Statistical inference for high-dimensional data

by

Yingli Qin

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Song Xi Chen, Major Professor
Jack C. M. Dekkers
Kenneth J. Koehler
Dan Nettleton
Dan Nordman

Iowa State University

Ames, Iowa

2009

DEDICATION

To my family

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

CHAPTER 1.   Introduction

High-dimensional data, where the number of variables p is large compared to the sample size n, are widely available from microarray studies, finance and many other sources. This dissertation focuses on the effects of high dimensionality on some aspects of statistical inference. A two-sample test for means of high-dimensional data proposed in this dissertation allows p to be much larger than n. We will show that in the simulation study the proposed test statistic performed consistently better than the other existing methods. Two distributions sharing the same mean may differ in many other aspects. We therefore consider a two-sample test for high-dimensional distributions. The proposed test statistic is based on empirical distribution functions and is a natural extension to our two-sample test statistic for means. Empirical likelihood has many important applications in nonparametric or semiparametric statistical inference. In this dissertation, we further study the effects of data dimension on the asymptotic normality of the empirical likelihood ratio for high-dimensional data under a general multivariate model.

In the remaining of Chapter 1, we shall briefly review some issues arising from high-dimensional data analysis and other topics relevant and important to the work in this dissertation. More detailed literature reviews will be available in each chapter.

1.1   Hypothesis Testing in Microarray Data Analysis

Microarray technology plays a key role in molecular biology and in medicine for discovering certain diseases and developing new drugs. It is very common that microarray data contain gene expression values measured on thousands of genes from much fewer biological objects. In order to detect a small proportion of differentially expressed genes across different treatment groups given very few biological observations, statisticians are in urgent need to develop powerful and approriate multiple testing procedures to maintain the level of certain type of error rate. For comparison between numerous tests used to identify differentially expressed genes, see Jeffery et al. (2006) and many other references. The most commonly used multiple testing procedures include the Bonferroni procedure which controls the family wise error rate (FWER) and the false discovery rate (FDR) approach proposed in Benjamini and Hochberg (1995). Also see the q-value method as an FDR-based measure of significance for genomewide studies in Storey and Tibshirani (2003).

Biologically speaking, each gene does not function individually in isolation. Rather, one gene tends to work with other genes to achieve certain biological tasks. The recent development of the Gene Ontology Consortium (Ashburner et al., 2000) allows researchers to carry out statistical inference for well defined gene sets. The Gene Ontology Consortium provides a vocabulary of defined terms representing gene sets. Identifying sets of genes which are differentially expressed with respect to certain treatments is a recent development in genetics research; see Gene Set Enrichment Analysis in Subramanian et al. (2005) and Significance Analysis of Function and Expression proposed in Barry et al. (2005), Efron and Tibshirani (2007), Newton et al. (2007) and Nettleton et al. (2008).

The original motivation of my work on high-dimensional hypothesis testing is to

develop novel statistical methods for identifying gene sets (dimension p can range from a moderate to a very large number) whose expression levels or joint distributions of expression values differ across two treatment groups. However, the methods proposed in this dissertation are widely applicable to high-dimensional data from many other sources.

1.2 High-Dimensional Data Problems

Traditional statistical data analyses are carried out for observations measured on a fixed number of variables. Due to availability of massive data from finance, microarrays, and climatology etc., people are exposed to situations where the number of variables increases dramatically, but, the number of observations increases much more slowly. Especially in most microarray studies, due to budgetary constraints and other experimental restrictions, the number of observations is relatively small compared to thousands of variables measured on each observation. Consequently, many classical statistical inference procedures, which require fixed data dimensions are not suitable anymore.

1.2.1 High-dimensional hypothesis testing for means

To demonstrate some challenges arising from high-dimensional hypothesis testing, let us consider two random samples $X_{i1}, X_{i2}, \cdots, X_{in_i} \in R^p$ for $i = 1$ and 2, which have means $\mu_i = (\mu_{i1}, \mu_{i2}, \cdots, \mu_{ip})'$ and covariance matrices Σ_i. The data, for example, can be gene expression values measured on mRNA microarrays. We therefore have n_i observations for the i-th group; the vector X_{ij} contains gene expression values measured on p genes for the j-th observation in the i-th group. To identify gene sets whose expression levels differ across two groups, we are interested in testing the

hypothesis

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_1 : \mu_1 \neq \mu_2.$$

Traditionally, if data with fixed dimension are normally distributed, people would use the Hotelling's T^2 test which is defined as

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)^T S_n^{-1} (\bar{X}_1 - \bar{X}_2),$$

where $\bar{X}_i$ is the i-th sample mean vector, for $i = 1, 2$ and S_n is the pooled sample covariance matrix defined as

$$S_n = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^T.$$

Then under the null hypothesis, $\frac{n-p+1}{np} T^2$ has a central F-distribution (Anderson, 2003) with p and $n - p + 1$ degrees of freedom where $n = n_1 + n_2 - 2$. Unfortunately, if $p > n$, which happens very commonly with high-dimensional data, the Hotelling's T^2 test is not well defined because the sample covariance matrix becomes singular. The singularity is due to the fact that the dimension p of S_n is larger than its degrees of freedom.

In an important work, Bai and Saranadasa (1996) proposed a test (BS test) to replace the Hotelling's T^2 test. The BS test statistic is based on the statistic

$$M_n = (\bar{X}_1 - \bar{X}_2)^T (\bar{X}_1 - \bar{X}_2) - \tau tr S_n,$$

where $\tau = \frac{n_1 + n_2}{n_1 n_2}$ and $tr(\cdot)$ is the trace operator of a matrix. And the authors assumed a general multivariate model

$$X_{ij} = \Gamma Z_{ij} + \mu_i \quad \text{for } j = 1, \cdots, n_i, \; i = 1 \text{ and } 2, \tag{1.1}$$

where Γ is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma\Gamma' = \Sigma$ and Σ is the common covariance matrix. And $\{Z_{ij}\}_{j=1}^{n_i}$ are m-variate independent and identically distributed

(i.i.d.) random vectors satisfying $E(Z_{ij}) = 0$, $Var(Z_{ij}) = I_m$ (the $m \times m$ identity matrix), and $E(z_{ijk}^4) = 3 + \Delta < \infty$. Here Δ describes the difference between the fourth moments of z_{ijk} and $N(0,1)$. Bai and Saranadasa (1996) further showed that given $p/n \to y > 0$ and some other mild conditions,

$$\frac{M_n}{\sqrt{Var(M_n)}} \xrightarrow{d} N(0,1), \text{ as } n \to \infty.$$

However, by assuming $p/n \to y > 0$, it basically requires p and n to increase to $+\infty$ at the same rate. Consequently, the BS test is not attractive in "large p, small n" case. In Chapter 2, we propose a new test which allows $p >> n$, i.e. p can be much larger than n. The proposed test statistic also has the asymptotic normal property and performed more powerfully than the BS test in our simulation study.

Many other important works have been published on hypothesis testing for means when both p and n go to infinity. Srivastava (2009) proposed a test for mean vectors with fewer observations than the dimension by assuming the same multivariate model as in (1.1). The test proposed by Srivastava can be treated as a standardized Hotelling's T^2 test without using a normality assumption. Schott (2007) considered high-dimensional tests for a one-way MANOVA as a generalization of the BS test statistic. Fan et al. (2007) evaluated approximation of the overall level of significance for simultaneous testing of means. They demonstrated that the bootstrap method can accurately approximate the overall level of significance if $\log p = o(n^{1/3})$ when the marginal tests are performed based on the normal or the t-distributions. See also Fan et al. (2005) and Huang et al. (2005) for high-dimensional estimation and testing in semiparametric regression models.

1.2.2 High-dimensional hypothesis testing for distributions

Let us consider two random samples $X_{i1}, X_{i2}, \cdots, X_{in_i} \in R^p$ for $i = 1$ and 2 drawn independently from multivariate continuous distributions F_1 and F_2, respectively. The hypothesis we are interested in becomes

$$H_0 : F_1 = F_2 \text{ vs. } H_1 : F_1 \neq F_2.$$

In fixed dimension settings, a rank test, multivariate Kolmogrorov-Smirnov test (Peacock, 1983; Fasano and Franceschini, 1987; Bickel, 1969), or a multivariate Cramér-von Mises test (Anderson and Darling, 1952; Ahmad, 1996) may be used. For arbitrary dimensions, Baringhaus and Franz (2004) proposed a test statistic, T_{n_1, n_2}, based on between and within sample Euclidean distances, where

$$T_{n_1, n_2} = \frac{n_1 n_2}{n_1 + n_2} \left[\frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} ||X_{1i} - X_{2j}|| - \frac{1}{2n_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} ||X_{1i} - X_{1j}|| \right.$$
$$\left. - \frac{1}{2n_2^2} \sum_{i=1}^{n_2} \sum_{j=1}^{n_2} ||X_{2i} - X_{2j}|| \right]. (1.2)$$

The test rejects the null hypothesis if T_{n_1, n_2} is large. As the distribution of T_{n_1, n_2} converges to a Brownian bridge which depends on an unknown distribution, the authors suggested a bootstrap method to simulate critical values. Alba Fernández et al. (2008) constructed a test statistic similar to this T_{n_1, n_2}. The test statistic is defined as

$$D_{n_1, n_2} = \frac{1}{n_1^2} \sum_{i,j=1}^{n_1} u(X_{1i} - X_{1j}) + \frac{1}{n_2^2} \sum_{i,j=1}^{n_2} u(X_{2i} - X_{2j}) - \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} u(X_{1i} - X_{2j}), (1.3)$$

where $u(t) = \int cos(x't) dG(x)$, which is the real part of the characteristic function of a distribution function G on R^p. The authors also suggested bootstrap and permutation procedures to estimate its limiting distribution.

Hall and Tajvidi (2002) proposed a permutation test of equal distributions for arbitrarily high-dimensional data. The critical values are determined by using per-

mutation conditional on the pairwise distances between pooled data, which is defined as

$$\mathfrak{L} = \{X_{11}, X_{12} \cdots, X_{1n_1}\} \bigcup \{X_{21}, X_{22} \cdots, X_{2n_2}\}.$$

Note that the measure of pairwise distance has to be symmetric but not necessarily be a metric. Another nonparametric multivariate test motivated by the application of identifying differentially expressed gene sets was investigated in Nettleton et al. (2008). In this dissertation, we propose a test for high-dimensional distributions based on empirical distribution functions. The limiting distribution of the test statistic will be established.

1.3   Empirical Likelihood

Using the empirical likelihood method to construct confidence regions was first introduced in Owen (1988, 1990) for means and some other parameters. It is a nonparametric method of statistical inference that does not require that the data come from a certain known family of distributions. Empirical likelihood is generally applicable to many areas of statistical analysis, see Hall (1990), Hall and Owen (1993), DiCiccio et al. (1989), Chen (1993, 1994a,b, 1996), Chen and Qin (2000), Chen and Cui (2003, 2006), and Qin (1993, 1999), Qin and Jing (2001), Qin and Zhang (2007). For a comprehensive reference, see Owen (2001).

Definition 1.1 *Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. random variables (univariate or multivariate) with cumulative distribution function (cdf) F. Let π_i be a probability (such that $\sum_{i=1}^{n} \pi_i = 1$ and all $\pi_i \geq 0$) assigned to the observed data value X_i. The nonparametric likelihood is*

$$L_n(F, \pi) = \prod_{i=1}^{n} \pi_i.$$

It has been proven (Owen, 2001) that $L_n(F, \pi)$ is maximized by F_n (that is $\pi_i = 1/n$ for $i = 1, 2, \cdots, n$), where F_n is the empirical cumulative distribution function.

We are interested in specifically the mean vector of a multivariate cdf F, say parameter μ. The next two definitions introduce the empirical likelihood and empirical likelihood ratio for the mean, respectively.

Definition 1.2 *The empirical likelihood function for the mean is as*

$$L_n(\mu) = \sup\left\{\prod_{i=1}^{n} \pi_i \Big| \pi_i \geq 0, \sum_{i=1}^{n} \pi_i = 1, \sum_{i=1}^{n} \pi_i X_i = \mu\right\}.$$

Using Lagrange multipliers, Owen (1988) showed that $L_n(\mu)$ is maximized when

$$\pi_i(\mu) = n^{-1}\{1 + \lambda^T(X_i - \mu)\}^{-1}(1 \leq i \leq n),$$

and $\lambda = \lambda(\mu)$ is determined by

$$\sum_{i=1}^{n}\{1 + \lambda^T(X_i - \mu)\}^{-1}(X_i - \mu) = 0.$$

Definition 1.3 *The empirical likelihood ratio for μ is*

$$w_n(\mu) = -2\log\{n^n L_n(\mu)\}. \tag{1.4}$$

When p is fixed, the Wilks' theorem indicates that

$$w_n(\mu) \to \chi_p^2 \tag{1.5}$$

in distribution as $n \to +\infty$.

In this dissertation, we consider the effects of high-dimensionality on the empirical likelihood ratio test. Let $X_1, \ldots, X_n$ be i.i.d. random vectors in R^p with mean vector $\mu = (\mu_1, \ldots, \mu_p)^T$ and non-singular variance matrix $\Sigma \in R^{p \times p}$. As $p \to +\infty$ for high-dimensional data, the natural substitute for (1.5) is

$$(2p)^{-1/2}\{w_n(\mu) - p\} \to N(0, 1) \tag{1.6}$$

in distribution as $n \to +\infty$, since χ_p^2 is asymptotic normal with mean p and variance $2p$. A key question is how large the dimension p can be while (1.6) is valid.

In a recent study, Hjort et al. (2009) established that it is $p = o(n^{1/3})$ together with some other assumptions to ensure (1.6) is valid. We evaluate in this dissertation the effects of data dimension on the asymptotic normality of the empirical likelihood ratio for high-dimensional data under a general multivariate model. Data dimension and dependence among components of the multivariate random vector affect the empirical likelihood directly throught the trace and the eigenvalues of the covariance matrix. The growth rates we obtain for the data dimension improve the rates of Hjort et al. (2009).

In an important study, Tsao (2004) found that, when p is moderately large but fixed, the distribution of $w_n(\mu)$ has an atom at infinity for fixed n: the probability of $w_n(\mu) = \infty$ is non-zero. Tsao further showed that, if p and n increase at the same rate such that $p/n \geq 0.5$, the probability of $w_n(\mu) = \infty$ converges to 1 since the probability of μ being contained in the convex hull of the sample converges to 0. These reveal effects of p on the empirical likelihood from another perspective.

1.4   Martingale Central Limit Theorem

Martingale theory has remarkable applications in economics, game theory, U-statistics, survival analysis and many other areas. The martingale central limit theorem generalizes limit results for sums of independent random variables and plays a key role in probability theory. In this dissertation, we apply the martingale central limit theorem to achieve the asymptotic normality for the tests proposed in Chapter 2, Chapter 3 and Chapter 4. We therefore introduce here the general definition of a martingale, martingale differences, martingale array, and the martingale central limit

theorem.

Definition 1.4 *Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is a set, $\mathcal{F}$ is a σ-field of subset of Ω and P a probability measure defined on $\mathcal{F}$. Let $\{\mathcal{F}_n, n \geq 1\}$ be an increasing sequence of σ-fields of $\mathcal{F}$ sets. Suppose that $\{S_n, n \geq 1\}$ is a sequence of random variables on Ω satisfying*

$$(i) \quad S_n \text{ is measurable with respect to } \mathcal{F}_n;$$

$$(ii) \quad E|S_n| < \infty;$$

$$(iii) \quad E(S_n|\mathcal{F}_m) = S_m \text{ a.s. for } m < n.$$

Then ,the sequence $\{S_n, \mathcal{F}_n\}$ is said to be a martingale and $X_n = S_n - S_{n-1}, n \geq 2$ denote martingale differences.

Definition 1.5 *Let $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n\}$ be a zero-mean and square-integrable martingale for each $n \geq 1$. Denote the martingale differences as $X_{ni} = S_{ni} - S_{n,i-1}, 1 \leq i \leq k_n(S_{n0} = 0)$. It is assumed that $k_n \to \infty$ as $n \to \infty$. Then $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ is called a martingale array.*

We present the martingale central limit theorem (Hall and Heyde, 1980) in the following theorem.

Theorem 1.1 *Let $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ be a zero-mean and square-integrable martingale array with differences X_{ni} and let η^2 be an a.s. finite random variable. Suppose that the $\sigma-$fields are nested: $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n,i+1}$ for $1 \leq i \leq k_n, n \geq 1$ and for all $\varepsilon > 0, \quad \sum_i E[X_{ni}^2 I(|X_{ni}| > \varepsilon)|\mathcal{F}_{n,i-1}] \xrightarrow{P} 0, \text{ and}$*

$$V_{nk_n}^2 = \sum E(X_{ni}^2|\mathcal{F}_{n,i-1}) \xrightarrow{P} \eta^2.$$

Then

$$S_{nk_n} = \sum X_{ni} \xrightarrow{d} Z,$$

where Z is a normally distributed random variable with zero mean and variance η^2.

1.5 Dissertation Organization

This dissertation consists of three main chapters. In Chapter 2, we propose a two-sample test for means when the data dimension is high. The test requires no explicit conditions between sample size n and data dimension p. The proposed test therefore provides great flexibility to carry out hypothesis testing in "large p, small n" situations. The simulation study shows that the proposed test performs more powerfully than other existing tests. A short version of this chapter has been accepted for publication by *The Annals of Statistics*.

In Chapter 3, we evaluate the effects of data dimension on the asymptotic normality of the empirical likelihood ratio for high-dimensional data under a general multivariate model. An abbreviated version of this chapter (Chen et al., 2009) has been published in *Biometrika*.

Chapter 4 focuses on a test of equality of two continuous high-dimensional cdfs. It is shown that the proposed test statistic is a weighted average of distance measures between two continuous cumulative distribution functions. The asymptotic normality of the test statistic is established. This manuscript will be submitted to *The Annals of Applied Statistics*.

CHAPTER 2.   A Two-Sample Test for High-Dimensional Data with Applications to Gene-set Testing

2.1   Introduction

High-dimensional data are increasingly encountered in many applications of statistics and most prominently in biological and financial studies. A common feature of high-dimensional data is that, while the data dimension is high, the sample size is relatively small. This is the so-called "large p, small n" phenomenon where $p/n \to \infty$; here p is the data dimension and n is the sample size. The high data dimension ("large p") alone has created the need to renovate and rewrite some of the conventional multivariate analysis procedures; these needs only get much greater for "large p small n" situations.

A specific "large p, small n" situation arises when simultaneously testing a large number of hypotheses which is largely motivated by the identification of significant genes in microarray and genetic sequence studies. A natural question is how many hypotheses can be tested simultaneously. This chapter tries to answer this question in the context of two-sample simultaneous tests for means. Consider two random samples $X_{i1}, \cdots, X_{in_i} \in R^p$ for $i = 1$ and 2 which have means $\mu_1 = (\mu_{11}, \cdots, \mu_{1p})^T$ and $\mu_2 = (\mu_{21}, \cdots, \mu_{2p})^T$ and covariance matrices Σ_1 and Σ_2, respectively. We consider

testing the following high-dimensional hypothesis:

$$H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_1 : \mu_1 \neq \mu_2. \tag{2.1}$$

The hypothesis H_0 consists of the p marginal hypotheses $H_{0l} : \mu_{1l} = \mu_{2l}$ for $l = 1, \cdots, p$ regarding the means on each data dimension.

There have been a series of important studies on the high-dimensional problem. Van der Laan and Bryan (2001) showed that the sample mean of p-dimensional data can consistently estimate the population mean uniformly across p dimensions if $\log(p) = o(n)$ for bounded random variables. In a major generalization, Kosorok and Ma (2007) considered uniform convergence for a range of univariate statistics constructed for each data dimension, which included the marginal empirical distribution, sample mean and sample median. They established the uniform convergence across p dimensions when $\log(p) = o(n^{1/2})$ or $\log(p) = o(n^{1/3})$ depending on the nature of the marginal statistics. Fan et al. (2007) evaluated approximation of the overall level of significance for simultaneous testing of means. They demonstrated that the bootstrap can accurately approximate the overall level of significance if $\log(p) = o(n^{1/3})$ when the marginal tests are performed based on the normal or the t-distributions. See also Fan et al. (2005) and Huang et al. (2005) for high-dimensional estimation and testing in semiparametric regression models.

In an important work, Bai and Saranadasa (1996) proposed using $||\bar{X}_1 - \bar{X}_2||$ to replace $(\bar{X}_1 - \bar{X}_2)^T S_n^{-1} (\bar{X}_1 - \bar{X}_2)$ in Hotelling's T^2-statistic where $\bar{X}_1$ and $\bar{X}_2$ are the two-sample means, S_n is the pooled sample covariance by assuming $\Sigma_1 = \Sigma_2 = \Sigma$, and $|| \cdot ||$ denotes the Euclidean norm in R^p. They established asymptotic normality of the test statistic and showed that it has attractive power when $p/n \to c < \infty$ and the maximum eigenvalue of Σ is constrained in a suitable way. However, the requirement of p and n being of the same order is too restrictive to be used in the "large p, small

n" situation.

To allow simultaneous testing for ultra high-dimensional data, we construct a test which allows p to be arbitrarily large and independent of the sample size as long as, in the case of common covariance Σ, $tr(\Sigma^4) = o\{tr^2(\Sigma^2)\}$ where $tr(\cdot)$ is the trace operator of a matrix. The above condition on Σ is trivially true for any p if either all the eigenvalues of Σ are bounded or the largest eigenvalue is of smaller order than $(p-b)^{1/2}b^{-1/4}$ where b is the number of unbounded eigenvalues. We establish the asymptotic normality of a test statistic which leads to a two-sample test for high-dimensional data.

Testing significance for gene-sets rather than a single gene is the latest development in genetic data analysis. A critical need for gene-set testing is to have a multivariate test that is applicable to a wide range of data dimensions (the number of genes in a set). It requires P-values for all gene-sets to allow procedures based on either the Bonferroni correction or the false discovery rate (Benjamini and Hochberg, 1995) to take into account the multiplicity in the test. We demonstrate in this chapter how to use the proposed test for testing significance for gene-sets. An advantage of the proposed test is that it readily produces P-values for the significance for each gene-set under study so that the multiplicity of multiple testing can be taken into consideration.

This chapter is organized as follows. We outline in Section 2.2 the framework of the two-sample test for high-dimensional data and introduce the proposed test statistic. Section 2.3 provides the theoretical properties of the test. The application of the proposed test of significance for gene-sets is demonstrated in Section 2.4, which includes an analysis of an Acute Lymphoblastic Leukemia data set. Results of simulation studies are reported in Section 2.5. All the technical details are given in Section 2.6.

2.2 Test Statistic

Suppose we have two independent and identically distributed random samples in R^p,

$$\{X_{i1}, X_{i2}, \cdots, X_{in_i}\} \overset{\text{i.i.d.}}{\sim} F_i \quad \text{for} \quad i = 1 \text{ and } 2,$$

where F_i is a distribution in R^p with mean μ_i and covariance Σ_i. A well-pursued interest in high-dimensional data analysis is to test if the two high-dimensional populations have the same mean or not; namely,

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_1 : \mu_1 \neq \mu_2. \tag{2.2}$$

The above hypothesis consists of p marginal hypotheses regarding the means of each data dimension. An important question from the point view of multiple testing is how many marginal hypotheses can be tested simultaneously. Van der Laan and Bryan (2001), Kosorok and Ma (2007) and Fan et al. (2007) addressed this question. Their results show that p can reach the rate of $e^{\alpha n^\beta}$ for some positive constants α and β. In establishing a rate of the above form, both Van der Laan and Bryan (2001) and Kosorok and Ma (2007) assumed that the marginal distributions of F_1 and F_2 are all supported on bounded intervals.

