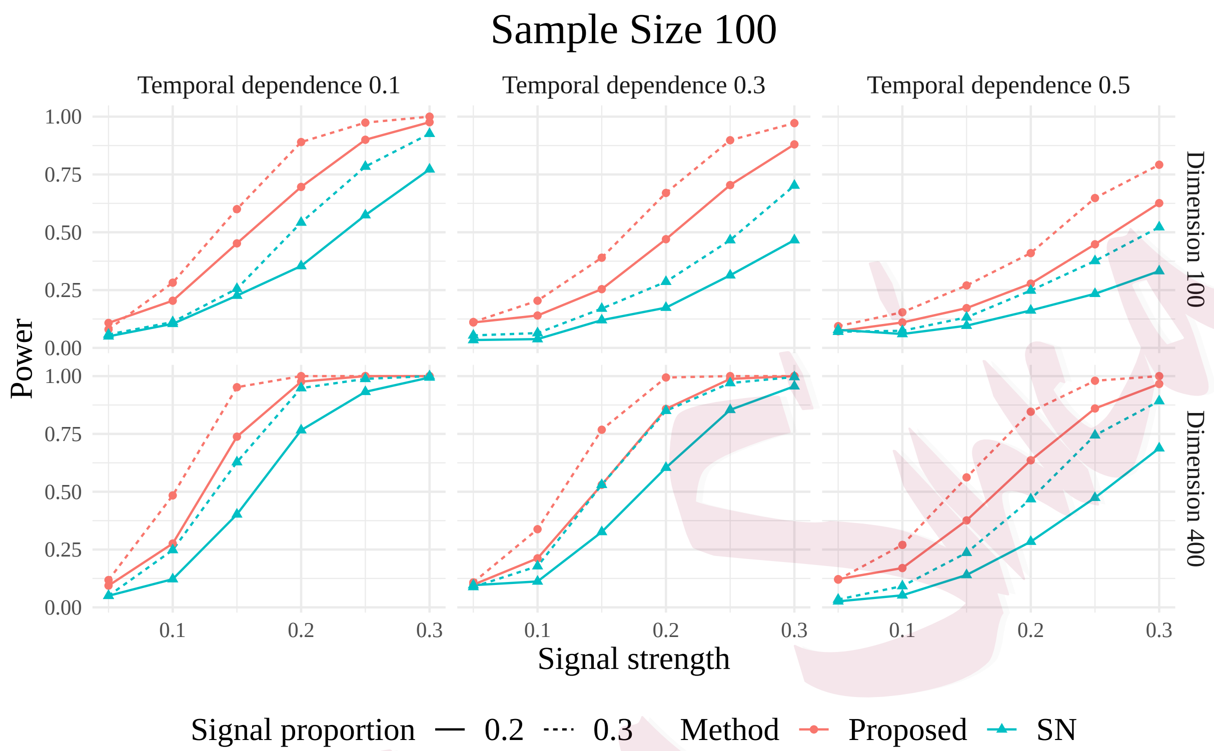
Figure 4: Empirical powers of the proposed test (red) and the APR test (blue) with respect to the signal strength r_0 (horizontal axis) for the two-sample hypotheses under the AR model with multivariate-t-distributed errors and gamma-distributed errors, sample size $n_0 = 100$, dimension $p = 400$, three levels of temporal dependence ρ_{time}, and two values of signal proportions.

of our proposed test.

