ARTICLE IN PRESS
Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Testing the martingale difference hypothesis in high dimension[☆]

Jinyuan Chang[a,b], Qing Jiang[c,b], Xiaofeng Shao[d,*]

[a] *School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, Zhejiang Province, China*
[b] *Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Chengdu, Sichuan Province, China*
[c] *Center for Statistics and Data Science, Beijing Normal University, Zhuhai, Guangdong Province, China*
[d] *Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, USA*

ARTICLE INFO

Article history:
Received 10 July 2021
Received in revised form 26 August 2022
Accepted 9 September 2022
Available online xxxx

JEL classification:
C12
C15
C55

Keywords:
α-mixing
Gaussian approximation
High-dimensional statistical inference
Martingale difference hypothesis
Parametric bootstrap

ABSTRACT

In this paper, we consider testing the martingale difference hypothesis for high-dimensional time series. Our test is built on the sum of squares of the element-wise max-norm of the proposed matrix-valued nonlinear dependence measure at different lags. To conduct the inference, we approximate the null distribution of our test statistic by Gaussian approximation and provide a simulation-based approach to generate critical values. The asymptotic behavior of the test statistic under the alternative is also studied. Our approach is nonparametric as the null hypothesis only assumes the time series concerned is martingale difference without specifying any parametric forms of its conditional moments. As an advantage of Gaussian approximation, our test is robust to the cross-series dependence of unknown magnitude. To the best of our knowledge, this is the first valid test for the martingale difference hypothesis that not only allows for large dimension but also captures nonlinear serial dependence. The practical usefulness of our test is illustrated via simulation and a real data analysis. The test is implemented in a user-friendly R-function.

1. Introduction

Testing the martingale difference hypothesis is a fundamental problem in econometrics and time series analysis. The concept of martingale difference plays an important role in many areas of economics and finance. Several economic and financial theories such as the efficient markets hypothesis (Fama, 1970, 1991; LeRoy, 1989; Lo, 1997), rational expectations (Hall, 1978) and optimal asset pricing (Cochrane, 2005; Fama, 2013), yield such dependence restrictions on the underlying economic and financial variables. More formally, let $\{\mathbf{x}_t\}$ be a p-dimensional time series with $\mathbb{E}(\mathbf{x}_t) = \mathbf{0}$ for any $t \in \mathbb{Z}$. Write $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,p})^\top$ and denote by $\mathscr{F}_s$ the σ-field generated by $\{\mathbf{x}_t\}_{t\leqslant s}$. We call $\{\mathbf{x}_t\}_{t\in\mathbb{Z}}$ a martingale difference sequence (MDS) if and only if $\mathbb{E}(\mathbf{x}_t \mid \mathscr{F}_{t-1}) = \mathbf{0}$ for any $t \in \mathbb{Z}$. Given the observations $\{\mathbf{x}_t\}_{t=1}^n$, we are interested in the hypothesis testing problem:

$$H_0 : \{\mathbf{x}_t\}_{t\in\mathbb{Z}} \text{ is a MDS} \quad \text{versus} \quad H_1 : \{\mathbf{x}_t\}_{t\in\mathbb{Z}} \text{ is not a MDS.} \tag{1}$$

☆ The authors equally contributed to the paper. Chang and Jiang were supported in part by the National Natural Science Foundation of China (grant nos. 71991472, 72125008, 11871401 and 12001442). Chang was also supported by the Center of Statistical Research at Southwestern University of Finance and Economics. We would like to thank the editor, the associate editor and two reviewers for their constructive suggestions which led to substantial improvements.

* Corresponding author.
E-mail addresses: changjinyuan@swufe.edu.cn (J. Chang), jiangqing@bnu.edu.cn (Q. Jiang), xshao@illinois.edu (X. Shao).

https://doi.org/10.1016/j.jeconom.2022.09.001
0304-4076/© 2022 Elsevier B.V. All rights reserved.

Please cite this article as: J. Chang, Q. Jiang and X. Shao, Testing the martingale difference hypothesis in high dimension. Journal of Econometrics (2022), https://doi.org/10.1016/j.jeconom.2022.09.001.

The MDS hypothesis implies that the past information does not help to improve the prediction of future values of a MDS, so the best nonlinear predictor of the future values of a MDS given the current information set is just its unconditional expectation. The theme of the lack of predictability is of central interest in economics and finance and has stimulated a huge literature in both econometrics and time series analysis.

So far most of the work on MDS testing is restricted to the univariate case, i.e., $p = 1$. In one strand of literature, the MDS testing problem is reduced to testing the uncorrelatedness in either time domain or spectral domain. See Box and Pierce (1970), Ljung and Box (1978), Durlauf (1991), Hong (1996), Deo (2000), Lobato et al. (2001), and Shao (2011a,b), among others. These tests target on serial correlation but are unable to capture nonlinear serial dependence. There are examples of uncorrelated processes that are not MDS such as certain bilinear processes and nonlinear moving average processes, see Domínguez and Lobato (2003) for specific examples. Hence, it is important to develop tests that can go beyond linear serial dependence. In the specification testing literature, the exponential function based approach, pioneered by Bierens (1984, 1990), de Jong (1996) and Bierens and Ploberger (1997), is capable of detecting nonlinear serial dependence. Using the characteristic function, Hong (1999) proposed the generalized spectral density as a new tool for specification testing in a nonlinear time series framework; see Hong and Lee (2003) and Hong and Lee (2005) for further developments. As an interesting extension of Hong (1999), Escanciano and Velasco (2006) developed a MDS test based on the generalized spectral distribution function to capture nonlinear serial dependence at all lags. Parallel to the exponential/characteristic function based approach, the indicator/distribution function based approach has been taken by Stute (1997), Koul and Stute (1999), Domínguez and Lobato (2003), and Park and Whang (2005) among others. We refer to Escanciano and Lobato (2009) for a comprehensive review.

For the multivariate time series, i.e., $p > 1$, the literature for the MDS testing is scarce. Although it is expected that most of the above-mentioned tests can be extended to relatively low dimensional case, the theoretical and empirical properties of these tests are unknown. Recently, Hong et al. (2017) proposed a multivariate extension of the classical univariate variance ratio test (Lo and MacKinlay, 1988; Poterba and Summers, 1988; Chen and Deo, 2006) to test a weak form of the efficient markets hypothesis, i.e., uncorrelatedness of $\mathbf{x}_t$. As argued in Hong et al. (2017), the rationale to consider the MDS test for multivariate time series is that even if the MDS hypothesis holds for each component series $\{x_{t,j}\}_{t\in\mathbb{Z}}$, the MDS hypothesis could be violated at the multivariate level. In particular, the current return on the ith asset may be predicted by past observations of the jth asset. A univariate test may fail to detect this kind of cross-serial dependence, which can be captured by a multivariate test. Since it is well known that the variance ratio test only targets on serial correlation, the test of Hong et al. (2017) is unable to capture nonlinear serial dependence.

Nowadays, time series of moderate or high dimension are routinely collected or generated owing to the advance in science and technology. For example, S&P 500 index measures the stock performance of 500 large companies listed on stock exchanges in the United States, and it is tempting to ask whether the stock returns of the 500 companies are predictable at the daily or weekly frequency for a given time period (say, 5 years). The same question can be asked for the stocks within the same sector, such as those in the real estate sector (see Section 6 for data illustration). This naturally leads us to the regime where the dimension p is comparable to or exceeds the sample size n. To the best of our knowledge, there is no MDS testing procedure available in the literature that allows the dimension p to exceed the sample size n. Most of the aforementioned tests developed in the univariate setting require nontrivial modification to accommodate the high-dimensionality. The multivariate variance ratio test in Hong et al. (2017) allows for growing dimension p in their theory (i.e., $1/p+p/n = o(1)$) but is quite limited since their test cannot be implemented when $p > n$ and may encounter computational problems when p is large (say, $p > 120$); see Section 5 for more details.

To fill this gap, we introduce a new test for the MDS hypothesis of multivariate and possibly high-dimensional time series. We first use the element-wise max-norm of a sample-based matrix to characterize the nonlinear dependence of underlying p-dimensional time series $\{\mathbf{x}_t\}$ at a given lag $j \geqslant 1$, and then combine such information at different lags to propose our test statistic. Owing to the high-dimensionality and unknown temporal and cross-series dependence, the limiting null distribution of our test statistic is hard to derive, and it may even not have a closed form. To circumvent such difficulty, we employ the celebrated Gaussian approximation technique (Chernozhukov et al., 2013), which has undergone a rapid development recently, to establish the asymptotic equivalence between the null distribution of our test statistic and that of a certain function of a multivariate Gaussian random vector. Our theoretical analysis shows that our proposed test works even if p grows exponentially with respect to the sample size n, provided that some suitable regularity assumptions hold. To facilitate feasible inference, we propose a simulation-based approach to generate critical values. We also investigate the power behavior of our test under some local alternatives.

Since the seminal contribution of Chernozhukov et al. (2013), the literature on Gaussian approximation in the high-dimensional setting has been growing rapidly. For the sample mean of independent random vectors, we mention Chernozhukov et al. (2013, 2017), Deng and Zhang (2020), Fang and Koike (2021), Kuchibhotla et al. (2021), Chernozhukov et al. (2022a), and Chernozhukov et al. (2022b). For high-dimensional U-statistics and U-processes, see Chen (2018) and Chen and Kato (2019) for recent developments. The applicability of Gaussian approximation has also been extended to high-dimensional time series setting by Zhang and Wu (2017), Zhang and Cheng (2018), Chernozhukov et al. (2019) and Chang et al. (2021b). Also see Chang et al. (2017a,b,c, 2018b), and Yu and Chen (2021) among others for the use of Gaussian approximation or variants in high-dimensional statistical inference.

Zhang and Wu (2017) and Zhang and Cheng (2018) considered the Gaussian approximation for $\max_{1\leqslant j\leqslant p} n^{-1/2}\sum_{t=1}^{n} x_{t,j}$ with the physical dependence measure (Wu, 2005) imposed on $\{\mathbf{x}_t\}$, and Chernozhukov et al. (2019) considered the same problem when $\{\mathbf{x}_t\}$ is a β-mixing sequence. Chang et al. (2021b) studied the Gaussian approximations for $\mathbb{P}(n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t \in A)$ over some general classes of the set A (hyper-rectangles, simple convex sets and sparsely convex sets) under three different dependency framework (α-mixing, m-dependent, and physical dependence measure), which include the results obtained in Zhang and Wu (2017), Zhang and Cheng (2018) and Chernozhukov et al. (2019) as special cases. Compared to the use of Gaussian approximation results for high-dimensional time series in the existing works, our test statistic is considerably more involved and motivates us to develop new techniques for establishing the asymptotic equivalence between the null distribution of our test statistic and that of a certain function of a multivariate Gaussian random vector. More specifically, the theoretical analysis in this paper targets on the Gaussian approximation for some function of the high-dimensional vector $(n-K)^{-1/2}\sum_{t=1}^{n-K}\boldsymbol{\eta}_t$, where $\boldsymbol{\eta}_t$ is a newly defined vector based on $\{\mathbf{x}_t, \mathbf{x}_{t+1}, \ldots, \mathbf{x}_{t+K}\}$ and K is the number of lags involved in our test statistic. Since K is allowed to grow with the sample size n in our setting, the dependence structure among $\{\boldsymbol{\eta}_t\}$ will vary with K which cannot be covered in the frameworks of above mentioned works, and the existing Gaussian approximation results cannot be applied here. Some nontrivial technical challenges need to be addressed in our theoretical analysis.

From a methodological and practical viewpoint, we highlight a few appealing features of our proposed test:

(a) Our approach is nonparametric as the null hypothesis only assumes the time series concerned is martingale difference without specifying any parametric forms of its conditional moments. Hence, it is robust to second-order and higher-order conditional moments of unknown forms, including conditional heteroscedasticity, a prominent feature of many financial time series.

(b) It allows the dimension p to grow exponentially with respect to the sample size n, and works well for a broad range of dimension p even at a medium sample size (e.g., $n = 300$) as shown in our simulation studies. We have developed an R-function `MartG_test` in the package `HDTSA` which implements the test in an automatic manner.

(c) There is no particular requirement on the strength of cross-series dependence in our theory, so our test is applicable to time series with cross-series dependence of unknown magnitude. Strong cross-series dependence has been commonly observed in many real high-dimensional time series data.

The rest of this paper is organized as follows. The methodology and theoretical analysis are given in Sections 2 and 3, respectively. Section 4 extends the proposed test to more general settings. Section 5 studies the finite sample performance of our proposed test. A real data analysis is presented in Section 6. Section 7 concludes the paper. Section 8 includes the mathematical proofs of our main results. Some additional technical arguments and numerical studies are given in the supplementary material. At the end of this section, we introduce some notation that is used throughout the paper. For any positive integer $q \geqslant 2$, we write $[q] = \{1, \ldots, q\}$ and denote by $\mathbb{S}^{q-1}$ the q-dimensional unit sphere. For any $q_1 \times q_2$ matrix $\mathbf{M} = (m_{i,j})_{q_1\times q_2}$, let $|\mathbf{M}|_\infty = \max_{i\in[q_1],j\in[q_2]}|m_{i,j}|$ and $|\mathbf{M}|_0 = \sum_{i=1}^{q_1}\sum_{j=1}^{q_2} I(m_{i,j} \neq 0)$, where $I(\cdot)$ denotes the indicator function. Specifically, if $q_2 = 1$, we use $|\mathbf{M}|_\infty = \max_{i\in[q_1]}|m_{i,1}|$ and $|\mathbf{M}|_0 = \sum_{i=1}^{q_1} I(m_{i,1} \neq 0)$ to denote the L_∞-norm and L_0-norm of the q_1-dimensional vector $\mathbf{M}$, respectively. For any q-dimensional vector $\mathbf{a} = (a_1, \ldots, a_q)^\top$, write $\psi(\mathbf{a})$ as the q-dimensional vector $\{\psi(a_1), \ldots, \psi(a_q)\}^\top$ for given function $\psi : \mathbb{R} \to \mathbb{R}$, and denote by $\mathbf{a}_{\mathcal{L}}$ the subvector of $\mathbf{a}$ collecting the components indexed by a given index set $\mathcal{L} \subset [q]$.

2. Methodology

2.1. Test statistic and the associated critical values

Let $\{\mathbf{x}_t\}$ be a p-dimensional time series with $\mathbb{E}(\mathbf{x}_t) = \mathbf{0}$ for any t. Given the observations $\{\mathbf{x}_t\}_{t=1}^n$, we shall develop a martingale difference hypothesis test that can capture certain nonlinear dependence between $\mathbf{x}_t$ and $\mathbf{x}_{t+j}$ for $j \in \mathbb{N}_+$. To this end, we let $\boldsymbol{\phi}(\cdot) : \mathbb{R}^p \to \mathbb{R}^d$ represent a map that is provided by the user. For example, $\boldsymbol{\phi}(\mathbf{x}) = \mathbf{x}$ is the linear identity map; $\boldsymbol{\phi}(\mathbf{x}) = \{\mathbf{x}^\top, (\mathbf{x}^2)^\top\}^\top$ includes both linear and quadratic terms, where $\mathbf{x}^2 = (x_1^2, \ldots, x_p^2)^\top$ with $\mathbf{x} = (x_1, \ldots, x_p)^\top$; $\boldsymbol{\phi}(\mathbf{x}) = \cos(\mathbf{x})$ captures certain type of nonlinear dependence, where $\cos(\mathbf{x}) = \{\cos(x_1), \ldots, \cos(x_p)\}^\top$ with $\mathbf{x} = (x_1, \ldots, x_p)^\top$.

Denote $\boldsymbol{\gamma}_j = (n-j)^{-1}\sum_{t=1}^{n-j}\mathbb{E}[\mathrm{vec}\{\boldsymbol{\phi}(\mathbf{x}_t)\mathbf{x}_{t+j}^\top\}]$ for each $j \geqslant 1$. Our proposal for testing the martingale difference hypothesis consists in checking all the pairwise covariance between $\boldsymbol{\phi}(\mathbf{x}_t)$ and $\mathbf{x}_{t+j}$, namely, our null hypothesis is now

$$H_0' : \boldsymbol{\gamma}_j = \mathbf{0} \text{ for all } j \geqslant 1\,. \tag{2}$$

It is easy to see that H_0 in (1) implies H_0' in (2) but not vice versa. In theory, it would be ideal to develop a test that is consistent with any violation of H_0 but this is very challenging in a model free setting, since the alternative we target is huge owing to the high-dimensionality and nonlinear serial dependence at all lags. As argued in Phillips and Jin (2014), "*Typically, the information set includes the infinite past history of the series, ... If a finite number of lagged values is included in the conditioning set, some dependence structure in the process may be missed due to omitted lags. However, tests that are designed to cope with the infinite lag case may have very low power (e.g., de Jong, 1996) and may not be feasible in empirical applications.*" Thus even in the low-dimensional setting, it is not clear whether there is a practical benefit for a test that is consistent with all alternatives. This motivates us to relax the null hypothesis H_0 and focus on the directional alternatives encoded by the function $\boldsymbol{\phi}(\cdot)$, which is pre-specified by the user and can incorporate some prior information.

Note that if the time series $\{\mathbf{x}_t\}$ is strictly stationary, then $\boldsymbol{\gamma}_j = \mathbb{E}[\mathrm{vec}\{\boldsymbol{\phi}(\mathbf{x}_0)\mathbf{x}_j^\top\}]$ which represents the population-level nonlinear dependence measure at lag j. In our asymptotic theory, no stationarity assumption needs to be imposed. To test H_0', it is natural to consider a test statistic with the following form

$$T_n = n\sum_{j=1}^{K} |\hat{\boldsymbol{\gamma}}_j|_\infty^2 \,, \tag{3}$$

where $\hat{\boldsymbol{\gamma}}_j = (n-j)^{-1}\sum_{t=1}^{n-j}\mathrm{vec}\{\boldsymbol{\phi}(\mathbf{x}_t)\mathbf{x}_{t+j}^\top\}$ is the estimator of $\boldsymbol{\gamma}_j$. Here $K = o(n)$ is a truncation lag and is allowed to grow with respect to the sample size n. This flexibility is important when there exists nonlinear serial dependence at large lags.

Intuitively, a large value of T_n provides evidence against H_0' in (2) and then we can reject H_0 in (1) if

$$T_n > \mathrm{cv}_\alpha \,, \tag{4}$$

where $\mathrm{cv}_\alpha > 0$ is the critical value at the significance level $\alpha \in (0,1)$. To determine cv_α, we need to derive the distribution of T_n under H_0. Write $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1^\top,\ldots,\hat{\boldsymbol{\gamma}}_K^\top)^\top$ and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top,\ldots,\boldsymbol{\gamma}_K^\top)^\top$. For fixed (p,d,K) and under suitable moment and weak dependence conditions, it follows from the central limit theorem that $\sqrt{n}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) \to_d \mathcal{N}(\mathbf{0}, \mathring{\boldsymbol{\Sigma}}_K)$ as $n\to\infty$ for some positive definite matrix $\mathring{\boldsymbol{\Sigma}}_K \in \mathbb{R}^{(Kpd)\times(Kpd)}$. Let $\mathring{\mathbf{g}} := (\mathring{g}_1,\ldots,\mathring{g}_{Kpd})^\top \sim \mathcal{N}(\mathbf{0},\mathring{\boldsymbol{\Sigma}}_K)$. By the continuous mapping theorem, the distribution of T_n under H_0 can be approximated by that of its Gaussian analogue $\mathring{G}_K = \sum_{j=1}^K |\mathring{\mathbf{g}}_{\mathcal{L}_j}|_\infty^2$ in the scenario with fixed (p,d,K), where $\mathcal{L}_j = \{(j-1)pd+1,\ldots,jpd\}$. Write $\tilde{n} = n-K$ and let

$$\boldsymbol{\eta}_t = ([\mathrm{vec}\{\boldsymbol{\phi}(\mathbf{x}_t)\mathbf{x}_{t+1}^\top\}]^\top,\ldots,[\mathrm{vec}\{\boldsymbol{\phi}(\mathbf{x}_t)\mathbf{x}_{t+K}^\top\}]^\top)^\top \tag{5}$$

for any $t\in[\tilde{n}]$. Define

$$\boldsymbol{\Sigma}_{n,K} = \mathrm{Cov}\bigg(\frac{1}{\sqrt{\tilde{n}}}\sum_{t=1}^{\tilde{n}}\boldsymbol{\eta}_t\bigg)\,, \tag{6}$$

which is the long-run covariance matrix of the sequence $\{\boldsymbol{\eta}_t\}_{t=1}^{\tilde{n}}$. For fixed (p,d,K), the asymptotic covariance $\mathring{\boldsymbol{\Sigma}}_K$ of $\sqrt{n}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})$ is essentially the limit of $\boldsymbol{\Sigma}_{n,K}$ specified in (6) as $n\to\infty$. In the high-dimensional scenarios, i.e., when (p,d,K) is diverging with respect to n, Proposition 1 indicates that such approximation for the null distribution of T_n is still valid even when p and d grow exponentially with respect to the sample size n.

Proposition 1. *Assume Conditions 1–3 in Section 3 hold and* $G_K = \sum_{j=1}^K |\mathbf{g}_{\mathcal{L}_j}|_\infty^2$, *where* $\mathbf{g} = (g_1,\ldots,g_{Kpd})^\top \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{n,K})$ *and* $\mathcal{L}_j = \{(j-1)pd+1,\ldots,jpd\}$. *Let* $K = O(n^\delta)$ *for some constant* $0 \leqslant \delta < f_1(\tau_1,\tau_2)$ *with* $f_1(\tau_1,\tau_2)$ *defined as (14) in Section 3. Then it holds that* $\sup_{x>0}|\mathbb{P}_{H_0}(T_n > x) - \mathbb{P}(G_K > x)| = o(1)$ *as* $n\to\infty$, *provided that* $\log(pd) = o(n^c)$ *for some constant* $c > 0$ *only depending on* (τ_1,τ_2,δ).

Proposition 1 reveals that the Kolmogorov–Smirnov distance between the null distribution of the proposed test statistic T_n and the distribution of G_K converges to zero, even when p and d diverge at some exponential rate of n. Letting

$$\mathrm{cv}_\alpha = \inf\{x>0 : \mathbb{P}(G_K \leqslant x) \geqslant 1-\alpha\} \tag{7}$$

in (4), Proposition 1 yields that $\mathbb{P}_{H_0}(T_n > \mathrm{cv}_\alpha) \to \alpha$ as $n\to\infty$. Since the long-run covariance matrix $\boldsymbol{\Sigma}_{n,K}$ is usually unknown in practice, we need to replace it by some estimate $\widehat{\boldsymbol{\Sigma}}_{n,K}$ and then use $\hat{\mathrm{cv}}_\alpha$ defined below to approximate the desired critical value cv_α specified in (7):

$$\hat{\mathrm{cv}}_\alpha = \inf\{x>0 : \mathbb{P}(\hat{G}_K \leqslant x \,|\, \mathcal{X}_n) \geqslant 1-\alpha\}\,, \tag{8}$$

where $\mathcal{X}_n = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ and $\hat{G}_K = \sum_{j=1}^K |\hat{\mathbf{g}}_{\mathcal{L}_j}|_\infty^2$ with $\hat{\mathbf{g}} := (\hat{g}_1,\ldots,\hat{g}_{Kpd})^\top \sim \mathcal{N}(\mathbf{0},\widehat{\boldsymbol{\Sigma}}_{n,K})$ and $\mathcal{L}_j = \{(j-1)pd+1,\ldots,jpd\}$. Then we reject the null hypothesis H_0 specified in (1) if

$$T_n > \hat{\mathrm{cv}}_\alpha \,. \tag{9}$$

We defer the details of $\widehat{\boldsymbol{\Sigma}}_{n,K}$ to Section 2.2.

Remark 1. If we select the function $\boldsymbol{\phi}(\mathbf{x}) = \mathbf{x}$, the test statistic T_n defined in (3) can also be applied for testing the high-dimensional white noise hypothesis, i.e., $H_0 : \{\mathbf{x}_t\}_{t\in\mathbb{Z}}$ is white noise versus $H_1 : \{\mathbf{x}_t\}_{t\in\mathbb{Z}}$ is not white noise. Chang et al. (2017a) considered this hypothesis testing problem with L_∞-type test statistic using the maximum absolute autocorrelations and cross-correlations of the component series in $\mathbf{x}_t$ over all lags $k\in[K]$. It is well known that the L_∞-type test statistic is powerful against the sparse alternatives, that is, only a small fraction of the elements in $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top,\ldots,\boldsymbol{\gamma}_K^\top)^\top$ are nonzero, while it can be powerless for the dense but faint alternatives, i.e., when most elements in $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top,\ldots,\boldsymbol{\gamma}_K^\top)^\top$ are nonzero but with very small magnitudes. To remedy such weakness, our proposed T_n in (3) combines the signals from different lags together using the sum of squares and is expected to improve the power performance in case of dense but faint alternatives. On the technical side, constructing the Gaussian approximation to the null distribution of T_n defined in (3) is more challenging than that for the L_∞-type statistic used in Chang et al.

(2017a). Chang et al. (2017a) only considered the case with fixed K under the β-mixing assumption. The null distribution of their test statistic can be easily obtained by the associated Gaussian approximation results developed in Chernozhukov et al. (2019). In this paper, we only impose the α-mixing assumption on $\{\mathbf{x}_t\}$ and the corresponding α-mixing coefficients of $\{\boldsymbol{\eta}_t\}$ become a triangular array owing to the divergence of K. To the best of our knowledge, our paper is the first attempt to derive the Gaussian approximation results in such a complex setting.

Remark 2. As we mentioned earlier, the only paper that allows growing dimension for the martingale difference hypothesis testing is Hong et al. (2017), which generalized the variance ratio test to multivariate time series. In their asymptotic theory, they considered both finite/fixed horizon (i.e., fixed K) and increasing horizon (i.e., $K \to \infty$ but $K^2/n \to 0$), which is also allowed in our theory. In their Theorem 7, they presented the limiting null distribution of a particular test statistic Zd_{tr} under the restriction that the dimension p grows but $p/n \to 0$. Their another two test statistics Z_{tr} and Z_{det} for the setting of fixed p cannot be implemented in practice when $p > \sqrt{n}$. By contrast, our test statistic can work for a much broader range of p, including the case $p \gg n$, and thus is advantageous in dealing with the martingale difference hypothesis testing for high-dimensional time series. In addition, we can capture nonlinear serial dependence owing to the flexibility of user-chosen $\boldsymbol{\phi}(\cdot)$, which yields a nonlinear dependence measure. In practice, we need to set the lags K and the user-chosen map $\boldsymbol{\phi}(\cdot)$, which can incorporate some prior information we have. For example, if the time series is expected to exhibit seasonal dependence, then K should be large enough to include some seasonal lags. If we are dealing with stock return data, then including quadratic terms in $\boldsymbol{\phi}(\cdot)$ might help to capture potential nonlinear dependence.

Remark 3. If the time series $\{\mathbf{x}_t\}$ is strictly stationary, we know the transformed data $\{\boldsymbol{\eta}_t\}$ is also strictly stationary and our test statistic T_n given in (3) essentially converts the MDS testing problem for $\mathbf{x}_t$ to testing zero mean for the transformed data $\boldsymbol{\eta}_t$. There are indeed several papers in the literature of Gaussian approximation that tackle the mean testing problem for high-dimensional time series; see Zhang and Wu (2017), Zhang and Cheng (2018), Chernozhukov et al. (2019) and Chang et al. (2021b). Zhang and Wu (2017) and Zhang and Cheng (2018) considered the Gaussian approximation theory in the framework that assumes the physical dependence (Wu, 2005) for $\{\boldsymbol{\eta}_t\}$. Chernozhukov et al. (2019) and Chang et al. (2021b) considered the Gaussian approximation theory, respectively, in the frameworks that assume the β-mixing assumption and α-mixing assumption for $\{\boldsymbol{\eta}_t\}$. Notice that the dependence structure among $\{\boldsymbol{\eta}_t\}$ will vary with K. The dependence frameworks for $\{\boldsymbol{\eta}_t\}$ assumed in these existing works do not cover our current setting, thus the existing Gaussian approximation results cannot be used for approximating the null distribution of our proposed test statistic T_n.

2.2. Estimation of long-run covariance matrix

In the low-dimensional setting, long-run covariance matrix estimation (or heteroscedastic–autocorrelation-consistent estimation) is a classic problem in econometrics and time series analysis and there is a rich literature. We refer the readers to two foundational papers by Newey and West (1987) and Andrews (1991). In the high-dimensional setting, the estimator proposed in the low-dimensional environment can still be used, but establishing the proper probabilistic bounds for the difference is very challenging. Recall $\tilde{n} = n - K$. Following Chang et al. (2017a), we adopt the following estimate for the long-run covariance matrix $\boldsymbol{\Sigma}_{n,K}$:

$$\widehat{\boldsymbol{\Sigma}}_{n,K} = \sum_{j=-\tilde{n}+1}^{\tilde{n}-1} \mathcal{K}\left(\frac{j}{b_n}\right)\widehat{\mathbf{H}}_j\,, \tag{10}$$

where $\widehat{\mathbf{H}}_j = \tilde{n}^{-1}\sum_{t=j+1}^{\tilde{n}}(\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}})(\boldsymbol{\eta}_{t-j} - \bar{\boldsymbol{\eta}})^\top$ if $j \geqslant 0$ and $\widehat{\mathbf{H}}_j = \tilde{n}^{-1}\sum_{t=-j+1}^{\tilde{n}}(\boldsymbol{\eta}_{t+j} - \bar{\boldsymbol{\eta}})(\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}})^\top$ otherwise, with $\bar{\boldsymbol{\eta}} = \tilde{n}^{-1}\sum_{t=1}^{\tilde{n}}\boldsymbol{\eta}_t$. Here $\mathcal{K}(\cdot)$ is a symmetric kernel function that is continuous at 0, and b_n is the bandwidth diverging with n. The theoretical property of $\widehat{\boldsymbol{\Sigma}}_{n,K}$ defined as (10) is summarized in Proposition 2 in Section 3. As indicated in Andrews (1991), to make $\widehat{\boldsymbol{\Sigma}}_{n,K}$ given in (10) be positive semi-definite, we can require the kernel function $\mathcal{K}(\cdot)$ to satisfy $\int_{-\infty}^{\infty}\mathcal{K}(x)e^{-\mathrm{i}x\lambda}\,\mathrm{d}x \geqslant 0$ for any $\lambda \in \mathbb{R}$, where $\mathrm{i} = \sqrt{-1}$. The Bartlett kernel, Parzen kernel and Quadratic Spectral kernel all satisfy this requirement. See Section 5 for the explicit forms of these kernels.

Given $\widehat{\boldsymbol{\Sigma}}_{n,K}$, to compute $\hat{\mathrm{cv}}_\alpha$ given in (8), we need to generate $\hat{\mathbf{g}} := (\hat{g}_1, \ldots, \hat{g}_{Kpd})^\top \sim \mathcal{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_{n,K})$. Notice that $\widehat{\boldsymbol{\Sigma}}_{n,K}$ is a $(Kpd) \times (Kpd)$ matrix. The standard procedure is based on the Cholesky decomposition of $\widehat{\boldsymbol{\Sigma}}_{n,K}$ and generating $\hat{\mathbf{g}}$ is a computationally $(nK^2p^2d^2 + K^3p^3d^3)$-hard problem that requires a large storage space for $\widehat{\boldsymbol{\Sigma}}_{n,K}$. In practice, p and d can be quite large. As suggested in Chang et al. (2017a), we can generate $\hat{\mathbf{g}}$ as follows:

Algorithm 1 Procedure for generating $\hat{\mathbf{g}}$

Step 1. Let $\boldsymbol{\Theta}$ be a $\tilde{n} \times \tilde{n}$ matrix with (i, j)th element $\mathcal{K}\{(i - j)/b_n\}$.

Step 2. Generate $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{\tilde{n}})^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta})$ independent of $\mathcal{X}_n$.

