THE UNIVERSITY OF CHICAGO

HIGH DIMENSIONAL INFERENCE BASED ON QUADRATIC FORMS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS

BY
ZHIPENG LOU

CHICAGO, ILLINOIS
JUNE 2020

To my parents, Jianfeng Lou and Huiling Chen

TABLE OF CONTENTS

LIST OF FIGURES . . . vi

LIST OF TABLES . . . vii

ACKNOWLEDGMENTS . . . viii

ABSTRACT . . . ix

1 INTRODUCTION . . . 1

2 HALF SAMPLING: NEW IMPLEMENTATION FOR TWO-SAMPLE TEST FOR HIGH DIMENSIONAL MEANS . . . 3
- 2.1 Introduction . . . 3
- 2.2 Theoretical results . . . 6
- 2.3 Half sampling procedure . . . 10
- 2.4 Simulation Study . . . 13
- 2.5 Proofs . . . 17

3 HIGH DIMENSIONAL ANALYSIS OF VARIANCE IN MULTIVARIATE LINEAR REGRESSION . . . 25
- 3.1 Introduction . . . 25
- 3.2 Theoretical results . . . 28
  - 3.2.1 Asymptotic distribution of the conventional MANOVA test statistics 29
  - 3.2.2 Modified U type test statistics . . . 33
- 3.3 Applications . . . 39
  - 3.3.1 High dimensional one-way MANOVA . . . 39
  - 3.3.2 High dimensional nonparametric one-way MANOVA . . . 41
- 3.4 Practical implementation . . . 43
- 3.5 A simulation study . . . 48
  - 3.5.1 Size accuracy . . . 49
  - 3.5.2 Power comparison . . . 50
- 3.6 Proofs . . . 51
  - 3.6.1 Proofs of Lemmas 3.2.4 and 3.2.7 . . . 54
  - 3.6.2 Proofs of Theorems 3.2.3 and 3.2.6 . . . 56
  - 3.6.3 Proofs of Theorem 3.2.8 and Lemma 3.4.1 . . . 60

4 PORTMANTEAU TEST FOR HIGH DIMENSIONAL WHITE NOISES . . . 63
- 4.1 Introduction . . . 63
- 4.2 A New Portmanteau Test . . . 65
  - 4.2.1 Asymptotic distribution under the null hypothesis . . . 66
  - 4.2.2 Asymptotic distribution under the alternative hypothesis . . . 70
  - 4.2.3 Practical implementation . . . 71
- 4.3 Simulation Study . . . 72

4.3.1 Empirical sizes 72
4.3.2 Empirical power 73
4.4 Proofs 75

REFERENCES 85

LIST OF FIGURES

3.1 Empirical averages of the values of $|\hat{f}/f - 1|$ 47
3.2 Density plots of $Q_n^{\diamond}/\sqrt{\mathrm{var}(Q_n^{\diamond})}$ and $N(0,1)$ 51

4.1 Empirical powers of $L_{n,\nu}$ and $G_{n,\nu}$ for vector moving average model 75
4.2 Empirical powers of $L_{n,\nu}$ and $G_{n,\nu}$ for vector autoregressive model 76

LIST OF TABLES

2.1 Empirical sizes and powers of the four considered tests with significance levels $\alpha = 0.01$ and 0.05 in Example 2.4.1 15
2.2 Empirical sizes and powers of the four considered tests with significance levels $\alpha = 0.01$ and 0.05 in Example 2.4.2 16

3.1 Empirical sizes for Example 3.4.1 with $\alpha = 0.05$ 49
3.2 Empirical sizes for Example 3.5.1 with $\alpha = 0.05$ 50
3.3 Empirical powers for Example 3.4.1 with $\alpha = 0.05$ 52
3.4 Empirical powers for Example 3.5.1 with $\alpha = 0.05$ 52

4.1 Empirical sizes for tests based $L_{n,\nu}$ and $G_{n,\nu}$ 73

ACKNOWLEDGMENTS

At this moment of accomplishment I would like to express my sincere gratitude to my research advisors, Professor Wei Biao Wu and Professor Ruey S Tsay, who accepted me as their PhD student and offered me their mentorship and love. This work would not have been possible without their supervision and involvement, support and encouragement. They helped me find a solution when what I could see was darkness and self-doubt. Under their guidance, I successfully overcame many difficulties and learnt a lot.

My earnest thanks to Professor Stephen Stigler. I am grateful for his valuable advice, constructive criticism, positive appreciation and counsel throughout the entire process which led to the successful completion of the research work. I would like to thank the faculty and staff in the department for their kind help during my graduate study. I would also like to thank my fellow students who I have met in Chicago, especially Yuefeng Han, Likai Chen, Mengyu Xu, Wanrong Zhu, Bumeng Zhuo, Pinhan Chen and Yuhan Liu.

I wish to thank my wife, Danna Zhang, who has been "great Professor Zhang". She has been extremely supportive of me and has made countless sacrifices to help me get to this point. I am keeping up the pace with her and I believe there is someday we can climb the highest mountain and stand together, all the scenery in sight and only she in my eyes.

Most of all, I want to thank my parents, Jianfeng Lou and Huiling Chen, for all their endless love, patience, support and encouragement throughout my life. I am greatly indebted to them. Without them I would never have enjoyed so many opportunities to do what I love for a career.

ABSTRACT

There is a well-developed statistical inference theory for classical one-dimensional models. However, many important inference problems are still unanswered for high dimensional models where the dimension or the number of involved parameters can be much larger than the sample size. In the thesis, we solve three problems on hypothesis testing for high dimensional data based on quadratic form test statistics.

In the first work, we investigate high dimensional two sample mean test for independent observations. The theoretical contribution is the new distributional theory of quadratic forms of mean vectors by Gaussian approximation. The primary methodological contribution is a new half sampling calibration procedure as a model-free tool for valid inference which are not susceptible to model choice or model misspecification.

In the second work, we test general linear hypotheses of the multivariate linear regression model with independent observations. We study asymptotic behaviors of the conventional MANOVA test statistic and find its asymptotic distribution is dichotomous in high dimensions, which creates much difficulty to use in practice. We solved this open problem by proposing a new U type test statistic and laying a theoretical foundation for high dimensional MANOVA.

In the third work, we study the portmanteau test for high dimensional white noises. It is a popular choice to test for temporal dependence in low dimensional processes. The methodology and theory in the high dimensional case is less investigated. We propose a test statistic that is workable for high dimensional processes. A key technical component of our analysis is a new Gaussian approximation result for quadratic forms of high dimensional martingale differences, which is of independent interest in the statistical inference of high dimensional time series.

CHAPTER 1

INTRODUCTION

During the past several decades, there has been great interest in hypothesis testing for high dimensional data. Various classical methods have been generalized and applied in the high dimensional setting. However, most of the generalization do require some structural assumptions on either the underlying observations or the covariance structure. In real applications, these assumptions are typically nontrivial to be verified. In this thesis, we consider three fundamental inference problems and propose new test procedures which greatly relax the assumptions.

In Chapter 2, we consider two sample mean test in the high dimensional setting based on a simple L^2 type test statistic. The two populations can have unequal covariances. Under mild moment conditions, we develop a new invariance principle which provides distributional approximation of the test statistics by its Gaussian analogue. To approximate the asymptotic distribution of the test statistic, we propose a half sampling procedure, which avoids estimation of the unknown underlying covariance matrices. The simulation results show that the proposed test outperforms the exist methods.

In Chapter 3, we develop a systematic theory for high dimensional analysis of variance in multivariate linear regression, where the dimension and the number of unknown coefficients can both grow with the sample size. We propose a new U type test statistic to test linear hypotheses and establish a high dimensional Gaussian approximation result under fairly mild moment assumptions. Our general framework and theory can be applied to deal with the classical one-way multivariate ANOVA and the nonparametric one-way MANOVA in high dimensions. To implement the test procedure in practice, we introduce a sample-splitting based estimator of the second moment of the error covariance and establish its consistency.

In Chapter 4, we propose a new portmanteau test for detecting serial correlations of high dimensional time series based on a modified Box-Pierce test statistic. Under mild moment conditions on the underlying random vectors, we establish an invariance principle of the test statistic in the high dimensional regime where the dimension can be much larger than the sample size. To estimate the cutoff values, we propose a permutation procedure which avoids estimating the cross-sectional dependencies.

We now introduce some notation. Let $\mathbb{I}\{E\}$ denote the indicator function of an event E. For two random variables X and Y, the Kolmogorov distance is defined by

$$\rho(X,Y) = \sup_{z\in\mathbb{R}} |\mathbb{P}(X \le z) - \mathbb{P}(Y \le z)|.$$

For any $q > 0$, we write $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}$ if $\mathbb{E}|X|^q < \infty$. For two matrices A and B, $A \circ B$ denotes their Hadamard product. Let $\mathrm{tr}(A) = \sum_{i=1}^n A_{ii}$ and $|A|_{\mathbb{F}} = \sqrt{\sum_{i,j=1}^n |A_{ij}|^2}$ denote its Frobenius norm. For any positive integer m, we use I_k to denote $k \times k$ identity matrix. For two sequences of positive numbers $(a_n)_{n\ge1}$ and $(b_n)_{n\ge1}$, we write $a_n \lesssim b_n$ if there exists some constant C such that $a_n \le Cb_n$ for all large n. We use $C, C_1, C_2, \ldots$ to denote positive constants whose value may vary at different places. Let $\mathbf{1}_n = (1, 1, \ldots, 1)^\top$ denote n-dimensional vector consists of 1's. For any random variable X, we write $\mathbb{E}_0(X) = X - \mathbb{E}(X)$. For two constants $a, b \in \mathbb{R}$, denote $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For any positive integer n, denote $[n] = \{1, 2, \ldots, n\}$. For any set $\mathcal{A}$, denote by $|\mathcal{A}|$ its cardinality.

CHAPTER 2

HALF SAMPLING: NEW IMPLEMENTATION FOR TWO-SAMPLE TEST FOR HIGH DIMENSIONAL MEANS

2.1 Introduction

Let $X_1, X_2, \ldots, X_n \in \mathbb{R}^p$ and $Y_1, Y_2, \ldots, Y_m \in \mathbb{R}^p$ be independent and identically distributed (i.i.d.) random vectors from two populations with mean vectors μ_X and μ_Y and covariance matrices Σ_X and Σ_Y, respectively. Assume that $(X_k)_{k\in\mathbb{N}}$ and $(Y_k)_{k\in\mathbb{N}}$ are independent. In this chapter, we consider testing equality of the two population mean vectors μ_X and μ_Y, that is, testing the hypotheses

$$H_0 : \mu_X = \mu_Y \ \text{ versus } \ H_1 : \mu_X \neq \mu_Y. \tag{2.1}$$

The above two sample testing problem is of fundamental importance in statistical inference. It arises in many scientific applications, including genetics, econometrics and signal processing. In conventional multivariate analysis where the dimension p is finite and the two populations have the same covariance matrices, one can employ the classical Hotelling's T^2 test of which the properties have been well studied in the literature. Specifically, let $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ and $\bar{Y}_m = m^{-1} \sum_{j=1}^m Y_j$ be the sample mean vectors. Then the test statistic is defined as

$$T_H^2 = \frac{nm}{n+m} (\bar{X}_n - \bar{Y}_m)^\top \hat{\Sigma}_n^{-1} (\bar{X}_n - \bar{Y}_m),$$

where $\hat{\Sigma}_n$ is the pooled sample covariance matrix. It is well known that Hotelling's T^2 test enjoys desirable properties, see, for example, Anderson [2003].

However, when the dimension p is larger than the sample size $n + m$, the sample covariance matrix $\hat{\Sigma}_n$ is not invertible and consequently the test statistic T_H^2 is not well defined.

Recently, there have been many proposals accommodating the Hotelling's T^2 test into the high dimensional setting. For instance, Bai and Saranadasa [1996] proposed to remove $\hat{\Sigma}_n^{-1}$ in T_H^2 and introduced a non-exact test based on the modified test statistic

$$T_{BS} = \frac{nm}{n+m}\|\bar{X}_n - \bar{Y}_m\|^2 - \mathrm{tr}(\hat{\Sigma}_n).$$

They derived the asymptotic normality of T_{BS} under a linear process model. To accommodate the possible unequal covariances, Chen and Qin [2010] proposed a U type test statistic by removing the cross-product terms $\sum_{i=1}^n X_i^\top X_i$ and $\sum_{j=1}^m Y_j^\top Y_j$ and derived the asymptotic normality under the similar linear model. Replacing $\hat{\Sigma}_n$ in T_H^2 by its diagonal part, Srivastava and Kubokawa [2013] introduced a scale-invariant test for normally distributed observations. In a recent work, Zhang et al. [2019] proposed to use the simple L^2 type test statistic (2.2), which is shown to be distributed as a mixture of independent chi-squared distributions asymptotically. To approximate the asymptotic distribution, Zhang et al. [2019] proposed to employ the Welch-Satterthwaite χ^2-approximation when $\Sigma_X = \Sigma_Y$. However, all the aforementioned work assumed that the underlying observations $(X_k)_{k\in\mathbb{N}}$ and $(Y_k)_{k\in\mathbb{N}}$ either follow the linear model proposed by Bai and Saranadasa [1996] or have a multivariate Gaussian distribution. In real applications, such structural assumptions are hard to be verified. The primary goal of this paper is to propose a test for (2.1) that works under very general assumptions without Gaussianity or linearity.

Observe that the null hypothesis $\mu_X = \mu_Y$ holds if and only if $\|\mu_X - \mu_Y\|^2 = 0$. Hence a simple and natural test statistic is

$$Q_n = \|\bar{X}_n - \bar{Y}_m\|^2. \tag{2.2}$$

We shall reject the null hypothesis H_0 whenever Q_n is larger than some critical value. The first goal of this chapter is to derive the asymptotic distribution of the test statistic Q_n. In the conventional setting where the dimension p is finite, by the multivariate central limit theorem, we have that as $n \wedge m \to \infty$,

$$\sqrt{\frac{mn}{m+n}}(\bar{X}_n - \bar{Y}_m) \Rightarrow N(0, \theta_0\Sigma_X + (1-\theta_0)\Sigma_Y), \tag{2.3}$$

where θ_0 is given in (2.4) below. Then the asymptotic distribution of Q_n can be easily derived via the continuous mapping theorem. However, when the dimension p is large such that $n = o(p^2)$, the above multivariate central limit theorem (2.3) can fail, see, for example, Portnoy [1986]. Obtaining the asymptotic distribution of Q_n then becomes highly nontrivial. To overcome this circumstance, we develop a new Gaussian approximation result which provides distributional approximation of Q_n by its Gaussian analogue. In particular, unlike the papers mentioned above, we do not assume the Gaussianity or any particular structure of the underlying random vectors $(X_k)_{k\in\mathbb{N}}$ and $(Y_k)_{k\in\mathbb{N}}$.

Under fairly moment conditions, the null distribution of Q_n is asymptotically a mixture of independent chi-squared random variables, weighted by the unknown eigenvalues of the covariance matrix $\mathrm{cov}(\bar{X}_n - \bar{Y}_m)$. The second goal of this chapter is then to approximate the null distribution of Q_n. To this end, we propose a new half-sampling procedure which avoids estimating the unknown covariance matrices. In the univariate case, Wu [1990] established the validity of resampling procedure for sample means of i.i.d. random variables where the number of observations retained is of the same order ot the total sample size. Although the half sampling procedure is a special case of subsampling, the approach in Wu [1990] which depends the classical central limit theorem is not applicable to establish the validity of the half sampling procedure in our setting as the asymptotic distribution of Q_n now is a mixture of independent chi-squared distribution which is non-normal. To show the validity of the half sampling procedure, we develop new techniques which involves a more involved Gaussian approximation and is of independent interest by itself.

2.2 Theoretical results

In this section, we develop a systematic theory for the asymptotic distribution of the test statistic Q_n. Throughout this chapter, we assume that there exists a positive constant $\theta_0 \in (0,1)$ such that

$$\frac{m}{m+n} \to \theta_0. \tag{2.4}$$

We shall first derive the asymptotic distribution of the test statistic Q_n under the null hypothesis H_0. Define

$$Q_{n,0} = \|\mathbb{E}_0(\bar{X}_n - \bar{Y}_m)\|^2.$$

Note that $Q_n = Q_{n,0}$ under the null hypothesis H_0. It suffices to study the distribution of $Q_{n,0}$.

As mentioned in the introduction, the multivariate central limit theorem of the random vector $\mathbb{E}_0(\bar{X}_n - \bar{Y}_m)$ can fail when the dimension is much larger than the sample size. In this section, we develop a new Gaussian approximation result which bounds the Kolmogorov distance between the distributions of $Q_{n,0}$ and its Gaussian analogue. Specifically, let $Z \in \mathbb{R}^p$ be a centered Gaussian random vector with the same covariance matrix as $\bar{X}_n - \bar{Y}_m$, that is,

$$\text{cov}(Z) = \text{cov}(\bar{X}_n - \bar{Y}_m) =: \Sigma_\theta.$$

Then the Gaussian analogue of $Q_{n,0}$ is given by $Z^\top Z$, which is distributed as a mixture of independent chi-squared distributions. Specifically, let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ be the eigenvalues of the covariance matrix Σ_θ. Then

$$Z^\top Z \overset{\mathcal{D}}{=} \sum_{j=1}^{p} \lambda_j \chi_{j,1},$$

where $(\chi_{j,1})_{j\in\mathbb{N}}$ are i.i.d. chi-squared random variables with one degree of freedom. For simplicity of notation, denote

$$D_\theta^2 = \frac{\mathrm{tr}(\Sigma_X)}{n^3} + \frac{\mathrm{tr}(\Sigma_Y)}{m^3}.$$

The following theorem establishes the asymptotic distribution of $Q_{n,0}$.

Theorem 2.2.1. *Let $F_\theta = \|\Sigma_\theta\|_{\mathbb{F}}$ and $2 < q \leq 3$. Define*

$$K_q = \frac{\mathbb{E}|X_1^\top X_2|^q}{n^{2q-2}} + \frac{\mathbb{E}|X_1^\top Y_1|^q}{(nm)^{q-1}} + \frac{\mathbb{E}|Y_1^\top Y_2|^q}{m^{2q-2}},$$
$$M_q = \frac{\mathbb{E}|X_1^\top \Sigma_X X_1|^{q/2}}{n^{3q/2-1}} + \frac{\mathbb{E}|X_1^\top \Sigma_Y X_1|^{q/2}}{n^{q-1}m^{q/2}} + \frac{\mathbb{E}|Y_1^\top \Sigma_X Y_1|^{q/2}}{n^{q/2}m^{q-1}} + \frac{\mathbb{E}|Y_1^\top \Sigma_Y Y_1|^{q/2}}{m^{3q/2-1}}.$$

(i) *Assume that $D_\theta/F_\theta \to 0$ and $(K_q \vee M_q)/F_\theta^q \to 0$, then*

$$\rho(Q_{n,0}, Z^\top Z) \leq C_q \left(\frac{K_q + M_q}{F_\theta^q}\right)^{1/(2q+1)} + 8(n\wedge m)^{-1/5} + 3(D_\theta/F_\theta)^{2/5} \to 0. \tag{2.5}$$

(ii) *Assume that $F_\theta/D_\theta \to 0$ and that for any $\nu > 0$,*

$$\frac{\mathbb{E}|\mathbb{E}_0(X_1^\top X_1)|^2\mathbb{I}\{|\mathbb{E}_0(X_1^\top X_1)| \geq n^2\nu D_\theta\}}{n^3 D_\theta^2} + \frac{\mathbb{E}|\mathbb{E}_0(Y_1^\top Y_1)|^2\mathbb{I}\{|\mathbb{E}_0(Y_1^\top Y_1)| \geq m^2\nu D_\theta\}}{m^3 D_\theta^2} \to 0, \tag{2.6}$$

then we have the central limit theorem

$$\frac{Q_{n,0} - \mathrm{tr}(\Sigma_\theta)}{D_\theta} \Rightarrow N(0,1).$$

Remark 1. In the first case where $D_\theta/F_\theta \to 0$, the sufficient condition for $\rho(Q_{n,0}, Z^\top Z) \to 0$ is $(K_q \vee M_q)/F_\theta^q \to 0$, which can be viewed as generalized Lyapunov type moment condition in the high dimensional setting. Moreover, the condition $(K_q \vee M_q)/F_\theta^q \to 0$ is fairly mild and easy to verify for some model. To illustrate this, we consider the commonly used linear process model proposed by Bai and Saranadasa [1996]. More specifically, let $(V_{k,l})_{l\in\mathbb{N},k=1,2}$ be i.i.d. ℓ-dimensional random vectors with $\mathbb{E}(V_{k,l}) = 0$ and $\operatorname{cov}(V_{k,l}) = I_\ell$. Assume that $(X_k)_{k\in\mathbb{N}}$ and $(Y_k)_{k\in\mathbb{N}}$ follow the linear model

$$X_k = \Gamma_1 V_{1,k} \text{ and } Y_k = \Gamma_2 V_{2,k}, \tag{2.7}$$

where Γ_1 and Γ_2 are $p \times \ell$ matrices. Moreover, assume that $\mathbb{E}|V_{k,il}|^4 = 3 + \Delta$ with $\Delta < \infty$ and $\mathbb{E}\prod_{k=1}^{\ell} V_{ik}^{\tau_k} = 0$ (resp. 1) when there is at least one $\tau_k = 1$ (resp. there are two τ_k's that are 2), whenever $\sum_{k=1}^{\ell}\tau_k = 4$, where $\tau_1, \tau_2, \dots, \tau_\ell$ are non-negative integers.

Proposition 2.2.2. *Under the linear model (2.7) and the corresponding assumptions,*

$$\frac{D_\theta^2}{F_\theta^2} \le (2+\Delta)(n \wedge m)^{-1} \text{ and } \frac{K_q \vee M_q}{F_\theta^q} \le (3+\Delta)^{q/2}(n \wedge m)^{-\delta/2}.$$

Proposition 2.2.2 indicates that the conditions of the first case in Theorem 2.2.1 is satisfied for the linear process model (2.7). Then, by Theorem 2.2.1, we establish the Gaussian approximation of $Q_{n,0}$ for model (2.7).

Corollary 2.2.3. *Under the linear model (2.7), then we have*

$$\rho(Q_{n,0}, Z^\top Z) \le C_q(3+\Delta)^{q/(4q+2)}(n \wedge m)^{-(q-2)/(4q+2)}.$$

Remark 2. In a recent work, Zhang et al. [2019] derived that $\rho(Q_{n,0}, Z^\top Z) \to 0$ under the model (2.7) with equal covariances, that is, $\Gamma_1 = \Gamma_2$. In view of the discussions above, Theorem 1 of Zhang et al. [2019] can be viewed as a special case of our Theorem 2.2.1.

Recall that $Z^\top Z$ is distributed as a mixture of independent chi-squared random variables. Applying the Lindeberg-Feller central limit theorem, we establish the asymptotic normality of $Q_{n,0}$.

Theorem 2.2.4. *Assume that $D_\theta/F_\theta \to 0$ and $(K_q \vee M_q)/F_\theta^q \to 0$. Then the central limit* *theorem* $\{Q_{n,0} - \mathrm{tr}(\Sigma_\theta)\}/F_\theta \Rightarrow N(0,2)$ *holds if and only if*

$$\frac{\lambda_1}{F_\theta} \to 0. \tag{2.8}$$

Remark 3. Condition (2.8) is a common assumption to ensure asymptotic normality of high dimensional quadratic forms, see, for example, Bai and Saranadasa [1996]. It is worth mentioning that if (2.8) is violated, the asymptotic distribution of $Q_{n,0}$ is non-normal. For instance, suppose that $\Sigma_\theta = (1-\varpi)I_p + \varpi \mathbf{1}_p \mathbf{1}_p^\top$ for some constant $\varpi \in (0,1)$, then Theorem 2.2.4 implies that

$$\rho(Q_{n,0}/\lambda_1, \chi_1^2) \to 0, \ \text{ where } \ \lambda_1 = (p-1)\varpi + 1.$$

Remark 4. As mentioned in the introduction, Chen and Qin [2010] proposed to use a U type test statistic, which is as follows

$$U_n = \frac{\sum_{i\neq j}\sum_{k\neq l}(X_i - Y_k)^\top (X_j - Y_l)}{n(n-1)m(m-1)}.$$

Following Theorems 2.2.1 and 2.2.4, we can easily derive the Gaussian approximation and the central limit theorem of U_n.

Corollary 2.2.5. *Assume that* $(K_q \vee M_q)/F_\theta^q \to 0$*, then*

$$\rho\{U_n, \mathbb{E}_0(Z^\top Z)\} \leq C_q \left(\frac{K_q + M_q}{F_\theta^q}\right)^{1/(2q+1)} + 8(n\wedge m)^{-1/5} \to 0.$$

Corollary 2.2.6. *Assume that* $(K_q \vee M_q)/F_\theta^q \to 0$*. Then the central limit theorem* $U_n/F_\theta \Rightarrow N(0,2)$ *holds if and only if condition (2.8) is satisfied.*

2.3 Half sampling procedure

Recall that the first part of Theorem 2.2.1 reveals that the asymptotic distribution of $Q_{n,0}$ is a linear combination of independent chi-squared random variables, weighted by the unknown eigenvalues $(\lambda_j)_{j=1}^p$ of the covariance matrix Σ_θ. In the high dimensional setting when the dimension can be much larger than the sample size, it is well known that consistently estimate the eigenvalues of Σ_θ is highly nontrivial, unless some structural assumptions are imposed. In this section, we propose a new half-sampling procedure to evaluate the critical value of the asymptotic distribution of $Q_{n,0}$. The proposed procedure is extremely convenient to implement in practice and avoids estimating the unknown covariance matrix Σ_θ. For ease of presentation, we assume in this section that both the sample sizes n and m are even. Define two sets

$$\mathcal{A} = \{A \subseteq [n] : |A| = n/2\} \text{ and } \mathcal{B} = \{B \subseteq [m] : |B| = m/2\}.$$

Based on the subsamples $(X_i)_{i\in A}$ and $(Y_j)_{j\in B}$, the sampling statistic is defined as

$$Q_{A,B} = \|\bar{X}_A - \bar{Y}_B - (\bar{X}_n - \bar{Y}_m)\|^2, \tag{2.9}$$

where

$$\bar{X}_A = \frac{1}{|A|}\sum_{i\in A} X_i \text{ and } \bar{Y}_B = \frac{1}{|B|}\sum_{j\in B} Y_j.$$

Then the sampling distribution is given by

$$\mathcal{I}(z) = \frac{1}{N}\sum_{A\in\mathcal{A},B\in\mathcal{B}} \mathbb{I}\{Q_{A,B} \le z\}, \text{ where } N = \binom{n}{n/2}\binom{m}{m/2}.$$

Theorem 2.3.1. *Assume that* $(K_q \vee M_q)/F_\theta^q \to 0$, *then*

$$\sup_{z\in\mathbb{R}} |\mathbb{P}(Q_{n,0} \le z) - \mathcal{I}(z)| \xrightarrow{\mathbb{P}} 0.$$

Theorem 2.3.1 reveals that the half sampling distribution $\mathcal{I}(z)$ consistently estimate the null distribution of Q_n regardless of whether the null hypothesis is true or not. For any significance level $\alpha \in (0,1)$, denote by ξ_α the $(1-\alpha)$th quantile of $\mathcal{I}(z)$, that is,

$$\xi_\alpha = \inf \left\{ z \in \mathbb{R} : \mathcal{I}(z) \ge 1 - \alpha \right\}.$$

Then our test is defined as

$$\Phi_\alpha = \mathbb{I}\{Q_n > \xi_\alpha\}.$$

We shall reject the null hypothesis whenever $\Phi_\alpha = 1$. Under the conditions of Theorem 2.3.1, we have $\mathbb{E}(\Phi_\alpha | H_0) \to \alpha$ for any $\alpha \in (0,1)$.