Hotelling's T^2 test is the conventional test for the above hypothesis when the dimension p is fixed and is less than $n =: n_1 + n_2 - 2$ and when $\Sigma_1 = \Sigma_2 = \Sigma$, say. Its performance for high-dimensional data was evaluated in Bai and Saranadasa (1996) when $p/n \to c \in [0, 1)$, and they report a decreasing power as c gets larger. A reason for the negative effect of high-dimension is the presence of the inverse of the sample covariance matrix in the T^2 statistic. While standardizing by the covariance matrix brings benefits for data with a fixed dimension, it becomes a liability for high-dimensional data. In particular, the sample covariance matrix S_n may not converge

to the population covariance when p and n are of the same order. Indeed, Yin et al. (1988) showed that when $p/n \to c$, the smallest and the largest eigenvalues of the sample covariance S_n do not converge to the respective eigenvalues of Σ. The same phenomenon, but on the weak convergence of the extreme eigenvalues of the sample covariance, is found in Tracy and Widom (1996). When $p > n$, Hotelling's T^2 statistic is not defined as S_n may not be invertible.

Our proposed test is motivated by Bai and Saranadasa (1996), who proposed testing hypothesis (2.2) under $\Sigma_1 = \Sigma_2 = \Sigma$ based on

$$M_n = (\bar{X}_1 - \bar{X}_2)'(\bar{X}_1 - \bar{X}_2) - \tau tr(S_n), \tag{2.3}$$

where $S_n = \frac{1}{n} \sum_{i=1}^{2} \sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$ and $\tau = \frac{n_1 + n_2}{n_1 n_2}$. The key feature of the Bai and Saranadasa proposal is removing S_n^{-1} in Hotelling's T^2 since having S_n^{-1} is no longer beneficial when $p/n \to c > 0$. The subtraction of $tr(S_n)$ in (2.3) makes $E(M_n) = ||\mu_1 - \mu_2||^2$. As

$$
\begin{aligned}
E(M_n) &= \frac{tr(\Sigma)}{n_1} + \mu_1'\mu_1 + \frac{tr(\Sigma)}{n_2} + \mu_2'\mu_2 - 2\mu_1'\mu_2 - \tau tr(S_n) \\
&= \tau tr(\Sigma) + \mu_1'\mu_1 + \mu_2'\mu_2 - 2\mu_1'\mu_2 - \tau \left\{ \frac{(n_1 - 1)tr(\Sigma) + (n_2 - 1)tr(\Sigma)}{n_1 + n_2 - 2} \right\} \\
&= ||\mu_1 - \mu_2||^2.
\end{aligned}
$$

The asymptotic normality of M_n is established and a test statistic is formulated by standardizing M_n with an estimate of its standard deviation.

The following are the main conditions assumed in the Bai-Saranadasa test:

$$p/n \to c < \infty \quad \text{and} \quad \lambda_p = o(p^{1/2}); \tag{2.4}$$

$$n_1/(n_1 + n_2) \to k \in (0, 1) \quad \text{and} \quad (\mu_1 - \mu_2)'\Sigma(\mu_1 - \mu_2) = o\{tr(\Sigma^2)/n\}, \tag{2.5}$$

where λ_p denotes the largest eigenvalue of Σ.

A careful study of the M_n statistic reveals that the restrictions on p and n, and on λ_p in (2.4) are needed to control terms $\sum_{j=1}^{n_i} X'_{ij} X_{ij}$, $i = 1$ and 2, in $||\bar{X}_1 - \bar{X}_2||^2$. However, these two terms are not useful in the test. To appreciate this point, let us consider

$$T_n =: \frac{\sum_{i \neq j}^{n_1} X'_{1i} X_{1j}}{n_1(n_1 - 1)} + \frac{\sum_{i \neq j}^{n_2} X'_{2i} X_{2j}}{n_2(n_2 - 1)} - 2 \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} X'_{1i} X_{2j}}{n_1 n_2}$$

after removing $\sum_{j=1}^{n_i} X'_{ij} X_{ij}$ for $i = 1$ and 2 from $||\bar{X}_1 - \bar{X}_2||^2$. Elementary derivations show that

$$E(T_n) = ||\mu_1 - \mu_2||^2.$$

Hence, T_n is basically all we need for testing. Bai and Saranadasa used $tr(S_n)$ to offset the two diagonal terms. However, $tr(S_n)$ itself imposes demands on the dimensionality too.

A derivation in Section 2.6 shows that under H_1 and similar conditions as the second condition in (2.5),

$$Var(T_n) = \left\{ \frac{2}{n_1(n_1-1)} tr(\Sigma_1^2) + \frac{2}{n_2(n_2-1)} tr(\Sigma_2^2) + \frac{4}{n_1 n_2} tr(\Sigma_1 \Sigma_2) \right\} \{1 + o(1)\},$$

where the $o(1)$ term vanishes under H_0.

2.3   Main Results

2.3.1   Model assumptions

We assume, like Bai and Saranadasa (1996), the following general multivariate model,

$$X_{ij} = \Gamma_i Z_{ij} + \mu_i \quad \text{for } j = 1, \cdots, n_i, \ i = 1 \text{ and } 2, \tag{2.6}$$

where each Γ_i is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma_i \Gamma'_i = \Sigma_i$, and $\{Z_{ij}\}_{j=1}^{n_i}$ are m-variate independent and identically distributed (i.i.d.) random vectors satisfying

$E(Z_{ij}) = 0$, $Var(Z_{ij}) = I_m$, the $m \times m$ identity matrix. Furthermore, if we write $Z_{ij} = (z_{ij1}, \ldots, z_{ijm})'$, we assume $E(z_{ijk}^4) = 3 + \Delta < \infty$, and

$$E\left(z_{ijl_1}^{\alpha_1} z_{ijl_2}^{\alpha_2} \cdots z_{ijl_q}^{\alpha_q}\right) = E(z_{ijl_1}^{\alpha_1})E(z_{ijl_2}^{\alpha_2}) \cdots E(z_{ijl_q}^{\alpha_q}) \tag{2.7}$$

for a positive integer q such that $\sum_{l=1}^{q} \alpha_l \leq 8$ and $l_1 \neq l_2 \neq \cdots \neq l_q$. Here Δ describes the difference between the fourth moments of z_{ijl} and $N(0,1)$. Model (2.6) says that X_{ij} can be expressed as a linear transformation of a m-variate Z_{ij} with zero mean and unit variance that satisfies (2.7). Model (2.6) is similar to factor models in multivariate analysis. However, instead of having the number of factors $m < p$ as in the conventional multivariate analysis, we require $m \geq p$. This is to allow the basic characteristics of the covariance Σ_i, for instance its rank and eigenvalues, to not be affected by the transformation. The rank and eigenvalues could be affected if $m < p$. The fact that m is arbitrary offers much flexibility in generating a rich collection of dependence structures. Condition (2.7) means that each Z_{ij} has a kind of pseudo-independence among its components $\{z_{ijl}\}_{l=1}^{m}$. Obviously, if Z_{ij} does have independent components, then (2.7) is trivially true.

We do not assume $\Sigma_1 = \Sigma_2$, as it is a rather strong assumption, and more importantly such an assumption is difficult to verify for high-dimensional data. Testing certain special structures of the covariance matrix when p and n are of the same order has been considered in Ledoit and Wolf (2002) and Schott (2005).

We assume

$$n_1/(n_1 + n_2) \to k \in (0,1) \quad \text{as} \quad n \to \infty \tag{2.8}$$

$$(\mu_1 - \mu_2)'\Sigma_i(\mu_1 - \mu_2) = o[n^{-1}tr\{(\Sigma_1 + \Sigma_2)^2\}] \quad \text{for } i = 1 \text{ or } 2 \tag{2.9}$$

which generalizes (2.5) to unequal covariances. Condition (2.9) is obviously satisfied under H_0 and implies that the difference between μ_1 and μ_2 is small relative to

$n^{-1}tr\{(\Sigma_1 + \Sigma_2)^2\}$ so that a workable expression for the variance of T_n under H_0 and the specified local alternative can be derived. It can be viewed as a high-dimensional version of the local alternative hypotheses. When p is fixed, if we use a standard test for two population means, for instance Hotelling's T^2 test, the local alternative hypotheses has the form of $\mu_1 - \mu_2 = \tau n^{-1/2}$ for a non-zero constant vector $\tau \in R^p$. Hotelling's test has non-trivial power under such local alternatives (Anderson, 2003). If we assume each component of $\mu_1 - \mu_2$ is the same, say δ, then the local alternatives imply $\delta = O(n^{-1/2})$ for a fixed p. When the difference is $o(n^{-1/2})$, Hotelling's test has no power beyond the level of significance.

To gain insight into (2.9) for high-dimensional situations, let us assume all eigenvalues of Σ_i are bounded above from infinity and below away from zero so that $\Sigma_i = I_p$ is a special case of such a regime. Let us also assume, like above, that each component of $\mu_1 - \mu_2$ is the same as a fixed δ, namely $\mu_{1l} - \mu_{2l} = \delta$ for $l = 1, \cdots, p$. Then (2.9) implies $\delta = o(n^{-1/2})$ which is a smaller order than $\delta = O(n^{-1/2})$ for the fixed p case. This can be understood as the high-dimensional data $(p \to \infty)$ containing more information for differentiating the two mean vectors than that in the fixed p case.

To understand the performance of the test when (2.9) is not valid, we reverse the local alternative condition (2.9) to

$$n^{-1}tr\{(\Sigma_1 + \Sigma_2)^2\} = o\{(\mu_1 - \mu_2)'\Sigma_i(\mu_1 - \mu_2)\} \quad \text{for } i = 1 \text{ or } 2, \qquad (2.10)$$

implying that the Mahanalobis distance between μ_1 and μ_2 is a larger order than that of $n^{-1}tr\{(\Sigma_1 + \Sigma_2)^2\}$. This condition can be viewed as a version of fixed alternatives. We will establish asymptotic normally of T_n under either (2.9) or (2.10) in Theorem 2.1.

The condition we impose on p to replace the first part of (2.4) is

$$tr(\Sigma_i\Sigma_j\Sigma_l\Sigma_h) = o[tr^2\{(\Sigma_1 + \Sigma_2)^2\}] \quad \text{for } i, j, l, h = 1 \text{ or } 2, \qquad (2.11)$$

as $p \to \infty$. To appreciate this condition, consider the case of $\Sigma_1 = \Sigma_2 = \Sigma$. Then (2.11) becomes

$$tr(\Sigma^4) = o\{tr^2(\Sigma^2)\}. \tag{2.12}$$

Let $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_p$ be the eigenvalues of Σ. If all eigenvalues are bounded, then (2.12) is trivially true. If, otherwise, there are b unbounded eigenvalues with respect to p, and the remaining $p - b$ eigenvalues are bounded above by a finite constant M such that $(p - b) \to \infty$ and $(p - b)\lambda_1^2 \to \infty$, and sufficient conditions for (2.12) are

$$\lambda_p = o\{(p - b)^{1/2}\lambda_1 b^{-1/4}\} \quad \text{or} \quad \lambda_p = o\{(p - b)^{1/4}\lambda_1^{1/2}\lambda_{p-b+1}^{1/2}\}, \tag{2.13}$$

where b can be either bounded or diverging to infinity, and the smallest eigenvalue λ_1 can converge to zero. To appreciate this, we note that

$$\frac{tr(\Sigma^4)}{tr^2(\Sigma^2)} \leq \frac{(p - b)M^4 + b\lambda_p^4}{(p - b)^2\lambda_1^4 + b^2\lambda_{p-b+1}^4 + 2(p - b)b\lambda_1^2\lambda_{p-b+1}^2}.$$

Hence, the ratio converges to 0 under either condition in (2.13).

2.3.2 Asymptotic normality of T_n

The following theorem establishes the asymptotic normality of T_n.

Theorem 2.1 *Under the assumptions (2.6), (2.7), (2.8), (2.11) and either (2.9) or (2.10),*

$$\frac{T_n - ||\mu_1 - \mu_2||^2}{\sqrt{Var(T_n)}} \xrightarrow{d} N(0, 1) \quad as \ p \to \infty \ and \ n \to \infty.$$

The asymptotic normality is attained without imposing any explicit restriction between p and n directly. The only restriction on the dimension is (2.11) or (2.12). As the discussion given just before Theorem 2.1 suggests, (2.12) is satisfied provided that

the number of divergent eigenvalues of Σ are not too many, and the divergence is not too fast. The reason for attaining this in the case of high data-dimension is because the statistic T_n is univariate, despite the fact that the hypothesis H_0 is of high-dimension. This is different from using a high-dimensional statistic. Indeed, Portnoy (1986) considered the central limit theorem for the p-dimensional sample mean $\bar{X}$ and found that the central limit theorem is not valid if p is not a smaller order of $\sqrt{n}$.

As shown in Section 2.6, $Var(T_n) = \sigma_n^2\{1 + o(1)\}$ where, under (2.9),

$$\sigma_n^2 =: \sigma_{n1}^2 = \frac{2}{n_1(n_1-1)}tr(\Sigma_1^2) + \frac{2}{n_2(n_2-1)}tr(\Sigma_2^2) + \frac{4}{n_1 n_2}tr(\Sigma_1\Sigma_2), \qquad (2.14)$$

and under (2.10),

$$\sigma_n^2 =: \sigma_{n2}^2 = \frac{4}{n_1}(\mu_1 - \mu_2)'\Sigma_1(\mu_1 - \mu_2) + \frac{4}{n_2}(\mu_1 - \mu_2)'\Sigma_2(\mu_1 - \mu_2). \qquad (2.15)$$

2.3.3 A ratio consistent estimator for $Var(T_n)$

In order to formulate a test procedure based on Theorem 2.1, σ_{n1}^2 in (2.14) needs to be estimated. Bai and Saranadasa (1996) used the following estimators for $tr(\Sigma^2)$ under $\Sigma_1 = \Sigma_2 = \Sigma$:

$$\widehat{tr(\Sigma^2)} = \frac{n^2}{(n+2)(n-1)}\{trS_n^2 - \tfrac{1}{n}(trS_n)^2\}.$$

Motivated by the benefits of excluding terms like $\sum_{j=1}^{n_i} X'_{ij}X_{ij}$ in the formulation of T_n, we propose the following estimator of $tr(\Sigma_i^2)$ and $tr(\Sigma_1\Sigma_2)$:

$$\widehat{tr(\Sigma_i^2)} = \{n_i(n_i - 1)\}^{-1}tr\{\Sigma_{j\neq k}^{n_i}(X_{ij} - \bar{X}_{i(j,k)})X'_{ij}(X_{ik} - \bar{X}_{i(j,k)})X'_{ik}\},$$

and

$$\widehat{tr(\Sigma_1\Sigma_2)} = (n_1 n_2)^{-1}tr\{\Sigma_{l=1}^{n_1}\Sigma_{k=1}^{n_2}(X_{1l} - \bar{X}_{1(l)})X'_{1l}(X_{2k} - \bar{X}_{2(k)})X'_{2k}\},$$

where $\bar{X}_{i(j,k)}$ is the ith sample mean after excluding X_{ij} and X_{ik}, and $\bar{X}_{i(l)}$ is the ith sample mean without X_{il}. These are similar to the idea of cross-validation, in that when we construct the deviations of X_{ij} and X_{ik} from the sample mean, both X_{ij} and X_{ik} are excluded from the sample mean calculation. By doing so, the above estimators $\widehat{tr(\Sigma_i^2)}$ and $\widehat{tr(\Sigma_1\Sigma_2)}$ can be written as the trace of sums of products of independent matrices. We also note that subtraction of only one sample mean per observation is needed in order to avoid a term like $||X_{ij}||^4$ which is harder to control asymptotically without an explicit assumption between p and n.

The next theorem shows that the above estimators are ratio-consistent to $tr(\Sigma_i^2)$ and $tr(\Sigma_1\Sigma_2)$, respectively.

Theorem 2.2 *Under the assumptions (2.6)-(2.9) and (2.11), for $i = 1$ or 2,*

$$\frac{\widehat{tr(\Sigma_i^2)}}{tr(\Sigma_i^2)} \xrightarrow{P} 1 \quad and \quad \frac{\widehat{tr(\Sigma_1\Sigma_2)}}{tr(\Sigma_1\Sigma_2)} \xrightarrow{P} 1 \quad as \ p \ and \ n \to \infty.$$

A ratio-consistent estimator of σ_{n1}^2 under H_0 is

$$\hat{\sigma}_{n1}^2 = \frac{2}{n_1(n_1-1)}\widehat{tr(\Sigma_1^2)} + \frac{2}{n_2(n_2-1)}\widehat{tr(\Sigma_2^2)} + \frac{4}{n_1 n_2}\widehat{tr(\Sigma_1\Sigma_2)}.$$

This together with Theorem 2.1 leads to the test statistic,

$$Q_n = T_n/\hat{\sigma}_{n1} \xrightarrow{d} N(0,1) \quad as \ p \ and \ n \to \infty$$

under H_0. The proposed test with an α level of significance rejects H_0 if $Q_n > \xi_\alpha$ where ξ_α is the upper α quantile of $N(0,1)$.

2.3.4   Power properties of the proposed test

Theorem 2.1 and Theorem 2.2 allow us to discuss the power properties of the proposed test. The discussion is made under (2.9) and (2.10), respectively. The power under the local alternative (2.9) is

$$\beta_{n1}(||\mu_1 - \mu_2||) = \Phi\left(-\xi_\alpha + \frac{nk(1-k)||\mu_1 - \mu_2||^2}{\sqrt{2tr\{\tilde{\Sigma}(k)^2\}}}\right), \tag{2.16}$$

where $\tilde{\Sigma}(k) = k\Sigma_1 + (1-k)\Sigma_2$ and Φ is the standard normal distribution function. The power of the Bai-Saranadasa test has the same form if $\Sigma_1 = \Sigma_2$ and if p and n are of the same order.

The power under (2.10) is

$$\beta_{n2}(||\mu_1 - \mu_2||) = \Phi\left(-\frac{\sigma_{n1}}{\sigma_{n2}}\xi_\alpha + \frac{||\mu_1 - \mu_2||^2}{\sigma_{n1}}\right) = \Phi\left(\frac{||\mu_1 - \mu_2||^2}{\sigma_{n1}}\right)$$

as $\sigma_{n1}/\sigma_{n2} \to 0$. Substitute the expression for σ_{n1}, and we have

$$\beta_{n2}(||\mu_1 - \mu_2||) = \Phi\left(\frac{nk(1-k)||\mu_1 - \mu_2||^2}{\sqrt{2tr\{\tilde{\Sigma}(k)^2\}}}\right). \tag{2.17}$$

Both (2.16) and (2.17) indicate that the proposed test has non-trivial power under the two types of the alternative hypothesis as long as

$$n||\mu_1 - \mu_2||^2/\sqrt{tr\{\tilde{\Sigma}(k)^2\}}$$

does not vanish to 0 as n and $p \to \infty$. The flavor of the proposed test is different from tests formulated by combining p marginal tests on H_{0l} (defined after (2.1)) for $l = 1, \ldots, p$. Such tests are usually constructed via $\max_{1 \le l \le p} T_{nl}$ where T_{nl} is a marginal test statistic for H_{0l}. This is the case of Kosorok and Ma (2007) and Fan, Hall and Yao (2007). A condition on p and n is needed to ensure (i) convergence of $\max_{1 \le l \le p} T_{nl}$,

and (ii) p can reach an order of $exp(\alpha n^\beta)$ for positive constants α and β. Usually some additional assumptions are needed; for instance, Kosorok and Ma (2007) assumed each component of the random vector has compact support for testing means.

Naturally, if the number of significant univariate hypotheses ($\mu_{1l} \neq \mu_{2l}$) is a lot less than p, which is the so-called sparsity scenario, a simultaneous test like the one we propose may encounter a loss of power. This is actually quantified by the power expression (2.16). Without loss of generality, suppose that each μ_i can be partitioned as $(\mu_i^{(1)'}, \mu_i^{(2)'})'$ so that under $H_1 : \mu_1^{(1)} = \mu_2^{(1)}$ and $\mu_1^{(2)} \neq \mu_2^{(2)}$, where $\mu_i^{(1)}$ is of p_1 dimensional and $\mu_i^{(2)}$ is of p_2 dimensional and $p_1 + p_2 = p$. Then $||\mu_1 - \mu_2|| = p_2\delta^2$ for some positive constant δ^2. Suppose that λ_{m_0} be the smallest non-zero eigenvalue of $\tilde{\Sigma}(k)$. Then under the local alternative (2.9), the asymptotic power is bounded above and below by

$$\Phi\left(-\xi_\alpha + \frac{nk(1-k)p_2\delta^2}{\sqrt{2p}\lambda_p}\right) \leq \beta(||\mu_1 - \mu_2||) \leq \Phi\left(-\xi_\alpha + \frac{nk(1-k)p_2\delta^2}{\sqrt{2(p-m_0)}\lambda_{m_0}}\right).$$

If p is very large relative to n and p_2 under both high-dimensionality and sparsity, so that $nk(1-k)p_2\eta^2/\sqrt{2(p-m_0)} \to 0$, the test could endure low power. With this in mind, we check on the performance of the test under sparsity in simulation studies in Section 2.5. The simulations show that the proposed test has a robust power and is in fact more powerful than tests based on multiple comparisons with either the Bonferroni and false discovery rate (FDR) procedures. We note here that, due to the multivariate nature of the test and the hypothesis, the proposed test cannot identify which components are significant after the null multivariate hypothesis is rejected. Additional follow-up procedures have to be employed for that purpose. The proposed test becomes very useful when the purpose is to identify significant groups of components like sets of genes, as illustrated in Section 2.4. The above discussion can be readily extended to the case of (2.10) due to the similarity in the two power

functions.

The proposed two-sample test can be modified for paired observations $\{(Y_{i1}, Y_{i2})\}_{i=1}^n$ where Y_{i1} and Y_{i2} are two measurements of p-dimensions on a subject i before and after a treatment. Let $X_i = Y_{i2} - Y_{i1}$, $\mu = E(X_i)$ and $\Sigma = Var(X_i)$. This is effectively a one-sample problem with high-dimensional data. The hypothesis of interest is

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0.$$

We can use $F_n = \Sigma_{i \neq j}^n X_i' X_j / \{n(n-1)\}$ as the test statistic. It is readily shown that $E(F_n) = \mu' \mu$ and $Var(F_n) = \frac{2}{n(n-1)} tr(\Sigma_1^2)\{1 + o(1)\}$ under both H_0 and H_1 if we assume a condition similar to (2.9) so that $\mu' \Sigma \mu = o\{n^{-1} tr(\Sigma^2)\}$, and the asymptotic normality of F_n by adding $tr(\Sigma^4) = o\{tr^2(\Sigma^2)\}$, a variation of (2.11), can be established by utilizing part of the proof on the asymptotic normality of T_n. The $tr(\Sigma^2)$ can be ratio-consistently estimated with n_1 replaced by n in $\widehat{tr(\Sigma_1^2)}$ which leads to a ratio-consistent variance estimation for F_n. Then the test and its power can be expressed in similar ways as those for the two-sample test.