8. Real-Data Analysis

In this section, we apply the proposed test to detect changes in stock returns and volatility before and after the financial crisis of 2008. We analyze the daily returns of S&P 500 stocks from January 2, 2007 to December 31, 2010, and use the capital asset pricing model to compare the performance of individual stocks against the market index. As a result of acquisitions and companies growing or shrinking in value, the list of companies on the S&P 500 changes over time. After excluding new and drop-out stocks in the S&P 500 from

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147



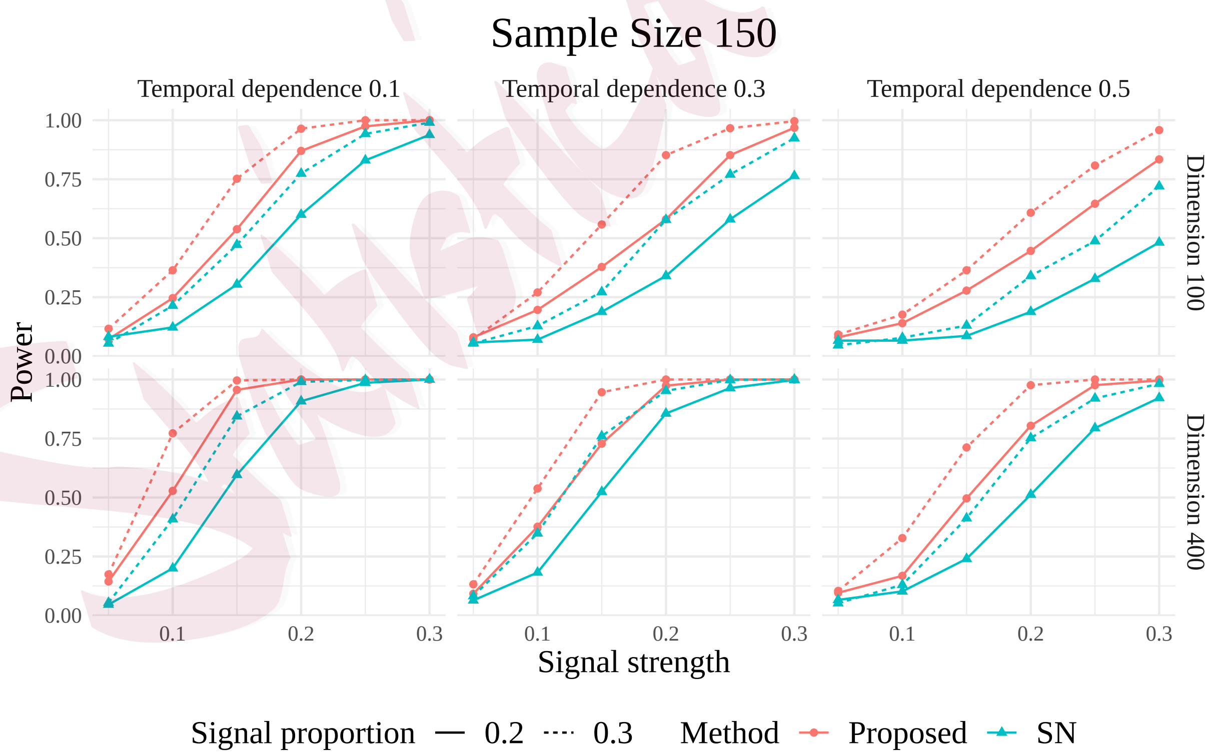


Figure 5: Empirical powers of the proposed test (red) and the self-normalized test (blue, denoted by SN) with respect to the signal strength r_0 (horizontal axis) for the one-sample hypotheses under the AR model, sample size $n_0 = 100, 150$, dimension $p = 100, 400$, three levels of temporal dependence ρ_{time}, and two values of signal proportions.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