Let $\beta_n = \mathbb{E}(\Phi_\alpha | H_1)$ denote the power function of the proposed test Φ_α. In the following theorem, we establish the asymptotic expression of β_n under the local alternative hypothesis (2.10).

Theorem 2.3.2. *Let* ξ_α^* *be the* $(1-\alpha)$*th quantile of* $Z^\top Z$*. Assume that*

$$\frac{(\mu_X - \mu_Y)^\top D_\theta (\mu_X - \mu_Y)}{F_\theta^2} \to 0. \tag{2.10}$$

Then, under the conditions of Theorem 2.3.1,

$$\left| \beta_n - \mathbb{P}\{\mathbb{E}_0(Z^\top Z) \ge \xi_\alpha^* - \|\mu_X - \mu_Y\|^2\} \right| \to 0. \tag{2.11}$$

Moreover, we have $\beta_n \to 1$ *if*

$$\frac{\|\mu_X - \mu_Y\|^2}{F_\theta} \to \infty.$$

Remark 5. Condition (2.10) can be viewed as the high dimensional local alternative, which is commonly used in the literature, see, for example, Bai and Saranadasa [1996] and Chen and Qin [2010]. It is worth mentioning that if condition (2.8) is also satisfied, then

$$\mathbb{E}(\Phi_\alpha | H_1) \to \Phi\left(-z_\alpha + \frac{\|\mu_X - \mu_Y\|^2}{\sqrt{2}F_\theta}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard normal random variable and z_α is $(1-\alpha)$th quantile of the standard normal distribution. In this case, the asymptotic power of our test Φ_α is same as that of the test proposed by Chen and Qin [2010].

Remark 6. In practice, the half sampling distribution $\mathcal{I}(z)$ may be difficult to compute as the total number of combinations N can be quite large. Instead, one can employ some stochastic approximation of $\mathcal{I}(z)$, see, for example, Politis and Romano [1994]. More specifically, let $A_1, A_2, \ldots, A_H$ and $B_1, B_2, \ldots, B_H$ be independently sampled from $\mathcal{A}$ and $\mathcal{B}$, respectively.

$$\mathcal{I}_H(z) = \frac{1}{H}\sum_{k=1}^{H} \mathbb{I}\left\{Q_{A_k, B_k} \le z\right\}.$$

Let $\xi_{\alpha,H}$ be the $(1-\alpha)$th quantile of $\mathcal{I}_H(z)$, that is,

$$\xi_{\alpha,H} = \inf\left\{z \in \mathbb{R} : \mathcal{I}_H(z) \ge 1 - \alpha\right\}.$$

In practice, one can implement the following test

$$\Phi_{\alpha,H} = \mathbb{I}\{Q_n > \xi_{\alpha,H}\} \tag{2.12}$$

2.4 Simulation Study

In this section, we conduct Monte Carlo simulations to illustrate the finite sample performance of the proposed test $\Phi_{\alpha,H}$ (2.12) with the half sampling size fixed at $H = 10000$. For comparison, we also implement the tests proposed by Chen and Qin [2010], Zhang et al. [2019] and Srivastava and Kubokawa [2013] denoted as CQ, ZGZC and SKK, respectively.

Throughout this section we take $\mu_X = 0$ and $\mu_{Y,j} = \nu_j \times \beta_j$ for each $1 \leq j \leq p$, where $(\beta_j)_{j \in \mathbb{N}}$ are i.i.d. uniform random variables $\mathcal{U}(-\delta, \delta)$ for some constant $\delta > 0$ and $(\nu_j)_{j \in \mathbb{N}}$ are i.i.d. binomial random variables with $\mathbb{P}(\nu_1 = 1) = \mathbb{P}(\nu_1 = 0) = 0.5$. When $\delta = 0$, the null hypothesis $\mu_X = \mu_Y$ is satisfied. Then the random vectors $(X_i)_{i=1}^n$ and $(Y_j)_{j=1}^m$ are generated via

$$X_i = E_i \text{ and } Y_j = E_{j+n} + \mu_Y,$$

where $(E_k)_{k \in \mathbb{N}}$ are i.i.d. p-dimensional random vectors coming from the following two different models.

Example 2.4.1. Let $L \in \mathbb{R}^{p \times p}$ be a lower triangular matrix such that $LL^\top = \Sigma$, where $\Sigma = (\Sigma_{ij})_{i,j=1}^p$ has entries $\Sigma_{ij} = 2^{-|i-j|}$. Let $V_i = (V_{i1}, V_{i2}, \ldots, V_{ip})^\top$, where $(V_{ij})_{i,j \in \mathbb{N}}$ are i.i.d. random variables. We generate the E_i's from a scale mixture of two independent multivariate distributions as follows,

$$E_i = \varpi_i \times LV_i + 3(1 - \varpi_i) \times LV_i',$$

where V_i' is independent copy of V_i and $(\varpi_i)_{i \in \mathbb{N}}$ are i.i.d. binomial random variables with $\mathbb{P}(\varpi_1 = 1) = \mathbb{P}(\varpi_1 = 0) = 0.5$.

Example 2.4.2. Let z_1, z_2, z_3 be non-negative constants such that $1 - z_2 - z_3 > z_1 \geq z_2 \geq$

$z_3 \geq 0$. We generate the E_i's from the following factor model

$$E_i = (1 - z_1 - z_2 - z_3)^{1/2} V_i + z_1^{1/2} V_{i,p+1} \mathbf{1}_p + z_2^{1/2} V_{i,p+2} \mathbf{1}_{p,\diamond} + z_3^{1/2} V_{i,p+3} \mathbf{1}_{p,\natural},$$

where $\mathbf{1}_{p,\diamond} = (\mathbf{1}_{p/2}^\top, -\mathbf{1}_{p/2}^\top)^\top$ and $\mathbf{1}_{p,\natural} = (\mathbf{1}_{p/4}^\top, -\mathbf{1}_{p/4}^\top, \mathbf{1}_{p/4}^\top, -\mathbf{1}_{p/4}^\top)^\top$. Note that the covariance matrix of (E_i) is

$$\mathrm{cov}(E_1) = (1 - z_1 - z_2 - z_3) I_p + z_1 \mathbf{1}_p \mathbf{1}_p^\top + z_2 \mathbf{1}_{p,\diamond} \mathbf{1}_{p,\diamond}^\top + z_3 \mathbf{1}_{p,\natural} \mathbf{1}_{p,\natural}^\top$$

and its eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ are

$$\lambda_j = 1 - z_1 - z_2 - z_3 + z_1 p \times \mathbb{I}\{j = 1\} + z_2 p \times \mathbb{I}\{j = 2\} + z_3 p \times \mathbb{I}\{j = 3\}. \tag{2.13}$$

In this section, we set $(z_1, z_2, z_3) = (0.15, 0.10, 0.05)$.

In both of the two examples, we consider two different distributions of the V_{ij}'s: one is standardized t distribution with 3 degrees of freedom and the other one is standardized chi-squared distribution with 5 degrees of freedom. In our simulation, we consider two cases of dimension $p = 800, 1200$ and two cases of sample sizes $(n, m) = (60, 80), (90, 120)$. The nominal significance levels are set to be $\alpha = 0.01$ and 0.05. The empirical sizes and powers of the four considered tests are calculated from 10000 replications.

It can be seen from Table 4.1 that the empirical sizes of the proposed test $\Phi_{\alpha,H}$ are very close to the nominal levels in all the cases. Both the ZGZC test and the SKK test are extremely conservative. The empirical sizes and powers of the CQ test are quite close to that of the proposed test. This is consistent with the discussions in Remark 5.

It can be seen from Table 2.2 that the proposed test $\Phi_{\alpha,H}$ controls the type I error very well with in all the cases. The CQ test has inflated empirical sizes for both the nominal levels $\alpha = 0.01$ and 0.05. This is expected as the CQ test relies on the asymptotic normality of U_n, which is not valid for this model. We refer to Remark 3 for more discussions. The SKK

Table 2.1: Empirical sizes and powers of the four considered tests with significance levels $\alpha = 0.01$ and 0.05 in Example 2.4.1

$\alpha = 0.01$			t_3				χ^2_5			
(n, m)	p	δ	NEW	CQ	ZGZC	SKK	NEW	CQ	ZGZC	SKK
(60, 80)	800	0.0	0.0102	0.0133	0.0005	0.0000	0.0110	0.0146	0.0007	0.0001
	1200		0.0108	0.0135	0.0000	0.0000	0.0094	0.0112	0.0003	0.0000
(90, 120)	800		0.0099	0.0141	0.0012	0.0000	0.0106	0.0139	0.0025	0.0002
	1200		0.0097	0.0123	0.0004	0.0000	0.0102	0.0126	0.0007	0.0000
(60, 80)	800	0.1	0.0168	0.0205	0.0009	0.0000	0.0181	0.0220	0.0014	0.0001
	1200		0.0212	0.0238	0.0001	0.0000	0.0174	0.0214	0.0004	0.0000
(90, 120)	800		0.0207	0.0255	0.0018	0.0002	0.0198	0.0251	0.0044	0.0004
	1200		0.0243	0.0293	0.0011	0.0000	0.0228	0.0278	0.0012	0.0001
(60, 80)	800	0.2	0.0538	0.0628	0.0045	0.0008	0.0550	0.0666	0.0060	0.0003
	1200		0.0728	0.0849	0.0022	0.0000	0.0721	0.0820	0.0028	0.0000
(90, 120)	800		0.1006	0.1174	0.0180	0.0047	0.0936	0.1124	0.0238	0.0043
	1200		0.1479	0.1660	0.0158	0.0024	0.1421	0.1609	0.0236	0.0016
(60, 80)	800	0.3	0.2267	0.2545	0.0288	0.0071	0.2159	0.2459	0.0431	0.0048
	1200		0.3362	0.3639	0.0274	0.0017	0.3319	0.3565	0.0385	0.0007
(90, 120)	800		0.4906	0.5299	0.1829	0.1272	0.4682	0.5082	0.2126	0.0805
	1200		0.6812	0.7082	0.2348	0.1224	0.6727	0.7005	0.2922	0.0698
$\alpha = 0.05$										
(60, 80)	800	0.0	0.0515	0.0544	0.0094	0.0018	0.0541	0.0570	0.0149	0.0015
	1200		0.0528	0.0550	0.0047	0.0002	0.0524	0.0562	0.0065	0.0002
(90, 120)	800		0.0516	0.0568	0.0150	0.0036	0.0475	0.0516	0.0199	0.0043
	1200		0.0532	0.0560	0.0075	0.0012	0.0521	0.0552	0.0127	0.0013
(60, 80)	800	0.1	0.0702	0.0745	0.0151	0.0028	0.0748	0.0798	0.0218	0.0024
	1200		0.0775	0.0812	0.0083	0.0005	0.0786	0.0829	0.0110	0.0005
(90, 120)	800		0.0859	0.0909	0.0257	0.0089	0.0825	0.0874	0.0342	0.0091
	1200		0.0958	0.1003	0.0197	0.0030	0.0951	0.0981	0.0259	0.0035
(60, 80)	800	0.2	0.1724	0.1811	0.0459	0.0135	0.1752	0.1828	0.0610	0.0111
	1200		0.2230	0.2302	0.0383	0.0041	0.2168	0.2238	0.0474	0.0038
(90, 120)	800		0.2754	0.2856	0.1173	0.0683	0.2631	0.2732	0.1370	0.0528
	1200		0.3563	0.3658	0.1208	0.0506	0.3564	0.3662	0.1470	0.0396
(60, 80)	800	0.3	0.4730	0.4828	0.1896	0.1018	0.4615	0.4727	0.2222	0.0708
	1200		0.6025	0.6117	0.2030	0.0789	0.5917	0.6002	0.2419	0.0481
(90, 120)	800		0.7397	0.7519	0.4964	0.4791	0.7365	0.7482	0.5465	0.3617
	1200		0.8789	0.8841	0.5951	0.5432	0.8714	0.8769	0.6538	0.4097

Table 2.2: Empirical sizes and powers of the four considered tests with significance levels $\alpha = 0.01$ and 0.05 in Example 2.4.2

$\alpha = 1\%$			t_3				χ^2_5			
(n, m)	p	δ	NEW	CQ	ZGZC	SKK	NEW	CQ	ZGZC	SKK
(60, 80)	800	0.0	0.0112	0.0350	0.0217	0.0103	0.0111	0.0357	0.0227	0.0130
	1200		0.0113	0.0329	0.0223	0.0083	0.0124	0.0369	0.0240	0.0124
(90, 120)	800		0.0107	0.0348	0.0224	0.0113	0.0113	0.0388	0.0255	0.0139
	1200		0.0107	0.0316	0.0224	0.0092	0.0112	0.0334	0.0227	0.0112
(60, 80)	800	0.3	0.1393	0.3835	0.2627	0.1433	0.0867	0.2780	0.1874	0.1035
	1200		0.1336	0.3849	0.2649	0.1214	0.0895	0.2756	0.1873	0.0890
(90, 120)	800		0.3707	0.7649	0.6148	0.3915	0.2345	0.6727	0.4944	0.2780
	1200		0.3686	0.7734	0.6231	0.3414	0.2234	0.6752	0.4909	0.2259
(60, 80)	800	0.35	0.3155	0.6800	0.5216	0.3205	0.1900	0.5535	0.3974	0.2216
	1200		0.3195	0.6934	0.5360	0.2861	0.1849	0.5435	0.3917	0.1850
(90, 120)	800		0.7469	0.9518	0.8841	0.7092	0.6034	0.9774	0.9085	0.6665
	1200		0.7533	0.9525	0.8863	0.6667	0.6120	0.9811	0.9194	0.6008
(60, 80)	800	0.4	0.6062	0.8905	0.7856	0.5765	0.4249	0.8772	0.7375	0.4758
	1200		0.6207	0.8977	0.7963	0.5351	0.4137	0.8852	0.7468	0.4041
(90, 120)	800		0.9361	0.9874	0.9634	0.8750	0.9552	0.9999	0.9986	0.9602
	1200		0.9354	0.9871	0.9629	0.8522	0.9604	0.9999	0.9994	0.9373
$\alpha = 5\%$										
(60, 80)	800	0.0	0.0542	0.0739	0.0623	0.0294	0.0511	0.0697	0.0594	0.0349
	1200		0.0504	0.0712	0.0588	0.0259	0.0540	0.0721	0.0611	0.0326
(90, 120)	800		0.0510	0.0728	0.0606	0.0301	0.0563	0.0767	0.0658	0.0382
	1200		0.0468	0.0638	0.0544	0.0259	0.0476	0.0653	0.0552	0.0288
(60, 80)	800	0.3	0.5232	0.6482	0.5749	0.3543	0.4047	0.5345	0.4753	0.2796
	1200		0.5285	0.6636	0.5884	0.3198	0.3940	0.5289	0.4661	0.2397
(90, 120)	800		0.8693	0.9330	0.8955	0.7015	0.8454	0.9450	0.9082	0.6560
	1200		0.8769	0.9368	0.9034	0.6656	0.8549	0.9546	0.9235	0.5933
(60, 80)	800	0.35	0.8034	0.8841	0.8368	0.6164	0.7307	0.8636	0.8129	0.5447
	1200		0.8126	0.8942	0.8459	0.5819	0.7231	0.8661	0.8112	0.4783
(90, 120)	800		0.9758	0.9870	0.9760	0.8864	0.9970	0.9998	0.9995	0.9616
	1200		0.9738	0.9868	0.9750	0.8663	0.9981	1.0000	0.9998	0.9440
(60, 80)	800	0.4	0.9417	0.9676	0.9464	0.8111	0.9599	0.9908	0.9816	0.8464
	1200		0.9437	0.9667	0.9493	0.7891	0.9670	0.9927	0.9850	0.8048
(90, 120)	800		0.9933	0.9955	0.9915	0.9502	1.0000	1.0000	1.0000	0.9992
	1200		0.9929	0.9961	0.9920	0.9403	1.0000	1.0000	1.0000	0.9989

test is conservative at the nominal level $\alpha = 0.05$ and the ZGZC test leads to an inflated size at the nominal level $\alpha = 0.01$.

2.5 Proofs

In this section, we provide the technical proofs of the theoretical results in previous sections. We first introduce some notation and definition. For any random variable $X \in \mathbb{R}$ and positive constant $z > 0$, the Levy concentration function (cf. Rudelson and Vershynin [2009]) is given as

$$\mathcal{L}(X, z) = \sup_{t \in \mathbb{R}} \mathbb{P}(t \leq X \leq t + z).$$

Let $(V_k)_{k \in \mathbb{N}}$ be independent p-dimensional random vectors with $\mathbb{E}(V_k) = 0$ and $\mathrm{cov}(V_k) = \Sigma_k$. For any non-random vector $b = (b_1, b_2, \ldots, b_n)^\top$, define

$$F_b = \left\| \sum_{k=1}^{n} b_k^2 \Sigma_k \right\|_{\mathbb{F}} \quad \text{and} \quad Q_b = \sum_{k \neq l} b_k b_l V_k^\top V_l.$$

Let $Z_b \in \mathbb{R}^p$ be a Gaussian random vector such that $\mathbb{E}(Z_b) = 0$ and $\mathrm{cov}(Z_b) = \sum_{k=1}^{n} b_k^2 \Sigma_k$.

Lemma 2.5.1. *Let $(\chi_{k,1})_{k \in \mathbb{N}}$ be independent chi-squared random variables with one degree of freedom. Let $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_m \geq 0$ be non-negative constants such that $\sum_{k=1}^{m} \alpha_k^2 = 1$. Then, for any $z > 0$,*

$$\mathcal{L}(\alpha_1 \chi_{1,1} + \alpha_2 \chi_{2,1} + \ldots + \alpha_m \chi_{m,1}, z) \leq \sqrt{\frac{4z}{\pi}}.$$

Lemma 2.5.2. *Define*

$$\Delta_{q,b} = \sum_{k \neq l} |b_k b_l|^q \mathbb{E} |V_k^\top V_l|^q + \sum_{k=1}^{n} \mathbb{E} \left(\sum_{l \neq k} |b_k b_l|^2 V_k^\top \Sigma_l V_k \right)^{q/2}.$$

Then we have

$$\rho\{Q_b, \mathbb{E}_0(Z_b^\top Z_b)\} \le C_q(\Delta_{q,b}/F_b^q)^{1/(2q+1)} + 4\left(\frac{\sum_{k=1}^n b_k^4 \|\Sigma_k\|_{\mathbb{F}}^2}{\pi^2 F_b^2}\right)^{1/5}.$$

Lemma 2.5.3. *For $a, b \in \mathbb{R}^n$, we have*

$$\rho\{Q_b \vee Q_a, \mathbb{E}_0(Z_b^\top Z_b) \vee \mathbb{E}_0(Z_a^\top Z_a)\} \le 4\left(\frac{\sum_{k=1}^n b_k^4 \|\Sigma_k\|_{\mathbb{F}}^2}{\pi^2 F_b^2}\right)^{1/5} + 4\left(\frac{\sum_{k=1}^n a_k^4 \|\Sigma_k\|_{\mathbb{F}}^2}{\pi^2 F_a^2}\right)^{1/5}$$
$$+ C_q(\Delta_{q,b} + \Delta_{q,a})^{1/(2q+1)}(F_b^{-1/2} + F_a^{-1/2})^{2q/(2q+1)}.$$

Proof of Lemma 2.5.3. For $\phi > 0$ and $\nu \in \mathbb{R}$, define $g_{\phi,\nu}(x,y) = g_0[\phi\{f_\phi(x,y) - \nu\}]$, where

$$f_\phi(x,y) = \phi^{-1}\log\{\exp(\phi x) + \exp(\phi y)\} \text{ and } g_0(x) = [1 - \min\{1, \max(x,0)\}^4]^4.$$

Then for any ν, we have

$$\mathbb{I}\{\max(x,y) \le \nu - \phi^{-1}\log 2\} \le g_{\phi,\nu}(x,y) \le \mathbb{I}\{\max(x,y) \le \nu + \phi^{-1}\}. \tag{2.14}$$

Let $\{W_k\}_{k\in\mathbb{N}}$ be independent Gaussian random vectors with $\mathbb{E}(W_k) = 0$ and $\mathrm{cov}(W_k) = \Sigma_k$. Define the Gaussian analogues of Q_b and Q_a respectively with the V_k's replaced by the W_k's,

$$Q_b^\diamond = \sum_{k\ne l} b_k b_l W_k^\top W_l \text{ and } Q_a^\diamond = \sum_{k\ne l} a_k a_l W_k^\top W_l.$$

By Lemma 2.5.1, if follows that

$$\rho\{Q_b^\diamond, \mathbb{E}_0(Z_b^\top Z_b)\} \le 4\left(\frac{\sum_{k=1}^n b_k^4 \|\Sigma_k\|_{\mathbb{F}}^2}{\pi^2 F_b^2}\right)^{1/5}$$

and

$$\rho\{Q_a^\diamond, \mathbb{E}_0(Z_a^\top Z_a)\} \leq 4\left(\frac{\sum_{k=1}^n a_k^4 \|\Sigma_k\|_{\mathbb{F}}^2}{\pi^2 F_a^2}\right)^{1/5}.$$

Following arguments in Xu et al. [2014] and, it follows that

$$\sup_{\nu\in\mathbb{R}} |\mathbb{E}\{g_{\phi,\nu}(Q_b, Q_a)\} - \mathbb{E}\{g_{\phi,\nu}(Q_b^\diamond, Q_a^\diamond)\}| \leq C_q(\Delta_{q,b} + \Delta_{q,a})\phi^q.$$

For simplicity, write $\nu_\phi = \nu - \phi^{-1}\log 2$ and $\phi_\star = (1+\log 2)/\phi$. By (2.14), it follows that

$$\begin{aligned}
\mathbb{P}(Q_b \vee Q_a \leq \nu_\phi) &\leq \mathbb{E}\{g_{\phi,\nu}(Q_b, Q_a)\} \leq \mathbb{E}\{g_{\phi,\nu}(Q_b^\diamond, Q_a^\diamond)\} + C_q(\Delta_{q,b} + \Delta_{q,a})\phi^q \\
&\leq \mathbb{P}(Q_b^\diamond \vee Q_a^\diamond \leq \nu + \phi^{-1}) + C_q(\Delta_{q,b} + \Delta_{q,a})\phi^q \\
&\leq \mathbb{P}\{\mathbb{E}_0(Z_b^\top Z_b) \vee \mathbb{E}_0(Z_a^\top Z_a) \leq \nu_\phi\} + C_q(\Delta_{q,b} + \Delta_{q,a})\phi^q \\
&+ \rho\{Q_b^\diamond, \mathbb{E}_0(Z_b^\top Z_b)\} + \rho\{Q_a^\diamond, \mathbb{E}_0(Z_a^\top Z_a)\} + \sqrt{\frac{4\phi_\star}{\pi F_b}} + \sqrt{\frac{4\phi_\star}{\pi F_a}}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}(Q_b \vee Q_a \leq \nu_\phi) &\geq \mathbb{P}\{\mathbb{E}_0(Z_b^\top Z_b) \vee \mathbb{E}_0(Z_a^\top Z_a) \leq \nu_\phi\} - C_q(\Delta_{q,b} + \Delta_{q,a})\phi^q \\
&- \rho\{Q_b^\diamond, \mathbb{E}_0(Z_b^\top Z_b)\} - \rho\{Q_a^\diamond, \mathbb{E}_0(Z_a^\top Z_a)\} - \sqrt{\frac{4\phi_\star}{\pi F_b}} - \sqrt{\frac{4\phi_\star}{\pi F_a}}.
\end{aligned}$$

□

Proof of Proposition 2.2.2. Elementary calculations imply that

$$\begin{aligned}
\mathbb{E}|X_1^\top X_2|^4 &= 3\|\Sigma_X\|_{\mathbb{F}}^4 + 6\mathrm{tr}(\Sigma_X^4) + \Delta^2 \sum_{1\leq i,j\leq \ell} (\Gamma_1^\top \Gamma_1)_{ij}^4 + 6\Delta \sum_{1\leq i,j,k\leq \ell} (\Gamma_1^\top\Gamma_1)_{ij}^2 (\Gamma_1^\top \Gamma_1)_{ik}^2 \\
&\leq (9 + \Delta^2 + 6\Delta)\|\Sigma_X\|_{\mathbb{F}}^4 = (\Delta+3)^2\|\Sigma_X\|_{\mathbb{F}}^4
\end{aligned}$$

and

$$\mathbb{E}|X_1^\top X_1 - \mathrm{tr}(\Sigma_X)|^2 = 2\|\Sigma_X\|_{\mathbb{F}}^2 + \Delta \sum_{1\le k\le \ell} (\Gamma_1^\top \Gamma_1)_{kk}^2 \le (2+\Delta)\|\Sigma_X\|_{\mathbb{F}}^2.$$

□

Proof of Theorem 2.2.1. Decompose $Q_n = D_0 + Q_n - D_0 =: D_0 + Q_n^\star$, where

$$D_0 = \frac{1}{n^2}\sum_{i=1}^n X_i^\top X_i + \frac{1}{m^2}\sum_{j=1}^m Y_j^\top Y_j.$$

Case 1*:* $D_\theta/F_\theta \to 0$.

We shall apply Lemma 2.5.2. Let $b = (n^{-1}\mathbf{1}_n^\top, m^{-1}\mathbf{1}_m^\top)^\top$ and

$$V_k = X_k \times \mathbb{I}\{1 \le k \le n\} + Y_k \times \mathbb{I}\{n+1 \le k \le n+m\}.$$

Elementary calculations imply that $\Delta_{q,b} \le C_q(K_q + M_q)$. Then, by Lemma 2.5.2,

$$\rho\{Q_n^\star, \mathbb{E}_0(Z^\top Z)\} \le C_q \left(\frac{K_q + M_q}{F_\theta^q}\right)^{1/(2q+1)} + 4(n\wedge m)^{-1/5}\pi^{-2/5}. \tag{2.15}$$

Consequently, as $\mathrm{var}(D_0) = D_\theta^2$, it follows that

$$\rho\{\mathbb{E}_0(Q_n), Q_n^\star\} \le C_q \left(\frac{K_q + M_q}{F_\theta^q}\right)^{1/(2q+1)} + 8(n\wedge m)^{-1/5}\pi^{-2/5} + 3(D_\theta/F_\theta)^{2/5}.$$

Then (2.5) follows in view of the triangle inequality

$$\rho(Q_n, Z^\top Z) \le \rho\{\mathbb{E}_0(Q_n), Q_n^\star\} + \rho\{Q_n^\star, \mathbb{E}_0(Z^\top Z)\}.$$

Case 2*:* $F_\theta/D_\theta \to 0$.

