TEST FOR SERIAL CORRELATION UNDER HIGH DIMENSIONALITY AND IMPROVED CONVERGENCE RATE FOR NORMAL EXTREMES

By

YIJUN ZHU

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Statistics

Written under the direction of

Han Xiao

And approved by

New Brunswick, New Jersey

May, 2020

ABSTRACT OF THE DISSERTATION

TEST FOR SERIAL CORRELATION UNDER HIGH DIMENSIONALITY AND IMPROVED CONVERGENCE RATE FOR NORMAL EXTREMES

By YIJUN ZHU

Dissertation Director:

Han Xiao

In this dissertation, we proposed a new test for the serial correlation under high dimensionality, based on the maximum self-normalized autocovariances. We show that the asymptotic distribution of the test statistics is the extreme value distribution of type I. To calibrate the size of test, we use a multiplier bootstrap procedure, and prove the consistency under mixing conditions. Our new test has a more accurate empirical rejection rate under the null hypothesis, compared to the white noise test using the maximum cross correlation proposed by Chang et al. (2017). We also consider a second test statistic, which is the sum of squared maximum and minimum self-normalized autocovariances. It aims at killing two birds with one stone: to have an empirical size that is closer to the nominal one, and to gain more power for detecting non-zero autocorrelations. We demonstrate the sizes and powers of the proposed tests through extensive numerical studies and a real example on economic indicators, which confirm their superiority over existing methods. Since the convergence rate of the normal extreme is of critical importance for hypothesis tests based on extreme type test statistics, we consider a transform of the normal extreme, with improved convergence rates. In this second project, we show that after a monotone transformation, the convergence rate of the squared normal extreme is of the order $(\log n)^{-3}$, which is faster than the existing results, of the order $(\log n)^{-2}$ at their best. More strikingly, we demonstrate that the empirical convergence speed is uniformly improved, especially at the tails, even when the sample is of a moderate size.

Acknowledgements

In the first place, I would like to express my special appreciation and thanks to my advisor Professor Han Xiao, for all the support and encouragement he gave me. Thank him for inspiring me to think critically, guiding me through many difficulties and also providing valuable instructions on my thesis writing and presentations. It would not have been possible to accomplish this dissertation without his wise guidance and constant feedback. It is my honor to work with him. I would also like to thank my committee members, Professor Rong Chen, Professor Ying Hung, and Professor Xiaodong Lin, for their time and effort serving on my thesis committee, as well as all their brilliant comments and suggestions.

Many thanks to the Department of Statistics of Rutgers University, for educating me in a nourishing and inspiring environment through the years and providing the financial support. I would also like to thank our graduate directors Professor Harry Crane and Professor Tirthankar Dasgupta for supervising and coordinating my graduate studies, and also Professor John Kolassa, our former graduate director, for his devotion and effort to help all graduate students year after year during his service.

I will forever be thankful to my family, especially my father Li Zhu and my mother Zhikun Yao, for always caring about me and encouraging me to pursuit my dream. Finally, I would like to express my appreciation to my husband, Sheng Xu, who has always been by my side during this journey.

Dedication

This dissertation is dedicated to the memory of my beloved grandfather, Yaoming Zhu, who always believed in my ability to realize my potential and encouraged me to persist despite of any difficulties. This is for him.

Table of Contents

Abstract ii

Acknowledgements iv

Dedication v

List of Tables viii

List of Figures ix

1. Introduction 1

2. Improved Convergence Rates of Normal Extremes 4

2.1. Introduction 4

2.2. Main Results 7

2.2.1. Pointwise Convergence Rate 7

2.2.2. Uniform Convergence Rate 10

2.2.3. Comparisons of Different Convergence Rates 16

2.2.4. k-th Maxima 18

2.3. Applications and Numerical Comparisons 19

2.3.1. Numerical Comparisons 19

2.3.2. An Example 20

2.4. Appendix 21

2.4.1. Transformation of M_n 21

2.4.2. Expansion of b_n 23

2.4.3. k-th Maxima 25

2.4.4. Additional Figures 26

2.4.5. Additional Tables 26

3. Test for Serial Correlation under High Dimensionality 35

3.1. Introduction 35

3.2. Test Statistics 38

3.3. Main Results 40

3.3.1. Limiting Distribution 41

3.3.2. Dependent Gaussian Multiplier Bootstrap 54

3.3.3. Asymptotic Independence 61

3.4. Simulations 68

3.4.1. Empirical Size 69

3.4.2. Empirical Power 72

3.4.3. Calibrated Critical Values 75

3.4.4. Sensitivity to Dependence 76

3.5. Real Data Analysis 77

3.6. Further Discussion 79

4. Concluding Remarks 91

List of Tables

2.1. Smallest sample size to reach $\pm 10\%$ of the norminal level. 20

2.2. The empirical rejection probabilities $(\%)$ when $\boldsymbol{x}_i$ is $N(0, I_p)$. 23

2.3. The empirical rejection probabilities $(\%)$ when $\boldsymbol{x}_i$ has iid t_7 entries. 24

2.4. The empirical sizes $(\%)$ at nominal level $= 10\%$ where x_i are independent and $N(0, I_p)$ 29

2.5. The empirical sizes $(\%)$ at nominal level $= 5\%$ where x_i are independent and $N(0, I_p)$ 30

2.6. The empirical sizes $(\%)$ at nominal level $= 1\%$ where x_i are independent and $N(0, I_p)$ 31

2.7. The empirical sizes $(\%)$ at nominal level $= 10\%$ where x_i are i.i.d. t_7 32

2.8. The empirical sizes $(\%)$ at nominal level $= 5\%$ where x_i are i.i.d. t_7 33

2.9. The empirical sizes $(\%)$ at nominal level $= 1\%$ where x_i are i.i.d. t_7 34

3.1. The empirical sizes (%) of all tests for testing white noise series generated from model 1. 71

3.2. The empirical sizes (%) of all tests for testing white noise series generated from model 2. 71

3.3. The empirical sizes (%) of all tests for testing white noise series generated from model 3. 72

3.4. The empirical sizes (%) of all tests for testing white noise series generated from model 4 72

3.5. The p-value(%) of white noise tests for testing residuals 78

List of Figures

2.1. Comparison of the CDFs. 20

2.2. Comparison of the CDFs in the upper tail. 21

2.3. Full plot of different convergence rates for second maximum 22

2.4. Full plot of different convergence rates for third maximum 27

2.5. Full plot of different convergence rates for fourth maximum 28

3.1. Plots of empirical powers of all tests where data is generated from model 5 . . 74

3.2. Plots of empirical powers of all tests where data is generated from model 6 . . 81

3.3. Plots of empirical powers of all tests using calibrated critical values where data is generated from model 5 82

3.4. Plots of empirical powers of all tests using calibrated critical values where data is generated from model 6 83

3.5. Plots of empirical powers of all tests against dependence using calibrated critical values where data is generated from model 5 84

3.6. Plots of empirical powers of all tests against dependence using calibrated critical values where data is generated from model 6 85

3.7. Plots of empirical powers of all tests against dependence using calibrated critical values where data is generated from modified model 6 86

3.8. ACF of original series 87

3.9. ACF of residuals from 1-MAR(1) 88

3.10. ACF of residuals from 2-MAR(1) 89

3.11. Plots of empirical powers of all tests against dependence using calibrated critical values where data is generated from model 7 90

Chapter 1

Introduction

Testing for serial correlation is a fundamental problem in time series analysis. One of the most important application is to test the whiteness of the residuals, after a model has been fitted to the data. Many testing procedures for univariate time series (see for example Box and Pierce, 1970; Ljung and Box, 1979; Hong, 1996; Romano and Thombs, 1996; Horowitz et al., 2006; Xiao and Wu, 2014, among many others) have been extended to examine the whiteness of the multivariate series. In particular, Hosking (1980), Li and McLeod (1981) and Poskitt and Tremayne (1982) considered the multivariate portmanteau statistics, Hosking (1981) introduced the Lagrange multiplier test, and Tiao and Box (1981) proposed the likelihood ratio test. The asymptotic distributions of these test statistics are all established under the assumption that the dimension of the time series is fixed. Furthermore, in most cases the time series under the null hypothesis is assumed to be not only the white noise, but also iid for the validity of the theories.

For contemporary high dimensional time series, the traditional white noise tests often cannot be implemented directly, or may lead to distorted sizes. Recently, Chang et al. (2017) proposed an omnibus test, based on the maximum absolute auto- and cross-correlation of all component time series. Since the distribution of the test statistic is not analytically tractable, the critical values are given by dependent Gaussian multiplier bootstrapping. Despite the fact that this new test and its variants show superior performances over the classical methods, the tests themselves can be too conservative when the dimension is very high, and the empirical rejecting probabilities are very close to zero when the time series are generated from various white noise models.

We propose two new tests for the high dimensional white noise. The first one is based on the maximum absolute self-normalized autocovariances, and the second one is based on the sum of squared maximum and minimum of the self-normalized autocovariances. The choice of the test statistics are motivated by a few reasons. First, it is natural to put the variables on the same scale, before looking at the extreme values from them, so we use the self-normalized version of the sample autocovariances. Second, if the sample autocovariances are at the same scale, the asymptotic distributions of the extremes become analytically tractable. Third, since there is usually no prior knowledge about the sign of the autocorrelations, including both the maximum and the minimum in the test can be more adaptive to the unknown pattern under the alternative.

We show that the asymptotic distribution of the maximum-based test statistic is the extreme value distribution of type I. In particular, we allow the dimension to grow exponentially with the sample size. Furthermore, the white noise under the null hypothesis needs not to be iid, and only mixing conditions are required. We also find that under very mild dependence conditions, the maximum and minimum sample autocovariances are asymptotically independent, which implies that the limiting distribution of the second test statistic is a convolution of two extreme value distributions. To calibrate the sizes of the proposed tests for finite samples, we employ the dependent Gaussian multiplier bootstrap, which is similar to the one used in Chang et al. (2018). The consistency of the bootstrap is also established.

We conduct an extensive numerical analysis to compare the sizes and powers of the proposed tests with other methods. It is observed that our tests, especially the second one, are uniformly more accurate in terms of the empirical rejection probabilities under the null, comparing with all other methods. At the same time, the powers of the proposed tests are comparable with others. We use an economic dataset to illustrate the empirical performance of the tests. More specifically, we use the white noise tests as diagnostics to identify a suitable autoregressive model of the matrix-valued time series. While the tests in Chang et al. (2018) fail to detect the autocorrelations, our tests are more sensitive, and direct us to use an autoregressive model with two terms.

It is well known that usually the convergence rates of the extremes are very slow. For example, as a classical result, Fisher and Tippett (1928) pointed out that the normal extreme converges to the limiting distribution with a speed no faster than $(\log\log n)^2/\log n$. In a sequence of papers (Hall, 1979, 1980), Peter Hall proved that the convergence rate of the squared normal extreme can be improved to $1/\log n$ by choosing a better centering constant, and it can be further improved to $1/(\log n)^2$ if the squared normal extreme is also rescaled properly. Interestingly, these improved rates are not reflected through the empirical performances, unless the sample size is extremely large. For example, when comparing the distributions of the squared normal extremes, with two different centering constants, corresponding to the convergence rates $(\log\log n)^2/\log n$ and $1/\log n$ respectively, we find that the former is always better approximated by the limiting distribution, even when the sample size is as large as 10^7, although it leads to a slower rate theoretically. This is however not in contradiction with the theories: only that a much larger sample size (perhaps unrealistic) is required!

Our major finding is that after a monotone transform, the convergence rate of the normal extreme can be increased to $1/(\log n)^3$. Both the point-wise and uniform rates are derived. The most interesting findings are as follows. First, the actual distribution is significantly and uniformly closer to the asymptotic one than existing results, even for moderate sample sizes of hundreds or thousands. Second, the actual distribution is always stochastically dominated by the asymptotic one. This second phenomenon has important implications for the hypothesis tests based on the maximum: the corresponding asymptotic test is guaranteed to be conservative.

This dissertation is organized as follows. In Chapter 2 we prove a faster convergence rate of the squared normal extreme, after a suitable monotone transform. Both the point-wise and uniform convergence rates are shown to be $1/(\log n)^3$. In addition, similar results are obtained for the k-th maxima. We also illustrate the finite sample performances through some numerical studies. Chapter 3 focuses on the new white noise tests for high dimensional time series. The asymptotic distributions of both test statistics are derived. We also introduce a dependent Gaussian multiplier bootstrap to calibrate the sizes of the test, and prove the consistency of the bootstrapping procedure. A thorough numerical study is carried out to compare the sizes and powers of the proposed tests with existing methods. We also apply our tests to identify a suitable model for an economic dataset. At the end, Chapter 4 concludes with a short summary.

Chapter 2

Improved Convergence Rates of Normal Extremes

2.1 Introduction

Let $X_1, X_2, \ldots,$ be a sequence of independent standard normal random variables, and let $M_n := \max\{X_1, X_2, \ldots, X_n\}$ be the maximum of the first n of them. According to the extreme value theory (see Leadbetter et al., 1983, for an overview), after proper centering and rescaling, the limiting distribution of M_n is the extreme value distribution of type I, or the so called Gumbel distribution, with the distribution function $G_1(x) = \exp(-e^{-x})$. In fact, if we define

$$\begin{aligned}\alpha_n &= (2\log n)^{-1/2}\\ \beta_n &= \sqrt{2\log n} - \frac{\log(\log n) + \log(4\pi)}{2\sqrt{2\log n}}\end{aligned}$$

then $\alpha_n^{-1}(M_n - \beta_n)$ converges to G_1 in distribution, i.e.

$$\lim_{n\to\infty} \mathbb{P}\left[\alpha_n^{-1}(M_n - \beta_n) \le x\right] = \lim_{n\to\infty} \Phi^n(\alpha_n x + \beta_n) = \exp\left(-e^{-x}\right), \qquad -\infty < x < \infty, \tag{2.1}$$

where $\Phi(\cdot)$ is the distribution function of $N(0,1)$.

The rate of convergence in (2.1) is extremely slow. The fact was noted by Fisher and Tippett (1928), and studied more precisely by Hall (1979), who proved that the convergence rate in (2.1) is no better than $(\log\log n)^2/\log n$. Hall (1979) also found that if β_{n1} is the solution of the equation

$$2\pi\beta_{n1}^2 \exp(\beta_{n1}^2) = n^2, \tag{2.2}$$

and $\alpha_{n1} = \beta_{n1}^{-1}$, then

$$\frac{C_1}{\log n} < \sup_{-\infty<x<\infty} \left|\mathbb{P}[\alpha_{n1}^{-1}(M_n - \beta_{n1}) \leq x] - G_1(x)\right| < \frac{C_2}{\log n}, \tag{2.3}$$

where C_1 and C_2 are absolute constants. In other words, the convergence rate can be improved to $(\log n)^{-1}$ by choosing a better centering constant β_{n1}. They further proved that the rate can not be better than $(\log n)^{-1}$ by choosing a different sequence of normalizing constants.

It is equivalent and sometimes more convenient to study the limiting behavior of M_n through its squared version M_n^2. There are counterparts of (2.1) and (2.3) for M_n^2. More importantly, Hall (1980) found that with a suitably chosen constants a_n and b_n, the normalized sequence $a_n^{-1}(M_n^2 - b_n)$ converges to $G_1(x)$ with the rate $(\log n)^{-2}$. A detailed overview of the progression regarding the convergence rates of normal extremes will be provided in Section 2.2.3 via the squared version M_n^2.

While the aforementioned results are all on the uniform convergence rates, the convergence to G_1 in the upper tail is of particular interests when performing hypothesis tests using maximum type statistics. For example, the stepdown procedure of Romano and Wolf (2005) for multiple testing requires the knowledge about the upper quantiles of the maximum test statistic. Cai et al. (2014) used the maximum coordinate-wise difference of two transformed sample mean vectors to test the equality of two high dimensional means.

In Figure 2.1 we plot the empirical distributions of M_n^2 with different choices of normalizing sequences. The black line is the theoretical cumulative distribution function (CDF) G_1, the dashed, red and blue lines are empirical CDF corresponding to convergence rates in (2.1), (2.3) and $(\log n)^{-2}$ respectively. Figure 2.2 zooms in on the upper tails. Despite the fact that the red line is associated with a faster convergence rate than that of the dashed one, Figure 2.2 shows that it is consistently farther from the theoretical CDF in the upper tail, even when the sample size is as large as 10^5. This needs not contradicts the theories on the uniform convergence rates, because we see in Figure 2.1 that the dashed line deviates apparently from the black one in the lower tail. However, tests based on the statistic in (2.3) will be quite off, and have no advantage over the statistic in (2.1). On the other hand, the green line, corresponding to the rate $(\log n)^{-2}$, shows the potential to outperform the dashed one, when the sample size is sufficiently large, as shown in the bottom right panel of Figure 2.2. The issue is that the green line is below the theoretical CDF, indicating that the corresponding asymptotic test is not conservative.

Our major finding is that the convergence rate can be further improved to $(\log n)^{-3}$ by applying a monotone transform to M_n^2. Let $b_n := \frac{1}{2}[\Phi^{-1}(1-1/n)]^2$. Define Y_n through the following transform of M_n^2:

$$Y_n := \left[1 - \left(1 + \frac{M_n^2 - 2b_n}{8b_n^2}\right)^{-1}\right]\left(4b_n^2 + 2b_n - 2\right).$$

The results in Section 2.2 imply the following rate of convergence

$$\sup_{-\infty < x < \infty} \left|\mathbb{P}\left(Y_n \leq x\right) - G_1(x)\right| < \frac{C_3}{(\log n)^3}.$$

The blue lines in Figure 2.1 give empirical CDF of Y_n, which are almost identical with G_1 even when the sample size is as small as 200. When zoomed into the upper tail in Figure 2.2, the faster convergence of Y_n is more clearly seen. Furthermore, if Y_n is used as the test statistic for the asymptotic test, it is not only more accurate, but also always conservative, since the blue curve sits above the black one (for G_1) in the upper tail.

The rest of this article is organized as follows. We present and prove the point-wise and uniform convergence rates of Y_n in Section 2.2.1 and Section 2.2.2 respectively. In Section 2.2.3 we demonstrate how the faster convergence rate is achieved by comparing with existing results. Similar convergence rates regarding the k-th maxima are presented in Section 2.2.4. Numerical analysis and an application on testing the covariance structure are given in Section 2.3. Additional figures, tables, and some technical results are relegated in the Appendix.

We conclude this section by a brief review of the literature on the convergence rates of normal extremes. Cohen (1982) showed that the penultimate approximation can achieve the $(\log n)^{-2}$ rate. Rootzén (1983) investigated the convergence rates of the extremes from a stationary Gaussian process. Hall (1991) found that the extreme of a continuous time Gaussian process also has a logarithmic convergence rate. For convergence rates of extremes from a non-Gaussian sequence, we refer to Hall and Wellner (1979), Smith (1982), Leadbetter et al. (1983), de Haan and Resnick (1996), Peng et al. (2010) and references therein.

2.2 Main Results

We will first consider the pointwise convergence rates in Section 2.2.1, and then illustrate how the faster rates are achieved by modifying the normalizing constants and applying a transform of M_n^2 in Section 2.2.3. The uniform convergence rates are given in Section 2.2.2. In Section 2.2.4 we present the corresponding results for the k-th maxima. We make the convention that $C, C_1, C_2, \ldots$ are generic absolute constants, whose values may vary from place to place.

2.2.1 Pointwise Convergence Rate

Let b_n be the solution of the equation $1 - \Phi(\sqrt{2b_n}) = 1/n$. Recall that Y_n is defined as:

$$Y_n := \left[1 - \left(1 + \frac{M_n^2 - 2b_n}{8b_n^2}\right)^{-1}\right] \left(4b_n^2 + 2b_n - 2\right). \tag{2.4}$$

According to the definition, $\sqrt{2b_n}$ is the $(1 - 1/n)$-th quantile of the standard normal distribution. Since $M_n^2 \geq 0$ and $b_5 \approx .35$, the transform given in (2.4) is strictly monotone when $n \geq 5$, which we shall assume in the sequel.

Using the Newton-Raphson approximation (see Appendix 2.4.2 for detailed derivations), it can be shown that

$$b_n = \log n - \tfrac{1}{2} \log \log n - \tfrac{1}{2} \log 4\pi + O(\log \log n / \log n).$$

We first prove the pointwise convergence rate of Y_n to G_1. It is convenient to express the result through b_n, which is of the order $\log n$.

Theorem 2.1. *For each fixed $-\infty < x < \infty$,*

$$\mathbb{P}(Y_n \leq x) - G_1(x) = G_1(x)e^{-x} \cdot \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + O(b_n^{-4}).$$

Proof. Define the function $g_n(x)$ as the inverse transform of (2.4)

$$g_n(x) = \left[\left(1 - \frac{x}{4b_n^2 + 2b_n - 2}\right)^{-1} - 1\right] \cdot 8b_n^2 + 2b_n \tag{2.5}$$

Since (2.4) is a monotone transform, the event $[Y_n \le x]$ is equivalent to $[M_n^2 \le g_n(x)]$. It can be shown that

$$g_n(x) = 2b_n + 2x - \frac{x}{b_n} + \frac{x^2+3x}{2b_n^2} - \frac{2x^2+5x}{4b_n^3} + O(b_n^{-4}). \tag{2.6}$$

When n is large enough, $g_n(x) > 0$, and we let $x_n = [g_n(x)]^{1/2}$. Note that

$$\mathbb{P}(M_n \le x_n) > \mathbb{P}(M_n^2 \le x_n^2) = \mathbb{P}(M_n \le x_n) - \mathbb{P}(M_n < -x_n) > \mathbb{P}(M_n \le x_n) - 2^{-n}. \tag{2.7}$$

According to Lemma 2.4.1 in Leadbetter et al. (1983), for any $0 \le z \le n$,

$$0 \le e^{-z} - \left(1 - \frac{z}{n}\right)^n \le \frac{z^2 e^{-z}}{2} \cdot \frac{1}{n-1}. \tag{2.8}$$

Let $\tau_n(x) = n\,[1 - \Phi(x_n)]$, it follows that

$$\mathbb{P}(M_n \le x_n) = [1 - (1 - \Phi(x_n))]^n = \exp[-\tau_n(x)] + O(n^{-1}). \tag{2.9}$$

To evaluate $\tau_n(x)$, we make use the following series expansion of the normal tail probability (Abramowitz and Stegun, 1964): for any $z > 0$, and any positive integer m,

$$1 - \Phi(z) = \frac{\phi(z)}{z}\left\{1 - \frac{1}{z^2} + \frac{1\cdot 3}{z^4} + \cdots + \frac{(-1)^m 1\cdot 3 \ldots (2n-1)}{z^{2m}} + R_m\right\},$$

where

$$R_m = (-1)^{m+1}(2m+1)! \int_z^\infty \frac{\phi(t)}{t^{2m+2}} dt,$$

which is less in absolute value than the first neglected term. In particular, when $m = 3$, it holds that for any $z > 0$,

$$\left(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5} - \frac{15}{z^7}\right)\phi(z) < 1 - \Phi(z) < \left(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5} - \frac{15}{z^7} + \frac{105}{z^9}\right)\phi(z) \tag{2.10}$$

According to the definition of $\tau_n(x)$ and (2.10), we first do the Taylor expansion (up to the order b_n^{-4}) for

$$\begin{aligned}\phi(x_n) &= \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-b_n - x + \frac{x}{2b_n} - \frac{x^2+3x}{4b_n^2} + \frac{2x^2+5x}{8b_n^3}\right) \\ &= \frac{e^{-x}e^{-b_n}}{\sqrt{2\pi}} \cdot \left(1 + \frac{x}{2b_n} - \frac{x^2+6x}{8b_n^2} - \frac{5x^3+6x^2-30x}{48b_n^3} + O(b_n^{-4})\right),\end{aligned}$$

and

$$\begin{aligned}\frac{1}{x_n} &= \left(2b_n + 2x - \frac{x}{b_n} + \frac{x^2+3x}{2b_n^2} - \frac{2x^2+5x}{4b_n^3}\right)^{-1/2} \\ &= \frac{1}{\sqrt{2b_n}}\left(1 - \frac{x}{2b_n} + \frac{3x^2+2x}{8b_n^2} - \frac{5x^3+8x^2+6x}{16b_n^3} + O(b_n^{-4})\right).\end{aligned}$$

Combining the two preceding equations and rearranging the terms, we have

$$\frac{\phi(x_n)}{x_n} = \frac{e^{-x}e^{-b_n}}{\sqrt{4\pi b_n}} \cdot \left(1 - \frac{x}{2b_n^2} - \frac{4x^3+3x^2-6x}{24b_n^3} + O(b_n^{-4})\right).$$

According to (2.10), we also calculate

$$\begin{aligned}1 - \frac{1}{x_n^2} + \frac{3}{x_n^4} - \frac{15}{x_n^6} &= 1 - \frac{1}{2b_n} + \frac{2x+3}{4b_n^2} - \frac{4x^2+14x+15}{8b_n^3} + O(b_n^{-4}) \\ &= \left(1 - \frac{1}{2b_n} + \frac{3}{4b_n^2} - \frac{15}{8b_n^3}\right) \cdot \left(1 + \frac{x}{2b_n^2} - \frac{x^2+3x}{2b_n^3} + O(b_n^{-4})\right)\end{aligned}$$

Recall b_n is the solution of the equation $1 - \Phi(\sqrt{2b_n}) = 1/n$. According to the approximation to normal probability function in (2.10), we have

$$\frac{ne^{-b_n}}{\sqrt{4\pi b_n}} = \left(1 - \frac{1}{2b_n} + \frac{3}{4b_n^2} - \frac{15}{8b_n^3} + O(b_n^{-4})\right)^{-1}. \tag{2.11}$$

Therefore,

$$\begin{aligned}
&\left(1 - \frac{1}{x_n^2} + \frac{3}{x_n^4} - \frac{15}{x_n^6}\right) \frac{n\phi(x_n)}{x_n} \\
&\quad = e^{-x} \cdot \left(1 - \frac{x}{2b_n^2} - \frac{4x^3 + 3x^2 - 6x}{24b_n^3} + O(b_n^{-4})\right) \cdot \left(1 - \frac{1}{2b_n} + \frac{3}{4b_n^2} - \frac{15}{8b_n^3} + O(b_n^{-4})\right)^{-1} \\
&\qquad \cdot \left(1 - \frac{1}{2b_n} + \frac{3}{4b_n^2} - \frac{15}{8b_n^3}\right) \cdot \left(1 + \frac{x}{2b_n^2} - \frac{x^2 + 3x}{2b_n^3} + O(b_n^{-4})\right) \\
&\quad = e^{-x} \left(1 - \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + O(b_n^{-4})\right)
\end{aligned}$$

Since $n\phi(x_n)/x_n^9 = O(b_n^{-4})$, we have by (2.10)

$$\tau_n(x) = e^{-x} \left(1 - \frac{4x^3 + 15x^2 + 30x}{24b_n^3}\right) + O(b_n^{-4}).$$

According to (2.9), it follows that

$$\begin{aligned}
\mathbb{P}(Y_n \le x) - G_1(x) &= \exp(-\tau_n(x)) + O(n^{-1}) - G_1(x) \\
&= G_1(x)e^{-x} \cdot \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + O(b_n^{-4}).
\end{aligned}$$

The proof is complete. □

Using (2.10) and Newton-Ralphson method, we have the following expansions for b_n

$$b_n = \log n - \frac{\Delta}{2} + \frac{\Delta - 2}{4\log n} + \frac{\Delta^2 - 6\Delta + 14}{16(\log n)^2} + O\left(\frac{(\log\log n)^3}{(2\log n)^3}\right), \tag{2.12}$$

where

$$\Delta = \log\log n + \log 4\pi.$$

Therefore, Theorem 2.1 implies that Y_n converges to G_1 with the rate $(\log n)^{-3}$. The detailed derivation of (2.12) is given in the Appendix.

2.2.2 Uniform Convergence Rate

In this section we establish the uniform convergence rate.

Theorem 2.2. *There exists an absolute constant c_1, such that*

$$\sup_{-\infty<x<\infty} |\mathbb{P}(Y_n \le x) - G_1(x)| < \frac{c_1}{(\log n)^3}.$$

We prove Theorem 2.2 using two lemmas. Recall that $g_n(x)$, defined in (2.5), is the inverse transform of (2.4).

Lemma 2.1. *Let $\{c_n\}$ be an increasing sequence of positive integers such that $c_n^4/b_n \to 0$, then*

$$g_n(x) = 2b_n + 2x - \frac{x}{b_n} + \frac{x^2+3x}{2b_n^2} - \frac{2x^2+5x}{4b_n^3} + \frac{d_{1n}(x)}{b_n^3},$$

where $\lim_{n\to\infty} \sup_{-c_n \le x \le c_n} |d_{1n}(x)| = 0$.

Proof of Lemma 2.1. According to (2.5), for $-c_n \le x \le c_n$, we can obtain the following expansion:

$$\begin{aligned} g_n(x) &= 2b_n + 8b_n^2 \cdot \left[\left(1 - \frac{x}{4b_n^2+2b_n+2}\right)^{-1} - 1\right] \\ &= 2b_n + 2x\gamma_n + \frac{x^2\gamma_n^2}{2b_n^2} + \frac{x^3\gamma_n^3}{8b_n^4} \cdot \left(1 - \frac{x\gamma_n}{4b_n^2}\right)^{-1}, \end{aligned} \tag{2.13}$$

where

$$\gamma_n = \left(1 + \frac{1}{2b_n} - \frac{1}{2b_n^2}\right)^{-1}.$$

When $n \ge 13$, $b_n > 1$, by series expansion of γ_n, we have

$$\begin{aligned} \gamma_n &= 1 - \frac{1}{2b_n} + \frac{3}{4b_n^2} - \frac{5}{8b_n^3} + \frac{e_{1n}}{b_n^4} \\ \gamma_n^2 &= 1 - \frac{1}{b_n} + \frac{e_{2n}}{b_n^2} \\ \gamma_n^3 \left(1 - \frac{x\gamma_n}{4b_n^2}\right)^{-1} &= 1 + e_{3n}. \end{aligned}$$

The following bounds can be easily verified: $|e_{1n}| \le 1$, $|e_{2n}| \le 2$ and $|e_{3n}| \le 1$. Then by simplifying (2.13) we have

$$g_n(x) = 2b_n + 2x - \frac{x}{b_n} + \frac{x^2+3x}{2b_n^2} - \frac{2x^2+5x}{4b_n^3} + \frac{16xe_{1n} + 4x^2e_{2n} + x^3(1+e_{3n})}{8b_n^4}.$$

The proof is completed by noting that

$$\sup_{-c_n \le x \le c_n} \left| \frac{16xe_{1n} + 4x^2 e_{2n} + x^3(1+e_{3n})}{8b_n} \right| \le \frac{8c_n + 4c_n^2 + c_n^3}{4b_n} \to 0$$

under the condition $c_n^4/b_n \to 0$. $\square$

Lemma 2.2. *Let* $\{c_n\}$ *be the same sequence as used in Lemma 2.1, then*

$$\tau_n(x) = e^{-x}\left(1 - \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + \frac{d_{2n}(x)}{b_n^3}\right),$$

where $\lim_{n\to\infty} \sup_{-c_n \le x \le c_n} |d_{2n}(x)| = 0$ *for all* $-c_n \le x \le c_n$.