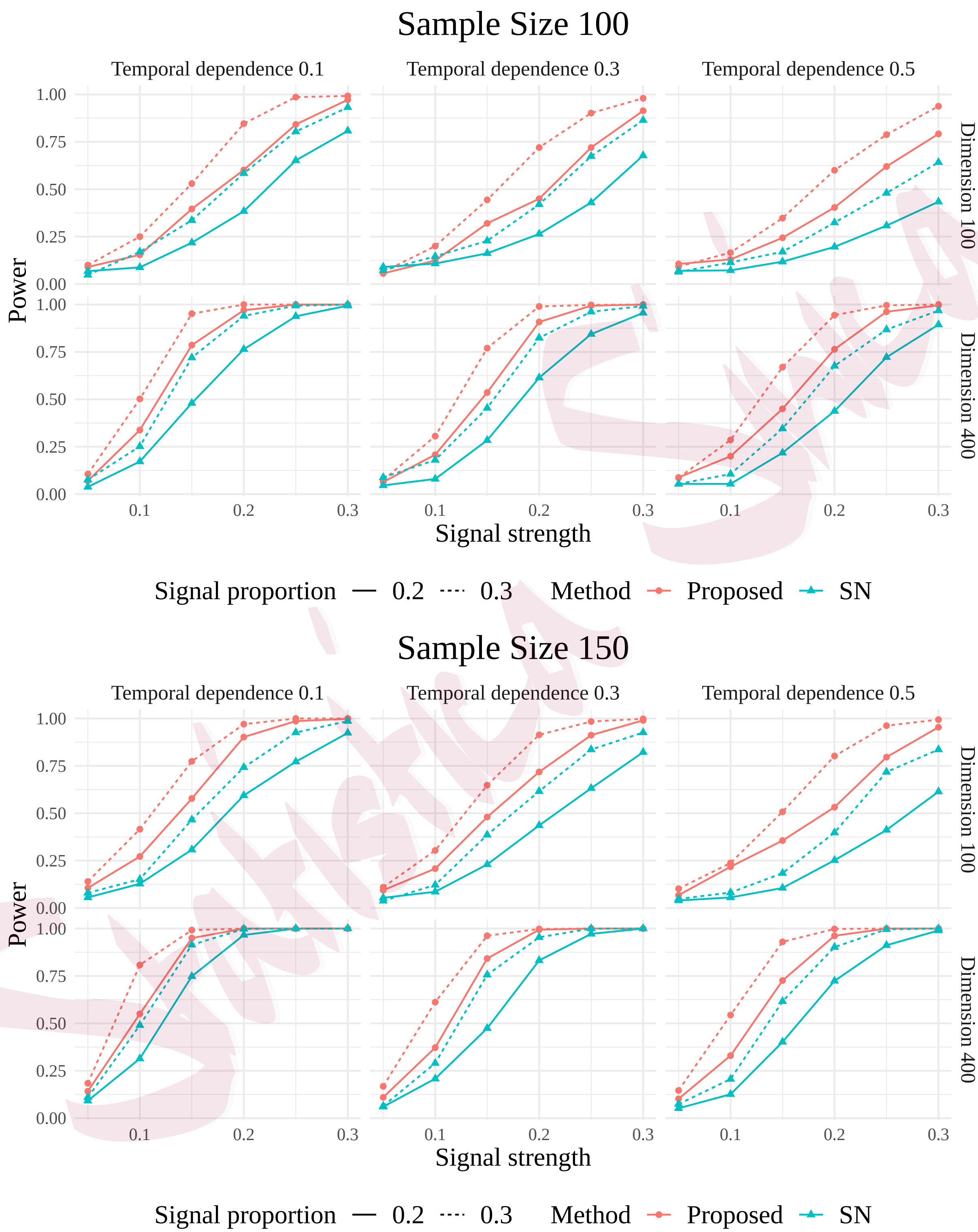


Figure 6: Empirical powers of the proposed test (red) and the self-normalized test (blue, denoted by SN) with respect to the signal strength r_0 (horizontal axis) for the one-sample hypotheses under the MA model, sample size $n_0 = 100, 150$, dimension $p = 100, 400$, three levels of temporal dependence ρ_{time}, and two values of signal proportions.

Table 3: Empirical sizes of the proposed test and the SN test (Wang and Shao, 2020) for the one-sample hypotheses under the temporal dependence parameter $\rho_{\text{time}} = 0.1, 0.3, 0.5$, $n_0 = 100, 150$, $p = 100, 400$, and the MA and AR models.

