The Pennsylvania State University
The Graduate School

TOPICS ON POWER ENHANCEMENT IN HIGH-DIMENSIONAL HYPOTHESIS TESTS

A Dissertation in
Statistics
by
Xiufan Yu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2021

The dissertation of Xiufan Yu was reviewed and approved by the following:

Runze Li
Eberly Family Chair Professor of Statistics
Dissertation Co-Advisor, Co-Chair of Committee

Lingzhou Xue
Associate Professor of Statistics
Dissertation Co-Advisor, Co-Chair of Committee

Bing Li
Verne M. Willaman Professor of Statistics

Jingzhi Huang
Professor of Finance and Mathematics
Faculty Chair in Finance

Ephraim Hanks
Associate Professor of Statistics
Chair of Graduate Studies

Abstract

In recent years, power-enhanced tests with high-dimensional data have received growing attention in theoretical and applied statistics. Many scientific research questions can be converted into hypothesis testing problems, for example, the discovery of association between gene-sets and disease outcomes, the evaluation on the validity of a pricing model for financial market. Various tests possess different high-power regions. In practice, we may lack prior knowledge about the alternatives when testing for a problem of interest. It is important to develop powerful testing procedures against more general alternatives.

In this dissertation, we propose new methods to achieve power enhancement (PE) in tests for high-dimensional data. In particular, we consider the problem of enhancing test power in three topics: (1) a one-sample test on multi-factor pricing models for large panels, (2) a two-sample test on the equality of high-dimensional covariance matrices, and (3) a simultaneous test on the equality of two-sample mean vectors and covariance matrices of high-dimensions.

Methodologically, we provide a new perspective to the literature by studying and utilizing the asymptotic joint distribution of different statistics. We show two PE techniques of (i) aggregating information via the combination of p-values, and (ii) constructing PE components, to achieve enhanced test power in two aspects: (a) expanding high-power regions towards a wider alternative space with respect to one parameter of interest, and (b) expanding test capability to alternative spaces with respect to more parameters. Theoretically, we derive joint limiting laws of the corresponding test statistics. We prove that the proposed power-enhanced tests achieve the desired PE properties following the guidance of the three general PE principles (Fan, Liao and Yao, 2015). Practically, the test efficacy is demonstrated by Monte Carlo simulations as well as empirical studies on testing market efficiency and identifying differentially expressed gene-sets.

Table of Contents

List of Figures **vii**

List of Tables **viii**

Acknowledgments **x**

Chapter 1
Introduction **1**
1.1 Power Enhancement for Testing Multi-Factor Pricing Models of Large Panels . . . 3
1.2 Power Enhancement for Testing High-Dimensional Covariance Matrices . . . 4
1.3 Power Enhancement for Simultaneously Testing High-Dimensional Mean Vectors and Covariance Matrices . . . 5
1.4 Organization . . . 6

Chapter 2
Literature Review **7**
2.1 Testing Multi-Factor Pricing Models . . . 7
2.1.1 Multiple t-tests . . . 8
2.1.2 Multivariate F-test . . . 8
2.1.3 Operational J_α Test for Large Panels . . . 9
2.1.4 Power Enhancement Test for Sparse Alternatives . . . 11
2.2 Testing Two-Sample Mean Vectors . . . 12
2.2.1 Hotelling's T^2 Test . . . 13
2.2.2 Quadratic-Form Test . . . 13
2.2.3 Maximum-Form Test . . . 14

2.2.4 Projection Test 15
2.3 Testing Two-Sample Covariance Matrices 15
2.3.1 Likelihood Ratio Test (LRT) 16
2.3.2 Corrected Likelihood Ratio Test 16
2.3.3 L_2-norm-based Test 17
2.3.4 Extreme-value-based Test 20
2.4 Two-Sample Simultaneous Tests on Mean Vectors and Covariance Matrices 21
2.4.1 Likelihood Ratio Test 21
2.4.2 Modified Likelihood Ratio Test 22

Chapter 3
Innovated Power Enhancement for Testing Multi-Factor Asset Pricing Models **23**
3.1 Introduction 23
3.2 Methodology 27
3.2.1 Problem Setup 27
3.2.2 Existing Methods 28
3.2.3 A New Maximum Test Statistic 29
3.2.4 Innovated Power Enhancement Test 31
3.3 Asymptotic Properties 33
3.3.1 Assumptions 33
3.3.2 Asymptotic Size and Power of L_{max} 34
3.3.3 Joint Limiting Law for S_{Wald} and L_{max} 35
3.3.4 Asymptotic Size and Power of F_{joint} 37
3.4 Numerical Properties 39
3.4.1 Data Generating Process and Parameter Calibration 39
3.4.2 Size and Power Analysis 41
3.5 Application to Testing Market Efficiency 44
3.6 Conclusion 49
3.7 Lemmas and Proofs 49
3.8 Supplement on Generalized Positive-Definite Penalized Estimation of Large Covariance Matrices 62

Chapter 4
Fisher's Combined Probability Test for High-Dimensional Covariance Matrices **70**
4.1 Introduction 70
4.2 Preliminaries 73
4.3 Fisher's Combined Probability Test 75

4.4 Asymptotic Properties 77
4.4.1 Assumptions 77
4.4.2 Asymptotic Joint Distribution 79
4.4.3 Asymptotic Size and Power 82
4.5 Simulation Studies 83
4.6 Application to Gene-Set Testing 88
4.7 Conclusion 92
4.8 Lemmas and Proofs 93

Chapter 5
Power-Enhanced Simultaneous Test of High-Dimensional Mean Vectors and Covariance Matrices 112
5.1 Introduction 112
5.2 Preliminaries 116
5.3 Methodology 117
5.3.1 Power-Enhanced Tests 117
5.3.2 Joint Distribution of the Two Power-Enhanced Test Statistics 123
5.3.3 Power-Enhanced Simultaneous Test 124
5.4 Simulation Studies 127
5.5 Application to Gene-Set Testing 140
5.6 Conclusion 144
5.7 Assumptions, Lemmas and Proofs 145

Chapter 6
Summary and Future Work 163
6.1 Summary 163
6.2 Future Work 164

Bibliography 165

List of Figures

3.1 Time series plot of the quadratic and maximum test statistics 46
3.2 Time series plot on p-values of the quadratic, maximum and joint tests . 46
3.3 Boxplots of p-values of the quadratic, maximum and joint tests for each sector 47
3.4 p-values of the quadratic, maximum and joint tests for sectors EN and HC 48

4.1 Histograms of the dimension of gene-sets for three GO categories . . 90
4.2 Boxplots of the LC and CLX test statistics for three GO categories 90

5.1 Power-enhanced simultaneous testing procedure 125
5.2 Empirical power comparison of the seven tests under H_b with Gaussian distributed $\{Z_{k,i}\}$ and $N = 100$, $p = 500$ 131
5.3 Boxplots of the dimension of gene-sets for three GO categories . . . 141
5.4 Boxplots of the $\widehat{\sigma}_{01}^{-1} M_{n_1,n_2}$ and $\widehat{\sigma}_{02}^{-1} T_{n_1,n_2}$ test statistics for three GO categories 142

List of Tables

3.1 Comparison of empirical size and power for L_{max} and S_{Wald} in a simulation study. 31
3.2 The calibrated mean vectors and covariance matrices used to generate factor loadings $\{\mathbf{b}_j\}_{j=1}^{p}$ in the data generating process. 40
3.3 The calibrated mean vectors, coefficients and covariance matrices used to generate factors $\{\mathbf{f}_t\}_{t=1}^{T}$ in the data generating process. . . . 40
3.4 The empirical size and power (%) of different testing procedures with Gaussian and non-Gaussian errors and covariance structure $\mathbf{\Sigma}^{\star}_{(1)}$. 43
3.5 The empirical size and power (%) of different testing procedures with Gaussian and non-Gaussian errors and covariance structure $\mathbf{\Sigma}^{\star}_{(2)}$. 44

4.1 Comparison of empirical size and power (%) for model (i) 85
4.2 Comparison of empirical size and power (%) for models (ii) and (iii) 86
4.3 Comparison of empirical size and power (%) comparisons for models (iv) and (v) 87
4.4 Summary of the dimension of gene-sets for three GO categories . . 89
4.5 Gene-set testing results at the nominal level $\alpha = 0.05$ 91
4.6 Gene-set testing results with the FDR control at $\alpha = 0.05$ 91

5.1 Empirical size (%) with Gaussian and Gamma distributed $\{Z_{k,i}\}$ in the data generating process 129
5.2 Empirical power (%) against H_m with Gaussian distributed $\{Z_{k,i}\}$ in the data generating process 132
5.3 Empirical power (%) against H_c with Gaussian distributed $\{Z_{k,i}\}$ in the data generating process 133
5.4 Empirical power (%) against H_b with Gaussian distributed $\{Z_{k,i}\}$ in the data generating process 134
5.5 Empirical power (%) against H_b with Gaussian distributed $\{Z_{k,i}\}$ in the data generating process (continued) 135
5.6 Empirical power (%) against H_m with Gamma distributed $\{Z_{k,i}\}$ in the data generating process 136

5.7 Empirical power (%) against H_c with Gamma distributed $\{Z_{k,i}\}$ in the data generating process 137
5.8 Empirical power (%) against H_b with Gamma distributed $\{Z_{k,i}\}$ in the data generating process 138
5.9 Empirical power (%) against H_b with Gamma distributed $\{Z_{k,i}\}$ in the data generating process (continued) 139
5.10 The number of significant gene-sets declared by different tests after BH control with nominal level $\alpha = 0.05$ 142

Acknowledgments

I am sincerely grateful to everyone who has helped me along the way.

First of all, I would like to express my deepest gratitude to my Ph.D. advisors, Dr. Runze Li and Dr. Lingzhou Xue, for their invaluable guidance and tremendous support. They have always been guiding me with their broad knowledge and inspiring me with their enthusiasm. The mentorship helps me grow not only professionally but also personally. Their unwavering encouragement gives me strength, and their kindness has been the driving force that motivates me to strive for better. They have taught me so many things that words are not enough to express how grateful I am. Thank you so much for everything.

I would like to extend my heartfelt gratitude to my doctoral committee members, Dr. Bing Li and Dr. Jingzhi Huang. I am deeply thankful for their substantial support to my career development, and for their constructive suggestions about this dissertation. I want to give special thanks to my collaborators, Dr. Jiawei Yao and Dr. Danning Li, for their precious input and helpful discussions. I would also like to thank all my friends for making this journey enjoyable.

Last but not least, I want to thank my parents and my fiancé for their unconditional love and endless support.

This dissertation research was partially supported by National Science Foundation grants CCF 2007823, DMS 1811552, DMS 1820702, DMS 1953189, DMS 1953196 and DMS 2015539, National Institute on Drug Abuse NIH grant P50 DA039838, and National Institute of Allergy and Infectious Diseases R21 grant AI144765. The findings and conclusions expressed in this dissertation do not necessarily reflect the views of the funding agencies.

Chapter 1

Introduction

Hypothesis testing plays a fundamental role in statistics. In contemporary scientific research, rapid technological innovations facilitate the collection of massive features for each individual. In many circumstances, the sample size is much smaller than the number of features, forming a "large p, small n" paradigm. The high-dimensionality fundamentally challenges classical theory. For such high-dimensional data, the traditional likelihood-based methods become largely defective or even ill-defined, motivating a critical need of developing new statistical methods to deal with the high-dimensionality.

Over the past decade, researchers have devoted a lot of efforts to tackle the challenges in high-dimensional settings. Let $\boldsymbol{\theta}$ denote a high-dimensional parameter of interest and suppose we are interested in testing whether $\boldsymbol{\theta}$ equals to zero, i.e., $H_0 : \boldsymbol{\theta} = \mathbf{0}$. Note that $\boldsymbol{\theta}$ may be a vector or a matrix depending on the problem of interest. One common solution to develop a test for such high-dimensional problem is to first choose a well-defined distance function to measure the distance between the null and alternative hypotheses, and then construct a test statistic based on an estimator of the distance. Two dominated choices of the distance functions are the L_2-norm $\|\boldsymbol{\theta}\|_2$ and the L_∞-norm $\|\boldsymbol{\theta}\|_\infty$. The null hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$ is transformed into its equivalent forms of $H_0' : \|\boldsymbol{\theta}\|_2 = 0$ and $H_0'' : \|\boldsymbol{\theta}\|_\infty = 0$, respectively. The two popular choices of distance functions correspond to two types of tests: (a) *the L_2-norm-based tests*, sometimes named as quadratic-form tests, (Bai and Saranadasa, 1996; Chen and Qin, 2010; Chen et al., 2010; Wang et al., 2015; Chen et al., 2018), in which the test statistics are constructed based on $\widehat{\|\boldsymbol{\theta}\|_2}$, and (b) *the* *extreme-value-based tests*, sometimes named as maximum-form tests, (Cai et al., 2013, 2014; Chang et al., 2017; Xue and Yao, 2020), in which the test statistics are constructed based on $\widehat{\|\boldsymbol{\theta}\|_\infty}$. It has been recognized that the L_2-norm-based tests are powerful against *dense alternatives* with only a few large non-zero signals, but suffer from low power against *sparse alternatives* in which the null hypothesis is violated by a large number of small disturbances. In contrast, the extreme-value-based tests are powerful against sparse alternatives, but perform poorly against dense alternatives.

Various tests have different high-power regions. In practice, lacking prior knowledge about the alternatives, it would be challenging for us to decide which test to use. Therefore, it is of great importance to develop testing procedures that remain powerful against general alternatives. Fan, Liao and Yao (2015) is regarded as one of the pioneering works in this field. They introduced three general *power enhancement (PE) principles* that a power-enhanced test is expected to follow:

(a) no size distortion;

(b) the power-enhanced test is at least as power as the original test;

(c) the testing power is substantially enhanced under a more general alternatives,

One natural way to combine the strengths of both types of tests is via a weighted sum of the two test statistics (Li and Xue, 2015; Yang and Pan, 2017; Li et al., 2018). However, the weighted sum tests bear some drawbacks. On the one hand, the performance of such tests depends critically on a proper choice of weights. Different test statistics are often of distinct magnitudes. An inappropriate choice of the combination weights may lead to test inefficiency, causing substantial loss of power. On the other hand, the distributions of the weighted sum statistics are usually challenging to obtain, resulting in difficulty in choosing critical values.

In this dissertation, we provide new perspectives to the literature of power enhancement in hypothesis tests of high-dimensional data. We propose novel power-enhanced tests for various research questions, including a one-sample test on multi-factor pricing models for large panels (Chapter 3), a two-sample test on the equality of covariance matrices of high-dimensions (Chapter 4), and a simultaneous test on the equality of two-sample high-dimensional mean vectors and covariances matrices (Chapter 5).

1.1 Power Enhancement for Testing on Multi-Factor Pricing Models of Large Panels

The multi-factor pricing model plays a vital role in arbitrage theory and practice of asset pricing. It employs multiple macroeconomic factors to explain the market phenomena, portfolio of securities, or equilibrium asset prices. A key implication from the multi-factor pricing theory is that the intercept should be zero for any asset. The essential of the factor pricing model has made this testing problem received considerable attention over the past few decades. Cross sectional regressions as well as time series regressions are applied to testing in many different contexts. Jensen (1968) was the first to propose the test of zero-intercept with the help of standardized t-statistics using ordinary least squares regression for each asset. Various improved tests have been investigated since then, for instance, Gibbons et al. (1989), MacKinlay and Richardson (1991).

During recent years, there has been a growing body of literature focusing on the high-dimensional settings, in which the number of portfolio is either comparable or much larger than the length of time periods. In this data-rich environment, the rapidly changing market requires us to model a large number of assets over a relatively short time period, so that the potential time variations in model coefficients can be mitigated. Most existing tests are based on quadratic forms, for example, the Wald test (Pesaran and Yamagata, 2012, 2017). These tests are powerful in detecting systematic mis-pricing phenomenon of the financial market. However, due to accumulation of errors in estimating high-dimensional parameters, the quadratic-form tests perform poorly when there exist a small portion of mis-priced assets. To boost the testing power over the Wald test, Fan et al. (2015) introduced a power enhancement component in high-dimensional cross-section tests based on a screening technique. By adding the constructed component to the Wald statistic, the testing power is strengthened in the presence of a few mis-priced assets.

In Chapter 3, we propose an innovated power-enhanced test for multi-factor pricing models of large panels. To provide a promising alternative to Fan et al. (2015), we first introduce a new maximum-form test statistic and then study the asymptotic joint distribution of the Wald test statistic and the proposed maximum test statistic. We prove that these two statistics are asymptotically independent. Given their asymptotic independence, we propose an innovated power enhancement testing procedure to combine their respective power via the combination of their p-values. Theoretically, we prove that the innovated power enhancement test retains the desired nominal significance level and achieves asymptotically consistent power against more general alternative, in the sense that the test is powerful in detecting market abnormalities in the presence of either systematic mis-pricing or a small portion of mis-priced assets. Furthermore, we demonstrate the finite-sample performance of our proposed test using simulation studies and an empirical study for testing market efficiency using asset returns of the Russel-2000 portfolio.

1.2 Power Enhancement for Testing High-Dimensional Covariance Matrices

Testing large covariance matrices is of great importance in statistical analysis. Many statistical procedures rely on the covariance matrices, for example, the inferences on mean vectors, discriminant analysis, principal component analysis, and clustering analysis. The likelihood ratio test has been extensively studied in the classical low-dimensional settings. (Anderson, 2003). The rapidly advancing technologies lead to an era of big data, motivating a critical need of developing new methods for high-dimensional data.

Most existing literature develop their tests based on two types of statistics: quadratic-form statistics (Schott, 2007; Li and Chen, 2012) and maximum-form statistics (Cai et al., 2014; Chang et al., 2017). However, it has been well-known that the quadratic-form tests suffer from low power against sparse alternatives in which the null hypothesis is violated by only a few components, whereas the maximum-form tests suffer from low power against dense alternatives where the null hypothesis is violated by many small elements. Recently, the weighted combination methods (Li and Xue, 2015; Yang and Pan, 2017; Li et al., 2018) are introduced to improve the power of quadratic-form tests or maximum-form tests provided that the weights are appropriately chosen.

In Chapter 4, we provide a new perspective to exploit the full potential of quadratic-form statistics and maximum-form statistics by studying their joint limiting distribution. After proving the two types of statistics are asymptotically independent, we propose a scale-invariant power-enhanced test based on the Fisher's combination method. Taking advantage of the merits of both types of statistics, we show that the proposed test can retain the correct asymptotic size and enlarge high-power regions to the union of sparse alternatives and dense alternatives. The test efficacy is further demonstrated by simulation studies an empirical studies on identifying differentially expressed gene-sets among various types of tumors.

1.3 Power Enhancement for Simultaneously Testing High-Dimensional Mean Vectors and Covariance Matrices

Inference on the equality of two distributions has attracted considerable interest in a wide range of real applications. As an example, identifying sets of genes that are expressing differentially with respect to various treatments has become increasingly popular in recent genetic studies. Statistical evidence on such discoveries can shed light on the underlying biological mechanism and provide helpful guidance on scientific experiments. Since each gene does not work individually but rather tend to function in groups to achieve complex biological tasks, researchers look into gene expression profiles based on groups of genes in a systematic way depending on their functional characteristics. The number of genes in one gene-set, i.e., the dimension of a gene-set, can be much larger than sample size. Therefore, the method we need should be capable of handling high-dimensional data.

The task of identifying differentially expressed gene-sets can be formulated as a hypothesis testing problem on whether gene expression levels of a gene-set are identically distributed across different phenotypes. To test the equality of two distributions, researchers usually consider mean vectors or covariance matrices that characterize most prevalent distributions. In recent years, a number of literature have been dedicated to testing mean vectors (Bai and Saranadasa, 1996; Chen et al., 2010; Cai et al., 2014; Wang et al., 2015; Chen, Li and Zhong, 2019) or testing covariance matrices (Schott, 2007; Li and Chen, 2012; Cai et al., 2013; Chen, Guo and Qiu, 2019) of high dimensions, yet few of them are capable of examining both aspects simultaneously. Both types of discrepancies constitute an important part in inferring significant gene-sets. However, in practice, without sufficient prior knowledge, it would be challenging to specify whether the differences come from mean vectors or covariance matrices. This motivates us to develop one test that is capable of letting us know the existence of a difference no matter the difference comes from the mean aspect or the covariance aspect.

In Chapter 5, we propose a new power-enhanced simultaneous test on two-sample mean vectors and covariance matrices of high-dimensional data. Methodologically, anchored on two quadratic statistics respectively for the means and covariances, we construct two power enhancement components in an effort to boost their respective testing power. After that, the two ingredients are integrated via various combination approaches, forming a new power-enhanced simultaneous test. Theoretically, we investigate the asymptotic independence of two constructed test statistics, and prove that the new power-enhanced simultaneous test is not only powerful in detecting either mean or covariance differences but also powerful under either sparse or dense alternatives. Theoretical analyses prove the accurate asymptotic size and consistent asymptotic power of the proposed test against more general alternatives, and simulation studies demonstrate the finite-sample performance. Moreover, we apply the proposed test in a real application to find differentially expressed gene-sets in cancer studies, which confirms prodigious performance of the proposed test with the support of biological evidence.

1.4 Organization

The rest of this dissertation is organized as follows. In Chapter 2, we provide a literature review of tests on multi-factor pricing models, tests on the equality of two-sample mean vectors, tests on the equality two-sample covariance matrices, and simultaneous tests on the means and covariances. In Chapter 3, we propose our innovated power-enhanced tests for multi-factor pricing models of large panels. In Chapter 4, we introduce our Fisher's combined probability test for two-sample high-dimensional covariance matrices. In Chapter 5, we focus on power enhancement in jointly testing mean vectors and covariance matrices of high-dimensional data. In Chapter 6, we summarize this dissertation and discuss some future work.

Chapter 2

Literature Review

2.1 Testing Multi-Factor Pricing Models

Testing zero pricing errors for linear multi-factor pricing models has received growing attention in empirical finance. The factor pricing model explains market phenomena by employing multiple common factors to model the excess return y_{jt} into the following decomposition:

$$y_{jt} = \theta_j + \mathbf{b}_j'\mathbf{f}_t + \epsilon_{jt}, \quad j = 1, \ldots, p, \quad t = 1, \ldots, T,$$

where p is the number of assets and T is the time dimension, y_{jt} is the excess return of $j-$th asset at time t and $\mathbf{f}_t = (f_{1t}, \ldots, f_{Kt})'$ is a K-dimensional vector of observable market factors. $\mathbf{b}_j = (b_{j1}, \ldots, b_{jK})'$ is a vector of factor loadings and ϵ_{jt} represents the idiosyncratic error.
To be deemed successful, the common factors of a multi-factor pricing model should explain all the expected excess return. In another word, not only any additional factor is redundant, but also that there should be no role for intercepts when regressing the excess returns on the existing factors. A key implication is to test the hypothesis of zero-intercept

$$H_0 : \boldsymbol{\theta} = \mathbf{0},$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ is the vector of intercepts for all p financial assets.
Most existing works address this testing problem by assuming the errors $\boldsymbol{\varepsilon}_t = (\epsilon_{1t}, \cdots, \epsilon_{pt})'$ identically and independently (i.i.d.) follow Gaussian distribution with zero mean vector and covariance matrix matrix $\boldsymbol{\Sigma}^\star$, and construct test statistics based on the Ordinary Least Squares (OLS) estimator of $\boldsymbol{\theta}$. Let $a_{f,T} = 1 - \left(\frac{1}{T}\sum_{t=1}^T \mathbf{f}_t'\right)\left(\frac{1}{T}\sum_{t=1}^T \mathbf{f}_t\mathbf{f}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T \mathbf{f}_t\right)$, the OLS estimator of $\boldsymbol{\theta}$ can be explicitly expressed as

$$\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_p)', \text{ where } \widehat{\theta}_j = \frac{1}{Ta_{f,T}} \sum_{t=1}^T y_{jt} \left[1 - \mathbf{f}_t' \left(\sum_{z=1}^T \mathbf{f}_z \mathbf{f}_z' \right)^{-1} \left(\sum_{z=1}^T \mathbf{f}_z \right) \right].$$

Under the assumption of $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}^\star)$, we have

$$\widehat{\boldsymbol{\theta}} | \{\mathbf{f}_t\}_{t=1}^T \sim N_p \left(\boldsymbol{\theta}, \ \frac{\boldsymbol{\Sigma}^\star}{Ta_{f,T}} \right).$$

2.1.1 Multiple t-tests

Jensen (1968) was the first to propose using individual standard t-statistics to test each θ_i. It is assumed the idiosyncratic errors ϵ_{jt} follows normal distribution. Then the test statistics are derived by

$$t_j = \frac{\widehat{\theta}_j}{se(\widehat{\theta}_j)}, \quad j = 1, \cdots, p.$$

with $\widehat{\theta}_j$ and $se(\widehat{\theta}_j)$ obtained by OLS estimation from time series regressions for each given asset j. Under the null hypothesis $H_0 : \theta_j = 0, \ j = 1, \cdots, p$, each t_j follows a t distribution with degree of freedom $T - K - 1$. However, the cross-sectional dependence of the errors makes it difficult to control the overall size of the test, in the meantime, causes trouble in interpreting the outcomes of the individual tests.

2.1.2 Multivariate F-test

To address the cross-sectional dependence, Gibbons, Ross and Shanken (1989) proposed a multivariate F-test on the intercepts that has exact small-sample properties. With the normal assumption, they construct a quadratic-form statistic

$$GRS = \frac{(T-p-K)a_{f,T}}{p}\widehat{\boldsymbol{\theta}}'\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\theta}}$$

where $\widehat{\boldsymbol{\Sigma}}$ is an unbiased estimation of the residual covariance matrix $\boldsymbol{\Sigma}^{\star}$. It is noteworthy that the condition $p \leqslant T-K-1$ is required to guarantee the invertibility of $\widehat{\boldsymbol{\Sigma}}$. This test has been confined to testing the market efficiency of a relatively small number of portfolios over a relatively long time periods. When dealing with the opposite situation, one feasible approach is to group the assets into portfolios. Yet it is likely to result in loss of information. To illustrate this, suppose that the p assets are grouped into P portfolios using the weights $\mathbf{w}_m$, $m = 1, \cdots, P$, when $P < T$. The test is then applied to the P portfolio excess return defined by $\mathbf{w}_p'\mathbf{y}_t$ so as to test the null hypothesis

$$H_0^p : \mathbf{w}_p'\boldsymbol{\theta} = \mathbf{0}, \quad p = 1, \cdots, P.$$

Compare H_0^p with H_0, one can notice that $\boldsymbol{\theta} = \mathbf{0}$ implies $\mathbf{w}_p'\boldsymbol{\theta} = \mathbf{0}$ for $p = 1, \cdots, P$, but not vice versa. As a result, it is possible that H_0^p is not rejected but H_0 is rejected. In this way, the grouping procedure leads to loss of power.

2.1.3 Operational J_α Test for Large Panels

Pesaran and Yamagata (2012) paid much attention on the development of tests in high-dimensional setting $p > T$. To deal with the inverse of the covariance matrix, they further assume the eigenvalues of $\boldsymbol{\Sigma}^{\star}$ are bounded away from both zero and infinity. Taking the advantage of recent advances from the analysis of large panels, they develop an infeasible quadratic test statistic

$$J(\boldsymbol{\Sigma}^{\star}) = \frac{Ta_{f,T}\widehat{\boldsymbol{\theta}}'(\boldsymbol{\Sigma}^{\star})^{-1}\widehat{\boldsymbol{\theta}} - p}{\sqrt{2p}}$$

which converges to a standard normal distribution as p goes to infinity. To make $J(\boldsymbol{\Sigma}^{\star})$ operational, a suitable estimator of $\boldsymbol{\Sigma}^{\star}$ is desired. It is well-known that the sample covariance matrix is no longer positive definite in high-dimensional scenario so that cannot be an appropriate substitution to make $J(\boldsymbol{\Sigma}^{\star})$ feasible.

Instead of directly estimating the entire p by p matrix, they establish a new statistic based on the diagonal elements of $\mathbf{\Sigma}^\star$, denoted by $J(\mathbf{D}^\star)$, where $\mathbf{D}^\star$ is a diagonal matrix consisting of the same diagonal entries of $\mathbf{\Sigma}^\star$. Moreover, they correct the resultant test statistic for the non-zero off-diagonal elements of $\mathbf{\Sigma}^\star$. Intuitively, the $J(\mathbf{D}^\star)$ serves as a robust version of $J(\mathbf{\Sigma}^\star)$ when the off-diagonal elements of $\mathbf{\Sigma}^\star$ is relatively less important. This condition prevails when the errors are cross-sectionally weakly correlated, and shares the same spirit as the underlying assumption of Chamberlain's approximate factor model (Chamberlain, 1983). It is established that considering $\mathbf{R}^\star$ as the correlation matrix of $\boldsymbol{\varepsilon}_t$, in another word, $\mathbf{R}^\star = (\mathbf{D}^\star)^{-1/2}\mathbf{\Sigma}^\star(\mathbf{D}^\star)^{-1/2}$, then,

$$J(\mathbf{D}^\star) = \frac{T a_{f,T} \widehat{\boldsymbol{\theta}}' (\mathbf{D}^\star)^{-1} \widehat{\boldsymbol{\theta}} - p}{\sqrt{2\text{tr}(\mathbf{R}^2)}}$$

converges in distribution to $N(0,1)$ under the null hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$.
In order to make the above test statistic feasible, they use the diagonal elements of the unbiased sample covariance to estimate $\mathbf{D}^\star$, i.e.,

$$\widehat{\mathbf{D}} = \frac{T}{T-K-1}\text{diag}\left(\widehat{\sigma}_{11}, \ldots, \widehat{\sigma}_{pp}\right), \ \widehat{\sigma}_{jj} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\epsilon}_{jt}^2$$

The coefficient $T/(T-K-1)$ is introduced to correct for small sample bias of the test. As for the estimation of $\text{tr}(\mathbf{R}^2)$, they construct a threshold estimator of the average squares of pair-wise error correlations. Consider

$$\widehat{\overline{\rho^2}} = \frac{2}{p(p-1)}\sum_{i=2}^{p}\sum_{j=1}^{i=1}\widehat{\rho}_{ij}^2 I\{(p-K-1)\widehat{\rho}_{ij}^2 \geqslant \delta_p\}$$

where $\widehat{\rho}_{ij} = \widehat{\boldsymbol{\varepsilon}}_{i\cdot}'\widehat{\boldsymbol{\varepsilon}}_{j\cdot}/\sqrt{\widehat{\boldsymbol{\varepsilon}}_{i\cdot}'\widehat{\boldsymbol{\varepsilon}}_{i\cdot}\widehat{\boldsymbol{\varepsilon}}_{j\cdot}'\widehat{\boldsymbol{\varepsilon}}_{j\cdot}}$, $I(\cdot)$ is an indicator function, and δ_p is the threshold whose value is chosen such that $P\left(\rho_{ij} \neq 0\right)$ declines steadily with p. Their test therefore becomes

$$J_\alpha = \frac{T a_{f,T} \widehat{\boldsymbol{\theta}}' \widehat{\mathbf{D}}^{-1} \widehat{\boldsymbol{\theta}} - \frac{T-K-1}{T-K-3}\cdot p}{\frac{T-K-1}{T-K-3}\sqrt{\frac{2(T-K-2)p}{T-K-5}\left[1 + (p-1)\widehat{\overline{\rho^2}}\right]}} \xrightarrow{d} N(0,1) \text{ under } H_0$$

2.1.4 Power Enhancement Test for Sparse Alternatives

It is well-known that due to the accumulation of errors in estimating high-dimensional parameters, the test based on quadratic forms, for example, the J_α test proposed in Pesaran and Yamagata (2012), often suffer from low power when against sparse alternatives where the null hypothesis is violated by only a few components. To boost the testing power against sparse alternatives, Fan, Liao and Yao (2015) introduced a novel power enhancement test.

Starting with the infeasible quadratic test statistic $J(\mathbf{\Sigma}^\star)$ (Pesaran and Yamagata, 2012), they constructed a feasible quadratic statistic

$$J_{Wald} = \frac{T a_{f,T} \widehat{\boldsymbol{\theta}}' \widehat{\mathbf{\Sigma}}^{-1} \widehat{\boldsymbol{\theta}} - p}{\sqrt{2p}}$$

with $\widehat{\mathbf{\Sigma}}$ being estimated through Fan, Liao and Mincheva (2011). They shows that the effect of replacing $\mathbf{\Sigma}^\star$ by such $\widehat{\mathbf{\Sigma}}$ is asymptotically negligible, namely, under H_0, $T\widehat{\boldsymbol{\theta}}'((\mathbf{\Sigma}^\star)^{-1} - \widehat{\mathbf{\Sigma}}^{-1})\widehat{\boldsymbol{\theta}}/\sqrt{p} = o_p(1)$. Hence,

$$J_{Wald} \xrightarrow{d} N(0,1) \quad \text{as } p, T \to \infty \quad \text{under } H_0.$$

On top of J_{wald}, Fan, Liao and Yao (2015) introduced a power enhancement component J_0, which is zero under the null hypothesis with high probability, but diverges quickly under sparse alternatives. They then proposed a novel statistic $J = J_0 + J_{Wald}$ by combing the power enhancement component J_0 with the asymptotically pivotal statistic J_{Wald} to strengthen the power under sparse alternatives. More formally speaking, they propose J_0 as a screening statistic defined by

$$J_0 = \sqrt{p} \sum_{j \in \widehat{S}} \widehat{\theta}_j^2 \widehat{v}_j^{-1}$$

with $\widehat{S} = \{j : |\widehat{\theta}_j| > \widehat{v}_j^{1/2} \delta_{p,T}, j = 1, ..., p\}$ and $\widehat{v}_j$ denoting a data-dependent normalizing factor, taken as the estimated asymptotic variance of $\widehat{\theta}_j$. The critical value $\delta_{p,T}$ is taken as $c \log(\log T)\sqrt{\log p}$, chosen to be slightly larger than the noise level $\max_{j \leqslant p} |\widehat{\theta}_j - \theta_j|/\widehat{v}_j$. So that $J_0 = 0$ with probability approaching 1. This property guarantees that the asymptotic null distribution of J is determined by that of J_{Wald}, and the size distortion due to adding J_0 is negligible. In addition, it can be seen that $J_0 \geqslant 0$ almost surely, besides, the nonnegativity ensures J is at least as powerful as J_{Wald}.

They demonstrate the asymptotic properties of controlling the size and boosting the power as follows: (1) under the null hypothesis H_0, $P(J_0 = 0|H_0) \to 0$ and hence $J = J_0 + J_{wald} \xrightarrow{d} N(0,1)$ as $T, p \to \infty$; (2) there exists $\Theta(J_{wald})$ and $\Theta(J_{wald})$ such that for any $0 < q < 1$, , $\inf_{\boldsymbol{\theta} \in \Theta_s} P(J_0 > \sqrt{p}|\boldsymbol{\theta}) \to 1$ and $\inf_{\boldsymbol{\theta} \in \Theta(J_{wald})} P(J_{wald} > z_q|\boldsymbol{\theta}) \to 1$ as $T, p \to \infty$. As a result, we have $\inf_{\boldsymbol{\theta} \in \Theta_s \cup \Theta(J_{wald})} P(J > z_q) \to 1$.

2.2 Testing Two-Sample Mean Vectors

Hypothesis testing on mean vectors is a fundamental problem in statistical inference theory and attracts considerable interest in numerous scientific applications. For example, neuroscientists make inferences on the average signals of fMRI data to monitor brain activities and diagnose abnormal tissues. Geneticists analyze gene expression levels between different phenotypes to understand the mechanism of how genes are related to diseases. Of interest is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2.$$

Before proceeding, we first set up some notations. Let $\mathbf{X}$ and $\mathbf{Y}$ be two p-dimensional random vectors with mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and covariance matrices $\boldsymbol{\Sigma}_1 = (\sigma_{ij1})_{p\times p}$ and $\boldsymbol{\Sigma}_2 = (\sigma_{ij2})_{p\times p}$, respectively. Let $\{\mathbf{X}_1, \cdots, \mathbf{X}_{n_1}\}$ be i.i.d. copies of $\mathbf{X}$, and $\{\mathbf{Y}_1, \cdots, \mathbf{Y}_{n_2}\}$ be i.i.d. copies of $\mathbf{Y}$ that are independent of $\{\mathbf{X}_1, \cdots, \mathbf{X}_{n_1}\}$. We assume $n_1/(n_1 + n_2) \to \gamma$ for some $\gamma \in (0, 1)$ as $n_1, n_2 \to \infty$. The related sample covariance matrices are

$$\widehat{\boldsymbol{\Sigma}}_1 = \frac{1}{n_1} \sum_{u=1}^{n_1} (\mathbf{X}_u - \bar{\mathbf{X}})(\mathbf{X}_u - \bar{\mathbf{X}})' := (\widehat{\sigma}_{ij1})_{p\times p}$$
$$\widehat{\boldsymbol{\Sigma}}_2 = \frac{1}{n_2} \sum_{v=1}^{n_2} (\mathbf{Y}_v - \bar{\mathbf{Y}})(\mathbf{Y}_v - \bar{\mathbf{Y}})' := (\widehat{\sigma}_{ij2})_{p\times p}$$

where $\bar{\mathbf{X}} = \frac{1}{n_1} \sum_{u=1}^{n_1} \mathbf{X}_u$ and $\bar{\mathbf{Y}} = \frac{1}{n_2} \sum_{v=1}^{n_2} \mathbf{Y}_v$ are the sample means of $\{\mathbf{X}_u\}_{u=1}^{n_1}$ and $\{\mathbf{Y}_v\}_{v=1}^{n_2}$.

2.2.1 Hotelling's T^2 Test

There is a large body of literature on the topic of mean tests. One of the most well-known approach is the Hotelling's T^2 test. It is the multivariate analogue of two-sample t-test in univariate statistics. Assuming the samples are normally distributed, then under the null hypothesis, the Hotelling's T^2 statistic

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{\mathbf{X}} - \bar{\mathbf{Y}})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \bar{\mathbf{Y}})$$

is distributed as an F distribution with degrees of freedom being p and n_1+n_2-p-1, where $\mathbf{S} = (n_1+n_2-2)^{-1}\left[\sum_{u=1}^{n_1}(\mathbf{X}_u - \bar{\mathbf{X}})(\mathbf{X}_u - \bar{\mathbf{X}})' + \sum_{v=1}^{n_2}(\mathbf{Y}_v - \bar{\mathbf{Y}})(\mathbf{Y}_v - \bar{\mathbf{Y}})'\right]$ is the pooled sample covariance matrix.

The Hotelling's T^2 test prevails in the traditional low-dimensional when p is fixed and small compared to n_1, n_2. However, a growing p brings great challenges for T^2 to function in high dimensions. As p exceeds $n_1 + n_2 - 2$, the sample covariance matrix becomes singular and hence T^2 is not well-defined. Even in the case when $p < n_1 + n_2 - 2$, the testing power of T^2 is largely defective if p diverges proportionally with $\min\{n_1, n_2\}$ (Bai and Saranadasa, 1996).

2.2.2 Quadratic-Form Test

To address the singularity issue of $\mathbf{S}$ in tests of high-dimensional data, many researchers proposed modifications to the Hotelling's T^2 statistic by replacing the singular sample covariance matrix with an invertible matrix. Bai and Saranadasa (1996) proposed to use an identity matrix as a substitution for $\mathbf{S}$. They established the asymptotic distribution of the test statistic

$$T_{BS} = \frac{n_1 n_2}{n_1 + n_2}(\bar{\mathbf{X}} - \bar{\mathbf{Y}})'(\bar{\mathbf{X}} - \bar{\mathbf{Y}}) - \mathrm{tr}(\mathbf{S})$$

where the subtraction of $\mathrm{tr}(\mathbf{S})$ is to make $E(T_{BS}) = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$, and proved the asymptotic normality

$$\frac{T_{BS}}{\sqrt{\frac{2(n_1+n_2-2)(n_1+n_2-1)}{(n_1+n_2-3)(n_1+n_2)}\left(\mathrm{tr}\mathbf{S}^2 - \frac{1}{n_1+n_2-2}\mathrm{tr}^2\mathbf{S}\right)}} \xrightarrow{d} N(0,1) \quad \text{as } n_1, n_2, p \to \infty$$

assuming $p/(n_1+n_2-2) \to c \in (0,\infty)$ under some restrictions on the maximum eigenvalue of covariance matrices.

To alleviate the restrictive assumptions on the relationship between p and (n_1, n_2), Chen and Qin (2010) provided a modification by removing $\sum_{u=1}^{n_1} \mathbf{X}_u'\mathbf{X}_u$ and $\sum_{v=1}^{n_2} \mathbf{Y}_v'\mathbf{Y}_v$ from $\|\bar{\mathbf{X}}-\bar{\mathbf{Y}}\|^2$. The resultant statistic is a U-statistic in the form of

$$T_{CQ} = \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1}(\mathbf{X}_u'\mathbf{X}_v) + \frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_2}(\mathbf{Y}_u'\mathbf{Y}_v) - \frac{2}{n_1n_2}\sum_u^{n_1}\sum_v^{n_2}(\mathbf{X}_u'\mathbf{Y}_v).$$

It is easy to verify that $E(T_{CQ}) = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$. Chen and Qin (2010) proved that under the null hypothesis,

$$T_{CQ}/\widehat{\sigma}_{n1} \xrightarrow{d} N(0,1) \quad \text{as } n_1, n_2, p \to \infty$$

where $\widehat{\sigma}_{n1}^2 = \frac{2}{n_1(n_1-1)}\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)} + \frac{2}{n_2(n_2-1)}\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_2^2)} + \frac{4}{n_1n_2}\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}$. The test impose no explicit restriction between p and (n_1, n_2), but one restriction on covariance matrices that $\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_l\boldsymbol{\Sigma}_h) = o(\mathrm{tr}^2(\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2)^2)$ for $i,j,k,l = 1$ or 2 as $p\to\infty$.

2.2.3 Maximum-Form Test

Both T_{BS} and T_{CQ} are developed based on an estimation of $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$. They can be regarded as a result of transforming the null hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ into an equivalent form $H_0' : \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 = 0$. Cai et al. (2014) considered the problem from a different view by converting H_0 to $H_0'' : \max_{1\leqslant i\leqslant p}|\mu_{1i}-\mu_{2i}| = 0$. They proposed a maximum-form statistic

$$T_{CLX} = \frac{n_1n_2}{n_1+n_2}\max_{1\leqslant i\leqslant p}\frac{\bar{Z}_i^2}{w_{i,i}}$$

where $\bar{\mathbf{Z}} := \boldsymbol{\Omega}(\bar{\mathbf{X}}-\bar{\mathbf{Y}})$, $\overline{\mathbf{W}} := \boldsymbol{\Omega}^{1/2}(\bar{\mathbf{X}}-\bar{\mathbf{Y}})$, in which $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ can be estimated using the CLIME estimator (Cai et al., 2011). They proved that with some regularity conditions, under the null hypothesis, for any x,

$$P\left(T_{CLX} - 2\log p + \log\log p \leqslant x\right) \to \exp\left(-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{x}{2}\right)\right) \qquad \text{as } n_1, n_2, p \to \infty.$$

2.2.4 Projection Test

An emerging way to handle high-dimensional data is via projection. The main idea of projection tests is to map the high-dimensional sample to a low-dimensional space, and subsequently apply traditional methods (e.g., Hotelling's T^2) to the projected sample. Intuitively, the projection procedure seeks to transform the data in such a way that the dimension is reduced, while the statistical distance between the null and alternative hypotheses is mostly preserved through the transformed distributions.

Huang (2015) and Li and Li (2021) proved that when $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, the optimal projection dimension is 1, and the optimal projection direction is $\mathbf{P} = \mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. Let $\widetilde{\mathbf{X}}_u = \mathbf{P}'\mathbf{X}_u$ for $u = 1, \cdots, n_1$, and $\widetilde{\mathbf{Y}}_v = \mathbf{P}'\mathbf{Y}_v$ for $v = 1, \cdots, n_2$. The projection test statistic is defined by a two-sample t-test carried out on $\{\widetilde{\mathbf{X}}_u\}_{u=1}^{n_1}$ and $\{\widetilde{\mathbf{Y}}_v\}_{v=1}^{n_2}$. Huang (2015) proposed a data-splitting strategy to estimate the optimal direction and conduct the test. They randomly partition the data into two disjoint parts, then use one for estimating the projection direction and the other for constructing the test. To address the potentially power loss brought by the data splitting, Li and Li (2021) further proposed a U-projection statistic that is similar to a U-statistic with the kernel of the data-splitting projection test statistic.

2.3 Testing Two-Sample Covariance Matrices

Testing the equality of two covariance matrices is of great importance in multivariate data analysis. Many statistical procedures rely on the fundamental assumption of equal covariance matrices, for example, discriminant analysis, principal component analysis, and clustering. We are interested in testing whether these two covariance matrices are equal. In another word, let $\mathbf{\Sigma} = (\sigma_{ij})_{p\times p}$ denote the common covariance matrix of $\mathbf{X}$ and $\mathbf{Y}$ under the null hypothesis, the problem of interest can be formularized as follows.

$$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$$

2.3.1 Likelihood Ratio Test (LRT)

The conventional likelihood ratio test has been commonly used and well studied in the low-dimensional settings where the dimension is small whereas the sample size is large. The likelihood ratio (LR) for H_0 is

$$\lambda_N = \frac{|\widehat{\mathbf{\Sigma}}_1|^{n_1/2}|\widehat{\mathbf{\Sigma}}_2|^{n_2/2}}{\left|\frac{1}{N}(n_1\widehat{\mathbf{\Sigma}}_1 + n_2\widehat{\mathbf{\Sigma}}_2)\right|^{N/2}}$$

with $N = n_1 + n_2$. When p is held fixed, as $n_1, n_2 \to \infty$, we have the LR statistic converging to the following asymptotic distribution:

$$L_N = -2\log\lambda_N \xrightarrow{d} \chi^2_{p(p+1)/2}$$

Though LRT yields to satisfactory performance under certain regularity conditions(Anderson, 2003), it subjects to two major concerns. On the one hand, the above asymptotic distribution holds only when $p < \min\{n_1, n_2\}$. Once the assumption is violated, at least one of the sample covariance matrices would become singular (Dykstra, 1970). Such singularity leads to consequence of L_n being either infinite or undefined, which fundamentally alters the limiting behavior of the LRT. On the other hand, even though p is smaller than both n_1 and n_2 and fixed as desired, the test is still likely to have distorted size for moderate and large values of p. As pointed out by Bai et al. (2009), modern random matrix theory (RMT) indicates that when both dimension and sample size are large, the LR statistic L_n drifts to infinity almost surely. Therefore, the classical χ^2 approximation leads to many false rejections of H_0, causing severe size distortion problem, in addition, artificially high power.

2.3.2 Corrected Likelihood Ratio Test

To address the aforementioned size distortion problem for large p, Bai et al. (2009) proposed a corrected likelihood ratio test, called CLRT. Employing the central limit theorem for linear spectral statistics from RMT (Bai and Silverstein, 2004), they allow the dimension goes to infinity as sample size grows under the asymptotic scheme $c_{n_1} = p/n_1 \to c_1 \in (0,1)$ and $c_{n_2} = p/n_2 \to c_2 \in (0,1)$ with $\min\{n_1, n_2\} \to \infty$. Based on the result of Zheng (2012), they proposed the CLRT as follows. Under H_0, as $n_1, n_2 \to \infty$,

$$-\frac{2\log\lambda_N}{N} - pF_{c_1,c_2}(f) \xrightarrow{d} N(m(f), \nu(f)),$$

where

$$\begin{aligned}
F_{c_1,c_2}(f) =& \frac{c_{n_1}+c_{n_2}-c_{n_1}c_{n_2}}{c_{n_1}c_{n_2}} \log\left(\frac{c_{n_1}+c_{n_2}}{c_{n_1}+c_{n_2}-c_{n_1}c_{n_2}}\right) \\
&+ \frac{c_{n_1}(1-c_{n_2})}{c_{n_2}(c_{n_1}+c_{n_2})}\log(1-c_{n_2}) + \frac{c_{n_2}(1-c_{n_1})}{c_{n_1}(c_{n_1}+c_{n_1})}\log(1-c_{n_1}), \\
m(f) =& \frac{1}{2}\left[\log\left(\frac{c_1+c_2-c_1c_2}{c_1+c_2}\right) - \frac{c_1}{c_1+c_2}\log(1-c_2) - \frac{c_2}{c_1+c_2}\log(1-c_1)\right], \\
\nu(f) =& -\frac{2c_2^2}{(c_1+c_2)^2}\log(1-c_1) - \frac{2c_1^2}{(c_1+c_2)^2}\log(1-c_2) - 2\log\frac{c_1+c_2}{c_1+c_2-c_1c_2}.
\end{aligned}$$

Jiang et al. (2012) further extended Bai's result to cover the case of $\max\{c_1, c_2\} = 1$.

2.3.3 L_2-norm-based Test

Both LRT and CLRT are limited to the situations where the dimension cannot be larger than the sample size. However, driven by a wide range of contemporary scientific applications, analysis of high-dimensional data is of great importance. Li and Chen (2012) proposed a quadratic-form test that allows the dimension to be diverging and much larger than the sample size. The intuition is to choose a suitable metric to measure the difference between two covariance matrices, and construct a proper estimator to quantify the difference from sample-level of view. So that we are able to conduct a test on the basis of behavior of the estimator. They formulate their quadratic-form test targeting on $\mathrm{tr}\{(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2\}$, the squared Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$. Even though the Frobenius norm has larger magnitude compared with other matrix norms, using it for testing brings two advantages. First, a test statistic based on the Frobenius norm is relatively easier for deriving limiting distribution, which provides theoretical evidence for size and power analysis. From another perspective, using Frobenius norm allows flexible modification

to directly target on a subset of the covariance matrix, which would be hard to accomplish with other norms.
To construct the test statistic, by

$$\operatorname{tr}\{(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2\} = \operatorname{tr}(\mathbf{\Sigma}_1^2) + \operatorname{tr}(\mathbf{\Sigma}_2^2) - 2\operatorname{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2),$$

they propose a test statistic T_{n_1,n_2} in the form of linear combination of unbiased estimators for each term. More explicitly, let

$$\begin{aligned} A_{n_1} =& \frac{1}{n_1(n_1-1)} \sum_{u \neq v} (\mathbf{X}_u'\mathbf{X}_v)^2 - \frac{2}{n_1(n_1-1)(n_1-2)} \sum_{u,v,k}^{\star} \mathbf{X}_u'\mathbf{X}_v\mathbf{X}_v'\mathbf{X}_k \\ &+ \frac{2}{n_1(n_1-1)(n_1-2)(n_1-3)} \sum_{u,v,k,l}^{\star} \mathbf{X}_u'\mathbf{X}_v\mathbf{X}_k'\mathbf{X}_l \end{aligned}$$

$$\begin{aligned} B_{n_2} =& \frac{1}{n_2(n_2-1)} \sum_{u \neq v} (\mathbf{Y}_u'\mathbf{Y}_v)^2 - \frac{2}{n_2(n_2-1)(n_2-2)} \sum_{u,v,k}^{\star} \mathbf{Y}_u'\mathbf{Y}_v\mathbf{Y}_v'\mathbf{Y}_k \\ &+ \frac{2}{n_2(n_2-1)(n_2-2)(n_2-3)} \sum_{u,v,k,l}^{\star} \mathbf{Y}_u'\mathbf{Y}_v\mathbf{Y}_k'\mathbf{Y}_l \end{aligned}$$

$$\begin{aligned} C_{n_1,n_2} =& \frac{1}{n_1 n_2} \sum_{u} \sum_{v} (\mathbf{X}_u'\mathbf{Y}_v)^2 - \frac{1}{n_1 n_2(n_1-1)} \sum_{u,k}^{\star} \sum_{v} \mathbf{X}_u'\mathbf{Y}_v\mathbf{Y}_v'\mathbf{X}_k \\ &- \frac{1}{n_1 n_2(n_2-1)} \sum_{u,k}^{\star} \sum_{v} \mathbf{Y}_u'\mathbf{X}_v\mathbf{X}_v'\mathbf{Y}_k \\ &+ \frac{2}{n_1 n_2(n_1-1)(n_2-1)} \sum_{u,k}^{\star} \sum_{v,l}^{\star} \mathbf{X}_u'\mathbf{Y}_v\mathbf{X}_k'\mathbf{Y}_l \end{aligned}$$

where $\sum^\star$ denotes summation over mutually distinct indices. They show that A_{n_1}, B_{n_2} and C_{n_1,n_2} are unbiased estimators for $\operatorname{tr}(\mathbf{\Sigma}_1^2)$, $\operatorname{tr}(\mathbf{\Sigma}_2^2)$ and $\operatorname{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)$ respectively, and then define the test statistic as

$$T_{n_1,n_2} = A_{n_1} + B_{n_2} - 2C_{n_1,n_2}. \tag{2.3.1}$$

Under certain regularity conditions together with Gaussian assumption, they prove

$$\frac{T_{n_1,n_2} - \text{tr}\{(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2\}}{\sigma_{n_1,n_2}} \xrightarrow{d} N(0,1) \tag{2.3.2}$$

as $\min\{n_1, n_2\} \to \infty$, where

$$\sigma^2_{n_1,n_2} = \sum_{k=1}^{2} \frac{4}{n_k^2}\text{tr}^2(\mathbf{\Sigma}_k^2) + \sum_{k=1}^{2} \frac{8}{n_k}\text{tr}\{(\mathbf{\Sigma}_k^2 - \mathbf{\Sigma}_1\mathbf{\Sigma}_2)^2\} + \frac{8}{n_1 n_2}\text{tr}^2(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)$$

is the leading order variance of T_{n_1,n_2}. Under the null hypothesis, $\sigma^2_{n_1,n_2}$ becomes

$$\sigma^2_{0,n_1,n_2} = 4\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^2 \text{tr}^2(\mathbf{\Sigma}^2).$$

Since A_{n_1} and B_{n_2} are unbiased estimators of $\text{tr}(\mathbf{\Sigma}_1^2)$ and $\text{tr}(\mathbf{\Sigma}_2^2)$, they use

$$\widehat{\sigma}_{0,n_1,n_2} = \frac{2}{n_2}A_{n_1} + \frac{2}{n_1}B_{n_2}$$

to approximate σ_{0,n_1,n_2}. In addition, they show that

$$\frac{A_{n_1}}{\text{tr}(\mathbf{\Sigma}_1^2)} \xrightarrow{p} 1, \quad \frac{B_{n_2}}{\text{tr}(\mathbf{\Sigma}_2^2)} \xrightarrow{p} 1, \text{ and } \quad \frac{\widehat{\sigma}_{0,n_1,n_2}}{\sigma_{0,n_1,n_2}} \xrightarrow{p} 1.$$

As a result, under the null hypothesis,

$$\frac{T_{n_1,n_2}}{\widehat{\sigma}_{0,n_1,n_2}} \xrightarrow{d} N(0,1) \tag{2.3.3}$$

Hence, the proposed test with a nominal significance level α rejects the null hypothesis if

$$T_{n_1,n_2} \geqslant \widehat{\sigma}_{0,n_1,n_2} z_\alpha$$

where z_α is the upper-α quantile of standard normal distribution.

2.3.4 Extreme-value-based Test

Other than using Frobenius norm to measure the difference between two covariance matrices, Cai, Liu and Xia (2013) considered this problem from another point of view. They point out that the null hypothesis $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ is equivalent to $H_0' : \max_{1 \leqslant i \leqslant j \leqslant p} |\sigma_{ij1} - \sigma_{ij2}| = 0$. A natural approach to testing this hypothesis is to compare the sample covariances $\widehat{\sigma}_{ij1}$' and $\widehat{\sigma}_{ij2}$'s and make use of their maximum differences. Before making a comparison among different entries, it is of great importance to standardize $\widehat{\sigma}_{ij1} - \widehat{\sigma}_{ij2}$ so as to eliminate the impact of variability resulting from the heteroscedasticity of $\widehat{\sigma}_{ij1}$'s and $\widehat{\sigma}_{ij2}$'s. To be specific, consider the variances

$$\theta_{ij1} = \text{var}\left((X_i - \mu_{1i})(X_j - \mu_{1j})\right) \quad \text{and} \quad \theta_{ij2} = \text{var}\left((Y_i - \mu_{2i})(Y_j - \mu_{2j})\right)$$

as well as their corresponding estimation

$$\begin{aligned} \widehat{\theta}_{ij1} &= \frac{1}{n_1} \sum_{u=1}^{n_1} \left[(X_{u,i} - \bar{X}_i)(X_{u,j} - \bar{X}_j) - \widehat{\sigma}_{ij1}\right]^2 \\ \text{and} \quad \widehat{\theta}_{ij2} &= \frac{1}{n_2} \sum_{v=1}^{n_2} \left[(Y_{v,i} - \bar{Y}_i)(Y_{v,j} - \bar{Y}_j) - \widehat{\sigma}_{ij2}\right]^2 . \end{aligned}$$

Hence, the variance of $\widehat{\sigma}_{ij1} - \widehat{\sigma}_{ij2}$ can be estimated by $\widehat{\theta}_{ij1}/n_1 + \widehat{\theta}_{ij1}/n_2$. They define the test statistic based on the maximum value of standardized differences between $\widehat{\sigma}_{ij1}$'s and $\widehat{\sigma}_{ij2}$'s, i.e.,

$$M_{n_1,n_2} = \max_{1 \leqslant i \leqslant j \leqslant p} \frac{\left(\widehat{\sigma}_{ij1} - \widehat{\sigma}_{ij2}\right)^2}{\widehat{\theta}_{ij1}/n_1 + \widehat{\theta}_{ij2}/n_2} \tag{2.3.4}$$

They prove that under the null hypothesis and certain regularity conditions,

$$P\left(M_{n_1,n_2} - 4\log p + \log\log p \leqslant x\right) \to \exp\left(-\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{x}{2}\right)\right) \tag{2.3.5}$$

for any x as $n_1, n_2, p \to \infty$. Hence, the proposed test with a nominal significance level α rejects the null hypothesis if

$$M_{n_1,n_2} \geqslant q_\alpha + 4\log p - \log\log p$$

where q_α is the upper-α quantile of the Type I extreme value distribution with the cumulative distribution function.

Sharing the commonality of focusing on the maximum of element-wise differences, Chang et al. (2017) proposed a perturbed-based maximum test in which the rejection region is determined using data-driven approaches, which greatly improved the finite-sample performance of the maximum test (Cai et al., 2013).

2.4 Two-Sample Simultaneous Tests on Mean Vectors and Covariance Matrices

Testing the equality of two distributions is of significant interest in a wide range of real applications. When making inferences on the discrepancies between two distributions, researchers usually consider mean vectors or covariance matrices which characterize most prevalent distributions, for example, the elliptical distributions. We are interested in the simultaneous test of mean vectors and covariance matrices. That is,

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \ \text{ and } \ \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2.$$

2.4.1 Likelihood Ratio Test

In the classical setting with a fixed dimension p, the likelihood ratio test (LRT) has been well studied in the multivariate analysis (Anderson, 2003). With the assumption that samples are normally distributed, the likelihood ratio (LR) statistic

$$\Lambda_{n_1,n_2} = \frac{|\widehat{\boldsymbol{\Sigma}}_1|^{\frac{1}{2}n_1}|\widehat{\boldsymbol{\Sigma}}_2|^{\frac{1}{2}n_2}}{|\widehat{\boldsymbol{\Sigma}}|^{\frac{1}{2}(n_1+n_2)}}$$

asymptotically follows the distribution $-2\log\Lambda_{n_1,n_2} \xrightarrow{d} \chi^2_{p(p+3)/2}$ as $\min\{n_1, n_2\} \to \infty$, where $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}_1 + \widehat{\boldsymbol{\Sigma}}_2 + \frac{n_1 n_2}{n_1+n_2}(\bar{\mathbf{X}} - \bar{\mathbf{Y}})(\bar{\mathbf{X}} - \bar{\mathbf{Y}})'$. Similar to the LRTs for testing mean vectors or covariance matrices separately (see Section 2.2.1 and Section 2.3.1), this test is only valid when p is fixed and p is smaller than $\min\{n_1, n_2\}$. Recent developments in technology facilitates the collection of massive information from a relatively small number of samples, motivating a critical need of developing new methods to cope with high-dimensional data.