Proof of Lemma 2.2. Recall that $x_n := [g_n(x)]^{1/2}$. Using the normal tail probability bound in (2.10), we have

$$\left|\tau_n(x) - n\phi(x_n)\left(\frac{1}{x_n} - \frac{1}{x_n^3} + \frac{3}{x_n^5} - \frac{15}{x_n^7}\right)\right| \le \frac{105 n\phi(x_n)}{x_n^9}. \tag{2.14}$$

Write

$$n\phi(x_n)\left(\frac{1}{x_n} - \frac{1}{x_n^3} + \frac{3}{x_n^5} - \frac{15}{x_n^7}\right) = \left(\frac{x_n}{\sqrt{2b_n}}\right)^{-1} \cdot \frac{n\phi(x_n)}{\sqrt{2b_n}} \cdot \left(1 - \frac{1}{x_n^2} + \frac{3}{x_n^4} - \frac{15}{x_n^6}\right). \tag{2.15}$$

Let

$$x_{1n} := \frac{x}{b_n} - \frac{x}{2b_n^2} + \frac{x^2+3x}{4b_n^3} - \frac{2x^2+5x}{8b_n^4} + \frac{d_{1n}(x)}{2b_n^4},$$

where $d_{1n}(x)$ is defined in Lemma 2.1. For the first term on the right hand side of (2.15), by Lemma 2.1,

$$\left(\frac{x_n}{\sqrt{2b_n}}\right)^{-1} = (1+x_{1n})^{-1/2} = 1 - \frac{x_{1n}}{2} + \frac{3x_{1n}}{8} - \frac{5x_{1n}^3}{16} + R_{1n}(x_{1n}). \tag{2.16}$$

Under the condition $c_n^4/b_n \to 0$, it holds that $\sup_{-c_n \le x \le c_n} |x_{1n}| \le 5c_n/b_n$, and thus

$$\sup_{-c_n \le x \le c_n} |R_{1n}(x)| = \frac{o(1)}{b_n^3}.$$

The terms on the right hand side of (2.16) except for $R_{1n}(x_{1n})$ can be expanded as

$$\left(\frac{x_n}{\sqrt{2b_n}}\right)^{-1} - R_{1n}(x_{1n}) = 1 - \frac{x}{2b_n} + \frac{3x^2+2x}{8b_n^2} - \frac{5x^3+8x^2+6x}{16b_n^3} + \frac{d_{3n}(x)}{b_n^3}.$$

Note that for each fractional term in x_{1n}, the power of x is no greater than that of b_n, and the same claim holds for the series $d_{3n}(x)/b_n^3$. Furthermore, the first term (of the smallest power of x) in the expansion of $d_{3n}(x)$ is x^3/b_n, which goes to 0 uniformly over $-c_n \leq x \leq c_n$. Therefore, we conclude

$$\lim_{n\to\infty} \sup_{-c_n \leq x \leq c_n} |d_{3n}(x)| = 0$$

The other two terms in (2.15) can be treated similarly:

$$\frac{n\phi(x_n)}{\sqrt{2b_n}} = \frac{ne^{-x}e^{-b_n}}{\sqrt{4\pi b_n}} \cdot \left(1 + \frac{x}{2b_n} - \frac{x^2+6x}{8b_n^2} - \frac{5x^3+6x^2-30x}{48b_n^3} + \frac{d_{4n}(x)}{b_n^3} + R_{2n}(x)\right),$$
$$1 - \frac{1}{x_n^2} + \frac{3}{x_n^4} - \frac{15}{x_n^6} = \left(1 - \frac{1}{2b_n} + \frac{3}{4b_n^2} - \frac{15}{8b_n^3}\right) \cdot \left(1 + \frac{x}{2b_n^2} - \frac{x^2+3x}{2b_n^3} + \frac{d_{5n}(x)}{b_n^3} + R_{3n}(x)\right),$$

where

$$\sup_{-c_n \leq x \leq c_n} |d_{4n}(x)| \to 0 \quad \text{and} \quad |R_{2n}(x)| = \frac{o(1)}{b_n^3},$$
$$\sup_{-c_n \leq x \leq c_n} |d_{5n}(x)| \to 0 \quad \text{and} \quad |R_{3n}(x)| = \frac{o(1)}{b_n^3}.$$

Combining all the preceding bounds together with (2.11), we have

$$n\phi(x_n)\left(\frac{1}{x_n} - \frac{1}{x_n^3} + \frac{3}{x_n^5} - \frac{15}{x_n^7}\right) = e^{-x}\left(1 - \frac{4x^3+15x^2+30x}{24b_n^3} + \frac{d_{6n}(x)}{b_n^3}\right).$$

Using similar arguments as those for d_{3n}, we can verify that

$$\lim_{n\to\infty} \sup_{-c_n \leq x \leq c_n} |d_{6n}(x)| = 0.$$

It is easy to show that $\sup_{-c_n \leq x \leq c_n} n\phi(x_n)/x_n^9 = o(b_n^{-3})$. So the proof is complete in view of (2.14).

$\square$

We are now ready to prove Theorem 2.2.

Proof of Theorem 2.2. Let c_1 be a generic absolute constant which may vary from place to place. We consider three scenarios: $x < -c_n$, $-c_n \le x \le c_n$ and $x > c_n$, with $c_n = 4\log b_n$. Obviously, this choice of c_n satisfies the condition $c_n^4/b_n \to 0$.

We begin with the situation $-c_n \le x \le c_n$. By (2.7), it holds that

$$\left|\mathbb{P}(Y_n \le x) - \left(1 - \frac{\tau_n(x)}{n}\right)^n\right| \le 2^{-n}.$$

By (2.8) and Lemma 2.2, we have

$$|\mathbb{P}(Y_n \le x) - G_1(x)| \le 2G_1(x)e^{-x}\left(\frac{|4x^3 + 15x^2 + 30x|}{24b_n^3} + \frac{|d_{2n}(x)|}{b_n^3}\right) + \frac{1}{2^n} + \frac{1}{n},$$

when n is large enough. Since $\sup_{-c_n \le x \le c_n} |d_{2n}(x)| \to 0$, it suffices to show that

$$\sup_{-c_n \le x \le c_n} \left|G_1(x)e^{-x}(4x^3 + 15x^2 + 30x)\right| < \infty.$$

Numerical evaluations show that

$$\sup_{-\infty < x < \infty} \left|G_1(x)e^{-x}(4x^3 + 15x^2 + 30x)\right| < 20$$

Therefore, we have

$$\sup_{-c_n < x < c_n} |\mathbb{P}(Y_n \le x) - G_1(x)| < \frac{c_1}{(\log n)^3}.$$

Now we consider the second scenario $x > c_n$. We will show that both $G_1(x)$ and $\mathbb{P}(Y_n \le x)$ are close to 1, and their differences from 1 are of the order $1/(\log n)^3$. Since $x > c_n = 4\log b_n$,

$$G_1(x) = \exp(-e^{-x}) > \exp(-b_n^4) \ge 1 - 1/b_n^4. \tag{2.17}$$

On the other hand, recall the definition of $g(\cdot)$ in (2.5)

$$\begin{aligned} 1-\mathbb{P}(Y_n \le x) &\le \mathbb{P}(Y_n \ge 4\log b_n) = \mathbb{P}\left[M_n^2 \ge g(4\log b_n)\right] \\ &\le \mathbb{P}\left(M_n^2 \ge 2b_n + 4\log b_n \cdot \frac{8b_n^2}{4b_n^2+2b_n-2}\right) \end{aligned}$$

Note that $8b_n^2/(4b_n^2+2b_n-2) > 1.5$ for $n \ge 33$. Let $y_n^2 = 2b_n + 6\log b_n$, then

$$\mathbb{P}(M_n^2 \ge y_n^2) \le \mathbb{P}(M_n \ge y_n) + 1/2^n.$$

Let $\tau_n = n[1-\Phi(y_n)]$. Using the normal tail probability bounds (2.10), we have

$$\begin{aligned} \tau_n &\le \frac{n}{\sqrt{2\pi}}\,(2b_n + 6\log b_n)^{-1/2} \cdot \exp(-b_n - 3\log b_n) \\ &= \frac{ne^{-b_n}}{\sqrt{3\pi b_n}}\left(1+\frac{3\log b_n}{b_n}\right)^{-1/2} \cdot \exp(-3\log b_n) \end{aligned}$$

Recall $1-\Phi(\sqrt{2b_n}) = 1/n$, so that by (2.10)

$$\frac{ne^{-b_n}}{\sqrt{4\pi b_n}}\left(1-\frac{1}{2b_n}\right) < 1.$$

When $n \ge 33$, we have

$$\left(1+\frac{3\log b_n}{b_n}\right)^{-1/2} \cdot \left(1-\frac{1}{2b_n}\right)^{-1} < 1,$$

and it follows that

$$\tau_n < \exp(-3\log b_n) = 1/b_n^3.$$

Using (2.8), we deduce that when n is large enough

$$\mathbb{P}(M_n \ge y_n) = 1-(\Phi(y_n))^n = 1-\left(1-\frac{\tau_n}{n}\right)^n \le 1-e^{-\tau_n} + \frac{1}{n-1} \le \tau_n + \frac{1}{n-1}.$$

Therefore, we conclude

$$1-\mathbb{P}(Y_n \le x) < \frac{1}{b_n^3} + \frac{1}{n-1} + \frac{1}{2^n} < \frac{c_1}{(\log n)^3},$$

for some absolute constant c_1. The preceding inequality, together with (2.17), completes the proof for $x > c_n$.

Finally we consider $x < -c_n$ by showing that both $G_1(x)$ and $\mathbb{P}(Y_n \leq x)$ converge to 0 faster than $1/(\log n)^3$. Using the definition of b_n, we have when $n \geq 33$, and $x < -c_n = -4 \log b_n$,

$$G_1(x) = \exp(-e^{-x}) < \exp(-b_n^4) < 1/b_n^4.$$

On the other hand, when $x \leq -4 \log b_n$,

$$\mathbb{P}(Y_n \leq x) \leq \mathbb{P}[M_n^2 \leq g(-4 \log b_n)] \leq \mathbb{P}\left(M_n^2 \leq 2b_n - 4 \log b_n \cdot \frac{8b_n^2}{4b_n^2 + 2b_n - 2}\right).$$

Again since $8b_n^2/(4b_n^2 + 2b_n - 2) > 1.5$ when $n \geq 33$, if we let ${y'_n}^2 = 2b_n - 6 \log b_n$, then

$$\mathbb{P}(Y_n \leq x) \leq P(M_n \leq y_n).$$

Let $\tau'_n = n[1 - \Phi(y'_n)]$, we have by (2.10)

$$\begin{aligned}
\exp(-\tau'_n) &< \exp\left\{-\frac{ne^{-b_n}}{\sqrt{4\pi b_n}}\left(1 - \frac{3 \log b_n}{b_n}\right)^{-1/2} \cdot \left(1 - \frac{1}{(2b_n - 6 \log b_n)^2}\right) \cdot \exp(3 \log b_n)\right\} \\
&< \exp\left\{-\exp(3 \log b_n)\right\} \\
&< 1/b_n^3,
\end{aligned}$$

when n is large enough. We conclude by (2.8)

$$\mathbb{P}(Y_n \leq x) < \frac{1}{b_n^3} + \frac{1}{n} < \frac{c_1}{(\log n)^3},$$

which completes the proof. □

2.2.3 Comparisons of Different Convergence Rates

The best uniform convergence rate that can be obtained for M_n^2, if only centering and rescaling is allowed, is $(\log n)^{-2}$. We will give a summary of the progression in the literature. We also explain why the transformed M_n^2 can have a faster convergence rate $(\log n)^{-3}$.

In order for M_n^2 to have the limiting distribution G_1, the simplest option is to choose

$$b_{n1} = \log n - \log(\log n)/2 - \log(4\pi)/2;$$

then as a counterpart of (2.1), it holds that $\frac{1}{2}(M_n^2 - 2b_{n1}) \Rightarrow G_1$, where we use $\Rightarrow$ to denote the convergence in distribution. Using similar arguments as given in Hall (1979), it can be shown that the convergence rate is $(\log\log n)^2/\log n$. Similarly as (2.2), if b_{n1} is the solution of the equation

$$4\pi b_{n2}\exp(2b_{n2}) = n^2,$$

and M_n^2 is centered by b_{n2}, then the rate of convergence is analogous to (2.3)

$$\frac{C_1}{\log n} < \sup_{-\infty<x<\infty} \left|\mathbb{P}\left[\tfrac{1}{2}(M_n^2 - 2b_{n2}) \le x\right] - G_1(x)\right| < \frac{C_2}{\log n}. \tag{2.18}$$

Again (2.18) can be established following the proof in Hall (1979).

We note that $\sqrt{2b_{n1}}$ is an approximation of the $(1-1/n)$-th quantile of standard normal distribution obtained by using the following approximation of the tail probability:

$$1 - \Phi\left(\sqrt{2b_{n1}}\right) \approx \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2\log n}} \cdot \exp(-b_{n1}) = \frac{1}{n};$$

and b_{n2} is obtained by the following approximation of $1 - \Phi(\sqrt{2b_{n2}})$:

$$1 - \Phi\left(\sqrt{2b_{n2}}\right) \approx \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2b_{n2}}} \cdot \exp(-b_{n2}) = \frac{1}{n}.$$

If we choose b_{n3} through a more precise approximation of $1 - \Phi(\sqrt{2b_{n3}})$:

$$1 - \Phi\left(\sqrt{2b_{n3}}\right) \approx \frac{1}{\sqrt{4\pi b_{n3}}}\left(1 - \frac{1}{2b_{n3}}\right)\exp\left(-b_{n3}\right) = \frac{1}{n},$$

and set $a_{n3} = 2 - 1/b_{n3}$, then $a_{n3}^{-1}(M_n^2 - 2b_{n3}) \Rightarrow G_1$ with the convergence rate

$$\frac{C_1}{(\log n)^2} < \sup_{-\infty<x<\infty} \left|\mathbb{P}\left[a_{n3}^{-1}(M_n^2 - 2b_{n3}) \le x\right] - G_1(x)\right| < \frac{C_2}{(\log n)^2}. \tag{2.19}$$

The way we represent the preceding result is slightly different from the original one given by Hall (1980). The choices of a_{n3} and b_{n3} differ from those in Hall (1980) by smaller order terms, which do not affect the convergence rates. We choose the current formulation in order to have a better comparison with our main result.

To achieve a better rate of convergence, we first choose b_n precisely through $1-\Phi(\sqrt{2b_n}) = 1/n$. Second, observe that the events in (2.18) and (2.19) can be written as

$$M_n^2 \leq 2b_{n2} + 2x$$
$$M_n^2 \leq 2b_{n3} + 2x - x/b_{n3}$$

respectively. According to (2.6), the event $[Y_n \leq x]$ implies that

$$M_n^2 \leq 2b_n + 2x - \frac{x}{b_n} + \frac{x^2+3x}{2b_n^2} + O\left(b_n^{-3}\right).$$

We see that a term of order $O(b_n^{-2})$ is needed on the right hand side to achieve the convergence rate $(\log n)^{-3}$ in Theorem 2.1.

2.2.4 k-th Maxima

In this section we present pointwise and uniform convergence rates for the k-th maxima $M_{n,k}$, defined as the k-th largest among the first n variables $\{X_1, X_2, \ldots, X_n\}$. These results follow from almost the same arguments as those for the maxima, so we state them without proofs.

Theorem 2.3. *Let b_n be the solution of the equation $1-\Phi(\sqrt{2b_n}) = 1/n$. For an given positive integer k, define*

$$Y_{n,k} := \left[1 - \left(1 + \frac{M_{n,k}^2 - 2b_n}{8b_n^2}\right)^{-1}\right]\left(4b_n^2 + 2b_n - 2\right).$$

(i) For each fixed $-\infty < x < \infty$, it holds that

$$\mathbb{P}(Y_{n,k} \leq x) - G_k(x) = G_1(x)\frac{e^{-kx}}{(k-1)!} \cdot \frac{4x^3+15x^2+30x}{24b_n^3} + O(b_n^{-4}),$$

where $G_k(x) := G_1(x)\sum_{j=0}^{k-1} e^{-jx}/j!$.

(ii) There exists a constant $c_2 > 0$*, such that*

$$\sup_{-\infty<x<\infty} |\mathbb{P}(Y_{n,k} \le x) - G_k(x)| < \frac{c_2}{(\log n)^3}.$$

2.3 Applications and Numerical Comparisons

2.3.1 Numerical Comparisons

In this section, we numerically compare the convergence rates of different versions of the normalized M_n^2, introduced in Section 2.2.3. Specifically, we compare $G_1(x)$ with the CDF of $Y_{n1} := \frac{1}{2}(M_n^2 - 2b_{n1})$, $Y_{n2} := \frac{1}{2}(M_n^2 - 2b_{n2})$, $Y_{n3} := (2 - 1/b_{n3})^{-1}(M_n^2 - 2b_{n3})$, and Y_n, labeled by b_{n1}, b_{n2}, b_{n3} and b_n respectively in Figure 2.1. The vertical lines mark 90%, 95% and 99% quantiles of Gumbel distribution. We see that the distribution of Y_n (blue curve) is uniformly closer to $G_1(x)$, no matter what the sample size is. Figure 2.2 zooms into the upper tail for a clearer visualization. An interesting finding is that the faster theoretical convergence rates of Y_{n2} and Y_{n3} over Y_{n1}, are not reflected through the plots for Y_{n2} even when the sample size is as large as 10^5. The distribution of Y_{n3} starts to be closer to $G_1(x)$ in the upper tail when $n = 10^5$. We remark that the inferior performances of Y_{n2} and Y_{n3} need not necessarily contradicts the theoretical convergence rates: from Figure 2.1 it is seen that the convergence of Y_{n1} is much slower in the left tail. On the other hand, in Figure 2.2 it is more clearly seen that Y_n always has a faster convergence rate, compared with the rest. Furthermore, the CDF of Y_n lies above $G_1(x)$, indicating that if a hypothesis test is based on the maximum type statistic, then it is guaranteed to be conservative by using Y_n. This is in contrast to Y_{n3}, which is always below $G_1(x)$. Similar patterns are observed for the second maxima in Figure 2.3. Two additional figures for the 3rd and 4th maxima are given in the Appendix.

Let c_α be the $(1-\alpha)$-th quantile of $G_1(x)$. We find the smallest sample size n such that $\mathbb{P}(Y_n > c_\alpha)$ reaches $\pm 10\%$ of α. The results are summarized in Table 2.1 for all of Y_{ni}, $i = 1, 2, 3$ and Y_n. Overall Y_n needs much smaller sample sizes. Such sizes do not exist for Y_{n2} when $n \le 10^6$, so we choose not to report them.

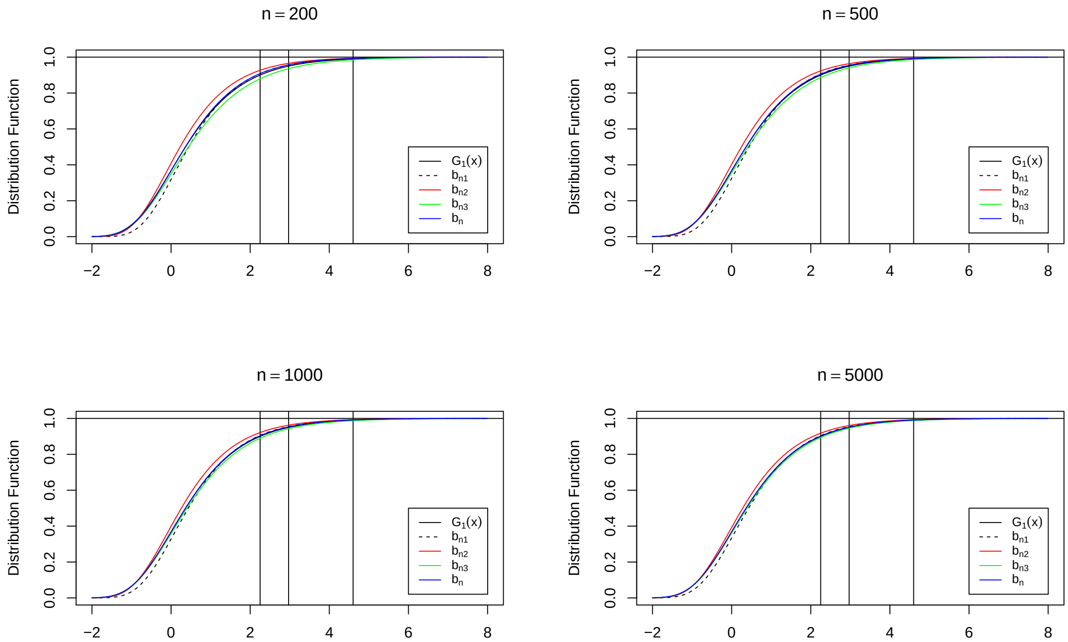


Figure 2.1: Comparison of the CDFs.

Table 2.1: Smallest sample size to reach $\pm 10\%$ of the norminal level.

α	Y_{n1}	Y_{n2}	Y_{n3}	Y_n
10%	92	-	1230	293
5%	995	-	3639	686
1%	359965	-	38208	4126

2.3.2 An Example

In this section, we consider an example on testing the covariance structure. Suppose $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ is a sequence of independently and identically distributed p-dimensional random vectors. Let $R = \{\rho_{ij}\}_{1 \le i,j \le p}$ be the correlation matrix of $\boldsymbol{x}_1$. Consider the hypothesis testing problem:

$$H_0 : R = I_p \quad \text{vs} \quad H_1 : R \neq I_p.$$

Jiang et al. (2004) proposed to use the maximum absolute sample correlation $L_N = \max_{1 \le i < j \le p} |\hat{\rho}_{ij}|$ as the test statistic, and proved that $\frac{1}{2}(NL_N^2 - 2b_{n1})$ converges in distribution to G_1, where $n = p(p-1)/2$. We consider the test statistics T_{Ni}, $i = 1, 2, 3$ and T_N, which are defined in the same way as Y_{ni} and Y_n in Section 2.3.1, but replacing M_n^2 therein by NL_N^2. The p-values

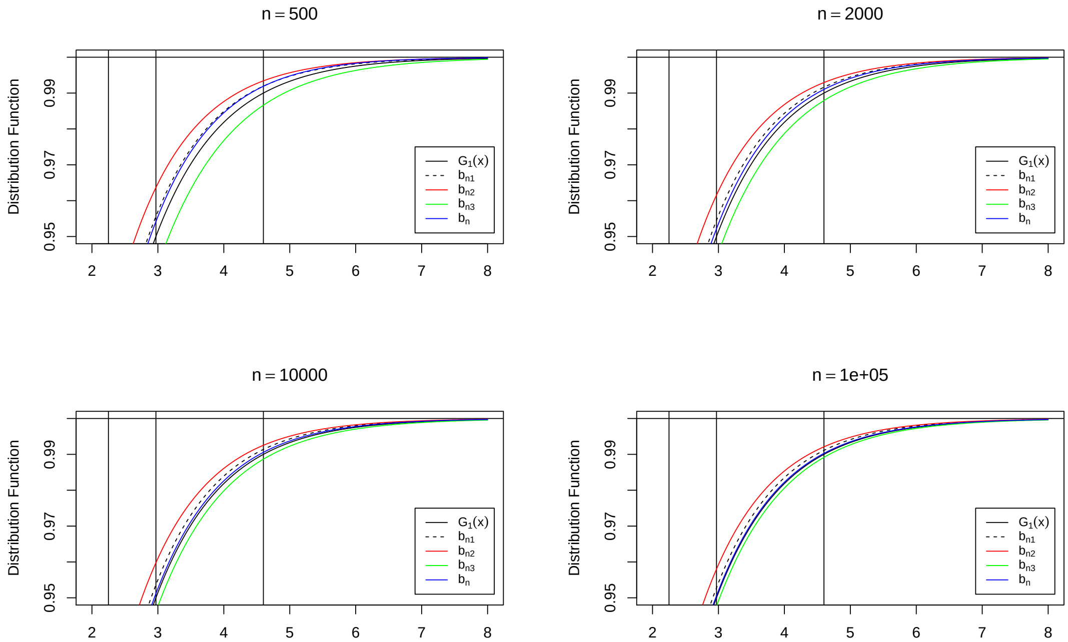


Figure 2.2: Comparison of the CDFs in the upper tail.

are calculated by comparing the test statistics with the Gumbel distribution G_1. By treating the sample correlations $N\hat{\rho}_{ij}$ as iid standard normal random variables, we obtain another approximation of the p-value, given by $1 - [\Phi(NL_N^2)]^n$. The test done this way is named as T_0. We report the empirical rejection probabilities based on 5000 repetitions in Table 2.2 and Table 2.3, where $\boldsymbol{x}_i \sim N(\mathbf{0}, I_p)$, and $\boldsymbol{x}_i$ has iid t_7 entries, respectively. We see that the empirical sizes of T_N, T_{N1} and T_0 are in general close to the nominal ones, and their performances are stable across different sample sizes and dimensions. The results are also consistent with our findings in Section 2.3.1. Six more extensive tables, covering more sample sizes and dimensions, are given in the Appendix.

2.4 Appendix

2.4.1 Transformation of M_n

Recall

$$Y_n := \left[1 - \left(1 + \frac{M_n^2 - 2b_n}{8b_n^2}\right)^{-1}\right] \left(4b_n^2 + 2b_n - 2\right).$$

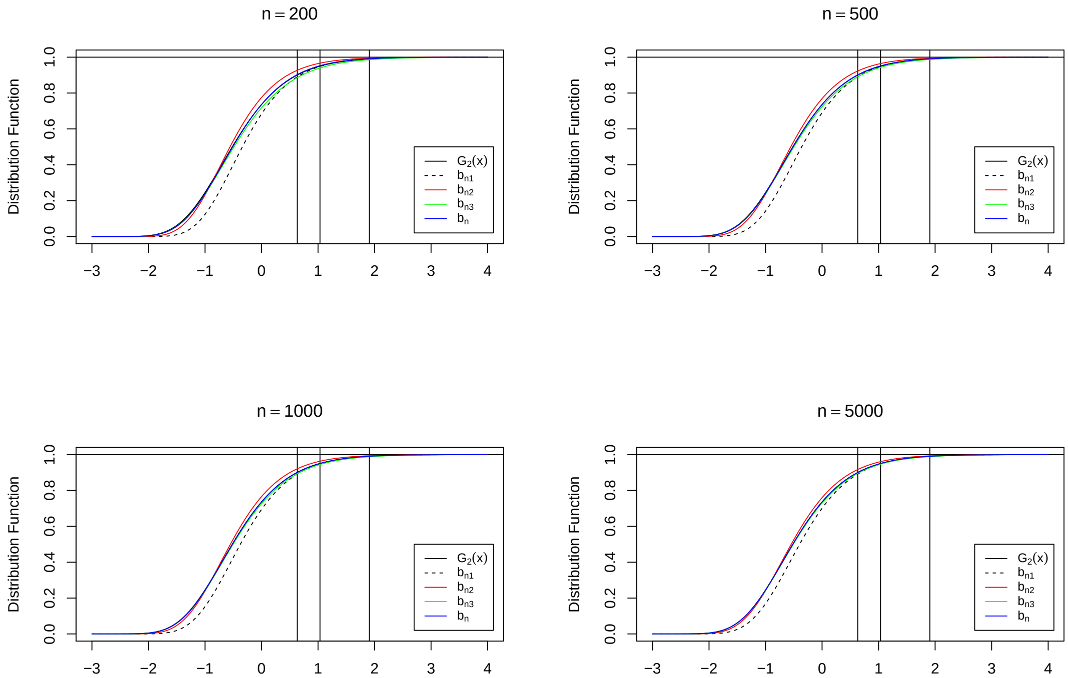


Figure 2.3: Comparison of CDFs for second maxima.

For $n \geq 5$, this transformation is strictly monotone. Then $[Y_n \leq x]$ implies that

$$\frac{M_n^2 - 2b_n}{8b_n^2} \leq \frac{x}{4b_n^2 + 2b_n - 2} + \frac{x^2}{(4b_n^2 + 2b_n - 2)^2}$$

which further implies:

$$\begin{aligned} M_n^2 &\leq 2b_n + 2x \cdot \left(1 + \frac{1}{2b_n} - \frac{1}{2b_n^2}\right)^{-1} + \frac{x^2}{2b_n^2} \cdot \left(1 + \frac{1}{2b_n} - \frac{1}{2b_n^2}\right)^{-2} \\ &= 2b_n + 2x \cdot \left(1 - \frac{1}{2b_n} + \frac{3}{4b_n^2} - \frac{5}{8b_n^3} + O(b_n^{-4})\right) + \frac{x^2}{2b_n^2} \cdot \left(1 - \frac{1}{b_n} + O(b_n^{-2})\right) \\ &= 2b_n + 2x - \frac{x}{b_n} + \frac{x^2 + 3x}{2b_n^2} - \frac{2x^2 + 5x}{4b_n^3} + O(b_n^{-4}) \end{aligned}$$

Table 2.2: The empirical rejection probabilities $(\%)$ when $\boldsymbol{x}_i$ is $N(0, I_p)$.

p	Test	$n=256$			$n=512$			$n=1024$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
32	T_0	8.96	4.42	0.72	9.64	4.86	0.94	10.74	5.60	1.28
	T_{N1}	8.62	4.02	0.62	8.94	4.48	0.78	10.28	5.02	1.10
	T_{N2}	7.10	3.36	0.48	7.72	3.80	0.54	8.48	4.18	0.80
	T_{N3}	10.06	5.12	1.02	10.64	5.46	1.16	11.92	6.42	1.62
	T_N	8.94	4.28	0.66	9.46	4.72	0.88	10.64	5.36	1.20
64	T_0	7.68	3.78	0.80	9.94	5.34	0.80	9.42	4.74	1.00
	T_{N1}	7.48	3.30	0.66	9.46	4.72	0.70	9.02	4.44	0.84
	T_{N2}	6.26	2.76	0.62	8.42	3.96	0.66	7.88	3.84	0.70
	T_{N3}	8.44	4.10	0.96	10.60	5.88	0.90	10.06	5.12	1.10
	T_N	7.68	3.76	0.72	9.88	5.24	0.80	9.36	4.70	0.96
128	T_0	7.60	3.32	0.62	9.30	4.86	0.80	9.86	4.82	0.98
	T_{N1}	7.34	3.12	0.60	8.90	4.52	0.72	9.56	4.58	0.82
	T_{N2}	6.26	2.72	0.60	7.86	3.82	0.66	8.20	4.00	0.68
	T_{N3}	8.16	3.82	0.66	9.78	5.14	0.90	10.14	5.30	1.14
	T_N	7.60	3.32	0.62	9.30	4.78	0.78	9.86	4.76	0.92
256	T_0	6.44	2.94	0.34	8.64	3.96	0.62	8.54	4.22	0.74
	T_{N1}	6.08	2.70	0.28	8.46	3.78	0.58	8.38	3.98	0.64
	T_{N2}	5.40	2.38	0.24	7.42	3.28	0.52	7.48	3.64	0.42
	T_{N3}	6.80	3.10	0.42	8.92	4.26	0.72	9.16	4.42	0.80
	T_N	6.44	2.92	0.34	8.66	3.94	0.60	8.54	4.20	0.70

2.4.2 Expansion of b_n

Recall b_n is the solution of the equation $1-\Phi(\sqrt{2b_n}) = 1/n$. We use the following approximation to the normal density:

$$1-\Phi(z) = \left(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5} - \frac{15}{z^7}\right)\phi(z).$$

Then $\sqrt{2b_n}$ is the solution of the following equation:

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(\frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5} - \frac{15}{x^7}\right) = 1/n. \tag{2.20}$$

Our goal is to use three consecutive applications of the Newton-Raphson approximation method to obtain the solution of (2.20) and then calculate b_n accordingly. Let

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(\frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5} - \frac{15}{x^7}\right).$$

Table 2.3: The empirical rejection probabilities $(\%)$ when $\boldsymbol{x}_i$ has iid t_7 entries.

p	Test	$n=256$			$n=512$			$n=1024$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
32	T_0	9.84	4.82	1.02	9.18	4.70	1.24	10.22	5.28	1.12
	T_{N1}	9.28	4.36	0.76	8.82	4.34	1.06	9.58	4.66	0.92
	T_{N2}	7.84	3.74	0.66	7.28	3.72	0.92	8.26	3.90	0.68
	T_{N3}	10.74	5.70	1.36	10.04	5.32	1.40	11.38	6.12	1.20
	T_N	9.70	4.66	0.84	9.14	4.44	1.16	10.02	5.02	0.98
64	T_0	9.28	4.28	0.94	10.18	5.02	0.82	9.02	4.78	1.00
	T_{N1}	8.96	3.88	0.70	9.80	4.50	0.68	8.46	4.54	0.78
	T_{N2}	7.70	3.44	0.52	8.42	3.94	0.56	7.44	3.82	0.70
	T_{N3}	9.72	4.74	1.04	10.76	5.44	0.98	9.82	5.28	1.14
	T_N	9.24	4.22	0.84	10.18	4.94	0.78	9.00	4.74	0.88
128	T_0	9.14	4.64	0.82	9.74	4.90	1.30	9.96	4.76	0.94
	T_{N1}	8.90	4.32	0.76	9.32	4.60	1.14	9.58	4.44	0.84
	T_{N2}	7.82	3.90	0.56	8.10	3.76	0.84	8.26	3.84	0.70
	T_{N3}	9.64	4.84	0.90	10.20	5.16	1.50	10.32	5.10	1.08
	T_N	9.14	4.58	0.80	9.74	4.80	1.20	9.96	4.66	0.90
256	T_0	9.08	4.32	0.94	10.20	4.98	0.98	9.98	5.36	1.30
	T_{N1}	8.80	4.02	0.88	9.96	4.74	0.84	9.68	5.02	1.18
	T_{N2}	7.92	3.50	0.86	8.82	4.18	0.80	8.84	4.52	1.10
	T_{N3}	9.36	4.72	1.04	10.58	5.30	1.04	10.48	5.62	1.38
	T_N	9.08	4.30	0.94	10.20	4.98	0.94	9.98	5.30	1.20

then the derivative of $f(x)$ is:

$$f'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(-1 + \frac{105}{x^8}\right).$$

We start from

$$x_0 = \sqrt{2\log n} - \frac{\Delta}{2\sqrt{2\log n}},$$

where

$$\Delta = \log\log n + \log 4\pi.$$

By Newton-Raphson approximation method,

$$f(x_0) + f'(x_0)(x_1 - x_0) = 1/n.$$

Then we can obtain:

$$x_1 = \sqrt{2\log n} - \frac{\Delta}{2\sqrt{2\log n}} - \frac{\Delta^2 - 4\Delta + 8}{8\,(2\log n)^{3/2}}.$$

Repeat this procedure for two more times, we have

$$x_2 = \sqrt{2\log n} - \frac{\Delta}{2\sqrt{2\log n}} - \frac{\Delta^2 - 4\Delta + 8}{8\,(2\log n)^{3/2}} - \frac{\Delta^3 - 8\Delta^2 + 32\Delta - 56}{16\,(2\log n)^{5/2}}.$$

$$\begin{aligned} x_3 = & \sqrt{2\log n} - \frac{\Delta}{2\sqrt{2\log n}} - \frac{\Delta^2 - 4\Delta + 8}{8\,(2\log n)^{3/2}} - \frac{\Delta^3 - 8\Delta^2 + 32\Delta - 56}{16\,(2\log n)^{5/2}} \\ & - \frac{15\Delta^4 - 184\Delta^3 + 1152\Delta^2 - 4128\Delta + 7040}{384\,(2\log n)^{7/2}}. \end{aligned}$$

Then by $b_n = x_3^2/2$, it can be easily calculated:

$$b_n = \log n - \frac{\Delta}{2} + \frac{\Delta - 2}{4\log n} + \frac{\Delta^2 - 6\Delta + 14}{16(\log n)^2} + O\left(\frac{(\log\log n)^3}{(2\log n)^3}\right).$$

2.4.3 k-th Maxima

Recall that x_n is defined as

$$x_n^2 := 2b_n + 2x - \frac{x}{b_n} + \frac{x^2 + 3x}{2b_n^2} - \frac{2x^2 + 5x}{4b_n^3} + O(b_n^{-4}).$$

Event $[Y_{n,k} \leq x]$ is equivalent to $[M_{n,k}^2 \leq x_n^2]$. Following the similar argument in Hall (1980), we have:

$$
\begin{aligned}
P(Y_{n,k} \leq x) - G_k(x) =& P\left(M_{n,k} \leq x_n\right) - G_k(x) + O(n^{k-1}2^{-n}) \\
=& \sum_{j=0}^{k-1} \binom{n}{j} \Phi^{n-j}(x_n)\left(1 - \Phi(x_n)\right)^j - G_k(x) + O(n^{k-1}2^{-n}) \\
=& \sum_{j=0}^{k-1} \Phi^{n-j}(x_n)\left(n\left(1 - \Phi(x_n)\right)\right)^j / j! - G_1(x) \sum_{j=0}^{k-1} e^{-jx}/j! + O(n^{-1}) \\
=& \exp\left\{-e^{-x}\left(1 - \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + O(b_n^{-4})\right)\right\} \\
& \cdot \sum_{j=0}^{k-1} e^{-jx}\left(1 - \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + O(b_n^{-4})\right)^j / j! \\
& - G_1(x) \sum_{j=0}^{k-1} e^{-jx}/j! + O(n^{-1}) \\
=& G_1(x) \Bigg\{ \left(1 + e^{-x} \cdot \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + O(b_n^{-4})\right) \\
& \cdot \sum_{j=0}^{k-1} e^{-jx}\left(1 - \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + O(b_n^{-4})\right)^j / j! - \sum_{j=0}^{k-1} e^{-jx}/j! \Bigg\} \\
& + O(n^{-1}) \\
=& G_1(x) \cdot \frac{4x^3 + 15x^2 + 30x}{24b_n^3} \cdot \sum_{j=0}^{k-1} e^{-jx}(e^{-x} - j)/j! + O(b_n^{-4}) \\
=& G_1(x) \frac{e^{-kx}}{(k-1)!} \cdot \frac{4x^3 + 15x^2 + 30x}{24b_n^3} + O(b_n^{-4})
\end{aligned}
$$

2.4.4 Additional Figures

In this section we provide comparisons of the CDFs of the third and fourth maxima.