When $p = O(1)$, which may be viewed as having finite dimension, the asymptotic normality as conveyed in Theorem 2.1 may not be valid anymore. It may be shown under conditions (2.6)-(2.9) without (2.11), as condition (2.11) is no longer relevant when p is bounded, that the test statistic $(n_1 + n_2)T_n$ converges to $\sum_{l=1}^{2p} \eta_l \chi_{1,l}^2$ where $\{\chi_{1,l}^2\}_{l=1}^{2p}$ are independent χ_1^2 distributed random variables, and $\{\eta_l\}_{l=1}^{2p}$ is a set of constants. The conclusion of Theorem 2.2 remains valid when p is bounded. The proposed test can still be used for testing in this situation of bounded dimension with estimated critical values via estimation of $\{\eta_l\}_{l=1}^{2p}$. However, people may like to use a test specially catered for such a case, for instance, Hotelling's test.

2.4   Gene-set Testing

Identifying sets of genes which are significant with respect to certain treatments is the latest development in genetics research (Barry et al., 2005; Nettleton et al., 2008; Efron and Tibshirani, 2007; Newton et al., 2007). Biologically speaking, each gene does not function individually in isolation. Rather, one gene tends to work with other genes to achieve certain biological tasks.

Suppose that $\mathcal{S}_1, \cdots, \mathcal{S}_q$ are q sets of genes, where the gene-set $\mathcal{S}_g$ consists of p_g genes. Let $F_{1\mathcal{S}_g}$ and $F_{2\mathcal{S}_g}$ be the distribution functions corresponding to $\mathcal{S}_g$ under the treatment and control, and $\mu_{1\mathcal{S}_g}$ and $\mu_{2\mathcal{S}_g}$ be their respective means. The hypothesis of interest is

$$H_{0g} : \mu_{1\mathcal{S}_g} = \mu_{2\mathcal{S}_g} \quad \text{for} \quad g = 1, \cdots, q.$$

The gene sets $\{\mathcal{S}_g\}_{g=1}^q$ can overlap as a gene can belong to several functional groups, and p_g, the number of genes in a set, can range from a moderate to a very large number. So, there are issues of both multiplicity and high-dimensionality in gene-set testing.

We propose applying the proposed test for the significance of each gene-set $\mathcal{S}_g$ when p_g is large. When p_g is of low-dimension, Hotelling's test may be used. Let pv_g, $g = 1, \cdots, q$ be the P-values obtained from these tests. To control the overall family-wise error rate, we can employ the Bonferroni procedure; to control FDR, we can use Benjamini and Hochberg (1995) method or its variations as in Benjamini and Yekutieli (2001) and Storey et al. (2004). These lead to control of the family-wise error rate or FDR in the context of gene-sets testing. In contrast, tests based on univariate testing have difficulties in producing P-values for gene-sets.

Acute Lymphoblastic Leukemia (ALL) is a form of leukemia, a cancer of white

blood cells. The ALL data (Chiaretti et al., 2004) contains microarray expressions for 128 patients with either T-cell or B-cell type Leukemia. Within the B-cell type leukemia, there are two sub-classes representing two molecular classes: the BCR/ABL class and NEG class. The data set has been analyzed by Dudoit et al. (2008) using a different methodology.

Gene-sets are technically defined in the Gene Ontology (GO) system that provides structured and controlled vocabularies producing names of gene-sets (also called GO terms). There are three groups of Gene ontologies of interest: Biological Processes (BP), Cellular Components (CC) and Molecular Functions (MF). We carried out preliminary screening for gene-filtering using the approach in Gentleman et al. (2005), which left 2391 genes for analysis. There are 575 unique GO terms which have more than 10 genes in BP category, 221 in MF and 154 in CC for the ALL data. The largest gene-set contains 2059 genes in BP, 2112 genes in MF and 2078 genes in CC; and the GO terms of the three categories share 1861 common genes. We are interested in detecting differences in the expression levels of gene-sets within a subset of B-cell ALL data between the BCR/ABL molecular sub-class ($n_1 = 37$) and the NEG molecular sub-class ($n_2 = 42$) for each of the three categories.

We applied the proposed two-sample test with a 5% significance level to test each of the gene-sets in conjunction with the Bonferroni correction to control the family-wise error rate at 0.05 level. It was found that there are 259 gene-sets declared significant in the BP group, 110 in the MF group and 53 in the CC group. Figure 2.1 displays the histograms of the P-values and the values of test statistic Q_n for the three gene-categories. It shows a strong non-uniform distribution of the P-values with a large number of P-values cluster near 0. At the same time, the Q_n-value plots indicate the average Q_n-values are much larger than zero. These explain the large number of

significant gene-sets detected by the proposed test.

The number of the differentially expressed gene-sets may seem to be high. This is mainly due to overlapping gene-sets. To appreciate this point, we computed for each (say ith) significant gene-set, the number of other significant gene-sets which overlap with it, say b_i; and obtained the average of $\{b_i\}$ and their standard deviation. The average number of overlaps (standard deviation) for BP group was 198.9(51.3), 55.6 (25.2) for MF and 41.6 (9.5) for CC. These numbers are indeed very high and reveal the gene-sets and their P-values are highly dependent.

Finally, we carried out back-testing for the same hypothesis by randomly splitting the 42 NEG class into two sub-classes of equal sample size and testing for mean differences. This set-up led to the situation of H_0. Figures 2.2 reports the P-values and Q_n-values for the three Gene Ontology groups. We note that the distributions of the P-values are much closer to the uniform distribution than Figure 2.1. It is observed that the histograms of Q_n-values are centered close to zero and are much closer to the normal distribution than their counterparts in Figure 2.1 which is reassuring.

2.5 Simulation Studies

In this section, we report results from simulation studies which were designed to evaluate the performance of the proposed two-sample test for high-dimensional data. For comparison, we also conducted the test proposed by Bai and Saranadasa (1996)(BS test), and two tests based on multiple comparison procedures by employing the Bonferroni and the FDR control (Benjamini and Hochberg, 1995). The Bonferroni procedure controls the family-wise error rate at a level of significance α which coincides with the significance for the FDR control, the proposed test and the BS test. In the two multiple comparison procedures, we conducted univariate two-sample t-tests for

the univariate hypotheses $H_{0l} : \mu_{1l} = \mu_{2l}$ vs. $H_{1l} : \mu_{1l} \neq \mu_{2l}$ for $l = 1, 2, \cdots, p$. We reject the null $H_0 : \mu_1 = \mu_2$ if there exists an $l \in \{1, \cdots p\}$ such that H_{0l} is rejected.

Two simulation models for X_{ij} are considered. One has a moving average structure that allows a general dependent structure; the other could allocate the the alternative hypotheses sparsely which enables us to evaluate the performance of the tests under sparsity.

2.5.1 Moving average model

The first simulation model has the following moving average structure:

$$X_{ijk} = \rho_1 Z_{ijk} + \rho_2 Z_{ijk+1} + \cdots + \rho_p Z_{ijk+p-1} + \mu_{ij}$$

for $i = 1$ and 2, $j = 1, 2, \cdots, n_i$ and $k = 1, 2, \cdots, p$ where $\{Z_{ijk}\}$ are respectively i.i.d. random variables. We consider two distributions for the innovations $\{Z_{ijk}\}$. One is a centralized Gamma$(4, 1)$ so that it has zero mean, and the other is $N(0, 1)$.

For each distribution of $\{Z_{ijk}\}$, we consider two configurations of dependence among components of X_{ij}. One has weaker dependence with $\rho_l = 0$ for $l > 3$. This prescribes a "two dependence" moving average structure where X_{ijk_1} and X_{ijk_2} are dependent only if $|k_1 - k_2| \leq 2$. The $\{\rho_l\}_{l=1}^3$ are generated independently from $U(2, 3)$ which are $\rho_1 = 2.883$, $\rho_2 = 2.794$ and $\rho_3 = 2.849$ and are kept fixed throughout the simulation. The second configuration has all ρ_ls generated from $U(2, 3)$, and again remain fixed throughout the simulation. We call this the "full dependence case". The above dependence structures assign equal covariance matrices $\Sigma_1 = \Sigma_2 = \Sigma$ which allows a meaningful comparison with the BS test.

Without loss of generality, we fix $\mu_1 = 0$ and choose μ_2 in the same fashion as Benjamini and Hochberg (1995). Specifically, the percentage of true null hypotheses $\mu_{1l} = \mu_{2l}$ for $l = 1, \cdots, p$ are chosen to be 0%, 25%, 50%, 75%, 95% and 99% and

100% , respectively. By experimenting with 95% and 99% we gain information on the performance of the test when $\mu_{1l} \neq \mu_{2l}$ are sparse. It provides empirical checks on the potential concerns of the power of the simultaneous high-dimensional tests as made at the end of Section 2.3. At each percentage level of true null, three patterns of allocation are considered for the non-zero μ_{2l} in $\mu_2 = (\mu_{21}, \cdots, \mu_{2p})'$: (i) the equal allocation where all the non-zero μ_{2l} are equal; (ii) linearly increasing and (iii) linearly decreasing allocations as specified in Benjamini and Hochberg (1995). To make the power comparable among the configurations of H_1, we set $\eta =: ||\mu_1 - \mu_2||^2/\sqrt{tr(\Sigma^2)} = 0.1$ throughout the simulation. We choose $p = 500$ and 1000 and $n = [20 \log(p)] = 124$ and 138, respectively.

Tables 2.1 and Table 2.2 report the empirical power and size of the four tests with Gamma innovations at a 5% nominal significance level or family-wise error rate or FDR based on 5000 simulations. The results for the Normal innovations have a similar pattern and are shown in Table 2.3 and Table 2.4. The simulation results in Tables 2.1, 2.2, 2.3 and 2.4 can be summarized as follows. The proposed test is much more powerful than the BS test for all cases considered in the simulation while maintaining a reasonably-sized approximation to the nominal 5% level. Both the proposed test and the BS test are more powerful than the two tests based on the multiple univariate testing using the Bonferroni and FDR procedures. This is a little expected as both the proposed and the BS test are designed to test for the entire p-dimensional hypotheses while the multiple testing procedures are targeted at the individual univariate hypothesis. What is surprising is that when the percentage of true null is high at 95% and 99%, the proposed test still is much more powerful than the two multiple testing procedures for all three allocations of the non-zero components in μ_2. It is observed that the sparsity (95% and 99% true null) does reduce the power

of the proposed test a little. However, the proposed test still enjoys good power, especially when compared with the other three tests.

We also observe that when there is more dependence among multivariate components of the data vectors in the full dependence model, there is a drop in the power for each of the tests. The power of the tests based on the Bonferroni and FDR procedures is alarmingly low and is only slightly larger than the nominal significance level.

We also collected information on the quality of $tr(\Sigma^2)$ estimation. Table (2.5) reports empirical averages and standard deviation of $\widehat{tr(\Sigma^2)}/tr(\Sigma^2)$. It shows that the proposed estimator for $tr(\Sigma^2)$ has a much smaller bias and standard deviation than those proposed in Bai and Saranadasa (1996) in all cases, and provides an empirical verification for Theorem 2.2.

2.5.2 Sparse model

An examination of the previous simulation setting reveals that the strength of the "signals" $\mu_{2l} - \mu_{1l}$ corresponding to the alternative hypotheses are low relative to the level of noise (variance) which may not be a favorable situation for the two tests based on multiple univariate testing. To gain more information on the performance of the tests under sparsity, we consider the following simulation model such that

$$X_{1il} = Z_{1il} \quad \text{and} \quad X_{2il} = \mu_l + Z_{2il} \quad \text{for } l = 1, \ldots, p$$

where $\{Z_{1il}, Z_{2il}\}_{l=1}^{p}$ are mutually independent $N(0, 1)$ random variables, and the "signals",

$$\mu_l = \varepsilon\sqrt{2\log(p)} \quad \text{for } l = 1, \ldots, q = [p^c] \text{ and } \quad \mu_l = 0 \quad \text{for } l > q,$$

for some $c \in (0, 1)$. Here q is the number of significant alternative hypotheses. The sparsity of the hypotheses is determined by c: the smaller the c is, the more sparse

the alternative hypotheses with $\mu_l \neq 0$. This simulation model is similar to the one used in Abramovich et al. (2006).

According to (2.16), the power of the proposed test has asymptotic power

$$\beta(||\mu||) = \Phi\left(-\xi_\alpha + \frac{np^{(c-1/2)}\varepsilon^2 log(p)}{2\sqrt{2}}\right)$$

which indicates that the test has a much reduced power if $c < 1/2$ with respect to p. We, therefore, chose $p = 1000$ and $c = 0.25, 0.35, 0.45$ and 0.55, respectively, which leads to $q = 6, 11, 22$, and 44, respectively. We call $c = 0.25, 0.35$ and 0.45 the sparse cases.

In order to prevent trivial powers of α or 1 in the simulation, we set $\varepsilon = 0.25$ for $c = 0.25$ and 0.45; and $\varepsilon = 0.15$ for $c = 0.35$ and 0.55. Table 2.6 summarizes the simulations results based on 500 simulations. It shows that in the extreme sparse cases of $c = 0.25$, the FDR and Bonferroni tests did have higher power than the proposed test. The power were largely similar among the three tests for $c = 0.35$. However, when the sparsity is moderated to $c = 0.45$, the proposed test starts to surpass the FDR and Bonferroni procedures. The gap in power performance is further increased when $c = 0.55$. Table 2.7 reports the quality of the variance estimation in Table 2.6 which shows the proposed variance estimators incur very little bias and variance for even very small sample sizes of $n_1 = n_2 = 10$.

2.6 Technical Proofs

Derivations for $E(T_n)$ and $Var(T_n)$: As

$$T_n = \frac{\Sigma_{i\neq j}^{n_1} X'_{1i} X_{1j}}{n_1(n_1 - 1)} + \frac{\Sigma_{i\neq j}^{n_2} X'_{2i} X_{2j}}{n_2(n_2 - 1)} - 2\frac{\Sigma_{i=1}^{n_1}\Sigma_{j=1}^{n_2} X'_{1i} X_{2j}}{n_1 n_2},$$

it is straightforward to show that $E(T_n) = \mu'_1\mu_1 + \mu'_2\mu_2 - 2\mu'_1\mu_2 = ||\mu_1 - \mu_2||^2$.

Let $P_1 = \frac{\sum_{i\neq j}^{n_1} X'_{1i} X_{1j}}{n_1(n_1-1)}$, $P_2 = \frac{\sum_{i\neq j}^{n_2} X'_{2i} X_{2j}}{n_2(n_2-1)}$ and $P_3 = -2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} X'_{1i} X_{2j}}{n_1 n_2}$. It can be shown that

$$
\begin{aligned}
Var(P_1) &= \frac{1}{n_1^2(n_1-1)^2}\bigg[2\Sigma_{(i=s)\neq(j=t)}^{n_1} tr\Big\{(\Sigma_1 + \mu_1\mu_1')'(\Sigma_1 + \mu_1\mu_1')\Big\} \\
&\quad + 4\Sigma_{(i=s)\neq(j\neq t)}^{n_1}\Big\{\mu_1'\Sigma_1\mu_1 + (\mu_1'\mu_1)^2\Big\} + \Sigma_{i\neq s\neq j\neq t}^{n_1}(\mu_1'\mu_1)^2\bigg] - (\mu_1'\mu_1)^2 \\
&= \frac{2}{n_1(n_1-1)} tr(\Sigma_1^2) + \frac{4\mu_1'\Sigma_1\mu_1}{n_1}, \\
Var(P_2) &= \frac{2}{n_2(n_2-1)} tr(\Sigma_2^2) + \frac{4\mu_2'\Sigma_2\mu_2}{n_2}, \\
Var(P_3) &= \frac{4}{n_1 n_2} tr(\Sigma_1\Sigma_2) + \frac{4\mu_2'\Sigma_1\mu_2}{n_1} + \frac{4\mu_1'\Sigma_2\mu_1}{n_2},
\end{aligned}
$$

and

$$
Cov(P_1, P_3) = -\frac{4\mu_1'\Sigma_1\mu_2}{n_1} \quad \text{and} \quad Cov(P_2, P_3) = -\frac{4\mu_1'\Sigma_2\mu_2}{n_2}.
$$

Because the two samples are independent, $Cov(P_1, P_2) = 0$. In summary,

$$
\begin{aligned}
Var(T_n) = \ & \frac{2}{n_1(n_1-1)} tr(\Sigma_1^2) + \frac{2}{n_2(n_2-1)} tr(\Sigma_2^2) + \frac{4}{n_1 n_2} tr(\Sigma_1\Sigma_2) \\
& + \frac{4}{n_1}(\mu_1-\mu_2)'\Sigma_1(\mu_1-\mu_2) + \frac{4}{n_2}(\mu_1-\mu_2)'\Sigma_2(\mu_1-\mu_2).
\end{aligned}
$$

Thus, under H_0

$$
Var(T_n) = \sigma_{n1}^2 =: \frac{2}{n_1(n_1-1)} tr(\Sigma_1^2) + \frac{2}{n_2(n_2-1)} tr(\Sigma_2^2) + \frac{4}{n_1 n_2} tr(\Sigma_1\Sigma_2).
$$

Under $H_1 : \mu_1 \neq \mu_2$, with (2.9),

$$
Var(T_n) = \sigma_{n1}^2\{1 + o(1)\};
$$

and with (2.10),

$$
Var(T_n) = \sigma_{n2}^2\{1 + o(1)\},
$$

where $\sigma_{n2}^2 = \frac{4}{n_1}(\mu_1-\mu_2)'\Sigma_1(\mu_1-\mu_2) + \frac{4}{n_2}(\mu_1-\mu_2)'\Sigma_2(\mu_1-\mu_2)$.

Asymptotic Normality of T_n: We note that $T_n = T_{n1} + T_{n2}$ where

$$
\begin{aligned}
T_{n1} &= \frac{\Sigma_{i\neq j}^{n_1}(X_{1i} - \mu_1)'(X_{1j} - \mu_1)}{n_1(n_1 - 1)} + \frac{\Sigma_{i\neq j}^{n_2}(X_{2i} - \mu_2)'(X_{2j} - \mu_2)}{n_2(n_2 - 1)} \\
&\quad -2\frac{\Sigma_{i=1}^{n_1}\Sigma_{j=1}^{n_2}(X_{1i} - \mu_1)'(X_{2j} - \mu_2)}{n_1 n_2}
\end{aligned}
\tag{2.18}
$$

and

$$
T_{n2} = \frac{2\Sigma_{i=1}^{n_1}(X_{1i} - \mu_1)'(\mu_1 - \mu_2)}{n_1} + \frac{2\Sigma_{i=1}^{n_2}(X_{2i} - \mu_2)'(\mu_2 - \mu_1)}{n_2} + ||\mu_1 - \mu_2||^2.
$$

It is easy to show that $E(T_{n1}) = 0$ and $E(T_{n2}) = ||\mu_1 - \mu_2||^2$, and

$$
Var(T_{n2}) = \frac{4(\mu_1 - \mu_2)'\Sigma_1(\mu_1 - \mu_2)}{n_1} + \frac{4(\mu_2 - \mu_1)'\Sigma_2(\mu_2 - \mu_1)}{n_2};
$$

$$
Var(T_{n1}) = \frac{2}{n_1(n_1 - 1)}tr(\Sigma_1^2) + \frac{2}{n_2(n_2 - 1)}tr(\Sigma_2^2) + \frac{4}{n_1 n_2}tr(\Sigma_1\Sigma_2).
$$

Under (2.9), as

$$
Var\left(\frac{T_{n2} - ||\mu_1 - \mu_2||^2}{\sigma_{n1}}\right) = o(1),
$$

$$
\frac{T_n - ||\mu_1 - \mu_2||^2}{\sqrt{Var(T_n)}} = \frac{T_{n1}}{\sigma_{n1}} + o_p(1).
\tag{2.19}
$$

Under (2.10),

$$
\frac{T_n - ||\mu_1 - \mu_2||^2}{\sqrt{Var(T_n)}} = \frac{T_{n2} - ||\mu_1 - \mu_2||^2}{\sigma_{n2}} + o_p(1).
\tag{2.20}
$$

As T_{n2} are independent sample averages, its asymptotic normality is readily attainable as showed later. The main task of the following proof is for the case under (2.9) when T_{n1} is the contributor of the asymptotic distribution. From (2.18), in the derivation for the asymptotic normality of T_{n1}, we can assume without loss of generality that $\mu_1 = \mu_2 = 0$.

Let $Y_i = X_{1i}$ for $i = 1, \cdots, n_1$ and $Y_{j+n_1} = X_{2j}$ for $j = 1, \cdots, n_2$, and for $i \neq j$, define ϕ_{ij} as follows,

$$\phi_{ij} = \begin{cases} n_1^{-1}(n_1 - 1)^{-1} Y_i' Y_j & \text{if } i, j \in \{1, 2, \cdots, n_1\}, \\ -n_1^{-1} n_2^{-1} Y_i' Y_j & \text{if } i \in \{1, 2, \cdots, n_1\} \text{ and } j \in \{n_1 + 1, \cdots, n_1 + n_2\} \\ n_2^{-1}(n_2 - 1)^{-1} Y_i' Y_j, & \text{if } i, j \in \{n_1 + 1, \cdots, n_1 + n_2\}. \end{cases}$$

Define $V_{nj} = \sum_{i=1}^{j-1} \phi_{ij}$ for $j = 2, 3, \cdots, n =: n_1 + n_2$, $S_{nm} = \sum_{j=2}^{m} V_{nj}$ and $\mathcal{F}_{nm} = \sigma\{Y_1, Y_2, \cdots, Y_m\}$ which is the σ-algebra generated by $\{Y_1, Y_2, \cdots, Y_m\}$. Now T_n can be rewritten as $T_n = 2 \sum_{j=2}^{n_1+n_2} V_{nj}$.

Lemma 2.1 *For each n, $\{S_{nm}, \mathcal{F}_{nm}\}_{m=1}^{n}$ is the sequence of zero mean and a square integrable martingale.*

Proof:

It's obvious that $\mathcal{F}_{nj-1} \subseteq \mathcal{F}_{nj}$, for any $1 \leq j \leq n$ and S_{nm} is of zero mean and square integrable. We only need to show $E(S_{nq}|\mathcal{F}_{nm}) = S_{nm}$ for any $q \geq m$. We note that if $j \leq m \leq n$, then $E(V_{nj}|\mathcal{F}_{nm}) = \Sigma_{i=1}^{j-1} E(\phi_{ij}|\mathcal{F}_{nm}) = \Sigma_{i=1}^{j-1} \phi_{ij} = V_{nj}$. If $j > m$, then $E(\phi_{ij}|\mathcal{F}_{nm}) = E(Y_i' Y_j|\mathcal{F}_{nm})$.

If $i > m$, as Y_i and Y_j are both independent of $\mathcal{F}_{nm}$,

$$E(\phi_{ij}|\mathcal{F}_{nm}) = E(\phi_{ij}) = 0.$$

If $i \leq m$, $E(\phi_{ij}|\mathcal{F}_{n,m}) = E(Y_i' Y_j|\mathcal{F}_{n,m}) = Y_i' E(Y_j) = 0$. Hence,

$$E(V_{nj}|\mathcal{F}_{n,m}) = 0.$$

In summary, for $q > m$, $E(S_{nq}|\mathcal{F}_{nm}) = \Sigma_{j=1}^{q} E(V_{nj}|\mathcal{F}_{nm}) = \Sigma_{j=1}^{m} V_{nj} = S_{nm}$. This completes the proof of the lemma.