Step 3. Let $\hat{\mathbf{g}} = (\hat{g}_1, \ldots, \hat{g}_{Kpd})^\top = \tilde{n}^{-1/2}\sum_{t=1}^{\tilde{n}}\xi_t(\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}})$.

We can show that $\hat{\mathbf{g}}$ obtained in Algorithm 1 satisfies $\hat{\mathbf{g}}\,|\,\mathcal{X}_n \sim \mathcal{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_{n,K})$. The computational complexity of Step 2 in Algorithm 1 is just $O(n^3)$ which is independent of (p, d). When p and d are large, the required storage space of Algorithm 1 is also much smaller than that of the standard procedure since it only requires to store $\{\boldsymbol{\eta}_t\}_{t=1}^{\tilde{n}}$ and $\bar{\boldsymbol{\eta}}$ rather than $\widehat{\boldsymbol{\Sigma}}_{n,K}$. In practice, we can draw $\hat{\mathbf{g}}_1, \ldots, \hat{\mathbf{g}}_B$ independently by Algorithm 1 for some large integer B and then take the $\lfloor B\alpha \rfloor$th largest value among $\hat{G}_{K,1}, \ldots, \hat{G}_{K,B}$ to approximate $\hat{\text{cv}}_\alpha$ defined as (8), where $\hat{G}_{K,i} = \sum_{j=1}^K |\hat{\mathbf{g}}_{i,\mathcal{L}_j}|_\infty^2$ with $\hat{\mathbf{g}}_i = (\hat{g}_{i,1}, \ldots, \hat{g}_{i,Kpd})^\top$ and $\mathcal{L}_j = \{(j-1)pd + 1, \ldots, jpd\}$.

3. Theoretical property

Recall $T_n = n\sum_{j=1}^K |\hat{\boldsymbol{\gamma}}_j|_\infty^2$. Since the distribution of $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1^\top, \ldots, \hat{\boldsymbol{\gamma}}_K^\top)^\top$ can be well approximated by that of $\bar{\boldsymbol{\eta}} = \tilde{n}^{-1}\sum_{t=1}^{\tilde{n}} \boldsymbol{\eta}_t$ with $\tilde{n} = n - K$, the difference between the distributions of T_n and $\tilde{T}_n := \tilde{n}\sum_{j=1}^K |\bar{\boldsymbol{\eta}}_{\mathcal{L}_j}|_\infty^2$ is expected to be asymptotically negligible. See Lemma L2 in Section 8 for details. The key step in our theoretical analysis is to approximate the null distribution of $\tilde{T}_n$ by Gaussian approximation.

For any $j_1, \ldots, j_K \in [pd]$ and $x > 0$, let $\mathcal{A}_{j_1,\ldots,j_K}(x) = \{\mathbf{b} \in \mathbb{R}^{Kpd} : \mathbf{b}_{S_{j_1,\ldots,j_K}}^\top \mathbf{b}_{S_{j_1,\ldots,j_K}} \leqslant x\}$ with $S_{j_1,\ldots,j_K} = \{j_1, j_2+pd, \ldots, j_K + (K-1)pd\}$. Define $\mathcal{A}(x;K) = \bigcap_{j_1=1}^{pd}\cdots\bigcap_{j_K=1}^{pd}\mathcal{A}_{j_1,\ldots,j_K}(x)$. We then have $\{\tilde{T}_n \leqslant x\} = \{\tilde{n}^{1/2}\bar{\boldsymbol{\eta}} \in \mathcal{A}(x;K)\}$. Note that the set $\mathcal{A}_{j_1,\ldots,j_K}(x)$ is convex that only depends on the components in $S_{j_1,\ldots,j_K}$. We can reformulate $\mathcal{A}_{j_1,\ldots,j_K}(x)$ as follows:

$$\mathcal{A}_{j_1,\ldots,j_K}(x) = \bigcap_{\mathbf{a}\in\mathbb{S}^{Kpd-1}:\,\mathbf{a}_{S_{j_1,\ldots,j_K}}\in\mathbb{S}^{K-1}} \{\mathbf{b}\in\mathbb{R}^{Kpd} : \mathbf{a}^\top\mathbf{b} \leqslant x^{1/2}\}.$$

Define $\mathcal{F} = \bigcup_{j_1=1}^{pd}\cdots\bigcup_{j_K=1}^{pd}\{\mathbf{a}\in\mathbb{S}^{Kpd-1} : \mathbf{a}_{S_{j_1,\ldots,j_K}} \in \mathbb{S}^{K-1}\}$. Then $\mathcal{A}(x;K) = \bigcap_{\mathbf{a}\in\mathcal{F}}\{\mathbf{b}\in\mathbb{R}^{Kpd} : \mathbf{a}^\top\mathbf{b}\leqslant x^{1/2}\}$ and

$$\{\tilde{T}_n \leqslant x\} = \left\{\frac{1}{\sqrt{\tilde{n}}}\sum_{t=1}^{\tilde{n}} \mathbf{a}^\top\boldsymbol{\eta}_t \leqslant x^{1/2} \text{ for any } \mathbf{a}\in\mathcal{F}\right\} \tag{11}$$

for any $x > 0$. As indicated in (11), to construct the Gaussian approximation of $\mathbb{P}_{H_0}(\tilde{T}_n \leqslant x)$, we need to impose the following assumption on the tail behavior of $\mathbf{a}^\top\boldsymbol{\eta}_t$. See also Chernozhukov et al. (2017) and Chang et al. (2021b).

Condition 1. *There exist some universal constants $C_1 > 1$, $C_2 > 0$ and $\tau_1 \in (0, 1]$ independent of (K, p, d, n) such that*

$$\sup_{t\in[n]}\sup_{\mathbf{a}\in\mathcal{F}} \mathbb{P}(|\mathbf{a}^\top\boldsymbol{\eta}_t| > x) \leqslant C_1\exp(-C_2x^{\tau_1})$$

for any $x > 0$.

Condition 1 is stronger than necessary for the theoretical justification of our proposed method, and it can be weakened at the expense of much lengthier proofs. For example, Condition 1 can be replaced by the assumption:

$$\max_{t\in[n]}\max_{\ell\in[Kpd]} \mathbb{P}(|\eta_{t,\ell}| > x) \leqslant C_1\exp(-C_2x^{\tau_1}) \tag{12}$$

for any $x > 0$. Recall $\eta_{t,\ell} = \phi_{l_1}(\mathbf{x}_t)x_{t+k,l_2}$ for some $l_1 \in [d]$, $l_2 \in [p]$ and $k \in [K]$. If $\boldsymbol{\phi}(\cdot)$ is selected as some bounded functions, then (12) holds provided that $\max_{t\in[n]}\max_{l_2\in[p]}\mathbb{P}(|x_{t,l_2}| > x) \leqslant C_*\exp(-C_{**}x^{\tau_1})$ for any $x > 0$. If $\boldsymbol{\phi}(\cdot)$ and $\mathbf{x}_t$ satisfy $\max_{t\in[n]}\max_{l_1\in[d]}\mathbb{P}\{|\phi_{l_1}(\mathbf{x}_t)| > x\} \leqslant C_*\exp(-C_{**}x^{\tau_*})$ and $\max_{t\in[n]}\max_{l_2\in[p]}\mathbb{P}(|x_{t,l_2}| > x) \leqslant C_*\exp(-C_{**}x^{\tau_{**}})$ for any $x > 0$, by Lemma 2 of Chang et al. (2013), we know (12) holds with $\tau_1 = \tau_*\tau_{**}/(\tau_* + \tau_{**})$. For any $\mathbf{a}\in\mathcal{F}$, there exists $(j_1, \ldots, j_K) \in [pd]^K$ such that $\sum_{\ell=1}^K a_{j_\ell+(\ell-1)pd}^2 = 1$ and $a_j = 0$ for $j \notin S_{j_1,\ldots,j_K}$, which implies $\sum_{\ell=1}^K |a_{j_\ell+(\ell-1)pd}| \leqslant \sqrt{K}$. By Bonferroni inequality and (12), for any given $\mathbf{a}\in\mathcal{F}$, it holds that

$$\begin{aligned}\mathbb{P}(|\mathbf{a}^\top\boldsymbol{\eta}_t| > x) &\leqslant \sum_{\ell=1}^K \mathbb{P}\left\{|\eta_{t,j_\ell+(\ell-1)pd}| > \frac{x}{\sum_{\ell=1}^K |a_{j_\ell+(\ell-1)pd}|}\right\}\\ &\leqslant \sum_{\ell=1}^K \mathbb{P}\left\{|\eta_{t,j_\ell+(\ell-1)pd}| > \frac{x}{\sqrt{K}}\right\} \leqslant C_*K\exp(-C_{**}K^{-\tau_1/2}x^{\tau_1})\end{aligned} \tag{13}$$

for any $x > 0$, which provides a rough upper bound for $\max_{t\in[n]}\sup_{\mathbf{a}\in\mathcal{F}}\mathbb{P}(|\mathbf{a}^\top\boldsymbol{\eta}_t| > x)$. When K is a fixed positive integer, by (13), we know Condition 1 is satisfied provided that (12) holds. If we only assume (12), we can still establish the associated Gaussian approximation results based on (13) rather than Condition 1 but the associated arguments will be quite cumbersome.

Condition 2. *Assume that $\{\mathbf{x}_t\}$ is α-mixing in the sense that*

$$\alpha(k) := \sup_t \sup_{(A,B)\in\mathscr{F}_{-\infty}^t\times\mathscr{F}_{t+k}^{+\infty}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \to 0 \ \ \textit{as } k\to\infty,$$

where $\mathscr{F}_{-\infty}^{u}$ and $\mathscr{F}_{u+k}^{+\infty}$ are the σ-fields generated respectively by $\{\mathbf{x}_t\}_{t\leqslant u}$ and $\{\mathbf{x}_t\}_{t\geqslant u+k}$. Furthermore, there exist some universal constants $C_3 > 1$, $C_4 > 0$ and $\tau_2 \in (0, 1]$ independent of (K, p, d, n) such that $\alpha(k) \leqslant C_3 \exp(-C_4 k^{\tau_2})$ for all $k \geqslant 1$.

The α-mixing assumption in Condition 2 is weaker than the β-mixing assumption considered in Chernozhukov et al. (2019). Restricting $\tau_2 \in (0, 1]$ is just to simplify the presentation. If the α-mixing coefficients satisfy Condition 2 with some constant $\tau_2 > 1$, then Condition 2 will be satisfied automatically with $\tau_2 = 1$. Under certain conditions, VAR processes, multivariate ARCH processes, and multivariate GARCH processes all satisfy Condition 2 with $\tau_2 = 1$; see Hafner and Preminger (2009), Boussama et al. (2011) and Wong et al. (2020). In addition, if we only require $\sup_{t\in[n]} \sup_{\mathbf{a}\in\mathcal{F}} \mathbb{P}(|\mathbf{a}^\top \boldsymbol{\eta}_t| > x) = O\{x^{-(\nu+\epsilon)}\}$ for any $x > 0$ in Condition 1 and $\alpha(k) = O\{k^{-\nu(\nu+\epsilon)/(2\epsilon)}\}$ for all $k \geqslant 1$ in Condition 2 with some constants $\nu > 2$ and $\epsilon > 0$, we can also apply the Fuk-Nagaev-type inequalities to construct the upper bounds for the tail probabilities of certain statistics for which our testing procedure still works for Kpd diverging at some polynomial rate of n. We refer to section 3.2 of Chang et al. (2018a) for the implementation of the Fuk–Nagaev-type inequalities in such a scenario.

Condition 3. *There exists a universal constant $C_5 > 0$ independent of (K, p, d, n) such that*

$$\inf_{\mathbf{a}\in\mathcal{F}} \mathrm{Var}\bigg(\frac{1}{\sqrt{\tilde{n}}}\sum_{t=1}^{\tilde{n}} \mathbf{a}^\top \boldsymbol{\eta}_t\bigg) \geqslant C_5.$$

Condition 4. *The kernel function $\mathcal{K}(\cdot)$ is continuously differentiable with bounded derivatives on $\mathbb{R}$ satisfying* (i) $\mathcal{K}(0) = 1$, (ii) $\mathcal{K}(x) = \mathcal{K}(-x)$ *for any $x \in \mathbb{R}$, and* (iii) $|\mathcal{K}(x)| \leqslant C_6 |x|^{-\vartheta}$ *as $|x| \to \infty$ for some universal constants $C_6 > 0$ and $\vartheta > 1$.*

Condition 3 is a mild technical assumption for the validity of the Gaussian approximation which requires the long-run variance of the sequence $\{\mathbf{a}^\top \boldsymbol{\eta}_t\}$ to be non-degenerate. Note that there are no explicit requirements on the cross-series dependence, and both weak and strong cross-series dependence are allowed in our theory. Condition 4 is commonly used for the nonparametric estimation of the long-run covariance matrix; see Newey and West (1987) and Andrews (1991). For the kernel functions with bounded support such as Parzen kernel and Bartlett kernel, we have $\vartheta = \infty$ in Condition 4.

For τ_1 and τ_2 specified in Conditions 1 and 2, we define

$$f_1(\tau_1, \tau_2) = \min\bigg(\frac{1}{15}, \frac{7\tau_1\tau_2}{18\tau_1 + 18\tau_2 - 3\tau_1\tau_2}, \frac{\tau_2}{9 - 3\tau_2}\bigg). \tag{14}$$

Such defined $f_1(\tau_1, \tau_2)$ is used to control the divergence rate of K which is determined from the technical proofs of Gaussian approximation theory. See Proposition 1 in Section 2. Notice that $\tau_1, \tau_2 \in (0, 1]$. When $\tau_1 = \tau_2 = 1$, then $f_1(\tau_1, \tau_2) = 1/15$.

Assume that the bandwidth b_n involved in (10) satisfies $b_n \asymp n^\rho$ for some constant $0 < \rho < (\vartheta - 1)/(3\vartheta - 2)$ with ϑ specified in Condition 4. Let

$$f_2(\rho, \vartheta) = \min\bigg(\frac{\rho}{5}, \frac{2\rho + \vartheta - 1 - 3\rho\vartheta}{6\vartheta - 3}\bigg). \tag{15}$$

Such defined $f_2(\rho, \vartheta)$ is also used to control the divergence rate of K which is obtained from the estimation of long-run covariance matrix $\boldsymbol{\Sigma}_{n,K}$. See Proposition 2 below. For given kernel function $\mathcal{K}(\cdot)$, the parameter ϑ is determined. Since $\vartheta = \infty$ if $\mathcal{K}(\cdot)$ is selected as the kernel functions with bounded support such as Parzen kernel and Bartlett kernel, then $f_2(\rho, \infty) = \min\{\rho/5, (1 - 3\rho)/6\}$. For given $\vartheta > 1$, the optimal selection of ρ that maximizes $f_2(\rho, \vartheta)$ with respect to ρ is $(5\vartheta - 5)/(21\vartheta - 13)$ and the associated $f_2(\rho, \vartheta) = (\vartheta - 1)/(21\vartheta - 13)$.

Proposition 2. *Assume that Conditions 1, 2 and 4 hold. Let $b_n \asymp n^\rho$ for some constant $0 < \rho < (\vartheta - 1)/(3\vartheta - 2)$, and $K = O(n^\delta)$ for some constant $0 \leqslant \delta < f_2(\rho, \vartheta)$ with $f_2(\rho, \vartheta)$ defined as (15). Then $|\widehat{\boldsymbol{\Sigma}}_{n,K} - \boldsymbol{\Sigma}_{n,K}|_\infty = o_{\mathrm{p}}[K^{-3}\{\log(npd)\}^{-2}]$ provided that $\log(pd) = o(n^c)$ for some constant $c > 0$ only depending on $(\tau_1, \tau_2, \rho, \vartheta, \delta)$.*

Different from the existing literature of high-dimensional covariance matrix estimation, our procedure does not require $\widehat{\boldsymbol{\Sigma}}_{n,K}$ to be consistent under the matrix L_2-operator norm and therefore it can work without imposing any structural assumptions on the underlying long-run covariance matrix $\boldsymbol{\Sigma}_{n,K}$. More specifically, our procedure only requires $|\widehat{\boldsymbol{\Sigma}}_{n,K} - \boldsymbol{\Sigma}_{n,K}|_\infty = o_{\mathrm{p}}[K^{-3}\{\log(npd)\}^{-2}]$, which is a quite mild requirement and our proposed $\widehat{\boldsymbol{\Sigma}}_{n,K}$ in Section 2.2 satisfies this even when p and d grow exponentially with n. Now we are ready to present the theoretical guarantees of our testing procedure (9).

Theorem 1. *Assume Conditions 1–4 hold. Let $b_n \asymp n^\rho$ for some constant $0 < \rho < (\vartheta - 1)/(3\vartheta - 2)$. Select $K = O(n^\delta)$ for some constant $0 \leqslant \delta < \min\{f_1(\tau_1, \tau_2), f_2(\rho, \vartheta)\}$ with $f_1(\tau_1, \tau_2)$ and $f_2(\rho, \vartheta)$ defined as (14) and (15), respectively. Then $\mathbb{P}_{H_0}(T_n > \hat{\mathrm{cv}}_\alpha) \to \alpha$ as $n \to \infty$, provided that $\log(pd) = o(n^c)$ for some constant $c > 0$ only depending on $(\tau_1, \tau_2, \rho, \vartheta, \delta)$.*

Theorem 1 reveals the validity of our proposed test in the sense that the testing procedure maintains the nominal significance level asymptotically under the null hypothesis, where pd is allowed to diverge exponentially with respect to the sample size n. In Theorem 2, the asymptotic power of the proposed tests is analyzed.

Theorem 2. *Assume the conditions of Theorem 1 hold. Let ϱ be the largest element in the main diagonal of $\boldsymbol{\Sigma}_{n,K}$, and write $\lambda(K,p,d,\alpha)=\{2\log(pd)\}^{1/2}+\{2\log(4K/\alpha)\}^{1/2}$. If $\sum_{j=1}^{K}|\boldsymbol{\gamma}_j|_\infty^2\geqslant n^{-1}K\varrho\lambda^2(K,p,d,\alpha)(1+\epsilon_n)^2$ under the alternative hypothesis for some $\epsilon_n>0$ satisfying $\epsilon_n\to 0$ and $\varrho\lambda^2(K,p,d,\alpha)K^{-1}(\log K)^{-1}\epsilon_n^2\to\infty$, then $\mathbb{P}_{H_1}(T_n>\hat{\mathrm{cv}}_\alpha)\to 1$ as $n\to\infty$.*

Theorem 2 shows that our proposed test is consistent under local alternatives. Recall $\boldsymbol{\gamma}=(\boldsymbol{\gamma}_1^\top,\ldots,\boldsymbol{\gamma}_K^\top)^\top$ with each $\boldsymbol{\gamma}_j\in\mathbb{R}^{pd}$. When K is fixed and $\varrho=O(1)$, the latter of which holds under suitable assumptions on the data generating process, the condition that $|\boldsymbol{\gamma}|_\infty\geqslant Cn^{-1/2}\{\log(Kpd)\}^{1/2}$ for some positive constant C, is sufficient for $\sum_{j=1}^{K}|\boldsymbol{\gamma}_j|_\infty^2\geqslant n^{-1}K\varrho\lambda^2(K,p,d,\alpha)(1+\epsilon_n)^2$. As we have discussed in Remark 3, if the time series $\{\mathbf{x}_t\}$ is strictly stationary, we know the transformed data $\{\boldsymbol{\eta}_t\}$ is also strictly stationary and the proposed test statistic T_n given in (3) essentially tests whether $\boldsymbol{\gamma}=\mathbb{E}(\boldsymbol{\eta}_t)=\mathbf{0}$ or not. As shown in Theorem 3 of Cai et al. (2014), $n^{-1/2}\{\log(Kpd)\}^{1/2}$ is the minimax optimal separation rate of any tests for the (Kpd)-dimensional mean vector hypothesis testing problem $H_0:\boldsymbol{\gamma}=\mathbf{0}$ versus $H_1:\boldsymbol{\gamma}\neq\mathbf{0}$ based on the data $\{\boldsymbol{\eta}_t\}_{t=1}^n$ if the smallest eigenvalues of $\mathrm{Var}(\boldsymbol{\eta}_t)$ are uniformly bounded away from zero. That is, for any $\alpha,\ \beta>0$ satisfying $\alpha+\beta<1$, there exists a constant $\delta_0>0$ such that $\inf_{\boldsymbol{\gamma}\in\mathcal{M}(\delta_0)}\sup_{\xi_\alpha\in\mathcal{T}_\alpha}\mathbb{P}_{H_1}(\text{reject } H_0 \text{ based on } \xi_\alpha)\leqslant 1-\beta$ for all sufficiently large n, p and d, where $\mathcal{M}(\delta_0)=\{\boldsymbol{\gamma}\in\mathbb{R}^{Kpd}:|\boldsymbol{\gamma}|_\infty\geqslant\delta_0 n^{-1/2}\{\log(Kpd)\}^{1/2}\}$, and $\mathcal{T}_\alpha$ is the set of all α-level tests for the test $H_0:\boldsymbol{\gamma}=\mathbf{0}$ versus $H_1:\boldsymbol{\gamma}\neq\mathbf{0}$. Hence, if the time series $\{\mathbf{x}_t\}$ is strictly stationary, our proposed testing procedure with fixed K will share some minimax optimal property.

4. General martingale difference hypothesis and specification testing

Our test procedure can also be extended to a more general martingale difference hypothesis, that is

$$H_0:\mathbb{E}(\mathbf{x}_t\mid\mathscr{F}_{t-1})=\boldsymbol{\mu}_x \text{ for any } t\in\mathbb{Z}\,, \tag{16}$$

where $\boldsymbol{\mu}_x\in\mathbb{R}^p$ is an unknown vector. In this scenario, we can consider the test statistic

$$T_n^{\mathrm{new}}=n\sum_{j=1}^{K}|\hat{\boldsymbol{\gamma}}_j^{\mathrm{new}}|_\infty^2\,, \tag{17}$$

where $\hat{\boldsymbol{\gamma}}_j^{\mathrm{new}}=(n-j)^{-1}\sum_{t=1}^{n-j}\mathrm{vec}\{\boldsymbol{\phi}(\mathbf{x}_t)(\mathbf{x}_{t+j}-\bar{\mathbf{x}})^\top\}$ with $\bar{\mathbf{x}}=n^{-1}\sum_{t=1}^{n}\mathbf{x}_t$. Write $\mathring{\mathbf{x}}_t=\mathbf{x}_t-\boldsymbol{\mu}_x$. In comparison to T_n given in (3), we replace $\mathbf{x}_{t+j}$ there by its mean-centered version $\mathbf{x}_{t+j}-\bar{\mathbf{x}}$ in T_n^{new}. Notice that

$$\begin{aligned}\hat{\boldsymbol{\gamma}}_j^{\mathrm{new}}=&\underbrace{\frac{1}{n-j}\sum_{t=1}^{n-j}\mathrm{vec}\bigg(\boldsymbol{\phi}(\mathbf{x}_t)\mathring{\mathbf{x}}_{t+j}^\top-\bigg[\frac{1}{n-j}\sum_{s=1}^{n-j}\mathbb{E}\{\boldsymbol{\phi}(\mathbf{x}_s)\}\bigg]\mathring{\mathbf{x}}_t^\top\bigg)}_{\mathrm{I}_j}\\&+\underbrace{\mathrm{vec}\bigg(\bigg[\frac{1}{n-j}\sum_{t=1}^{n-j}\mathbb{E}\{\boldsymbol{\phi}(\mathbf{x}_t)\}\bigg]\bigg(\frac{1}{n-j}\sum_{t=1}^{n-j}\mathring{\mathbf{x}}_t-\frac{1}{n}\sum_{t=1}^{n}\mathring{\mathbf{x}}_t\bigg)^\top\bigg)}_{\mathrm{II}_j}\\&-\underbrace{\mathrm{vec}\bigg\{\bigg(\frac{1}{n-j}\sum_{t=1}^{n-j}[\boldsymbol{\phi}(\mathbf{x}_t)-\mathbb{E}\{\boldsymbol{\phi}(\mathbf{x}_t)\}]\bigg)\bigg(\frac{1}{n}\sum_{t=1}^{n}\mathring{\mathbf{x}}_t\bigg)^\top\bigg\}}_{\mathrm{III}_j}\,.\end{aligned}$$

Since $K=o(n)$ and $j\in[K]$, I_j is the leading term of $\hat{\boldsymbol{\gamma}}_j^{\mathrm{new}}$, and II_j and III_j are the negligible terms in comparison to I_j. Define

$$\boldsymbol{\eta}_t^{\mathrm{new}}=\begin{pmatrix}\mathrm{vec}(\boldsymbol{\phi}(\mathbf{x}_t)\mathring{\mathbf{x}}_{t+1}^\top-[(n-1)^{-1}\sum_{s=1}^{n-1}\mathbb{E}\{\boldsymbol{\phi}(\mathbf{x}_s)\}]\mathring{\mathbf{x}}_t^\top)\\ \vdots\\ \mathrm{vec}(\boldsymbol{\phi}(\mathbf{x}_t)\mathring{\mathbf{x}}_{t+K}^\top-[(n-K)^{-1}\sum_{s=1}^{n-K}\mathbb{E}\{\boldsymbol{\phi}(\mathbf{x}_s)\}]\mathring{\mathbf{x}}_t^\top)\end{pmatrix}.$$

Write $\tilde{n}=n-K$. If Conditions 1 and 3 hold for $\boldsymbol{\eta}_t^{\mathrm{new}}$, together with Condition 2, we know the null distribution of T_n^{new} can be approximated by that of its Gaussian analogue $G_K^{\mathrm{new}}=\sum_{j=1}^{K}|\mathbf{g}_{\mathcal{L}_j}^{\mathrm{new}}|_\infty^2$, where $\mathcal{L}_j=\{(j-1)pd+1,\ldots,jpd\}$ and $\mathbf{g}^{\mathrm{new}}=(g_1^{\mathrm{new}},\ldots,g_{Kpd}^{\mathrm{new}})^\top\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{n,K}^{\mathrm{new}})$ with $\boldsymbol{\Sigma}_{n,K}^{\mathrm{new}}=\mathrm{Cov}(\tilde{n}^{-1/2}\sum_{t=1}^{\tilde{n}}\boldsymbol{\eta}_t^{\mathrm{new}})$. Write $\hat{\mathring{\mathbf{x}}}_t=\mathbf{x}_t-\bar{\mathbf{x}}$ and

$$\hat{\boldsymbol{\eta}}_t^{\mathrm{new}}=\begin{pmatrix}\mathrm{vec}[\boldsymbol{\phi}(\mathbf{x}_t)\hat{\mathring{\mathbf{x}}}_{t+1}^\top-\{(n-1)^{-1}\sum_{s=1}^{n-1}\boldsymbol{\phi}(\mathbf{x}_s)\}\hat{\mathring{\mathbf{x}}}_t^\top]\\ \vdots\\ \mathrm{vec}[\boldsymbol{\phi}(\mathbf{x}_t)\hat{\mathring{\mathbf{x}}}_{t+K}^\top-\{(n-K)^{-1}\sum_{s=1}^{n-K}\boldsymbol{\phi}(\mathbf{x}_s)\}\hat{\mathring{\mathbf{x}}}_t^\top]\end{pmatrix}.$$

Identical to (10), we can adopt the following estimate for $\boldsymbol{\Sigma}_{n,K}^{\text{new}}$:

$$\widehat{\boldsymbol{\Sigma}}_{n,K}^{\text{new}} = \sum_{j=-\tilde{n}+1}^{\tilde{n}-1} \mathcal{K}\bigg(\frac{j}{b_n}\bigg)\widehat{\mathbf{H}}_j^{\text{new}},$$

where $\widehat{\mathbf{H}}_j^{\text{new}} = \tilde{n}^{-1}\sum_{t=j+1}^{\tilde{n}}(\hat{\boldsymbol{\eta}}_t^{\text{new}} - \bar{\hat{\boldsymbol{\eta}}}^{\text{new}})(\hat{\boldsymbol{\eta}}_{t-j}^{\text{new}} - \bar{\hat{\boldsymbol{\eta}}}^{\text{new}})^\top$ if $j \geqslant 0$ and $\widehat{\mathbf{H}}_j^{\text{new}} = \tilde{n}^{-1}\sum_{t=-j+1}^{\tilde{n}}(\hat{\boldsymbol{\eta}}_{t+j}^{\text{new}} - \bar{\hat{\boldsymbol{\eta}}}^{\text{new}})(\hat{\boldsymbol{\eta}}_t^{\text{new}} - \bar{\hat{\boldsymbol{\eta}}}^{\text{new}})^\top$ otherwise, with $\bar{\hat{\boldsymbol{\eta}}}^{\text{new}} = \tilde{n}^{-1}\sum_{t=1}^{\tilde{n}}\hat{\boldsymbol{\eta}}_t^{\text{new}}$. Algorithm 2 states how to implement the proposed general martingale difference hypothesis test in practice.

Algorithm 2 Testing procedure for general martingale difference hypothesis

Step 1. Compute the test statistic T_n^{new} as in (17), and let $\boldsymbol{\Theta}$ be a $\tilde{n}\times\tilde{n}$ matrix with (i,j)th element $\mathcal{K}\{(i-j)/b_n\}$.

Step 2. Generate $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_{\tilde{n}})^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta})$ independent of $\mathcal{X}_n$, and let $\hat{\mathbf{g}}^{\text{new}} = \tilde{n}^{-1/2}\sum_{t=1}^{\tilde{n}}\xi_t(\hat{\boldsymbol{\eta}}_t^{\text{new}} - \bar{\hat{\boldsymbol{\eta}}}^{\text{new}})$.

Step 3. Draw $\hat{\mathbf{g}}_1^{\text{new}},\ldots,\hat{\mathbf{g}}_B^{\text{new}}$ independently by Step 2 for some large integer B.

Step 4. For given significance level $\alpha\in(0,1)$, take the $\lfloor B\alpha\rfloor$th largest value among $\hat{G}_{K,1}^{\text{new}},\ldots,\hat{G}_{K,B}^{\text{new}}$ as the critical value $\hat{\text{cv}}_\alpha$, where $\hat{G}_{K,i}^{\text{new}} = \sum_{j=1}^K |\hat{\mathbf{g}}_{i,\mathcal{L}_j}^{\text{new}}|_\infty^2$ with $\hat{\mathbf{g}}_i^{\text{new}} = (\hat{g}_{i,1}^{\text{new}},\ldots,\hat{g}_{i,Kpd}^{\text{new}})^\top$ and $\mathcal{L}_j = \{(j-1)pd+1,\ldots,jpd\}$.

Step 5. We reject H_0 defined as (16) if $T_n^{\text{new}} > \hat{\text{cv}}_\alpha$.

Below we shall provide some detailed discussion about potential extension of our test to the specification testing framework. Let $\mathbf{y}_t$ and $\mathbf{u}_t$ be observable p-dimensional and q-dimensional time series, respectively. Consider the time series model

$$\mathbf{y}_t = \mathbf{h}(\mathbf{u}_t;\boldsymbol{\theta}_0) + \mathbf{x}_t\,, \tag{18}$$

where $\mathbf{x}_t$ is the error process, and $\mathbf{h}(\cdot;\cdot)\in\mathbb{R}^p$ is a known link function with unknown truth $\boldsymbol{\theta}_0\in\mathbb{R}^m$. Without loss of generality, we assume $\mathbb{E}(\mathbf{x}_t\,|\,\mathbf{u}_t) = \mathbf{0}$. Model (18) is quite general for our analysis where we can select $\mathbf{u}_t$ as $\mathbf{y}_{t-1},\ldots,\mathbf{y}_{t-\ell}$ for some integer $\ell\geqslant 1$. For the model diagnosis, we are interested in the hypothesis testing problem:

$$H_0:\{\mathbf{x}_t\}_{t\in\mathbb{Z}} \text{ is a MDS} \quad \text{versus} \quad H_1:\{\mathbf{x}_t\}_{t\in\mathbb{Z}} \text{ is not a MDS.} \tag{19}$$

Based on the conditional moment restrictions $\mathbb{E}(\mathbf{x}_t\,|\,\mathbf{u}_t) = \mathbf{0}$, for given basis functions $\boldsymbol{\psi}(\cdot):\mathbb{R}^q\to\mathbb{R}^l$ with $pl\geqslant m$, we can identify the unknown truth $\boldsymbol{\theta}_0$ by the pl unconditional moment restrictions

$$\mathbb{E}[\{\mathbf{y}_t - \mathbf{h}(\mathbf{u}_t;\boldsymbol{\theta}_0)\}\otimes\boldsymbol{\psi}(\mathbf{u}_t)] = \mathbf{0}\,,$$

where $\otimes$ denotes the Kronecker product.