2.4.5 Additional Tables

In this section we provide a more extensive simulation results for the example in Section 2.3.2, covering more sample sizes and dimensions.

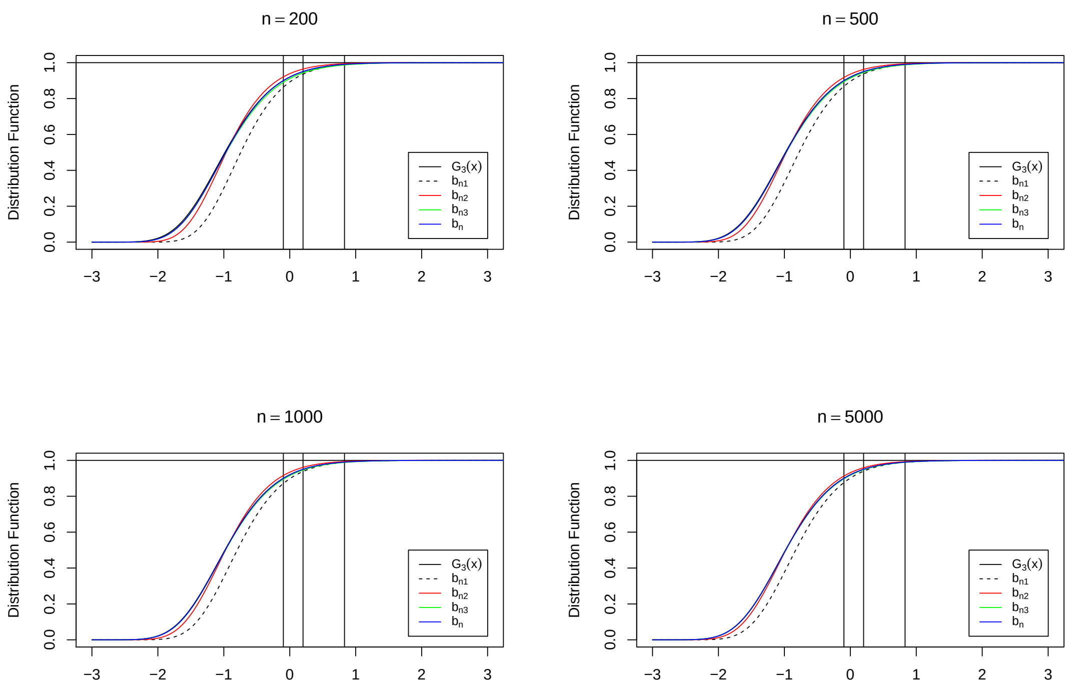


Figure 2.4: Full plot of different convergence rates for third maximum

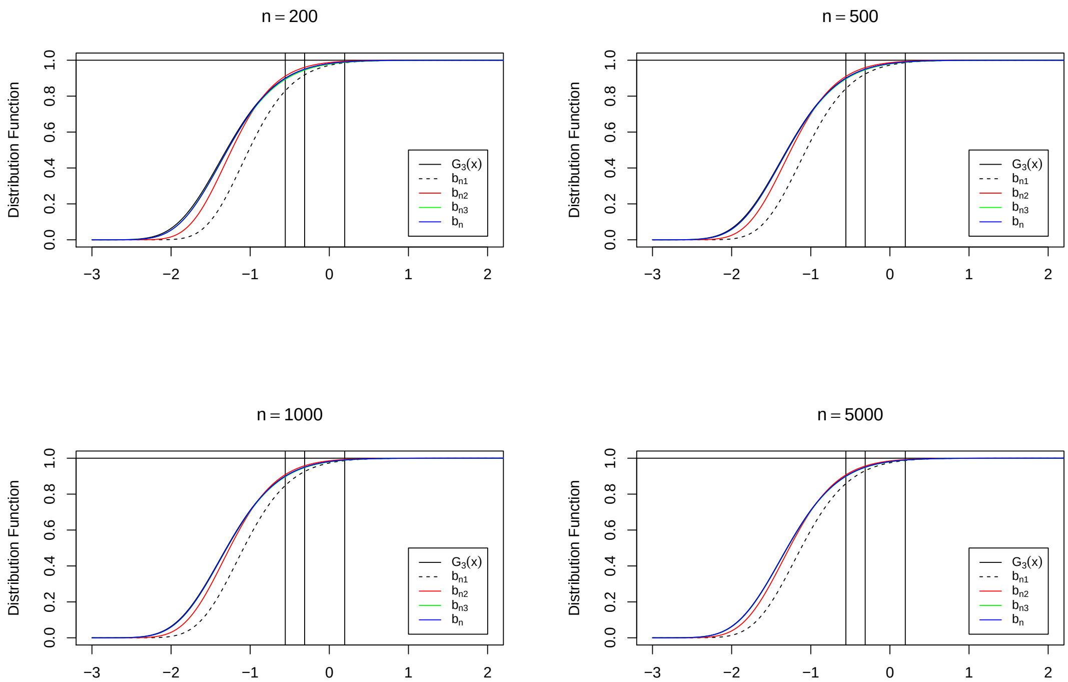


Figure 2.5: Full plot of different convergence rates for fourth maximum

Table 2.4: The empirical sizes $(\%)$ at nominal level $=$ 10% where x_i are independent and $N(0, I_p)$

p	Test	$n = 128$	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
16	T_0	8.26	9.82	9.68	10.24	9.96
	b_{n1}	7.80	8.92	8.96	9.20	9.22
	b_{n2}	6.04	7.14	7.64	7.52	7.36
	b_{n3}	10.42	11.52	11.68	12.10	11.78
	b_n	8.02	9.50	9.32	9.66	9.64
32	T_0	7.58	8.96	9.64	10.74	10.18
	b_{n1}	7.00	8.62	8.94	10.28	9.46
	b_{n2}	5.46	7.10	7.72	8.48	7.86
	b_{n3}	8.80	10.06	10.64	11.92	11.32
	b_n	7.46	8.94	9.46	10.64	10.08
64	T_0	5.80	7.68	9.94	9.42	10.36
	b_{n1}	5.50	7.48	9.46	9.02	9.98
	b_{n2}	4.56	6.26	8.42	7.88	8.78
	b_{n3}	6.36	8.44	10.60	10.06	11.06
	b_n	5.80	7.68	9.88	9.36	10.30
128	T_0	5.70	7.60	9.30	9.86	9.46
	b_{n1}	5.48	7.34	8.90	9.56	9.12
	b_{n2}	4.70	6.26	7.86	8.20	8.14
	b_{n3}	6.00	8.16	9.78	10.14	9.76
	b_n	5.70	7.60	9.30	9.86	9.46
256	T_0	4.50	6.44	8.64	8.54	9.44
	b_{n1}	4.28	6.08	8.46	8.38	9.16
	b_{n2}	3.72	5.40	7.42	7.48	8.30
	b_{n3}	4.74	6.80	8.92	9.16	9.84
	b_n	4.50	6.44	8.66	8.54	9.44

Table 2.5: The empirical sizes $(\%)$ at nominal level $=5\%$ where x_i are independent and $N(0, I_p)$

p	Test	$n=128$	$n=256$	$n=512$	$n=1024$	$n=2048$
16	T_0	3.84	4.66	5.22	5.14	4.94
	b_{n1}	3.44	4.00	4.60	4.36	4.34
	b_{n2}	2.72	3.22	3.76	3.50	3.58
	b_{n3}	5.04	6.18	6.68	6.74	6.34
	b_n	3.58	4.14	4.76	4.52	4.50
32	T_0	3.34	4.42	4.86	5.60	5.06
	b_{n1}	3.02	4.02	4.48	5.02	4.68
	b_{n2}	2.26	3.36	3.80	4.18	3.92
	b_{n3}	3.96	5.12	5.46	6.42	5.64
	b_n	3.26	4.28	4.72	5.36	4.90
64	T_0	2.52	3.78	5.34	4.74	5.46
	b_{n1}	2.32	3.30	4.72	4.44	4.94
	b_{n2}	1.88	2.76	3.96	3.84	4.20
	b_{n3}	2.70	4.10	5.88	5.12	6.04
	b_n	2.48	3.76	5.24	4.70	5.26
128	T_0	2.48	3.32	4.86	4.82	4.56
	b_{n1}	2.30	3.12	4.52	4.58	4.22
	b_{n2}	1.88	2.72	3.82	4.00	3.68
	b_{n3}	2.84	3.82	5.14	5.30	4.98
	b_n	2.48	3.32	4.78	4.76	4.52
256	T_0	1.94	2.94	3.96	4.22	4.72
	b_{n1}	1.78	2.70	3.78	3.98	4.38
	b_{n2}	1.54	2.38	3.28	3.64	3.88
	b_{n3}	2.02	3.10	4.26	4.42	4.96
	b_n	1.94	2.92	3.94	4.20	4.70

Table 2.6: The empirical sizes $(\%)$ at nominal level $= 1\%$ where x_i are independent and $N(0, I_p)$

p	Test	$n = 128$	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
16	T_0	0.52	1.00	1.00	1.10	0.98
	b_{n1}	0.42	0.78	0.78	0.86	0.70
	b_{n2}	0.32	0.60	0.64	0.72	0.56
	b_{n3}	0.86	1.48	1.52	1.50	1.58
	b_n	0.38	0.76	0.74	0.80	0.64
32	T_0	0.46	0.72	0.94	1.28	1.10
	b_{n1}	0.38	0.62	0.78	1.10	0.84
	b_{n2}	0.30	0.48	0.54	0.80	0.68
	b_{n3}	0.52	1.02	1.16	1.62	1.48
	b_n	0.38	0.66	0.88	1.20	0.88
64	T_0	0.46	0.80	0.80	1.00	1.02
	b_{n1}	0.36	0.66	0.70	0.84	0.88
	b_{n2}	0.30	0.62	0.66	0.70	0.68
	b_{n3}	0.50	0.96	0.90	1.10	1.16
	b_n	0.38	0.72	0.80	0.96	1.00
128	T_0	0.30	0.62	0.80	0.98	0.94
	b_{n1}	0.26	0.60	0.72	0.82	0.78
	b_{n2}	0.26	0.60	0.66	0.68	0.66
	b_{n3}	0.38	0.66	0.90	1.14	1.04
	b_n	0.28	0.62	0.78	0.92	0.90
256	T_0	0.28	0.34	0.62	0.74	0.92
	b_{n1}	0.18	0.28	0.58	0.64	0.84
	b_{n2}	0.18	0.24	0.52	0.42	0.74
	b_{n3}	0.34	0.42	0.72	0.80	1.04
	b_n	0.24	0.34	0.60	0.70	0.92

Table 2.7: The empirical sizes $(\%)$ at nominal level $= 10\%$ where x_i are i.i.d. t_7

p	Test	$n = 128$	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
16	T_0	8.54	9.62	9.88	9.86	9.62
	b_{n1}	7.74	8.98	8.88	9.10	8.82
	b_{n2}	6.12	7.02	7.08	7.40	6.94
	b_{n3}	10.58	11.40	11.68	11.86	11.18
	b_n	8.26	9.42	9.32	9.46	9.22
32	T_0	8.86	9.84	9.18	10.22	9.82
	b_{n1}	8.38	9.28	8.82	9.58	9.26
	b_{n2}	7.00	7.84	7.28	8.26	7.72
	b_{n3}	9.88	10.74	10.04	11.38	10.74
	b_n	8.76	9.70	9.14	10.02	9.66
64	T_0	7.74	9.28	10.18	9.02	9.74
	b_{n1}	7.32	8.96	9.80	8.46	9.52
	b_{n2}	6.10	7.70	8.42	7.44	8.10
	b_{n3}	8.42	9.72	10.76	9.82	10.38
	b_n	7.72	9.24	10.18	9.00	9.72
128	T_0	7.84	9.14	9.74	9.96	9.84
	b_{n1}	7.60	8.90	9.32	9.58	9.54
	b_{n2}	6.58	7.82	8.10	8.26	8.36
	b_{n3}	8.32	9.64	10.20	10.32	10.32
	b_n	7.84	9.14	9.74	9.96	9.84
256	T_0	7.40	9.08	10.20	9.98	10.62
	b_{n1}	7.20	8.80	9.96	9.68	10.10
	b_{n2}	6.48	7.92	8.82	8.84	9.08
	b_{n3}	7.64	9.36	10.58	10.48	10.94
	b_n	7.40	9.08	10.20	9.98	10.62

Table 2.8: The empirical sizes $(\%)$ at nominal level $= 5\%$ where x_i are i.i.d. t_7

p	Test	$n = 128$	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
16	T_0	4.02	4.66	4.46	5.16	4.30
	b_{n1}	3.56	3.96	3.98	4.54	3.66
	b_{n2}	2.98	2.98	3.16	3.46	2.96
	b_{n3}	5.12	5.92	6.00	6.44	5.98
	b_n	3.70	4.16	4.10	4.72	3.88
32	T_0	4.34	4.82	4.70	5.28	4.96
	b_{n1}	3.92	4.36	4.34	4.66	4.68
	b_{n2}	3.22	3.74	3.72	3.90	3.84
	b_{n3}	5.08	5.70	5.32	6.12	5.66
	b_n	4.20	4.66	4.44	5.02	4.90
64	T_0	3.66	4.28	5.02	4.78	4.98
	b_{n1}	3.38	3.88	4.50	4.54	4.40
	b_{n2}	2.82	3.44	3.94	3.82	3.70
	b_{n3}	4.20	4.74	5.44	5.28	5.54
	b_n	3.66	4.22	4.94	4.74	4.90
128	T_0	4.06	4.64	4.90	4.76	4.62
	b_{n1}	3.90	4.32	4.60	4.44	4.30
	b_{n2}	3.56	3.90	3.76	3.84	3.82
	b_{n3}	4.28	4.84	5.16	5.10	5.08
	b_n	4.02	4.58	4.80	4.66	4.56
256	T_0	3.84	4.32	4.98	5.36	5.32
	b_{n1}	3.60	4.02	4.74	5.02	5.10
	b_{n2}	3.36	3.50	4.18	4.52	4.64
	b_{n3}	4.00	4.72	5.30	5.62	5.54
	b_n	3.80	4.30	4.98	5.30	5.26

Table 2.9: The empirical sizes $(\%)$ at nominal level $= 1\%$ where x_i are i.i.d. t_7

p	Test	$n = 128$	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
16	T_0	0.80	0.78	0.88	1.04	1.04
	b_{n1}	0.64	0.60	0.78	0.90	0.86
	b_{n2}	0.48	0.50	0.68	0.76	0.68
	b_{n3}	1.18	1.24	1.36	1.46	1.40
	b_n	0.60	0.58	0.74	0.88	0.82
32	T_0	0.76	1.02	1.24	1.12	0.84
	b_{n1}	0.58	0.76	1.06	0.92	0.58
	b_{n2}	0.50	0.66	0.92	0.68	0.52
	b_{n3}	1.06	1.36	1.40	1.20	1.22
	b_n	0.62	0.84	1.16	0.98	0.66
64	T_0	0.56	0.94	0.82	1.00	0.98
	b_{n1}	0.52	0.70	0.68	0.78	0.84
	b_{n2}	0.44	0.52	0.56	0.70	0.60
	b_{n3}	0.66	1.04	0.98	1.14	1.14
	b_n	0.54	0.84	0.78	0.88	0.94
128	T_0	0.76	0.82	1.30	0.94	0.98
	b_{n1}	0.74	0.76	1.14	0.84	0.86
	b_{n2}	0.66	0.56	0.84	0.70	0.64
	b_{n3}	0.92	0.90	1.50	1.08	1.12
	b_n	0.76	0.80	1.20	0.90	0.94
256	T_0	0.68	0.94	0.98	1.30	1.10
	b_{n1}	0.64	0.88	0.84	1.18	1.00
	b_{n2}	0.58	0.86	0.80	1.10	0.94
	b_{n3}	0.80	1.04	1.04	1.38	1.18
	b_n	0.68	0.94	0.94	1.20	1.10

Chapter 3

Test for Serial Correlation under High Dimensionality

3.1 Introduction

White noise or serial correlation test is of fundamental importance in time series analysis, and has been extensively studied in both statistics and econometrics. It is one of the most important diagnostics to assess the adequacy of a fitted model. For the univariate case, classical portmanteau test (Box and Pierce, 1970; Ljung and Box, 1979) has been a standard procedure. For an overview of its variants, see Escanciano and Lobato (2009). The classical portmanteau test involves a fixed number sample autocovariances. Many tests have also been proposed to take account of possible serial correlations at large lags, including Deo (2000), Durlauf (1991), Hong (1996), Robinson (1991) and Shao (2011), among others. For many of these tests, the asymptotic distributions are only valid when the time series under the null hypothesis is i.i.d., and the sizes of the tests are distorted if the underlying series is uncorrelated but not independent. The performance of the tests can be improved using bootstrap methods, see for example Horowitz et al. (2006) and Romano and Thombs (1996). Recently, Xiao and Wu (2014) proposed to use the maximum absolute sample autocovariances as the test statistic, and showed that it is powerful to the alternative autocovariance sequence with a few spikes. Due to the slow convergence rates of normal extremes, they proposed to use the blocks of blocks bootstrap (Horowitz et al., 2006) to improve the finite sample performance. Hill and Motegi (2016) considered bootstrapping similar statistics using dependent wild bootstrap (Shao, 2010).

For multivariate white noise test, portmanteau procedures were proposed and studied by Hosking (1980), Li and McLeod (1981) and Poskitt and Tremayne (1982), among others. Hosking (1981) considered the score or Lagrange multiplier test when the alternative is a vector autoregressive moving average process. These tests are designed in the classical setting where the dimension of the series is treated as fixed.

Recently there has been an emerging interest on modeling high dimensional time series. Roughly speaking, these works fall into two major categories: (i) vector autoregressive modeling with regularization (Davis et al., 2016; Basu et al., 2015; Guo et al., 2016; Han et al., 2015, 2016; Nicholson et al., 2017; Song and Bickel, 2011; Kock and Callot, 2015; Negahban and Wainwright, 2011; Nardi and Rinaldo, 2011, among others), and (ii) dynamic factor models (Bai, 2003; Forni et al., 2005; Lam et al., 2011, 2012; Wang et al., 2019; Chen et al., 2018; Ghosh et al., 2019; Chen et al., 2019, among others).

It is of great interest to study the white noise test for high dimensional time series, either as an initial step before any modeling, or as a diagnostic after a model has been fitted. However, for the contemporary high dimensional data, the aforementioned classical tests often cannot be implemented directly, or may lead to distorted sizes. Chang et al. (2018) considered the white noise test for high dimensional time series, and proposed to use the maximum sample cross correlation as the test statistic. The distribution of the test statistic is not tractable analytically. They adopted a wild bootstrap procedure, where the critical value was obtained by sampling the maximum of a very high dimensional Gaussian random vector. Using the result from Chernozhukov et al. (2013), they showed that the test is consistent if the covariance matrix of this Gaussian vector is chosen as the kernel estimate of the covariance matrix of all sample cross correlations involved in the test. Despite of the better performance than classical procedures, this test can be conservative itself under high dimensionality, i.e. the empirical rejection probabilities are very close to zero when the time series are generated from various white noise models.

We propose two new tests for the high dimensional white noise. The first one is based on the maximum absolute self-normalized autocovariances, and the second one is based on the sum of squared maximum and minimum of the self-normalized autocovariances. The choice of the test statistics are motivated by a few reasons. First, it is natural to put the variables on the same scale, before looking at the extreme values from them, so we use the self-normalized version of the sample autocovariances. Second, if the sample autocovariances are at the same scale, the asymptotic distributions of the extremes become analytically tractable. Third, since there is usually no prior knowledge about the sign of the autocorrelations, including both the maximum and the minimum in the test can be more adaptive to the unknown pattern under the alternative.

We show that the asymptotic distribution of the maximum-based test statistic is the extreme value distribution of type I. In particular, we allow the dimension to grow exponentially with the sample size. Furthermore, the white noise under the null hypothesis needs not to be iid, and only mixing conditions are required. We also find that under very mild dependence conditions, the maximum and minimum sample autocovariances are asymptotically independent, which implies that the limiting distribution of the second test statistic is a convolution of two extreme value distributions. To calibrate the sizes of the proposed tests for finite samples, we employ the dependent Gaussian multiplier bootstrap, which is similar to the one used in Chang et al. (2018). The consistency of the bootstrap is also established.

We conduct an extensive numerical analysis to compare the sizes and powers of the proposed tests with other methods. It is observed that our tests, especially the second one, are uniformly more accurate in terms of the empirical rejection probabilities under the null, comparing with all other methods. At the same time, the powers of the proposed tests are comparable with others. We use an economic dataset to illustrate the empirical performance of the tests. More specifically, we use the white noise tests as diagnostics to identify a suitable autoregressive model of the matrix-valued time series. While the tests in Chang et al. (2018) fail to detect the autocorrelations, our tests are more sensitive, and direct us to use an autoregressive model with two terms.

The rest of the chapter is organized as follows. Section 3.2 introduces the new test statistics. Then the main theoretical results are presented in Section 3.3, followed by an extensive numerical study in Section 3.4. A real example in economics is analyzed in Section 3.5 for the illustration purpose. A discussion of the further research directions is given in Section 3.6.

3.2 Test Statistics

Consider a p-dimensional centered stationary time series $\{x_t\}$, and a hypothesis testing problem:

$$H_0 : \{x_t\} \text{ is white noise} \quad \text{vs} \quad H_1 : \{x_t\} \text{ is not white noise}$$

Given the observations $x_1, x_2, \ldots, x_n$, the sample cross covariance between i-th and j-th time series at lag k is denoted by

$$\hat{\gamma}_{ij}(k) = \frac{1}{n} \sum_{t=1}^{n-k} x_{it} x_{j+k,t}.$$

For each lag k, there are p^2 sample cross covariances. A maximum number of lags K is pre-selected before we do test. So in total there are $p^2 K$ sample cross covariances involved. Intuitively, when the time series is not white noise, there must be some non-zero correlation appearing among p series. It is straightforward to use the maximized sample cross covariance as the test statistic, i.e. when the maximized sample cross covariance exceeds some threshold, it is a sign that the time series is not white noise. Moreover, it is more analytically tractable to take the maximum value of all sample cross covariance at the same scale. Therefore, we carry out the test based on the maximized self-normalized sample cross covariance. Specifically, the variance of $\hat{\gamma}_{ij}(k)$ can be estimated by

$$\hat{\tau}_{ijk} = \frac{1}{n} \sum_{t,t'} \mathcal{K}\left(\frac{t-t'}{w_n}\right) \left[x_{it} x_{j,t+k} - \hat{\gamma}_{ij}(k)\right] \left[x_{it'} x_{j,t'+k} - \hat{\gamma}_{ij}(k)\right], \tag{3.1}$$

where $\mathcal{K}(\cdot)$ is a positive definite kernel function, and w_n is the bandwidth parameter. The reason for using estimated standard deviation to normalize sample cross covariance rather than using sample cross correlation directly is that, the asymptotic distribution of such self-normalized extremes is more analytically tractable, which is shown to be Gumbel distribution. More importantly, in general, sample cross correlations themselves can be less sensitive to the alternative. Concretely, if the time series is i.i.d. across time, then the asymptotic variance of $\sqrt{n}\hat{\rho}_{ij}(k)$ is one. However, if the process is not white noise, the variance would be smaller than one. For instance, for an AR(1) process $x_t = \phi x_{t-1} + e_t$ with coefficient $-1 < \phi < 1$ and i.i.d. innovations e_t, the asymptotic variance of lag-1 sample autocorrelation $\sqrt{n}\hat{\rho}(1)$ is $1 - \phi^2$, and for

MA(1) process $x_t = e_t + \theta e_{t-1}$ with coefficient $-1 < \theta < 1$, the asymptotic variance of lag-1 sample autocorrelation $\sqrt{n}\hat{\rho}(1)$ is $1 - 3\theta^2/(1+\theta^2)^2 + 4\theta^4/(1+\theta^2)^4$. Both are smaller than 1. Similar conclusion can be made for multivariate linear processes with i.i.d. innovations using Bartlett formula (Bartlett, 1955). Therefore, the sample version of a nonzero cross correlation is less likely to be larger than other sample cross correlations due to its small variance.

The following notations are used in the proposed test. Let ξ_t be the p^2K-dimensional vector:

$$\xi_t = \text{vec}(x_t x_{t+1}' - \hat{\Gamma}_1, \dots, x_t x_{t+K}' - \hat{\Gamma}_K)$$

where $\hat{\Gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k} x_t x_{t+k}'$ is the lag-k sample autocovariance matrix. There are in total $\tilde{n}$ such vector for $1 \le t \le \tilde{n}$. Define $\chi(\cdot) = \{\chi_1(\cdot), \chi_2(\cdot), \chi_3(\cdot)\}$ to be a mapping from $\{1, 2, \dots, p^2K\}$ to $\{(i,j,k) : 1 \le i,j \le p, 1 \le k \le K\}$ such that the l-th element for ξ_t is $x_{\chi_1(l),t} x_{\chi_2(l),t+\chi_3(l)} - \hat{\gamma}_{\chi(l)}$. Define

$$\Xi = \mathbb{E}\left\{ \left(\frac{1}{\tilde{n}^{1/2}} \sum_{t=1}^{\tilde{n}} \xi_t \right) \left(\frac{1}{\tilde{n}^{1/2}} \sum_{t=1}^{\tilde{n}} \xi_t \right)' \right\}. \tag{3.2}$$

In practice, Ξ is usually unknown and is estimated from data. Specifically, denote the diagonal element of Ξ as $\tau_{\chi(l)}$, which is estimated by (3.1). Then our proposed test statistic is

$$T_n = n^{1/2} \max_{1 \le l \le p^2 K} |\hat{\eta}_l|,$$

where $\hat{\eta}_l = \hat{\gamma}_{\chi(l)}/\sqrt{\hat{\tau}_{\chi(l)}}$. We reject H_0 when $T_n > \text{cv}_\alpha$, where $\alpha \in (0,1)$ is the significance level of the test. In order to improve the finite sample performance, we use a bootstrap procedure to generate critical values. Specifically, let $G \sim N(0, \Xi)$ and normalize each component of G_l with the corresponding estimated standard deviation $\sqrt{\hat{\tau}_{\chi(l)}}$. Write the normalized vector as Z whose l-th element is $\frac{G_l}{\sqrt{\hat{\tau}_{\chi(l)}}}$. In practice, we can draw $G_1, \dots, G_B$ from $N(0, \Xi)$ for a large number B, and calculate the normalized $Z_1, \dots, Z_B$, and then take the $\lfloor B\alpha \rfloor$-th largest among $|Z_1|_\infty, \dots, |Z_B|_\infty$ as the critical value. According to Xiao and Wu (2013), it can be shown that the asymptotic distribution of the test statistic converges to Gumbel distribution under some dependence conditions.

In addition, since there is usually no prior knowledge about the sign of the autocorrelations, including both the maximum and the minimum in the test can be more adaptive to the unknown pattern under the alternative. Therefore, we propose a second test statistics:

$$S_n = \hat{M}^2 + \hat{m}^2,$$

where

$$\hat{M} = n^{1/2} \max_{1 \le l \le p^2 K} \hat{\eta}_l, \quad \hat{m} = n^{1/2} \max_{1 \le l \le p^2 K} (-\hat{\eta}_l),$$

where $\hat{\eta}_l = \hat{\gamma}_{\chi(l)} / \sqrt{\hat{\tau}_{\chi(l)}}$. The critical values can be obtained similarly. Specifically, write

$$\hat{M}_Z = \max_{1 \le l \le p^2 K} \hat{Z}_l, \quad \hat{m}_Z = \max_{1 \le l \le p^2 K} (-\hat{Z}_l),$$

where $\hat{Z}_l = G_l / \sqrt{\hat{\tau}_{\chi(l)}}$. Then the critical value is simply the $\lfloor B\alpha \rfloor$-th largest among the sum of square of $\hat{M}_B$ and $\hat{m}_B$. Similarly, the asymptotic distribution of S_n can be derived accordingly. We proved the limiting distribution of S_n is a convolution of two independent Gumbel distributions based on the results from Marques et al. (2015).

However, in practice, Ξ is unknown and is estimated from data. When we have a high-dimensional time series, there might be computational issues to estimate such huge matrix. Therefore, instead, we use the dependent Gaussian multiplier bootstrap to estimate critical values. See more details in Section 3.3.2.

3.3 Main Results

To study the theoretical properties of $\hat{M}$ and $\hat{m}$, we need the following regularity conditions.

Condition 1 There exists constants $C_1, C_2 > 0$ and $0 < \lambda_1 \le 2$ such that for any $x > 0$,

$$\sup_t \sup_{1 \le i \le p} \mathbb{P}(|x_{i,t}| > u) \le C_1 \exp(-C_2 u^{\lambda_1}).$$

Condition 2 Assume $\{x_t\}$ is β-mixing in the sense that $\beta_k \to 0$ as $k \to \infty$ where $\beta_k = \sup_t \mathbb{E}\left\{\sup_{B \in \mathcal{F}_{t+k}^{\infty}} |\mathbb{P}(B|\mathcal{F}_{-\infty}^{t}) - \mathbb{P}(B)|\right\}$ and $\mathcal{F}_{-\infty}^{t}$ and $\mathcal{F}_{u+k}^{\infty}$ are the σ-fields generated respectively by $\{x_u\}_{u \le t}$ and $\{x_u\}_{u \ge t+k}$. In addition, assume there exists constants $C_3 > 0$ and

$0 < \lambda_2 \leq 1$ independent of p and n such that $\beta_k \leq \exp(-C_3 k^{\lambda_2})$ for any $k > 0$.

Remark 3.1. *Han and Wu (2019) argued that β-mixing condition is dimension-dependent under a high-dimensional triangular array time series setting. Thus, Condition 2 needs to be verified with caution under high dimensionality.*

3.3.1 Limiting Distribution

In order to derive the limiting distribution of the test statistics, it is crucial to bound the Kolmogorov distance between the distribution of the test statistics and that of the extreme values from Gaussian distribution. To achieve this goal, we first look at the extreme values of autocovariances normalized by its true standard deviation, which are denoted as

$$M = n^{1/2} \max_{1 \leq l \leq p^2 K} \eta_l, \quad m = n^{1/2} \max_{1 \leq l \leq p^2 K} (-\eta_l);$$

where $\eta_l = \hat{\gamma}_{\chi(l)} / \sqrt{\tau_{\chi(l)}}$ and $\tau_{\chi(l)}$ is the diagonal element from the true covariance matrix Ξ defined in (3.2). Proposition 3.1 bounds the difference between the joint distribution of M and m with that of M_Z and m_Z. Then by substituting the true standard deviation with the estimated standard deviation, we can obtain the Kolmogorov distance between the joint distribution of $\hat{M}$,$\hat{m}$ and that of $\hat{M}_Z$ and $\hat{m}_Z$ in Proposition 3.2. Then our main result Theorem 3.1 follows accordingly.

Proposition 3.1. *Under Conditions 1-2, it holds that*

$$\sup_{t_1, t_2 \in \mathbb{R}} |\mathbb{P}(M \leq t_1, m \leq t_2) - \mathbb{P}(M_Z \leq t_1, M_Z \leq t_2)| \rightarrow o(1), \tag{3.3}$$

as $n \rightarrow \infty$, provided that $\log p = o(n^{\lambda_1/(4+9\lambda_1)})$.