As $\mathbb{E}|Q_n^\star|^2/D_\theta^2 \leq 2F_\theta^2/D_\theta^2 \to 0$, we have $Q_n^\star/D_\theta \xrightarrow{\mathbb{P}} 0$. By (2.6), we have

$$\frac{D_0 - \mathrm{tr}(\Sigma_\theta)}{D_\theta} \Rightarrow N(0,1).$$

Consequently, it follows that

$$\frac{Q_n - \mathrm{tr}(\Sigma_\theta)}{D_\theta} \Rightarrow N(0,1).$$

□

Proof of Theorem 2.3.1. For any $A \in \mathcal{A}$ and $B \in \mathcal{B}$, define $Q_{A,B}^\star = Q_{A,B} - D_0$ and

$$\mathcal{I}_\star(z) = \frac{1}{N} \sum_{A \in \mathcal{A}, B \in \mathcal{B}} \mathbb{I}\left\{Q_{A,B}^\star \leq z\right\}.$$

Note that it is equivalent to show that

$$\sup_{z \in \mathbb{R}} |\mathbb{P}(Q_n^\star \leq z) - \mathcal{I}_\star(z)| \xrightarrow{\mathbb{P}} 0.$$

By (2.15), we have $\rho\{Q_n^\star, \mathbb{E}_0(Z^\top Z)\} \to 0$. Hence it suffices to show that

$$\sup_{z \in \mathbb{R}} |\mathbb{P}\{\mathbb{E}_0(Z^\top Z) \leq z\} - \mathcal{I}_\star(z)| \xrightarrow{\mathbb{P}} 0.$$

By Lemma 2.5.1, it suffices to show that

$$\sup_{z \in \mathbb{R}} \mathbb{E}|\mathbb{P}\{\mathbb{E}_0(Z^\top Z) \leq z\} - \mathcal{I}_\star(z)|^2 \to 0. \tag{2.16}$$

By similar arguments as in the proof of Theorem 2.2.1, it follows that for any $A \in \mathcal{A}$ and

$B \in \mathcal{B}$,

$$\rho\{Q^{\star}_{A,B}, \mathbb{E}_0(Z^\top Z)\} \le C_q \left(\frac{K_q + M_q}{F_\theta^q} \right)^{1/(2q+1)} + 4(n \wedge m)^{-1/5} \pi^{-2/5} \to 0.$$

Consequently, we have

$$\sup_{z \in \mathbb{R}} |\mathbb{P}\{\mathbb{E}_0(Z^\top Z) \le z\} - \mathbb{E}\{\mathcal{I}_\star(z)\}| \to 0. \tag{2.17}$$

For any subsets $A_1, A_2 \in \mathcal{A}$ and $B_1, B_2 \in \mathcal{B}$, define

$$\mathcal{Q}_1 = \|\bar{X}_{A_1} - \bar{Y}_{B_1} - (\bar{X}_n - \bar{Y}_m)\|^2 - D_0 \text{ and } \mathcal{Q}_2 = \|\bar{X}_{A_2} - \bar{Y}_{B_2} - (\bar{X}_n - \bar{Y}_m)\|^2 - D_0.$$

Let $(Z_k)_{k \in \mathbb{N}}$ and $(G_k)_{k \in \mathbb{N}}$ be independent Gaussian random vectors $N(0, \Sigma_X)$ and $N(0, \Sigma_Y)$, respectively. Define $\mathcal{V}_1 = \mathcal{Z}_1^\top \mathcal{Z}_1 - \operatorname{tr}(\Sigma_\theta)$ and $\mathcal{V}_2 = \mathcal{Z}_2^\top \mathcal{Z}_2 - \operatorname{tr}(\Sigma_\theta)$, where

$$\mathcal{Z}_1 = \bar{Z}_{A_1} - \bar{G}_{B_1} - (\bar{Z}_n - \bar{G}_m) \text{ and } \mathcal{Z}_2 = \bar{Z}_{A_2} - \bar{G}_{B_2} - (\bar{Z}_n - \bar{G}_m).$$

Then $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are jointly Gaussian with

$$\operatorname{cov}(\mathcal{Z}_1, \mathcal{Z}_2) = \left(\frac{4\mathcal{C}_A - n}{n^2} \right) \Sigma_X + \left(\frac{4\mathcal{C}_B - m}{m^2} \right) \Sigma_Y,$$

where $\mathcal{C}_A = |A_1 \cap A_2|$ and $\mathcal{C}_B = |B_1 \cap B_2|$. By Lemma 2.5.3 and $(K_q \vee M_q)/F_\theta^q \to 0$,

$$\rho(\mathcal{Q}_1 \vee \mathcal{Q}_2, \mathcal{V}_1 \vee \mathcal{V}_2) \le C_q \left(\frac{K_q + M_q}{F_\theta^q} \right)^{1/(2q+1)} + 8(n \wedge m)^{-1/5} \pi^{-2/5} \to 0.$$

Define $\mathcal{V}_\diamond = \mathcal{Z}_\diamond^\top \mathcal{Z}_\diamond - \mathbb{E}(\mathcal{Z}_\diamond^\top \mathcal{Z}_\diamond)$, where

$$\mathcal{Z}_\diamond = \mathcal{Z}_1 - \left(\frac{4\mathcal{C}_A - n}{n} \right) (\bar{Z}_{A_2} - \bar{Z}_n) + \left(\frac{4\mathcal{C}_B - m}{m} \right) (\bar{G}_{B_2} - \bar{G}_m).$$

Elementary calculations imply that $\text{cov}(\mathcal{Z}_\diamond, \mathcal{Z}_2) = 0$ and

$$\mathbb{E}|\mathcal{V}_1 - \mathcal{V}_\diamond|^2 = 4\text{tr}(\Sigma_\theta \Sigma_{\mathcal{C}}) - 2\|\Sigma_{\mathcal{C}}\|_{\mathbb{F}}^2,$$

where

$$\Sigma_{\mathcal{C}} = \frac{(4\mathcal{C}_A - n)^2}{n^3}\Sigma_X + \frac{(4\mathcal{C}_B - m)^2}{m^3}\Sigma_Y.$$

Consequently, by Lemma 2.5.1 and the Markov inequality, it follows that

$$\rho(\mathcal{V}_1, \mathcal{V}_\diamond) \leq 3\left\{\frac{\text{tr}(\Sigma_\theta \Sigma_{\mathcal{C}})}{F_\theta^2}\right\}^{1/5}.$$

Therefore, for any $x \in \mathbb{R}$, we have

$$\begin{aligned}\mathbb{P}(\mathcal{V}_1 \vee \mathcal{V}_2 \leq z) &\leq \mathbb{P}(\mathcal{V}_\diamond \vee \mathcal{V}_2 \leq z) + \rho(\mathcal{V}_1, \mathcal{V}_\diamond) = \mathbb{P}(\mathcal{V}_\diamond \leq z)\mathbb{P}(\mathcal{V}_2 \leq z) + \rho(\mathcal{V}_1, \mathcal{V}_\diamond) \\ &\leq \mathbb{P}(\mathcal{V}_1 \leq z)\mathbb{P}(\mathcal{V}_2 \leq z) + 2\rho(\mathcal{V}_1, \mathcal{V}_\diamond)\end{aligned}$$

and

$$\mathbb{P}(\mathcal{V}_1 \vee \mathcal{V}_2 \leq z) \geq \mathbb{P}(\mathcal{V}_1 \leq z)\mathbb{P}(\mathcal{V}_2 \leq z) - 2\rho(\mathcal{V}_1, \mathcal{V}_\diamond).$$

Consequently, it follows that

$$\sup_{z \in \mathbb{R}} |\mathbb{P}(\mathcal{V}_1 \vee \mathcal{V}_2 \leq z) - \mathbb{P}(\mathcal{V}_1 \leq z)\mathbb{P}(\mathcal{V}_2 \leq z)| \leq 2\rho(\mathcal{V}_1, \mathcal{V}_\diamond).$$

By the Markov inequality, it follows that for any $\delta \in (0, 1)$,

$$\mathbb{P}(|4\mathcal{C}_A - n| > n^\delta) \leq \frac{n^{2-2\delta}}{n-1} \text{ and } \mathbb{P}(|4\mathcal{C}_B - m| > m^\delta) \leq \frac{m^{2-2\delta}}{m-1}.$$

Define an event $\mathcal{E}_\delta = \{|4\mathcal{C}_A - n| \le n^\delta, |4\mathcal{C}_B - m| \le m^\delta\}$. Then

$$\mathbb{P}(\mathcal{E}_\delta) \ge 1 - 2n^{1-2\delta} - 2m^{1-2\delta}.$$

Under event $\mathcal{E}_\delta$, we have $\mathrm{tr}(\Sigma_{\mathcal{C}}\Sigma_\theta) \le (n \wedge m)^{2(\delta-1)} F_\theta^2$ and consequently

$$\rho(\mathcal{V}_1, \mathcal{V}_\diamond) \le 3(n \wedge m)^{2(\delta-1)/5}.$$

Taking $\delta > 1/2$, it follows that $\rho(\mathcal{V}_1, \mathcal{V}_\diamond) \to 0$ and consequently,

$$\sup_{z \in \mathbb{R}} \mathrm{var}\{\mathcal{I}_\star(z)\} \le 2\mathbb{P}(\mathcal{E}_\delta)\rho(\mathcal{V}_1, \mathcal{V}_\diamond) + 1 - \mathbb{P}(\mathcal{E}_\delta) \to 0. \tag{2.18}$$

Then (2.16) follows from (2.17) and (2.18). □

Proof of Theorem 2.3.2. Decompose

$$Q_n = Q_{n,0} + \|\mu_X - \mu_Y\|^2 + 2(\mu_X - \mu_Y)^\top \mathbb{E}_0(\bar{X}_n - \bar{Y}_m).$$

By (2.10),

$$\frac{(\mu_X - \mu_Y)^\top \mathbb{E}_0(\bar{X}_n - \bar{Y}_m)}{F_\theta} \xrightarrow{\mathbb{P}} 0.$$

Hence, under the conditions of Theorem 2.2.1, it follows that

$$\rho(Q_n - \|\mu_X - \mu_Y\|^2, Z^\top Z) \to 0.$$

□

CHAPTER 3

HIGH DIMENSIONAL ANALYSIS OF VARIANCE IN MULTIVARIATE LINEAR REGRESSION

3.1 Introduction

In statistical inference of multivariate linear regression, a fundamental problem is to investigate the relationships between the covariates and the responses. In this chapter, we aim to test whether a given set of covariates are associated with the responses by multivariate analysis of variance (MANOVA). To fix the idea, we build the multivariate linear regression model with p predictors as

$$Y_i = B^\top X_i + V_i, \ \ i = 1, 2, \ldots, n, \tag{3.1}$$

where $Y_i = (Y_{i1}, \ldots, Y_{id})^\top$ and $X_i = (X_{i1}, \ldots, X_{ip})^\top$ are respectively the response vector and predictor vector for the ith sample, $B^\top = (B_1, \ldots, B_p)$ is the unknown coefficient matrix with $B_k \in \mathbb{R}^d$, $1 \le k \le p$, consisting of coefficients on the kth covariate, and the innovation vectors $V_i \in \mathbb{R}^d$, $1 \le i \le n$, are i.i.d. with $\mathbb{E}(V_i) = 0$ and $\text{cov}(V_i) = \Sigma$. The first element of X_i can be set to be 1 to reflect an intercept term. Equivalently we can write (3.1) in compact matrix form as

$$Y = XB + V \tag{3.2}$$

for $Y = (Y_1, \ldots, Y_n)^\top$, $X = (X_1, \ldots, X_n)^\top$ and $V = (V_1, \ldots, V_n)^\top$. Let $\mathcal{C} \in \mathbb{R}^{m \times p}$ be a matrix of rank m, where $1 \le m \le p$. We are interested in testing a collection of linear constraints on the coefficient matrix

$$H_0 : \mathcal{C}B = 0 \ \text{ versus } \ H_1 : \mathcal{C}B \ne 0. \tag{3.3}$$

This testing problem has been extensively studied in the low dimensional setting where both d and p are fixed. A natural and popular choice is the likelihood ratio test when the errors are normally distributed; see Chapter 8 in Anderson [2003] for a review of theoretical investigations. In recent years, high dimensional data are increasingly encountered in various applications. Over the past decade, there have been tremendous efforts to develop new methodologies and theories for high dimensional regression. The paradigm where d is 1 or small and p can increase with n has received considerable attention, while the one where d is very large and p is relatively small has been less studied. The model (3.2) in the latter setting has been applied to a number of research problems involving high-dimensional data types such as DNA sequence data, gene expression microarray data, and imaging data; see for example Zapala and Schork [2006], Wessel and Schork [2006] and Zapala and Schork [2012]. Those related studies typically generate huge amounts of data (responses) that, due to their expense and sophistication, are often collected on a relatively small number of individuals, and investigate how the data can be explained by a certain number of predictor variables such as the ages of individuals assayed, clinical diagnoses, strain memberships, cell line types, or genotype information (Zapala and Schork [2006]). Owing to inappropriateness of applying the standard MANOVA strategy and shortage of high-dimensional MANOVA theory, biological researchers often considered some form of data reduction such as cluster analysis and factor analysis, which can suffer from many problems, as pointed out by Zapala and Schork [2012]. In the works Zapala and Schork [2006, 2012], the authors incorporated a distance matrix to modify the standard MANOVA, but they commented that there is very little published material that can be used to guide a researcher as to which distance measure is the most appropriate for a given situation. Motivated by these real-world applications, we aim to develop a general methodology for high dimensional MANOVA and lay a theoretical foundation for assessing statistical significance.

The testing problem (3.3) for the model (3.2) is closely related to a group of high dimensional hypothesis tests. Two-sample mean test, for testing $H_0 : \mu_1 = \mu_2$ where $\mu_1, \mu_2 \in \mathbb{R}^d$, is a special case with $p = 2$, $B = (\mu_1, \mu_2)^\top$ and $\mathcal{C} = (1, -1)$. There is a large literature accommodating the Hotelling T^2 type statistic into the high-dimensional situation where d is large; see for example, Bai and Saranadasa [1996], Chen and Qin [2010], Srivastava et al. [2013] among many others. It can be generalized to test the equality of multiple mean vectors in high dimensions. Some notable work includes Schott [2007], Cai and Xia [2014], Hu et al. [2017], Li et al. [2017], Zhang et al. [2017] and Zhou et al. [2017]. In most existing work, the random samples were assumed to be Gaussian or follow some linear structure as that of Bai and Saranadasa [1996]. The testing problem we are concerned is much more general. For one thing, all the aforementioned high dimensional mean test problems can be fitted into our framework, apart from which, we can deal with the more general multivariate linear regression in the presence of an increasing number of predictor variables. For another, we do not assume the Gaussianity or any particular structure of the random vectors $(V_k)_{k\in\mathbb{N}}$.

Throughout the paper, we assume that $p < n$ and the design matrix X is of full column rank such that $X^\top X$ is invertible. The conventional MANOVA test statistic for (3.3) is given by

$$Q_n = |PY|_{\mathbb{F}}^2 = \sum_{1\le i,j\le n} P_{ij} Y_i^\top Y_j, \tag{3.4}$$

where

$$P = X(X^\top X)^{-1}\mathcal{C}^\top\{\mathcal{C}(X^\top X)^{-1}\mathcal{C}^\top\}^{-1}\mathcal{C}(X^\top X)^{-1}X^\top = (P_{ij})_{i,j=1}^n$$

is the orthogonal projection matrix onto the linear space spanned by the column vectors of $X(X^\top X)^{-1}\mathcal{C}^\top$. We shall reject the null hypothesis H_0 if Q_n is larger than some critical value. In the univariate case where $d = 1$, the asymptotic behavior of Q_n has been extensively studied in literature; see Götze and Tikhomirov [1999] and Götze and Tikhomirov [2002] for detailed discussions. The validity to perform a test for (3.3) using Q_n when d is large has been open for a long time. The first goal of the paper is to provide a solution to this open problem by rigorously establishing a distributional approximation of the traditional

MANOVA test statistic when d is allowed to grow with n. Our key tool is the Gaussian approximation for degenerate U statistics: quadratic functionals of non-Gaussian random vectors can be approximated by those of Gaussian vectors with the same covariance structure. It is worth mentioning that Chen [2018] established a Gaussian approximation result for high dimensional non-degenerate U statistics by Stein's method, which can not be applied to the degenerate case. From a technical point of view, we employ completely different arguments to derive a finite sample error bound.

The main contributions of this chapter are three-fold. Firstly, we develop a systematic theory for the conventional MANOVA test statistic Q_n in the high dimensional setting. Specifically, we shall establish a dichotomy result: Q_n can either be approximated by a mixture of chi-sqaured distributions or by a normal distribution under different conditions, see Corollary 3.2.3. Due to this dichotomous nature of the asymptotic distribution of Q_n, we recommend using a new U type test statistic, which is the second contribution of our paper. Using the modified test statistic, such a dichotomy does not appear; see Theorem 3.2.6 for the asymptotic result. Thirdly, we will propose a new estimator for the second spectral moment of the covariance matrix via a data-splitting technique. To the best of our knowledge, it is the first work concerning an unbiased and ratio consistent estimator in the multivariate linear regression model.

3.2 Theoretical results

In this section, we shall first establish an asymptotic distribution theory for the MANOVA test statistic given by (3.4) in the high dimensional setting where the dimension d is large. We start with some notational definitions. Recall that $\Sigma = \mathrm{cov}(V_1)$. For simplicity of notation, let $f = |\Sigma|_{\mathbb{F}}$ denote the Frobenius norm of the covariance matrix Σ. For any $q > 2$, define

$$M_q = \mathbb{E}\left|\frac{V_1^\top V_2}{f}\right|^q \text{ and } L_q = \mathbb{E}\left|\frac{V_1^\top \Sigma V_1}{f^2}\right|^{q/2}.$$

3.2.1 Asymptotic distribution of the conventional MANOVA test statistics

Under the null hypothesis H_0, $\mathcal{C}B = 0$ and consequently

$$PXB = X(X^\top X)^{-1}\mathcal{C}^\top\{\mathcal{C}(X^\top X)^{-1}\mathcal{C}^\top\}^{-1}\mathcal{C}B = 0.$$

Therefore the MANOVA test statistic $Q_n = |PXB + PV|_{\mathbb{F}}^2 = |PV|_{\mathbb{F}}^2$, which can be further decomposed as

$$Q_n = \sum_{1\le i,j\le n} P_{ij}V_i^\top V_j = \sum_{i=1}^n P_{ii}V_i^\top V_i + \sum_{i\neq j} P_{ij}V_i^\top V_j =: D_n + Q_n^\star.$$

Observe that the two terms D_n and $Q_n^\star$ are uncorrelated. Hence $\mathbb{E}(Q_n) = m\mathrm{tr}(\Sigma)$ and

$$\mathrm{var}(Q_n) = \mathrm{var}(D_n) + \mathrm{var}(Q_n^\star) = \sum_{i=1}^n P_{ii}^2\|\mathbb{E}_0(V_1^\top V_1)\|^2 + 2\Big(m - \sum_{i=1}^n P_{ii}^2\Big)f^2,$$

where $\mathbb{E}_0(\cdot) = \cdot - \mathbb{E}(\cdot)$ and $\|\cdot\| = \|\cdot\|_2$. In the high dimensional setting where the dimension d can be much larger than the sample size n, the magnitudes of $\mathrm{var}(D_n)$ and $\mathrm{var}(Q_n^\star)$ will be quite different for non-Gaussian random vectors; cf. Example 3.4.1. As a consequence, Q_n can exhibit different asymptotic distributions. More precisely, to quantify the discrepancy between $\mathrm{var}(D_n)$ and $\mathrm{var}(Q_n^\star)$, we define

$$\Lambda_n^2 = \frac{\sum_{i=1}^n P_{ii}^2\|\mathbb{E}_0(V_1^\top V_1)\|^2}{mf^2}.$$

If $\Lambda_n \to 0$, $\mathbb{E}_0(D_n)/\sqrt{\mathrm{var}(Q_n)} \xrightarrow{\mathbb{P}} 0$ asymptotically. In this case, we first derive the asymptotic distribution of $Q_n^\star$ based on a new invariance principle; cf. Lemma 3.2.1. Consequently we establish a Gaussian approximation result for Q_n which upper bounds the Kolmogorov distance between the distribution functions of Q_n and its Gaussian analogue. Specifically, let $Z_1, Z_2, \ldots, Z_n \in \mathbb{R}^d$ be i.i.d. Gaussian random vectors $N(0,\Sigma)$ and $Z =$

$(Z_1, Z_2, \ldots, Z_n)^\top$. Then the Gaussian analogue of Q_n is defined with V replaced by Z,

$$G_n = |PZ|_{\mathbb{F}}^2 = \sum_{1 \le i,j \le n} P_{ij} Z_i^\top Z_j. \tag{3.5}$$

Recall that $\mathcal{C} \in \mathbb{R}^{m \times p}$ is of full row rank m. Hence P is an orthogonal projection matrix with rank m and the eigenvalues of P are $\lambda_1(P) = \cdots = \lambda_m(P) = 1$ and $\lambda_{m+1}(P) = \cdots = \lambda_n(P) = 0$. As a result, G_n is distributed as a linear combination of independent chi-squared random variables. Specifically, let $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_d \ge 0$ be the eigenvalues of the covariance matrix Σ, then

$$G_n \stackrel{\mathcal{D}}{=} \sum_{k=1}^{d} \lambda_k \eta_{k,m},$$

where $\eta_{k,m}$, $k \in \mathbb{N}$, are i.i.d. chi-squared random variables with m degrees of freedom.

Before presenting the main Gaussian approximation result, we first introduce a regular condition on the design matrix X.

Assumption 3.2.1. Recall that $P_{11}, P_{22}, \ldots, P_{nn}$ are diagonal elements of the matrix P. Assume that

$$\frac{1}{m} \sum_{i=1}^{n} P_{ii}^2 \to 0.$$

Remark 7. Since $P \in \mathbb{R}^{n \times n}$ is positive semi-definite and $\mathrm{tr}(P) = \sum_{i=1}^n P_{ii} = m$, Assumption 3.2.1 is satisfied as long as $\max_{i \le n} |P_{ii}| \to 0$, which is a natural condition; cf. Example 3.2.1.

Lemma 3.2.1. *Let $q = 2 + \delta$, where $0 < \delta \le 1$. Assume $M_q < \infty$ and that as $n \to \infty$,*

$$\Delta_q = \frac{\sum_{i \ne j} |P_{ij}|^q}{m^{q/2}} M_q + \frac{\sum_{i=1}^n P_{ii}^{q/2}}{m^{q/2}} L_q \to 0. \tag{3.6}$$

Then, under Assumptions 3.2.1,

$$\rho\{Q_n^\star, \mathbb{E}_0(G_n)\} \le C_q \Delta_q^{1/(2q+1)} + C\left(\frac{\sum_{i=1}^n P_{ii}^2}{m}\right)^{1/5} \to 0.$$

The following lemma establishes an upper bound for Δ_q.

Lemma 3.2.2. *Assume that $M_q < \infty$, then we have*

$$\Delta_q \le 2 \max_{i \le n} |P_{ii}|^{\delta/2} m^{-\delta/2} M_q.$$

Remark 8. Condition (3.6) can be viewed as the Lyapunov condition for high dimensional Gaussian approximation. It is quite natural and does not impose any explicit restriction on the relation between the dimension d and the sample size n directly. In particular, condition (3.6) can be dimension free for some commonly used models, namely, (3.6) is satisfied for arbitrary dimension d. For example, suppose that the V_i's follow the linear process model

$$V_i = A\varepsilon_i, \tag{3.7}$$

where A is a $d \times \ell$ matrix for some positive integer $\ell \ge 1$, $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{i\ell})^\top$ and ε_{ij}, $i, j \in \mathbb{N}$, are independent random variables with $\mathbb{E}(\varepsilon_{ij}) = 0$ and uniformly bounded qth moment

$$\max_{j \le \ell} \mathbb{E}|\varepsilon_{ij}|^q \le C < \infty.$$

Applying the Burkholder inequality, we can show that

$$M_q \le (1+\delta)^q \max_{j \le \ell} \|\varepsilon_{ij}\|_q^{2q}.$$

Consequently, by Lemma 3.2.2, a sufficient condition for $\Delta_q \to 0$ is

$$\frac{1}{m} \max_{i \le n} |P_{ii}| \to 0. \tag{3.8}$$

It is worth mentioning that condition (3.8) depends only on the design matrix X and does not impose any restriction on the dimension d. Furthermore, under Assumption 3.2.1, condition (3.8) is automatically satisfied in view of

$$\left(\frac{1}{m} \max_{i \le n} |P_{ii}|\right)^2 \le \frac{1}{m^2} \sum_{i=1}^{n} P_{ii}^2 \to 0.$$

In another case where $\Lambda_n \to \infty$, Q_n is stochastically dominant by the diagonal term $D_n = \sum_{i=1}^n P_{ii} V_i^\top V_i$, which is a weighted sum of independent random variables. Under natural Lindeberg condition, we derive a central limit theorem of Q_n; cf. (3.9). Together with Lemma 3.2.1, we establish a detailed characteriztion of the asymptotic distribution of Q_n under the null hypothesis H_0.

Corollary 3.2.3. *Assume that the null hypothesis H_0 is satisfied.*

1. *Assume $\Lambda_n \to 0$. Then, under (3.6) and Assumption 3.2.1,*

$$\rho(Q_n, G_n) \le C_1 \Lambda_n^{2/5} + C_q \Delta_q^{1/(2q+1)} + C_2 \left(\frac{\sum_{i=1}^n P_{ii}^2}{m}\right)^{1/5} \to 0.$$

2. *Assume $\Lambda_n \to \infty$ and that the Lindeberg condition holds for*

$$U_i = \frac{\mathbb{E}_0(P_{ii} V_i^\top V_i)}{\Lambda_n f \sqrt{m}},$$

that is, $\sum_{i=1}^n \mathbb{E}(U_i^2 \mathbb{I}\{|U_i| > \epsilon\}) \to 0$ for any $\epsilon > 0$. Then we have the central limit theorem

$$\frac{Q_n - m\mathrm{tr}(\Sigma)}{\Lambda_n f \sqrt{m}} \Rightarrow N(0, 1). \tag{3.9}$$

Remark 9. Corollary 3.2.3 illustrates an interesting dichotomy: the conventional MANOVA test statistic Q_n can have one of the two different asymptotic distributions under the null hypothesis, depending on the magnitude of the unknown quantity Λ_n. This nature of dichotomy poses extra difficulty for testing (3.3) in practical implementation as we need to predetermine which asymptotic distribution to use. Any subjective choice may lead to unreliable conclusion. To illustrate this, suppose now $\Lambda_n \to 0$. For any significance level $\alpha \in (0,1)$, let $G_n^{-1}(\alpha)$ denote the $(1-\alpha)$th quantile of G_n, that is,

$$G_n^{-1}(\alpha) = \inf\{t : \mathbb{P}(G_n \leq t) \geq 1 - \alpha\}.$$

Based on Corollary 3.2.3, an α level test for (3.3) is given by

$$\Phi_0 = \mathbb{I}\{Q_n \geq G_n^{-1}(\alpha)\}.$$

However, if one implements Φ_0 under the case where $\Lambda \to \infty$, then the type I error of Φ_0 satisfies that

$$\mathbb{E}(\Phi_0 | H_0) \to \frac{1}{2},$$

which implies that Φ_0 in this scenario ($\Lambda \to \infty$) is no better than random guessing.