2.4.2 Modified Likelihood Ratio Test

In the literature, there only exist a few works on simultaneous tests of mean vectors and covariance matrices in high dimensions. Most of these tests are based on modifications of LRTs. Jiang and Yang (2013) studied a central limit theorem of the LRT statistic with the normal assumption and $p/n_1 \to c_1$, $p/n_2 \to c_2$ for some $0 < c_1, c_2 \leqslant 1$. Under this assumption, Jiang and Yang (2013) proved under the null hypothesis, the LRT statistic would no longer converge in distribution to a chi-square distribution, but rather a normal distribution with explicit mean and covariance. To be more specific, under H_0, they proved

$$\frac{\log \Lambda_{n_1,n_2} - \mu_N}{N\sigma_N} \xrightarrow{d} N(0,1),$$

where $\widetilde{n_1} = n_1 - 1, \widetilde{n_2} = n_2 - 1$, $\mu_N = -\frac{1}{4}\left[-4p - c_1 - c_2 + Nr_N^2(2p - 2N + 3) -n_1 r_{\widetilde{n_1}}^2(2p - 2n_1 + 3) - n_2 r_{\widetilde{n_2}}^2(2p - 2n_2 + 3)\right]$, $\sigma_N^2 = \frac{1}{2}\left[\frac{n_1^2}{N^2} r_{\widetilde{n_1}}^2 + \frac{n_2^2}{N^2} r_{\widetilde{n_2}}^2 - r_N^2\right]$, and $r_x = \sqrt{1 - \log\left(1 - \frac{p}{x}\right)}$ for $x > p$.

The normal assumption was recently relaxed by Niu et al. (2019), where they provided a new modification to the LRT by applying the central limit theorem for linear spectral statistics of sample covariance matrices. To allow p to diverge at a comparable rate as the sample size tends to infinity, i.e., $0 < c < \infty$, Liu et al. (2017) proposed a new approach by replacing the entropy loss with the quadratic loss for covariance matrix estimation. Both works only studied the tests of one single population and may not be easily extended to two-sample tests.

Chapter 3

Innovated Power Enhancement for Testing Multi-Factor Asset Pricing Models

3.1 Introduction

The multi-factor pricing model plays a fundamental role in the arbitrage theory and practice of capital asset pricing (Ross, 1976). It employs multiple common factors to capture the systematic risk and explain financial market phenomena including the co-movements of securities and equilibrium of asset prices. Testing the validity of pricing models and has always been essential to the asset pricing theory and practice. A key implication from the multi-factor pricing theory is that the cross-sectional intercept, i.e. the "alpha" which is commonly used in the finance literature, should be zero*.

The problem of testing zero-alpha can also be interpreted as testing mean-variance efficiency in linear factor pricing models and has attracted considerable attention over the past few decades. About half a century ago, Jensen (1968) proposed a validity test based on standardized t-statistics using ordinary least squares regression for each asset. Gibbons et al. (1989) proposed an exact multivariate F-test under the assumptions that the errors follow a normal distribution and the

*To make the notation consistent, throughout this Chapter, we use $\boldsymbol{\theta}$ to represent the commonly used "alpha" in the finance literature.

number of assets (p) is small compared to the size of the asset return time series (T). Beaulieu et al. (2007) proposed an exact Monte-Carlo-based test to allow for possibly non-Gaussian errors but still requiring $p < T$.

In recent years, the rapidly changing financial market motivates the investigation of a large number of assets over a relatively short time period, where the cross-sectional dimension p is as large as or even larger than the time-series dimension T. Pesaran and Yamagata (2012) proposed a quadratic-form test statistic based on an adaptive thresholding estimator of the error covariance matrix to study a large panel of assets assuming the errors are normally distributed. Gungor and Luger (2013) adopted an adaptive sample-splitting technique to group the assets into one portfolio and proposed two sign-based tests allowing for non-normally distributed disturbances. Gungor and Luger (2016) introduced Monte-Carlo-based distributed-free F-tests for mean-variance efficiency within the framework of multivariate linear regressions.

In practice of examining market efficiency with large panels, the inefficient pricing may be reflected by a small fraction of mis-priced assets or a systematic mispricing phenomenon in the financial market. Overall speaking, the market is generally efficient on the whole. With a large universe of assets, there may exist a small number of inefficiently priced assets from time to time or within a specific sector, yet it is reasonable to posit that the whole market runs efficiently in a systematic way for most of the time. However, during a financial crisis or market correction, the entire market declines drastically over a short period, implying the occurrence of mispricing across a multitude of assets. The former situation implies the existence of a small portion of nonzero-alpha assets, while the latter one implies many (possibly small) nonzero alphas in the cross section. The two types of market inefficiency can be formulated as two different alternative hypotheses in the tests of multi-factor pricing models: (1) *the sparse alternative*, in which the intercept alpha has only a few large nonzero components to imply a few mis-priced assets, and (2) *the dense alternative*, in which the intercept alpha consists of many small nonzero elements to imply the systematic mispricing of assets. The mathematical details about the alternatives will be presented in Section 3.3.

Practically, since we have limited knowledge about the specific sparse alternative or dense alternative in advance, it is important to derive a powerful testing procedure against more general alternatives such as the union of sparse and dense alternatives. As it is known, most existing tests based on the quadratic-form statistics, for example, the Wald test, are powerful against the dense alternative but suffer from low power against the sparse alternative (Pesaran and Yamagata, 2012; Fan et al., 2015; Li and Xue, 2015; Li et al., 2018). Fan et al. (2015) studied the problem of boosting test power against sparse alternatives. They introduced a power enhancement test by adding a screening-based power enhancement component to the Wald test statistic (denoted by S_{Wald}), and proved the following *power enhancement properties*: (i) the power enhancement component is non-negative, (ii) the distortion of asymptotic size is negligible after adding the power enhancement component, and (iii) the asymptotic power of S_{Wald} is enhanced against the designated sparse alternative.

In this work, we provide an innovated power enhancement testing procedure as an alternative to Fan et al. (2015). To boost the power against the sparse alternative, we first introduce a new maximum-form test statistic (denoted by L_{max}) and derive its null limiting distribution given a large number of assets. It is known that such extreme-value-based tests are powerful against the sparse alternative but perform poorly against the dense alternative (Arias-Castro et al., 2011; Zhong et al., 2013; Li and Xue, 2015; Li et al., 2018). To achieve enhanced power against the more general alternatives, we propose a new testing procedure based on the Fisher's method (Fisher, 1925) that combines the respective power of S_{Wald} and L_{max}. To this end, we shall answer the following three questions along the way:

- Q1: What is the joint limiting law of S_{Wald} and L_{max} under the null hypothesis?
- Q2: How to effectively combine S_{Wald} and L_{max} to boost power against the alternatives?
- Q3: What is the possibly optimal combination of S_{Wald} and L_{max}?

Specifically, we study the joint limiting distribution of S_{Wald} and L_{max} under the null hypothesis and prove that they are asymptotically independent when both p and T are diverging. Taking advantage of the asymptotic independence, we propose our innovated power enhancement test based on Fisher's method to combine the asymptotic p-values into one test statistic. Our theoretical studies show that the proposed test is not only in line with the power enhancement properties discussed by Fan et al. (2015), but also achieve more. We prove that the proposed test retains the correct asymptotic size, and what's more, achieves asymptotically consistent power against the union of sparse and dense alternatives. Furthermore, we discuss its optimality in terms of Bahadur efficiency.

In comparison with existing methods in the literature, our proposed test not only achieves theoretical improvements of enhanced power and asymptotic optimality, but also enjoys some practical merits. On the one hand, compared to many existing market efficiency tests (Pesaran and Yamagata, 2012; Gungor and Luger, 2013, 2016), our test yields higher testing power over a wider alternative space. Moreover, the proposed test is more computationally efficient as the limiting distribution can be explicitly derived. On the other hand, our test benefits from the construction of using Fisher's method to combine the p-values. Fisher's method has been widely used in data fusion or meta-analysis for combining p-values from independent studies (Hedges and Olkin, 2014; Singh et al., 2005), however, it receives limited attention in high-dimensional hypothesis testing. Compared to other approaches for gathering and aggregating information, such as the power enhancement test (Fan et al., 2015) and weighted sum test (Li and Xue, 2015; Li et al., 2018), the proposed innovated power enhancement test based on the Fisher's method enjoys the asymptotic optimality with respect to Bahadur efficiency, and avoids the potential scaling issue between the Wald statistic and the maximum statistic.

The rest of this chapter is organized as follows. We first present complete methodological details in Section 3.2, including the new maximum test and the proposed innovated power enhancement test. Section 3.3 studies the asymptotic properties, such as the limiting null distribution, asymptotic power and optimality. Section 3.4 uses simulation studies to demonstrate the finite-sample performance of the proposed tests. Section 3.5 presents an empirical study on testing market efficiency using asset returns from the Russel-2000 portfolio. Section 3.6 includes a few concluding remarks. The technical proofs are presented in Section 3.7. Section 3.8 provides a supplement on generalized positive-definite penalized estimation of large covariance matrices.

3.2 Methodology

Section 3.2.1 sets up the problem of interest. In Section 3.2.2, we present existing methods including the Wald test statistic S_{Wald} and the power enhancement test statistic PE. Then, we introduce a new maximum-form test statistic L_{max} in Section 3.2.3. Motivated by the respective power of S_{Wald} and L_{max} against different alternatives, we propose the innovated power enhancement test in Section 3.2.4.

3.2.1 Problem Setup

Motivated by the Arbitrage Pricing Theory (Ross, 1976), the multi-factor pricing model explains the excess return by employing multiple factors. Let y_{jt} be the excess return of j-th financial asset for $j = 1, \ldots, p$ at time points $t = 1, \ldots, T$, and $\mathbf{f}_t = (f_{1t}, \ldots, f_{Kt})'$ be a K-dimensional vector consisting of observable common factors. More specifically, the excess return y_{jt} is explained through common factors $\mathbf{f}_t$ in a multi-factor pricing model, that is:

$$y_{jt} = \theta_j + \mathbf{b}_j'\mathbf{f}_t + \epsilon_{jt}, \quad j = 1, \ldots, p, \quad t = 1, \ldots, T, \tag{3.2.1}$$

where $\mathbf{b}_j = (b_{j1}, \ldots, b_{jK})'$ is the vector of factor loadings and ϵ_{jt} denotes the idiosyncratic error. As a key implication for tradable factors, the intercept θ_j should be zero for any asset j. Let $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)'$ include the intercepts of all assets. The problem of interest is to test whether a factor pricing model is consistent with the empirical data especially when the number of asset p is larger than time period T in the model. In another word, we are interested in testing the validity of multi-factor pricing models, or equivalently zero intercepts in the null hypothesis:

$$H_0 : \boldsymbol{\theta} = \mathbf{0}. \tag{3.2.2}$$

Before proceeding, we define several helpful notations. For matrix $\mathbf{A}$, use $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ to denote its smallest and largest eigenvalues. Let $\|\mathbf{A}\|_F = \sqrt{\sum_i \sum_j A_{ij}^2}$, $\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})}$, $\|\mathbf{A}\|_\infty = \max_i \sum_j |A_{ij}|$, and $\|\mathbf{A}\|_{\max} = \max_{i,j} |A_{ij}|$ be the Frobenius norm, spectral norm, matrix l_∞ norm and entrywise l_∞ norm of

$\mathbf{A}$, respectively. Define $\mathcal{F}_{-\infty}^{0}$ and $\mathcal{F}_{T}^{\infty}$ as the corresponding σ-algebras generated by $\{\mathbf{f}_t : -\infty \leqslant t \leqslant 0\}$ and $\{\mathbf{f}_t : T \leqslant t \leqslant \infty\}$, and $\alpha(T) = \sup_{A\in\mathcal{F}_{-\infty}^{0}, B\in\mathcal{F}_{T}^{\infty}} |P(A)P(B) - P(AB)|$ as the mixing coefficient. Throughout the paper, we use C to denote a generic positive constant that does not depend on p, T and may differ at difference places.

3.2.2 Existing Methods

To test $H_0 : \boldsymbol{\theta} = \mathbf{0}$ when $p > T$ or $p \gg T$, many existing methods such as Gibbons et al. (1989), MacKinlay and Richardson (1991), and Pesaran and Yamagata (2012) are essentially based on the quadratic-form statistic $\widehat{\boldsymbol{\theta}}'\mathbf{V}\widehat{\boldsymbol{\theta}}$, where $\widehat{\boldsymbol{\theta}}$ is the OLS estimator for $\boldsymbol{\theta}$ and $\mathbf{V}$ is a given positive-definite matrix. Recall that the OLS estimate $\widehat{\boldsymbol{\theta}}$ can be written as

$$\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_p)', \text{ where } \widehat{\theta}_j = \frac{1}{Ta_{f,T}} \sum_{t=1}^{T} y_{jt} \left[1 - \mathbf{f}_t' \left(\sum_{z=1}^{T} \mathbf{f}_z \mathbf{f}_z' \right)^{-1} \left(\sum_{z=1}^{T} \mathbf{f}_z \right) \right]. \tag{3.2.3}$$

where $a_{f,T} = 1 - \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t\mathbf{f}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t\right)$. With the assumption that the errors are homoscedastically normally distributed with no serial correlations among them (Assumption 3.1 presented in Section 3.3.1), we know that $a_{f,T}$ consistently estimate $a_f = 1 - E\mathbf{f}_t'(E\mathbf{f}_t\mathbf{f}_t')^{-1}E\mathbf{f}_t$, and

$$\widehat{\boldsymbol{\theta}}|\{\mathbf{f}_t\}_{t=1}^{T} \sim N_p\left(\boldsymbol{\theta},\ \frac{\boldsymbol{\Sigma}^{\star}}{Ta_{f,T}}\right), \quad \text{under } H_0 : \boldsymbol{\theta} = \mathbf{0}. \tag{3.2.4}$$

Based on (3.2.4), the ideal Wald test statistic $(Ta_{f,T}\widehat{\boldsymbol{\theta}}'\,(\boldsymbol{\Sigma}^{\star})^{-1}\,\widehat{\boldsymbol{\theta}} - p)/\sqrt{2p}$ converges in distribution to $N(0,1)$ as T go to infinity under $H_0 : \boldsymbol{\theta} = \mathbf{0}$. Yet in real-world applications, neither can we observe the error terms nor do we know underlying error covariance matrix $\boldsymbol{\Sigma}^{\star}$. Let $\widehat{\boldsymbol{\Sigma}}_{pd}$ denote a positive-definite estimator of $\boldsymbol{\Sigma}^{\star}$. Fortunately, with some regularity conditions, the effect of replacing $\boldsymbol{\Sigma}^{\star}$ with $\widehat{\boldsymbol{\Sigma}}_{pd}$ is asymptotically negligible under the null hypothesis as long as the estimation by $\widehat{\boldsymbol{\Sigma}}_{pd}$ achieves a satisfactory convergence rate to estimate $\boldsymbol{\Sigma}^{\star}$ (Fan et al., 2015).

Therefore, a feasible Wald test statistic can be obtained by

$$S_{Wald} = \frac{Ta_{f,T}\widehat{\boldsymbol{\theta}}'\widehat{\boldsymbol{\Sigma}}_{pd}^{-1}\widehat{\boldsymbol{\theta}} - p}{\sqrt{2p}}, \tag{3.2.5}$$

which converges in distribution to $N(0,1)$ as both p and T diverge under H_0. The positive-definite estimation and asymptotic null distribution are presented in Section 3.8.

The Wald statistic S_{Wald} is asymptotically pivotal and enjoys the asymptotic normality under the null. However, it is know that S_{Wald} performs poorly against the sparse alternative that consists of a small portion of nonzero values in $\boldsymbol{\theta}$ (Fan et al., 2015). To boost power under the sparse alternative, Fan et al. (2015) introduced a screening-based statistic J_0 as a power enhancement component, where J_0 depends on a high-criticism threshold. Fan et al. (2015) proposed the power enhancement test by adding the component J_0 to S_{Wald}, namely,

$$PE = S_{Wald} + J_0 \tag{3.2.6}$$

which enjoys the *power enhancement properties* that the asymptotic null distribution of PE is determined by S_{Wald} and the asymptotic power of S_{Wald} is improved under the designated sparse alternative. In practice, PE critically depends on the scale of the threshold: a small threshold may fail to control the type-I error, whereas a large threshold would cause a nonignorable loss of power. In Section 3.2.3, we introduce a new maximum test statistic L_{max} as a promising alternative to J_0 without requiring a pre-specified threshold.

3.2.3 A New Maximum Test Statistic

In this subsection, we consider a different approach from Fan et al. (2015) by using the powerful maximum test statistic against the sparse alternative. Over the past decade, there has been much attention to the development of extreme-type test statistics for high-dimensional hypothesis testing. For example, Donoho and Jin (2004) and Hall and Jin (2010) studied the higher criticism statistic and its innovated version for testing sparse normal mean models. Zhong et al. (2013) proposed the maximum-type thresholding tests for high-dimensional means. Cai et al. (2013, 2014) studied the maximum-type test statistics to test the equality of two-sample mean vectors and covariance matrices of high-dimensional data respectively. Chernozhukov et al. (2019) proposed the maximum of inequality-specific statistics to carry out inference on causal and structural parameters using many moment inequalities.

In the sequel, we propose a new maximum statistic to test $H_0 : \boldsymbol{\theta} = \mathbf{0}$ in multi-factor pricing models. Recall that $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_p)'$ is the OLS estimator and $\widehat{\boldsymbol{\theta}}|\{\mathbf{f}_t\}_{t=1}^T \sim N_p\left(\boldsymbol{\theta}, \frac{\boldsymbol{\Sigma}^\star}{Ta_{f,T}}\right)$ under the null hypothesis. For $j = 1, \ldots, p$, the conditional variance of $\widehat{\theta}_j$ (given $\{\mathbf{f}_t\}_{t=1}^T$) converges in probability to $\nu_j = \frac{\sigma^\star_{jj}}{Ta_f}$, which can be estimated by its sample counterpart

$$\widehat{\nu}_j = \frac{1}{Ta_{f,T}} \cdot \frac{1}{T-K-1} \sum_{t=1}^T \widehat{\epsilon}_{jt}^2.$$

Next, we follow the spirit of Cai et al. (2013, 2014) to propose the maximum test statistic after taking the squares of the standardized statistics $\widehat{\theta}_1 \widehat{\nu}_1^{-1/2}, \ldots, \widehat{\theta}_p \widehat{\nu}_p^{-1/2}$ as follows:

$$L_{max} = \max_{1 \leqslant j \leqslant p} \frac{\widehat{\theta}_j^2}{\widehat{\nu}_j} - 2\log p + \log\log p. \tag{3.2.7}$$

We will show that the asymptotic null distribution of L_{max} is an explicit Gumbel distribution

$$G(y) = \exp(-\frac{1}{\sqrt{\pi}} \exp(-\frac{y}{2}))$$

when both p and T go to infinity. To prove this asymptotic null distribution, we make use of the Poisson approximation result from Arratia et al. (1989) and the limiting distribution derived by Berman (1964), and the intermediate limiting distribution recommended by Liu et al. (2008). More details will be provided in Theorem 3.1 of Section 3.2.

Let q_α be the upper α-quantile of the Gumbel distribution $G(y)$. The proposed maximum test rejects the null hypothesis at the significance level α if

$$L_{max} \geqslant q_\alpha.$$

In Theorem 3.2 of Section 3.3.2, we will prove that L_{max} is powerful against the sparse alternative.

3.2.4 Innovated Power Enhancement Test

Now, we are ready to introduce the innovated power enhancement test based on the Wald test statistic S_{Wald} and the new maximum test statistic L_{max}. Before proceeding, we first take a look at the finite-sample performance of S_{Wald} and L_{max} under different alternatives in a simulation study. The simulation data are generated from a three-factor model under both the sparse alternative and the dense alternative, whose details are presented in Section 4. For both test statistics, the percentage of rejection of the null hypothesis is reported based on $1,000$ independent replications at the significance level of 5%. As illustrated in Table 3.1, both S_{Wald} and L_{max} achieve reasonably good empirical size. But, S_{Wald} is powerful against the dense alternative but performs poorly against the sparse alternative, whereas L_{max} is powerful against the sparse alternative but less powerful against the dense alternative.

Table 3.1: Comparison of empirical size and power for L_{max} and S_{Wald} in a simulation study.

	p	T	L_{max}	S_{Wald}
Null Hypothesis	500	100	5.6	5.7
	500	200	5.2	6.4
	500	500	6.1	4.1
Sparse Alternative	500	100	61.5	38.1
	500	200	84.9	29.7
	500	500	99.5	38.7
Dense Alternative	500	100	79.3	97.7
	500	200	83.3	98.8
	500	500	86.1	99.2

Since L_{max} and S_{Wald} have their respective power, we plan to effectively construct a much more powerful test statistic based on an effective and possibly optimal combination of L_{max} and S_{Wald}. In most of existing combination methods, it seems natural to take the weighted sum of L_{max} and S_{Wald} to obtain the weighted test statistic

$$W_{joint} = w_S S_{Wald} + w_L L_{max},$$

for instance, Fan et al. (2015); Li and Xue (2015); Yang and Pan (2017); Li et al. (2018) and others. The asymptotic null distribution of W_{joint} could be determined by S_{Wald} when $w_L L_{max}$ is asymptotically diminishing in probability under the null hypothesis (Fan et al., 2015; Yang and Pan, 2017) or obtained from the convolution of their marginal asymptotic distributions after deriving their joint limiting laws (Li and Xue, 2015; Li et al., 2018).

However, this seemingly natural weighted sum statistic W_{joint} may run into the challenging issue that S_{Wald} and L_{max} may have very different scales. In practice, L_{max} tends to have a much larger scale than S_{Wald} as it aims to capture the most extreme event. Hence, it is more desired to combine L_{max} and S_{Wald} after transforming them into the same scale. To achieve this, we propose the innovated power enhancement testing procedure based on the celebrated Fisher's method (Fisher, 1925) by combining their corresponding asymptotic p-values (i.e., p_{Wald} and p_{max}). Recall that G is the limiting null Gumbel distribution of L_{max} and Φ is the limiting null standard normal distribution of S_{Wald}. To be more specific, we use the Fisher combination to construct the following test statistic:

$$F_{joint} = -2\log(p_{Wald}) - 2\log(p_{max}), \tag{3.2.8}$$

where $p_{Wald} = 1 - \Phi(S_{Wald})$ and $p_{max} = 1 - G(L_{max})$. Due to the fact that p_{Wald} and p_{max} take values between 0 and 1, the proposed test statistic F_{joint} has the desired scale-invariant property. Moreover, the proposed testing procedure based on F_{joint} can be more efficiently computed than the weighted testing procedure based on W_{joint}. After showing the surprisingly asymptotic independence between L_{max} and S_{Wald} under H_0 in Theorem 3.3 of Section 3.3.3, we know that p_{Wald} and p_{max} have the asymptotically independent and identical uniform distribution on $[0, 1]$ under H_0. As a result, $-2\log(p_{Wald})$ and $-2\log(p_{max})$ have asymptotically independent and identical chi-squared distribution with 2 degrees of freedom under H_0. Thus, we have the explicit asymptotic null distribution of F_{joint} as follows:

$$F_{joint} \xrightarrow{d} \chi^2_4 \text{ under } H_0 \tag{3.2.9}$$

when both p and T go to infinity. Let c_α be the upper α-quantile of the χ^2_4 distribution. The proposed innovated power enhancement test rejects H_0 at the significance level α if

$$F_{joint} \geqslant c_\alpha. \tag{3.2.10}$$

Last, but not the least, as we will show in Remark 3.5, the proposed Fisher combination test enjoys the asymptotic optimality with respect to Bahadur efficiency.

3.3 Asymptotic Properties

This section establishes the asymptotic properties for the proposed maximum test statistic L_{max} and innovated power enhancement test statistic F_{joint}. Section 3.3.1 presents regularity assumptions on multi-factor pricing models. Section 3.3.2studies the asymptotic size and power of L_{max}. Section 3.3.3 proves the surprising asymptotic independence between L_{max} and S_{Wald}. Section 3.3.4 studies the asymptotic size and power of F_{joint}.

3.3.1 Assumptions

First, we present the assumptions on the strong mixing condition of the factors $\mathbf{f}_t$ and the normality of $\boldsymbol{\varepsilon}_t = (\epsilon_{1t}, \cdots, \epsilon_{pt})'$ for $t = 1, \cdots, T$.

Assumption 3.1. $\{\boldsymbol{\varepsilon}_t\}_{t\geqslant 1}$ are *i.i.d.* $N_p(\mathbf{0}, \boldsymbol{\Sigma}^\star)$ random vectors. There are constants $c_1, c_2 > 0$ such that $\lambda_{\min}(\boldsymbol{\Sigma}^\star) \geqslant c_1$, $\|\boldsymbol{\Sigma}^\star\|_\infty \leqslant c_2$. Moreover, there exists $0 < r < 1$, such that all the correlations are bounded away from ± 1, that is $\max_{1\leqslant i<j\leqslant p} \frac{|\sigma^\star_{ij}|}{\sqrt{\sigma^\star_{ii}\sigma^\star_{jj}}} \leqslant r < 1$.

Assumption 3.2. Suppose that the factors $\{\mathbf{f}_t\}_{t\geqslant 1}$ satisfy the following conditions:

(i) The factors $\{\mathbf{f}_t\}_{t\geqslant 1}$ are stationary, ergodic, and independent from $\{\boldsymbol{\varepsilon}_t\}_{t\geqslant 1}$.

(ii) There exist $r_1, b_1 > 0$ such that for any $s > 0$ and $j \leqslant K$, $P(|f_{jt}| > s) \leqslant \exp(-(s/b_1)^{r_1})$.

(iii) There exist $r_2, b_2 > 0$ such that $\alpha(t) \leqslant \exp(-b_2 t^{r_2})$ holds for any $t \in \mathbb{Z}^+$.

(iv) There exists $C > 0$, such that $\lambda_{\min}(\text{cov}(\mathbf{f}_t)) > C$.

(v) There exists $M > 0$ such that $Ey_{it}^2 < M$, $Ef_{it}^2 < M$ and $|b_{ij}| < M$ for any i, j and t.

Assumptions 3.1 and 3.2 are commonly used in the literature of multi-factor pricing models. Here, we follow Pesaran and Yamagata (2012) and Fan et al. (2015) to assume the normal distribution of $\{\boldsymbol{\varepsilon}_t\}_{t\geqslant 1}$ in Assumption 3.1, which may be relaxed by semiparametric copula assumption (Xue and Zou, 2012, 2014*a*; Fan et al., 2016). The strong mixing condition has been widely used in the study of factor models such as Fan et al. (2015, 2017), Luo et al. (2021) and Yu et al. (2021).

3.3.2 Asymptotic Size and Power of L_{max}

Now, we derive the asymptotic null distribution for the proposed maximum test statistic L_{max}. To this intent, we shall use the Poisson approximation result from Arratia et al. (1989) and the limiting theorem derived by Berman (1964) to prove the following theorem.

Theorem 3.1. Given Assumptions 3.1–3.2 and $\log p = o(T^{1/3})$, under $H_0 : \boldsymbol{\theta} = \mathbf{0}$, we have

$$P\left(L_{max} \leqslant y\right) \to G(y) \quad \text{as } p, T \to \infty \tag{3.3.1}$$

for any $y \in \mathbb{R}$, where $G(y) = \exp(-\frac{1}{\sqrt{\pi}}\exp(-\frac{y}{2}))$ is the Gumbel distribution.

On top of Theorem 3.1, we can further show that L_{max} is able to control the asymptotic Type-I error and achieves the asymptotically consistent power over the following sparse alternative:

$$\Theta_s = \{\boldsymbol{\theta} :\ \max_{j\leqslant p}\ \sqrt{T}\frac{|\theta_j|}{\sqrt{\sigma^\star_{jj}}} > Ca_f^{-1/2}\sqrt{\log p}\},$$

where $\sigma^\star_{11}, \ldots, \sigma^\star_{pp}$ are the true variances of random errors. Note that the parameter space Θ_s is specified by the maximum norm to represent the sparse alternative (Cai et al., 2013).

Let q_α be the upper α-quantile of the Gumbel distribution $G(y)$. The following theorem proves the correct asymptotic size and consistent asymptotic power of L_{max} against Θ_s.

Theorem 3.2. Under the same assumptions of Theorem 3.1, the maximum test L_{max} achieves

(1) the accurate asymptotic size, that is, under $H_0 : \boldsymbol{\theta} = \mathbf{0}$,

$$P\left(L_{max} \geqslant q_\alpha\right) \to \alpha \quad \text{as} \quad p, T \to \infty,$$

(2) and the consistent asymptotic power uniformly over Θ_s, namely,

$$\inf_{\boldsymbol{\theta} \in \Theta_s} P(L_{max} \geqslant q_\alpha) \to 1 \quad \text{as} \quad p, T \to \infty.$$

Remark 3.1. Fan et al. (2015) used a screening statistic to boost the power against the sparse alternative. Yet the power enhancement test in Fan et al. (2015) critically depends on the sure screening property of the thresholding step to select significant signals in $S(\boldsymbol{\theta}) = \{j : \sqrt{T}\frac{|\theta_j|}{\sqrt{\sigma^\star_{jj}}} > Ca_f^{-1/2}\sqrt{\log p} \cdot \log(\log T)\}$. As shown in Theorem 3.1, Assumption 4.2 and Theorem 4.1 of Fan et al. (2015), the minimal signal condition and the additional factor of $\log(\log T)$ are required to make sure that the high criticism threshold dominates the maximum noise level and the significant signals are retained after the thresholding with high probability.

3.3.3 Joint Limiting Law for S_{Wald} and L_{max}

In what follows, we study the joint limiting law for S_{Wald} and L_{max}. We begin with a simple but important fact that S_{Wald} can be regraded as the sum of dependent variables $\widehat{\boldsymbol{\theta}}'\widehat{\boldsymbol{\Sigma}}_{pd}^{-1}\widehat{\boldsymbol{\theta}}$ and L_{max} is the maximum of dependent variables $\widehat{\theta}_1^2\widehat{\nu}_1^{-1}, \ldots, \widehat{\theta}_p^2\widehat{\nu}_p^{-1}$. The joint limiting law for S_{Wald} and L_{max} is related to the asymptotic joint distribution of the sum and maximum of a sequence of random variables, which has been investigated by several seminal papers in the literature of statistics and probability. In the classical setting that $u_1, \cdots, u_n$ are independently and identically distributed random variables, Chow and Teugels (1978) proved that the sum statistic $\sum_{i=1}^n u_i$ and the maximum statistic $\max_{1 \leqslant i \leqslant n} u_i$ are asymptotically independent when n goes to infinity. Hsing (1995) and Ho and Hsing (1996) proved the asymptotic independence of the sum and the maximum in a more challenging setting that u_i's are strongly mixing stationary random variables or stationary normal random variables. Recently, Li and Xue (2015) and Li et al. (2018) studied the joint limiting laws of the sum and the maximum and proved their asymptotic independence when testing the covariance matrix of high-dimensional data.

The aforementioned existing results do not directly apply to S_{Wald} and L_{max} for testing pricing models since their dependencies between the sum and maximum do not simply fall into the dependence category of Chow and Teugels (1978); Hsing (1995); Ho and Hsing (1996); Li and Xue (2015); Li et al. (2018). Fortunately, we may follow their philosophy to obtain the surprising asymptotic independence between S_{Wald} and L_{max} under the null hypothesis. To this end, we study the asymptotic behavior of the joint distribution $P\big(\{S_{Wald} \leqslant z\} \cap \{L_{max} \geqslant y\}\big)$. Let A denote the event associated with S_{Wald}, and let B be the event corresponding to L_{max}. Define $B_1, \ldots, B_p$ denote the events associated with $\widehat{\theta}_1, \ldots, \widehat{\theta}_p$ respectively. Using a simple fact that $B = \cup_j B_j$, we can rewrite the joint distribution of S_{Wald} and L_{max} as $P\big(A \cap B\big) = P\big(\cup_j (A \cap B_j)\big)$. Note that we can further decompose the sum statistic A into two partial sums, that is, $A = A_{j1} + A_{j2}$, where A_{j1} that is associated B_j and A_{j2} that is independent with B_j. Hence, we can use Bonferroni inequality to bound the difference between $P\big(\{S_{Wald} \leqslant z\} \cap \{L_{max} \geqslant y\}\big)$ and $P\big(\{S_{Wald} \leqslant z\}\big)P\big(\{L_{max} \geqslant y\}\big)$ and then obtain the desired asymptotic independence under the null hypothesis.

In the following theorem, given a positive-definite sparse covariance estimator (see Section 3.8), we prove the convergence of the joint limiting distribution of S_{Wald} and L_{max} to a closed-form distribution over the two-dimensional Euclidean space under H_0.

Theorem 3.3. Let $\widehat{\mathbf{\Sigma}}_{pd}$ be a positive-definite covariance estimator of $\mathbf{\Sigma}^\star$ such that

$$\|\widehat{\mathbf{\Sigma}}_{pd}^{-1} - (\mathbf{\Sigma}^\star)^{-1}\|_{\max} = O_p\left(\sqrt{(\log p)/T}\right).$$

Let $d = \max_{1 \leqslant i \leqslant T} \sum_{j=1}^{T} \mathbb{1}\{\sigma_{ij}^\star \neq 0\}$ denote the maximum number of nonzeros in any row of $\mathbf{\Sigma}^\star$. Suppose $\sqrt{p}(\log p)^2 = o(T)$ and $d = O(1)$, and with Assumptions 3.1 and 3.2, we have

$$\sup_{z,y} \Big| P\big(\{S_{Wald} \leqslant z\} \cap \{L_{max} \geqslant y\}\big) - \Phi(z)\big(1 - G(y)\big)\Big| \xrightarrow{p} 0$$

under $H_0 : \boldsymbol{\theta} = \mathbf{0}$, as $p, T \to \infty$.

Remark 3.2. The asymptotic independence between S_{Wald} and L_{max} immediately follows from the fact that the joint distribution $P\big(\{S_{Wald} \leqslant z\} \cap \{L_{max} \geqslant y\}\big)$

converges to $\Phi(z)\big(1 - G(y)\big)$, which is a product of two marginal asymptotic distribution functions.

Remark 3.3. The asymptotic joint distribution of the quadratic statistic S_{Wald} and the maximum statistic L_{Wald} share similar philosophy with Li and Xue (2015) but differ in many aspects. Li and Xue (2015) studied the joint distribution of a quadratic statistic and a maximum statistics in high-dimensional independence tests, in which they assumed the samples are i.i.d. distributed. Here, the financial returns are not independent but cross-sectionally and longitudinally related, requiring additional efforts in handling such dependency.

Remark 3.4. We will introduce an improved positive-definite estimation of the sparse covariance matrix in Section 3.8. The proposed covariance estimator extends the methodology and theory of Xue et al. (2012) and Fan et al. (2015) to solve a folded-concave penalized estimation problem with the strongly mixing data. As shown in Section 3.8, the unique optimal solution turns out to be the oracle solution $\widehat{\boldsymbol{\Sigma}}^{orcl}$ that is given by

$$\widehat{\boldsymbol{\Sigma}}^{orcl} = \underset{\boldsymbol{\Sigma} \geqslant \delta \mathbf{I}}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}} \right\|_F^2 : \operatorname{supp}(\boldsymbol{\Sigma}) \subseteq \operatorname{supp}(\boldsymbol{\Sigma}^{\star}) \right\}$$

Thus, by analyzing $\widehat{\boldsymbol{\Sigma}}^{orcl}$, we show that the improved positive-definite covariance estimation achieves the desired estimation bound required by Theorem 3.3

3.3.4 Asymptotic Size and Power of F_{joint}

Given the explicit joint limiting law for S_{Wald} and L_{max} in Section 3.3, we study the asymptotic properties of our proposed innovated power enhancement test F_{joint} in the sequel.

Recall that c_α is the upper α quantile of the χ^2_4 distribution and $F_{joint} = -2\log(p_{Wald}) - 2\log(p_{max})$ rejects $H_0 : \boldsymbol{\theta} = \mathbf{0}$ when F_{joint} is as extreme as c_α. In what follows, we prove F_{joint} achieves the accurate asymptotic size in Theorem 3.4.

Theorem 3.4. Under the same assumptions as in Theorem 3.3, the Fisher's combined test achieves the accurate asymptotic size calculation, that is,

$$P(F_{joint} \geqslant c_\alpha) \to \alpha$$

under $H_0 : \boldsymbol{\theta} = \mathbf{0}$, as $p, T \to \infty$.

Moreover, we show that F_{joint} achieves the consistent asymptotic power against the more general alternative $\Theta_s \cup \Theta_d$ in Theorem 3.5, where $\Theta_s = \{\boldsymbol{\theta} : \max_{j \leqslant p} \frac{\sqrt{T}|\theta_j|}{\sigma^{\star}_{jj}} > C a_f^{-1/2} \sqrt{\log p}\}$ denotes the sparse alternative studied in Section 3.2 and

$$\Theta_d = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\|^2 > C\sqrt{p} \log\log p / T\}$$

denotes the designated dense alternative. The parameter space Θ_d uses the Frobenius norm to specify the dense alternative as in Fan et al. (2015); Li and Xue (2015) and Li et al. (2018). Hence, $\Theta_s \cup \Theta_d$ denotes the composite of both sparse and dense alternatives.

Theorem 3.5. Under the same assumptions as in Theorem 3.3, the Fisher's combined test achieves the consistent asymptotic power, that is, as $p, T \to \infty$,

$$\inf_{\boldsymbol{\theta} \in \Theta_s \cup \Theta_d} P(F_{joint} \geqslant c_\alpha) \to 1.$$

Remark 3.5. In addition to the correct size and consistent power, F_{joint} also achieves the asymptotic optimality in terms of Bahadur efficiency. Littell and Folks (1971, 1973) studied the asymptotic optimality of Fisher's methods for combining independent tests. For the two asymptotically independent tests $p_{Wald} = 1 - \Phi(S_{Wald})$ and $p_{max} = 1 - G(L_{max})$, Littell and Folks (1973) proved that the Fisher's method achieves a larger exact Bahadur slope (Bahadur, 1967) than any other reasonable combination procedure. To be more specific, let $T(\cdot, \cdot)$ be any function of (p_{Wald}, p_{max}) that is non-decreasing in each coordinate, and suppose $c_T(\boldsymbol{\alpha})$ and $c_F(\boldsymbol{\alpha})$ are the exact Bahadur slopes for $T(p_{Wald}, p_{max})$ and F_{joint}. As a direct result of Theorem 1 in Littell and Folks (1973), we have $c_T(\boldsymbol{\alpha}) \leqslant c_F(\boldsymbol{\alpha})$. The larger Bahadur slope further implies the Fisher's method gives a faster rate of decay of the p-value under the alternatives. In summary, our proposed innovated power enhancement test is optimal among all combination procedures with respect to Bahadur relative efficiency.

3.4 Numerical Properties

This section examines finite-sample performance of the proposed maximum test and innovated power enhancement test in various simulation models with different sparsity and correlation structures. Section 3.4.1 describes data generating process together with simulation settings, and Section 3.4.2 presents analysis on empirical size and power.

3.4.1 Data Generating Process and Parameter Calibration

The data generating process (DGP) mimics the Fama and French three-factor model (Fama and French, 1993) to describe excess returns:

$$y_{jt} = \theta_j + \mathbf{b}_j'\mathbf{f}_t + \epsilon_{jt}, \quad \text{for } j = 1, \ldots, p;\ t = 1, \ldots, T,$$

where $K = 3$. The factor loadings $\{\mathbf{b}_j\}_{j=1}^p$ are drawn independently from $N_3(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B)$. The factor returns $\mathbf{f}_t$ are generated following a stationary vector autoregressive model $\mathbf{f}_t = \boldsymbol{\mu}_f + \boldsymbol{\Phi}\mathbf{f}_{t-1} + \mathbf{u}_t$ with $\mathbf{u}_t \sim N_3(\mathbf{0}, \boldsymbol{\Sigma}_u)$.

To make the DGP more empirically realistic, we calibrate the model parameters to match those in the three-factor model following similar calibration procedures as in Fan et al. (2011). The values of these mean vectors, covariance matrices, and coefficients are displayed in Tables 3.2 and 3.3. The parameters are estimated using daily returns of Fama-French 3 Factors (i.e., Market Risk, SMB, and HML) and 30 industry portfolios from January 1st, 2018 to December 31st, 2019 ($T = 503$). The datasets are available at Kenneth French's data library. In what follows, we present an outline of our calibration process.

1. Given the excess returns and factors $\{\mathbf{y}_t, \mathbf{f}_t\}_{t=1}^T$ as input data, we calculate the least square estimator $\widehat{\mathbf{B}}$ of $\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \boldsymbol{\varepsilon}_t$, $t = 1, \cdots, T = 503$. We then compute the sample mean vector $\boldsymbol{\mu}_B$ and sample covariance matrix $\boldsymbol{\Sigma}_B$ from row vectors of $\widehat{\mathbf{B}}$.

2. Assuming the factors follow a stationary VAR(1) model $\mathbf{f}_t = \boldsymbol{\mu}_f + \Phi\mathbf{f}_{t-1} + \mathbf{u}_t$ with $\mathbf{u}_t \sim N_3(\mathbf{0}, \boldsymbol{\Sigma}_u)$, we estimate the parameters $\boldsymbol{\mu}_f$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Sigma}_u$ from the data. Note that all eigenvalues of $\boldsymbol{\Phi}$ fall within the unit circle, therefore the

model is stationary. The covariance matrix $\boldsymbol{\Sigma}_f$ can be obtained by solving the linear equation $\boldsymbol{\Sigma}_f = \boldsymbol{\Phi}\boldsymbol{\Sigma}_f\boldsymbol{\Phi}' + \boldsymbol{\Sigma}_u$.

Table 3.2: The calibrated mean vectors and covariance matrices used to generate factor loadings $\{\mathbf{b}_j\}_{j=1}^p$ in the data generating process.

$\boldsymbol{\mu}_B$	$\boldsymbol{\Sigma}_B$		
0.9537	0.0695	0.0599	0.0233
0.1438	0.0599	0.1146	0.0582
0.1788	0.0233	0.0582	0.1113

Table 3.3: The calibrated mean vectors, coefficients and covariance matrices used to generate factors $\{\mathbf{f}_t\}_{t=1}^T$ in the data generating process.

$\boldsymbol{\mu}_f$	$\boldsymbol{\Phi}$			$\boldsymbol{\Sigma}_u$			$\boldsymbol{\Sigma}_f$		
0.0400	-0.0404	0.1547	-0.0488	0.9308	0.0434	-0.1366	0.9387	0.0434	-0.1396
-0.0149	-0.0249	-0.0043	0.0247	0.0434	0.2548	-0.0307	0.0434	0.2558	-0.0315
-0.0383	0.0546	-0.0221	0.0474	-0.1366	-0.0307	0.3453	-0.1396	-0.0315	0.3483

We draw the random shocks $\{\boldsymbol{\varepsilon}_t\}_{t=1}^T$ independently from $N_p(\mathbf{0}, \boldsymbol{\Sigma}^\star)$ and consider two different covariance structures: the independent structure $\boldsymbol{\Sigma}^\star_{(1)} = \mathbf{I}_p$ and the block-diagonal structure $\boldsymbol{\Sigma}^\star_{(2)} = \text{diag}\{\mathbf{B}_1, \cdots, \mathbf{B}_{p/4}\}$. For $i = 1, \ldots, p/4$, each block $\mathbf{B}_i$ has a 4×4 compound symmetry correlation structure whose off-diagonal entry ρ_i is generated from $U[0, 0.5]$. With the intention of investigating test robustness to non-Gaussian errors, we also consider another generating mechanism by letting $\boldsymbol{\varepsilon}_t = (\boldsymbol{\Sigma}^\star)^{1/2}\mathbf{z}_t$, where each element of $\mathbf{z}_t$ follows a standardized t_5 distribution.

To compare the numerical performance of various testing procedures, we consider the null hypothesis H_0 that $\theta_1 = \cdots = \theta_p = 0$ and also two different alternative hypotheses: the sparse and strong alternative H_s and the dense and mild alternative H_d, where

$$H_s: \quad \theta_j = \begin{cases} 0.3 & \text{if } j \leqslant p/T \\ 0 & \text{otherwise} \end{cases}$$

and

$$H_d: \quad \theta_j = \begin{cases} \sqrt{\dfrac{\log p}{T}} & \text{if } j \leqslant p^{0.5} \\ 0 & \text{otherwise} \end{cases}$$

The sparse alternative H_s specifies a few mis-priced assets in the pricing model and favors the extreme-type test statistic, while the dense alternative H_d implies the systematic mis-pricing.

Let $p = 100, 500$ and $T = 100, 200, 500$. For each simulation model, we generate $1,000$ independent replications for each (p, T) pair under the null hypothesis H_0 as well as two alternative hypotheses H_s and H_d respectively. Out of these $1,000$ replications, we compute the frequency of rejections for each test at the significance level of $\alpha = 0.05$. The frequency of rejections under H_0 studies the empirical size and the frequency of rejections under H_s or H_d reflects the empirical power against the sparse or dense alternative hypothesis.

3.4.2 Size and Power Analysis

We consider four testing procedures in simulation studies: the Wald test S_{Wald}, the proposed maximum test L_{max}, the power enhancement test PE, and the proposed Fisher's combined probability test F_{joint}. As suggested by Pesaran and Yamagata (2012), for the purpose of mitigating finite sample biases, we conduct small sample corrections for S_{Wald} and L_{max}. Since the statistic $(\widehat{\theta}_j - \theta_j)^2/\widehat{\nu}_j$ is distributed as a t-distribution with $T - K - 1$ degrees of freedom, under H_0 we have,

$$E\left(\frac{\widehat{\theta}_j^2}{\widehat{\nu}_j}\right) = \frac{T-K-1}{T-K-3}, \quad \text{var}\left(\frac{\widehat{\theta}_j^2}{\widehat{\nu}_j}\right) = \frac{2(T-K-1)^2(T-K-2)}{(T-K-3)^2(T-K-5)}$$

As a result, we conduct small-sample corrections to S_{Wald} and L_{max} as follows,

$$S_{Wald} = \frac{T a_{f,T} \widehat{\boldsymbol{\theta}}' \widehat{\boldsymbol{\Sigma}}_{pd}^{-1} \widehat{\boldsymbol{\theta}} - p \cdot \frac{T-K-3}{T-K-1}}{\sqrt{2p} \cdot \frac{T-K-3}{T-K-1} \cdot \sqrt{\frac{T-K-2}{T-K-5}}} \tag{3.4.1}$$

and

$$L_{max} = \max_{1 \leqslant j \leqslant p} \frac{\widehat{\theta}_j^2}{\widehat{\nu}_j} \cdot \frac{T-K-3}{T-K-1} - 2\log p + \log\log p, \tag{3.4.2}$$

Note that in comparison with definitions (3.2.5) and (3.2.7), the coefficients we add to (3.4.1) and (3.4.2) are converging to 1 as T approaches infinity, hence such corrections do not influence the asymptotic properties established in Section 3. Therefore, the critical values of the tests remain the same. That is, taking account of the small-sample correction, the Wald test rejects H_0 if $S_{Wald} \geqslant z_\alpha$ and the maximum test rejects H_0 if $L_{max} \geqslant q_\alpha$.

Table 3.4 and Table 3.5 present empirical size and power of each testing procedure with normally and non-normally distributed errors for the independent covariance structure $\mathbf{\Sigma}^\star_{(1)}$ and the block-diagonal covariance structure $\mathbf{\Sigma}^\star_{(2)}$, respectively. The empirical results of non-Gaussian errors exhibit a similar pattern to the Gaussian cases. As shown in the tables, the empirical size of S_{Wald} and L_{max} is close to the nominal level 0.05 after the small-sample correction under the null hypothesis. The empirical size of PE and F_{joint} is also close to the nominal level but slightly larger than that of S_{Wald} and L_{max} in these finite-sample cases.

As for the empirical power, we observe that S_{Wald} suffers from low power against the sparse alternative H_s and L_{max} is not powerful against the dense alternative H_d. PE test maintains the power of S_{Wald} against H_d and enhances the power of S_{Wald} against H_s. But, the power enhancement of PE test over S_{Wald} is not significant enough and much less powerful than L_{max} against H_s, especially in high-dimensional setting where p is greater than T. The proposed Fisher's test F_{joint} is built on the combined strength of S_{Wald} and L_{max} to achieve the comparable empirical power as L_{max} against H_s or as S_{Wald} against H_d. To sum up, F_{joint} achieves reasonably well empirical size and the best empirical power among four testing procedures against both sparse alternative H_s and dense alternative H_d.

Table 3.4: The empirical size and power (%) of different testing procedures with Gaussian and non-Gaussian errors and covariance structure $\boldsymbol{\Sigma}^{\star}_{(1)}$.

			Gaussian				Non-Gaussian			
H	p	T	F_{joint}	L_{max}	S_{Wald}	PE	F_{joint}	L_{max}	S_{Wald}	PE
H_0	100	100	7.5	5.6	5.7	6.0	7.8	5.3	6.1	6.2
	100	200	6.7	3.9	5.0	5.0	7.1	4.1	5.7	5.7
	100	500	6.8	5.2	6.0	6.0	7.2	4.1	5.9	5.9
	500	100	6.7	5.1	5.0	5.2	7.1	4.4	5.3	5.5
	500	200	6.8	5.8	4.1	4.1	6.1	4.3	5.6	5.6
	500	500	6.4	4.5	5.1	5.1	6.0	4.7	4.6	4.6
H_s	100	100	37.3	34.5	17.7	26.7	37.7	34.6	18.2	23.9
	100	200	75.4	76.5	36.0	49.7	74.4	75.1	35.1	49.1
	100	500	99.9	100.0	84.9	96.6	99.7	99.8	87.0	94.8
	500	100	70.6	61.8	39.1	45.4	71.3	65.4	39.2	44.5
	500	200	84.4	85.2	28.8	39.7	80.8	81.9	27.9	38.8
	500	500	99.6	99.5	38.9	80.3	99.4	99.7	42.5	78.8
H_d	100	100	88.9	57.8	85.3	85.7	90.1	59.8	86.8	87.3
	100	200	88.4	57.1	86.3	86.5	89.6	58.4	86.3	86.3
	100	500	91.0	59.0	88.4	88.4	89.4	57.7	86.2	86.2
	500	100	99.0	81.9	97.5	97.7	99.2	87.3	98.3	98.4
	500	200	98.9	81.4	97.9	97.9	99.4	84.7	98.9	98.9
	500	500	99.7	85.5	99.3	99.3	99.5	83.5	99.2	99.2

Note: This table reports the frequencies of rejection by each method under the null and alternative hypotheses based on 1000 independent replications at the significance level of 5%.

Table 3.5: The empirical size and power (%) of different testing procedures with Gaussian and non-Gaussian errors and covariance structure $\boldsymbol{\Sigma}^{\star}_{(2)}$.

			Gaussian				Non-Gaussian			
H	p	T	F_{joint}	L_{max}	S_{Wald}	PE	F_{joint}	L_{max}	S_{Wald}	PE
H_0	100	100	7.3	5.1	6.0	6.3	6.4	4.7	5.1	5.1
	100	200	6.1	4.2	5.2	5.2	7.6	5.9	5.9	5.9
	100	500	6.4	4.4	5.6	5.6	6.5	4.2	5.1	5.1
	500	100	6.2	6.2	3.6	3.6	5.8	5.3	3.9	3.9
	500	200	5.6	6.0	3.5	3.5	5.1	4.2	3.5	3.5
	500	500	6.6	5.0	5.9	5.9	6.3	4.0	6.1	6.1
H_s	100	100	34.0	29.8	17.6	24.0	35.2	32.4	17.4	23.1
	100	200	73.6	75.2	34.7	46.6	75.8	73.6	36.1	48.5
	100	500	99.9	100.0	90.2	97.3	99.9	99.9	91.3	96.8
	500	100	59.7	57.0	28.5	34.6	62.8	62.0	28.4	36.7
	500	200	75.2	80.0	19.3	31.9	79.6	83.6	19.0	31.7
	500	500	99.8	99.7	50.1	87.1	99.6	99.4	53.1	82.3
H_d	100	100	79.3	53.5	72.6	72.9	79.0	57.8	71.7	72.7
	100	200	80.4	57.2	71.8	71.8	78.3	56.1	70.2	70.5
	100	500	76.3	51.7	65.9	65.9	78.4	55.7	67.8	67.8
	500	100	94.5	77.4	88.2	88.9	95.6	82.2	88.8	89.5
	500	200	93.3	78.3	84.5	84.5	94.7	81.2	86.1	86.2
	500	500	95.2	77.5	87.5	87.5	95.7	79.3	88.3	88.3

Note: This table reports the frequencies of rejection by each method under the null and alternative hypotheses based on 1000 independent replications at the significance level of 5%.

3.5 Application to Testing Market Efficiency

This section studies the market efficiency test using daily stock returns (y_{jt}) with a broad universe. On each day, our universe is comprised of the symbols in Russel 2000 portfolio that have at least 504-day history, i.e., $p = 2000$. The observable factors ($\mathbf{f}_t$) contain the popular Fama-French 5 factors and 68 industry factors defined by Global Industry Classification Standard (GICS), i.e., $K = 73$. The daily factor returns of Fama-French 5 factors are readily available at Ken French's data library (see `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`). For industry factors, we collect stocks from each industry and build a cap-weighted industry portfolio in the beginning of each month, from which we obtain daily returns of that industry factor for the month. To be more specific, the factor loadings of each industry for each stock is a 0-1 indicator. The industry factor returns are constructed by grouping stocks in the same industry as defined by GICS and calculate their average returns (weighted by their market capitalization). These factors tend to capture most systematic drivers of the asset returns, hence segmenting total returns into common factor returns and specific returns as two distinct sources. Our factor test is to detect whether such factor structure explains the cross-section of asset returns as fully as possible.

We conduct our test each day in a rolling fashion, using the trailing 504-day asset and factor returns, i.e., $T = 504$.. The test results would imply the level of market efficiency in the past two year, so to speak. As individual stock returns are known to be fat-tailed and might have some extreme values (say, larger than 50% in absolute value) on some day, we uniformly cap the daily returns at $\pm 20\%$, roughly their 99% quantile. The Wald test, the maximum test and the Fisher's test are computed on each day, respectively.

Figure 3.1 plots the time series of both quadratic and max test statistics. The test statistics have different scales, but looking at the shape of each one, we see that both tests capture the anomaly of specific asset returns from time to time. The maximum test is more volatile than the quadratic test, as it manages to detect large movement of a single stock, but the two largely go in tandem with each other. It is worth noticing that there is a sharp spike in November, 2011, which can be traced back to the bankruptcy of a stock, MF Global Holdings Ltd, at that time. The maximum test immediately responds to this abnormally large drop, whereas the Wald test fails to detect such anomalies in a single stock.

Figure 3.2 shows the p-values of the Wald, maximum, and joint tests. The joint test successfully detect the market abnormality caused by MFGLQ in November 2011, even though the Wald test fails to do so. Another interesting finding is that the tests reject the model at around 2016, which correspond to the periods when the US financial market experienced significant draw-down. A financial market draw-down means a large number of stocks tumble at the same time. During the crisis, the turmoil often gives rise to mispricing of individual stocks, whose returns cannot be fully explained by common factors. Mathematically, the expectation of cross-sectional intercept may not be zero during the period. As mean-variance efficiency is equivalent to zero cross-sectional intercept, the rejection of mean-variance efficiency is consistent with the financial market draw-down phenomenon in 2016.

Figure 3.1: Time series plot of the quadratic and maximum test statistics

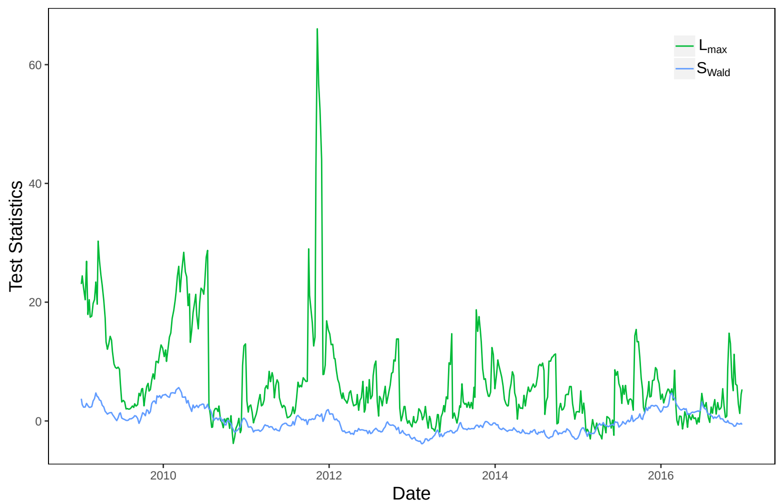
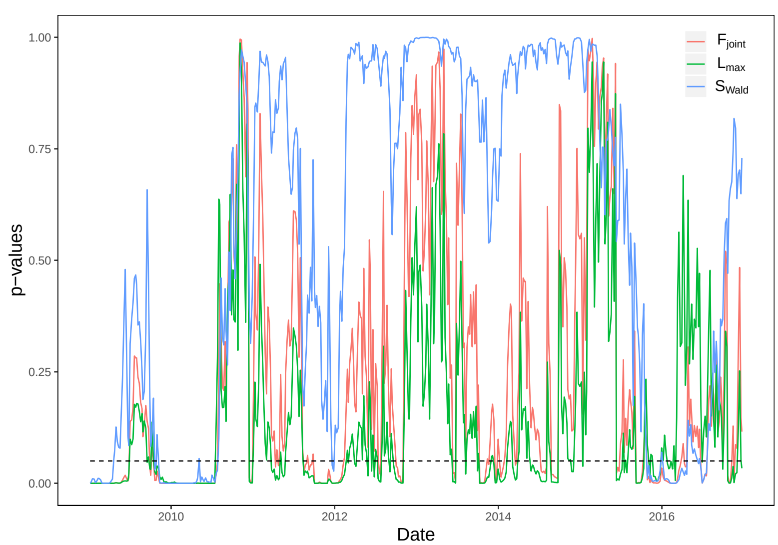
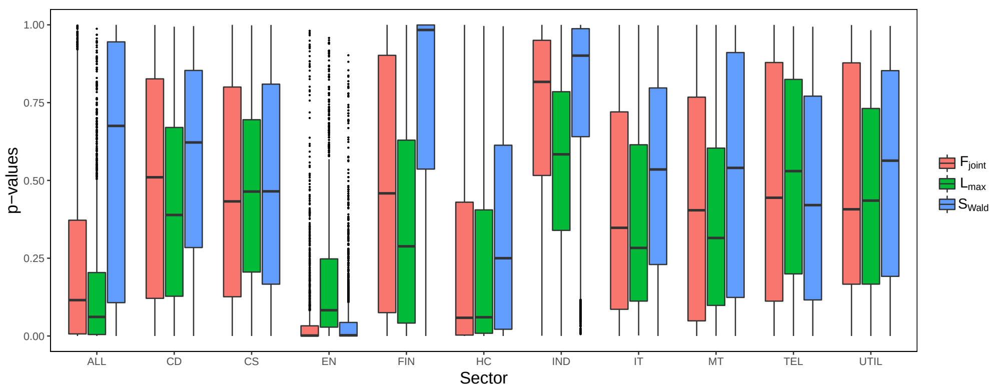


Figure 3.2: Time series plot on p-values of the quadratic, maximum and joint tests



In addition, we observe from Figure 3.2 that the Wald test does not reject the factor model mostly, indicating those factors explain the cross section of the asset returns pretty well in an average sense. In contrast, the maximum test and the Fisher's joint test are more powerful, and reject the model more frequently. This resounds with the findings in Fan et al. (2015) that the sparse alternatives are present in a large universe of stocks, but with a good factor model at hand, it does not happen that often. The Fisher's joint test, as an aggregation of the Wald and the maximum test, is also able to detect such alternatives and produce better testing results.