Method	(n_0, p)	ρ_{time} under MA			ρ_{time} under AR		
		0.1	0.3	0.5	0.1	0.3	0.5
Proposed	(100,100)	0.065	0.077	0.068	0.062	0.073	0.076
	(150,100)	0.066	0.062	0.071	0.082	0.079	0.077
	(100,400)	0.066	0.063	0.036	0.075	0.054	0.046
	(150,400)	0.081	0.054	0.044	0.071	0.077	0.050
SN	(100,100)	0.045	0.038	0.047	0.052	0.058	0.051
	(150,100)	0.052	0.055	0.066	0.045	0.052	0.065
	(100,400)	0.061	0.057	0.058	0.072	0.046	0.088
	(150,400)	0.044	0.051	0.050	0.072	0.045	0.058

January 2, 2007 to December 31, 2010, our sample includes data on 429 stocks. These stocks are divided into 11 sectors: Consumer Discretionary (64 stocks), Consumer Staples (31), Energy (17), Financials (60), Health Care (55), Industrials (58), Information Technology (66), Materials (22), Real Estate (25), Telecommunications Services (4), and Utilities (27). We apply the proposed high-dimensional test to all stocks in the sample, and to 10 sectors, excluding the Telecommunications Services sector, owing to its rather small dimension.

To evaluate the short-, medium- and long-term effects of the financial crisis on stock returns and volatility, we consider three designs for the time periods: (i) Design 1: March to August 2008 as period 1, and November 2008 to April 2009 as period 2; (ii) Design 2: January to August 2008 as period 1, and the complete 2009 as period 2; and (iii) Design 3: the complete 2007 and 2010 as periods 1 and 2, respectively. In Designs 1 and 2, September and October 2008 are excluded to avoid the extremely high volatility in the midst of the financial crisis. Design 3 offers a baseline setting, with the study periods far away from the midst of the crisis. The sample sizes of the two periods under the three designs are $n_1 = 126, 164, 242$ and $n_2 = 121, 244, 242$, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

For each of Designs 1–3, let $\boldsymbol{Y}_{i,t} = (Y_{i,t,1}, \ldots, Y_{i,t,p})^{\mathrm{T}}$ be the closing prices of the stocks on the tth day of the ith period, for $i = 1, 2$, $t = 1, \ldots, n_i$, and the dimension p equal to 429. Let $\tilde{X}_{i,t,j} = \log Y_{i,t,j} - \log Y_{i,t-1,j}$ be the return of the jth stock, and $X_{i,t,j}$ be the excess return of $\tilde{X}_{i,t,j}$, which is equal to $\tilde{X}_{i,t,j}$ minus the risk-free cash interest rate at the time. Similarly, let $\{Z_{i,t}\}_{t=1}^{n_i}$ be the excess return of the S&P 500 index in the ith period. We consider the single-index model (Sharpe, 1963)

$$X_{i,t,j} = \alpha_{i,j} + \beta_{i,j} Z_{i,t} + \epsilon_{i,t,j} \quad \text{and} \quad \mathrm{Var}(\epsilon_{i,t,j}) = \sigma_{i,j} \tag{8.1}$$

to adjust the portfolio return by the S&P 500 market index, for $j = 1, \ldots, p$. Under this model, the stock excess return is influenced by the market index through the beta coefficient of this stock, the alpha coefficient $\alpha_{i,j}$ indicates how the stock performs after accounting for the market risk, and the error variance $\sigma_{i,j}$ refers to the stock-specific risk. Let $\mathcal{S}_0$ and $\mathcal{S}_k$ for $k = 1, \ldots, 10$ denote the index set of all stocks and the stocks in the kth sector, respectively.