3.2.2 Modified U *type test statistics*

The discussion in Remark 9 illustrates that the conventional MANOVA test statistic Q_n is not suitable for testing (3.3) in the high dimensional setting due to the dichotomous nature of the asymptotic distribution. In this section, we propose a modified U type test statistic of Q_n for which such a dichotomuy does not occur.

To fix the idea, let $B_0 \in \mathbb{R}^{p \times d}$ denote the coefficient matrix of model (3.2) under the null hypothesis H_0. Hence $\mathcal{C}B_0 = 0$. Motivated by Lemma 3.2.1 and Corollary 3.2.3, a natural candidate of the test statistic is given by

$$\tilde{Q}_n = Q_n - \sum_{i=1}^{n} P_{ii}(Y_i - B_0^\top X_i)^\top (Y_i - B_0^\top X_i).$$

Note that $\tilde{Q}_n = Q_n^\star$ under the null hypothesis H_0. However, the coefficient matrix B_0 is unknown and $\tilde{Q}_n$ is infeasible. In practice, we propose to use an empirical approximation $Q_n - \hat{D}_n$ of $\tilde{Q}_n$. In particular, the modified test statistic $Q_n - \hat{D}_n$ should satisfy that under the null hypothesis H_0,

$$\hat{D}_n = \sum_{i=1}^{n} P_{ii} V_i^\top V_i + \sum_{i \neq j} K_{ij} V_i^\top V_j, \tag{3.10}$$

where $K_{ij} = K_{ji}$ for any $1 \leq i \neq j \leq n$ and

$$\frac{Q_n - \hat{D}_n - Q_n^\star}{\sqrt{\mathrm{var}(Q_n^\star)}} \xrightarrow{\mathbb{P}} 0. \tag{3.11}$$

The latter reveals that the modified test statistic $Q_n - \hat{D}_n$ is asymptotically distributed as $Q_n^\star$.

Let $P_0 = X(X^\top X)^{-1} X^\top - P$ be the projection matrix under the null hypothesis H_0 and $\bar{P}_0 = I_n - P_0$. Then the residual matrix under the null hypothesis is

$$\hat{V}_0 = \bar{P}_0 Y = (\hat{V}_{1,0}, \hat{V}_{2,0}, \ldots, \hat{V}_{n,0})^\top.$$

Then a natural candidate of $\hat{D}_n$ is $\sum_{k=1}^{n} \theta_k \hat{V}_{k,0}^\top \hat{V}_{k,0}$, which can be decomposed as

$$\sum_{k=1}^{n} \theta_k \hat{V}_{k,0}^\top \hat{V}_{k,0} = \sum_{i=1}^{n} \left(\sum_{k=1}^{n} \theta_k \bar{P}_{ik,0}^2 \right) V_i^\top V_i + \sum_{i \neq j} \left(\sum_{k=1}^{n} \theta_k \bar{P}_{ik,0} \bar{P}_{jk,0} \right) V_i^\top V_j.$$

To satisfy (3.10), we shall choose $\theta = (\theta_1, \theta_2, \ldots, \theta_n)^\top$ such that $\sum_{k=1}^{n} \theta_k \bar{P}_{ik,0}^2 = P_{ii}$ for each $1 \le i \le n$, which equivalently can be written in compact matrix form as

$$(\bar{P}_0 \circ \bar{P}_0)\theta = (P_{11}, \ldots, P_{nn})^\top.$$

Then the modified U type test statistic is given by

$$Q_n^\diamond = Q_n - \hat{D}_n = Q_n - \sum_{i=1}^n \theta_i \hat{V}_{i,0}^\top \hat{V}_{i,0}.$$

Let $P_\diamond = P - \bar{P}_0 D_\theta \bar{P}_0 = (P_{ij,\diamond})_{i,j=1}^n$, where $D_\theta = \mathrm{diag}(\theta_1, \ldots, \theta_n)$ is a diagonal matrix. Then, under the null hypothesis H_0,

$$Q_n^\diamond = \mathrm{tr}(V^\top P_\diamond V) = \sum_{i \neq j} P_{ij,\diamond} V_i^\top V_j.$$

In the following lemma, we introduce a sufficient condition such that the proposed test statistic $Q_n^\diamond$ exists and is well defined.

Lemma 3.2.4. *Assume that there exists a positive constant $\varpi_0 < 1/2$ such that*

$$\max_{i \le n} |P_{ii,0}| \le \varpi_0. \tag{3.12}$$

Then $\bar{P}_0 \circ \bar{P}_0$ is strictly diagonally dominant and

$$\sum_{i=1}^n \theta_i P_{ii} \le \frac{\sum_{i=1}^n P_{ii}^2}{(1 - 2\varpi_0)(1 - \varpi_0)}.$$

Remark 10. Condition (3.12) ensures that $\bar{P}_0 \circ \bar{P}_0$ is invertible and consequently the solution θ exists and is unique. Under Assumption 3.2.1 and (3.12), it follows that $m^{-1} \sum_{i=1}^n \theta_i P_{ii} \to 0$ and

$$\mathrm{var}(Q_n^\diamond) = 2\Big(m - \sum_{i=1}^n \theta_i P_{ii}\Big) f^2 > 0,$$

which implies that the proposed U test statistic $Q_n^\diamond$ is non-degenerate and well defined. Furthermore, a careful inspection of the proof of Lemma 3.2.4 reveals that

$$\frac{\|Q_n^\diamond - Q_n^\star\|^2}{\mathrm{var}(Q_n^\star)} \lesssim \frac{\max_{i\le n}|P_{ii}|}{m}\sum_{i=1}^n P_{ii}^2 \to 0.$$

As a result, the modified test statistic $Q_n^\diamond$ is asymptotically distributed as $Q_n^\star$ under the null hypothesis. Hence (3.11) is satisfied.

Now we briefly comment on condition (3.12). Since both P and $X(X^\top X)^{-1}X^\top - P$ are projection matrices, it follows that $\max\{P_{ii,0}, P_{ii}\} \le X_i^\top (X^\top X)^{-1} X_i$ for each $1 \le i \le n$. Hence a sufficient condtion for (3.12) is

$$\max_{i\le n} X_i^\top (X^\top X)^{-1} X_i \le \varpi_0.$$

In the following example, we shall verify this condition for (3.2) under the Gaussian random design.

Example 3.2.1. Suppose $X_1, X_2, \ldots, X_n \in \mathbb{R}^p$ are i.i.d. Gaussian random vectors $N(0, \Gamma)$, where the covariance matrix $\Gamma \in \mathbb{R}^{p\times p}$ has minimal eigenvalue $\lambda_{\min}(\Gamma) > 0$. Then with probability at least $1 - 2\exp(-n/2) - n^{-1}$,

$$\max_{i\le n} X_i^\top (X^\top X)^{-1} X_i \le \frac{9p + 18\sqrt{2p\log n} + 36\log n}{n}.$$

In this case, condition (3.12) is satisfied with high probability as long as p/n is sufficiently small.

Proposition 3.2.5. *Under condition (3.12), we have* $\mathbb{E}(Q_n^\diamond) \ge 0$. *In particular,*

$$\mathbb{E}(Q_n^\diamond) = 0 \text{ if and only if } \mathcal{C}B = 0.$$

Similar as (3.5), the Gaussian analogue of $Q_n^\diamond$ is defined with V replaced by Z,

$$G_n^\diamond = \operatorname{tr}(Z^\top P_\diamond Z) = \sum_{i \neq j} P_{ij,\diamond} Z_i^\top Z_j.$$

Let $\lambda_1(P_\diamond), \lambda_2(P_\diamond), \ldots, \lambda_n(P_\diamond)$ be the eigenvalues of the matrix $P_\diamond$. Since $P_\diamond$ is symmetric, all the $\lambda_i(P_\diamond)$'s are real and consequently $G_n^\diamond$ is distributed as

$$G_n^\diamond \stackrel{\mathcal{D}}{=} \sum_{k=1}^{d} \sum_{i=1}^{n} \lambda_k \lambda_i(P_\diamond) \eta_{ik,1},$$

where $\eta_{ik,1}$, $i, k \in \mathbb{N}$, are i.i.d. chi-squared random variables with one degree of freedom. The following theorem establishes a Gaussian approximation which upper bounds the Kolmogorov distance between the distribution functions of $Q_n^\diamond$ and its Gaussian analogue $G_n^\diamond$. It is shown that the modification of the test statistic Q_n removes the dichotomous nature of the asymptotic distribution.

Theorem 3.2.6. *Assume that (3.12) holds and that*

$$\Delta_{q,\diamond} = \frac{\sum_{i \neq j} |P_{ij,\diamond}|^q}{m^{q/2}} M_q + \frac{\sum_{i=1}^n |\sum_{j=1}^n P_{ij,\diamond}^2|^{q/2}}{m^{q/2}} L_q \to 0.$$

Then, under Assumptions 3.2.1 and the null hypothesis H_0,

$$\rho(Q_n^\diamond, G_n^\diamond) \leq C_q \Delta_{q,\diamond}^{1/(2q+1)} + C \left(\frac{\sum_{i=1}^n P_{ii}^2}{m} \right)^{1/5} \to 0.$$

Similar to Lemma 3.2.2, we establish an upper bound for $\Delta_{q,\diamond}$ in the following lemma.

Lemma 3.2.7. *Under condition (3.12), we have*

$$\Delta_{q,\diamond} \lesssim \max_{i \leq n} |P_{ii}|^{\delta/2} m^{-\delta/2} M_q.$$

Remark 11. Note that the upper bound for $\Delta_{q,\diamond}$ is similar with that of Δ_q in Lemma 3.2.2.

As discussed in Remark 8, the Lyapunov type condition $\Delta_{q,\diamond} \to 0$ is fairly mild.

For any significance level $\alpha \in (0,1)$, let C_α be the $(1-\alpha)$th quantile of $G_n^\diamond/\sqrt{\mathrm{var}(G_n^\diamond)}$. In view of Theorem 3.2.6 and Proposition 3.2.5, an α level test for (3.3) is given by

$$\Phi_\diamond = \mathbb{I}\left\{\frac{Q_n^\diamond}{f\sqrt{2m_\diamond}} > C_\alpha\right\}, \text{ where } m_\diamond = m - \sum_{i=1}^{n} \theta_i P_{ii}.$$

We shall reject the null hypothesis H_0 whenever $\Phi_\diamond = 1$.

To perform $\Phi_\diamond$ in practice, one needs to estimate the Frobenius norm $f = |\Sigma|_{\mathbb{F}}$ and the critical value C_α. An unbiased and ratio consistent estimator for f^2 is proposed in Section 3.4; cf. (3.15). To approximate the critical value C_α, we shall first derive a central limit theorem for $Q_n^\diamond$ in the following theorem. Then one can take C_α to be the $(1-\alpha)$th quantile of the standard normal distribution. Consequently, we obtain a feasible test (3.16).

Theorem 3.2.8. *Assume that there exists a positive constant $\varpi_1 < 1$ such that*

$$\max_{i \le n} |P_{ii}| \le \varpi_1(1-2\varpi_0)(1-\varpi_0).$$

Under the conditions of Theorem 3.2.6, the central limit theorem $Q_n^\diamond/\sqrt{\mathrm{var}(Q_n^\diamond)} \Rightarrow N(0,1)$ holds if and only if

$$\frac{\lambda_1}{f\sqrt{m}} \to 0. \tag{3.13}$$

Remark 12. If $m \to \infty$, as $\lambda_1 \le f$, condition (3.13) is automatically satisfied and $Q_n^\diamond$ is asymptotically normal. In another case where m is bounded, (3.13) is equivalent to $\mathrm{tr}(\Sigma^4)/f^4 \to 0$, which is a common assumption to ensure the asymptotic normality of high dimensional quadratic statistics; see, for example, Bai and Saranadasa [1996], Chen and Qin [2010], Cai and Ma [2013] and Zhang et al. [2018] among others.

It's worth mentioning that if (3.13) is violated, Theorem 3.2.8 implies that the asymptotic distribution of $Q_n^\diamond$ can be non-normal. For example, consider testing the hypotheses $\mu_Y = 0$ versus $\mu_Y \ne 0$, where $\mu_Y = \mathbb{E}(Y_1)$ is the population mean vector of i.i.d. random vectors

$Y_1, Y_2, \ldots, Y_n$. Assume that the covariance matrix $\Sigma = \mathrm{cov}(Y_1) = (\Sigma_{jk})_{j,k=1}^d$ has entries $\Sigma_{jk} = \vartheta + (1-\vartheta)\mathbb{I}\{j = k\}$ for some constant $\vartheta \in (0,1)$. Then $\lambda_1/(f\sqrt{m}) \to 1$ and Theorem 3.2.6 implies that

$$\frac{Q_n^\diamond}{\sqrt{\mathrm{var}(Q_n^\diamond)}} = \frac{\sum_{i\neq j} Y_i^\top Y_j}{f\sqrt{2n(n-1)}} \Rightarrow \frac{\chi_1^2 - 1}{\sqrt{2}}.$$

3.3 Applications

As mentioned in the introduction, our paradigm (3.3) is actually fairly general and it can be applied to many commonly studied hypothesis testing problems. In this section, we consider two specific examples to illustrate the utility of the proposed U type test statistic and the corresponding distribution theory.

3.3.1 High dimensional one-way MANOVA

Let $K \geq 2$ be a positive integer. Let $\mathcal{Y}_{i1}, \mathcal{Y}_{i2}, \ldots, \mathcal{Y}_{iN_i} \in \mathbb{R}^d$, $i = 1, \ldots, K$, be K independent samples following the model

$$\mathcal{Y}_{ij} = \mu_i + \mathcal{V}_{ij}, \quad j = 1, \ldots, N_i, i = 1, \ldots, K,$$

where $\mu_i = \mathbb{E}(\mathcal{Y}_{ij})$, $i = 1, \ldots, K$, are unknown mean vectors, $\mathcal{V}_{ij} \in \mathbb{R}^d, i, j \in \mathbb{N}$, are i.i.d. random vectors with $\mathbb{E}(\mathcal{V}_{11}) = 0$ and $\mathrm{cov}(\mathcal{V}_{11}) = \Sigma$. We consider testing the equality of the μ_i's, namely, testing the hypotheses

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_K \text{ versus } H_1 : \mu_i \neq \mu_j \text{ for some } i \neq j.$$

Let $N = \sum_{i=1}^K N_i$ denote the total sample size. Following the steps in Section 3.2.2, we derive the test statistic

$$Q_n^\diamond = \sum_{i=1}^K P_{ii,\natural} \sum_{j\neq k} \mathcal{Y}_{ij}^\top \mathcal{Y}_{ik} + \sum_{i\neq l} P_{il,\natural} \sum_{j=1}^{N_i}\sum_{k=1}^{N_l} \mathcal{Y}_{ij}^\top \mathcal{Y}_{lk}, \tag{3.14}$$

where

$$P_{ii,\natural} = \frac{1}{N-2}\left(\frac{N}{N_i} - \frac{N+K-2}{N-1}\right),\ \ 1\le i\le K,$$
$$P_{il,\natural} = \frac{1}{N-2}\left(\frac{1}{N_i} + \frac{1}{N_l} - \frac{N+K-2}{N-1}\right),\ \ 1\le i\neq l\le K.$$

In the context of two sample mean test where $K=2$, $Q_n^\diamond$ reduces to

$$Q_n^\diamond = \frac{\sum_{i\neq j}\sum_{k\neq l}(\mathcal{Y}_{1i}-\mathcal{Y}_{2k})^\top(\mathcal{Y}_{1j}-\mathcal{Y}_{2l})}{(N-1)(N-2)N_1N_2/N},$$

which coincides with the common U type test statistics used in Chen and Qin [2010].

To investigate the asymptotic behavior of $Q_n^\diamond$, we impose a natural assumption.

Assumption 3.3.1. Assume that there exist positive constants $\pi_1, \pi_2, \ldots, \pi_K \in (0,1)$ such that $\pi_1 + \pi_2 + \cdots + \pi_K = 1$ and

$$\frac{N_i}{N} \to \pi_i,\ \ i = 1, 2, \ldots, K.$$

Let $\mathcal{Z}_{ij} \in \mathbb{R}^d$, $i, j \in \mathbb{N}$, be i.i.d. Gaussian random vectors $N(0,\Sigma)$. Similar to (3.5), the Gaussian analogue of $Q_n^\diamond$ is defined with the $\mathcal{V}_{ij}$'s replaced by the $\mathcal{Z}_{ij}$'s,

$$G_n^\diamond = \sum_{i=1}^K P_{ii,\natural} \sum_{j\neq k} \mathcal{Z}_{ij}^\top \mathcal{Z}_{ik} + \sum_{i\neq l} P_{il,\natural} \sum_{j=1}^{N_i}\sum_{k=1}^{N_l} \mathcal{Z}_{ij}^\top \mathcal{Z}_{lk}.$$

Following Theorem 3.2.6, we establish the asymptotic distribution of $Q_n^\diamond$.

Proposition 3.3.1. *Let $q = 2 + \delta$, where $0 < \delta \leq 1$. Assume that*

$$M_{q,\mathcal{V}} = \mathbb{E}\left|\frac{\mathcal{V}_{11}^{\top}\mathcal{V}_{12}}{f}\right|^{q} < \infty.$$

Then, under the null hypothesis H_0 and Assumption 3.3.1,

$$\rho(Q_n^{\diamond}, G_n^{\diamond}) \leq C_q(M_{q,\mathcal{V}} N^{-\delta/2})^{1/(2q+1)} + C(K/N)^{1/5}.$$

Remark 13. It is worth mentioning that both the dimension d and the number of groups K can grow with the total sample size N. For ease of illustration, we assume that the $\mathbb{E}_0(Y_{ij})$'s are identically distributed. However, following the proof of Theorem 3.2.6, one can easily derive the asymptotic distribution of $Q_n^{\diamond}$ under the heteroscedastic setting where the K samples can have different distributions.

3.3.2 High dimensional nonparametric one-way MANOVA

For each $1 \leq i \leq K$, let F_i denote the cumulative distribution function of the random vector $\mathcal{Y}_{ij}$. In this section, we consider testing if the K independent samples are equally distributed, namely, testing the hypotheses

$$H_0 : F_1 = F_2 = \ldots = F_K \text{ versus } H_1 : F_i \neq F_j \text{ for some } i \neq j.$$

It is very fundamental and important in statistical inference and it has been extensively studied in the literature; see, for example, Kruskal and Wallis [1952], Akritas and Arnold [1994], Brunner and Puri [2001], Rizzo and Székely [2010] and Thas [2010] among many others. We shall compute the modified U test statistic for this specific testing problem and then derive the asymptotic distribution. In particular, our asymptotic framework allows both d and K to grow with N.

To begin with, let $\phi_i(t) = \mathbb{E}\{\exp(it^{\top}\mathcal{Y}_{ij})\}$ denote the characteristic function of $\mathcal{Y}_{ij}$. Then it is equivalent to test

$$H_0 : \phi_1(t) = \phi_2(t) = \ldots = \phi_K(t) \text{ versus } H_1 : \phi_i(t) \neq \phi_j(t) \text{ for some } i \neq j.$$

For each $1 \leq i \leq K$ and $1 \leq j \leq N_i$, define a two dimensional random process

$$\mathcal{Y}_{ij}(t) = \left\{ \sqrt{w(t)} \cos(t^\top \mathcal{Y}_{ij}), \sqrt{w(t)} \sin(t^\top \mathcal{Y}_{ij}) \right\}^\top, \ \ t \in \mathbb{R}^d,$$

where $w(t)$ is a suitable positive weight function defined on $\mathbb{R}^d$. The mean function and the covariance function of $\mathcal{Y}_{ij}(t)$ are respectively given by

$$\mu_i(t) = \mathbb{E}\{\mathcal{Y}_{ij}(t)\} \text{ and } \Sigma(t,s) = \text{cov}\{\mathcal{Y}_{11}(t), \mathcal{Y}_{11}(s)\}, \ \ t, s \in \mathbb{R}^d.$$

Observe that $\mu_i(t) = \mu_j(t)$ if and only if $\phi_i(t) = \phi_j(t)$ for each $i \neq j$. Hence it is equivalent to test the hypotheses

$$H_0 : \mu_1(t) = \mu_2(t) = \ldots = \mu_K(t) \text{ versus } H_1 : \mu_i(t) \neq \mu_j(t) \text{ for some } i \neq j.$$

Similar to (3.14), we derive the test statistic

$$\mathcal{Q}_n^\diamond = \sum_{i=1}^K P_{ii,\natural} \sum_{j \neq k} \int \mathcal{Y}_{ij}(t)^\top \mathcal{Y}_{ik}(t) dt + \sum_{i \neq l} P_{il,\natural} \sum_{j=1}^{N_i} \sum_{k=1}^{N_l} \int \mathcal{Y}_{ij}(t)^\top \mathcal{Y}_{lk}(t) dt.$$

Let $\mathcal{Z}_{ij}(t)$, $i, j \in \mathbb{N}$, be i.i.d. two dimensional Gaussian processes such that $\mathbb{E}\{\mathcal{Z}_{ij}(t)\} = 0$ and $\text{cov}\{\mathcal{Z}_{11}(t), \mathcal{Z}_{11}(s)\} = \Sigma(t,s)$. Then the Gaussian analogue of $\mathcal{Q}_n^\diamond$ is defined as

$$\mathcal{G}_n^\diamond = \sum_{i=1}^K P_{ii,\natural} \sum_{j \neq k} \int \mathcal{Z}_{ij}(t)^\top \mathcal{Z}_{ik}(t) dt + \sum_{i \neq l} P_{il,\natural} \sum_{j=1}^{N_i} \sum_{k=1}^{N_l} \int \mathcal{Z}_{ij}(t)^\top \mathcal{Z}_{lk}(t) dt.$$

Proposition 3.3.2. *Let $q = 2 + \delta$, where $0 < \delta \le 1$. Define*

$$\mathcal{M}_q = \mathbb{E}\left|\frac{\int \mathcal{Y}_{11}(t)^\top \mathcal{Y}_{12}(t)dt}{\mathcal{F}}\right|^q, \text{ where } \mathcal{F}^2 = \int\int |\Sigma(t,s)|_{\mathbb{F}}^2 dtds.$$

Then, under Assumption 3.3.1 and the null hypothesis that the K independent samples are equally distributed, we have

$$\rho(\mathcal{Q}_n^\diamond, \mathcal{G}_n^\diamond) \le C_q(\mathcal{M}_q N^{-\delta/2})^{1/(2q+1)} + C(K/N)^{-1/5}.$$

Remark 14. To compute the test statistic $\mathcal{Q}_n^\diamond$, we need to calculate the high dimensional integral over $t \in \mathbb{R}^d$, which can be computational intractable. To facilitate the computation, we shall choose suitable weight function $w(t)$ such that the test statistic $\mathcal{Q}_n^\diamond$ has a simple closed-form expression. For example, suppose that $w(t) = (\kappa^2 + |t|^2)^{-(d+1)/2}$ for some $\kappa > 0$, where $|\cdot|$ stands for the Euclidean distance. Then for any i, j, k, l, it follows that

$$\int_{t\in\mathbb{R}^d} \mathcal{Y}_{ij}(t)^\top \mathcal{Y}_{lk}(t)dt = \int_{t\in\mathbb{R}^d} \cos\{t^\top(Y_{ij} - Y_{lk})\}w(t)dt = \exp(-\kappa|Y_{ij} - Y_{lk}|).$$

Consequently, $\mathcal{Q}_n^\diamond$ reduces to

$$\mathcal{Q}_n^\diamond = \sum_{i=1}^K P_{ii,\natural} \sum_{j\ne k} \exp(-\kappa|Y_{ij} - Y_{ik}|) + \sum_{i\ne l} P_{il,\natural} \sum_{j=1}^{N_i}\sum_{k=1}^{N_l} \exp(-\kappa|Y_{ij} - Y_{lk}|).$$

3.4 Practical implementation

To construct feasible test for (3.3) based on the test statistic $Q_n^\diamond$, we need to estimate the Frobenius norm $f = |\Sigma|_{\mathbb{F}}$ of the unknown covariance matrix Σ. In this section, we propose an unbiased estimator of f^2, which is shown to be ratio-consistent under fairly mild moment conditions. Combined with the central limit theorem of $Q_n^\diamond$ in Theorem 3.2.8, we then establish a feasible test; cf. (3.16).

To begin with, we observe that $\mathbb{E}(V_i^\top V_j)^2 = f^2$ for any $i \ne j$. Hence, given $V_1, V_2, \ldots, V_n$, one can estimate f^2 via an unbiased U type estimator

$$\hat{f}_o^2 = \frac{1}{n(n-1)} \sum_{i \neq j} (V_i^\top V_j)^2.$$

As the V_i's are unobservable, we shall replace V by the residual matrix $\hat{V}$ of the model (3.2), which is given by

$$\hat{V} = \bar{P}_1 Y = (\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_n)^\top,$$

where

$$\bar{P}_1 = I_n - X(X^\top X)^{-1} X^\top = (\bar{P}_{ij,1})_{i,j=1}^n.$$

However, the resulting estimator is generally biased. More precisely, note that for any $i \neq j$,

$$\begin{aligned}\mathbb{E}(\hat{V}_i^\top \hat{V}_j)^2 = \bar{P}_{ii,1}\bar{P}_{jj,1} f^2 + \sum_{k=1}^n \bar{P}_{ik,1}^2 \bar{P}_{jk,1}^2 \left[\|\mathbb{E}_0(V_1^\top V_1)\|^2 - 2f^2 \right] \\ + \bar{P}_{ij,1}^2 f^2 + \bar{P}_{ij,1}^2 \mathbb{E}(V_1^\top V_1)(V_2^\top V_2).\end{aligned}$$

which implies that $(\hat{V}_i^\top \hat{V}_j)^2$ is no longer an unbiased estimator of f^2 even after properly scaled. Motivated by this, we shall exclude the bias terms $(V_i^\top V_i)^2$ and $(V_i^\top V_i)(V_j^\top V_j)$'s. To this end, we propose a new estimator for f^2 via a data-splitting procedure as follows. For simplicity, we assume that the sample size n is even in what follows.

1. Randomly split $[n] = \{1, 2, \ldots, n\}$ into two halves $\mathcal{A}$ and $\mathcal{A}^c$. Let $\mathcal{M}_\mathcal{A} = \{(X_i, Y_i), i \in \mathcal{A}\}$ and $\mathcal{M}_{\mathcal{A}^c} = \{(X_i, Y_i), i \in \mathcal{A}^c\}$ denote the two data sets.