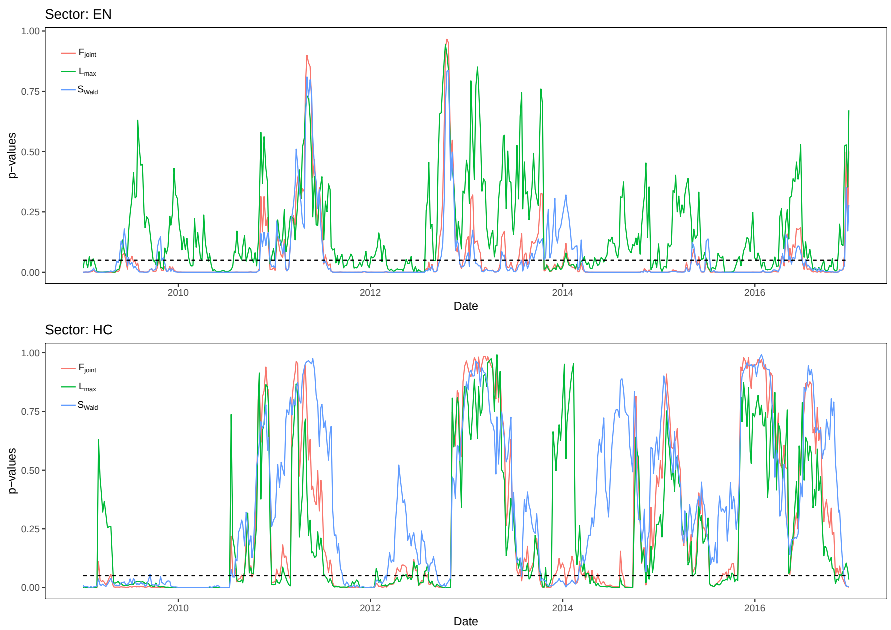
Figure 3.3: Boxplots of p-values of the quadratic, maximum and joint tests for each sector



Sector explanation: All (all sectors), EN (energy), MT (materials), IND (industrials), CD (consumer discretionary), CS (consumer staples), HC (health care), FIN (financials), IT (information technology), TEL (tele-communication services), UTIL (utilities).

Figure 3.3 displays the boxplots of p-values of our tests applied to different industry sectors. We observe that our factor model is also able to explain the cross section of asset returns in each sector. The explanatory powers vary among each sector, e.g., the small p-values of the HC (health care) and the EN (energy) sectors indicate that they are not as efficient as other sectors. What's more, the plot shows that for the EN sector the Wald test yields smaller p-values than the maximum tests, while an opposite situation happens in the HC sector. Such results reflect different market patterns of various sectors, specifically, the market inefficiency within the EN sector mostly originates from systematic mispricing whereas the market inefficiency within the HC sector is more likely caused by a few mis-priced stocks. This agrees with the common sense that energy-related stocks are often systematically impacted by extrinsic factors such as government policy, other than market factors. On the contrary, stocks in the health care sector are rarely influ-

Figure 3.4: p-values of the quadratic, maximum and joint tests for sectors EN and HC



enced by external forces in a systematic manner, but are more often affected on a case-by-case basis such as an approval for a new product developed by a certain company.

Last but not the least, we take a closer look at the time series plot of p-values for EN (energy) and HC (health care) sectors, i.e., Figure 3.4. From the plot, we are able to see different behaviors of the maximum test and the Wald test, as well as our Fisher's joint test. For the energy sector, the Wald test detects strong significance around 2010 and 2012, however, the maximum test only declare significance occasionally. For the health care sector, the maximum test reveals violations of null hypothesis in the year of 2012, whereas the Wald test has relatively large p-values during the entire year. In both cases, one of the maximum and Wald tests fails to detect abnormalities, yet Fisher's joint test successfully declares significance, suggesting that our proposed test is powerful against both sparse and dense alternatives.

3.6 Conclusion

In this work, we provide a promising innovated power enhancement procedure for testing the presence of zero intercepts in multi-factor pricing models. In order to boost power over the traditional Wald test, we introduce a new maximum test statistic and then derive the explicit joint limiting law of the Wald statistic and the proposed maximum statistic. Built on their asymptotic independence, we propose Fisher's combined probability test to combine their respective p-values. Both theoretical and numerical studies show that the proposed method is at least as powerful as the Wald statistic or the proposed maximum statistic to achieve the correct asymptotic size and further enhances their asymptotic power against the more general alternative. Compared with Fan et al. (2015), Li and Xue (2015) and Li et al. (2018), the proposed test provides a scare-invariant, hyperparameter-free, and computationally efficient procedure, which does not need to choose any threshold to ensure the consistency of a thresholding statistic. The numerical performance of the proposed innovated power enhancement test is demonstrated in simulation studies and a real application to testing market efficiency of the Russel 2000 portfolio.

3.7 Lemmas and Proofs

3.7.1 Lemmas

Before proceeding, we give some helpful technical lemmas. Let $\mathbf{Z} = (Z_1, \ldots, Z_p)'$ be a multivariate normal random vector with mean $\boldsymbol{\mu}$, variance $\boldsymbol{\Sigma}$ and $\sigma_{ii} = 1$ for $i = 1, \cdots p$. The following Poisson approximation result is a special case of Theorem 1 from Arratia et al. (1989).

Lemma 3.1. (Arratia et al., 1989) Let I be an index set and $\{B_\alpha, \alpha \in I\}$ be a set of subsets of I, that is, $B_\alpha \subset I$ for each $\alpha \in I$. Let also $\{\eta_\alpha, \alpha \in I\}$ be random variables. For a given $t \in \mathbb{R}$, set $\lambda = \sum_{\alpha \in I} P(\eta_\alpha > t)$. Then

$$|P(\max_{\alpha \in I} \eta_\alpha \leqslant t) - e^{-\lambda}| \leqslant (1 \wedge \lambda^{-1})(b_1 + b_2 + b_3)$$

where

$$b_1 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} P(\eta_\alpha > t) P(\eta_\beta > t),$$
$$b_2 = \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} P(\eta_\alpha > t, \eta_\beta > t),$$
$$b_3 = \sum_{\alpha \in I} E \left| P\left(\eta_\alpha > t | \sigma(\eta_\beta, \beta \notin B_\alpha)\right) - P\left(\eta_\alpha > t\right) \right|,$$

and $\sigma(\eta_\beta, \beta \notin B_\alpha)$ is the σ-algebra generated by $\{\eta_\beta, \beta \notin B_\alpha\}$. In particular, if η_α is independent of $\{\eta_\beta, \beta \notin B_\alpha\}$ for each α, then $b_3 = 0$.

Lemma 3.2. (Berman, 1962) If X and Y follow a bivariate normal distribution with zero mean, unit variance and correlation coefficient ρ, then

$$\lim_{C \to \infty} \frac{P(X > C, Y > C)}{\{2\pi(1-\rho)^{1/2} C^2\}^{-1} \exp\{-C^2/(1+\rho)\}(1+\rho)^{1/2}} = 1,$$

uniformly for all ρ such that $|\rho| \leqslant \delta$ for any $0 < \delta < 1$.

Lemma 3.3. Suppose $\max_{i \neq j} |\sigma_{ij}| \leqslant r < 1$, $\lambda_{\min}(\boldsymbol{\Sigma}) \geqslant c_0$, and $\max_i \sum_{j=1}^p |\sigma_{ij}| \leqslant C_0$ for some constants $c_0, C_0 > 0$. As $p \to \infty$,

$$\sup_y \left| P\left(\max_i (Z_i - \mu_i)^2 - 2\log p + \log\log p \leqslant y\right) - \exp\left(-\lambda_p(y)\right) \right| \to 0.$$

where $\lambda_p(y) = 2p\{1 - \Phi(\sqrt{2\log p - \log\log p + y})\}$.

Proof. Without loss of generality, we assume $\boldsymbol{\mu} = 0$. Define

$$B_i = \{1 \leqslant j \leqslant p : |\sigma_{ij}| \geqslant p^{-(1+r)/2}\}.$$

Let $y_p = \sqrt{2\log p - \log\log p + y}$. By Lemma 3.1, for any $y \in \mathbb{R}$,

$$|P(\max_i |Z_i| \leqslant y_p) - e^{-\lambda_p(y_p)}| \leqslant b_1 + b_2 + b_3 \tag{3.7.1}$$

where $\lambda_p(y_p) = pP(|Z_1| \geqslant y_p)$ and

$$b_1 = \sum_{i=1}^p \sum_{j \in B_i} P(|Z_i| \geqslant y_p) P(|Z_j| \geqslant y_p),$$

$$b_2 = \sum_{i=1}^{p} \sum_{i \neq j \in B_i} P(|Z_i| \geqslant y_p, |Z_j| \geqslant y_p),$$

$$b_3 = \sum_{i=1}^{p} E \left| P\left(|Z_i| \geqslant y_p | \sigma(Z_j, j \notin B_i)\right) - P\left(|Z_i| \geqslant y_p\right) \right|.$$

In the sequel, we bound b_1, b_2 and b_3 respectively. It is straightforward to bound b_1 as

$$b_1 \leqslant Cp \cdot \max_i |B_i| \cdot p^{-2} \leqslant C\ p^{-(1-r)/2} \to 0 \text{ as } p \to \infty.$$

To bound b_2, we define $B_i' = \{j : |\sigma_{ij}| \geqslant \varepsilon\}$ for $\varepsilon = (3-r)/(9+r)$. Then, we have

$$b_2 = \sum_{i=1}^{p} \sum_{i \neq j \in B_i \backslash B_i'} P(|Z_i| \geqslant y_p, |Z_j| \geqslant y_p) + \sum_{i=1}^{p} \sum_{i \neq j \in B_i'} P(|Z_i| \geqslant y_p, |Z_j| \geqslant y_p) \quad (3.7.2)$$

Let $Z_j = \sigma_{ij} Z_i + \sqrt{1-\sigma_{ij}^2} Z_j'$, then we have $Z_j' \perp\!\!\!\perp Z_i$, $Z_j' \sim N(0,1)$. For $j \in B_i \backslash B_i'$,

$$\begin{aligned}
&P\left(|Z_i| \geqslant y_p, |Z_j| \geqslant y_p\right) = P\left(|Z_i| \geqslant y_p, |\sigma_{ij} Z_i + \sqrt{1-\sigma_{ij}^2} Z_j'| \geqslant y_p\right) \\
\leqslant & P\left(|Z_i| \geqslant 2y_p\right) + P\left(y_p \leqslant |Z_i| < 2y_p, |\sigma_{ij}||Z_i| + \sqrt{1-\sigma_{ij}^2} |Z_j'| \geqslant y_p\right) \\
\leqslant & P\left(|Z_i| \geqslant 2y_p\right) + P\left(y_p \leqslant |Z_i| < 2y_p, \sqrt{1-\sigma_{ij}^2} |Z_j'| \geqslant (1-2|\sigma_{ij}|) y_p\right) \\
\leqslant & P\left(|Z_i| \geqslant 2y_p\right) + P\left(|Z_i| \geqslant y_p\right) P\left(|Z_j'| \geqslant \frac{1-2|\sigma_{ij}|}{\sqrt{1-\sigma_{ij}^2}} y_p\right) \\
\leqslant & C(\log p)^{3/2} p^{-4} + Cp^{-1} p^{-(1-2\varepsilon)^2/(1-\varepsilon^2)},
\end{aligned}$$

where the third inequality follows the independence between Z_i and Z_j', and the last inequality is a direct result of the upper bounds for tail probabilities of normal distributed variables. As for $j \in B_i'$, by Lemma 3.2, $P\left(|Z_i| \geqslant y_p, |Z_j| \geqslant y_p\right) \leqslant C(\log p)^{\varepsilon/(1-\varepsilon)} p^{-2/(1+r)}$. Therefore, (3.7.2) becomes

$$\begin{aligned}
b_2 &\leqslant Cp \cdot p^{\frac{1+r}{2}} \cdot \left(C(\log p)^{3/2} p^{-4} + Cp^{-1} p^{-(1-2\varepsilon)^2/(1-\varepsilon^2)}\right) + Cp \cdot (\log p)^{\varepsilon/(1-\varepsilon)} p^{-2/(1+r)} \\
&= C(\log p)^{3/2} p^{-3+\frac{1+r}{2}} + Cp^{-\frac{(1-2\varepsilon)^2}{1-\varepsilon^2}+\frac{1+r}{2}} + C(\log p)^{\frac{\varepsilon}{1-\varepsilon}} p^{1-\frac{2}{1+r}} \to 0 \text{ as } p \to \infty.
\end{aligned}$$

Let $\mathbf{Z}^i$ be the vector consisting of all Z_j with $j \notin B_i$ and the covariance of $\mathbf{Z}^i$ is denoted by $\mathbf{\Sigma}^i$. Let $\boldsymbol{\sigma}^i$ be the covariance between Z_i and $\mathbf{Z}^i$. In addition, let

$U_i = \boldsymbol{\sigma}^i(\boldsymbol{\Sigma}^i)^{-1}\mathbf{Z}^i$ and $V_i = Z_i - U_i$, then $\mathrm{var}\,(U_i) \leqslant c_0^{-1}\sum_{j\notin B_i}\sigma_{ij}^2 \leqslant c_0^{-1}p^{-r}$. It is well known that V_i and $\mathbf{Z}^i$ are independent and $V_i \sim N\left(0, 1-\boldsymbol{\sigma}^i(\boldsymbol{\Sigma}^i)^{-1}(\boldsymbol{\sigma}^i)'\right)$ with $\mathrm{var}\,(V_i) < 1$. It is easy to get that for $i = 1, \cdots, p$,

$$P\left(|U_i| \geqslant 2p^{-r/2}\sqrt{c_0^{-1}\log p}\right) \leqslant Cp^{-2}.$$

Also we have

$$\begin{aligned}
& P\left(|U_i + V_i| \geqslant y_p|U_i\right) - P\left(|U_i + V_i| \geqslant y_p\right) \\
\leqslant\; & P\left(|U_i| + |V_i| \geqslant y_p|U_i\right) - P\left(|V_i| - |U_i| \geqslant y_p\right) \\
=\; & P\left(|V_i| \geqslant y_p - |U_i||U_i\right) - P\left(|V_i| \geqslant y_p - |U_i|\right) + P\left(y_p - |U_i| \leqslant |V_i| < y_p + |U_i|\right) \\
\triangleq\; & g_1(U_i) - Eg_1(U_i) + P\left(y_p - |U_i| \leqslant |V_i| < y_p + |U_i|\right)
\end{aligned}$$

Similarly,

$$\begin{aligned}
& P\left(|U_i + V_i| \geqslant y_p\right) - P\left(|U_i + V_i| \geqslant y_p|U_i\right) \\
\leqslant\; & P\left(|V_i| \geqslant y_p + |U_i|\right) - P\left(|V_i| \geqslant y_p + |U_i||U_i\right) + P\left(y_p - |U_i| \leqslant |V_i| < y_p + |U_i|\right) \\
\triangleq\; & Eg_2(U_i) - g_2(U_i) + P\left(y_p - |U_i| \leqslant |V_i| < y_p + |U_i|\right)
\end{aligned}$$

where $g_1(U_i) = P\left(|V_i| \geqslant y_p - |U_i||U_i\right)$ and $g_2(U_i) = P\left(|V_i| \geqslant y_p + |U_i||U_i\right)$. Now we can bound b_3 as follows:

$$\begin{aligned}
b_3 &= \sum_{i=1}^{p} E\left|P\left(|U_i + V_i| \geqslant y_p|U_i\right) - P\left(|U_i + V_i| \geqslant y_p\right)\right| \\
&\leqslant \max\left\{\sum_{i=1}^{p} E|g_1(U_i) - Eg_1(U_i)|, \sum_{i=1}^{p} E|Eg_2(U_i) - g_2(U_i)|\right\} \\
&\quad + \sum_{i=1}^{p} P\left(y_p - |U_i| \leqslant |V_i| < y_p + |U_i|\right) \\
&\triangleq \max\{I, II\} + III
\end{aligned} \tag{3.7.3}$$

Note that for $k = 1, 2$,

$$I \text{ or } II \leqslant \sum_{i=1}^{p} E\left|g_k(U_i) - Eg_k(U_i)\right| \leqslant \sum_{i=1}^{p}\sqrt{\mathrm{var}\,(g_k(U_i))} \leqslant \sum_{i=1}^{p}\sqrt{E\left(g_k^2(U_i)\right)}$$

$$
\begin{aligned}
&= \sum_{i=1}^{p} P\left(|V_i| \geqslant y_p \pm |U_i|\right) \leqslant \sum_{i=1}^{p}\left[P\left(|V_i| \geqslant y_p - 2p^{-r/2}\sqrt{c_0^{-1}\log p}\right) + Cp^{-2}\right] \\
&\leqslant Cp \cdot \exp\left(-\frac{1}{2\max\{\mathrm{var}(V_i)\}}\left(y_p - 2p^{-r/2}\sqrt{c_0^{-1}\log p}\right)^2\right) + Cp^{-1} \\
&\leqslant Cp^{1-\frac{1}{\max\{\mathrm{var}(V_i)\}}} + Cp^{-1} \qquad (3.7.4)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
III &\leqslant \sum_{i=1}^{p}\left[P\left(y_p - 2p^{-r/2}\sqrt{c_0^{-1}\log p} < |V_i| \leqslant y_p + 2p^{-r/2}\sqrt{c_0^{-1}\log p}\right) + Cp^{-2}\right] \\
&\leqslant Cp \cdot p^{-r/2}\sqrt{\log p} \cdot \exp\left(-\frac{1}{2}\left(y_p - 2p^{-r/2}\sqrt{c_0^{-1}\log p}\right)^2\right) + Cp^{-1} \\
&\leqslant C(\log p)p^{-r/2} + Cp^{-1} \qquad (3.7.5)
\end{aligned}
$$

Combine (3.7.3) (3.7.4) and (3.7.5) together,

$$
b_3 \leqslant \max\{I, II\} + III \to 0 \text{ as } p \to \infty.
$$

The proof of Lemma 3.3 is complete. □

Lemma 3.4. Under the same assumption of Lemma 3.3, $S_z = \frac{1}{\sqrt{2p}}((\mathbf{Z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Z}-\boldsymbol{\mu})-p)$ and $L_z = \max_i(Z_i-\mu_i)^2 - 2\log p + \log\log p$ are asymptotically independent. Namely,

$$
\sup_{z,y}\left|P\big(\{S_z \leqslant z\} \cap \{L_z \geqslant y\}\big) - \Phi(z)\Big(1 - G(y)\Big)\right| \to 0 \text{ as } p \to \infty.
$$

where $G(y) = \exp(-\frac{1}{\sqrt{\pi}}\exp(-\frac{y}{2}))$.

Proof. Without loss of generality, we assume $\boldsymbol{\mu} = 0$. Define $y_p = (2\log p - \log\log p + y)^{1/2}$. By Bonferroni inequality, we have that, for any fixed $k \leqslant [\frac{p}{2}]$,

$$
\sum_{i=1}^{2k}(-1)^{i-1}E_i \leqslant P\left(\{\max_{1\leqslant i\leqslant p}|Z_i| \geqslant y_p\} \cap \{(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} - p)/\sqrt{2p} \leqslant z\}\right) \leqslant \sum_{i=1}^{2k-1}(-1)^{i-1}E_i,
$$

where

$$E_i = \sum_{1\leqslant j_1<\ldots<j_i\leqslant p} P\left(|Z_{j_1}| \geqslant y_p, \ldots, |Z_{j_i}| \geqslant y_p, (\mathbf{Z}'\mathbf{\Sigma}^{-1}\mathbf{Z} - p)/\sqrt{2p} \leqslant z\right).$$

Define $k \leqslant C\sqrt{\log p}$ and $J_1 = (j_1, \ldots, j_i)$. We separate $\mathbf{Z}$ in to two parts $\mathbf{Z}_{J_1} = (Z_{j_1}, \ldots, Z_{j_i})'$ and $\mathbf{Z}_{J_2} = (Z_1, \ldots, Z_p)'\backslash\mathbf{Z}_{J_1}$ with covariance $\mathbf{\Sigma}_{11}$ and $\mathbf{\Sigma}_{22}$ respectively. We call the covariance between $\mathbf{Z}_{J_1}$ and $\mathbf{Z}_{J_2}$ as $\mathbf{\Sigma}_{12}$. Let $\gamma_p = \frac{8\sqrt{2}(\log p)^2}{\sqrt{p}}$ and $\boldsymbol{\tau} = \mathbf{Z}_{J_2} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{Z}_{J_1}$. Then using the simple fact that

$$\mathbf{Z}'\mathbf{\Sigma}^{-1}\mathbf{Z} - p = \mathbf{Z}_{J_1}'\mathbf{\Sigma}_{11}^{-1}\mathbf{Z}_{J_1} + \boldsymbol{\tau}'(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\boldsymbol{\tau} - p,$$

we have

$$\begin{aligned}
& P\left(|Z_{j_1}| \geqslant y_p, \ldots, |Z_{j_i}| \geqslant y_p, \frac{1}{\sqrt{2p}}(\mathbf{Z}'\mathbf{\Sigma}^{-1}\mathbf{Z} - p) \leqslant z\right) \\
\leqslant\ & P\left(|Z_{j_1}| \geqslant y_p, \ldots, |Z_{j_i}| \geqslant y_p\right) P\Big(\frac{1}{\sqrt{2p}}(\boldsymbol{\tau}'(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\boldsymbol{\tau} - p + i) \leqslant z + \gamma_p\Big) \\
& + P\big(|\mathbf{Z}_{J_1}'\mathbf{\Sigma}_{11}^{-1}\mathbf{Z}_{J_1} - i| > \sqrt{2p}\gamma_p\big)
\end{aligned}$$

and

$$\begin{aligned}
& P\left(|Z_{j_1}| \geqslant y_p, \ldots, |Z_{j_i}| \geqslant y_p, \frac{1}{\sqrt{2p}}(\mathbf{Z}'\mathbf{\Sigma}^{-1}\mathbf{Z} - p) \leqslant z\right) \\
\geqslant\ & P\left(|Z_{j_1}| \geqslant y_p, \ldots, |Z_{j_i}| \geqslant y_p\right) P\Big(\frac{1}{\sqrt{2p}}(\boldsymbol{\tau}'(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\boldsymbol{\tau} - p + i) \leqslant z - \gamma_p\Big) \\
& - P\big(|\mathbf{Z}_{J_1}'\mathbf{\Sigma}_{11}^{-1}\mathbf{Z}_{J_1} - i| > \sqrt{2p}\gamma_p\big)
\end{aligned}$$

Next, it is easy to see that

$$P\left(|\mathbf{Z}_{J_1}'\mathbf{\Sigma}_{11}^{-1}\mathbf{Z}_{J_1} - i| > \sqrt{2p}\gamma_p\right) \leqslant C\exp\left(-2(\log p)^2\right).$$

Furthermore, by Berry-Esseen inequality, we have

$$\left|P\left(\frac{\boldsymbol{\tau}'(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\boldsymbol{\tau} - p + i}{\sqrt{2p}} \leqslant z \pm \gamma_p\right) - \Phi\left(\sqrt{\frac{p}{p-i}}(z \pm \gamma_p)\right)\right| \leqslant Cp^{-\frac{1}{2}}.$$

What's more,

$$\left| P\left(\frac{\boldsymbol{\tau}'(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\tau} - p + i}{\sqrt{2p}} \leqslant z \pm \gamma_p \right) - \Phi(z) \right| \leqslant C\gamma_p.$$

Define $V_i = \sum_{1 \leqslant j_1 < \ldots < j_i \leqslant p} P\left(|Z_{j_1}| \geqslant y_p, \ldots, |Z_{j_i}| \geqslant y_p\right)$. Now we can show that

$$\begin{aligned}
& P\left(\{\max_{1\leqslant i\leqslant p} |Z_i| \geqslant y_p\} \cap \{(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} - p)/\sqrt{2p} \leqslant z\} \right) \leqslant \sum_{i=1}^{2k-1} (-1)^{i-1} E_i \\
\leqslant\ & \sum_{i=1}^{2k-1} (-1)^{i-1} V_i \ P\left((\boldsymbol{\tau}'(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\tau} - p + i)/\sqrt{2p} \leqslant z + (-1)^{i-1}\gamma_p \right) \\
& + \sum_{i=1}^{2k-1} \sum_{1\leqslant j_1<\ldots<j_i\leqslant p} P\left(|\mathbf{Z}'_{J_1}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{Z}_{J_1} - i| > \sqrt{2p}\gamma_p \right) \\
\leqslant\ & \Big(\sum_{i=1}^{2k}(-1)^{i-1}V_i + V_{2k}\Big) P\left((\boldsymbol{\tau}'(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\tau} - p + i)/\sqrt{2p} \leqslant z + (-1)^{i-1}\gamma_p \right) \\
& + Cp^{2k-1}\exp(-2(\log p)^2) \\
\leqslant\ & P(\{\max_{1\leqslant i\leqslant p} |Z_i| \geqslant y_p\})\,(\Phi(z) + C\gamma_p) + (\Phi(z) + C\gamma_p)V_{2k} + Cp^{2k-1}\exp(-2(\log p)^2)
\end{aligned}$$

and

$$\begin{aligned}
& P\left(\{\max_{1\leqslant i\leqslant p} |Z_i| \geqslant y_p\} \cap \{(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} - p)/\sqrt{2p} \leqslant z\} \right) \\
\geqslant\ & \sum_{i=1}^{2k}(-1)^{i-1}E_i \\
\geqslant\ & P(\{\max_{1\leqslant i\leqslant p} |Z_i| \geqslant y_p\})\,(\Phi(z) - C\gamma_p) - (\Phi(z) + C\gamma_p)V_{2k} - Cp^{2k}\exp(-2(\log p)^2)
\end{aligned}$$

From the proof of Lemma 6 in Cai et al. (2014), we know that

$$V_{2k} = (1 + o(1))\pi^{-k}p^{-2k}e^{-ky}.$$

If let $p \to \infty$, we complete the proof of Lemma 3.4. □

3.7.2 Proof of Theorem 3.1

Let $y_p = \sqrt{2\log p - \log\log p + y}$, note that

$$\lambda_p(y) \leqslant \frac{2p}{\sqrt{2\pi} y_p} \exp\left(-\frac{y_p^2}{2}\right) = \frac{2\sqrt{\log p}}{\sqrt{2\pi} y_p} \exp\left(-\frac{y}{2}\right).$$

On the other hand,

$$\lambda_p(y) \geqslant \frac{2p}{\sqrt{2\pi} y_p} \left(1 - \frac{1}{y_p^2}\right) \exp\left(-\frac{y_p^2}{2}\right) = \frac{2\sqrt{\log p}}{\sqrt{2\pi} y_p} \exp\left(-\frac{y}{2}\right) - O\left(\frac{1}{\log p}\right)$$

Hence, for any $y \in \mathbb{R}$, as $p \to \infty$,

$$\exp\left(-\lambda_p(y)\right) \to G(y) = \exp\left(-\frac{1}{\sqrt{\pi}} \exp\left(-\frac{y}{2}\right)\right).$$

Together with Lemma 3.3, for any $y \in \mathbb{R}$,

$$P\left(\max_{1\leqslant j\leqslant p} \frac{(\widehat{\theta}_j - \theta_j)^2}{\nu_j} - 2\log p + \log\log p \leqslant y\right) \to G(y) \quad \text{as } p \to \infty. \tag{3.7.6}$$

It suffices to prove

$$\left| P\left(\max_{1\leqslant j\leqslant p} \frac{|\widehat{\theta}_j - \theta_j|}{\nu_j^{1/2}} \leqslant y_p\right) - P\left(\max_{1\leqslant j\leqslant p} \frac{|\widehat{\theta}_j - \theta_j|}{\widehat{\nu}_j^{1/2}} \leqslant y_p\right)\right| \to 0 \quad \text{as } p, T \to \infty. \tag{3.7.7}$$

Note that

$$\max_{1\leqslant j\leqslant p} \frac{|\widehat{\theta}_j - \theta_j|}{\widehat{\nu}_j^{1/2}} \leqslant C\sqrt{\log p} \quad \text{and} \quad \max_{1\leqslant j\leqslant p} \left|\frac{\widehat{\nu}_j^{1/2}}{\nu_j^{1/2}} - 1\right| \leqslant C\sqrt{\frac{\log p}{T}} \tag{3.7.8}$$

hold almost surely. Therefore,

$$\max_{1\leqslant j\leqslant p} \left|\frac{|\widehat{\theta}_j - \theta_j|}{\nu_j^{1/2}} - \frac{|\widehat{\theta}_j - \theta_j|}{\widehat{\nu}_j^{1/2}}\right| \leqslant \max_{1\leqslant j\leqslant p} \frac{|\widehat{\theta}_j - \theta_j|}{\widehat{\nu}_j^{1/2}} \cdot \max_{1\leqslant j\leqslant p} \frac{|\widehat{\nu}_j^{1/2} - \nu_j^{1/2}|}{\nu_j^{1/2}} \leqslant C\sqrt{\frac{(\log p)^2}{T}}. \tag{3.7.9}$$

In addition,

$$P\left(\max_{1\leqslant j\leqslant p}\frac{|\widehat{\theta}_j-\theta_j|}{\widehat{\nu}_j^{1/2}}\leqslant y_p\right)\leqslant P\left(\max_{1\leqslant j\leqslant p}\frac{|\widehat{\theta}_j-\theta_j|}{\nu_j^{1/2}}\leqslant y_p+\frac{C\log p}{\sqrt{T}}\right)$$

and

$$P\left(\max_{1\leqslant j\leqslant p}\frac{|\widehat{\theta}_j-\theta_j|}{\widehat{\nu}_j^{1/2}}\leqslant y_p\right)\geqslant P\left(\max_{1\leqslant j\leqslant p}\frac{|\widehat{\theta}_j-\theta_j|}{\nu_j^{1/2}}\leqslant y_p-\frac{C\log p}{\sqrt{T}}\right).$$

Also,

$$\begin{aligned}&P\left(\max_{1\leqslant j\leqslant p}\frac{|\widehat{\theta}_j-\theta_j|}{\nu_j^{1/2}}\leqslant y_p+\frac{C\log p}{\sqrt{T}}\right)\\&=\exp\left(-2p\left(1-\Phi(y_p+C\log p/\sqrt{T})\right)\right)(1+o(1))\\&=\exp\left(-2p\left(1-\Phi(y_p)\right)\cdot\exp\left(2p\left(\Phi(y_p+C\log p/\sqrt{T})-\Phi(y_p)\right)\right)(1+o(1)),\end{aligned}$$

where

$$\begin{aligned}\Phi(y_p+C\log p/\sqrt{T})-\Phi(y_p)\leqslant\frac{C\log p/\sqrt{T}}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y_p^2\right)\\\leqslant\frac{C\log p/\sqrt{T}}{\sqrt{2\pi}}\exp\left(-\log p+\frac{1}{2}\log\log p+\frac{1}{2}y\right)=o(p^{-1})\end{aligned}$$

as long as $\log p=o(T^{1/3})$. Similarly, we have

$$P\left(\max_{1\leqslant j\leqslant p}\frac{|\widehat{\theta}_j-\theta_j|}{\nu_j^{1/2}}\geqslant y_p-\frac{C\log p}{\sqrt{T}}\right)=\exp\left(-2p\left(1-\Phi(y_p)\right)(1+o(1)\right).$$

This proves (3.7.7) for any $y\in\mathbb{R}$.

3.7.3 Proof of Theorem 3.2

The accurate asymptotic size directly follows the asymptotic null distribution of L_{max} in Theorem 3.1. As for the consistent asymptotic power, following Fan et al. (2015), we have

$$\inf_{\boldsymbol{\theta}}P\left(4/9<\widehat{\nu}_j/\nu_j<9/4,\ \forall j=1,\cdots,p\right)\to 1.$$

hence on Θ_s,

$$\frac{\widehat{\theta}_j^2}{\widehat{\nu}_j} \geqslant \frac{\theta_j^2}{2\widehat{\nu}_j} - \max_{1\leqslant j\leqslant p}\frac{(\widehat{\theta}_j-\theta_j)^2}{\widehat{\nu}_j} \geqslant \frac{2\theta_j^2}{9\nu_j} - \max_{1\leqslant j\leqslant p}\frac{(\widehat{\theta}_j-\theta_j)^2}{\widehat{\nu}_j} > C\log p.$$

We immediately obtain the uniform lower bound that

$$\inf_{\boldsymbol{\theta}\in\Theta_s} P(\max_{1\leqslant j\leqslant p}\frac{\widehat{\theta}_j^2}{\widehat{\nu}_j} - 2\log p + \log\log p \geqslant q_\alpha) \to 1.$$

3.7.4 Proof of Theorem 3.3

Recall that $\widehat{\theta}_j = \frac{1}{Ta_{f,T}}\sum_{t=1}^T y_{jt}(1-\mathbf{f}_t'\mathbf{W})$. Then, we have

$$\widehat{\theta}_j = \theta_j + \frac{1}{Ta_{f,T}}\sum_{t=1}^T \epsilon_{jt}(1-\mathbf{f}_t'\mathbf{W})$$

Given the observed factors $\mathbf{f}_t$, it is easy to see that $\widehat{\boldsymbol{\theta}}$ follows the multivariate normal distribution with mean vector $\boldsymbol{\theta}$ and variance matrix $\boldsymbol{\Sigma}^\star/Ta_{f,T}$. Furthermore,

$$\left|\max_{j\leqslant p}\frac{\widehat{\theta}_j^2}{\widehat{\nu}_j} - \max_{j\leqslant p}\frac{\widehat{\theta}_j^2}{\nu_j}\right| \leqslant \max_{j\leqslant p}\frac{\widehat{\theta}_j^2}{\widehat{\nu}_j}\cdot\max_{j\leqslant p}\left|\frac{\widehat{\nu}_j}{\nu_j}-1\right|. \quad (3.7.10)$$

Together with (3.7.8),

$$\max_{1\leqslant j\leqslant p}\left|\frac{\widehat{\nu}_j}{\nu_j}-1\right| \leqslant \max_{1\leqslant j\leqslant p}\left|\frac{\widehat{\nu}_j^{1/2}+\nu_j^{1/2}}{\nu_j^{1/2}}\right|\cdot\max_{1\leqslant j\leqslant p}\left|\frac{\widehat{\nu}_j^{1/2}-\nu_j^{1/2}}{\nu_j^{1/2}}\right| \leqslant C\sqrt{\frac{\log p}{T}}.$$

Moreover, under the null hypothesis H_0,

$$\max_{1\leqslant j\leqslant p}\frac{\widehat{\theta}_j^2}{\widehat{\nu}_j} \leqslant C\log p.$$

Let $L_{max}^\star = \max_{j\leqslant p}\widehat{\theta}_j^2/\nu_j - 2\log p + \log\log p$. Now, we have

$$|L_{max} - L_{max}^\star| \leqslant C\sqrt{\frac{(\log p)^3}{T}}. \quad (3.7.11)$$

Note that $\|\widehat{\boldsymbol{\Sigma}}_{pd} - \boldsymbol{\Sigma}^{\star}\|_2 \leqslant d\|\widehat{\boldsymbol{\Sigma}}_{pd} - \boldsymbol{\Sigma}^{\star}\|_{\max} = O_p\left(d\sqrt{(\log p)/T}\right)$. Then, Fan et al. (2015) show that under the null hypothesis, we have $|\widehat{\boldsymbol{\theta}}'(\widehat{\boldsymbol{\Sigma}}_{pd}^{-1} - (\boldsymbol{\Sigma}^{\star})^{-1})\widehat{\boldsymbol{\theta}}| = o_p(1)$. Thus, from Lemma 3.4, we obtain the asymptotic independence between S_{Wald} and $L^{\star}_{max}$, that is

$$\sup_{z,y}\Big|P\big(\{S_{Wald} \leqslant z\} \cap \{L^{\star}_{max} \geqslant y\}\big) - \Phi(z)\big(1 - G(y)\big)\Big| \xrightarrow{p} 0 \quad \text{under } H_0.$$

In view of (3.7.11), it further implies the asymptotic independence between S_{Wald} and L_{max} that

$$\sup_{z,y}\Big|P\big(\{S_{Wald} \leqslant z\} \cap \{L_{max} \geqslant y\}\big) - \Phi(z)\big(1 - G(y)\big)\Big| \xrightarrow{p} 0 \quad \text{under } H_0.$$

Therefore, the proof of Theorem 3.3 is complete.

3.7.5 Proof of Theorem 3.4

The accurate asymptotic size directly follows the asymptotic independence of S_{Wald} and L_{max} in Theorem 3.3. The proof of Theorem 3.4 is then complete.

3.7.6 Proof of Theorem 3.5

Now, it remains to prove the consistent asymptotic power. Recall that $F_{joint} = -2\log(1 - G(L_{max})) - 2\log(1 - \Phi(S_{Wald}))$. We can provide the lower bound for the power function as follows.

$$\begin{aligned}
&\inf_{\boldsymbol{\theta}\in\Theta_s\cup\Theta_d} P(F_{joint} \geqslant c_\alpha) \\
\geqslant\ &\min\left\{\inf_{\boldsymbol{\theta}\in\Theta_s} P(F_{joint} \geqslant c_\alpha), \inf_{\boldsymbol{\theta}\in\Theta_d} P(F_{joint} \geqslant c_\alpha)\right\} \\
\geqslant\ &\min\left\{\inf_{\boldsymbol{\theta}\in\Theta_s} P(-2\log(1 - G(L_{max})) \geqslant c_\alpha), \inf_{\boldsymbol{\theta}\in\Theta_d} P(-2\log(1 - \Phi(S_{Wald})) \geqslant c_\alpha)\right\}
\end{aligned}$$

where the second inequality uses the fact that $-2\log(1-G(L_{max}))$ and $-2\log(1-\Phi(S_{Wald}))$ are always non-negative. Following the same arguments as in the proof of Theorem 3.2, we immediately obtain the uniform lower bound that

$$\inf_{\boldsymbol{\theta}\in\Theta_s} P(-2\log(1-G(L_{max}))\geqslant c_\alpha)\to 1.$$

Now, to complete the proof, it remains to show that

$$\inf_{\boldsymbol{\theta}\in\Theta_d} P(-2\log(1-\Phi(S_{Wald}))\geqslant c_\alpha)\to 1.$$

Let $g_\alpha = \Phi^{-1}(1-\exp(-\frac{1}{2}c_\alpha))$, then we consider its equivalent form as

$$P(-2\log(1-\Phi(S_{Wald}))\geqslant c_\alpha) = P\left(\frac{1}{\sqrt{2p}}(Ta_{f,T}\widehat{\boldsymbol{\theta}}'\widehat{\boldsymbol{\Sigma}}_{pd}^{-1}\widehat{\boldsymbol{\theta}}-p)\geqslant g_\alpha\right).$$

Note that S_{Wald} can be rewritten as follows:

$$\begin{aligned} S_{Wald} &= \frac{Ta_{f,T}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\widehat{\boldsymbol{\Sigma}}_{pd}^{-1}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})-p}{\sqrt{2p}}+\frac{\sqrt{2}Ta_{f,T}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\widehat{\boldsymbol{\Sigma}}_{pd}^{-1}\boldsymbol{\theta}}{\sqrt{p}}+\frac{Ta_{f,T}\boldsymbol{\theta}'\widehat{\boldsymbol{\Sigma}}_{pd}^{-1}\boldsymbol{\theta}}{\sqrt{2p}} \\ &:= S_1+S_2+S_3. \end{aligned}$$

To study the asymptotic power, we also introduce

$$S_1^\star = \frac{Ta_{f,T}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\boldsymbol{\Sigma}^\star)^{-1}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})-p}{\sqrt{2p}},$$

$$S_2^\star = \frac{\sqrt{2}Ta_{f,T}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\boldsymbol{\Sigma}^\star)^{-1}\boldsymbol{\theta}}{\sqrt{p}},$$

and

$$S_3^\star = \frac{Ta_{f,T}\boldsymbol{\theta}'(\boldsymbol{\Sigma}^\star)^{-1}\boldsymbol{\theta}}{\sqrt{2p}}.$$

Next, we consider the lower bound

$$P(\frac{Ta_{f,T}\widehat{\boldsymbol{\theta}}'\widehat{\boldsymbol{\Sigma}}_{pd}^{-1}\widehat{\boldsymbol{\theta}}-p}{\sqrt{2p}}\geqslant g_\alpha)\geqslant P(S_2+S_3\geqslant c\log p\log p)-P(S_1\leqslant g_\alpha-c\log p\log p)$$

where c is a small positive constant that does not depend on p and T. Pesaran and Yamagata (2012) showed that $S_1^\star$ converges to a standard normal distribution. Furthermore, Fan et al. (2015) pointed out that the consistent estimation of $(\boldsymbol{\Sigma}^\star)^{-1}$

will guarantee that S_1 also converges to a standard normal distribution. As for S_2 and S_3, note that $p^{1/4}S_2^\star/(2\sqrt{2}S_3^\star)^{1/2}$ follows a standard normal distribution. Then we have

$$\frac{S_2^\star}{p^{-1/4}\sqrt{S_3^\star}} = O_p(1).$$

On the event $\{\widehat{\boldsymbol{\theta}} \neq \boldsymbol{\theta}\}$, we have

$$1 - \frac{||\widehat{\boldsymbol{\Sigma}}_{pd}^{-1} - (\boldsymbol{\Sigma}^\star)^{-1}||_2}{\lambda_{\min}((\boldsymbol{\Sigma}^\star)^{-1})} \leqslant \left|\frac{S_2}{S_2^\star}\right| = \left|\frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\widehat{\boldsymbol{\Sigma}}_{pd}^{-1}\boldsymbol{\theta}}{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\boldsymbol{\Sigma}^\star)^{-1}\boldsymbol{\theta}}\right| \leqslant \frac{||\widehat{\boldsymbol{\Sigma}}_{pd}^{-1} - (\boldsymbol{\Sigma}^\star)^{-1}||_2}{\lambda_{\min}((\boldsymbol{\Sigma}^\star)^{-1})} + 1$$

Therefore, $|S_2/S_2^\star| \xrightarrow{p} 1$. Similarly, $|S_3/S_3^\star| \xrightarrow{p} 1$. On the event $\{\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}\}$, we have $S_2 = S_2^\star = 0$, and hence $S_2/(p^{-1/4}\sqrt{S_3}) = 0$. Then, taking both events into consideration, we have

$$\frac{S_2}{p^{-1/4}\sqrt{S_3}} = O_p(1).$$

Now given that $\boldsymbol{\theta} \in \Theta_d$,

$$\begin{aligned}
S_3 &= \frac{Ta_{f,T}\boldsymbol{\theta}'\widehat{\boldsymbol{\Sigma}}_{pd}^{-1}\boldsymbol{\theta}}{\sqrt{2p}} = \frac{Ta_{f,T}\boldsymbol{\theta}'(\boldsymbol{\Sigma}^\star)^{-1}\boldsymbol{\theta}}{\sqrt{2p}} + \frac{Ta_{f,T}\boldsymbol{\theta}'(\widehat{\boldsymbol{\Sigma}}_{pd}^{-1} - (\boldsymbol{\Sigma}^\star)^{-1})\boldsymbol{\theta}}{\sqrt{2p}} \\
&\geqslant \left(\lambda_{\min}((\boldsymbol{\Sigma}^\star)^{-1}) - \|\widehat{\boldsymbol{\Sigma}}_{pd}^{-1} - (\boldsymbol{\Sigma}^\star)^{-1}\|_2\right)\frac{Ta_{f,T}}{\sqrt{2p}}\|\boldsymbol{\theta}\|^2 \\
&\geqslant C\log\log p
\end{aligned} \tag{3.7.12}$$

where in the last inequality we use the fact that $\|\widehat{\boldsymbol{\Sigma}}_{pd}^{-1} - (\boldsymbol{\Sigma}^\star)^{-1}\|_2 \leqslant d\|\widehat{\boldsymbol{\Sigma}}_{pd}^{-1} - (\boldsymbol{\Sigma}^\star)^{-1}\|_{\max}$. We obtain the following uniform lower bound over Θ_d as

$$\begin{aligned}
&\inf_{\boldsymbol{\theta}\in\Theta_d} P(\frac{Ta_{f,T}\widehat{\boldsymbol{\theta}}'\widehat{\boldsymbol{\Sigma}}_{pd}^{-1}\widehat{\boldsymbol{\theta}} - p}{\sqrt{2p}} \geqslant g_\alpha) \\
\geqslant\ &\inf_{\boldsymbol{\theta}\in\Theta_d} P(S_2 + S_3 \geqslant c\log\log p) - \sup_{\boldsymbol{\theta}\in\Theta_d} P(S_1 \leqslant g_\alpha - c\log\log p).
\end{aligned}$$

where the right hand side goes to 1 as p diverges. Hence,

$$\inf_{\boldsymbol{\theta}\in\Theta_d} P(\frac{Ta_{f,T}\widehat{\boldsymbol{\theta}}'\widehat{\boldsymbol{\Sigma}}_{pd}^{-1}\widehat{\boldsymbol{\theta}} - p}{\sqrt{2p}} \geqslant g_\alpha) \to 1 \quad \text{as } p, T \to \infty.$$

Now, we combine both uniform lower bounds to obtain that

$$\inf_{\boldsymbol{\theta} \in \Theta_s \cup \Theta_d} P(F_{joint} \geqslant c_\alpha) \to 1 \quad \text{as } p, T \to \infty.$$

Therefore, the proof of Theorem 3.5 is complete.

3.8 Supplement on Generalized Positive-Definite Penalized Estimation of Large Covariance Matrices

3.8.1 Methodology

The sparse error variance matrix in factor models has been discussed by Fan et al. (2011), Pesaran and Yamagata (2012) and Gungor and Luger (2016). Since the common factors contributes to dictating the co-movements across the whole panel, it is reasonable to say a particular asset's idiosyncratic shock is correlated with only a few other assets, rather than a large amount of them. To make sure the positive definiteness property, we follow Xue et al. (2012) to consider the following penalized estimation problem

$$\widehat{\boldsymbol{\Sigma}}_{pd} = \operatorname*{argmin}_{\boldsymbol{\Sigma} \geqslant \delta \mathbf{I}} \frac{1}{2} \left\| \boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}} \right\|_F^2 + \sum_{i \neq j} P_\lambda(|\sigma_{ij}|). \tag{3.8.1}$$

where $\delta > 0$ is some arbitrary small value to ensure that the smallest eigenvalue of the estimator is greater than δ, $P_\lambda(|\cdot|)$ is the point-wise penalty function with λ being the tuning parameter, and $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix generated residuals $\widehat{\epsilon}_{jt}$'s, namely,

$$\widehat{\boldsymbol{\Sigma}} = (\widehat{\sigma}_{ij})_{p \times p} = \left(\frac{1}{T - K - 1} \sum_{t=1}^{T} \widehat{\epsilon}_{it} \widehat{\epsilon}_{jt} \right)_{p \times p} \quad \text{where} \quad \widehat{\epsilon}_{jt} = y_{jt} - \widehat{\theta}_j - \widehat{\mathbf{b}}_j' \mathbf{f}_t.$$

3.8.2 Alternating Direction Algorithm

Now the estimation problem becomes an optimization problem. We consider to use the alternating direction method of multipliers (ADMM) to solve the problem directly. We first introduce an additional block $\mathbf{\Theta}$ and an equality constrain to reformulate (3.8.1) as follows.

$$\begin{gathered}(\widehat{\mathbf{\Sigma}}, \widehat{\mathbf{\Theta}}) = \underset{\mathbf{\Sigma},\mathbf{\Theta}}{\operatorname{argmin}} \frac{1}{2} \left\| \mathbf{\Sigma} - \widehat{\mathbf{\Sigma}} \right\|_F^2 + \sum_{i \neq j} P_\lambda(|\sigma_{ij}|) \\ \text{subject to} \quad \mathbf{\Theta} \geqslant \delta \mathbf{I}, \mathbf{\Sigma} = \mathbf{\Theta}.\end{gathered} \tag{3.8.2}$$

The solution to (3.8.2) gives solution to the original problem (3.8.1). To handle the equality constrain, we consider to minimize its augmented Lagrangian function defined as follows.

$$L_\beta(\mathbf{\Theta}, \mathbf{\Sigma}; \mathbf{\Lambda}) = \frac{1}{2} \left\| \mathbf{\Sigma} - \widehat{\mathbf{\Sigma}} \right\|_F^2 + \sum_{i \neq j} P_\lambda(|\sigma_{ij}|) - \langle \mathbf{\Lambda}, \mathbf{\Sigma} - \mathbf{\Theta} \rangle + \frac{\beta}{2} \|\mathbf{\Sigma} - \mathbf{\Theta}\|_F^2$$

where β is a penalty parameter and $\mathbf{\Lambda}$ is the Lagrange multiplier. We iteratively perform the following steps till converge:

- $\mathbf{\Theta}$ step: $\mathbf{\Theta}^{k+1} = \underset{\mathbf{\Theta} \geqslant \delta \mathbf{I}}{\operatorname{argmin}} L_\beta(\mathbf{\Theta}, \mathbf{\Sigma}^k; \Lambda^k)$
- $\mathbf{\Sigma}$ step: $\mathbf{\Sigma}^{k+1} = \underset{\mathbf{\Sigma}}{\operatorname{argmin}} \; L_\beta(\mathbf{\Theta}^{k+1}, \mathbf{\Sigma}; \Lambda^k)$
- $\mathbf{\Lambda}$ step: $\mathbf{\Lambda}^{k+1} = \mathbf{\Lambda}^k - \beta(\mathbf{\Sigma}^{k+1} - \mathbf{\Theta}^{k+1})$

In fact, each step can be further simplified into closed-form solutions. For $k = 0, 1, 2, \cdots$,

- $\mathbf{\Theta}$ step:

$$\begin{aligned}\mathbf{\Theta}^{k+1} &= \underset{\mathbf{\Theta} \geqslant \delta \mathbf{I}}{\operatorname{argmin}} L_\beta(\mathbf{\Theta}, \mathbf{\Sigma}^k; \mathbf{\Lambda}^k) = \underset{\mathbf{\Theta} \geqslant \delta \mathbf{I}}{\operatorname{argmin}} \left\langle \mathbf{\Lambda}^k, \mathbf{\Theta} \right\rangle + \frac{\beta}{2} \left\| \mathbf{\Sigma}^k - \mathbf{\Theta} \right\|_F^2 \\ &= \underset{\mathbf{\Theta} \geqslant \delta \mathbf{I}}{\operatorname{argmin}} \left\| \mathbf{\Theta} - \frac{1}{\beta}(-\mathbf{\Lambda}^k + \beta \mathbf{\Sigma}^k) \right\|_F^2 = \left(-\frac{1}{\beta} \mathbf{\Lambda}^k + \mathbf{\Sigma}^k \right)_{proj}\end{aligned}$$

where $(\mathbf{Z})_{proj}$ is the projection of a matrix $\mathbf{Z}$ onto the convex set $\{\boldsymbol{\Theta} \geqslant \delta \mathbf{I}\}$. Assumen that $\mathbf{Z}$ has the eigen-decomposition $\sum_{j=1}^{d} \lambda_j \mathbf{v}_j \mathbf{v}_j'$, then $(\mathbf{Z})_{proj}$ can be obtained by $\sum_{j=1}^{d} \max(\lambda_j, \delta) \mathbf{v}_j \mathbf{v}_j'$.

- $\boldsymbol{\Sigma}$ step:

$$\begin{aligned}
\boldsymbol{\Sigma}^{k+1} &= \underset{\boldsymbol{\Sigma}}{\operatorname{argmin}}\ L_\beta(\boldsymbol{\Theta}^{k+1}, \boldsymbol{\Sigma}; \boldsymbol{\Lambda}^k) \\
&= \underset{\boldsymbol{\Sigma}}{\operatorname{argmin}} \frac{1}{2} \left\| \boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}} \right\|_F^2 + \sum_{i \neq j} P_\lambda(|\sigma_{ij}|) - \left\langle \boldsymbol{\Lambda}^k, \boldsymbol{\Sigma} \right\rangle + \frac{\beta}{2} \left\| \boldsymbol{\Sigma} - \boldsymbol{\Theta}^{k+1} \right\|_F^2 \\
&= \underset{\boldsymbol{\Sigma}}{\operatorname{argmin}} \frac{1}{2} \left\| \boldsymbol{\Sigma} - \frac{\widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Lambda}^k + \beta \boldsymbol{\Theta}^{k+1}}{1+\beta} \right\|_F^2 + \frac{1}{1+\beta} \sum_{i \neq j} P_\lambda(|\sigma_{ij}|) \\
&= \mathbf{M}_{pen}\left(\frac{1}{1+\beta} \left(\widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Lambda}^k + \beta \boldsymbol{\Theta}^{k+1} \right),\ \lambda\ ,\ \frac{1}{1+\beta} \right)
\end{aligned}$$

where $\mathbf{M}_{pen}(\mathbf{A}, \lambda, w)$ is an entry-wise thresholding rule for $\mathbf{A} = \{a_{ij}\}_{1 \leqslant i,j \leqslant p}$ in terms of the specific penalty, with λ being tuning parameter and w being the weight of regularizers, which is defined by

$$\{\mathbf{M}_{pen}(\mathbf{A}, \lambda, w)\}_{ij} = \begin{cases} m_{pen}(a_{ij}, \lambda, w) & \text{for } i \neq j \\ a_{ij} \quad \text{for } i = j \end{cases}$$

Here, $m_{pen}(a_{ij}, \lambda, w) := \operatorname{argmin}_z \frac{1}{2}(z - a_{ij})^2 + w P_\lambda(|a_{ij}|)$ has closed-form solution, whose explicit form depends on the penalty function $P_\lambda(\cdot)$.

Before implementing the algorithm, it is always better to check the unconstrained optimal solution $\widetilde{\boldsymbol{\Sigma}}$ first, where $\widetilde{\boldsymbol{\Sigma}}$ is the solution to (3.8.2) and can be simplified into closed form for given penalty.

$$\widetilde{\boldsymbol{\Sigma}} = \underset{\boldsymbol{\Sigma}}{\operatorname{argmin}} \frac{1}{2} \left\| \boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}} \right\|_F^2 + \sum_{i \neq j} p_\lambda(|\sigma_{ij}|)$$

If $\widetilde{\boldsymbol{\Sigma}}$ is positive definite, then $\widehat{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{\Sigma}}$ is exactly the final solution to (3.8.1); else if not, we will use $\widetilde{\boldsymbol{\Sigma}}$ to construct the initial values for Algorithm 1 by setting $\left(\boldsymbol{\Sigma}^0, \boldsymbol{\Theta}^0, \boldsymbol{\Lambda}^0\right)$ to be $\boldsymbol{\Sigma}^0 = \tilde{\boldsymbol{\Sigma}},\ \boldsymbol{\Theta}^0 = \delta \mathbf{I}_p,\ \boldsymbol{\Lambda}^0 = \mathbf{0}$.

Algorithm 1 ADMM Algorithm for solving (3.8.2)

Input $(\mathbf{\Sigma}^0, \mathbf{\Theta}^0, \mathbf{\Lambda}^0) \in \mathbb{R}^{p\times p} \times \{\mathbf{\Theta} : \mathbf{\Theta} \geqslant \delta \mathbf{I}\} \times \mathbb{R}^{p\times p}$
repeat
  $\mathbf{\Theta}$-step: $\mathbf{\Theta}^{k+1} = \left(-\frac{1}{\beta}\mathbf{\Lambda}^k + \mathbf{\Sigma}^k\right)_{proj}$
  $\mathbf{\Sigma}$-step: $\mathbf{\Sigma}^{k+1} = \mathbf{M}_{pen}\left(\frac{1}{1+\beta}\left(\widehat{\mathbf{\Sigma}} + \mathbf{\Lambda}^k + \beta\mathbf{\Theta}^{k+1}\right),\ \lambda\ ,\ \frac{1}{1+\beta}\right)$
  $\mathbf{\Lambda}$-step: $\mathbf{\Lambda}^{k+1} = \mathbf{\Lambda}^k - \beta(\mathbf{\Sigma}^{k+1} - \mathbf{\Theta}^{k+1})$
until converge

3.8.3 Theoretical Results

In what follows, we present the assumptions and a new theorem regarding the convergence rate of $\widehat{\mathbf{\Sigma}}_{pd}$ in the multi-factor pricing models. The proposed positive-definite and sparse estimation does remove the undesired bias of the positive-definite ℓ_1-penalized estimation Xue et al. (2012). To this end, we study the landscape of the objective function and prove the uniqueness of the optimal solution with high probability.

Assumption 3.3. $P_\lambda(|t|)$ is a folded-concave penalty function on $t \in (-\infty, \infty)$ satisfying

(i) $P_\lambda(t)$ is nondecreasing in $t \in (0, \infty)$ with $P_\lambda(0) = 0$.

(ii) $\frac{P_\lambda(t)}{t}$ is nonincreasing in $t \in (0, \infty)$.

(iii) $P_\lambda(t)$ is differentiable in $t \in (0, \infty)$ with $P'_\lambda(0+) = \lambda$.

(iv) There exists a constant $0 < \mu \leqslant 1$, such that $P_\lambda(t) + \frac{\mu}{2}t^2$ is convex.

(v) There exists a scalar $\gamma > 0$ such that $P'_\lambda(t) = 0$, for all $t \geqslant \gamma\lambda$.

Remark 3.6. Assumption 3.3 can be easily satisfied by many popular regularizers such as hard-thresholding rule (Antoniadis, 1997) with $(\mu, \gamma) = (1, 1)$, SCAD (Fan and Li, 2001) with $(\mu, \gamma) = ((a-1)^{-1}, a)$ and MCP (Zhang, 2010) with $(\mu, \gamma) = (b^{-1}, b)$.

Assumption 3.4. There exist positive sequences $\kappa(p,T) = o(1)$, $a_T = o(1)$, and a constant $M > 0$, such that for any $C > M$,

$$P\left(\max_{j\leqslant p} \frac{1}{T}\sum_{t=1}^{T} |\epsilon_{jt} - \widehat{\epsilon}_{jt}|^2 > Ca_T^2\right) = O(\kappa(p,T))$$

Let $d = \max\limits_{1\leqslant i\leqslant T} \sum_{j=1}^{T} \mathbb{1}\{\sigma_{ij}^\star \neq 0\}$ denote the maximum number of nonzeros in any row of $\mathbf{\Sigma}^\star$, and let $s = \sum_{i=1}^{T}\sum_{j=1}^{T} \mathbb{1}\{\sigma_{ij}^\star \neq 0\}$ denote the total number of nonzero entries.

Theorem 3.6. Suppose that $\min_{(i,j)\in\text{supp}(\mathbf{\Sigma}^\star)} |\sigma_{ij}^\star| \geqslant \lambda(\gamma+1)+Cw_{p,T}$ and $\lambda_{\min}(\mathbf{\Sigma}^\star)$ $\geqslant \delta + C\min\left(\sqrt{s}, d\right) w_{p,T}$. Let $w_{p,T} = \sqrt{\frac{\log p}{T}} + a_T$, and $(\log p)^2 = o(T)$. Given Assumptions 3.3-3.4 and $\lambda > Cw_{p,T}$, with probability at least $1 - O\left(\frac{1}{p^2} + \kappa_1(p,T)\right)$, the program (3.8.1) has a unique optimal solution $\widehat{\mathbf{\Sigma}}_{pd}$, which agrees with the oracle estimator $\widehat{\mathbf{\Sigma}}^{orcl}$ given by

$$\widehat{\mathbf{\Sigma}}^{orcl} = \underset{\mathbf{\Sigma}\geqslant\delta\mathbf{I}}{\text{argmin}}\left\{\frac{1}{2}\left\|\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}\right\|_F^2 : \text{supp}(\mathbf{\Sigma}) \subseteq \text{supp}(\mathbf{\Sigma}^\star)\right\} \tag{3.8.3}$$

and satisfying that

(i) $\text{supp}(\widehat{\mathbf{\Sigma}}_{pd}) = \text{supp}(\mathbf{\Sigma}^\star)$

(ii) $\|\widehat{\mathbf{\Sigma}}_{pd} - \mathbf{\Sigma}^\star\|_{\max} \leqslant Cw_{p,T}$

(iii) $\|\widehat{\mathbf{\Sigma}}_{pd} - \mathbf{\Sigma}^\star\|_F \leqslant C\sqrt{s}w_{p,T}$

(iv) $\|\widehat{\mathbf{\Sigma}}_{pd} - \mathbf{\Sigma}^\star\|_2 \leqslant C\min\{\sqrt{s}, d\}w_{p,T}$

In addition, if $\min\{\sqrt{s}, d\}w_{p,T} = o(1)$, then with probability at least $1 - O\Big(\frac{1}{p^2}$ $+\kappa_1(p,T)\Big)$,

$$\|\widehat{\mathbf{\Sigma}}_{pd}^{-1} - (\mathbf{\Sigma}^\star)^{-1}\|_2 = O_p\left(\min\{\sqrt{s}, d\}w_{p,T}\right)$$

Proof. We first show that the objective function in (3.8.1) is strictly convex over the constrained set provided that $P_\lambda(|t|)$ satisfies conditions in Assumption 3.3 $(i)-(iv)$. Let $\mathcal{L}_n(\mathbf{\Sigma}) = \frac{1}{2}\left\|\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}\right\|_F^2$ denote the Frobenius loss, we have $\nabla^2\mathcal{L}_n(\mathbf{\Sigma}) = I_p \otimes I_p = I_{p^2}$. Hence,

$$\lambda_{\min}\left(\nabla^2\mathcal{L}_n(\mathbf{\Sigma})\right) = 1 \geqslant \mu.$$

It immediately implies that the function $\mathcal{L}_n(\boldsymbol{\Sigma}) - \frac{\mu}{2}\|\boldsymbol{\Sigma}\|_F^2$ is convex. Further by assumption 3.3 (iv), the function $\sum_{i\neq j} P_\lambda(|\sigma_{ij}|) + \frac{\mu}{2}\|\boldsymbol{\Sigma}\|_F^2$ is strictly convex. Thus the overall objective function is strictly convex over the feasible set.