Case 1. If m is fixed or diverges slowly with the sample size n, applying the estimation procedure suggested in Chang et al. (2015), we can obtain a consistent estimator $\hat{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta}_0$ and it admits the following asymptotic expansion:

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = \frac{1}{n}\sum_{t=1}^n \mathbf{w}(\mathbf{y}_t,\mathbf{u}_t) + \text{high order term}\,, \tag{20}$$

where $\mathbf{w}(\cdot)$ is the influence function such that $\mathbb{E}\{\mathbf{w}(\mathbf{y}_t,\mathbf{u}_t)\} = \mathbf{0}$. Write $\hat{\mathbf{x}}_t = \mathbf{y}_t - \mathbf{h}(\mathbf{u}_t;\hat{\boldsymbol{\theta}}_n)$. Together with (20), it holds that

$$\hat{\mathbf{x}}_t = \mathbf{x}_t - \nabla_{\boldsymbol{\theta}}\mathbf{h}(\mathbf{u}_t;\boldsymbol{\theta}_0)\cdot\frac{1}{n}\sum_{s=1}^n \mathbf{w}(\mathbf{y}_s,\mathbf{u}_s) + \text{high order term}\,.$$

Based on obtained $\{\hat{\mathbf{x}}_t\}_{t=1}^n$, we can propose the following test statistic for (19):

$$T_n^\natural = n\sum_{j=1}^K |\boldsymbol{\gamma}_j^\natural|_\infty^2\,, \tag{21}$$

where $\boldsymbol{\gamma}_j^\natural = (n-j)^{-1}\sum_{t=1}^{n-j}\text{vec}\{\boldsymbol{\phi}(\hat{\mathbf{x}}_t)\hat{\mathbf{x}}_{t+j}^\top\}$. In comparison to the original test statistic T_n given in (3) based on observed $\{\mathbf{x}_t\}_{t=1}^n$, we replace $\mathbf{x}_t$ there by its estimate $\hat{\mathbf{x}}_t$. By Taylor expansion, under some regularity conditions, it holds that

$$\boldsymbol{\gamma}_j^\natural = \frac{1}{n-j}\sum_{t=1}^{n-j}\text{vec}\{\boldsymbol{\phi}(\mathbf{x}_t)\mathbf{x}_{t+j}^\top\} - \frac{1}{n}\sum_{t=1}^n \mathbf{A}_j\mathbf{w}(\mathbf{y}_t,\mathbf{u}_t) + \text{high order term}\,,$$

where $\mathbf{A}_j = (n-j)^{-1}\sum_{t=1}^{n-j}\mathbb{E}\{\mathbf{x}_{t+j}\otimes[\nabla_{\mathbf{x}}\boldsymbol{\phi}(\mathbf{x}_t)\nabla_{\boldsymbol{\theta}}\mathbf{h}(\mathbf{u}_t;\boldsymbol{\theta}_0)]+[\nabla_{\boldsymbol{\theta}}\mathbf{h}(\mathbf{u}_{t+j};\boldsymbol{\theta}_0)]\otimes\boldsymbol{\phi}(\mathbf{x}_t)\}$. Define

$$\boldsymbol{\eta}_t^{\natural} = \begin{pmatrix} \mathrm{vec}\{\boldsymbol{\phi}(\mathbf{x}_t)\mathbf{x}_{t+1}^{\top}\} - \mathbf{A}_1\mathbf{w}(\mathbf{y}_t,\mathbf{u}_t) \\ \vdots \\ \mathrm{vec}\{\boldsymbol{\phi}(\mathbf{x}_t)\mathbf{x}_{t+K}^{\top}\} - \mathbf{A}_K\mathbf{w}(\mathbf{y}_t,\mathbf{u}_t) \end{pmatrix}.$$

Recall $\tilde{n} = n-K$. Following the same arguments in Section 2.1, the null distribution of $T_n^{\natural}$ can be approximated by that of its Gaussian analogue $G_K^{\natural} = \sum_{j=1}^{K}|\mathbf{g}_{\mathcal{L}_j}^{\natural}|_{\infty}^2$, where $\mathcal{L}_j = \{(j-1)pd+1,\ldots,jpd\}$ and $\mathbf{g}^{\natural} = (g_1^{\natural},\ldots,g_{Kpd}^{\natural})^{\top}\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{n,K}^{\natural})$ with $\boldsymbol{\Sigma}_{n,K}^{\natural} = \mathrm{Cov}(\tilde{n}^{-1/2}\sum_{t=1}^{\tilde{n}}\boldsymbol{\eta}_t^{\natural})$. The key challenge here is to construct a valid estimate $\widehat{\boldsymbol{\Sigma}}_{n,K}^{\natural}$ satisfying $|\widehat{\boldsymbol{\Sigma}}_{n,K}^{\natural}-\boldsymbol{\Sigma}_{n,K}^{\natural}|_{\infty} = o_{\mathrm{p}}[K^{-3}\{\log(npd)\}^{-2}]$ with unknown $\mathbf{A}_1,\ldots,\mathbf{A}_K$ and unobserved $\{\mathbf{x}_t\}$.

Case 2. If $m \gg n$, we need to assume the unknown truth $\boldsymbol{\theta}_0 = (\theta_{0,1},\ldots,\theta_{0,m})^{\top}$ in (18) is sparse. Let $\mathcal{S} = \{k\in[m]:\theta_{0,k}\neq 0\}$. Using the penalized estimation procedure, for example, Chang et al. (2018c), we can obtain a sparse estimate $\hat{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta}_0$ satisfying the oracle property: (i) $\mathbb{P}(\hat{\boldsymbol{\theta}}_{n,\mathcal{S}^c}=\mathbf{0})\to 1$ as $n\to\infty$, and (ii) $\hat{\boldsymbol{\theta}}_{n,\mathcal{S}}$ follows the asymptotic expansion:

$$\hat{\boldsymbol{\theta}}_{n,\mathcal{S}} - \boldsymbol{\theta}_{0,\mathcal{S}} - \boldsymbol{\xi}_n = \frac{1}{n}\sum_{t=1}^{n}\tilde{\mathbf{w}}(\mathbf{y}_t,\mathbf{u}_t) + \text{high order term}\,, \tag{22}$$

where $\tilde{\mathbf{w}}(\cdot)$ is the influence function such that $\mathbb{E}\{\tilde{\mathbf{w}}(\mathbf{y}_t,\mathbf{u}_t)\} = \mathbf{0}$, and $\boldsymbol{\xi}_n$ is the asymptotic bias satisfying $|\boldsymbol{\xi}_n|_{\infty} = O_{\mathrm{p}}(\delta_n)$ for some $\delta_n = o(1)$ but $\delta_n \gg n^{-1/2}$. To propose the testing procedure in the setting with $m\gg n$, we need to do the next three steps first: (a) identify the index set $\mathcal{S}$, (b) estimate the asymptotic bias $\boldsymbol{\xi}_n$, (c) obtain the bias-corrected estimate $\tilde{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta}_0$ based on the estimate of $\boldsymbol{\xi}_n$. Write $\hat{\mathbf{x}}_t = \mathbf{y}_t - \mathbf{h}(\mathbf{u}_t;\tilde{\boldsymbol{\theta}}_n)$. We can still use the test statistic $T_n^{\natural}$ given in (21) in current setting. To determine the associated critical value, we only need to replace $\mathbf{w}(\cdot)$ and $\nabla_{\boldsymbol{\theta}}\mathbf{h}(\cdot;\boldsymbol{\theta}_0)$ by $\tilde{\mathbf{w}}(\cdot)$ and $\nabla_{\boldsymbol{\theta}_{\mathcal{S}}}\mathbf{h}(\cdot;\boldsymbol{\theta}_0)$, respectively, in the procedure for the setting with fixed or slowly diverging m. However, as commented in Chang et al. (2021a), if $\mathbf{h}(\cdot;\boldsymbol{\theta})$ is a nonlinear function of $\boldsymbol{\theta}$, the asymptotic bias $\boldsymbol{\xi}_n$ may include some unknown information which makes the estimation of $\boldsymbol{\xi}_n$ extremely difficult (if not impossible). How to address this problem requires further study.

5. Simulation studies

In this section, we examine the finite sample performance of our proposed test in comparison with the ones proposed by Hong et al. (2017). All tests in our simulation are implemented at the 5% significance level using 4000 Monte Carlo replications, and the number of bootstrap replications used to determine the critical value $\hat{\mathrm{cv}}_{\alpha}$ in our procedure is chosen as $B = 2000$. We set the sample size $n\in\{100,300\}$ and lags $K\in\{2,4,6,8\}$. The dimension p is set according to the ratio $p/n\in\{0.04,0.08,0.15,0.4,1.2\}$, which covers low-, moderate- and high-dimensional scenarios. Two types of maps are considered, i.e., (i) linear function ($d=p$), $\boldsymbol{\phi}(\mathbf{x}_t)=\mathbf{x}_t$; (ii) both linear and quadratic functions ($d=2p$), $\boldsymbol{\phi}(\mathbf{x}_t) = \{\mathbf{x}_t^{\top},(\mathbf{x}_t^2)^{\top}\}^{\top}$. Furthermore, we use three kernel functions for the estimation of long-run covariance matrix $\boldsymbol{\Sigma}_{n,K}$, i.e.,

(a) Quadratic Spectral (QS) kernel: $\mathcal{K}_{\mathrm{QS}}(x) = 25(12\pi^2x^2)^{-1}\{(6\pi x/5)^{-1}\sin(6\pi x/5)-\cos(6\pi x/5)\}$.
(b) Parzen (PR) kernel: $\mathcal{K}_{\mathrm{PR}}(x) = (1-6x^2+6|x|^3)I(0\leqslant|x|\leqslant 1/2)+2(1-|x|)^3I(1/2<|x|\leqslant 1)$.
(c) Bartlett (BT) kernel: $\mathcal{K}_{\mathrm{BT}}(x) = (1-|x|)I(|x|\leqslant 1)$.

Recall $\tilde{n} = n-K$. We use the data-driven bandwidth formulas developed in Andrews (1991) to determine the associated bandwidth b_n involved in these three kernel functions, that is, $b_{\mathrm{QS}} = 1.3221\{\hat{a}(2)\tilde{n}\}^{1/5}$, $b_{\mathrm{PR}} = 2.6614\{\hat{a}(2)\tilde{n}\}^{1/5}$ and $b_{\mathrm{BT}} = 1.1447\{\hat{a}(1)\tilde{n}\}^{1/3}$, where $\hat{a}(2) = \{\sum_{\ell=1}^{Kpd}4\hat{\rho}_{\ell}^2\hat{\sigma}_{\ell}^4(1-\hat{\rho}_{\ell})^{-8}\}\{\sum_{\ell=1}^{Kpd}\hat{\sigma}_{\ell}^4(1-\hat{\rho}_{\ell})^{-4}\}^{-1}$ and $\hat{a}(1) = \{\sum_{\ell=1}^{Kpd}4\hat{\rho}_{\ell}^2\hat{\sigma}_{\ell}^4(1-\hat{\rho}_{\ell})^{-6}(1+\hat{\rho}_{\ell})^{-2}\}\{\sum_{\ell=1}^{Kpd}\hat{\sigma}_{\ell}^4(1-\hat{\rho}_{\ell})^{-4}\}^{-1}$, with $\hat{\rho}_{\ell}$ and $\hat{\sigma}_{\ell}^2$ being, respectively, the estimated autoregressive coefficient and innovation variance from fitting an AR(1) model to time series $\{\eta_{t,\ell}\}_{t=1}^{\tilde{n}}$, the ℓth component sequence of $\{\boldsymbol{\eta}_t\}_{t=1}^{\tilde{n}}$ defined in (5). Denote the test statistics based on the three kernels with linear map by T_{QS}^{l}, T_{PR}^{l} and T_{BT}^{l}, respectively, and denote the ones with both linear and quadratic map by T_{QS}^{q}, T_{PR}^{q} and T_{BT}^{q}, respectively. Note that the data-driven formulas by Andrews (1991) are based on AR(1) model assumption and also deliver an estimation-optimal bandwidth in the low-dimensional setting. Here we apply it to determine the associated bandwidth b_n in both moderate- and high-dimensional settings since there are no other known formulas and the numerical studies in Chang et al. (2017a) show such formula seems to work well when the dimension is large. We also include three tests proposed by Hong et al. (2017) in our simulation comparison, i.e., the trace-based test Z_{tr}, the determinant-based test $Z_{\det}$, and the large-dimensional test Zd_{tr}. Note that Hong et al. (2017) only examined the finite sample performance of Z_{tr} and $Z_{\det}$, which cannot be implemented when $p>\sqrt{n}$, whereas Zd_{tr} is shown to be valid under the assumption $p/n\to 0$ and its implementation becomes infeasible when $p>n$. The tests of Hong et al. (2017) require the matrix normalization which is computationally prohibitive in the high-dimensional setting. See Section S.4 in the supplementary material for the comparison of computational cost between our test and the tests of Hong et al. (2017).

5.1. Empirical size

To examine the empirical size, we consider the following models:

Model 1. i.i.d. normal sequence: $\mathbf{x}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{A})$ where $\mathbf{A} = (a_{kl})_{p\times p}$ with $a_{kl} = 0.995^{|k-l|}$ for any $k, l \in [p]$.

Model 2. Stochastic volatility model: $\mathbf{x}_t = \boldsymbol{\varepsilon}_t \exp(\boldsymbol{\sigma}_t)$ with $\boldsymbol{\sigma}_t = 0.25\boldsymbol{\sigma}_{t-1} + 0.05\mathbf{u}_t$, $\boldsymbol{\varepsilon}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_\varepsilon)$ and $\mathbf{u}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_u)$, where $\boldsymbol{\Omega}_\varepsilon = (\omega_{\varepsilon,kl})_{p\times p}$ and $\boldsymbol{\Omega}_u = (\omega_{u,kl})_{p\times p}$ with $\omega_{\varepsilon,kl} = I(k=l) + 0.4I(k \neq l)$ and $\omega_{u,kl} = 0.9^{|k-l|}$ for any $k, l \in [p]$.

Model 3. Bivariate constant conditional correlation GARCH(1,1) model: $\mathbf{x}_t = \mathbf{b}_t^{1/2} \circ \boldsymbol{\varepsilon}_t$ with $\mathbf{b}_t = \mathbf{a}_0 + \mathbf{A}_1\mathbf{b}_{t-1} + \mathbf{A}_2\mathbf{x}_{t-1}^2$ and $\boldsymbol{\varepsilon}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_\varepsilon)$, where $\circ$ denotes the Hadamard product, $\mathbf{a}_0 = (0.2, 0.1\,\mathbf{1}_{p-1}^\top)^\top$, $\mathbf{A}_1 = 0.9\,\mathbf{I}_p$, $\mathbf{A}_2 = \mathrm{diag}(0.05, 0.08, 0.03\,\mathbf{1}_{p-2}^\top)$, and $\boldsymbol{\Omega}_\varepsilon = (\omega_{\varepsilon,kl})_{p\times p}$ with $\omega_{\varepsilon,kl} = I(k=l) + 0.5I(k\neq l)$ for any $k, l \in [p]$. Here $\mathbf{1}_q$ and $\mathbf{I}_q$ denote, respectively, the q-dimensional vector with all components being 1 and q-dimensional identity matrix for any given integer q.

A few comments are in order. Model 1 was used by Chang et al. (2017a) in their simulation for high-dimensional white noise testing problem. Model 2 is the multivariate extension of the univariate stochastic volatility model considered in Escanciano and Velasco (2006) for the univariate martingale difference hypothesis testing problem. Model 3 is motivated from Hong et al. (2017), which reduces to the bivariate GARCH model considered in Hong et al. (2017) when $p = 2$.

As seen from Table 1, our tests have quite accurate size when the dimension p is low for all models. For a fixed sample size n, the rejection rates tend to decrease as the dimension p increases, showing the impact on the bootstrap-based approximation from the dimension p. For a fixed dimension p, enlarging sample size from $n = 100$ to $n = 300$ helps to bring down the size distortion to some extent for most kernels and maps, e.g., the empirical sizes for Models 1–3 are undersized when $n = 100$ and $p/n = 1.2$ ($p = 120$), and the empirical sizes increase and become much closer to the 5% nominal level when $n = 300$ and $p/n = 0.4$ ($p = 120$). Overall our tests show reasonably good size control and the undersize phenomenon for the moderate- and high-dimensional scenarios could be due to the bandwidth choice, which is always a difficult issue in practice. The three tests of Hong et al. (2017) also show quite accurate size for Models 1 and 2, and there is some noticeable over-rejection for Model 3 when $n = 100$. When $n = 300$ and $p/n = 0.4$, we are unable to implement the test Zd_{tr} even though $p < n$. The reason is that the computation of Zd_{tr} requires to store five $120^2 \times 120^2$ matrices, and product of three $120^2 \times 120^2$ matrices during the calculation, which results in running out of the memory (RAM: 8158 MB). This indicates the difficulty of implementing their tests for $p = 120$ and beyond.

In order to investigate the influence of the data-driven bandwidth used in our simulation, we examine the sensitivity of our size and power results by replacing the data-driven bandwidth b_n by its scaled version $c \cdot b_n$ with $c \in \{2^{-3}, 2^{-2}, 2^{-1}, 2^1, 2^2, 2^3\}$. Simulation results for Bartlett kernel are displayed in Tables 2 and 4. Simulation results for Quadratic Spectral kernel and Parzen kernel are reported in the supplementary material. For different multiples c, the sizes and powers are relatively robust. In addition, we find that the results for $c < 1$ perform a little better than these for $c > 1$ in general, but not by much. Therefore, the choice of $c = 1$ in our simulation is reasonable.

5.2. Empirical power

To study the empirical power of the proposed method, we consider the following models:

Model 4. First-order exponential autoregressive model: $\mathbf{x}_t = 0.15\mathbf{x}_{t-1} + \exp(-2\mathbf{x}_{t-1}^2) + \boldsymbol{\varepsilon}_t$ with $\boldsymbol{\varepsilon}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_\varepsilon)$, where $\boldsymbol{\Omega}_\varepsilon = (\omega_{\varepsilon,kl})_{p\times p}$ with $\omega_{\varepsilon,kl} = I(k=l) + 0.25I(k\neq l)$ for any $k, l \in [p]$.

Model 5. The sum of a white noise and cosine of the first difference of an autoregressive process: $\mathbf{x}_t = \boldsymbol{\varepsilon}_t + 0.8\cos(\mathbf{z}_t - \mathbf{z}_{t-1})$ with $\mathbf{z}_t = 0.85\mathbf{z}_{t-1} + \mathbf{u}_t$, $\boldsymbol{\varepsilon}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_\varepsilon)$ and $\mathbf{u}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_u)$, where $\boldsymbol{\Omega}_\varepsilon = (\omega_{\varepsilon,kl})_{p\times p}$ and $\boldsymbol{\Omega}_u = (\omega_{u,kl})_{p\times p}$ with $\omega_{\varepsilon,kl} = I(k=l) + 0.3I(k\neq l)$ and $\omega_{u,kl} = 0.7^{|k-l|}$ for any $k, l \in [p]$.

Model 6. Threshold autoregressive model of order one: $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,p})^\top$ with $x_{t,j} = -0.45x_{t-1,j}I(x_{t-1,j} \geqslant 1) + 0.6x_{t-1,j}I(x_{t-1,j} < 1) + \varepsilon_{t,j}$ for each $j \in [p]$, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{t,1}, \ldots, \varepsilon_{t,p})^\top \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$.

Models 4–6 can be regarded as multivariate extensions of the univariate models considered in Escanciano and Velasco (2006) (see Models 7–9 there). Table 3 shows that for Models 4–6, the powers based on three different kernels are similar for the same map with the use of Bartlett kernel exhibiting slightly more power in most cases. When $n = 100$ and for Model 4, using the linear and quadratic map leads to more power when $p/n \leqslant 0.15$, but less power when $p/n > 0.15$. This can be explained by the impact from the high dimension. The additional nonlinear serial dependence captured by the quadratic map is apparent when $p \leqslant 15$, but as the dimension p increases to 120, the signal related to nonlinear dependence is likely dominated by that related to linear dependence and possibly the noise, so using linear map alone yields more power. Similar phenomena occur for Models 5 and 6. As expected, when we increase the sample size n from 100 to 300, we see the appreciation of the power as both linear and nonlinear serial dependence get strengthened at the

Table 1
Empirical sizes (%) of the tests T^l_{QS}, T^l_{PR}, T^l_{BT}, T^q_{QS}, T^q_{PR}, T^q_{BT}, Z_{tr}, Z_{det} and Zd_{tr} for Models 1–3 at the 5% nominal level.

			Model 1									Model 2									Model 3								
n	p/n	K	T^l_{QS}	T^l_{PR}	T^l_{BT}	T^q_{QS}	T^q_{PR}	T^q_{BT}	Z_{tr}	Z_{det}	Zd_{tr}	T^l_{QS}	T^l_{PR}	T^l_{BT}	T^q_{QS}	T^q_{PR}	T^q_{BT}	Z_{tr}	Z_{det}	Zd_{tr}	T^l_{QS}	T^l_{PR}	T^l_{BT}	T^q_{QS}	T^q_{PR}	T^q_{BT}	Z_{tr}	Z_{det}	Zd_{tr}
100	0.04	2	4.2	4.5	4.5	4.3	4.3	4.4	5.2	4.4	5.5	4.2	4.4	4.7	2.2	2.4	2.5	5.2	4.7	4.8	3.5	3.6	4.2	2.9	2.8	3.2	6.7	6.3	5.2
		4	5.1	5.0	5.2	4.6	4.3	4.5	5.2	5.2	5.4	3.1	3.3	3.5	2.9	2.8	3.2	4.3	4.9	6.3	3.2	3.2	3.5	3.0	3.1	3.4	6.9	6.9	6.3
		6	4.6	4.4	4.8	4.5	4.5	4.6	5.0	4.7	6.0	3.0	2.9	3.5	2.8	2.7	2.9	5.4	4.4	6.0	3.1	3.1	3.6	3.9	4.0	4.1	6.5	6.8	6.1
		8	4.4	4.4	4.7	5.0	4.9	5.1	5.6	4.9	6.2	2.8	2.8	3.3	2.9	2.8	3.1	5.4	4.5	6.5	3.3	3.3	4.0	4.5	4.4	4.9	6.7	7.9	6.9
	0.08	2	4.1	4.1	4.1	3.1	3.2	3.4	4.8	4.9	5.2	3.8	3.9	4.0	1.9	1.8	1.8	4.4	4.9	5.2	3.3	3.2	3.5	2.7	2.5	2.7	5.9	8.0	5.2
		4	3.9	3.8	4.0	4.4	4.3	4.7	6.0	4.8	5.2	3.0	3.0	3.5	1.8	1.9	2.1	5.4	5.1	6.0	2.6	2.5	3.0	2.7	2.7	2.7	7.1	8.7	5.2
		6	4.6	4.4	4.6	4.4	4.4	4.7	6.7	6.3	6.2	2.2	2.1	2.5	2.2	2.2	2.4	5.5	5.9	5.6	2.2	2.3	2.8	4.4	4.2	4.5	7.5	8.2	7.1
		8	4.8	4.9	5.2	4.0	3.9	4.2	7.4	5.7	5.4	2.3	2.0	3.0	2.1	2.2	2.4	7.2	5.5	6.9	2.8	2.8	3.3	4.4	4.4	4.9	8.2	8.9	7.6
	0.15	2	4.2	4.4	4.4	4.0	3.8	4.2	NA	NA	4.6	3.4	3.4	3.7	1.9	1.9	2.3	NA	NA	4.8	3.4	3.4	4.1	1.9	1.9	1.9	NA	NA	5.1
		4	4.3	4.2	4.6	3.2	3.2	3.5	NA	NA	4.3	2.5	2.5	2.7	1.7	1.8	2.1	NA	NA	5.9	2.2	2.2	2.7	2.5	2.4	2.7	NA	NA	5.8
		6	4.2	4.0	4.4	3.8	3.8	4.0	NA	NA	5.2	2.7	2.9	3.3	1.7	1.6	1.9	NA	NA	5.7	2.2	2.1	2.8	2.8	2.7	3.0	NA	NA	7.4
		8	4.0	4.1	4.5	4.4	4.5	4.8	NA	NA	6.3	2.7	2.6	3.1	2.1	2.2	2.4	NA	NA	6.4	1.8	1.8	2.6	3.4	3.6	3.6	NA	NA	8.3
	0.40	2	3.8	4.0	4.1	2.1	2.3	2.6	NA	NA	4.8	3.4	3.5	3.9	2.0	2.3	2.5	NA	NA	4.7	2.7	2.5	3.0	1.8	1.8	1.9	NA	NA	5.5
		4	2.8	2.9	3.2	2.5	2.6	2.6	NA	NA	5.3	3.2	3.2	3.6	2.1	2.1	2.3	NA	NA	5.4	1.7	1.7	2.1	2.0	2.2	1.9	NA	NA	6.6
		6	2.9	3.0	3.4	2.8	2.8	3.2	NA	NA	5.6	2.6	2.6	3.1	2.1	2.0	2.1	NA	NA	6.0	1.1	1.1	1.7	2.4	2.6	2.3	NA	NA	7.9
		8	3.2	3.0	3.4	3.2	3.1	3.4	NA	NA	5.8	2.8	2.8	3.2	2.6	2.5	2.8	NA	NA	5.2	1.5	1.4	2.2	3.3	3.6	3.4	NA	NA	9.2
	1.20	2	2.2	2.3	2.5	1.1	1.2	1.3	NA	NA	NA	3.8	3.8	4.3	2.6	2.8	2.7	NA	NA	NA	1.6	1.6	2.0	2.9	3.3	2.5	NA	NA	NA
		4	2.0	2.1	2.6	1.1	1.2	1.2	NA	NA	NA	2.9	3.1	3.2	2.5	2.5	2.7	NA	NA	NA	1.1	1.1	1.8	3.9	4.1	3.1	NA	NA	NA
		6	2.0	2.1	2.4	1.1	1.1	1.3	NA	NA	NA	3.3	3.3	3.9	2.3	2.4	2.4	NA	NA	NA	1.1	1.1	1.5	4.9	5.3	3.9	NA	NA	NA
		8	1.4	1.7	2.0	1.2	1.2	1.4	NA	NA	NA	3.1	3.2	3.8	2.9	3.1	3.2	NA	NA	NA	1.1	1.0	1.5	6.3	6.7	5.5	NA	NA	NA
300	0.04	2	5.6	5.5	5.8	4.2	4.1	4.4	5.1	5.5	5.8	4.1	4.1	4.2	3.8	3.7	3.9	4.9	5.6	4.7	4.0	4.0	4.0	3.6	3.4	3.8	6.2	7.5	5.2
		4	3.9	4.2	4.5	4.7	4.6	5.0	5.9	5.4	5.0	3.7	3.9	4.2	2.9	2.8	3.2	6.1	5.9	5.6	3.8	3.7	4.1	3.3	3.4	3.9	6.2	6.6	5.5
		6	4.2	4.1	4.2	5.5	5.2	5.5	6.4	6.7	5.6	3.9	3.6	3.9	3.9	4.0	4.2	6.6	6.8	6.4	3.7	3.7	4.1	4.7	4.7	4.9	7.3	7.9	5.6
		8	4.7	4.8	5.0	6.0	6.0	6.3	7.1	6.9	5.8	3.7	3.8	4.0	4.1	4.0	4.3	7.1	6.4	5.1	3.2	3.0	3.4	4.4	4.4	4.8	8.6	8.0	6.3
	0.08	2	4.8	4.8	5.0	4.0	4.0	4.1	NA	NA	5.5	4.2	4.3	4.4	3.5	3.6	3.8	NA	NA	4.8	3.6	3.5	3.8	3.2	3.2	3.2	NA	NA	5.6
		4	3.8	3.8	3.9	4.1	4.0	4.2	NA	NA	5.2	3.8	3.5	3.8	3.7	3.6	3.9	NA	NA	5.0	3.7	3.5	4.0	3.2	3.0	3.4	NA	NA	5.4
		6	4.6	4.4	5.0	5.0	5.0	5.4	NA	NA	5.0	3.6	3.3	3.8	4.1	4.2	4.4	NA	NA	5.4	3.3	3.2	3.7	3.4	3.3	3.7	NA	NA	5.3
		8	3.9	4.2	4.3	5.7	5.9	6.1	NA	NA	5.4	3.7	3.6	4.1	3.8	3.7	4.0	NA	NA	5.4	3.0	3.0	3.4	3.6	3.7	4.1	NA	NA	5.9
	0.15	2	4.7	4.6	4.8	3.8	3.9	4.3	NA	NA	6.0	4.4	4.4	4.7	3.7	3.6	3.9	NA	NA	4.6	3.9	4.0	4.2	3.1	2.9	3.3	NA	NA	5.0
		4	4.4	4.4	4.6	4.4	4.3	4.6	NA	NA	4.6	3.7	3.9	3.9	4.2	4.2	4.4	NA	NA	5.1	3.2	3.2	3.6	3.2	3.0	3.3	NA	NA	5.3
		6	3.9	3.9	4.1	4.6	4.4	4.8	NA	NA	5.1	3.5	3.5	3.8	3.4	3.7	3.8	NA	NA	5.3	3.2	3.0	3.4	3.6	3.5	3.8	NA	NA	6.8
		8	3.9	4.0	4.2	4.4	4.4	4.7	NA	NA	5.6	3.5	3.5	3.7	4.2	4.3	4.4	NA	NA	5.6	3.0	3.0	3.4	3.6	3.4	4.1	NA	NA	5.9
	0.40	2	4.2	4.2	4.3	2.6	2.6	2.8	NA	NA	NA	4.5	4.6	4.8	3.1	3.0	3.4	NA	NA	NA	3.8	3.8	4.1	2.8	2.7	3.1	NA	NA	NA
		4	3.5	3.5	3.6	3.2	3.2	3.5	NA	NA	NA	4.2	4.3	4.5	3.8	3.7	4.0	NA	NA	NA	3.1	3.1	3.5	2.7	2.6	3.0	NA	NA	NA
		6	3.7	3.9	4.2	4.1	4.1	4.7	NA	NA	NA	4.1	4.0	4.4	3.9	4.0	4.2	NA	NA	NA	2.7	2.6	3.0	2.4	2.3	2.7	NA	NA	NA
		8	3.2	3.3	3.8	4.1	4.0	4.6	NA	NA	NA	4.1	4.1	4.2	4.2	4.2	4.4	NA	NA	NA	2.3	2.3	2.8	2.8	2.9	3.1	NA	NA	NA
	1.20	2	3.1	3.0	3.4	1.8	1.7	2.0	NA	NA	NA	4.0	4.2	4.2	3.9	3.9	4.0	NA	NA	NA	3.8	3.8	4.0	2.4	2.4	2.8	NA	NA	NA
		4	2.3	2.2	2.4	1.8	1.8	2.0	NA	NA	NA	3.8	3.9	4.0	3.9	3.9	4.2	NA	NA	NA	2.7	2.7	3.1	1.8	1.9	2.0	NA	NA	NA
		6	1.3	1.2	1.7	1.7	1.7	1.9	NA	NA	NA	3.8	3.5	3.9	4.2	4.4	4.7	NA	NA	NA	2.1	2.2	2.6	2.1	2.1	2.2	NA	NA	NA
		8	1.1	1.3	1.8	1.7	1.8	2.1	NA	NA	NA	4.2	4.4	4.6	4.0	3.9	4.2	NA	NA	NA	1.8	1.7	2.4	2.0	2.0	2.5	NA	NA	NA

sample level. Overall, the powers of our tests are quite encouraging for the three models, and all combinations of kernel and map under consideration.