2. For both $\mathcal{M}_\mathcal{A}$ and $\mathcal{M}_{\mathcal{A}^c}$, fit model (3.1) with the least squares estimates and calculate

the sample covariance matrices

$$\hat{\Sigma}_{\mathcal{A}} = \frac{1}{n/2 - p} \hat{V}_{\mathcal{A}}^{\top} \hat{V}_{\mathcal{A}} \text{ and } \hat{\Sigma}_{\mathcal{A}^c} = \frac{1}{n/2 - p} \hat{V}_{\mathcal{A}^c}^{\top} \hat{V}_{\mathcal{A}^c}.$$

3. Compute the estimator of f^2

$$\hat{f}_{\mathcal{A}}^2 = \operatorname{tr}(\hat{\Sigma}_{\mathcal{A}} \hat{\Sigma}_{\mathcal{A}^c}). \tag{3.15}$$

Note that $\hat{\Sigma}_{\mathcal{A}}$ and $\hat{\Sigma}_{\mathcal{A}^c}$ are independent and both of them are unbiased estimators of the covariance matrix Σ. Hence $\hat{f}_{\mathcal{A}}^2$ is unbiased for f^2 in view of

$$\mathbb{E}(\hat{f}_{\mathcal{A}}^2) = \operatorname{tr}\{\mathbb{E}(\hat{\Sigma}_{\mathcal{A}})\mathbb{E}(\hat{\Sigma}_{\mathcal{A}^c})\} = \operatorname{tr}(\Sigma\Sigma) = f^2.$$

Theorem 3.4.1. *Assume that $p/n < \varpi_2$ for some positive constant $\varpi_2 < 1/2$ and that the least squares estimates are well defined for both $\mathcal{M}_{\mathcal{A}}$ and $\mathcal{M}_{\mathcal{A}^c}$. Then we have*

$$\left\| \frac{\hat{f}_{\mathcal{A}}}{f} - 1 \right\|^2 \lesssim \frac{M_4}{n^2} + \frac{p \times \operatorname{tr}(\Sigma^4)}{n^2 f^4} + \frac{\|\mathbb{E}_0(V_1^{\top} \Sigma V_1)\|^2}{n f^4}.$$

Remark 15. The proof of Theorem 3.4.1 is given in Section 3.6, where a more general bound of $\mathbb{E}|\hat{f}_{\mathcal{A}}/f - 1|^{\beta}$ is established for $1 < \beta \le 2$. Theorem 3.4.1 ensures that the proposed estimator $\hat{f}_{\mathcal{A}}$ is ratio consistent under mild moment conditions. Suppose now $(V_i)_{i \in \mathbb{N}}$ follow the linear process model (3.7) with $\max_{k \le \ell} \mathbb{E}|\varepsilon_{ik}|^4 \le C < \infty$. Elementary calculation shows that M_4 is bounded and

$$\|\mathbb{E}_0(V_1^{\top} \Sigma V_1)\|^2 \lesssim \operatorname{tr}(\Sigma^4).$$

Consequently, Theorem 3.4.1 implies that

$$\left\| \frac{\hat{f}_{\mathcal{A}}}{f} - 1 \right\|^2 \lesssim n^{-2} + \frac{\operatorname{tr}(\Sigma^4)}{n f^4}.$$

In this case, $\hat{f}_{\mathcal{A}}$ is ratio consistent as long as $n \to \infty$.

Remark 16. There are totally $\binom{n}{n/2}$ different ways of splitting $[n]$ into two halves. To reduce the influence of randomness of an arbitrary splitting, we can repeat the procedure independently for multiple times and then take the average of the resulting estimators. We refer to Fan et al. [2012] for more discussions about data-splitting and repeated data-splitting.

Remark 17. Let $\hat{\Sigma} = (n-p)^{-1}\hat{V}^\top \hat{V}$ be the sample covariance matrix. Note that $\mathbb{E}(\hat{V}_i^\top \hat{V}_j) = \bar{P}_{ij,1}\mathrm{tr}(\Sigma)$. Hence we estimate f^2 via

$$\hat{f}_S^2 = \frac{\sum_{i,j=1}^n |\hat{V}_i^\top \hat{V}_j - \bar{P}_{ij,1}\mathrm{tr}(\hat{\Sigma})|^2}{(n-p+2)(n-p-1)} = \frac{(n-p)^2}{(n-p+2)(n-p-1)}\left[|\hat{\Sigma}|_{\mathbb{F}}^2 - \frac{\{\mathrm{tr}(\hat{\Sigma})\}^2}{n-p}\right].$$

It is same with the estimator proposed by Srivastava and Fujikoshi [2006], where the V_i's are assumed to be normally distributed. See also Bai and Saranadasa [1996].

When the V_i's are non-Gaussian such that $\|\mathbb{E}_0(V_1^\top V_1)\|^2 \neq 2f^2$, this estimator is generally biased as

$$\mathbb{E}(\hat{f}_S^2) - f^2 = \frac{\sum_{i=1}^n \bar{P}_{ii,1}^2}{(n-p)(n-p+2)}\left[\|\mathbb{E}_0(V_1^\top V_1)\|^2 - 2f^2\right].$$

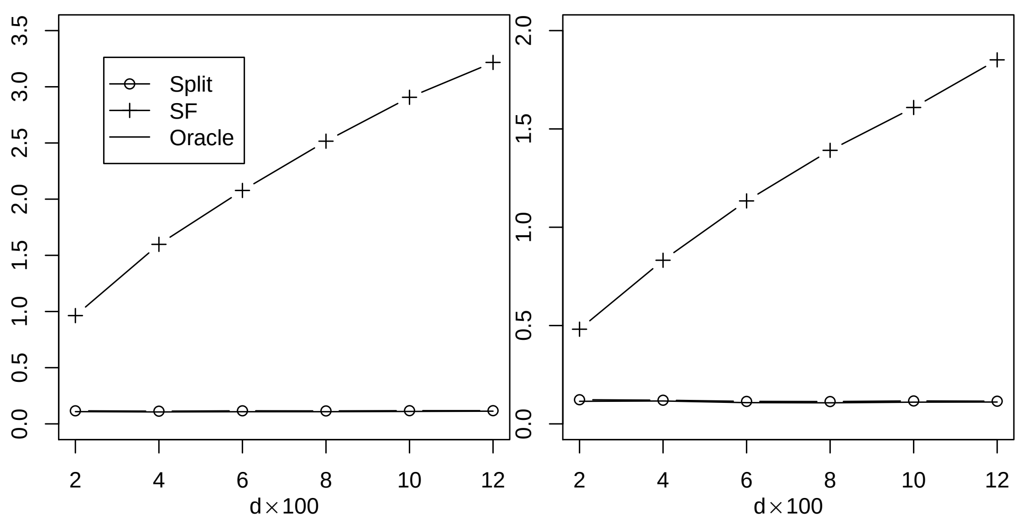
We also note that the bias of the estimator $\hat{f}_S^2$ can diverge when $\|\mathbb{E}_0(V_1^\top V_1)\|^2$ is much larger than f^2. Below we provide an example that typifies the diverging bias.

Example 3.4.1. Let the i.i.d. random vectors ϵ_i, ϵ_i', $i \in \mathbb{N}$, be $N(0, \Sigma)$ distributed, where $\Sigma = (\Sigma_{ij})_{i,j=1}^d$ has entries $\Sigma_{ij} = \theta^{|i-j|}$ for some $\theta \in (0, 1)$. Following Wang et al. [2015], we draw i.i.d. innovations $V_1, V_2, \ldots, V_n$, from a scale mixture of two independent multivariate Gaussian distributions, i.e.,

$$V_i = \nu_i \times \epsilon_i + 3(1-\nu_i) \times \epsilon_i',$$

where ν_i, $i \in \mathbb{N}$, are i.i.d. bernoulli random variables with $\mathbb{P}(\nu_i = 1) = 0.9$ and $\mathbb{P}(\nu_i = 0) = 0.1$. A simulation study is given in Section 3.5 by setting $\theta = 0.3$ and 0.7. We report in Figure 3.1 the average values of $|\hat{f}/f-1|$ for $\hat{f}_{\mathcal{A}}$, $\hat{f}_{\text{o}}$ and $\hat{f}_S$, based on 1000 replications with the numerical setup $(n,p,m)=(100,20,10)$ and $d=200,400,800,1000,1200$. For both cases of θ, $|\hat{f}_{\mathcal{A}}/f-1|$ and $|\hat{f}_o/f-1|$ are very close to 0, while $|\hat{f}_S/f-1|$ is quite large. More precisely, we can derive that $\|\mathbb{E}_0(V_1^\top V_1)\|^2 \approx (18+d)f^2$.

Figure 3.1: Empirical averages of the values of $|\hat{f}/f-1|$



Replacing f by the ratio-consistent estimator $\hat{f}_{\mathcal{A}}$ in Theorem 3.2.8, we establish the following central limit theorem.

Corollary 3.4.2. *Under the conditions of Theorem 3.2.8 and Lemma 3.4.1,*

$$\frac{Q_n^\diamond}{\hat{f}_{\mathcal{A}}\sqrt{m_\diamond}} \Rightarrow N(0,2).$$

The proof of Corollary 3.4.2 is straightforward and thus omitted. For any $\alpha \in (0,1)$, let Z_α be the $(1-\alpha)$th quantile of the standard normal distribution. Based on Corollary 3.4.2, an α level test is given by

$$\Phi_Z = \mathbb{I}\left\{\frac{Q_n^\diamond}{\hat{f}_{\mathcal{A}}\sqrt{2m_\diamond}} > Z_\alpha\right\}. \tag{3.16}$$

The null hypothesis H_0 is rejected whenever $\Phi_Z = 1$.

3.5 A simulation study

In this section, we conduct a Monte Carlo simulation study to assess the finite sample performance of the proposed tests. In the model (3.1), we write $X_i = (1, \mathbf{x}_i^\top)^\top$ to include an intercept. Here $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{R}^{p-1}$ are i.i.d. $N(0, I_{p-1})$ random vectors. Let $m < p$. For $1 \le k \le p - m$, all entries of the coefficient vector B_k are i.i.d. uniform random variables in the interval $(1, 2)$. After those B_k's are generated, we keep their values throughout the simulation. Our goal is to identify the zero B_k's by testing the hypothesis

$$H_0 : B_{p-m+1} = B_{p-m+2} = \cdots = B_p = 0.$$

In our simulation, we set $(p, m) = (20, 10)$, $n = 100, 200$ and $d = 400, 800, 1200$. We consider two designs of the innovations (V_i): the one introduced in Example 3.4.1 and the one in Example 3.5.1 below. In both examples, the parameter θ is set to be 0.3 and 0.7.

Example 3.5.1. Let ξ_{ij}, $i, j \in \mathbb{N}$, be i.i.d. random variables with $\mathbb{E}(\xi_{11}) = 0$ and $\mathrm{var}(\xi_{11}) = 1$. In particular, we consider two cases for the ξ_{ij}'s; they are drawn from the standardized t_5 distribution and the standardized χ_5^2 distribution, respectively. For some $\theta \in (0, 1)$, we generate

$$V_i = \sqrt{1-\theta} \times \xi_i + \sqrt{\theta} \times (\xi_{i0}, \xi_{i0}, \ldots, \xi_{i0})^\top, \;\; i \in \mathbb{N}.$$

We shall apply a Gaussian multiplier bootstrap (GMB) approach to implement our proposed test. The procedure is as follows.

1. Compute the residual matrix $\hat{V} = (\hat{V}_1, \ldots, \hat{V}_n)^\top = \bar{P}_1 Y$. Generate i.i.d. $N(0, 1)$ random variables ω_{ij}, $i, j \in \mathbb{N}$, and compute the bootstrap residuals $V^\star = (V_1^\star, V_2^\star, \ldots, V_n^\star)^\top$ by the Gaussian multiplier, where

$$V_i^\star = \frac{1}{\sqrt{n-p}} \sum_{j=1}^{n} \omega_{ij} \hat{V}_i, \;\; i = 1, 2, \ldots, n.$$

2. Use $V^\star$ to compute $\hat{f}^\star_{\mathcal{A}}$ via (3.15) and the bootstrap test statistic $Q_n^{\diamond\star} = \mathrm{tr}(V^{\star\top} P_\diamond V^\star)$.

3. Repeat the first two steps independently for $\mathcal{B}$ times and collect $Q^{\diamond\star}_{n,k}$ and $\hat{f}^\star_{\mathcal{A},k}$, $k = 1, 2, \ldots, \mathcal{B}$.

4. Let α be a nominal level and let $\hat{C}_\alpha$ be the $(1-\alpha)$th quantile of the sequence $\{Q^{\diamond\star}_{n,k}/(\hat{f}^\star_{\mathcal{A},k}\sqrt{2m_\diamond})\}_{k\le\mathcal{B}}$. Then our test is given by

$$\Phi_B = \mathbb{I}\left\{\frac{Q_n^\diamond}{\hat{f}_{\mathcal{A}}\sqrt{2m_\diamond}} > \hat{C}_\alpha\right\}, \tag{3.17}$$

   and we shall reject the null hypothesis whenever $\Phi_B = 1$.

In our simulation, we set the bootstrap size $\mathcal{B} = 1000$. As comparison, we also perform the test suggested in (3.16) based on the central limit theorem and the one proposed in Srivastava and Kubokawa [2013] which we denote by SK.

3.5.1 *Size accuracy*

We first evaluate the size accuracy by using the three tests. For each case, we report the empirical size based on 2000 replications as displayed in Table 3.1 and Table 3.2. The results suggest that our proposed test by using the bootstrap procedure provides the best size accuracy in general as the empirical sizes are close to the nominal level α.

Table 3.1: Empirical sizes for Example 3.4.1 with $\alpha = 0.05$

		$\theta = 0.3$			$\theta = 0.7$		
n	d	CLT	GMB	SK	CLT	GMB	SK
100	400	0.057	0.047	0.041	0.059	0.051	0.036
	800	0.049	0.045	0.033	0.063	0.056	0.026
	1200	0.062	0.055	0.021	0.048	0.045	0.028
200	400	0.056	0.052	0.042	0.052	0.047	0.037
	800	0.052	0.049	0.037	0.053	0.050	0.033
	1200	0.045	0.044	0.029	0.050	0.046	0.035

For Example 3.4.1, both of the test by CLT and our Gaussian multiplier bootstrap method have better performance than the SK test since the latter is too conservative as d is large.

As expected from our theoretical results, normal approximation can work reasonably well in this design.

Table 3.2: Empirical sizes for Example 3.5.1 with $\alpha = 0.05$

			t_5			χ^2_5		
θ	n	d	CLT	GMB	SK	CLT	GMB	SK
0.3	100	400	0.068	0.058	0.023	0.083	0.065	0.036
		800	0.082	0.066	0.023	0.074	0.058	0.016
		1200	0.082	0.068	0.015	0.067	0.053	0.011
	200	400	0.073	0.059	0.022	0.067	0.054	0.018
		800	0.071	0.057	0.012	0.074	0.058	0.014
		1200	0.076	0.059	0.011	0.077	0.058	0.011
0.7	100	400	0.074	0.055	0.002	0.082	0.062	0.002
		800	0.084	0.066	0.001	0.085	0.071	0.000
		1200	0.073	0.057	0.000	0.076	0.062	0.001
	200	400	0.083	0.067	0.001	0.080	0.064	0.000
		800	0.068	0.050	0.000	0.075	0.062	0.000
		1200	0.070	0.051	0.001	0.074	0.056	0.000

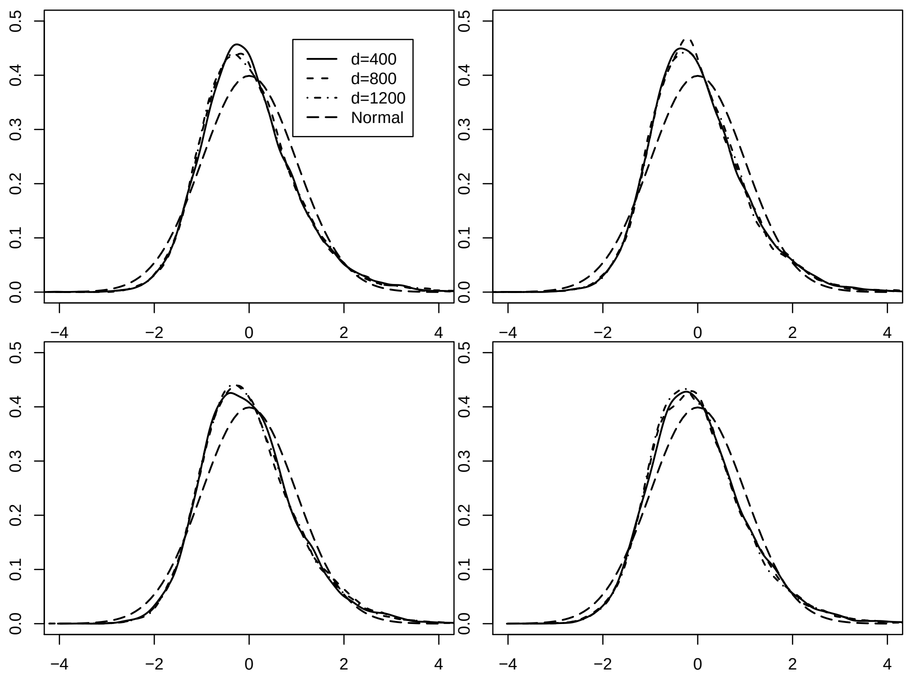
For Example 3.5.1, the Gaussian multiplier bootstrap method outperforms other two procedures in size accuracy for all cases. The SK test suffers from size distortion. The test by CLT inflates the size more than the GMB method, which can be explained by Theorem 3.2.8. More specifically, for both $\theta = 0.3$ and $\theta = 0.7$, elementary calculations show that $\lambda_1/f \to 1$. As a result, the condition (3.13) is violated as $m = 10$; see also Remark 12 for some discussion on the non-normality of $Q_n^\diamond$. To have more insight, we display in Figure 3.2 the density plots of $Q_n^\diamond/\sqrt{\mathrm{var}(Q_n^\diamond)}$ for $n = 100$ as well as the density of $N(0,1)$. As we can see from the plots, the distribution of $Q_n^\diamond/\sqrt{\mathrm{var}(Q_n^\diamond)}$ is skewed to the right for all cases, which explains the inflated sizes of the CLT test.

3.5.2 Power comparison

We conduct a power analysis for the three tests with the following alternative hypothesis

$$H_1 : B_{p-m+2} = B_{p-m+3} = \cdots = B_p = 0 \text{ and } B_{p-m+1} = (\delta, \ldots, \delta)^\top,$$

Figure 3.2: Density plots of $Q_n^\diamond/\sqrt{\mathrm{var}(Q_n^\diamond)}$ and $N(0,1)$



for some $\delta > 0$. Table 3.3 and Table 3.4 present empirical powers for some cases of δ as examples. We observe that our proposed test and the test using CLT have competing powers, while the SK test suffers from low power in all cases. We also remark that the test based on CLT inflates the size more though it can induce similar power as our test.

3.6 Proofs

In this section, we provide detailed proofs for the results presented in the previous sections.

Lemma 3.6.1. *For any $\varepsilon > 0$, we have*

$$\mathcal{L}(G_n^\diamond, \varepsilon) \leq 8\left(\frac{\sum_{i=1}^n \theta_i P_{ii}}{\pi^2 m}\right)^{1/5} + \sqrt{\frac{4\varepsilon}{\pi m^{1/2} f}}.$$

Table 3.3: Empirical powers for Example 3.4.1 with $\alpha = 0.05$

			$\theta = 0.3$			$\theta = 0.7$		
δ	n	d	CLT	GMB	SK	CLT	GMB	SK
0.050	100	400	0.144	0.128	0.059	0.089	0.084	0.038
		800	0.172	0.163	0.039	0.148	0.136	0.043
	200	400	0.281	0.269	0.097	0.179	0.171	0.100
		800	0.444	0.430	0.106	0.260	0.244	0.101
0.075	100	400	0.338	0.312	0.094	0.173	0.151	0.083
		800	0.471	0.454	0.078	0.266	0.243	0.055
	200	400	0.723	0.718	0.285	0.462	0.449	0.236
		800	0.930	0.923	0.303	0.653	0.640	0.258
0.100	100	400	0.631	0.614	0.169	0.355	0.336	0.139
		800	0.829	0.816	0.136	0.572	0.552	0.132
	200	400	0.977	0.974	0.643	0.814	0.803	0.486
		800	1.000	1.000	0.655	0.968	0.966	0.569

Table 3.4: Empirical powers for Example 3.5.1 with $\alpha = 0.05$

				t_5			χ_5^2		
δ	θ	n	d	CLT	GMB	SK	CLT	GMB	SK
0.15	0.3	100	400	0.417	0.377	0.217	0.371	0.320	0.214
			800	0.395	0.351	0.178	0.401	0.357	0.180
		200	400	0.744	0.702	0.540	0.741	0.707	0.532
			800	0.762	0.725	0.483	0.769	0.731	0.480
	0.7	100	400	0.194	0.163	0.009	0.213	0.175	0.015
			800	0.210	0.179	0.006	0.191	0.161	0.003
		200	400	0.391	0.344	0.028	0.387	0.333	0.020
			800	0.395	0.353	0.015	0.387	0.340	0.013
0.20	0.3	100	400	0.655	0.618	0.475	0.654	0.611	0.447
			800	0.665	0.633	0.390	0.626	0.576	0.403
		200	400	0.939	0.932	0.835	0.954	0.947	0.866
			800	0.971	0.965	0.840	0.957	0.943	0.817
	0.7	100	400	0.345	0.299	0.034	0.344	0.295	0.025
			800	0.349	0.313	0.015	0.330	0.286	0.015
		200	400	0.664	0.624	0.094	0.629	0.587	0.097
			800	0.656	0.608	0.054	0.619	0.562	0.035
0.25	0.3	100	400	0.861	0.836	0.706	0.849	0.827	0.684
			800	0.872	0.849	0.646	0.842	0.803	0.629
		200	400	0.994	0.993	0.978	0.997	0.994	0.976
			800	0.994	0.994	0.960	0.996	0.994	0.966
	0.7	100	400	0.484	0.423	0.055	0.465	0.420	0.070
			800	0.515	0.457	0.044	0.501	0.455	0.038
		200	400	0.835	0.807	0.256	0.837	0.805	0.236
			800	0.854	0.829	0.168	0.821	0.794	0.133

Proof of Lemma 3.6.1. By Lemma 7.2 in Xu et al. [2014], for any $\varepsilon > 0$,

$$\mathcal{L}(G_n, \varepsilon) \leq \sqrt{\frac{4\varepsilon}{\pi m^{1/2} f}}.$$

Note that $|\bar{P}_0 D_\theta \bar{P}_0|_{\mathbb{F}}^2 = \sum_{i=1}^n \theta_i P_{ii}$, hence

$$\|\mathbb{E}_0(G_n - G_n^\diamond)\|^2 = 2|\bar{P}_0 D_\theta \bar{P}_0|_{\mathbb{F}}^2 f^2 = 2\sum_{i=1}^n \theta_i P_{ii} f^2.$$

Consequently, it follows that

$$\rho\{\mathbb{E}_0(G_n), G_n^\diamond\} \leq 4\left(\frac{\sum_{i=1}^n \theta_i P_{ii}}{\pi^2 m}\right)^{1/5}.$$

Therefore

$$\mathcal{L}(G_n^\diamond, \varepsilon) \leq 2\rho\{\mathbb{E}_0(G_n), G_n^\diamond\} + \mathcal{L}(G_n, \varepsilon) \leq 8\left(\frac{\sum_{i=1}^n \theta_i P_{ii}}{\pi^2 m}\right)^{1/5} + \sqrt{\frac{4\varepsilon}{\pi m^{1/2} f}}.$$

□

Proof of Lemma 3.2.2. By Jensen's inequality,

$$\mathbb{E}|V_1^\top \Sigma V_1|^{q/2} \leq \mathbb{E}|V_1^\top \mathbb{E}(V_2 V_2^\top) V_1|^{q/2} \leq \mathbb{E}|V_1^\top V_2|^q. \tag{3.18}$$

Hence $L_q \leq M_q$. Since P is positive semi-definite and $\text{rank}(P) = m$, it follows that $\text{tr}(P) = |P|_{\mathbb{F}}^2 = m$ and $\max_{i \neq j} |P_{ij}| \leq \max_{i \leq n} |P_{ii}|$. Consequently, we have

$$\sum_{i \neq j} |P_{ij}|^q \leq m \max_{i \leq n} |P_{ii}|^\delta \text{ and } \sum_{i=1}^n P_{ii}^{q/2} \leq m \max_{i \leq n} |P_{ii}|^{\delta/2}.$$

Combined with (3.18), it follows that

$$\Delta_q \leq 2 \max_{i \leq n} |P_{ii}|^{\delta/2} m^{-\delta/2} M_q.$$

$\square$

3.6.1 Proofs of Lemmas 3.2.4 and 3.2.7

Proof of Lemma 3.2.4. For simplicity of notation, let $\varphi = (P_{11}, P_{22}, \ldots, P_{nn})^\top$ and $A = \bar{P}_0 \circ \bar{P}_0$. Let D_A denote the diagonal matrix of A. Since P_0 is a projection matrix, it follows that by (3.12)

$$|I_n - D_A^{-1} A|_\infty = \max_{i \leq n} \frac{\sum_{j \neq i} |P_{ij,0}|^2}{(1 - P_{ii,0})^2} = \max_{i \leq n} \frac{P_{ii,0}}{1 - P_{ii,0}} \leq \frac{\varpi_0}{1 - \varpi_0}.$$

Hence

$$\begin{aligned} |A^{-1}|_\infty &\leq |A^{-1} - D_A^{-1}|_\infty + |D_A^{-1}|_\infty \leq |I_n - D_A^{-1} A|_\infty |A^{-1}|_\infty + |D_A^{-1}|_\infty \\ &\leq \frac{\varpi_0}{1 - \varpi_0} |A^{-1}|_\infty + \frac{1}{(1 - \varpi_0)^2}, \end{aligned}$$

which implies that

$$|A^{-1}|_\infty \leq \frac{1}{(1 - 2\varpi_0)(1 - \varpi_0)}.$$

Consequently, it follows that

$$|\theta|_\infty = |A^{-1} \varphi|_\infty \leq |A^{-1}|_\infty |\varphi|_\infty \leq \frac{\max_{i \leq n} |P_{ii}|}{(1 - 2\varpi_0)(1 - \varpi_0)} \tag{3.19}$$

and

$$\sum_{i=1}^{n} \theta_i P_{ii} = \varphi^\top \theta = \varphi^\top A^{-1} \varphi \leq \frac{\sum_{i=1}^{n} P_{ii}^2}{(1-2\varpi_0)(1-\varpi_0)}.$$

□

Proof of Lemma 3.2.7. Recall that

$$P_{ij,\diamond} = P_{ij} + (\theta_i + \theta_j) P_{ij,0} - \sum_{k=1}^{n} \theta_k P_{ik,0} P_{jk,0}.$$