Given the strict convexity of (3.8.1) over the constrained set, we only need to show that $\widetilde{\boldsymbol{\Sigma}}^{orcl}$ is a zero-subgradient point of the composite objective function in (3.8.1) within the feasible set. Then by strict convexity, $\widetilde{\boldsymbol{\Sigma}}^{orcl}$ must be the unique global minimum. Let $\mathcal{S} = \text{supp}(\boldsymbol{\Sigma}^\star)$ denote the support of $\boldsymbol{\Sigma}^\star$, and $\mathcal{D}$ represent the diagonal of $\boldsymbol{\Sigma}^\star$. Following the primal-dual witness (PDW) technique developed by Loh and Wainwright (2017), it remains to construct the zero-subgradient point $\widetilde{\boldsymbol{\Sigma}}^{orcl}$ such that $\widetilde{\boldsymbol{\Sigma}}^{orcl} \geqslant \delta\mathbf{I}$, $\text{diag}(\widetilde{\boldsymbol{\Sigma}}^{orcl}) = \mathbf{1}$ and $\text{vec}(\widetilde{\boldsymbol{\Sigma}}^{orcl})$ is a zero-subgradient point of the corresponding objective function

$$\begin{aligned}
&\mathcal{L}_n(\boldsymbol{\beta}) - q_\lambda(\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1 \\
\triangleq&\frac{1}{2}\left(\boldsymbol{\beta} - \text{vec}(\widehat{\boldsymbol{\Sigma}})\right)^T\left(\boldsymbol{\beta} - \text{vec}(\widehat{\boldsymbol{\Sigma}})\right) - \left(\lambda\|\boldsymbol{\beta}\|_1 - \sum_{k\notin\mathcal{D}} P_\lambda(|\boldsymbol{\beta}_k|)\right) + \lambda\|\boldsymbol{\beta}\|_1
\end{aligned} \tag{3.8.4}$$

Construct $\widetilde{\boldsymbol{\Sigma}}^{orcl}$ in its vector form as follows:

$$\text{vec}(\widetilde{\boldsymbol{\Sigma}}^{orcl}_{\mathcal{D}}) = 1, \text{vec}(\widetilde{\boldsymbol{\Sigma}}^{orcl}_{\mathcal{S}^c}) = 0, \text{vec}(\widetilde{\boldsymbol{\Sigma}}^{orcl}_{\mathcal{S}\setminus\mathcal{D}}) = \underset{\boldsymbol{\beta}\in\mathcal{S}\setminus\mathcal{D}}{\text{argmin}}\ \mathcal{L}_n(\boldsymbol{\beta}) - q_\lambda(\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1$$

The zero-subgradient condition now becomes

$$\nabla\mathcal{L}_n\left(\text{vec}(\widetilde{\boldsymbol{\Sigma}}^{orcl})\right) - \nabla q_\lambda\left(\text{vec}(\widetilde{\boldsymbol{\Sigma}}^{orcl})\right) + \lambda\widehat{\mathbf{z}} = 0$$

where $\widehat{\mathbf{z}}$ is the corresponding dual variable satisfying $||\widehat{\mathbf{z}}||_\infty < 1$. Namely, for $||\widehat{\mathbf{z}}||_\infty < 1$,

$$\nabla\mathcal{L}_n\left(\text{vec}(\widetilde{\boldsymbol{\Sigma}}^{orcl})\right) - \nabla\mathcal{L}_n\left(\text{vec}(\boldsymbol{\Sigma}^\star)\right) + \nabla\mathcal{L}_n\left(\text{vec}(\boldsymbol{\Sigma}^\star)\right) - \nabla q_\lambda\left(\text{vec}(\widetilde{\boldsymbol{\Sigma}}^{orcl})\right) + \lambda\widehat{\mathbf{z}} = 0.$$

Rewrite the above equation in the block form and plug in the constructed estimator,

$$\begin{pmatrix}\text{vec}(\widetilde{\boldsymbol{\Sigma}}^{orcl}_{\mathcal{S}}) - \text{vec}(\boldsymbol{\Sigma}^\star_{\mathcal{S}}) \\ \mathbf{0}\end{pmatrix} + \begin{pmatrix}\text{vec}(\widetilde{\boldsymbol{\Sigma}}^\star_{\mathcal{S}}) - \text{vec}(\widehat{\boldsymbol{\Sigma}}_{\mathcal{S}}) - \nabla q_\lambda\left(\text{vec}(\widetilde{\boldsymbol{\Sigma}}^{orcl}_{\mathcal{S}})\right) \\ \text{vec}(\boldsymbol{\Sigma}^\star_{\mathcal{S}^c}) - \text{vec}(\widehat{\boldsymbol{\Sigma}}_{\mathcal{S}^c})\end{pmatrix} + \lambda\begin{pmatrix}\widehat{\mathbf{z}}_{\mathcal{S}} \\ \widehat{\mathbf{z}}_{\mathcal{S}^c}\end{pmatrix} = 0.$$

Now, it suffices to show that

$$\|\widehat{\mathbf{z}}_{\mathcal{S}^c}\|_\infty = \left\|-\frac{1}{\lambda}\left(\mathrm{vec}(\mathbf{\Sigma}^\star_{\mathcal{S}^c}) - \mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{S}^c})\right)\right\|_\infty = \frac{1}{\lambda}\left\|\mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{S}^c}) - \mathrm{vec}(\mathbf{\Sigma}^\star_{\mathcal{S}^c})\right\|_\infty < 1 \tag{3.8.5}$$

As shown in Fan et al. (2011), under Assumptions 3.1 and 3.4, with probability at least $1 - \frac{1}{p^2} - \kappa_1(p, T)$, we have

$$\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}^\star\right\|_{\max} \leqslant C\left(\sqrt{\frac{\log p}{T}} + a_T\right) \triangleq C w_{p,T}$$

Thus (3.8.5) naturally holds when $\lambda > \left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}^\star\right\|_{\max}$, so that $\mathrm{vec}(\widetilde{\mathbf{\Sigma}}^{orcl}_{\mathcal{S}})$ is a zero-subgradient point of (3.8.4) satisfying dual feasibility. Moreover, from (3.8.3), we have

$$\begin{aligned}
&\left\|\mathrm{vec}(\widetilde{\mathbf{\Sigma}}^{orcl}) - \mathrm{vec}(\mathbf{\Sigma}^\star)\right\|_\infty \\
=&\left\|\mathrm{vec}(\widetilde{\mathbf{\Sigma}}^{orcl}_{\mathcal{S}}) - \mathrm{vec}(\mathbf{\Sigma}^\star_{\mathcal{S}})\right\|_\infty \\
=&\left\|-\mathrm{vec}(\mathbf{\Sigma}^\star_{\mathcal{S}}) + \mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{S}}) + \nabla q_\lambda\left(\mathrm{vec}(\widetilde{\mathbf{\Sigma}}^{orcl}_{\mathcal{S}})\right) - \lambda\widehat{\mathbf{z}}_{\mathcal{S}}\right\|_\infty \\
\leqslant&\left\|\mathrm{vec}(\mathbf{\Sigma}^\star_{\mathcal{S}}) - \mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{S}})\right\|_\infty + \left\|\nabla q_\lambda\left(\mathrm{vec}(\widetilde{\mathbf{\Sigma}}^{orcl}_{\mathcal{S}})\right) - \lambda\widehat{\mathbf{z}}_{\mathcal{S}}\right\|_\infty \\
\leqslant&\left\|\mathrm{vec}(\mathbf{\Sigma}^\star_{\mathcal{S}}) - \mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{S}})\right\|_\infty + \lambda
\end{aligned}$$

For any $(i, j) \in \mathrm{supp}(\mathbf{\Sigma}^\star)$,

$$|\widetilde{\sigma}^{orcl}_{ij}| \geqslant |\sigma^\star_{ij}| - |\sigma^\star_{ij} - \widetilde{\sigma}^{orcl}_{ij}| \geqslant \lambda(\gamma+1) + C w_{p,T} - \left\|\mathrm{vec}(\widetilde{\mathbf{\Sigma}}^{orcl}) - \mathrm{vec}(\mathbf{\Sigma}^\star)\right\|_\infty \geqslant \gamma\lambda \tag{3.8.6}$$

Hence, for any $(i, j) \in \mathrm{supp}(\mathbf{\Sigma}^\star)$, $P'_\lambda(\widetilde{\sigma}^{orcl}_{ij}) = P'_\lambda(\sigma^\star_{ij}) = 0$, i.e. the constructed $\widetilde{\mathbf{\Sigma}}^{orcl}$ agrees with $\widehat{\mathbf{\Sigma}}^{orcl}$ obtained by (3.8.3). In addition,

$$\left\|\widetilde{\mathbf{\Sigma}}^{orcl} - \mathbf{\Sigma}^\star\right\|_{\max} = \left\|\mathrm{vec}(\widetilde{\mathbf{\Sigma}}^{orcl}_{\mathcal{S}}) - \mathrm{vec}(\mathbf{\Sigma}^\star_{\mathcal{S}})\right\|_\infty = \left\|\mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{S}}) - \mathrm{vec}(\mathbf{\Sigma}^\star_{\mathcal{S}})\right\|_\infty \leqslant C w_{p,T}$$

Note that

$$\left\|\widetilde{\mathbf{\Sigma}}^{orcl} - \mathbf{\Sigma}^\star\right\|_2 \leqslant \left\|\widetilde{\mathbf{\Sigma}}^{orcl} - \mathbf{\Sigma}^\star\right\|_F \leqslant \sqrt{s}\left\|\widetilde{\mathbf{\Sigma}}^{orcl} - \mathbf{\Sigma}^\star\right\|_{\max} \tag{3.8.7}$$

$$\left\|\widetilde{\mathbf{\Sigma}}^{orcl} - \mathbf{\Sigma}^{\star}\right\|_2 \leqslant \left\|\widetilde{\mathbf{\Sigma}}^{orcl} - \mathbf{\Sigma}^{\star}\right\|_{\infty} \leqslant d \left\|\widetilde{\mathbf{\Sigma}}^{orcl} - \mathbf{\Sigma}^{\star}\right\|_{\max} \tag{3.8.8}$$

Combining (3.8.7) and (3.8.8) together, we have

$$\left\|\widetilde{\mathbf{\Sigma}}^{orcl} - \mathbf{\Sigma}^{\star}\right\|_2 \leqslant \min\left(\sqrt{s}, d\right) \left\|\widetilde{\mathbf{\Sigma}}^{orcl} - \mathbf{\Sigma}^{\star}\right\|_{\max} \leqslant C \min\left(\sqrt{s}, d\right) w_{p,T}$$

Since

$$\begin{aligned}\lambda_{\min}(\widetilde{\mathbf{\Sigma}}^{orcl}) &\geqslant \lambda_{\min}(\mathbf{\Sigma}^{\star}) + \lambda_{\min}(\widetilde{\mathbf{\Sigma}}^{orcl} - \mathbf{\Sigma}^{\star}) \\ &\geqslant \delta + C \min\left(\sqrt{s}, d\right) w_{p,T} + \lambda_{\min}(\widetilde{\mathbf{\Sigma}}^{orcl} - \mathbf{\Sigma}^{\star}) \geqslant \delta,\end{aligned}$$

then $\widetilde{\mathbf{\Sigma}}^{orcl}$ is the unique global optimum of (3.8.1), and $\widehat{\mathbf{\Sigma}}_{pd} = \widetilde{\mathbf{\Sigma}}^{orcl}$ with high probability.

□

Remark 3.7. Fan et al. (2011) prove that in a linear multi-factor pricing model (3.2.1), if $\{\epsilon_{jt}\}_{j \leqslant p, t \leqslant T}$ are estimated by OLS estimator, then under Assumptions 3.1 and 3.2, Assumption 3.4 is automatically satisfied with $a_T = \sqrt{(\log p)/T}$ and $\kappa_1(p, T) = p^{-2} + T^{-2}$.

Chapter 4

Fisher's Combined Probability Test for High-Dimensional Covariance Matrices

4.1 Introduction

Hypothesis testing on large covariance matrices has received considerable attention in the past decade. The covariance matrices not only have the fundamental importance in multivariate statistics such as discriminant analysis and principal component analysis (Zou and Xue, 2018; Ke et al., 2020; Chen et al., 2020), but also play a vital role in various research topics, for example, portfolio allocation (Goldfarb and Iyengar, 2003), gene-set testing (Chen and Qin, 2010), and gene-set clustering (Chang et al., 2017).

Let $\mathbf{X}$ and $\mathbf{Y}$ represent two independent p-dimensional random vectors with covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ respectively. We are interested in testing whether these two covariance matrices are equal, that is, $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$. This test is well studied in the classical setting where the dimension is fixed and the sample size diverges (Anderson, 2003). For instance, the likelihood ratio test was shown to enjoy the optimality under mild conditions (Sugiura and Nagao, 1968; Perlman, 1980). However, the likelihood function is not well-defined due to the singular sample covariance matrix in the high-dimensional setting where the dimension is no longer fixed but diverges at a possibly faster rate than the sample size.

Over the past decade, statisticians have made a lot of efforts to tackle the challenges in the high-dimensional setting and proposed three different types of statistics for testing large covariance matrices. Firstly, quadratic form statistics were studied to test against the dense alternatives, which can be written in terms of the Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ with many small differences between two covariance matrices. When the dimension is on the same order of the sample size, Schott (2007) proposed a test statistic based on the sum of squared differences between two sample covariance matrices, and Srivastava and Yanagihara (2010) used a consistent estimator of $\mathrm{tr}(\mathbf{\Sigma}_1^2)/\left[\mathrm{tr}(\mathbf{\Sigma}_1)\right]^2 - \mathrm{tr}(\mathbf{\Sigma}_2^2)/\left[\mathrm{tr}(\mathbf{\Sigma}_2)\right]^2$ to construct a new test statistic. Li and Chen (2012) introduced an unbiased estimator of the Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ to allow for the ultra-high dimensionality that the dimension grows much faster than the sample size. Recently, He et al. (2021) proposed the adaptive testing to combine the finite-order U-statistics that includes the variants of quadratic form statistics. Secondly, maximum form statistics were explored to account for the sparse alternatives with only a few large differences between two covariance matrices, which can be written in terms of the entry-wise maximum norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$. Cai et al. (2013) studied the maximal standardized differences between two sample covariance matrices to test against the sparse alternative, and Chang et al. (2017) proposed a perturbed-based maximum test using a data-driven approach to determine the rejection region. Thirdly, Li and Xue (2015), Yang and Pan (2017) and Li et al. (2018) used a weighted combination of quadratic form statistics and maximum form statistics to test against the dense or sparse alternatives, which shares the similar philosophy with the power enhancement method (Fan et al., 2015) for testing cross-sectional dependence.

Similar to these weighted combination tests, we are motivated by combining the strengths of quadratic form statistics and maximum form statistics to boost the power against the dense or sparse alternatives. It is also of great importance to combine the power of these two different statistics in real-world applications such as financial studies and genetic association studies. For instance, the anomalies in financial markets may come from the mispricing of a few assets or a systematic market mispricing (Fan et al., 2015), and the phenotype may be affected by a few causal variants or a large number of mutants (Liu et al., 2019).

It is worth pointing out that these weighted combination tests critically depend on the proper choice of weights to combine two different types of test statistics. There may exist a non-negligible discrepancy on the different magnitudes between quadratic form statistics and maximum form statistics in practice, which makes the choice of weights a very challenging task. As a promising alternative to Fan et al. (2015), Li and Xue (2015), Yang and Pan (2017) and Li et al. (2018), we provide a new perspective to exploit the full potential of quadratic form statistics and maximum form statistics for testing high-dimensional covariance matrices.

We propose a scale-invariant power enhancement test based on Fisher's method (Fisher, 1925) to combine the p-values of quadratic form statistics and maximum form statistics. To study the asymptotic property, we need to solve several non-trivial challenges in the theoretical analysis and then derive the asymptotic joint distribution of quadratic form statistics and maximum form statistics under the null hypothesis. We prove that the asymptotic null distribution of the proposed combination test statistic does not depend on the unknown parameters. More specifically, the proposed statistic follows a chi-squared distribution with 4 degrees of freedom asymptotically under the null hypothesis. We also show the consistent asymptotic power against the union of dense alternatives and sparse alternatives, which is more general than the designated alternative in the weighted combination test. It is worth pointing out that Fisher's method achieves the asymptotic optimality with respect to Bahadur relative efficiency. Moreover, we demonstrate the numerical properties in simulation studies and a real application to gene-set testing (Dudoit et al., 2008; Ritchie et al., 2015). In the real example, we find that the proposed test can detect the important gene-sets more effectively, and our findings are supported by biological evidences.

In recent literature, Liu and Xie (2020) proposed the Cauchy combination of p-values for testing high-dimensional mean vectors, and He et al. (2021) proved the joint normal limiting distribution of finite-order U-statistics with an identity covariance matrix and used the minimum combination of their p-values. Their methods and theories cannot be directly applied to the challenging setting for testing two-sample high-dimensional covariance matrices. Specifically, Li and Xue (2015) and He et al. (2021) considered the one-sample test for large covariance matrices that $H_0 : \mathbf{\Sigma} = \mathbf{I}$ under the restricted complete independence assumption among entries of $\mathbf{X}$, and Li et al. (2018) studied the one-sample test that H_0 : $\mathbf{\Sigma}$ is a banded matrix under the Gaussian assumption. Li and Xue (2015), Li et al. (2018), and He et al. (2021) studied the one-sample covariance test and did not prove the asymptotic independence result for testing two-sample covariance matrices. However, it is significantly more challenging to deal with the complicated dependence in the two-sample tests for large covariance matrices. To the best of our knowledge, our work presents the first proof of the asymptotic independence result of quadratic form statistics and maximum form statistics for testing two-sample covariance matrices, which provides the essential theoretical guarantee for Fisher's method to combine their p-values.

In the theoretical analysis, we use a non-trivial decorrelation technique to address the complex nonlinear dependence in high dimensional covariances. Recently, Shi et al. (2019) used the decorrelation to study the linear hypothesis testing for high-dimensional generalized linear models. But the nonlinear dependence in the two-sample covariance testing is much more challenging than the linear hypothesis testing. Moreover, we develop a new concentration inequality for two-sample degenerate U-statistics of high-dimensional data, which makes a separate contribution to the literature. This result is an extension of the concentration inequality for one-sample degenerate U-statistics (Arcones and Gine, 1993).

The rest of this chapter is organized as follows. After presenting the preliminaries in Section 4.2, we introduce the Fisher's method for testing two-sample large covariance matrices in Section 4.3. Section 4.4 studies the asymptotic size and asymptotic power, and Section 4.5 demonstrates the numerical properties in simulation studies. Section 4.6 evaluates the proposed test in an empirical study on testing gene-sets. Section 4.7 includes the concluding remarks. The technical details are presented in Section 4.8.

4.2 Preliminaries

Let $\mathbf{X}$ and $\mathbf{Y}$ be p-dimensional random vectors with covariance matrices $\mathbf{\Sigma}_1 = (\sigma_{ij1})_{p\times p}$ and $\mathbf{\Sigma}_2 = (\sigma_{ij2})_{p\times p}$ respectively. Without loss of generality, we assume both $\mathbf{X}$ and $\mathbf{Y}$ have zero means. Let $\{\mathbf{X}_1, \cdots, \mathbf{X}_{n_1}\}$ be independently and identically distributed ($i.i.d.$) random samples of $\mathbf{X}$, and $\{\mathbf{Y}_1, \cdots, \mathbf{Y}_{n_2}\}$ be $i.i.d.$ samples of $\mathbf{Y}$ that are independent of $\{\mathbf{X}_1, \cdots, \mathbf{X}_{n_1}\}$. The problem of interest is to test whether two covariance matrices are equal,

$$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2. \tag{4.2.1}$$

We first revisit the quadratic form statistic (Li and Chen, 2012) to test against the dense alternative and the maximum form statistic (Cai et al., 2013) to test against the sparse alternative. The dense alternative can be written in terms of the Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ and the sparse alternative can be written using the entry-wise maximum norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$.

Li and Chen (2012) proposed a quadratic-form test after reformulating the null hypothesis (4.2.1) into its equivalent form based on the squared Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$, that is,

$$H_0 : \|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_F^2 = 0.$$

To construct the test statistic, given the simple fact that

$$\|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_F^2 = \mathrm{tr}\{(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2\} = \mathrm{tr}(\mathbf{\Sigma}_1^2) + \mathrm{tr}(\mathbf{\Sigma}_2^2) - 2\mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2),$$

Li and Chen (2012) proposed a test statistic T_{n_1,n_2} in the form of linear combination of unbiased estimators for each term, specifically,

$$T_{n_1,n_2} = A_{n_1} + B_{n_2} - 2C_{n_1,n_2}, \tag{4.2.2}$$

where A_{n_1}, B_{n_2} and C_{n_1,n_2} are the unbiased estimators under H_0 for $\mathrm{tr}(\mathbf{\Sigma}_1^2)$, $\mathrm{tr}(\mathbf{\Sigma}_2^2)$ and $\mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)$ respectively. Then, the expected value of T_{n_1,n_2} is zero under the null hypothesis. For details about A_{n_1}, B_{n_2} and C_{n_1,n_2}, please refer to Section 2 of Li and Chen (2012). Li and Chen (2012) proved that the asymptotic distribution of T_{n_1,n_2} is a normal distribution. Let z_α be the upper α quantile of the standard normal distribution, and $\widehat{\sigma}_{0,n_1,n_2}$ is a consistent estimator of the leading term σ_{0,n_1,n_2} in the standard deviation of T_{n_1,n_2} under H_0. The test rejects the null hypothesis at the significance level α if

$$T_{n_1,n_2} \geqslant \widehat{\sigma}_{0,n_1,n_2} z_\alpha. \tag{4.2.3}$$

As an alternative to the quadratic form statistic (Li and Chen, 2012), Cai et al. (2013) studied the null hypothesis (4.2.1) in terms of the maximal absolute difference of two covariance matrices, i.e.,

$$H_0 : \max_{1 \leqslant i \leqslant j \leqslant p} |\sigma_{ij1} - \sigma_{ij2}| = 0.$$

Cai et al. (2013) proposed a maximum test statistic M_{n_1,n_2} based on the maximum of standardized differences between $\widehat{\sigma}_{ij1}$'s and $\widehat{\sigma}_{ij2}$'s. The maximum form statistic is written as

$$M_{n_1,n_2} = \max_{1 \leqslant i \leqslant j \leqslant p} \frac{(\widehat{\sigma}_{ij1} - \widehat{\sigma}_{ij2})^2}{\widehat{\theta}_{ij1}/n_1 + \widehat{\theta}_{ij2}/n_2}, \tag{4.2.4}$$

where the denominator $\widehat{\theta}_{ij1}/n_1 + \widehat{\theta}_{ij1}/n_2$ estimates the variance of $\widehat{\sigma}_{ij1} - \widehat{\sigma}_{ij2}$ to account for the heteroscedasticity of $\widehat{\sigma}_{ij1}$'s and $\widehat{\sigma}_{ij2}$'s among different entries. Cai et al. (2013) proved that the asymptotic null distribution of M_{n_1,n_2} is a Type I extreme value distribution (also known as the Gumbel distribution). Thus, the test rejects the null hypothesis at a significance level α if

$$M_{n_1,n_2} \geqslant q_\alpha + 4 \log p - \log \log p, \tag{4.2.5}$$

where q_α is the upper α quantile of the Gumbel distribution.

4.3 Fisher's Combined Probability Test

Li and Chen (2012) and Cai et al. (2013) have their respective power for testing high-dimensional covariance matrices. The quadratic form statistic T_{n_1,n_2} is powerful against the dense alternative, where the difference between $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ under the squared Frobenius norm is no smaller than the order of $\mathrm{tr}(\mathbf{\Sigma}_1^2)/n_1 + \mathrm{tr}(\mathbf{\Sigma}_2^2)/n_2$. The maximum form statistic M_{n_1,n_2} is powerful against the sparse alternative, where at least one entry of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ has the magnitude larger than the order of $\sqrt{\log p/n}$. However, T_{n_1,n_2} performs poorly against the sparse alternative and M_{n_1,n_2} performs poorly against the dense alternative. More details will be presented in Section 4.4.3 and Section 4.5.

The weighted combination $J = J_0 + J_1$ was studied to achieve the power enhancement (Fan et al., 2015; Li and Xue, 2015; Yang and Pan, 2017; Li et al., 2018), where J_0 is built on the extreme value form statistic and J_1 is constructed from the asymptotically pivotal statistic. It is worth pointing out that, with the proper weighted combination, J enjoys the so-called *power enhancement properties* (Fan et al., 2015): (i) J is at least as powerful as J_1, (ii) the size distortion due to the addition of J_0 is asymptotically negligible, and (iii) power is improved under the designated alternatives. For testing large covariance matrices, Yang and Pan (2017) proposed $J_1 = (1 - (s_p + \xi_1)^{-1})M_n$ and $J_0 = n^{\frac{1}{s_p+\xi_1}+\frac{1}{\xi_2}} \cdot \max_{1 \leqslant i,j \leqslant p}(\widehat{\sigma}_{ij1} - \widehat{\sigma}_{ij2})^2$, where M_n is a macro-statistic which performs well against the dense alternative, and s_p is the number of distinct entries in two covariance matrices. Note that the quantities ξ_1 and ξ_2 are carefully chosen such that $J_0 \to 0$ under H_0.

As a promising alternative, we propose a scale-invariant combination procedure based on Fisher's method (Fisher, 1925) to combine both strengths of T_{n_1,n_2} and M_{n_1,n_2}. Let $\Phi(\cdot)$ be the cumulative distribution function of $N(0,1)$ and $G(x) = \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right)$ be the cumulative distribution function of the Gumbel distribution. More specifically, we combine the p-values of T_{n_1,n_2} and M_{n_1,n_2} after the negative natural logarithm transformation, that is,

$$F_{n_1,n_2} = -2\log p_T - 2\log p_M, \tag{4.3.1}$$

where

$$p_T = 1 - \Phi\left(T_{n_1,n_2}/\widehat{\sigma}_{0,n_1,n_2}\right)$$

and

$$p_M = 1 - G(M_{n_1,n_2} - 4\log p + \log\log p)$$

are the p-values associated with the test statistics T_{n_1,n_2} and M_{n_1,n_2}, respectively.

Let c_α denote the upper α quantile of a chi-squared distribution with 4 degrees of freedom (i.e., χ_4^2). We reject the null hypothesis at the significance level α if

$$F_{n_1,n_2} \geqslant c_\alpha. \tag{4.3.2}$$

Unlike the weighted statistic $J = J_0 + J_1$, F_{n_1,n_2} does not need to estimate s_p or choose ξ_1 and ξ_2 to construct the proper weights, which may be non-trivial to deal with in practice. The inappropriate choice of s_p, ξ_1 and ξ_2 may lead to the size distortion or loss of power. In contrast, F_{n_1,n_2} is scale-invariant as the p-values always take values between 0 and 1, and the asymptotic null distribution of F_{n_1,n_2} (i.e., χ_4^2) does not depend on any hyper-parameters. As we will show in Section 4.4.3, F_{n_1,n_2} achieves the desired nominal significance level asymptotically while boosting the power against either sparse or dense alternatives. Moreover, Fisher's method achieves the asymptotic optimality with respect to Bahadur relative efficiency (Littell and Folks, 1971, 1973).

Remark 4.1. The idea of combining p-values has been widely used as an important technique for data fusion or meta analysis (Hedges and Olkin, 2014). Recently, the Cauchy combination of p-values was used for testing high-dimensional mean vectors in (Liu and Xie, 2020), and the minimum combination of p-values from the finite-order U-statistics was used for testing two-sample high-dimensional covariance matrices in (He et al., 2021). However, neither Liu and Xie (2020) nor He et al. (2021) studied the combination of p-values of T_{n_1,n_2} and M_{n_1,n_2}, and it is fundamentally challenging to study the asymptotic joint distribution of T_{n_1,n_2} and M_{n_1,n_2}. We will solve this open problem in Section 4.4.2.

4.4 Asymptotic Properties

This section presents the asymptotic properties of our proposed Fisher's combined probability test F_{n_1,n_2}. Section 4.4.1 presents the assumptions. Section 4.4.2 studies the joint limiting distribution of two test statistics M_{n_1,n_2} and T_{n_1,n_2} under the null hypothesis. Section 4.4.3 proves the correct asymptotic size and consistent asymptotic power of our proposed method.

4.4.1 Assumptions

We define some useful notations. For any matrix $\mathbf{A}$, let $\lambda_i(\mathbf{A})$ be the i-th largest eigenvalue of $\mathbf{A}$. For any set $\mathcal{A}$, card($\mathcal{A}$) represents the cardinality of $\mathcal{A}$. For $0 < r < 1$, let

$$\mathcal{V}_i(r) = \left\{1 \leqslant j \leqslant p : \frac{|\sigma_{ij1}|}{\sqrt{\sigma_{ii1}\sigma_{jj2}}} \geqslant r \text{ or } \frac{|\sigma_{ij2}|}{\sqrt{\sigma_{ii2}\sigma_{jj2}}} \geqslant r\right\}$$

be the set of indices j such that X_j (or Y_j) is highly correlated (whose correlation $> r$) with X_i (or Y_i) for a given $i \in \{1, \ldots, p\}$. And for any $\alpha > 0$, let

$$s_i(\alpha) = \text{card}(\mathcal{V}_i((\log p)^{-1-\alpha})), \; i = 1, \cdots, p$$

denote the number of indices j in the set $\mathcal{V}_i((\log p)^{-1-\alpha})$. Moreover, define

$$\mathcal{W}(r) = \{1 \leqslant i \leqslant p : \mathcal{V}_i(r) \neq \varnothing\}$$

such that, $\forall i \in \mathcal{W}(r)$, X_i (or Y_i) is highly correlated with some other variable of $\mathbf{X}$ (or $\mathbf{Y}$).

Throughout the rest of this section, we assume that $\mathbf{X}$ and $\mathbf{Y}$ are both Gaussian random vectors. The Gaussian assumption facilitates the use of a new decorrelation technique to address the complex nonlinear dependence in high dimensional covariances in the theoretical analysis of the proposed scale-invariant combination test.

Remark 4.2. Li and Xue (2015), Li et al. (2018) and He et al. (2021) studied the asymptotic joint distribution of the maximum test statistic and the quadratic test statistic for one-sample covariance test under the Gaussian assumption or restricted complete independence assumption. Please see the first paragraph of Section 2 in Li and Xue (2015), the first paragraph of Section 2 in Li et al. (2018), and Condition 2.3 in He et al. (2021) for more details. However, the nonlinear dependence in two-sample covariance test is fundamentally more challenging than the dependence in the one-sample covariance test.

Assumption 4.1. As $\min\{n_1, n_2\} \to \infty$ and $p \to \infty$,

(i) $n_1/(n_1 + n_2) \to \gamma$, for some constant $\gamma \in (0, 1)$.

(ii) $\sum_{i=1}^{q} \lambda_i^2(\mathbf{\Sigma}_j)/\sum_{i=1}^{p} \lambda_i^2(\mathbf{\Sigma}_j) \to 0$ for any integer $q = O(\log p)$ and $j = 1, 2$.

Remark 4.3. Assumption 4.1 is analogous to (A1) and (A2) in Li and Chen (2012), where the first condition is standard for two-sample asymptotic analysis, and the second one describes the extent of high dimensionality and the dependence which can be accommodated by the proposed tests. Sharing the spirit, Assumption

4.1 does not impose explicit requirements on relationships between p and n_1, n_2, but rather requires a mild condition (ii) regarding the covariances, which can be satisfied if eigenvalues of two covariance matrices are bounded.

Assumption 4.2. There exists a subset $\Upsilon \subset \{1, 2, \cdots, p\}$ with $\text{card}\,(\Upsilon) = o(p)$ and some constant $\alpha_0 > 0$, such that for all $\kappa > 0$, $\max_{1 \leqslant i \leqslant p, i \notin \Upsilon} s_i(\alpha_0) = o(p^\kappa)$. In addition, there exists a constant $0 < r_0 < 1$, such that $\text{card}(\mathcal{W}(r_0)) = o(p)$.

Remark 4.4. Assumption 4.2 was introduced by Cai et al. (2013) such that $\max_{1 \leqslant i \leqslant p, i \notin \Upsilon} \; s_i(\alpha_0)$ and $\mathcal{W}(r_0)$ are moderate for $\alpha_0 > 0$ and $0 < r_0 < 1$. It is satisfied if the eigenvalues of covariance matrices are bounded from above and correlations are bounded away from ± 1.

4.4.2 Asymptotic Joint Distribution

Now, we present the joint limiting law for M_{n_1,n_2} and T_{n_1,n_2} under the null hypothesis.

Theorem 4.1. Suppose Assumptions 4.1 and 4.2 hold, and $\log p = o(n^{\frac{1}{5}})$ for $n = n_1 + n_2$, then under the null hypothesis H_0, for any $x, t \in \mathbb{R}$, we have

$$P\left(\frac{T_{n_1,n_2}}{\widehat{\sigma}_{0,n_1,n_2}} \leqslant t,\; M_{n_1,n_2} - 4\log p + \log\log p \leqslant x\right) \to \Phi(t) \cdot G(x) \tag{4.4.1}$$

as $n_1, n_2, p \to \infty$, where $G(x) = \exp\left(-\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{x}{2}\right)\right)$ is the cdf of Gumbel distribution, and $\Phi(t)$ is the cdf of standard normal distribution.

Remark 4.5. Together with Theorems 1 and 2 from Li and Chen (2012) and Theorem 1 from Cai et al. (2013), Theorem 4.1 implies that M_{n_1,n_2} and T_{n_1,n_2} are asymptotically independent.

In the sequel, we provide a high-level intuition to prove the asymptotic independence result (4.4.1). First of all, it is worth mentioning that under Assumption 4.1, all the third-moment and fourth-moment terms in A_{n_1}, B_{n_2} and C_{n_1,n_2} are of small order than the leading second-moment terms, which may be neglected when deriving the asymptotic normality. Hence in theoretical analysis, we may consider the simplified statistic of T_{n_1,n_2} defined by

$$\widetilde{T}_{n_1,n_2} = \frac{1}{n_1(n_1-1)}\sum_{u\neq v}(\mathbf{X}_u'\mathbf{X}_v)^2 + \frac{1}{n_2(n_2-1)}\sum_{u\neq v}(\mathbf{Y}_u'\mathbf{Y}_v)^2 - \frac{2}{n_1n_2}\sum_u\sum_v(\mathbf{X}_u'\mathbf{Y}_v)^2. \tag{4.4.2}$$

As pointed out by Li and Chen (2012), $\widetilde{T}_{n_1,n_2}$ and T_{n_1,n_2} shares the same asymptotic behavior.

Compared with the simple one-sample covariance test in Li and Xue (2015), Li et al. (2018), and He et al. (2021), it is significantly more difficult to analyze the asymptotic joint distribution given the complicated dependence in the two-sample tests for large covariance matrices. To address this challenge, we use a decorrelation technique to address the complex nonlinear dependence in high dimensional covariances. Specifically, we introduce a decorrelated statistic $T^*_{n_1,n_2}$. Under $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$, we may partition $\mathbf{X}$ and $\mathbf{Y}$ as follows:

$$\mathbf{X}_{p\times 1} = \begin{pmatrix}\mathbf{X}^{(1)}\\ \mathbf{X}^{(2)}\end{pmatrix} \text{ and } \mathbf{Y}_{p\times 1} = \begin{pmatrix}\mathbf{Y}^{(1)}\\ \mathbf{Y}^{(2)}\end{pmatrix} \sim N_p\left(\begin{pmatrix}\mathbf{0}_{p-q}\\ \mathbf{0}_q\end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22}\end{pmatrix}\right).$$

where $\mathbf{X}^{(1)}, \mathbf{Y}^{(1)} \in \mathbb{R}^{p-q}$, $\mathbf{X}^{(2)}, \mathbf{Y}^{(2)} \in \mathbb{R}^q$ for integer q satisfying $q = O(\log p)$. Let $\mathbf{Z}_1 = \mathbf{X}^{(1)} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{X}^{(2)}$, $\mathbf{Z}_2 = \mathbf{X}^{(2)}$, $\mathbf{W}_1 = \mathbf{Y}^{(1)} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{Y}^{(2)}$, $\mathbf{W}_2 = \mathbf{Y}^{(2)}$. It's easy to see that $\mathbf{Z}_1$ is independent of $\mathbf{Z}_2$, and the same results hold for $\mathbf{W}_1$ and $\mathbf{W}_2$. Back to the sample level, we have that $\{\mathbf{Z}_{1u}\}_{u=1}^{n_1}$ and $\{\mathbf{W}_{1v}\}_{v=1}^{n_2}$ i.i.d. follow $N_{p-q}(\mathbf{0}, \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})$. Following the pattern of $\widetilde{T}_{n_1,n_2}$ in (5.7.32), we define

$$\begin{aligned} T^*_{n_1,n_2} =& \frac{1}{n_1(n_1-1)}\sum_{u\neq v}(\mathbf{Z}_{1u}'\mathbf{Z}_{1v})^2 + \frac{1}{n_2(n_2-1)}\sum_{u\neq v}(\mathbf{W}_{1u}'\mathbf{W}_{1v})^2 \\ &- \frac{2}{n_1n_2}\sum_u\sum_v(\mathbf{Z}_{1u}'\mathbf{W}_{1v})^2. \end{aligned} \tag{4.4.3}$$

$\{\mathbf{Z}_{1u}\}_{u=1}^{n_1}$ and $\{\mathbf{W}_{1v}\}_{v=1}^{n_2}$ are regarded as a decorrelated version of $\{\mathbf{X}_u\}_{u=1}^{n_1}$ and $\{\mathbf{Y}_v\}_{v=1}^{n_2}$, respectively. $T^*_{n_1,n_2}$ is regarded as the $\widetilde{T}_{n_1,n_2}$ statistic derived from the decorrelated samples. The above deccorelation shares a similar philosophy with Shi et al. (2019). We should point out that Shi et al. (2019) used the decorrelation to study the linear hypothesis testing for high-dimensional generalized linear models, but the nonlinear dependence in the two-sample covariance testing is much more challenging than the linear hypothesis testing.

In what follows, we study the joint distribution of M_{n_1,n_2} and $\widetilde{T}_{n_1,n_2}$. Let A denote the event associated with the maximum statistic M_{n_1,n_2}, and let B be the event corresponding to the quadratic statistic $\widetilde{T}_{n_1,n_2}$. We use the simple but very helpful fact that $A = \cup_i A_i$. Then, we may rewrite the joint probability $P\left(A \cap B\right)$ into the probability for a union of events, that is, $P\left(A \cap B\right) = P\left(\left(\cup_i A_i\right) \cap B\right)$. In what follows, we give the proof sketch to derive the upper bound $P(A \cap B) - P(A)P(B) \leqslant o(1)$. We begin with a union bound to obtain that $P\left(\cup_i (A_i \cap B)\right) \leqslant \sum_i P(A_i \cap B)$. In order to deal with the joint probability of $A_i \cap B$, we further decompose the quadratic statistic into two parts: $T^*_{n_1,n_2}$ is independent of A_i, and the remaining term $\widetilde{T}_{n_1,n_2} - T^*_{n_1,n_2}$ is associated with A_i. Consequently, B can be written as $B = B_i^c \cup B_i$, in which B_i^c represents to the event corresponding to $T^*_{n_1,n_2}$. Therefore, $\sum_i P(A_i \cap B) \leqslant \sum_i P(A_i \cap B_i^c) + \sum_i P(A_i \cap B_i) \leqslant \sum_i P(A_i)P(B_i^c) + \sum_i P(B_i)$. Lemma 4.2 suggests $T^*_{n_1,n_2}$ is sufficiently close to $\widetilde{T}_{n_1,n_2}$ so that we have $P(B_i^c) \approx P(B)$, $\sum_i P(A_i) \to P(A)$ and $\sum_i P(B_i) = o(1)$. The lower bound $o(1) \leqslant P(A \cap B) - P(A)P(B)$ can be similarly derived from the Bonferroni inequality. Therefore, we can prove the asymptotic independence given that $|P(A \cap B) - P(A)P(B)| = o(1)$.

In the following, we present three useful lemmas to prove (4.4.1) in Theorem 1.

Lemma 4.1 (Asymptotic Normality)**.** Under Assumption 4.1, as $n_1, n_2, p \to \infty$,

$$\frac{T^*_{n_1,n_2}}{2\left(n_1^{-1} + n_2^{-1}\right) \operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \xrightarrow{d} N(0,1). \tag{4.4.4}$$

Lemma 4.2 (Exponential Decay)**.** Under Assumption 4.1, for any $\epsilon > 0$, there exists positive constants C, c that do not depend on p, n_1, n_2, such that

$$P\left(\frac{\left|\widetilde{T}_{n_1,n_2} - T^*_{n_1,n_2}\right|}{2\left(n_1^{-1} + n_2^{-1}\right) \operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \geqslant \epsilon\right) \leqslant C \exp\{-c\epsilon n^{\beta}\}, \tag{4.4.5}$$

with $1/5 < \beta < 1/3$.

Remark 4.6. Lemma 4.2 presents a new concentration inequality for two-sample degenerate U-statistics. It extends the well-known concentration inequality for one-sample degenerate U-statistics (Arcones and Gine, 1993) and makes a separate contribution to the literature.

As a final step, Lemma 4.3 derives the joint limiting distribution of the test statistic M_{n_1,n_2} and the simplified statistic $\widetilde{T}_{n_1,n_2}$, which directly implies Theorem 4.1.

Lemma 4.3. Under the same assumptions as in Theorem 4.1,

$$P\left(\frac{\widetilde{T}_{n_1,n_2}}{\widehat{\sigma}_{0,n_1,n_2}} \leqslant t,\ M_{n_1,n_2} - 4\log p + \log\log p \leqslant x\right) \to \Phi(t) \cdot G(x) \tag{4.4.6}$$

for any $x, t \in \mathbb{R}$, as $n_1, n_2, p \to \infty$.

Lemma 4.1 shows that such decorrelation procedure does not affect the asymptotic behavior of the quadratic test statistic. Lemma 4.2 depicts the tail behavior of the difference between $\widetilde{T}_{n_1,n_2}$ and $T^*_{n_1,n_2}$ with explicit decaying rate. Lemma 4.1 and Lemma 4.2 lay the foundation of replacing $\widetilde{T}_{n_1,n_2}$ with $T^*_{n_1,n_2}$ in the theoretical analysis.

4.4.3 Asymptotic Size and Power

Given the explicit joint distribution of M_{n_1,n_2} and T_{n_1,n_2}, we proceed to present the asymptotic properties of our proposed Fisher's test F_{n_1,n_2}. Recall that c_α is the upper α-quantile of χ^2_4 distribution and $F_{n_1,n_2} = -2\log(p_M) - 2\log(p_T)$ rejects H_0 if F_{n_1,n_2} is as extreme as c_α. On top of the asymptotic independence established in Section 4.4.2 and by simple probability transformation, it's easy to obtain the null distribution of F_{n_1,n_2}, and therefore, the asymptotic size of the test. The results are formally presented in Theorem 4.2.

Theorem 4.2 (Asymptotic Size)**.** Under the same assumptions as in Theorem 4.1, the Fisher's test achieves accurate asymptotic size, that is, under the null hypothesis,

$$P\left(F_{n_1,n_2} \geqslant c_\alpha\right) \to \alpha \quad \text{as } n_1, n_2, p \to \infty.$$

Remark 4.7. Besides Fisher's method, the asymptotic independence result makes it feasible to combine p-values using other approaches such as Tippett's method

(Tippett, 1931), Stouffer's method (Stouffer et al., 1949), and Cauchy combination (Liu and Xie, 2020).

Li and Chen (2012) and Cai et al. (2013) provided power analysis of tests T_{n_1,n_2} and M_{n_1,n_2} over the dense alternative $\mathcal{G}_d$ and the sparse alternative $\mathcal{G}_s$ respectively.

$$\mathcal{G}_d = \left\{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) : \mathbf{\Sigma}_1 > 0, \mathbf{\Sigma}_2 > 0, \frac{1}{n_1}\mathrm{tr}(\mathbf{\Sigma}_1^2) + \frac{1}{n_2}\mathrm{tr}(\mathbf{\Sigma}_2^2) = o\left(\mathrm{tr}\{(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2\}\right)\right\}, \tag{4.4.7}$$

$$\mathcal{G}_s = \left\{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) : \mathbf{\Sigma}_1 > 0, \mathbf{\Sigma}_2 > 0, \max_{1 \leqslant i \leqslant j \leqslant p} \frac{|\sigma_{ij1} - \sigma_{ij2}|}{\sqrt{\theta_{ij1}/n_1 + \theta_{ij2}/n_2}} \geqslant 4\sqrt{\log p}\right\}. \tag{4.4.8}$$

Taking advantage of the combination, we shall show that our proposed combined test F_{n_1,n_2} makes the most of merits from the two tests and successfully boost the power against either dense or sparse alternatives.

Theorem 4.3 (Asymptotic Power)**.** Under the same assumptions as in Theorem 4.1, the Fisher's test achieves consistent asymptotic power, that is, under the alternative hypothesis,

$$\inf_{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \mathcal{G}_d \cup \mathcal{G}_s} P\left(F_{n_1,n_2} \geqslant c_\alpha\right) \to 1 \quad \text{as } n_1, n_2, p \to \infty.$$

Remark 4.8. (Bahadur Efficiency) Littell and Folks (1971, 1973) studied the asymptotic optimality of using Fisher's method to combine independent tests. They proved for the two asymptotically independent tests $p_T = 1 - \Phi\left(T_{n_1,n_2}/\widehat{\sigma}_{0,n_1,n_2}\right)$ and $p_M = 1 - G(M_{n_1,n_2} - 4\log p + \log\log p)$, the Fisher's method delivers the largest exact Bahadur slope among all reasonable combination approaches, indicating the fastest decay rate for the p-values under the alternatives. Therefore, the Fisher's combined test is asymptotically optimal in terms of Bahadur relative efficiency.

4.5 Simulation Studies

This section examines the finite-sample performance of our Fisher's combined probability test, compared to the tests proposed by Cai et al. (2013) (refer as the

CLX test in the following context) and Li and Chen (2012) (refer as the LC test). We generate $\{\mathbf{X}_1, \cdots, \mathbf{X}_{n_1}\}$ *i.i.d.* from $N_p(\mathbf{0}, \mathbf{\Sigma}_1)$ and $\{\mathbf{Y}_1, \cdots, \mathbf{Y}_{n_2}\}$ *i.i.d.* from $N_p(\mathbf{0}, \mathbf{\Sigma}_2)$. The sample sizes are taken to be $n_1 = n_2 = N$ with $N = 100$ and 200, while the dimension p varies over the values 100, 200, 500, 800 and 1000. For each simulation setting, the average number of rejections are reported based on 1000 replications. The significance level is set to be 0.05 for all the tests.

Under the null hypothesis H_0, we set $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}^{*(i)}, i = 1, \cdots, 5$, and consider the following five models to evaluate the testing size.

(i) $\mathbf{\Sigma}^{*(1)} = \mathbf{I}_p$.

(ii) $\mathbf{\Sigma}^{*(2)} = (\mathbf{\Omega}^{*(2)})^{-1}$, where $\omega_{ij}^{*(2)} = 0.5^{|i-j|}$.

(iii) $\mathbf{\Sigma}^{*(3)}$ is a block diagnoal matrix given by each block being $0.5\mathbf{I}_5 + 0.5\mathbb{1}_5\mathbb{1}_5'$.

(iv) $\mathbf{\Sigma}^{*(4)} = \{\sigma_{ij}^{*(4)}\}_{p\times p}$, $\sigma_{ij}^{*(4)} = (-1)^{i+j}0.4^{|i-j|^{1/10}}$.

(v) $\mathbf{\Sigma}^{*(5)} = (\mathbf{\Sigma}^{(5)} + \delta\mathbf{I})/(1+\delta)$, where $\sigma_{ii}^{(5)} = 1$, $\sigma_{ij}^{(5)} = 0.5 * Bernoulli(1, 0.05)$ for $i < j$ and $\sigma_{ij}^{(5)} = \sigma_{ji}^{(5)}$, $\delta = |\lambda_{\min}(\mathbf{\Sigma}^{(5)})| + 0.05$.

Model (i) is the most commonly used multivariate standard normal distribution. Model (ii) and Model (iii) are the cases when the true covariance matrices have certain banded-type and block-type sparsity. Model (iv) was first proposed by Srivastava and Yanagihara (2010) and further studied in Cai et al. (2013). Model (v) is also a sparse matrix yet without any specific sparsity pattern.

To evaluate the power of the tests, we consider the scenarios when the differences of the two covariance matrices satisfy certain structure. There are two types of alternatives we desire to look into: the sparse alternative H_s and the dense alternative H_d.

Generally speaking, the sparse alternative shares commonality among different models. Let $\mathbf{U}$ denote the difference between $\mathbf{\Sigma}_2$ and $\mathbf{\Sigma}_1$, i.e. $\mathbf{U} = \mathbf{\Sigma}_2 - \mathbf{\Sigma}_1$. Inspired by Cai et al. (2013), we consider the situation when $\mathbf{U}$ is a symmetric sparse matrix with eight random nonzero entries. The locations of four nonzero entries are randomly selected from the upper triangle of $\mathbf{U}$, each with a magnitude of Unif(0,4)$\times \max_{1\leqslant j\leqslant p}\sigma_{jj}^*$. The other four are determined by symmetry. Then we generate samples from these covariance pairs $\left(\mathbf{\Sigma}_1^{(i)}, \mathbf{\Sigma}_2^{(i)}\right)$, $i = 1, \cdots, 5$, in order to evaluate the power of the tests against sparse alternative, where $\mathbf{\Sigma}_1^{(i)} = \mathbf{\Sigma}^{*(i)} + \delta\mathbf{I}$ and $\mathbf{\Sigma}_2^{(i)} = \mathbf{\Sigma}^{*(i)} + \delta\mathbf{I} + \mathbf{U}$, with $\delta = |\min\{\lambda_{\min}(\mathbf{\Sigma}^{*(i)} + \mathbf{U}), \lambda_{\min}(\mathbf{\Sigma}^{*(i)})\}| + 0.05$.

In terms of the dense alternative setting, since the five models differ a lot from each other, we shall discuss their corresponding alternative settings separately afterwards. To begin with, we shall take a look at the simplest case in Model (i). We consider its dense alternative to be the AR(1) model with parameter $\rho = 0.2$ and 0.3, denoted by $\mathbf{\Sigma}_\rho^{AR}$. In another word, we generate the copies of $\mathbf{X}$ from the p-dimensional standard normal while copies of $\mathbf{Y}$ from $N_p\left(\mathbf{0}, \mathbf{\Sigma}_\rho^{AR}\right)$. We follow the same alternative hypothesis as in Srivastava and Yanagihara (2010) for Model (iv), which is $\sigma_{ij}^{(4)} = (-1)^{i+j}0.6^{|i-j|^{1/10}}$, whereas we use the identity matrix $\mathbf{I}_p$ for Models (ii), (iii) and (v).

Table 4.1: Comparison of empirical size and power (%) for model (i)

N	p	100	200	500	800	1000	100	200	500	800	1000
				Size					Power under sparse alternative		
	Proposed	5.6	5.0	5.0	5.2	5.6	98.0	96.6	87.3	83.9	80.2
100	CLX	4.3	5.2	4.5	4.4	4.5	98.5	98.3	91.1	89.8	85.8
	LC	4.8	5.0	5.1	4.5	4.2	20.6	11.2	5.9	5.7	5.0
	Proposed	4.6	4.7	4.8	4.9	4.3	100.0	100.0	100.0	100.0	100.0
200	CLX	3.6	4.2	4.5	5.5	5.0	100.0	100.0	100.0	100.0	100.0
	LC	5.4	3.2	4.6	4.8	5.3	50.5	22.2	8.0	7.6	7.3
						Power under dense alternative					
				$\rho = 0.2$					$\rho = 0.3$		
	Proposed	59.8	56.3	55.7	53.1	53.1	99.7	99.8	99.7	100.0	99.9
100	CLX	13.9	8.9	8.1	6.9	6.6	51.5	45.7	38.3	31.9	27.2
	LC	60.7	63.2	64.8	62.4	63.3	99.7	99.8	100.0	99.9	99.8
	Proposed	98.6	99.3	99.3	98.8	98.9	100.0	100.0	100.0	100.0	100.0
200	CLX	46.5	40.1	30.9	28.0	25.3	99.8	99.9	100.0	99.8	99.9
	LC	98.6	99.3	99.0	99.1	98.9	100.0	100.0	100.0	100.0	100.0

Note: This table reports the frequencies of rejection by each method under the null and alternative hypotheses based on 1000 independent replications at significance level 5%.

For each covariance model, we generate samples independently from $N_p(\mathbf{0}, \mathbf{\Sigma}^{*(i)})$ to evaluate the size, and use different covariance pairs described above to examine the power against dense and sparse alternatives. The empirical size and power are calculated based on 1,000 replications at significance level 5% and the results are reported in Tables 4.1, 4.2 and 4.3.

Table 4.2: Comparison of empirical size and power (%) for models (ii) and (iii)

		Model (ii)					Model (iii)				
N	p	100	200	500	800	1000	100	200	500	800	1000
		Size									
100	Proposed	4.9	5.5	4.2	5.6	5.3	6.0	6.1	4.8	4.9	3.9
	CLX	4.6	5.4	4.9	5.5	4.5	4.5	4.4	5.1	4.6	4.0
	LC	4.6	5.3	3.8	4.5	5.2	5.3	5.6	4.7	5.1	4.3
200	Proposed	6.5	5.4	4.1	3.8	4.3	6.3	6.5	4.8	4.1	4.9
	CLX	4.5	4.3	5.8	4.0	4.3	4.3	6.5	4.1	3.8	4.8
	LC	5.8	4.9	4.1	3.7	5.1	5.6	5.2	4.3	4.3	4.8
		Power under sparse alternative									
100	Proposed	98.4	96.1	87.5	85.3	79.8	98.1	95.7	88.1	82.3	81.3
	CLX	98.8	97.7	92.3	90.2	85.9	98.7	97.5	91.3	88.0	86.6
	LC	19.7	11.4	6.8	5.8	5.7	20.0	11.6	6.6	5.4	5.3
200	Proposed	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	CLX	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	LC	50.1	22.5	8.7	7.2	6.1	53.7	23.0	10.1	6.9	6.0
		Power under dense alternative									
100	Proposed	85.7	83.0	84.7	83.7	81.7	97.6	98.0	97.6	96.3	98.2
	CLX	15.9	11.7	7.0	7.7	6.2	36.0	27.5	21.5	17.0	14.8
	LC	88.5	87.7	89.6	89.2	89.8	97.9	98.5	98.5	97.4	99.1
200	Proposed	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	CLX	59.6	50.4	37.5	33.7	31.1	90.7	91.8	87.7	86.1	83.6
	LC	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: This table reports the frequencies of rejection by each method under the null and alternative hypotheses based on 1000 independent replications at significance level 5%.

Table 4.3: Comparison of empirical size and power (%) comparisons for models (iv) and (v)

		Model (iv)					Model (v)				
N	p	100	200	500	800	1000	100	200	500	800	1000
						Size					
100	Proposed	9.8	9.5	10.4	9.6	9.3	5.7	5.2	4.0	4.8	4.3
	CLX	4.1	4.1	3.8	4.2	4.0	4.6	4.9	4.6	4.9	4.2
	LC	9.5	9.3	10.7	10.3	9.3	5.4	5.2	4.7	4.6	3.7
200	Proposed	10.1	10.8	9.0	10.1	8.2	6.3	6.0	3.6	4.3	4.4
	CLX	3.2	4.5	3.0	3.4	4.8	5.1	4.0	3.7	4.6	4.3
	LC	8.8	10.6	9.0	10.7	8.2	5.7	5.2	4.1	3.8	5.0
						Power under sparse alternative					
100	Proposed	97.6	96.4	88.1	84.8	81.5	99.9	85.2	78.9	72.5	86.7
	CLX	98.8	98.1	92.4	89.3	86.5	100.0	90.0	83.5	77.8	90.9
	LC	19.3	12.0	6.8	5.9	5.0	33.1	11.3	6.9	5.2	4.6
200	Proposed	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	CLX	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	LC	52.3	22.1	8.8	7.2	7.3	80.3	20.4	8.0	8.6	6.9
						Power under dense alternative					
100	Proposed	84.1	89.7	92.2	95.5	96.8	100.0	100.0	100.0	100.0	100.0
	CLX	57.4	62.8	67.3	76.3	76.4	34.9	14.0	6.9	5.3	5.1
	LC	84.5	89.4	92.4	95.8	96.5	100.0	100.0	100.0	100.0	100.0
200	Proposed	98.9	98.7	99.8	99.9	100.0	100.0	100.0	100.0	100.0	100.0
	CLX	88.6	90.3	95.6	97.1	98.0	94.2	52.0	12.8	8.8	6.9
	LC	99.1	98.9	99.9	99.8	100.0	100.0	100.0	100.0	100.0	100.0

Note: This table reports the frequencies of rejection by each method under the null and alternative hypotheses based on 1000 independent replications at significance level 5%.

The size and power comparisons from Tables 4.1, 4.2 and 4.3 give us some intriguing findings:

(1) Under H_0, the sizes of all three tests are well retained close to the nominal level 0.05, except for Model (iv), in which both the LC test and our proposed test suffer from the size distortion, because of the violation of the test assumptions on covariance matrices.

(2) As can be seen from Model (i), the CLX test is demonstrated to be powerful under the sparse alternative H_s, however, its performance is not satisfactory under the dense alternative. Even though in Models (ii)-(iv), the CLX test

still has competitive powers, it fails with a decaying power as dimension grows in Model (v).

(3) In the meantime, the LC test remains a high power under the dense alternative H_d, whereas performs poorly against the sparse alternative with a tendency of decaying as dimension p grows large.

(4) In comparison, our proposed Fisher's combined test exhibits competent results. Our proposed test performs as good as the CLX test under the sparse alternative, together with the comparable performance to the LC test when against the dense alternative.

In a summary, based on the simulation results in this section, we are able to say that the proposed Fisher test boost the power tremendously against more general alternatives, in the meanwhile, retaining the desired nominal significance level.

4.6 Application to Gene-Set Testing

We further demonstrate the power of our proposed test by applying the test to identify those sets of genes which potentially have significant differences in covariance matrices across different types of tumors. In biology, each gene does not work individually, but rather tends to function as groups to achieve complex biological tasks. Sets of genes are interpreted by Gene Ontology (GO) terms making use of the Gene Ontology system, in which genes are assigned to a set of predefined bins depending on their functional characteristics. The Gene Ontology covers three domains: biological process (BP), cellular component (CC) and molecular function (MF).

We consider the Acute Lymphoblastic Leukemia(ALL) data from the Ritz Laboratory at the Dana-Farber Cancer Institute (DFCI). The latest data is accessible at the ALL package (version 1.24.0) on Bioconductor website, including the original version published by Chiaretti et al. (2004). The ALL dataset consists of microarrays expression measures of 12,625 probes on Affymetrix chip series HG-U95Av2 for 128 different individuals with acute lymphoblastic leukemia, which is a type of blood cancer in that bone marrow affects white blood cells. Based on the type of lymphocyte that the leukemia cells come from, the disease is classified into subgroups of T-cell ALL and B-cell ALL. In our study, we focus on a subset of the ALL data of 79 patients with the B-cell ALL. We are interested in two types of B-cell tumors: BCR/ABL and NEG, with sample sizes being 37 and 42 respectively.