By contrast, the three tests of Hong et al. (2017) mostly fail to reject the martingale difference hypothesis for Models 4 and 5 in all settings. This is presumably due to the inability of their tests to capture nonlinear serial dependence. For Model 6, their tests exhibit great power, which is probably due to the fact that the model implies strong linear serial dependence although it is a nonlinear model per se. Indeed, the sample ACF at lag 1, 2, 3 are 0.324, 0.120 and 0.046, respectively, based on our simulation. Again their tests cannot be implemented when p is too large relative to n, as their ability of handling the high dimension is quite limited.

5.3. Power curve

In this subsection, we perturb Models 1–3 so that the new sequence is not a MDS and present power curves. For given constant $a \in \{0, 0.5, 1, 1.5, 2, 2.5\}$, the model settings are as follows:

Model 1′. Let $\mathbf{x}_t$ follow Model 1 and $\mathbf{y}_t = \mathbf{x}_t + a\exp(-2\mathbf{x}^2_{t-1})$.
Model 2′. Let $\mathbf{x}_t$ follow Model 2 and $\mathbf{y}_t = \mathbf{x}_t + a\cos(\boldsymbol{\varepsilon}_{t-1} \circ \boldsymbol{\sigma}_{t-1})$, where $\boldsymbol{\varepsilon}_{t-1}$ and $\boldsymbol{\sigma}_{t-1}$ are specified in Model 2.
Model 3′. Let $\mathbf{x}_t$ follow Model 3 and $\mathbf{y}_t = \mathbf{x}_t + a\log(\mathbf{x}^2_{t-2})$.

We aim to test whether $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ defined in Models 1′–3′ is a MDS. When $a = 0$, $\mathbf{y}_t = \mathbf{x}_t$ and Models 1′–3′ become Models 1–3, respectively, which follow the null hypothesis. Figs. 1–3 display the empirical sizes and powers of our proposed tests $(T^l_{\mathrm{BT}}, T^q_{\mathrm{BT}})$ and Hong et al. (2017)'s test (Zd_{tr}) when the sample size $n = 100$. Notice that Zd_{tr} is feasible when $p < n$. Thus when $p/n = 1.2$, there is no power curve for Zd_{tr}. As seen from Fig. 1, our tests and Hong et al. (2017)'s test control the empirical sizes well under the null hypothesis with $a = 0$ and the empirical powers increase for larger values of the distance parameter a. But our tests outperform Hong et al. (2017)'s test especially for large K. In Fig. 2, Hong et al. (2017)'s test almost cannot detect the alternative hypotheses, but our tests still work well. This is presumably due to the inability of their test to capture nonlinear serial dependence. Based on Fig. 3, similar phenomenon is observed that the empirical powers increase as the distance a grows. Somewhat counter-intuitively, the empirical powers of Zd_{tr} decrease when a increases from 2 to 2.5, which means the power is non-monotonic. In addition, comparing the results of our tests for two maps, we find that the test based on linear and quadratic map is more powerful than the test only based on

Table 2
Empirical sizes (%) of the tests T_{BT}^{l} and T_{BT}^{q} for Models 1–3 at the 5% nominal level, where c represents the constant which is multiplied by Andrews' bandwidth.

			Model 1 with T_{BT}^{l}							Model 1 with T_{BT}^{q}						
			c							c						
n	p/n	K	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}
100	0.04	2	4.6	4.9	4.7	4.3	5.3	6.1	9.4	4.2	4.3	4.2	4.0	4.3	4.5	7.9
		4	5.1	5.0	4.8	4.5	4.3	5.3	6.6	5.0	4.2	4.4	4.6	3.8	4.8	6.4
		6	5.1	5.0	4.1	4.8	4.5	4.1	6.0	5.3	5.0	4.9	5.1	5.1	4.5	5.8
		8	5.7	5.3	4.9	5.4	4.5	4.1	4.0	5.3	5.6	5.0	6.1	4.8	5.5	5.1
	0.08	2	4.9	4.2	4.3	4.6	4.9	5.9	9.2	4.2	4.3	4.1	4.2	4.3	5.2	7.2
		4	5.4	4.8	5.1	4.2	3.7	4.6	5.0	4.5	4.0	4.4	4.6	3.8	3.9	5.7
		6	4.7	4.7	5.1	4.1	3.7	4.5	4.7	4.9	5.0	5.5	4.6	4.5	3.8	5.3
		8	4.8	5.2	4.6	4.0	4.5	3.3	4.3	4.8	5.0	5.3	5.2	5.0	4.2	4.4
	0.15	2	5.3	4.5	4.4	4.5	4.5	5.6	8.9	4.2	3.2	3.3	4.0	3.7	4.1	6.1
		4	5.0	4.7	4.3	3.9	3.7	3.8	5.6	4.3	4.0	4.0	3.8	3.2	3.5	4.6
		6	4.0	4.5	4.7	3.9	4.3	3.6	3.6	3.7	4.1	3.9	4.1	3.9	3.9	3.8
		8	4.8	4.7	4.5	3.9	3.9	3.0	3.1	4.5	4.9	4.5	4.8	4.3	3.7	4.1
	0.40	2	4.6	4.1	4.2	4.2	4.7	4.5	6.4	3.1	3.1	2.7	2.6	2.7	2.9	5.1
		4	3.5	3.6	3.8	3.2	3.4	2.6	3.5	2.7	2.9	3.1	2.9	3.0	2.7	3.1
		6	4.0	4.2	4.1	3.8	2.8	2.2	2.8	3.0	3.7	2.9	3.2	2.2	2.3	3.2
		8	3.8	3.7	4.0	3.8	3.1	2.3	1.7	3.5	3.2	3.3	3.5	3.0	2.8	2.5
	1.20	2	3.9	4.4	3.1	3.1	2.9	3.1	4.1	1.8	1.8	1.4	1.4	1.1	1.4	2.2
		4	2.9	3.0	2.5	2.4	1.4	1.3	2.1	1.8	1.7	1.6	1.3	1.2	1.2	1.6
		6	2.8	3.0	2.3	2.2	1.6	0.9	0.9	1.8	1.8	1.8	1.7	1.2	1.4	1.6
		8	2.9	2.8	2.4	2.3	1.4	0.8	0.5	1.9	1.9	2.0	1.7	1.4	1.4	1.1
300	0.04	2	5.2	4.8	5.2	4.7	4.9	4.6	5.6	4.7	4.6	5.0	4.4	4.8	4.8	5.6
		4	4.5	4.6	5.2	4.3	4.8	4.6	4.6	5.6	5.4	4.3	4.9	4.8	4.6	4.1
		6	4.9	4.6	4.3	4.7	4.6	3.4	3.6	5.7	5.7	5.1	5.0	4.3	4.5	3.9
		8	5.1	5.0	5.6	4.3	3.7	3.5	3.2	6.8	6.5	6.3	5.6	5.4	5.4	4.7
	0.08	2	4.8	4.7	4.1	4.9	4.2	4.6	5.1	4.4	4.8	4.4	4.3	4.5	4.5	4.3
		4	5.2	4.6	4.6	4.1	4.3	3.7	4.3	5.1	5.1	5.0	4.8	4.4	3.8	3.7
		6	4.3	4.4	5.1	4.6	3.9	3.5	2.7	6.2	5.8	5.6	5.6	5.1	4.4	3.1
		8	5.1	4.8	5.0	4.1	3.8	2.7	2.0	6.2	5.7	5.3	5.4	4.5	4.7	4.0
	0.15	2	4.3	4.3	5.4	5.0	4.4	5.1	5.0	4.5	4.2	4.0	3.6	3.6	3.8	4.3
		4	4.6	4.3	4.4	4.5	3.3	3.1	3.4	4.4	4.4	4.9	4.2	4.3	3.5	3.5
		6	5.2	4.3	4.1	4.6	3.7	2.9	2.9	5.1	5.0	5.0	5.0	4.0	3.5	2.8
		8	4.3	4.1	4.2	3.7	2.7	2.5	2.2	5.3	5.0	5.7	5.5	4.9	3.5	3.1
	0.40	2	4.2	4.2	4.7	3.8	3.7	4.2	3.5	3.8	3.8	2.9	3.2	3.1	2.7	3.0
		4	3.8	3.8	3.9	3.9	3.4	3.0	1.7	4.2	4.6	3.2	3.2	2.7	2.4	2.1
		6	3.8	4.1	3.1	3.4	2.9	2.1	1.4	3.7	4.1	4.4	3.3	2.9	2.9	1.7
		8	3.8	3.2	3.8	3.4	2.6	1.5	0.7	3.7	4.0	4.3	3.8	3.5	2.6	1.9
	1.20	2	3.0	3.6	3.4	3.4	3.2	2.4	1.9	2.0	1.9	1.8	1.8	1.3	1.5	1.2
		4	3.6	2.9	2.9	3.5	2.2	1.2	0.8	2.0	1.9	2.2	1.6	1.4	1.0	0.6
		6	2.6	2.4	2.6	2.1	1.5	0.6	0.2	2.3	2.3	1.8	1.6	1.4	0.9	0.7
		8	2.1	2.7	2.6	1.6	1.2	0.6	0.1	2.2	2.0	2.1	2.2	1.9	0.8	0.7

			Model 2 with T_{BT}^{l}							Model 2 with T_{BT}^{q}						
			c							c						
n	p/n	K	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}
100	0.04	2	4.2	4.4	4.1	4.0	4.0	4.1	5.5	3.0	3.0	2.8	2.5	2.4	2.8	4.3
		4	4.3	4.3	3.4	3.8	2.9	2.4	3.4	2.8	3.5	3.1	3.4	2.3	2.2	2.7
		6	3.8	4.0	3.2	2.8	2.0	2.1	2.1	3.1	3.2	3.5	3.0	2.9	2.3	2.7
		8	3.5	3.6	4.1	3.2	2.3	2.0	1.5	3.7	3.3	3.5	3.7	3.3	3.1	3.2
	0.08	2	4.3	4.0	3.4	4.1	3.5	4.1	5.7	2.5	2.6	2.2	2.2	2.3	2.2	3.7
		4	3.9	3.5	3.5	3.0	2.3	2.1	2.5	2.8	2.7	1.8	2.0	2.2	1.9	2.4
		6	3.3	3.6	3.7	3.1	2.7	1.3	2.1	2.5	2.5	2.6	2.2	2.2	1.7	2.2
		8	3.2	3.0	3.4	2.6	2.1	1.5	0.8	2.5	2.9	2.7	2.1	2.0	2.3	2.1
	0.15	2	4.3	4.4	4.3	3.3	3.9	4.8	5.1	2.5	2.2	2.1	2.4	2.2	2.0	2.7
		4	3.8	3.9	3.6	3.0	2.4	2.6	2.7	2.4	1.9	2.2	1.9	2.0	1.9	2.2
		6	3.2	3.6	3.3	2.9	2.3	1.4	1.6	2.3	2.0	2.2	1.6	2.6	1.8	2.0
		8	3.3	3.5	3.1	3.0	2.0	1.0	0.9	2.4	2.9	2.4	2.1	2.0	1.8	2.0
	0.40	2	4.1	3.8	4.0	3.6	3.8	4.0	5.9	2.4	2.6	2.2	2.6	2.0	2.9	3.8
		4	4.0	3.5	4.1	3.6	2.9	2.5	3.3	2.5	2.7	2.5	2.3	2.2	2.1	2.5
		6	3.9	4.0	3.2	3.4	2.8	2.0	2.2	2.5	2.9	2.4	2.4	2.1	2.0	2.2
		8	3.5	3.9	3.7	3.2	2.1	1.9	1.4	2.8	2.9	3.2	2.1	2.1	2.5	2.5
	1.20	2	4.5	4.0	3.9	4.1	4.3	5.1	6.5	3.1	2.4	2.7	2.6	2.5	2.9	4.7
		4	4.5	3.7	4.2	3.6	3.8	4.0	4.1	2.7	2.8	3.0	2.8	2.4	2.4	3.3
		6	3.8	4.2	3.8	3.2	3.0	3.0	3.2	2.7	3.3	2.9	3.1	2.7	2.2	2.7
		8	4.0	4.1	4.1	3.7	2.9	2.4	2.4	3.2	3.2	3.2	2.9	3.2	2.7	2.8
300	0.04	2	4.4	4.9	4.0	4.0	3.8	4.4	4.3	3.7	3.9	3.5	3.6	3.2	3.7	3.4
		4	4.2	4.0	3.8	3.7	3.3	3.2	3.0	4.2	3.6	4.6	4.0	3.4	2.9	2.9
		6	4.9	3.5	4.0	4.0	3.3	2.3	1.8	3.9	4.1	4.7	4.2	3.9	3.8	2.4
		8	4.4	4.5	4.1	3.5	3.3	1.7	1.7	5.3	4.6	4.4	4.4	3.9	3.1	3.4
	0.08	2	4.6	4.9	4.9	4.4	4.5	4.4	4.8	4.0	4.1	3.8	3.1	3.3	3.3	3.7
		4	4.2	4.1	3.9	4.4	3.7	3.1	3.1	4.8	3.9	4.4	4.2	3.6	3.1	2.9
		6	4.7	4.2	4.1	4.2	3.0	2.7	2.4	4.9	4.8	4.2	4.3	3.5	3.2	2.7
		8	4.0	4.3	4.4	3.6	3.4	2.1	1.6	4.8	4.4	4.5	3.6	4.5	3.3	3.1
	0.15	2	5.0	4.0	4.4	4.4	4.6	4.8	4.7	3.7	4.2	3.8	3.4	3.2	3.6	3.8
		4	4.8	4.8	4.7	4.2	3.6	4.3	3.6	4.4	4.1	3.7	3.9	3.1	2.8	2.9
		6	4.6	4.8	4.8	3.9	3.3	2.9	2.8	5.5	4.3	4.6	3.6	3.9	3.1	3.1
		8	3.8	4.2	4.3	4.8	3.4	2.7	1.7	4.3	4.5	4.7	4.5	3.9	3.7	3.0
	0.40	2	5.2	4.5	5.1	4.8	5.0	4.9	5.3	3.5	4.2	4.4	3.5	3.3	3.6	3.6
		4	4.9	4.4	4.3	4.5	4.6	3.7	3.7	4.9	4.6	4.6	4.2	4.1	3.4	2.6
		6	4.6	4.5	4.2	4.4	4.0	3.2	3.1	4.9	4.1	4.4	4.2	4.1	3.2	3.4
		8	4.2	4.7	4.3	3.7	3.5	2.8	2.4	5.8	5.1	4.6	4.4	4.2	4.0	3.6
	1.20	2	4.6	5.4	4.9	5.2	4.8	4.4	5.7	4.6	4.4	4.6	4.3	4.0	4.2	4.5
		4	4.6	4.8	4.5	4.1	3.8	3.7	3.7	4.1	4.3	4.2	4.1	3.4	3.8	3.8
		6	4.7	4.1	4.4	4.4	3.6	3.5	3.5	5.4	4.4	4.4	4.3	4.2	3.9	3.7
		8	4.5	3.6	4.0	4.5	4.3	3.6	2.5	5.1	4.5	5.0	4.5	4.6	3.8	3.7

			Model 3 with T_{BT}^{l}							Model 3 with T_{BT}^{q}						
			c							c						
n	p/n	K	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}
100	0.04	2	4.1	4.0	3.6	3.6	3.7	3.9	5.0	3.9	3.7	2.8	2.8	2.9	3.0	3.9
		4	4.4	4.4	3.8	3.5	2.6	2.5	2.7	3.9	4.0	4.0	3.2	2.9	3.0	3.4
		6	4.2	4.3	4.0	3.8	2.8	1.9	2.2	4.4	4.6	4.5	4.2	3.7	3.6	3.6
		8	3.8	4.1	4.7	3.9	2.7	1.9	1.5	6.0	5.8	5.3	5.2	4.0	4.3	4.2
	0.08	2	4.0	3.8	4.2	3.4	3.2	4.0	4.5	3.1	3.3	3.0	2.7	3.0	2.5	4.3
		4	3.7	3.7	3.7	3.1	2.6	2.2	2.4	3.6	3.8	3.2	3.3	2.8	2.3	3.1
		6	4.3	3.9	3.7	2.9	2.2	1.5	1.4	4.3	4.5	3.4	4.0	3.4	2.3	2.7
		8	4.5	4.0	3.7	3.4	2.0	1.3	1.2	5.0	5.5	4.8	4.4	4.1	3.4	4.0
	0.15	2	4.2	3.8	3.6	4.0	3.1	2.9	3.9	3.3	3.0	2.5	2.4	1.8	2.4	3.1
		4	2.9	3.3	3.5	2.8	1.9	1.6	1.8	3.4	3.6	2.9	2.5	2.3	2.5	2.6
		6	3.2	3.6	3.5	3.2	1.7	1.3	1.0	3.8	3.7	2.6	3.5	2.4	3.1	3.2
		8	3.0	3.6	3.2	2.3	1.5	0.9	0.7	3.9	4.1	3.6	3.9	3.4	3.0	3.4
	0.40	2	3.5	3.9	3.0	3.1	2.7	2.2	2.3	1.8	2.6	2.3	1.5	2.2	2.4	3.2
		4	3.0	2.8	2.6	2.2	1.4	1.0	0.9	2.3	2.2	2.4	2.0	2.1	2.4	3.1
		6	3.2	3.2	2.9	2.0	1.2	0.6	0.5	2.9	2.8	3.2	2.6	2.7	2.9	4.3
		8	2.9	3.1	2.3	1.9	1.3	0.7	0.6	3.0	3.1	3.7	3.4	4.0	4.0	4.8
	1.20	2	3.1	3.4	2.6	2.4	1.8	1.8	1.5	1.8	2.1	2.1	2.1	3.0	4.6	5.0
		4	2.3	2.1	1.7	1.4	0.8	0.6	0.4	1.4	2.0	2.4	2.9	4.2	5.4	6.3
		6	2.0	1.7	1.6	1.2	0.7	0.5	0.5	1.9	2.4	2.9	3.9	4.7	6.5	8.7
		8	1.7	2.0	1.7	1.4	0.6	0.5	0.4	2.3	3.0	3.7	5.3	6.9	8.4	10.6
300	0.04	2	4.2	5.0	4.4	4.4	3.8	4.1	3.9	4.0	4.1	3.8	3.4	2.9	3.0	3.0
		4	4.5	4.1	4.6	3.5	3.3	2.7	2.4	4.5	4.1	4.6	4.0	3.1	2.3	2.4
		6	4.3	5.0	3.8	3.5	2.6	2.1	1.7	5.0	5.7	4.9	4.3	3.2	2.5	2.2
		8	3.7	4.2	3.8	3.3	3.2	1.7	1.2	5.2	5.7	5.0	4.9	3.5	2.9	2.1
	0.08	2	4.5	4.7	4.2	4.3	3.4	3.8	2.9	3.9	3.2	3.5	3.2	3.6	3.0	2.6
		4	4.7	4.5	3.9	3.9	3.3	2.4	2.3	4.2	4.6	4.2	3.2	3.3	2.3	2.0
		6	4.2	3.6	4.0	4.0	3.0	1.9	1.5	4.5	4.8	4.3	4.1	3.4	2.4	2.1
		8	4.1	4.2	3.9	2.7	2.5	1.6	0.7	5.6	5.2	5.0	4.0	3.5	3.4	1.7
	0.15	2	4.0	4.8	3.8	4.0	4.0	2.9	2.7	4.4	3.6	3.7	2.9	2.6	2.6	2.7
		4	4.1	4.0	4.0	3.5	2.9	2.4	1.7	4.0	3.5	3.2	3.3	2.1	2.2	1.6
		6	4.2	3.7	3.6	3.5	2.0	1.6	1.0	4.5	4.3	4.3	4.3	2.8	2.1	1.5
		8	4.0	3.7	3.7	3.1	2.2	1.4	0.4	4.9	5.0	4.2	3.9	3.6	2.2	1.4
	0.40	2	4.3	4.5	3.8	3.8	4.1	3.0	2.8	3.9	3.4	3.5	2.8	2.5	2.0	2.0
		4	4.0	3.9	3.6	2.6	3.3	2.1	1.3	3.8	3.4	3.5	2.9	2.4	1.7	1.6
		6	3.6	3.9	3.6	3.4	2.0	1.4	0.9	3.9	3.7	3.1	3.1	2.6	1.8	1.5
		8	3.4	3.4	3.1	2.6	1.8	1.0	0.4	4.5	3.9	3.9	3.0	2.4	1.6	1.3
	1.20	2	4.3	4.5	3.9	3.5	2.7	3.2	2.1	3.1	2.7	2.8	2.3	2.2	2.0	1.6
		4	3.2	3.6	4.0	2.8	2.2	1.3	0.9	2.8	2.8	2.5	2.3	1.7	1.3	0.9
		6	3.2	2.9	2.7	2.6	1.7	1.0	0.3	3.2	3.2	2.7	2.3	2.2	1.3	1.3
		8	2.9	2.8	2.8	2.1	1.4	0.8	0.2	3.5	3.6	3.0	2.6	2.3	1.6	1.0

Table 3
Empirical powers (%) of the tests T^l_{QS}, T^l_{PR}, T^l_{BT}, T^q_{QS}, T^q_{PR}, T^q_{BT}, Z_{tr}, Z_{det} and Zd_{tr} for Models 4–6 at the 5% nominal level.

			Model 4									Model 5									Model 6								
n	p/n	K	T^l_{QS}	T^l_{PR}	T^l_{BT}	T^q_{QS}	T^q_{PR}	T^q_{BT}	Z_{tr}	Z_{det}	Zd_{tr}	T^l_{QS}	T^l_{PR}	T^l_{BT}	T^q_{QS}	T^q_{PR}	T^q_{BT}	Z_{tr}	Z_{det}	Zd_{tr}	T^l_{QS}	T^l_{PR}	T^l_{BT}	T^q_{QS}	T^q_{PR}	T^q_{BT}	Z_{tr}	Z_{det}	Zd_{tr}
100	0.04	2	79.0	78.1	81.2	93.5	93.5	94.5	6.1	5.5	6.0	65.7	65.2	69.1	94.2	94.3	95.5	4.5	4.7	5.6	77.0	77.4	84.4	80.5	81.3	85.4	100	63.9	100
		4	87.8	87.5	89.2	97.8	97.8	98.3	6.7	6.0	5.9	81.5	80.3	83.4	98.3	98.4	98.6	4.5	4.6	6.6	66.7	66.5	77.7	75.7	76.2	81.8	99.7	62.4	99.4
		6	91.1	90.9	92.6	98.7	98.7	99.0	6.4	5.0	6.6	87.1	86.8	89.1	99.0	99.0	99.2	5.0	5.4	7.3	64.7	63.8	76.8	75.6	76.8	82.3	97.7	59.6	97.0
		8	93.5	93.0	94.5	99.3	99.1	99.4	4.6	5.7	6.8	90.9	90.8	92.3	99.2	99.1	99.4	4.9	5.3	7.9	66.6	65.1	78.0	77.8	77.9	84.8	94.2	52.5	91.1
	0.08	2	82.2	81.6	84.8	93.0	93.2	94.7	6.4	7.0	5.7	71.8	71.7	76.3	94.9	95.0	96.0	4.3	5.0	4.9	75.0	75.1	85.2	71.0	72.3	78.7	100	51.7	100
		4	91.8	91.4	93.0	97.7	97.8	98.2	6.7	6.9	6.3	86.7	86.0	89.2	97.9	97.8	98.4	5.7	5.6	6.9	65.8	65.2	79.9	65.6	67.6	75.3	100	63.3	100
		6	93.9	93.6	95.3	98.6	98.7	98.9	5.4	7.4	6.7	90.7	90.4	92.7	99.0	99.1	99.3	6.5	6.3	8.2	63.6	62.7	78.6	65.0	65.7	75.0	100	66.0	99.9
		8	95.3	95.0	96.3	99.1	99.2	99.3	5.8	6.7	7.0	92.6	92.6	94.9	99.0	98.9	99.2	6.6	6.1	9.1	61.6	60.4	77.8	65.8	66.5	75.9	99.7	61.4	99.7
	0.15	2	84.2	84.1	87.2	90.8	91.0	92.7	NA	NA	5.9	74.7	74.6	80.1	91.7	91.9	93.7	NA	NA	5.3	71.3	71.6	83.3	60.3	62.4	69.9	NA	NA	100
		4	92.2	91.6	94.4	96.8	96.7	97.3	NA	NA	6.3	88.1	87.5	91.3	96.8	96.8	97.6	NA	NA	6.6	58.4	58.5	76.7	53.1	54.8	64.9	NA	NA	100
		6	95.2	95.0	96.6	97.7	97.7	97.9	NA	NA	7.9	91.5	91.4	94.2	97.7	97.9	98.1	NA	NA	9.4	55.5	54.6	75.4	50.1	51.5	62.9	NA	NA	100
		8	96.0	95.8	97.1	98.3	98.2	98.6	NA	NA	8.2	94.4	94.4	96.2	98.4	98.5	98.8	NA	NA	11.2	54.3	53.2	75.3	48.9	50.8	61.7	NA	NA	100
	0.40	2	85.0	84.5	88.7	78.8	79.7	82.4	NA	NA	6.3	73.7	73.4	79.6	80.5	81.4	84.7	NA	NA	5.9	62.0	62.8	81.7	44.5	47.7	55.3	NA	NA	100
		4	91.5	91.1	94.3	88.4	89.0	90.4	NA	NA	7.0	87.6	87.0	91.4	90.0	90.7	92.6	NA	NA	9.2	47.2	47.0	72.9	34.6	37.7	44.9	NA	NA	100
		6	94.6	94.2	96.9	91.7	92.2	93.4	NA	NA	6.7	91.7	91.2	94.5	91.8	92.2	93.7	NA	NA	12.8	40.4	38.8	69.5	32.3	34.1	41.1	NA	NA	100
		8	95.2	94.9	97.2	92.2	92.7	93.3	NA	NA	9.4	92.9	92.8	95.8	93.6	93.9	95.2	NA	NA	14.5	36.3	35.1	65.4	32.0	34.1	39.8	NA	NA	100
	1.20	2	78.5	78.0	83.9	44.3	45.6	48.8	NA	NA	NA	67.6	68.3	77.9	52.5	55.0	58.8	NA	NA	NA	49.3	49.1	77.3	43.7	46.8	45.8	NA	NA	NA
		4	87.6	86.9	91.8	57.0	58.6	61.8	NA	NA	NA	82.8	82.1	88.8	64.3	66.2	70.1	NA	NA	NA	30.0	29.4	60.8	46.9	49.4	43.5	NA	NA	NA
		6	89.2	88.8	93.2	58.3	60.0	62.7	NA	NA	NA	87.1	86.4	92.2	68.6	71.0	73.6	NA	NA	NA	21.7	20.9	52.3	49.7	53.1	44.8	NA	NA	NA
		8	90.8	90.3	94.5	62.6	64.2	66.9	NA	NA	NA	88.9	88.6	92.9	69.4	71.4	74.8	NA	NA	NA	17.5	16.7	44.4	53.9	56.2	47.9	NA	NA	NA
300	0.04	2	100	100	100	100	100	100	13.6	8.3	8.6	100	100	100	100	100	100	5.0	5.3	5.2	100	100	100	99.4	99.5	99.7	100	99.0	100
		4	100	100	100	100	100	100	15.1	8.9	10.4	100	100	100	100	100	100	5.1	5.9	5.7	99.8	99.8	99.9	98.9	99.0	99.3	100	98.6	100
		6	100	100	100	100	100	100	10.4	8.9	7.9	100	100	100	100	100	100	6.1	6.5	6.1	99.8	99.7	99.9	98.6	98.5	99.1	100	97.0	100
		8	100	100	100	100	100	100	9.7	9.2	6.2	100	100	100	100	100	100	6.5	6.9	6.2	100	100	100	99.2	99.2	99.6	100	95.7	100
	0.08	2	100	100	100	100	100	100	NA	NA	9.3	100	100	100	100	100	100	NA	NA	4.8	100	100	100	98.6	98.6	99.3	NA	NA	100
		4	100	100	100	100	100	100	NA	NA	11.7	100	100	100	100	100	100	NA	NA	5.9	99.9	99.9	100	98.2	98.2	99.1	NA	NA	100
		6	100	100	100	100	100	100	NA	NA	8.3	100	100	100	100	100	100	NA	NA	6.0	99.9	99.9	100	98.3	98.4	99.1	NA	NA	100
		8	100	100	100	100	100	100	NA	NA	7.3	100	100	100	100	100	100	NA	NA	7.0	100	100	100	98.3	98.3	99.0	NA	NA	100
	0.15	2	100	100	100	100	100	100	NA	NA	10.2	100	100	100	100	100	100	NA	NA	5.6	100	100	100	98.2	98.4	98.8	NA	NA	100
		4	100	100	100	100	100	100	NA	NA	11.1	100	100	100	100	100	100	NA	NA	6.1	99.9	99.9	100	97.0	97.0	98.3	NA	NA	100
		6	100	100	100	100	100	100	NA	NA	9.9	100	100	100	100	100	100	NA	NA	6.8	99.9	99.9	100	97.2	97.3	98.6	NA	NA	100
		8	100	100	100	100	100	100	NA	NA	7.0	100	100	100	100	100	100	NA	NA	7.3	99.9	100	100	96.8	97.0	98.6	NA	NA	100
	0.40	2	100	100	100	100	100	100	NA	NA	NA	100	100	100	100	100	100	NA	NA	NA	100	100	100	96.4	96.4	98.2	NA	NA	NA
		4	100	100	100	100	100	100	NA	NA	NA	100	100	100	100	100	100	NA	NA	NA	100	99.9	100	94.3	94.4	97.4	NA	NA	NA
		6	100	100	100	100	100	100	NA	NA	NA	100	100	100	100	100	100	NA	NA	NA	99.9	99.9	100	92.0	92.1	96.9	NA	NA	NA
		8	100	100	100	100	100	100	NA	NA	NA	100	100	100	100	100	100	NA	NA	NA	100	100	100	92.5	92.3	96.9	NA	NA	NA
	1.20	2	100	100	100	100	100	100	NA	NA	NA	100	100	100	100	100	100	NA	NA	NA	100	100	100	92.9	92.7	97.0	NA	NA	NA
		4	100	100	100	100	100	100	NA	NA	NA	100	100	100	100	100	100	NA	NA	NA	100	100	100	82.9	83.2	93.0	NA	NA	NA
		6	100	100	100	100	100	100	NA	NA	NA	100	100	100	100	100	100	NA	NA	NA	100	100	100	76.0	76.1	89.6	NA	NA	NA
		8	100	100	100	100	100	100	NA	NA	NA	100	100	100	100	100	100	NA	NA	NA	100	100	100	69.3	69.3	87.1	NA	NA	NA

linear map for the three models. This should not be surprising. Since the alternatives in the three models are nonlinear transformations, the linear and quadratic map can capture both linear and nonlinear dependence. Generally speaking, both of the two maps perform well in the three models.