Hence, by (3.12), (3.19) and the triangle inequality,

$$\max_{i \neq j} |P_{ij,\diamond}| \leq \max_{i \leq n} |P_{ii}| + 3 \max_{i \leq n} |\theta_i| \max_{i \leq n} |P_{ii,0}| \leq \frac{2 \max_{i \leq n} |P_{ii}|}{(1-2\varpi_0)(1-\varpi_0)}.$$

Together with the fact that $|P_\diamond|_{\mathbb{F}}^2 = m - \sum_{i=1}^{n} \theta_i P_{ii} < m$, it follows that

$$\sum_{i \neq j} |P_{ij,\diamond}|^q \leq m \max_{i \neq j} |P_{ij,\diamond}|^\delta \lesssim m \max_{i \leq n} |P_{ii}|^\delta. \tag{3.20}$$

For simplicity of notation, let $\bar{P}_{k,0}$ denote the k-th column of $\bar{P}_0$. Hence $P_{il,\diamond} = P_{il} - \bar{P}_{i,0}^\top D_\theta \bar{P}_{l,0}$ and

$$\begin{aligned} \sum_{i=1}^{n} |P_{il,\diamond}|^2 &= P_{ll} - 2 \sum_{i=1}^{n} P_{il} \bar{P}_{i,0}^\top D_\theta \bar{P}_{l,0} + \bar{P}_{l,0}^\top D_\theta \bar{P}_0 D_\theta \bar{P}_{l,0} \\ &=: P_{ll} - 2\Delta_{1,l} + \Delta_{2,l}. \end{aligned}$$

Since $\bar{P}_0$ is a projection matrix, it follows that

$$\max_{l \leq n} \Delta_{2,l} \leq \max_{l \leq n} |D_\theta \bar{P}_{l,0}|^2 \leq |\theta|_\infty^2 \max_{l \leq n} |\bar{P}_{l,0}|^2 < |\theta|_\infty^2$$

and

$$\max_{l\le n}|\Delta_{1,l}| \le |\theta|_\infty|\varphi|_\infty + |\theta|_\infty|\varphi|_\infty^{1/2}\max_{i\le n}|P_{ii,0}|^{1/2} \le 2|\theta|_\infty.$$

Therefore $\max_{l\le n}\sum_{i=1}^n |P_{il,\diamond}|^2 \lesssim \max_{i\le n}|P_{ii}|$. Together with (3.20) and $L_q \le M_q$, it follows that

$$\Delta_{q,\diamond} \lesssim \max_{i\le n}|P_{ii}|^{\delta/2}m^{-\delta/2}M_q.$$

□

3.6.2 Proofs of Theorems 3.2.3 and 3.2.6

Let $g(x) = (1-\min(1,\max(x,0))^4)^4$ and $g_{\psi,t}(x) = g(\psi(x-t))$, where $\psi > 0$. Hence, for any $\psi > 0$,

$$\mathbb{I}\{x \le t\} \le g_{\psi,t}(x) \le \mathbb{I}\{x \le t+\psi^{-1}\}. \tag{3.21}$$

Define $c_g = \sup_{x\in\mathbb{R}}(|g'(x)| + |g''(x)| + |g'''(x)|) < \infty$. Since $g(x)$ is three times continuously differentiable, we have

$$\sup_{x\in\mathbb{R}}|g'_{\psi,t}(x)| \le c_g\psi,\ \ \sup_{x\in\mathbb{R}}|g''_{\psi,t}(x)| \le c_g\psi^2 \ \text{ and } \ \sup_{x\in\mathbb{R}}|g'''_{\psi,t}(x)| \le c_g\psi^3. \tag{3.22}$$

Lemma 3.6.2. *Let $\Delta_{q,\diamond}$ be defined in Theorem 3.2.6. Then*

$$\sup_{t\in\mathbb{R}}|\mathbb{E}\{g_{\psi,t}(Q_n^\diamond)\} - \mathbb{E}\{g_{\psi,t}(G_n^\diamond)\}| \le C_q\Delta_{q,\diamond}\psi^q f^q m^{q/2}. \tag{3.23}$$

Proof of Lemma 3.6.2. For each $1 \le l \le n$, define $\mathbf{w}_l = \sum_{i<l}P_{il,\diamond}V_i + \sum_{i>l}P_{il,\diamond}Z_i$,

$$\mathcal{K}_l = 2V_l^\top\mathbf{w}_l,\ \ \tilde{\mathcal{K}}_l = 2Z_l^\top\mathbf{w}_l \ \text{ and } \ W_l = q(V_1,\ldots,V_{l-1},0,Z_{l+1},\ldots,Z_n),$$

where $q(\nu_1,\ldots,\nu_n)=\sum_{i\neq j}P_{ij,\diamond}\nu_i^\top\nu_j$. Decompose

$$g_{\psi,t}(W_l+\mathcal{K}_l)-g_{\psi,t}(W_l+\tilde{\mathcal{K}}_l)=R_1+R_2+R_3,$$

where $R_1=g'_{\psi,t}(W_l)(\mathcal{K}_l-\tilde{\mathcal{K}}_l)$, $R_2=g''_{\psi,t}(W_l)(\mathcal{K}_l^2-\tilde{\mathcal{K}}_l^2)/2$ and

$$R_3=g_{\psi,t}(W_l+\mathcal{K}_l)-g_{\psi,t}(W_l+\tilde{\mathcal{K}}_l)-R_1-R_2.$$

Since $\mathrm{cov}(V_l)=\mathrm{cov}(Z_l)$ and $\{W_l,\mathbf{w}_l\}$ is independent of $\{V_l,Z_l\}$, it follows that $\mathbb{E}(R_1)=0$ and $\mathbb{E}(R_2)=0$. By (3.22) and the fact that $0\le g_{\psi,t}(x)\le 1$,

$$\mathbb{E}|R_3|\le C_q\psi^q\mathbb{E}|\mathcal{K}_l|^q+C_q\psi^q\mathbb{E}|\tilde{\mathcal{K}}_l|^q. \tag{3.24}$$

By Rosenthal's inequality,

$$\mathbb{E}\Big|\sum_{i<l}P_{il,\diamond}V_i^\top V_l\Big|^q\le C_q\sum_{i<l}|P_{il,\diamond}|^q\mathbb{E}|V_i^\top V_l|^q+C_q\mathbb{E}\Big|\sum_{i<l}P_{il,\diamond}^2V_l^\top\Sigma V_l\Big|^{q/2}$$

and

$$\begin{aligned}\mathbb{E}\Big|\sum_{i>l}P_{il,\diamond}Z_i^\top V_l\Big|^q&\le C_q\sum_{i>l}|P_{il,\diamond}|^q\mathbb{E}|Z_i^\top V_l|^q+C_q\mathbb{E}\Big|\sum_{i>l}P_{il,\diamond}^2V_l^\top\Sigma V_l\Big|^{q/2}\\&\le C_q\Big|\sum_{i>l}P_{il,\diamond}^2\Big|^{q/2}\mathbb{E}|V_1^\top\Sigma V_1|^{q/2}.\end{aligned}$$

Consequently, it follows that

$$\begin{aligned}\mathbb{E}|\mathcal{K}_l|^q&\le C_q\sum_{i<l}|P_{il,\diamond}|^q\mathbb{E}|V_1^\top V_2|^q+C_q\Big|\sum_{i\neq l}P_{il,\diamond}^2\Big|^{q/2}\mathbb{E}|V_1^\top\Sigma V_1|^{q/2}\\,\mathbb{E}|\tilde{\mathcal{K}}_l|^q&\le C_q\Big|\sum_{i\neq l}P_{il,\diamond}^2\Big|^{q/2}\mathbb{E}|V_1^\top\Sigma V_1|^{q/2}.\end{aligned} \tag{3.25}$$

Then, (3.23) follows from (3.25) in view of

$$\sup_{t\in\mathbb{R}} |\mathbb{E}\{g_{\psi,t}(Q_n^\diamond)\} - \mathbb{E}\{g_{\psi,t}(G_n^\diamond)\}| \le \sum_{l=1}^n \sup_{t\in\mathbb{R}} |\mathbb{E}\{g_{\psi,t}(W_l + \mathcal{K}_l)\} - \mathbb{E}\{g_{\psi,t}(W_l + \tilde{\mathcal{K}}_l)\}|.$$

□

Proof of Theorem 3.2.6. By (3.21) and Lemma 3.6.2, it follows that

$$\begin{aligned}\rho(Q_n^\diamond, G_n^\diamond) &\le \inf_{\psi>0}\left\{\sup_{t\in\mathbb{R}} |\mathbb{E}\{g_{\psi,t}(Q_n^\diamond)\} - \mathbb{E}\{g_{\psi,t}(G_n^\diamond)\}| + \mathcal{L}(G_n^\diamond, \psi^{-1})\right\} \\ &\le \inf_{\psi>0}\left\{C_q\Delta_{q,\diamond}\psi^q f^q m^{q/2} + 8\left(\frac{\sum_{i=1}^n \theta_i P_{ii}}{\pi^2 m}\right)^{1/5} + \sqrt{\frac{4}{\pi m^{1/2} f\psi}}\right\} \\ &= C_q\Delta_{q,\diamond}^{1/(2q+1)} + 8\left(\frac{\sum_{i=1}^n \theta_i P_{ii}}{\pi^2 m}\right)^{1/5} \\ &\le C_q\Delta_{q,\diamond}^{1/(2q+1)} + C\left(\frac{\sum_{i=1}^n P_{ii}^2}{m}\right)^{1/5}.\end{aligned}$$

□

Proof of Theorem 3.2.3. Let $G_n^\star = \sum_{i\ne j} P_{ij} Z_i^\top Z_j$. Similar as Lemma 3.6.1, it follows that

$$\rho\{G_n^\star, \mathbb{E}_0(G_n)\} \le 4\left(\frac{\sum_{i=1}^n P_{ii}^2}{\pi^2 m}\right)^{1/5}$$

and for any $\varepsilon > 0$,

$$\mathcal{L}(G_n^\star, \varepsilon) \le 8\left(\frac{\sum_{i=1}^n P_{ii}^2}{\pi^2 m}\right)^{1/5} + \sqrt{\frac{4\varepsilon}{\pi m^{1/2} f}}.$$

Case 1: $\Lambda_n \to 0$.

By similar argument as that of the proof of Theorem 3.2.6, we have

$$\rho(Q_n^\star, G_n^\star) \le C_q\Delta_q^{1/(2q+1)} + 8\left(\frac{\sum_{i=1}^n P_{ii}^2}{\pi^2 m}\right)^{1/5}.$$

Note that $\|\mathbb{E}_0(D_n)\|^2 = \sum_{i=1}^n P_{ii}^2 \|\mathbb{E}_0(V_1^\top V_1)\|^2$. Hence, by the Markov inequality,

$$\begin{aligned}\rho\{\mathbb{E}_0(Q_n), Q_n^\star\} &\le \inf_{\varepsilon>0}\left\{\frac{\|\mathbb{E}_0(D_n)\|^2}{\varepsilon^2} + \mathcal{L}(G_n^\star, \varepsilon)\right\} + 2\rho(Q_n^\star, G_n^\star) \\ &\le C_1\Lambda_n^{2/5} + C_q\Delta_q^{1/(2q+1)} + C_2\left(\frac{\sum_{i=1}^n P_{ii}^2}{m}\right)^{1/5}.\end{aligned}$$

Consequently,

$$\begin{aligned}\rho(Q_n, G_n) &\le \rho\{\mathbb{E}_0(Q_n), Q_n^\star\} + \rho(Q_n^\star, G_n^\star) + \rho\{G_n^\star, \mathbb{E}_0(G_n)\} \\ &\le C_1\Lambda_n^{2/5} + C_q\Delta_q^{1/(2q+1)} + C_2\left(\frac{\sum_{i=1}^n P_{ii}^2}{m}\right)^{1/5}.\end{aligned}$$

Case 2: $\Lambda_n \to \infty$.

By the Lindeberg-Feller Central Limit Theorem,

$$\frac{Q_n - m\mathrm{tr}(\Sigma) - Q_n^\star}{\Lambda_n f\sqrt{m}} = \sum_{i=1}^n U_i \Rightarrow N(0,1).$$

Note that $\mathbb{E}(Q_n^\star) = 0$ and

$$\frac{\mathrm{var}(Q_n^\star)}{\Lambda_n^2 m f^2} = \frac{2(m - \sum_{i=1}^n P_{ii}^2)f^2}{\Lambda_n^2 m f^2} \le \frac{2}{\Lambda_n^2} \to 0.$$

Hence

$$\frac{Q_n - m\mathrm{tr}(\Sigma)}{\Lambda_n f\sqrt{m}} \Rightarrow N(0,1).$$

□

3.6.3 Proofs of Theorem 3.2.8 and Lemma 3.4.1

Proof of Theorem 3.2.8. By the Lindeberg central limit theorem, $G_n^\diamond/\sqrt{\mathrm{var}(G_n^\diamond)} \Rightarrow N(0,1)$ holds if and only if

$$\frac{\max_{i\le n}|\lambda_i^\diamond|\lambda_1}{f\sqrt{m-\sum_{i=1}^n \theta_i P_{ii}}} \to 0. \tag{3.26}$$

Under the conditions of Theorem 3.2.6, it follows that $\rho(Q_n^\diamond, G_n^\diamond) \to 0$. Hence the central limit theorem $Q_n^\diamond/\sqrt{\mathrm{var}(Q_n^\diamond)} \Rightarrow N(0,1)$ holds if and only if (3.26) is satisfied. Then it suffices to show that (3.26) is equivalent with (3.13). By (3.19),

$$\left|\max_{i\le n}|\lambda_i^\diamond| - 1\right| \le \frac{\max_{i\le n}|P_{ii}|}{(1-2\varpi_0)(1-\varpi_0)} \le \varpi_1.$$

By Lemma 3.2.4 and Assumption 3.2.1,

$$\left|\frac{m-\sum_{i=1}^n \theta_i P_{ii}}{m} - 1\right| \to 0.$$

Therefore (3.26) is equivalent with (3.13). □

Proof of Lemma 3.4.1. Let $2 < q \le 4$. Define $H_{\mathcal{A}} = I_{n/2} - P_{\mathcal{A}}$ and $H_{\mathcal{A}^c} = I_{n/2} - P_{\mathcal{A}^c}$, where

$$P_{\mathcal{A}} = X_{\mathcal{A}}(X_{\mathcal{A}}^\top X_{\mathcal{A}})^{-1}X_{\mathcal{A}}^\top \text{ and } P_{\mathcal{A}^c} = X_{\mathcal{A}^c}(X_{\mathcal{A}^c}^\top X_{\mathcal{A}^c})^{-1}X_{\mathcal{A}^c}.$$

Then $\mathrm{tr}(P_{\mathcal{A}}) = \mathrm{tr}(P_{\mathcal{A}^c}) = p$ and $|P_{\mathcal{A}}|_{\mathbb{F}}^2 = |P_{\mathcal{A}^c}|_{\mathbb{F}}^2 = p$. Let $\Gamma = (n/2-p)^2(\hat{f}_{\mathcal{A}}^2 - f^2)$, which can be decomposed as $\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4$, where

$$\Gamma_1 = \sum_{k\in\mathcal{A}} \sum_{l\in\mathcal{A}^c} H_{\mathcal{A},kk} H_{\mathcal{A}^c,ll} \{(V_k^\top V_l)^2 - f^2\},$$

$$\Gamma_2 = \sum_{k\in\mathcal{A}} \sum_{l\neq l'\in\mathcal{A}^c} H_{\mathcal{A},kk} H_{\mathcal{A}^c,ll'} V_k^\top V_l V_k^\top V_{l'},$$

$$\Gamma_3 = \sum_{k\neq k'\in\mathcal{A}} \sum_{l\in\mathcal{A}^c} H_{\mathcal{A},kk'} H_{\mathcal{A}^c,ll} V_k^\top V_l V_{k'}^\top V_l,$$

$$\Gamma_4 = \sum_{k\neq k'\in\mathcal{A}} \sum_{l\neq l'\in\mathcal{A}^c} H_{\mathcal{A},kk'} H_{\mathcal{A}^c,ll'} V_{k'}^\top V_l V_k^\top V_{l'}.$$

Decompose

$$\begin{aligned}\Gamma_1 &= \sum_{k\in\mathcal{A}} \sum_{l\in\mathcal{A}^c} H_{\mathcal{A},kk} H_{\mathcal{A}^c,ll} \{(V_k^\top V_l)^2 - V_l^\top \Sigma V_l\} + (n/2 - p) \sum_{l\in\mathcal{A}^c} H_{\mathcal{A}^c,ll} (V_l^\top \Sigma V_l - f^2) \\ &= \Gamma_{11} + (n/2 - p)\Gamma_{12}.\end{aligned}$$

By the Burkholder inequality, we have

$$\mathbb{E}|\Gamma_{12}|^{q/2} \leq C_q \sum_{l\in\mathcal{A}^c} |H_{\mathcal{A}^c,ll}|^{q/2} \mathbb{E}|V_1^\top \Sigma V_1 - f^2|^{q/2}$$

and

$$\begin{aligned}\mathbb{E}|\Gamma_{11}|^{q/2} &\leq C_q \sum_{k\in\mathcal{A}} |H_{\mathcal{A},kk}|^{q/2} \mathbb{E}\left|\sum_{l\in\mathcal{A}^c} H_{\mathcal{A}^c,ll} \{(V_k^\top V_l)^2 - V_l^\top \Sigma V_l - V_k^\top \Sigma V_k + f^2\}\right|^{q/2} \\ &\quad + C_q (n/2 - p)^{q/2} \sum_{k\in\mathcal{A}} |H_{\mathcal{A},kk}|^{q/2} \mathbb{E}|V_1^\top \Sigma V_1 - f^2|^{q/2} \\ &\leq C_q \sum_{k\in\mathcal{A}} |H_{\mathcal{A},kk}|^{q/2} \sum_{l\in\mathcal{A}^c} |H_{\mathcal{A}^c,ll}|^{q/2} \left(\mathbb{E}|(V_k^\top V_l)^2 - f^2|^{q/2} + \mathbb{E}|V_1^\top \Sigma V_1 - f^2|^{q/2}\right) \\ &\quad + C_q (n/2 - p)^{q/2} \sum_{k\in\mathcal{A}} |H_{\mathcal{A},kk}|^{q/2} \mathbb{E}|V_1^\top \Sigma V_1 - f^2|^{q/2}.\end{aligned}$$

Consequently, it follows that

$$\mathbb{E}|\Gamma_1|^{q/2} \lesssim n^2\mathbb{E}|(V_1^\top V_2)^2 - f^2|^{q/2} + n^{q/2+1}\mathbb{E}|V_1^\top \Sigma V_1 - f^2|^{q/2}.$$

Similarly, we derive that

$$\mathbb{E}|\Gamma_2|^{q/2} \leq \sum_{l\neq l' \in \mathcal{A}^c} |P_{\mathcal{A}^c,ll'}|^{q/2} \left(n\mathbb{E}|V_1^\top V_2 V_1^\top V_3|^{q/2} + n^{q/2}\mathbb{E}|V_1^\top \Sigma V_2|^{q/2} \right),$$
$$\mathbb{E}|\Gamma_3|^{q/2} \leq \sum_{k\neq k' \in \mathcal{A}} |P_{\mathcal{A},kk'}|^{q/2} \left(n\mathbb{E}|V_1^\top V_2 V_1^\top V_3|^{q/2} + n^{q/2}\mathbb{E}|V_1^\top \Sigma V_2|^{q/2} \right),$$
$$\mathbb{E}|\Gamma_4|^{q/2} \leq \sum_{k\neq k' \in \mathcal{A}} |P_{\mathcal{A},kk'}|^{q/2} \sum_{l\neq l' \in \mathcal{A}^c} |P_{\mathcal{A}^c,ll'}|^{q/2} \left(\mathbb{E}|V_1^\top V_2|^{q/2} \right)^2.$$

Then we have

$$\mathbb{E}\left|\frac{\hat{f}_{\mathcal{A}}^2}{f^2} - 1\right|^{q/2} \leq \frac{\mathbb{E}|\Gamma|^q}{n^q f^q} \leq \frac{(1+p^{q/4}n^{1-q/2})^2 M_q}{n^{q-2}} + \frac{\mathbb{E}|\mathbb{E}_0(V_1^\top \Sigma V_1)|^{q/2}}{n^{q/2-1}f^q} + \frac{p^{q/4}\mathbb{E}|V_1^\top \Sigma V_2|^{q/2}}{n^{q-2}f^q}.$$

When $q=4$, we have $\mathbb{E}|V_1^\top \Sigma V_2|^2 = \mathrm{tr}(\Sigma^4)$,

$$\sum_{k\neq k' \in \mathcal{A}} |P_{\mathcal{A},kk'}|^{q/2} \leq |P_{\mathcal{A}}|_{\mathbb{F}}^2 = p \ \text{ and } \sum_{l\neq l' \in \mathcal{A}^c} |P_{\mathcal{A}^c,ll'}|^{q/2} \leq |P_{\mathcal{A}^c}|_{\mathbb{F}}^2 = p.$$

Consequently, it follows that

$$\left\|\frac{\hat{f}_{\mathcal{A}}^2}{f^2} - 1\right\|^2 \leq \frac{\mathbb{E}|\Gamma|^{q/2}}{n^4 f^4} \leq \frac{M_4}{n^2} + \frac{p\times \mathrm{tr}(\Sigma^4)}{n^2 f^4} + \frac{\|\mathbb{E}_0(V_1^\top \Sigma V_1)\|^2}{nf^4}.$$

$\square$

CHAPTER 4

PORTMANTEAU TEST FOR HIGH DIMENSIONAL WHITE NOISES

4.1 Introduction

Testing for white noise in the time domain is of great importance in diagnostic checking for time series models. It has historically attracted much attention in statistics and econometrics with low dimensional time series data. To fix the idea, let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a sequence of p-dimensional stationary process with mean $\mathbb{E}(\varepsilon_t) = 0$ and covariance matrix $\mathrm{cov}(\varepsilon_t) = \Sigma$. For integer l, let $\Sigma_l = \mathbb{E}(\varepsilon_l \varepsilon_0^\top)$ denote the autocovariance (function) matrix at lag l. We say that (ε_t) is a white noise sequence if $\Sigma_l = 0$ for all $l \neq 0$. In the univariate case when $p = 1$, let $\gamma_l = \mathbb{E}(\varepsilon_l \varepsilon_0)$. To test the hypothesis $H_0 : \gamma_1 = \ldots = \gamma_\nu = 0$ based on the data $\varepsilon_1, \ldots, \varepsilon_n$, where the lag $\nu \geq 1$ is a prescribed integer, Box and Pierce [1970] made a fundamental contribution for proposing the portmanteau test statistic

$$Q_{\mathrm{BP}} = n \sum_{l=1}^{\nu} \hat{\rho}_l^2, \text{ where } \hat{\rho}_l = \frac{\sum_{t=l+1}^{n} \varepsilon_t \varepsilon_{t-l}}{\sum_{t=1}^{n} \varepsilon_t^2}$$

are sample autocorrelations. Ljung and Box [1978] showed that the modified statistic

$$Q_{\mathrm{LB}} = n \sum_{l=1}^{\nu} \frac{n+2}{n-l} \hat{\rho}_l^2$$

has a slightly better finite sample performance. Under the assumption that ε_t are independent and identically distributed, both Q_{BP} and Q_{LB} are asymptotically χ^2_ν, chi-square distribution with ν degrees of freedom. Equipped with the chi-square convergence result, one can perform the Box-Pierce test and the Ljung-Box test easily in practice. Those tests and their variants have become popular choices of testing white noises. When $p > 1$ but fixed, Chitturi [1974] and Hosking [1980] generalized the test statistics Q_{BP} and Q_{LB} to the multivariate setting by introducing respectively

$$Q_{\mathrm{C}} = n\sum_{l=1}^{\nu}\mathrm{tr}(\hat{\Sigma}_l\hat{\Sigma}^{-1}\hat{\Sigma}_l^{\top}\hat{\Sigma}^{-1}) \text{ and } Q_{\mathrm{H}} = n\sum_{l=1}^{\nu}\frac{n}{n-l}\mathrm{tr}(\hat{\Sigma}_l\hat{\Sigma}^{-1}\hat{\Sigma}_l^{\top}\hat{\Sigma}^{-1}), \tag{4.1}$$

where $\hat{\Sigma}_l$ is the sample autocovariance matrix at lag l and $\hat{\Sigma} = \hat{\Sigma}_0$ is the sample covariance matrix. When ε_t are independent and identically distributed random vectors, both Q_{C} and Q_{H} are also asymptotically $\chi^2_{p^2\nu}$. See Tsay [2014] and Lütkepohl [2005] for a comprehensive overview of multivariate white noise test.

High dimensional time series data are increasingly encountered in a wide range of disciplines. As pointed out in Li et al. [2018], the aforementioned conventional multivariate portmanteau tests are no longer reliable or even not well-defined in the high dimensional regime where the dimension p is comparable or even larger than the sample size n. Testing white noise for high dimensional data is becoming a crucial problem in various applications. Under the so-called Marchenko-Pastur regime where $p/n \to c \in (0, \infty)$, Li et al. [2018] proposed a portmanteau-type test statistic for linear processes concerning the sum of the squared singular values of the first ν lagged sample autocovariance matrices and established the asymptotic normality of the test statistic by random matrix theory. As another line of research, Chang et al. [2017] proposed a new omnibus test based on the maximum absolute autocorrelations and cross-correlations. Under suitable conditions on the tail probability and the mixing coefficients, the dimension p can grow exponentially with the sample size n in a way that $\log p = o(n^{\kappa})$ for some constant $\kappa > 0$. Using the extreme value theory, Tsay [2019] proposed a similar test statistic for heavy-tailed time series by investigating the Spearman's rank autocorrelations and derived its asymptotic distribution under the null hypothesis and contemporaneous independence.

In this chapter, we shall develop a new portmanteau-type white noise test for high dimensional processes under a general framework by equivalently testing for the Frobenius norms of autocovariance matrices. Differently from Chang et al. [2017] who detected the most significant one correlation, our test can be more powerful in the sense that we can deal with the cases where the autocovariance matrices consist of many small but non-zero signals that are of similar magnitudes. In particular, we shall propose an asymptotically unbiased estimator of the sum of the squared Frobenius norms of the first ν lagged autocovariance matrices as our test statistic and derive its asymptotic distribution under both the null hypothesis and a specific alternative hypothesis. Our key tool is the high dimensional invariance principle for quadratic forms of martingale differences. In the special case of linear processes, Li et al. [2018] assumed that p is comparable with n and derived a central limit theorem for their test statistic. Their central limit theorem may result in size distortion. In comparison, in our test the dimension p can be arbitrarily large and the approximating distribution is a linear combination of independent χ^2 random variables and the size is generally quite accurate.