Let $\boldsymbol{\alpha}_{i,(k)} = (\alpha_{i,j} : j \in \mathcal{S}_k)$ and $\boldsymbol{\sigma}_{i,(k)} = (\sigma_{i,j} : j \in \mathcal{S}_k)$ be vectors of the alpha coefficients and error variances, respectively, of the kth sector. During the financial crisis in 2007–2008, many financial markets suffered their worst stock crash in history, reflected by the sudden dramatic decline in stock prices and extreme increase in volatility across almost all sections of the stock markets (Bates, 2012; Bardgett et al., 2019). We are interested in testing the change of the stock-adjusted return and specific volatility before and after the start of the financial crisis for each sector. That is, we wish to test the hypotheses

$$H_{0,\alpha,k} : \boldsymbol{\alpha}_{1,(k)} = \boldsymbol{\alpha}_{2,(k)} \quad \text{vs.} \quad H_{a,\alpha,k} : \boldsymbol{\alpha}_{1,(k)} \neq \boldsymbol{\alpha}_{2,(k)} \quad \text{and} \tag{8.2}$$

$$H_{0,\sigma,k} : \boldsymbol{\sigma}_{1,(k)} = \boldsymbol{\sigma}_{2,(k)} \quad \text{vs.} \quad H_{a,\sigma,k} : \boldsymbol{\sigma}_{1,(k)} \neq \boldsymbol{\sigma}_{2,(k)}, \tag{8.3}$$

for $k = 0, \ldots, 10$.

Using the estimated beta coefficient $\hat{\beta}_{i,j}$ from fitting (8.1), let $R_{i,t,j} = X_{i,t,j} - \hat{\beta}_{i,j} Z_{i,t}$

be the adjusted return of the jth stock, and $\tilde{R}_{i,t,j} = R_{i,t,j} - \bar{R}_{i,j}$ be the centered adjusted return, where $\bar{R}_{i,j} = \sum_{t=1}^{n_i} R_{i,t,j}/n_i$. We apply the proposed method on $\{R_{i,t,j}\}$ and $\{\tilde{R}^2_{i,t,j}\}$ to test hypotheses (8.2) and (8.3), respectively. Here, we treat the estimation of the beta coefficient $\beta_{i,j}$ as sufficiently accurate that the estimation error can be ignored when testing $\alpha_{i,j}$ and $\sigma_{i,j}$. The proposed work may be extended to testing for regression coefficients under high-dimensional time series data, which is left to future research.