6. Real data analysis

In this section, we apply our proposed tests to a real dataset, which collects weekly closing prices from 17 September 2004 to 26 December 2008 for 394 stocks. The returns of the stocks are obtained by the log difference of the data. And the sample size n for the returns is 223. These stocks can be classified into 9 major sectors, which consist of materials (22 stocks), real estate (25 stocks), utilities (26 stocks), consumer staples (30 stocks), healthcare (55 stocks), industrials (56 stocks), financials (58 stocks), IT (60 stocks), and consumer discretionary (62 stocks). Here we examine the validity of the martingale difference hypothesis within each sector and for all stocks using our tests and the ones proposed in Hong et al. (2017). Note that neither Z_{tr} nor Z_{det} is applicable here, since $p < \sqrt{n}$ is violated for each sector. Hence we only present the results of Zd_{tr} for each sector, as it is not usable when we apply to all stock returns. Denote by $\mathbf{x}_t$ the returns of these stocks at time t. Financial theory usually assumes the stock prices follow geometric Brownian Motion which implies $\mathbb{E}(\mathbf{x}_t) = \mathbf{0}$ under the efficient markets hypothesis. We can propose the test statistic $T_{\mathrm{mean}} = |n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t|_\infty$ for the null hypothesis $H_0 : \mathbb{E}(\mathbf{x}_t) = \mathbf{0}$. Using the method given in Section 4.1 of Chang et al. (2021b) with three kernels (QS, PR, BT) to estimate the associated long-run covariance matrix, the associated p-values for such null hypothesis are 0.759, 0.749 and 0.753, respectively, which means there is no strong evidence against the zero-mean assumption of $\mathbf{x}_t$ in our real data.

Table 5 reports the p-values of Zd_{tr} and our tests with assuming $\mathbb{E}(\mathbf{x}_t) = \mathbf{0}$ and without assuming $\mathbb{E}(\mathbf{x}_t) = \mathbf{0}$. It appears that there is no strong evidence against the martingale difference hypothesis based on all tests, except for a marginally significant p-value of Zd_{tr} when $K = 2$ for the sector of consumer staples. Generally speaking, the martingale difference hypothesis is expected to hold for the weekly returns data, so in a sense both our tests and Zd_{tr} help confirming this property. For the same map, the use of different kernels do not seem to affect the p-values much, indicating the insensitivity of our results with respect to the kernel. For this particular dataset, the use of linear and quadratic maps also produces p-values that are not far away from the use of linear maps alone, for most sectors. The p-values corresponding to Zd_{tr} seem to monotonically decrease as K goes down from 8 to 2 for all sectors, an interesting phenomenon worthy of some theoretical investigation. In addition, the results of our tests with assuming $\mathbb{E}(\mathbf{x}_t) = \mathbf{0}$ and without assuming $\mathbb{E}(\mathbf{x}_t) = \mathbf{0}$

Table 4
Empirical powers (%) of the tests T^l_{BT} and T^q_{BT} for Models 4–6 at the 5% nominal level, where c represents the constant which is multiplied by Andrews' bandwidth.

			Model 4 with T^l_{BT}							Model 4 with T^q_{BT}						
			c							c						
n	p/n	K	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}
100	0.04	2	81.2	81.8	80.5	78.8	76.6	73.2	72.1	95.7	95.9	95.6	93.5	93.8	92.8	92.4
		4	89.9	89.8	89.3	88.6	86.7	80.6	78.3	99.1	98.8	98.9	98.0	97.7	96.6	97.0
		6	93.2	93.2	94.0	91.3	89.8	85.1	80.6	99.3	99.2	99.4	99.0	99.0	98.2	98.2
		8	94.4	94.7	94.4	93.6	91.2	88.3	82.7	99.6	99.6	99.2	99.4	99.1	99.1	98.8
	0.08	2	87.0	86.7	86.6	85.0	80.9	76.6	72.7	95.7	95.7	95.2	94.9	94.2	92.6	92.4
		4	93.7	93.5	93.8	92.6	89.8	83.9	77.3	98.4	98.9	98.7	98.2	97.5	97.1	96.4
		6	96.7	96.2	96.1	94.6	93.5	87.2	80.4	99.2	99.2	99.3	99.0	98.5	98.2	97.9
		8	97.1	96.9	97.2	96.2	93.9	90.2	83.5	99.4	99.4	99.1	99.2	98.6	98.9	98.4
	0.15	2	88.4	88.5	89.5	87.0	82.2	75.4	69.6	94.4	93.6	93.7	92.8	90.5	90.3	90.3
		4	95.6	95.4	94.7	94.6	91.4	83.6	74.9	97.9	97.7	97.4	97.3	97.1	95.8	94.7
		6	97.2	96.8	97.2	96.9	94.2	87.8	76.9	98.3	98.4	98.3	98.0	98.2	97.0	95.7
		8	97.8	98.0	98.3	97.0	94.9	89.5	80.8	98.6	98.6	99.0	98.7	98.3	97.1	96.4
	0.40	2	89.7	90.9	89.5	88.2	81.8	71.2	63.0	82.9	84.6	82.3	81.4	79.2	76.2	78.4
		4	95.6	96.3	95.7	93.9	88.7	79.3	66.2	90.4	90.2	90.7	89.6	88.0	84.7	84.0
		6	97.2	97.8	97.0	96.1	92.3	82.3	68.2	92.4	92.6	92.0	92.1	91.7	88.6	87.6
		8	97.8	97.8	97.9	96.7	93.0	85.7	71.0	93.3	93.1	93.7	93.1	92.4	90.7	88.7
	1.20	2	86.7	86.2	88.1	83.9	76.3	62.3	49.4	51.2	49.4	49.8	48.7	45.3	43.6	46.9
		4	93.6	93.6	93.0	90.4	84.4	69.1	51.9	60.3	60.9	60.7	58.4	55.2	51.9	52.5
		6	95.3	96.1	96.0	93.9	87.1	72.2	53.4	63.6	63.9	63.1	63.7	60.1	56.4	54.5
		8	96.4	96.0	96.3	94.7	89.3	75.2	55.3	63.8	65.7	65.5	65.8	61.0	58.6	57.2
300	0.04	2	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		4	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		6	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		8	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	0.08	2	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		4	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		6	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		8	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	0.15	2	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		4	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		6	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		8	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	0.40	2	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		4	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		6	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		8	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	1.20	2	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		4	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		6	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		8	100	100	100	100	100	100	100	100	100	100	100	100	100	100

			Model 5 with T^l_{BT}							Model 5 with T^q_{BT}						
			c							c						
n	p/n	K	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}
100	0.04	2	73.2	73.7	72.0	69.0	64.3	61.5	61.2	96.0	96.3	96.0	95.8	95.0	94.0	93.9
		4	85.9	85.6	85.9	83.3	79.6	72.9	69.7	99.0	98.9	98.8	98.8	97.9	97.5	96.7
		6	90.2	90.4	90.5	89.3	85.9	79.9	75.9	99.0	99.5	99.1	99.0	98.8	98.6	98.2
		8	93.1	93.4	92.5	91.8	89.1	84.7	78.3	99.5	99.2	99.4	99.5	99.4	98.9	98.9
	0.08	2	80.6	80.8	78.9	76.6	69.9	65.3	63.2	96.0	96.4	95.7	95.3	93.9	93.5	93.3
		4	90.4	90.7	91.1	88.9	84.1	76.9	70.5	98.9	98.4	98.6	98.7	97.6	96.8	96.3
		6	93.9	94.3	93.8	92.6	89.9	82.0	74.6	99.1	99.2	99.3	99.0	98.6	98.1	98.1
		8	95.9	95.8	95.6	95.1	92.1	86.6	79.1	99.5	99.5	99.4	99.3	99.0	99.0	98.4
	0.15	2	84.7	83.8	83.9	79.0	72.5	65.9	59.6	95.1	95.4	94.4	93.2	92.1	91.1	90.5
		4	93.0	92.6	92.5	91.2	86.1	75.5	69.4	97.9	97.9	98.1	97.3	96.9	95.3	94.9
		6	95.6	95.6	95.5	93.8	90.7	81.4	74.8	98.9	98.6	98.5	98.2	97.8	96.6	96.8
		8	96.9	96.9	96.6	95.7	93.1	85.4	77.3	98.8	98.8	98.8	98.6	98.5	97.7	97.2
	0.40	2	85.7	85.3	85.3	81.4	72.6	61.7	53.4	87.1	87.3	86.9	85.2	81.8	79.7	81.6
		4	93.2	94.2	93.1	91.2	84.8	73.1	59.2	93.6	93.4	93.2	91.4	90.4	87.7	86.9
		6	95.7	96.5	96.3	93.5	88.9	76.9	64.9	94.8	94.6	94.6	93.5	93.0	89.7	89.1
		8	97.5	97.4	97.0	95.6	92.3	81.1	67.4	94.9	95.1	95.2	94.1	93.6	91.3	90.2
	1.20	2	84.2	84.1	84.0	78.5	65.1	52.3	41.8	64.0	63.2	63.0	58.6	55.2	51.8	58.8
		4	91.4	91.5	91.2	88.3	79.5	59.5	45.3	72.4	71.0	71.7	69.7	65.4	60.1	61.8
		6	94.5	94.6	94.5	91.8	84.5	67.7	50.3	75.2	74.7	74.1	73.7	69.4	64.4	66.4
		8	96.1	95.5	95.4	93.3	86.7	73.2	54.2	75.6	74.6	74.8	74.8	70.8	68.1	67.5
300	0.04	2	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		4	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		6	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		8	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	0.08	2	100	100	100	100	100	100	99.9	100	100	100	100	100	100	100
		4	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		6	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		8	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	0.15	2	100	100	100	100	100	100	99.9	100	100	100	100	100	100	100
		4	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		6	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		8	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	0.40	2	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		4	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		6	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		8	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	1.20	2	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		4	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		6	100	100	100	100	100	100	100	100	100	100	100	100	100	100
		8	100	100	100	100	100	100	100	100	100	100	100	100	100	100

			Model 6 with T^l_{BT}							Model 6 with T^q_{BT}						
			c							c						
n	p/n	K	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}
100	0.04	2	99.5	98.2	93.2	84.3	79.1	79.9	88.1	99.6	96.7	90.6	84.4	84.7	87.3	95.6
		4	98.5	97.7	91.6	78.6	68.2	68.0	78.2	99.4	96.9	91.2	82.8	78.8	84.0	93.1
		6	98.7	97.6	91.8	78.5	65.0	64.2	72.5	99.7	97.7	91.0	84.3	79.7	82.6	92.0
		8	98.6	97.7	91.2	78.9	65.4	61.6	70.3	99.5	98.0	91.8	85.1	80.3	83.2	92.4
	0.08	2	99.9	99.3	94.6	85.1	77.4	77.6	88.7	99.7	95.3	86.8	78.4	76.5	84.0	94.3
		4	99.9	99.2	94.0	78.3	66.9	66.9	79.4	99.3	96.1	85.7	74.2	71.9	79.8	92.6
		6	99.7	99.3	94.0	78.5	62.2	59.3	73.7	99.3	95.8	85.9	76.4	70.0	77.7	91.8
		8	99.8	99.5	94.8	78.5	61.2	56.5	70.6	99.1	96.3	87.9	76.0	69.7	76.2	90.5
	0.15	2	100	99.8	95.4	83.1	76.1	76.3	87.1	99.0	93.9	81.1	71.2	70.1	79.7	93.2
		4	99.9	99.5	94.8	76.5	61.1	60.5	77.3	98.3	93.0	78.6	63.0	59.7	71.7	90.7
		6	99.9	99.4	94.6	75.3	58.6	53.0	69.7	98.1	93.3	78.7	63.0	58.3	68.4	89.5
		8	99.9	99.5	94.9	78.2	55.2	48.3	66.7	97.9	93.6	79.2	65.0	56.9	65.7	87.5
	0.40	2	99.9	99.5	96.2	82.6	67.3	67.3	83.4	95.6	87.2	69.5	55.1	55.0	68.6	90.4
		4	99.9	99.3	94.8	73.1	49.7	47.0	68.7	92.6	84.6	63.1	44.1	43.0	56.7	86.5
		6	99.8	99.4	95.3	69.5	41.9	36.9	60.5	89.4	80.8	58.7	40.1	41.6	52.0	83.8
		8	99.8	99.4	95.0	66.2	38.8	33.1	54.8	87.6	79.8	56.9	39.6	38.1	51.9	83.1
	1.20	2	99.7	99.3	97.0	77.6	54.1	54.0	75.4	78.9	67.6	51.2	46.8	50.3	64.6	89.0
		4	99.1	99.0	95.2	63.0	30.7	28.2	53.1	64.1	56.3	41.5	42.9	49.9	59.9	85.0
		6	99.1	98.0	93.0	53.2	23.1	20.5	43.1	55.7	49.7	40.3	45.1	54.0	64.0	84.4
		8	98.3	96.8	92.8	48.7	20.8	18.3	40.4	50.8	46.6	41.2	48.1	57.5	68.9	85.1
300	0.04	2	100	100	100	100	99.9	99.9	99.9	100	99.9	99.8	99.6	99.5	99.4	99.7
		4	100	100	100	100	99.9	99.7	99.5	100	100	99.8	99.5	99.1	99.2	99.4
		6	100	100	100	100	99.8	99.6	99.2	100	99.9	99.8	99.4	99.0	99.1	99.5
		8	100	100	100	100	99.9	99.6	99.3	100	100	99.8	99.5	99.4	99.2	99.0
	0.08	2	100	100	100	100	99.9	99.9	99.9	100	99.9	99.7	99.3	99.0	99.0	99.5
		4	100	100	100	100	99.9	99.6	99.3	100	99.9	99.5	98.9	98.9	98.5	99.1
		6	100	100	100	100	99.9	99.7	99.3	100	100	99.5	98.9	98.5	98.4	99.1
		8	100	100	100	100	100	99.6	99.2	100	99.8	99.7	99.3	98.5	98.5	99.1
	0.15	2	100	100	100	100	100	99.9	100	100	100	99.3	99.1	98.8	99.0	99.5
		4	100	100	100	100	99.8	99.8	99.5	100	99.8	99.5	98.7	97.8	97.7	98.8
		6	100	100	100	100	99.9	99.7	99.1	100	100	99.4	98.6	97.8	98.1	98.5
		8	100	100	100	100	100	99.7	99.2	100	100	99.4	98.7	98.1	97.3	98.5
	0.40	2	100	100	100	100	100	100	99.9	100	99.8	99.3	97.9	97.3	97.6	98.7
		4	100	100	100	100	100	99.7	99.4	99.8	99.8	98.9	97.5	95.3	94.9	96.6
		6	100	100	100	100	100	99.7	98.8	99.8	99.8	98.5	97.3	94.2	93.3	95.5
		8	100	100	100	100	100	99.7	98.4	99.8	99.7	98.9	96.6	92.7	91.5	94.5
	1.20	2	100	100	100	100	100	100	99.9	99.7	99.5	98.9	96.9	95.2	94.3	95.8
		4	100	100	100	100	100	99.8	98.7	98.5	98.3	97.4	92.9	87.0	84.6	88.8
		6	100	100	100	100	100	99.5	97.4	98.0	97.5	96.1	89.5	79.0	76.0	82.5
		8	100	100	100	100	100	98.8	95.5	96.6	96.0	94.5	86.5	72.4	68.5	77.6

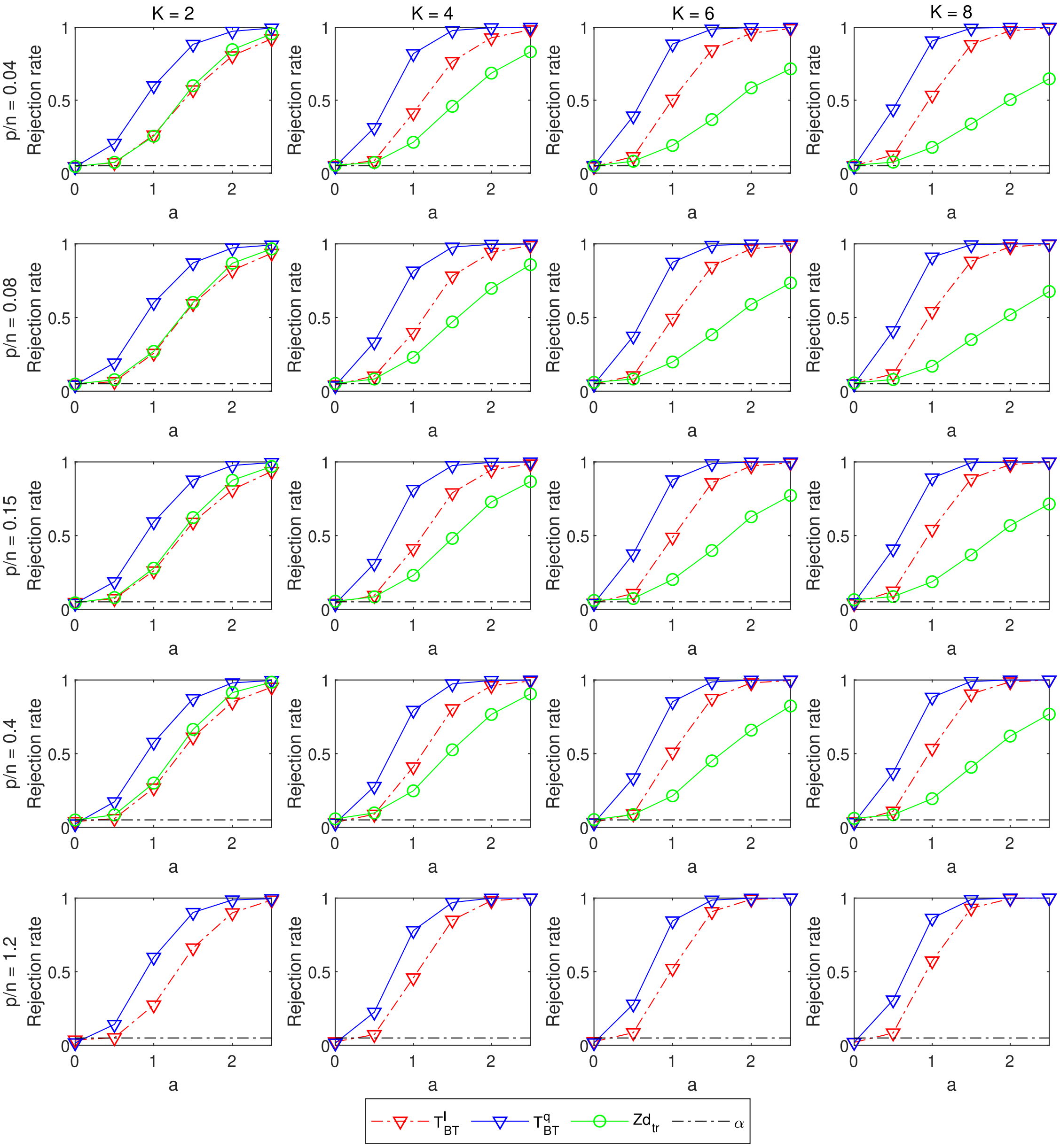

Fig. 1. Empirical sizes and powers of T^l_{BT}, T^q_{BT} and Zd_{tr} for Model 1′ at the nominal level $\alpha = 0.05$, where the sample size $n = 100$.

are quite similar, which is consistent with the aforementioned conclusion that $\mathbb{E}(\mathbf{x}_t)$ is not significantly different from zero. Overall, our tests are preferred to the ones proposed in Hong et al. (2017) due to the fact that they can be used regardless of whether the dimension p exceeds the sample size n.

7. Discussion

In this paper, we propose a new martingale difference test that captures nonlinear serial dependence and works in the high-dimensional environment, as motivated by the increasing availability of high-dimensional nonlinear time series from economics and finance. Under mild moment and weak temporal dependence assumptions, we establish the validity of Gaussian approximation and provide a simulation-based approach for critical values. In addition to its built-in capability of accommodating both low and high dimensions, our test also has a number of appealing features such as being robust to conditional moments of unknown forms and strong/weak cross-series dependence. From our numerical simulations and a real data analysis, we observe quite encouraging finite sample performance. Therefore we feel confident to recommend

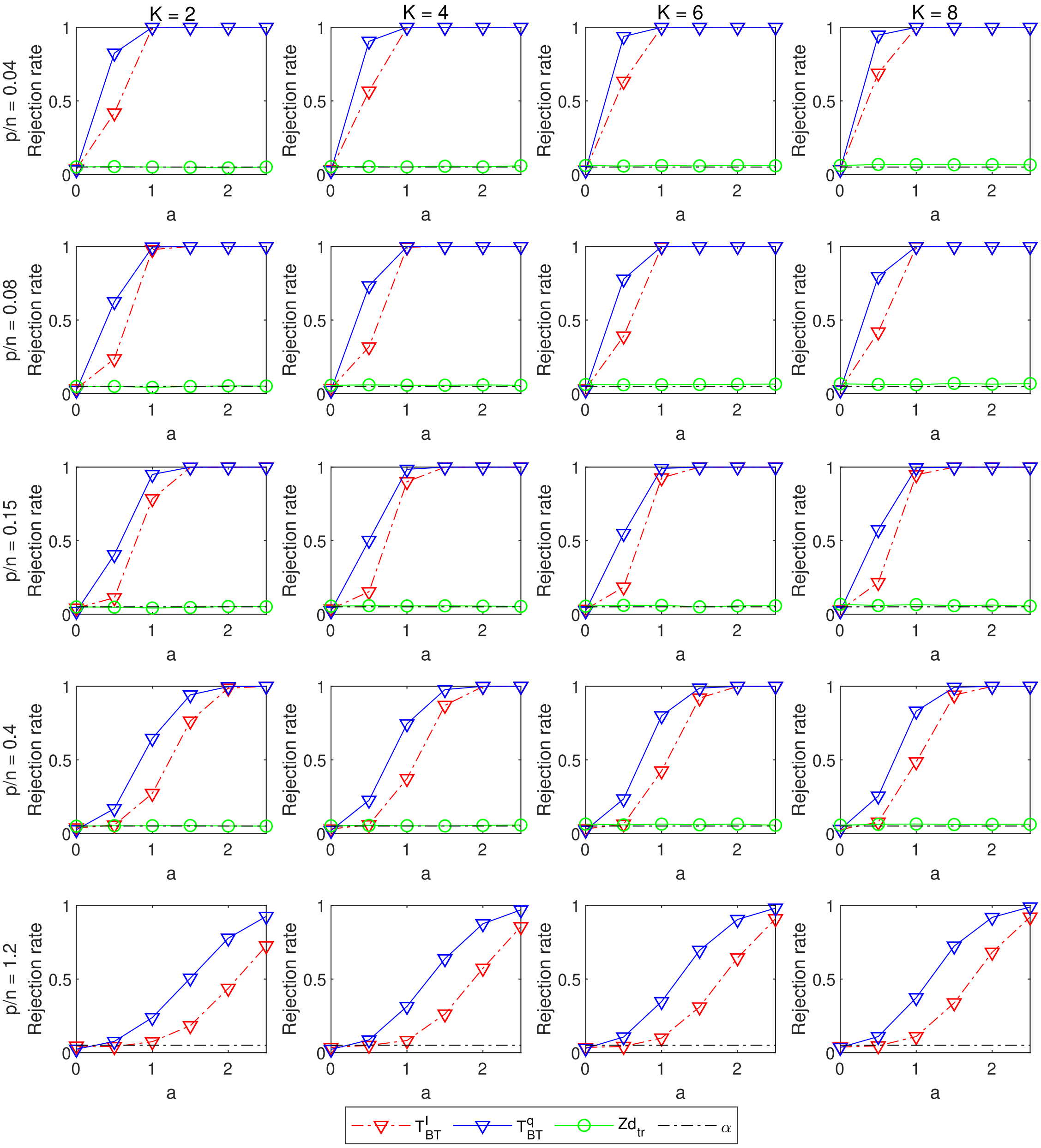

Fig. 2. Empirical sizes and powers of T^{l}_{BT}, T^{q}_{BT} and Zd_{tr} for Model 2′ at the nominal level $\alpha = 0.05$, where the sample size $n = 100$.

its use by the practitioners when there is a need to assess the martingale difference hypothesis for econometric/financial time series of moderate or high dimension.

In the literature, testing quantile/directional predictability has been studied for low-dimensional time series; see Han et al. (2016). It would be also interesting to extend their test to the high-dimensional setting. A sound data-driven bandwidth choice in our simulation-based approach for generating the critical values merits additional research, especially from a testing-optimal viewpoint. We leave these topics for future investigation.

8. Technical proofs

In this section, we provide the detailed proofs for all theoretical results stated in the paper, and also introduce necessary lemmas and propositions with proofs. Throughout this section, we use C to denote a generic positive finite constant that does not depend on (p, d, n, K) and may be different in different uses. For two sequences of positive numbers $\{a_n\}$ and

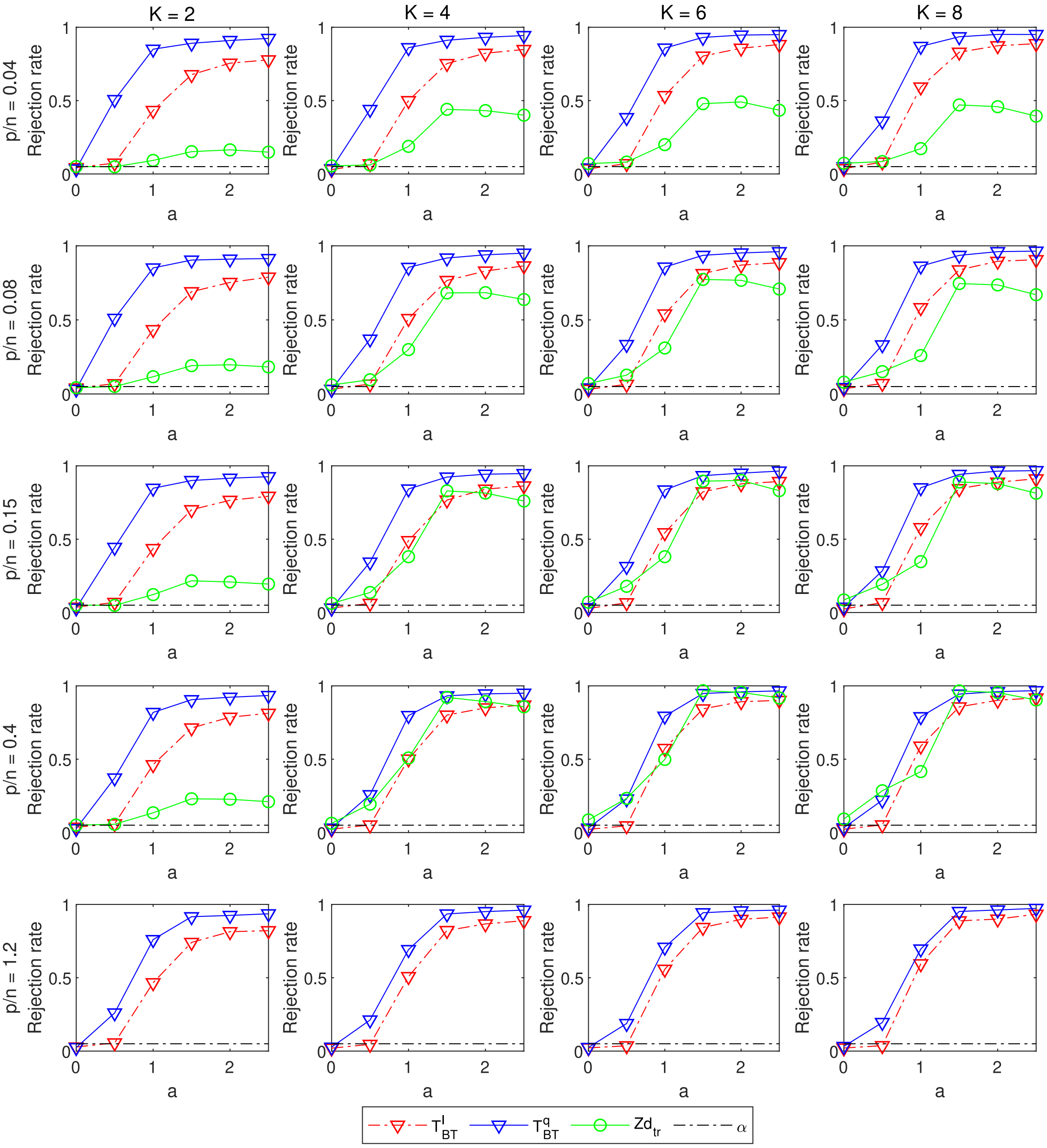

Fig. 3. Empirical sizes and powers of T^{l}_{BT}, T^{q}_{BT} and Zd_{tr} for Model 3′ at the nominal level $\alpha = 0.05$, where the sample size $n = 100$.

$\{b_n\}$, we write $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ if $\limsup_{n\to\infty} a_n/b_n \leqslant c_0$ for some positive constant c_0. We write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold simultaneously. We write $a_n \ll b_n$ or $b_n \gg a_n$ if $\limsup_{n\to\infty} a_n/b_n = 0$. For a countable set $\mathcal{F}$, we use $|\mathcal{F}|$ to denote the cardinality of $\mathcal{F}$.

Write $\mathbf{u} := (u_1, \ldots, u_{Kpd})^\top = (\hat{\boldsymbol{\gamma}}_1^\top, \ldots, \hat{\boldsymbol{\gamma}}_K^\top)^\top$ with $\hat{\boldsymbol{\gamma}}_j = (n-j)^{-1}\sum_{t=1}^{n-j} \mathrm{vec}\{\boldsymbol{\phi}(\mathbf{x}_t)\mathbf{x}_{t+j}^\top\}$ for any $j \in [K]$. Let $\tilde{n} = n - K$. Recall $\boldsymbol{\eta}_t = ([\mathrm{vec}\{\boldsymbol{\phi}(\mathbf{x}_t)\mathbf{x}_{t+1}^\top\}]^\top, \ldots, [\mathrm{vec}\{\boldsymbol{\phi}(\mathbf{x}_t)\mathbf{x}_{t+K}^\top\}]^\top)^\top$. Since $\{\mathbf{x}_t\}$ is an α-mixing process satisfying Condition 2, we know the newly defined process $\{\boldsymbol{\eta}_t\}$ is also α-mixing with the α-mixing coefficients $\{\tilde{\alpha}_K(k)\}_{k\geqslant 1}$ satisfying

$$\tilde{\alpha}_K(k) \leqslant C_3 \exp(-C_4|k-K|_+^{\tau_2}), \tag{23}$$

where the positive constants τ_2, C_3 and C_4 are specified in Condition 2. Write $\bar{\boldsymbol{\eta}} := (\bar{\eta}_1, \ldots, \bar{\eta}_{Kpd})^\top = \tilde{n}^{-1}\sum_{t=1}^{\tilde{n}} \boldsymbol{\eta}_t$. For each $j \in [K]$, define $Z_j = n\max_{\ell\in\mathcal{L}_j} u_\ell^2$ and $\tilde{Z}_j = \tilde{n}\max_{\ell\in\mathcal{L}_j} \bar{\eta}_\ell^2$ with $\mathcal{L}_j := \{(j-1)pd+1, \ldots, jpd\}$. Then the test statistic can be written as $T_n = n\sum_{j=1}^K |\hat{\boldsymbol{\gamma}}_j|_\infty^2 = \sum_{j=1}^K Z_j$. Furthermore, we let $\tilde{T}_n := \sum_{j=1}^K \tilde{Z}_j$.

Table 5
P-values of our tests and Hong et al.'s test for the weekly stock returns.