Let us consider K gene sets $S_1, \cdots, S_K$, and $\mathbf{\Sigma}_{1S_k}$ and $\mathbf{\Sigma}_{2S_k}$ be the covariance matrices of two types of tumors respectively. The null hypotheses we are interested are

$$H_{0,category} : \mathbf{\Sigma}_{1S_k} = \mathbf{\Sigma}_{2S_k}, \quad k = 1, \cdots, K$$

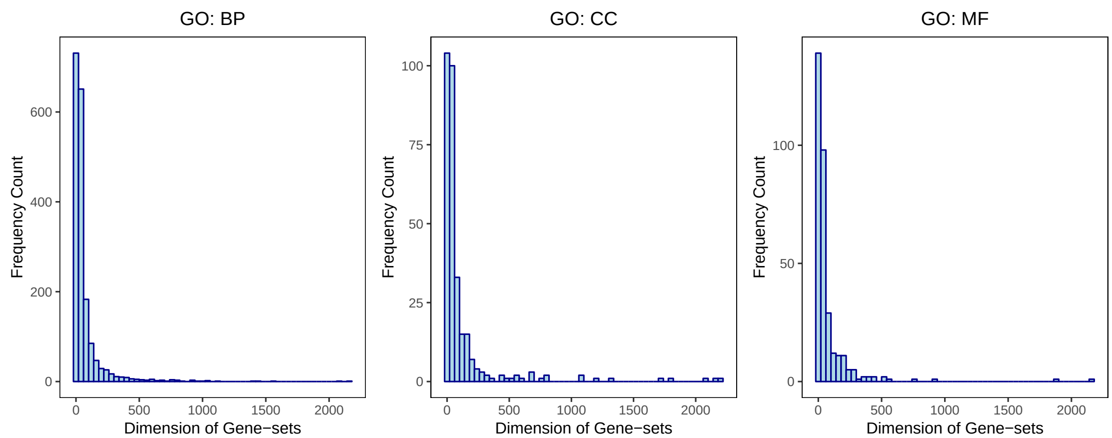
where $category \in \{BP, CC, MF\}$ because we classify the gene sets into three different GO categories and shall test each GO category separately.

To control the computational costs, we first perform a pre-screening procedure following the same criteria as in Dudoit et al. (2008) by choosing those probes that satisfy (i) the fluorescence intensities greater than 100 (absolute scale) for at least 25% of the 79 cell samples; (ii) the interquartile range (IQR) of the fluorescence intensities for the 79 cell samples greater than 0.5 (log base 2 scale). The preliminary gene-filtering retains 2,391 probes. After that we then identify those GO terms annotating at least 10 of the 2,391 filtered probes, which gives us 1849 unique GO terms in BP category, 306 in CC and 324 in MF for further analysis. Table 4.4 and Figure 4.1 summarize the dimension of gene-sets contained in each category.

Table 4.4: Summary of the dimension of gene-sets for three GO categories

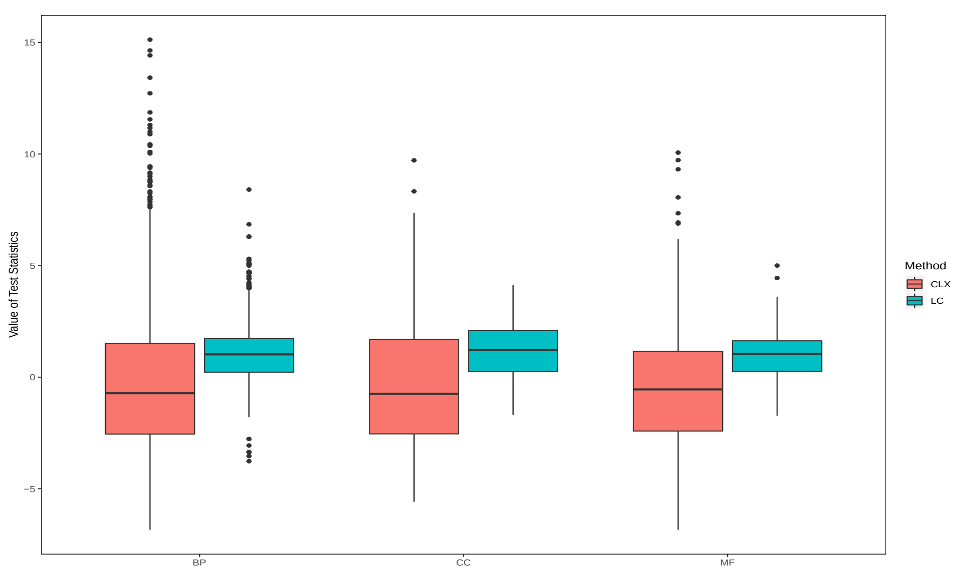
GO Category	Total number	Min	1st-Quantile	Median	3rd-Quantile	Max
BP	1849	10	15	27	62	2153
CC	306	10	17	32	85	2181
MF	324	10	14	26	68	2148

Figure 4.1: Histograms of the dimension of gene-sets for three GO categories



We first take a look at the performance of the CLX test and the LC test. Figure 4.2 displays boxplots of both test statistics. It can be observed that test statistics have quite different magnitudes, indicating difficulty in the approach of weighted summation combination of the two statistics.

Figure 4.2: Boxplots of the LC and CLX test statistics for three GO categories



We then apply our proposed Fisher's method to test the hypothesis, together with comparisons to the CLX and LC tests. We also compare our test with the natural Bonferroni combination. The test outcomes are reported in Table 4.5, with nominal level $\alpha = 0.05$ for each test. Furthermore, in order to control the false discovery rate (FDR), we apply the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg, 1995) to each GO category, and the results are listed in Table 5.10, with nominal level $\alpha = 0.05$ for every category.

Table 4.5: Gene-set testing results at the nominal level $\alpha = 0.05$

GO Category	Total number of Gene-sets	Number of Significant Gene-sets			
		CLX	LC	Bonferroni	Proposed
BP	1849	297	505	451	615
CC	306	52	111	96	116
MF	324	38	78	61	96

Table 4.6: Gene-set testing results with the FDR control at $\alpha = 0.05$

GO Category	Total number of Gene-sets	Number of Significant Gene-sets			
		CLX	LC	Bonferroni	Proposed
BP	1849	0	126	81	254
CC	306	0	55	24	68
MF	324	0	20	4	26

As shown in Table 4.6, our proposed test identifies much more significant gene-sets than the other methods. The LC identifies a few while the Bonferroni test identifies fewer significant gene-sets than the LC test does. This illustrates that the Bonferroni test is relatively conservative, which is consistent with what we expect. Unfortunately, the CLX test fails to declare any significance after we control the FDR using BH procedure. This is possibly because the signals in the differences are not strong enough for the CLX test to detect.

Biological evidence supports that such improvement is quite meaningful and very helpful in cancer research. To clarify this, we further investigate those gene-sets that are not declared significant by the CLX and LC tests but are identified by our proposed Fisher test. Taking the GO term "GO:0005905" as an example, it refers to the clathrin-coated pit which functions in the cellular component (CC) gene ontology category. Protein evidence by Ezkurdia et al. (2014) confirms that the clathrin-coated pit works with several protein-coding genes, such as CLTCL1, PICALM, etc., that are closely related to human cancers. We also take a deep look at "GO:0035259", the glucocorticoid receptor binding, in the molecular function (MF) gene ontology category. Many genes contribute to this gene-set, among them, we pay special attention to STAT3, a protein-coding gene which plays an important role in the immune system by transmitting signals for the maturation of immune system cells, especially T-cells and B-cells. Researchers have observed that STAT3 gene mutations are highly correlated with cancers, especially blood cancers (Hodge et al., 2005; Jerez et al., 2012; Haapaniemi et al., 2015; Milner et al., 2015). In a short summary, our proposed test incorporates the information from the CLX statistic, which successfully enhances the power over the LC test, even though the LC test itself may not declare any significance.

4.7 Conclusion

This work studies the fundamental problem of testing high-dimensional covariance matrices. Unlike the existing quadratic form statistics, maximum form statistics, and their weighted combination, we provide a new perspective to exploit the full potential of quadratic form statistics and maximum form statistics. We propose a scale-invariant and computationally efficient power enhancement test based on Fisher's method to combine their respective p-values. Theoretically, after deriving their joint limiting null distribution, we prove that the proposed combination method retains the correct asymptotic size and boosts the power against more general alternatives. Numerically, we demonstrate the finite-sample properties in simulation studies and the practical relevance through an empirical study on gene-set testing problem.

It is still an open question to relax the Gaussian assumption when deriving the asymptotic joint distribution of quadratic form statistics and maximum form statistics in the two-sample covariance tests. There are several potential directions to relax the Gaussian assumption. For instance, we may use the semiparametric Gaussian copula distribution (Liu et al., 2012; Xue and Zou, 2012, 2014*b*). Alternatively, we may use the Gaussian approximation theory (Chernozhukov et al., 2013, 2014) to bridge this gap. We will leave this open question for future work.

4.8 Lemmas and Proofs

4.8.1 Lemmas

Before proceeding, we first introduce some necessary notation. Throughout the proofs, we use C and c to denote some generic positive constants chosen sufficiently large and small, respectively, which may have different values at each appearance. And we use $\{c_1, c_2, \cdots\}$ to represent some positive absolute constants that does not depend on p or n. For any matrix $\mathbf{A}$, let $\lambda_i(\mathbf{A})$ be the i-th largest eigenvalue and $s_i(\mathbf{A})$ be the i-th largest sigular value of $\mathbf{A}$. Moreover, for ease of notation, let $\nu_i = \lambda_i(\mathbf{\Sigma})$ be the i-th largest eigenvalue of the common covariance matrix $\mathbf{\Sigma}$ under the null hypothesis.

Lemma 4.4. Consider $\boldsymbol{\xi}_1 = (\xi_{11}, \cdots, \xi_{1m})' \in \mathbb{R}^m$ and $\boldsymbol{\xi}_2 = (\xi_{21}, \cdots, \xi_{2r})' \in \mathbb{R}^r$, where ξ_{ij} are independently distributed random variables with mean 0, variance 1, and even-order moments bounded by

$$E|\xi_{ij}|^{2k} \leqslant \frac{1}{2}k!K^{2(k-1)}, \quad k \geqslant 2 \tag{4.8.1}$$

for some constant $K > 0$. Let $\mathbf{A}$ be any $m \times r$ real matrix. Then, for any $t > 0$,

$$P\left(\boldsymbol{\xi}_1'\mathbf{A}\boldsymbol{\xi}_2 > t\right) \leqslant \exp\left(-\min\left\{\frac{t^2}{12K^2\mathrm{tr}(\mathbf{A}'\mathbf{A})}, \frac{t}{32K^2\sqrt{\lambda_1(\mathbf{A}'\mathbf{A})}}\right\}\right). \tag{4.8.2}$$

Remark 4.9. As a special case, if $\xi_{ij} \sim N(0,1)$, we can choose $K^2 = 3$. However, taking advantage of the nice properties of Gaussian distribution, we can obtain a tighter upper bound as follows.

$$P\left(\boldsymbol{\xi}_1'\mathbf{A}\boldsymbol{\xi}_2 > t\right) \leqslant \exp\left(-\min\left\{\frac{t^2}{6\mathrm{tr}(\mathbf{A}'\mathbf{A})}, \frac{t}{16\sqrt{\lambda_1(\mathbf{A}'\mathbf{A})}}\right\}\right). \tag{4.8.3}$$

Lemma 4.5 (Bai and Silverstein (2010), Theorem A.15)**.** Let $\mathbf{A}_1$ and $\mathbf{A}_2$ be complex matrices of order $m \times r$, we have

$$\sum_{i=1}^{m\wedge r} s_i(\mathbf{A}_1^H\mathbf{A}_2) \leqslant \sum_{i=1}^{m\wedge r} s_i(\mathbf{A}_1)s_i(\mathbf{A}_2). \tag{4.8.4}$$

where $\mathbf{A}_1^H$ indicates conjugate transpose of matrix $\mathbf{A}_1$.

4.8.2 Proof of Lemma 4.1

Proof. It's easy to verify that $\{\mathbf{Z}_{1u}\}_{u=1}^{n_1}$ and $\{\mathbf{W}_{1u}\}_{u=1}^{n_2}$ are identically and independently following $N_{p-q}\left(\mathbf{0}, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)$. Based on Li and Chen (2012), in order to obtain

$$\frac{T^*_{n_1,n_2}}{2\left(n_1^{-1} + n_2^{-1}\right)\operatorname{tr}\left(\left(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)^2\right)} \xrightarrow{d} N(0,1), \tag{4.8.5}$$

we only need to show that

$$\frac{\operatorname{tr}\left(\left(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)^4\right)}{\operatorname{tr}^2\left(\left(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)^2\right)} \to 0. \tag{4.8.6}$$

Note that

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{p-q} & \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{p-q} & \mathbf{0} \\ \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I}_q \end{pmatrix}.$$

Then,

$$\begin{aligned}
\operatorname{tr}(\boldsymbol{\Sigma}^2) &= \operatorname{tr}\left\{\left[\begin{pmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{p-q} & \mathbf{0} \\ \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I}_q \end{pmatrix} \begin{pmatrix} \mathbf{I}_{p-q} & \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix}\right]^2\right\} \\
&= \operatorname{tr}\left\{\begin{pmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} + \boldsymbol{\Sigma}_{22} \end{pmatrix}^2\right\} \\
&:= \operatorname{tr}\left\{\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^2\right\} = \operatorname{tr}\left(\mathbf{A}^2 + \mathbf{BC}\right) + \operatorname{tr}(\mathbf{CB} + \mathbf{D}^2),
\end{aligned}$$

where $\mathbf{A} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$, $\mathbf{B} = (\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}$, $\mathbf{C} = \mathbf{\Sigma}_{21}$ and $\mathbf{D} = \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1} + \mathbf{\Sigma}_{22}$. Similarly,

$$\mathrm{tr}(\mathbf{\Sigma}^4) = \mathrm{tr}\left\{\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{pmatrix}^4\right\}.$$

Define $\mathbf{E} = \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$. As a result,

$$\mathrm{tr}\left(\mathbf{\Sigma}^2\right) - \mathrm{tr}\left(\left(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right)^2\right) = 2\mathrm{tr}\left(\mathbf{AE}\right) + \mathrm{tr}\left(\mathbf{E}^2\right) + 2\mathrm{tr}\left(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\right) + \mathrm{tr}\left(\mathbf{\Sigma}_{22}^2\right).$$

By construction, it is easy to see that the matrix $\mathbf{E}$ has at most q non-zero eigenvalues. By Weyl's inequality and interlacing theorem,

$$\lambda_i\left(\mathbf{E}\right) \leqslant \lambda_i\left(\mathbf{\Sigma}_{11}\right) \leqslant \lambda_i\left(\mathbf{\Sigma}\right) = \nu_i \quad \text{for } i = 1, \cdots, q.$$

Therefore, $\mathrm{tr}\left(\mathbf{E}^2\right) \leqslant \sum_{i=1}^q \nu_i^2$. Similarly, we obtain that $\mathrm{tr}\left(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\right) + \mathrm{tr}\left(\mathbf{\Sigma}_{22}^2\right) \leqslant \sum_{i=1}^q \nu_i^2$. Moreover, by Lemma 4.5,

$$\mathrm{tr}\left(\mathbf{AE}\right) \leqslant \sum_{i=1}^q \lambda_i\left(\mathbf{AE}\right) \leqslant \sum_{i=1}^q \lambda_i^2\left(\mathbf{\Sigma}_{11}\right) \leqslant \sum_{i=1}^q \nu_i^2.$$

Then we have,

$$\mathrm{tr}\left(\mathbf{\Sigma}^2\right) - \mathrm{tr}\left(\left(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right)^2\right) \leqslant C\sum_{i=1}^q \nu_i^2.$$

With Assumption 4.1(ii), $\sum_{i=1}^q \nu_i^2 / \sum_{i=1}^p \nu_i^2 \to 0$, we prove that

$$\frac{\mathrm{tr}\left(\left(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right)^2\right)}{\mathrm{tr}\left(\mathbf{\Sigma}^2\right)} \to 1. \tag{4.8.7}$$

Furthermore, note that

$$\mathrm{tr}\left(\mathbf{\Sigma}^4\right) - \mathrm{tr}\left(\left(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right)^4\right) \geqslant 0.$$

Hence, (4.8.6) holds, and so does (4.8.5). Together with (4.8.7), we complete the proof of asymptotic normality of $T^*_{n_1,n_2}$ shown in (4.4.4). □

4.8.3 Proof of Lemma 4.2

Proof. To begin with, we rewrite $\widetilde{T}_{n_1,n_2}$ into the form of two-sample U-statistic

$$\widetilde{T}_{n_1,n_2} = \frac{1}{n_1(n_1-1)n_2(n_2-1)} \sum_{u \neq v}^{n_1} \sum_{k \neq l}^{n_2} H(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l), \tag{4.8.8}$$

where

$$H(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l) = (\mathbf{X}_u'\mathbf{X}_v)^2 + (\mathbf{Y}_k'\mathbf{Y}_l)^2 - \frac{(\mathbf{X}_u'\mathbf{Y}_k)^2 + (\mathbf{X}_v'\mathbf{Y}_l)^2 + (\mathbf{X}_u'\mathbf{Y}_l)^2 + (\mathbf{X}_v'\mathbf{Y}_k)^2}{2}.$$

Similarly, $T^*_{n_1,n_2}$ can be written as

$$T^*_{n_1,n_2} = \frac{1}{n_1(n_1-1)n_2(n_2-1)} \sum_{u \neq v}^{n_1} \sum_{k \neq l}^{n_2} H(\mathbf{Z}_{1u}, \mathbf{Z}_{1v}, \mathbf{W}_{1k}, \mathbf{W}_{1l}). \tag{4.8.9}$$

In addition, we define

$$H_0(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l) := H(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l) - H(\mathbf{Z}_{1u}, \mathbf{Z}_{1v}, \mathbf{W}_{1k}, \mathbf{W}_{1l}) \tag{4.8.10}$$

Moreover, let

$$\begin{aligned} H_1(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l) :=& H_0(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l) I\{|H_0(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l)| \leqslant c_n\} \\ & - E\left[H_0(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l) I\{|H_0(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l)| \leqslant c_n\}\right] \end{aligned} \tag{4.8.11}$$

where c_n is a constant depending on $n = n_1 + n_2$. For ease of the notation, in the following context, we simplify $H_0(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l)$ by $H_{0,uvkl}$, and $H_1(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l)$ by $H_{1,uvkl}$. It's easy to verify both $H_{0,uvkl}$ and $H_{1,uvkl}$ are symmetric and degenerate. The difference between $\widetilde{T}_{n_1,n_2}$ and $T^*_{n_1,n_2}$, denoted by U_{n_1,n_2}, can be further written as

$$U_{n_1,n_2} := \widetilde{T}_{n_1,n_2} - T^*_{n_1,n_2} = \frac{1}{n_1(n_1-1)n_2(n_2-1)} \sum_{u \neq v}^{n_1} \sum_{k \neq l}^{n_2} H_{0,uvkl}$$

$$= \frac{\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2}\left(H_{1,uvkl} + H_{0,uvkl}I\{|H_{0,uvkl}| > c_n\} - E\left[H_{0,uvkl}I\{|H_{0,uvkl}| > c_n\}\right]\right)}{n_1(n_1-1)n_2(n_2-1)}.$$

The problem of interest (4.4.5) becomes to prove

$$P\left(\frac{|U_{n_1,n_2}|}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)} \geqslant \epsilon\right) = P\left(|n_1 n_2 U_{n_1,n_2}| \geqslant 2\epsilon(n_1+n_2)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)\right)$$
$$\leqslant C\exp(-c\epsilon n^{\beta}). \tag{4.8.12}$$

Furthermore,

$$(4.8.12) \leqslant P\left(\left|\frac{1}{(n_1-1)(n_2-1)}\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2} H_{1,uvkl}\right| \geqslant \epsilon(n_1+n_2)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)\right) \tag{4.8.13}$$
$$+P\left(\left|\frac{\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2} H_{0,uvkl}I\{|H_{0,uvkl}| > c_n\}}{(n_1-1)(n_2-1)}\right| \geqslant \frac{\epsilon}{2}(n_1+n_2)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)\right) \tag{4.8.14}$$
$$+P\left(\left|\frac{\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2} E\left[H_{0,uvkl}I\{|H_{0,uvkl}| > c_n\}\right]}{(n_1-1)(n_2-1)}\right| \geqslant \frac{\epsilon}{2}(n_1+n_2)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)\right). \tag{4.8.15}$$

We first deal with the first term (4.8.13). In here, we generalize the proof of Proposition 2.3 in Arcones and Gine (1993) to two-sample degenerate U-statistics. Let $a_{uvkl} := H_{1,uvkl}$, $\{\epsilon_i\}_{i=1}^{n_1+n_2}$ are i.i.d. Rademacher variables that are independent of $\{\mathbf{X}_u\}_{u=1}^{n_1}$ and $\{\mathbf{Y}_v\}_{v=1}^{n_2}$. Now define

$$M = \sum_{u<v}\sum_{k<l}\epsilon_u\epsilon_v\epsilon_{k+n_1}\epsilon_{l+n_1}a_{uvkl}.$$

Let $v^2 = \sum_{u<v}\sum_{k<l}a_{uvkl}^2$, then $E\left(M^2/v^2\right) = 1$. We now show that for any $0 < \lambda < (2e)^{-1}$, there exists $\mu(\lambda) < \infty$ independent of v, such that

$$Ee^{\lambda\left|\frac{M}{v}\right|^{1/2}} \leqslant \mu(\lambda). \tag{4.8.16}$$

This is because

$$\begin{aligned} Ee^{\lambda|\frac{M}{v}|^{1/2}} &= \sum_{k=0}^{\infty} E\left(\frac{\lambda^k}{k!}\left|\frac{M}{v}\right|^{k/2}\right) = \sum_{k=0}^{4}\frac{\lambda^k}{k!}E\left|\frac{M}{v}\right|^{k/2} + \sum_{k=5}^{\infty}\frac{\lambda^k}{k!}E\left|\frac{M}{v}\right|^{k/2} \\ &\leqslant \sum_{k=0}^{4}\frac{\lambda^k}{k!}\left(E\left|\frac{M}{v}\right|^{2}\right)^{k/4} + \sum_{k=5}^{\infty}\frac{\lambda^k}{k!}(k-1)^k\left(E\left|\frac{M}{v}\right|^{2}\right)^{k/4} \\ &= \sum_{k=0}^{4}\frac{\lambda^k}{k!} + \sum_{k=5}^{\infty}\frac{\lambda^k}{k!}(k-1)^k. \end{aligned}$$

The inequality holds as a result of Holder's inequality and the inequality proved by Borell (1979) that for all $1 < q < p < \infty$,

$$(E|M|^p)^{\frac{1}{p}} \leqslant \left(\frac{p-1}{q-1}\right)^2 \cdot (E|M|^q)^{\frac{1}{q}}.$$

By Stirling's formula,

$$Ee^{\lambda|\frac{M}{v}|^{1/2}} \leqslant \sum_{k=0}^{4}\frac{\lambda^k}{k!} + \sum_{k=5}^{\infty}(\lambda e)^k. \tag{4.8.17}$$

The right-hand side of (4.8.17) converges to $\mu(\lambda)$ when $\lambda < (2e)^{-1}$.
What's more, for any $0 < \alpha < 1/2$, $c > 0$, and $t > 0$,

$$t|M|^{\alpha} \leqslant 2\alpha \cdot \left|c\frac{M}{v}\right|^{1/2} + (1-2\alpha)\cdot\left(\frac{tv^{\alpha}}{c^{\alpha}}\right)^{\frac{1}{1-2\alpha}}. \tag{4.8.18}$$

Taking c such that $2\alpha c^{1/2} < (2e)^{-1}$ and combining (4.8.16) with (4.8.18), we are able to say, there exists $c_1, c_2 > 0$ depending on α, such that for all $t > 0$ and $0 < \alpha < 1/2$,

$$Ee^{t|M|^{\alpha}} \leqslant c_1 E\exp\left\{c_2(tv^{\alpha})^{\frac{1}{1-2\alpha}}\right\}. \tag{4.8.19}$$

Take $\alpha = 0.4$, (4.8.19) becomes

$$Ee^{t|M|^{\frac{2}{5}}} \leqslant c_1 E\exp\left\{c_2 t^5 v^2\right\}.$$

Then apply the above inequality to (4.8.13),

$$
\begin{aligned}
& P\left(\left|\frac{1}{(n_1-1)(n_2-1)}\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2} H_{1,uvkl}\right| \geqslant \epsilon(n_1+n_2)\mathrm{tr}\left(\mathbf{\Sigma}^2\right)\right) \\
= & P\left(\left|\frac{\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2} H_{1,uvkl}}{\sqrt{n_1(n_1-1)n_2(n_2-1)}}\right| \geqslant \epsilon(n_1+n_2)\mathrm{tr}\left(\mathbf{\Sigma}^2\right)\sqrt{\frac{(n_1-1)(n_2-1)}{n_1n_2}}\right) \\
\leqslant & E\exp\left\{t\cdot\left|\frac{\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2} H_{1,uvkl}}{\sqrt{n_1(n_1-1)n_2(n_2-1)}}\right|^{\frac{2}{5}} - tu^{\frac{2}{5}}\right\} \\
\leqslant & c_1E\exp\left\{\frac{16c_2t^5}{n_1(n_1-1)n_2(n_2-1)}\sum_{u<v}^{n_1}\sum_{k<l}^{n_2} H_{1,uvkl}^2\right\}\cdot\exp\{-tu^{\frac{2}{5}}\} \\
\leqslant & c_1E\exp\left\{\frac{4c_2t^5}{n_1(n_1-1)n_2(n_2-1)}\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2}(H_{1,uvkl}^2-\sigma^2)\right\}\cdot\exp\{4c_2t^5\sigma^2 - tu^{\frac{2}{5}}\},
\end{aligned}
\tag{4.8.20}
$$

where $\sigma^2 = EH_{1,uvkl}^2$ and $u = \sqrt{n_1^{-1}n_2^{-1}(n_1-1)(n_2-1)}(n_1+n_2)\cdot\epsilon\mathrm{tr}\left(\mathbf{\Sigma}^2\right)$. Sharing the spirit with Hoeffding (1963), we rewrite

$$
\frac{1}{n_1(n_1-1)n_2(n_2-1)}\sum_{u\neq v}\sum_{k\neq l}(H_{1,uvkl}^2-\sigma^2) \tag{4.8.21}
$$

as the average of $V(\mathbf{X}_{u_1},\cdots,\mathbf{X}_{u_{n_1}},\mathbf{Y}_{v_1},\cdots,\mathbf{Y}_{v_{n_2}})$ over all the permutations $(u_1,\ldots,u_{n_1})$ of $\{1,\ldots,n_1\}$ and $(v_1,\ldots,v_{n_2})$ of $\{1,\ldots,n_2\}$, i.e.,

$$
(4.8.21) = \frac{1}{n_1!n_2!}\sum_{(u_1,\cdots,u_{n_1})}\sum_{(v_1,\cdots,v_{n_2})} V(\mathbf{X}_{u_1},\cdots,\mathbf{X}_{u_{n_1}},\mathbf{Y}_{v_1},\cdots,\mathbf{Y}_{v_{n_2}}).
$$

where $V(\mathbf{X}_1,\cdots,\mathbf{X}_{n_1},\mathbf{Y}_1,\cdots,\mathbf{Y}_{n_2})$ are defined by

$$
V(\mathbf{X}_1,\cdots,\mathbf{X}_{n_1},\mathbf{Y}_1,\cdots,\mathbf{Y}_{n_2}) = \frac{1}{m}\sum_{i=1}^{m} g(\mathbf{X}_{2i-1},\mathbf{X}_{2i},\mathbf{Y}_{2i-1},\mathbf{Y}_{2i}),
$$

with $m := \min\{[n_1/2],[n_2/2]\}$, and $g(\mathbf{X}_u,\mathbf{X}_v,\mathbf{Y}_k,\mathbf{Y}_l) := H_{1,uvkl}^2 - \sigma^2$.

Because of the convexity of exponential function, we have

$$\begin{aligned}&E\exp\left\{\frac{4c_2t^5}{n_1(n_1-1)n_2(n_2-1)}\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2}(H_{1,uvkl}^2-\sigma^2)\right\}\\ \leqslant&E\exp\left\{\frac{4c_2t^5}{m}\sum_{i=1}^{m}g(\mathbf{X}_{2i-1},\mathbf{X}_{2i},\mathbf{Y}_{2i-1},\mathbf{Y}_{2i})\right\}.\end{aligned} \tag{4.8.22}$$

Since

$$Eg=0,\quad |g|\leqslant 8c_n^2,\quad Eg^2\leqslant EH_{1,uvkl}^4\leqslant 4c_n^2\sigma^2,$$

by Bernstein inequality for i.i.d. random variables,

$$(4.8.22)\leqslant\exp\left\{\frac{16c_2^2t^{10}/m\cdot 4c_n^2\sigma^2}{2-\frac{2}{3}\cdot\frac{4c_2t^5}{\sqrt{m}}\cdot 8c_n^2m^{-1/2}}\right\}=\exp\left\{\frac{64c_2^2t^{10}c_n^2\sigma^2}{2m-\frac{64}{3}c_2t^5c_n^2}\right\},\quad |t|^5<\frac{3m}{32c_2c_n^2}.$$

Together with (4.8.20),

$$P\left(\left|\frac{\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2}H_1}{\sqrt{n_1(n_1-1)n_2(n_2-1)}}\right|\geqslant u\right)\leqslant\ c_1\exp\left\{-tu^{\frac{2}{5}}+4c_2t^5\sigma^2+\frac{64c_2^2t^{10}c_n^2\sigma^2}{2m-\frac{64}{3}c_2t^5c_n^2}\right\},$$

where $u=\sqrt{n_1^{-1}n_2^{-1}(n_1-1)(n_2-1)}(n_1+n_2)\cdot\epsilon\mathrm{tr}\left(\mathbf{\Sigma}^2\right)$ and $0<t<(3m/(32c_2c_n^2))^{\frac{1}{5}}$.

Following similar arguments to Proposition 2.3 of Arcones and Gine (1993),

$$P\left(\left|\frac{\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2}H_{1,uvkl}}{\sqrt{n_1(n_1-1)n_2(n_2-1)}}\right|\geqslant u\right)\leqslant c_1\exp\left\{-\frac{c_2u^{1/2}}{\sigma^{1/2}+(c_nu^{1/4}m^{-1/2})^{\frac{2}{5}}}\right\}.$$

Let $n:=n_1+n_2$, note that $\sigma^2\leqslant EH_{0,uvkl}^2\leqslant C\mathrm{tr}^2(\mathbf{\Sigma}^2)$, if we choose c_n such that

$$c_n=o\left(um^{1/2}\epsilon^{-\frac{5}{2}}n^{-\frac{5}{2}\beta}\right),\quad \frac{1}{5}<\beta<\frac{1}{2}, \tag{4.8.23}$$

then,

$$(4.8.13)\leqslant\exp\{-cn^{\beta}\}. \tag{4.8.24}$$

Now we will focus on (4.8.14) and (4.8.15). For (4.8.15), note that

$$E\left[H_{0,uvkl}I\{H_{0,uvkl}>c_n\}\right]\leqslant\sqrt{EH_{0,uvkl}^2}\cdot\sqrt{P(|H_{0,uvkl}|>c_n)}.$$

Hence if we have

$$P(|H_{0,uvkl}| > c_n) = o\left(\frac{\epsilon^2 n^2 \mathrm{tr}^2\left(\mathbf{\Sigma}^2\right)}{4n_1^2 n_2^2 \cdot EH_{0,uvkl}^2}\right),$$

then (4.8.15)=0.
Also for (4.8.14),

$$\begin{aligned}(4.8.14) &= P\left(\left|\frac{\sum_{u\neq v}^{n_1}\sum_{k\neq l}^{n_2} H_{0,uvkl} I\{H_{0,uvkl} > c_n\}}{(n_1-1)(n_2-1)}\right| \geqslant \frac{\epsilon}{2}(n_1+n_2)\mathrm{tr}\left(\mathbf{\Sigma}^2\right)\right)\\ &\leqslant P\left(\max_{1\leqslant u\neq v\leqslant n_1}\max_{1\leqslant k\neq l\leqslant n_2}|H_{0,uvkl}| > c_n\right) \leqslant n_1^2 n_2^2 P\left(|H_{0,uvkl}| > c_n\right).\end{aligned}$$

From the above discussion, it turns out in order to handle (4.8.14) and (4.8.15), it suffices to choose a proper c_n such that the probability of $|H_{0,uvkl}|$ larger than c_n is sufficiently small. More explicitly, we want to prove

$$P\left(|H_{0,uvkl}| > c_n\right) \leqslant C\exp\{-c\epsilon n^{\beta}\}. \tag{4.8.25}$$

By definition,

$$\begin{aligned}\mathbf{X}_u'\mathbf{X}_v &= \left(\mathbf{Z}_{1u}' + \left(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{Z}_{2u}\right)', \quad \mathbf{Z}_{2u}'\right)\begin{pmatrix}\mathbf{Z}_{1v} + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{Z}_{2v}\\ \mathbf{Z}_{2v}\end{pmatrix}\\ &= \mathbf{Z}_{1u}'\mathbf{Z}_{1v} + \mathbf{Z}_{2u}'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{Z}_{1v} + \mathbf{Z}_{1u}'\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{Z}_{2v} + \mathbf{Z}_{2u}'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{Z}_{2v} + \mathbf{Z}_{2u}'\mathbf{Z}_{2v}\\ &:= \mathbf{Z}_{1u}'\mathbf{Z}_{1v} + \mathrm{res}\left(\mathbf{X}_{uv}\right).\end{aligned}$$

Similarly, we can obtain

$$\mathbf{Y}_u'\mathbf{Y}_v = \mathbf{W}_{1u}'\mathbf{W}_{1v} + \mathrm{res}\left(\mathbf{Y}_{uv}\right),$$

$$\mathbf{X}_u'\mathbf{Y}_v = \mathbf{Z}_{1u}'\mathbf{W}_{1v} + \mathrm{res}\left(\mathbf{XY}_{uv}\right),$$

with $\mathrm{res}\left(\mathbf{Y}_{uv}\right) = \mathbf{W}_{2u}'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{W}_{1v} + \mathbf{W}_{1u}'\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{W}_{2v} + \mathbf{W}_{2u}'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{W}_{2v} + \mathbf{W}_{2u}'\mathbf{W}_{2v}$, $\mathrm{res}\left(\mathbf{XY}_{uv}\right) = \mathbf{Z}_{2u}'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{W}_{1v} + \mathbf{Z}_{1u}'\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{W}_{2v} + \mathbf{Z}_{2u}'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{W}_{2v}$ $+\mathbf{Z}_{2u}'\mathbf{W}_{2v}$.

Then $H_{0,uvkl}$ can be written as

$$\begin{aligned} H_{0,uvkl} =& 2\mathbf{Z}_{1u}'\mathbf{Z}_{1v}\mathrm{res}\,(\mathbf{X}_{uv}) + \mathrm{res}\,(\mathbf{X}_{uv})^2 + 2\mathbf{W}_{1k}'\mathbf{W}_{1l}\mathrm{res}\,(\mathbf{Y}_{kl}) + \mathrm{res}\,(\mathbf{Y}_{kl})^2 \\ & - \frac{1}{2}\left[2\mathbf{Z}_{1u}'\mathbf{W}_{1k}\mathrm{res}\,(\mathbf{XY}_{uk}) + \mathrm{res}\,(\mathbf{XY}_{uk})^2\right] - \frac{1}{2}\left[2\mathbf{Z}_{1v}'\mathbf{W}_{1l}\mathrm{res}\,(\mathbf{XY}_{vl}) + \mathrm{res}\,(\mathbf{XY}_{vl})^2\right] \\ & - \frac{1}{2}\left[2\mathbf{Z}_{1u}'\mathbf{W}_{1l}\mathrm{res}\,(\mathbf{XY}_{ul}) + \mathrm{res}\,(\mathbf{XY}_{ul})^2\right] - \frac{1}{2}\left[2\mathbf{Z}_{1v}'\mathbf{W}_{1k}\mathrm{res}\,(\mathbf{XY}_{vk}) + \mathrm{res}\,(\mathbf{XY}_{vk})^2\right] \end{aligned}$$

Without loss of generality, we only need to show

$$P\left(|2\mathbf{Z}_{11}'\mathbf{Z}_{12}\mathrm{res}\,(\mathbf{X}_{12})\,| \geqslant \frac{c_n}{8}\right) \leqslant C\exp\{-c\epsilon n^{\beta}\} \tag{4.8.26}$$

and

$$P\left(\mathrm{res}\,(\mathbf{X}_{12})^2 \geqslant \frac{c_n}{8}\right) \leqslant C\exp\{-c\epsilon n^{\beta}\}. \tag{4.8.27}$$

For (4.8.26),

$$P\left(|2\mathbf{Z}_{11}'\mathbf{Z}_{12}\mathrm{res}\,(\mathbf{X}_{12})\,| > \frac{c_n}{8}\right) \leqslant P\left(|\mathbf{Z}_{11}'\mathbf{Z}_{12}| > \frac{c_n}{16d_n}\right) + P\left(|\mathrm{res}\,(\mathbf{X}_{12})\,| > d_n\right). \tag{4.8.28}$$

Taking advantage of Gaussian assumption, we consider $\mathbf{Z}_{1u}$ and $\mathbf{Z}_{2u}$ as

$$\mathbf{Z}_{1u} = \mathbf{\Gamma}_1\boldsymbol{\xi}_u, \quad \mathbf{Z}_{2u} = \mathbf{\Gamma}_2\boldsymbol{\eta}_u,$$

respectively, where $\boldsymbol{\xi}_u = (\xi_{u1}, \cdots, \xi_{u(p-q)})'$, $\boldsymbol{\eta}_u = (\eta_{u1}, \cdots, \eta_{uq})'$, and $\{\xi_{ui}\}_{i=1}^{p-q}$, $\{\eta_{uj}\}_{j=1}^{q}$ are independent standard normal random variables. Furthermore, $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ take into account the covariances such that

$$\mathbf{\Gamma}_1\mathbf{\Gamma}_1' = \mathrm{cov}(\mathbf{Z}_{1u}) = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21},$$

and

$$\mathbf{\Gamma}_2\mathbf{\Gamma}_2' = \mathrm{cov}(\mathbf{Z}_{2u}) = \mathbf{\Sigma}_{22}.$$

In this way, $\mathbf{Z}_{11}'\mathbf{Z}_{12}$ and $\mathrm{res}(\mathbf{X}_{12})$ can be reformulated into

$$\mathbf{Z}_{11}'\mathbf{Z}_{12} = \boldsymbol{\xi}_1'\mathbf{\Gamma}_1'\mathbf{\Gamma}_1\boldsymbol{\xi}_2 := \boldsymbol{\xi}_1'\mathbf{A}_1\boldsymbol{\xi}_2$$

and

$$\begin{aligned}
&\text{res}(\mathbf{X}_{12}) \\
&= \mathbf{Z}_{21}'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{Z}_{12} + \mathbf{Z}_{11}'\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{Z}_{22} + \mathbf{Z}_{21}'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{Z}_{22} + \mathbf{Z}_{21}'\mathbf{Z}_{22} \\
&= \boldsymbol{\eta}_1'\mathbf{\Gamma}_2'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Gamma}_1\boldsymbol{\xi}_2 + \boldsymbol{\xi}_1'\mathbf{\Gamma}_1'\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Gamma}_2\boldsymbol{\eta}_2 + +\boldsymbol{\eta}_1'\mathbf{\Gamma}_2'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Gamma}_2\boldsymbol{\eta}_2 + \boldsymbol{\eta}_1'\mathbf{\Gamma}_2'\mathbf{\Gamma}_2\boldsymbol{\eta}_2 \\
&:= \boldsymbol{\eta}_1'\mathbf{A}_2\boldsymbol{\xi}_2 + \boldsymbol{\eta}_2'\mathbf{A}_2\boldsymbol{\xi}_1 + \boldsymbol{\eta}_1'\mathbf{A}_3\boldsymbol{\eta}_2 + \boldsymbol{\eta}_1'\mathbf{A}_4\boldsymbol{\eta}_2.
\end{aligned}$$

where $\mathbf{A}_1 = \mathbf{\Gamma}_1'\mathbf{\Gamma}_1$, $\mathbf{A}_2 = \mathbf{\Gamma}_2'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Gamma}_1$, $\mathbf{A}_3 = \mathbf{\Gamma}_2'\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Gamma}_2$, and $\mathbf{A}_4 = \mathbf{\Gamma}_2'\mathbf{\Gamma}_2$. What's more, using the same arguments as in part(i), it's easy to verify that the squares of the Frobenius norm and operator norm of A_i (i=$1, 2, 3, 4$) can be bounded by

$$\begin{aligned}
&\text{tr}(\mathbf{A}_1'\mathbf{A}_1) \leqslant \text{tr}(\mathbf{\Sigma}^2), \ \lambda_1(\mathbf{A}_1'\mathbf{A}_1) \leqslant \nu_1^2, \\
\text{and} \quad &\text{tr}(\mathbf{A}_j'\mathbf{A}_j) \leqslant \sum_{j=1}^{q} \nu_i^2, \ \lambda_1(\mathbf{A}_j'\mathbf{A}_j) \leqslant \nu_1^2 \quad \text{for } j = 2, 3, 4.
\end{aligned}$$

By Lemma 4.4, take $t_1 = \max\left(\sqrt{6}\text{tr}^{1/2}\left(\mathbf{\Sigma}^2\right)\epsilon^{1/2}n^{\beta/2}, 16\nu_1\epsilon n^{\beta}\right)$, then

$$P\left(|\mathbf{Z}_{11}'\mathbf{Z}_{12}| > t_1\right) \leqslant 2\exp\{-\epsilon n^{\beta}\}. \tag{4.8.29}$$

Similarly, consider $t_2 = \max\left(\sqrt{6}(\sum_{i=1}^{q}\nu_i^2)^{1/2}\epsilon^{1/2}n^{\beta/2}, 16\nu_1\epsilon n^{\beta}\right)$, then

$$\begin{aligned}
&P\left(|\boldsymbol{\eta}_1'\mathbf{A}_2\boldsymbol{\xi}_2| > t_2\right) \leqslant 2\exp\{-\epsilon n^{\beta}\}, \quad P\left(|\boldsymbol{\eta}_2'\mathbf{A}_2\boldsymbol{\xi}_1| > t_2\right) \leqslant 2\exp\{-\epsilon n^{\beta}\}, \\
&P\left(|\boldsymbol{\eta}_1'\mathbf{A}_3\boldsymbol{\eta}_2| > t_2\right) \leqslant 2\exp\{-\epsilon n^{\beta}\}, \quad P\left(|\boldsymbol{\eta}_1'\mathbf{A}_4\boldsymbol{\eta}_2| > t_2\right) \leqslant 2\exp\{-\epsilon n^{\beta}\}.
\end{aligned}$$

As a result,

$$P\left(|\text{res}(\mathbf{X}_{12})| > 4t_2\right) \leqslant 8\exp\{-\epsilon n^{\beta}\}. \tag{4.8.30}$$

In combination with (4.8.27), (4.8.28) and (4.8.29), if we choose c_n such that

$$c_n \geqslant \max\{64t_1t_2, 128t_2^2\},$$

then we can conclude (4.8.25). Further because $\max\{64t_1t_2, 128t_2^2\} \leqslant 128t_1t_2 \leqslant C\text{tr}(\mathbf{\Sigma}^2)n^{2\beta}$, we let

$$c_n = C\epsilon\text{tr}(\mathbf{\Sigma}^2)n^{2\beta}.$$

It is easy to verify that c_n satisfies the condition (4.8.23) for $1/5 < \beta < 1/3$. Therefore we finish the proof of (4.4.5). $\square$

4.8.4 Proof of Lemma 4.3

Proof. Equipped with Lemmas 4.1 and 4.2 , we shall proceed to the proof of Lemma 4.3. Under the null hypothesis H_0, $\mathbf{X}, \mathbf{Y}$ i.i.d. follow $N_p(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = (\sigma_{ij})_{1\leqslant i,j\leqslant p}$. Therefore, for $1 \leqslant i, j \leqslant p$, $\theta_{ij1} = \text{var}(X_iX_j) = \text{var}(Y_iY_j) = \theta_{ij2}$. From now on, under the null hypothesis, we use θ_{ij} to denote both θ_{ij1} and θ_{ij2}. Let $\mathcal{A} = \{(i,j) : 1 \leqslant i \leqslant j \leqslant p\}$ be the set of two-dimensional indices. Following the idea of Cai et al. (2013), we first address those pairs of variables whose correlations are relatively high. More specifically, let $\mathcal{B}_1 = \left\{(i,j) \in \mathcal{A} : |\sigma_{ij}|/\sqrt{\sigma_{ii}\sigma_{jj}} \geqslant (\log p)^{-1-\alpha_0}\right\}$, and $\mathcal{B}_2 = \{(i,j) \in \mathcal{A} : i \in \mathcal{W}(r) \cup \Upsilon \text{ or } j \in \mathcal{W}(r) \cup \Upsilon, i \neq j\}$. As discussed by Cai et al. (2013), the effects of elements in sets $\mathcal{B}_1 \cup \mathcal{B}_2$ is negligible under Assumption 4.2. Therefore we only need to pay attention to the set of $\mathcal{A}\backslash\{\mathcal{B}_1 \cup \mathcal{B}_2\}$, denoted by $\mathcal{S}$. Reorder the two-dimensional indices $\{(i,j) \in \mathcal{S}\}$ into $\{(i_k, j_k) : 1 \leqslant k \leqslant s := \text{card}(\mathcal{S})\}$. Let $\theta_k = \theta_{i_kj_k}$, $k = 1, \cdots, s$, and define

$$U_{lk} = \begin{cases} \dfrac{1}{n_1}\left(X_{l,i_k}X_{l,j_k} - \sigma_{i_kj_k}\right) & \text{for } 1 \leqslant l \leqslant n_1 \\ -\dfrac{1}{n_2}\left(Y_{l,i_k}Y_{l,j_k} - \sigma_{i_kj_k}\right) & \text{for } n_1 + 1 \leqslant l \leqslant n_1 + n_2 \end{cases}$$

$$\text{and} \qquad V_{lk} = \frac{U_{lk}}{\sqrt{(n_1^{-1} + n_2^{-1})\theta_k}}.$$

To control the tail behavior, we further define

$$\widehat{U}_{lk} = U_{lk}I\left\{U_{lk} \leqslant \tau_n\right\} - EU_{lk}I\left\{U_{lk} \leqslant \tau_n\right\},$$

where $\tau_n = C\log(p+n)/n$, and correspondingly, $\widehat{V}_{lk} = \widehat{U}_{lk}/\sqrt{(n_1^{-1} + n_2^{-1})\theta_k}$. For ease of notation, we let $n := n_1 + n_2$, and define $V_k = \sum_{l=1}^n V_{lk}$ and $\widehat{V}_k = \sum_{l=1}^n \widehat{V}_{lk}$. Following the same arguments as in Cai et al. (2013), it suffices to prove that for any $x, t \in \mathbb{R}$, as $n, p \to \infty$,

$$P\left(\max_{1\leqslant k\leqslant s} \widehat{V}_k^2 \geqslant y_p,\ \frac{\widetilde{T}_{n_1,n_2}}{2\left(n_1^{-1} + n_2^{-1}\right)\text{tr}\left(\mathbf{\Sigma}^2\right)} \leqslant t\right) \to (1 - G(x))\,\Phi(t), \qquad (4.8.31)$$

where $y_p = x + 4\log p - \log\log p$ and $G(x) = \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right)$.
According to Bonferroni inequality, for any fixed integer m satisfying $0 < m < s/2$,

$$\begin{aligned}
&\sum_{d=1}^{2m}(-1)^{d-1}\sum_{1\leqslant k_1<\cdots<k_d\leqslant s} P\left(\bigcap_{j=1}^{d}\left\{\widehat{V}_{k_j}^2 \geqslant y_p\right\}\bigcap\left\{\frac{\widetilde{T}_{n_1,n_2}}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)}\leqslant t\right\}\right)\\
\leqslant& P\left(\max_{1\leqslant k\leqslant s}\widehat{V}_k^2 \geqslant y_p,\ \frac{\widetilde{T}_{n_1,n_2}}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)}\leqslant t\right)\\
\leqslant& \sum_{d=1}^{2m-1}(-1)^{d-1}\sum_{1\leqslant k_1<\cdots<k_d\leqslant s} P\left(\bigcap_{j=1}^{d}\left\{\widehat{V}_{k_j}^2 \geqslant y_p\right\}\bigcap\left\{\frac{\widetilde{T}_{n_1,n_2}}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)}\leqslant t\right\}\right).
\end{aligned} \tag{4.8.32}$$

We first remove all the variables that are related with $(k_1, \cdots, k_d)$ from $\widetilde{T}_{n_1,n_2}$. Suppose $k_1, \cdots, k_d$ corresponds to the pairs $(i_{k_1}, j_{k_1}), \cdots, (i_{k_d}, j_{k_d})$. Partition $\mathbf{X}_{p\times 1}$ and $\mathbf{Y}_{p\times 1}$ into $\mathbf{X}_{p\times 1} = (\mathbf{X}^{(1)'}, \mathbf{X}^{(2)'})'$ and $\mathbf{Y}_{p\times 1} = (\mathbf{Y}^{(1)'}, \mathbf{Y}^{(2)'})'$ such that $\mathbf{X}^{(2)}$ and $\mathbf{Y}^{(2)}$ contain all the distinct elements in $\{i_{k_1}, j_{k_1}, \cdots, i_{k_d}, j_{k_d}\}$. We use $T^{*(k_1\cdots k_d)}_{n_1,n_2}$ to represent the $T^*_{n_1,n_2}$ statistic defined in (4.4.3). What's more, $T^{*(k_1\cdots k_d)}_{n_1,n_2}$ is independent from $\bigcap_{j=1}^{d}\left\{\widehat{V}_{k_j}^2 \geqslant y_p\right\}$ because of the decorrelation procedure by partitioning. Take $\beta = 1/4$ in (4.4.5), and let $\epsilon = \epsilon_p$ satisfy $\epsilon_p(\log p)^{\frac{1}{4}} \to \infty$. Since $\log p = o(n^{1/5})$, then by Lemma 4.2, there exists $\delta > 0$, such that

$$P\left(\frac{|\widetilde{T}_{n_1,n_2} - T^{*(k_1\cdots k_d)}_{n_1,n_2}|}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)} > \epsilon_p\right) \leqslant C\exp(-(\log p)^{1+\delta}). \tag{4.8.33}$$

It's easy to verity that for any $\epsilon_p > 0$ and random variables A and B,

$$P(A \leqslant t - \epsilon_p) - P(|B| \geqslant \epsilon_p) \leqslant P(A + B \leqslant t) \leqslant P(A \leqslant t + \epsilon_p) + P(|B| \geqslant \epsilon_p).$$

Therefore,

$$\begin{aligned}
&\sum_{d=1}^{2m}(-1)^{d-1}\sum_{1\leqslant k_1<\cdots<k_d\leqslant s} P\left(\bigcap_{j=1}^{d}\left\{\widehat{V}_{k_j}^2 \geqslant y_p\right\}\bigcap\left\{\frac{\widetilde{T}_{n_1,n_2}}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)}\leqslant t\right\}\right)\\
\geqslant& \sum_{d=1}^{2m}(-1)^{d-1}\sum_{1\leqslant k_1<\cdots<k_d\leqslant s} P\left(\bigcap_{j=1}^{d}\left\{\widehat{V}_{k_j}^2 \geqslant y_p\right\}\bigcap\left\{\frac{T^{*(k_1\cdots k_d)}_{n_1,n_2}}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)}\leqslant t-(-1)^{d-1}\epsilon_p\right\}\right)\\
&-\sum_{d=1}^{2m}\sum_{1\leqslant k_1<\cdots<k_d\leqslant s} P\left(\bigcap_{j=1}^{d}\left\{\widehat{V}_{k_j}^2 \geqslant y_p\right\}\bigcap\frac{|\widetilde{T}_{n_1,n_2} - T^{*(k_1\cdots k_d)}_{n_1,n_2}|}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\mathbf{\Sigma}^2\right)}\geqslant \epsilon_p\right)
\end{aligned}$$

$$\geqslant \sum_{d=1}^{2m} (-1)^{d-1} \sum_{1\leqslant k_1<\cdots<k_d\leqslant s} P\left(\bigcap_{j=1}^{d}\left\{\widehat{V}_{k_j}^2 \geqslant y_p\right\}\right) P\left(\frac{T_{n_1,n_2}^{*(k_1\cdots k_d)}}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \leqslant t-(-1)^{d-1}\epsilon_p\right)$$
$$-\sum_{d=1}^{2m} \sum_{1\leqslant k_1<\cdots<k_d\leqslant s} P\left(\frac{|\widetilde{T}_{n_1,n_2}-T_{n_1,n_2}^{*(k_1\cdots k_d)}|}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \geqslant \epsilon_p\right). \tag{4.8.34}$$

Together with (4.8.33), the last term can be bounded in the following way.

$$\begin{aligned}
&\sum_{d=1}^{2m} \sum_{1\leqslant k_1<\cdots<k_d\leqslant s} P\left(\frac{|\widetilde{T}_{n_1,n_2}-T_{n_1,n_2}^{*(k_1\cdots k_d)}|}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \geqslant \epsilon_p\right)\\
\leqslant&\sum_{d=1}^{2m} \sum_{1\leqslant k_1<\cdots<k_d\leqslant s} C\exp(-(\log p)^{1+\delta}) \leqslant Cs^{2m}\cdot\exp(-(\log p)^{1+\delta})\\
\leqslant& C\exp\left(2m\log p+2m\log(p-1)-(\log p)^{1+\delta}\right)=o(1). \qquad (4.8.35)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\sum_{d=1}^{2m-1} (-1)^{d-1} \sum_{1\leqslant k_1<\cdots<k_d\leqslant s} P\left(\bigcap_{j=1}^{d}\left\{\widehat{V}_{k_j}^2 \geqslant y_p\right\} \bigcap \left\{\frac{\widetilde{T}_{n_1,n_2}}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \leqslant t\right\}\right)\\
\leqslant& \sum_{d=1}^{2m-1} (-1)^{d-1} \sum_{1\leqslant k_1<\cdots<k_d\leqslant s} \left[P\left(\bigcap_{j=1}^{d}\left\{\widehat{V}_{k_j}^2 \geqslant y_p\right\}\right)\right.\\
&\left.P\left(\frac{T_{n_1,n_2}^{*(k_1\cdots k_d)}}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \leqslant t+(-1)^{d-1}\epsilon_p\right)\right]+o(1). \qquad (4.8.36)
\end{aligned}$$

Define $\mathbf{W}_l^{(d)} = (\widehat{V}_{lk_1},\cdots,\widehat{V}_{lk_d}) \in \mathbb{R}^d$, $l=1,\cdots,n$, then

$$P\left(\bigcap_{j=1}^{d}\left\{\widehat{V}_{k_j}^2 \geqslant y_p\right\}\right) = P\left(\left|\sum_{l=1}^{n}\mathbf{W}_l^{(d)}\right|_{\min} \geqslant \sqrt{y_p}\right),$$

where $|\mathbf{w}|_{\min} := \min_{1\leqslant i\leqslant d}|w_i|$ indicates the the minimum absolute value in a vector $\mathbf{w}\in\mathbb{R}^d$. By Theorem 1.1. in Zaitsev (1987), we have

$$\begin{aligned}
&\left|P\left(\left|\sum_{l=1}^{n}\mathbf{W}_l^{(d)}\right|_{\min} \geqslant \sqrt{y_p}\right) - P\left(|\mathbf{Z}_d|_{\min} \geqslant \sqrt{y_p}\pm\delta_p(\log p)^{-\frac{1}{2}}\right)\right|\\
\leqslant& c_1 d^{\frac{5}{2}}\exp\left(-\frac{\delta_p}{c_2 d^3\sqrt{n\log p}\tau_n}\right),
\end{aligned}$$

where $\mathbf{Z}_d$ is a d-dimensional normal vector with mean zero and covaraince $\mathrm{cov}(\mathbf{Z}_d)$ $= \mathrm{cov}(\sum_{l=1}^{n} \mathbf{W}_l^{(d)})$. Since d is a fixed integer and $\log p = o(n^{\frac{1}{5}})$, by letting $\delta_p \to 0$ sufficiently slow, we have

$$\sum_{d=1}^{2m} \sum_{1 \leqslant k_1 < \cdots < k_d \leqslant s} c_1 d^{\frac{5}{2}} \exp\left(-\frac{\delta_p}{c_2 d^3 \sqrt{n \log p} \tau_n}\right) = o(1). \tag{4.8.37}$$

In combination with (4.8.34) and (4.8.36), (4.8.32) becomes

$$\begin{aligned}
&\sum_{d=1}^{2m} (-1)^{d-1} \sum_{1 \leqslant k_1 < \cdots < k_d \leqslant s} \Bigg[P\left(|\mathbf{Z}_d|_{\min} \geqslant \sqrt{y_p} + (-1)^{d-1} \delta_p (\log p)^{-\frac{1}{2}} \right) \\
&\quad \times P\left(\frac{T_{n_1,n_2}^{*(k_1 \cdots k_d)}}{2\left(n_1^{-1} + n_2^{-1}\right) \operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \leqslant t - (-1)^{d-1} \epsilon_p \right) \Bigg] - o(1) \\
\leqslant &P\left(\max_{1 \leqslant k \leqslant s} \widehat{V}_k^2 \geqslant y_p,\ \frac{\widetilde{T}_{n_1,n_2}}{2\left(n_1^{-1} + n_2^{-1}\right) \operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \leqslant t \right) \\
\leqslant &\sum_{d=1}^{2m-1} (-1)^{d-1} \sum_{1 \leqslant k_1 < \cdots < k_d \leqslant s} \Bigg[P\left(|\mathbf{Z}_d|_{\min} \geqslant \sqrt{y_p} - (-1)^{d-1} \delta_p (\log p)^{-\frac{1}{2}} \right) \\
&\quad \times P\left(\frac{T_{n_1,n_2}^{*(k_1 \cdots k_d)}}{2\left(n_1^{-1} + n_2^{-1}\right) \operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \leqslant t + (-1)^{d-1} \epsilon_p \right) \Bigg] + o(1).
\end{aligned} \tag{4.8.38}$$

According to Lemmas 2 and 3 in Chen and Qin (2010) and using the Berry-Esseen bound for martingales (Haeusler, 1988), we have

$$\left| P\left(\left\{ \frac{T_{n_1,n_2}^{*(k_1 \cdots k_d)}}{2\left(n_1^{-1} + n_2^{-1}\right) \operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \leqslant t \pm \epsilon_p \right\} \right) - \Phi(t \pm \epsilon_p) \right| \leqslant C \left(\frac{\operatorname{tr}(\boldsymbol{\Sigma}^4)}{\operatorname{tr}^2(\boldsymbol{\Sigma}^2)} + \frac{1}{n} \right)^{1/5}.$$

Combine with Lemma 5 in Cai et al. (2013), for $x, t \in \mathbb{R}$ and any fixed integer $d \geqslant 1$,

$$\begin{aligned}
&\sum_{1 \leqslant k_1 < \cdots < k_d \leqslant s} P\left(|\mathbf{Z}_d|_{\min} \geqslant \sqrt{y_p} + (-1)^{d-1} \delta_p (\log p)^{-\frac{1}{2}} \right) \times \\
&P\left(\frac{T_{n_1,n_2}^{*(k_1 \cdots k_d)}}{2\left(n_1^{-1} + n_2^{-1}\right) \operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} \leqslant t - (-1)^{d-1} \epsilon_p \right) \\
&= \frac{1}{d!} \left(\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{x}{2}\right) \right)^d \Phi(t)(1 + o(1)).
\end{aligned}$$

Therefore, we prove (4.8.31). What's more, Li and Chen (2012) and Cai et al. (2013) prove

$$\left| M_{n_1,n_2} - \max_{1\leqslant k\leqslant s} \widehat{V}_k^2 \right| \to 0 \quad \text{and} \quad \left| \frac{\widetilde{T}_{n_1,n_2}}{2\left(n_1^{-1}+n_2^{-1}\right)\operatorname{tr}\left(\boldsymbol{\Sigma}^2\right)} - \frac{\widetilde{T}_{n_1,n_2}}{\widehat{\sigma}_{0,n_1,n_2}} \right| \to 0.$$

By Slutsky's theorem, we reach the conclusion (4.4.1). □

4.8.5 Proof of Lemma 4.4

Proof. Following the main idea from Bellec (2019), we start with considering

$$\begin{aligned} E\exp\left(\lambda\boldsymbol{\xi}_1'\mathbf{A}\boldsymbol{\xi}_2\right) &= E\exp\left\{\lambda\sum_{i=1}^{m}\sum_{j=1}^{r}a_{ij}\xi_{1i}\xi_{2j}\right\} = E\exp\left\{\sum_{i=1}^{m}\xi_{1i}\left(\lambda\sum_{j=1}^{r}a_{ij}\xi_{2j}\right)\right\} \\ &\leqslant E\exp\left\{\lambda^2K^2\sum_{i=1}^{m}\left(\sum_{j=1}^{r}a_{ij}\xi_{2j}\right)^2\right\} = E\exp\left\{\lambda^2K^2\|\mathbf{A}\boldsymbol{\xi}_2\|_2^2\right\}, \end{aligned} \tag{4.8.39}$$

where λ satisfies $0<\lambda\leqslant 1/(16K^2\lambda_1(\mathbf{A}^T\mathbf{A})^{\frac{1}{2}})$.