Lemma 2.2 *Under Condition (2.9),*

$$\frac{\sum_{j=2}^{n_1+n_2} E[V_{nj}^2|\mathcal{F}_{n,j-1}]}{\sigma_{n1}^2} \xrightarrow{P} \frac{1}{4}.$$

Proof:

Note that

$$
\begin{aligned}
E(V_{nj}^2|\mathcal{F}_{nj-1}) &= E\left\{ \left(\sum_{i=1}^{j-1} Y_i'Y_j \right)^2 |\mathcal{F}_{nj-1} \right\} = E\left(\sum_{i_1,i_2=1}^{j-1} Y_{i_1}'Y_jY_j'Y_{i_2}|\mathcal{F}_{nj-1} \right)\\
&= \sum_{i_1,i_2=1}^{j-1} Y_{i_1}'E(Y_jY_j'|\mathcal{F}_{nj-1})Y_{i_2} = \sum_{i_1,i_2=1}^{j-1} Y_{i_1}'E(Y_jY_j')Y_{i_2}\\
&= \sum_{i_1,i_2=1}^{j-1} Y_{i_1}'\frac{\tilde{\Sigma}_j}{\tilde{n}_j(\tilde{n}_j-1)}Y_{i_2},
\end{aligned}
$$

where $\tilde{\Sigma}_j = \Sigma_1$, and $\tilde{n}_j = n_1$, for $j \in [1, n_1]$ and $\tilde{\Sigma}_j = \Sigma_2$, $\tilde{n}_j = n_2$, if $j \in [n_1 + 1, n_1 + n_2]$.

Define $\eta_n = \sum_{j=2}^{n_1+n_2} E(V_{nj}^2|\mathcal{F}_{nj-1})$. Then

$$
\begin{aligned}
E(\eta_n) &= \frac{tr(\Sigma_1^2)}{2n_1(n_1-1)} + \frac{tr(\Sigma_2^2)}{2n_2(n_2-1)} + \frac{tr(\Sigma_1\Sigma_2)}{(n_1-1)(n_2-1)}\\
&= \tfrac{1}{4}\sigma_{n1}^2\{1 + o(1)\}. \quad\quad\quad (2.21)
\end{aligned}
$$

Now consider

$$
\begin{aligned}
E(\eta_n^2) &= E\left\{ \sum_{j=2}^{n_1+n_2} \sum_{i_1,i_2=1}^{j-1} Y_{i_1}'\frac{\tilde{\Sigma}_j}{\tilde{n}_j(\tilde{n}_j-1)}Y_{i_2} \right\}^2\\
&= E\left\{ 2 \sum_{2\leq j_1<j_2} \sum_{i_1,i_2=1}^{j_1-1} \sum_{i_3,i_4=1}^{j_2-1} Y_{i_1}'\frac{\tilde{\Sigma}_{j_1}}{\tilde{n}_{j_1}(\tilde{n}_{j_1}-1)}Y_{i_2}Y_{i_3}'\frac{\tilde{\Sigma}_{j_2}}{\tilde{n}_{j_2}(\tilde{n}_{j_2}-1)}Y_{i_4} \right. \quad (2.22)\\
&\quad\quad \left. + \sum_{j=2}^{n_1+n_2} \sum_{i_1,i_2=1}^{j-1} \sum_{i_3,i_4=1}^{j-1} Y_{i_1}'\frac{\tilde{\Sigma}_j}{\tilde{n}_j(\tilde{n}_j-1)}Y_{i_2}Y_{i_3}'\frac{\tilde{\Sigma}_j}{\tilde{n}_j(\tilde{n}_j-1)}Y_{i_4} \right\}\\
&= 2E(A) + E(B), \quad \text{say,}
\end{aligned}
$$

where

$$A = \sum_{2 \leq j_1 < j_2}^{n_1+n_2} \sum_{i_1,i_2=1}^{j_1-1} \sum_{i_3,i_4=1}^{j_2-1} Y_{i_1}' \frac{\tilde{\Sigma}_{j_1}}{\tilde{n}_{j_1}(\tilde{n}_{j_1}-1)} Y_{i_2} Y_{i_3}' \frac{\tilde{\Sigma}_{j_2}}{\tilde{n}_{j_2}(\tilde{n}_{j_2}-1)} Y_{i_4},$$

$$B = \sum_{j=2}^{n_1+n_2} \sum_{i_1,i_2=1}^{j-1} \sum_{i_3,i_4=1}^{j-1} Y_{i_1}' \frac{\tilde{\Sigma}_{j}}{\tilde{n}_{j}(\tilde{n}_{j}-1)} Y_{i_2} Y_{i_3}' \frac{\tilde{\Sigma}_{j}}{\tilde{n}_{j}(\tilde{n}_{j}-1)} Y_{i_4}. \qquad (2.23)$$

The term B can be further partitioned as $B = B_1 + B_2$ and

$$E(B_1) \equiv E\left\{ \sum_{j=2}^{n_1} \sum_{i_1,i_2=1}^{j-1} \sum_{i_3,i_4=1}^{j-1} Y_{i_1}' \frac{\Sigma_1}{n_1(n_1-1)} Y_{i_2} Y_{i_3}' \frac{\Sigma_1}{n_1(n_1-1)} Y_{i_4} \right\},$$

$$E(B_2) \equiv E\left\{ \sum_{j=n_1+1}^{n_1+n_2} \sum_{i_1,i_2=1}^{j-1} \sum_{i_3,i_4=1}^{j-1} Y_{i_1}' \frac{\Sigma_2}{n_2(n_2-1)} Y_{i_2} Y_{i_3}' \frac{\Sigma_2}{n_2(n_2-1)} Y_{i_4} \right\}.$$

We only prove here that $E(B_1) = o(\sigma_{n1}^4)$ as $E(B_2) = o(\sigma_{n1}^4)$ can be proved by following the same procedure.

Consider different combinations of i_1, i_2, i_3, i_4 such that

$$\begin{aligned} E(B_1) &= E\left[\sum_{j=2}^{n_1} \sum_{i_1 \neq i_2}^{j-1} \left\{ Y_{i_1}' \frac{\Sigma_1}{n_1(n_1-1)} Y_{i_1} Y_{i_2}' \frac{\Sigma_1}{n_1(n_1-1)} Y_{i_2} \right\} \right. \\ &\quad + 2 \sum_{j=2}^{n_1} \sum_{i_1 \neq i_2}^{j-1} \left\{ Y_{i_1}' \frac{\Sigma_1}{n_1(n_1-1)} Y_{i_2} Y_{i_2}' \frac{\Sigma_1}{n_1(n_1-1)} Y_{i_1} \right\} \\ &\quad \left. + \sum_{j=2}^{n_1} \sum_{i=1}^{j-1} \left\{ Y_i' \frac{\Sigma_1}{n_1(n_1-1)} Y_i Y_i' \frac{\Sigma_1}{n_1(n_1-1)} Y_i \right\} \right] \\ &= : E(B_{1a}) + 2E(B_{1b}) + E(B_{1d}), \end{aligned}$$

where

$$B_{1a} = \sum_{j=2}^{n_1} \sum_{i_1 \neq i_2}^{j-1} \left\{ Y_{i_1}' \frac{\Sigma_1}{n_1(n_1-1)} Y_{i_1} Y_{i_2}' \frac{\Sigma_1}{n_1(n_1-1)} Y_{i_2} \right\},$$

$$B_{1b} = \sum_{j=2}^{n_1} \sum_{i_1 \neq i_2}^{j-1} \left\{ Y_{i_1}' \frac{\Sigma_1}{n_1(n_1-1)} Y_{i_2} Y_{i_2}' \frac{\Sigma_1}{n_1(n_1-1)} Y_{i_1} \right\},$$

$$B_{1d} = \sum_{j=2}^{n_1} \sum_{i=1}^{j-1} \left\{ Y_i' \frac{\Sigma_1}{n_1(n_1-1)} Y_i Y_i' \frac{\Sigma_1}{n_1(n_1-1)} Y_i \right\}$$

and

$$E(B_{1a}) = \sum_{j=2}^{n_1}\sum_{i_1 \neq i_2}^{j-1} \frac{tr^2(\Sigma_1^2)}{n_1^4(n_1-1)^4} = O(n_1^{-5})tr^2(\Sigma_1^2) = o(\sigma_{n1}^4),$$

$$E(B_{1b}) = \sum_{j=2}^{n_1}\sum_{i_1 \neq i_2}^{j-1} \frac{tr(\Sigma_1^4)}{n_1^4(n_1-1)^4} = O(n_1^{-5})tr(\Sigma_1^4) = o(\sigma_{n1}^4),$$

$$E(B_{1d}) = E\sum_{j=2}^{n_1}\sum_{i=1}^{j-1}\left\{Y_i'\frac{\Sigma_1}{n_1(n_1-1)}Y_iY_i'\frac{\Sigma_1}{n_1(n_1-1)}Y_i\right\}$$

$$= n_1^{-4}(n_1-1)^{-4}\sum_{j=2}^{n_1}\sum_{i=1}^{j-1}E\left(Z_i'\Gamma_1'\Sigma_1\Gamma_1 Z_i Z_i'\Gamma_1'\Sigma_1\Gamma_1 Z_i\right)$$

$$= n_1^{-4}(n_1-1)^{-4}\sum_{j=2}^{n_1}\sum_{i=1}^{j-1}\left\{\Delta\sum_{l=1}^{m}\tilde{\sigma}_{ll}^{11^2} + 2tr(\Sigma_1^4) + tr^2(\Sigma_1^2)\right\}$$

$$= O(n_1^{-5})\left\{\Delta\sum_{l=1}^{m}\tilde{\sigma}_{ll}^{11^2} + 2tr(\Sigma_1^4) + tr^2(\Sigma_1^2)\right\} = o(\sigma_{n1}^4). \qquad (2.24)$$

Note that:

$$\begin{cases} \Gamma_s'\Sigma_t\Gamma_s = (\tilde{\sigma}_{lk}^{(st)})_{m \times m}; \\ \\ \sum_{l=1}^{m}(\tilde{\sigma}_{ll}^{(11)})^2 \leq \sum_{s,t}^{m}(\tilde{\sigma}_{st}^{(11)})^2 = tr(\Gamma_1'\Sigma_1\Gamma_1\Gamma_1'\Sigma_1\Gamma_1) = tr(\Sigma_1^4). \end{cases}$$

Then we have proven that $E(B_1) = o(\sigma_{n1}^4)$ and $E(B_2) = o(\sigma_{n1}^4)$ can be proved similarly.

To show $Var(\eta_n)/\sigma_{n1}^4 = o(1)$, we proceed with $E(A)$.

$$A = \sum_{2 \leq j_1 < j_2}^{n_1+n_2}\sum_{i_1,i_2=1}^{j_1-1}\sum_{i_3,i_4=1}^{j_2-1} Y_{i_1}'\frac{\Sigma_{j_1}}{n_{j_1}(n_{j_1}-1)}Y_{i_2}Y_{i_3}'\frac{\Sigma_{j_2}}{n_{j_2}(n_{j_2}-1)}Y_{i_4}.$$

As $\mu_1 = \mu_2 = 0$, we only need to consider i_1, i_2, i_3 and i_4 in these four cases: (a) $(i_1 = i_2) \neq (i_3 = i_4)$; (b) $(i_1 = i_3) \neq (i_2 = i_4)$;(c) $(i_1 = i_4) \neq (i_2 = i_3)$; (d) $i_1 = i_2 = i_3 = i_4$. For j_1 and j_2, we can have three possible combinations: $(1)j_1 < j_2 \leq n_1; (2)j_1 \leq n_1 < j_2; (3)n_1 < j_1 < j_2$. Then

$$2E(A) = \left\{\frac{tr^2(\Sigma_1^2)}{4n_1^2(n_1-1)^2} + \frac{tr^2(\Sigma_2^2)}{4n_2^2(n_2-1)^2} + \frac{tr(\Sigma_1^2)tr(\Sigma_1\Sigma_2)}{n_1^2(n_1-1)(n_2-1)} + \frac{tr(\Sigma_2^2)tr(\Sigma_1\Sigma_2)}{n_2^2(n_1-1)(n_2-1)}\right.$$

$$\left. + \frac{tr^2(\Sigma_1\Sigma_2)}{n_1n_2(n_1-1)(n_2-1)} + \frac{tr(\Sigma_1^2)tr(\Sigma_2^2)}{2n_1(n_1-1)n_2(n_2-1)}\right\}\{1+o(1)\}.$$

Hence, from (2.22) and (2.23),

$$
\begin{aligned}
E(\eta_n^2) &= \left\{ \frac{tr^2(\Sigma_1^2)}{4n_1^2(n_1-1)^2} + \frac{tr^2(\Sigma_2^2)}{4n_2^2(n_2-1)^2} + \frac{tr(\Sigma_1^2)tr(\Sigma_1\Sigma_2)}{n_1^2(n_1-1)(n_2-1)} + \frac{tr(\Sigma_2^2)tr(\Sigma_1\Sigma_2)}{n_2^2(n_1-1)(n_2-1)} \right. \\
&\quad \left. + \frac{tr^2(\Sigma_1\Sigma_2)}{n_1 n_2(n_1-1)(n_2-1)} + \frac{tr(\Sigma_1^2)tr(\Sigma_2^2)}{2n_1(n_1-1)n_2(n_2-1)} \right\} + o(\sigma_{n1}^4).
\end{aligned} \tag{2.25}
$$

Based on (2.21) and (2.25),

$$
Var(\eta_n) = E(\eta_n^2) - E^2(\eta_n) = o(\sigma_{n1}^4). \tag{2.26}
$$

Combine (2.21) and (2.26), and we have

$$
\sigma_{n1}^{-2} E\left\{ \Sigma_{j=1}^{n_1+n_2} E(V_{nj}^2|\mathcal{F}_{n,j-1}) \right\} = \tfrac{1}{4} \text{ and } \sigma_{n1}^{-4} Var\left\{ \Sigma_{j=1}^{n_1+n_2} E(V_{nj}^2|\mathcal{F}_{n,j-1}) \right\} = o(1).
$$

This completes the proof of Lemma 2.2.

Lemma 2.3 *Under condition (2.9),*

$$
\sum_{j=2}^{n_1+n_2} \sigma_{n1}^{-2} E\{V_{nj}^2 I(|V_{nj}| > \epsilon\sigma_{n1})|F_{nj-1}\} \xrightarrow{p} 0.
$$

Proof:

We note that $\sum_{j=2}^{n_1+n_2} \sigma_{n1}^{-2} E\{V_{nj}^2 I(|V_{nj}| > \epsilon\sigma_{n1})|F_{nj-1}\} \leq \sigma_{n1}^{-q}\epsilon^{2-q} \sum_{j=2}^{n_1+n_2} E(V_{nj}^q|F_{nj-1})$

for some $q > 2$. By choosing $q = 4$, the conclusion of the lemma is true if we can show

$$
E\left\{ \sum_{j=2}^{n_1+n_2} E(V_{nj}^4|F_{nj-1}) \right\} = o(\sigma_{n1}^4). \tag{2.27}
$$

We notice that

$$
\begin{aligned}
E\left\{ \sum_{j=2}^{n_1+n_2} E(V_{nj}^4|F_{nj-1}) \right\} &= \sum_{j=2}^{n_1+n_2} E(V_{nj}^4) = \sum_{j=2}^{n_1+n_2} E\left(\sum_{i=1}^{j-1} \phi_{ij} \right)^4 \\
&= \sum_{j=2}^{n_1+n_2} \sum_{i_1,i_2,i_3,i_4}^{j-1} E(\phi_{i_1 j}\phi_{i_2 j}\phi_{i_3 j}\phi_{i_4 j}).
\end{aligned}
$$

The last term can be decomposed as $3Q + P$ where

$$Q = \sum_{j=2}^{n_1+n_2} \sum_{s \neq t}^{j-1} E(\phi_{sj}^2 \phi_{tj}^2)$$

and $P = \sum_{j=2}^{n_1+n_2} \sum_{s=1}^{j-1} E(\phi_{sj})^4$. Now (2.27) is true if $3Q + P = o(\sigma_{n1}^4)$. We consider the

term Q and P separately. Note that

$$Q = O(n^{-8}) \sum_{j=2}^{n_1+n_2} \sum_{s \neq t}^{j-1} E\Big\{ tr(Y_j Y_j' Y_t Y_t' Y_j Y_j' Y_s Y_s') \Big\} = o(\sigma_{n1}^4).$$

The last equation follows the similar procedure in (2.24) in Lemma 2.2. It remains to

show $P = \sum_{j=2}^{n_1+n_2} \sum_{s=1}^{j-1} E(\phi_{sj})^4 = o(\sigma_{n1}^4)$. Note that

$$\begin{aligned}
P &= \sum_{j=2}^{n_1} \sum_{s=1}^{j-1} E(\phi_{sj})^4 + \sum_{j=n_1+1}^{n_1+n_2} \sum_{s=1}^{j-1} E(\phi_{sj})^4 \\
&= \sum_{j=2}^{n_1} \sum_{s=1}^{j-1} E(\phi_{sj})^4 + \sum_{j=n_1+1}^{n_1+n_2} \sum_{s=1}^{n_1} E(\phi_{sj})^4 + \sum_{j=n_1+1}^{n_1+n_2} \sum_{s=n_1+1}^{j-1} E(\phi_{sj})^4 \\
&= P_1 + P_2 + P_3,
\end{aligned}$$

where $P_1 = \sum_{j=2}^{n_1} \sum_{s=1}^{j-1} E(\phi_{sj})^4$, $P_2 = \sum_{j=n_1+1}^{n_1+n_2} \sum_{s=1}^{n_1} E(\phi_{sj})^4$ and

$$P_3 = \sum_{j=n_1+1}^{n_1+n_2} \sum_{s=n_1+1}^{j-1} E(\phi_{sj})^4.$$

Define $\Gamma_1' \Gamma_2 =: (v_{ij})_{m \times m}$, $\Gamma_1' \Sigma_2 \Gamma_1 =: (v_{ij}^{(2)})_{m \times m}$ and $(\Gamma_1' \Sigma_2 \Gamma_1)^2 =: (v_{ij}^{(4)})_{m \times m}$. Note

that the following facts which will be used repeatedly in the rest of this section:

$$\sum_{i,j=1}^{m} v_{ij}^4 \leq (\sum_{i,j=1}^{m} v_{ij}^2)^2 = tr^2\Big(\Gamma_1' \Gamma_2 \Gamma_2' \Gamma_1\Big) = tr^2(\Sigma_1 \Sigma_2),$$

$$\sum_{i=1}^{m} \sum_{j_1 \neq j_2}^{m} (v_{ij_1}^2 v_{ij_2}^2) \leq (\sum_{i,j=1}^{m} v_{ij}^2)^2 = tr^2(\Sigma_1 \Sigma_2),$$

$$\sum_{i_1 \neq i_2}^{m} \sum_{j_1 \neq j_2}^{m} v_{i_1 j_1} v_{i_1 j_2} v_{i_2 j_1} v_{i_2 j_2} \leq \sum_{i_1 \neq i_2}^{m} v_{i_1 i_2}^{(2)} v_{i_1 i_2}^{(2)} \leq \sum_{i_1, i_2=1}^{m} v_{i_1 i_2}^{(2)} v_{i_1 i_2}^{(2)},$$

$$\sum_{i_1, i_2=1}^{m} v_{i_1 i_2}^{(2)} v_{i_1 i_2}^{(2)} = \sum_{i_1=1}^{m} v_{i_1 i_1}^{(4)} = tr\Big(\Gamma_1' \Sigma_2 \Gamma_1 \Gamma_1' \Sigma_2 \Gamma_1\Big) = tr\Big\{ (\Sigma_1 \Sigma_2)^2 \Big\}.$$

Let us consider P_2 first. Note that

$$P_2 = \sum_{j=n_1+1}^{n_1+n_2} \sum_{s=1}^{n_1} E(\phi_{sj})^4 = O(n^{-8}) \sum_{j=n_1+1}^{n_1+n_2} \sum_{s=1}^{n_1} E\left(X_{1s}'X_{2j-n_1}\right)^4.$$

For P_2 term, let us focus on $X_{11}'X_{21}$ only (i.i.d.). Then, from (2.6),

$$E(X_{11}'X_{21})^4 = E(Z_{11}'\Gamma_1'\Gamma_2 Z_{21})^4 = E\left(\sum_{i,j=1}^{m} z_{11i}v_{ij}z_{21j}\right)^4$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{m}(3+\Delta)^2 v_{ij}^4 + \sum_{i=1}^{m}(3+\Delta)\sum_{j_1\neq j_2}^{m} v_{ij_1}^2 v_{ij_2}^2$$

$$+ \sum_{j=1}^{m}(3+\Delta)\sum_{i_1\neq i_2}^{m} v_{i_1j}^2 v_{i_2j}^2 + 9\sum_{i_1\neq i_2}^{m}\sum_{j_1\neq j_2}^{m} v_{i_1j_1}v_{i_1j_2}v_{i_2j_1}v_{i_2j_2}$$

$$= O\left\{tr^2(\Sigma_1\Sigma_2)\right\} + O\left[tr\left\{(\Sigma_1\Sigma_2)^2\right\}\right].$$

Then we conclude

$$P_2 = O(n^{-8}) \sum_{j=n_1+1}^{n_1+n_2} \sum_{s=1}^{n_1} \left(O\left\{tr^2(\Sigma_1\Sigma_2)\right\} + O\left[tr\left\{(\Sigma_1\Sigma_2)^2\right\}\right]\right)$$

$$= O(n^{-6})\left(O\left\{tr^2(\Sigma_1\Sigma_2)\right\} + O\left[tr\left\{(\Sigma_1\Sigma_2)^2\right\}\right]\right) = o(\sigma_{n1}^4).$$

We can also prove that $P_1 = o(\sigma_{n1}^4)$ and $P_3 = o(\sigma_{n1}^4)$ by going through the similar procedure. This completes the proof of the lemma.

Proof of Theorem 2.1:

We note equations (2.19) and (2.20) under conditions (2.9) and (2.10), respectively. Based on Corollary 3.1 of Hall and Heyde (1980), Lemma 2.1, Lemma 2.2 and Lemma 2.3, it can be concluded that $T_{n1}/\sigma_{n1} \xrightarrow{d} N(0,1)$. This implies the desired asymptotic normality of T_n under (2.9). Under (2.10), as T_{n2} is the sum of two independent averages, its asymptotic normality can be attained by following the standard means. Hence the theorem is proved.

Proof of Theorem 2.2:

We only present the proof for the ratio consistency of $\widehat{tr(\Sigma_1^2)}$ as the proofs of the other two follow the same route. We want to show

$$E\{\widehat{tr(\Sigma_1^2)}\} = tr(\Sigma_1^2)\{1 + o(1)\} \text{ and } Var\{\widehat{tr(\Sigma_1^2)}\} = o\{tr^2(\Sigma_1^2)\}. \qquad (2.28)$$

For notation simplicity, we denote X_{1j} as X_j, n_1 as n and Σ_1 as Σ, since we are effectively in a one-sample situation.