Proof. Write $d_0 = \sup_{t_1, t_2 \in \mathbb{R}} |\mathbb{P}(M \leq t_1, m \leq t_2) - \mathbb{P}(M_Z \leq t_1, M_Z \leq t_2)|$. Observe that

$$\zeta = (\zeta_1, \ldots, \zeta_{p^2 K})' = n^{-1} \sum_{t=1}^{\tilde{n}} \mu_t + R_n,$$

where μ_t is a $p^2 K$-dimensional vector with $\mu_{l,t} = \xi_{l,t} / \sqrt{\tau_{\chi(l)}}$ and R_n is the reminder term. We can define $u = \tilde{n}^{-1} \sum_{t=1}^{\tilde{n}} \mu_t \equiv (u_1, \ldots, u_{p^2K})'$ and let

$$\tilde{M} = \tilde{n}^{1/2} \max_{1 \le l \le p^2 K} u_l, \quad \tilde{m} = \tilde{n}^{1/2} \max_{1 \le l \le p^2 K} (-u_l).$$

The first step is to show $d_0 \le d_1 + o(1)$ where

$$d_1 := \sup_{t_1, t_2} \left| \mathbb{P}(\tilde{M} \le t_1, \tilde{m} \le t_2) - \mathbb{P}(M_Z \le t_1, M_Z \le t_2) \right|.$$

Observe that for any $t_1, t_2 \in \mathbb{R}$ and $\epsilon > 0$,

$$d_0 \le d_1 + \mathbb{P}(|M - \tilde{M}| > \epsilon) + \mathbb{P}(|m - \tilde{m}| > \epsilon) + \mathbb{P}(|M_Z - t_1| \le \epsilon, |m_Z - t_2| \le \epsilon).$$

By the anti-concentration inequality of Gaussian random variables, it holds that

$$\mathbb{P}(|M_Z - t_1| \le \epsilon, |m_Z - t_2| \le \epsilon) \le \mathbb{P}(|M_Z - t_1| \le \epsilon) + \mathbb{P}(|m_Z - t_2| \le \epsilon) \le C\epsilon(\log(p/\epsilon))^{1/2}.$$

The rest of the proof is the same with the proof of Lemma 4 in Chang et al. (2017). Thus, it suffices to show $d_1 = o(1)$. Similar to the proof of Theorem 1 in Chang et al. (2018), we can decompose the sequence $\{1, 2, \ldots, \tilde{n}\}$ to $H + 1$ blocks, where $H = \lfloor \tilde{n}/s \rfloor$ and s is a positive integer satisfying $s = o(n)$ and $s \le n/2$. Let q and r be two positive integers (depending on n) such that $s = q + r$ and $q = o(n)$, $r = o(q)$. Then we can further decompose each of the first H blocks into a large block I_h and a small block J_h, where $h = 1, \ldots, H$. Specifically, $I_1 = \{1, \ldots, q\}, J_1 = \{q + 1, \ldots, q + r\}, \ldots, I_H = \{(H - 1)(q + r) + 1, \ldots, (H - 1)(q + r) + q\}, J_H = \{(H - 1)(q + r) + q + 1, \ldots, H(q + r)\}$, and $J_{H+1} = \{H(q + r) + 1, \ldots, \tilde{n}\}$ is the reminder block. Given $D_n \to \infty$, write $u^+_{l,t} = u_{l,t}\mathbb{1}\{|u_{l,t}| \le D_n\} - \mathbb{E}(u_{l,t}\mathbb{1}\{|u_{l,t}| \le D_n\})$ and $u^-_{l,t} = u_{l,t}\mathbb{I}\{|u_{l,t}| > D_n\} - \mathbb{E}(u_{l,t}\mathbb{I}\{|u_{l,t}| > D_n\})$. Write $u^+_t = (u^+_{1,t}, \ldots, u^+_{p^2K,t})^T$ and $u^-_t = (u^-_{1,t}, \ldots, u^-_{p^2K,t})^T$ for each $t = 1, \ldots, \tilde{n}$. In addition, define

$$S_h = \sum_{t \in I_h} u^+_t, \quad S'_h = \sum_{t \in J_h} u^+_t.$$

Let $W = (W_1, \ldots, W_{p^2K})'$ be a centered normal random vector with covariance matrix

$E(WW^T) = \frac{1}{qH}\sum_{h=1}^{H} E(S_h S_h^T)$, and let M_W and m_W be the corresponding maximum and minimum of W. Then we can proceed the proof by two parts. First is to show

$$d_2 := \sup_{t_1,t_2\in\mathbb{R}} \left|\mathbb{P}(\tilde{M} \le t_1, \tilde{m} \le t_2) - \mathbb{P}(M_W \le t_1, M_W \le t_2)\right| = o(1), \tag{3.4}$$

and then to show

$$d_3 := \sup_{t_1,t_2\in\mathbb{R}} |\mathbb{P}(M_W \le t_1, m_W \le t_2) - \mathbb{P}(M_Z \le t_1, M_Z \le t_2)| = o(1). \tag{3.5}$$

We first prove $d_2 = o(1)$. Observe that

$$\sqrt{\tilde{n}}u = \frac{1}{\tilde{n}^{1/2}}\sum_{t=1}^{\tilde{n}} u_t^+ + \frac{1}{\tilde{n}^{1/2}}\sum_{t=1}^{\tilde{n}} u_t^-.$$

Define

$$\tilde{M}^+ = \frac{1}{\tilde{n}^{1/2}} \max_{1\le l\le p^2K} \sum_{t=1}^{\tilde{n}} u_{l,t}^+, \quad \tilde{m}^+ = \frac{1}{\tilde{n}^{1/2}} \max_{1\le l\le p^2K} \sum_{t=1}^{\tilde{n}} -u_{l,t}^+.$$

By triangle inequality, we have

$$|\tilde{M} - \tilde{M}^+| \le \max_{1\le l\le p^2K} \left|\frac{1}{\tilde{n}^{1/2}}\sum_{t=1}^{\tilde{n}} u_{l,t}^-\right|.$$

For any $\epsilon_1 > 0$, it follows that

$$d_2 \le d_4 + \sup_{t_1,t_2\in\mathbb{R}} \mathbb{P}(|M_W - t_1| \le \epsilon_1, |m_W - t_2| \le \epsilon_1) + 2\mathbb{P}\left(\max_{1\le l\le p^2K}\left|\frac{1}{\tilde{n}^{1/2}}\sum_{t=1}^{\tilde{n}} u_{l,t}^-\right| > \epsilon_1\right),$$

where

$$d_4 := \sup_{t_1,t_2\in\mathbb{R}} \left|\mathbb{P}(\tilde{M}^+ \le t_1, \tilde{m}^+ \le t_2) - \mathbb{P}(M_W \le t_1, M_W \le t_2)\right|. \tag{3.6}$$

By anti-concentration inequality in Theorem 3 of Chernozhukov et al. (2015), it holds that

$$\begin{aligned}\mathbb{P}(|M_W - t_1| \le \epsilon_1, |m_W - t_2| \le \epsilon_1) &\le \mathbb{P}(|M_W - t_1| \le \epsilon_1) + \mathbb{P}(|m_W - t_2| \le \epsilon_1)\\ &\le C\epsilon_1(\log p)^{1/2}\end{aligned} \tag{3.7}$$

for any $\epsilon_1 \to 0$. According to Davydov inequality from Davydov (1968), for each $l = 1, \ldots, p^2K$, we have

$$\mathbb{E}\left(\left|\frac{1}{\tilde{n}^{1/2}}\sum_{t=1}^{\tilde{n}} u_{l,t}^{-}\right|^2\right) \leq \frac{1}{\tilde{n}}\sum_{t=1}^{\tilde{n}} \mathbb{E}((u_{l,t}^{-})^2) + \frac{C}{\tilde{n}}\sum_{t_1 \neq t_2}[\mathbb{E}((u_{l,t_1}^{-})^4)]^{1/4}[\mathbb{E}((u_{l,t_2}^{-})^4)]^{1/4}\beta_{t_1-t_2}.$$

Moreover, it follows from condition 1 that

$$\begin{aligned}\mathbb{P}(|u_{l,t}| > u) &= \mathbb{P}\left(\left|\frac{x_{it}x_{j,t+k}}{\sqrt{\tau_{ijk}}}\right| > u, |x_{it}| > u^{\omega}\right) + \mathbb{P}\left(\left|\frac{x_{it}x_{j,t+k}}{\sqrt{\tau_{ijk}}}\right| > u, |x_{it}| \leq u^{\omega}\right) \\ &\leq \mathbb{P}(|x_{it}| > u^{\omega}) + \mathbb{P}\left(\left|\frac{x_{it}}{\sqrt{\tau_{ijk}}}\right| > u^{1-\omega}\right) \\ &\leq C\exp(-Cu^{\omega\lambda_1}) + C\exp(-Cu^{(1-\omega)\lambda_1})\end{aligned}$$

for any $u > 0$. This is bounded by $C\exp(-Cu^{\lambda_1/2})$ when $\omega = 1/2$. Then following the same argument in proof of Theorem 1 of Chang et al. (2018), we have

$$\mathbb{E}[(u_{l,t}^{-})^4] \leq CD_n^4\exp(-CD_n^{\lambda_1/2}).$$

This further implies that

$$\begin{aligned}\sup_{1\leq l\leq p^2K}\mathbb{E}\left(\left|\frac{1}{\tilde{n}^{1/2}}\sum_{t=1}^{\tilde{n}} u_{l,t}^{-}\right|^2\right) &\leq CD_n^2\exp(-CD_n^{\lambda_1/2}) + CD_n^4\exp(-CD_n^{\lambda_1/2})\sum_{k=1}^{N-1}\exp(-Ck_2^{\lambda}) \\ &\leq CD_n^2\exp(-CD_n^{\lambda_1/2}).\end{aligned}$$

Thus, by Markov's inequality, we have

$$\mathbb{P}\left(\max_{1\leq l\leq p^2K}\left|\frac{1}{\tilde{n}^{1/2}}\sum_{t=1}^{\tilde{n}} u_{l,t}^{-}\right| > \epsilon_1\right) \leq \frac{\tilde{n}}{\epsilon_1^2}\sup_{1\leq l\leq p^2K}\mathbb{E}\left(\left|\frac{1}{\tilde{n}^{1/2}}\sum_{t=1}^{\tilde{n}} u_{l,t}^{-}\right|^2\right) \leq C\tilde{n}D_n^2\exp(-CD_n^{\lambda_1/2})/\epsilon_1^2.$$

Combining the result from (3.7), by taking $\epsilon_1 = (\log p)^{-1}$ and $D_n = C(\log p)^{2/\lambda_1}$ for some sufficiently large C, we have $d_2 \leq d_4 + o(1)$. Now it left to show $d_4 = o(1)$. We wish to apply Lemma 3.1 to this case, so we need to verify the conditions of Lemma 3.1. If we assume $(r/q)\log^2 p \leq Cn^{-3c_2}$ and $rD_n\log^{3/2} p + qD_n\log^{1/2} p \leq Cn^{1/2-3c_2/2}$ for some constant

$c_2 \in (0, 1/3)$, it holds that

$$d_4 \leq C\left\{n^{-c_2/2} + \left(\frac{qD_n^2\log^7(pn)}{n}\right)^{1/6}\right\} + 2(H-1)b_r. \tag{3.8}$$

To make p diverges as fast as possible, we take $r \asymp (\log n)^{c_3}$ for some large constant $c_3 > 0$. Since $D_n = C(\log p)^{2/\lambda_1}$, the above conditions can be simplified as the following:

$$C(\log n)^{c_3}(\log p)^2 n^{3c_2} \leq q$$
$$C(\log n)^{c_3}(\log p)^{5/2+2/\lambda_1} \leq n^{1/2-9c_2/2}$$

Thus, we need $\log p \leq Cn^{\omega_1}$ where $\omega_1 = \frac{\lambda_1(1-9c_2)}{5\lambda_1+4}$. Moreover, according to Condition 2, it holds that $2(H-1)b_r \leq Cq\exp(-C_3 r^{\lambda_2}) \to o(1)$. Then the right side of (3.8) converges to zero provided that $\log p \leq Cn^{\omega_2}$ where $\omega_2 = \frac{\lambda_1(1-9c_2)}{9\lambda_1+4}$. Therefore, by combining these two conditions, we have $d_4 = o(1)$ if $\log p = o(n^{\lambda_1/(4+9\lambda_1)})$. Now it left to bound d_3. Let Ξ_u and Ξ_Z be the covariance matrices for u and Z respectively. According to Lemma 3.3, it holds that

$$d_3 \leq C|\Xi_u - \Xi_Z|_\infty^{1/3}\{1 \vee \log(p/(|\Xi_u - \Xi_Z|_\infty))\}^{2/3}.$$

This is the same as to bound $|\tilde{W} - W|_\infty$ in the proof for Theorem 1 of Chang et al. (2018). Verifying the choice of q, r, D_n above, we can conclude that $d_3 = o(1)$. This completes the proof of Proposition 3.1.

$\square$

Below is the Lemma used in the proof of Proposition 3.1.

Lemma 3.1. *Let $X_1, \ldots, X_n$ be dependent random variables in $\mathbb{R}^p$ with zero mean and $M = \max_{1\leq j\leq p}\sqrt{n}\sum_{i=1}^n x_{ij}$, $m = -\min_{1\leq j\leq p}\sqrt{n}\sum_{i=1}^n x_{ij}$. Assume there exists $D_n \geq 1$ such that $|X_{ij}| \leq D_n$. Let $S_l = \sum_{i\in I_l} X_i$, $S_l' = \sum_{i\in J_l} X_i$ and $Y = (Y_1, \ldots, Y_p)^T$ be a normal random vector with mean zero and covariance matrix $E(YY') = \frac{1}{mq}\sum_{l=1}^m E(S_l S_l')$. Let $M_Y = \max_{1\leq j\leq p} Y_j$ and $m_Y = -\min_{1\leq j\leq p} Y_j$. Suppose that there exist constants $0 < c_1 \leq C_1$ and $0 < c_2 < 1/9$ such that $c_1 \leq \underline{\sigma}^2(q) \leq \bar{\sigma}^2(r) \vee \bar{\sigma}^2(q) \leq C_1$, $(r/q)\log^2 p \leq C_1 n^{-3c_2}$ and $rD_n\log^{\frac{3}{2}} p + qD_n\log^{\frac{1}{2}} p \leq C_1 n^{\frac{1}{2}-\frac{3c_2}{2}}$. Then there exist a constant C depending only on*

c_1, c_2, C_1 *such that*

$$\sup_{t_1,t_2\in\mathbb{R}} |\mathbb{P}(M \le t_1, m \le t_2) - \mathbb{P}(M_Y \le t_1, m_Y \le t_2)| = C\left\{ n^{-c_2/2} + \left(\frac{qD_n^2\log^7(pn)}{n}\right)^{1/6}\right\} + 2(H-1)b_r. \tag{3.9}$$

Proof. Let $\{\tilde{S}_h\}$ and $\{\tilde{S}'_h\}$, $h = 1, \ldots, H$ be two independent sequences of p-dimensional random vectors such that $\tilde{S}_h$ has the identical distribution with S_h and $\tilde{S}'_h$ has the identical distribution with S'_h. We can proceed the proof in the similar manner as the proof for Theorem B.1. of Chernozhukov et al. (2016). First we want to reduce the sum of dependent data to sum of independent blocks. Notice that

$$\left|\max_{1\le j\le p}\sum_{i=1}^n x_{ij} - \max_{1\le j\le p}\sum_{h=1}^H S_{lj}\right| \le \max_{1\le j\le p}\left|\sum_{h=1}^H S'_{hj}\right| + \max_{1\le j\le p}\left|S'_{H+1,j}\right|.$$

If we write $-S_h = \sum_{i\in I_h} -X_i$ and $-S'_h = \sum_{i\in J_h} -X_i$ and define $-\tilde{S}_h$ and $-\tilde{S}'{}_h$ accordingly, the similar inequality can be obtained:

$$\left|\max_{1\le j\le p}\sum_{i=1}^n (-x_{ij}) - \max_{1\le j\le p}\sum_{h=1}^H (-S_{hj})\right| \le \max_{1\le j\le p}\left|\sum_{h=1}^m S'_{hj}\right| + \max_{1\le j\le p}\left|S'_{H+1,j}\right|.$$

Thus, for every $\delta_1, \delta_2 > 0$,

$$\begin{aligned}\mathbb{P}(M \le t_1, m \le t_2) \le &\mathbb{P}\left(\max_{1\le j\le p}\frac{1}{\sqrt{n}}\sum_{h=1}^H \tilde{S}_{hj} \le t_1 + \delta_1 + \delta_2, \max_{1\le j\le p}\frac{1}{\sqrt{n}}\sum_{h=1}^H -\tilde{S}_{hj} \le t_2 + \delta_1 + \delta_2\right)\\ &+ \mathbb{P}\left(\max_{1\le j\le p}\left|\frac{1}{\sqrt{n}}\sum_{h=1}^H \tilde{S}'_{hj}\right| > \delta_1\right) + \mathbb{P}\left(\max_{1\le j\le p}\left|\frac{1}{\sqrt{n}}\sum_{h=1}^H -\tilde{S}'_{hj}\right| > \delta_1\right)\\ &+ \mathbb{P}\left(\max_{1\le j\le p}\left|\frac{1}{\sqrt{n}}S'_{H+1,j}\right| > \delta_2\right) + \mathbb{P}\left(\max_{1\le j\le p}\left|\frac{1}{\sqrt{n}} - S'_{H+1,j}\right| > \delta_2\right)\\ &+ 2(H-1)b_r.\end{aligned}$$

Specifically, the fourth part is a result from Corollary 2.7 of Eberlein (1984),

$$\sup_{t_1,t_2\in\mathbb{R}}\Bigg|\mathbb{P}\Bigg(\max_{1\leq j\leq p}\sum_{h=1}^{H}S_{hj}\leq t_1,\max_{1\leq j\leq p}\sum_{h=1}^{H}(-S_{hj})\leq t_2\Bigg) - \mathbb{P}\Bigg(\max_{1\leq j\leq p}\sum_{h=1}^{H}\tilde{S_{hj}}\leq t_1,\max_{1\leq j\leq p}\sum_{h=1}^{H}(-\tilde{S}_{hj})\leq t_2\Bigg)\Bigg|\leq (H-1)b_r;$$

$$\sup_{t_1,t_2\in\mathbb{R}}\Bigg|\mathbb{P}\Bigg(\max_{1\leq j\leq p}\sum_{h=1}^{H}S'_{hj}\leq t_1,\max_{1\leq j\leq p}\sum_{h=1}^{H}(-S'_{hj})\leq t_2\Bigg) - \mathbb{P}\Bigg(\max_{1\leq j\leq p}\sum_{h=1}^{H}\tilde{S'_{hj}}\leq t_1,\max_{1\leq j\leq p}\sum_{h=1}^{H}(-\tilde{S}'_{hj})\leq t_2\Bigg)\Bigg|\leq (H-1)b_q.$$

Since $|S_{H+1,j}|\leq (q+r-1)D_n$, by taking $\delta_2=(q+r-1)D_n/\sqrt{n}$, the third part becomes

$$\mathbb{P}\left(\max_{1\leq j\leq p}\left|\frac{1}{\sqrt{n}}S'_{H+1,j}\right|>\delta_2\right)=\mathbb{P}\left(\max_{1\leq j\leq p}\left|\frac{1}{\sqrt{n}}-S'_{H+1,j}\right|>\delta_2\right)=0.$$

Note that since $qD_n\log^{1/2}p\leq Cn^{1/2-3c_2/2}$, it holds that $\delta_2\leq Cn^{-c_2}\log^{-1/2}p$. Moreover, by Markov's inequality and taking $\delta_1=\epsilon_2^{-1}\mathbb{E}(\max_{1\leq j\leq p}|n^{-1/2}\sum_{h=1}^{H}\tilde{S'_{hj}}|)$, we have

$$\mathbb{P}\left(\max_{1\leq j\leq p}\left|\frac{1}{\sqrt{n}}\sum_{h=1}^{H}\tilde{S'_{hj}}\right|>\delta_1\right)\leq\epsilon_2,$$

for any $\epsilon_2>0$. According to Lemma A.3. of Chernozhukov et al. (2016),

$$\mathbb{E}\left(\max_{1\leq j\leq p}\left|n^{-1/2}\sum_{h=1}^{H}\tilde{S'_{hj}}\right|\right)\leq K\left(\sqrt{(r/q)\hat{\sigma}^2(r)\log p}+n^{-1/2}rD_n\log p\right),$$

where K is a universal constant. Based on assumptions that $rD_n\log^{\frac{3}{2}}p\leq n^{1/2-3c_2/2}$ and $(r/q)\log^2 p\leq Cn^{-3c_2}$, the right side is bounded by $Cn^{-3c_2/2}\log^{-1/2}p$. Therefore, by taking $\epsilon_2=n^{-c_2/2}$, we have $\delta_1\leq Cn^{-c_2}\log^{-1/2}p$, so that

$$\mathbb{P}(M\leq t_1,m\leq t_2)\leq\mathbb{P}\Bigg(\max_{1\leq j\leq p}\frac{1}{\sqrt{n}}\sum_{h=1}^{H}\tilde{S_{hj}}\leq t_1+Cn^{-c_2}\log^{-1/2}p, \max_{1\leq j\leq p}\frac{1}{\sqrt{n}}\sum_{h=1}^{H}-\tilde{S}_{hj}\leq t_2+Cn^{-c_2}\log^{-1/2}p\Bigg)+n^{-c_2/2}+2(H-1)b_r.$$

The other direction can be proved in the similar manner. The next step is to adopt normal approximation to the sum of independent blocks. Since $\tilde{S_h}$ are independent and the covariance matrix of $\sqrt{mq/n}Y$ is the same as the covariance matrix of $n^{-1/2}\sum_{h=1}^{H}\tilde{S}_h$, we can directly apply the high-dimensional CLT for Hyperrectangles in Proposition 2.1 of Chernozhukov et al. (2017) if the conditions are satisfied. Specifically, these conditions can be verified by taking $B_n = \sqrt{q}D_n$. It follows that

$$\begin{aligned}\sup_{t_1,t_2\in\mathbb{R}}\Bigg|\mathbb{P}\Bigg(&\max_{1\le j\le p}\frac{1}{\sqrt{n}}\sum_{h=1}^{H}\tilde{S_{hj}}\le t_1,\max_{1\le j\le p}\frac{1}{\sqrt{n}}\sum_{h=1}^{H}-\tilde{S}_{hj}\le t_2\Bigg)\\ &-\mathbb{P}\Bigg(\max_{1\le j\le p}\sqrt{mq/n}Y_j\le t_1,\max_{1\le j\le p}\sqrt{mq/n}(-Y_j)\le t_2\Bigg)\Bigg|\le C\left(\frac{qD_n^2\log^7(pn)}{n}\right)^{1/6}.\end{aligned}$$

Next we need to verify the anti-concentration of M_Y and m_Y. Similar to (3.7), we may apply the anti-concentration inequality from Chernozhukov et al. (2015) again. For any $\delta\to 0$, we have

$$\sup_{t_1,t_2\in\mathbb{R}}\mathbb{P}\left(|M_Y-t_1|\le\delta,|m_Y-t_2|\le\delta\right)\le C\delta\log^{1/2}p. \tag{3.10}$$

Thus, by taking $\delta = Cn^{-c_2}\log^{-1/2}p$, the right side is bounded by Cn^{-c_2}. Now it left to replace $\sqrt{mq/n}$ by 1. For any $\epsilon_3>0$ and $t_1,t_2\in\mathbb{R}$,

$$\begin{aligned}&\left|\mathbb{P}\left(\sqrt{\frac{mq}{n}}Y_M\le t_1,\sqrt{\frac{mq}{n}}Y_m\le t_2\right)-\mathbb{P}\left(Y_M\le t_1,Y_m\le t_2\right)\right|\\ \le&\mathbb{P}\left(|M_Y-t_1|\le\epsilon_3,|m_Y-t_2|\le\epsilon_3\right)\\ &+\mathbb{P}\left(\left(1-\sqrt{\frac{mq}{n}}\right)|Y_M|>\epsilon_3\right)+\mathbb{P}\left(\left(1-\sqrt{\frac{mq}{n}}\right)|Y_m|>\epsilon_3\right).\end{aligned}\tag{3.11}$$

Observe that

$$1-\sqrt{\frac{mq}{n}}\le 1-\frac{mq}{n}\le 1-\left(\frac{n}{q+r}-1\right)\left(\frac{q}{n}\right)=\frac{r}{q+r}+\frac{q}{n}.$$

It is obvious that $r/(q+r)\le Cn^{-2c_2}\log^{-1}p$. Moreover, since $qD_n\log^{\frac{1}{2}}p\le Cn^{1/2-3c_2/2}$, we have $q\log p\le q^2D_n^2\log p\le Cn^{1-2c_2}$. Therefore, it holds that $1-\sqrt{\frac{mq}{n}}\le Cn^{-c2_2}\log^{-1}p$.

In addition, since $\mathbb{E}(|\max_{1\le j\le p} Y_j|) \le C\log^{1/2} p$, by Markov's inequality, we have

$$\mathbb{P}\left(\left|\max_{1\le j\le p} Y_j\right| > n^{c_2}\log^{1/2} p\right) \le Cn^{-c_2}.$$

It follows that

$$\mathbb{P}\left(\left(1-\sqrt{\frac{mq}{n}}\right)|Y_M| > Cn^{-c_2}\log^{-1/2} p\right) \le Cn^{-c_2}.$$

Finally, combining the above result and (3.10) with $\epsilon_3 = Cn^{-c_2/4}\log^{-1/2} p$, (3.11) becomes

$$\sup_{t_1,t_2\in\mathbb{R}}\left|\mathbb{P}\left(\sqrt{\frac{mq}{n}}Y_M \le t_1, \sqrt{\frac{mq}{n}}Y_m \le t_2\right) - \mathbb{P}\left(Y_M \le t_1, Y_m \le t_2\right)\right| \le Cn^{-c_2}.$$

This completes the proof for Lemma 3.1. □

Proposition 3.2. *Assume the kernel function $\mathcal{K}(\cdot)$ satisfies $|\mathcal{K}(\cdot)| \asymp |x|^{-\tau}$ as $x\to\infty$ for some $\tau > 1$, and the bandwidth $w_n \asymp n^\rho$ for some $0 < \rho < \min\{\frac{\tau-1}{3\tau}, \frac{\lambda_2}{2\lambda_2+1}\}$. Under conditions in Proposition 3.1, it holds that*

$$\sup_{t_1,t_2\in\mathbb{R}}\left|\mathbb{P}(\hat{M}\le t_1, \hat{m}\le t_2) - \mathbb{P}(\hat{M}_Z \le t_1, \hat{m}_Z \le t_2)\right| \to 0. \tag{3.12}$$

provided that $\log p = o(n^\omega)$ where ω is a positive constant specified in the proof of this Proposition.

Proof. Combining the result from Proposition 3.1, we proceed the proof in two parts. First we want to show

$$\sup_{t_1,t_2\in\mathbb{R}}\left|\mathbb{P}(\hat{M}\le t_1, \hat{m}\le t_2) - \mathbb{P}(M_Z \le t_1, m_Z \le t_2)\right| \le d_0 + o_p(1).$$

The second step is to show

$$\sup_{t_1,t_2\in\mathbb{R}}\left|\mathbb{P}(M_Z\le t_1, m_Z\le t_2) - \mathbb{P}(\hat{M}_Z \le t_1, \hat{m}_Z \le t_2)\right| = o_p(1).$$

Following the similar argument in Proposition 3.1, it holds that for any $\epsilon > 0$

$$\sup_{t_1,t_2\in\mathbb{R}} |\mathbb{P}(\hat{M} \le t_1, \hat{m} \le t_2) - \mathbb{P}(M_Z \le t_1, M_Z \le t_2)|$$
$$\le d_0 + \mathbb{P}(|\hat{M} - M| > \epsilon) + \mathbb{P}(|\hat{m} - m| > \epsilon) + C\epsilon(\log(p/\epsilon))^{1/2}. \quad (3.13)$$

According to Lemma 2 of Chang et al. (2017), there exists a constant c_1 satisfying $n^{1-c_1}(\log p)^{1/2} \to \infty$, such that

$$\mathbb{P}\left(\max_{\substack{1\le i,j\le p\\ 1\le k\le K}} |\hat{\gamma}_{ij}(k)| > Cn^{-c_1}(\log p)^{1/2}\right) \le Cp^{-1}.$$

It suffices to bound $\left|[\hat{\tau}_{\chi(l)}]^{-1/2} - [\tau_{\chi(l)}]^{-1/2}\right|$ for all $1 \le l \le p^2K$. Recall

$$\Xi = \mathbb{E}\left\{\left(\frac{1}{\tilde{n}^{1/2}}\sum_{t=1}^{\tilde{n}}\xi_t\right)\left(\frac{1}{\tilde{n}^{1/2}}\sum_{t=1}^{\tilde{n}}\xi_t,\right)'\right\},$$

and the diagonal element of Ξ is $\tau_{\chi(l)}$. Write

$$\hat{\Xi} = \sum_{k=-\tilde{n}+1}^{\tilde{n}-1}\mathcal{K}\left(\frac{k}{w_n}\right)\hat{\Gamma}_k,$$

where

$$\hat{\Gamma}_k = \begin{cases} \frac{1}{\tilde{n}}\sum_{t=k+1}^{\tilde{n}}\xi_t\xi_{t-k}', & k \ge 0; \\ \frac{1}{\tilde{n}}\sum_{t=-k+1}^{\tilde{n}}\xi_{t+k}\xi_t', & k < 0. \end{cases}$$

Our goal is to show $|\hat{\Xi} - \Xi|_\infty = o_p(1)$. Following the similar argument in the proof for Theorem 2 of Chang et al. (2018), we can define

$$\tilde{\Xi} = \sum_{k=-\tilde{n}+1}^{\tilde{n}-1}\mathcal{K}\left(\frac{k}{w_n}\right)\Gamma_k,$$

where

$$\Gamma_k = \begin{cases} \frac{1}{\tilde{n}}\sum_{t=k+1}^{\tilde{n}}\mathbb{E}(\xi_t\xi_{t-k}'), & k \ge 0; \\ \frac{1}{\tilde{n}}\sum_{t=-k+1}^{\tilde{n}}\mathbb{E}(\xi_{t+k}\xi_t'), & k < 0. \end{cases}$$

Proposition 1(b) of Andrews (1991) shows that $|\tilde{\Xi} - \Xi|_\infty \to 0$. According to Lemma 4 of

Chang et al. (2018), provided that $\log p \le Cn^{C\delta}$ for $\delta = \min\{\frac{\lambda_1}{\lambda_1+8}(2\alpha\lambda_2+\alpha-1), \frac{\lambda_1}{8}[(\alpha-\rho)\tau+\alpha+\alpha\lambda_2+\rho-2]\}$, it holds that

$$\mathbb{P}\left(|\hat{\Xi}-\tilde{\Xi}|_\infty > C\{\log(pn)\}^{4/\lambda_1} n^{-f(\alpha_0)/2}\right) \le Cp^{-1},$$

where $\kappa = \max\{\frac{1}{2\lambda_2+1}, \frac{\rho\tau-\rho+2}{\tau+1+\lambda_2}, \frac{\rho\tau+1}{\tau}\}$ and α_0 is the maximizer for the function $f(\alpha) = \min\{1-\alpha-2\rho, 2(\alpha-\rho)\tau-2\}$ over $\kappa < \alpha < 1-2\rho$. Therefore, $|\hat{\Xi}-\Xi|_\infty$ goes to zero with the same convergence rate. Note this convergence rate is uniform for each component of $\hat{\Xi}-\Xi$. Observe that

$$\hat{\tau}_{\chi(l)} - \tau_{\chi(l)} = \left([\hat{\tau}_{\chi(l)}]^{1/2} - [\tau_{\chi(l)}]^{1/2}\right)^2 + 2\tau_{\chi(l)}\left(\hat{\tau}_{\chi(l)} - \tau_{\chi(l)}\right),$$

then it follows that

$$\mathbb{P}\left(\max_{1\le l\le p^2K}\left|[\hat{\tau}_{\chi(l)}]^{1/2} - [\tau_{\chi(l)}]^{1/2}\right| > C\{\log(pn)\})^{4/\lambda_1} n^{-f(\alpha_0)/2}\right) \le Cp^{-1}.$$

Moreover, since $[\hat{\tau}_{\chi(l)}]^{-1/2} - [\tau_{\chi(l)}]^{-1/2} = -\left([\hat{\tau}_{\chi(l)}]^{1/2} - [\tau_{\chi(l)}]^{1/2}[\hat{\tau}_{\chi(l)}]^{-1/2}\right)[\tau_{\chi(l)}]^{-1/2}$, it holds with at least probability $1-Cp^{-1}$ that

$$\max_{1\le l\le p^2K}\left|[\hat{\tau}_{\chi(l)}]^{-1/2} - [\tau_{\chi(l)}]^{-1/2}\right| \le C\{\log(pn)\}^{4/\lambda_1} n^{-f(\alpha_0)/2}.$$

Therefore, it holds with probability at least $1-Cp^{-1}$ that

$$\begin{aligned}|\hat{M}-M| &\le \max_{1\le l\le p^2K}\left|[\hat{\tau}_{\chi(l)}]^{-1/2} - [\tau_{\chi(l)}]^{-1/2}\right| \cdot \max_{\substack{1\le i,j\le p\\1\le k\le K}}|\hat{\gamma}_{ij}(k)| \\ &\le (\log(pn))^{4/\lambda_1+1/2} n^{-f(\alpha_0)/2-c_1}.\end{aligned}$$