			MDS test						General MDS test						Hong et al.'s test
Sectors	p	K	$T^l_{\rm QS}$	$T^l_{\rm PR}$	$T^l_{\rm BT}$	$T^q_{\rm QS}$	$T^q_{\rm PR}$	$T^q_{\rm BT}$	$T^l_{\rm QS}$	$T^l_{\rm PR}$	$T^l_{\rm BT}$	$T^q_{\rm QS}$	$T^q_{\rm PR}$	$T^q_{\rm BT}$	$Zd_{\rm tr}$
Joint test	394	2	0.334	0.308	0.311	0.403	0.393	0.364	0.322	0.303	0.292	0.371	0.349	0.357	NA
		4	0.523	0.517	0.546	0.440	0.462	0.443	0.524	0.508	0.520	0.465	0.465	0.450	NA
		6	0.558	0.543	0.570	0.522	0.526	0.499	0.574	0.513	0.589	0.506	0.510	0.512	NA
		8	0.612	0.602	0.643	0.553	0.538	0.521	0.639	0.600	0.643	0.518	0.527	0.546	NA
Materials	22	2	0.687	0.676	0.723	0.723	0.691	0.687	0.707	0.697	0.696	0.719	0.701	0.696	0.500
		4	0.741	0.729	0.772	0.749	0.726	0.763	0.749	0.732	0.776	0.753	0.730	0.779	0.759
		6	0.582	0.563	0.608	0.586	0.566	0.607	0.602	0.565	0.609	0.595	0.572	0.594	0.844
		8	0.513	0.510	0.562	0.511	0.513	0.559	0.530	0.519	0.542	0.520	0.533	0.546	0.879
Real estate	25	2	0.610	0.610	0.603	0.629	0.591	0.646	0.622	0.603	0.614	0.610	0.600	0.643	0.305
		4	0.537	0.518	0.594	0.578	0.535	0.599	0.570	0.512	0.612	0.554	0.541	0.600	0.385
		6	0.464	0.440	0.491	0.475	0.440	0.488	0.474	0.454	0.498	0.476	0.450	0.475	0.510
		8	0.457	0.421	0.472	0.465	0.447	0.472	0.467	0.450	0.489	0.458	0.418	0.489	0.612
Utilities	26	2	0.710	0.683	0.721	0.706	0.687	0.707	0.679	0.659	0.698	0.716	0.691	0.674	0.166
		4	0.755	0.744	0.766	0.761	0.737	0.800	0.750	0.745	0.773	0.746	0.743	0.792	0.173
		6	0.756	0.736	0.782	0.761	0.736	0.770	0.755	0.747	0.757	0.752	0.732	0.777	0.171
		8	0.565	0.561	0.579	0.577	0.545	0.569	0.561	0.564	0.588	0.594	0.560	0.587	0.193
Consumer staples	30	2	0.804	0.803	0.838	0.650	0.648	0.685	0.804	0.786	0.843	0.683	0.650	0.713	0.042
		4	0.411	0.400	0.458	0.417	0.394	0.421	0.420	0.426	0.438	0.393	0.404	0.444	0.170
		6	0.446	0.466	0.462	0.358	0.346	0.385	0.446	0.436	0.501	0.380	0.393	0.380	0.226
		8	0.498	0.516	0.520	0.412	0.406	0.422	0.504	0.493	0.518	0.387	0.415	0.409	0.291
Healthcare	55	2	0.835	0.808	0.846	0.813	0.816	0.855	0.803	0.794	0.838	0.809	0.796	0.848	0.131
		4	0.611	0.603	0.618	0.549	0.543	0.544	0.590	0.626	0.590	0.537	0.542	0.559	0.172
		6	0.636	0.626	0.665	0.592	0.578	0.591	0.625	0.616	0.642	0.597	0.616	0.605	0.188
		8	0.661	0.641	0.657	0.615	0.621	0.636	0.626	0.626	0.656	0.618	0.595	0.614	0.351
Industrials	56	2	0.588	0.541	0.595	0.579	0.547	0.603	0.575	0.549	0.588	0.553	0.547	0.590	0.365
		4	0.642	0.640	0.677	0.665	0.617	0.696	0.676	0.625	0.697	0.670	0.657	0.686	0.485
		6	0.637	0.630	0.676	0.650	0.626	0.665	0.666	0.629	0.678	0.639	0.637	0.692	0.573
		8	0.697	0.698	0.739	0.694	0.680	0.730	0.706	0.692	0.742	0.705	0.683	0.723	0.631
Financials	58	2	0.675	0.641	0.676	0.265	0.254	0.244	0.677	0.656	0.675	0.273	0.268	0.250	0.148
		4	0.715	0.704	0.726	0.360	0.379	0.347	0.719	0.703	0.734	0.367	0.361	0.362	0.290
		6	0.710	0.708	0.724	0.429	0.441	0.416	0.706	0.674	0.730	0.429	0.422	0.406	0.370
		8	0.740	0.715	0.763	0.485	0.497	0.478	0.737	0.728	0.739	0.486	0.501	0.472	0.498
IT	60	2	0.276	0.293	0.293	0.296	0.292	0.307	0.295	0.277	0.306	0.283	0.267	0.288	0.121
		4	0.550	0.541	0.586	0.531	0.537	0.595	0.551	0.545	0.569	0.532	0.541	0.590	0.299
		6	0.610	0.599	0.615	0.611	0.577	0.623	0.593	0.566	0.634	0.583	0.583	0.610	0.454
		8	0.637	0.588	0.636	0.622	0.583	0.613	0.624	0.599	0.619	0.596	0.586	0.629	0.629
Consumer discretionary	62	2	0.273	0.273	0.264	0.286	0.316	0.303	0.267	0.274	0.260	0.318	0.306	0.308	0.407
		4	0.350	0.344	0.351	0.366	0.359	0.377	0.355	0.363	0.355	0.384	0.335	0.362	0.648
		6	0.372	0.342	0.385	0.407	0.393	0.409	0.358	0.363	0.360	0.387	0.395	0.390	0.800
		8	0.377	0.360	0.359	0.405	0.401	0.405	0.358	0.351	0.378	0.406	0.391	0.403	0.888

8.1. A key proposition

Let $\{\mathbf{z}_t\}_{t=1}^n$ be a d_z-dimensional dependent sequence with $\mathbb{E}(\mathbf{z}_t)=\mathbf{0}$ for any $t\in[n]$. Define $\mathbf{s}_{n,z}=n^{-1/2}\sum_{t=1}^n\mathbf{z}_t$ and $\Xi=\mathrm{Var}(n^{-1/2}\sum_{t=1}^n\mathbf{z}_t)$. Write $\mathbf{z}_t=(z_{t,1},\ldots,z_{t,d_z})^\top$. We assume $\{\mathbf{z}_t\}_{t=1}^n$ satisfy the following three assumptions:

AS1. There exist universal constants $b_1>1$, $b_2>0$ and $r_1\in(0,1]$ such that $\sup_{t\in[n]}\sup_{j\in[d_z]}\mathbb{P}(|z_{t,j}|>u)\leqslant b_1\exp(-b_2u^{r_1})$ for any $u>0$.

AS2. There exist universal constants $a_1>1$, $a_2>0$ and $r_2\in(0,1]$ such that the α-mixing coefficients of the sequence $\{\mathbf{z}_t\}_{t=1}^n$, denoted by $\{\alpha_z(k)\}_{k\geqslant1}$, satisfying $\alpha_z(k)\leqslant a_1\exp(-a_2|k-m|_+^{r_2})$ for any $k\geqslant1$ and some $m=m(n)>0$, where $m=o(n)$ may diverge with n.

AS3. There exists a universal constant $c>0$ such that $\mathbb{E}(|n^{-1/2}\sum_{t=1}^n z_{t,j}|^2)\geqslant c$ for any $j\in[d_z]$.

Let $\mathbf{s}_{n,y}\sim\mathcal{N}(\mathbf{0},\Xi)$ be independent of $\mathcal{Z}_n=\{\mathbf{z}_1,\ldots,\mathbf{z}_n\}$. Define

$$\varrho_n:=\sup_{\mathbf{u}\in\mathbb{R}^{d_z},\nu\in[0,1]}\left|\mathbb{P}(\sqrt{\nu}\mathbf{s}_{n,z}+\sqrt{1-\nu}\mathbf{s}_{n,y}\leqslant\mathbf{u})-\mathbb{P}(\mathbf{s}_{n,y}\leqslant\mathbf{u})\right|. \tag{24}$$

Chang et al. (2021b) gives an upper bound for ϱ_n when m is a fixed constant. Proposition 3 presents a more general result that allows m diverging with n, whose proof is presented in the supplementary material.

Proposition 3. *Assume $d_z \geqslant n^{\varpi}$ for some sufficiently small constant $\varpi > 0$. Under* AS1–AS3*, it holds that*

$$\varrho_n \lesssim \frac{m^{1/3}(\log d_z)^{2/3}}{n^{1/9}}\{m^{1/6}(\log d_z)^{1/2} + m^{1/3} + (\log d_z)^{1/(3r_2)}\}$$

provided that $\log d_z \ll \min\{m^{3r/(6+2r)}n^{7r/(18+6r)}, m^{-3r_1/(6+2r_1)}n^{7r_1/(18+6r_1)}, n^{r_2/(9-3r_2)}\}$ *with* $m \lesssim n^{1/9}(\log n)^{1/3}$, *where* $r = r_1r_2/(r_1+r_2)$ *and* r_1 *and* r_2 *are specified in* AS1 *and* AS2*, respectively.*

Proposition 3 requires m involved in Assumption AS2 cannot diverge faster than $n^{1/9}(\log n)^{1/3}$. The proof of Proposition 3 is based on the widely used "large-and-small-blocks" technique in time series analysis. The key step for the proof of Proposition 3 is to establish the associated Gaussian approximation result for the partial sum over the large blocks, see Lemma L4 in the supplementary material. The restrictions on $\log d_z$ given in Proposition 3 are derived from the conditions of Lemma L4 with suitable selections of the lengths of large and small blocks. In the proofs of Propositions 2 and 3, and Theorem 2, we need the following lemma whose proof is given in the supplementary material.

Lemma L1. *Under* AS1–AS3*, it holds that*

$$\max_{0\leqslant a\leqslant n-q}\max_{j\in[d_z]}\mathbb{P}\bigg(\max_{k\in[q]}\bigg|\sum_{t=a+1}^{a+k} z_{t,j}\bigg| \geqslant x\bigg) \lesssim \exp(-Cq^{-1}m^{-1}x^2) + qx^{-1}\exp(-Cx^r) + qx^{-1}\exp(-Cm^{-r_1}x^{r_1}) \tag{25}$$

for any $x > 0$ and $m \leqslant q \leqslant n$, where $r = r_1r_2/(r_1+r_2)$.

8.2. Proof of Proposition 1

Recall $T_n = \sum_{j=1}^K Z_j$ and $\tilde{T}_n := \sum_{j=1}^K \tilde{Z}_j$. To construct Proposition 1, we need the following lemma whose proof is given in the supplementary material.

Lemma L2. *Assume Conditions 1–3 hold. Let $\tau = \tau_1\tau_2/(\tau_1+\tau_2)$. If $\log(Kpd) = o(n^{\tau/2})$ and $K^{\tau_1}\log(Kpd) = o(n^{\tau_1/2})$, then*

$$|T_n - \tilde{T}_n| \lesssim \frac{K^{3/2}\{\log(Kpd)\}^{1/2}}{n^{1/2}}\max[\{\log(Kpd)\}^{1/\tau}, K\{\log(Kpd)\}^{1/\tau_1}]$$

with probability at least $1 - C(Kpd)^{-1}$ under H_0.

Recall $\bar{\boldsymbol{\eta}} = \tilde{n}^{-1}\sum_{t=1}^{\tilde{n}}\boldsymbol{\eta}_t$ and $G_K = \sum_{j=1}^K \max_{\ell\in\mathcal{L}_j}|g_\ell|^2$ with $\mathbf{g} = (g_1,\ldots,g_{Kpd})^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{n,K})$ where $\boldsymbol{\Sigma}_{n,K} = \tilde{n}\mathbb{E}\{(\bar{\boldsymbol{\eta}} - \boldsymbol{\mu})(\bar{\boldsymbol{\eta}} - \boldsymbol{\mu})^\top\}$ and $\boldsymbol{\mu} = \tilde{n}^{-1}\sum_{t=1}^{\tilde{n}}\mathbb{E}(\boldsymbol{\eta}_t)$. Under H_0, we have $\boldsymbol{\mu} = \mathbf{0}$. Thus $\boldsymbol{\Sigma}_{n,K} = \tilde{n}\mathbb{E}(\bar{\boldsymbol{\eta}}\bar{\boldsymbol{\eta}}^\top)$. Define $\mathbf{v} := (v_1,\ldots,v_{Kpd})^\top = \tilde{n}^{1/2}\bar{\boldsymbol{\eta}}$. Our proof includes two steps: (i) using Proposition 3 to show $\sup_{x>0}|\mathbb{P}(\tilde{T}_n \leqslant x) - \mathbb{P}(G_K \leqslant x)| = o(1)$, and (ii) using Lemma L2 to show $\sup_{x>0}|\mathbb{P}(T_n \leqslant x) - \mathbb{P}(G_K \leqslant x)| = o(1)$.

Step 1. For any $j_1,\ldots,j_K \in [pd]$ and $x > 0$, let $\mathcal{A}_{j_1,\ldots,j_K}(x) = \{\mathbf{b} \in \mathbb{R}^{Kpd} : \mathbf{b}_{S_{j_1,\ldots,j_K}}^\top \mathbf{b}_{S_{j_1,\ldots,j_K}} \leqslant x\}$ with $S_{j_1,\ldots,j_K} = \{j_1, j_2+pd, \ldots, j_K+(K-1)pd\}$. Define $\mathcal{A}(x;K) = \bigcap_{j_1=1}^{pd}\cdots\bigcap_{j_K=1}^{pd}\mathcal{A}_{j_1,\ldots,j_K}(x)$. We then have $\{\tilde{T}_n \leqslant x\} = \{\mathbf{v} \in \mathcal{A}(x;K)\}$ and $\{G_K \leqslant x\} = \{\mathbf{g} \in \mathcal{A}(x;K)\}$. Note that the set $\mathcal{A}_{j_1,\ldots,j_K}(x)$ is convex that only depends on the components in $S_{j_1,\ldots,j_K}$. For a generic integer $q \geqslant 2$, denote by $\mathbb{S}^{q-1}$ the q-dimensional unit sphere. We can reformulate $\mathcal{A}_{j_1,\ldots,j_K}(x)$ as follows:

$$\mathcal{A}_{j_1,\ldots,j_K}(x) = \bigcap_{\mathbf{a}\in\{\mathbf{a}\in\mathbb{S}^{Kpd-1}:\mathbf{a}_{S_{j_1,\ldots,j_K}}\in\mathbb{S}^{K-1}\}}\{\mathbf{b}\in\mathbb{R}^{Kpd} : \mathbf{a}^\top\mathbf{b} \leqslant \sqrt{x}\}.$$

Define $\mathcal{F} = \bigcup_{j_1=1}^{pd}\cdots\bigcup_{j_K=1}^{pd}\{\mathbf{a} \in \mathbb{S}^{Kpd-1} : \mathbf{a}_{S_{j_1,\ldots,j_K}} \in \mathbb{S}^{K-1}\}$. Then $\mathcal{A}(x;K) = \bigcap_{\mathbf{a}\in\mathcal{F}}\{\mathbf{b}\in\mathbb{R}^{Kpd} : \mathbf{a}^\top\mathbf{b} \leqslant \sqrt{x}\}$. For the unit sphere $\mathbb{S}^{K-1}$ equipped with $|\cdot|_2$, it is well-known that its ϵ-covering number $N_{\mathbb{S}^{K-1},\epsilon}$ satisfies $\epsilon^{-K} \leqslant N_{\mathbb{S}^{K-1},\epsilon} \leqslant (1+2\epsilon^{-1})^K$, see Lemma 5.2 of Vershynin (2012). Let $\mathcal{S}_\epsilon$ be an ϵ-net of $\mathbb{S}^{K-1}$ with cardinality $N_{\mathbb{S}^{K-1},\epsilon}$. Without loss of generality, we assume $\mathcal{S}_\epsilon \subset \mathbb{S}^{K-1}$. Then $\tilde{\mathcal{S}}_\epsilon^{(j_1,\ldots,j_K)} := \{\mathbf{a} \in \mathbb{S}^{Kpd-1} : \mathbf{a}_{S_{j_1,\ldots,j_K}} \in \mathcal{S}_\epsilon\}$ provides an ϵ-net of $\{\mathbf{a} \in \mathbb{S}^{Kpd-1} : \mathbf{a}_{S_{j_1,\ldots,j_K}} \in \mathbb{S}^{K-1}\}$ for any given $(j_1,\ldots,j_K) \in [pd]^K$, and $|\tilde{\mathcal{S}}_\epsilon^{(j_1,\ldots,j_K)}| = N_{\mathbb{S}^{K-1},\epsilon}$. Furthermore, we know $\mathcal{F}_\epsilon = \bigcup_{j_1=1}^{pd}\cdots\bigcup_{j_K=1}^{pd}\tilde{\mathcal{S}}_\epsilon^{(j_1,\ldots,j_K)} \subset \mathcal{F}$ is an ϵ-net of $\mathcal{F}$ with $|\mathcal{F}_\epsilon|$ satisfying $\epsilon^{-K} \leqslant |\mathcal{F}_\epsilon| \leqslant \{(2+\epsilon)\epsilon^{-1}pd\}^K$. Recall $\mathcal{A}(x;K) = \bigcap_{\mathbf{a}\in\mathcal{F}}\{\mathbf{b}\in\mathbb{R}^{Kpd} : \mathbf{a}^\top\mathbf{b} \leqslant \sqrt{x}\}$. Define $A_1(x) = \bigcap_{\mathbf{a}\in\mathcal{F}_\epsilon}\{\mathbf{b}\in\mathbb{R}^{Kpd} : \mathbf{a}^\top\mathbf{b} \leqslant (1-\epsilon)\sqrt{x}\}$ and $A_2(x) = \bigcap_{\mathbf{a}\in\mathcal{F}_\epsilon}\{\mathbf{b}\in\mathbb{R}^{Kpd} : \mathbf{a}^\top\mathbf{b} \leqslant \sqrt{x}\}$. We can show that $A_1(x) \subset \mathcal{A}(x;K) \subset A_2(x)$. Define

$$\rho_{1,g}(x) := |\mathbb{P}\{\mathbf{v}\in A_1(x)\} - \mathbb{P}\{\mathbf{g}\in A_1(x)\}| \vee |\mathbb{P}\{\mathbf{v}\in A_2(x)\} - \mathbb{P}\{\mathbf{g}\in A_2(x)\}|\,,$$
$$\rho_{2,g}(x) := |\mathbb{P}\{\mathbf{g}\in A_2(x)\} - \mathbb{P}\{\mathbf{g}\in A_1(x)\}|\,.$$

It then holds that

$$\begin{aligned}\mathbb{P}\{\mathbf{v}\in\mathcal{A}(x;K)\} &\leqslant \mathbb{P}\{\mathbf{v}\in A_2(x)\} \leqslant \mathbb{P}\{\mathbf{g}\in A_2(x)\} + \rho_{1,g}(x)\\ &\leqslant \mathbb{P}\{\mathbf{g}\in A_1(x)\} + \rho_{2,g}(x) + \rho_{1,g}(x)\\ &\leqslant \mathbb{P}\{\mathbf{g}\in\mathcal{A}(x;K)\} + \rho_{1,g}(x) + \rho_{2,g}(x)\,.\end{aligned}$$

Analogously, we also have $\mathbb{P}\{\mathbf{v} \in \mathcal{A}(x; K)\} \geqslant \mathbb{P}\{\mathbf{g} \in \mathcal{A}(x; K)\} - \rho_{1,g}(x) - \rho_{2,g}(x)$. Hence, we have

$$|\mathbb{P}\{\mathbf{v} \in \mathcal{A}(x; K)\} - \mathbb{P}\{\mathbf{g} \in \mathcal{A}(x; K)\}| \leqslant \rho_{1,g}(x) + \rho_{2,g}(x) . \tag{26}$$

We set $\epsilon = n^{-1}$ throughout the following arguments. Then $|\mathcal{F}_\epsilon| \geqslant n^K$. Due to $\tau_2 \in (0, 1]$, it holds that $K \lesssim (\log |\mathcal{F}_\epsilon|)^{1/\tau_2}$. Note that $K \lesssim n^{1/9}(\log n)^{1/3}$. By Proposition 3 with $m = K$, $d_z \lesssim (npd)^K$ and $(r_1, r_2) = (\tau_1, \tau_2)$, we have

$$\begin{aligned}\sup_{x>0} \rho_{1,g}(x) =& \sup_{x>0}\left|\mathbb{P}\left(\max_{\mathbf{a}\in\mathcal{F}_\epsilon} \mathbf{a}^\top \mathbf{v} \leqslant x\right) - \mathbb{P}\left(\max_{\mathbf{a}\in\mathcal{F}_\epsilon} \mathbf{a}^\top \mathbf{g} \leqslant x\right)\right| \\ &\lesssim n^{-1/9}K^{5/3}\{\log(npd)\}^{7/6} + n^{-1/9}K^{(1+3\tau_2)/(3\tau_2)}\{\log(npd)\}^{(1+2\tau_2)/(3\tau_2)} ,\end{aligned}$$

provided that $\log(npd) \ll \min\{K^{(\tau-6)/(6+2\tau)}n^{7\tau/(18+6\tau)}, K^{-(6+5\tau_1)/(6+2\tau_1)}n^{7\tau_1/(18+6\tau_1)}, K^{-1}n^{\tau_2/(9-3\tau_2)}\}$. To make $\sup_{x>0}\rho_{1,g}(x) = o(1)$, we need to require $\log(npd) \ll \min\{n^{2/21}K^{-10/7}, n^{\tau_2/(3+6\tau_2)}K^{-(1+3\tau_2)/(1+2\tau_2)}\}$. Notice that $\rho_{2,g}(x) = \mathbb{P}\{(1-\epsilon)\sqrt{x} < \max_{\mathbf{a}\in\mathcal{F}_\epsilon} \mathbf{a}^\top\mathbf{g} \leqslant \sqrt{x}\}$. If $x \leqslant K^3\{\log(npd)\}^3$, by Nazarov's inequality (Lemma A.1, Chernozhukov et al., 2017), we have $\rho_{2,g}(x) \leqslant C\epsilon\sqrt{x\log|\mathcal{F}_\epsilon|} \lesssim n^{-1}K^2\{\log(npd)\}^2$. If $x > K^3\{\log(npd)\}^3$, by Markov inequality, we have

$$\rho_{2,g}(x) \leqslant \mathbb{P}\left\{(1-\epsilon)\sqrt{x} \leqslant \max_{\mathbf{a}\in\mathcal{F}_\epsilon} \mathbf{a}^\top\mathbf{g}\right\} \leqslant \frac{\mathbb{E}(\max_{\mathbf{a}\in\mathcal{F}_\epsilon}|\mathbf{a}^\top\mathbf{g}|)}{(1-\epsilon)K^{3/2}\{\log(npd)\}^{3/2}} \lesssim \{\log(npd)\}^{-1},$$

where the last step is based on Lemma 7.4 in Fan et al. (2018). Hence, $\sup_{x>0}\rho_{2,g}(x) = o(1)$ if $\log(npd) \ll \min\{n^{2/21}K^{-10/7}, n^{\tau_2/(3+6\tau_2)}K^{-(1+3\tau_2)/(1+2\tau_2)}\}$. Due to $|\mathbb{P}(\tilde{T}_n \leqslant x) - \mathbb{P}(G_K \leqslant x)| = |\mathbb{P}\{\mathbf{v} \in \mathcal{A}(x; K)\} - \mathbb{P}\{\mathbf{g} \in \mathcal{A}(x; K)\}|$, (26) implies

$$\sup_{x>0}|\mathbb{P}(\tilde{T}_n \leqslant x) - \mathbb{P}(G_K \leqslant x)| = o(1)$$

provided that $\log(npd) \ll \min\{K^{(\tau-6)/(6+2\tau)}n^{7\tau/(18+6\tau)}, K^{-(6+5\tau_1)/(6+2\tau_1)}n^{7\tau_1/(18+6\tau_1)}, K^{-1}n^{\tau_2/(9-3\tau_2)}, K^{-(1+3\tau_2)/(1+2\tau_2)}n^{\tau_2/(3+6\tau_2)}, K^{-10/7}n^{2/21}\}$.

Step 2. For any $\zeta > 0$, we have

$$\sup_{x>0}|\mathbb{P}(T_n \leqslant x) - \mathbb{P}(G_K \leqslant x)| \leqslant \sup_{x>0}|\mathbb{P}(\tilde{T}_n \leqslant x) - \mathbb{P}(G_K \leqslant x)| + \mathbb{P}(|T_n - \tilde{T}_n| > \zeta) + \sup_{x>0}\mathbb{P}(x-\zeta < G_K \leqslant x+\zeta) . \tag{27}$$

Note that $K = o(n)$. Selecting $\zeta = CK^{3/2}\{\log(npd)\}^{1/2}n^{-1/2}\max[\{\log(npd)\}^{1/\tau}, K\{\log(npd)\}^{1/\tau_1}]$ for some sufficiently large constant $C > 0$, Lemma L2 yields that $\mathbb{P}(|T_n - \tilde{T}_n| > \zeta) = o(1)$. In the sequel, we will consider $\mathbb{P}(x-\zeta < G_K \leqslant x+\zeta)$ under the scenarios $x \leqslant \zeta$ and $x > \zeta$, respectively. Notice that $(1, 0, \ldots, 0)^\top \in \mathcal{F}$ and $(-1, 0, \ldots, 0)^\top \in \mathcal{F}$. Recall that $\mathbf{g} = (g_1, \ldots, g_{Kpd})^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{n,K})$ and $\{G_K \leqslant x\} = \{\max_{\mathbf{a}\in\mathcal{F}} \mathbf{a}^\top\mathbf{g} \leqslant \sqrt{x}\}$ for any $x > 0$. Then we have

$$\begin{aligned}\sup_{x\leqslant\zeta}\mathbb{P}(x-\zeta < G_K \leqslant x+\zeta) &\leqslant \sup_{x\leqslant\zeta}\mathbb{P}(G_K \leqslant x+\zeta) = \sup_{x\leqslant\zeta}\mathbb{P}\left(\max_{\mathbf{a}\in\mathcal{F}}\mathbf{a}^\top\mathbf{g} \leqslant \sqrt{x+\zeta}\right) \\ &\leqslant \sup_{x\leqslant\zeta}\mathbb{P}(-\sqrt{x+\zeta} \leqslant g_1 \leqslant \sqrt{x+\zeta}) \lesssim \sqrt{\zeta} ,\end{aligned} \tag{28}$$

where the last step is due to the anti-concentration inequality of normal random variable. For any $x > \zeta$, it holds that

$$\begin{aligned}\mathbb{P}(x-\zeta < G_K \leqslant x+\zeta) &= \mathbb{P}(G_K \leqslant x+\zeta) - \mathbb{P}(G_K \leqslant x-\zeta) \\ &\leqslant \mathbb{P}\left(\max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\mathbf{g} \leqslant \sqrt{x+\zeta}\right) - \mathbb{P}\left\{\max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\mathbf{g} \leqslant (1-\epsilon)\sqrt{x-\zeta}\right\} \\ &\leqslant \mathbb{P}\left(\max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\mathbf{g} \leqslant \sqrt{x}+\sqrt{\zeta}\right) - \mathbb{P}\left\{\max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\mathbf{g} \leqslant (1-\epsilon)(\sqrt{x}-\sqrt{\zeta})\right\} \\ &\leqslant \mathbb{P}\left\{(1-\epsilon)(\sqrt{x}-\sqrt{\zeta}) < \max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\mathbf{g} \leqslant (1-\epsilon)\sqrt{x}\right\} \\ &\quad + \mathbb{P}\left\{(1-\epsilon)\sqrt{x} < \max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\mathbf{g} \leqslant \sqrt{x}+\sqrt{\zeta}\right\} .\end{aligned}$$

Recall $|\mathcal{F}_\epsilon| \leqslant \{(2+\epsilon)\epsilon^{-1}pd\}^K$ with $\epsilon = n^{-1}$. By Nazarov's inequality, we have $\sup_{x>\zeta}\mathbb{P}\{(1-\epsilon)(\sqrt{x}-\sqrt{\zeta}) < \max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\mathbf{g} \leqslant (1-\epsilon)\sqrt{x}\} \lesssim \sqrt{\zeta K\log(npd)}$ and $\sup_{x>\zeta}\mathbb{P}(\sqrt{x} < \max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\mathbf{g} \leqslant \sqrt{x}+\sqrt{\zeta}) \lesssim \sqrt{\zeta K\log(npd)}$. Due to $\mathbb{P}\{(1-\epsilon)\sqrt{x} < \max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\mathbf{g} \leqslant \sqrt{x}+\sqrt{\zeta}\} = \rho_{2,g}(x) + \mathbb{P}(\sqrt{x} < \max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\mathbf{g} \leqslant \sqrt{x}+\sqrt{\zeta})$, together with (28), we have

$$\sup_{x>0}\mathbb{P}(x-\zeta < G_K \leqslant x+\zeta) \lesssim \sup_{x>0}\rho_{2,g}(x) + \sqrt{\zeta K\log(npd)} = o(1) + \sqrt{\zeta K\log(npd)} .$$

If $\log(npd) \ll \min\{K^{-5\tau/(3\tau+2)}n^{\tau/(3\tau+2)}, K^{-7\tau_1/(3\tau_1+2)}n^{\tau_1/(3\tau_1+2)}\}$, then $\zeta K\log(npd) = o(1)$. By (27), to make $\sup_{x>0}|\mathbb{P}(T_n \leqslant x) - \mathbb{P}(G_K \leqslant x)| = o(1)$, we need to require $K \lesssim n^{1/9}(\log n)^{1/3}$ and

$$\log(npd) \ll \begin{cases} K^{-(6-\tau)/(6+2\tau)}n^{7\tau/(18+6\tau)}, \\ K^{-(6+5\tau_1)/(6+2\tau_1)}n^{7\tau_1/(18+6\tau_1)}, \\ K^{-1}n^{\tau_2/(9-3\tau_2)}, \\ K^{-10/7}n^{2/21}, \\ K^{-(1+3\tau_2)/(1+2\tau_2)}n^{\tau_2/(3+6\tau_2)}, \\ K^{-5\tau/(3\tau+2)}n^{\tau/(3\tau+2)}, \\ K^{-7\tau_1/(3\tau_1+2)}n^{\tau_1/(3\tau_1+2)}. \end{cases}$$

Due to $\log(npd) \to \infty$ as $n \to \infty$, K should satisfy the restriction $K \ll n^{f_1(\tau_1,\tau_2)}$ with $f_1(\tau_1,\tau_2)$ specified in (14). If $K = O(n^\delta)$ for some constant $0 \leqslant \delta < f_1(\tau_1,\tau_2)$, there exists a constant $c > 0$ only depending on (τ_1,τ_2,δ) such that $\sup_{x>0}|\mathbb{P}(T_n \leqslant x) - \mathbb{P}(G_K \leqslant x)| = o(1)$ provided that $\log(pd) \ll n^c$. □

8.3. Proof of Proposition 2

Write $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_{Kpd})^\top = \tilde{n}^{-1}\sum_{t=1}^{\tilde{n}}\mathbb{E}(\boldsymbol{\eta}_t)$. Define $\boldsymbol{\Sigma}_{n,K}^* = \sum_{j=-\tilde{n}+1}^{\tilde{n}-1}\mathcal{K}(j/b_n)\mathbf{H}_j$, where $\mathbf{H}_j = \tilde{n}^{-1}\sum_{t=j+1}^{\tilde{n}}\mathbb{E}\{(\boldsymbol{\eta}_t - \boldsymbol{\mu})(\boldsymbol{\eta}_{t-j} - \boldsymbol{\mu})^\top\}$ if $j \geqslant 0$ and $\mathbf{H}_j = \tilde{n}^{-1}\sum_{t=-j+1}^{\tilde{n}}\mathbb{E}\{(\boldsymbol{\eta}_{t+j} - \boldsymbol{\mu})(\boldsymbol{\eta}_t - \boldsymbol{\mu})^\top\}$ if $j < 0$. By the triangle inequality, we have

$$|\widehat{\boldsymbol{\Sigma}}_{n,K} - \boldsymbol{\Sigma}_{n,K}|_\infty \leqslant |\widehat{\boldsymbol{\Sigma}}_{n,K} - \boldsymbol{\Sigma}_{n,K}^*|_\infty + |\boldsymbol{\Sigma}_{n,K}^* - \boldsymbol{\Sigma}_{n,K}|_\infty.$$