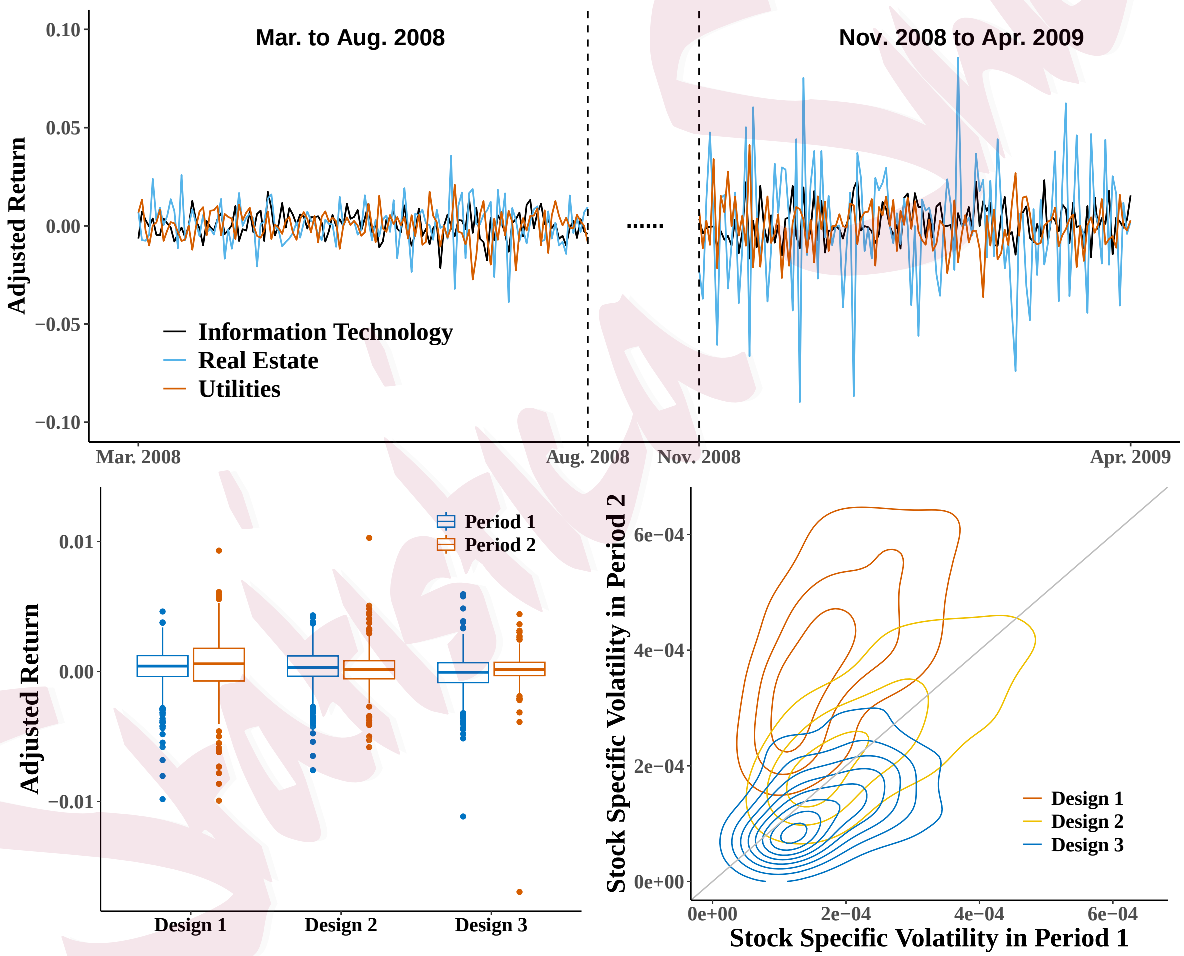


Figure 7: Time series plots of the average adjusted returns $\{\bar{R}^{\text{sec}}_{i,t,k}\}_{t=1}^{n_i}$ for three selected sectors in Design 1 (top panel) over two periods, box plots of the estimated alpha coefficients $\hat{\alpha}_{i,j}$ (bottom left panel), and a density contour plot of the estimated stock specific variance $\hat{\sigma}_{i,j}$ (bottom right panel) with the 45° line. The two lower panels are based on all selected 429 stocks.

Figure 7 displays a time series plot of the average adjusted return $\bar{R}^{\text{sec}}_{i,t,k} = |\mathcal{S}_k|^{-1} \sum_{j\in\mathcal{S}_k} R_{i,t,j}$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0147

for three selected sectors, a box plot of the estimated alpha coefficients $\hat{\alpha}_{i,j}$, and a contour plot of the estimated variance $\hat{\sigma}_{i,j}$ for all 429 selected stocks. Figure 7 shows that the overall means of the adjusted returns are centered around zero, both before and after the start of the economic crisis. The top panel also indicates an obvious increase in volatility in the first six months after the crisis, especially in the real estate sector. The box plot and the contour plot also show that the economic crisis led to an extremely volatile market in the short term, as reflected by Design 1. However, the volatility then gradually decreased to slightly lower than the pre-crisis level, as shown in Designs 2 and 3.

Table 4: Average differences of the estimates $\hat{\alpha}_{i,j}$ and $\hat{\sigma}_{i,j}$ between the two periods in Designs 1–3 within each sector, and the significance level of testing the hypotheses (8.2) for equality of the alpha coefficients and hypotheses (8.3) for equality of stock-specific volatility for the 10 sectors. The number of $*$ (one to three) represents the p-value of the proposed test within $[0.025, 0.05)$, $[0.01, 0.025)$ and $[0, 0.01)$, respectively.