Thus, by taking $\epsilon = (\log(pn))^{4/\lambda_1+1/2} n^{-f(\alpha_0)/2-c_1}$, (3.13) implies that

$$\begin{aligned}&\sup_{t_1,t_2\in\mathbb{R}}|\mathbb{P}(\hat{M}\le t_1, \hat{m}\le t_2) - \mathbb{P}(M_Z\le t_1, M_Z\le t_2)| \\ &\qquad \le d_0 + Cp^{-1} + (\log(pn))^{4/\lambda_1+1} n^{-f(\alpha_0)/2-c_1}.\end{aligned}$$

To make right side can decay to zero and p can diverge at exponential rate of n, we need to assume $-\frac{f(\alpha_0)}{2} < c_1 < 1 + \frac{\lambda_1(2+f(\alpha_0))}{16+2\lambda_1}$ and $\log p = o(n^{\omega_3})$ where $\omega_3 = \frac{\lambda_1(2c_1+f(\alpha_0))}{8+2\lambda_1}$. Now it left to prove $\sup_{t_1,t_2\in\mathbb{R}} \left|\mathbb{P}(M_Z \le t_1, m_Z \le t_2) - \mathbb{P}(\hat{M}_Z \le t_1, \hat{m}_Z \le t_2)\right| = o_p(1)$. Similarly, for any $\epsilon > 0$,

$$\begin{aligned}
&\sup_{t_1,t_2\in\mathbb{R}} \left|\mathbb{P}(M_Z \le t_1, m_Z \le t_2) - \mathbb{P}(\hat{M}_Z \le t_1, \hat{m}_Z \le t_2)\right| \\
\le&\mathbb{P}(|\hat{M}_Z - M_Z| > \epsilon) + \mathbb{P}(|\hat{m}_Z - m_Z| > \epsilon) + \sup_{t_1,t_2\in\mathbb{R}} \mathbb{P}(|M_Z - t_1| \le \epsilon, |m_Z - t_2| \le \epsilon).
\end{aligned}$$

Since $\mathbb{E}(|\max_{1\le l\le p^2K} G_l|) \le C\log^{1/2} p$, by Markov's inequality, it holds that

$$\mathbb{P}\left(\left|\max_{1\le l\le p^2K} G_l\right| > n^{c_2}(\log p)^{1/2}\right) \le Cn^{-c_2},$$

where c_2 is a positive constant. Thus, it holds with probability at least $1 - Cn^{-c_2}$ that

$$\begin{aligned}
|\hat{M}_Z - M_Z| &\le \max_{1\le l\le p^2K} \left|[\hat{\tau}_{\chi(l)}]^{-1/2} - [\tau_{\chi(l)}]^{-1/2}\right| \cdot \left|\max_{1\le l\le p^2K} G_l\right| \\
&\le (\log(pn))^{4/\lambda_1+1/2} n^{-f(\alpha_0)/2+c_2}.
\end{aligned}$$

Therefore, by taking $\epsilon = (\log(pn))^{4/\lambda_1+1} n^{-f(\alpha_0)/2+c_2}$, it follows from anti-concentration that

$$\begin{aligned}
\sup_{t_1,t_2\in\mathbb{R}} &|\mathbb{P}(M_Z \le t_1, m_Z \le t_2) - \mathbb{P}(\hat{M}_Z \le t_1, \hat{m}_Z \le t_2)| \\
&\le Cn^{-c_2} + C(\log(pn))^{4/\lambda_1+1} n^{-f(\alpha_0)/2+c_2}.
\end{aligned}$$

The right side decays to zero when $\log p \le Cn^{\omega_4}$ where $\omega_4 = \frac{\lambda_1 f(\alpha_0)}{8+2\lambda_1}$ (note that ω_4 takes its supremum when $c_2 = 0$). As a result, (3.12) holds provided that $\log p = o(n^{\omega})$ where $\omega = \min\{\lambda_1 f(\alpha_0)/(8+2\lambda_1), \lambda_1/(4+9\lambda_1)\}$. This completes the proof for Proposition 3.2. $\square$

In summary, the difference between the joint distribution of $\hat{M}, \hat{m}$ and that of $\hat{M}_Z, \hat{m}_Z$ converges to 0 under some mild conditions. However, in practice, the test statistics we use is $\hat{M}^2 + \hat{m}^2$. There is still some gap to be shown. First of all, according to Lemma 6 of Xiao and Wu (2013), the asymptotic distribution of the maximum self-normalized autocovariance converges to Gumbel distribution, provided that some dependence conditions on the covariance matrix Ξ are satisfied. Specifically, let $\mathcal{I}_n = \{(i,j,k) : 1 \leq i,j \leq p_n, 1 \leq k \leq K\}$ and define $v_\alpha = \sum_t (x_{i,t+k}x_{jt})/\sqrt{n}$ for some $\alpha = (i,j,k) \in \mathcal{I}_n$. Note that for simplicity, we assume all the covariance pair are centered with mean zero. Then we need the following technical condition:

Condition 3.

$$\sum_{\alpha \neq \beta \in \mathcal{I}_n} \mathrm{Cov}(v_\alpha, v_\beta)^2 = O(p_n^{4-\delta}), \tag{3.14}$$

for some $\delta > 0$ and

$$\lim_n \sup_{\alpha \neq \beta \in \mathcal{I}_n} \mathrm{Cov}(v_\alpha, v_\beta) < 1. \tag{3.15}$$

Recall n is the number of observed data and p is the dimension which grows to infinity as n goes to infinity. Under this condition, there exists some constant sequence $\{b_n\}$ such that

$$\lim_{n \to \infty} \mathbb{P}(\hat{M}_Z^2 - b_n \leq y) = \exp(-e^{y/2}).$$

Remark 3.2. *For convenience reason, we require* $\mathrm{Cov}(v_\alpha, v_\beta) < 1$ *for all* $\{\alpha \neq \beta \in \mathcal{I}_n\}$ *when* $n \to \infty$*. However, it is possible that some covariances to be 1 but the above result still holds.*

Similarly, the limiting distribution of the minimum extreme $\hat{m}_Z$ can be shown to be Gumbel distribution. Furthermore, Marques et al. (2015) derived the near-exact approximations for the distribution of linear combinations of independent Gumbel random variables. According to Section 3.3.3, the limiting distribution of $\hat{M}_Z$ and that of $\hat{m}_Z$ are asymptotically independent. Therefore, it is feasible to derive the limiting distribution of $\hat{M}_Z^2 + \hat{m}_Z^2$ according to Marques et al. (2015). Note that there is no closed form for the distribution of convolution of Gumbels. For convenience, we denote it as $G_1 * G_1'$, where $G_1(y) = \exp(-e^{-y})$. Then it comes to our main result regarding the limiting distribution of the test statistic $\hat{M}^2 + \hat{m}^2$.

Theorem 3.1. *Let* $N := p^2K$*. Assume the conditions of Proposition 3.2 hold, then under condition 3, it holds that for any* $y \in \mathbb{R}$*,*

$$\mathbb{P}(\hat{M}^2 + \hat{m}^2 - 4\log N + 2\log(\log N) + 2\log(4\pi) \leq y) \to G_1(y/2) * G_1'(y/2), \tag{3.16}$$

where $G_1(y) = \exp(-e^{-y})$, provided that $\log p = o(n^{\omega})$ for some positive constant ω specified in Proposition 3.2.

The above Theorem is an immediate result from Proposition 3.2.

Remark 3.3. *The asymptotic distribution of T_n is the extreme value distribution of type I. We can proceed the proof in the similar manner as Theorem 3.1. Here we state the result without proof. Under the same conditions as Theorem 3.1, it holds that*

$$\mathbb{P}(T_n^2 - 4\log N + \log(\log N) + \log(8\pi) \leq y) \to G_1(y/2).$$

3.3.2 Dependent Gaussian Multiplier Bootstrap

In this section, we introduce the dependent Gaussian multiplier bootstrap to evaluate the critical values. First, there are some preliminary results for comparison bound for distributions of Gaussian maximums. Let $X = (X_1, \ldots, X_p)^T$ and $Z = (Z_1, \ldots, Z_p)^T$ be centered Gaussian random vectors in $\mathbb{R}^p$ with covariance matrices $\Sigma_X = (\sigma_{jk}^X)_{1\leq j,k\leq p}$ and $\Sigma_Z = (\sigma_{jk}^Z)_{1\leq j,k\leq p}$ respectively. Let $M = \max_{1\leq j\leq p} X_j, m = \min_{1\leq j\leq p} X_j, M_Z = \max_{1\leq j\leq p} Z_j$ and $m_Z = \min_{1\leq j\leq p} Z_j$. Similar to Theorem 1 of (Chernozhukov et al., 2015), consider a smooth function $F : \mathbb{R}^p \to \mathbb{R}$, which approximate the maximum function:

$$F_\beta(w) := \beta^{-1}\log\left(\sum_{j=1}^{p}\exp(\beta w_j)\right),$$

where $w = (w_1, \cdots, w_p)^T$ and $\beta > 0$ is the smoothing parameter. In Lemma 3.2, we derive the bound for the difference between smooth functions. Based on Lemma 3.2, we can derive the upper bound on the Kolmogorov distance between the joint distribution of M, m and that of M_Z, m_Z, which is stated in Lemma 3.3.

Lemma 3.2 (Comparison bounds for smooth functions)**.** *For every $g \in \mathbb{C}^2(\mathbb{R}^2)$ and $\beta > 0$,*

$$\begin{aligned}&|\mathbb{E}[g(F_\beta(X), F_\beta(-X)) - g(F_\beta(Z), F_\beta(-Z))]|\\ \leq&\triangle \times \left\{\frac{1}{2}\Big(||g_{11}||_\infty + ||g_{12}||_\infty + ||g_{21}||_\infty + ||g_{22}||_\infty\Big) + \beta\Big(||g_1||_\infty + ||g_2||_\infty\Big)\right\},\end{aligned}$$

where $g_1 = \frac{\partial g(\cdot)}{\partial F_\beta(w)}$, $g_2 = \frac{\partial g(\cdot)}{\partial F_\beta(-w)}$ *and* $g_{11} = \frac{\partial^2 g(\cdot)}{\partial (F_\beta(w))^2}$, $g_{12} = \frac{\partial^2 g(\cdot)}{\partial F_\beta(w)\partial F_\beta(-w)}$, $g_{21} = \frac{\partial^2 g(\cdot)}{\partial F_\beta(-w)\partial F_\beta(w)}$, $g_{22} = \frac{\partial^2 g(\cdot)}{\partial (F_\beta(-w))^2}$.

Proof. Following the similar argument in proof of Theorem 1 of Chernozhukov et al. (2015), here $g : \mathbb{R}^2 \to \mathbb{R}$ is a bivariate function with finite first and second partial derivatives. Write $\tilde{F}_\beta(W) = (F_\beta(W), F_\beta(-W))$. Without loss of generality, assume X and Z are independent. The Slepian interpolation between X and Z is $W(t) = \sqrt{t}X + \sqrt{1-t}Z$, $t \in [0,1]$. Let $m := g \circ \tilde{F}_\beta$ and $\Psi(t) := \mathbb{E}[m(W(t))]$. Then we have

$$|\mathbb{E}[m(X)] - \mathbb{E}[m(Z)]| = |\Psi(1) - \Psi(0)| = \left| \int_0^1 \Psi'(t)dt \right|.$$

Taking the first derivative of $\Psi(t)$ and then applying the Lemma 2 (Stein's identity) of (Chernozhukov et al., 2015), we have

$$\begin{aligned}\Psi'(t) &= \frac{1}{2}\sum_{j=1}^{p} \mathbb{E}\left[\frac{\partial m(W(t))}{\partial w_j} \cdot (t^{-1/2}X_j - (1-t)^{-1/2}Z_j)\right] \\ &= \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p}(\sigma_{jk}^X - \sigma_{jk}^Z)\mathbb{E}\left[\frac{\partial^2 m(W(t))}{\partial w_j \partial w_k}\right].\end{aligned}$$

It follows that

$$\begin{aligned}\left|\int_0^1 \Psi'(t)dt\right| &\le \frac{1}{2}\sum_{j,k=1}^{p} \left|\sigma_{jk}^X - \sigma_{jk}^Z\right| \cdot \left|\int_0^1 \mathbb{E}\left[\frac{\partial^2 m(W(t))}{\partial w_j \partial w_k}\right] dt\right| \\ &\le \frac{\triangle}{2}\int_0^1 \sum_{j,k=1}^{p}\left|\mathbb{E}\left[\frac{\partial^2 m(W(t))}{\partial w_j \partial w_k}\right]\right| dt. \qquad (3.17)\end{aligned}$$

For any function $f : \mathbb{R}^p \to \mathbb{R}$, write $\partial_j f(w) = \partial f(w)/\partial w_j$ and $|f|_\infty = \sup_{x \in \mathbb{R}} |f(x)|$. For every $1 \le j,k \le p$, taking the first and second derivatives:

$$\partial_j F_\beta(w) = \pi_j(w), \quad \partial_j \partial_k F_\beta(w) = \beta\theta_{jk}(w),$$

$$\partial_j F_\beta(-w) = -\pi_j(-w), \quad \partial_j \partial_k F_\beta(-w) = \beta\theta_{jk}(-w),$$

where

$$\pi_j(w) := e^{\beta w_j} / \sum_{m=1}^{p} e^{\beta w_m}, \quad \theta_{jk}(w) := 1(j = k)\pi_j(w) - \pi_j(w)\pi_k(w).$$

Then, following the chain rule, we have

$$\begin{aligned}\partial_j \partial_k m(w) =& \{g_{11}\pi_k(w) + g_{12}(-\pi_k(-w))\} \cdot \pi_j(w) + \beta g_1 \theta_{jk}(w) \\ &+ \{g_{21}\pi_k(w) + g_{22}(-\pi_k(-w))\} \cdot (-\pi_j(-w)) + \beta g_2 \theta_{jk}(-w)\end{aligned}$$

Since $\sum_{j=1}^{p} \pi_j(w) = 1$ and $\sum_{j,k=1}^{p} |\theta_{jk}(w)| \leq 2$, it holds that

$$\begin{aligned}\sum_{j,k=1}^{p} |\partial_j \partial_k m(W(t))| \leq& |g_{11}(W(t))| + |g_{12}(W(t))| + 2\beta |g_1(W(t))| \\ &+ |g_{21}(W(t))| + |g_{22}(W(t))| + 2\beta |g_2(W(t))|.\end{aligned}$$

Plugging back to (3.17), we can obtain the following bound:

$$\begin{aligned}&|\mathbb{E}[g(F_\beta(X), F_\beta(-X))] - \mathbb{E}[g(F_\beta(Z), F_\beta(-Z))]| \\ \leq& \frac{\triangle}{2} \times \left\{ \int_0^1 \mathbb{E}[|g_{11}(W(t))|]dt + \int_0^1 \mathbb{E}[|g_{12}(W(t))|]dt + 2\beta \int_0^1 \mathbb{E}[|g_1(W(t))|]dt \right. \\ &\left. + \int_0^1 \mathbb{E}[|g_{21}(W(t))|]dt + \int_0^1 \mathbb{E}[|g_{22}(W(t))|]dt + 2\beta \int_0^1 \mathbb{E}[|g_2(W(t))|]dt \right\} \\ \leq& \triangle \times \left\{ \frac{1}{2} \left(||g_{11}||_\infty + ||g_{12}||_\infty + ||g_{21}||_\infty + ||g_{22}||_\infty \right) + 2\beta \left(||g_1||_\infty + ||g_2||_\infty \right) \right\}.\end{aligned}$$

□

Lemma 3.3. *Suppose that $p \geq 2$ and $\sigma_{jj}^Z > 0$ for all $1 \leq j \leq p$. Then*

$$\sup_{x,y \in \mathbb{R}} |\mathbb{P}(M \leq x, m \leq y) - \mathbb{P}(M_Z \leq x, m_Z \leq y)| \leq C \triangle^{1/3} \{1 \vee a_p^2 \vee \log(1/\triangle)\}^{1/3} \log^{1/3} p,$$

where $\triangle := \max_{1 \leq j,k \leq p} |\sigma_{jk}^X - \sigma_{jk}^Z|$, $a_p := \mathbb{E}[\max_{1 \leq j \leq p}(Z_j / \sigma_{jj}^Z)]$ and C is a positive constant only depending on $\min_{1 \leq j \leq p} \sigma_{jj}^Z$ and $\max_{1 \leq j \leq p} \sigma_{jj}^Z$.

Proof. Similar to proof in Theorem 2 of Chernozhukov et al. (2015), define a bivariate step function g to approximate the joint distributional function of M and m. Take

$$g_{x,y,\beta,\delta} = g_0\left(\frac{t_1 - x - e_{p,\beta}}{\delta}, \frac{t_2 - y - e_{p,\beta}}{\delta}\right),$$

where $\beta > 0$ and $e_{p,\beta} := \beta^{-1}\log p$. The base function $g_0 : \mathbb{R}^2 \to [0,1]$ is a bivariate C^2 function such that $g_0(t_1,t_2) = 1$ for $t \leq 0$ and $g_0(t_1,t_2) = 0$ for $t \geq 1$. Note that for $\forall t_1, t_2 \in \mathbb{R}$, it holds that

$$1(t_1 \leq x + e_{p,\beta}, t_2 \leq x + e_{p,\beta}) \leq g_{x,y,\beta,\delta}(t_1,t_2) \leq 1(t_1 \leq x + e_{p,\beta} + \delta, t_2 \leq x + e_{p,\beta} + \delta).$$

Moreover, for every z, observe that

$$0 \leq F_\beta(z) - M \leq \beta^{-1}\log p \quad \text{and} \quad 0 \leq F_\beta(-z) - m \leq \beta^{-1}\log p.$$

It follows that

$$\mathbb{P}(M \leq x, m \leq y) \leq \mathbb{P}(F_\beta(X) \leq x+e_{p,\beta}, F_\beta(-X) \leq x+e_{p,\beta}) \leq \mathbb{E}[g_{x,y,\beta,\delta}(F_\beta(X), F_\beta(-X))],$$

for any $x \in \mathbb{R}$, $\beta > 0$ and $\delta > 0$. For the selected function g, we have $|g_.|_\infty = \delta^{-1}$ and $|g_{..}|_\infty = \delta^{-2}$. According to the bound for smooth function in Lemma 3.2,

$$|\mathbb{E}[g(F_\beta(X), F_\beta(-X)) - g(F_\beta(Z), F_\beta(-Z))]| \leq C\triangle(\delta^{-2} + \beta\delta^{-1})$$

Therefore, we have

$$\begin{aligned}
\mathbb{P}(M \leq x, m \leq y) &\leq \mathbb{E}[g_{x,y,\beta,\delta}(F_\beta(Z), F_\beta(-Z))] + C\triangle(\delta^{-2} + \beta\delta^{-1}) \\
&\leq \mathbb{P}(F_\beta(Z) \leq x + e_{p,\beta} + \delta, F_\beta(-Z) \leq y + e_{p,\beta} + \delta) + C\triangle(\delta^{-2} + \beta\delta^{-1}) \\
&\leq \mathbb{P}(M_Z \leq x + e_{p,\beta} + \delta, m_Z \leq y + e_{p,\beta} + \delta) + C\triangle(\delta^{-2} + \beta\delta^{-1}).
\end{aligned}$$

Now it left to bound the difference between $\mathbb{P}(M_Z \leq x + e_{p,\beta} + \delta, m_Z \leq y + e_{p,\beta} + \delta)$ and $\mathbb{P}(M_Z \leq x, m_Z \leq y)$. By anti-concentration inequality in Theorem 3 of Chernozhukov et al.

(2015), we have

$$\begin{aligned}
&\mathbb{P}(M_Z \leq x + e_{p,\beta} + \delta, m_Z \leq y + e_{p,\beta} + \delta) - \mathbb{P}(M_Z \leq x, m_Z \leq y) \\
\leq &\mathbb{P}(x \leq M_Z \leq x + e_{p,\beta} + \delta) + \mathbb{P}(y \leq m_Z \leq y + e_{p,\beta} + \delta) \\
\leq &C(e_{p,\beta} + \delta)\sqrt{1 \vee a_p^2 \vee \log\{1/\delta\}}.
\end{aligned}$$

Hence, it follows that

$$\begin{aligned}
\mathbb{P}(M \leq x, m \leq y) - \mathbb{P}(M_Z \leq x, &m_Z \leq y) \\
&\leq C\left\{(\delta^{-2}\triangle + \beta\delta^{-1}) + (e_{p,\beta} + \delta)\sqrt{1 \vee a_p^2 \vee \log\{1/\delta\}}\right\}.
\end{aligned}$$

The other direction can be shown in the similar manner.

$$\begin{aligned}
\mathbb{P}(M \leq x, m \leq y) &\geq \mathbb{P}(F_\beta(X) \leq x, F_\beta(-X) \leq y) \\
&\geq \mathbb{E}[g_{x-e_{p,\beta}-\delta, y-e_{p,\beta}-\delta,\beta,\delta}(F_\beta(X), F_\beta(-X))] \\
&\geq \mathbb{E}[g_{x-e_{p,\beta}-\delta, y-e_{p,\beta}-\delta,\beta,\delta}(F_\beta(Z), F_\beta(-Z))] - C\triangle(\delta^{-2} + \beta\delta^{-1}) \\
&\geq \mathbb{P}(F_\beta(Z) \leq x - \delta, F_\beta(-Z) \leq y - \delta) - C\triangle(\delta^{-2} + \beta\delta^{-1}) \\
&\geq \mathbb{P}(M_Z \leq x - e_{p,\beta} - \delta, m_Z \leq y - e_{p,\beta} - \delta) - C\triangle(\delta^{-2} + \beta\delta^{-1})
\end{aligned}$$

Apply the anti-concentration inequality again, we have

$$\begin{aligned}
&\mathbb{P}(M_Z \leq x - e_{p,\beta} - \delta, m_Z \leq y - e_{p,\beta} - \delta) - \mathbb{P}(M_Z \leq x, m_Z \leq y) \\
\geq &-\mathbb{P}(x - e_{p,\beta} - \delta \leq M_Z \leq x) - \mathbb{P}(y - e_{p,\beta} - \delta \leq m_Z \leq y) \\
\geq &-C(e_{p,\beta} + \delta)\sqrt{1 \vee a_p^2 \vee \log\{1/\delta\}}.
\end{aligned}$$

As a result, we can derive the following upper bound:

$$\begin{aligned}
\sup_{x,y\in\mathbb{R}} |\mathbb{P}(M \leq x, m \leq y) - \mathbb{P}(M_Z \leq x, m_Z \leq y)| \leq C\Big\{&(\delta^{-2}\triangle + \beta\delta^{-1}) \\
&+ (e_{p,\beta} + \delta)\sqrt{1 \vee a_p^2 \vee \log\{1/\delta\}}\Big\}.
\end{aligned}$$

Finally, choose $\delta = \triangle^{1/3}(1 \vee a)^{-1/3}(2\log p)^{1/3}$ and $\beta = \delta^{-1}\log p$, we can obtain the bound in Lemma 3.3 as

$$\sup_{x,y\in\mathbb{R}} |\mathbb{P}(M \leq x, m \leq y) - \mathbb{P}(M_Z \leq x, m_Z \leq y)| \leq C\triangle^{1/3}\{1 \vee a_p^2 \vee \log(1/\triangle)\}^{1/3} \log^{1/3} p.$$

This completes the proof. □

In practice, we use the dependent Gaussian multiplier bootstrap to obtain the critical values, which is similar to the one used in Chang et al. (2017). Specifically, let $\beta = (\beta_1, \ldots, \beta_{\tilde{n}})^T \sim N(0, \Theta)$ be a random vector independent of $\{x_t\}$, where $\tilde{n} = n - K$ and Θ is an $\tilde{n} \times \tilde{n}$ matrix with (i, j)-th element $\mathcal{K}\{(i-j)/w_n\}$. Conditioning on the whole dataset $\mathcal{D}$, it is obvious that $\tilde{G} = \frac{1}{\tilde{n}^{1/2}} \sum_{t=1}^{\tilde{n}} \beta_t \xi_t$ has a normal distribution with mean zero and covariance matrix $\hat{\Xi}$, where $\hat{\Xi}$ is the kernel estimate of long-run covariance matrix of autocovariances. Then normalize the l-th component of $\tilde{G}$ with corresponding $\hat{\tau}_{\chi(l)}$ and let

$$\hat{M}_B = \max_{1\leq l\leq p^2K} \hat{B}_l, \quad \hat{m}_B = \max_{1\leq l\leq p^2K} (-\hat{B}_l),$$

where $\hat{B}_l = \tilde{G}_l / \sqrt{\hat{\tau}_{\chi(l)}}$. This is one bootstrap sample. We can repeat this procedure for a large number of times and use the empirical quantile as the critical values. The below Proposition 3.3 shows the difference between the joint distribution of maximum and minimum of normalized Gaussian random vector from G and the joint distribution from $\tilde{G}$ conditioning on the whole dataset $\mathcal{D}$. As a result, we can establish the consistency of this dependent multiplier bootstrap procedure.

Proposition 3.3. *Assume the conditions of Proposition 3.2 and Lemma 3.3 holds, then*

$$\sup_{t_1,t_2\in\mathbb{R}} \left|\mathbb{P}(\hat{M}_B \leq t_1, \hat{m}_B \leq t_2 \mid \mathcal{D}) - \mathbb{P}(\hat{M}_Z \leq t_1, \hat{m}_Z \leq t_2)\right| \xrightarrow{p} 0. \tag{3.18}$$

Proof. Similar to the proof of Proposition 3.2, we first replace the estimated $\hat{\tau}_{\chi(l)}$ with the true standard deviation $\tau_{\chi(l)}$. Then according to Lemma 3.3, the bootstrap estimation error depends mainly on the difference between the empirical and population covariance matrix $|\Xi - \hat{\Xi}|_\infty$, which was bounded in Proposition 3.2. Note that in the worst case, $a_p \leq \sqrt{2\log p}$. Thus we have

$$\sup_{t_1,t_2\in\mathbb{R}} |\mathbb{P}(M_B \le t_1, m_B \le t_2) - \mathbb{P}(M_Z \le t_1, m_Z \le t_2)| \le C|\hat{\Xi}-\Xi|_\infty^{1/3}\{\log(p/|\hat{\Xi}-\Xi|_\infty)\}^{2/3}.$$

Then it left to show

$$\sup_{t_1,t_2\in\mathbb{R}} \left|\mathbb{P}(M_B \le t_1, m_B \le t_2) - \mathbb{P}(\hat{M}_B \le t_1, \hat{m}_B \le t_2)\right| = o_p(1) \tag{3.19}$$

$$\sup_{t_1,t_2\in\mathbb{R}} \left|\mathbb{P}(M_Z \le t_1, m_Z \le t_2) - \mathbb{P}(\hat{M}_Z \le t_1, \hat{m}_Z \le t_2)\right| = o_p(1) \tag{3.20}$$

where $M_B = \max_{1\le l\le p^2K} B_l, m_B = \max_{1\le l\le p^2K}(-B_l)$ and $B_l = \tilde{G}_l/\sqrt{\tau_{\chi(l)}}$. Note that (3.20) has been shown in Proposition 3.2. So it suffices to show (3.19), which can be done in the similar manner as (3.20). Observe that B is drawn from normal distribution. According to Markov's inequality, for constant $c_3 > 0$ we have

$$\mathbb{P}\left(\left|\max_{1\le l\le p^2K} B_l\right| > n^{c_3}(\log p)^{1/2}\right) \le Cn^{-c_3}.$$

Thus, it holds with probability at least $1 - Cn^{-c_3}$ that

$$\begin{aligned}|\hat{M}_B - M_B|_\infty &\le \max_{1\le l\le p^2K}\left|[\hat{\tau}_{\chi(l)}]^{-1/2} - [\tau_{\chi(l)}]^{-1/2}\right| \cdot \left|\max_{1\le l\le p^2K} B_l\right| \\ &\le (\log(pn))^{4/\lambda_1+1/2} n^{-f(\alpha_0)/2+c_3}.\end{aligned}$$

By taking $\epsilon = (\log(pn))^{4/\lambda_1+1} n^{-f(\alpha_0)/2+c_3}$, where $f(\cdot)$ and α_0 are defined in the proof of Proposition 3.2, it follows that

$$\begin{aligned}\sup_{t_1,t_2\in\mathbb{R}} |\mathbb{P}(M_B \le t_1, m_B \le t_2) - \mathbb{P}(\hat{M}_B \le & t_1, \hat{m}_B \le t_2)| \\ & \le Cn^{-c_3} + C(\log(pn))^{4/\lambda_1+1} n^{-f(\alpha_0)/2+c_3}.\end{aligned}$$

In summary, we have

$$\sup_{t_1,t_2\in\mathbb{R}}\left|\mathbb{P}(\hat{M}_B\le t_1,\hat{m}_B\le t_2)-\mathbb{P}(\hat{M}_Z\le t_1,\hat{m}_Z\le t_2)\right|\le C|\hat{\Xi}-\Xi|_\infty^{1/3}(\log p)^{2/3}$$
$$+Cn^{-c_2}+C(\log(pn))^{4/\lambda_1+1}n^{-f(\alpha_0)/2+c_2}+Cn^{-c_3}+C(\log(pn))^{4/\lambda_1+1}n^{-f(\alpha_0)/2+c_3}$$

Since we want the right side decays to zero and p diverges as fast as possible, take $\log p = o(n^{\omega})$ where $\omega_5=\min\left\{\frac{3\lambda_1 f(\alpha_0)}{24+4\lambda_1},\frac{\lambda_1(f(\alpha_0)-c_2)}{8+2\lambda_1},\frac{\lambda_1(f(\alpha_0)-c_3)}{8+2\lambda_1}\right\}$. We want to find the minimum rate of each maximized possible exponentials of n. So we take $c_2\to 0$ and $c_3\to 0$. In conclusion, if $\log p = o\left(n^{\lambda_1 f(\alpha_0)/(8+2\lambda_1)}\right)$, then (3.18) follows. This completes the proof for Proposition 3.3. □

Furthermore, according to the continuous mapping theorem for sign measure, Proposition 3.3 implies the following Theorem:

Theorem 3.2. *Assume the conditions of Proposition 3.2 and Lemma 3.3 holds, then*

$$\sup_{y\in\mathbb{R}}\left|\mathbb{P}(\hat{M}^2+\hat{m}^2\le y)-\mathbb{P}(\hat{M}_B^2+\hat{m}_B^2\le y\mid\mathcal{D})\right|\xrightarrow{p}0. \tag{3.21}$$

This result implies that it is valid to use the quantiles of $\hat{M}_B^2+\hat{m}_B^2$ from dependent multiplier bootstrap samples to approximate that of $\hat{M}^2+\hat{m}^2$, and furthermore, our proposed test is consistent.

3.3.3 Asymptotic Independence

In this section, we investigate the conditions under which the maximum and minimum of dependent Gaussian random variables are asymptotically independent. We first look at the asymptotic independence when those Gaussian variables are independent, and then extend it to dependent case.

Lemma 3.4 (Asymptotic independence for i.i.d)**.** *If $X_1,X_2,\ldots,X_n$ is a sequence of independently and identically distributed Gaussian random variables and write M_n and m_n for the maximum and minimum of the sequence, then M_n and m_n are asymptotically independent.*

Specifically, write

$$\mathbb{P}(a_n(M_n - b_n) \le x) \xrightarrow{w} G(x),$$

$$\mathbb{P}(\alpha_n(m_n - \beta_n) \le x) \xrightarrow{w} H(x),$$

then

$$\mathbb{P}(a_n(M_n - b_n) \le x, \alpha_n(m_n - \beta_n) \le y) \xrightarrow{w} G(x)H(y),$$

where $\{a_n > 0\}$*,* $\{b_n\}$ *and* $\{\alpha_n > 0\}$ *and* $\{\beta_n\}$ *are some normalizing constants and* G *and* H *are the limiting distribution functions.*

Proof. According to Theorem 1.5.3 of Leadbetter et al. (1983), for any i.i.d normal random variables, the asymptotic distribution of M_n is extreme value distribution of type I:

$$G(x) = \exp(-e^{-x}),$$

with

$$a_n = (2\log n)^{\frac{1}{2}} \quad \text{and} \quad b_n = (2\log n)^{\frac{1}{2}} - \frac{1}{2}(2\log n)^{-\frac{1}{2}}(\log\log n + \log 4\pi)$$

The asymptotic distribution of m_n can be derived accordingly. Since $m_n = \min(x_1, x_2, \ldots, x_n) = -\max(-x_1, -x_2, \ldots, -x_n)$, then

$$\begin{aligned}
\mathbb{P}(\alpha_n(-m_n + \beta_n) \le x) &= 1 - \mathbb{P}(\alpha_n(-m_n + \beta_n) \ge x) \\
&= 1 - \mathbb{P}(\alpha_n(m_n - \beta_n) \le -x) \\
&\to 1 - H(-x) \\
&= G(x).
\end{aligned}$$

Thus, let $\alpha_n = a_n$ and $\beta_n = -b_n$, we have

$$\mathbb{P}(\alpha_n(m_n - \beta_n) \le x) \to H(x),$$

where

$$H(x) = 1 - G(-x) = 1 - \exp(-e^{x}).$$

Then the result follows from Theorem 1.8.3 of Leadbetter et al. (1983) after identifying the normalizing constants and d.f.'s G and H. $\square$

Then it comes to our main result of the asymptotic independence between M_n and m_n from dependent normal random variables.