4.2 A New Portmanteau Test

Let $Y_{t,l} = \mathrm{vec}(\varepsilon_t \varepsilon_{t-l}^\top)$. Then $\mathbb{E}(Y_{t,l}) = \mathrm{vec}(\Sigma_l)$. Define

$$\Delta_\nu = \sum_{l=1}^{\nu} |\Sigma_l|_{\mathbb{F}}^2 = \sum_{l=1}^{\nu} |\mathrm{vec}(\Sigma_l)|^2,$$

where $|\cdot|_{\mathbb{F}}$ denotes the Frobenius norm. Testing the white noise hypothesis $H_0 : \Sigma_1 = \ldots = \Sigma_\nu = 0$ is equivalent to testing $\Delta_\nu = 0$. Let $\bar{Y}_{n,l} = n^{-1} \sum_{t=1}^{n} Y_{t,l}$, which is an unbiased estimate of $\mathrm{vec}(\Sigma_l)$. A natural portmanteau test statistic would be

$$Q = n \sum_{l=1}^{\nu} \bar{Y}_{n,l}^\top \bar{Y}_{n,l} = \frac{1}{n} \sum_{l=1}^{\nu} \sum_{t,s=1}^{n} Y_{t,l}^\top Y_{s,l},$$

which is related to Q_{C} in (4.1) by removing $\hat{\Sigma}$. In this paper, however, we do not use Q as our test statistic since it is not an unbiased estimator of Δ_ν. Instead, we consider the following modification using the idea of blocking and de-diagonalization. Let $b \geq \nu/3$ be an integer. For technical simplicity, assume that n/b is an integer. Partition the interval $[1, n]$

into consecutive blocks of size b and let $\mathcal{Y}_{k,l} = \sum_{t=(k-1)b+1}^{kb} Y_{t,l}$. We can also write

$$\sum_{t=1}^{n} Y_{t,l} = \sum_{k=1}^{n/b} \mathcal{Y}_{k,l} \text{ and } Q = \frac{1}{n} \sum_{l=1}^{\nu} \sum_{k,k'=1}^{n/b} \mathcal{Y}_{k,l}^{\top} \mathcal{Y}_{k',l}.$$

We consider the following modified test statistic

$$Q^* = \frac{1}{n} \sum_{l=1}^{\nu} \sum_{|k-k'| \geq 2} \mathcal{Y}_{k,l}^{\top} \mathcal{Y}_{k',l} = \frac{2}{n} L_{n,\nu}, \text{ where } L_{n,\nu} = \sum_{l=1}^{\nu} \sum_{k-k' \geq 2} \mathcal{Y}_{k,l}^{\top} \mathcal{Y}_{k',l}.$$

In Q^* we exclude the case $|k - k'| \leq 1$ from Q. Intuitively, when b is large, $\mathcal{Y}_{k,l}$ and $\mathcal{Y}_{k',l}$ are asymptotically independent when $|k - k'| \geq 2$. Hence $\mathbb{E}(\mathcal{Y}_{k,l}^{\top} \mathcal{Y}_{k',l}) \approx \mathbb{E}\mathcal{Y}_{k,l}^{\top} \mathbb{E}\mathcal{Y}_{k',l} = b^2 |\Sigma_l|_{\mathbb{F}}^2$ and $\mathbb{E}(Q^*) \approx (n-3) b^2 \Delta_\nu$. The dependence between $\mathcal{Y}_{k,l}$ and $\mathcal{Y}_{k',l}$ is stronger when $|k - k'| \leq 1$ and $\mathbb{E}(\mathcal{Y}_{k,l}^{\top} \mathcal{Y}_{k',l})$ can be very differently from $b^2 |\Sigma_l|_{\mathbb{F}}^2$. The latter suggests that our modified Q^* can be a better test statistic. One can reject the null white noise hypothesis H_0 if Q^* exceeds certain critical value. To this end, we will need to develop an distributional approximation theory for Q^*. An invariance principle is presented in Section 4.2.1. Unlike Q_C in (4.1) which is asymptotically $\chi_{p^2\nu}^2$ under null hypothesis due to the normalization $\hat{\Sigma}^{-1}$, the approximate distribution of Q^* depends on a high dimensional covariance matrix. Estimation of the latter is highly challenging. To overcome the latter difficulty, in Section 4.2.3 we propose a permutation based approach for practically feasible implementations. Section 4.2.2 provides an approximate formula for the power under the moving average alternative hypothesis.

4.2.1 Asymptotic distribution under the null hypothesis

In this section, we shall present a distributional approximation of $L_{n,\nu}$. Let

$$X_{t,\nu} = (Y_{t,1}^{\top}, Y_{t,2}^{\top}, \ldots, Y_{t,\nu}^{\top})^{\top}.$$

Assume that the block size b is a multiple of 3 and let $m = b/3$. Write

$$L_{n,\nu} = \sum_{k-k' \geq 2} B_{k,\nu}^{\top} B_{k',\nu} \text{ and } B_{k,\nu} = \sum_{t=3(k-1)m+1}^{3km} X_{t,\nu}, \tag{4.2}$$

As in Li et al. [2018] and in Tsay [2019], we study the asymptotic distribution of $L_{n,\nu}$ under the assumption that $(\varepsilon_t)_{t\in\mathbb{Z}}$ are independent and identically distributed, which is a common practice in the literature of white noise tests. Let $(Z_{k,\nu})_{k\in\mathbb{Z}}$ be p-dimensional independent and identically distributed Gaussian random vectors with $\mathbb{E}(Z_{k,\nu}) = 0$ and $\mathrm{cov}(Z_{k,\nu}) = \mathrm{cov}(B_{k,\nu})$. Define the Gaussian analogue of $L_{n,\nu}$ as

$$V_{n,\nu} = \sum_{k-k' \geq 2} Z_{k,\nu}^{\top} Z_{k',\nu}.$$

The invariance principle, Theorem 4.2.1 below, indicates that the distributions of $L_{n,\nu}$ and $V_{n,\nu}$ are asymptotically close. For a vector $v = (v_1, \ldots, v_d)$ let $|v| = (\sum_{i=1}^{d} v_i^2)^{1/2}$.

Theorem 4.2.1. *Let $\vartheta = |\Sigma|_{\mathbb{F}}$. Assume that ε_t are i.i.d., $\mathbb{E}|\varepsilon_t|^q < \infty$, $2 < q \leq 3$, and*

$$(m/n)^{\delta/4} M_q \to 0, \text{ where } M_q = \mathbb{E}\left|\frac{\varepsilon_1^{\top}\varepsilon_2}{\vartheta}\right|^q < \infty \tag{4.3}$$

and $\delta = q - 2$. Then we have

$$\sup_{t\in\mathbb{R}} |\mathbb{P}(L_{n,\nu} \leq t) - \mathbb{P}(V_{n,\nu} \leq t)| \to 0.$$

Note that $M_2 = 1$. The Lyapunov type condition (4.3) is fairly mild. Example 4.2.1 shows that it holds for the widely used linear process models.

Example 4.2.1. Consider the linear process model $\varepsilon_t = (\varepsilon_{t1}, \ldots, \varepsilon_{tp})^{\top}$ with

$$\varepsilon_t = D\eta_t,$$

where D is a $p \times \ell$ matrix, $\ell \geq 1$, $\eta_t = (\eta_{t1}, \ldots, \eta_{t\ell})^\top$ and $(\eta_{tk})_{t,k\in\mathbb{Z}}$ are independent zero-mean random variables having uniformly bounded qth moment

$$\max_{k\leq \ell} E|\eta_{tk}|^q \leq C < \infty.$$

Elementary calculations imply that

$$M_q \leq (1+\delta)^q \max_{k\leq \ell} (E|\eta_{tk}|^q)^2.$$

Consequently, (4.3) is reduced to the natural and mild condition $m/n \to 0$. It is worth mentioning that in this case there is *no restriction on the dimension* p. In a recent work, Li et al. [2018] proposed a portmanteau test under the same linear process model with $\ell = p$ and $q = 4$. They needed the requirement that $p/n \to c \in (0, \infty)$, that is, the dimension p should be comparable with the sample size n, while we can allow any p. □

Remark 18. Note that $V_{n,\nu}$ is a quadratic form of Gaussian vectors. Then it is distributed as a linear combination of independent chi-squared random variables. Theorem 4.2.1 reveals that the asymptotic approximating distribution of $L_{n,\nu}$ is a linear combination of chi-squared random variables. More precisely, let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ be the eigenvalues of Σ. Note that $(X_{t,\nu})_{t\in\mathbb{Z}}$ is a sequence of martingale differences with respect to σ-field generated by $\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$. Hence

$$\operatorname{cov}(Z_{k,\nu}) = \operatorname{cov}(B_{k,\nu}) = 3m(I_\nu \otimes \Sigma \otimes \Sigma),$$

where $\otimes$ denotes the Kronecker product and I_ν denotes the $\nu \times \nu$ identity matrix. Then it follows that as $m/n \to 0$,

$$\rho(2V_{n,\nu}, R) \to 0, \text{ where } R = \sum_{j=1}^{p} \sum_{k=1}^{p} \lambda_j \lambda_k (\chi_{jk,\nu} - \nu). \tag{4.4}$$

Here $(\chi_{jk,\nu})_{j,k\in\mathbb{N}}$ are independent chi-square random variables with ν degrees of freedom. Note that $\mathrm{var}(R) = 2\nu(\sum_{j=1}^{p}\lambda_j^2)^2$. In general $R/(\mathrm{var}(R))^{1/2}$ is not asymptotically $N(0,1)$ unless some assumptions are made on the structure of the covariance matrix Σ. In the following corollary, we derive the sufficient and necessary condition for the asymptotic normality of R. Consequently, we establish the asymptotic normality of the test statistic $L_{n,\nu}$. □

Corollary 4.2.2. *Let λ_1 be the largest eigenvalue of Σ. Then the central limit theorem $R/(\mathrm{var}(R))^{1/2} \Rightarrow N(0,1)$ holds if and only if*

$$\frac{\lambda_1}{\nu^{1/4}\vartheta} \to 0. \tag{4.5}$$

In such case, under the conditions of Theorem 4.2.1, we have

$$\frac{2L_{n,\nu}}{n\nu^{1/2}\vartheta^2} \Rightarrow N(0,2). \tag{4.6}$$

The first part of the Corollary easily follows from the Lindeberg-Feller central limit theorem. The central limit theorem for $L_{n,\nu}$ is an immediate consequence of Theorem 4.2.1.

Remark 19. If $\nu \to \infty$, since $\lambda_1 \le \vartheta$, (4.5) automatically holds and $L_{n,\nu}$ is asymptotically Gaussian. However, if ν is bounded, condition (4.5) is satisfied if $\lambda_1/\vartheta \to 0$, which is a common assumption to ensure the asymptotic normality of high-dimensional quadratic forms, see Bai and Saranadasa [1996]. When (4.5) is violated, the asymptotic distribution of $L_{n,\nu}$ can be non-normal. For example, suppose now ν is finite and $\Sigma_{jk} = \varpi + (1-\varpi)\mathbb{I}\{j=k\}$ for some constant $\varpi \in (0,1)$, then we have that as $p \to \infty$,

$$\frac{2L_{n,\nu}}{n} \Rightarrow \lambda_1^2(\chi_\nu^2 - \nu), \text{ where } \lambda_1 = 1 + (p-1)\varpi.$$

In the extremal case with $\varpi = 1$, we have the complete dependence $\varepsilon_{tj} = \varepsilon_{t1}, 1 \le j \le p$ with $\mathbb{E}(\varepsilon_{t1}^2) = 1$, $Q^*/p^2 \Rightarrow \chi_\nu^2 - \nu$. Our Q^* can be viewed as a variant of the Box-Pierce test statistic by removing diagonal and one-off-diagonal elements in $\sum_{l=1}^{\nu}\hat{\gamma}_l^2$.

4.2.2 Asymptotic distribution under the alternative hypothesis

Now we study the asymptotic distribution of $L_{n,\nu}$ under the moving average alternatives. Let $(\epsilon_t)_{t\in\mathbb{Z}}$ be p-dimensional i.i.d. random vectors with mean $\mathbb{E}(\epsilon_t) = 0$ and covariance matrix $\mathrm{cov}(\epsilon_t) = \Sigma_\epsilon$. We assume that $(\varepsilon_t)_{t\in\mathbb{Z}}$ follow from the vector moving average model

$$H_1 : \varepsilon_t = A\epsilon_{t-1} + \epsilon_t,$$

where A is a $p \times p$ coefficient matrix. Note that $\Sigma_1 = A\Sigma_\epsilon$ and $\Sigma_l = 0$ for $l \geq 2$. As in Li et al. [2018], we study the asymptotic distribution of $L_{n,\nu}$ with $\nu = 1$. To this end, we shall adopt the martingale approximation of the sequence $(X_{t,1})_{t\in\mathbb{Z}}$. More precisely, define

$$D_t = \epsilon_t \otimes \varepsilon_{t-1} + \mathbb{E}_0(A\epsilon_t \otimes \varepsilon_t).$$

Then D_t are stationary martingale differences with respect to the σ-field generated by $\{\epsilon_t, \epsilon_{t-1}, \ldots\}$ and $\mathrm{cov}(D_0)$ is equal to the long run covariance matrix of $(\varepsilon_t \otimes \varepsilon_{t-1})_{t\in\mathbb{Z}}$. We refer to Wu [2007] for more discussions about the martingale approximation for stationary sequence.

Theorem 4.2.3. *Let $\vartheta_\diamond = |\mathrm{cov}(D_0)|_{\mathbb{F}}$ and $m \geq 2$. Assume $\mathbb{E}|\epsilon_t|^{2q} < \infty$ and*

$$(m/n)^{\delta/2}\mathcal{M}_q \to 0, \text{ where } \mathcal{M}_q = \mathbb{E}\left|\frac{D_0^\top D_m}{\vartheta_\diamond}\right|^q, \tag{4.7}$$

$$\frac{\mathrm{vec}(\Sigma_1)^\top\{n\mathrm{cov}(D_0) + d\mathrm{cov}(A\epsilon_0 \otimes \varepsilon_0)\}^\top \mathrm{vec}(\Sigma_1)}{\vartheta_\diamond^2} \to 0, \tag{4.8}$$

$$\frac{\mathbb{E}|D_0^\top(\epsilon_m \otimes \varepsilon_{m-1})|^q}{n^{q/2}\vartheta_\diamond^q} \to 0, \quad \frac{\mathbb{E}|\epsilon_0^\top \epsilon_m|^q \mathbb{E}|\varepsilon_0^\top \varepsilon_m|^q}{(nm)^{q/2}\vartheta_\diamond^q} \to 0. \tag{4.9}$$

Define $\mathcal{V}_n = \sum_{k-k'\geq 2} \mathcal{Z}_k^\top \mathcal{Z}_{k'}$, where $(\mathcal{Z}_k)_{k\in\mathbb{Z}}$ are independent zero-mean Gaussian random *vectors with* $\mathrm{cov}(\mathcal{Z}_k) = 3m\mathrm{cov}(D_0)$*. Then under the alternative hypothesis* H_1*,*

$$\rho\{\mathbb{E}_0(L_{n,1}), \mathcal{V}_n\} \to 0.$$

Remark 20. When $A = 0$, ε_t are independent and identically distributed and consequently Theorem 4.2.3 reduces to Theorem 4.2.1. Condition (4.8) can be viewed as a high dimensional local alternative hypothesis.

4.2.3 Practical implementation

In view of Theorem 4.2.1, R-the approximating distribution of $L_{n,\nu}$ depends on eigenvalues of the unknown covariance matrix $\Sigma = \mathrm{cov}(\varepsilon_1)$. When the dimension p is large, it is highly challenging to estimate Σ or its eigenvalues consistently without imposing structural assumptions. Here we propose an alternative permutation procedure to calculate the p-value of $L_{n,\nu}$, so that we can avoid the estimation of the covariance matrix Σ. The rationale for this procedure is that the permutation operation can weaken the temporal dependence. For simplicity assume that, besides $(\varepsilon_t)_{t=1}^n$, $\varepsilon_{1-\nu}, \ldots, \varepsilon_0$ are also available.

1. Let $\pi = \{\pi(1-\nu), \pi(2-\nu), \ldots, \pi(n)\}$ be a random permutation of $\{1-\nu, 2-\nu, \ldots, n\}$. Calculate the statistic $L_{n,\nu}^{\pi}$ via (4.2) based on the data $\{\varepsilon_{\pi(1-\nu)}, \varepsilon_{\pi(2-\nu)}, \ldots, \varepsilon_{\pi(n)}\}$.

2. Repeat step 1 for B times and obtain the statistics $L_{n,\nu,1}^{\pi}, \ldots, L_{n,\nu,B}^{\pi}$.

3. Calculate the p-value via $B^{-1}\sum_{b=1}^{B} \mathbb{I}\{L_{n,\nu} > L_{n,\nu,b}^{\pi}\}$.

Then, for any prespecified significance level $\alpha \in (0, 1)$, we shall reject the null hypothesis H_0 whenever the p-value is smaller than α. Our simulation study in Section 4.3.1 shows that this procedure has an accurate size and power.

4.3 Simulation Study

4.3.1 Empirical sizes

In this section, we conduct Monte Carlo simulations to assess the finite sample performance of the proposed test in Section 4.2.3. Motivated by Wang et al. [2015], we generate a sequence of independent and identically distributed random vectors ε_t from a scale mixture of two independent distributions:

$$\varepsilon_t = \delta_t \times H\eta_t + 3(1-\delta_t) \times H\eta'_t, \tag{4.10}$$

where $\eta_{tk}, \eta'_{t'k'}, t, k, t', k' \in \mathbb{Z}$, are independent and identically distributed random variables, $(\delta_t)_{t\in\mathbb{Z}}$ are independent Bernoulli random variables with $\mathbb{P}(\delta_t = 1) = \mathbb{P}(\delta_t = 0) = 0.5$, $H \in \mathbb{R}^{p\times p}$ is a lower triangular matrix such that $HH^\top = \Sigma$ with $\Sigma_{ij} = w^{|i-j|}$, $w \in (0,1)$. We consider two different distributions of η_{tk}: standard normal and the standardized t distribution with 4 degrees of freedom. We take the numerical setup: $(\nu, m, d) = (1,1,33)$, $(2,2,17)$; $w = 0.3$ or 0.7 (weak or strong cross-sectional dependence, resp.), and $p = 400, 800$.

At each run, we compute the test statistics in (4.2) with the generated data and then calculate the corresponding p-values via the permutation procedure in Section 4.2.3. The permutation sample size B is set to be 5000. We reject H_0 if the resulting p-value is less than the nominal level.

As a comparison, we also implement the portmanteau test proposed by Li et al. [2018]. More precisely, Li et al. [2018] derived the central limit theorem

$$G_{n,\nu} = \frac{\sum_{l=1}^{\nu} |\hat{\Sigma}_l|_{\mathbb{F}}^2 - \nu n c_{p,n}^2 \hat{s}_1^2}{\sqrt{(2\nu)} c_{p,n}(\hat{s}_2 - c_{p,n}\hat{s}_1^2)} \Rightarrow N(0,1),$$

where $c_{p,n} = p/n$, $\hat{\Sigma}_l = n^{-1}\sum_{t=1}^{n} \varepsilon_t \varepsilon_{t-l}^\top$, and $\hat{s}_1 = p^{-1}\mathrm{tr}(\hat{\Sigma})$ and $\hat{s}_2 = p^{-1}|\hat{\Sigma}|_{\mathbb{F}}^2$ are empirical estimates of the first two spectral moments of the sample covariance matrix $\hat{\Sigma}$. Their test rejects the null hypothesis H_0 if $G_{n,\nu} \geq z_\alpha$, where z_α is the $(1-\alpha)$th quantile of the

Table 4.1: Empirical sizes for tests based $L_{n,\nu}$ and $G_{n,\nu}$

		$N(0,1)$				t_4			
$\alpha = 1\%$									
p	w	$L_{n,1}$	$G_{n,1}$	$L_{n,2}$	$G_{n,2}$	$L_{n,1}$	$G_{n,1}$	$L_{n,2}$	$G_{n,2}$
400	0.3	1.45	28.75	1.30	27.45	1.05	30.70	1.10	27.95
	0.7	1.05	23.60	0.85	20.80	0.70	21.95	0.90	20.15
800	0.3	0.80	29.40	1.05	27.45	0.75	30.40	0.80	29.45
	0.7	0.90	27.50	1.35	26.55	0.95	27.90	1.40	27.90
$\alpha = 5\%$									
400	0.3	5.50	33.50	4.85	30.80	5.75	35.30	5.65	32.45
	0.7	5.50	29.60	4.40	26.95	4.45	27.85	4.30	25.70
800	0.3	5.50	33.35	6.20	31.35	4.95	34.10	4.35	33.20
	0.7	5.05	32.20	5.10	31.10	5.55	32.30	5.40	29.00
$\alpha = 10\%$									
400	0.3	11.20	35.90	10.80	33.70	11.20	37.30	9.80	34.80
	0.7	10.00	32.70	9.95	30.25	9.45	30.70	9.40	28.90
800	0.3	10.45	35.80	10.85	33.75	11.30	36.10	9.40	35.15
	0.7	10.45	34.65	10.10	34.30	10.45	34.40	9.05	32.80

standard normal distribution. Given $\alpha = 0.01$, 0.05 or 0.1, we concern the empirical sizes by calculating the proportions of wrong rejections based on 2000 replications.

Table 4.1 reports the empirical sizes for the test using $L_{n,\nu}$ and $G_{n,\nu}$ respectively. It suggests that our proposed test outperforms the one by Li et al. [2018] in size accuracy for all cases as the empirical sizes by our test are close to the nominal level α, while the latter substantially inflates the size. Their central limit theorem leads to tests which may have size distortions.

4.3.2 Empirical power

Here we shall conduct a power analysis for the tests performed in Section 4.3.1. To calculate the empirical power, we consider two different alternative hypotheses where there exists some serial correlations. Let $(\epsilon_t)_{t\in\mathbb{Z}}$ be independent random vectors generated via (4.10) where η_{kt} follows from the standard normal distribution.

Case 1: Consider the vector moving average model $\varepsilon_t = A\epsilon_{t-1} + \epsilon_t$ with A having entries

$$A_{ij} = \delta_{ij}\theta_{ij} \times \mathbb{I}\{|i-j| \le 1\},$$

where $(\delta_{ij})_{i,j\in\mathbb{N}}$ are independent Bernoulli random variables and $(\theta_{ij})_{i,j\in\mathbb{N}}$ are independent uniform random variables $\mathcal{U}(-v, v)$ with $v = 0.05, 0.10, \ldots, 0.30$.

Case 2: Assume that $(\varepsilon_t)_{t\in\mathbb{Z}}$ follow from the vector autoregressive model

$$\varepsilon_t = A\varepsilon_{t-1} + \epsilon_t.$$

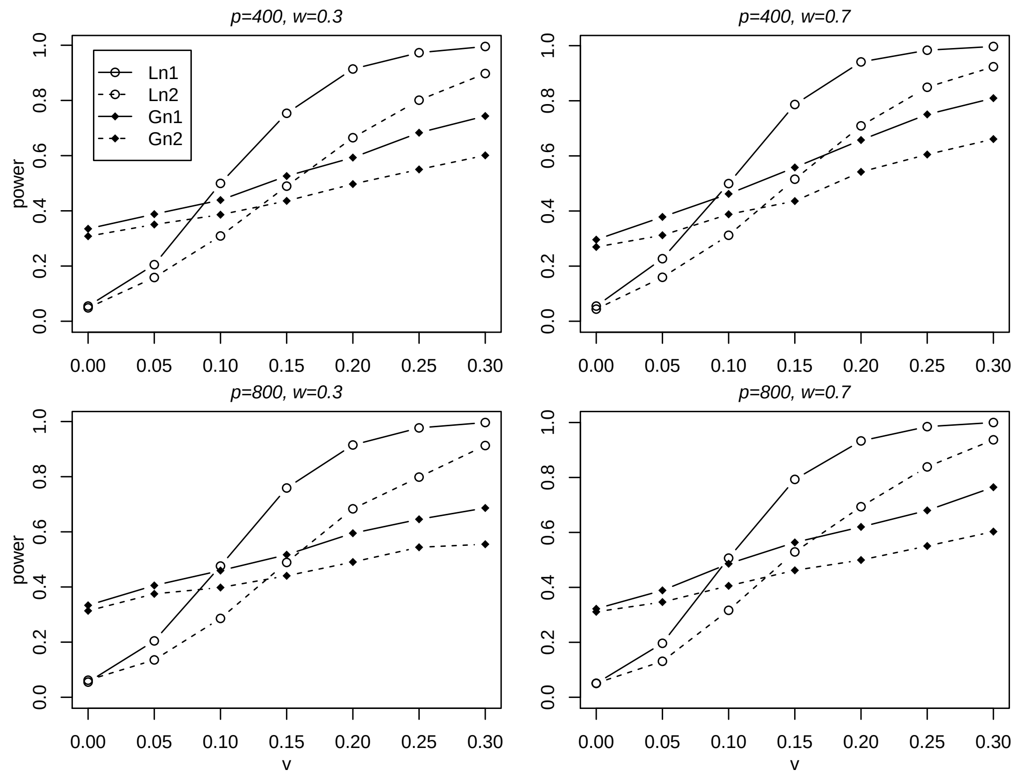
To generate the coefficient matrix A, we first randomly select a subset $\mathcal{I}$ from $[p] = \{1, 2, \ldots, p\}$ with cardinality $|\mathcal{I}| = 20$ and then set

$$A_{ij} = \theta_i \times \mathbb{I}\{i \in \mathcal{I}, j \in \mathcal{I}, i = j\} + \theta_{ij} \times \mathbb{I}\{i \in \mathcal{I}, j \in \mathcal{I}, i < j\},$$

where $(\theta_i)_{i\in\mathbb{N}}$ are independent uniform random variables $\mathcal{U}(-0.5, 0.5)$ and $(\theta_{ij})_{i,j\in\mathbb{N}}$ are the same as in Case 1 with $v = 0.1, 0.2, \ldots, 0.5$.

Given $\alpha = 0.05$ and $\nu = 1$ or 2, we display in Figure 4.1 and Figure 4.2 the empirical power curves for the tests based on $L_{n,\nu}$ and $G_{n,\nu}$ for Case 1 and Case 2, respectively. In Case 1 with the vector moving average model, the empirical power of $L_{n,1}$ is higher than that of $L_{n,2}$. This is quite reasonable, as $\mathbb{E}(L_{n,1}) = \mathbb{E}(L_{n,2})$ and $L_{n,2}$ introduce more randomness. Our test has a good size and power performance, while the test by Li et al. [2018] suffers from size and power inaccuracy.