Let $\tau_* = (\tau_1\tau_2)/(\tau_1 + 2\tau_2)$. As we will show later in Sections 8.3.1 and 8.3.2, $|\boldsymbol{\Sigma}_{n,K}^* - \boldsymbol{\Sigma}_{n,K}|_\infty \lesssim n^{-\rho}K^2$, and

$$|\widehat{\boldsymbol{\Sigma}}_{n,K} - \boldsymbol{\Sigma}_{n,K}^*|_\infty = O_{\rm p}\bigg[\frac{\{\log(npd)\}^{(2+\tau_1\vartheta-\tau_1)/(2\tau_1\vartheta-\tau_1)}}{n^{(2\rho+\vartheta-1-3\rho\vartheta)/(2\vartheta-1)}}\bigg] + O_{\rm p}\bigg[\frac{\{\log(npd)\}^{2/\tau_1}}{n^{(\rho+\vartheta-2\rho\vartheta-1)/\vartheta}}\bigg] + O_{\rm p}\bigg[\frac{\{\log(npd)\}^{1/\tau_*}}{n^{1-\rho}}\bigg]$$

provided that $K \lesssim n^{(2\rho\vartheta+1-2\rho)/(2\vartheta-1)}\{\log(npd)\}^{(4-\tau_1)/(2\tau_1\vartheta-\tau_1)} \wedge n^{(1-\rho+\rho\vartheta)/\vartheta}$. Therefore, $K^3\{\log(npd)\}^2|\widehat{\boldsymbol{\Sigma}}_{n,K} - \boldsymbol{\Sigma}_{n,K}|_\infty = o_{\rm p}(1)$ provided that $0 < \rho < (\vartheta-1)/(3\vartheta-2)$ and

$$\log(npd) \ll \begin{cases} K^{-5/2}n^{\rho/2}, \\ \{K^{-(6\vartheta-3)}n^{(2\rho+\vartheta-1-3\rho\vartheta)}\}^{\tau_1/(2+5\tau_1\vartheta-3\tau_1)}, \\ \{K^{-3\vartheta}n^{\rho+\vartheta-2\rho\vartheta-1}\}^{\tau_1/(2\vartheta+2\tau_1\vartheta)}, \\ K^{-3\tau_*/(1+2\tau_*)}n^{(\tau_*-\rho\tau_*)/(1+2\tau_*)}. \end{cases}$$

Due to $\log(npd) \to \infty$ as $n \to \infty$, K should satisfy the restriction $K \ll n^{f_2(\rho,\vartheta)}$ with $f_2(\rho,\vartheta)$ specified in (15). If $K = O(n^\delta)$ for some constant $0 \leqslant \delta < f_2(\rho,\vartheta)$, there exists a constant $c > 0$ only depending on $(\tau_1,\tau_2,\rho,\vartheta,\delta)$ such that $K^3\{\log(npd)\}^2|\widehat{\boldsymbol{\Sigma}}_{n,K} - \boldsymbol{\Sigma}_{n,K}|_\infty = o_{\rm p}(1)$ provided that $\log(pd) \ll n^c$. □

8.3.1. Convergence rate of $|\widehat{\boldsymbol{\Sigma}}_{n,K} - \boldsymbol{\Sigma}_{n,K}^*|_\infty$

Without loss of generality, we can assume $\boldsymbol{\mu} = \mathbf{0}$. Recall that $\widehat{\boldsymbol{\Sigma}}_{n,K} = \sum_{j=-\tilde{n}+1}^{\tilde{n}-1}\mathcal{K}(j/b_n)\widehat{\mathbf{H}}_j$, where $\widehat{\mathbf{H}}_j = \tilde{n}^{-1}\sum_{t=j+1}^{\tilde{n}}(\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}})(\boldsymbol{\eta}_{t-j} - \bar{\boldsymbol{\eta}})^\top$ if $j \geqslant 0$, $\widehat{\mathbf{H}}_j = \tilde{n}^{-1}\sum_{t=-j+1}^{\tilde{n}}(\boldsymbol{\eta}_{t+j} - \bar{\boldsymbol{\eta}})(\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}})^\top$ otherwise, $\tilde{n} = n - K$ and $\bar{\boldsymbol{\eta}} = \tilde{n}^{-1}\sum_{t=1}^{\tilde{n}}\boldsymbol{\eta}_t$. By the triangle inequality, it holds that

$$\begin{aligned}\bigg|\sum_{j=0}^{\tilde{n}-1}\mathcal{K}\Big(\frac{j}{b_n}\Big)(\widehat{\mathbf{H}}_j - \mathbf{H}_j)\bigg|_\infty &\leqslant \underbrace{\bigg|\sum_{j=0}^{\tilde{n}-1}\mathcal{K}\Big(\frac{j}{b_n}\Big)\bigg[\frac{1}{\tilde{n}}\sum_{t=j+1}^{\tilde{n}}\{\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-j}^\top - \mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-j}^\top)\}\bigg]\bigg|_\infty}_{\rm I} \\ &\quad + \underbrace{\bigg|\sum_{j=0}^{\tilde{n}-1}\mathcal{K}\Big(\frac{j}{b_n}\Big)\Big(\frac{1}{\tilde{n}}\sum_{t=j+1}^{\tilde{n}}\boldsymbol{\eta}_t\Big)\bar{\boldsymbol{\eta}}^\top\bigg|_\infty}_{\rm II} + \underbrace{\bigg|\sum_{j=0}^{\tilde{n}-1}\mathcal{K}\Big(\frac{j}{b_n}\Big)\bar{\boldsymbol{\eta}}\Big(\frac{1}{\tilde{n}}\sum_{t=j+1}^{\tilde{n}}\boldsymbol{\eta}_{t-j}\Big)^\top\bigg|_\infty}_{\rm III} \\ &\quad + \underbrace{\bigg|\sum_{j=0}^{\tilde{n}-1}\Big(\frac{\tilde{n}-j}{\tilde{n}}\Big)\mathcal{K}\Big(\frac{j}{b_n}\Big)\bar{\boldsymbol{\eta}}^{\otimes 2}\bigg|_\infty}_{\rm IV}.\end{aligned}$$

In the sequel, we will specify the convergence rates of I, II, III and IV respectively. Recall $\boldsymbol{\eta}_t = (\eta_{t,1},\ldots,\eta_{t,Kpd})^\top$.

Convergence rate of I. Given $\ell_1,\ell_2\in[Kpd]$, we define $\psi_{t,j}=\eta_{t+j,\ell_1}\eta_{t,\ell_2}-\mathbb{E}(\eta_{t+j,\ell_1}\eta_{t,\ell_2})$. For any $M=o(n)\to\infty$ satisfying $M\gtrsim K$ and $b_n=o(M)$, we have

$$\begin{aligned}&\mathbb{P}\bigg(\bigg|\sum_{j=0}^{\tilde n-1}\mathcal{K}\bigg(\frac{j}{b_n}\bigg)\bigg[\frac{1}{\tilde n}\sum_{t=j+1}^{\tilde n}\{\eta_{t,\ell_1}\eta_{t-j,\ell_2}-\mathbb{E}(\eta_{t,\ell_1}\eta_{t-j,\ell_2})\}\bigg]\bigg|>x\bigg)\\&\quad\leqslant\mathbb{P}\bigg\{\sum_{j=0}^{M}\bigg|\mathcal{K}\bigg(\frac{j}{b_n}\bigg)\bigg|\bigg|\frac{1}{\tilde n}\sum_{t=1}^{\tilde n-j}\psi_{t,j}\bigg|>\frac{x}{2}\bigg\}+\mathbb{P}\bigg\{\sum_{j=M+1}^{\tilde n-1}\bigg|\mathcal{K}\bigg(\frac{j}{b_n}\bigg)\bigg|\bigg|\frac{1}{\tilde n}\sum_{t=1}^{\tilde n-j}\psi_{t,j}\bigg|>\frac{x}{2}\bigg\}\end{aligned}\tag{29}$$

for any $x>0$. Lemma 2 of Chang et al. (2013) yields $\max_{0\leqslant j\leqslant\tilde n-1}\max_{t\in[\tilde n-j]}\mathbb{P}(|\psi_{t,j}|>x)\leqslant C\exp(-Cx^{\tau_1/2})$ for any $x>0$. By Condition 4 and $b_n\asymp n^{\rho}$ for some $\rho\in(0,1)$, we have $\sum_{j=M+1}^{\tilde n-1}\mathcal{K}(j/b_n)\lesssim\sum_{j=M+1}^{\tilde n-1}(j/b_n)^{-\vartheta}\lesssim n^{\rho\vartheta}M^{1-\vartheta}$. Analogous to Lemma 4 of Chang et al. (2018b), we can show that

$$\begin{aligned}\mathbb{P}\bigg\{\sum_{j=M+1}^{\tilde n-1}\bigg|\mathcal{K}\bigg(\frac{j}{b_n}\bigg)\bigg|\bigg|\frac{1}{\tilde n}\sum_{t=1}^{\tilde n-j}\psi_{t,j}\bigg|>\frac{x}{2}\bigg\}&\leqslant\sum_{j=M+1}^{\tilde n-1}\mathbb{P}\bigg(\bigg|\frac{1}{\tilde n}\sum_{t=1}^{\tilde n-j}\psi_{t,j}\bigg|>\frac{CM^{\vartheta-1}x}{n^{\rho\vartheta}}\bigg)\\&\leqslant\sum_{j=M+1}^{\tilde n-1}\sum_{t=1}^{\tilde n-j}\mathbb{P}\bigg(|\psi_{t,j}|>\frac{CM^{\vartheta-1}x}{n^{\rho\vartheta}}\bigg)\leqslant Cn^2\exp\bigg\{-\frac{CM^{\tau_1(\vartheta-1)/2}x^{\tau_1/2}}{n^{\rho\vartheta\tau_1/2}}\bigg\}\end{aligned}\tag{30}$$

for any $x>0$. Write $D_n=\sum_{j=0}^{M}|\mathcal{K}(j/b_n)|$ and $\tau_*=(\tau_1\tau_2)/(\tau_1+2\tau_2)$. It is easy to see $D_n\lesssim b_n\asymp n^{\rho}$. For each given j, we observe that $\{\psi_{t,j}\}$ is also an α-mixing sequence and its α-mixing coefficients $\tilde\alpha_{\psi_{t,j}}(k)\leqslant\tilde\alpha_K(|k-j|_+)\leqslant C_3\exp(-C_4|k-j-K|_+^{\tau_2})$, where $\tilde\alpha_K(\cdot)$ is the α-mixing coefficients of the process $\{\boldsymbol{\eta}_t\}$ defined in (23). By Bonferroni inequality and Lemma L1 with $q=\tilde n-j$, $m=j+K$, $r_1=\tau_1/2$, $r_2=\tau_2$ and $r=\tau_*$, we have

$$\begin{aligned}\mathbb{P}\bigg\{\sum_{j=0}^{M}\bigg|\mathcal{K}\bigg(\frac{j}{b_n}\bigg)\bigg|\bigg|\frac{1}{\tilde n}\sum_{t=1}^{\tilde n-j}\psi_{t,j}\bigg|>\frac{x}{2}\bigg\}&\leqslant\sum_{j=0}^{M}\mathbb{P}\bigg(\bigg|\frac{1}{\tilde n}\sum_{t=1}^{\tilde n-j}\psi_{t,j}\bigg|>\frac{x}{2D_n}\bigg)\\&\lesssim M\exp\bigg(-\frac{Cn^{1-2\rho}x^2}{M}\bigg)+\frac{Mn^{\rho}}{x}\bigg[\exp\{-Cn^{(1-\rho)\tau_*}x^{\tau_*}\}+\exp\bigg\{-\frac{Cn^{(1-\rho)\tau_1/2}x^{\tau_1/2}}{M^{\tau_1/2}}\bigg\}\bigg]\end{aligned}$$

for any $x>0$. Together with (29) and (30), it holds that

$$\begin{aligned}\mathbb{P}(\mathrm{I}>x)&\lesssim\sum_{\ell_1\in[Kpd]}\sum_{\ell_2\in[Kpd]}\mathbb{P}\bigg(\bigg|\sum_{j=0}^{\tilde n-1}\mathcal{K}\bigg(\frac{j}{b_n}\bigg)\bigg[\frac{1}{\tilde n}\sum_{t=j+1}^{\tilde n}\{\eta_{t,\ell_1}\eta_{t-j,\ell_2}-\mathbb{E}(\eta_{t,\ell_1}\eta_{t-j,\ell_2})\}\bigg]\bigg|>x\bigg)\\&\lesssim(Kpdn)^2\exp\bigg\{-\frac{CM^{\tau_1(\vartheta-1)/2}x^{\tau_1/2}}{n^{\rho\vartheta\tau_1/2}}\bigg\}+M(Kpd)^2\exp\bigg(-\frac{Cn^{1-2\rho}x^2}{M}\bigg)\\&\quad+\frac{Mn^{\rho}(Kpd)^2}{x}\bigg[\exp\{-Cn^{(1-\rho)\tau_*}x^{\tau_*}\}+\exp\bigg\{-\frac{Cn^{(1-\rho)\tau_1/2}x^{\tau_1/2}}{M^{\tau_1/2}}\bigg\}\bigg]\end{aligned}$$

for any $x>0$, which implies that

$$\mathrm{I}=O_{\mathrm{p}}\bigg[\frac{n^{\rho\vartheta}\{\log(npd)\}^{2/\tau_1}}{M^{\vartheta-1}}\bigg]+O_{\mathrm{p}}\bigg[\frac{M^{1/2}\{\log(npd)\}^{1/2}}{n^{(1-2\rho)/2}}\bigg]+O_{\mathrm{p}}\bigg[\frac{M\{\log(npd)\}^{2/\tau_1}}{n^{1-\rho}}\bigg]+O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{1/\tau_*}}{n^{1-\rho}}\bigg].\tag{31}$$

To make I converge as fast as possible, we need to specify the optimal M in (31). If $\log(npd)\leqslant n^{(1-\rho)(\vartheta-1)\tau_1/\{\vartheta(4-\tau_1)\}}$, with selecting $M\asymp n^{(2\rho\vartheta+1-2\rho)/(2\vartheta-1)}\{\log(npd)\}^{(4-\tau_1)/(2\tau_1\vartheta-\tau_1)}$, we have

$$\mathrm{I}=O_{\mathrm{p}}\bigg\{\bigg[\frac{\{\log(npd)\}^{(2+\tau_1\vartheta-\tau_1)/\tau_1}}{n^{2\rho+\vartheta-1-3\rho\vartheta}}\bigg]^{1/(2\vartheta-1)}\bigg\}+O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{1/\tau_*}}{n^{1-\rho}}\bigg].$$

If $\log(npd)>n^{(1-\rho)(\vartheta-1)\tau_1/\{\vartheta(4-\tau_1)\}}$, with selecting $M\asymp n^{(1-\rho+\rho\vartheta)/\vartheta}$, we have

$$\mathrm{I}=O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{2/\tau_1}}{n^{(\rho+\vartheta-2\rho\vartheta-1)/\vartheta}}\bigg]+O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{1/\tau_*}}{n^{1-\rho}}\bigg].$$

Therefore, we can conclude that

$$\mathrm{I}=O_{\mathrm{p}}\bigg\{\bigg[\frac{\{\log(npd)\}^{(2+\tau_1\vartheta-\tau_1)/\tau_1}}{n^{2\rho+\vartheta-1-3\rho\vartheta}}\bigg]^{1/(2\vartheta-1)}\bigg\}+O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{2/\tau_1}}{n^{(\rho+\vartheta-2\rho\vartheta-1)/\vartheta}}\bigg]+O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{1/\tau_*}}{n^{1-\rho}}\bigg]$$

provided that $K\lesssim n^{(2\rho\vartheta+1-2\rho)/(2\vartheta-1)}\{\log(npd)\}^{(4-\tau_1)/(2\tau_1\vartheta-\tau_1)}\wedge n^{(1-\rho+\rho\vartheta)/\vartheta}$.

Convergence rates of II **and** III. Given $\ell_1, \ell_2 \in [Kpd]$, write

$$\mathrm{II}(\ell_1, \ell_2) = \bigg|\sum_{j=0}^{\tilde{n}-1} \mathcal{K}\bigg(\frac{j}{b_n}\bigg)\bigg(\frac{1}{\tilde{n}}\sum_{t=j+1}^{\tilde{n}} \eta_{t,\ell_1}\bigg)\bar{\eta}_{\ell_2}\bigg|.$$

By Bonferroni inequality and the triangle inequality, it holds that

$$\begin{aligned}\mathbb{P}\{\mathrm{II}(\ell_1, \ell_2) > x\} \leqslant \underbrace{\mathbb{P}\bigg[\sum_{j=0}^{\tilde{n}-1}\bigg|\mathcal{K}\bigg(\frac{j}{b_n}\bigg)\bigg|\bigg|\frac{1}{\tilde{n}}\sum_{t=j+1}^{\tilde{n}}\{\eta_{t,\ell_1} - \mathbb{E}(\eta_{t,\ell_1})\}\bigg||\bar{\eta}_{\ell_2}| > \frac{x}{2}\bigg]}_{\mathrm{II}_{1,\ell_1,\ell_2}(x)} \\ + \underbrace{\mathbb{P}\bigg[\sum_{j=0}^{\tilde{n}-1}\bigg|\mathcal{K}\bigg(\frac{j}{b_n}\bigg)\bigg|\bigg|\frac{1}{\tilde{n}}\sum_{t=j+1}^{\tilde{n}}\mathbb{E}(\eta_{t,\ell_1})\bigg||\bar{\eta}_{\ell_2}| > \frac{x}{2}\bigg]}_{\mathrm{II}_{2,\ell_1,\ell_2}(x)}\end{aligned}$$

for any $x > 0$. Note that $\sum_{j=0}^{\tilde{n}-1}|\mathcal{K}(j/b_n)| \lesssim b_n \asymp n^{\rho}$. By Bonferroni inequality, the triangle inequality and Lemma L1, we have

$$\begin{aligned}\mathrm{II}_{1,\ell_1,\ell_2}(x) &\leqslant \sum_{j=0}^{\tilde{n}-1}\mathbb{P}\bigg[\bigg|\frac{1}{\tilde{n}}\sum_{t=j+1}^{\tilde{n}}\{\eta_{t,\ell_1} - \mathbb{E}(\eta_{t,\ell_1})\}\bigg||\bar{\eta}_{\ell_2}| > \frac{Cx}{n^{\rho}}\bigg] \\ &\leqslant \sum_{j=0}^{\tilde{n}-1}\mathbb{P}\bigg[\bigg|\frac{1}{\tilde{n}}\sum_{t=j+1}^{\tilde{n}}\{\eta_{t,\ell_1} - \mathbb{E}(\eta_{t,\ell_1})\}\bigg| > \frac{Cx^{1/2}}{n^{\rho/2}}\bigg] + n\mathbb{P}\bigg(|\bar{\eta}_{\ell_2}| > \frac{Cx^{1/2}}{n^{\rho/2}}\bigg) \\ &\lesssim n\exp\bigg(-\frac{Cn^{1-\rho}x}{K}\bigg) + \frac{n^{\rho/2+1}}{x^{1/2}}\bigg[\exp\{-Cn^{\tau(2-\rho)/2}x^{\tau/2}\} + \exp\bigg\{-\frac{Cn^{\tau_1(2-\rho)/2}x^{\tau_1/2}}{K^{\tau_1}}\bigg\}\bigg]\end{aligned}$$

for any $x > 0$, where $\tau = \tau_1\tau_2/(\tau_1 + \tau_2)$. Condition 1 yields that $\sup_{t\in[\tilde{n}]}\sup_{\ell\in[Kpd]}\mathbb{E}(|\eta_{t,\ell}|) \leqslant C$. Analogously, it holds that

$$\mathrm{II}_{2,\ell_1,\ell_2}(x) \lesssim n\exp\bigg(-\frac{Cn^{1-2\rho}x^2}{K}\bigg) + \frac{n^{1+\rho}}{x}\bigg[\exp\{-Cn^{(1-\rho)\tau}x^{\tau}\} + \exp\bigg\{-\frac{Cn^{(1-\rho)\tau_1}x^{\tau_1}}{K^{\tau_1}}\bigg\}\bigg]$$

for any $x > 0$. Therefore, by Bonferroni inequality, we have

$$\begin{aligned}\mathbb{P}(\mathrm{II} > x) &\leqslant \sum_{\ell_1,\ell_2\in[Kpd]}\mathbb{P}\{\mathrm{II}(\ell_1, \ell_2) > x\} \\ &\lesssim n(Kpd)^2\bigg\{\exp\bigg(-\frac{Cn^{1-\rho}x}{K}\bigg) + \exp\bigg(-\frac{Cn^{1-2\rho}x^2}{K}\bigg)\bigg\} \\ &\quad + \frac{n^{1+\rho}(Kpd)^2}{x}\bigg[\exp\{-Cn^{(1-\rho)\tau}x^{\tau}\} + \exp\bigg\{-\frac{Cn^{(1-\rho)\tau_1}x^{\tau_1}}{K^{\tau_1}}\bigg\}\bigg] \\ &\quad + \frac{n^{\rho/2+1}(Kpd)^2}{x^{1/2}}\bigg[\exp\{-Cn^{\tau(2-\rho)/2}x^{\tau/2}\} + \exp\bigg\{-\frac{Cn^{\tau_1(2-\rho)/2}x^{\tau_1/2}}{K^{\tau_1}}\bigg\}\bigg]\end{aligned}$$

for any $x > 0$, which implies that

$$\begin{aligned}\mathrm{II} = O_{\mathrm{p}}\bigg\{\frac{K\log(npd)}{n^{1-\rho}}\bigg\} + O_{\mathrm{p}}\bigg[\frac{K^{1/2}\{\log(npd)\}^{1/2}}{n^{(1-2\rho)/2}}\bigg] + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{1/\tau}}{n^{1-\rho}}\bigg] + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{2/\tau}}{n^{2-\rho}}\bigg] \\ + O_{\mathrm{p}}\bigg[\frac{K\{\log(npd)\}^{1/\tau_1}}{n^{1-\rho}}\bigg] + O_{\mathrm{p}}\bigg[\frac{K^2\{\log(npd)\}^{2/\tau_1}}{n^{2-\rho}}\bigg].\end{aligned}$$

Note that $M \gtrsim K$, $\tau_1 \in (0, 1]$ and $\tau_* < \tau$ in (31) and $K = o(n)$. Then

$$\mathrm{II} = O_{\mathrm{p}}\bigg\{\bigg[\frac{\{\log(npd)\}^{(2+\tau_1\vartheta-\tau_1)/\tau_1}}{n^{2\rho+\vartheta-1-3\rho\vartheta}}\bigg]^{1/(2\vartheta-1)}\bigg\} + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{2/\tau_1}}{n^{(\rho+\vartheta-2\rho\vartheta-1)/\vartheta}}\bigg] + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{1/\tau_*}}{n^{1-\rho}}\bigg]$$

provided that $K \lesssim n^{(2\rho\vartheta+1-2\rho)/(2\vartheta-1)}\{\log(npd)\}^{(4-\tau_1)/(2\tau_1\vartheta-\tau_1)} \wedge n^{(1-\rho+\rho\vartheta)/\vartheta}$. Similarly, we also have

$$\mathrm{III} = O_{\mathrm{p}}\bigg\{\bigg[\frac{\{\log(npd)\}^{(2+\tau_1\vartheta-\tau_1)/\tau_1}}{n^{2\rho+\vartheta-1-3\rho\vartheta}}\bigg]^{1/(2\vartheta-1)}\bigg\} + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{2/\tau_1}}{n^{(\rho+\vartheta-2\rho\vartheta-1)/\vartheta}}\bigg] + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{1/\tau_*}}{n^{1-\rho}}\bigg]$$

provided that $K \lesssim n^{(2\rho\vartheta+1-2\rho)/(2\vartheta-1)}\{\log(npd)\}^{(4-\tau_1)/(2\tau_1\vartheta-\tau_1)} \wedge n^{(1-\rho+\rho\vartheta)/\vartheta}$.

Convergence rate of IV. Given $\ell_1, \ell_2 \in [Kpd]$, write

$$\mathrm{IV}(\ell_1, \ell_2) = \bigg|\sum_{j=0}^{\tilde{n}-1}\bigg(\frac{\tilde{n}-j}{\tilde{n}}\bigg)\mathcal{K}\bigg(\frac{j}{b_n}\bigg)\bar{\eta}_{\ell_1}\bar{\eta}_{\ell_2}\bigg|.$$

By Bonferroni inequality and the triangle inequality, it holds that

$$\mathbb{P}\{\mathrm{IV}(\ell_1, \ell_2) > x\} \leqslant \mathbb{P}\bigg\{\sum_{j=0}^{\tilde{n}-1}\bigg|\mathcal{K}\bigg(\frac{j}{b_n}\bigg)\bigg||\bar{\eta}_{\ell_1}||\bar{\eta}_{\ell_2}| > x\bigg\}$$

for any $x > 0$. Identical to the arguments for deriving the upper bound of $\mathrm{II}_{1,\ell_1,\ell_2}(x)$, we know the same upper bound also holds for $\mathbb{P}\{\mathrm{IV}(\ell_1, \ell_2) > x\}$. Hence, we have

$$\mathrm{IV} = O_{\mathrm{p}}\bigg\{\bigg[\frac{\{\log(npd)\}^{(2+\tau_1\vartheta-\tau_1)/\tau_1}}{n^{2\rho+\vartheta-1-3\rho\vartheta}}\bigg]^{1/(2\vartheta-1)}\bigg\} + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{2/\tau_1}}{n^{(\rho+\vartheta-2\rho\vartheta-1)/\vartheta}}\bigg] + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{1/\tau_*}}{n^{1-\rho}}\bigg]$$

provided that $K \lesssim n^{(2\rho\vartheta+1-2\rho)/(2\vartheta-1)}\{\log(npd)\}^{(4-\tau_1)/(2\tau_1\vartheta-\tau_1)} \wedge n^{(1-\rho+\rho\vartheta)/\vartheta}$.

Therefore, we can conclude that

$$\begin{aligned}\bigg|\sum_{j=0}^{\tilde{n}-1}\mathcal{K}\bigg(\frac{j}{b_n}\bigg)(\widehat{\mathbf{H}}_j - \mathbf{H}_j)\bigg|_\infty &\leqslant \mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}\\ &= O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{(2+\tau_1\vartheta-\tau_1)/(2\tau_1\vartheta-\tau_1)}}{n^{(2\rho+\vartheta-1-3\rho\vartheta)/(2\vartheta-1)}}\bigg] + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{2/\tau_1}}{n^{(\rho+\vartheta-2\rho\vartheta-1)/\vartheta}}\bigg] + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{1/\tau_*}}{n^{1-\rho}}\bigg].\end{aligned}$$

Identically, we can also show

$$\bigg|\sum_{j=-\tilde{n}+1}^{-1}\mathcal{K}\bigg(\frac{j}{b_n}\bigg)(\widehat{\mathbf{H}}_j - \mathbf{H}_j)\bigg|_\infty = O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{(2+\tau_1\vartheta-\tau_1)/(2\tau_1\vartheta-\tau_1)}}{n^{(2\rho+\vartheta-1-3\rho\vartheta)/(2\vartheta-1)}}\bigg] + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{2/\tau_1}}{n^{(\rho+\vartheta-2\rho\vartheta-1)/\vartheta}}\bigg] + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{1/\tau_*}}{n^{1-\rho}}\bigg].$$

Hence, we have

$$|\widehat{\boldsymbol{\Sigma}}_{n,K} - \boldsymbol{\Sigma}^*_{n,K}|_\infty = O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{(2+\tau_1\vartheta-\tau_1)/(2\tau_1\vartheta-\tau_1)}}{n^{(2\rho+\vartheta-1-3\rho\vartheta)/(2\vartheta-1)}}\bigg] + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{2/\tau_1}}{n^{(\rho+\vartheta-2\rho\vartheta-1)/\vartheta}}\bigg] + O_{\mathrm{p}}\bigg[\frac{\{\log(npd)\}^{1/\tau_*}}{n^{1-\rho}}\bigg]$$

provided that $K \lesssim n^{(2\rho\vartheta+1-2\rho)/(2\vartheta-1)}\{\log(npd)\}^{(4-\tau_1)/(2\tau_1\vartheta-\tau_1)} \wedge n^{(1-\rho+\rho\vartheta)/\vartheta}$. □

8.3.2. Convergence rate of $|\boldsymbol{\Sigma}^*_{n,K} - \boldsymbol{\Sigma}_{n,K}|_\infty$

Note that $\boldsymbol{\Sigma}_{n,K} = \tilde{n}\mathbb{E}\{(\bar{\boldsymbol{\eta}}-\boldsymbol{\mu})(\bar{\boldsymbol{\eta}}-\boldsymbol{\mu})^\top\}$, $\mathbf{H}_j = \tilde{n}^{-1}\sum_{t=j+1}^{\tilde{n}}\mathbb{E}\{(\boldsymbol{\eta}_t-\boldsymbol{\mu})(\boldsymbol{\eta}_{t-j}-\boldsymbol{\mu})^\top\}$ if $j \geqslant 0$ and $\mathbf{H}_j = \tilde{n}^{-1}\sum_{t=-j+1}^{\tilde{n}}\mathbb{E}\{(\boldsymbol{\eta}_{t+j}-\boldsymbol{\mu})(\boldsymbol{\eta}_t-\boldsymbol{\mu})^\top\}$ if $j < 0$, where $\bar{\boldsymbol{\eta}} = \tilde{n}^{-1}\sum_{t=1}^{\tilde{n}}\boldsymbol{\eta}_t$, $\boldsymbol{\mu} = \tilde{n}^{-1}\sum_{t=1}^{\tilde{n}}\mathbb{E}(\boldsymbol{\eta}_t)$ and $\boldsymbol{\eta}_t = (\eta_{t,1},\ldots,\eta_{t,Kpd})^\top$. We write $\boldsymbol{\Sigma}_{n,K} = \{\sigma_{n,K}(\ell_1,\ell_2)\}_{(Kpd)\times(Kpd)}$, $\mathbf{H}_j = \{H_j(\ell_1,\ell_2)\}_{(Kpd)\times(Kpd)}$ and $\mathring{\eta}_{t,\ell} = \eta_{t,\ell} - \mathbb{E}(\eta_{t,\ell})$. For any $\ell_1, \ell_2 \in [Kpd]$, it holds that

$$\begin{aligned}\sigma_{n,K}(\ell_1,\ell_2) &= \tilde{n}\mathbb{E}\bigg\{\bigg(\frac{1}{\tilde{n}}\sum_{t=1}^{\tilde{n}}\mathring{\eta}_{t,\ell_1}\bigg)\bigg(\frac{1}{\tilde{n}}\sum_{t=1}^{\tilde{n}}\mathring{\eta}_{t,\ell_2}\bigg)\bigg\}\\ &= \frac{1}{\tilde{n}}\sum_{t=1}^{\tilde{n}}\mathbb{E}(\mathring{\eta}_{t,\ell_1}\mathring{\eta}_{t,\ell_2}) + \frac{1}{\tilde{n}}\sum_{t_1=1}^{\tilde{n}-1}\sum_{j=1}^{\tilde{n}-t_1}\mathbb{E}(\mathring{\eta}_{t_1,\ell_1}\mathring{\eta}_{t_1+j,\ell_2}) + \frac{1}{\tilde{n}}\sum_{t_2=1}^{\tilde{n}-1}\sum_{j=1}^{\tilde{n}-t_2}\mathbb{E}(\mathring{\eta}_{t_2+j,\ell_1}\mathring{\eta}_{t_2,\ell_2})\\ &= H_0(\ell_1,\ell_2) + \sum_{j=1}^{\tilde{n}-1}H_{-j}(\ell_1,\ell_2) + \sum_{j=1}^{\tilde{n}-1}H_j(\ell_1,\ell_2).\end{aligned}$$