Note that

$$
\begin{aligned}
\widehat{tr(\Sigma^2)} &= \{n(n-1)\}^{-1}tr\Big\{\Sigma_{j\neq k}^n(X_j - \mu + \mu - \bar{X}_{(j,k)})(X_j - \mu + \mu)' \\
&\quad (X_k - \mu + \mu - \bar{X}_{(j,k)})(X_k - \mu + \mu)'\Big\} \\
&= \{n(n-1)\}^{-1}tr\Big[\sum_{j\neq k}^n \Big\{(X_j - \mu)(X_j - \mu)'(X_k - \mu)(X_k - \mu)' \\
&\quad - 2(\bar{X}_{(j,k)} - \mu)(X_j - \mu)'(X_k - \mu)(X_k - \mu)' + 2(X_j - \mu)\mu'(X_k - \mu)(X_k - \mu)' \\
&\quad - 2(\bar{X}_{(j,k)} - \mu)\mu'(X_k - \mu)(X_k - \mu)' + (\bar{X}_{(j,k)} - \mu)(X_j - \mu)'(\bar{X}_{(j,k)} - \mu)(X_k - \mu)' \\
&\quad - \{2(X_j - \mu)\mu'(\bar{X}_{(j,k)} - \mu)(X_k - \mu)' - 2(\bar{X}_{(j,k)} - \mu)\mu'(\bar{X}_{(j,k)} - \mu)(X_k - \mu)' \\
&\quad + (X_j - \mu)\mu'(X_k - \mu)\mu' - 2(\bar{X}_{(j,k)} - \mu)\mu'(X_k - \mu)\mu' \\
&\quad + (\bar{X}_{(j,k)} - \mu)\mu'(\bar{X}_{(j,k)} - \mu)\mu'\Big\}\Big] =: \sum_{l=1}^{10} tr(A_l), \text{say}.
\end{aligned}
$$

where A_l for $l = 1, 2, \cdots, 10$ are defined as:

$$A_1 = \frac{1}{n(n-1)}\sum_{j\neq k}^n (X_j - \mu)(X_j - \mu)'(X_k - \mu)(X_k - \mu)',$$

$$A_2 = -\frac{2}{n(n-1)}\sum_{j\neq k}^n (\bar{X}_{(j,k)} - \mu)(X_j - \mu)'(X_k - \mu)(X_k - \mu)',$$

$$A_3 = \frac{2}{n(n-1)}\sum_{j\neq k}^n (X_j - \mu)\mu'(X_k - \mu)(X_k - \mu)',$$

$$A_4 = -\frac{2}{n(n-1)}\sum_{j\neq k}^n (\bar{X}_{(j,k)} - \mu)\mu'(X_k - \mu)(X_k - \mu)',$$

$$A_5 = \frac{1}{n(n-1)} \sum_{j \neq k}^{n} (\bar{X}_{(j,k)} - \mu)(X_j - \mu)'(\bar{X}_{(j,k)} - \mu)(X_k - \mu)',$$

$$A_6 = -\frac{2}{n(n-1)} \sum_{j \neq k}^{n} (X_j - \mu)\mu'(\bar{X}_{(j,k)} - \mu)(X_k - \mu)',$$

$$A_7 = \frac{2}{n(n-1)} \sum_{j \neq k}^{n} (\bar{X}_{(j,k)} - \mu)\mu'(\bar{X}_{(j,k)} - \mu)(X_k - \mu)',$$

$$A_8 = \frac{1}{n(n-1)} \sum_{j \neq k}^{n} (X_j - \mu)\mu'(X_k - \mu)\mu',$$

$$A_9 = -\frac{2}{n(n-1)} \sum_{j \neq k}^{n} (\bar{X}_{(j,k)} - \mu)\mu'(X_k - \mu)\mu',$$

$$A_{10} = \frac{1}{n(n-1)} \sum_{j \neq k}^{n} (\bar{X}_{(j,k)} - \mu)\mu'(\bar{X}_{(j,k)} - \mu)\mu'.$$

It can be shown that $E\{tr(A_1)\} = tr(\Sigma^2)$, $E\{tr(A_i)\} = 0$ for $i = 2, \cdots, 9$ and $E\{tr(A_{10})\} = \mu'\Sigma\mu/(n-2) = o\{tr(\Sigma^2)\}$. The last equation is based on (2.9). This leads to the first part of (2.28). The second part is true given the sufficient conditions that

$$Var\Big\{tr(A_1)\Big\} = o\Big\{tr^2(\Sigma^2)\Big\} \text{ and } \frac{tr(A_l)}{tr(\Sigma^2)} = o_p(1), \text{ for } l = 2, 3, \cdots, 10.$$

Note that $tr(A_{10})$ is non-negative and $E\{tr(A_{10})\} = o\{tr(\Sigma^2)\}$. Then $tr(A_{10}) = o_p\{tr(\Sigma^2)\}$ since

$$P\Big\{\frac{tr(A_{10})}{tr(\Sigma^2)} > \epsilon\Big\} < \frac{E\{tr(A_{10})\}}{\epsilon tr(\Sigma^2)} = o(1).$$

We shall only show $Var\{tr(A_1)\} = o\{tr(\Sigma^2)\}$ here. Derivations for other $Var\{tr(A_i)\}$ are quite similar. Note that

$$Var\{tr(A_1)\} = \frac{1}{n^2(n-1)^2} E\Big[tr\Big\{ \sum_{j_1 \neq k_1}^{n} (X_{j_1} - \mu)(X_{j_1} - \mu)'(X_{k_1} - \mu)(X_{k_1} - \mu)'\Big\}$$

$$\times tr\Big\{ \sum_{j_2 \neq k_2}^{n} (X_{j_2} - \mu)(X_{j_2} - \mu)'(X_{k_2} - \mu)(X_{k_2} - \mu)'\Big\}\Big] - tr^2(\Sigma^2).$$

It can be shown, by considering the possible combinations of the subscripts j_1, k_1, j_2 and k_2, that

$$
\begin{aligned}
Var\{tr(A_1)\} \ &= \ \frac{2E\{(X_1 - \mu)'(X_2 - \mu)\}^4}{n(n-1)} + \frac{4(n-2)}{n(n-1)}E\{(X_1 - \mu)'\Sigma(X_1 - \mu)\}^2 \\
&\quad + \frac{(n-2)(n-3)}{n(n-1)}tr^2(\Sigma^2) - tr^2(\Sigma^2) \\
&=: \ \frac{2}{n(n-1)}B_{11} + \frac{4(n-2)}{n(n-1)}B_{12} + o\{tr^2(\Sigma^2)\}, \qquad\qquad (2.29)
\end{aligned}
$$

where $\frac{4(n-2)}{n(n-1)}B_{12} = o\{tr^2(\Sigma^2)\}$ which can be shown based on (2.24). Note that

$$
\begin{aligned}
B_{11} &= E(Z_1'\Gamma'\Gamma Z_2)^4 = E\Big(\sum_{s,t=1}^{m} z_{1s}\nu_{st}z_{2t}\Big)^4 \\
&= E\Bigg(\sum_{s_1,s_2,s_3,s_4,t_1,t_2,t_3,t_4=1}^{m} \nu_{s_1t_1}\nu_{s_2t_2}\nu_{s_3t_3}\nu_{s_4t_4}z_{1s_1}z_{1s_2}z_{1s_3}z_{1s_4}z_{2t_1}z_{2t_2}z_{2t_3}z_{2t_4}\Bigg) \\
&= \sum_{j=1}^{16} B_{11}^{(j)}.
\end{aligned}
$$

Define $tr(\Gamma\Gamma') = (\nu_{st})_{m \times m}$. The following combinations of the subscripts s_1,s_2,s_3,s_4 and t_1,t_2,t_3,t_4 lead to none zero $B_{11}^{(j)}$:

(1) if $(s_1 = s_2 = s_3 = s_4), (t_1 = t_2 = t_3 = t_4) : B_{11}^{(1)} = \displaystyle\sum_{s,t=1}^{m} \nu_{st}^4(3 + \Delta)^2;$

(2) if $(s_1 = s_2 \neq s_3 = s_4), (t_1 = t_2 = t_3 = t_4) : B_{11}^{(2)} = \displaystyle\sum_{s_1 \neq s_2, t=1}^{m} \nu_{s_1t}^2\nu_{s_2t}^2(3 + \Delta);$

(3) if $(s_1 = s_3 \neq s_2 = s_4), (t_1 = t_2 = t_3 = t_4) : B_{11}^{(3)} = \displaystyle\sum_{s_1 \neq s_2, t=1}^{m} \nu_{s_1t}^2\nu_{s_2t}^2(3 + \Delta);$

(4) if $(s_1 = s_4 \neq s_2 = s_3), (t_1 = t_2 = t_3 = t_4) : B_{11}^{(4)} = \displaystyle\sum_{s_1 \neq s_2, t=1}^{m} \nu_{s_1t}^2\nu_{s_2t}^2(3 + \Delta);$

(5) if $(s_1 = s_2 = s_3 = s_4), (t_1 = t_2 \neq t_3 = t_4) : B_{11}^{(5)} = \displaystyle\sum_{t_1 \neq t_2, s=1}^{m} \nu_{st_1}^2\nu_{st_2}^2(3 + \Delta);$

(6) if $(s_1 = s_2 = s_3 = s_4), (t_1 = t_3 \neq t_2 = t_4) : B_{11}^{(6)} = \displaystyle\sum_{t_1 \neq t_2, s=1}^{m} \nu_{st_1}^2\nu_{st_2}^2(3 + \Delta);$

(7) if $(s_1 = s_2 = s_3 = s_4), (t_1 = t_4 \neq t_2 = t_3) : B_{11}^{(7)} = \sum_{t_1 \neq t_2, s=1}^{m} \nu_{st_1}^2 \nu_{st_2}^2 (3 + \Delta);$

(8) if $(s_1 = s_2 \neq s_3 = s_4), (t_1 = t_2 \neq t_3 = t_4) : B_{11}^{(8)} = \sum_{t_1 \neq t_2, s_1 \neq s_2 = 1}^{m} \nu_{s_1 t_1}^2 \nu_{s_2 t_2}^2;$

(9) if $(s_1 = s_2 \neq s_3 = s_4), (t_1 = t_3 \neq t_2 = t_4) : B_{11}^{(9)} = \sum_{t_1 \neq t_2, s_1 \neq s_2 = 1}^{m} \nu_{s_1 t_1} \nu_{s_1 t_2} \nu_{s_2 t_1} \nu_{s_2 t_2};$

(10) if $(s_1 = s_2 \neq s_3 = s_4), (t_1 = t_4 \neq t_2 = t_3) : B_{11}^{(10)} = \sum_{t_1 \neq t_2, s_1 \neq s_2 = 1}^{m} \nu_{s_1 t_1} \nu_{s_1 t_2} \nu_{s_2 t_2} \nu_{s_2 t_1};$

(11) if $(s_1 = s_3 \neq s_2 = s_4), (t_1 = t_2 \neq t_3 = t_4) : B_{11}^{(11)} = \sum_{t_1 \neq t_2, s_1 \neq s_2 = 1}^{m} \nu_{s_1 t_1} \nu_{s_2 t_1} \nu_{s_1 t_2} \nu_{s_2 t_2};$

(12) if $(s_1 = s_3 \neq s_2 = s_4), (t_1 = t_3 \neq t_2 = t_4) : B_{11}^{(12)} = \sum_{t_1 \neq t_2, s_1 \neq s_2 = 1}^{m} \nu_{s_1 t_1}^2 \nu_{s_2 t_2}^2;$

(13) if $(s_1 = s_3 \neq s_2 = s_4), (t_1 = t_4 \neq t_2 = t_3) : B_{11}^{(13)} = \sum_{t_1 \neq t_2, s_1 \neq s_2 = 1}^{m} \nu_{s_1 t_1} \nu_{s_2 t_2} \nu_{s_1 t_2} \nu_{s_2 t_1};$

(14) if $(s_1 = s_4 \neq s_2 = s_3), (t_1 = t_2 \neq t_3 = t_4) : B_{11}^{(14)} = \sum_{t_1 \neq t_2, s_1 \neq s_2 = 1}^{m} \nu_{s_1 t_1} \nu_{s_2 t_1} \nu_{s_2 t_2} \nu_{s_1 t_2};$

(15) if $(s_1 = s_4 \neq s_2 = s_3), (t_1 = t_3 \neq t_2 = t_4) : B_{11}^{(15)} = \sum_{t_1 \neq t_2, s_1 \neq s_2 = 1}^{m} \nu_{s_1 t_1}^2 \nu_{s_2 t_2}^2;$

(16) if $(s_1 = s_4 \neq s_2 = s_3), (t_1 = t_4 \neq t_2 = t_3) : B_{11}^{(16)} = \sum_{t_1 \neq t_2, s_1 \neq s_2 = 1}^{m} \nu_{s_1 t_1}^2 \nu_{s_2 t_2}^2.$

Note that $tr^2(\Sigma^2) = tr^2(\Gamma \Gamma' \Gamma \Gamma') = (\sum_{s,t=1}^{m} \nu_{st}^2)^2 = \sum_{s_1, s_2, t_1, t_2 = 1}^{m} \nu_{s_1 t_1}^2 \nu_{s_2 t_2}^2$ and

$$tr(\Sigma^4) = tr(\Gamma' \Gamma \Gamma' \Gamma \Gamma' \Gamma \Gamma' \Gamma) = \sum_{t_1, t_2, s_1, s_2 = 1}^{m} \nu_{s_1 t_1} \nu_{s_1 t_2} \nu_{s_2 t_1} \nu_{s_2 t_2}.$$

It can be shown that $B_{11} \leq c \, tr^2(\Sigma^2)$ for a finite positive number c and hence $2\left\{n(n-1)\right\}^{-1} B_{11} = o\left\{tr^2(\Sigma^2)\right\}$. Therefore, from (2.29), $Var\left\{tr(A_1)\right\} = o\left\{tr^2(\Sigma^2)\right\}$. By following the similar procedure, we can also prove that $Var\{tr(A_j)\} = o\{tr^2(\Sigma^2)\}$, for $j = 2, \cdots, 9$. In conclusion, $Var(\widehat{tr(\Sigma^2)}) = o(tr^2(\Sigma^2))$. Therefore, $\widehat{tr(\Sigma^2)}$ is a ratio consistent estimator of $tr(\Sigma^2)$. This completes the proof.

Table 2.1   Empirical power and size for the 2-dependence model with gamma innovation

Type of	% of	$p = 500, n = 124$				$p = 1000, n = 138$			
Allocation	True Null	NEW	BS	Bonf	FDR	NEW	BS	Bonf	FDR
Equal	0%	.511	.399	.13	.16	.521	.413	.11	.16
	25%	.521	.387	.14	.16	.518	.410	.12	.16
	50%	.513	.401	.13	.17	.531	.422	.12	.17
	75%	.522	.389	.13	.18	.530	.416	.11	.17
	95%	.501	.399	.14	.16	.500	.398	.13	.17
	99%	.499	.388	.13	.15	.507	.408	.15	.18
	100%(size)	.043	.043	.040	.041	.043	.042	.042	.042
Increasing	0%	.520	.425	.11	.13	.522	.409	.12	.15
	25%	.515	.431	.12	.15	.523	.412	.14	.16
	50%	.512	.412	.13	.15	.528	.421	.15	.17
	75%	.522	.409	.15	.17	.531	.431	.16	.19
	95%	.488	.401	.14	.15	.500	.410	.15	.17
	99%	.501	.409	.15	.17	.511	.412	.15	.16
	100%(size)	.042	.041	.040	.041	.042	.040	.039	.041
Decreasing	0%	.522	.395	.11	.15	.533	.406	.09	.15
	25%	.530	.389	.11	.15	.530	.422	.11	.17
	50%	.528	.401	.12	.17	.522	.432	.12	.17
	75%	.533	.399	.13	.18	.519	.421	.12	.17
	95%	.511	.410	.12	.15	.508	.411	.15	.18
	99%	.508	.407	.14	.15	.507	.418	.16	.17
	100%(size)	.041	.042	.041	.042	.042	.040	.040	.042

Table 2.2 Empirical power and size for the full-dependence model with gamma innovation

Type of Allocation	% of True Null	$p = 500, n = 124$				$p = 1000, n = 138$			
		NEW	BS	Bonf	FDR	NEW	BS	Bonf	FDR
Equal	0%	.322	.120	.08	.10	.402	.216	.09	.11
	25%	.318	.117	.08	.10	.400	.218	.08	.11
	50%	.316	.115	.09	.11	.409	.221	.09	.10
	75%	.307	.113	.10	.12	.410	.213	.09	.13
	95%	.233	.128	.11	.14	.308	.215	.10	.13
	99%	.225	.138	.12	.15	.316	.207	.11	.12
	100%(size)	.041	.041	.043	.043	.042	.042	.040	.041
Increasing	0%	.331	.121	.09	.12	.430	.225	.10	.11
	25%	.336	.119	.10	.12	.423	.231	.12	.12
	50%	.329	.123	.12	.14	.422	.226	.13	.14
	75%	.330	.115	.12	.15	.431	.222	.14	.15
	95%	.219	.120	.12	.13	.311	.218	.14	.15
	99%	.228	.117	.13	.15	.315	.217	.15	.17
	100%(size)	.041	.040	.042	.043	.042	.042	.040	.042
Decreasing	0%	.320	.117	.08	.11	.411	.213	.08	.10
	25%	.323	.119	.09	.11	.408	.210	.08	.11
	50%	.327	.120	.11	.12	.403	.208	.09	.10
	75%	.322	.122	.12	.12	.400	.211	.12	.13
	95%	.217	.109	.12	.15	.319	.207	.12	.15
	99%	.224	.111	.13	.16	.327	.205	.11	.13
	100%(size)	.042	.043	.039	.041	.042	.211	.040	.041

Table 2.3   Empirical power and size for the 2-dependence model with normal innovation

Type of	% of	$p = 500, n = 124$				$p = 1000, n = 138$			
Allocation	True Null	NEW	BS	Bonf	FDR	NEW	BS	Bonf	FDR
Equal	0%	.533	.411	.11	.16	.533	.405	.13	.16
	25%	.526	.409	.14	.18	.524	.399	.12	.16
	50%	.522	.421	.13	.19	.531	.411	.15	.18
	75%	.529	.416	.15	.18	.529	.413	.15	.18
	95%	.501	.395	.14	.16	.511	.410	.15	.16
	99%	.500	.401	.15	.16	.509	.411	.16	.17
	100%(size)	.044	.042	.042	.043	.044	.043	.042	.043
Increasing	0%	.524	.388	.12	.16	.531	.413	.11	.17
	25%	.530	.390	.11	.17	.530	.420	.12	.16
	50%	.531	.402	.10	.18	.522	.421	.14	.17
	75%	.528	.406	.12	.18	.529	.409	.13	.18
	95%	.511	.411	.11	.13	.513	.416	.14	.15
	99%	.510	.412	.12	.13	.509	.413	.13	.15
	100%(size)	.043	.041	.041	.042	.044	.043	.041	.042
Decreasing	0%	.529	.417	.13	.17	.529	.420	.11	.15
	25%	.527	.413	.13	.16	.535	.409	.12	.17
	50%	.531	.420	.12	.18	.520	.422	.13	.18
	75%	.532	.421	.15	.17	.531	.418	.13	.18
	75%	.508	.412	.13	.15	.512	.400	.14	.16
	99%	.500	.399	.13	.16	.503	.410	.15	.16
	100%(size)	.042	.042	.041	.043	.042	.041	.042	.042

Table 2.4   Empirical power and size for the full-dependence model with normal innovation

Type of	% of	$p = 500, n = 124$				$p = 1000, n = 138$			
Allocation	True Null	NEW	BS	Bonf	FDR	NEW	BS	Bonf	FDR
Equal	0%	.303	.111	.06	.08	.410	.220	.08	.11
	25%	.301	.109	.07	.08	.421	.225	.07	.10
	50%	.302	.109	.07	.10	.418	.231	.09	.11
	75%	.289	.111	.08	.11	.326	.233	.10	.12
	95%	.277	.104	.09	.10	.321	.162	.11	.15
	99%	.279	.100	.08	.11	.318	.171	.12	.16
	100%(size)	.041	.039	.041	.041	.042	.040	.041	.042
Increasing	0%	.318	.108	.09	.10	.401	.210	.07	.09
	25%	.321	.110	.08	.09	.403	.215	.08	.10
	50%	.320	.112	.09	.10	.399	.213	.08	.10
	75%	.315	.115	.10	.12	.339	.209	.10	.12
	95%	.234	.101	.10	.11	.322	.155	.11	.14
	99%	.219	.102	.09	.11	.308	.169	.13	.14
	100%(size)	.042	.041	.040	.041	.041	.043	.039	.040
Decreasing	0%	.226	.107	.07	.09	.379	.218	.08	.09
	25%	.229	.116	.07	.09	.385	.209	.08	.10
	50%	.232	.119	.08	.10	.380	.216	.09	.11
	75%	.230	.111	.08	.10	.391	.220	.10	.12
	95%	.217	.124	.09	.12	.324	.204	.09	.13
	99%	.221	.120	.10	.11	.322	.208	.10	.14
	100%(size)	.042	.042	.041	.041	.041	.040	.039	.040

Table 2.5 Empirical averages of $\widehat{tr(\Sigma^2)}/tr(\Sigma^2)$ with standard deviations in the parentheses

Type of Innovation	Type of Dependence	NEW	BS	$tr(\Sigma^2)$
		$p = 500, n = 124$		
Normal	2-Dependence	1.03 (0.015)	1.39 (0.016)	3102
	Full-Dependence	1.008 (0.00279)	1.17 (0.0032)	35911
Gamma	2-Dependence	1.03 (0.006)	1.10 (0.007)	14227
	Full-Dependence	1.108 (0.0019)	1.248 (0.0017)	152248
		$p = 1000, n = 138$		
Normal	2-Dependence	0.986 (0.0138)	1.253 (0.0136)	6563
	Full-Dependence	0.995 (0.0026)	1.072 (0.0033)	76563
Gamma	2-Dependence	1.048 (0.005)	1.138 (0.006)	32104
	Full-Dependence	1.088 (0.00097)	1.231 (0.0013)	325879

Table 2.6   Empirical power and size for the sparse model

Sample Size ($n_1 = n_2$)	Methods	$\varepsilon=.25$				$\varepsilon=.15$			
		c=.25		c=.45		c=.35		c=.55	
		Power	Size	Power	Size	Power	Size	Power	Size
10	FDR	.084	.056	.180	.040	.044	.034	.066	.034
	Bonf	.084	.056	.170	.040	.044	.034	.062	.032
	New	.100	.046	.546	.056	.072	.064	.344	.064
20	FDR	.380	.042	.855	.044	.096	.036	.326	.058
	Bonf	.368	.038	.806	.044	.092	.034	.308	.056
	New	.238	.052	.976	.042	.106	.052	.852	.046
30	FDR	.864	.042	1	.060	.236	.048	.710	.038
	Bonfe	.842	.038	.996	.060	.232	.048	.660	.038
	New	.408	.050	.998	.058	.220	.054	.988	.042

Table 2.7   Average ratios of $\widehat{\sigma^2_M}/\sigma^2_M$ and their standard deviation (in parenthesis) for the sparse model