Figure 4.1: Empirical powers of $L_{n,\nu}$ and $G_{n,\nu}$ for vector moving average model



4.4 Proofs

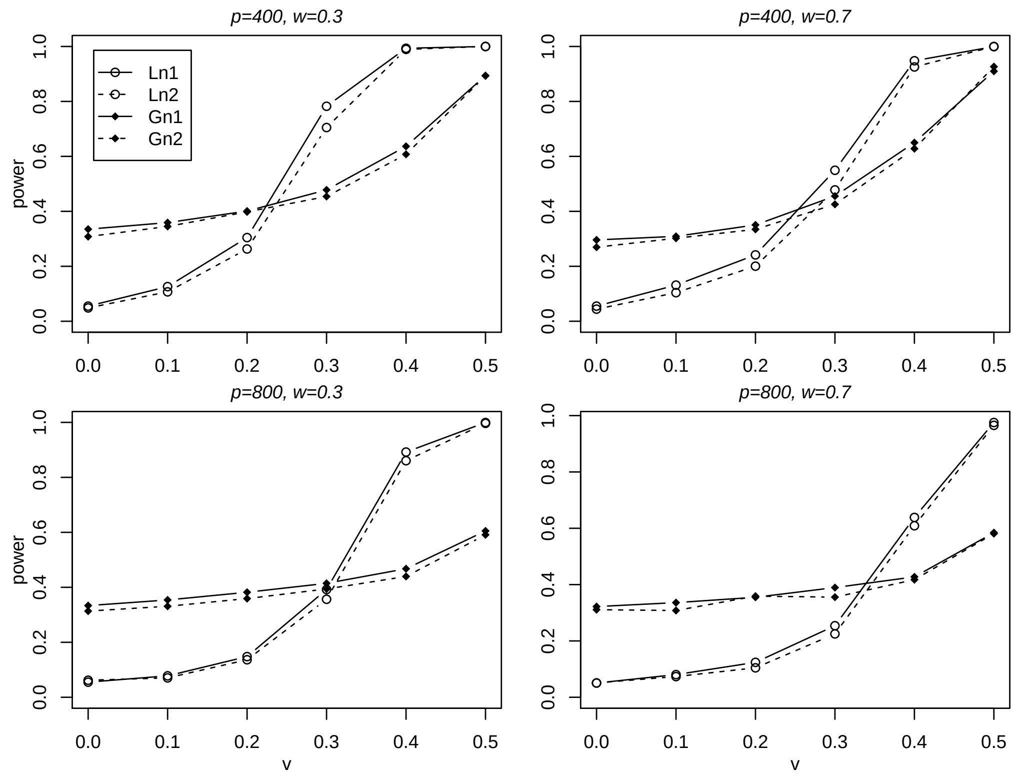
In this section, we provide technical proofs for theorems stated in Section 4.2. We first introduce some notation and basic definitions. For $\tau > 0$, define

$$h_{\tau,t}(x) = h_0\{\tau(x-t)\}, \text{ where } h_0(x) = [1 - \min\{1, \max(x, 0)\}^4]^4.$$

Then, for any $\tau > 0$,

$$\mathbb{I}\{x \leq t\} \leq h_{\tau,t}(x) \leq \mathbb{I}\{x \leq t + \tau^{-1}\}. \tag{4.11}$$

Figure 4.2: Empirical powers of $L_{n,\nu}$ and $G_{n,\nu}$ for vector autoregressive model



For a random variable $X \in \mathbb{R}$ and $\varpi > 0$, recall that the Levy concentration is given by

$$\mathcal{L}(X, \varpi) = \sup_{t \in \mathbb{R}} \mathbb{P}(t \leq X \leq t + \varpi).$$

For a positive semidefinite matrix A, let $A^{1/2}$ denote its principal square root matrix such that $A = A^{1/2}A^{1/2}$. For a random variable X, write $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}$, $q > 0$, if $\mathbb{E}|X|^q < \infty$. Throughout this section, we use C_q denote positive constant depends only on q and its value may vary at different places. Recall $\delta = q - 2$.

Lemma 4.4.1. *Let $X_1, \ldots, X_n \in \mathbb{R}^p$ be independent random vectors with $\mathbb{E}(X_i) = 0$. Let $\mu_1, \mu_2, \ldots, \mu_n \in \mathbb{R}^p$ be non-random vectors. Define $L_{n,X} = \sum_{i-j\geq 2}(X_i + \mu_i)^\top (X_j + \mu_j)$ and its Gaussian analogue $L_{n,Y} = \sum_{i-j\geq 2}(Y_i + \mu_i)^\top (Y_j + \mu_j)$, where $Y_1, Y_2, \ldots, Y_n \in \mathbb{R}^p$ are independent Gaussian random vectors with $\mathbb{E}(Y_i) = 0$ and $\mathrm{cov}(Y_i) = \mathrm{cov}(X_i)$ for each*

$1 \le i \le n$. *Then we have*

$$\sup_{t\in\mathbb{R}} |\mathbb{E}\{h_{\tau,t}(L_{n,X})\} - \mathbb{E}\{h_{\tau,t}(L_{n,Y})\}| \le C_q K_q \tau^q,$$

where

$$K_q = \sum_{i=1}^n \mathbb{E}\Big| \sum_{j\le i-2} X_i^\top (X_j + \mu_j)\Big|^q + \sum_{i=1}^n \mathbb{E}\Big| \sum_{j\ge i+2} X_i^\top (Y_j + \mu_j)\Big|^q + \sum_{i=1}^n \mathbb{E}\Big| \sum_{j\le i-2} Y_i^\top (X_j + \mu_j)\Big|^q + \sum_{i=1}^n \mathbb{E}\Big| \sum_{j\ge i+2} Y_i^\top (Y_j + \mu_j)\Big|^q.$$

Moreover, if $X_i, i \in \mathbb{Z}$, *are independent and identically distributed, then*

$$\rho(L_{n,X}, L_{n,Y}) \le C_q (n^{-\delta/2} M_{q,X})^{1/(2q+1)},$$

where $M_{q,X} = \|X_1^\top X_2\|_q^q / |\mathrm{cov}(X_1)|_{\mathbb{F}}^q$.

Lemma 4.4.2. *Let* $V_{n,\nu}$ *be defined in Theorem 4.2.1. For any* $\varpi > 0$, *we have*

$$\mathcal{L}(V_{n,\nu}, \varpi) \le \left(\frac{8\varpi}{n\pi\nu^{1/2}\vartheta^2}\right)^{1/2} + 8(m/n)^{1/5}.$$

Lemma 4.4.3. *Let* $(X_n)_{n\ge1}$ *and* $(Y_n)_{n\ge1}$ *be two sequences of random variables. Assume that for some* $\alpha > 0$ *and* $q > 0$,

$$\|X_n - Y_n\|_q \le A_{n,q}, \;\text{ and }\; \mathcal{L}(Y_n, \varpi) \le \frac{\varpi^\alpha}{B_n^\alpha} + C_n, \tag{4.12}$$

where $A_{n,q}$, B_n *and* C_n *satisfy that*

$$\frac{A_{n,q}}{B_n} \to 0, \;\text{ and }\; C_n \to 0, \tag{4.13}$$

then as $n \to \infty$, we have

$$\rho(X_n, Y_n) \to 0.$$

Proof of Lemma 4.4.3. Let $\delta_n = A_n^{\chi} B_n^{1-\chi}$, where constant $0 < \chi < 1$. By the Markov inequality, (4.12) and (4.13), it follows that

$$\begin{aligned}
\rho(X_n, Y_n) &\le \mathbb{P}(|X_n - Y_n| > \delta_n) + \mathcal{L}(Y_n, \delta_n) \\
&\le \frac{\|X_n - Y_n\|_q^q}{\delta_n^q} + (\delta_n/B_n)^{\alpha} + C_n \\
&\le (A_{n,q}/B_n)^{(1-\chi)q} + (A_{n,q}/B_n)^{\alpha\chi} + C_n \to 0.
\end{aligned}$$

□

Proof of Theorem 4.2.1. For simplicity of notation, we suppress the subscript ν in what follows. Decompose

$$\begin{aligned}
B_k &= \sum_{i=(3k-3)m+1}^{(3k-2)m} X_i + \sum_{i=(3k-2)m+1}^{(3k-1)m} X_i + \sum_{i=(3k-1)m+1}^{3km} X_i \\
&= B_{k,1} + B_{k,2} + B_{k,3}.
\end{aligned}$$

Define $\xi = (\xi_k)_{k\in\mathbb{Z}}$, where $\xi_k = (\varepsilon_{3km-4m+1}, \varepsilon_{3km-4m+2}, \ldots, \varepsilon_{3km-3m})$. As $(X_i)_{i\in\mathbb{Z}}$ are ν-dependent martingale differences with respect to the σ-field generated by $\{\varepsilon_i, \varepsilon_{i-1}, \ldots\}$, it follows that

$$\mathrm{cov}(B_k|\xi) = \Sigma_{1,\xi_k} + \Sigma_2 + \Sigma_{3,\xi_{k+1}}, \text{ and } \mathbb{E}(B_k|\xi) = \mu_{k+1},$$

where

$$\Sigma_{1,\xi_k} = \mathrm{cov}(B_{k,1}|\xi_k), \;\; \Sigma_2 = \mathrm{cov}(B_{k,2}), \;\; \Sigma_{3,\xi_{k+1}} = \mathrm{cov}(B_{k,3}|\xi_{k+1}), \;\; \mu_{k+1} = \mathbb{E}(B_{k,3}|\xi_{k+1}).$$

Let $\{Z_{k,i}, 1 \le i \le 3, k \in \mathbb{Z}\}$ be independent p-dimensional standard Gaussian random vectors. For each k, define

$$B_{\natural,k} = \Sigma_{1,\xi_k}^{1/2} Z_{k,1} + \Sigma_2^{1/2} Z_{k,2} + \Sigma_{3,\xi_{k+1}}^{1/2} Z_{k,3} + \mu_{k+1}.$$

Conditioning on ξ, we note that

$$\mathrm{cov}(B_{\natural,k}|\xi) = \mathrm{cov}(B_k|\xi) \text{ and } \mathbb{E}(B_{\natural,k}|\xi) = \mathbb{E}(B_k|\xi).$$

Hence, conditioning on ξ, we define the Gaussian analogue of L_n as

$$L_{n,\natural} = \sum_{k-k' \ge 2} B_{\natural,k}^\top B_{\natural,k'}.$$

Note that $(B_{\natural,k})_{k\in\mathbb{Z}}$ are one-dependent random vectors. Following the idea of Berkes et al. [2014], we rearrange the sums of $(B_{\natural,k})_{k\in\mathbb{Z}}$ and obtain a sequence of independent random vectors

$$B_{\star,k} = \Sigma_{1,\xi_{k+1}}^{1/2} Z_{k+1,1} + \Sigma_2^{1/2} Z_{k,2} + \Sigma_{3,\xi_{k+1}}^{1/2} Z_{k,3} + \mu_{k+1}, \ \ k \in \mathbb{Z}.$$

Similar as $L_{n,\natural}$, we define

$$L_{n,\star} = \sum_{k-k' \ge 2} B_{\star,k}^\top B_{\star,k'}.$$

Step 1: Distribution approximation of L_n by $L_{n,\natural}$.

We are to bound

$$\Delta_{\tau,L} = \sup_{t\in\mathbb{R}} |\mathbb{E}\{h_{\tau,t}(L_n)\} - \mathbb{E}\{h_{\tau,t}(L_{n,\natural})\}|.$$

Note that $B_k - \mu_{k+1}$ is independent with $(\varepsilon_i)_{i \le 3(k-2)m}$ for each $k \in \mathbb{Z}$. Hence, by the Burkholder inequality,

$$\begin{aligned}\sum_{k\ge 3} \mathbb{E}\Big|\sum_{l\le k-2}(B_k-\mu_{k+1})^\top B_l\Big|^q &\le C_q\sum_{k\ge 3}(km)^{q/2}\|(B_k-\mu_{k+1})^\top X_1\|_q^q\\ &\le C_q\sum_{k\ge 3}(km)^{q/2}m^{q/2}\|X_0^\top X_m\|_q^q \le C_q d^{q/2+1}m^q\|X_0^\top X_m\|_q^q.\end{aligned}$$

Similarly, it follows that

$$\sum_{k\ge 3} \mathbb{E}\Big|\sum_{l\ge k+2}(B_k-\mu_{k+1})^\top B_{\natural,l}\Big|^q \le C_q d^{q/2+1}m^q\|X_0^\top X_m\|_q^q.$$

Then, by Lemma 4.4.1,

$$\Delta_{\tau,L} \le C_q d^{q/2+1}m^q\tau^q\|X_0^\top X_m\|_q^q \le C_q d^{q/2+1}m^q\nu^{q/2}\tau^q\|\varepsilon_1^\top\varepsilon_2\|_q^{2q}. \tag{4.14}$$

Step 2: Asymptotic distribution of $L_{n,\star}$.

We are to bound the Kolmogorov distance between the distributions of $L_{n,\star}$ and V_n,

$$\rho_\star = \rho(L_{n,\star}, V_n).$$

To this end, we shall apply Lemma 4.4.1. Note that $B_{\star,1}, B_{\star,2}, \ldots, B_{\star,d} \in \mathbb{R}^p$ are independent and identically distributed with $\|\mathrm{cov}(B_{\star,1})\|_{\mathbb{F}} = 3m\nu^{1/2}\vartheta^2$ and

$$\|B_{\star,1}^\top B_{\star,3}\|_q^2 \le C_q m^2\|X_0^\top X_m\|_q^2 \le C_q m^2\nu\|\varepsilon_1^\top\varepsilon_2\|_q^4.$$

Hence, by Lemma 4.4.1 and (4.3), it follows that

$$\rho_\star \le C_q\bigg(\frac{\|X_0^\top X_m/\vartheta^2\|_q^q}{d^{\delta/2}\nu^{q/2}}\bigg)^{1/(2q+1)} \le C_q\bigg(\frac{m^{\delta/4}\|\varepsilon_1^\top\varepsilon_2/\vartheta\|_q^q}{n^{\delta/4}}\bigg)^{2/(2q+1)} \to 0. \tag{4.15}$$

Step 3: Distribution approximation of $L_{n,\natural}$.

Now we bound the Kolmogorov distance between the distribution functions of $L_{n,\natural}$ and $L_{n,\star}$:

$$\rho_\natural = \rho(L_{n,\natural}, L_{n,\star}).$$

Decompose $L_{n,\natural} - L_{n,\star} = \mathcal{I}_1 + \mathcal{I}_2$, where

$$\mathcal{I}_1 = \sum_{k\geq 3} B_{\natural,k}^\top(\Sigma_{1,\xi_1}^{1/2} Z_{1,1} - \Sigma_{1,\xi_{k-1}}^{1/2} Z_{k-1,1}),$$
$$\mathcal{I}_2 = \sum_{l\leq d-2} B_{\star,l}^\top(\Sigma_{1,\xi_{l+2}}^{1/2} Z_{l+2,1} - \Sigma_{1,\xi_{d+1}}^{1/2} Z_{d+1,1}).$$

Elementary calculations imply that

$$\|\mathcal{I}_1\|_q^2 \leq C_q dm^2 \|X_0^\top X_m\|_q^2, \text{ and } \|\mathcal{I}_2\|_q^2 \leq C_q dm^2 \|X_0^\top X_m\|_q^2.$$

Therefore, by (4.3),

$$\frac{\|L_{n,\natural} - L_{n,\star}\|_q^2}{n^2\nu\vartheta^4} \leq \frac{C_q dm^2 \|X_0^\top X_m\|_q^2}{n^2\nu\vartheta^4} \leq \frac{C_q dm^2 \nu \|\varepsilon_1^\top \varepsilon_2\|_q^4}{m^2 d^2 \nu\vartheta^4}$$
$$= C_q(d^{-q/4} M_q)^{4/q} \to 0.$$

By (4.15) and Lemma 4.4.2, for any $\delta > 0$,

$$\mathcal{L}(L_{n,\star}, \delta) \leq 2\rho_\star + \mathcal{L}(V_n, \delta) \leq 2\rho_\star + \left(\frac{8\delta}{n\pi\nu^{1/2}\vartheta^2}\right)^{1/2} + 8(m/n)^{1/5}.$$

Consequently, by Lemma 4.4.3,

$$\rho_\natural \to 0. \tag{4.16}$$

Step 4: Completion of the proof.

By (4.11) and (4.14)−(4.16), it follows that for any $t > 0$,

$$\begin{aligned}\mathbb{P}(L_n \le t) &\le \mathbb{E}h_{\tau,t}(L_n) \le \mathbb{E}h_{\tau,t}(L_{n,\natural}) + \Delta_{\tau,L} \le \mathbb{P}(L_{n,\natural} \le t + \tau^{-1}) + \Delta_{\tau,L} \\ &\le \rho_{\natural} + \rho_{\star} + \Delta_{\tau,L} + \mathbb{P}(V_n \le t) + \mathcal{L}(V_n, \tau^{-1}),\end{aligned}$$

and similarly that

$$\mathbb{P}(L_n \le t) \ge \mathbb{P}(V_n \le t) - \rho_{\natural} - \rho_{\star} - \Delta_{\tau,L} - \mathcal{L}(V_n, \tau^{-1}).$$

Then, taking $\tau^{-1} = d^{\chi/q+1-\chi/2} m \nu^{1/2} \vartheta^2$ for some positive constant $\chi \in (0,1)$, it follows by (4.3) that $\Delta_{\tau,L} \to 0$ and $\mathcal{L}(V_n, \tau^{-1}) \to 0$. Consequently,

$$\rho \le \rho_{\natural} + \rho_{\star} + \Delta_{\nu,L} + \mathcal{L}(V_n, \tau^{-1}) \to 0.$$

□

Lemma 4.4.4. *Define*

$$\tilde{L}_{n,1} = \sum_{k-k' \ge 2} \mathbb{E}_0(B_k)^\top \mathbb{E}_0(B_{k'}) \text{ and } L_{n,\diamond} = \sum_{k-k' \ge 2} W_k^\top W_{k'},$$

where

$$W_k = \sum_{i=3(k-1)m+1}^{3km} D_i.$$

Then under the alternative hypothesis H_1,

$$\|\tilde{L}_{n,1} - L_{n,\diamond}\|_q^2 \le C_q n \|D_m^\top E_0(A\epsilon_0 \otimes \varepsilon_0)\|_q^2 + C_q d \|E_0(A\epsilon_0 \otimes \varepsilon_0)^\top E_0(A\epsilon_m \otimes \varepsilon_m)\|_q^2.$$

Proof of Lemma 4.4.4. Note that $\mathbb{E}_0(X_i) - D_i = \mathbb{E}_0(A\epsilon_{i-1} \otimes \varepsilon_{i-1}) - \mathbb{E}_0(A\epsilon_i \otimes \varepsilon_i)$ for each

$i \in \mathbb{Z}$. Hence for any $k \leq l \in \mathbb{Z}$,

$$\sum_{k \leq i \leq l} (X_i - \mathbb{E}X_i - D_i) = \mathbb{E}_0(A\epsilon_{k-1} \otimes \varepsilon_{k-1}) - \mathbb{E}_0(A\epsilon_l \otimes \varepsilon_l). \tag{4.17}$$

Decompose $\tilde{L}_{n,1} - L_{n,\diamond} = \Delta_{1,\diamond} + \Delta_{2,\diamond}$, where

$$\Delta_{1,\diamond} = \sum_{k-l \geq 2} \mathbb{E}_0(B_k)^\top \{\mathbb{E}_0(B_l) - W_l\} \text{ and } \Delta_{2,\diamond} = \sum_{k-l \geq 2} \{\mathbb{E}_0(B_k) - W_k\}^\top W_l.$$

By the Burkholder inequality, it follows that

$$\begin{aligned} \|\Delta_{1,\diamond}\|_q^2 &\leq C_q \sum_{k \geq 3} \|\mathbb{E}_0(B_k)^\top \mathbb{E}_0(A\epsilon_{3(k-2)m} \otimes \varepsilon_{3(k-2)m})\|_q^2 \\ &\quad + C_q \sum_{k \geq 3} \|\mathbb{E}_0(B_k)^\top \mathbb{E}_0(A\epsilon_0 \otimes \varepsilon_0)\|_q^2 \\ &\leq C_q n \|D_m^\top \mathbb{E}_0(A\epsilon_0 \otimes \varepsilon_0)\|_q^2 + C_q d \|\mathbb{E}_0(A\epsilon_0 \otimes \varepsilon_0)^\top \mathbb{E}_0(A\epsilon_m \otimes \varepsilon_m)\|_q^2. \end{aligned}$$

Similarly, we have $\|\Delta_{2,\diamond}\|_q^2 \leq C_q n \|D_m^\top \mathbb{E}_0(A\epsilon_0 \otimes \varepsilon_0)\|_q^2$. □

Proof of Theorem 4.2.3. By (4.8), we have

$$\begin{aligned} \|\mathbb{E}_0(L_{n,1}) - \tilde{L}_{n,1}\|_2^2 &\leq 12n^2 d \text{vec}(\Sigma_1)^\top \text{cov}(A\epsilon_0 \otimes \varepsilon_0) \text{vec}(\Sigma_1) \\ &\quad + 6n^3 \text{vec}(\Sigma_1)^\top \text{cov}(D_0) \text{vec}(\Sigma_1), \end{aligned}$$

which implies $\|\mathbb{E}_0(L_{n,1}) - \tilde{L}_{n,1}\|_2 = o(n\vartheta_\diamond)$ in view of (4.8). By Lemma 4.4.4, (4.7) and (4.9), it follows that

$$\|\tilde{L}_{n,1} - L_{n,\diamond}\|_q = o(n\vartheta_\diamond).$$

By a similar argument as the proof of Theorem 4.2.1, under (4.7),

$$\rho_\diamond = \rho(L_{n,\diamond}, \mathcal{V}_n) \to 0.$$

Combined with Lemma 4.4.2, it follows that for any $\delta > 0$,

$$\mathcal{L}(L_{n,\diamond}, \delta) \leq 2\rho_\diamond + \mathcal{L}(\mathcal{V}_n, \delta) \leq 2\rho_\diamond + \left(\frac{8\delta}{n\pi\vartheta_\diamond}\right)^{1/2} + 8(m/n)^{1/5}.$$

Consequently, by Lemma 4.4.3, we have $\tilde{\rho} \to 0$. □

REFERENCES

Michael G Akritas and Steven F Arnold. Fully nonparametric hypotheses for factorial designs I: Multivariate repeated measures designs. *J. Amer. Statist. Assoc.*, 89(425):336–343, 1994.

T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley Series in Probability and Statistics. 2003.

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

István Berkes, Weidong Liu, and Wei Biao Wu. Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.*, 42(2):794–817, 2014.

G. E. P. Box and David A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.*, 65:1509–1526, 1970.

Edgar Brunner and Madan L. Puri. Nonparametric methods in factorial designs. *Statist. Papers*, 42(1):1–52, 2001.

T. Tony Cai and Zongming Ma. Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, 19(5B):2359–2388, 2013.

T. Tony Cai and Yin Xia. High-dimensional sparse MANOVA. *J. Multivariate Anal.*, 131: 174–196, 2014.

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

Song Xi Chen and Ying-Li Qin. A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.*, 38(2):808–835, 2010.

Xiaohui Chen. Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.*, 46(2):642–678, 2018.

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

Jianqing Fan, Shaojun Guo, and Ning Hao. Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 74 (1):37–65, 2012.

F. Götze and A. Tikhomirov. Asymptotic distribution of quadratic forms and applications. *J. Theoret. Probab.*, 15(2):423–475, 2002.

F. Götze and A. N. Tikhomirov. Asymptotic distribution of quadratic forms. *Ann. Probab.*, 27(2):1072–1098, 1999.

J. R. M. Hosking. The multivariate portmanteau statistic. *J. Amer. Statist. Assoc.*, 75(371): 602–608, 1980.

Jiang Hu, Zhidong Bai, Chen Wang, and Wei Wang. On testing the equality of high dimensional mean vectors with unequal covariance matrices. *Annals of the Institute of Statistical Mathematics*, 69(2):365–387, 2017.

William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *J. Amer. Statist. Assoc.*, 47(260):583–621, 1952.

Huiqin Li, Jiang Hu, Zhidong Bai, Yanqing Yin, and Kexin Zou. Test on the linear combinations of mean vectors in high-dimensional data. *TEST*, 26(1):188–208, 2017.

Zeng Li, Clifford Lam, Jianfeng Yao, and Qiwei Yao. On testing for high-dimensional white noise. *arXiv preprint arXiv:1808.03545*, 2018.

G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

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

Dimitris N. Politis and Joseph P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.*, 22(4):2031–2050, 1994.

Stephen Portnoy. On the central limit theorem in $\mathbf{R}^p$ when $p \to \infty$. *Probab. Theory Related Fields*, 73(4):571–583, 1986.

Maria L. Rizzo and Gábor J. Székely. DISCO analysis: a nonparametric extension of analysis of variance. *Ann. Appl. Stat.*, 4(2):1034–1055, 2010.

Mark Rudelson and Roman Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62(12):1707–1739, 2009.

James R. Schott. Some high-dimensional tests for a one-way MANOVA. *J. Multivariate Anal.*, 98(9):1825–1839, 2007.

Muni S. Srivastava and Yasunori Fujikoshi. Multivariate analysis of variance with fewer observations than the dimension. *J. Multivariate Anal.*, 97(9):1927–1940, 2006.

Muni S. Srivastava and Tatsuya Kubokawa. Tests for multivariate analysis of variance in high dimension under non-normality. *J. Multivariate Anal.*, 115:204–216, 2013.

Muni S. Srivastava, Shota Katayama, and Yutaka Kano. A two sample test in high dimensional data. *J. Multivariate Anal.*, 114:349–358, 2013.

Olivier Thas. *Comparing distributions.* Springer Series in Statistics. Springer, New York, 2010.

Ruey S. Tsay. *Multivariate time series analysis.* Wiley, Hoboken, NJ, 2014.

Ruey S. Tsay. Testing for serial correlations in high-dimensional time series via extreme value theory. *Technical report*, 2019.

Lan Wang, Bo Peng, and Runze Li. A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist. Assoc.*, 110(512):1658–1669, 2015.

Jennifer Wessel and Nicholas J Schork. Generalized genomic distance–based regression methodology for multilocus association analysis. *The American Journal of Human Genetics*, 79(5):792–806, 2006.

C.-F. J. Wu. On the asymptotic properties of the jackknife histogram. *Ann. Statist.*, 18(3): 1438–1452, 1990.

Wei Biao Wu. Strong invariance principles for dependent random variables. *Ann. Probab.*, 35(6):2294–2320, 2007.

Mengyu Xu, Danna Zhang, and Wei Biao Wu. L^2 asymptotics for high-dimensional data. *arXiv preprint arXiv:1405.7244*, 2014.

Matthew A. Zapala and Nicholas J. Schork. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proceedings of the National Academy of Sciences*, 103(51):19430–19435, 2006.

Matthew A Zapala and Nicholas J Schork. Statistical properties of multivariate distance matrix regression for high-dimensional data analysis. *Frontiers in genetics*, 3:190, 2012.

Jin-Ting Zhang, Jia Guo, and Bu Zhou. Linear hypothesis testing in high-dimensional one-way MANOVA. *J. Multivariate Anal.*, 155:200–216, 2017.

Jin-Ting Zhang, Jia Guo, Bu Zhou, and Ming-Yen Cheng. A simple two-sample test in high dimensions based on L^2-norm. *Journal of the American Statistical Association*, pages 1–42, 2019.

Xianyang Zhang, Shun Yao, and Xiaofeng Shao. Conditional mean and quantile dependence testing in high dimension. *Ann. Statist.*, 46(1):219–246, 2018.

Bu Zhou, Jia Guo, and Jin-Ting Zhang. High-dimensional general linear hypothesis testing under heteroscedasticity. *J. Statist. Plann. Inference*, 188:36–54, 2017.