By Davydov's inequality, $|H_j(\ell_1,\ell_2)| \leqslant \tilde{n}^{-1}\sum_{t=j+1}^{\tilde{n}}|\mathbb{E}(\mathring{\eta}_{t,\ell_1}\mathring{\eta}_{t-j,\ell_2})| \lesssim \tilde{n}^{-1}(\tilde{n}-j)\exp(-C|j-K|_+^{\tau_2})$ for any $j \geqslant 1$. This bound also holds for $|H_{-j}(\ell_1,\ell_2)|$ with $j \geqslant 1$. Observe that $\boldsymbol{\Sigma}^*_{n,K} := \{\sigma^*_{n,K}(\ell_1,\ell_2)\}_{(Kpd)\times(Kpd)} = \sum_{j=-\tilde{n}+1}^{\tilde{n}-1}\mathcal{K}(j/b_n)\mathbf{H}_j$ and $\mathcal{K}(\cdot)$ is symmetric with $\mathcal{K}(0) = 1$. By the triangle inequality and Condition 4,

$$\begin{aligned}|\sigma^*_{n,K}(\ell_1,\ell_2) - \sigma_{n,K}(\ell_1,\ell_2)| &\leqslant \sum_{j=1}^{\tilde{n}-1}\bigg|\mathcal{K}\bigg(\frac{j}{b_n}\bigg) - 1\bigg|\big\{|H_j(\ell_1,\ell_2)| + |H_{-j}(\ell_1,\ell_2)|\big\}\\ &\lesssim \sum_{j=1}^{\tilde{n}-1}\frac{j(\tilde{n}-j)}{b_n\tilde{n}}\exp(-C|j-K|_+^{\tau_2})\end{aligned}$$

$$\lesssim \frac{1}{b_n}\left[\sum_{j=1}^{K} j + \sum_{j=K+1}^{\tilde{n}-1} j\exp\{-C(j-K)^{\tau_2}\}\right]$$
$$\lesssim b_n^{-1}K^2\,.$$

Thus $|\boldsymbol{\Sigma}_{n,K}^{*} - \boldsymbol{\Sigma}_{n,K}|_\infty \lesssim b_n^{-1}K^2$. □

8.4. Proof of Theorem 1

Recall $G_K = \sum_{j=1}^K \max_{\ell\in\mathcal{L}_j} |g_\ell|^2$ and $\hat{G}_K = \sum_{j=1}^K \max_{\ell\in\mathcal{L}_j} |\hat{g}_\ell|^2$ with $\mathbf{g} = (g_1,\ldots,g_{Kpd})^\top \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{n,K})$ and $\hat{\mathbf{g}} = (\hat{g}_1,\ldots,\hat{g}_{Kpd})^\top \sim \mathcal{N}(\mathbf{0},\widehat{\boldsymbol{\Sigma}}_{n,K})$. As shown in Proposition 1, $\sup_{x>0}|\mathbb{P}(T_n\leqslant x)-\mathbb{P}(G_K\leqslant x)| = o(1)$. Write $\mathcal{X}_n = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$. To construct Theorem 1, it suffices to show $\sup_{x>0}|\mathbb{P}(G_K\leqslant x)-\mathbb{P}(\hat{G}_K\leqslant x\,|\,\mathcal{X}_n)| = o(1)$. Recall $\rho_{2,g}(x) = |\mathbb{P}\{\mathbf{g}\in A_2(x)\} - \mathbb{P}\{\mathbf{g}\in A_1(x)\}|$ for $A_1(x)$ and $A_2(x)$ defined in Section 8.2. Here we also define

$$\rho_{3,g}(x) := |\mathbb{P}\{\mathbf{g}\in A_1(x)\} - \mathbb{P}\{\hat{\mathbf{g}}\in A_1(x)\,|\,\mathcal{X}_n\}| \vee |\mathbb{P}\{\mathbf{g}\in A_2(x)\} - \mathbb{P}\{\hat{\mathbf{g}}\in A_2(x)\,|\,\mathcal{X}_n\}|\,.$$

Identical to the result $\{G_K\leqslant x\} = \{\mathbf{g}\in\mathcal{A}(x;K)\}$ stated in Section 8.2, we also have $\{\hat{G}_K\leqslant x\} = \{\hat{\mathbf{g}}\in\mathcal{A}(x;K)\}$ for any $x>0$, where $\mathcal{A}(x;K)$ is defined in Section 8.2. Then it holds that

$$\begin{aligned}\mathbb{P}(\hat{G}_K\leqslant x\,|\,\mathcal{X}_n) &= \mathbb{P}\{\hat{\mathbf{g}}\in\mathcal{A}(x;K)\,|\,\mathcal{X}_n\} \leqslant \mathbb{P}\{\hat{\mathbf{g}}\in A_2(x)\,|\,\mathcal{X}_n\}\\ &\leqslant \mathbb{P}\{\mathbf{g}\in A_2(x)\} + \rho_{3,g}(x)\\ &\leqslant \mathbb{P}\{\mathbf{g}\in A_1(x)\} + \rho_{2,g}(x) + \rho_{3,g}(x)\\ &\leqslant \mathbb{P}\{\mathbf{g}\in \mathcal{A}(x;K)\} + \rho_{2,g}(x) + \rho_{3,g}(x)\\ &\leqslant \mathbb{P}(G_K\leqslant x) + \rho_{2,g}(x) + \rho_{3,g}(x)\end{aligned}$$

for any $x>0$. Similarly, we can also obtain the reverse inequality. Notice that we have shown in Section 8.2 that $\sup_{x>0}\rho_{2,g}(x) = o(1)$. Therefore,

$$\sup_{x>0}|\mathbb{P}(G_K\leqslant x) - \mathbb{P}(\hat{G}_K\leqslant x\,|\,\mathcal{X}_n)| \leqslant \sup_{x>0}\rho_{2,g}(x) + \sup_{x>0}\rho_{3,g}(x) = o(1) + \sup_{x>0}\rho_{3,g}(x).$$

By Lemma 13 of Chang et al. (2021b), it holds that

$$\begin{aligned}&\sup_{x>0}|\mathbb{P}\{\mathbf{g}\in A_1(x)\} - \mathbb{P}\{\hat{\mathbf{g}}\in A_1(x)\,|\,\mathcal{X}_n\}|\\ &= \sup_{x>0}\left|\mathbb{P}\left\{\max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\mathbf{g}\leqslant(1-\epsilon)\sqrt{x}\right\} - \mathbb{P}\left\{\max_{\mathbf{a}\in\mathcal{F}_\epsilon}\mathbf{a}^\top\hat{\mathbf{g}}\leqslant(1-\epsilon)\sqrt{x}\,\bigg|\,\mathcal{X}_n\right\}\right|\\ &\lesssim \Delta_n^{1/3}\{K\log(npd)\}^{2/3}\end{aligned}$$

with $\Delta_n = \max_{\mathbf{a}_1,\mathbf{a}_2\in\mathcal{F}}|\mathbf{a}_1^\top(\boldsymbol{\Sigma}_{n,K}-\widehat{\boldsymbol{\Sigma}}_{n,K})\mathbf{a}_2|$, where $\mathcal{F}$ is defined in Section 8.2. Recall $|\mathbf{a}|_0\leqslant K$ and $|\mathbf{a}|_2 = 1$ for any $\mathbf{a}\in\mathcal{F}$. Thus, $|\mathbf{a}_1^\top(\boldsymbol{\Sigma}_{n,K}-\widehat{\boldsymbol{\Sigma}}_{n,K})\mathbf{a}_2| \leqslant |\mathbf{a}_1|_1|\mathbf{a}_2|_1|\boldsymbol{\Sigma}_{n,K}-\widehat{\boldsymbol{\Sigma}}_{n,K}|_\infty \leqslant K|\boldsymbol{\Sigma}_{n,K}-\widehat{\boldsymbol{\Sigma}}_{n,K}|_\infty$. Then we have $\sup_{x>0}|\mathbb{P}\{\mathbf{g}\in A_1(x)\} - \mathbb{P}\{\hat{\mathbf{g}}\in A_1(x)\,|\,\mathcal{X}_n\}| \lesssim K|\boldsymbol{\Sigma}_{n,K}-\widehat{\boldsymbol{\Sigma}}_{n,K}|_\infty^{1/3}\{\log(npd)\}^{2/3}$. Analogously, we also have $\sup_{x>0}|\mathbb{P}\{\mathbf{g}\in A_2(x)\} - \mathbb{P}\{\hat{\mathbf{g}}\in A_2(x)\,|\,\mathcal{X}_n\}| \lesssim K|\boldsymbol{\Sigma}_{n,K}-\widehat{\boldsymbol{\Sigma}}_{n,K}|_\infty^{1/3}\{\log(npd)\}^{2/3}$. Hence,

$$\sup_{x>0}|\mathbb{P}(G_K\leqslant x) - \mathbb{P}(\hat{G}_K\leqslant x\,|\,\mathcal{X}_n)| \lesssim K|\boldsymbol{\Sigma}_{n,K}-\widehat{\boldsymbol{\Sigma}}_{n,K}|_\infty^{1/3}\{\log(npd)\}^{2/3} + o(1)\,. \tag{32}$$

By Proposition 2, we complete the proof. □

8.5. Proof of Theorem 2

Recall that $\mathcal{X}_n = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ and $\hat{G}_K = \sum_{j=1}^K\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell|^2$ with $\hat{\mathbf{g}} = (\hat{g}_1,\ldots,\hat{g}_{Kpd})^\top$. By Bonferroni inequality, we have

$$\mathbb{P}(\hat{G}_K > x\,|\,\mathcal{X}_n) \leqslant \sum_{j=1}^K\mathbb{P}\left(\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell|^2 > \frac{x}{K}\,\bigg|\,\mathcal{X}_n\right) = \sum_{j=1}^K\mathbb{P}\left(\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell| > \frac{x^{1/2}}{K^{1/2}}\,\bigg|\,\mathcal{X}_n\right)$$

for any $x>0$. Since $\hat{\mathbf{g}}\sim\mathcal{N}(\mathbf{0},\widehat{\boldsymbol{\Sigma}}_{n,K})$ with $\widehat{\boldsymbol{\Sigma}}_{n,K} = \{\hat{\sigma}_{n,K}(\ell_1,\ell_2)\}_{Kpd\times Kpd}$, then

$$\mathbb{E}\left(\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell|\,\bigg|\,\mathcal{X}_n\right) \leqslant [1+\{2\log(pd)\}^{-1}]\{2\log(pd)\}^{1/2}\max_{\ell\in\mathcal{L}_j}\{\hat{\sigma}_{n,K}(\ell,\ell)\}^{1/2}$$

for any $j\in[K]$. Recall $\boldsymbol{\Sigma}_{n,K} = \{\sigma_{n,K}(\ell_1,\ell_2)\}_{Kpd\times Kpd}$ and $\varrho = \max_{\ell\in[Kpd]}\sigma_{n,K}(\ell,\ell)$. Define an event

$$\mathcal{E}_0(\nu) = \left\{\max_{\ell\in[Kpd]}\left|\frac{\hat{\sigma}_{n,K}(\ell,\ell)}{\sigma_{n,K}(\ell,\ell)} - 1\right| \leqslant \nu\right\},$$

where $\nu > 0$ and $\nu \asymp \{K\log(pd)\}^{-1}$. As shown in Proposition 2, $\max_{\ell\in[Kpd]}|\hat{\sigma}_{n,K}(\ell,\ell)-\sigma_{n,K}(\ell,\ell)| = o_{\rm p}[K^{-3}\{\log(npd)\}^{-2}] = o_{\rm p}(\nu)$. From Condition 3, we have $\min_{\ell\in[Kpd]}\sigma_{n,K}(\ell,\ell)\geqslant C$, where C is a positive constant. It holds that

$$\max_{\ell\in[Kpd]}\left|\frac{\hat{\sigma}_{n,K}(\ell,\ell)}{\sigma_{n,K}(\ell,\ell)}-1\right| \leqslant \frac{\max_{\ell\in[Kpd]}|\hat{\sigma}_{n,K}(\ell,\ell)-\sigma_{n,K}(\ell,\ell)|}{\min_{\ell\in[Kpd]}\sigma_{n,K}(\ell,\ell)} = o_{\rm p}(\nu).$$

Thus $\mathbb{P}\{\mathcal{E}_0(\nu)^{\rm c}\}\to 0$ as $n\to\infty$. Restricted on $\mathcal{E}_0(\nu)$, it holds that

$$\max_{j\in[K]}\mathbb{E}\bigg(\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell|\,\bigg|\,\mathcal{X}_n\bigg) \leqslant (1+\nu)^{1/2}\varrho^{1/2}\big[1+\{2\log(pd)\}^{-1}\big]\{2\log(pd)\}^{1/2}.$$

By Borell inequality for Gaussian process, it holds that

$$\mathbb{P}\bigg\{\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell| \geqslant \mathbb{E}\bigg(\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell|\,\bigg|\,\mathcal{X}_n\bigg)+x\,\bigg|\,\mathcal{X}_n\bigg\} \leqslant 2\exp\bigg\{-\frac{x^2}{2\max_{\ell\in\mathcal{L}_j}\hat{\sigma}_{n,K}(\ell,\ell)}\bigg\}$$

for any $x>0$. Let $x_* = K(1+\nu)\varrho([1+\{2\log(pd)\}^{-1}]\{2\log(pd)\}^{1/2}+\{2\log(4K/\alpha)\}^{1/2})^2$. Restricted on $\mathcal{E}_0(\nu)$, we have

$$\frac{x_*^{1/2}}{K^{1/2}} \geqslant \max_{j\in[K]}\mathbb{E}\bigg(\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell|\,\bigg|\,\mathcal{X}_n\bigg)+(1+\nu)^{1/2}\varrho^{1/2}\bigg\{2\log\bigg(\frac{4K}{\alpha}\bigg)\bigg\}^{1/2},$$

which yields that

$$\begin{aligned}
\mathbb{P}\{\hat{G}_K > x_*,\ \mathcal{E}_0(\nu)\,|\,\mathcal{X}_n\} &\leqslant \sum_{j=1}^K\mathbb{P}\bigg\{\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell|-\mathbb{E}\bigg(\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell|\,\bigg|\,\mathcal{X}_n\bigg) > \frac{x_*^{1/2}}{K^{1/2}}-\mathbb{E}\bigg(\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell|\,\bigg|\,\mathcal{X}_n\bigg),\ \mathcal{E}_0(\nu)\,\bigg|\,\mathcal{X}_n\bigg\}\\
&\leqslant \sum_{j=1}^K\mathbb{P}\bigg[\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell|-\mathbb{E}\bigg(\max_{\ell\in\mathcal{L}_j}|\hat{g}_\ell|\,\bigg|\,\mathcal{X}_n\bigg) > (1+\nu)^{1/2}\varrho^{1/2}\bigg\{2\log\bigg(\frac{4K}{\alpha}\bigg)\bigg\}^{1/2},\ \mathcal{E}_0(\nu)\,\bigg|\,\mathcal{X}_n\bigg]\\
&\leqslant 2K\exp\bigg\{-\frac{2(1+\nu)\varrho\log(4K/\alpha)}{2(1+\nu)\varrho}\bigg\} = \frac{\alpha}{2}\,.
\end{aligned}$$

Since $\mathbb{P}\{\mathcal{E}_0(\nu)^{\rm c}\,|\,\mathcal{X}_n\} = o_{\rm p}(1)$, then $\mathbb{P}\{\mathcal{E}_0(\nu)^{\rm c}\,|\,\mathcal{X}_n\}\leqslant\alpha/4$ with probability approaching one. Hence, $\mathbb{P}(\hat{G}_K > x_*\,|\,\mathcal{X}_n)\leqslant 5\alpha/6$ with probability approaching one. Following the definition of $\hat{\rm cv}_\alpha$, it holds with probability approaching one that

$$\hat{\rm cv}_\alpha \leqslant (1+\nu)K\varrho\lambda^2(K,p,d,\alpha)\big[1+\{2\log(pd)\}^{-1}\big]^2 \tag{33}$$

with $\lambda(K,p,d,\alpha) = \{2\log(pd)\}^{1/2}+\{2\log(4K/\alpha)\}^{1/2}$.

We next specify the lower bound of T_n. Recall that $T_n = n\sum_{j=1}^K|\hat{\boldsymbol\gamma}_j|_\infty^2 = \sum_{j=1}^K\max_{\ell\in\mathcal{L}_j}(n^{1/2}u_\ell)^2$, where $\mathbf{u} = (u_1,\ldots,u_{Kpd})^\top = (\hat{\boldsymbol\gamma}_1^\top,\ldots,\hat{\boldsymbol\gamma}_K^\top)^\top$ with $\hat{\boldsymbol\gamma}_j = (n-j)^{-1}\sum_{t=1}^{n-j}{\rm vec}\{\boldsymbol\phi(\mathbf{x}_t)\mathbf{x}_{t+j}^\top\}$. Let $\tilde{\mathbf{u}} = (\tilde{u}_1,\ldots,\tilde{u}_{Kpd})^\top = (\boldsymbol\gamma_1^\top,\ldots,\boldsymbol\gamma_K^\top)^\top$ with $\boldsymbol\gamma_j = (n-j)^{-1}\sum_{t=1}^{n-j}\mathbb{E}[{\rm vec}\{\boldsymbol\phi(\mathbf{x}_t)\mathbf{x}_{t+j}^\top\}]$. Define $\ell_j^* = \arg\max_{\ell\in\mathcal{L}_j}|\tilde{u}_\ell|$ for $j\in[K]$. By Cauchy–Schwarz inequality, it holds that

$$\begin{aligned}
T_n = \sum_{j=1}^K\max_{\ell\in\mathcal{L}_j}(n^{1/2}u_\ell)^2 &\geqslant \sum_{j=1}^K(n^{1/2}u_{\ell_j^*})^2 = \sum_{j=1}^K\big(n^{1/2}u_{\ell_j^*}-n^{1/2}\tilde{u}_{\ell_j^*}+n^{1/2}\tilde{u}_{\ell_j^*}\big)^2\\
&= n\sum_{j=1}^K(u_{\ell_j^*}-\tilde{u}_{\ell_j^*})^2+n\sum_{j=1}^K\tilde{u}_{\ell_j^*}^2+2n\sum_{j=1}^K\tilde{u}_{\ell_j^*}(u_{\ell_j^*}-\tilde{u}_{\ell_j^*})\\
&\geqslant n\sum_{j=1}^K(u_{\ell_j^*}-\tilde{u}_{\ell_j^*})^2+n\sum_{j=1}^K\tilde{u}_{\ell_j^*}^2-2n\bigg(\sum_{j=1}^K\tilde{u}_{\ell_j^*}^2\bigg)^{1/2}\bigg\{\sum_{j=1}^K(u_{\ell_j^*}-\tilde{u}_{\ell_j^*})^2\bigg\}^{1/2}.
\end{aligned}$$

According to the definition of $\mathbf{u}$ and $\tilde{\mathbf{u}}$, we have $n^{1/2}(u_{\ell_j^*}-\tilde{u}_{\ell_j^*}) = n^{1/2}(n-j)^{-1}\sum_{t=1}^{n-j}[\phi_{l_1^*}(\mathbf{x}_t)x_{t+j,l_2^*}-\mathbb{E}\{\phi_{l_1^*}(\mathbf{x}_t)x_{t+j,l_2^*}\}]$ for some $l_1^*\in[d]$ and $l_2^*\in[p]$. Note that $K\ll n^{1/7}$. By Bonferroni inequality and Lemma L1, it holds that for any $x>0$

$$\begin{aligned}
\mathbb{P}\bigg\{n\sum_{j=1}^K(u_{\ell_j^*}-\tilde{u}_{\ell_j^*})^2 > x\bigg\} &\leqslant \sum_{j=1}^K\mathbb{P}\bigg(\frac{n^{1/2}}{n-j}\bigg|\sum_{t=1}^{n-j}[\phi_{l_1^*}(\mathbf{x}_t)x_{t+j,l_2^*}-\mathbb{E}\{\phi_{l_1^*}(\mathbf{x}_t)x_{t+j,l_2^*}\}]\bigg| > \frac{x^{1/2}}{K^{1/2}}\bigg)\\
&\lesssim \frac{n^{1/2}K^{3/2}}{x^{1/2}}\bigg\{\exp\bigg(-\frac{Cn^{\tau/2}x^{\tau/2}}{K^{\tau/2}}\bigg)+\exp\bigg(-\frac{Cn^{\tau_1/2}x^{\tau_1/2}}{K^{3\tau_1/2}}\bigg)\bigg\}+K\exp\bigg(-\frac{Cx}{K^2}\bigg)
\end{aligned}$$

with $\tau = \tau_1\tau_2/(\tau_1+\tau_2)$, which implies that $n\sum_{j=1}^K(u_{\ell_j^*}-\tilde{u}_{\ell_j^*})^2 = O_{\rm p}(K^2\log K)$. Choose $u>0$ such that $(1+\nu)^{1/2}[1+\{2\log(pd)\}^{-1}+u] = 1+\epsilon_n$ for some $\epsilon_n>0$. Due to $\sum_{j=1}^K\tilde{u}_{\ell_j^*}^2\geqslant n^{-1}K\varrho\lambda^2(K,p,d,\alpha)(1+\epsilon_n)^2$ and $n\sum_{j=1}^K(u_{\ell_j^*}-\tilde{u}_{\ell_j^*})^2 =$

$O_p(K^2 \log K)$, by (33), it holds with probability approaching one that

$$\begin{aligned} T_n &\geqslant n \sum_{j=1}^{K} (u_{\ell_j^*} - \tilde{u}_{\ell_j^*})^2 + (1+\nu) K \varrho \lambda^2(K,p,d,\alpha)[1 + \{2\log(pd)\}^{-1} + u]^2 \\ &\quad - O_p\{K^{3/2}(\log K)^{1/2}\varrho^{1/2}\lambda(K,p,d,\alpha)(1+\epsilon_n)\} \\ &> (1+\nu)K\varrho\lambda^2(K,p,d,\alpha)\big[1+\{2\log(pd)\}^{-1}\big]^2 + 2K\varrho\lambda^2(K,p,d,\alpha)u \\ &\quad - O_p\{K^{3/2}(\log K)^{1/2}\varrho^{1/2}\lambda(K,p,d,\alpha)(1+\epsilon_n)\} \\ &> \hat{\mathrm{cv}}_\alpha + 2K\varrho\lambda^2(K,p,d,\alpha)u - O_p\{K^{3/2}(\log K)^{1/2}\varrho^{1/2}\lambda(K,p,d,\alpha)\}\,. \end{aligned}$$

Notice that $\epsilon_n \to 0$ and $\varrho\lambda^2(K,p,d,\alpha)K^{-1}(\log K)^{-1}\epsilon_n^2 \to \infty$. Then it holds that

$$\epsilon_n \gg \frac{K^{1/2}(\log K)^{1/2}}{\{\log(pd)\}^{1/2} + (\log K)^{1/2}} \gg \frac{1}{\log(pd)} \gtrsim \frac{1}{K\log(pd)} = \nu,$$

which implies that $u \asymp \epsilon_n$. It yields that $K\varrho\lambda^2(K,p,d,\alpha)u \gg K^{3/2}(\log K)^{1/2}\varrho^{1/2}\lambda(K,p,d,\alpha)$ and $K\varrho\lambda^2(K,p,d,\alpha)u \to \infty$. Therefore, we have $T_n - \hat{\mathrm{cv}}_\alpha > K\varrho\lambda^2(K,p,d,\alpha)u$ with probability approaching one. Hence, $\mathbb{P}_{H_1}(T_n > \hat{\mathrm{cv}}_\alpha) \to 1$ as $n \to \infty$. □

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2022.09.001.

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.
Bierens, H.J., 1984. Model specification testing of time series regressions. J. Econometrics 26, 323–353.
Bierens, H.J., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443–1458.
Bierens, H.J., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65, 1129–1151.
Boussama, F., Fuchs, F., Stelzer, R., 2011. Stationary and geometric ergodicity of BEKK multivariate GARCH models. Stochastic Process. Appl. 121, 2331–2360.
Box, G.E.P., Pierce, D.A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J. Amer. Statist. Assoc. 65, 1509–1526.
Cai, T.T., Liu, W., Xia, Y., 2014. Two-sample test of high dimensional means under dependence. J. R. Stat. Soc. Ser. B Stat. Methodol. 76, 349–372.
Chang, J., Chen, S.X., Chen, X., 2015. High dimensional generalized empirical likelihood for moment restrictions with dependent data. J. Econometrics 185, 283–304.
Chang, J., Chen, S.X., Tang, C.Y., Wu, T.T., 2021a. High-dimensional empirical likelihood inference. Biometrika 108, 127–147.
Chang, J., Chen, X., Wu, M., 2021b. Central limit theorems for high dimensional dependent data. arXiv:2104.12929.
Chang, J., Guo, B., Yao, Q., 2018a. Principal component analysis for second-order stationary vector time series. Ann. Statist. 46, 2094–2124.
Chang, J., Qiu, Y., Yao, Q., Zou, T., 2018b. Confidence regions for entries of a large precision matrix. J. Econometrics 206, 57–82.
Chang, J., Tang, C.Y., Wu, Y., 2013. Marginal empirical likelihood and sure independence feature screening. Ann. Statist. 41, 2123–2148.
Chang, J., Tang, C.Y., Wu, T.T., 2018c. A new scope of penalized empirical likelihood with high-dimensional estimating equations. Ann. Statist. 46, 3185–3216.
Chang, J., Yao, Q., Zhou, W., 2017a. Testing for high-dimensional white noise using maximum cross correlations. Biometrika 104, 111–127.
Chang, J., Zheng, C., Zhou, W.-X., Zhou, W., 2017b. Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. Biometrics 73, 1300–1310.
Chang, J., Zhou, W., Zhou, W.-X., Wang, L., 2017c. Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. Biometrics 73, 31–41.
Chen, X., 2018. Gaussian and bootstrap approximations for high-dimensional u-statistics and their applications. Ann. Statist. 46, 642–678.
Chen, W.W., Deo, R.S., 2006. The variance ratio statistic at large horizons. Econom. Theory 22, 206–234.
Chen, X., Kato, K., 2019. Randomized incomplete u-statistics in high dimensions. Ann. Statist. 47, 3127–3156.
Chernozhukov, V., Chetverikov, D., Kato, K., 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. Ann. Statist. 41, 2786–2819.
Chernozhukov, V., Chetverikov, D., Kato, K., 2017. Central limit theorems and bootstrap in high dimensions. Ann. Probab. 45, 2309–2352.
Chernozhukov, V., Chetverikov, D., Kato, K., 2019. Inference on causal and structural parameters using many moment inequalities. Rev. Econom. Stud. 86, 1867–1900.
Chernozhukov, V., Chetverikov, D., Kato, K., Koike, Y., 2022a. Improved central limit theorem and bootstrap approximations in high dimensions. Ann. Statist. in press.
Chernozhukov, V., Chetverikov, D., Koike, Y., 2022b. Nearly optimal central limit theorem and bootstrap approximations in high dimensions. Ann. Appl. Probab. in press.
Cochrane, J.H., 2005. Asset Pricing. Princeton University Press.
de Jong, R.M., 1996. The Bierens tests under data dependence. J. Econometrics 72, 1–32.
Deng, H., Zhang, C.-H., 2020. Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. Ann. Statist. 48, 3643–3671.
Deo, R.S., 2000. Spectral tests of the martingale hypothesis under conditional heteroskedasticity. J. Econometrics 99, 291–315.
Domínguez, M.A., Lobato, I.N., 2003. A consistent test for the martingale difference hypothesis. Econometric Rev. 22, 351–377.
Durlauf, S.N., 1991. Spectral-based test for the martingale hypothesis. J. Econometrics 50, 355–376.
Escanciano, J.C., Lobato, I.N., 2009. Testing the martingale hypothesis. In: Mills, T.C., Patterson, K. (Eds.), Palgrave Handbook of Econometrics. Palgrave Macmillan, London.

Escanciano, J.C., Velasco, C., 2006. Generalized spectral tests for the martingale difference hypothesis. J. Econometrics 134, 151–185.
Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. J. Finance 25, 383–417.
Fama, E.F., 1991. Efficient capital markets: II. J. Finance 46, 1575–1617.
Fama, E.F., 2013. Two Pillars of Asset Pricing. In: Nobel Prize Lecture.
Fan, J., Shao, Q.-M., Zhou, W.-X., 2018. Are discoveries spurious? Distribution of maximum spurious correlations and their applications. Ann. Statist. 46, 989–1017.
Fang, X., Koike, Y., 2021. High-dimensional central limit theorems by Stein's method. Ann. Appl. Probab. 31, 1660–1686.
Hafner, C.M., Preminger, A., 2009. On asymptotic theory for multivariate GARCH models. J. Multivariate Anal. 100, 2044–2054.
Hall, R.E., 1978. Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. J. Polit. Econ. 86, 971–987.
Han, H., Linton, O., Oka, T., Whang, Y.-J., 2016. The cross-quantilogram: measuring quantile dependence and testing directional predictability between time series. J. Econometrics 193, 251–270.
Hong, Y., 1996. Consistent testing for serial correlation of unknown form. Econometrica 64, 837–864.
Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. J. Amer. Statist. Assoc. 94, 1201–1220.
Hong, Y., Lee, T.-H., 2003. Inference on predictability of foreign exchange rate changes via generalized spectrum and nonlinear time series models. Rev. Econ. Stat. 85, 1048–1062.
Hong, Y., Lee, Y.-J., 2005. Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form. Rev. Econom. Stud. 72, 499–541.
Hong, S., Linton, O., Zhang, H., 2017. An investigation into multivariate variance ratio statistics and their application to stock market predictability. J. Financ. Econom. 15, 173–222.
Koul, H.L., Stute, W., 1999. Nonparametric model checks for time series. Ann. Statist. 27, 204–236.
Kuchibhotla, A.K., Mukherjee, S., Banerjee, D., 2021. High-dimensional CLT: Improvements, non-uniform extensions and large deviations. Bernoulli 27, 192–217.
LeRoy, S.F., 1989. Efficient capital markets and martingales. J. Econ. Lit. 27, 1583–1621.
Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 65, 297–303.
Lo, A.W., 1997. Market Efficiency: Stock Market Behaviour in Theory and Practice, Vols. I and II. Edward Elgar.
Lo, A.W., MacKinlay, A.C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. Rev. Financ. Stud. 1, 41–66.
Lobato, I., Nankervis, J.C., Savin, N.E., 2001. Testing for autocorrelation using a modified Box–Pierce q test. Internat. Econom. Rev. 42, 187–205.
Newey, W., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Park, J.Y., Whang, Y.-J., 2005. A test of the martingale hypothesis. Stud. Nonlinear Dyn. Econom. 9, 2.
Phillips, P.C.B., Jin, S., 2014. Testing the martingale hypothesis. J. Bus. Econom. Statist. 32, 537–554.
Poterba, J.M., Summers, L.H., 1988. Mean reversion in stock prices: Evidence and implications. J. Financ. Econ. 22, 27–59.
Shao, X., 2011a. A bootstrap-assisted spectral test of white noise under unknown dependence. J. Econometrics 162, 213–224.
Shao, X., 2011b. Testing for white noise under unknown dependence and its applications to goodness-of-fit for time series models. Econom. Theory 27, 312–343.
Stute, W., 1997. Nonparametric model checks for regression. Ann. Statist. 25, 613–641.
Vershynin, R., 2012. Inroduction to the non-asymptotic analysis of random matrices. In: Eldar, Y.C., Kutyniok, G. (Eds.), Compressed Sensing: Theory and Applications. Cambridge University Press.
Wong, K.C., Li, Z., Tewari, A., 2020. Lasso guarantees for β-mixing heavy-tailed time series. Ann. Statist. 48, 1124–1142.
Wu, W.-B., 2005. Nonlinear system theory: Another look at dependence. Proc. Natl. Acad. Sci. USA 102, 14150–14154.
Yu, M., Chen, X., 2021. Finite sample change point inference and identification for high dimensional mean vectors. J. R. Stat. Soc. Ser. B Stat. Methodol. 83, 247–270.
Zhang, X., Cheng, G., 2018. Gaussian approximation for high dimensional vector under physical dependence. Bernoulli 24, 2640–2675.
Zhang, D., Wu, W.-B., 2017. Gaussian approximation for high dimensional time series. Ann. Statist. 45, 1895–1919.