Let $\mathbf{B}=\mathbf{A}'\mathbf{A}=(b_{ij})_{1\leqslant i,j\leqslant r}$ and let $\mathbf{B}_0$ be the matrix $\mathbf{B}$ with diagonal entries set to 0, i.e. $\mathbf{B}_0=\mathbf{B}-\operatorname{diag}(b_{11},\cdots,b_{rr})$. Then for any $i=1,\cdots,r$,

$$0\leqslant b_{ii}=\sum_{j=1}^{m}a_{ji}^2\leqslant\lambda_1\left(\mathbf{A}'\mathbf{A}\right).$$

Using Cauchy-Schwarz inequality,

$$(4.8.39)\leqslant\sqrt{E\exp\left\{2\lambda^2K^2\sum_{i=1}^{r}b_{ii}\xi_{2i}^2\right\}}\cdot\sqrt{E\exp\left\{2\lambda^2K^2\sum_{i\neq j}b_{ij}\xi_{2i}\xi_{2j}\right\}}. \tag{4.8.40}$$

By decoupling inequality (Vershynin, 2011),

$$E\exp\left\{2\lambda^2K^2\sum_{i\neq j}b_{ij}\xi_{2i}\xi_{2j}\right\}\leqslant E\exp\left\{8\lambda^2K^2\sum_{i\neq j}b_{ij}\widetilde{\xi}_{2i}\xi_{2j}\right\}, \tag{4.8.41}$$

in which $\widetilde{\xi}_{2i}$ is an independent copy of ξ_{2i}. Following the same spirit as (4.8.39),

$$(4.8.41) \leqslant E\exp\left\{64\lambda^4K^6\sum_{i=1}^{r}\left(\sum_{j\neq i}b_{ij}\xi_{2j}\right)^2\right\} = E\exp\left\{64\lambda^4K^6\|\mathbf{B}_0\boldsymbol{\xi}_2\|_2^2\right\}. \tag{4.8.42}$$

Note that

$$\|\mathbf{B}_0\boldsymbol{\xi}_2\|_2^2 \leqslant 2\|\mathbf{B}\boldsymbol{\xi}_2\|_2^2 + 2\sum_{i=1}^{r}b_{ii}^2\xi_{2i}^2 \leqslant 2\lambda_1(\mathbf{A}'\mathbf{A})\|\mathbf{A}\boldsymbol{\xi}_2\|_2^2 + 2\lambda_1(\mathbf{A}'\mathbf{A})\sum_{i=1}^{r}b_{ii}\xi_{2i}^2.$$

Then (4.8.42) becomes

$$\begin{aligned}
&E\exp\left\{8\lambda^2K^2\sum_{i\neq j}b_{ij}\widetilde{\xi}_{2i}\xi_{2j}\right\} \leqslant E\exp\left\{64\lambda^4K^6\|\mathbf{B}_0\boldsymbol{\xi}_2\|_2^2\right\}\\
\leqslant &E\exp\left\{256\lambda^2K^4\lambda_1(\mathbf{A}'\mathbf{A})\left(\frac{1}{2}\lambda^2K^2\|\mathbf{A}\boldsymbol{\xi}_2\|_2^2 + \frac{1}{2}\lambda^2K^2\sum_{i=1}^{r}b_{ii}\xi_{2i}^2\right)\right\}\\
&\leqslant E\exp\left\{\frac{1}{2}\lambda^2K^2\|\mathbf{A}\boldsymbol{\xi}_2\|_2^2 + \frac{1}{2}\lambda^2K^2\sum_{i=1}^{r}b_{ii}\xi_{2i}^2\right\}\\
&\leqslant \sqrt{E\exp\left\{\lambda^2K^2\|\mathbf{A}\boldsymbol{\xi}_2\|_2^2\right\}}\cdot\sqrt{E\exp\left\{\lambda^2K^2\sum_{i=1}^{r}b_{ii}\xi_{2i}^2\right\}}.
\end{aligned} \tag{4.8.43}$$

Plug (4.8.43) into (4.8.40),

$$\begin{aligned}
&\left(E\exp\left\{\lambda^2K^2\|A\boldsymbol{\xi}_2\|_2^2\right\}\right)^2 \leqslant E\exp\left\{2\lambda^2K^2\sum_{i=1}^{r}b_{ii}\xi_{2i}^2\right\}\cdot E\exp\left\{2\lambda^2K^2\sum_{i\neq j}b_{ij}\xi_{2i}\xi_{2j}\right\}\\
\leqslant &E\exp\left\{2\lambda^2K^2\sum_{i=1}^{n}b_{ii}\xi_{2i}^2\right\}\cdot\sqrt{E\exp\left\{\lambda^2K^2\|A\boldsymbol{\xi}_2\|_2^2\right\}}\cdot\sqrt{E\exp\left\{\lambda^2K^2\sum_{i=1}^{n}b_{ii}\xi_{2i}^2\right\}}.
\end{aligned}$$

Together with (4.8.39) and Proposition 4 in Bellec (2019),

$$\begin{aligned}
E\exp\left\{\lambda\boldsymbol{\xi}_1'\mathbf{A}\boldsymbol{\xi}_2\right\} &\leqslant E\exp\left\{\lambda^2K^2\|\mathbf{A}\boldsymbol{\xi}_2\|_2^2\right\} \leqslant E\exp\left\{2\lambda^2K^2\sum_{i=1}^{r}b_{ii}\xi_{2i}^2\right\}\\
\leqslant \exp\left\{\frac{3}{2}\cdot 2\lambda^2K^2\sum_{i=1}^{r}b_{ii}\right\} &= \exp\left\{3\lambda^2K^2\mathrm{tr}(\mathbf{A}'\mathbf{A})\right\}.
\end{aligned} \tag{4.8.44}$$

Then we are able to obtain

$$P\left(\boldsymbol{\xi}_1'\mathbf{A}\boldsymbol{\xi}_2 > t\right) \leqslant \exp\left\{-\lambda t + 3\lambda^2 K^2 \mathrm{tr}(\mathbf{A}'\mathbf{A})\right\},\ 0 < \lambda \leqslant \frac{1}{16K^2\lambda_1(\mathbf{A}'\mathbf{A})^{\frac{1}{2}}}. \tag{4.8.45}$$

Without the constrain on λ, we can choose λ to be $\widetilde{\lambda} = t/(6K^2\mathrm{tr}(\mathbf{A}^T\mathbf{A}))$, which minimizes the right-hand side of (4.8.45). Therefore,

$$P\left(\boldsymbol{\xi}_1'\mathbf{A}\boldsymbol{\xi}_2 > t\right) \leqslant \exp\left\{-\frac{t^2}{12K^2\mathrm{tr}(\mathbf{A}'\mathbf{A})}\right\}.$$

If $\widetilde{\lambda} > \left(16K^2\sqrt{\lambda_1(\mathbf{A}'\mathbf{A})}\right)^{-1}$, then consider λ to be $\lambda_a = \left(16K^2\sqrt{\lambda_1(\mathbf{A}'\mathbf{A})}\right)^{-1}$. Note that

$$-t\lambda_a + 3\lambda_a^2K^2\mathrm{tr}(\mathbf{A}'\mathbf{A}) \leqslant -t\lambda_a + 3\lambda_a\widetilde{\lambda}\mathrm{tr}(\mathbf{A}'\mathbf{A})K^2 = -t\lambda_a + \lambda_a\frac{t}{2} \leqslant -\frac{t}{32K^2\sqrt{\lambda_1(\mathbf{A}'\mathbf{A})}}.$$

The proof of (4.8.2) is completed.

For Remark 4.9, As a special case of Lemma 4.4, we consider the scenario when $\xi_{ij} \sim N(0,1)$. In order for condition (4.8.1) to hold, we need to choose K^2 such that for $k \geqslant 2$,

$$K^2 \geqslant \left(\frac{2}{k!}E|\xi_{ij}|^{2k}\right)^{1/(k-1)} = \left(\frac{1}{2^{k-1}}\cdot\frac{(2k)!}{(k!)^2}\right)^{1/(k-1)} := f(k)$$

By Stirling's formula, we have

$$f(k) \leqslant 2\left(\frac{2e}{\pi}\right)^{\frac{1}{k-1}} \leqslant 2\sqrt{\frac{2e}{\pi}} < 3 \quad \text{for } k \geqslant 3. \tag{4.8.46}$$

Also note that $f(2) = 3$. Therefore, if $\xi_{ij} \sim N(0,1)$, we can choose $K^2 = 3$. However, note that for standard normal variable ξ_{ij}, it's easy to verify that

(i) For any $\lambda \in \mathbb{R}$, $E\exp\{\lambda\xi_{ij}\} = \exp\{\lambda^2/2\}$.

(ii) For any $0 < \lambda < 1/4$, $E\exp\{\lambda\xi_{ij}^2\} \leqslant \exp\{3\lambda/2\}$.

Thus, in the proof of Lemma 4.4, (4.8.39) and (4.8.42) hold with $K^2 = 1/2$, and (4.8.44) remains valid. In the meantime, the remaining parts of the proof stay the same. As a conclusion, for standard normal variable ξ_{ij}, (4.8.2) holds with $K^2 = 1/2$, even though condition (4.8.1) is not satisfied at this point. □

4.8.6 Proof of Theorem 4.1

Proof. This is a direct result of applying Slutsky's theorem to Lemma 4.3. □

4.8.7 Proof of Theorem 4.2

Proof. The accurate asymptotic size directly follows the asymptotic independence of M_{n_1,n_2} and T_{n_1,n_2} in Theorem 4.1. The proof of Theorem 4.2 is complete. □

4.8.8 Proof of Theorem 4.3

Proof. As for the consistent asymptotic power, note that

$$\begin{aligned}&\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_d\cup\mathcal{G}_s} P\left(F_{n_1,n_2}\geqslant c_\alpha\right)\\ \geqslant&\min\left\{\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_d} P\left(F_{n_1,n_2}\geqslant c_\alpha\right),\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_s} P\left(F_{n_1,n_2}\geqslant c_\alpha\right)\right\}.\end{aligned}$$

It suffices to show

$$\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_k} P\left(F_{n_1,n_2}\geqslant c_\alpha\right)\to 1\quad\text{for } k=d,s. \tag{4.8.47}$$

Since both $-2\log\left(1-G(M_{n_1,n_2}-4\log p+\log\log p)\right)$ and $-2\log\left(1-\Phi\left(T_{n_1,n_2}/\widehat{\sigma}_{0,n_1,n_2}\right)\right)$ are always non-negative, we have

$$\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_d} P\left(F_{n_1,n_2}\geqslant c_\alpha\right)\geqslant\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_d} P\left(\frac{T_{n_1,n_2}}{\widehat{\sigma}_{0,n_1,n_2}}\geqslant\Phi^{-1}\left(1-\exp(-c_\alpha/2)\right)\right)$$

and

$$\begin{aligned}&\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_s} P\left(F_{n_1,n_2}\geqslant c_\alpha\right)\\ \geqslant&\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_s} P\left(M_{n_1,n_2}-4\log p+\log\log p\geqslant G^{-1}\left(1-\exp(-c_\alpha/2)\right)\right).\end{aligned}$$

Together with discussions from Li and Chen (2012) and Theorem 2 in Cai et al. (2013), (5.7.33) is proved. The proof of Theorem 4.3 is complete. □

Chapter 5

Power-Enhanced Simultaneous Test of High-Dimensional Mean Vectors and Covariance Matrices

5.1 Introduction

Inferences on the equality of two distributions is of significant interest in a wide range of real applications. Genetic studies use the differential gene expression analysis to understand how genes are related to diseases (Wang et al., 2015). Medical image analysis examines the differential structure of image to diagnose abnormal tissues (Ginestet et al., 2017). Pharmaceutical researchers rely on the analysis of comparative clinical trial outcomes for drug discovery and development (Cummings et al., 2019).

To make inferences on the discrepancies between two distributions, researchers usually consider their mean vectors and covariance matrices that characterize commonly used distributions, for example, the elliptical distributions (Anderson, 2003). Over the past decade, there has been significant progress in testing the equality of two mean vectors (Chen and Qin, 2010; Cai et al., 2014; Wang et al., 2015) or covariance matrices (Li and Chen, 2012; Cai et al., 2013; Zhu et al., 2017; Chang et al., 2017) under the high-dimensional setting. Yet few works are capable of examining both mean vectors and covariance matrices simultaneously.

However, in practice, we often do not know whether the discrepancies reside in mean vectors or in covariance structure. It has been recognized that mean tests are powerful to detect the differences in mean vectors but cannot detect the different covariance structure. In contrast, covariance tests are powerful to identify the differences in covariance structure but are incompetent to distinguish the differential structure of two mean vectors. Thus, it is crucial to develop a new simultaneous testing procedure that is powerful to detect differences in either mean vectors or covariance matrices.

Let $\mathbf{X}$ and $\mathbf{Y}$ be two p-dimensional populations with mean vectors $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ and covariance matrices $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$, respectively. We consider the simultaneous test on the equality of mean vectors and covariance matrices of the two populations, i.e.,

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{and} \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2. \tag{5.1.1}$$

In real-world applications such as genetic studies, the sample size is often less than a hundred, but the number of features can be thousands or even larger (Clarke et al., 2008). Throughout this chapter, we assume that the dimension p is much larger than the sample size n_1 or n_2. The challenge of high dimensionality leads to fundamental difficulties in understanding the asymptotic behavior of test statistics.

In the literature, there only exist a few works on jointly testing means and covariances. In the classical setting with a fixed dimension p, the likelihood ratio test (LRT) was extensively studied in the multivariate analysis (Anderson, 2003) when the samples come from normal distributions. When p diverges proportionally as the sample size tends to infinity such that $p/\min\{n_1, n_2\} \to c$ for some $0 < c \leqslant 1$, Jiang and Yang (2013) studied the modified LRTs under the normal assumption and derived central limit theorems. The normal assumption was recently relaxed by Niu et al. (2019). To allow p to diverge at a comparable rate as the sample size tends to infinity, i.e., $0 < c < \infty$, Liu et al. (2017) proposed a new approach by replacing the entropy loss with the quadratic loss for covariance matrix estimation. Hyodo and Nishiyama (2018) proposed a new joint test using a weighted sum of multiple U-statistics to allow p to diverge faster than the sample size.

These existing testing procedures are essentially based on a weighted sum of one test statistic related to the mean difference and another test statistic regarding the covariance difference. The weighted sum is generally not an optimal combination due to potentially different scales of two test statistics. These tests could be driven by the test statistic of a larger scale, leading to undesired power loss in the corresponding alternative space (Xie et al., 2011). These joint testing procedures suffer from the power loss in detecting sparse signals as the errors in estimating high-dimensional parameters accumulate in quadratic-form test statistics (Fan et al., 2015), as they are based on the modified LRTs or the L_2-norm-based test. Moreover, most of these tests only allow for a moderate-high dimension in the asymptotic regime such that the dimension diverges at a slower rate than the sample size.

This work aims to develop a new power-enhanced simultaneous testing procedure, that is powerful to detect differences in either mean vectors or covariance structure against either sparse alternatives or dense alternatives under an ultra-high dimensional setting. Fan et al. (2015) introduced the power enhancement framework for high-dimensional hypothesis testing, which consists of the following *power enhancement (PE) principles*: (a) no size distortion; (b) the power-enhance test is at least at powerful as the original test; (c) the power is substantially enhanced under a more general alternative. In this work, we interpret the more general alternatives from the following two perspectives:

(i) expanding the high-power region of mean tests or covariance tests to a wider alternative space respectively. We aim to develop the power-enhanced tests against the union of their corresponding dense and sparse alternatives.

(ii) extending the test capability to alternative spaces with respect to both mean and covariance parameters. We aim to combine both strengths of two power-enhanced tests and develop a joint test that is capable of detecting the difference from either mean vectors or covariance matrices.

In this work, we first construct power-enhanced tests for mean vectors and covariance matrices separately to boost their respective power. The test statistics of Chen and Qin (2010) and Li and Chen (2012) are constructed based on estimators of the squared Euclidean distance of two sample means and the squared Frobenius distance of two covariance matrices, respectively. It is known that they are powerful to detect dense signals but unable to detect sparse signals (Chen, Li and

Zhong, 2019). We introduce their respective PE components to effectively enlarge the high-power regions to the union of sparse and dense alternatives. We show that the proposed power-enhanced tests satisfy three PE principles. It is worth pointing out that we need to deal with a more significant challenge than Fan et al. (2015), since the distributions of test statistics in Chen and Qin (2010) and Li and Chen (2012) are no longer symmetric under the null hypothesis.

After developing two individual power-enhanced tests, we need to combine their strengths to develop the power-enhanced simultaneous test to further enhance the test capability to jointly testing mean vectors and covariance matrices. We study the asymptotic joint distribution of two statistics, and then aggregate information from the two aspects via the combination of their respective p-values using Fisher's method (Fisher, 1925). We also show that the proposed power-enhanced simultaneous test satisfies three PE principles. Unlike Fan et al. (2015) and Yu et al. (2020), we do not require the stringent normal assumption when deriving the asymptotic independence result. Benefiting from the p-value combination, our proposed test is scale-invariant and computationally efficient.

We study the theoretical properties under an ultra-high dimensional setting where the dimension may grow at a nearly exponential rate of the sample size. Moreover, we conduct simulation studies to compare the proposed test's numerical performances against several benchmark tests under various alternatives. In a real application, we further demonstrate the power of the proposed test to find differentially expressed gene-sets using an acute lymphoblastic leukemia dataset. Our findings are supported by the biological literature.

The rest of this chapter is organized as follows. Section 5.2 introduces preliminaries on high-dimensional mean tests and covariance tests. Section 5.3 presents complete methodological details. Theoretical properties, including the power enhancement properties, the asymptotic size and power analysis as well as the asymptotic optimality, are also established in this section. Section 5.4 conducts simulation studies to demonstrate the finite-sample properties under different alternative hypotheses. Section 5.5 presents an empirical study on identifying differentially expressed gene-sets among various types of cancers. Section 5.6 includes a few concluding remarks. All technical details are presented in Section 5.7.

5.2 Preliminaries

Let $\mathbf{X}$ be a p-dimensional random vector with mean $\boldsymbol{\mu}_1 = (\mu_{11}, \cdots, \mu_{1p})'$ and covariance $\boldsymbol{\Sigma}_1 = (\sigma_{1,ij})_{p\times p}$, and $\mathbf{Y}$ be a p-dimensional random vector with mean $\boldsymbol{\mu}_2 = (\mu_{21}, \cdots, \mu_{2p})'$ and covariance $\boldsymbol{\Sigma}_2 = (\sigma_{2,ij})_{p\times p}$. Suppose that $\{\mathbf{X}_1, \cdots, \mathbf{X}_{n_1}\}$ are independently and identically distributed (i.i.d.) copies of $\mathbf{X}$, and $\{\mathbf{Y}_1, \cdots, \mathbf{Y}_{n_2}\}$ are i.i.d. copies of $\mathbf{Y}$ that are independent of $\{\mathbf{X}_1, \cdots, \mathbf{X}_{n_1}\}$. Now, we consider the high-dimensional mean test

$$H_{0m} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, \tag{5.2.1}$$

and the high-dimensional covariance test

$$H_{0c} : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2, \tag{5.2.2}$$

respectively. Chen and Qin (2010) proposed the following quadratic-form statistic M_{n_1,n_2} for testing if the two high-dimensional populations share the same mean vector in (5.2.1):

$$M_{n_1,n_2} = \frac{1}{n_1(n_1-1)} \sum_{u\neq v}^{n_1} (\mathbf{X}_u'\mathbf{X}_v) + \frac{1}{n_2(n_2-1)} \sum_{u\neq v}^{n_2} (\mathbf{Y}_u'\mathbf{Y}_v) - \frac{2}{n_1 n_2} \sum_{u}^{n_1}\sum_{v}^{n_2} (\mathbf{X}_u'\mathbf{Y}_v) . \tag{5.2.3}$$

Under the null hypothesis H_{0m}, Chen and Qin (2010) considered the standardized test statistic $M_{n_1,n_2}/\widehat{\sigma}_{01}$ and proved that,

$$\text{under } H_{0m} : \quad \frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}} \xrightarrow{d} N(0,1) \quad \text{as } n_1, n_2, p \to \infty, \tag{5.2.4}$$

where $\widehat{\sigma}_{01}$ is a consistent estimator of $\sigma_{01} = (\frac{2}{n_1(n_1-1)}\text{tr}(\boldsymbol{\Sigma}_1^2) + \frac{2}{n_2(n_2-1)}\text{tr}(\boldsymbol{\Sigma}_2^2) + \frac{4}{n_1 n_2}\text{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2))^{\frac{1}{2}}$, which is the standard deviation of M_{n_1,n_2} under H_{0m}. The test rejects H_{0m} with significance level α if $M_{n_1,n_2} \geqslant \widehat{\sigma}_{01} z_\alpha$, where z_α is the upper α-quantile of standard normal distribution.

To test the equality of two covariance matrices in (5.2.2), Li and Chen (2012) constructed their test statistic based on the squared Frobenius norm of $\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2$. Since $\|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|_F^2 = \text{tr}\left((\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)^2\right) = \text{tr}(\boldsymbol{\Sigma}_1^2) + \text{tr}(\boldsymbol{\Sigma}_2^2) - 2\text{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)$, they proposed a test statistic T_{n_1,n_2} in the form of linear combination of unbiased estimators for each term, that is,

$$T_{n_1,n_2} = A_{n_1} + B_{n_2} - 2C_{n_1,n_2}. \tag{5.2.5}$$

where A_{n_1}, B_{n_2} and C_{n_1,n_2} are unbiased estimators for $\mathrm{tr}(\mathbf{\Sigma}_1^2)$, $\mathrm{tr}(\mathbf{\Sigma}_2^2)$ and $\mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_2)$, respectively. Under H_{0c}, the leading variance of T_{n_1,n_2} is $\sigma_{02}^2 = 4(\frac{1}{n_1} + \frac{1}{n_2})^2\mathrm{tr}^2\left(\mathbf{\Sigma}^2\right)$. With $\widehat{\sigma}_{02}$ being a consistent estimator of σ_{02}, they conducted the test for H_{0c} on the basis of the test statistic $T_{n_1,n_2}/\widehat{\sigma}_{02}$ and proved that,

$$\text{under } H_{0c}: \quad \frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}} \xrightarrow{d} N(0,1) \quad \text{as } n_1, n_2, p \to \infty, \tag{5.2.6}$$

The test rejects H_{0c} with a nominal significance level α if $T_{n_1,n_2} \geqslant \widehat{\sigma}_{02} z_\alpha$.

These two test statistics (5.2.3) and (5.2.5) are cornerstones of our proposed simultaneous test on the equality of two-sample mean vectors and covariance matrices. Note that the null space under H_0 is, in fact, a subspace of null spaces under two separate null hypotheses H_{0m} and H_{0c}. Hence, even though their limiting null distributions (5.2.4) and (5.2.6) are derived under H_{0m} and H_{0c}, they remain valid under H_0 with the same assumptions.

5.3 Methodology

This section presents our proposed power-enhanced simultaneous test on jointly testing means and covariances in high dimensions. We elucidate the motivation of our approach along the way. In Section 5.3.1, we propose power-enhanced tests for the mean test and the covariance test respectively to boost their respective power. Anchored in the two power-enhanced statistics, in Section 5.3.2 we study their asymptotic joint distribution, and subsequently, in Section 5.3.3 we introduce our simultaneous test to expand the test capability for jointly testing high-dimensional mean vectors and covariance matrices.

5.3.1 Power-Enhanced Tests

Both M_{n_1,n_2} and T_{n_1,n_2} are quadratic-form statistics. It has been known that such type of statistics suffer from low power against sparse alternatives where the parameter of interest differs only in a small proportion of coordinates. One predominant approach to achieve high testing power against sparse alternatives is to utilize extreme values to construct test statistics (Cai et al., 2013; Chang et al., 2017; Chernozhukov et al., 2019), whereas another way continues with the quadratic-form statistics but rules out non-signal bearing dimensions via thresholding (Fan, 1996; Chen, Li and Zhong, 2019; Chen, Guo and Qiu, 2019). However, these tests generally require either stringent conditions or bootstrap to derive the limiting null distribution and are likely to suffer from size distortions due to slow convergence. Also, even though the extreme value tests and thresholding tests retain high power against sparse alternatives, they tend to lack the ability to detect dense and faint signals, in which circumstances the quadratic-form tests are favored.

To deal with the challenge mentioned above, we first explore power enhancement for testing high-dimensional mean vectors and covariance matrices based on $M_{n_1 n_2}$ and T_{n_1,n_2}, respectively. Fan et al. (2015) provides a helpful insight for us to enhance testing power against sparse alternatives and preserve the merits of existing quadratic-form tests at the same time. We construct two PE components J_m and J_c, which are designed to take zero values under the null hypothesis but diverge quickly under sparse alternatives. The PE components are designed delicately following the guidance of the three PE principles. By adding the PE components to the original statistics, the resultant tests $\widehat{\sigma}_{01}^{-1} M_{n_1,n_2} + J_m$ and $\widehat{\sigma}_{02}^{-1} T_{n_1,n_2} + J_c$ acquire substantially enhanced power under sparse alternatives with little size distortion under the null hypothesis.

Different from Fan et al. (2015), the distributions of our test statistics are no longer symmetric under H_0, causing difficulties in designing the PE components. After careful investigation, we prove that the marginal standardized statistics follow chi-squared distributions. Let $n = n_1 + n_2$ and δ_p and η_p be the thresholds chosen for the mean statistics and covariances statistics, respectively. We choose J_m and J_c to be the sum of marginal standardized statistics whose values exceed δ_p and η_p. By construction, the screening procedure rules out all the noises under the null hypothesis. Still, it makes it capable of capturing non-zero signals under sparse alternatives, implying that J_m and J_c equal to zero under the null hypothesis but diverge quickly under the sparse alternatives.

5.3.1.1 Power-Enhanced Mean Tests

We use $\mathbf{X} = (X_1, \cdots, X_p)'$ and $\mathbf{Y} = (Y_1, \cdots, Y_p)'$ to denote the random vectors of interest. Let $\mathbf{X}_u = (X_{u1}, \cdots, X_{up})'$ and $\mathbf{Y}_v = (Y_{v1}, \cdots, Y_{vp})'$ be the corresponding random samples. We rewrite the statistic M_{n_1,n_2} into $M_{n_1,n_2} = \sum_{i=1}^{p} M_i$, where

$$M_i = \frac{1}{n_1(n_1-1)} \sum_{u\neq v}^{n_1} (X_{ui}X_{vi}) + \frac{1}{n_2(n_2-1)} \sum_{u\neq v}^{n_2} (Y_{ui}Y_{vi}) - \frac{2}{n_1 n_2} \sum_{u}^{n_1}\sum_{v}^{n_2} (X_{ui}Y_{vi}) .$$

For each $i = 1, \cdots, p$, M_i consistently estimates $(\mu_{1i} - \mu_{2i})^2$ as $n_1, n_2 \to \infty$. Under the null hypothesis $H_{0m} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, the variance of M_i is

$$\nu_i := \frac{2}{n_1(n_1-1)}\sigma_{1,ii}^2 + \frac{2}{n_2(n_2-1)}\sigma_{2,ii}^2 + \frac{4}{n_1 n_2}\sigma_{1,ii}\sigma_{2,ii},$$

which can be consistently estimated by $\widehat{\nu}_i := \frac{2}{n_1(n_1-1)}\widehat{\sigma}_{1,ii}^2 + \frac{2}{n_2(n_2-1)}\widehat{\sigma}_{2,ii}^2$ $+\frac{4}{n_1 n_2}\widehat{\sigma}_{1,ii}\widehat{\sigma}_{2,ii}$, with $\widehat{\sigma}_{1,ii}$ and $\widehat{\sigma}_{2,ii}$ being sample variances of X_i and Y_i, respectively.

Define

$$J_m = \sqrt{p} \sum_{i=1}^{p} M_i \widehat{\nu}_i^{-1/2} \mathcal{I}\{\sqrt{2} M_i \widehat{\nu}_i^{-1/2} + 1 > \delta_p\} \tag{5.3.1}$$

with $\delta_p = 2\log p$ as the power enhancement component for the mean test. The theoretical analysis regarding J_m is established upon $\delta_p = 2\log p$. In practical implementations, we follow Fan et al. (2015) to choose a slightly larger thresholding value, specifically $\delta_{p,n} = 2\log p \log\log n$, to mitigate finite-sample biases.

Theorem 5.1. Suppose $n_1/(n_1+n_2) \to \gamma$ for some constant $\gamma \in (0,1)$ as $\min\{n_1, n_2\} \to \infty$ and $\log p = o(n^{1/3})$. With Assumptions 5.1-5.3 in the supplement, under the null hypothesis $H_{0m} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, as $n_1, n_2, p \to \infty$,

$$P\left(J_m = 0 | H_{0m}\right) \to 1, \quad M_{PE} = \frac{1}{\widehat{\sigma}_{01}} \sum_{i=1}^{p} M_i + J_m \xrightarrow{d} N(0,1). \tag{5.3.2}$$

Theorem 5.1 proves that $J_m = 0$ holds under H_{0m} with probability tending to 1. Thus, adding J_m to the mean statistic $\widehat{\sigma}_{01}^{-1} M_{n_1,n_2}$ will not affect its limiting null distribution. The proposed power-enhanced mean test rejects H_{0m} with the significance level α if $M_{PE} \geqslant z_\alpha$.

5.3.1.2 Power-Enhanced Covariance Tests

As for the covariance test statistic T_{n_1,n_2}, we first decompose T_{n_1,n_2} into

$$T_{n_1,n_2} = \sum_{i=1}^{p}\sum_{j=1}^{p} T_{ij} = \sum_{i=1}^{p}\sum_{j=1}^{p}(A_{ij} + B_{ij} - 2C_{ij}),$$

where

$$\begin{aligned}
A_{ij} &= \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1} X_{ui}X_{vi}X_{uj}X_{vj} - \frac{2}{n_1(n_1-1)(n_1-2)}\sum_{u\neq v\neq k}^{n_1} X_{ui}X_{vi}X_{vj}X_{kj} \\
&+ \frac{1}{n_1(n_1-1)(n_1-2)(n_1-3)}\sum_{u\neq v\neq k\neq l}^{n_1} X_{ui}X_{vi}X_{kj}X_{lj}, \\
B_{ij} &= \frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_2} Y_{ui}Y_{vi}Y_{uj}Y_{vj} - \frac{2}{n_2(n_2-1)(n_2-2)}\sum_{u\neq v\neq k}^{n_2} Y_{ui}Y_{vi}Y_{vj}Y_{kj} \\
&+ \frac{1}{n_2(n_2-1)(n_2-2)(n_2-3)}\sum_{u\neq v\neq k\neq l}^{n_2} Y_{ui}Y_{vi}Y_{kj}Y_{lj}, \\
C_{ij} &= \frac{1}{n_1 n_2}\sum_{u}^{n_1}\sum_{v}^{n_2} X_{ui}Y_{vi}X_{uj}Y_{vj} - \frac{1}{n_1 n_2(n_1-1)}\sum_{u\neq k}^{n_1}\sum_{v}^{n_2} X_{ui}Y_{vi}Y_{vj}X_{kj} \\
&- \frac{1}{n_1 n_2(n_2-1)}\sum_{u\neq k}^{n_2}\sum_{v}^{n_1} Y_{ui}X_{vi}X_{vj}Y_{kj} \\
&+ \frac{1}{n_1 n_2(n_1-1)(n_2-1)}\sum_{u\neq k}^{n_1}\sum_{v\neq l}^{n_2} X_{ui}Y_{vi}X_{kj}Y_{lj}.
\end{aligned}$$

The decomposition is essential to derive the power enhancement. For each $i, j = 1, \cdots, p$, T_{ij} consistently estimates the element-wise difference in covariances, i.e., $T_{ij} \xrightarrow{p} (\sigma_{1,ij} - \sigma_{2,ij})^2$ as $n_1, n_2 \to \infty$. Under the null hypothesis $H_{0c} : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, the variance of T_{ij} is

$$\begin{aligned}
\xi_{ij} := 2\Bigg(&\frac{1}{n_1}\left(\sigma_{1,ij}^2 + \sigma_{1,ii}\sigma_{1,jj} + \Delta_1 \mathrm{tr}(\boldsymbol{\gamma}_{1i}\boldsymbol{\gamma}_{1j}^T \circ \boldsymbol{\gamma}_{1i}\boldsymbol{\gamma}_{1j}^T)\right) \\
&+ \frac{1}{n_2}\left(\sigma_{2,ij}^2 + \sigma_{2,ii}\sigma_{2,jj} + \Delta_2 \mathrm{tr}(\boldsymbol{\gamma}_{2i}\boldsymbol{\gamma}_{2j}^T \circ \boldsymbol{\gamma}_{2i}\boldsymbol{\gamma}_{2j}^T)\right)\Bigg)^2 (1 + o(1)),
\end{aligned}$$

where $\boldsymbol{\gamma}_{ki}$, $k = 1, 2$ are explicitly presented in Assumption 5.2 in the supplement.

By method of moments, ξ_{ij} can be consistently estimated by

$$\widehat{\xi}_{ij} := 2\left(\frac{1}{n_1}\left(\frac{1}{n_1}\sum_{u=1}^{n_1} X_{ui}^2 X_{uj}^2 - \widehat{\sigma}_{1,ij}^2\right) + \frac{1}{n_2}\left(\frac{1}{n_2}\sum_{v=1}^{n_2} Y_{vi}^2 Y_{vj}^2 - \widehat{\sigma}_{2,ij}^2\right)\right)^2,$$

where $\widehat{\sigma}_{1,ij}$ and $\widehat{\sigma}_{2,ij}$ are sample covariances of (X_i, X_j) and (Y_i, Y_j), respectively. Define

$$J_c = \sqrt{p}\sum_{i=1}^{p}\sum_{j=1}^{p} T_{ij}\widehat{\xi}_{ij}^{-1/2}\mathcal{I}\{\sqrt{2}T_{ij}\widehat{\xi}_{ij}^{-1/2} + 1 > \eta_p\} \tag{5.3.3}$$

as the power enhancement component for the covariance test, with $\eta_p = 4\log p$. Similar to the previous section, the theoretical analysis regarding J_c is established upon $\eta_p = 4\log p$. In practical implementations, we use a slightly larger thresholding value, specifically $\eta_{p,n} = 4\log p \log\log n$, for the purpose of mitigating finite-sample biases.

Theorem 5.2. Suppose $n_1/(n_1+n_2) \to \gamma$ for some constant $\gamma \in (0,1)$ as $\min\{n_1, n_2\} \to \infty$ and $\log p = o(n^{1/5})$. With Assumptions 5.1-5.3 in the supplement, under the null hypothesis $H_{0c} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, as $n_1, n_2, p \to \infty$,

$$P\left(J_c = 0 | H_{0c}\right) \to 1, \quad T_{PE} = \frac{1}{\widehat{\sigma}_{02}}\sum_{i=1}^{p}\sum_{j=1}^{p} T_{ij} + J_c \xrightarrow{d} N(0,1). \tag{5.3.4}$$

Theorem 5.2 proves that under the null hypothesis H_{0c}, $J_c = 0$ with probability approaching 1. The power-enhanced covariance test rejects H_{0m} with significance level α if $T_{PE} \geqslant z_\alpha$.

5.3.1.3 Power Enhancement Properties

In this section, we study the power enhancement properties of our proposed power-enhanced tests M_{PE} and T_{PE}. Chen and Qin (2010) and Li and Chen (2012) provided power analysis of the mean test statistic M_{n_1,n_2} and the covariance test statistic T_{n_1,n_2}, respectively. Consider the following parameter spaces $\mathcal{G}_m^d$ and $\mathcal{G}_c^d$ for their alternative hypotheses:

$$\mathcal{G}_m^d = \left\{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) : \min\{n_1, n_2\}\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2/\sqrt{\max\{\operatorname{tr}(\mathbf{\Sigma}_1^2), \operatorname{tr}(\mathbf{\Sigma}_2^2)\}} \to \infty\right\},$$

$$\mathcal{G}_c^d = \left\{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) : \mathbf{\Sigma}_1 > 0, \mathbf{\Sigma}_2 > 0, \frac{1}{n_1}\text{tr}(\mathbf{\Sigma}_1^2) + \frac{1}{n_2}\text{tr}(\mathbf{\Sigma}_2^2) = o\left(\text{tr}\{(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2\}\right)\right\}.$$

Chen and Qin (2010) pointed out that as $n_1, n_2, p \to \infty$, the mean test statistic M_{n_1,n_2} would correctly reject the null hypothesis H_{0m} with probability approaching 1 if the mean differences $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ fall into the subspace $\mathcal{G}_m^d$. Li and Chen (2012) drew analogous conclusions with regards to the covariance alternative space $\mathcal{G}_c^d$ corresponding to the covariance test T_{n_1,n_2}. More specifically, as $n_1, n_2, p \to \infty$,

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d} P\left(M_{n_1,n_2} \geqslant \widehat{\sigma}_{01} z_\alpha\right) \to 1 \quad \text{and} \quad \inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^d} P\left(T_{n_1,n_2} \geqslant \widehat{\sigma}_{02} z_\alpha\right) \to 1. \tag{5.3.5}$$

Note that $\mathcal{G}_m^d$ and $\mathcal{G}_c^d$ use the squared Euclidean-norm $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$ and the squared Frobenius-norm $\|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_F^2$ to specify a large magnitude of differences in mean vectors and covariance matrices in order for the quadratic-form test statistics to be powerful in detecting the discrepancies.

In what follows, we present the power enhancement properties of our proposed tests. We will show that adding the power enhancement components J_m and J_c enables the tests to observe sparse signals which only differ in a few coordinates.

Theorem 5.3. With the same conditions as in Theorem 5.2, as $n_1, n_2, p \to \infty$, we have

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s} P\left(M_{PE} \geqslant z_\alpha\right) \to 1, \quad \text{and} \quad \inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s} P\left(T_{PE} \geqslant z_\alpha\right) \to 1,$$

with

$$\begin{aligned}
\mathcal{G}_m^s =& \left\{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) : \max_{1\leqslant i\leqslant p} \frac{(\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} \geqslant C\delta_p\right\} \\
\mathcal{G}_c^s =& \left\{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) : \mathbf{\Sigma}_1 > 0, \mathbf{\Sigma}_2 > 0, \max_{1\leqslant i,j\leqslant p} \frac{(\sigma_{1,ij} - \sigma_{2,ij})^2}{\xi_{ij}^{1/2}} \geqslant C\eta_p\right\}
\end{aligned}$$

where C is an absolute constant that does not depend on n_1, n_2 and p.

Theorem 5.3 shows that the power-enhanced tests have the same rejection regions as those of the original tests, but the high power regions are substantially expanded from $\mathcal{G}_m^d$ and $\mathcal{G}_s^d$ to $\mathcal{G}_m^d \cup \mathcal{G}_m^s$ and $\mathcal{G}_c^d \cup \mathcal{G}_c^s$, respectively.

Remark 5.1. Theorems 5.1-5.3 demonstrate that δ_p and η_p dominate the maximum noise level under the null hypothesis, and select signals under the designated alternatives. As long as n and p are not too small such that $\delta_p, \eta_p > 1$, which coincides with the high-dimensional framework. The theorems confirms the resultant power-enhanced mean test M_{PE} and power-enhanced covariance test M_{PE} satisfy the three PE principles introduced by Fan et al. (2015).

5.3.2 Joint Distribution of the Two Power-Enhanced Test Statistics

With the two power-enhanced tests, we have boosted the respective power of testing mean vectors and covariance matrices. Before heading to the aggregation of information from the two aspects, we study the joint limiting distribution of the two statistics M_{PE} and T_{PE}.

We begin with some insights on the joint distributions for statistics of the two aspects. Suppose we have a random sample i.i.d. drawn from a univariate normal distribution, then it is well-known that the sample mean and sample variance are independent. To a slightly more complex case, suppose we have a random sample i.i.d. drawn from a multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in the traditional statistical settings when p is fixed. We look into two likelihood ratio test (LRT) statistics. Let Λ_1 be the LRT statistic for testing $H_0 : \{\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \mathbf{I}_p\}$ versus $H_a : \{\boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} = \mathbf{I}_p\}$, and let Λ_2 be the LRT statistic for testing $H_0 : \{\boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} = \mathbf{I}_p\}$ versus $H_a : \{\boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} > \mathbf{0}\}$. In multivariate statistics, we know that Λ_1 and Λ_2 are independent (Lemma 10.3.1, Anderson (2003)).

The above discussion inspires us to conjecture on analogous propositions regarding the joint distribution of M_{PE} and T_{PE}. As a matter of fact, in the following theorem, we prove that the two statistics are indeed asymptotically independent.

Theorem 5.4. Suppose $n_1/(n_1 + n_2) \to \gamma$ for some constant $\gamma \in (0, 1)$ as $\min\{n_1, n_2\} \to \infty$ and $\log p = o(n^{1/5})$. With Assumptions 5.1-5.3, under H_0, as $n_1, n_2, p \to \infty$,

$$P\left(M_{PE} \leqslant x_1, T_{PE} \leqslant x_2\right) \to \Phi(x_1)\Phi(x_2) \tag{5.3.6}$$

for any $x_1, x_2 \in \mathbb{R}$.

5.3.3 Power-Enhanced Simultaneous Test

With the information of the two separate power-enhanced tests at hand, the next step is to reasonably aggregate the results for testing means and covariances simultaneously. Most existing works rely on the classical likelihood ratio test (Anderson, 2003) and its variants (Jiang and Yang, 2013; Liu et al., 2017; Niu et al., 2019). Their test statistics are in the form of a summation of two statistics, where one is designed for detecting discrepancies in covariance matrices, and the other is to catch signals of distinct mean vectors. We call this type of combined statistic as the weighted sum statistics (Li and Xue, 2015; Li et al., 2018).

However, the weighted sum statistics bear some drawbacks. The two components are usually of different magnitudes. The combined test would be mostly driven by the statistic with a larger scale, but insensitive to the statistic with a smaller scale. Such inefficiency in combination would lead to power loss in certain alternative spaces. Also, the distribution of the weighted sum statistic depends on the convolution of two marginal distributions, which is usually computationally challenging, resulting in difficulty in choosing critical value.

We propose a scale-invariant statistic to simultaneously test the equality of mean vectors and covariance matrices, by combining their separate p-values via Fisher's method:

$$J_{n_1,n_2} = -2\log(p_m) - 2\log(p_c), \tag{5.3.7}$$

where $p_m = 1 - \Phi\left(M_{PE}\right)$ and $p_c = 1 - \Phi\left(T_{PE}\right)$ are the p-values acquired from the power-enhanced mean test and the covariance test, respectively, and $\Phi(\cdot)$ is the cdf of $N(0,1)$.

As a matter of fact, the Fisher's method has been widely used in meta-analysis for combining the results of multiple scientific studies (Hedges and Olkin, 2014). It is worth noticing that meta-analysis is designed for combining studies coming from independent resources. Yet combining two test statistics which are constructed from the same sample would be a different story, and therefore requires careful investigation on the independence assumption.

Theorem 5.4 proves that under the null hypothesis H_0, the two test statistics M_{PE} and T_{PE} are asymptotically independent. Hence, under H_0, p_m and p_c asymptotically independently follow a uniform distribution on the interval $[0,1]$,

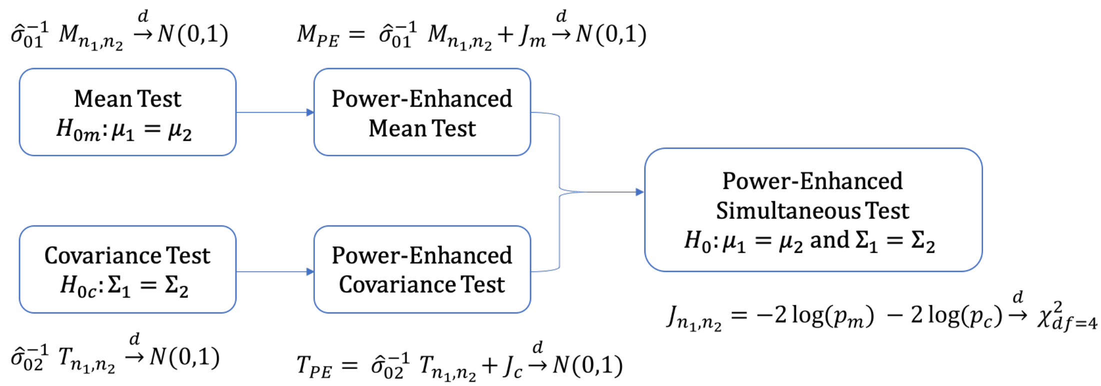


Figure 5.1: Power-enhanced simultaneous testing procedure

and therefore $-2\log(p_m)$ and $-2\log(p_c)$ asymptotically independently follow a chi-squared distribution with 2 degrees of freedom. As a result,

$$\text{under } H_0: \quad J_{n_1,n_2} \xrightarrow{d} \chi_4^2 \quad \text{as } n_1, n_2, p \to \infty. \tag{5.3.8}$$

Let q_α denote the upper-α quantile of χ_4^2 distribution, we reject the null hypothesis at the significance level α if

$$J_{n_1,n_2} \geqslant q_\alpha. \tag{5.3.9}$$

The procedures are summarized schematically in Figure 5.1. Equipped with the two key ingredients M_{PE} and T_{PE}, we proceed to investigate the size and power property of our proposed test J_{n_1,n_2} in Theorem 5.5. We show that our proposed test owns asymptotically accurate size approximation to the nominal significance level α and detects differences in either mean vectors or covariances over a wide range of alternatives.

Theorem 5.5 (Asymptotic Size and Power for Power-Enhanced Simultaneous Test)**.** Suppose $n_1/(n_1+n_2) \to \gamma$ for some constant $\gamma \in (0,1)$ as $\min\{n_1, n_2\} \to \infty$ and $\log p = o(n^{1/5})$. With Assumptions 5.1-5.3, as $n_1, n_2, p \to \infty$, the test J_{n_1,n_2} achieves (i) asymptotically accurate size, that is, under the null hypothesis H_0 : $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, $P\left(J_{n_1,n_2} \geqslant q_\alpha\right) \to \alpha$, and (ii) asymptotically consistent power, specifically, $\inf_{\{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s\}\cup\{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s\}} P\left(J_{n_1,n_2} \geqslant q_\alpha\right) \to 1$.

Remark 5.2. Theorem 5.5 confirms that our second PE procedure of expanding test capability from testing mean or covariances only to jointly testing mean vectors and covariance matrices satisfies the three PE principles.

There are other ways to aggregate information from the two aspects as the asymptotic independence permits the validity of many other combination methods. In what follows, we present two other tests using different methods to aggregate information to facilitate numerical comparison in the empirical studies. In Remark 5.3 and Section 5.4, we will show that the Fisher's combined test (5.3.9) outperforms other approaches as it is asymptotically optimal with respect to Bahadur efficiency.

One is a weighted statistics. Given the asymptotic independence, we may take the sum of squares of two statistics and transform the two asymptotic normal variables to an asymptotic χ_2^2 variable:

$$\text{under } H_0: \quad S_{n_1,n_2} = M_{PE} + T_{PE} \xrightarrow{d} \chi_2^2 \quad \text{as } n_1, n_2, p \to \infty. \tag{5.3.10}$$

The test rejects H_0 with a nominal significance level α if $S_{n_1,n_2} \geqslant c_\alpha$, where c_α is the upper-α quantile of χ_2^2 distribution. The other is an alternative p-value combination method. We consider the aggregation via Cauchy transformation (Liu and Xie, 2020; Liu et al., 2019). The Cauchy combination is appealing for its insensitiveness of dependence between the statistics to be combined. In here, even though we obtain the asymptotic independence between M_{PE} and T_{PE}, we introduce the Cauchy combination test as a promising alternative. We define the Cauchy combination statistic as follows.

$$C_{n_1,n_2} = \frac{1}{2}\tan\left((0.5 - p_m)\pi\right) + \frac{1}{2}\tan\left((0.5 - p_c)\pi\right). \tag{5.3.11}$$

Under H_0, the asymptotic independence ensures that C_{n_1,n_2} converges to a standard Cauchy distribution as $n_1, n_2, p \to \infty$. The test rejects H_0 with a nominal significance level α if $C_{n_1,n_2} \geqslant k_\alpha$, where k_α is the upper-α quantile of standard Cauchy distribution.

Remark 5.3. Littell and Folks (1971, 1973) established the asymptotic optimality of Fisher's methods for combining independent tests in terms of the Bahadur slope. To combine the two p-values $p_m = 1 - \Phi(M_{PE})$ and $p_c = 1 - \Phi(T_{PE})$, the Fisher's method yields the largest exact Bahadur slope among all reasonable methods for combining independent tests, leading to the fastest decay rate of the p-values under the alternative hypotheses. Such results imply that to attain equal test power, the Fisher's combination test requires the smallest sample size. No other combining method is superior to Fisher's method according to Bahadur relative efficiency. Therefore, J_{n_1,n_2} is asymptotically optimal with respect to Bahadur relative efficiency.

5.4 Simulation Studies

In this section, we conduct simulation studies to demonstrate the numerical performance of our proposed power-enhanced simultaneous test. To evaluate the power of the tests under different circumstances, we consider the following three types of alternative hypotheses.

(1) H_m: $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2,\ \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$.

(2) H_c: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2,\ \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$.

(3) H_b: $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2,\ \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$.

H_m describes the cases when the two populations share the same covariance matrix but have different means. H_c mimics the opposite situation in which the two populations have the same mean vector but differ in covariances. H_b considers the scenarios that there exist distinctions in both means and covariances among the two groups. For each alternative, we further consider two types of differences in the parameter of interest: the dense alternatives and the sparse alternatives. We use H_m^d and H_m^s to represent the existence of dense and sparse differences in $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, and analogously, H_c^d and H_c^s to denote those in $\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2$.

We simulate our samples from the moving average structure shown below, so that we are able to accommodate the complex alternative hypotheses in a general data generating process. For $i = 1, \cdots, p$, let

$$\begin{aligned} X_{u,i} &= \mu_{1,i} + Z_{u,i} + \theta_1 Z_{u,i+1}, & u = 1, \cdots, n_1, \\ Y_{v,i} &= \mu_{2,i} + Z_{v+n_1,i} + \theta_2 Z_{v+n_1,i+1}, & v = 1, \cdots, n_2, \end{aligned} \tag{5.4.1}$$

In such a way, the parameters $\{\mu_{1,i}, \mu_{2,i}\}$ alter the mean vectors of our simulated samples $\{\mathbf{X}_u\}_{u=1}^{n_1}$ and $\{\mathbf{Y}_v\}_{v=1}^{n_2}$ to generate H_m^d and H_m^s, and $\{\theta_1, \theta_2\}$ control the covariance structure to account for H_c^d. By assigning different values to these parameters, we obtain simulated samples with various means and covariances. For the sparse alternatives with respect to covariance matrices H_c^s, we generate samples from a different approach by letting $\mathbf{X}_u = \mathbf{\Sigma}_1^{1/2}\mathbf{Z}_u + \boldsymbol{\mu}_1$, $\mathbf{Y}_v = \mathbf{\Sigma}_2^{1/2}\mathbf{Z}_{v+n_1} + \boldsymbol{\mu}_2$ for $u = 1, \cdots, n_1$, $v = 1, \cdots, n_2$, where $\mathbf{Z}_k = (Z_{k,1}, \cdots, Z_{k,p})'$, $k = 1, \cdots, n_1 + n_2$.

To check the robustness to non-normally distributed data, we draw $\{Z_{k,i}, 1 \leqslant k \leqslant n_1 + n_2, 1 \leqslant i \leqslant p+1\}$ identically and independently (i.i.d.) from two distributions: one is standard normal $N(0, 1)$, and the other is centralized Gamma$(4, 2)$. We take the sample sizes as $n_1 = n_2 = N$ being 100 and 200, and let the dimension p take values in $\{100, 200, 500, 800, 1000\}$. For each setup, we compare our three proposed testing methods with four existing popular approaches: our proposed power-enhanced simultaneous test J_{n_1,n_2} as in (5.3.9), the proposed power-enhanced mean test M_{PE} as in (5.3.2), the proposed power-enhanced covariance test T_{PE} as in (5.3.4), the mean test M_{n_1,n_2} proposed by Chen and Qin (2010) as in (5.2.3), the covariance test T_{n_1,n_2} proposed by Li and Chen (2012) as in (5.2.5), the S_{n_1,n_2} approximation test as in (5.3.10), and the Cauchy combination test C_{n_1,n_2} as in (5.3.11). Fore each simulation setting, we report the frequencies of rejections over $5,000$ replications with significance level $\alpha = 0.05$.

To carry out $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2,\ \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, we set $\mu_{1i} = \mu_{2i} = 0$ for all $i = 1, \cdots, p$, and $\theta_1 = \theta_2 = 0$. Both samples are essentially i.i.d. from p-dimensional standard normal or multivariate gamma distribution. To evaluate the power, we fix $\{\mu_{1i}\}_{i=1}^p$ as zeros and $\theta_1 = 0$ for the first population, and vary $\{\mu_{2i}\}_{i=1}^p$ to set up the mean differences in H_m^d and H_m^s. As for the covariance alternatives, we change θ_2 to account for dense covariance differences in H_c^d and implement sparsely differed covariance matrix pair $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ to generate H_c^s.

As for H_m, we set $\theta_2 = 0$ to make sure the two samples share the same covariance matrix. In term of the mean vectors, for H_m^d, we follow Benjamini and Hochberg (1995) and consider a fixed percentage (pct) of violations in $\mu_{1,i} = \mu_{2,i}$ for $i = 1, \cdots, p$. The nonzero signal strength is determined in a similar fashion to Li and Chen (2012) as $\delta = \sqrt{\eta p^{-1/2}}$. To prevent trivial power of α and 1, we choose $\eta = 0.3$ and $pct = 15\%$. We set $\mu_{2,i} = \delta$ for $1 \leqslant i \leqslant [p \cdot pct]$ and zeros for the remaining ones. For sparse alternative H_m^s, we set the nonzero signal to be $\delta = 0.3\sqrt{\log p}$ and the number of non-zeros to be p^r with $r = 0.05$.

As for H_c, we ensure the two samples share equal means on every dimension. We set $\theta_2 = 0.2$ to create an MA(1) pattern of covariance as the dense alternative H_c^d. For the sparse alternative H_c^s, we follow Cai et al. (2013) to generate a symmetric sparse matrix $\mathbf{U}$ with 8 random nonzero entries, each with a magnitude of $\delta = 0.3\sqrt{\log p^2}$. The locations of 4 nonzero entries are randomly selected from the upper triangle of $\mathbf{U}$ while the other 4 are specified by symmetry. Then we generate samples from $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ with $\mathbf{\Sigma}_1 = (1+\varepsilon)\mathbf{I}_p$ and $\mathbf{\Sigma}_2 = (1+\varepsilon)\mathbf{I}_p + \mathbf{U}$, where $\varepsilon = |\min\{\lambda_{\min}(\mathbf{U} + \mathbf{I}_p), 1\}| + 0.05$ is to make sure both $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are positive definite. Finally, with respect to H_b, we adopt the same idea as in H_m for the mean differences, and the same approach as in H_c for the covariance differences.

Table 5.1: Empirical size (%) with Gaussian and Gamma distributed $\{Z_{k,i}\}$ in the data generating process

		Normal					Gamma				
n	Method	$p = 100$	200	500	800	1000	100	200	500	800	1000
100	M_{n_1,n_2}	5.24	5.24	5.12	5.06	5.32	5.10	4.72	5.00	5.04	5.10
	M_{PE}	5.96	5.84	5.36	5.46	5.48	5.64	5.18	5.32	5.36	5.34
	T_{n_1,n_2}	4.96	4.80	4.90	5.02	4.82	5.30	5.22	4.96	5.22	4.44
	T_{PE}	4.96	4.80	4.90	5.02	4.82	5.32	5.22	4.96	5.22	4.44
	S_{n_1,n_2}	5.70	5.84	5.98	5.12	5.40	5.92	5.60	5.34	5.10	4.84
	C_{n_1,n_2}	5.58	5.80	6.14	5.24	5.68	5.86	5.64	5.48	5.24	5.32
	J_{n_1,n_2}	5.60	5.56	5.22	5.42	5.12	5.58	5.56	5.16	5.54	5.06
200	M_{n_1,n_2}	5.48	5.30	5.46	5.16	5.22	4.94	5.06	5.06	5.24	5.26
	M_{PE}	5.68	5.56	5.62	5.18	5.30	5.34	5.32	5.18	5.32	5.34
	T_{n_1,n_2}	4.78	4.72	5.20	4.98	4.98	4.92	5.26	4.86	5.38	5.60
	T_{PE}	4.78	4.72	5.20	4.98	4.98	4.94	5.26	4.86	5.38	5.60
	S_{n_1,n_2}	5.14	4.80	5.46	5.22	5.46	5.64	5.22	5.18	5.56	5.54
	C_{n_1,n_2}	5.36	4.86	5.22	5.36	5.30	5.30	5.00	5.30	5.60	5.72
	J_{n_1,n_2}	5.50	5.12	5.24	5.30	5.08	5.50	5.22	5.56	5.54	5.54

Note: This table reports the frequencies of rejection by each method under the null hypothesis H_0 based on 5000 independent replications conducted at the significance level 5%.

Table 5.1 presents the empirical size of the seven tests with Normal and Gamma distributed $\{Z_{k,i}\}$ in the data generating process (5.4.1). Tables 5.2, 5.3, 5.4 and

5.5 report the empirical power of the seven methods for testing H_m, H_c and H_b with normal distributed $\{Z_{k,i}\}$. We also carry out studies on the power analysis for Gamma distributed $\{Z_{k,i}\}$. The results show a similar pattern to the Gaussian cases and are presented in Tables 5.6 - 5.9 in the supplementary materials. These numerical comparisons provide us with the following findings:

(1) Under H_0, all of the seven tests achieve reasonably accurate size approximation over a broad range of dimensionality. Besides, the empirical sizes with Gamma distribution illustrate that these tests are quite robust to non-Gaussianity.

(2) The numerical results of the power-enhanced tests M_{PE} and T_{PE} echo with the power enhancement properties presented in Theorems 5.1 and 5.2. Table 5.1 reveals that adding power enhancement components does not inflate the testing size under the null hypothesis H_0. On the other hand, Tables 5.2 - 5.5 reflect that the testing power is substantially enhanced under sparse alternatives H_m^s and H_c^s.

(3) As shown in Tables 5.2 and 5.3, the mean tests (M_{n_1,n_2} and M_{PE}) are powerful in detecting mean differences as in H_m, but have almost no power in discovering the covariance differences under H_c. In contrast, the covariance tests (T_{n_1,n_2} and T_{PE}) perform well in declaring significance for covariance alternative H_c, however, it is powerless to identify the unequal means under H_m.

(4) With respect to H_m and H_c, even though one of the M_{PE} test and T_{PE} test fails, the three combination tests remain powerful across all the experiments. This coincides with the power analysis shown in Theorem 5.5 that the combination of two tests makes use of their respective power under different alternatives, therefore successfully discover the discrepancies in either mean vectors or covariance matrices.