Sample Size	True σ^2_M	$\varepsilon=.25$		$\varepsilon=.15$	
		c=0.25	c=0.45	c=0.35	c=0.55
$n_1 = n_2 = 10$	84.4	1.003(.0123)	1.005 (.0116)	.998 (.0120)	.999(.0110)
$n_1 = n_2 = 20$	20.5	1.003(.0033)	1.000 (.0028)	1.003(.0028)	1.002(.0029)
$n_1 = n_2 = 30$	9.0	.996(.0013)	.998(.0013)	1.004(.0014)	.999(.0013)
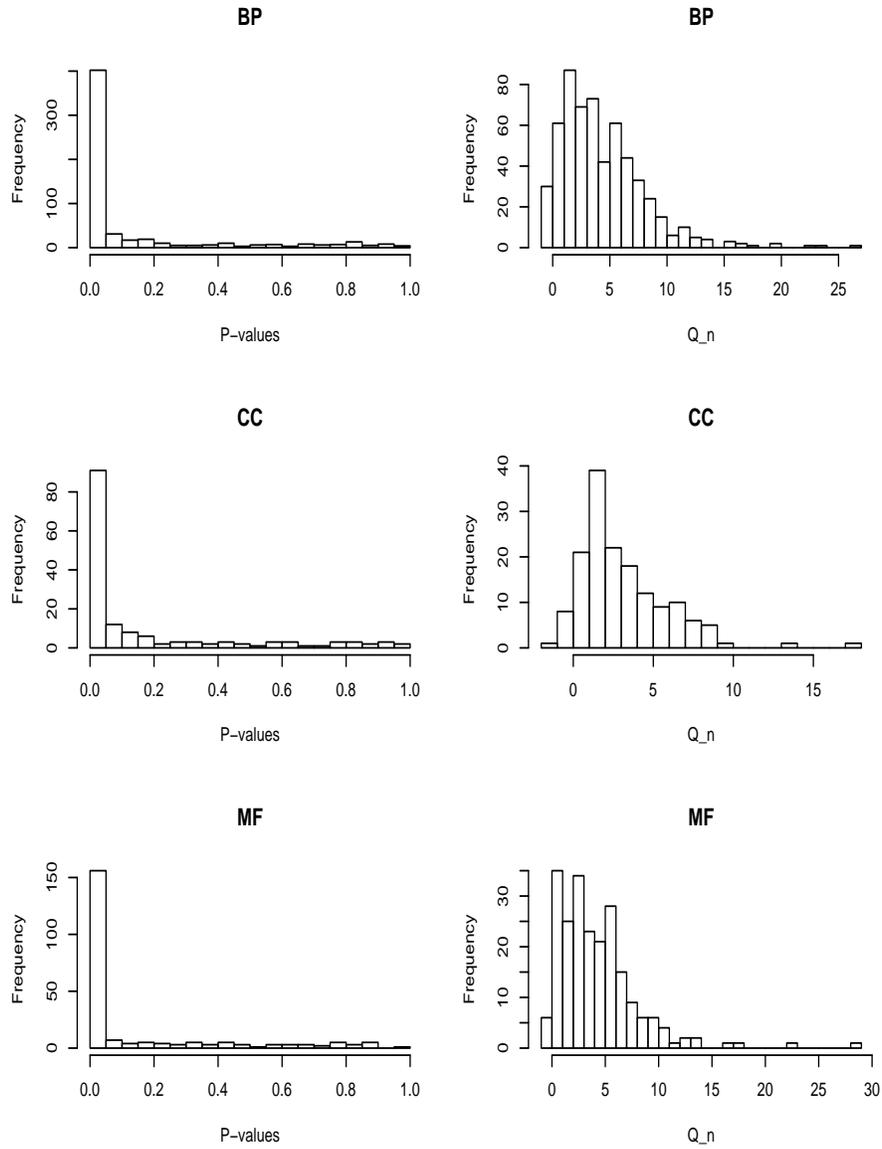
Figure 2.1   Two-sample tests for differentially expressed gene-sets between BCR/ABL and NEG class B-cell ALL: Histograms of P-values (left panels) and Q_n values (right panels) for BP, CC and MF gene categories.
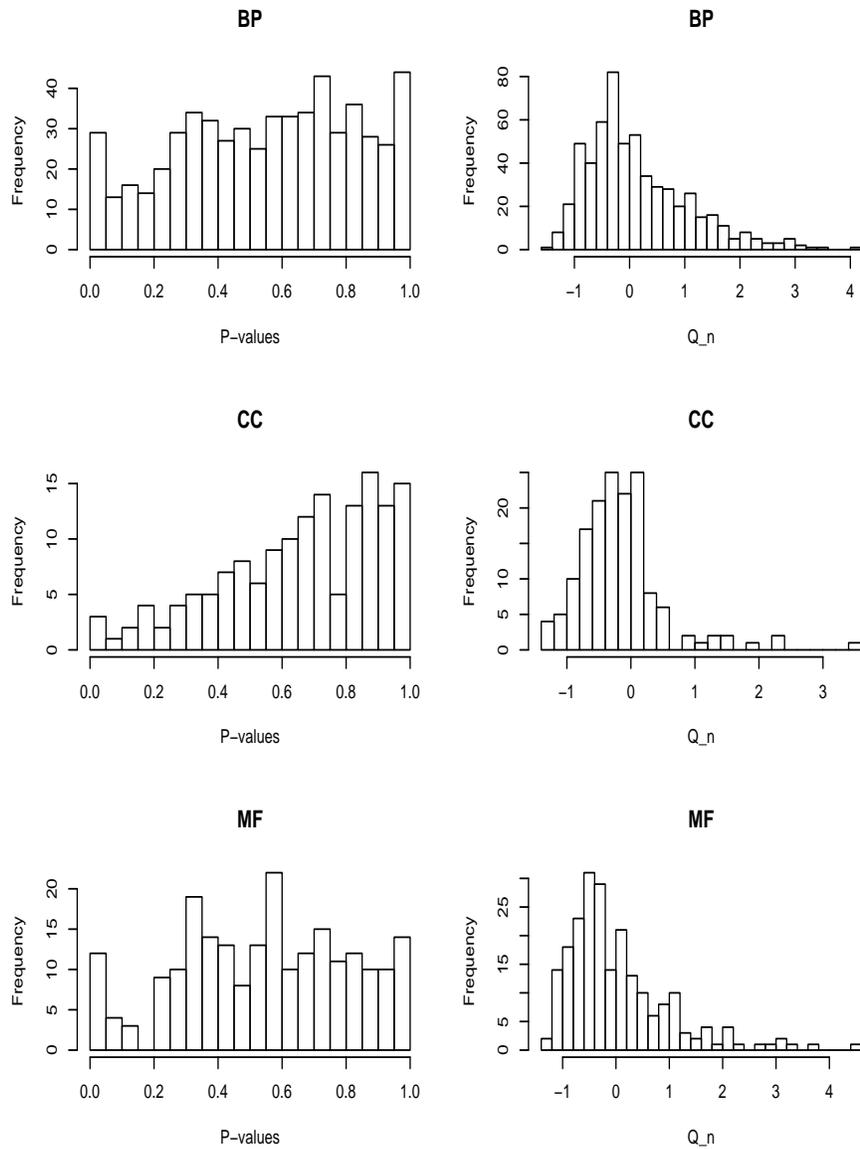
Figure 2.2   Back-testing for differentially expressed gene-sets between two randomly assigned B-cell NEG groups: Histograms of P-values (left panels) and Q_n values (right panels) for BP, CC and MF gene categories.

CHAPTER 3.   Effects of Data Dimension on Empirical Likelihood

3.1   Introduction

Since Owen (1988, 1990) introduced the empirical likelihood method, it has been extended to many settings as a tool for nonparametric and semiparametric statistical inference. Its most attractive property is its permitting likelihood-like inference in a nonparametric or semiparametric setting. This is largely due to its sharing two key features with the conventional likelihood: Wilks theorem and Bartlett correction (Hall and La Scala, 1990; DiCiccio et al., 1991; Chen and Cui, 2006). See Owen (2001) for an overview.

High-dimensional data are increasingly common; for instance, in genetic and mRNA microarray analysis, marketing research and financial applications. There is a rapidly expanding literature on multivariate analysis where the data dimension p depends on the sample size n and grows to infinity as $n \to \infty$; see, for example, Portnoy (1984, 1985) in the context of M-estimation, Bai and Saranadasa (1996) for a two-sample test for means, Ledoit and Wolf (2002) and Schott (2005) for testing a specific covariance structure, and Schott (2007) for tests with more than two samples.

Given the interest in both high-dimensional data and empirical likelihood, there is a need to evaluate the behavior of the latter when the data dimension and the sample size

increase simultaneously. In this chapter, we evaluate the effects of the data dimension and dependence on the asymptotic normality of the empirical likelihood ratio statistic for the mean.

Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random vectors in R^p with mean vector $\mu = (\mu_1, \ldots, \mu_p)^T$ and non-singular covariance matrix Σ. Let

$$L_n(\mu) = \sup\left\{ \Pi_{i=1}^n \pi_i : \pi_i \geq 0, \sum_{i=1}^n \pi_i = 1, \sum_{i=1}^n \pi_i X_i = \mu \right\} \tag{3.1}$$

be the empirical likelihood for μ and $w_n(\mu) = -2 \log\{n^n L_n(\mu)\}$ be the empirical likelihood ratio statistic. When p is fixed, Owen (1988, 1990) showed that

$$w_n(\mu) \to \chi_p^2 \tag{3.2}$$

in distribution as $n \to \infty$, which mimics Wilks' theorem for parametric likelihood ratios. An extension of the above result for parameters defined by general estimating equations is given in Qin and Lawless (1994).

As $p \to \infty$ for high-dimensional data, the natural substitute for (3.2) is

$$(2p)^{-1/2}\{w_n(\mu) - p\} \to N(0, 1) \tag{3.3}$$

in distribution as $n \to \infty$, since χ_p^2 is asymptotic normal with mean p and variance $2p$. A key question is how large the dimension p can be while (3.3) is valid. In a recent study, Hjort et al. (2009) have established that it is $p = o(n^{1/3})$ under the assumptions:

(A1). The eigen-values of Σ are uniformly bounded away from zero and infinity, and

(A2). Each component of X_i is a uniformly bounded random variable.

When *(A2)* is relaxed, we have:

(A2'). $E||X_i/\sqrt{p}||^q$ and $p^{-1} \sum_{j=1}^p E|X_i^{(j)} - \mu_j|^q$ are bounded for some $q \geq 4$, where $||\cdot||$ is the Euclidean norm. Hjort et al. (2009) showed that (3.3) is valid if $p^{3+6/(q-2)}/n \to 0$.

When $q = 4$ in $(A2')$, it means $p = o(n^{1/6})$. Hence, there is a significant decrease in the rate at which $p \to \infty$ when $(A2)$ is weakened. Tsao (2004) found that, when p is moderately large but fixed, the distribution of $w_n(\mu)$ has an atom at infinity for fixed n: the probability of $w_n(\mu) = \infty$ is non-zero. Tsao showed that, if p and n increase at the same rate such that $p/n \geq 0.5$, the probability of $w_n(\mu) = \infty$ converges to 1 since the probability of μ being contained in the convex hull of the sample converges to 0. These reveal the effects of p on the empirical likelihood from another perspective.

In this chapter, we analyze the empirical likelihood for high-dimensional data under a general multivariate model, which facilitates a more detailed analysis than Hjort et al. (2009) and allows less restrictive conditions. The analysis requires neither the largest eigenvalue of Σ nor $E||X_i/\sqrt{p}||^q$ to be bounded, and hence accommodates a wider range of dependencies among components of X_i.

Our main finding is that the effect of the dimensionality and the dependence among components of X_i on the empirical likelihood are leveraged through $\text{tr}(\Sigma)$, the trace of the covariance matrix Σ and its largest eigenvalue λ_p. We provide a general rate for the dimension p, which is shown to be dependent on $\text{tr}(\Sigma)$ and λ_p. In particular, under assumptions $(A1)$ and $(A2)$, $p = o(n^{1/2})$, which improves $p = o(n^{1/3})$ of Hjort et al. (2009). This is likely to be the best rate for p in the context of the empirical likelihood as $p = o(n^{1/2})$ is the sufficient and necessary condition for the convergence of the sample covariance matrix to Σ under the trace-norm when all the eigenvalues of Σ are bounded.

Empirical likelihood is known for manifesting its higher order terms in an elegant fashion so that it has attractive higher order properties, for instance the Bartlett correction, as recently shown in Chen and Cui (2006) for general estimating equations. While the involvement of the higher order terms is attractive for a fixed p, we find

for high-dimensional data these "so-called" higher order terms may not be of higher-order anymore as they can emerge as terms of the same magnitude or larger than the previous leading term in the fixed p case. This is the reason for imposing a restriction on the rate of increasing of p so that those higher order terms for the fixed p case stay as the higher order terms when p is allowed to increase as the sample size increases.

The above remark is well supported by our analysis on the performance of the least square empirical likelihood (Owen, 1991; Brown and Chen, 1998) for high-dimensional data. Least square empirical likelihood is a simplified version of the empirical likelihood. For fixed p, it is equivalent to the empirical likelihood in the leading order and easily computable. However, it is not Bartlett correctable due to an incomplete higher order structure. The latter (a lighter higher order term) turns out to be an advantage when the data dimension is high. Indeed, we find the least square empirical likelihood allows $p = o(n^{2/3})$ under $(A1)$ to ensure a least square version of (3.3) is valid. This improves the rate given by Theorem 3.3 for the empirical likelihood ratio under the corresponding condition.

This chapter is organized as follows. The outline of some preliminary formulation is provided in Section 3.2. Section 3.3 contains the main results which quantify the effects of dimension on the empirical likelihood. Section 3.4 reports some numerical results. An application to Dow Johns Industrial Average data is presented in Section 3.5. Some technical details are given in Section 3.6.

3.2   Preliminaries

Suppose that each of the i.i.d. observations $X_i \in R^p$ is specified by $X_i = \Gamma Z_i + \mu$, where Γ is a $p \times m$ matrix, $m \geq p$ and $Z_i = (Z_{i1}, \ldots, Z_{im})^T$ is a random vector such

that

$$E(Z_i) = 0, \text{var}(Z_i) = I_m, E(Z_{il}^{4k}) = m_{4k} \in (0, \infty), \tag{3.4}$$

$$E\left(Z_{il_1}^{\alpha_1} Z_{il_2}^{\alpha_2} \cdots Z_{il_q}^{\alpha_q}\right) = E(Z_{il_1}^{\alpha_1})E(Z_{il_2}^{\alpha_2}) \cdots E(Z_{il_q}^{\alpha_q})$$

whenever $\sum_{l=1}^{q} \alpha_l \leq 4k$ and $l_1 \neq l_2 \neq \cdots \neq l_q$. Here k is some positive integer and I_m is the m-dimensional identity matrix.

The above multivariate model, employed in Bai and Saranadasa (1996), means that each X_i is a linear transformation of some m-variate random vector Z_i. An important feature is that m, the dimension of Z_i, is left arbitrary provided $m \geq p$ and $\Gamma\Gamma^T = \Sigma$ which can generate a rich collection of X_i from Z_i with the given covariance Σ. It also requires that power transformations of different components of Z_i are uncorrelated, which is weaker than assuming that they are independent. The model (3.4) encompasses a rich collection of multivariate models. It includes the elliptically contoured distributions with $Z_i = RU^{(m)}$ where R is a non-negative random variable and $U^{(m)}$ is the uniform random vector on the unit sphere (Fang and Zhang, 1990). The multivariate normal and t-distribution are elliptically contoured and so is a mixture of normal distributions whose density is defined by $\int n(x|\mu, v^{-2}\Sigma)dw(v)$ where $n(x|\mu, \Sigma)$ is the density of $N(\mu, \Sigma)$ and $w(v)$ is the distribution function of a non-negative univariate random variable (Anderson, 2003). Both the moment and the correlation conditions are imposed on Z_i rather than X_i. This model structure allows the moments of $||X_i - \mu||^{2k}$ to be derived and allows us to conduct a more detailed analysis than possible in Hjort et al. (2009).

The integer k determines the number of finite moments for Z_{il}. As $k \geq 1$, each Z_{il} has at least finite fourth moment. This is the minimal moment condition to ensure the convergence of the largest eigenvalue of the sample covariance matrix to the largest eigenvalue of Σ (Yin et al., 1988; Bai et al., 1988), and hence the convergence of the

sample covariance matrix to Σ under the matrix norm based on the largest eigenvalue. By inspecting the proofs given in Section 3.6, we see that a divergent sample covariance matrix would dramatically alter the asymptotic mean and variance of the empirical likelihood ratio. Hence, it is unclear if (3.3) would remain true.

From the standard empirical likelihood solutions (Owen, 1988, 1990) which are valid for any p, fixed or growing, the optimal weights π_i for the optimization problem (3.1) are

$$\pi_i = \frac{1}{n}\frac{1}{1 + \lambda^T(X_i - \mu)},$$

where $\lambda \in R^p$ is a Lagrange multiplier satisfying

$$g(\lambda) = \sum_{i=1}^{n}\frac{X_i - \mu}{1 + \lambda^T(X_i - \mu)} = 0. \tag{3.5}$$

Hence, the empirical likelihood $L_n(\mu)$ equals $n^{-n}\prod_{i=1}^{n}\{1 + \lambda^T(X_i - \mu)\}^{-1}$. As the maximum empirical likelihood is attained at $\pi_i = n^{-1}$ $(i = 1, \ldots, n)$, the empirical likelihood ratio for μ is

$$w_n(\mu) = -2\log\{n^n L_n(\mu)\} = 2\sum_{i=1}^{n}\log\{1 + \lambda^T(X_i - \mu)\}. \tag{3.6}$$

Throughout the chapter we let $\gamma_1(A) \leq \cdots \leq \gamma_p(A)$ denote the eigenvalues and $\text{tr}(A)$ denote the trace operator of a matrix A. When $A = \Sigma$, we write $\gamma_j(\Sigma)$ as γ_j, $(j = 1, \ldots, p)$. It is assumed throughout the chapter that $\gamma_1 \geq C_1$ for some positive constant C_1.

3.3 Effects of High-Dimension

The Lagrange multiplier λ defined in (3.5) is a key element in any empirical likelihood formulation, and reflects the implicit nature of the methodology. When p is

fixed, Owen (1990) showed that

$$||\lambda|| = O_p(n^{-1/2}).\tag{3.7}$$

This has been the prevailing order for λ except in nonparametric curve estimation where n is replaced by the "effective sample size" (Chen, 1996). When p grows with n, (3.7) is no longer valid.

Theorem 3.1 *If* $\{tr(\Sigma)\}^{4k-1}\gamma_p = O(n^{2k-1})$ *and* $\gamma_p^2 p^2 = o(n)$, *then*

$$||\lambda|| = O_p[\{tr(\Sigma)/n\}^{1/2}].$$

Theorem 3.1 implies that the effect of the dimension and dependence among components of X_i on the Lagrange multiplier is directly determined through $tr(\Sigma)$ and γ_p. The rate for $||\lambda||$ can be regarded as a generalization of (3.7) for a fixed p since $O_p[\{tr(\Sigma)/n\}^{1/2}]$ degenerates to $O_p(n^{-1/2})$ in that case.

We first study the effects of dimension on the asymptotic normality of $w_n(\mu)$, assuming existence of the minimal fourth moment for each Z_{il}. Later, we will increase the number of moments. We assume for the time being that $k = 1$ in (3.4) and $tr^5(\Sigma)\gamma_p^5 = o(np)$. Since $p\gamma_1 \leq tr(\Sigma) \leq p\gamma_p$, this implies the conditions of Theorem 3.1.

We wish to establish an expansion for $w_n(\mu)$. Put $W_i = \lambda^T(X_i - \mu)$. From (3.22) of Section 3.6, $\max_{i=1,\ldots,n} |W_i| = o_p(1)$, which allows

$$\log\{1 + \lambda^T(X_i - \mu)\} = W_i - W_i^2/2 + W_i^3/(1 + \xi_{i1})^4\tag{3.8}$$

where $|\xi_{i1}| \leq |\lambda^T(X_i - \mu)|$. Expand (3.5) so that

$$0 = g(\lambda) = \bar{X} - \mu - S_n\lambda + \beta_n$$

where $\beta_n = n^{-1}\sum_{i=1}^n (X_i - \mu)W_i^2/(1 + \xi_i)^3$ for some $|\xi_i| \leq |\lambda^T(X_i - \mu)|$ and $S_n = n^{-1}\sum_{i=1}^n (X_i - \mu)(X_i - \mu)^T$. Hence,

$$\lambda = S_n^{-1}(\bar{X} - \mu) + S_n^{-1}\beta_n.\tag{3.9}$$

From (3.8) and (3.26), we obtain an expansion for $w_n(\mu)$:

$$
\begin{aligned}
w_n(\mu) &= n(\bar{X} - \mu)^T S_n^{-1}(\bar{X} - \mu) - n\beta_n S_n^{-1}\beta_n + \tfrac{2}{3}\sum_{i=1}^n \{\lambda^T(X_i - \mu)\}^3/(1 + \xi_i)^4 \\
&= n(\bar{X} - \mu)^T \Sigma^{-1}(\bar{X} - \mu) + n(\bar{X} - \mu)^T(S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) \\
&\quad - n\beta_n S_n^{-1}\beta_n + \tfrac{2}{3}R_n\{1 + o_p(1)\}
\end{aligned}
\tag{3.10}
$$

where $R_n = \sum_{i=1}^n \{\lambda^T(X_i - \mu)\}^3$. This expansion looks similar to that given in Owen (1990) for a fixed p, but the stochastic order of each term requires careful evaluation as p grows with n.

From Lemma 3.5 in Section 3.6, we have

$$
(2p)^{-1/2}\{n(\bar{X} - \mu)^T \Sigma^{-1}(\bar{X} - \mu) - p\} \to N(0, 1)
\tag{3.11}
$$

in distribution as $n \to \infty$, which is true under much weaker conditions, for instance $p/n \to c \geq 0$ by applying the martingale central limit theorem. Derivations given in Section 3.6 show that the other two terms on the right hand side of (3.10) are both $o_p(p^{1/2})$. These lead us to establish (3.3) as summarized in the following theorem.

Theorem 3.2 *If $k = 1$ in (3.4) and $\mathrm{tr}^5(\Sigma)\gamma_p^5 = o(np)$, then (3.3) is valid.*

Theorem 3.2 indicates that, when γ_p is bounded, (3.3) is true if $p = o(n^{1/4})$, which improves the order $p = o(n^{1/6})$ obtained by Hjort et al. (2009) under the finite fourth moment condition of X_i which we do not need in our study. The conditions assumed under Theorem 3.2 are liberal compared to (A1) and (A2), and there is no explicit restriction on γ_p, which may diverge to ∞ as $n \to \infty$.

Next we show that the dimension p can increase more rapidly if Z_{il} possesses more than fourth moment. Assuming higher order moments allows us to evaluate those terms in (3.10) more accurately. Specifically, we will assume Z_{il} has at least finite

12th moment, namely $k \geq 3$ in model (3.4). The case $k \geq 2$ can be considered as part of the case $k \geq 1$ whose analysis is covered by Theorem 3.2. The following theoremshows that $p = o(n^{1/2})$ is approachable.

Theorem 3.3 *If $k \geq 3$ in (3.4), $\{\text{tr}(\Sigma)\}^{4k-1}\gamma_p = O(n^{2k-1})$ and $p^2\gamma_p^5 = o\{n^{(4k-1)/(4k)}\}$, then (3.3) is valid.*

When γ_p is bounded, Theorem 3.3 implies that $w_n(\mu)$ is asymptotically normally distributed if $p = o(n^{1/2-1/(8k)})$, which is close to $o(n^{1/2})$ for $k \geq 3$ and improves the earlier rate $o(n^{1/3})$ attained in Hjort et al. (2009). By reviewing the proof of Theorem 3.3, we can see that, if Z_{ij} are all bounded random variables, the dimensionality p can reach $o(n^{1/2})$. We believe that $p = o(n^{1/2})$ is the best rate for the asymptotic normality of the empirical likelihood ratio with the normalizing constants p and $(2p)^{1/2}$. This is based on the following considerations. Lemma 3.4 in Section 3.6 implies that, when the largest eigenvalue of Σ is bounded, $||S_n - \Sigma||_{\text{tr}} \to 0$ in probability if and only if $p = o(n^{1/2})$. Here $||A||_{\text{tr}} = \{\text{tr}(A'A)\}^{1/2}$ is the trace norm. Bai and Yin (1993) established the convergence of S_n to Σ with probability one if $p = o(n)$ under the matrix norm based on the largest eigenvalue by assuming each Z_{nl} is independent and identically distributed. However, it can be seen from our proofs in Section 3.6 that the convergence of S_n to Σ under the trace norm is the one used in establishing various results for the empirical likelihood.