Theorem 3.3. *Let $\{\boldsymbol{X_n}, n = 1, 2, \dots\}$ be a sequence of centered Gaussian random vectors where $\boldsymbol{X_n} = (X_{n1}, X_{n2}, \dots, X_{n,p_n})$. The dimension of each $\boldsymbol{X_n}$ depends on n and goes to ∞ as $n \to \infty$. Let $\Sigma_{p_n} = (\sigma_{n,ij})_{1 \le i,j \le p_n}$ be the covariance matrix for each $\boldsymbol{X_n}$ and assume the diagonal elements are all 1's. Write the maximum and minimum of $\{X_i, i = 1, \dots, p_n\}$ as $M_{n,X}$, $m_{n,X}$. Then $M_{n,X}$ and $m_{n,X}$ are asymptotically independent if*

$$\lim_{n\to\infty} \sum_{1<i<j\le p_n} |\sigma_{n,ij}|(1-\sigma_{n,ij}^2)^{-\frac{1}{2}} p_n^{-\frac{2}{1+|\sigma_{n,ij}|}} (\log p_n) = 0. \tag{3.22}$$

Proof. In order to prove the asymptotic independence between M_X and m_X, we consider a sequence of standard Gaussian random vector $Y_{p_n \times 1}$ with mean **0** and covariance matrix I_{p_n}. Let $M_{n,Y}$, $m_{n,Y}$ be the corresponding maximum and minimum of $\{Y_i, i = 1, \dots, p_n\}$. According to Lemma 3.4, $M_{n,Y}$ and $m_{n,Y}$ are asymptotically independent. Thus, the difference between the joint distribution of $M_{n,X}$, $m_{n,Y}$ and that of $M_{n,Y}$, $m_{n,Y}$ is our main interest. We can show they are close enough when $\sigma_{n,ij}$ satisfies some asymptotic conditions. First, it is helpful to look at the multi-dimensional Gaussian density functions. Let $\phi_n(x_1, \cdots, x_n; \sigma_{ij}, 1 \le i, j \le n)$ be the n-dimensional Gaussian density function with mean vector **0** and covariance matrix Σ, where $\Sigma = (\sigma_{ij})$ is an $n \times n$ symmetric positive definite matrix with 1's along the diagonal. So ϕ_n is a function of the $x's$ and $n(n-1)/2$ parameters σ_{ij}. Q_n is defined to be:

$$Q_n(c_1, c_2; \{\sigma_{ij}\}) = \int_{c_2}^{c_1} \cdots \int_{c_2}^{c_1} \phi_n(x_1, \cdots, x_n; \{\sigma_{ij}\}) \prod_{j=1}^{n} dx_j.$$

Applying the same technique in Section 2.1 from Slepian (1962) to take the partial derivative of Q_n with respect to σ_{hl}:

$$\frac{\partial Q_n}{\partial \sigma_{hl}} = \int_{c_2}^{c_1} \cdots \int_{c_2}^{c_1} \prod_{j=1}^{n} dx_j \cdot \frac{\partial^2}{\partial x_h \partial x_l} \phi_n(x_1, \cdots, x_n; \{\sigma_{ij}\}).$$

Perform integration over x_h and x_l:

$$\begin{aligned}\frac{\partial Q_n}{\partial \sigma_{hl}} &= \int_{c_2}^{c_1} \cdots \int_{c_2}^{c_1} \prod_{j \neq h, j \neq l} dx_j \cdot \int_{c_2}^{c_1} \int_{c_2}^{c_1} \frac{\partial^2}{\partial x_h \partial x_l} \phi_n(x_1, \cdots, x_n; \{\sigma_{ij}\}) dx_h dx_l \\ &= \int_{c_2}^{c_1} \cdots \int_{c_2}^{c_1} \prod_{j \neq h, j \neq l} dx_j \cdot \left\{ \phi_n^{(hl)}(c_1, c_1) - \phi_n^{(hl)}(c_1, c_2) - \phi_n^{(hl)}(c_2, c_1) + \phi_n^{(hl)}(c_1, c_2) \right\},\end{aligned}$$

where

$$\phi_n^{(hl)}(c_1, c_2) = \phi_n(x_1, \cdots, x_{h-1}, c_1, x_{h+1}, \cdots, x_{l-1}, c_2, x_{l+1}, \cdots, x_n; \{\sigma_{ij}\}).$$

Since for each $\phi_n^{(hl)}(\cdot, \cdot)$ is a density function which is always positive, we can replace the integration limits to ∞ to obtain the upper bound of this partial derivative:

$$\begin{aligned}\left| \frac{\partial Q_n}{\partial \sigma_{hl}} \right| &\leq \int_{c_2}^{c_1} \cdots \int_{c_2}^{c_1} \phi_n^{(hl)}(c_1, c_1) + \phi_n^{(hl)}(c_1, c_2) + \phi_n^{(hl)}(c_2, c_1) + \phi_n^{(hl)}(c_1, c_2) \prod_{j \neq h, j \neq l} dx_j \\ &\leq \phi_2(c_1, c_1; \sigma_{hl}) + \phi_2(c_1, c_2; \sigma_{hl}) + \phi_2(c_2, c_1; \sigma_{hl}) + \phi_2(c_2, c_2; \sigma_{hl}),\end{aligned}$$

where

$$\phi_2(x, y; \sigma) = (2\pi)^{-1}(1 - \sigma^2)^{1/2} \exp\left\{ -\frac{x^2 - 2\sigma xy + y^2}{2(1 - \sigma^2)} \right\}.$$

Now we can derive the upper bound of the difference of those joint distributions. Based on the definition of Q_n, we can write

$$\mathbb{P}(M_{n,X} \leq c_1, m_{n,X} \geq c_2) = Q_{p_n}(c_1, c_2; \{\sigma_{n,ij}\}),$$

$$\mathbb{P}(M_{n,Y} \leq c_1, m_{n,Y} \geq c_2) = Q_{p_n}(c_1, c_2; \{0\}),$$

where $c_1 \geq 0$ and $c_2 \leq 0$. By the law of the mean, there exists some $\sigma'_{n,ij}$ which is between 0 and $\sigma_{n,ij}$, where $1 < i < j \leq p_n$, such that

$$
\begin{aligned}
&|\mathbb{P}(M_{n,X} \leq c_1, m_{n,X} \geq c_2) - \mathbb{P}(M_{n,Y} \leq c_1, m_{n,Y} \geq c_2)| \\
\leq &\sum_{1<i<j\leq p_n} \left| \sigma_{n,ij} \Big(\frac{\partial Q_{p_n}}{\partial \sigma_{n,ij}} \Big)(c_1, c_2; \{\sigma'_{n,ij}\}) \right| \\
\leq &\sum_{1<i<j\leq p_n} |\sigma_{n,ij}| \left\{ \phi_2(c_1, c_1; \sigma'_{n,ij}) + \phi_2(c_1, c_2; \sigma'_{n,ij}) + \phi_2(c_2, c_1; \sigma'_{n,ij}) + \phi_2(c_2, c_2; \sigma'_{n,ij}) \right\}.
\end{aligned}
$$

The sum has $p_n(p_n - 1)/2$ terms. Checking the monotonicity of each ϕ_2 regarding $|\sigma|$, it is obvious that $\phi_2(c_1, c_1; \sigma)$ is a monotonically increasing function of $|\sigma|$, so as $\phi_2(c_2, c_2; \sigma)$. As for $\phi_2(c_1, c_2; \sigma)$ and $\phi_2(c_2, c_1; \sigma)$, if $|c_2| \geq |c_1|$, we have the following inequality:

$$c_1^2 - 2\sigma c_1 c_2 + c_2^2 \geq c_1^2 - 2|\sigma| c_1^2 + c_1^2$$

Recall $c_1 \geq 0$ and $c_2 \leq 0$. This follows from

$$c_2^2 - c_1^2 - 2\sigma c_1 c_2 + 2|\sigma| c_1^2 \geq c_2^2 - c_1^2 + 2|\sigma| c_1 c_2 + 2|\sigma| c_1^2 = (c_2 + c_1)(c_2 - c_1 + 2|\sigma| c_1) \geq 0.$$

Thus, we have

$$
\begin{aligned}
\phi_2(c_1, c_2; \sigma) &= (2\pi)^{-1}(1-\sigma^2)^{1/2} \exp\left\{ -\frac{c_1^2 - 2\sigma c_1 c_2 + c_2^2}{2(1-\sigma^2)} \right\} \\
&\leq (2\pi)^{-1}(1-\sigma^2)^{1/2} \exp\left\{ -\frac{c_1^2 - 2|\sigma| c_1^2 + c_1^2}{2(1-\sigma^2)} \right\} \\
&= \phi_2(c_1, c_1; |\sigma|).
\end{aligned}
$$

Since $0 \leq |\sigma'_{n,ij}| \leq |\sigma_{n,ij}|$, we can obtain the upper bound as a function of $\sigma_{n,ij}$ as

$$
\begin{aligned}
&|\mathbb{P}(M_{n,X} \leq c_1, m_{n,X} \geq c_2) - \mathbb{P}(M_{n,Y} \leq c_1, m_{n,Y} \geq c_2)| \\
\leq &\sum_{1<i<j\leq p_n} |\sigma_{n,ij}| \left\{ 3\phi_2(c_1, c_1; |\sigma'_{n,ij}|) + \phi_2(c_2, c_2; |\sigma'_{n,ij}|) \right\} \\
\leq &\sum_{1<i<j\leq p_n} |\sigma_{n,ij}| \left\{ 3\phi_2(c_1, c_1; |\sigma_{n,ij}|) + \phi_2(c_2, c_2; |\sigma_{n,ij}|) \right\}.
\end{aligned}
$$

Similarly, if $|c_2| < |c_1|$, then

$$c_1^2 - 2\sigma c_1 c_2 + c_2^2 \geq c_2^2 - 2|\sigma| c_2^2 + c_2^2.$$

The similar upper bound can be obtained accordingly:

$$|\mathbb{P}(M_{n,X} \leq c_1, m_{n,X} \geq c_2) - \mathbb{P}(M_{n,Y} \leq c_1, m_{n,Y} \geq c_2)| \leq \sum_{1<i<j\leq p_n} |\sigma_{n,ij}|\{3\phi_2(c_2, c_2; |\sigma_{n,ij}|) + \phi_2(c_1, c_1; |\sigma_{n,ij}|)\}.$$

Let $c_1 = a_{p_n}^{-1}x + b_{p_n}$ and $c_2 = \alpha_{p_n}^{-1}x + \beta_{p_n}$, where a_{p_n}, b_{p_n} and α_{p_n}, β_{p_n} are the normalizing constants from Lemma 3.4:

$$a_{p_n} = (2\log p_n)^{\frac{1}{2}} \quad \text{and} \quad b_{p_n} = (2\log p_n)^{\frac{1}{2}} - \frac{1}{2}(2\log p_n)^{-\frac{1}{2}}(\log\log p_n + \log 4\pi),$$

$$\alpha_{p_n} = a_{p_n} \quad \text{and} \quad \beta_{p_n} = -b_{p_n}.$$

If $n \to \infty$, then $p_n \to \infty$. As a result, we have

$$c_1^2 = (a_{p_n}^{-1}x + b_{p_n})^2 = 2\log p_n - \log\log p_n + O(1),$$

$$c_2^2 = (\alpha_{p_n}^{-1}x + \beta_{p_n})^2 = 2\log p_n - \log\log p_n + O(1).$$

For either case, we can write the upper bound explicitly by

$$|\mathbb{P}(M_{n,X} \leq c_1, m_{n,X} \geq c_2) - \mathbb{P}(M_{n,Y} \leq c_1, m_{n,Y} \geq c_2)| \leq C \sum_{1<i<j\leq p_n} |\sigma_{n,ij}|(1-\sigma_{ij}^2)^{-\frac{1}{2}} p_n^{-\frac{2}{1+|\sigma_{n,ij}|}} (\log p_n)^{\frac{1}{1+|\sigma_{n,ij}|}},$$

where $C > 0$ is a constant. Then Theorem 3.3 follows immediately. □

Now we shall discuss the validity of (3.22). Here we give two examples of its sufficient condition.

Lemma 3.5. *If there exists* $0 < \alpha < 1$, $\delta > 0$ *such that* $\delta > 2\alpha/(1+\alpha)$ *and* *(i)* $\sup_n \sup_{1<i<j\le p_n} |\sigma_{n,ij}| = \alpha$,

(ii) $\sum_{1<i<j\le p_n} \sigma_{n,ij}^2 = O(p_n^{2-\delta})$,

then (3.22) *holds.*

Remark 3.4. *For convenience we write condition (i), but it is possible for some σ_{ij} to be very close to* 1*, where the asymptotic independence is still true as long as* (3.22) *holds.*

Proof. By condition (i), there exist some $\alpha > 0$ such that

$$(1-\sigma_{n,ij}^2)^{-1/2} \le (1-\alpha^2)^{-1/2},$$

for all $\sigma_{n,ij}$. Given $p_n \to \infty$ as $n \to \infty$, then (3.22) would be implied by

$$\lim_{p\to\infty} \sum_{1<i<j\le p} |\sigma_{ij}| p^{-\frac{2}{1+|\sigma_{ij}|}} (\log p) = 0.$$

Based on condition (ii), we have for any $k > 1$,

$$|\{\sigma_{ij} : |\sigma_{ij}| > 1/k, 1 \le i,j \le p_n\}| = O(p^{2-\delta}),$$

where $|A|$ is the cardinality of set A. Therefore, we split the sum into two parts. First sum is over $\{\sigma_{ij} : |\sigma_{ij}| > 1/k\}$ where the amount of σ_{ij} grows linearly with p, while the second sum is over $\{\sigma_{ij} : |\sigma_{ij}| \le 1/k\}$. In order to bound the sum, we would consider a large number k. Write

$$\mathcal{I}^{(1)} = \{(i,j) : |\sigma_{ij}| > 1/k, 1 < i < j \le p\} \quad \text{and} \quad \mathcal{I}^{(2)} = \{(i,j) : |\sigma_{ij}| \le 1/k, 1 < i < j \le p\}.$$

Consider the first sum. Since the number of σ_{ij} in set $\mathcal{I}^{(1)}$ is $O(p^{2-\delta})$, it follows that

$$\sum_{\mathcal{I}^{(1)}} |\sigma_{ij}| p^{-\frac{2}{1+|\sigma_{ij}|}} (\log p) \le C(\log p) \cdot \alpha p^{2-\delta-\frac{2}{1+\alpha}} \to 0,$$

which resulting from the assumption $\delta > 2\alpha/(1+\alpha)$. Applying Cauchy-Schwarz Inequality to the square of second sum. By condition ii, we have

$$\begin{aligned}\Big(\sum_{\mathcal{I}^{(2)}}(\log p)|\sigma_{ij}|p^{-\frac{2}{1+|\sigma_{ij}|}}\Big)^2 &\le (\log p)^2\cdot\sum_{\mathrm{II}_p}\sigma_{ij}^2\cdot\sum_{\mathrm{II}_p}p^{-\frac{4}{1+|\sigma_{ij}|}}\\ &\le (\log p)^2\cdot p^{2-\frac{4}{1+1/k}}\cdot\sum_{\mathrm{II}_p}\sigma_{ij}^2\\ &\le C(\log p)^2\cdot p^{4-\delta-\frac{4}{1+1/k}},\end{aligned}$$

where the exponent of p is also negative when $1/k\to 0$, so that the square of second sum also goes to 0. $\square$

Remark 3.5. *The other example would be the special scenario when random variables in* $\boldsymbol{X_n}$ *are stationary, i.e., the covariance* σ_{ij} *depends only on the difference between* i *and* j*. Let* $\sigma_{j-i}=\sigma_{ij}$*, then we can simplify* (3.22) *as*

$$\lim_{p\to\infty}\sum_{j=1}^{p-1}(p-j)|\sigma_j|(1-\sigma_j^2)^{-\frac{1}{2}}p^{-\frac{2}{1+|\sigma_j|}}(\log p)=0. \tag{3.23}$$

Berman (1964) studied the limiting extreme value distribution function for the maximum for Gaussian sequence. We can show that (3.23) *holds if either*

$$\lim_{p\to\infty}\sigma_p\log p=0$$

or

$$\sum_{i=1}^{p-1}\sigma_i^2=O(p^{1-\delta})$$

holds for some $\delta>0$.

3.4 Simulations

In this section, we compare our new tests T_n and S_n with the tests based on maximized cross correlation Y_n and $Y_n^\star$ from Chang et al. (2017), and three portmanteau tests: $Q_1 = n\sum_{k=1}^K \mathrm{tr}(\hat\Gamma(k)'\hat\Gamma(k))$, $Q_2=n^2\sum_{k=1}^K \mathrm{tr}(\hat\Gamma(k)'\hat\Gamma(k))/(n-k)$, $Q_3=n\sum_{k=1}^K \mathrm{tr}(\hat\Gamma(k)'\hat\Gamma(k))+p^2K(K+1)$ where $\hat\Gamma(k)=\mathrm{diag}\{\hat\Sigma(0)\}^{-1/2}\hat\Sigma(k)\mathrm{diag}\{\hat\Sigma(0)\}^{-1/2}$, and Lagrange multiplier test (LM) by Hosking (1981) as well as the likelihood ratio test (TB) by Tiao and Box (1981). Set the nominal significance level at $\alpha = 0.05$. The dimension of the time series is $p = 3, 15, 50, 150$ and sample size is $n = 300$. We select the lag to be $K = 1, 2, 3, 4, 5$. In addition, we use the quadratic kernel derived by Andrews (1991):

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

and the bandwidth is selected to be $w_n = 1.3221\{\hat{a}(2)\tilde{n}\}^{1/5}$ where

$$\hat{a}(2) = \frac{\sum_{l=1}^{p^2K} 4\hat{\rho}_l^2\hat{\sigma}_l^4(1-\hat{\rho}_l)^{-8}}{\sum_{l=1}^{p^2K} \hat{\sigma}_l^4(1-\hat{\rho}_l)^{-4}},$$

with $\hat{\rho}_l$ and $\hat{\sigma}_l^2$ being the estimated autoregressive coefficient and innovation variance from fitting an AR(1) model to $\{\xi_{l,t}\}$, $t = 1, \ldots, \tilde{n}$. This is the data-driven bandwidth suggested by Andrews (1991). For each setting, we replicate the test for 500 times. The critical values are obtained from the dependent Gaussian multiplier bootstrap procedure described in Section 3.3.2.

3.4.1 Empirical Size

First, we generate white noise series to examine the empirical rejection rates of the tests. Consider a white noise model $x_t = Az_t$, where z_t is a $p \times 1$ white noise and the loading matrices A can be as following:

Model 1. Let $S = (s_{kl})_{1\le k,l\le p}$ where $s_{kl} = 0.995^{|k-l|}$, then $A = S^{1/2}$.

Model 2. Let $r = \lceil p/2.5 \rceil$, $S = (s_{kl})_{1\le k,l\le p}$ where $s_{kk} = 1$, $s_{kl} = 0.8$ for $r(q-1)+1 \le k \ne l \le rq$ for $q = 1, \cdots, \lfloor p/r \rfloor$, and $s_{kl} = 0$ otherwise. Then let $A = S^{1/2}$.

Model 3. Let $A = (a_{kl})_{1\le k,l\le p}$, where $a_{kl} \sim U(-1, 1)$ independently.

For each model of loading matrix, there are also two different types of white noise z_t considered:

1. z_t are independent from $N(0, I_p)$.

2. z_t consists of p independent ARCH process. Each process is of the form $u_t = \sigma_t e_t$, where e_t are independent and $N(0, 1)$, and $\sigma_t^2 = \gamma_0 + \gamma_1 u_{t-1}^2$ with $\gamma_0 \sim U(0.25, 0.5)$

and $\gamma_1 \sim U(0, 0.5)$.

The above white noise models were proposed by Chang et al. (2017). Obviously all the p series are independent with each other in these models. So we also consider a different setting as the following:

Model 4. $x_t = \sigma_t e_t, t = 1, \ldots, p$, where e_t are independent and $N(0,1)$, and $\sigma_t^2 = \gamma_0 + \gamma_1 x_{t-1}^2$ with $\gamma_0 \sim U(0.25, 0.5)$ and $\gamma_1 \sim U(0, 0.5)$. Thus, we have n independent ARCH(1) process.

Remark 3.6. *The white noise test T_n^* described in Chang et al. (2017) is to apply the time series principal component analysis to the data first and then test on the transformed data. However, the T_n^* we compared with is from the R package* `HDtest`*, where the data is transformed by different methods. Specifically in this simulation, we choose the first option in the function wntest to transform the data using package* `fastclime` *with $\lambda = 0.05$ which estimates the contemporaneous correlations.*

Remark 3.7. *The likelihood ratio test (TB) does not involve lag parameter k, so there is only one value reported for each setting in the empirical size table. For Lagrange multiplier test (LM), as the test statistics is calculated from multivariate regression, there is no value reported when $pk \geq n$. To be specific, when $p = 150$, the test is applicable only at lag $k = 1$.*

Table 3.1-3.4 report the empirical size of our new test T_n and S_n along with other white noise tests. In general, T_n and S_n are able to better control the empirical size regardless of the increase of p, and T_n usually has a larger size than S_n. The test using maximum cross correlation (Y_n) performs well when p is small, however, it decays very fast when p increases. The test using maximum cross covariance after a transformation (Y_n^*) is not very stable to attain the nominal significance level, especially in Model 2 and Model 3. It becomes extremely large for some setting, such as Model 3 at $p = 150$. For Model 4, Y_n^* is expected to fail since the white noise series is already i.i.d, so performing a transformation on data is unnecessary. As for those three portmanteau tests, Q_2 and Q_3 perform very similarly, while Q_1 performs worse than Q_2 and Q_3, as it is almost 0 when $p = 150$. The Lagrange multiplier test (LM) fails badly to capture the nominal significance level as p increases, while the likelihood ratio test

(TB) always reject H_0 when $p = 150$ for all models. In summary, our proposed new tests are more robust regardless of the choice of white noise model, and they can calibrate the test more accurately compared to other tests.

Table 3.1: The empirical sizes (%) of all tests for testing white noise series generated from model 1.

(a) $z_t \sim N(0, I_p)$

p	K	T_n	S_n	Y_n	Y_n^*	Q_1	Q_2	Q_3	LM	TB
3	1	5.6	1.2	6.2	8.0	6.2	6.2	6.2	6.2	6.4
	2	6.2	4.8	5.6	9.0	5.0	5.2	5.2	5.2	
	3	4.8	6.2	5.2	9.2	3.6	4.0	4.0	4.0	
	4	5.2	6.0	5.0	8.0	4.8	5.0	4.8	5.0	
	5	5.2	5.6	5.0	8.6	4.0	5.0	4.2	4.8	
15	1	5.4	2.6	5.4	2.2	4.2	4.2	4.2	4.2	5.0
	2	6.4	5.0	5.2	2.6	4.0	5.2	5.2	4.2	
	3	5.0	5.6	4.4	2.0	2.2	3.0	3.0	2.6	
	4	5.4	4.6	4.6	1.8	2.8	3.6	3.4	3.2	
	5	5.8	5.2	5.2	2.2	2.6	3.6	3.6	2.6	
50	1	4.6	3.4	4.6	2.2	2.0	2.6	2.6	1.8	6.0
	2	4.8	4.4	3.0	2.0	0.6	2.0	2.0	0.4	
	3	4.4	3.2	4.2	1.0	1.0	3.2	3.0	0.2	
	4	3.8	4.8	3.6	0.4	0.6	2.8	2.4	0.0	
	5	4.0	4.2	3.6	0.4	0.0	3.4	3.2	0.0	
150	1	4.8	3.8	3.2	1.4	0.0	0.2	0.2	0.0	100.0
	2	7.0	6.6	4.8	1.6	0.0	1.4	1.4		
	3	7.6	6.6	4.8	0.8	0.0	1.8	1.6		
	4	7.0	6.2	3.8	1.0	0.0	3.4	3.2		
	5	6.2	4.2	2.4	1.0	0.0	4.4	3.8		

(b) $z_t \sim ARCH(1)$

p	K	T_n	S_n	Y_n	Y_n^*	Q_1	Q_2	Q_3	LM	TB
3	1	4.6	0.4	5.2	6.4	7.4	7.8	7.6	7.6	8.0
	2	5.2	4.8	5.0	7.8	7.4	7.4	7.4	7.2	
	3	5.4	4.8	6.0	7.8	5.4	5.8	5.6	7.0	
	4	5.4	5.4	5.4	7.2	4.2	5.2	4.8	5.8	
	5	5.2	5.2	5.0	7.6	4.0	4.8	4.4	4.2	
15	1	5.6	2.2	5.0	1.8	11.0	11.2	11.2	11.4	13.0
	2	4.4	4.0	4.8	1.2	10.0	11.0	11.0	9.8	
	3	6.8	5.2	5.8	0.8	8.4	9.4	9.4	7.4	
	4	6.2	5.0	5.6	1.2	4.8	7.2	6.6	5.2	
	5	6.8	5.2	5.0	1.2	3.2	7.0	6.8	4.0	
50	1	5.4	4.2	4.8	1.0	7.0	10.0	10.0	5.4	16.4
	2	4.4	3.4	3.6	0.6	5.2	8.0	8.0	2.4	
	3	5.4	5.0	4.4	0.0	3.8	9.8	9.8	0.0	
	4	4.6	5.0	3.6	0.2	2.2	8.2	8.0	0.0	
	5	5.4	5.2	3.6	0.2	1.4	8.6	8.6	0.0	
150	1	6.8	7.0	5.6	1.8	0.4	1.8	1.8	0.6	100.0
	2	8.6	8.6	6.2	0.6	0.2	2.6	2.4		
	3	9.6	8.4	5.4	0.4	0.0	3.8	3.6		
	4	9.4	8.6	5.4	0.6	0.0	6.0	6.0		
	5	8.6	8.4	4.4	1.0	0.0	8.0	7.6		

Table 3.2: The empirical sizes (%) of all tests for testing white noise series generated from model 2.

(a) $z_t \sim N(0, I_p)$

p	K	T_n	S_n	Y_n	Y_n^*	Q_1	Q_2	Q_3	LM	TB
3	1	5.2	5.0	4.2	6.6	4.8	4.8	4.8	4.8	5.0
	2	6.0	4.8	4.8	8.0	4.6	4.6	4.6	5.0	
	3	6.6	5.2	3.4	7.4	4.2	4.6	4.4	4.0	
	4	5.2	5.0	3.6	7.0	4.4	5.6	5.4	4.4	
	5	5.6	5.4	4.2	6.4	4.4	4.8	4.8	5.4	
15	1	4.6	5.4	4.8	10.2	4.6	4.8	4.6	4.6	5.6
	2	5.6	5.2	2.8	13.0	3.8	4.8	4.8	4.6	
	3	5.2	5.4	2.4	13.4	3.8	5.0	4.8	4.8	
	4	5.4	5.4	2.6	14.4	3.0	4.0	4.0	4.0	
	5	5.8	5.8	3.0	14.8	2.2	3.8	3.8	2.8	
50	1	6.2	7.0	2.8	10.8	3.0	3.4	3.4	3.6	8.0
	2	5.8	5.6	3.0	9.2	1.8	2.6	2.6	1.2	
	3	6.6	5.8	3.0	10.8	1.2	5.6	5.6	0.4	
	4	7.8	7.6	2.6	11.0	0.8	4.0	3.6	0.0	
	5	8.0	7.0	2.2	10.8	0.2	4.0	3.4	0.0	
150	1	8.0	7.0	1.2	5.0	0.2	0.6	0.6	0.0	100.0
	2	7.6	6.0	1.6	6.8	0.0	0.6	0.6		
	3	6.8	6.2	1.4	6.4	0.0	1.2	1.0		
	4	7.6	6.8	1.8	6.2	0.0	2.8	2.8		
	5	9.2	7.2	1.0	6.8	0.0	6.0	6.0		

(b) $z_t \sim ARCH(1)$

p	K	T_n	S_n	Y_n	Y_n^*	Q_1	Q_2	Q_3	LM	TB
3	1	4.4	4.4	4.6	9.4	8.8	8.8	8.8	9.0	9.0
	2	4.6	4.2	4.6	10.4	9.0	9.2	9.0	9.6	
	3	6.4	4.8	5.4	11.0	7.2	8.0	8.0	7.6	
	4	6.0	5.2	5.0	10.6	7.6	7.8	7.8	7.4	
	5	5.4	4.4	4.2	9.6	6.4	7.6	7.4	7.6	
15	1	6.6	4.8	2.6	3.8	10.4	10.8	10.8	10.6	12.2
	2	5.6	4.6	3.2	5.6	9.2	10.4	10.2	9.6	
	3	6.0	5.2	3.0	6.0	7.8	9.8	9.4	6.4	
	4	6.0	4.4	3.0	8.0	7.0	10.4	10.2	5.8	
	5	6.0	4.2	2.8	7.4	6.2	9.8	9.4	5.2	
50	1	6.2	5.8	1.0	3.8	6.6	7.4	7.4	5.4	11.6
	2	7.0	6.2	0.6	4.2	4.6	8.2	8.2	2.2	
	3	7.6	7.0	0.4	3.2	2.8	10.2	10.0	0.2	
	4	8.0	6.8	0.4	4.0	1.4	9.4	9.4	0.2	
	5	8.0	7.2	0.6	3.4	1.2	9.6	9.2	0.0	
150	1	8.8	7.6	0.0	2.2	0.0	1.2	1.2	0.0	100.0
	2	9.8	9.0	0.2	1.6	0.0	1.6	1.6		
	3	10.6	9.6	0.2	1.2	0.0	2.6	2.6		
	4	10.8	9.8	0.4	1.4	0.0	4.4	4.0		
	5	11.4	10.0	0.4	0.8	0.0	6.8	6.4		

Table 3.3: The empirical sizes (%) of all tests for testing white noise series generated from model 3.

(a) $z_t \sim N(0, I_p)$

p	K	T_n	S_n	Y_n	Y_n^*	Q_1	Q_2	Q_3	LM	TB
3	1	4.6	4.2	4.0	9.6	4.2	4.2	4.2	4.2	4.2
	2	5.4	4.6	4.4	9.6	4.8	5.0	4.8	4.4	
	3	4.8	5.4	5.4	8.4	4.4	4.6	4.4	5.2	
	4	5.6	5.8	4.6	10.4	5.6	5.8	5.6	5.0	
	5	5.8	5.0	4.6	9.6	5.2	6.4	6.4	5.4	
15	1	5.0	4.8	4.0	4.6	4.2	5.2	5.0	5.0	6.6
	2	6.6	4.8	3.2	5.0	4.0	6.0	5.8	4.2	
	3	8.2	7.0	4.6	3.8	3.8	5.4	5.2	4.0	
	4	8.6	6.6	3.8	4.2	3.6	5.8	5.8	4.0	
	5	8.0	6.6	4.2	4.8	2.6	5.8	5.6	3.6	
50	1	6.0	6.2	2.2	7.6	2.6	3.8	3.6	2.0	9.0
	2	6.4	5.4	1.6	9.0	2.8	5.0	4.8	1.8	
	3	6.8	4.6	1.4	11.6	1.6	4.2	4.2	0.0	
	4	5.2	3.6	1.6	11.0	1.4	4.6	4.4	0.2	
	5	5.0	4.0	1.8	11.6	0.8	4.6	4.6	0.0	
150	1	7.2	6.8	2.8	20.0	0.0	0.4	0.4	0.0	100.0
	2	7.0	6.2	2.0	25.8	0.0	1.0	1.0		
	3	9.0	7.2	1.6	26.4	0.0	2.0	2.0		
	4	8.4	7.6	1.6	30.8	0.0	2.8	2.4		
	5	9.4	7.6	1.4	32.4	0.0	3.6	3.6		

(b) $z_t \sim ARCH(1)$

p	K	T_n	S_n	Y_n	Y_n^*	Q_1	Q_2	Q_3	LM	TB
3	1	5.8	6.2	5.6	6.6	8.8	8.8	8.8	8.8	8.8
	2	4.8	4.6	4.0	8.0	7.2	7.4	7.4	7.2	
	3	6.0	5.2	4.4	8.6	7.8	8.4	8.4	7.8	
	4	5.8	4.8	4.6	9.4	9.0	9.4	9.4	7.8	
	5	5.2	5.0	3.8	9.6	6.2	6.8	6.8	6.2	
15	1	6.8	5.8	3.6	3.2	11.8	13.0	12.8	13.0	14.6
	2	6.4	5.2	2.2	3.4	8.0	8.6	8.4	9.0	
	3	6.6	5.2	3.2	2.8	7.0	8.4	8.2	7.0	
	4	7.0	5.0	3.4	2.6	6.0	8.2	7.8	5.8	
	5	6.0	5.0	2.6	1.8	6.0	8.8	8.6	5.4	
50	1	6.4	5.8	2.0	7.8	6.0	8.4	8.2	4.0	16.2
	2	7.2	5.4	1.8	8.0	6.0	9.4	9.4	3.0	
	3	9.2	7.6	2.2	7.8	3.4	10.6	10.0	1.4	
	4	8.0	7.2	1.8	8.2	2.4	8.8	8.4	0.0	
	5	8.4	5.4	1.6	8.0	1.4	9.8	9.4	0.0	
150	1	7.4	6.6	1.8	23.0	0.2	1.0	1.0	0.2	100.0
	2	9.4	7.2	1.2	28.4	0.0	2.6	2.4		
	3	9.6	7.4	1.4	32.8	0.0	3.2	3.2		
	4	9.2	6.0	1.2	33.6	0.0	4.8	4.6		
	5	9.4	6.0	0.4	35.6	0.0	6.2	5.8		

Table 3.4: The empirical sizes (%) of all tests for testing white noise series generated from model 4

p	K	T_n	S_n	Y_n	Y_n^*	Q_1	Q_2	Q_3	LM	TB
3	1	3.8	3.4	4.4	7.4	4.6	4.6	4.6	4.6	5.0
	2	5.2	4.6	3.6	8.0	5.8	6.2	6.2	6.4	
	3	5.4	5.0	4.0	8.4	5.4	5.8	5.6	5.6	
	4	5.4	4.6	3.8	7.4	5.2	5.4	5.4	5.0	
	5	4.8	3.6	2.8	7.6	4.0	4.8	4.6	5.0	
15	1	5.2	4.6	2.8	7.8	3.0	3.2	3.2	3.2	4.2
	2	4.4	4.2	1.4	8.6	3.4	4.6	4.6	3.0	
	3	5.0	3.2	1.8	8.4	2.8	4.0	4.0	4.2	
	4	5.6	4.2	1.4	9.4	3.0	5.0	5.0	4.2	
	5	5.4	3.4	1.0	9.6	2.6	4.2	4.0	2.8	
50	1	5.4	4.0	2.0	14.2	1.8	3.0	3.0	1.6	6.8
	2	7.0	5.4	1.0	15.2	2.2	4.8	4.6	1.2	
	3	7.0	5.4	0.4	16.4	1.2	5.2	4.8	0.0	
	4	6.4	4.8	0.2	16.6	1.6	5.4	5.2	0.2	
	5	7.0	5.4	0.2	17.0	0.6	6.0	5.6	0.0	
150	1	6.8	5.6	1.0	15.6	0.0	0.6	0.4	0.2	100.0
	2	6.0	4.6	0.2	17.2	0.0	1.4	1.4		
	3	5.8	3.6	0.0	17.8	0.0	1.8	1.8		
	4	5.6	4.2	0.0	16.2	0.0	4.4	4.4		
	5	5.6	4.6	0.0	17.2	0.0	5.4	4.8		

3.4.2 Empirical Power

In this section, our goal is to compare the empirical power of all white noise tests. Here, two types of non-white noise series are considered:

Model 5. $x_t = Ax_{t-1} + e_t$, where e_t are independent and each e_t consists of p independent t_8 random variables, and the coefficient matrix $A = (a_{kl})$ is generated by: $a_{kl} \sim U(-0.25, 0.25)$ independently for $1 \le k, l \le k_0$, and $a_{kl} = 0$ otherwise.