(5) Tables 5.4 and 5.5 illustrate that our proposed simultaneous test acquires additional gains when both mean differences and covariance differences exist. Under H_b, both M_{PE} test and and T_{PE} test successfully sense the differences

with regards to the means and covariances respectively. By combining the two tests together, our proposed approach yields to a higher testing power as it can simultaneously detect both types of differences.

(6) What's more, for each simulation setting, the proposed test J_{n_1,n_2} prevails with higher power compared with the S_{n_1,n_2} and C_{n_1,n_2} tests. This finding resounds with the asymptotically optimal property discussed in Remark 5.3.

In addition, Figure 5.2 provides a graphical representation of the testing power using seven approaches under H_b when both mean differences and covariances differences exist. We study four different hypotheses consisting of the combination of sparsely/densely differed means (H_m^s/H_m^d) and sparsely/densely differed covariances (H_c^s/H_c^d). The figure shows that the tests M_{n_1,n_2} and T_{n_1,n_2} favor dense alternatives H_m^d and H_c^d, respectively, because of its nature of quadratic forms, but lack the ability of detecting sparse alternatives such as H_m^s and H_c^s. Fortunately, the proposed power-enhanced tests M_{PE} and T_{PE} greatly promote the respective testing power under H_m^s and H_c^s. When it comes to jointly testing means and covariances, the plot clearly shows that our proposed Fisher's combined test J_{n_1,n_2} achieves the highest power among the three combination approach (F_{n_1,n_2}, S_{n_1,n_2} and C_{n_1,n_2}).

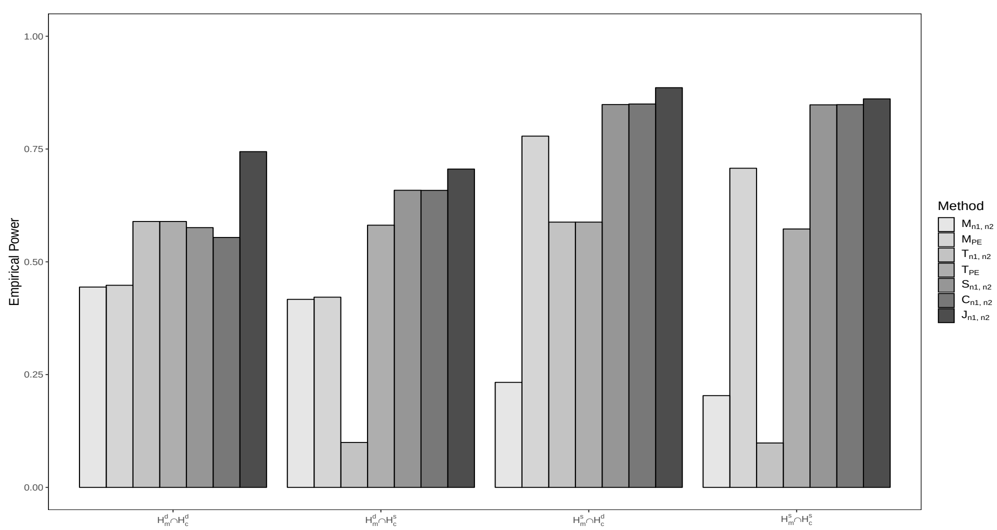


Figure 5.2: Empirical power comparison of the seven tests under H_b with Gaussian distributed $\{Z_{k,i}\}$ and $N = 100$, $p = 500$

Table 5.2: Empirical power (%) against H_m with Gaussian distributed $\{Z_{k,i}\}$ in the data generating process

H_m	N	Method	$p = 100$	200	500	800	1000
H_m^d	100	M_{n_1,n_2}	47.30	44.94	47.00	46.64	46.52
		M_{PE}	48.64	45.92	47.52	46.88	46.88
		T_{n_1,n_2}	5.38	5.36	4.98	5.12	4.48
		T_{PE}	5.38	5.36	4.98	5.12	4.48
		S_{n_1,n_2}	34.06	30.16	30.32	29.32	28.44
		C_{n_1,n_2}	34.22	29.88	29.96	28.88	27.98
		J_{n_1,n_2}	39.36	37.48	36.78	36.80	36.72
	200	M_{n_1,n_2}	87.04	88.92	90.76	91.54	91.32
		M_{PE}	87.34	89.10	90.86	91.58	91.34
		T_{n_1,n_2}	5.36	4.80	5.62	4.64	4.96
		T_{PE}	5.36	4.80	5.62	4.64	4.96
		S_{n_1,n_2}	75.18	76.74	77.74	79.46	79.26
		C_{n_1,n_2}	75.48	77.82	78.60	80.40	79.96
		J_{n_1,n_2}	80.32	82.20	83.32	84.46	84.64
H_m^s	100	M_{n_1,n_2}	42.24	33.38	54.22	44.68	17.74
		M_{PE}	79.00	79.54	96.30	95.58	78.98
		T_{n_1,n_2}	4.74	4.28	4.80	4.98	5.44
		T_{PE}	4.74	4.28	4.80	4.98	5.44
		S_{n_1,n_2}	76.80	78.10	95.66	95.22	77.88
		C_{n_1,n_2}	76.92	78.04	95.60	95.20	78.02
		J_{n_1,n_2}	77.68	78.92	95.86	95.50	78.50
	200	M_{n_1,n_2}	82.24	72.48	94.46	88.60	40.12
		M_{PE}	99.22	99.68	100.00	99.98	99.74
		T_{n_1,n_2}	5.10	4.64	4.76	4.66	5.04
		T_{PE}	5.10	4.64	4.76	4.66	5.04
		S_{n_1,n_2}	99.18	99.66	100.00	99.98	99.74
		C_{n_1,n_2}	99.18	99.62	100.00	99.98	99.74
		J_{n_1,n_2}	99.18	99.64	100.00	99.98	99.74

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%. (2) H_m stands for the type of alternative hypotheses (dense/sparse) in regards of mean differences of the two populations.

Table 5.3: Empirical power (%) against H_c with Gaussian distributed $\{Z_{k,i}\}$ in the data generating process

H_c	N	Method	$p = 100$	200	500	800	1000
H_c^d	100	M_{n_1,n_2}	4.80	4.92	5.12	4.84	5.10
		M_{PE}	5.50	5.60	5.38	5.26	5.48
		T_{n_1,n_2}	58.62	60.38	58.90	59.12	59.98
		T_{PE}	58.64	60.38	58.90	59.12	59.98
		S_{n_1,n_2}	37.22	37.52	36.12	36.52	38.16
		C_{n_1,n_2}	37.60	38.38	36.64	37.00	38.58
		J_{n_1,n_2}	45.20	46.52	46.02	45.94	47.24
	200	M_{n_1,n_2}	5.10	5.00	5.26	5.18	5.26
		M_{PE}	5.48	5.26	5.36	5.32	5.40
		T_{n_1,n_2}	97.46	98.24	98.52	98.22	98.46
		T_{PE}	97.46	98.24	98.52	98.22	98.46
		S_{n_1,n_2}	91.80	92.62	93.94	92.90	93.16
		C_{n_1,n_2}	92.62	92.98	94.24	93.76	94.08
		J_{n_1,n_2}	94.34	94.68	95.62	95.50	95.64
H_c^s	100	M_{n_1,n_2}	5.12	4.98	5.14	5.28	5.22
		M_{PE}	5.76	5.44	5.38	5.54	5.44
		T_{n_1,n_2}	30.94	19.48	9.78	7.66	8.14
		T_{PE}	65.60	66.76	58.20	45.00	37.90
		S_{n_1,n_2}	60.00	63.74	57.06	44.02	36.72
		C_{n_1,n_2}	60.08	63.80	57.10	43.88	36.96
		J_{n_1,n_2}	62.46	65.58	57.92	44.82	37.54
	200	M_{n_1,n_2}	5.08	5.38	5.40	4.90	5.16
		M_{PE}	5.56	5.60	5.46	5.02	5.18
		T_{n_1,n_2}	72.40	46.14	17.54	11.64	9.84
		T_{PE}	99.68	99.48	98.70	99.04	99.16
		S_{n_1,n_2}	99.60	99.44	98.80	98.98	99.16
		C_{n_1,n_2}	99.58	99.44	98.82	98.98	99.16
		J_{n_1,n_2}	99.62	99.44	98.74	98.98	99.18

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%. (2) H_c stands for the type of alternative hypotheses (dense/sparse) in regards of covariance differences of the two populations.

Table 5.4: Empirical power (%) against H_b with Gaussian distributed $\{Z_{k,i}\}$ in the data generating process

H_b	N	Method	$p = 100$	200	500	800	1000
$H_m^d \cap H_c^d$	100	M_{n_1,n_2}	44.78	44.80	44.40	45.26	45.60
		M_{PE}	46.40	45.44	44.80	45.54	45.84
		T_{n_1,n_2}	57.80	58.70	58.94	59.16	59.78
		T_{PE}	57.80	58.70	58.94	59.16	59.78
		S_{n_1,n_2}	58.80	58.12	57.58	58.08	59.30
		C_{n_1,n_2}	57.22	56.56	55.40	55.76	56.72
		J_{n_1,n_2}	73.24	74.68	74.42	75.94	76.24
	200	M_{n_1,n_2}	84.44	85.58	87.92	89.20	89.22
		M_{PE}	84.86	85.74	88.02	89.26	89.22
		T_{n_1,n_2}	98.16	98.26	98.26	98.38	98.48
		T_{PE}	98.16	98.26	98.26	98.38	98.48
		S_{n_1,n_2}	98.84	99.12	99.38	99.24	99.30
		C_{n_1,n_2}	98.50	98.88	98.98	98.98	99.08
		J_{n_1,n_2}	99.64	99.88	99.88	99.88	99.90
$H_m^d \cap H_c^s$	100	M_{n_1,n_2}	38.90	42.42	41.68	39.08	38.40
		M_{PE}	40.52	43.12	42.16	39.42	38.70
		T_{n_1,n_2}	32.48	18.16	9.96	7.96	7.12
		T_{PE}	72.36	72.62	58.12	42.76	37.58
		S_{n_1,n_2}	75.42	77.76	65.86	52.62	48.62
		C_{n_1,n_2}	75.02	77.84	65.84	52.84	48.82
		J_{n_1,n_2}	81.34	81.98	70.56	59.42	54.56
	200	M_{n_1,n_2}	77.06	82.92	85.12	83.38	82.46
		M_{PE}	77.54	83.14	85.22	83.40	82.48
		T_{n_1,n_2}	74.24	46.52	17.68	11.54	10.06
		T_{PE}	99.82	99.46	98.84	98.94	99.00
		S_{n_1,n_2}	99.90	99.72	99.40	99.38	99.34
		C_{n_1,n_2}	99.90	99.72	99.42	99.36	99.34
		J_{n_1,n_2}	99.92	99.82	99.52	99.42	99.42

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%. (2) H_b stands for the type of alternative hypotheses (dense/sparse) in regards of mean and covariance differences of the two populations.

Table 5.5: Empirical power (%) against H_b with Gaussian distributed $\{Z_{k,i}\}$ in the data generating process (continued)

H_b	N	Method	$p = 100$	200	500	800	1000
$H_m^s \cap H_c^d$	100	M_{n_1,n_2}	41.16	31.58	23.28	18.20	17.70
		M_{PE}	77.24	77.26	77.86	77.36	78.96
		T_{n_1,n_2}	58.18	59.96	58.80	58.26	58.78
		T_{PE}	58.20	59.96	58.80	58.26	58.78
		S_{n_1,n_2}	83.14	83.90	84.86	84.50	85.36
		C_{n_1,n_2}	83.14	84.06	84.98	84.64	85.72
		J_{n_1,n_2}	87.50	88.68	88.60	88.00	88.80
	200	M_{n_1,n_2}	81.52	69.82	51.78	41.72	38.26
		M_{PE}	99.10	99.46	99.60	99.74	99.74
		T_{n_1,n_2}	97.56	97.78	98.40	97.96	98.54
		T_{PE}	97.56	97.78	98.40	97.96	98.54
		S_{n_1,n_2}	99.96	99.96	100.00	99.98	99.98
		C_{n_1,n_2}	99.96	99.98	100.00	100.00	99.98
		J_{n_1,n_2}	99.98	100.00	100.00	100.00	99.98
$H_m^s \cap H_c^s$	100	M_{n_1,n_2}	34.84	29.34	20.34	16.94	15.04
		M_{PE}	66.12	71.62	70.74	67.76	68.60
		T_{n_1,n_2}	30.74	19.70	9.84	7.76	7.12
		T_{PE}	69.44	74.36	57.28	43.96	38.32
		S_{n_1,n_2}	85.56	89.86	84.80	78.64	78.14
		C_{n_1,n_2}	85.48	89.90	84.84	78.90	78.10
		J_{n_1,n_2}	87.30	91.14	86.12	79.82	79.44
	200	M_{n_1,n_2}	72.94	64.52	46.32	36.32	31.20
		M_{PE}	97.54	98.48	98.92	98.78	98.68
		T_{n_1,n_2}	74.22	46.46	17.76	11.72	10.94
		T_{PE}	99.72	99.36	99.00	98.86	98.88
		S_{n_1,n_2}	99.98	99.94	99.94	99.80	99.76
		C_{n_1,n_2}	99.98	99.94	99.94	99.84	99.76
		J_{n_1,n_2}	99.98	99.94	99.96	99.84	99.76

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%. (2) H_b stands for the type of alternative hypotheses (dense/sparse) in regards of mean and covariance differences of the two populations.

Table 5.6: Empirical power (%) against H_m with Gamma distributed $\{Z_{k,i}\}$ in the data generating process

H_m	N	Method	$p=100$	200	500	800	1000
H_m^d	100	M_{n_1,n_2}	46.88	46.22	47.72	46.80	46.88
		M_{PE}	48.46	46.98	48.22	47.06	47.34
		T_{n_1,n_2}	5.50	5.02	5.12	4.90	5.04
		T_{PE}	5.52	5.02	5.12	4.90	5.04
		S_{n_1,n_2}	33.60	30.44	30.64	29.80	29.40
		C_{n_1,n_2}	33.98	30.44	30.38	29.44	29.34
		J_{n_1,n_2}	40.22	37.10	38.12	37.38	37.66
	200	M_{n_1,n_2}	86.68	88.84	90.38	91.02	91.20
		M_{PE}	87.06	88.96	90.44	91.08	91.20
		T_{n_1,n_2}	5.34	5.38	4.94	5.24	4.96
		T_{PE}	5.34	5.38	4.94	5.24	4.96
		S_{n_1,n_2}	75.70	76.64	78.30	79.58	79.22
		C_{n_1,n_2}	76.06	77.04	79.22	80.26	80.20
		J_{n_1,n_2}	80.24	81.24	83.26	84.50	83.98
H_m^s	100	M_{n_1,n_2}	42.84	32.98	23.46	19.52	18.70
		M_{PE}	78.34	79.32	79.12	80.50	79.96
		T_{n_1,n_2}	5.18	5.10	5.18	5.22	4.98
		T_{PE}	5.22	5.10	5.18	5.22	4.98
		S_{n_1,n_2}	76.52	77.50	77.72	79.30	78.96
		C_{n_1,n_2}	76.48	77.52	77.84	79.08	79.10
		J_{n_1,n_2}	77.62	78.42	78.66	79.74	79.68
	200	M_{n_1,n_2}	82.98	73.38	52.78	45.22	39.70
		M_{PE}	99.50	99.64	99.68	99.82	99.80
		T_{n_1,n_2}	5.22	5.30	5.22	5.24	5.14
		T_{PE}	5.22	5.30	5.22	5.24	5.14
		S_{n_1,n_2}	99.38	99.58	99.64	99.80	99.76
		C_{n_1,n_2}	99.38	99.56	99.62	99.80	99.76
		J_{n_1,n_2}	99.40	99.58	99.62	99.80	99.76

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%. (2) H_m stands for the type of alternative hypotheses (dense/sparse) in regards of mean differences of the two populations.

Table 5.7: Empirical power (%) against H_c with Gamma distributed $\{Z_{k,i}\}$ in the data generating process

H_c	N	Method	$p = 100$	200	500	800	1000
H_c^d	100	M_{n_1,n_2}	5.00	4.90	5.02	4.98	5.08
		M_{PE}	5.58	5.36	5.32	5.22	5.26
		T_{n_1,n_2}	58.08	58.72	59.36	58.28	59.90
		T_{PE}	58.10	58.74	59.36	58.28	59.90
		S_{n_1,n_2}	36.16	37.14	37.28	36.48	38.54
		C_{n_1,n_2}	37.56	37.92	37.14	36.46	39.10
		J_{n_1,n_2}	45.54	46.50	45.62	44.68	48.40
	200	M_{n_1,n_2}	5.28	5.10	5.12	5.20	5.36
		M_{PE}	5.62	5.32	5.22	5.28	5.42
		T_{n_1,n_2}	97.12	97.94	98.54	98.32	98.62
		T_{PE}	97.12	97.94	98.54	98.32	98.62
		S_{n_1,n_2}	91.02	92.44	92.90	93.40	93.84
		C_{n_1,n_2}	91.56	92.92	93.44	94.24	94.86
		J_{n_1,n_2}	93.26	94.92	95.36	95.70	95.94
H_c^s	100	M_{n_1,n_2}	5.26	5.42	5.40	5.08	5.50
		M_{PE}	5.60	5.52	5.46	5.10	5.52
		T_{n_1,n_2}	36.26	23.22	11.70	8.88	8.56
		T_{PE}	44.98	30.20	14.74	10.92	10.28
		S_{n_1,n_2}	34.64	22.08	10.34	8.60	8.56
		C_{n_1,n_2}	35.68	22.06	10.42	8.94	9.16
		J_{n_1,n_2}	39.86	26.86	13.68	10.56	10.02
	200	M_{n_1,n_2}	5.16	5.50	5.20	5.50	5.32
		M_{PE}	5.36	5.56	5.22	5.54	5.36
		T_{n_1,n_2}	67.72	47.68	19.18	12.88	11.32
		T_{PE}	86.90	75.10	45.64	33.48	27.58
		S_{n_1,n_2}	81.02	67.20	40.14	29.66	24.32
		C_{n_1,n_2}	81.62	67.32	39.80	29.58	24.46
		J_{n_1,n_2}	83.52	70.46	43.26	32.72	26.32

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%. (2) H_c stands for the type of alternative hypotheses (dense/sparse) in regards of covariance differences of the two populations.

Table 5.8: Empirical power (%) against H_b with Gamma distributed $\{Z_{k,i}\}$ in the data generating process

H_b	N	Method	$p=100$	200	500	800	1000
$H_m^d \cap H_c^d$	100	M_{n_1,n_2}	45.42	43.76	44.96	45.08	46.04
		M_{PE}	46.72	44.48	45.30	45.48	46.30
		T_{n_1,n_2}	57.56	59.30	60.08	60.70	60.20
		T_{PE}	57.58	59.32	60.08	60.70	60.20
		S_{n_1,n_2}	57.62	57.46	57.24	58.92	57.90
		C_{n_1,n_2}	55.94	55.42	55.32	56.60	55.94
		J_{n_1,n_2}	72.64	73.44	74.46	76.24	75.94
	200	M_{n_1,n_2}	84.34	85.66	88.48	89.48	89.38
		M_{PE}	84.64	85.86	88.56	89.52	89.38
		T_{n_1,n_2}	97.52	98.28	98.26	98.56	98.34
		T_{PE}	97.52	98.28	98.26	98.56	98.34
		S_{n_1,n_2}	98.12	98.94	99.08	99.28	99.14
		C_{n_1,n_2}	97.74	98.60	98.88	99.06	99.02
		J_{n_1,n_2}	99.46	99.74	99.76	99.88	99.72
$H_m^d \cap H_c^s$	100	M_{n_1,n_2}	38.44	40.82	41.32	39.57	38.92
		M_{PE}	40.26	41.70	41.58	39.79	39.02
		T_{n_1,n_2}	36.34	22.80	11.84	8.71	8.36
		T_{PE}	44.40	29.18	14.46	10.29	9.46
		S_{n_1,n_2}	48.44	39.58	28.40	27.93	24.86
		C_{n_1,n_2}	47.86	39.04	28.52	27.64	24.54
		J_{n_1,n_2}	58.30	50.18	40.00	36.93	34.56
	200	M_{n_1,n_2}	77.00	83.46	84.40	83.39	82.84
		M_{PE}	77.90	83.72	84.54	83.39	82.90
		T_{n_1,n_2}	67.78	45.94	19.46	13.44	10.98
		T_{PE}	85.94	72.74	43.38	30.28	26.40
		S_{n_1,n_2}	91.26	87.96	78.74	74.72	71.94
		C_{n_1,n_2}	90.62	87.10	79.38	75.28	72.80
		J_{n_1,n_2}	94.34	92.36	85.74	81.28	79.42

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%. (2) H_b stands for the type of alternative hypotheses (dense/sparse) in regards of mean and covariance differences of the two populations.

Table 5.9: Empirical power (%) against H_b with Gamma distributed $\{Z_{k,i}\}$ in the data generating process (continued)

H_b	N	Method	$p = 100$	200	500	800	1000
$H_m^s \cap H_c^d$	100	M_{n_1,n_2}	41.38	32.02	22.14	19.02	16.66
		M_{PE}	76.86	77.74	78.46	78.32	79.40
		T_{n_1,n_2}	58.16	60.80	57.92	58.98	59.72
		T_{PE}	58.18	60.82	57.94	58.98	59.72
		S_{n_1,n_2}	83.56	84.80	85.26	84.86	86.18
		C_{n_1,n_2}	83.42	84.96	85.58	84.82	86.36
		J_{n_1,n_2}	87.96	88.68	88.98	88.38	89.44
	200	M_{n_1,n_2}	82.22	71.14	51.92	43.64	37.84
		M_{PE}	99.04	99.40	99.62	99.74	99.84
		T_{n_1,n_2}	97.18	97.64	98.42	98.60	98.22
		T_{PE}	97.20	97.64	98.42	98.60	98.22
		S_{n_1,n_2}	99.90	99.96	99.98	99.96	99.98
		C_{n_1,n_2}	99.94	99.98	99.98	99.98	99.98
		J_{n_1,n_2}	99.94	99.98	100.00	99.98	99.98
$H_m^s \cap H_c^s$	100	M_{n_1,n_2}	33.30	29.82	19.76	16.76	14.60
		M_{PE}	68.78	72.24	70.54	68.42	68.48
		T_{n_1,n_2}	36.72	22.80	11.56	8.52	8.02
		T_{PE}	45.42	30.00	14.88	10.38	9.24
		S_{n_1,n_2}	76.90	75.44	70.92	67.68	67.72
		C_{n_1,n_2}	77.02	75.18	70.62	67.68	67.54
		J_{n_1,n_2}	79.56	78.04	72.96	69.54	69.44
	200	M_{n_1,n_2}	73.42	64.90	46.66	36.84	33.44
		M_{PE}	96.56	97.46	98.08	98.16	98.18
		T_{n_1,n_2}	67.58	46.46	19.00	12.94	10.56
		T_{PE}	86.06	73.48	47.22	32.64	27.40
		S_{n_1,n_2}	98.84	98.72	98.52	98.58	98.40
		C_{n_1,n_2}	98.84	98.72	98.62	98.54	98.38
		J_{n_1,n_2}	99.06	99.02	98.74	98.74	98.46

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%. (2) H_b stands for the type of alternative hypotheses (dense/sparse) in regards of mean and covariance differences of the two populations.

5.5 Application to Gene-Set Testing

This section demonstrates the power of our proposed tests through a real application on an Acute Lymphoblastic Leukemia (ALL) dataset from the Ritz Laboratory at the Dana-Farber Cancer Institute (DFCI). The data was originally published by Chiaretti et al. (2004) and is now available at the Bioconductor website. The ALL dataset contains gene expression levels of 12,625 probes on Affymetrix chip series HG-U95Av2 from 128 individuals with either T-cell ALL or B-cell ALL, depending on the type of lymphocyte for the leukemia cells. This study focuses on a subset of the ALL data for 79 patients with the B-cell ALL. We further divide the patients into two groups according to their B-cell tumors' subtypes: the BCR/ABL fusion and the cytogenetically normal NEG, whose sample sizes are 37 and 42, respectively.

Identifying differentially expressed gene-sets has received considerable attention in genetic studies (Efron and Tibshirani, 2007; Goeman and Bühlmann, 2007). Since each gene does not work individually but rather tend to function groups to achieve complex biological tasks, researchers look into gene expression profiles based on groups of genes depending on their functional characteristics. To make full use of prior biological knowledge, we group sets of genes according to their Gene Ontology (GO) annotations. The GO system describes the biological domains with respect to three aspects: biological process (BP), cellular component (CC), and molecular function (MF). We follow the same criteria as in Dudoit and Van Der Laan (2007) to perform a pre-screening procedure by excluding those probes with low fluorescence intensities and narrowly spread, characterized by small absolute values and small interquantile ranges. The filtering step retains 2,391 probes, corresponding to 1849 unique GO terms in BP category, 306 in CC and 324 in MF.

Let $S_1, \cdots, S_K$ denote K gene-sets and $\{\boldsymbol{\mu}_{1S_k}, \boldsymbol{\Sigma}_{1S_k}\}$, $\{\boldsymbol{\mu}_{2S_k}, \boldsymbol{\Sigma}_{2S_k}\}$ be the mean vectors and covariance matrices of two types of tumors respectively. We are interested in testing

$$H_{0,category} : \boldsymbol{\mu}_{1S_k} = \boldsymbol{\mu}_{2S_k} \text{ and } \boldsymbol{\Sigma}_{1S_k} = \boldsymbol{\Sigma}_{2S_k}, \quad k = 1, \cdots, K$$

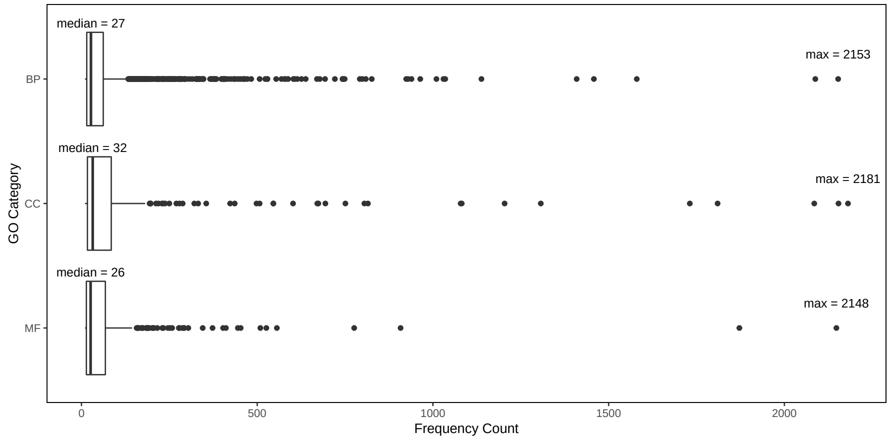


Figure 5.3: Boxplots of the dimension of gene-sets for three GO categories

where *category* $\in \{BP, CC, MF\}$. We classify gene-sets into three different GO categories and shall test each GO category separately. Figure 5.3 plots the dimension of gene-sets contained in each category. The dimension of gene-sets in each category can be as large as two thousand, which is much larger than the sample sizes $n_1 = 37$ and $n_2 = 42$.

Before proceeding, we explore the compuated values of power-enhanced test statistics M_{n_1,n_2} and T_{n_1,n_2} for all gene-sets. Figure 5.4 presents boxplots of M_{n_1,n_2} and T_{n_1,n_2} within each GO category. The M_{n_1,n_2} statistics have relatively larger values compared with the T_{n_1,n_2} statistics. Recall that under the null hypothesis, both statistics converge to $N(0,1)$ in distribution. The finding that the M_{n_1,n_2} statistics have larger absolute values indicates that for these gene-sets, their mean vectors are more different compared to the covariance matrices between the two groups. Moreover, considering significance level $\alpha = 0.05$ and the upper α-quantile of $N(0,1)$ $z_\alpha = 1.645$, a large number of M_{n_1,n_2} statistics fall above the threshold z_α. Therefore we would expect a lot of rejections for testing the equality of the mean vectors. The discussions in this paragraph give us an exploratory view of the dataset. Later on, we will present more precise comparisons among various test approaches.

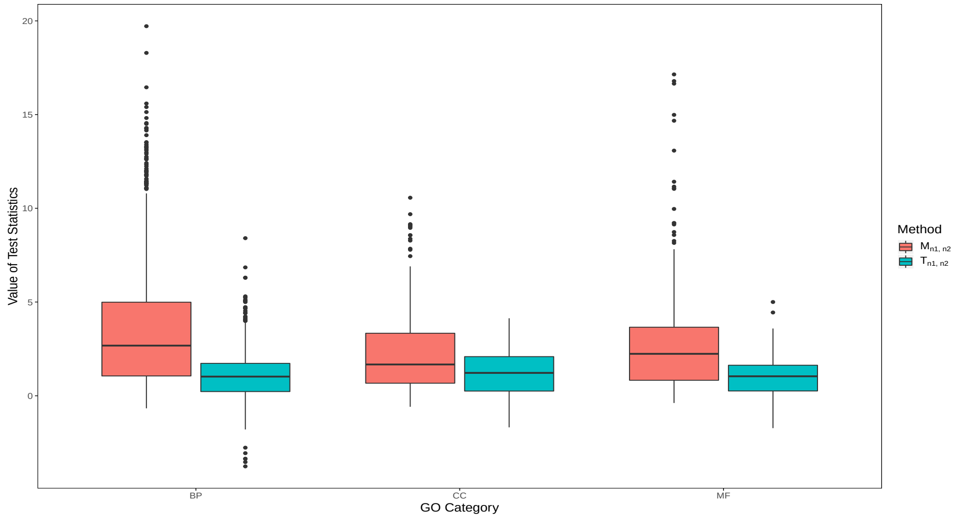


Figure 5.4: Boxplots of the $\widehat{\sigma}_{01}^{-1} M_{n_1,n_2}$ and $\widehat{\sigma}_{02}^{-1} T_{n_1,n_2}$ test statistics for three GO categories

We then apply our power-enhanced simultaneous test J_{n_1,n_2} to test the means and covariances simultaneously, together with the mean test M_{n_1,n_2}, the covariance test T_{n_1,n_2} and the two power-enhanced tests M_{PE} and T_{PE}. We compare our proposed approach J_{n_1,n_2} with the χ^2 approximation S_{n_1,n_2} as well as the Cauchy combination C_{n_1,n_2}. In order to control the false discovery rate (FDR), we apply the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg, 1995) to each GO category. Table 5.10 reports the number of significant gene-sets declared by different tests with nominal level $\alpha = 0.05$ for every category.

Table 5.10: The number of significant gene-sets declared by different tests after BH control with nominal level $\alpha = 0.05$

GO Category		BP	CC	MF
Total number of Gene-sets		1849	306	324
Number of Significant Gene-sets	M_{n_1,n_2}	1134	140	183
	M_{PE}	1469	216	236
	T_{n_1,n_2}	126	55	20
	T_{PE}	126	55	20
	S_{n_1,n_2}	1485	219	234
	C_{n_1,n_2}	1484	220	233
	J_{n_1,n_2}	1511	226	238

As shown in Table 5.10, J_{n_1,n_2} identifies more significant gene-sets than the other methods. The M_{n_1,n_2} test declares a lot of significance whereas the T_{n_1,n_2} test only identifies a few. The M_{PE} identifies a few more differentially expressed gene-sets with respect to mean vectors, while the T_{PE} does not yield additional power in detecting the differences among covariances. This indicates there exist a large number of unequal means between the two types of tumors, but not much differences in their covariance patterns. This phenomenon emphasizes the importance of developing a powerful method for jointly testing the means and covariances, so that we have a better chance to detect differences between two distributions even though we are in lack of prior knowledge about whether the differences reside in means or covariances.

The χ^2 approximation S_{n_1,n_2} and the Cauchy combination C_{n_1,n_2} yield comparative performance. As shown in Table 5.10, the two methods identify more differences than the covariance test T_{PE}, yet potentially miss some differentially expressed gene-sets compared to the mean test M_{PE}. In contrast, our proposed J_{n_1,n_2} is able to identify more discrepancies between the two groups, compared to the other three combination approaches and also compared to the original means test as well as the covariance tests. In a short summary, our proposed Fisher's combined simultaneous test J_{n_1,n_2} benefits from incorporating the information from the mean tests and covariance tests, and outperforms other combination methods in detecting the significant differences among the gene-sets.

Next, we study those gene-sets which are declared significant only by J_{n_1,n_2} but not any other method. Among those, we pay special attention to the GO-term "GO:0005125" in the MF category. This gene-set contributes to cytokine activity, including interleukins which are a group of cytokines that regulate inflammatory and immune responses (Okada and Pollack, 2004). Extensive scientific studies have revealed the close relationships between interleukins and leukemia (Touw et al., 1990; Paietta et al., 1997; Yoda et al., 2010; Canale et al., 2011). For another example, it is known microRNAs act complementarily to regulate disease-related mRNA modules in human diseases (Chavali et al., 2013). We observe that the expression levels of "GO:0006913" in the BP category are statistically different between the two groups. This GO-term refers to nucleocytoplasmic transport, whose association with leukemia has been validated by numerous cancer studies

(Chavali et al., 2013; Gravina et al., 2014; Takeda and Yaseen, 2014; Yan et al., 2017). These biological evidences suggest our power-enhanced simultaneous test J_{n_1,n_2} provides more useful information compared with other approaches, which further implies the importance of developing power-enhanced simultaneous tests.

5.6 Conclusion

In this work, we study the problem of jointly testing the equality of two-sample mean vectors and covariance matrices of high-dimensional data. We introduce a new power-enhanced simultaneous test, and prove the test achieves accurate asymptotic size, enhanced and consistent asymptotic power under a more general alternative, and asymptotic optimality with respect to Bahadur efficiency.

From a methodological perspective, we show two power enhancement technique via (i) constructing PE components, and (ii) aggregating information by the combination of p-values, to achieve enhanced test power in two aspects: (a) expanding high-power regions towards a wider alternative space with respect to one parameter, and (b) enhancing test capability to alternative spaces with respect to more parameters.

From a theoretical perspective, compared with existing simultaneous test, our method relaxes the stringent high-dimensionality assumption by allowing the dimension p to increase exponentially with the sample size n. In addition, our test is robust to non-Gaussian data. In theory, we derive the asymptotic independence of two PE statistics for the mean vectors and covariance matrices. And we prove that the proposed test satisfies PE properties following the guidance of the three general PE principles.

From a practical perspective, the proposed test is scale-invariant and computationally efficient. We demonstrate the finite-sample performance using simulation studies and a real-world application. The empirical studies on the gene-set testing problem further confirms that such power improvement is quite meaningful and very helpful in cancer research, with the support of biological evidence.

5.7 Assumptions, Lemmas and Proofs

5.7.1 Assumptions

This section lays out the fundamental assumptions underlying our results. Most of them are essentially needed to establish the asymptotic properties of the two test statistics M_{n_1,n_2} and T_{n_1,n_2}, and have also been discussed in Chen and Qin (2010) and Li and Chen (2012).

Assumption 5.1. For any $i, j, k, l \in \{1, 2\}$, as $n_1, n_2, p \to \infty$,

$$\operatorname{tr}(\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_l) \to \infty, \quad \operatorname{tr}\{\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_l\} = o\{\operatorname{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)\operatorname{tr}(\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_l)\}. \tag{5.7.1}$$

Assumption 5.2. The random vector $\{\mathbf{X}_u\}_{u=1}^{n_1}$, $\{\mathbf{Y}_v\}_{v=1}^{n_2}$ satisfy

$$\mathbf{X}_u = \boldsymbol{\Gamma}_1\mathbf{Z}_{1u} + \boldsymbol{\mu}_1,\ \mathbf{Y}_v = \boldsymbol{\Gamma}_2\mathbf{Z}_{2v} + \boldsymbol{\mu}_2 \quad 1 \leqslant u \leqslant n_1, 1 \leqslant v \leqslant n_2, \tag{5.7.2}$$

where $\boldsymbol{\Gamma}_i = (\boldsymbol{\gamma}_{i1}, \cdots, \boldsymbol{\gamma}_{ip})'$ is a $p \times m_i$ matrix for some $m_i \geqslant p$ such that $\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i' = \boldsymbol{\Sigma}_i$ for $i = 1, 2$, and $\{\mathbf{Z}_{ij}\}_{j=1}^{n_i} = \{(z_{ij1}, \cdots, z_{ijm_i})'\}_{j=1}^{n_i} \in \mathbb{R}^{m_i}$ are i.i.d. random vectors such that for any positive integers q and α_l such that $\sum_{l=1}^{q}\alpha_l \leqslant 8$, and for any $1 \leqslant k_1 \neq k_2 \neq \cdots \neq k_q \leqslant m_i$,

$$E(z_{ijk}) = 0,\ \operatorname{var}(z_{ijk}) = 1,\ \operatorname{Cov}(z_{ijk_1}, z_{ijk_2}) = 0,\ E(z_{ijk}^4) = 3 + \Delta_i,\ E(z_{ijk}^8) < \infty, \tag{5.7.3}$$

and

$$E(z_{ijk_1}^{\alpha_1}z_{ijk_2}^{\alpha_2}\cdots z_{ijk_q}^{\alpha_q}) = E(z_{ijk_1}^{\alpha_1})E(z_{ijk_2}^{\alpha_2})\cdots E(z_{ijk_q}^{\alpha_q}). \tag{5.7.4}$$

Assumption 5.3. There exists a positive constant H such that for all $h \in [-H, H]$,

$$Ee^{h(X_{ui}-\mu_{1i})^2} < \infty,\ Ee^{h(Y_{vi}-\mu_{2i})^2} < \infty \quad \text{for } i = 1, \cdots, p. \tag{5.7.5}$$

Remark 5.4. Assumptions 5.2 and 5.3 lay the foundation for our test to be robust to non-Gaussian data. (5.7.2) expresses the samples using a factor-model structure. (5.7.3) spells the moment condition needed for the factors z_{ijk}, in which the Δ_i measures the fourth-moment difference compared to a standard normal distribution. (5.7.4) depicts a pseudo-independence pattern among its components for each $\mathbf{Z}_{ij}$. The condition is trivially satisfied if $\mathbf{Z}_{ij}$ does have independent structure. (5.7.5) assumes the distributions of both populations are sub-Gaussian.

5.7.2 Lemmas

Lemma 5.1. Under the same conditions as in Theorem 5.5, as $n_1, n_2, p \to \infty$,

$$\mathrm{Cov}\left(\frac{M_{n_1,n_2}}{\sigma_{01}}, \frac{\widetilde{T}_{n_1,n_2}}{\sigma_{02}}\right) = o(1), \tag{5.7.6}$$

where $\widetilde{T}_{n_1,n_2}$ are defined by (5.7.32).

Proof. Without loss of generality, we assume the commmon mean vector under H_0 is $\boldsymbol{\mu} = \mathbf{0}$. We rewrite the statistics M_{n_1,n_2} and $\widetilde{T}_{n_1,n_2}$ into the form of two-sample U-statistics as follows.

$$M_{n_1,n_2} = \frac{4}{n_1(n_1-1)n_2(n_2-1)} \sum_{u<v}^{n_1} \sum_{k<l}^{n_2} H_1(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l), \tag{5.7.7}$$

and

$$\widetilde{T}_{n_1,n_2} = \frac{4}{n_1(n_1-1)n_2(n_2-1)} \sum_{u<v}^{n_1} \sum_{k<l}^{n_2} H_2(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l). \tag{5.7.8}$$

where

$$H_1(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l) = (\mathbf{X}_u'\mathbf{X}_v) + (\mathbf{Y}_k'\mathbf{Y}_l) - \frac{(\mathbf{X}_u'\mathbf{Y}_k) + (\mathbf{X}_v'\mathbf{Y}_l) + (\mathbf{X}_u'\mathbf{Y}_l) + (\mathbf{X}_v'\mathbf{Y}_k)}{2},$$

and

$$H_2(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l) = (\mathbf{X}_u'\mathbf{X}_v)^2 + (\mathbf{Y}_k'\mathbf{Y}_l)^2 - \frac{(\mathbf{X}_u'\mathbf{Y}_k)^2 + (\mathbf{X}_v'\mathbf{Y}_l)^2 + (\mathbf{X}_u'\mathbf{Y}_l)^2 + (\mathbf{X}_v'\mathbf{Y}_k)^2}{2}.$$

Let $\mathbf{A} = \boldsymbol{\Gamma}_1^T\boldsymbol{\Gamma}_1 = (a_{ij})_{1\leqslant i,j\leqslant m_1}$, $\mathbf{B} = \boldsymbol{\Gamma}_2^T\boldsymbol{\Gamma}_2 = (b_{ij})_{1\leqslant i,j\leqslant m_2}$, and $\mathbf{D} = \boldsymbol{\Gamma}_1^T\boldsymbol{\Gamma}_2 = (d_{ij})_{1\leqslant i\leqslant m_1, 1\leqslant j\leqslant m_2}$. Recall that under the null hypothesis, we have $\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1^T = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}$, and $\boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_2^T = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. The covariance of the two U-statistics $M_{n_1,n_2}/\sigma_{01}$ and $\widetilde{T}_{n_1,n_2}/\sigma_{02}$ can be obtained by

$$\mathrm{Cov}\left(\frac{M_{n_1,n_2}}{\sigma_{01}}, \frac{\widetilde{T}_{n_1,n_2}}{\sigma_{02}}\right) = \binom{n_1}{2}^{-1}\binom{n_2}{2}^{-1} \sum_{0\leqslant i,j\leqslant 2} \binom{2}{i}\binom{n_1-2}{2-i}\binom{2}{j}\binom{n_2-2}{2-j}\mathrm{cov}_{(i,j)}, \tag{5.7.9}$$

where $\mathrm{cov}_{(i,j)} = \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_{u_1}, \mathbf{X}_{v_1}, \mathbf{Y}_{k_1}, \mathbf{Y}_{l_1}), H_2(\mathbf{X}_{u_2}, \mathbf{X}_{v_2}, \mathbf{Y}_{k_2}, \mathbf{Y}_{l_2})\right)$ with i being the number of integers common to (u_1, v_1) and (u_2, v_2), and j being the number of integers common to (k_1, l_1) and (k_2, l_2). After simple calculation, we have

$$\begin{aligned}
\mathrm{cov}_{(0,0)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_3, \mathbf{X}_4, \mathbf{Y}_3, \mathbf{Y}_4)\right) = 0, \\
\mathrm{cov}_{(1,0)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_3, \mathbf{Y}_3, \mathbf{Y}_4)\right) = 0, \\
\mathrm{cov}_{(0,1)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_3, \mathbf{X}_4, \mathbf{Y}_1, \mathbf{Y}_3)\right) = 0, \\
\mathrm{cov}_{(1,1)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_3)\right) \\
&= \frac{\sigma_{01}^{-1}\sigma_{02}^{-1}}{4} \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3 E z_{11i}^3 E z_{21j}^3, \\
\mathrm{cov}_{(2,0)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_3, \mathbf{Y}_4)\right) \\
&= \sigma_{01}^{-1}\sigma_{02}^{-1} \sum_{i,j=1}^{m_1} a_{ij}^3 E z_{11i}^3 E z_{12j}^3, \\
\mathrm{cov}_{(0,2)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_3, \mathbf{X}_4, \mathbf{Y}_1, \mathbf{Y}_2)\right) \\
&= \sigma_{01}^{-1}\sigma_{02}^{-1} \sum_{i,j=1}^{m_2} b_{ij}^3 E z_{21i}^3 E z_{22j}^3, \\
\mathrm{cov}_{(2,1)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_3)\right) \\
&= \sigma_{01}^{-1}\sigma_{02}^{-1} \left(\sum_{i,j=1}^{m_1} a_{ij}^3 E z_{11i}^3 E z_{12j}^3 + \frac{1}{2} \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3 E z_{11i}^3 E z_{21j}^3 \right), \\
\mathrm{cov}_{(1,2)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_2)\right) \\
&= \sigma_{01}^{-1}\sigma_{02}^{-1} \left(\sum_{i,j=1}^{m_2} b_{ij}^3 E z_{21i}^3 E z_{22j}^3 + \frac{1}{2} \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3 E z_{11i}^3 E z_{21j}^3 \right), \\
\mathrm{cov}_{(2,2)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2)\right) \\
&= \sigma_{01}^{-1}\sigma_{02}^{-1} \left(\sum_{i,j=1}^{m_1} a_{ij}^3 E z_{11i}^3 E z_{12j}^3 + \sum_{i,j=1}^{m_2} b_{ij}^3 E z_{21i}^3 E z_{22j}^3 + \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3 E z_{11i}^3 E z_{21j}^3 \right).
\end{aligned}$$

Therefore, (5.7.9) becomes

$$\left| \mathrm{Cov}\left(\frac{M_{n_1,n_2}}{\sigma_{01}}, \frac{\widetilde{T}_{n_1,n_2}}{\sigma_{02}} \right) \right| \leqslant \frac{C}{n_1 n_2 \sigma_{01}\sigma_{02}} \left| \sum_{i,j=1}^{m_1} a_{ij}^3 + \sum_{i,j=1}^{m_2} b_{ij}^3 + \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3 \right|, \tag{5.7.10}$$

where C is a sufficiently large positive constant. What's more,

$$\left|\sum_{i,j=1}^{m_1} a_{ij}^3\right| \leqslant \max_{1\leqslant i,j\leqslant m_1} |a_{ij}| \cdot \sum_{i,j=1}^{m} a_{ij}^2 \leqslant \lambda_{\max}^{\frac{1}{2}}(\mathbf{A}^T\mathbf{A}) \cdot \operatorname{tr}(\mathbf{A}^2) = \lambda_{\max}(\mathbf{\Sigma}) \cdot \operatorname{tr}(\mathbf{\Sigma}^2).$$

Similarly,

$$\left|\sum_{i,j=1}^{m_2} b_{ij}^3\right| \leqslant \lambda_{\max}(\mathbf{\Sigma}) \cdot \operatorname{tr}(\mathbf{\Sigma}^2), \quad \left|\sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3\right| \leqslant \lambda_{\max}(\mathbf{\Sigma}) \cdot \operatorname{tr}(\mathbf{\Sigma}^2).$$

Further by Assumption 2, we have $\lambda_{\max}(\mathbf{\Sigma}) = o(\operatorname{tr}^{\frac{1}{2}}(\mathbf{\Sigma}^2))$, the proof of Lemma 5.1 is complete. □

Lemma 5.2. Suppose $n_1/(n_1+n_2) \to \gamma$ for some constant $\gamma \in (0,1)$ as $\min\{n_1, n_2\} \to \infty$. With Assumption 5.2, as $n_1, n_2 \to \infty$,

(i) if $\mu_{1i} = \mu_{2i}$,

$$\frac{\sqrt{2}M_i}{\nu_i^{1/2}} + 1 \xrightarrow{d} \chi_1^2, \tag{5.7.11}$$

where $\nu_i = \frac{2}{n_1(n_1-1)}\sigma_{1,ii}^2 + \frac{2}{n_2(n_2-1)}\sigma_{2,ii}^2 + \frac{4}{n_1 n_2}\sigma_{1,ii}\sigma_{2,ii}$.

(ii) if $\mu_{1i} \neq \mu_{2i}$,

$$\frac{M_i - (\mu_{1i}-\mu_{2i})^2}{\sqrt{4(\mu_{1i}-\mu_{2i})^2\left(\frac{\sigma_{1,ii}}{n_1} + \frac{\sigma_{2,ii}}{n_2}\right)}} \xrightarrow{d} N(0,1). \tag{5.7.12}$$

Proof. By definition, it is straightforward to show that

$$\begin{aligned}\operatorname{var}(M_i) =& \frac{2}{n_1(n_1-1)}\sigma_{1,ii}^2 + \frac{2}{n_2(n_2-1)}\sigma_{2,ii}^2 + \frac{4}{n_1 n_2}\sigma_{1,ii}\sigma_{2,ii} \\ &+ \frac{4}{n_1}(\mu_{1i}-\mu_{2i})^2\sigma_{1,ii} + \frac{4}{n_2}(\mu_{1i}-\mu_{2i})^2\sigma_{2,ii}.\end{aligned}$$

Note that M_i can be re-written as follows.

$$\begin{aligned}M_i &= \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1} X_{ui}X_{vi} + \frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_2} Y_{ui}Y_{vi} - \frac{2}{n_1 n_2}\sum_{u}^{n_1}\sum_{v}^{n_2} X_{ui}Y_{vi} \\ &= \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1}(X_{ui}-\mu_{1i})(X_{vi}-\mu_{1i}) + \frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_2}(Y_{ui}-\mu_{2i})(Y_{vi}-\mu_{2i})\end{aligned}$$

$$
\begin{aligned}
&-\frac{2}{n_1 n_2}\sum_u^{n_1}\sum_v^{n_2}(X_{ui}-\mu_{1i})(Y_{vi}-\mu_{2i})+\frac{2}{n_1}\sum_u^{n_1}(\mu_{1i}-\mu_{2i})(X_{ui}-\mu_{1i})\\
&+\frac{2}{n_2}\sum_v^{n_2}(\mu_{2i}-\mu_{1i})(Y_{vi}-\mu_{2i})+(\mu_{1i}-\mu_{2i})^2\\
:=\ & M_{i,0}+\frac{2}{n_1}\sum_u^{n_1}(\mu_{1i}-\mu_{2i})(X_{ui}-\mu_{1i})+\frac{2}{n_2}\sum_v^{n_2}(\mu_{2i}-\mu_{1i})(Y_{vi}-\mu_{2i})+(\mu_{1i}-\mu_{2i})^2,
\end{aligned}
\tag{5.7.13}
$$

where

$$
\begin{aligned}
M_{i,0} :=\ & \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1}(X_{ui}-\mu_{1i})(X_{vi}-\mu_{1i})+\frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_2}(Y_{ui}-\mu_{2i})(Y_{vi}-\mu_{2i})\\
&-\frac{2}{n_1 n_2}\sum_u^{n_1}\sum_v^{n_2}(X_{ui}-\mu_{1i})(Y_{vi}-\mu_{2i}).
\end{aligned}
\tag{5.7.14}
$$

We first show that $\sqrt{2}M_{i,0}\nu_i^{-1/2}+1 \xrightarrow{d} \chi_1^2$ as $n_1, n_2 \to \infty$. Without loss of generality, we consider $\mu_{1i}=\mu_{2i}=0$ in $M_{i,0}$. Therefore, (5.7.14) becomes

$$
\begin{aligned}
M_{i,0} =\ & \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1}X_{ui}X_{vi}+\frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_2}Y_{ui}Y_{vi}-\frac{2}{n_1 n_2}\sum_u^{n_1}\sum_v^{n_2}X_{ui}Y_{vi}\\
=\ & \frac{\sigma_{1,ii}}{n_1-1}\left[\left(\frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{X_{ui}}{\sigma_{1,ii}^{1/2}}\right)^2-\frac{1}{n_1}\sum_u^{n_1}\frac{X_{ui}^2}{\sigma_{1,ii}}\right]\\
&+\frac{\sigma_{2,ii}}{n_2-1}\left[\left(\frac{1}{\sqrt{n_2}}\sum_u^{n_2}\frac{Y_{ui}}{\sigma_{2,ii}^{1/2}}\right)^2-\frac{1}{n_2}\sum_u^{n_2}\frac{Y_{ui}^2}{\sigma_{2,ii}}\right]\\
&-\frac{2\sigma_{1,ii}^{1/2}\sigma_{2,ii}^{1/2}}{\sqrt{n_1 n_2}}\cdot\frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{X_{ui}}{\sigma_{1,ii}^{1/2}}\cdot\frac{1}{\sqrt{n_2}}\sum_v^{n_2}\frac{Y_{vi}}{\sigma_{2,ii}^{1/2}}\\
=\ & \left(\frac{\sigma_{1,ii}^{1/2}}{\sqrt{n_1}}\cdot\frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{X_{ui}}{\sigma_{1,ii}^{1/2}}-\frac{\sigma_{2,ii}^{1/2}}{\sqrt{n_2}}\cdot\frac{1}{\sqrt{n_2}}\sum_u^{n_2}\frac{Y_{ui}}{\sigma_{2,ii}^{1/2}}\right)^2-\frac{\sigma_{1,ii}}{n_1-1}\cdot\frac{1}{n_1}\sum_u^{n_1}\frac{X_{ui}^2}{\sigma_{1,ii}}\\
&-\frac{\sigma_{2,ii}}{n_2-1}\cdot\frac{1}{n_2}\sum_u^{n_2}\frac{Y_{ui}^2}{\sigma_{2,ii}}+\frac{\sigma_{1,ii}}{n_1(n_1-1)}\left(\frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{X_{ui}}{\sigma_{1,ii}^{1/2}}\right)^2\\
&+\frac{\sigma_{2,ii}}{n_2(n_2-1)}\left(\frac{1}{\sqrt{n_2}}\sum_u^{n_2}\frac{Y_{ui}}{\sigma_{2,ii}^{1/2}}\right)^2
\end{aligned}
\tag{5.7.15}
$$

By the central limit theorem and the law of large numbers, as $n_1, n_2 \to \infty$,

$$\frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{X_{ui}}{\sigma_{1,ii}^{1/2}} \xrightarrow{d} N(0,1),\ \ \frac{1}{\sqrt{n_2}}\sum_u^{n_2}\frac{Y_{ui}}{\sigma_{2,ii}^{1/2}} \xrightarrow{d} N(0,1),\ \ \frac{1}{n_1}\sum_u^{n_1}\frac{X_{ui}^2}{\sigma_{1,ii}} \xrightarrow{p} 1,\ \ \frac{1}{n_2}\sum_u^{n_2}\frac{Y_{ui}^2}{\sigma_{2,ii}} \xrightarrow{p} 1.$$

What's more,

$$\begin{aligned}
\frac{\sqrt{2}\sigma_{1,ii}}{n_1\nu_i^{1/2}} &= \frac{\sqrt{2}\sigma_{1,ii}}{n_1\cdot\sqrt{2}\left(\frac{\sigma_{1,ii}}{n_1}+\frac{\sigma_{2,ii}}{n_2}\right)(1+o(1))} \to \frac{\sigma_{1,ii}}{\sigma_{1,ii}+\frac{1}{1-\gamma}\sigma_{2,ii}},\\
\frac{\sqrt{2}\sigma_{2,ii}}{n_2\nu_i^{1/2}} &= \frac{\sqrt{2}\sigma_{2,ii}}{n_2\cdot\sqrt{2}\left(\frac{\sigma_{1,ii}}{n_1}+\frac{\sigma_{2,ii}}{n_2}\right)(1+o(1))} \to \frac{\sigma_{2,ii}}{(1-\gamma)\sigma_{1,ii}+\sigma_{2,ii}}.
\end{aligned}$$

Also because $\{X_{ui}\}_{u=1}^{n_1}$ and $\{Y_{vi}\}_{v=1}^{n_2}$ are independent, and note that

$$\frac{\sigma_{1,ii}}{\sigma_{1,ii}+\frac{1}{1-\gamma}\sigma_{2,ii}} + \frac{\sigma_{2,ii}}{(1-\gamma)\sigma_{1,ii}+\sigma_{2,ii}} = 1, \tag{5.7.16}$$

we have $\sqrt{2}M_{i,0}\nu_i^{-1/2}+1 \xrightarrow{d} \chi_1^2$ as $n_1, n_2 \to \infty$. According to (5.7.13), $M_i = M_{i,0}$ if $\mu_{1i} = \mu_{2i}$. We complete the proof of (5.7.11).

If $\mu_{1i} \neq \mu_{2i}$, based on (5.7.13),

$$\begin{aligned}
& M_i - (\mu_{1i}-\mu_{2i})^2 \\
=& M_{i,0} + \frac{2}{n_1}(\mu_{1i}-\mu_{2i})\sum_u^{n_1}(X_{ui}-\mu_{1i}) + \frac{2}{n_2}(\mu_{2i}-\mu_{1i})\sum_v^{n_2}(Y_{vi}-\mu_{2i}) \\
=& M_{i,0} + \frac{2\sigma_{1,ii}^{1/2}(\mu_{1i}-\mu_{2i})}{\sqrt{n_1}}\cdot\frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{X_{ui}-\mu_{1i}}{\sigma_{1,ii}^{1/2}} + \frac{2\sigma_{2,ii}^{1/2}(\mu_{2i}-\mu_{1i})}{\sqrt{n_2}}\cdot\frac{1}{\sqrt{n_2}}\sum_v^{n_2}\frac{Y_{vi}-\mu_{2i}}{\sigma_{2,ii}^{1/2}}
\end{aligned}$$

Since

$$\begin{aligned}
\frac{M_{i,0}}{\sqrt{4(\mu_{1i}-\mu_{2i})^2\left(\frac{\sigma_{1,ii}}{n_1}+\frac{\sigma_{2,ii}}{n_2}\right)}} &= \frac{M_{i,0}}{\nu_i^{1/2}}\cdot\left(\frac{2\left(\frac{\sigma_{1,ii}}{n_1}+\frac{\sigma_{2,ii}}{n_2}\right)^2(1+o(1))}{4(\mu_{1i}-\mu_{2i})^2\left(\frac{\sigma_{1,ii}}{n_1}+\frac{\sigma_{2,ii}}{n_2}\right)}\right)^{1/2} \to 0,\\
\frac{\left|2\sigma_{1,ii}^{1/2}(\mu_{1i}-\mu_{2i})/\sqrt{n_1}\right|}{\sqrt{4(\mu_{1i}-\mu_{2i})^2\left(\frac{\sigma_{1,ii}}{n_1}+\frac{\sigma_{2,ii}}{n_2}\right)}} &= \left(\frac{\sigma_{1,ii}}{n_1\left(\frac{\sigma_{1,ii}}{n_1}+\frac{\sigma_{2,ii}}{n_2}\right)}\right)^{1/2} \to \left(\frac{\sigma_{1,ii}}{\sigma_{1,ii}+\frac{1}{1-\gamma}\sigma_{2,ii}}\right)^{1/2},
\end{aligned}$$

$$\frac{\left|2\sigma_{2,ii}^{1/2}(\mu_{2i}-\mu_{1i})/\sqrt{n_2}\right|}{\sqrt{4(\mu_{1i}-\mu_{2i})^2\left(\frac{\sigma_{1,ii}}{n_1}+\frac{\sigma_{2,ii}}{n_2}\right)}} = \left(\frac{\sigma_{2,ii}}{n_2\left(\frac{\sigma_{1,ii}}{n_1}+\frac{\sigma_{2,ii}}{n_2}\right)}\right)^{1/2} \to \left(\frac{\sigma_{2,ii}}{(1-\gamma)\sigma_{1,ii}+\sigma_{2,ii}}\right)^{1/2},$$

and together with the central limit theorem,

$$\frac{M_i-(\mu_{1i}-\mu_{2i})^2}{\sqrt{4(\mu_{1i}-\mu_{2i})^2\left(\frac{\sigma_{1,ii}}{n_1}+\frac{\sigma_{2,ii}}{n_2}\right)}} \xrightarrow{d} N(0,1) \quad \text{as } n_1, n_2 \to \infty.$$

This finishes the proof of Lemma 5.2. □

Lemma 5.3. Suppose $n_1/(n_1+n_2) \to \gamma$ for some constant $\gamma \in (0,1)$ as $\min\{n_1, n_2\} \to \infty$. With Assumption 5.2, as $n_1, n_2 \to \infty$,

(i) if $\sigma_{1,ij} = \sigma_{2,ij}$,

$$\frac{\sqrt{2}T_{ij}}{\xi_{ij}^{1/2}} + 1 \xrightarrow{d} \chi_1^2, \tag{5.7.17}$$

where

$$\xi_{ij} = 2\left(\frac{1}{n_1}\left(\sigma_{1,ij}^2+\sigma_{1,ii}\sigma_{1,jj}+\Delta_1\mathrm{tr}(\gamma_{1i}\gamma_{1j}^T\circ\gamma_{1i}\gamma_{1j}^T)\right) + \frac{1}{n_2}\left(\sigma_{2,ij}^2+\sigma_{2,ii}\sigma_{2,jj}+\Delta_2\mathrm{tr}(\gamma_{2i}\gamma_{2j}^T\circ\gamma_{2i}\gamma_{2j}^T)\right)\right)^2(1+o(1)).$$

(ii) if $\sigma_{1,ij} \neq \sigma_{2,ij}$,

$$\frac{T_{ij}-(\sigma_{1,ij}-\sigma_{2,ij})^2}{\psi_{ij}^{1/2}} \xrightarrow{d} N(0,1), \tag{5.7.18}$$

where

$$\psi_{ij} = 4(\sigma_{1,ij}-\sigma_{2,ij})^2\left(\frac{1}{n_1}\left(\sigma_{1,ij}^2+\sigma_{1,ii}\sigma_{1,jj}+\Delta_1\mathrm{tr}(\gamma_{1i}\gamma_{1j}^T\circ\gamma_{1i}\gamma_{1j}^T)\right) + \frac{1}{n_2}\left(\sigma_{2,ij}^2+\sigma_{2,ii}\sigma_{2,jj}+\Delta_2\mathrm{tr}(\gamma_{2i}\gamma_{2j}^T\circ\gamma_{2i}\gamma_{2j}^T)\right)\right).$$

Proof. As discussed by Li and Chen (2012), the third and fourth-moment summation terms in A_{ij}, B_{ij} and C_{ij} are of the smaller order than the leading second-moment terms. After centering each variable, removing those terms from T_{ij} would not affect its asymptotic behavior. As a result, without loss of generality, we assume $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$ and study the simplified statistic

$$\begin{aligned}\widetilde{T}_{ij} =& \frac{1}{n_1(n_1-1)} \sum_{u\neq v}^{n_1} X_{ui}X_{uj}X_{vi}X_{vj} + \frac{1}{n_2(n_2-1)} \sum_{u\neq v}^{n_2} Y_{ui}Y_{uj}Y_{vi}Y_{vj} \\ & - \frac{2}{n_1 n_2} \sum_{u}^{n_1}\sum_{v}^{n_2} X_{ui}X_{uj}Y_{vi}Y_{vj}. \end{aligned} \tag{5.7.19}$$

$\widetilde{T}_{ij}$ and T_{ij} share the same asymptotic properties.