As shown by Theorems 3.2 and 3.3, when (3.3) is valid, the asymptotic mean and variance of the empirical likelihood ratio are respectively p and $2p$ which are known. This means that the empirical likelihood carries out internal studentising even when p increases along with n. However, it is apparent that the internal studentisation prevents p from growing faster as it brings in those higher-order terms.

The least-square empirical likelihood is a simplified version of the empirical likeli-hood. The least-square empirical likelihood ratio for μ is

$$q_n(\mu) = \min \sum (n\pi_i - 1)^2$$

subject to $\sum_{i=1}^{n} \pi_i = 1$ and $\sum \pi_i(X_i - \mu) = 0$. The least square empirical likelihood uses $\sum (n\pi_i - 1)^2$ to approximate $-2 \sum \log(n\pi_i)$. As shown in Brown and Chen (1998), the optimal weights π_i admit close-form solutions so that

$$q_n(\mu) = n(\bar{X} - \mu)^T H_n^{-1}(\bar{X} - \mu) \tag{3.12}$$

where $H_n = S_n - (\bar{X} - \mu)(\bar{X} - \mu)^T$. Hence, $q_n(\mu)$ can be readily computed without solving the non-linear equation (3.5) as for the full empirical likelihood. The least square empirical likelihood ratio is a first order approximation to the full empirical likelihood ratio, and $q_n(\mu) \to \chi_p^2$ in distribution when p is fixed.

The least square empirical likelihood is less affected by higher dimension. In par-ticular, if $k \geq 3$ in (3.4), then

$$(2p)^{-1/2}\{q_n(\mu) - p\} \to N(0, 1) \tag{3.13}$$

in distribution as $n \to \infty$ when $p = o(n^{2/3})$, which improves the rate given by Theorem 3.3 for the full empirical likelihood ratio $w_n(\mu)$.

To appreciate (3.13), we note from (3.12)

$$q_n(\mu) = n(\bar{X} - \mu)^T \Sigma^{-1}(\bar{X} - \mu) + n(\bar{X} - \mu)^T (H_n^{-1} - \Sigma^{-1})(\bar{X} - \mu). \tag{3.14}$$

Then, following a similar line to the proof of Lemma 3.6,

$$n(\bar{X} - \mu)^T (H_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) = O_p(p^2/n) = o_p(p^{1/2}).$$

As the first term on the right hand side of (3.14) is asymptotically normal with mean p and variance $2p$ as conveyed in (3.11), (3.13) is valid.

If we confine ourselves to specific distributions, faster rates for p can be established. For example if the data are normally distributed, the least square empirical likelihood ratio is the Hotelling's T^2 statistic, which is shown in Bai and Saranadasa (1996) to be asymptotically normal if $p/n \to c \in [0, 1)$.

3.4   Numerical Results

We report results from a simulation study designed to evaluate the asymptotic normality of the empirical likelihood ratio. The $p \times 1$ independent and identically distributed data vectors $\{X_i\}_{i=1}^n$ were generated from a moving average model:

$$X_{ij} = Z_{ij} + \rho Z_{ij+1} \quad (i = 1, \ldots, n, j = 1, \ldots, p)$$

where, for each i, the innovations $\{Z_{ij}\}_{j=1}^{p+1}$ were independent random variables with zero mean and unit variance. We considered two distributions for the innovation Z_{ij}. One is the standard normal distribution, and the other is a standardized version of a Pareto distribution with distribution function $(1 - x^{-4.5})I(x \geq 1)$. We standardized the Pareto random variables so that they had mean zero and unit variance. As the Pareto distribution has only the first four moments finite, we had $k = 1$ in (3.4), whereas $k = \infty$ for the normally distributed innovations. In both distributions, X_i is a multivariate random vector with zero mean and covariance $\Sigma = (\sigma_{ij})_{p \times p}$ where $\sigma_{ii} = 1 + \rho^2, \sigma_{ii\pm 1} = \rho$ and $\sigma_{ij} = 0$ for $|i - j| > 1$. We set ρ to be 0.5 throughout the simulation.

To make p and n increase simultaneously, we considered two growth rates for p with respect to n: (i) $p = c_1 n^{0.4}$ and (ii) $p = c_2 n^{0.24}$. We chose the sample size $n = 200, 400$ and 800. By assigning $c_1 = 3, 4$ and 5 in the faster growth rate setting (i), we obtained three dimensions for each sample size, which were $p = 25, 33$ and

43 for $n = 200$, $p = 33$, 44 and 58 for $n = 400$, and $p = 42$, 55 and 72 for $n = 800$, respectively. For the slower growth rate setting (ii), to maintain a certain amount of increase between successive dimensions when n was increased, we assigned larger $c_2 = 4, 6$ and 8, which led to $p = 14, 17$ and 20 for $n = 200$; $p = 21$, 25 and 30 for $n = 400$; and $p = 29$, 34 and 40 for $n = 800$, respectively.

We carried out 500 simulations for each of the (p, n)-combinations and for each of the two innovation distributions. Fig 3.1 displays the Q-Q plots between the standardized empirical likelihood ratio and $N(0, 1)$ for the faster growth rate (i), and those for the slower growth rate (ii) are presented in Fig. 3.2. As n and p were increased simultaneously, there was a general convergence of the standardized empirical likelihood ratio to $N(0, 1)$. We also observed that the convergence in Fig. 3.2 for the slower growth rate setting (ii) was faster than that in Fig. 3.1 for the faster growth rate setting. This is expected as the setting (i) ensured much higher-dimensionality. The convergence for the normal innovation was faster than that for the Pareto case when $p = c_1 n^{0.4}$ in Fig. 3.1. This may be explained by the fact that the Pareto distribution has only four finite moments, which corresponds to $k = 1$, whereas the normal innovation has all moments finite. According to Theorems 3.2 and 3.3, the growth rate for p depends on the value of k, the larger the k, the higher the rate. For the lower growth rate in setting (ii), Fig. 3.2 shows that, there was substantial improvement in the convergence in the Q-Q plots as p was increased at the slower rate for both distributions of innovations,

It is observed that the most of the lack-of-fit in the Gaussian Q-Q plots in Fig. 3.1 and Fig. 3.2 appeared at the lower and upper quantiles. This could be attributed to the lack-of-fit between χ_p^2 and $N(0, 1)$, as χ_p^2 may be viewed as the intermediate convergence of the empirical likelihood ratio.

To verify this point, we carried out further simulations by inverting settings (i) and (ii) so that for a given dimension p, three sample sizes were generated according to (iii) $n = (p/c_1)^{1/0.4}$ and (iv) $n = (p/c_2)^{1/0.24}$, with $c_1 = 3, 4$ and 5 and $c_2 = 4, 5$ and 6, respectively. We chose $p = 35, 45$ and 55 for the setting (iii) and $p = 17, 20$ and 25 for the setting (iv). Figure 3.3 and Figure 3.4 are the corresponding χ_p^2 Q-Q plots for (iii) and (iv). These two figures show that there was a substantial improvement in the overall fit of the Q-Q plots, and the lack-of-fit appeared in the Gaussian Q-Q plots was substantially reduced.

3.5   Case Study

We give in this section a financial application of the empirical likelihood ratio in analyzing stock returns for public companies included in the Dow Jones Industrial Average. We have daily closing prices for the thirty stocks in the Dow Jones from July 1st, 1986 to September 2nd, 2008. These prices had been adjusted for stock split, buy-back, dividend payouts and other distributions. Although stock prices typically exhibit dependence over time, their price changes are less so over time. To further reduce the time dependence, we consider monthly (four-weekly) returns which leads to 265 observations. Here the data-dimension p is 30 which is high relative to $n = 265$, the number of observations.

Let $Y_t = (Y_{t,1}, Y_{t,2}, \cdots, Y_{t,30})^T$ be the vector of closing prices of the Dow Jones stocks at the beginning of tth month, let

$$R_t = ((Y_{t,1} - Y_{t-1,1})/Y_{t-1,1}, \cdots, (Y_{t,30} - Y_{t-1,30})/Y_{t-1,30})^T$$

be the relative return for the tth month and let

$$X_t = (\log(Y_{t,1}/Y_{t-1,1}), \cdots, \log(Y_{t,30}/Y_{t-1,30}))^T$$

be the log-returns for $t = 1, 2, \cdots, 265$. It is assumed that $X_1, X_2, \cdots, X_n$ are independent and identically distributed random vectors. Let $r = E(R_t) = (r_1, r_2, \cdots, r_p)^T$ be the vector of the average relative return and $\mu = E(X_t) = (\mu_1, \cdots, \mu_p)^T$ be the average log-return and $\Sigma = Var(X_t)$. We wish to test

$$H_0 : \mu = 0 \text{ vs. } H_1 : \mu \neq 0$$

using the empirical likelihood method. The empirical likelihood test statistic is

$$T_n = (2p)^{-1/2}\{w_n(0) - p\}.$$

The hypothesis is rejected at a significant level α if $|T_n| \geq z_\alpha$ where z_α is the upper α-quantile of N(0, 1). We also carried out similar tests for sectors of stocks included in the Dow Jones. The sectors are basic materials, consumer goods, finance, health care, industrial goods, services and technology, respectively. Let $\mu^{(j)}$ denote the mean for components of X_t that correspond to a specific sector. We test for $H_0 : \mu^{(j)} = 0$ vs $H_1 : \mu^{(j)} \neq 0$. Table 3.1 reports values of T_n and the p-values for testing the mean log-return for the entire 30 stocks and sectors of the Dow Jones. It was found that the p-value for the average log-return of the 30 stocks is 0.059, and hence the null hypothesis cannot be rejected at $\alpha = 0.05$. Some sectors were found to have average log-return significantly different from zero, for instance basic material, consumer goods and service sectors.

We would like to add that, if a hypothesis $\mu_l = 0$ is not rejected, it does not necessarily imply that the average relative return r_l is zero. To appreciate this point, we assume the Black-Scholes (Black and Scholes, 1973) continuous-time diffusion model for stock price $Y_{t,l}$, $l = 1, \cdots, p$, such that

$$dY_{t,l}/Y_{t,l} = r_l dt + \sigma_l dB_{t,l}$$

where r_l and σ_l are respectively the instantaneous mean and standard deviation of the relative return of the lth stock, and $B_{t,l}$ is the standard Brownian motion. This model implies, via Ito formula, that $log(Y_{t,l}/Y_{t-1,l})$ is distributed as $N((r_l - \sigma_l^2/2)\delta_t, \sigma_l^2\delta_t)$ where δ_t is the length of sampling. Hence, for our analysis, $\mu_l = (r_l - \sigma_l^2/2)\delta_t$ with $\delta_t = 1/12$ as the returns are calculated monthly. If $\mu_l = 0$, the average relative return $r_l = \sigma_l^2/2$ implying a positive return due to holding a risky stock.

3.6 Technical Proofs

We present the proofs of Theorem 3.1, some lemmas used in the proof of Theorem 3.2, and Theorem 3.3.

We first establish some lemmas.

Lemma 3.1 *If $m_{4k} < \infty$ for some $k \geq 1$, then*

$$E||X_i - \mu||^{2k} = O\{tr^k(\Sigma)\} \quad and \quad Var(||X_i - \mu||^{2k}) = O\{tr^{2k-1}(\Sigma)\gamma_p\}.$$

Proof:

We only show the case of $k = 1$ since other cases can be done similarly. It is easy to check that

$$E||X_i - \mu||^2 = tr\{E(X_i - \mu)^T(X_i - \mu)\} = tr(\Sigma) \tag{3.15}$$

and

$$E||X_i - \mu||^4 = E||\Gamma Z_i||^4 = E\left(Z_i^T\Gamma^T\Gamma Z_i Z_i^T\Gamma^T\Gamma Z_i\right) = tr\left\{\Gamma^T\Gamma E(Z_i Z_i^T\Gamma^T\Gamma Z_i Z_i^T)\right\}.$$

Write $\Gamma^T\Gamma = (\nu_{sl})_{1\leq s,l\leq m}$. Then

$$Z_i Z_i^T\Gamma^T\Gamma Z_i Z_i^T = \left(\sum_{j=1}^{m}\sum_{l=1}^{m} Z_{ik_1}Z_{il}\nu_{lj}Z_{ij}Z_{ik_2}\right)_{1\leq k_1,k_2\leq m}.$$

When $k_1 = k_2 = s$,

$$E(\sum_{j=1}^m \sum_{l=1}^m Z_{ik_1} Z_{il} \nu_{lj} Z_{ij} Z_{ik_2}) = \sum_{l=1}^m E\{(Z_{is})^2 (Z_{il})^2\}\nu_{ll} = \nu_{ss} E(Z_{is})^4 + \sum_{l \neq s} \nu_{ll}.$$

When $k_1 \neq k_2$, $E(\sum_{j=1}^m \sum_{l=1}^m Z_{ik_1} Z_{il} \nu_{lj} Z_{ij} Z_{ik_2}) = 2\nu_{k_1 k_2}$. Hence

$$
\begin{aligned}
E||X_i - \mu||^4 &= \sum_{s=1}^m \nu_{ss}\{\nu_{ss} E(Z_{is})^4 + \sum_{l \neq s} \nu_{ll}\} + 2\sum_{s=1}^m \sum_{l \neq s} \nu_{sl}\nu_{ls} \\
&= \sum_{s=1}^m \nu_{ss}^2\{E(Z_{is})^4 - 3\} + tr^2(\Gamma^T \Gamma) + 2tr(\Gamma^T \Gamma \Gamma^T \Gamma) \\
&= (m_4 - 3)\sum_{s=1}^m \nu_{ss}^2 + tr^2(\Sigma) + 2tr(\Sigma^2). \quad (3.16)
\end{aligned}
$$

Note that $\sum_{s=1}^m \nu_{ss}^2 \leq \sum_{j=1}^m \sum_{s=1}^m \nu_{js}^2 = tr\{(\Gamma^T \Gamma)^2\} = tr(\Sigma^2)$. This together with (3.15) and (3.16) implies that $Var(||X_i - \mu||^2) = (m_4 - 3)\sum_{s=1}^m \nu_{ss}^2 + 2tr(\Sigma^2) = O\{tr(\Sigma^2)\}$. This completes the proof of Lemma 3.1.

Lemma 3.2 *If $m_{4k} < \infty$ for some $k \geq 1$, then, with probability one*

$$\max_{1 \leq i \leq n} ||X_i - \mu|| = o[\{tr(\Sigma)\}^{\frac{2k-1}{4k}} \gamma_p^{\frac{1}{4k}}(\Sigma) n^{\frac{1}{4k}}] + O(\sqrt{tr(\Sigma)}).$$

Proof:

We note that

$$
\begin{aligned}
&\max_{1 \leq i \leq n} ||X_i - \mu|| \\
&= \{\max_{1 \leq i \leq n} ||X_i - \mu||^{2k}\}^{1/(2k)} \\
&\leq \{|\max_{1 \leq i \leq n} |||X_i - \mu||^{2k} - E(||X_i - \mu||^{2k})| + E(||X_i - \mu||^{2k})\}^{1/(2k)} \\
&= \{\sqrt{Var(||X_i - \mu||^{2k})} \max_{1 \leq i \leq n} |\frac{||X_i - \mu||^{2k} - E(||X_i - \mu||^{2k})}{\sqrt{Var(||X_i - \mu||^{2k})}}| + E(||X_i - \mu||^{2k})\}^{1/(2k)}
\end{aligned}
$$

and

$$\max_{1 \leq i \leq n} |\frac{||X_i - \mu||^{2k} - E(||X_i - \mu||^{2k})}{\sqrt{Var(||X_i - \mu||^{2k})}}| = o(n^{1/2})$$

with probability one as $n \to \infty$. This lemma is proved by applying Lemma 3 of Owen (1990) and Lemma 3.1.

From now on, we let $Y_i = \Sigma^{-1/2}(X_i - \mu)$, $V_n = \frac{1}{n}\sum_{i=1}^n Y_i Y_i^T$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ and $D_n = V_n - I_p = (d_{sl})_{l,s=1,\cdots,p}$.

Lemma 3.3 . *Under the conditions of Theorem 3.1, $tr(D_n^2) = O_p(p^2/n)$.*

Proof:

We only need to show $E\{tr(D_n^2)\} = O(p^2/n)$. Note that $V_n = \Sigma^{-1/2}\Gamma S_z \Gamma^T \Sigma^{-1/2}$ where $S_z = n^{-1}\sum_{i=1}^n Z_i Z_i^T$. Let $\tilde{\Sigma} = \Gamma^T \Sigma^{-1}\Gamma = (\tilde{\sigma}_{jl})_{1 \le j, l \le m}$, say. Then

$$tr(D_n^2) = tr(S_z \tilde{\Sigma} S_z \tilde{\Sigma}) - 2tr(S_z \tilde{\Sigma}) + p \qquad (3.17)$$

and

$$E\{tr(S_z \tilde{\Sigma})\} = E\left(\sum_{j,l=1}^m n^{-1}\sum_{i=1}^n Z_{ij} Z_{il} \tilde{\sigma}_{lj}\right) = \sum_{j,l=1}^m \delta_{jl}\tilde{\sigma}_{lj} = \sum_{j=1}^m \tilde{\sigma}_{jj} = p \quad (3.18)$$

since $tr(\tilde{\Sigma}) = tr(I_p) = p$. By utilizing information of Z_i given in (3.4),

$$
\begin{aligned}
E[tr\{(S_Z \tilde{\Sigma})^2\}] &= E\left(\sum_{j,l=1}^m \sum_{l_1,l_2=1}^m n^{-2}\sum_{i_1,i_2=1}^n Z_{i_1j} Z_{i_1 l_1} Z_{i_2 l} Z_{i_2 l_2} \tilde{\sigma}_{l_1 l}\tilde{\sigma}_{l_2 j}\right) \\
&= m_4 n^{-1}\sum_{j=1}^m \tilde{\sigma}_{jj}^2 + n^{-1}\sum_{j\ne l}\left(2\tilde{\sigma}_{jl}^2 + \tilde{\sigma}_{jj}\tilde{\sigma}_{ll}\right) + (1 - n^{-1})\sum_{j,l=1}^m \tilde{\sigma}_{jl}^2 \\
&= \sum_{j,l=1}^m \tilde{\sigma}_{jl}^2 + (m_4 - 1)n^{-1}\sum_{j=1}^m \tilde{\sigma}_{jj}^2 + n^{-1}\sum_{j\ne l}\left(\tilde{\sigma}_{jl}^2 + \tilde{\sigma}_{jj}\tilde{\sigma}_{ll}\right).
\end{aligned}
$$

It is easy to check that $\sum_{j,l=1}^m \tilde{\sigma}_{jl}^2 = tr(\tilde{\Sigma}^2) = p$, $\sum_{j=1}^m \tilde{\sigma}_{jj}^2 \le \sum_{j,l=1}^m \tilde{\sigma}_{jl}^2 = p$, $\sum_{j\ne l}\tilde{\sigma}_{jl}^2 \le \sum_{j,l=1}^m \tilde{\sigma}_{jl}^2 = p$ and $|\sum_{j\ne l}\tilde{\sigma}_{jj}\tilde{\sigma}_{ll}| \le (\sum_{j=1}^m \tilde{\sigma}_{jj})^2 = p^2$. Thesis together with (3.17) to (3.18) imply $E\{tr(D_n^2)\} = O(p^2/n)$. This completes the proof of the lemma.

Lemma 3.4 *Under condition (3.4), $\max_{1 \leq i \leq p} |\gamma_i(S_n) - \gamma_i(\Sigma)| = O_p(\gamma_p p / \sqrt{n})$.*

Proof:

Note that

$$
\begin{aligned}
|\gamma_i(S_n) - \gamma_i(\Sigma)|^2 &= |\gamma_i^{1/2}(S_n^2) - \gamma_i^{1/2}(\Sigma^2)|^2 \qquad\qquad (3.19) \\
&\leq \sum_{i=1}^{p} |\gamma_i^{1/2}(S_n^2) - \gamma_i^{1/2}(\Sigma^2)|^2 \\
&= \sum_{i=1}^{p} \gamma_i(S_n^2) + \sum_{i=1}^{p} \gamma_i(\Sigma^2) - 2\sum_{i=1}^{p} \gamma_i^{1/2}(S_n^2)\gamma_i^{1/2}(\Sigma^2) \\
&= tr(S_n^2) + tr(\Sigma^2) - 2\sum_{i=1}^{p} \gamma_i(S_n)\gamma_i(\Sigma).
\end{aligned}
$$

By Von Neumann's inequality, $\sum_{i=1}^{p} \gamma_i(S_n)\gamma_i(\Sigma) \geq tr(S_n\Sigma)$. Hence

$$
\max_{1 \leq i \leq p} |\gamma_i(S_n) - \gamma_i(\Sigma)| \leq \sqrt{tr\{(S_n - \Sigma)^2\}}.
$$

Now

$$
\begin{aligned}
tr\{(S_n - \Sigma)^2\} &= tr[\{\Sigma^{1/2}(V_n - I_p)\Sigma^{1/2}\}^2] \\
&= tr(D_n \Sigma D_n \Sigma) \\
&\leq \gamma_p^2 tr(D_n^2) = O_p(\gamma_p^2 p^2 / n)
\end{aligned}
$$

by applying Lemma 3.3.

This lemma implies that all the eigenvalues of S_n converge to those of Σ uniformly at the rate of $O_p\{\gamma_p p / \sqrt{n}\}$.

Proof of Theorem 3.1:

By (3.5), $\lambda \in R^p$ satisfies

$$
0 = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i - \mu}{1 + \lambda^T(X_i - \mu)} = g(\lambda). \qquad\qquad (3.20)
$$

Write $\lambda = \rho\theta$ where $\rho \geq 0$ and $||\theta|| = 1$. Hence

$$
\begin{aligned}
0 \; &= \; ||g(\rho\theta)|| \geq |\theta^T g(\rho\theta)| \\
&= \; n^{-1}|\theta^T\{\sum_{i=1}^n (X_i - \mu) - \rho\sum_{i=1}^n \frac{(X_i-\mu)\theta^T(X_i-\mu)}{1+\rho\theta^T(X_i-\mu)}\}| \\
&\geq \; \rho\theta^T S_n\theta\{1 + \rho\max_{1\leq i\leq n}||X_i - \mu||\}^{-1} - n^{-1}|\sum_{i=1}^n \theta^T(X_i - \mu)|.
\end{aligned}
$$

Hence,

$$
\rho\left\{\theta^T S_n\theta - \max_{1\leq i\leq n}||X_i - \mu||n^{-1}|\sum_{i=1}^n \theta^T(X_i - \mu)|\right\} \leq n^{-1}|\sum_{i=1}^n \theta^T(X_i - \mu)|.
$$

Since $n^{-1}|\sum_{i=1}^n \theta^T(X_i - \mu)| = O_p(\sqrt{tr(\Sigma)/n})$, it follows from Lemma 3.2 that

$$
\max_{1\leq i\leq n}||X_i - \mu||n^{-1}|\sum_{i=1}^n \theta^T(X_i - \mu)|
$$

$$
= \; o_p\{tr^{1-\frac{1}{4k}}(\Sigma)\gamma_p^{\frac{1}{4k}}(\Sigma)n^{-\frac{1}{2}+\frac{1}{4k}} + O_p\{tr(\Sigma)n^{-1/2}\} = o_p(1). \qquad (3.21)
$$

By Lemma 3.4, for a positive constant C_1, $P(\theta^T S_n\theta \geq \frac{1}{2}C_1) \to 1$ as $n \to \infty$. Hence $||\lambda|| = \rho = O_p(\sqrt{tr(\Sigma)/n})$. This completes the proof of Theorem 3.1.