Model 6. $x_t = Az_t$, where $z_t = (z_{1,t}, \dots, z_{p,t})^T$. For $1 \le k \le k_0$, $z_t = (z_{k,1}, \dots, z_{k,n})^T \sim N(0, \Sigma)$, where $k_0 = \min(\lceil p/5 \rceil, 12)$ and Σ is an $n \times n$ matrix with 1 on the diagonal and $0.5|i-j|^{-0.6}$ on the (i,j)-th element for $1 \le |i-j| \le 7$ and 0 on all the other elements. For $k \ge k_0$, $z_{1,t}, \dots, z_{k,t}$ are independently generated from t_8. The coefficient matrix $A = (a_{kl})$ is generated by: $a_{kl} \sim U(-1, 1)$ with probability $1/3$ and $a_{kl} = 0$ with probability $2/3$ independently for $1 \le k \ne l \le p$, and $a_{kk} = 0.8$ for $1 \le k \le p$.

Remark 3.8. *The likelihood ratio test (TB) does not involve lag parameter k, so the power curve is flat. For Lagrange multiplier test (LM), it is not applicable when $pk \ge n$. However, for convenience reasons, we let the curve go to zero when $p = 150$ at lag $k = 2, 3, 4, 5$.*

Figure 3.1-3.2 display the empirical power of all white noise tests against the lag number K used in the test. When the autocorrelation decays very fast, as in Model 5, Y_n^* is more powerful compared to our proposed tests, especially when p is small. As p increases, Y_n, Y_n^* and T_n, S_n have similar performance, where the power curve is very close to 1. However, those traditional white tests, such as portmanteau tests, are close to powerless. As the autocorrelation remains relatively strong in Model 6, all tests lose power as p becomes larger. This results from the different model structures. Specifically, those tests based on maximum-type test statistics (T_n, S_n, Y_n and Y_n^*) are really close to zero. This is because the autocovariance do not change with respect to the dimension p, while the variance of autocovariance increases as p increases. If we divide the autocovairance by its estimated standard deviation, the test statistics will become even smaller, which makes it harder to reject the null hypothesis. In later section, we adjust Model 6 to validate the aforementioned argument. In addition, those portmanteau tests are relatively more powerful than new tests. Since Model 6 has widespread signals within first k_0 rows and portmanteau tests sum up all the autocorrelations, it is expected to observe more power from them. It should be mentioned that the power curves of most of these tests are less informative, since these tests have different sizes. It is more reasonable to compare the power of these tests when they have same size first. This can be achieved by calibrating the critical values from each of the corresponding white noise model. In addition, instead of plotting the power curve against the lag value K, we are more interested in how sensitive our tests are when the dependence in the data changes. These extensive experiments and more discussions are presented in the following sections.

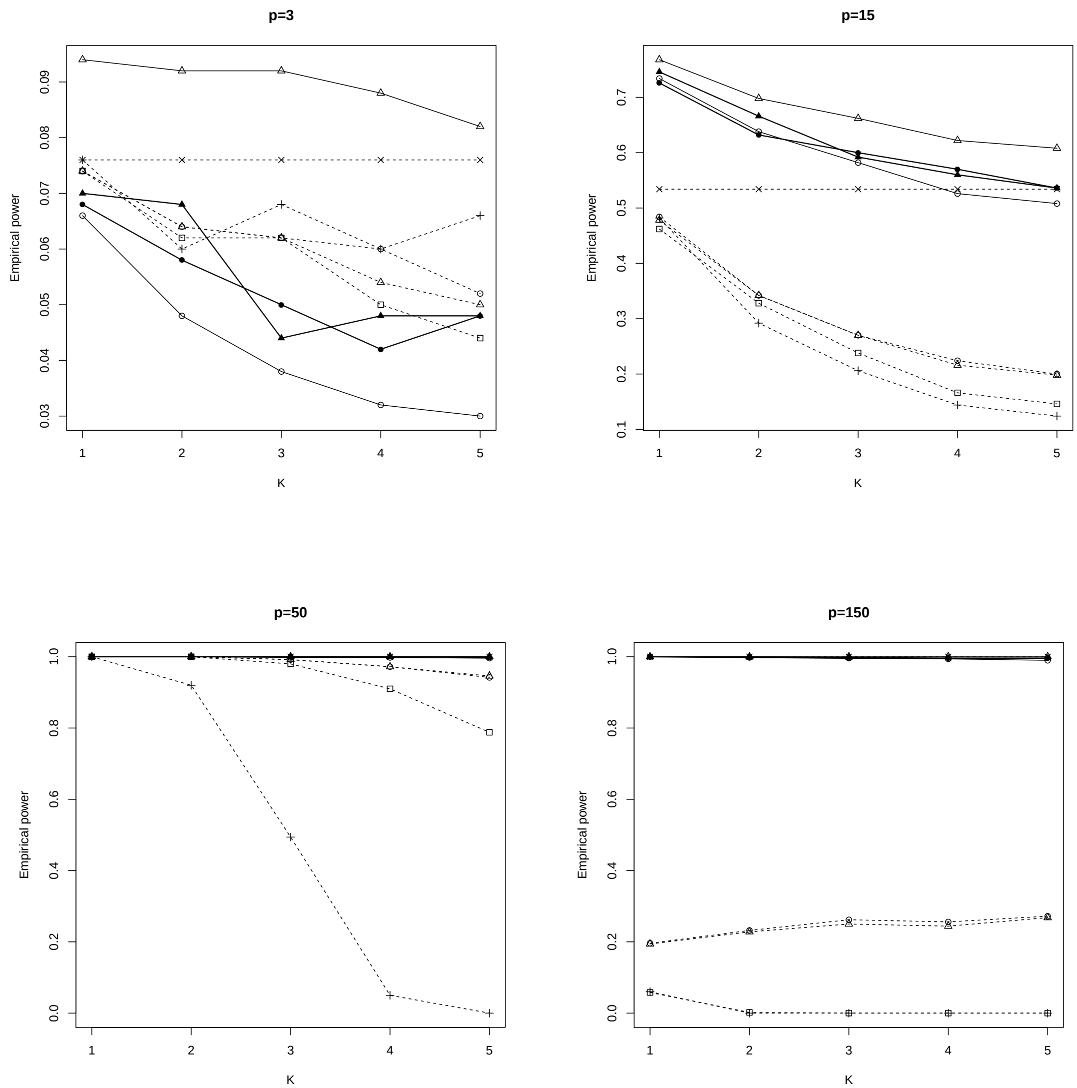


Figure 3.1: Plots of empirical power against lag K for new tests T_n (solid ●)and S_n(solid ▲), Y_n (solid ○) and $Y_n^\star$ (solid △), Q_1 (dashed □), Q_2 (dashed ○), Q_3 (dashed △), LM (dashed +), and TB (dashed ×). Data are generated from model 5.

3.4.3 Calibrated Critical Values

Instead of evaluating the critical values from dependent multiplier bootstrapping, we calibrate the critical values based on the corresponding white noise series. Specifically, consider two white noise models corresponding to Model 5 and Model 6 respectively:

Model 5 (white noise). $x_t = e_t$, where e_t are independent and each e_t consists of p independent t_8 random variables.

Model 6 (white noise). $x_t = Az_t$, where $z_t = (z_{1,t}, \ldots, z_{p,t})^T$. For $1 \le k \le k_0, z_t = (z_{k,1}, \ldots, z_{k,n})^T \sim N(0, \Sigma)$, where where $k_0 = \min(\lceil p/5 \rceil, 12)$ and Σ is an $n \times n$ matrix with 1 on the diagonal and 0 on all the other elements. For $k \ge k_0$, $z_{1,t}, \ldots, z_{k,t}$ are independently generated from t_8. The coefficient matrix $A = (a_{kl})$ is generated by: $a_{kl} \sim U(-1, 1)$ with probability $1/3$ and $a_{kl} = 0$ with probability $2/3$ independently for $1 \le k \ne l \le p$, and $a_{kk} = 0.8$ for $1 \le k \le p$.

Here, we try to remain the same structure with the original non-white noise model and remove the correlation at the same time. For instance, for model 5, the dependence comes from the lag-1 autoregressive term. So we remove the autoregressive term and use the innovation e_t as the corresponding white noise series for model 5. Similarly for model 6, the correlation only appears in the first k_0 rows. So we use an identity matrix as the covariance matrix for the normal distribution used to generated the first k_0 rows, and the rest rows are same with original non-white noise model. In this way, before doing white noise test, we can generate 2000 white noise samples, calculate the test statistics for each test, and take the α-th upper quantile as the threshold. Ideally, the empirical type I error is designed to be 5% if the calibrated critical values are used. Therefore, by using the calibrated critical values to do the test, we are able to compare all the white noise tests at the same size. Similarly, we reject the null hypothesis if the test statistics is greater than the calibrated critical values.

Figure 3.3-3.4 display the power plots of all white noise tests against lag K using calibrated critical values. Generally, the power of Y_n^* decreases due to the correction of significance level. While the general patterns seem to be the same as previously stated, it is worth observing the gap between our proposed tests T_n,S_n and Y_n,Y_n^* reduces significantly. In summary, our tests are still dominant in terms of a better control on the size. It is always more important to calibrate the size accurately first and then improve the power. In the last section, there are some further discussion where we propose a new method to improve the empirical power of our new tests.

3.4.4 Sensitivity to Dependence

In order to investigate how sensitive the white noise tests are when the dependence changes, we modify the level of dependence in the non-white noise model. For each model, how we measure the dependence of the data is different. For instance, in Model 5, the dependence of data comes from the autoregressive matrix A. Write A_0 as the original matrix described in Model 5. Let $|A| = d|A_0|$ where $d = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. When $d = 0$, this is exactly the same white noise model stated in Section 3.4.3. When d becomes larger, the dependence of data becomes stronger. While in Model 6, the covariance matrix of first k_0 rows is the source of dependence. Let Σ be an $n \times n$ matrix with 1 on the diagonal and $d|i-j|^{-0.6}$ on the (i, j)-th element for $1 \leq |i-j| \leq 7$, where $d = 0, 0.1, 0.2, 0.3, 0.4, 0.5$. When $d = 0$, the covariance matrix Σ becomes a diagonal matrix, which leads to the white noise model in Section 3.4.3. The dependence of the data becomes stronger as d increases. Then we are able to plot the empirical power of each tests against the dependence measure d. Here we only consider lag $k = 1$. In other words, there will be p^2 cross covariances involved in the test. Note that we also calibrate the critical values from the white noise model as Section 3.4.3 described. The results are presented in Figure 3.5 and Figure 3.6. First of all, when $d = 0$, the generated series is indeed a white noise. So the empirical rejecting probability is 5%, which is why all power curves start from around 5%. Notice that when $p = 150$, the power for those portmanteau tests are higher than expected. Although the expected value of the portmanteau statistic is the same with the truth, the variance for each $\hat{\rho_{ij}}$ is $O_p(1/\sqrt{n})$. Since portmanteau tests sum up all autocorrelation, its variance can be very large due to the accumulated variation from each of the p^2 autocorrelation. On the other hand, the empirical power increases as the dependency becomes stronger. When $d = 1$, the results are the same as shown in Figure 3.3 and Figure 3.4 at lag 1. Similar to the aforementioned discussion, as the autocorrelation remains relatively strong in Model 6, compared to Model 5, all the tests based on maximum-type statistics are relatively powerless when p is large. However, in Model 5, those tests are substantially powerful.

In order to investigate the interesting pattern in Model 6, we adjust the variance of loading matrix in Model 6, so that the variance of each series does not increase as p increases. Specifically, the variance of each series is partially influenced by the variance from coefficient matrix A, which is accumulated when the dimension p increases. Thus, we use the following modified model to generate the non-white series:

Model 6 (adjusted). $x_t = Az_t$, where $z_t = (z_{1,t}, \dots, z_{p,t})^T$. For $1 \le k \le k_0, z_t = (z_{k,1}, \dots, z_{k,n})^T \sim N(0, \Sigma)$, where where $k_0 = \min(\lceil p/5 \rceil, 12)$ and Σ is an $n \times n$ matrix with 1 on the diagonal and $0.5|i-j|^{-0.6}$ on the (i,j)-th element for $1 \le |i-j| \le 7$ and 0 on all the other elements. For $k \ge k_0$, $z_{1,t}, \dots, z_{k,t}$ are independently generated from t_8. The coefficient matrix $A = (a_{kl})$ is generated by: $a_{kl} \sim U(-1/\sqrt{p}, 1/\sqrt{p})$ with probability $1/3$ and $a_{kl} = 0$ with probability $2/3$ independently for $1 \le k \ne l \le p$, and $a_{kk} = 0.8$ for $1 \le k \le p$.

The results are displayed in Figure 3.7. As expected, there is a significant improvement from those solid curves compared to the power plots where data is generated from the original Model 6, especially for $p = 150$. These results confirm the discussion in Section 3.4.2.

3.5 Real Data Analysis

One of the most important application of white noise test is to check the model adequacy. In this section, we use the white noise tests as diagnostic tool to identify a suitable number of terms used in the autoregressive model of matrix-valued time series. We fit the p-MAR(1): p-term matrix autoregressive model with 1 lag, estimated using the least-square method (Chen et al., 2018):

$$X_t = A_1 X_{t-1} B_1' + \cdots + A_p X_{t-1} B_p' + E_t,$$

and test on the residuals using our proposed test T_n and S_n, as well as Y_n and Y_n^* proposed by Chang et al. (2017). We use the same dataset presented in Section 5.2 of Chen et al. (2018), the economic indicators from five countries: Canada, France, Germany, United Kingdom and United States. Four quarterly indicators are selected: 3-month interbank interest rate (first-order differenced series), GDP growth (first-order differenced log of GDP series), total manufacture production growth (first-order differenced log of production series) and total consumer price index (growth from the last period). The data was downloaded from Organisation for Economic

Co-operation and Development (OECD) at https://data.oecd.org/, and was pre-processed the same way as in Chen et al. (2018) before fitting the autoregressive model. To be specific, all series are normalized so that the combined variance of each indicator is 1, and the seasonality of CPI is adjusted by subtracting the sample quarterly means. The ACF plot of original series is shown in Figure 3.8.

There are several significant lags in Figure 3.8, so we continue to fit the 1-MAR(1) model, and apply those white noise tests on the residuals. The significance level is $\alpha = 0.05$. Select the lag to be $K = 1, 2, 3, 4, 5$. We generate $B = 2000$ bootstrap samples to evaluate the critical values. The p-value of each test is reported in Table 3.5. Specifically, T_n and S_n reject the null hypothesis that the residual series is white noise for all lags, while residuals pass Y_n and Y_n^* with pretty high p-value. If we add one more term to fit a 2-MAR(1) model, the residuals pass T_n and S_n when K is small. When $K = 4$ and $K = 5$, these two tests still reject the null hypothesis. However, the p-values are relatively close to 5%. So we stop here and select 2-MAR(1) model to fit the indicator data. We also plot the ACF of residuals from each model for reference.

In summary, for this example, the tests proposed by Chang et al. (2017) fail to detect the correlations, while our white noise tests direct us to use the MAR(1) model with two terms. Although there are still some significant lags in Figure 3.10, rather than adding more terms to MAR(1) model, we may consider a more sophisticated model such as MAR(2) or MAR(3) to obtain a much cleaner residuals.

Table 3.5: The p-value(%) of white noise tests for testing residuals

Model	Test	K				
		1	2	3	4	5
1-MAR(1)	T_n	0.10	0.20	0.40	0.65	0.70
	S_n	1.05	2.35	2.25	0.20	0.25
	Y_n	70.20	54.15	56.80	59.60	59.10
	Y_n^*	55.15	23.60	24.85	25.30	25.15
2-MAR(1)	T_n	7.30	11.35	14.85	3.05	3.70
	S_n	13.75	14.95	19.80	3.10	3.95
	Y_n	71.15	75.00	78.80	75.70	76.65
	Y_n^*	22.00	31.45	37.70	44.65	45.05

3.6 Further Discussion

In this section, we propose a new test using the self-normalized cross correlation in order to gain more power. The motivation is the variance of sample cross correlation tends to be smaller under the alternative hypothesis. In order to make the signal stronger, we can divide the sample cross correlation by its estimated standard deviation. Notice that under null hypothesis, this is equivalent to normalize the sample cross covariance. The estimation of variance for sample cross correlation is more complicated. Here we adopt a block estimates. Specifically, partition n samples into B blocks with block size m. The common practice is to select $m = n^{1/3}$. Then for each block, calculate the sample cross correlation for all pairs of $\{(i, j) : 1 \leq i, j \leq p\}$. In this way, we have B samples of cross correlations. Then use the sample variance of these B samples to estimate the variance of sample cross correlation. We also evaluate the critical values by calibration. The test based on maximum absolute self-normalized cross correlation is denoted as Z_n. There is also an improved version by using the sum of square of maximum and minimum of self-normalized cross correlation, which is denoted as Z'_n. In order to demonstrate the result clearly, we use a simple model to generate the non-white noise series.

Model 7. $x_t = Ax_{t-1} + e_t$, where e_t are independent and each e_t consists of p independent t_8 random variables, and the coefficient matrix $|A| = d|A_0|$ where $d = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. $A_0 = (a_{kl})$ is generated by: $a_{kk} \sim U(-0.5, 0.5)$ independently for $1 \leq k \leq k_0$, and $a_{kl} = 0$ for $1 \leq k, l \leq k_0, k \neq l$ and $k, l > k_0$.

The first k_0 rows in Model 7 are independent $\mathrm{AR}(1)$ process, and the rest rows are independent t_8 random variables. Recall for an $\mathrm{AR}(1)$ process $x_t = \phi x_{t-1} + e_t$with autoregressive coefficient $-1 < \phi < 1$ and i.i.d. innovations e_t, the asymptotic variance of lag-1 sample autocovariance $\sqrt{n}\hat{\rho}(1)$ is $1 - \phi^2$. Thus, by normalizing the sample cross and auto correlation by its corresponding standard deviation, the dependency signals are stronger to be detected. We demonstrate these findings by the power plots against dependence d in Figure 3.11. When $d = 0$, this is exactly the same white noise model stated in Section 3.4.3. When d becomes larger, the dependence of data becomes stronger. As shown in Figure 3.11, the tests based on self-normalized cross correlation Z_n and Z'_n (the red curves) are more powerful than other maximum-type tests as expected.

Last but not least, we remark that the most important application of the white noise tests is to assess the adequacy of a fitted model. Much of the literature on the classical portmanteau tests has focused on the impact of vector autoregressive and moving average modeling on the distribution of sample autocorrelations of the residuals, and on that of the portmanteau test statistics, see for example Li and McLeod (1981) and Lütkepohl (2005). For high dimensional time series, the model building itself becomes more complicated, and has to be performed with some regularization. The effect of the regularized estimation procedure is not yet clear for the high dimensional white noise tests based on residuals. Although it might not affect the asymptotic distributions of the test statistics, its influence on the finite sample performances should not be neglected. Deep understandings of this problem, both theoretical and empirical, are of our utmost interest in future research.

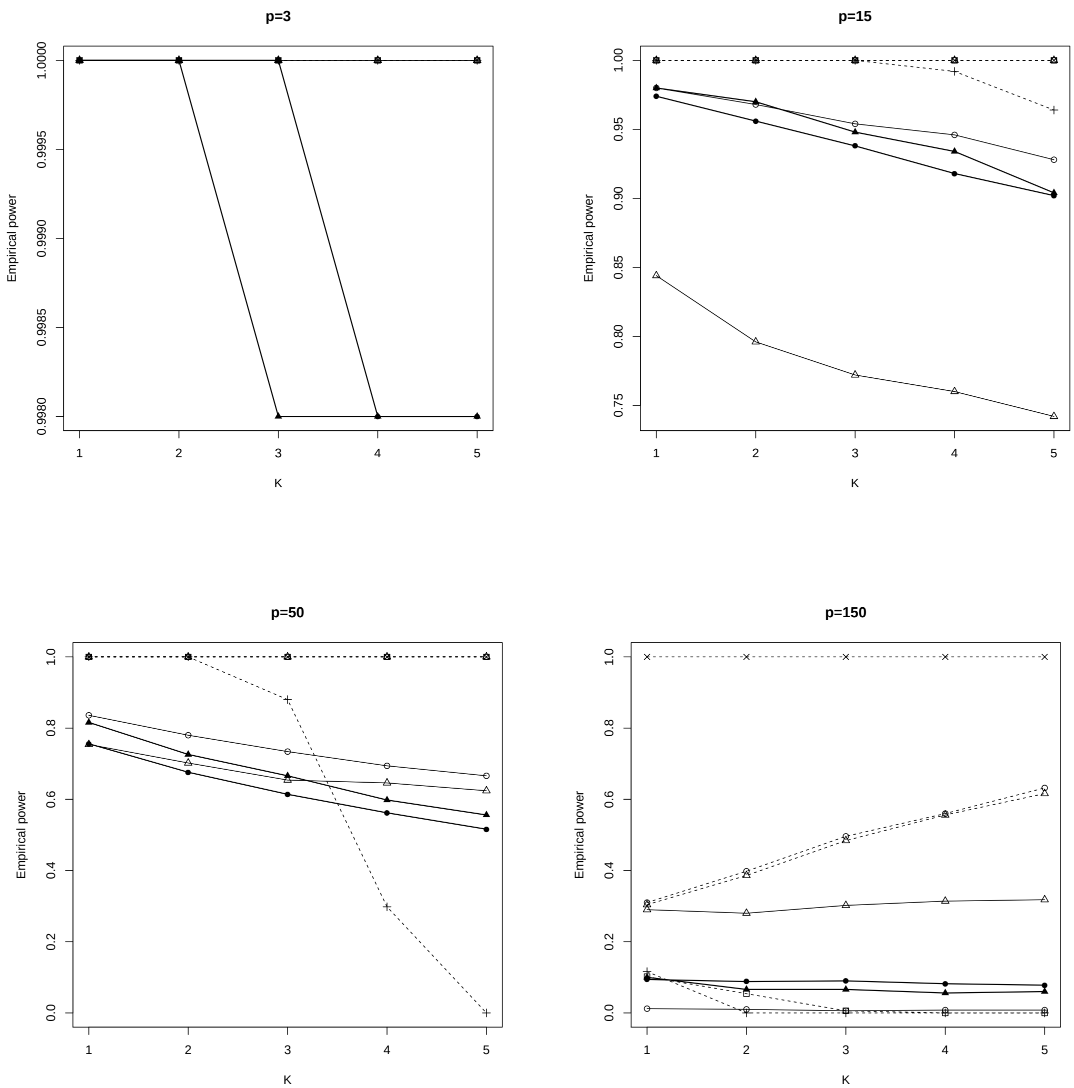


Figure 3.2: Plots of empirical power against lag K for new tests T_n (solid ●)and S_n(solid ▲), Y_n (solid ○) and $Y_n^{\star}$ (solid △), Q_1 (dashed □), Q_2 (dashed ○), Q_3 (dashed △), LM (dashed +), and TB (dashed ×). Data are generated from model 6.

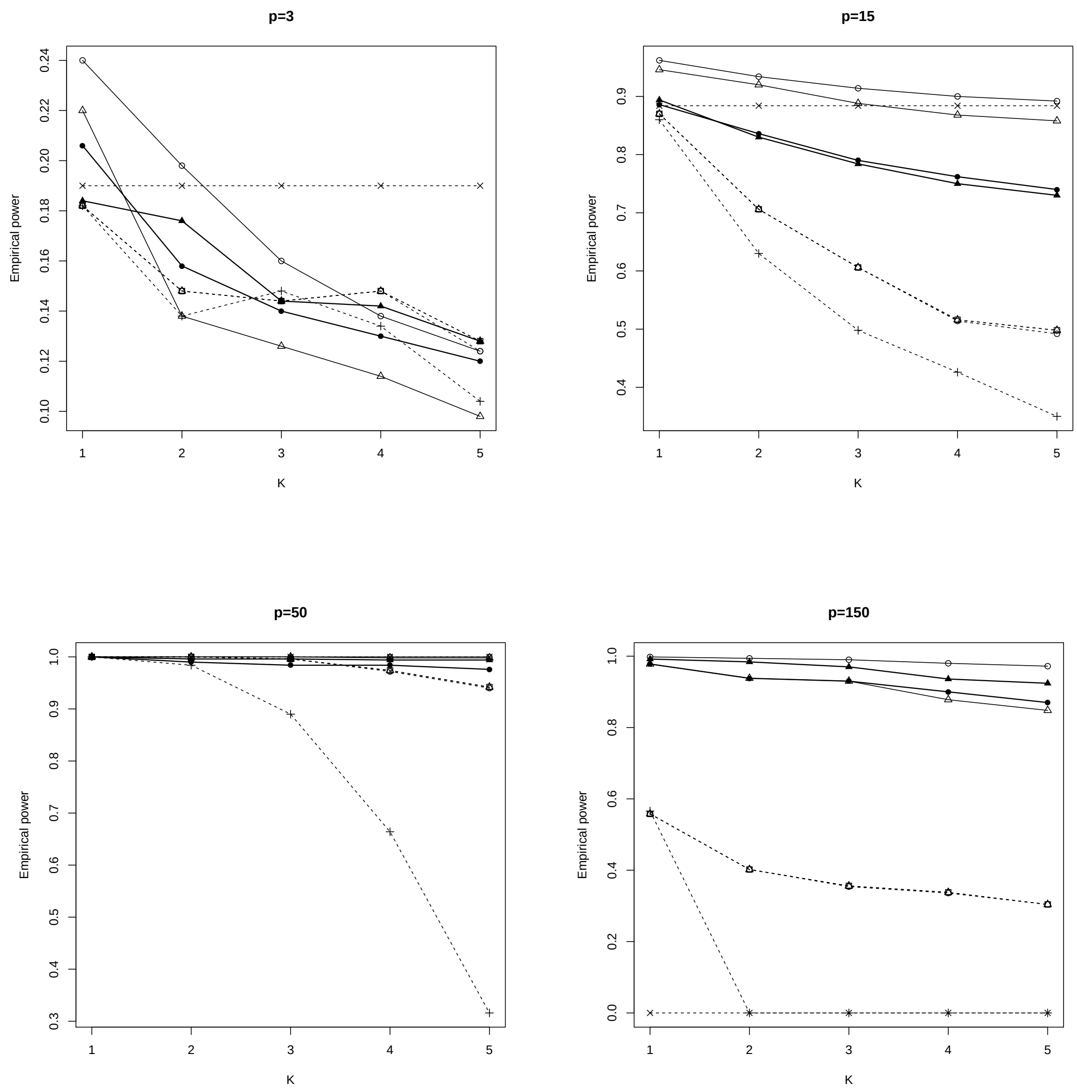


Figure 3.3: Plots of empirical power against lag K for new tests T_n (solid ●)and S_n(solid ▲), Y_n (solid ○) and $Y_n^\star$ (solid △), Q_1 (dashed □), Q_2 (dashed ○), Q_3 (dashed △), LM (dashed +), and TB (dashed ×). Data are generated from model 5. Calibrated critical values are used.

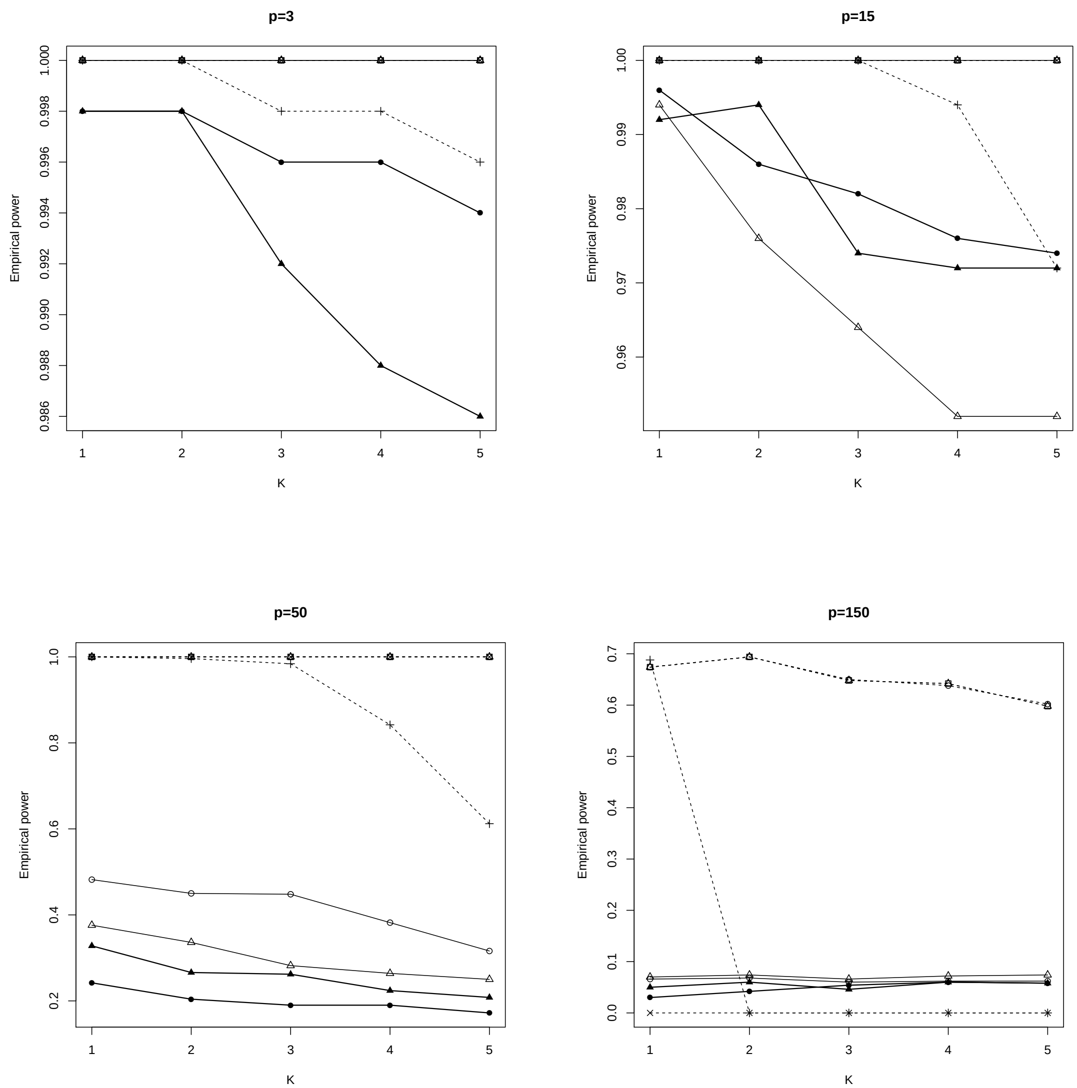


Figure 3.4: Plots of empirical power against lag K for new tests T_n (solid ●)and S_n(solid ▲), Y_n (solid ○) and $Y_n^{\star}$ (solid △), Q_1 (dashed □), Q_2 (dashed ○), Q_3 (dashed △), LM (dashed +), and TB (dashed ×). Data are generated from model 6. Calibrated critical values are used.