Sector	Diff. of average alpha coefficient			Diff. of average volatility		
	Design 1	Design 2	Design 3	Design 1	Design 2	Design 3
Consumer Discretionary	16.434	2.319	15.753***	9.851***	1.342***	-2.742**
Consumer Staples	-0.188	1.621	0.81	-2.131	-3.069	-0.626
Energy	1.346	-3.816	-6.152	2.071*	-2.346***	-4.056
Financials	-3.755	-9.998	1.977**	22.504***	13.591*	9.695
Health Care	-3.589	2.77	-2.744	2.546	-0.43	-1.709
Industrials	-2.652	-7.334	3.432	3.173***	-1.235***	-2.285
Information Technology	9.31	8.366	1.221	3.949***	-0.858	-2.302***
Materials	12.093	-0.736	-2.759***	5.292***	0.017***	-1.339
Real Estate	-9.889	-18.526	10.887	17.137***	4.593***	-0.676
Utilities	-6.282	3.39	-2.109	2.941***	-0.683	-2.366
Overall	2.324	-1.402	3.204**	7.381***	1.67***	-0.424

Table 4 reports the average differences of the estimates $\hat{\alpha}_{i,j}$ and $\hat{\sigma}_{i,j}$ between the two periods for each sector, including the level of significance of the test. It shows that in Designs 1 and 2, the changes of the expected adjusted returns (alpha coefficient) over the two periods were all not significant. This is expected, because the expected value of the alpha coefficient should be zero in an efficient market (Jensen, 1969). However, the financial

doi:10.5705/ss.202022.0147

crisis greatly affected stock volatility, as shown in Design 1, in which eight of ten sectors exhibited a significant increase in volatility, and in Design 2, in which six sectors showed a significant increase in volatility. There were also significant increases in volatility after the financial crisis for the overall stocks under Designs 1 and 2. In contrast, under the baseline Design 3, only two sectors showed significant differences between the two periods, with a decrease rather than an increase in volatility.

Supplementary Material

The online Supplementary Material contains proofs of all theorems and lemmas, and additional results not reported in the main paper to conserve space.

Acknowledgments

The authors thank the anonymous referees, associate editor, and editor for their helpful and constructive comments and suggestions. Chen, Qiu and Zhang were partially supported by National Natural Science Foundation of China Grant 92358303. Chen was partially supported by National Natural Science Foundation of China Grants 12292980 and 12292983. Zhang was partially supported by National Key R&D Program of China Grants 2021YFA1000100 and 2021YFA1000101, National Natural Science Foundation of China Grants 71931004, 71931005 and 72201101, and the Science and Technology Commission of Shanghai Municipality Grant 21JS1400501.

References

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

Arbia, G. (2016). Spatial econometrics: A broad view. *Foundations and Trends in Econometrics* **8**, 145-265.

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**, 311-32

Bardgett, C., Gourier, E. and Leippold, M. (2019). Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets. *Journal of Financial Economics* **131**, 593-618.

Bates, D. S. (2012). US stock market crash risk, 1926–2010. *Journal of Financial Economics* **105**, 229-259.

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

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York.

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**, 349-372.

Chang, J., Zheng, C., Zhou, W. -X. and Zhou, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73**, 1300-1310.

Chang, J., Qiu, Y., Yao, Q. and Zou, T. (2018). Statistical inference for large precision matrices with dependent data and their applications. *Journal of Econometrics* **206**, 57–82.

Chen, L., Paul, D., Prentice, R. and Wang, P. (2011). A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *Journal of the American Statistical Association* **106**, 1345-1360.

Chen, S. X., Li, J. and Zhong, P.-S. (2019). Two-sample and ANOVA tests for high dimensional means. *The Annals of Statistics* **47**, 1443-1474.

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

Chen, S. X., Qiu, Y. and Zhang, S. (2023). Sharp optimality for high-dimensional covariance testing under sparse signals. *The Annals of Statistics* **51**, 1921-1945.