By definition, it is straightforward to show that $\mathrm{var}(\widetilde{T}_{ij}) = \psi_{ij} + \xi_{ij}$. Following similar arguments as in the proof of Lemma 5.2, we first rewrite $\widetilde{T}_{ij}$ in the form of

$$\widetilde{T}_{ij} = T_{ij}^{(1)} + T_{ij}^{(2)}, \tag{5.7.20}$$

where

$$\begin{aligned} T_{ij}^{(1)} :=& \frac{1}{n_1(n_1-1)} \sum_{u\neq v}^{n_1} (X_{ui}X_{uj} - \sigma_{1,ij})(X_{vi}X_{vj} - \sigma_{1,ij}) \\ & + \frac{1}{n_2(n_2-1)} \sum_{u\neq v}^{n_1} (Y_{ui}Y_{uj} - \sigma_{2,ij})(Y_{vi}Y_{vj} - \sigma_{2,ij}) \\ & - \frac{2}{n_1 n_2} \sum_{u}^{n_1}\sum_{v}^{n_2} (X_{ui}X_{uj} - \sigma_{1,ij})(Y_{vi}Y_{vj} - \sigma_{2,ij}), \end{aligned} \tag{5.7.21}$$

and

$$\begin{aligned} T_{ij}^{(2)} :=& (\sigma_{1,ij} - \sigma_{2,ij})^2 + \frac{2}{n_1}(\sigma_{1,ij} - \sigma_{2,ij}) \sum_{u=1}^{n_1} (X_{ui}X_{uj} - \sigma_{1,ij}) \\ & + \frac{2}{n_2}(\sigma_{2,ij} - \sigma_{1,ij}) \sum_{v=1}^{n_2} (Y_{vi}Y_{vj} - \sigma_{2,ij}). \end{aligned} \tag{5.7.22}$$

Following the same procedure as in the proof of Lemma 5.2, we are able to prove $\sqrt{2}T_{ij}^{(1)}\xi_{ij}^{-1/2} + 1 \xrightarrow{d} \chi_1^2$ as $n_1, n_2 \to \infty$. When $\sigma_{1,ij} = \sigma_{2,ij}$, $\widetilde{T}_{ij} = T_{ij}^{(1)}$, which proves (5.7.17). If $\sigma_{1,ij} \neq \sigma_{2,ij}$, we have $\psi_{ij}^{-1/2}(T_{ij}^{(2)} - (\sigma_{1,ij} - \sigma_{2,ij})^2) \xrightarrow{d} N(0,1)$ and $\psi_{ij}^{-1/2}T_{ij}^{(1)} = (\xi_{ij}/\psi_{ij})^{1/2} \cdot T_{ij}^{(1)}\xi_{ij}^{-1/2} \xrightarrow{p} 0$ as $n_1, n_2 \to \infty$. Together with (5.7.20), we complete the proof of Lemma 5.3. $\square$

5.7.3 Proof of Theorem 5.2

Proof. By Slutsky's theorem and Lemma 5.3, under the null hypothesis H_{0c} : $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, we have $\sqrt{2}T_{ij}\widehat{\xi}_{ij}^{-1/2} + 1 \xrightarrow{d} \chi_1^2$ as $n_1, n_2 \to \infty$. Lemma 2 of Chen, Guo and Qiu (2019) proves that $P\left(\max_{1\leqslant i,j\leqslant p} \sqrt{2}T_{ij}\widehat{\xi}_{ij}^{-1/2} + 1 > 4\log p | H_{0c}\right) \to 0$ as $n_1, n_2, p \to \infty$. Therefore, with $\eta_p = 4\log\log n \log p$,

$$P\left(J_c = 0 | H_{0c}\right) = P\left(\max_{1\leqslant i,j\leqslant p} \sqrt{2}T_{ij}\widehat{\xi}_{ij}^{-1/2} + 1 \leqslant \eta_{n,p} | H_{0c}\right) \to 1 \quad \text{as } n_1, n_2, p \to \infty.$$

Li and Chen (2012) proved that under the null hypothesis H_{0c} , $\widehat{\sigma}_{02}^{-1}\sum_{i=1}^p\sum_{j=1}^p T_{ij}$ $\xrightarrow{d} N(0,1)$ as $n_1, n_2, p \to \infty$. As a result, under H_{0c},

$$T_{PE} = \frac{1}{\widehat{\sigma}_{02}}\sum_{i=1}^p\sum_{j=1}^p T_{ij} + J_c \xrightarrow{d} N(0,1) \quad \text{as } n_1, n_2, p \to \infty.$$

The proof of Theorem 5.2 is complete. □

5.7.4 Proof of Theorem 5.3

Proof. For the power enhancement properties of M_{PE},

$$\begin{aligned} &\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s} P\left(M_{PE} \geqslant z_\alpha\right) \\ &\geqslant \min\left\{\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d} P\left(M_{PE} \geqslant z_\alpha\right), \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(M_{PE} \geqslant z_\alpha\right)\right\}. \end{aligned}$$

By definition, $J_m = \sqrt{p}\sum_{i=1}^p M_i\widehat{\nu}_i^{-1/2}\mathcal{I}\{\sqrt{2}M_i\widehat{\nu}_i^{-1/2} + 1 > \delta_p\} > 0$ as long as n and p are not too small such that $\delta_p > 1$. Hence, as $n_1, n_2, p \to \infty$,

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d} P\left(M_{PE} \geqslant z_\alpha\right) \geqslant \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d} P\left(\frac{1}{\widehat{\sigma}_{01}}\sum_{i=1}^p M_i \geqslant z_\alpha\right) \tag{5.7.23}$$

where the right-hand-side of (5.7.23) approaches 1 as shown in Chen and Qin (2010). It suffices to show that

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(M_{PE} \geqslant z_\alpha\right) \to 1. \tag{5.7.24}$$

Let

$$S_m(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \left\{ 1 \leqslant i \leqslant p : \frac{(\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} \geqslant 31\delta_p \right\}$$

and

$$\widehat{S}_m = \left\{ 1 \leqslant i \leqslant p : \frac{\sqrt{2} M_i}{\widehat{\nu}_i^{1/2}} + 1 > \delta_p \right\},$$

then $J_m = \sqrt{p} \sum_{i=1}^p M_i \widehat{\nu}_i^{-1/2} \mathcal{I}\{\sqrt{2} M_i \widehat{\nu}_i^{-1/2} + 1 > \delta_p\} = \sqrt{p} \sum_{i \in \widehat{S}_m} M_i \widehat{\nu}_i^{-1/2}$. By definition, $\{J_m \leqslant \sqrt{p}(\delta_p - 1)/\sqrt{2}\} = \{\widehat{S}_m = \varnothing\}$ and $\mathcal{G}_m^s = \{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) : S_m(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \neq \varnothing\}$. It is easy to see that, for $\epsilon > 0$,

$$\sup_{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathcal{G}_m^s} \max_{1 \leqslant i \leqslant p} P\left(\left| \frac{\widehat{\nu}_i}{\nu_i} - 1 \right| \geqslant \epsilon \right) = o(p^{-1}).$$

Note that,

$$\begin{aligned}
&\inf_{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathcal{G}_m^s} P\left(S_m(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \subset \widehat{S}_m \right) = \inf_{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathcal{G}_m^s} P\left(\bigcap_{i \in S_m(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)} \left\{ \frac{\sqrt{2} M_i}{\widehat{\nu}_i^{1/2}} + 1 > \delta_p \right\} \right) \\
&\geqslant \inf_{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathcal{G}_m^s} P\left(\bigcap_{i \in S_m(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)} \left\{ \left\{ \frac{M_i - (\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} + \frac{(\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} > \frac{\delta_p \sqrt{1+\epsilon}}{\sqrt{2}} \right\} \right.\right. \\
&\qquad \left.\left. \bigcap \left\{ \left| \frac{\widehat{\nu}_i}{\nu_i} - 1 \right| \leqslant \epsilon \right\} \right\} \right) \\
&\geqslant \inf_{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathcal{G}_m^s} \left\{ P\left(\bigcap_{i \in S_m(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)} \left\{ \frac{M_i - (\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} + \frac{(\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} > \frac{\delta_p \sqrt{1+\epsilon}}{\sqrt{2}} \right\} \right) \right. \\
&\qquad \left. - P\left(\bigcup_{i \in S_m(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)} \left\{ \left| \frac{\widehat{\nu}_i}{\nu_i} - 1 \right| \geqslant \epsilon \right\} \right) \right\} \\
&\geqslant \inf_{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathcal{G}_m^s} P\left(\bigcap_{i \in S_m(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)} \left\{ \frac{M_i - (\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} > -\frac{(\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} + \frac{\delta_p \sqrt{1+\epsilon}}{\sqrt{2}} \right\} \right) \\
&\qquad - p \max_{1 \leqslant i \leqslant p} P\left(\left| \frac{\widehat{\nu}_i}{\nu_i} - 1 \right| \geqslant \epsilon \right) \\
&\geqslant 1 - \sup_{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathcal{G}_m^s} P\left(\bigcup_{i \in S_m(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)} \left\{ \left| \frac{M_i - (\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} \right| > \frac{(\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} - \delta_p \frac{\sqrt{1+\epsilon}}{\sqrt{2}} \right\} \right) \\
&\qquad - p \max_{1 \leqslant i \leqslant p} P\left(\left| \frac{\widehat{\nu}_i}{\nu_i} - 1 \right| \geqslant \epsilon \right) \\
&\geqslant 1 - \sup_{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathcal{G}_m^s} p \cdot \max_{1 \leqslant i \leqslant p} P\left(\left| \frac{M_i - (\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} \right| > \frac{(\mu_{1i} - \mu_{2i})^2}{\nu_i^{1/2}} - \delta_p \frac{\sqrt{1+\epsilon}}{\sqrt{2}} \right)
\end{aligned}$$

$$- p \max_{1 \leqslant i \leqslant p} P\left(\left| \frac{\widehat{\nu}_i}{\nu_i} - 1 \right| \geqslant \epsilon \right).$$

Further that

$$\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} \max_{1\leqslant i\leqslant p} P\left(\left| \frac{M_i - (\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}} \right| > \frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}} - \delta_p \frac{\sqrt{1+\epsilon}}{\sqrt{2}} \right)$$
$$\leqslant \max_{1\leqslant i\leqslant p} P\left(\left| \frac{M_{i,0}}{\nu_i^{1/2}} \right| > \delta_p \right)$$
$$+ \sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} \max_{1\leqslant i\leqslant p} P\left(\left| \frac{2\sigma_{1,ii}^{1/2}(\mu_{1i}-\mu_{2i})}{\sqrt{n_1}\nu_i^{1/2}} \cdot \frac{1}{\sqrt{n_1}} \sum_u^{n_1} \frac{X_{ui}-\mu_{1i}}{\sigma_{1,ii}^{1/2}} \right| > \frac{1}{2}\left(\frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}} - \delta_p(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}}) \right) \right)$$
$$+ \sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} \max_{1\leqslant i\leqslant p} P\left(\left| \frac{2\sigma_{2,ii}^{1/2}(\mu_{2i}-\mu_{1i})}{\sqrt{n_2}\nu_i^{1/2}} \cdot \frac{1}{\sqrt{n_2}} \sum_v^{n_2} \frac{Y_{vi}-\mu_{2i}}{\sigma_{2,ii}^{1/2}} \right| > \frac{1}{2}\left(\frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}} - \delta_p(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}}) \right) \right)$$
$$\leqslant 2\exp(-\frac{\delta_p}{2})$$
$$+ \sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} \max_{1\leqslant i\leqslant p} P\left(\left| \frac{1}{\sqrt{n_1}} \sum_u^{n_1} \frac{X_{ui}-\mu_{1i}}{\sigma_{1,ii}^{1/2}} \right| > \frac{\sqrt{n_1}v_i^{1/4}}{4\sigma_{1,ii}^{1/2}} \left(\frac{|\mu_{1i}-\mu_{2i}|}{\nu_i^{1/4}} - \delta_p^{1/2}(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}}) \right) \right)$$
$$+ \sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} \max_{1\leqslant i\leqslant p} P\left(\left| \frac{1}{\sqrt{n_2}} \sum_v^{n_2} \frac{Y_{vi}-\mu_{2i}}{\sigma_{2,ii}^{1/2}} \right| > \frac{\sqrt{n_2}v_i^{1/4}}{4\sigma_{2,ii}^{1/2}} \left(\frac{|\mu_{1i}-\mu_{2i}|}{\nu_i^{1/4}} - \delta_p^{1/2}(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}}) \right) \right)$$
$$\leqslant 2\exp(-\frac{\delta_p}{2}) + \sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} \max_{1\leqslant i\leqslant p} P\left(\left| \frac{1}{\sqrt{n_1}} \sum_u^{n_1} \frac{X_{ui}-\mu_{1i}}{\sigma_{1,ii}^{1/2}} \right| > \frac{1}{4}\left(\delta_p^{1/2}(\sqrt{31}-1-\frac{\sqrt{1+\epsilon}}{\sqrt{2}}) \right) \right)$$
$$+ \sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} \max_{1\leqslant i\leqslant p} P\left(\left| \frac{1}{\sqrt{n_2}} \sum_v^{n_2} \frac{Y_{vi}-\mu_{2i}}{\sigma_{2,ii}^{1/2}} \right| > \frac{1}{4}\left(\delta_p^{1/2}(\sqrt{31}-1-\frac{\sqrt{1+\epsilon}}{\sqrt{2}}) \right) \right) = o(p^{-1}).$$

It implies, $\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2) \subset \widehat{S}_m \right) \to 1$. Furthermore,

$$\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(J_m \leqslant \sqrt{p}\cdot\frac{\delta_p - 1}{\sqrt{2}} \right) = \sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(\widehat{S}_m = \varnothing \right)$$
$$= \sup_{\{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2): S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\neq\varnothing\}} P\left(\widehat{S}_m = \varnothing \right) \to 0,$$

i.e., $\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(J_m > \sqrt{p}(\delta_p - 1)/\sqrt{2}\right) \to 1$. Therefore, as $n_1, n_2, p \to \infty$,

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(M_{PE} \geqslant z_\alpha\right) \geqslant \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(\sqrt{p}\cdot\frac{\delta_p - 1}{\sqrt{2}} + \frac{1}{\widehat{\sigma}_{01}}\sum_{i=1}^{p} M_i \geqslant z_\alpha\right) \to 1.$$

(5.7.24) is proved. In terms of power enhancement for T_{PE},

$$\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s} P\left(T_{PE} \geqslant z_\alpha\right) \geqslant \min\left\{\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^d} P\left(T_{PE} \geqslant z_\alpha\right), \inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(T_{PE} \geqslant z_\alpha\right)\right\}.$$

Let

$$S_c(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2) = \left\{(i,j) : \frac{(\sigma_{1,ij} - \sigma_{2,ij})^2}{\xi_{ij}^{1/2}} \geqslant 31\eta_p, 1 \leqslant i,j \leqslant p\right\},$$

and

$$\widehat{S}_c = \left\{(i,j) : \frac{\sqrt{2}T_{ij}}{\widehat{\xi}_{ij}^{1/2}} + 1 > \eta_p, 1 \leqslant i,j \leqslant p\right\},$$

then $J_c = \sqrt{p}\sum_{i=1}^{p}\sum_{j=1}^{p} T_{ij}\widehat{\xi}_{ij}^{-1/2}\mathcal{I}\{\sqrt{2}T_{ij}\widehat{\xi}_{ij}^{-1/2} + 1 > \eta_p\} = \sqrt{p}\sum_{(i,j)\in\widehat{S}_c} T_{ij}\widehat{\xi}_{ij}^{-1/2}$. Note that $J_c > 0$ so long as n and p are not too small such that $\eta_{n,p} > 1$. As $n_1, n_2, p \to \infty$,

$$\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^d} P\left(T_{PE} \geqslant z_\alpha\right) \geqslant \inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^d} P\left(\frac{1}{\widehat{\sigma}_{02}}\sum_{i=1}^{p}\sum_{j=1}^{p} T_{ij} \geqslant z_\alpha\right) \tag{5.7.25}$$

where the right-hand-side of (5.7.25) approaches 1 as shown in Li and Chen (2012). It remains to prove that

$$\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(T_{PE} \geqslant z_\alpha\right) \to 1 \quad \text{as } n_1, n_2, p \to \infty. \tag{5.7.26}$$

Similarly,

$$\begin{aligned}
&\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(S_c(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2) \subset \widehat{S}_c\right) = \inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(\bigcap_{(i,j)\in S_c(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)}\left\{\frac{\sqrt{2}T_{ij}}{\widehat{\xi}_{ij}^{1/2}} + 1 > \eta_p\right\}\right) \\
&\geqslant \inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(\left\{\bigcap_{i\in S_c(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)}\left\{\frac{T_{ij} - (\sigma_{1,ij} - \sigma_{2,ij})^2}{\xi_{ij}^{1/2}} + \frac{(\sigma_{1,ij} - \sigma_{2,ij})^2}{\xi_{ij}^{1/2}} > \frac{\eta_p\sqrt{1+\epsilon}}{\sqrt{2}}\right\}\right.\right. \\
&\qquad\qquad \left.\left.\bigcap\left\{\left|\frac{\widehat{\xi}_{ij}}{\xi_{ij}} - 1\right| \leqslant \epsilon\right\}\right\}\right)
\end{aligned}$$

$$\geqslant 1 - \sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} p^2 \cdot \max_{1\leqslant i,j\leqslant p} P\left(\left|\frac{T_{ij}-(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}}\right| > \frac{(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}} - \eta_p\frac{\sqrt{1+\epsilon}}{\sqrt{2}}\right)$$
$$- p^2 \max_{1\leqslant i,j\leqslant p} P\left(\left|\frac{\widehat{\xi}_{ij}}{\xi_{ij}} - 1\right| \geqslant \epsilon\right).$$

Together with

$$\sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} \max_{1\leqslant i,j\leqslant p} P\left(\left|\frac{T_{ij}-(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}}\right| > \frac{(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}} - \eta_p\frac{\sqrt{1+\epsilon}}{\sqrt{2}}\right)$$
$$\leqslant \max_i P\left(\left|\frac{T_{ij}^{(1)}}{\xi_{ij}^{1/2}}\right| > \eta_p\right)$$
$$+ \sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} \max_{1\leqslant i,j\leqslant p} P\left(\left|\frac{2(\sigma_{1,ij}-\sigma_{2,ij})}{n_1\xi_{ij}^{1/2}}\sum_{u=1}^{n_1}(X_{ui}X_{uj}-\sigma_{1,ij})\right| > \frac{1}{2}\left(\frac{(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}} - \eta_p^{1/2}(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}})\right)\right)$$
$$+ \sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} \max_{1\leqslant i,j\leqslant p} P\left(\left|\frac{2(\sigma_{2,ij}-\sigma_{1,ij})}{n_2\xi_{ij}^{1/2}}\sum_{u=1}^{n_1}(Y_{ui}Y_{uj}-\sigma_{2,ij})\right| > \frac{1}{2}\left(\frac{(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}} - \eta_p^{1/2}(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}})\right)\right)$$
$$\leqslant 2\exp(-\frac{\delta_p}{2})$$
$$+ \sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} \max_{1\leqslant i,j\leqslant p} P\left(\left|\frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{X_{ui}X_{uj}-\sigma_{1,ij}}{\mathrm{var}(X_{ui}X_{uj})^{1/2}}\right| > \frac{1}{4}\left(\eta_p^{1/2}(\sqrt{31}-1-\frac{\sqrt{1+\epsilon}}{\sqrt{2}})\right)\right)$$
$$+ \sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} \max_{1\leqslant i,j\leqslant p} P\left(\left|\frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{Y_{ui}Y_{uj}-\sigma_{2,ij}}{\mathrm{var}(Y_{ui}Y_{uj})^{1/2}}\right| > \frac{1}{4}\left(\eta_p^{1/2}(\sqrt{31}-1-\frac{\sqrt{1+\epsilon}}{\sqrt{2}})\right)\right)$$
$$= o(p^{-2})$$

and

$$\sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} \max_{1\leqslant i,j\leqslant p} P\left(\left|\frac{\widehat{\xi}_{ij}}{\xi_{ij}} - 1\right| \geqslant \epsilon\right) = o(p^{-2}) \quad \text{for } \epsilon > 0,$$

we have, $\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(S_c(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2) \subset \widehat{S}_c\right) \to 1$. Furthermore,

$$\sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(J_c \leqslant \sqrt{p}\cdot\frac{\eta_p-1}{\sqrt{2}}\right) = \sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(\widehat{S}_c = \varnothing\right)$$
$$= \sup_{\{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2):S_c(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\neq\varnothing\}} P\left(\widehat{S}_c = \varnothing\right) \to 0,$$

i.e., $\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^s} P\left(J_c > \sqrt{p}(\eta_p - 1)/\sqrt{2}\right) \to 1$. Therefore, as $n_1, n_2, p \to \infty$,

$$\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^s} P\left(T_{PE} \geqslant z_\alpha\right) \geqslant \inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^s} P\left(\sqrt{p}\cdot\frac{\eta_p - 1}{\sqrt{2}} + \frac{1}{\widehat{\sigma}_{02}}\sum_{i=1}^{p}\sum_{j=1}^{p} T_{ij} \geqslant z_\alpha\right) \to 1.$$

(5.7.26) is proved. The proof of Theorem 5.3 is complete. □

5.7.5 Proof of Theorem 5.4

Proof. From Theorems 5.1 and 5.2, under the null hypothesis H_0,

$$P\left(J_m = 0\right) \to 1, \quad P\left(J_c = 0\right) \to 1 \quad \text{as } n_1, n_2, p \to \infty.$$

Therefore, for any $x_1, x_2 \in \mathbb{R}$,

$$\begin{aligned}
&\left|P\left(M_{PE} \leqslant x_1, T_{PE} \leqslant x_2\right) - P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}} \leqslant x_1, \frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}} \leqslant x_2\right)\right| \\
\leqslant &\left|P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}} + J_m \leqslant x_1, \frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}} + J_c \leqslant x_2\right) - P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}} + J_m \leqslant x_1, \frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}} \leqslant x_2\right)\right| \\
&+ \left|P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}} + J_m \leqslant x_1, \frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}} \leqslant x_2\right) - P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}} \leqslant x_1, \frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}} \leqslant x_2\right)\right| \\
\leqslant &\left|P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}} + J_m \leqslant x_1, \frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}} + J_c \leqslant x_2 \middle| J_c = 0\right)\right. \\
&\left. - P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}} + J_m \leqslant x_1, \frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}} \leqslant x_2 \middle| J_c = 0\right)\right| \times P\left(J_c = 0\right) + P\left(J_c \neq 0\right) \\
&+ \left|P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}} + J_m \leqslant x_1, \frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}} \leqslant x_2 \middle| J_m = 0\right)\right. \\
&\left. - P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}} \leqslant x_1, \frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}} \leqslant x_2 \middle| J_m = 0\right)\right| \times P\left(J_m = 0\right) + P\left(J_m \neq 0\right) \\
&\to 0 \quad \text{under } H_0 \quad \text{as } n_1, n_2, p \to \infty.
\end{aligned} \tag{5.7.27}$$

It remains to prove that under H_0,

$$P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}} \leqslant x_1, \frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}} \leqslant x_2\right) \to \Phi(x_1)\Phi(x_2) \quad \text{as } n_1, n_2, p \to \infty. \tag{5.7.28}$$

Together with Lemma 5.1, it suffices to show that for any $a, b \in \mathbb{R}$, $aM_{n_1,n_2}/\widehat{\sigma}_{01} + bT_{n_1,n_2}/\widehat{\sigma}_{02}$ converge to a normal distribution. From the discussions of Chen and Qin (2010) and Li and Chen (2012), we know that $\widehat{\sigma}_{02}^{-1}\left(T_{n_1,n_2} - \widetilde{T}_{n_1,n_2}\right) \xrightarrow{p} 0$ and

$\widehat{\sigma}_{0i}/\sigma_{0i} \xrightarrow{p} 1$ for $i = 1, 2$. As a result, we only need to show that under H_0, $aM_{n_1,n_2}/\sigma_{01} + b\widetilde{T}_{n_1,n_2}/\sigma_{02}$ is asymptotically normally distributed. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the common mean vector and covariance matrix under the null hypothesis. Without loss of generality, we assume $\boldsymbol{\mu} = \mathbf{0}$.
For ease of notation, let $\mathbf{W}_i$ be a new random variable taking values of

$$\mathbf{W}_i = \mathbf{X}_i,\ i = 1, \cdots, n_1; \quad \mathbf{W}_{n_1+i} = \mathbf{Y}_i,\ i = 1, \cdots, n_2.$$

Let $\mathcal{F}_0 = \{\varnothing, \Omega\}$, $\mathcal{F}_m = \sigma\{Y_1, \ldots, Y_m\}$ for $m = 1, 2, \ldots, n_1 + n_2$, and let $E_m(\cdot)$ denote the conditional expectation given $\mathcal{F}_m$., i.e., $E_m(\cdot) = E(\,\cdot\,|\mathcal{F}_m)$. Consider

$$D_{n,m} = (E_m - E_{m-1})M_{n_1,n_2}, \quad G_{n,m} = (E_m - E_{m-1})\widetilde{T}_{n_1,n_2}.$$

To be more specific,

$$\begin{aligned}
D_{n,m} =& \frac{2}{n_1(n_1-1)} \sum_{i=1}^{m-1} \mathbf{W}_i'\mathbf{W}_m \mathcal{I}\{m \leqslant n_1\} + \frac{2}{n_2(n_2-1)} \sum_{i=n_1+1}^{m-1} \mathbf{W}_i'\mathbf{W}_m \mathcal{I}\{m > n_1\} \\
&- \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \mathbf{W}_i'\mathbf{W}_m \mathcal{I}\{m > n_1\} \\
G_{n,m} =& \frac{2}{n_1(n_1-1)} \left[\mathbf{W}_m' Q_{1,m-1} \mathbf{W}_m - \mathrm{tr}(Q_{1,m-1}\boldsymbol{\Sigma})\right] \mathcal{I}\{m \leqslant n_1\} \\
&+ \frac{2}{n_2(n_2-1)} \left[\mathbf{W}_m' Q_{2,m-1} \mathbf{W}_m - \mathrm{tr}(Q_{2,m-1}\boldsymbol{\Sigma})\right] \mathcal{I}\{m > n_1\} \\
&- \frac{2}{n_1 n_2} \left[\mathbf{W}_m' Q_{1,n_1} \mathbf{W}_m - \mathrm{tr}(Q_{1,n_1}\boldsymbol{\Sigma})\right] \mathcal{I}\{m > n_1\},
\end{aligned}$$

where $Q_{1,m-1} = \sum_{i=1}^{m-1}(\mathbf{W}_i\mathbf{W}_i' - \boldsymbol{\Sigma})$ and $Q_{2,n_1+l-1} = \sum_{i=1}^{l-1}(\mathbf{W}_{n_1+i}\mathbf{W}_{n_1+i}' - \boldsymbol{\Sigma})$. It's easy to verify that $aM_{n_1,n_2}/\sigma_{01} + b\widetilde{T}_{n_1,n_2}/\sigma_{02} = \sum_{i=1}^{n_1+n_2}(aD_{n,m}/\sigma_{01} + bG_{n,m}/\sigma_{02})$, and for any n, $\{aD_{n,m}/\sigma_{01} + bG_{n,m}/\sigma_{02}, 1 \leqslant m \leqslant n\}$ is a martingale difference sequence with respect to the σ-fields $\{\mathcal{F}_m, 1 \leqslant m \leqslant n\}$. By Martingale central limit theorem, we only need to show that

$$\frac{\sum_{m=1}^{n_1+n_2} E_{m-1}(aD_{n,m}/\sigma_{01} + bG_{n,m}/\sigma_{02})^2}{s_n^2} \xrightarrow{p} 1, \tag{5.7.29}$$

and

$$\frac{\sum_{m=1}^{n_1+n_2} E(aD_{n,m}/\sigma_{01} + bG_{n,m}/\sigma_{02})^4}{s_n^4} \to 0, \tag{5.7.30}$$

where $s_n^2 = \text{var}(aM_{n_1,n_2}/\sigma_{01} + b\widetilde{T}_{n_1,n_2}/\sigma_{02}) = a^2 + b^2 + o(1)$.

Chen and Qin (2010) and Li and Chen (2012) proved that

$$\frac{1}{\sigma_{01}^2}\sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}^2) \xrightarrow{p} 1, \quad \frac{1}{\sigma_{02}^2}\sum_{m=1}^{n_1+n_2} E_{m-1}(G_{n,m}^2) \xrightarrow{p} 1.$$

and

$$\frac{1}{\sigma_{01}^4}\sum_{m=1}^{n_1+n_2} E(D_{n,m}^4) \to 0, \quad \frac{1}{\sigma_{02}^4}\sum_{m=1}^{n_1+n_2} E(G_{n,m}^4) \to 0.$$

Therefore,

$$\frac{1}{s_n^4}\sum_{m=1}^{n_1+n_2} E\left(\frac{aD_{n,m}}{\sigma_{01}} + \frac{bG_{n,m}}{\sigma_{02}}\right)^4 \leqslant \frac{8}{s_n^4}\left(\frac{a^4}{\sigma_{01}^4}\sum_{m=1}^{n_1+n_2} E(D_{n,m}^4) + \frac{b^4}{\sigma_{02}^4}\sum_{m=1}^{n_1+n_2} E(G_{n,m}^4)\right) \to 0,$$

(5.7.30) is proved. In order to prove (5.7.29), we only need to show

$$\frac{1}{\sigma_{01}\sigma_{02}}\sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}G_{n,m}) \xrightarrow{p} 0. \tag{5.7.31}$$

Rigorous calculation suggests for a sufficiently large positive constant C, we have

$$\begin{aligned}\left|E\left(\sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}G_{n,m})\right)\right| &= \left|\sum_{m=1}^{n_1+n_2} E\left(E_{m-1}(D_{n,m}G_{n,m})\right)\right| \\ &\leqslant C\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^2 \lambda_{\max}(\mathbf{\Sigma})\text{tr}(\mathbf{\Sigma}^2),\end{aligned}$$

and

$$E\left(\sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}G_{n,m})\right)^2 \leqslant (n_1+n_2)\sum_{m=1}^{n_1+n_2} E\left(E_{m-1}^2(D_{n,m}G_{n,m})\right) \leqslant \frac{C\lambda_{\max}^2(\mathbf{\Sigma})\text{tr}^2(\mathbf{\Sigma}^2)}{(n_1+n_2)^4}.$$

Therefore,

$$E\left(\sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}G_{n,m})\right) = o\left(\sigma_{01}\sigma_{02}\right), \text{var}\left(\sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}G_{n,m})\right) = o\left(\sigma_{01}^2\sigma_{02}^2\right).$$

Hence (5.7.31) holds, which gives (5.7.29). Further by the martingale central limit theorem (Hall and Heyde, 2014), we have (5.7.28). Together with (5.7.27), we finish the proof of Theorem 5.4. □

5.7.6 Proof of Theorem 5.5

Proof. (i) (Asymptotically accurate size) Without loss of generality, we assume the common mean vector under the null hypothesis H_0 equals to zero. As discussed by Li and Chen (2012), the third and fourth-moment summation terms in A_{n_1}, B_{n_2} and C_{n_1,n_2} are all of small order than the leading second-moment terms. As a result, after centering each datum, removing those terms from T_{n_1,n_2} would not affect its asymptotic behaviors. It brings us a lot of convenience for theoretical analysis and greatly alleviates the computational burden at the same time. We consider the simplified T_{n_1,n_2} statistic.

$$\widetilde{T}_{n_1,n_2} = \frac{1}{n_1(n_1-1)} \sum_{u \neq v}^{n_1} \left(\mathbf{X}_u' \mathbf{X}_v\right)^2 + \frac{1}{n_2(n_2-1)} \sum_{u \neq v}^{n_2} \left(\mathbf{Y}_u' \mathbf{Y}_v\right)^2 - \frac{2}{n_1 n_2} \sum_{u}^{n_1} \sum_{v}^{n_2} \left(\mathbf{X}_u' \mathbf{Y}_v\right)^2 . \tag{5.7.32}$$

Lemma 5.1 presents a crucial intermediate result that the statistics M_{n_1,n_2} and $\widetilde{T}_{n_1,n_2}$ are asymptotically uncorrelated. As an implication of Lemma 5.1 and further by the martingale central limit theorem (Hall and Heyde, 2014), we are able to obtain the asymptotic joint null distribution of M_{n_1,n_2} and T_{n_1,n_2}. In combination with Theorems 5.1 and 5.2, the joint limiting null distribution of M_{PE} and T_{PE} is obtained and summarized in Theorem 5.4. Then the asymptotically accurate size of the simultaneous test J_{n_1,n_2} directly follows from the asymptotic independence of Theorem 5.4.

(ii) (Asymptotically consistent power) Note that

$$\begin{aligned} &\inf_{\{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s\}\cup\{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s\}} P\left(J_{n_1,n_2} \geqslant q_\alpha\right) \\ &\geqslant \min\left\{\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s} P\left(J_{n_1,n_2} \geqslant q_\alpha\right), \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s} P\left(J_{n_1,n_2} \geqslant q_\alpha\right)\right\}. \end{aligned}$$

It suffices to show

$$\inf_{\mathcal{G}_k^d\cup\mathcal{G}_k^s} P\left(J_{n_1,n_2} \geqslant q_\alpha\right) \to 1, \tag{5.7.33}$$

for $k = c, m$. Since both $-2\log\left(1-\Phi(T_{PE})\right)$ and $-2\log\left(1-\Phi(M_{PE})\right)$ are always non-negative, we have

$$\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s} P\left(J_{n_1,n_2}\geqslant q_\alpha\right)\geqslant \inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s} P\left(T_{PE}\geqslant \Phi^{-1}\left(1-\exp(-q_\alpha/2)\right)\right),$$

and

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s} P\left(J_{n_1,n_2}\geqslant q_\alpha\right)\geqslant \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s} P\left(M_{PE}\geqslant \Phi^{-1}\left(1-\exp(-q_\alpha/2)\right)\right).$$

Together with the power analysis in Theorem 5.3, (5.7.33) is proved. □

Chapter 6

Summary and Future Work

6.1 Summary

This dissertation focuses on power enhancement (PE) in high-dimensional hypothesis tests with applications to financial econometrics and statistical genetics. We propose power-enhanced tests for various research questions, including one-sample and two-sample tests on one single parameter of interest, and hypothesis tests on multiple parameters. Two PE techniques via (i) constructing PE components, and (ii) aggregating information by the combination of p-values, are presented to boost testing power towards more general alternative, where the term of "more general alternatives" are interpreted in two ways:

(a) expanding high-power regions to a wider alternative space with respect to one parameter. To be more specific, in Chapters 3, 4 and 5.3.1, we propose power-enhanced tests on each parameter of interest to enlarge their respective high-power regions to the union of sparse and dense alternative spaces.

(b) extending the test capability to alternative spaces with respect to more parameters. In detail, in Chapter 5.3.3, we develop a power-enhanced simultaneous test to enhance the test capability to jointly testing mean vectors and covariance matrices.

It is worth pointing out our proposed tests are scale-invariant, hyperparameter-free, and computationally efficient. The proposed tests are proven to achieve desired PE properties following the guidance of the three general PE principles.

6.2 Future Work

There are many interesting directions for future work. First, from a methodological perspective, in this dissertation, the aggregation of information from multiple statistics largely rely on the (asymptotic) independence among the components. This limits the scope of its applicability in generalizing the method to solving other problems. It is of great importance to develop a framework that accommodates dependency among the components for combination. The recently proposed Cauchy combination (Liu and Xie, 2020) pioneers the exploration to this research area. We are eagerly pursuing the development of new methodologies for combination of dependent components, and extending the research of high-dimensional power enhancement tests upon the dependent combination.

Another direction is to extend the power enhancement methodology to other research areas. In this dissertation, we have covered the test for coefficients in a linear model, and tests on mean vectors and covariances matrices of i.i.d. samples. The power enhancement can be regarded as a technique in hypothesis testing. Such technique can be extended to other research areas like testing for functional data or testing in network and graphical models. The inherent complexity of data structure in such data type brings fundamental challenges in analyzing the asymptotic behavior of test statistics, leading to non-trivial but promising extensions to construct the power-enhanced tests with diverse types of data.

Last but not least, from an application perspective, the power-enhanced tests can be applied to solving other scientific problems. As an example, in this dissertation we demonstrate the test efficacy using empirical studies on testing market efficiency using asset returns of the Russel-2000 portfolio, and identifying differentially expressed gene-sets among various types of tumors. In fact, many research questions can be converted into hypothesis testing problems, for example, the discovery of association in microbiome research, the evaluation of indirect effects for genome-wide mediation analysis, the detection of activation in fMRI data. A power-enhanced test can help researchers find more useful information, providing helpful guidance on scientific experiments.

Bibliography

Anderson, T. (2003), ‘An introduction to multivariate statistical analysis’, *Wiley Series in Probability and Statistics* .

Antoniadis, A. (1997), ‘Wavelets in statistics: a review’, *Journal of the Italian Statistical Society* **6**(2), 97–130.

Arcones, M. A. and Gine, E. (1993), ‘Limit theorems for U-processes’, *The Annals of Probability* **21**(3), 1494–1542.

Arias-Castro, E., Candès, E. J. and Plan, Y. (2011), ‘Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism’, *The Annals of Statistics* **39**(5), 2533–2556.

Arratia, R., Goldstein, L. and Gordon, L. (1989), ‘Two moments suffice for poisson approximations: the chen-stein method’, *The Annals of Probability* **17**(1), 9–25.

Bahadur, R. R. (1967), ‘Rates of convergence of estimates and test statistics’, *The Annals of Mathematical Statistics* **38**(2), 303–324.

Bai, Z. D. and Silverstein, J. W. (2004), ‘CLT for linear spectral statistics of large-dimensional sample covariance matrices’, *The Annals of Probability* **32**(1), 553–605.

Bai, Z., Jiang, D., Yao, J.-F. and Zheng, S. (2009), ‘Corrections to LRT on large-dimensional covariance matrix by RMT’, *The Annals of Statistics* **37**(6B), 3822–3840.

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

Bai, Z. and Silverstein, J. W. (2010), *Spectral Analysis of Large Dimensional Random Matrices*, Vol. 20, Springer.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2007), 'Multivariate tests of mean–variance efficiency with possibly non-gaussian errors: An exact simulation-based approach', *Journal of Business & Economic Statistics* **25**(4), 398–410.

Bellec, P. C. (2019), 'Concentration of quadratic forms under a Bernstein moment assumption', *arXiv preprint arXiv:1901.08736* .

Benjamini, Y. and Hochberg, Y. (1995), 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society. Series B (Methodological)* **57**(1), 289–300.

Berman, S. M. (1962), 'Limiting distribution of the maximum term in sequences of dependent random variables', *The Annals of Mathematical Statistics* **33**(3), 894–908.

Berman, S. M. (1964), 'Limit theorems for the maximum term in stationary sequences', *The Annals of Mathematical Statistics* **35**(2), 502–516.

Borell, C. (1979), 'On the integrability of banach space valued Walsh polynomials', *Lecture Notes in Mathematics* **721**, 1–3.

Cai, T., Liu, W. and Luo, X. (2011), 'A constrained ℓ_1 minimization approach to sparse precision matrix estimation', *Journal of the American Statistical Association* **106**(494), 594–607.

Cai, T., Liu, W. and Xia, Y. (2013), 'Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings', *Journal of the American Statistical Association* **108**(501), 265–277.

Cai, T., Liu, W. and Xia, Y. (2014), 'Two-sample test of high dimensional means under dependence', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(2), 349–372.

Canale, S., Cocco, C., Frasson, C., Seganfreddo, E., Di Carlo, E., Ognio, E., Sorrentino, C., Ribatti, D., Zorzoli, A. and Basso, G. (2011), 'Interleukin-27 inhibits pediatric B-acute lymphoblastic leukemia cell spreading in a preclinical model', *Leukemia* **25**(12), 1815–1824.

Chamberlain, G. (1983), 'Funds, factors, and diversification in arbitrage pricing models', *Econometrica: Journal of the Econometric Society* **51**(5), 1305–1323.

Chang, J., Zhou, W., Zhou, W.-X. and Wang, L. (2017), 'Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering', *Biometrics* **73**(1), 31–41.

Chavali, S., Bruhn, S., Tiemann, K., Sætrom, P., Barrenäs, F., Saito, T., Kanduri, K., Wang, H. and Benson, M. (2013), 'MicroRNAs act complementarily to regulate disease-related mRNA modules in human diseases', *Rna* **19**(11), 1552–1562.

Chen, S., Ma, S., Xue, L. and Zou, H. (2020), 'An alternating manifold proximal gradient method for sparse principal component analysis and sparse canonical correlation analysis', *Informs Journal on Optimization* **2**(3), 192–208.

Chen, S. X., Guo, B. and Qiu, Y. (2019), 'Multi-level thresholding test for high dimensional covariance matrices', *arXiv preprint arXiv:1910.13074* .

Chen, S. X., Li, J. and Zhong, P.-S. (2019), 'Two-sample and ANOVA tests for high dimensional means', *The Annals of Statistics* **47**(3), 1443–1474.

Chen, S. X. and Qin, Y.-L. (2010), 'A two-sample test for high-dimensional data with applications to gene-set testing', *The Annals of Statistics* **38**(2), 808–835.

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

Chen, Z., Fan, J. and Li, R. (2018), 'Error variance estimation in ultrahigh-dimensional additive models', *Journal of the American Statistical Association* **113**(521), 315–327.

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

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014), 'Gaussian approximation of suprema of empirical processes', *Annals of Statistics* **42**(4), 1564–1597.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2019), 'Inference on causal and structural parameters using many moment inequalities', *The Review of Economic Studies* **86**(5), 1867–1900.

Chiaretti, S., Li, X., Gentleman, R., Vitale, A., Vignetti, M., Mandelli, F., Ritz, J. and Foa, R. (2004), 'Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival', *Blood* **103**(7), 2771–2778.

Chow, T. L. and Teugels, J. L. (1978), 'The sum and the maximum of iid random variables', *Proceedings of the 2nd Prague Symposium on Asymptotic Statistics* pp. 81–92.

Clarke, R., Ressom, H. W., Wang, A., Xuan, J., Liu, M. C., Gehan, E. A. and Wang, Y. (2008), 'The properties of high-dimensional data spaces: implications for exploring gene and protein expression data', *Nature reviews cancer* **8**(1), 37–49.

Cummings, J., Lee, G., Ritter, A., Sabbagh, M. and Zhong, K. (2019), 'Alzheimer's disease drug development pipeline: 2019', *Alzheimer's & Dementia: Translational Research & Clinical Interventions* **5**, 272–293.

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

Dudoit, S., Keles, S. and van der Laan, M. J. (2008), 'Multiple tests of association with biological annotation metadata', *Probability and Statistics: Essays in Honor of David A. Freedman* **2**, 153–218.

Dudoit, S. and Van Der Laan, M. J. (2007), *Multiple testing procedures with applications to genomics*, Springer Science & Business Media.

Dykstra, R. L. (1970), 'Establishing the positive definiteness of the sample covariance matrix', *The Annals of Mathematical Statistics* **41**(6), 2153–2154.

Efron, B. and Tibshirani, R. (2007), 'On testing the significance of sets of genes', *The Annals of Applied Statistics* **1**(1), 107–129.

Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A. and Tress, M. L. (2014), 'Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes', *Human Molecular Genetics* **23**(22), 5866–5878.

Fama, E. F. and French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

Fan, J. (1996), 'Test of significance based on wavelet thresholding and Neyman's truncation', *Journal of the American Statistical Association* **91**(434), 674–688.

Fan, J. and Li, R. (2001), 'Variable selection via nonconcave penalized likelihood and its oracle properties', *Journal of the American Statistical Association* **96**(456), 1348–1360.

Fan, J., Liao, Y. and Mincheva, M. (2011), 'High dimensional covariance matrix estimation in approximate factor models', *The Annals of Statistics* **39**(6), 3320–3356.

Fan, J., Liao, Y. and Yao, J. (2015), 'Power enhancement in high-dimensional cross-sectional tests', *Econometrica* **83**(4), 1497–1541.

Fan, J., Xue, L. and Yao, J. (2017), 'Sufficient forecasting using factor models', *Journal of Econometrics* **201**(2), 292–306.

Fan, J., Xue, L. and Zou, H. (2016), 'Multitask quantile regression under the transnormal model', *Journal of the American Statistical Association* **111**(516), 1726–1735.

Fisher, R. A. (1925), *Statistical Methods for Research Workers*, Vol. 1, Edinburgh by Oliver and Boyd.

Gibbons, M. R., Ross, S. A. and Shanken, J. (1989), 'A test of the efficiency of a given portfolio', *Econometrica: Journal of the Econometric Society* **57**(5), 1121–1152.

Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S. and Kolaczyk, E. D. (2017), 'Hypothesis testing for network data in functional neuroimaging', *The Annals of Applied Statistics* **11**(2), 725–750.

Goeman, J. J. and Bühlmann, P. (2007), 'Analyzing gene expression data in terms of gene sets: methodological issues', *Bioinformatics* **23**(8), 980–987.

Goldfarb, D. and Iyengar, G. (2003), 'Robust portfolio selection problems', *Mathematics of Operations Research* **28**(1), 1–38.

Gravina, G. L., Senapedis, W., McCauley, D., Baloglu, E., Shacham, S. and Festuccia, C. (2014), 'Nucleo-cytoplasmic transport as a therapeutic target of cancer', *Journal of hematology & oncology* **7**(1), 85.

Gungor, S. and Luger, R. (2013), 'Testing linear factor pricing models with large cross sections: A distribution-free approach', *Journal of Business & Economic Statistics* **31**(1), 66–77.

Gungor, S. and Luger, R. (2016), 'Multivariate tests of mean-variance efficiency and spanning with a large number of assets and time-varying covariances', *Journal of Business & Economic Statistics* **34**(2), 161–175.

Haapaniemi, E. M., Kaustio, M., Rajala, H. L., van Adrichem, A. J., Kainulainen, L., Glumoff, V., Doffinger, R., Kuusanmäki, H., Heiskanen-Kosma, T. and Trotta, L. (2015), 'Autoimmunity, hypogammaglobulinemia, lymphoproliferation, and mycobacterial disease in patients with activating mutations in STAT3', *Blood* **125**(4), 639–648.

Haeusler, E. (1988), 'On the rate of convergence in the central limit theorem for martingales with discrete and continuous time', *The Annals of Probability* **16**(1), 275–299.

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

Hall, P. and Jin, J. (2010), 'Innovated higher criticism for detecting sparse signals in correlated noise', *The Annals of Statistics* **38**(3), 1686–1732.

He, Y., Xu, G., Wu, C. and Pan, W. (2021), 'Asymptotically independent U-statistics in high-dimensional testing', *The Annals of Statistics, to appear* .

Hedges, L. V. and Olkin, I. (2014), *Statistical Methods for Meta-Analysis*, Academic press.

Ho, H. C. and Hsing, T. (1996), 'On the asymptotic joint distribution of the sum and maximum of stationary normal random variables', *Journal of Applied Probability* **33**(1), 138–145.

Hodge, D. R., Hurt, E. M. and Farrar, W. L. (2005), 'The role of IL-6 and STAT3 in inflammation and cancer', *European Journal of Cancer* **41**(16), 2502–2512.

Hoeffding, W. (1963), 'Probability inequalities for sums of bounded random variables', *Journal of the American Statistical Association* **58**(301), 13–30.

Hsing, T. (1995), 'A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables', *The Annals of Probability* **23**(2), 938–947.

Huang, Y. (2015), 'Projection test for high-dimensional mean vectors with optimal direction', *Department of Statistics, The Pennsylvania State University at University Park* .

Hyodo, M. and Nishiyama, T. (2018), 'A simultaneous testing of the mean vector and the covariance matrix among two populations for high-dimensional data', *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research* **27**(3), 680–699.

Jensen, M. C. (1968), 'The performance of mutual funds in the period 1945–1964', *The Journal of Finance* **23**(2), 389–416.

Jerez, A., Clemente, M. J., Makishima, H., Koskela, H., LeBlanc, F., Ng, K. P., Olson, T., Przychodzen, B., Afable, M. and Gomez-Segui, I. (2012), 'STAT3 mutations unify the pathogenesis of chronic lymphoproliferative disorders of NK cells and T-cell large granular lymphocyte leukemia', *Blood* **120**(15), 3048–3057.

Jiang, D., Jiang, T. and Yang, F. (2012), 'Likelihood ratio tests for covariance matrices of high-dimensional normal distributions', *Journal of Statistical Planning and Inference* **142**(8), 2241–2256.

Jiang, T. and Yang, F. (2013), 'Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions', *The Annals of Statistics* **41**(4), 2029–2074.

Ke, Z. T., Xue, L. and Yang, F. (2020), 'Diagonally dominant principal component analysis', *Journal of Computational and Graphical Statistics* **29**(3), 592–607.

Li, C. and Li, R. (2021), 'Linear hypothesis testing in linear models with high dimensional responses', *Journal of the American Statistical Association, to appear* .

Li, D. and Xue, L. (2015), 'Joint limiting laws for high-dimensional independence tests', *arXiv preprint arXiv:1512.08819* .

Li, D., Xue, L. and Zou, H. (2018), 'Applications of Peter Hall's martingale limit theory to estimating and testing high dimensional covariance matrices', *Statistica Sinica* **28**(4), 2657–2670.

Li, J. and Chen, S. X. (2012), 'Two sample tests for high-dimensional covariance matrices', *The Annals of Statistics* **40**(2), 908–940.

Littell, R. C. and Folks, J. L. (1971), 'Asymptotic optimality of Fisher's method of combining independent tests', *Journal of the American Statistical Association* **66**(336), 802–806.

Littell, R. C. and Folks, J. L. (1973), 'Asymptotic optimality of Fisher's method of combining independent tests II', *Journal of the American Statistical Association* **68**(341), 193–194.

Liu, H., Han, F., Yuan, M., Lafferty, J. and Wasserman, L. (2012), 'High-dimensional semiparametric Gaussian copula graphical models', *The Annals of Statistics* **40**(4), 2293–2326.

Liu, W.-D., Lin, Z. and Shao, Q.-M. (2008), 'The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization', *The Annals of Applied Probability* **18**(6), 2337–2366.

Liu, Y., Chen, S., Li, Z., Morrison, A. C., Boerwinkle, E. and Lin, X. (2019), 'ACAT: A fast and powerful p value combination method for rare-variant analysis in sequencing studies', *The American Journal of Human Genetics* **104**(3), 410–421.

Liu, Y. and Xie, J. (2020), 'Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures', *Journal of the American Statistical Association* **115**(529), 393–402.

Liu, Z., Liu, B., Zheng, S. and Shi, N.-Z. (2017), 'Simultaneous testing of mean vector and covariance matrix for high-dimensional data', *Journal of Statistical Planning and Inference* **188**, 82–93.

Loh, P.-L. and Wainwright, M. J. (2017), 'Support recovery without incoherence: A case for nonconvex regularization', *The Annals of Statistics* **45**(6), 2455–2482.

Luo, W., Xue, L., Yao, J. and Yu, X. (2021), 'Inverse moment methods for sufficient forecasting using high-dimensional predictors', *arXiv preprint arXiv:1705.00395v2* .

MacKinlay, A. C. and Richardson, M. P. (1991), 'Using generalized method of moments to test mean-variance efficiency', *The Journal of Finance* **46**(2), 511–527.

Milner, J. D., Vogel, T. P., Forbes, L., Ma, C. A., Stray-Pedersen, A., Niemela, J. E., Lyons, J. J., Engelhardt, K. R., Zhang, Y. and Topcagic, N. (2015), 'Early-onset lymphoproliferation and autoimmunity caused by germline STAT3 gain-of-function mutations', *Blood* **125**(4), 591–599.

Niu, Z., Hu, J., Bai, Z. and Gao, W. (2019), 'On LR simultaneous test of high-dimensional mean vector and covariance matrix under non-normality', *Statistics & Probability Letters* **145**, 338–344.

Okada, H. and Pollack, I. F. (2004), 'Cytokine gene therapy for malignant glioma', *Expert opinion on biological therapy* **4**(10), 1609–1620.

Paietta, E., Racevskis, J., Neuberg, D., Rowe, J., Goldstone, A. and Wiernik, P. (1997), 'Expression of CD25 (interleukin-2 receptor α chain) in adult acute lymphoblastic leukemia predicts for the presence of BCR/ABL fusion transcripts: results of a preliminary laboratory analysis of ECOG/MRC intergroup study E2993', *Leukemia* **11**(11), 1887–1890.

Perlman, M. D. (1980), 'Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations', *The Annals of Statistics* **8**(2), 247–263.

Pesaran, M. H. and Yamagata, T. (2012), 'Testing CAPM with a large number of assets', *March 2012, IZA Discussion Paper No.9469* .

Pesaran, M. H. and Yamagata, T. (2017), 'Testing for alpha in linear factor pricing models with a large number of securities', *USC-INET Research Paper No.17-13* .

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K. (2015), 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Research* **43**(7), e47–e47.

Ross, S. A. (1976), 'The arbitrage theory of capital asset pricing', *Journal of Economic Theory* **13**, 341–360.

Schott, J. R. (2007), 'A test for the equality of covariance matrices when the dimension is large relative to the sample sizes', *Computational Statistics & Data Analysis* **51**(12), 6535–6542.

Shi, C., Song, R., Chen, Z. and Li, R. (2019), 'Linear hypothesis testing for high dimensional generalized linear models', *The Annals of Statistics* **47**(5), 2671–2703.

Singh, K., Xie, M., Strawderman, W. E. et al. (2005), 'Combining information from independent sources through confidence distributions', *The Annals of Statistics* **33**(1), 159–183.

Srivastava, M. S. and Yanagihara, H. (2010), 'Testing the equality of several covariance matrices with fewer observations than the dimension', *Journal of Multivariate Analysis* **101**(6), 1319–1329.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A. and Williams Jr, R. M. (1949), *The American Soldier: Adjustment during Army Life*, Princeton Univ. Press.

Sugiura, N. and Nagao, H. (1968), 'Unbiasedness of some test criteria for the equality of one or two covariance matrices', *The Annals of Mathematical Statistics* **39**(5), 1686–1692.

Takeda, A. and Yaseen, N. R. (2014), Nucleoporins and nucleocytoplasmic transport in hematologic malignancies, *in* 'Seminars in cancer biology', Vol. 27, Elsevier, pp. 3–10.

Tippett, L. H. C. (1931), *The Methods of Statistics: An Introduction Mainly for Experimentalists*, London: Williams & Norgate Ltd.

Touw, I., Pouwels, K., Van Agthoven, T., Van Gurp, R., Budel, L., Hoogerbrugge, H., Delwel, R., Goodwin, R., Namen, A. and Lowenberg, B. (1990), 'Interleukin-7 is a growth factor of precursor B and T acute lymphoblastic leukemia', *Blood* **75**(11), 2097–2101.

Vershynin, R. (2011), 'A simple decoupling inequality in probability theory', *preprint* .

Wang, L., Peng, B. and Li, R. (2015), 'A high-dimensional nonparametric multivariate test for mean vector', *Journal of the American Statistical Association* **110**(512), 1658–1669.

Xie, M., Singh, K. and Strawderman, W. E. (2011), 'Confidence distributions and a unifying framework for meta-analysis', *Journal of the American Statistical Association* **106**(493), 320–333.

Xue, K. and Yao, F. (2020), 'Distribution and correlation-free two-sample test of high-dimensional means', *Annals of Statistics* **48**(3), 1304–1328.

Xue, L., Ma, S. and Zou, H. (2012), 'Positive-definite ℓ_1-penalized estimation of large covariance matrices', *Journal of the American Statistical Association* **107**(500), 1480–1491.

Xue, L. and Zou, H. (2012), 'Regularized rank-based estimation of high-dimensional nonparanormal graphical models', *The Annals of Statistics* **40**(5), 2541–2571.

Xue, L. and Zou, H. (2014*a*), 'Optimal estimation of sparse correlation matrices of semiparametric gaussian copulas', *Statistics and its Interface* **7**(2), 201–209.

Xue, L. and Zou, H. (2014*b*), 'Rank-based tapering estimation of bandable correlation matrices', *Statistica Sinica* **24**(1), 83–100.

Yan, Y., Qiu, S., Jin, Z., Gong, S., Bai, Y., Lu, J. and Yu, T. (2017), 'Detecting subnetwork-level dynamic correlations', *Bioinformatics* **33**(2), 256–265.

Yang, Q. and Pan, G. (2017), 'Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices', *Journal of the American Statistical Association* **112**(517), 188–200.

Yoda, A., Yoda, Y., Chiaretti, S., Bar-Natan, M., Mani, K., Rodig, S. J., West, N., Xiao, Y., Brown, J. R. and Mitsiades, C. (2010), 'Functional screening identifies CRLF2 in precursor B-cell acute lymphoblastic leukemia', *Proceedings of the National Academy of Sciences* **107**(1), 252–257.

Yu, X., Li, D. and Xue, L. (2020), 'Fisher's combined probability test for high-dimensional covariance matrices', *arXiv preprint arXiv:2006.00426* .

Yu, X., Yao, J. and Xue, L. (2019), 'Innovated power enhancement for testing multi-factor asset pricing models', *Available at SSRN 3809369* .

Yu, X., Yao, J. and Xue, L. (2021), 'Nonparametric estimation and conformal inference of the sufficient forecasting with a diverging number of factors', *Journal of Business & Economic Statistics, to appear* .

Zaitsev, A. Y. (1987), 'On the gaussian approximation of convolutions under multidimensional analogues of S.N. Bernstein's inequality conditions', *Probability Theory and Related Fields* **74**(4), 535–566.

Zhang, C.-H. (2010), 'Nearly unbiased variable selection under minimax concave penalty', *The Annals of Statistics* **38**(2), 894–942.

Zheng, S. (2012), 'Central limit theorems for linear spectral statistics of large dimensional F-matrices', *Annales de l'IHP Probabilités et statistiques* **48**(2), 444–476.

Zhong, P.-S., Chen, S. X. and Xu, M. (2013), 'Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence', *The Annals of Statistics* **41**(6), 2820–2851.

Zhu, L., Lei, J., Devlin, B. and Roeder, K. (2017), 'Testing high-dimensional covariance matrices, with application to detecting schizophrenia risk genes', *The Annals of Applied Statistics* **11**(3), 1810.

Zou, H. and Xue, L. (2018), 'A selective overview of sparse principal component analysis', *Proceedings of the IEEE* **106**(8), 1311–1320.

Vita

Xiufan Yu

Education

Ph.D. in Statistics, Pennsylvania State University	2016-2021
M.S. in Statistics, Pennsylvania State University	2016-2019
B.S. in Statistics, University of Science and Technology of China	2012-2016