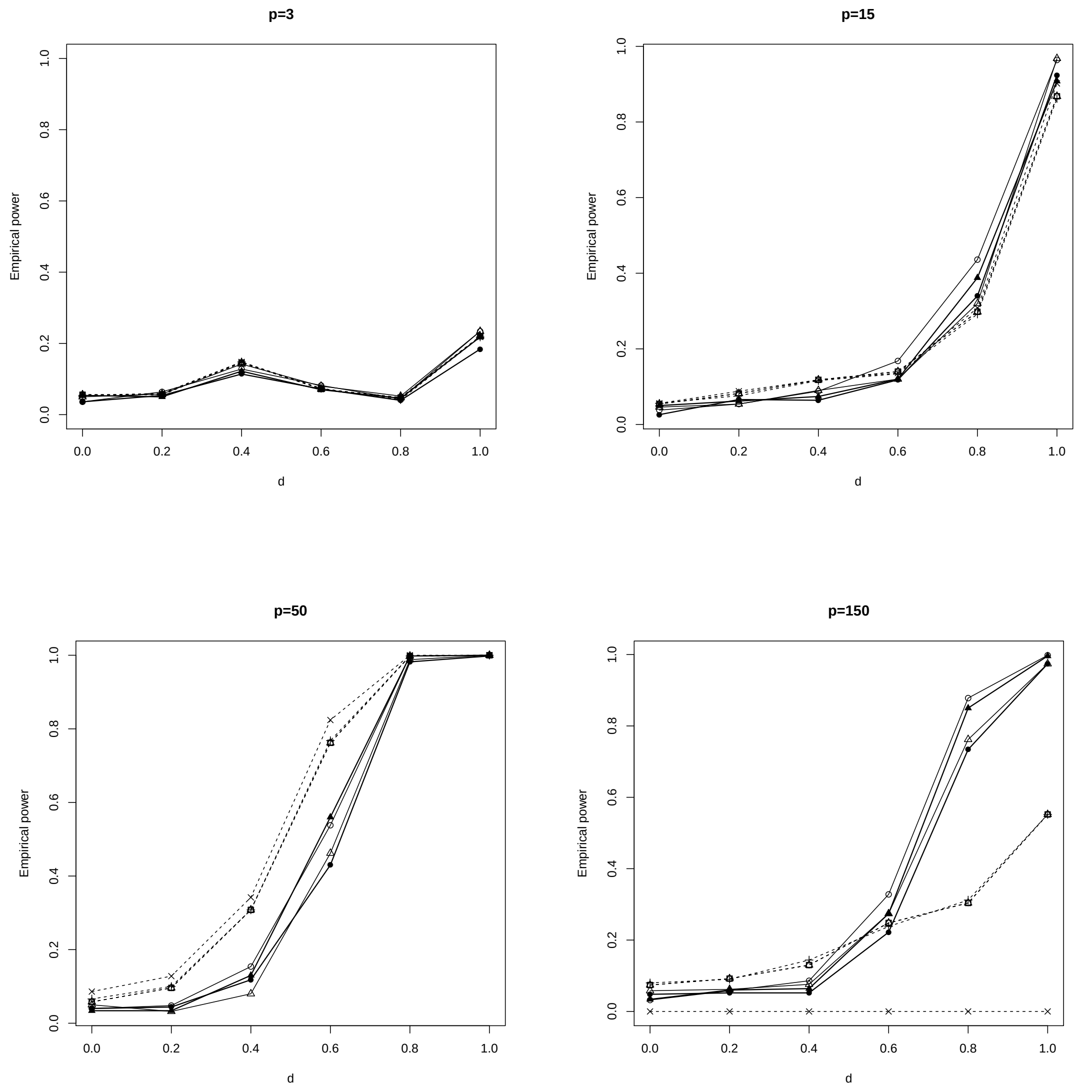


Figure 3.5: Plots of empirical power against dependence d for new tests T_n (solid ●)and S_n(solid ▲), Y_n (solid ○) and $Y_n^\star$ (solid △), Q_1 (dashed □), Q_2 (dashed ○), Q_3 (dashed △), LM (dashed +), and TB (dashed ×). Data are generated from model 5 with adjusted dependence. Calibrated critical values are used.

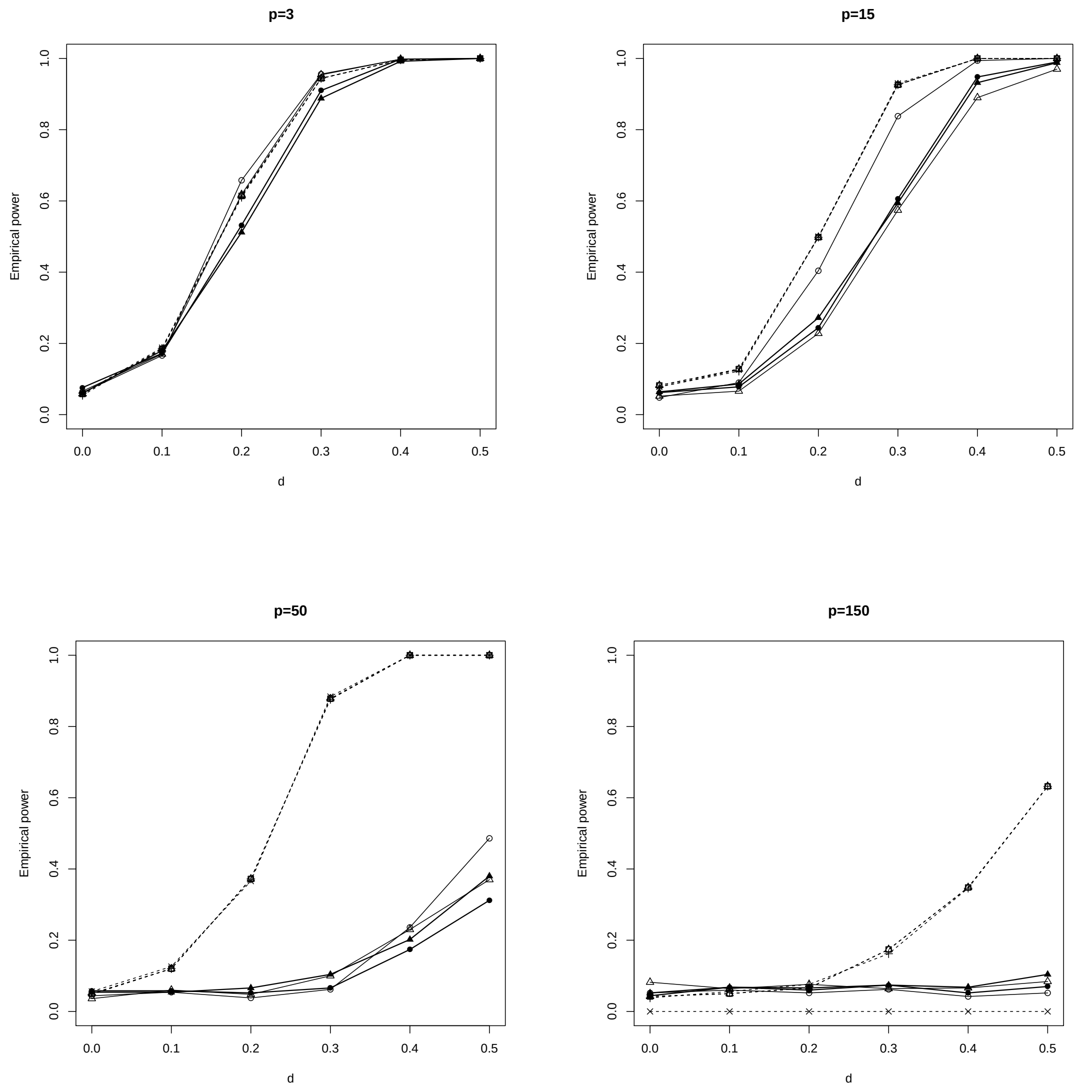


Figure 3.6: Plots of empirical power against dependence d for new tests T_n (solid •)and S_n(solid ▲), Y_n (solid ○) and $Y_n^\star$ (solid △), Q_1 (dashed □), Q_2 (dashed ○), Q_3 (dashed △), LM (dashed +), and TB (dashed ×). Data are generated from model 6 with adjusted dependence. Calibrated critical values are used.

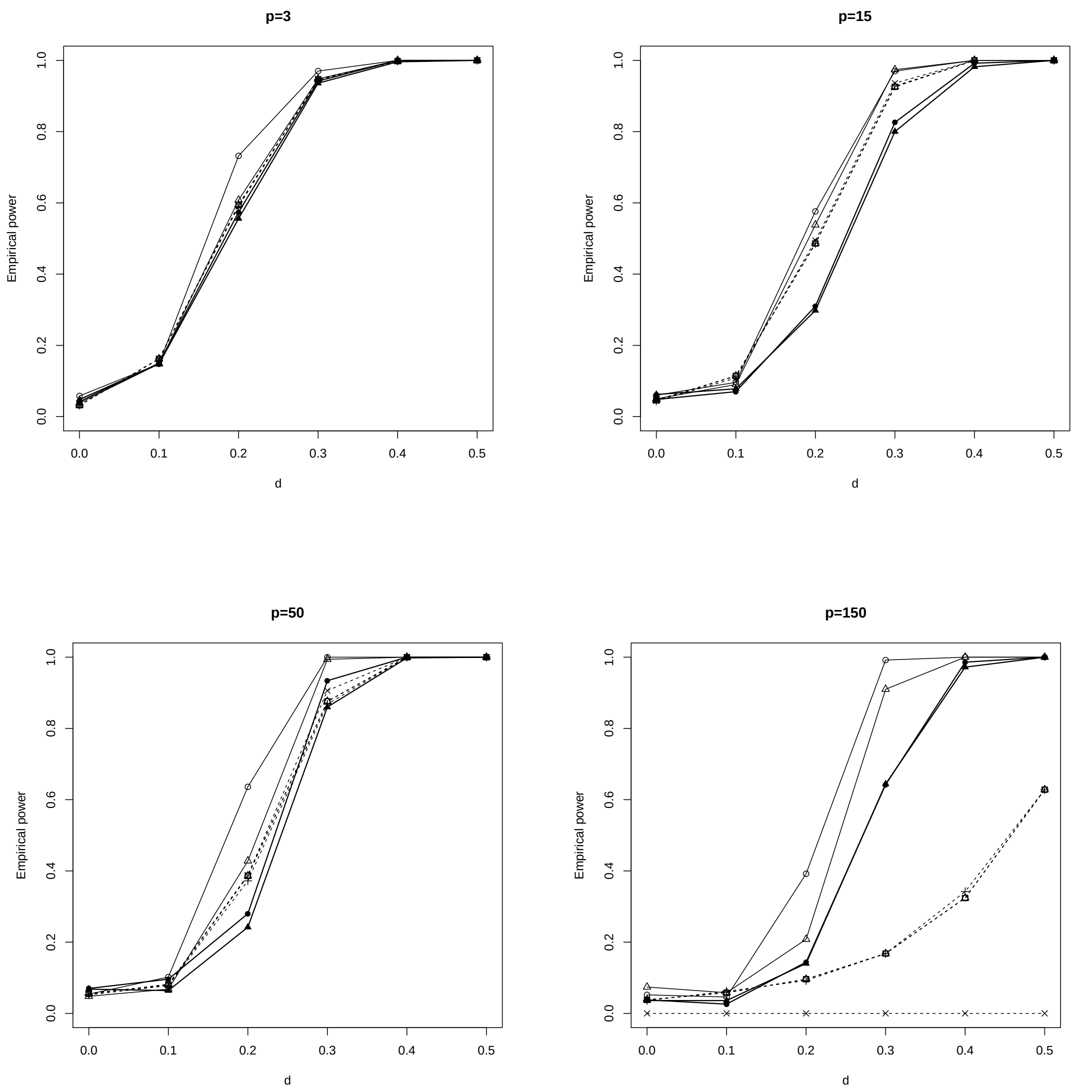


Figure 3.7: Plots of empirical power against dependence d for new tests T_n (solid ●)and S_n(solid ▲), Y_n (solid ○) and $Y_n^{\star}$ (solid △), Q_1 (dashed □), Q_2 (dashed ○), Q_3 (dashed △), LM (dashed +), and TB (dashed ×). Data are generated from model 6 (adjusted). Calibrated critical values are used.

Figure 3.8: ACF of original series

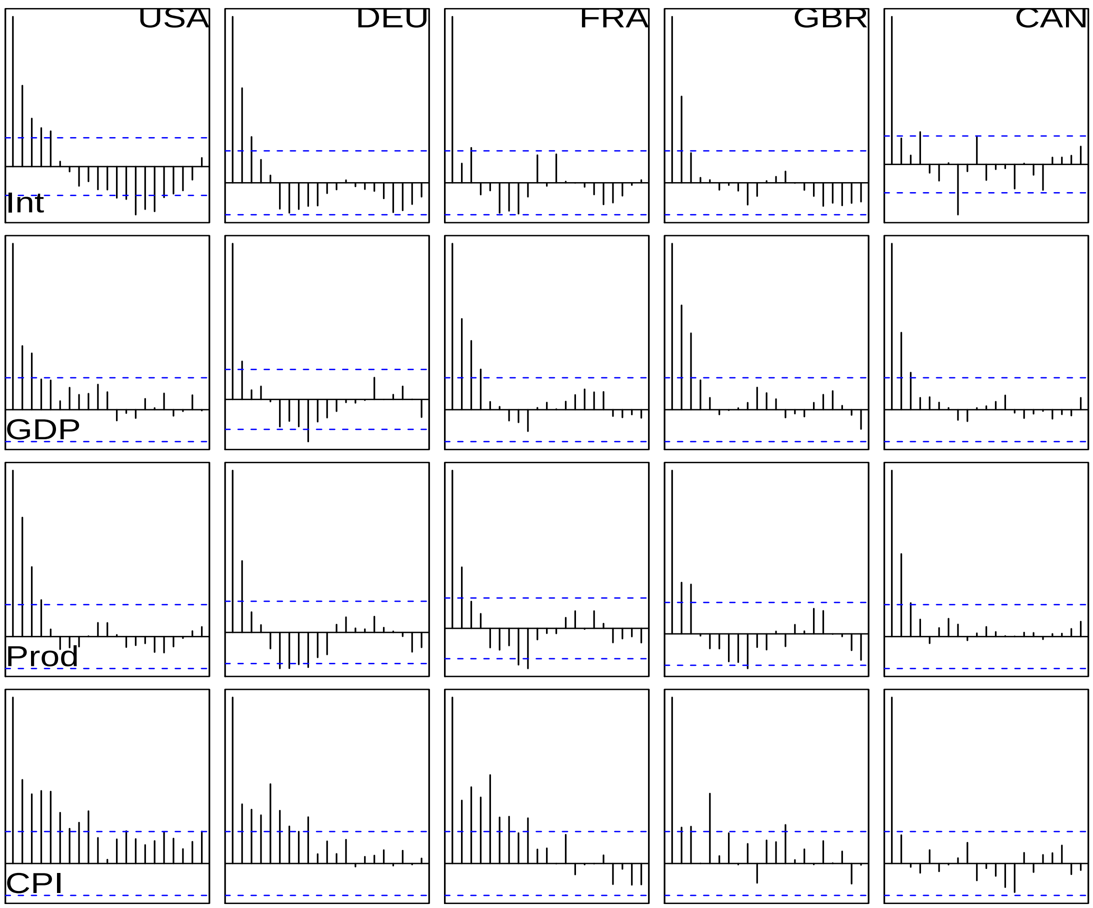


Figure 3.9: ACF of residuals from 1-MAR(1)

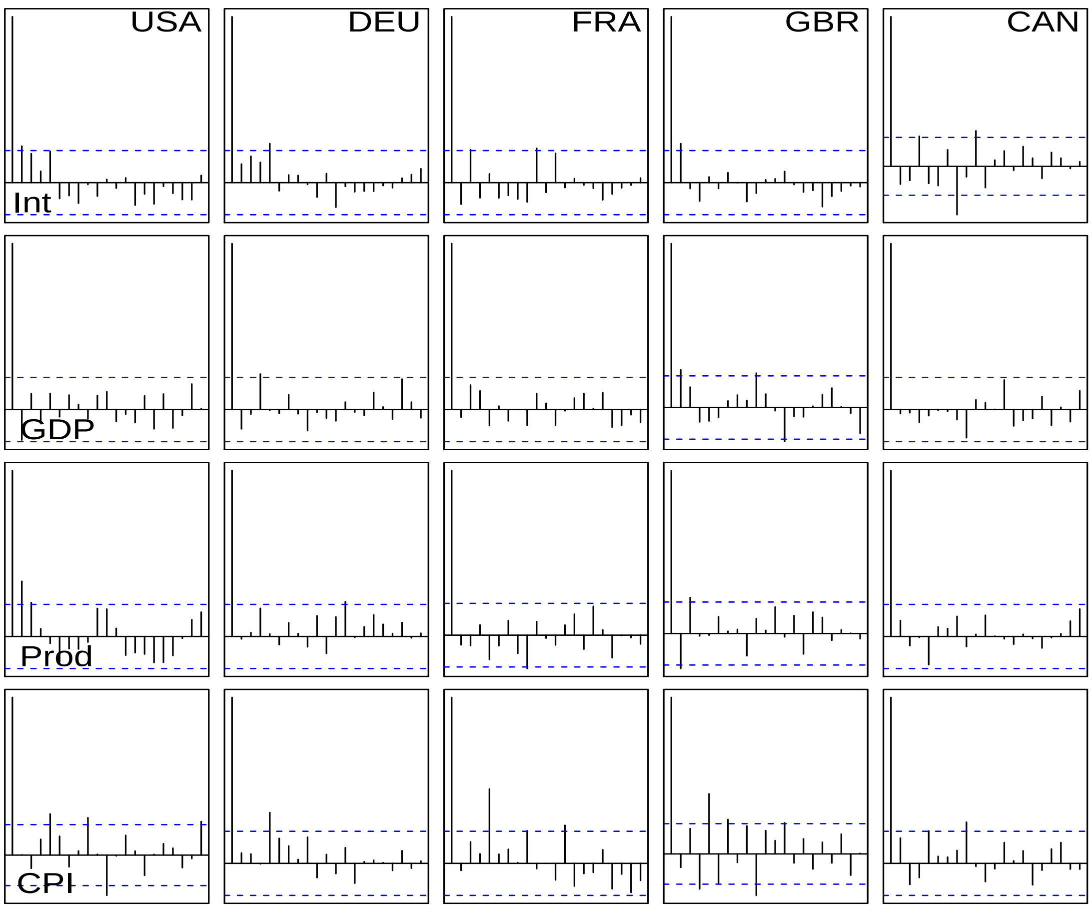
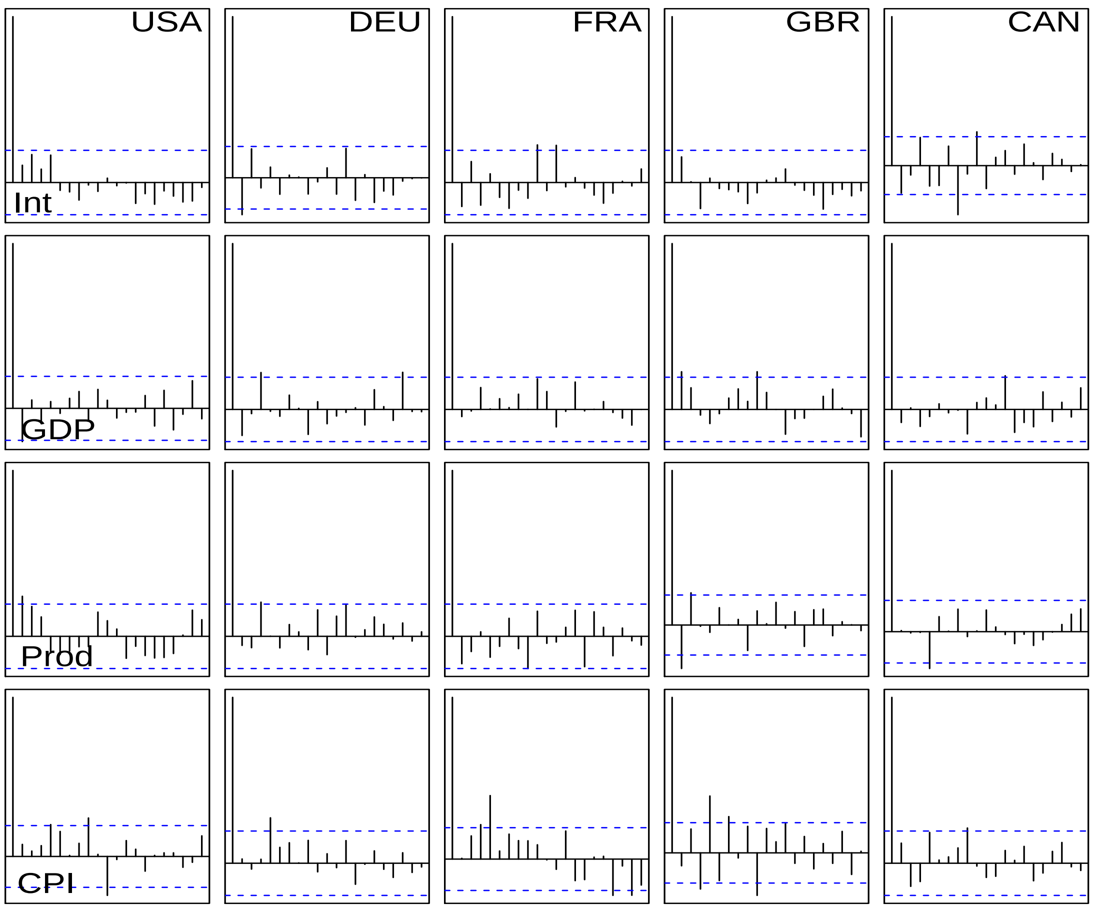


Figure 3.10: ACF of residuals from 2-MAR(1)



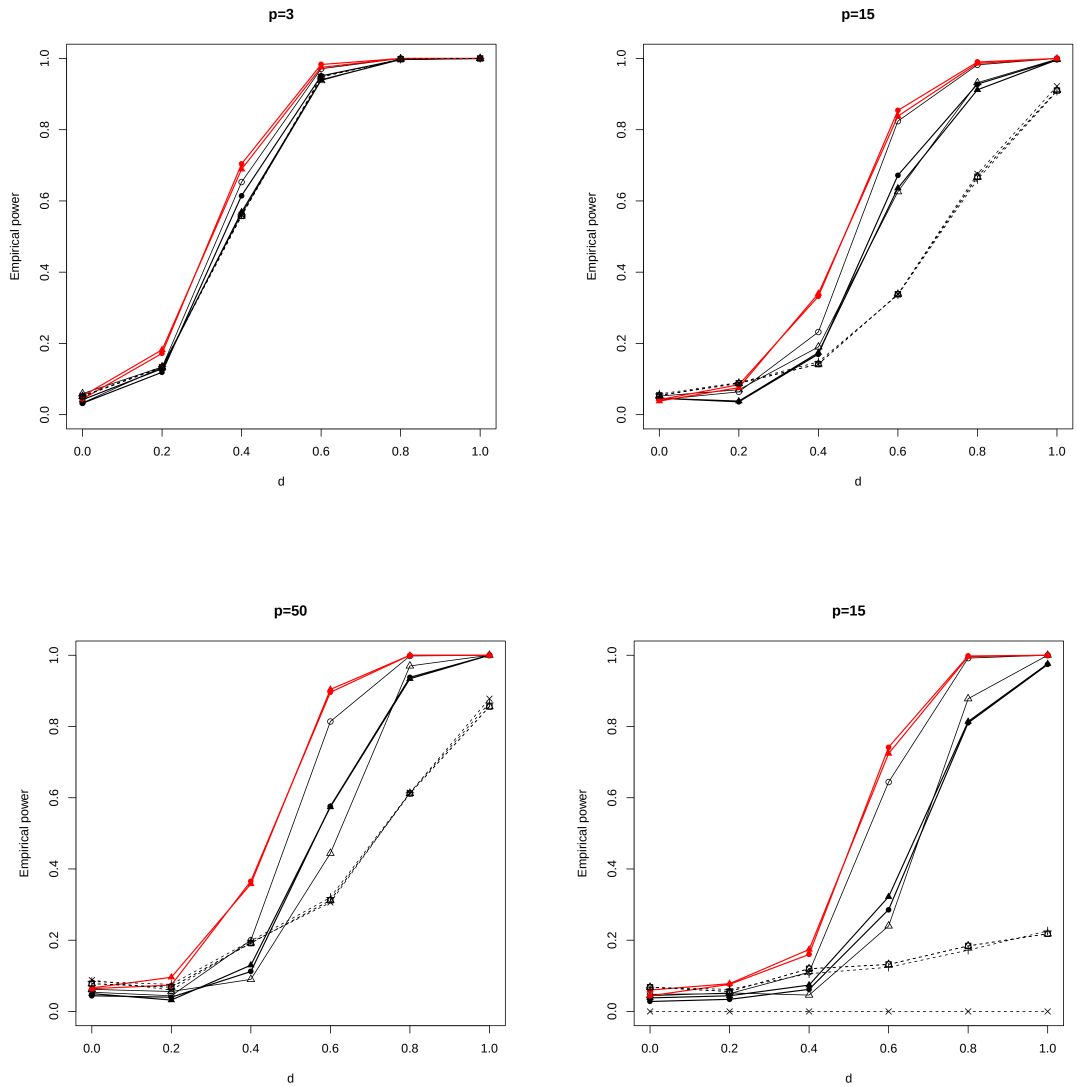


Figure 3.11: Plots of empirical power against lag K for new tests Z_n (red solid ●)and Z'_n(red solid ▲), T_n (solid ●)and S_n(solid ▲), Y_n (solid ○) and $Y_n^\star$ (solid △), Q_1 (dashed □), Q_2 (dashed ○), Q_3 (dashed △), LM (dashed +), and TB (dashed ×). Data are generated from model 7. Calibrated critical values are used.

Chapter 4

Concluding Remarks

In the first project, we propose a monotone transform of the squared normal extremes, which leads to a faster convergence to the limiting Gumbel distribution. We show that both the pointwise and the uniform convergence rates are of the order $(\log n)^{-3}$. It improves the best existing result, which is at the rate $(\log n)^{-2}$. While the faster convergence rates are often reflected when the sample size is large, after the proposed transform, the distribution of the squared normal extreme is very close to the limiting one even when the sample size is moderate, with only hundreds of observations. Furthermore, it is stochastically dominated by the limiting distribution. This is important because the asymptotic test based on the transformed maximum is conservative, so that the type I error is guaranteed to be controlled at the nominal level.

In the second project, we consider the white noise tests for high dimensional time series. Two tests are proposed: (i) maximum self-normalized sample autocovariance, and (ii) sum of squared maximum and minimum sample autocovariances (after normalization). Our theoretical analysis is under the paradigm of high dimensional statistics, where the dimension is allowed to grow exponentially with the length of the time series. We show that the limiting distribution of the first test statistic converges to the Gumbel distribution, and that of the second one converges to the convolution of two independent Gumbel random variables. To calibrate the sizes of the tests when the sample size is not very large, we propose to use the dependent Gaussian multiplier bootstrap, and establish its consistency. We conduct an extensive numerical studies, and find that the proposed tests do enjoy better type I error controls, compared with existing testing procedures.

We also find that while the sizes are better controlled, the proposed tests also have comparable powers with other ones. To further improve the power, one approach is to consider the self-normalized sample autocorrelations. Another approach is to use the idea of higher criticism tests (Donoho et al., 2004) so that the test is more adaptive to the unknown dependence pattern under the alternative. These directions will be investigated in future research.

Finally, for most tests based on the maximum, two approximations are involved, a normal approximation and an asymptotic approximation of the normal extreme by the Gumbel distribution. The second approximation is investigated in the first project of this dissertation. On the other hand, a thorough analysis of the rate of convergence for the first one is needed, and is also of our future interests.

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariant matrix estimation. *Econometrica*, 59(3):817–858.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bartlett, M. S. (1955). An introduction to stochastic processes, with special reference to methods and applications.

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

Berman, S. M. (1964). Limit theorems for the maximum term in stationary sequences. *The Annals of Mathematical Statistics*, pages 502–516.

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

Chang, J., Qiu, Y., Yao, Q., and Zou, T. (2018). Confidence regions for entries of a large precision matrix. *Journal of Econometrics*, 206(1):57–82.

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

Chen, R., Xiao, H., and Yang, D. (2018). Autoregressive models for matrix-valued time series. *arXiv preprint arXiv:1812.08916*.

Chen, R., Yang, D., and Zhang, C.-h. (2019). Factor models for high-dimensional tensor time series. *arXiv preprint arXiv:1905.07530*.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.*, 41(6):2786–2819.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, 162(1):47–70.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2016). Testing many moment inequalities. Technical report, cemmap working paper.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.*, 45(4):2309–2352.

Chitturi, R. V. (1974). Distribution of residual autocorrelations in multiple autoregressive schemes. *J. Amer. Statist. Assoc.*, 69:928–934.

Cohen, J. P. (1982). The penultimate form of approximation to normal extremes. *Adv. in Appl. Probab.*, 14(2):324–339.

Davis, R. A., Zang, P., and Zheng, T. (2016). Sparse vector autoregressive modeling. *J. Comput. Graph. Statist.*, 25(4):1077–1096.

Davydov, Y. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory of Probability & Its Applications*, 13(4):691–696.

de Haan, L. and Resnick, S. (1996). Second-order regular variation and rates of convergence in extreme-value theory. *Ann. Probab.*, 24(1):97–124.

Deo, R. S. (2000). Spectral tests of the martingale hypothesis under conditional heteroscedasticity. *J. Econometrics*, 99(2):291–315.

Donoho, D., Jin, J., et al. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, 32(3):962–994.

Durlauf, S. N. (1991). Spectral based testing of the martingale hypothesis. *J. Econometrics*, 50(3):355–376.

Eberlein, E. (1984). Weak convergence of partial sums of absolutely regular sequences. *Statistics & probability letters*, 2(5):291–293.

Escanciano, J. C. and Lobato, I. N. (2009). Testing the martingale hypothesis. In *Palgrave handbook of econometrics*, pages 972–1003. Springer.

Fan, J., Zhang, C., Zhang, J., et al. (2001). Generalized likelihood ratio statistics and wilks phenomenon. *The Annals of statistics*, 29(1):153–193.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *J. Amer. Statist. Assoc.*, 100(471):830–840.

Ghosh, S., Khare, K., and Michailidis, G. (2019). High-dimensional posterior consistency in bayesian vector autoregressive models. *Journal of the American Statistical Association*, 114(526):735–748.

Guo, S., Wang, Y., and Yao, Q. (2016). High-dimensional and banded vector autoregressions. *Biometrika*, 103(4):889–903.

Hall, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probab.*, 16(2):433–439.

Hall, P. (1980). Estimating probabilities for normal extremes. *Adv. in Appl. Probab.*, 12(2):491–500.

Hall, P. (1991). On convergence rates of suprema. *Probab. Theory Related Fields*, 89(4):447–455.

Hall, W. J. and Wellner, J. A. (1979). The rate of convergence in law of the maximum of an exponential sample. *Statist. Neerlandica*, 33(3):151–154.

Han, F., Lu, H., and Liu, H. (2015). A direct estimation of high dimensional stationary vector autoregressions. *J. Mach. Learn. Res.*, 16:3115–3150.

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

Han, F., Xu, S., and Liu, H. (2016). Rate-optimal estimation of high dimensional time series. Technical report, Washington University, Department of Statistics.

Hill, J. B. and Motegi, K. (2016). A Max-Correlation White Noise Test for Weakly Dependent Time Series. *ArXiv e-prints*.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica*, 64(4):837–864.

Horowitz, J. L., Lobato, I., Nankervis, J. C., and Savin, N. (2006). Bootstrapping the box–pierce q test: a robust test of uncorrelatedness. *Journal of Econometrics*, 133(2):841–862.

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

Hosking, J. R. M. (1981). Lagrange-multiplier tests of multivariate time-series models. *J. Roy. Statist. Soc. Ser. B*, 43(2):219–230.

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

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *J. Econometrics*, 186(2):325–344.

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

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

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York-Berlin.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. Roy. Statist. Soc. Ser. B*, 43(2):231–239.

Ljung, G. M. and Box, G. E. P. (1979). The likelihood function of stationary autoregressive-moving average models. *Biometrika*, 66(2):265–270.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag, Berlin.

Marques, F. J., Coelho, C. A., and De Carvalho, M. (2015). On the distribution of linear combinations of independent gumbel random variables. *Statistics and Computing*, 25(3):683–701.

Nardi, Y. and Rinaldo, A. (2011). Autoregressive process modeling via the Lasso procedure. *J. Multivariate Anal.*, 102(3):528–549.

Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.*, 39(2):1069–1097.

Nicholson, W. B., Matteson, D. S., and Bien, J. (2017). Varx-l: Structured regularization for large vector autoregressions with exogenous variables. *International Journal of Forecasting*, 33(3):627 – 651.

Peng, Z., Saralees, N., and Lin, F. (2010). Convergence rate of extremes for the general error distribution. *J. Appl. Probab.*, 47(3):668–679.

Poskitt, D. S. and Tremayne, A. R. (1982). Diagnostic test for multiple time series models. *Ann. Statist.*, 10(1):114–120.

Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Econometrics*, 47(1):67–84.

Romano, J. P. and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.*, 91(434):590–600.

Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108.

Rootzén, H. (1983). The rate of convergence of extremes of stationary normal sequences. *Adv. in Appl. Probab.*, 15(1):54–80.

Shao, X. (2010). The dependent wild bootstrap. *J. Amer. Statist. Assoc.*, 105(489):218–235.

Shao, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory*, 27(2):312–343.

Slepian, D. (1962). The one-sided barrier problem for gaussian noise. *Bell System Technical Journal*, 41(2):463–501.

Smith, R. L. (1982). Uniform rates of convergence in extreme-value theory. *Adv. in Appl. Probab.*, 14(3):600–622.

Song, S. and Bickel, P. J. (2011). Large Vector Auto Regressions. *ArXiv e-prints*.

Tiao, G. C. and Box, G. E. (1981). Modeling multiple time series with applications. *journal of the American Statistical Association*, 76(376):802–816.

Wang, D., Liu, X., and Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics*, 208(1):231 – 248. Special Issue on Financial Engineering and Risk Management.

Xiao, H. and Wu, W. B. (2013). Asymptotic theory for maximum deviations of sample covariance matrix estimates. *Stochastic Processes and their Applications*, 123(7):2899–2920.

Xiao, H. and Wu, W. B. (2014). Portmanteau test and simultaneous inference for serial covariances. *Statistica Sinica*, pages 577–599.