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.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2019). Inference on causal and structural parameters using many moment inequalities. *The Review of Economic Studies* **86**, 1867–1900.

Cui, X., Li, R., Yang, G. and Zhou, W. (2020). Empirical likelihood test for a large-dimensional mean vector. *Biometrika* **107**, 591-607.

Davydov, Y. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory of Probability & Its Applications* **13**, 691-696.

Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probability Theory and Related Fields* **132**, 203–236.

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics* **32**, 962 - 994.

Donoho, D. and Jin, J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science* **30**, 1–25.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Springer, New York.

Fan, J., Ke, Y., and Wang, K. (2020). Factor-adjusted regularized model selection. *Journal of Econometrics* **216**, 71–85.

Fan, J., Liao, Y. and Yao, J. (2015). Power enhancement in high dimensional cross-sectional tests. *Econometrica* **83**, 1497–1541.

Fan, J., Lv, J. and Qi, L. (2011). Sparse high-dimensional models in economics. *Annual Review of Economics* **3**, 291-317.

Feng, L., Zou, C., Wang, Z. and Zhu, L. (2015). Two-sample Behrens-Fisher problem for high-dimensional data. *Statistica Sinica* **25**, 1297-1312.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

Hall, P. and Jin, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *The Annals of Statistics* **38**, 1686-1732.

Han, F. and Wu, W. B. (2023). *Probability Inequalities for High Dimensional Time Series Under a Triangular Array Framework.* Springer, London.

Huang, Y., Li, C., Li, R. and Yang, S. (2021). An overview of tests on high-dimensional means. *Journal of Multivariate Analysis* **188**, 104813.

Jensen, M. C. (1969). Risk, the pricing of capital assets, and the evaluation of investment portfolios. *The Journal of Business* **42**: 167-247.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data.* Springer, Berlin.

Liu, X. and Chen, R. (2020). Threshold factor models for high-dimensional time series. *Journal of Econometrics* **216**, 53–70.

Nordman, D. J. and Lahiri, S. N. (2005). Validity of sampling window method for linear long-range dependent processes. *Econometric Theory* **21**, 1087–1111.

Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Academic Press, New York.

Qiu, Y., Chen, S. X. and Nettleton, D. (2018). Detecting rare and faint signals via thresholding maximum likelihood estimators. *The Annals of Statistics* **46**, 895-923.

Qiu, Y. and Zhou, X. H. (2022). Inference on multi-level partial correlations based on multi-subject time series data. *Journal of the American Statistical Association* **117**, 2268–2282.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science* **9**, 277-293.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* **97**, 1167-1179.

Wang, L., Peng, B. and Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association* **110**, 1658–1669.

Wang, R. and Shao, X. (2020). Hypothesis testing for high-dimensional time series via self-normalization. *The Annals of Statistics* **48**, 2728-2758.

Wong, K. C., Li, Z. and Tewari, A. (2020). Lasso guarantees for β-mixing heavy-tailed time series. *The Annals of Statistics* **48**, 1124–1142.

Xu, Z., Chen, S. X. and Wu, X. (2020). Meteorological change and impacts on air pollution: Results from North China. *Journal of Geophysical Research: Atmospheres* **125**, e2020JD032423.

Zhong, P. S., Chen, S. X. and Xu, M. Y. (2013). Tests alternative to higher criticism for high dimensional means under sparsity and column-wise dependence. *The Annals of Statistics* **41**, 2820-2851.

Shuyi Zhang

KLATASDS-MoE, School of Statistics, Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai 200062, China

E-mail: syzhang@fem.ecnu.edu.cn

Song Xi Chen

Guanghua School of Management and Center for Statistical Science, Peking University, Beijing 100871, China

E-mail: csx@gsm.pku.edu.cn

Yumou Qiu

School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing 100871, China

E-mail: qiuyumou@math.pku.edu.cn