By repeating (3.21) in the proof of the above theorem and Lemma 3.2, we have

$$
\max_{1\leq i\leq n}||\lambda^T(X_i - \mu)|| \leq ||\lambda||\max_{1\leq i\leq n}||X_i - \mu|| = o_p(1) \qquad (3.22)
$$

We need the following lemmas for proving Theorem 3.2.

Lemma 3.5 *If $p/n \to c \geq 0$, then*

$$
\frac{n(\bar{X} - \mu)^T\Sigma^{-1}(\bar{X} - \mu) - p}{\sqrt{2p}} \xrightarrow{d} N(0,1) \; as \; n \to \infty.
$$

Proof:

The proof entails applying the martingale central limit theorem as given in Hall and Hyde (1980). Bai and Saranadasa (1996) used this approach to establish the asymptotic normality for a two-sample test statistic for high-dimensional data. What we have here is easier due to the one sample nature.

Lemma 3.6 *Under the conditions of Theorem 3.2,*

$$n(\bar{X} - \mu)^T(S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) = o_p(p^2/\sqrt{n}).$$

Proof:

Recall that $D_n = V_n - I_p = (d_{sl})_{1 \leq s \leq p, 1 \leq l \leq p}$. It follows from Lemma 3.3 that

$$P(\max_{1 \leq k_1, k_2 \leq p} |d_{k_1 k_2}| > \epsilon) \leq \sum_{k_1=1}^{p} \sum_{k_2=1}^{p} \epsilon^{-2} E(d_{k_1 k_2}^2) = \epsilon^{-2} E\{tr(D_n^2)\} = O(p^2/n).$$

Hence, $d_{jl} = O_p(\sqrt{p^2/n}) = o_p(1)$ uniformly in $1 \leq j, l \leq p$. It is easy to check that

$$V_n^{-1} - I_p = -D_n + D_n^2 + D_n^2(V_n^{-1} - I_p)$$

and

$$n(\bar{X} - \mu)^T(S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) = n\bar{Y}^T(V_n^{-1} - I_p)\bar{Y}.$$

From Lemma 3.1, $E(||\bar{Y}||^2) = \frac{1}{n}E(||Y_1||^2) = p/n$. Since $|\bar{Y}^T A\bar{Y}| \leq ||\bar{Y}||^2\sqrt{tr(A^2)}$ for any symmetric matrix A, it follows from Lemma 3.4 and the condition $p = o(n^{1/3})$ that

$$|n\bar{Y}^T D_n \bar{Y}| \leq n||\bar{Y}||^2\sqrt{tr(D_n^2)} = O_p(p^2/\sqrt{n}) = o_p(\sqrt{p}).$$

Similarly, $|n\bar{Y}^T D_n^2 \bar{Y}| \leq n||\bar{Y}||^2 tr(D_n^2) = O_p(p^3/n) = o_p(\sqrt{p})$.

Furthermore, we note the following facts

$$|\bar{Y}^T D_n^3 \bar{Y}| \leq \max_{1 \leq i \leq p}\{|\gamma_i(D_n)|\}\bar{Y}^T D_n^2\bar{Y} = o_p(\bar{Y}^T D_n^2\bar{Y}) \tag{3.23}$$

since $\max_{1 \leq i \leq p}\{|\gamma_i(D_n)|\} \leq \sqrt{tr(D_n^2)} \to 0$, and

$$\bar{Y}^T D_n^4 \bar{Y} \leq \gamma_p(D_n^2)\bar{Y}^T D_n^2\bar{Y} = o_p(\bar{Y}^T D_n^2\bar{Y}).$$

In general, if $p = o(\sqrt{n})$, for any positive integer l, $\bar{Y}^T D_n^{2+l}\bar{Y} = o_p(\bar{Y}^T D_n^2\bar{Y})$ The lemma follows from summarizing the above results.

Proof of Theorem 3.2:

Put $W_i = \lambda^T(X_i - \mu)$. Then (3.22) implies that $\max_{1 \le i \le n} |W_i| = o_p(1)$. Expand equation (3.20),

$$0 = g(\lambda) = \bar{X} - \mu - S_n\lambda + \beta_n \tag{3.24}$$

where

$$\beta_n = n^{-1}\sum_{i=1}^n (X_i - \mu)\frac{W_i^2}{(1 + \xi_i)^3}$$

and $|\xi_i| \le |\lambda^T(X_i - \mu)|$. As $\max_{1 \le i \le n} |W_i| = o_p(1)$, $\max_{1 \le i \le n} |\xi_i| = o_p(1)$ as well. Hence $\beta_n = \beta_{n1}\{1 + o_p(1)\}$ where $\beta_{n1} = n^{-1}\sum_{i=1}^n (X_i - \mu)W_i^2$. Apply Theorem 3.1 and Lemma 3.2 with $k = 1$, we have, if $tr(\Sigma) = O\{\gamma_p^{5/3}(\Sigma)n^{1/3}\}$,

$$
\begin{aligned}
||\beta_{n1}|| &\le n^{-1}\sum_{i=1}^n ||X_i - \mu||W_i^2 \\
&\le \max_{1 \le i \le n} ||X_i - \mu||n^{-1}\sum_{i=1}^n \lambda^T(X_i - \mu)(X_i - \mu)^T\lambda \\
&= \max_{1 \le i \le n} ||X_i - \mu||\,||\lambda||^2 O_p(\gamma_p(\Sigma)) \\
&= o_p(||\lambda||\gamma_p^{1+\frac{1}{4}}(\Sigma)tr^{\frac{3}{4}}(\Sigma)n^{-\frac{1}{4}}) + O_p(||\lambda||tr(\Sigma)n^{-\frac{1}{2}}) = o_p(||\lambda||). \tag{3.25}
\end{aligned}
$$

It follows from (3.24) that

$$\lambda = S_n^{-1}(\bar{X} - \mu) + S_n^{-1}\beta_n \tag{3.26}$$

and $\log(1 + W_i) = W_i - W_i^2/2 + W_i^3/(1 + \xi_i)^4$ for some ξ_i such that $|\xi_i| \le |W_i|$. Therefore

$$
\begin{aligned}
w_n(\mu) &= n(\bar{X} - \mu)^T S_n^{-1}(\bar{X} - \mu) - n\beta_n S_n^{-1}\beta_n + \tfrac{2}{3}\sum_{i=1}^n \frac{\{\lambda^T(X_i - \mu)\}^3}{(1 + \xi_i)^4} \\
&= n(\bar{X} - \mu)^T \Sigma^{-1}(\bar{X} - \mu) + n(\bar{X} - \mu)^T(S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) \\
&\quad - n\beta_n S_n^{-1}\beta_n + \tfrac{2}{3}R_n\{1 + o_p(1)\} \tag{3.27}
\end{aligned}
$$

where $R_n = \sum_{i=1}^n \{\lambda^T (X_i - \mu)\}^3$. By (3.25) and Lemma 3.4,

$$
\begin{aligned}
|n\beta_n S_n^{-1} \beta_n| &\leq n\|\beta_n\|^2/\gamma_1(S_n) \\
&= O_p(\gamma_p^2(\Sigma) tr^3(\Sigma) n^{-1}) + o_p(\gamma_p^{5/2} tr^{5/2}(\Sigma) n^{-1/2}) = o_p(\sqrt{p}). \quad (3.28)
\end{aligned}
$$

We also note that

$$
\begin{aligned}
R_n &\leq \sum_{i=1}^n |W_i|^3 \leq \{\sum_{i=1}^n W_i^2 \sum_{i=1}^n W_i^4\}^{1/2} \\
&\leq \sqrt{n\lambda^T S_n \lambda}\{\sum_{i=1}^n \|\lambda\|^4 \|X_i - \mu\|^4\}^{1/2} \\
&= O_p(\sqrt{tr(\Sigma)\gamma_p(\Sigma)}) O_p(\sqrt{tr^2(\Sigma) n^{-2} n tr^2(\Sigma)}) \\
&= O_p(tr^{5/2}(\Sigma)\gamma_p^{1/2} n^{-1/2}) = o_p(\sqrt{p}). \quad (3.29)
\end{aligned}
$$

Hence the theorem follows from Lemmas 3.5 and 3.6.

The proof of Theorem 3.3 requires the following lemmas.

Lemma 3.7 *Under the conditions of Theorem 3.3, $tr\{(V_n^{-1} - I_p)^2\} = O_p(p^2/n)$.*

Proof:

We note that

$$
tr\{(V_n^{-1} - I_p)^2\} = \sum_{i=1}^p (\gamma_i^{-1}(V_n) - 1)^2 \leq \gamma_1^{-2}(V_n) \sum_{i=1}^p (\gamma_i(V_n) - 1)^2 = \gamma_1^{-2}(V_n) tr(D_n^2).
$$

Thus, the lemma follows from that $\gamma_1(V_n) \to 1$ in probability from Lemma 3.4 and $tr(D_n^2) = O_p(p^2/n)$ from Lemma 3.3.

Let

$$
\xi_{n1} = n^{-1} \sum_{i=1}^n (\bar{Y}^T Y_i)^4 \quad \text{and} \quad \beta_{n2} = n^{-1} \sum_{i=1}^n (X_i - \mu)(\bar{Y}^T Y_i)^2. \quad (3.30)
$$

Lemma 3.8 *Under the conditions of Theorem 3.3, $E(\xi_{n1}) = O(p^2/n^2)$ and $\xi_{n1} = O_p(p^2/n^2)$.*

Proof:

Note that

$$
\begin{aligned}
E(\xi_{n1}) &= n^{-5} \sum_{i=1}^{n} E(Y_i^T Y_i)^4 + n^{-5} \sum_{i \neq j} \Big[E(Y_i^T Y_j)^4 + 4E\{Y_i^T Y_i (Y_i^T Y_j)^3\} \quad (3.31) \\
&\quad + 6E\{(Y_i^T Y_i)^2 (Y_i^T Y_j)^2\} \Big] + 3n^{-5} \sum_{i \neq j_1 \neq j_2} E\{(Y_{j_1}^T Y_i)^2 (Y_{j_2}^T Y_i)^2\}. \quad (3.32)
\end{aligned}
$$

Put $\tilde{\Gamma} = \Sigma^{-1/2} \Gamma$ and write $\tilde{\Gamma}^T \tilde{\Gamma} = \Gamma^T \Sigma^{-1} \Gamma = (\tilde{\nu}_{jl})_{1 \leq j,l \leq m}$. Using a similar derivation of Lemma 3.1, we have

$$
\begin{aligned}
E(Y_i^T Y_i)^4 &= E(Z_i^T \tilde{\Gamma}^T \tilde{\Gamma} Z_i)^4 \\
&= \sum_{k_1, \cdots, k_4, l_1, \cdots, l_4} E(Z_{ik_1} Z_{ik_2} Z_{ik_3} Z_{ik_4} Z_{il_1} Z_{il_2} Z_{il_3} Z_{il_4}) \tilde{\nu}_{k_1,l_1} \tilde{\nu}_{k_2,l_2} \tilde{\nu}_{k_3,l_3} \tilde{\nu}_{k_4,l_4} \\
&= O\{tr^4(\Sigma)\}. \quad (3.33)
\end{aligned}
$$

When $i \neq j$,

$$
\begin{aligned}
E(Y_i^T Y_j)^4 &= E(Z_i^T \tilde{\Gamma}^T \tilde{\Gamma} Z_j)^4 \\
&= \sum_{k_1, \cdots, k_4, l_1, \cdots, l_4} E(Z_{ik_1} Z_{ik_2} Z_{ik_3} Z_{ik_4}) E(Z_{jl_1} Z_{jl_2} Z_{jl_3} Z_{jl_4}) \tilde{\nu}_{k_1,l_1} \tilde{\nu}_{k_2,l_2} \tilde{\nu}_{k_3,l_3} \tilde{\nu}_{k_4,l_4} \\
&= m_4^2 \sum_{s,l} \tilde{\nu}_{sl}^4 + 12 m_4 \sum_{k_1 \neq k_2, l} \tilde{\nu}_{k_1 l}^2 \tilde{\nu}_{k_2 l}^2 + 36 \sum_{k_1 \neq k_2, l_1 \neq l_2} \tilde{\nu}_{k_1 l_1}^2 \tilde{\nu}_{k_2 l_2}^2.
\end{aligned}
$$

Since $\sum_{s,l} \tilde{\nu}_{sl}^4$, $\sum_{k_1 \neq k_2, l} \tilde{\nu}_{k_1 l}^2 \tilde{\nu}_{k_2 l}^2$ and $\sum_{k_1 \neq k_2, l_1 \neq l_2} \tilde{\nu}_{k_1 l_1}^2 \tilde{\nu}_{k_2 l_2}^2$ are all bounded by

$$
(\sum_{s,l} \tilde{\nu}_{sl}^2)^2 = tr^2\{(\tilde{\Gamma}^T \tilde{\Gamma})^2\} = tr^2(\tilde{\Gamma}^T \Sigma^{-1} \tilde{\Gamma}) = tr^2(I_p) = p^2,
$$

we have

$$
E(Y_i^T Y_j)^4 = O(p^2) \quad \text{for} \quad i \neq j. \quad (3.34)
$$

When $i \neq j$,

$$E\{Y_i^T Y_i (Y_i^T Y_j)^3\}$$

$$= \sum_{k_1,k_2,l_1,l_2,s_1,s_2,t_1,t_2} E(Z_{ik_1} Z_{ik_2} Z_{il_2} Z_{is_2} Z_{it_2}) E(Z_{jl_1} Z_{js_1} Z_{jt_1}) \tilde{\nu}_{k_1,k_2} \tilde{\nu}_{l_1,l_2} \tilde{\nu}_{s_1,s_2} \tilde{\nu}_{t_1,t_2}$$

$$= m_3 m_5 \sum_{s,l} \tilde{\nu}_s \tilde{\nu}_{sl}^3 + m_3^2 \sum_{s \neq l,r} \{\tilde{\nu}_{ss} \tilde{\nu}_{rl}^3 + 9\tilde{\nu}_{sl} \tilde{\nu}_{rs} \tilde{\nu}_{rl}^2\}. \tag{3.35}$$

Since $|\tilde{\nu}_{sl}|^2 \leq \tilde{\nu}_{ss} \tilde{\nu}_{ll}$ and $\tilde{\nu}_{ll} \leq \gamma_p(\Gamma^T \Sigma^{-1} \Gamma) \leq tr(\Gamma^T \Sigma^{-1} \Gamma) = p$ for any $1 \leq l \leq m$,

$$|\sum_{s,l} \tilde{\nu}_{ss} \tilde{\nu}_{sl}^3| \leq p \sum_{s,l} \tilde{\nu}_{ss}^2 \tilde{\nu}_{ll} \leq ptr\{(\Gamma^T \Sigma^{-1} \Gamma)^2\} tr(\Gamma^T \Sigma^{-1} \Gamma) = p^3.$$

Using the same argument, we have the other terms on the right hand side of (3.35) are $O(p^3)$ as well. This leads to

$$E\{Y_i^T Y_i (Y_i^T Y_j)^3\} = O(p^3) \quad \text{for} \quad i \neq j. \tag{3.36}$$

When $i \neq j$,

$$E\{(Y_i^T Y_i)^2 (Y_i^T Y_j)^2\}$$

$$= \sum_{k_1,k_2,l_1,l_2,s_1,s_2,l_2,t_2} E(Z_{ik_1} Z_{ik_2} Z_{il_2} Z_{is_1} Z_{is_2} Z_{it_2}) \tilde{\nu}_{k_1 k_2} \tilde{\nu}_{ll_2} \tilde{\nu}_{s_1 s_2} \tilde{\nu}_{lt_2}$$

$$= m_6 \sum_{l,s} \tilde{\nu}_{ss}^2 \tilde{\nu}_{sl}^2 + m_3^2 \sum_{k_1 \neq k_2, l} \{J_1 \tilde{\nu}_{k_1 k_1} \tilde{\nu}_{k_2 k_2} \tilde{\nu}_{lk_1} \tilde{\nu}_{lk_2} + J_2 \tilde{\nu}_{k_1 k_2}^2 \tilde{\nu}_{lk_1} \tilde{\nu}_{lk_2} + J_3 \tilde{\nu}_{k_1 k_1} \tilde{\nu}_{k_2 k_2} \tilde{\nu}_{lk_1}^2\}$$

$$+ \sum_{l,k_1 \neq k_2 \neq k_3} \{J_4 \tilde{\nu}_{k_1 k_1} \tilde{\nu}_{k_2 k_2} \tilde{\nu}_{lk_3}^2 + J_5 \tilde{\nu}_{k_1 k_2}^2 \tilde{\nu}_{lk_3}^2 + J_6 \tilde{\nu}_{k_1 k_3} \tilde{\nu}_{k_2 k_3} \tilde{\nu}_{lk_1} \tilde{\nu}_{lk_2}\}. \tag{3.37}$$

Here and from now on we use J_j, $j \geq 1$ to denote positive integers representing the number of combinations of the subscripts and whose values have no effect on the order of magnitude of $E\{(Y_i^T Y_i)^2 (Y_i^T Y_j)^2\}$.

Write $(\tilde{\Gamma}^T \tilde{\Gamma})^2 = (\tilde{\nu}_{ls}^{(2)})_{1 \leq l,s \leq m}$. Then

$$\sum_{l,s} \tilde{\nu}_{ss}^2 \tilde{\nu}_{sl}^2 = \sum_s \tilde{\nu}_{ss}^2 \tilde{\nu}_{ss}^{(2)} \leq tr^2\{(\Gamma \Sigma^{-1} \Gamma)^2\} = p^2,$$

and

$$
\begin{aligned}
\sum_{k_1 \neq k_2, l} \tilde{\nu}_{k_1 k_1} \tilde{\nu}_{k_2 k_2} \tilde{\nu}_{l k_1} \tilde{\nu}_{l k_2} &= \sum_{k_1 \neq k_2} \tilde{\nu}_{k_1 k_1} \tilde{\nu}_{k_2 k_2} \tilde{\nu}_{k_1 k_2}^{(2)} \\
&= \sum_{k_1, k_2} \tilde{\nu}_{k_1 k_1} \tilde{\nu}_{k_2 k_2} \tilde{\nu}_{k_1 k_2}^{(2)} - \sum_{s} \tilde{\nu}_{ss}^2 \tilde{\nu}_{k_1 k_2}^{(2)} = O(p^3) \quad (3.38)
\end{aligned}
$$

since

$$
\left| \sum_{k_1, k_2} \tilde{\nu}_{k_1 k_1} \tilde{\nu}_{k_2 k_2} \tilde{\nu}_{k_1 k_2}^{(2)} \right| \leq \gamma_p \{ (\Gamma^T \Sigma^{-1} \Gamma)^2 \} \sum_{k_1, k_2} \tilde{\nu}_{k_1 k_1} \tilde{\nu}_{k_2 k_2} \leq p \left(\sum_s \tilde{\nu}_{ss} \right)^2 = p^3.
$$

It can be shown that the other terms on the RHS of (3.37) are at most of order p^3.

Hence

$$
E\{ (Y_i^T Y_i)^2 (Y_i^T Y_j)^2 \} = O(p^3) \quad \text{for} \quad i \neq j. \tag{3.39}
$$

When $j_1 \neq j_2 \neq i$,

$$
E\{ (Y_{j_1}^T Y_i)^2 (Y_{j_2}^T Y_i)^2 \} = m_4 \sum_{t,l,s} \tilde{\nu}_{ts}^2 \tilde{\nu}_{ls}^2 + \sum_{t,l,s_1 \neq s_2} \{ J_7 \tilde{\nu}_{ts_1}^2 \tilde{\nu}_{ls_2}^2 + J_8 \tilde{\nu}_{ts_1} \tilde{\nu}_{ts_2} \tilde{\nu}_{ls_1} \tilde{\nu}_{ls_2} \},
$$

$$
\sum_{t,l,s} \tilde{\nu}_{ts}^2 \tilde{\nu}_{ls}^2 = \sum_{t,s} \tilde{\nu}_{ts}^2 \tilde{\nu}_{ss}^{(2)} \leq \left(\sum_{t,s} \tilde{\nu}_{ts}^2 \right) \left(\sum_s \tilde{\nu}_{ss}^{(2)} \right) = p^2,
$$

$$
\sum_{t,l,s_1 \neq s_2} \tilde{\nu}_{ts_1}^2 \tilde{\nu}_{ls_2}^2 = \sum_{s_1 \neq s_2} \tilde{\nu}_{s_1 s_1}^{(2)} \tilde{\nu}_{s_2 s_2}^{(2)} \leq \left(\sum_s \tilde{\nu}_{ss}^{(2)} \right)^2 = p^2
$$

and

$$
\sum_{t,l,s_1 \neq s_2} \tilde{\nu}_{ts_1} \tilde{\nu}_{ts_2} \tilde{\nu}_{ls_1} \tilde{\nu}_{ls_2} = \sum_{s_1 \neq s_2} (\tilde{\nu}_{s_1 s_1}^{(2)})^2 \leq \sum_{s_1, s_2} (\tilde{\nu}_{s_1 s_2}^{(2)})^2 = p,
$$

we have

$$
E\{ (Y_{j_1}^T Y_i)^2 (Y_{j_2}^T Y_i)^2 \} = O(p^2) \quad \text{for} \quad j_1 \neq j_2 \neq i. \tag{3.40}
$$

In summary of (3.32), (3.33), (3.34), (3.36), (3.39) and (3.40), we have

$$
E(\xi_{n1}) = O(p^4/n^4 + p^3/n^3 + p^2/n^2) = O(p^2/n^2)
$$

as $p = o(\sqrt{n})$. This leads to the conclusion of the lemma.

Lemma 3.9 *Under the conditions of Theorem 3.3, $||\beta_{n2}|| = O_p(\sqrt{\gamma_p}p/n)$.*

Proof:

Note that

$$
\begin{aligned}
||\beta_{n2}||^2 &= n^{-6} \sum_{i,j_1,j_2,l_1,l_2} Y_{j_1}^T Y_i Y_{j_2}^T Y_i Y_{l_1}^T Y_i Y_{l_2}^T Y_i (X_i - \mu)^T (X_i - \mu) \\
&+ n^{-6} \sum_{i_1 \neq i_2, j_1, j_2, l_1, l_2} Y_{j_1}^T Y_{i_1} Y_{j_2}^T Y_{i_1} Y_{l_1}^T Y_{i_2} Y_{l_2}^T Y_{i_2} (X_{i_1} - \mu)^T (X_{i_2} - \mu) \\
&=: F_{n1} + F_{n2}
\end{aligned}
\tag{3.41}
$$

and

$$
\begin{aligned}
F_{n1} &= n^{-6} \sum_{i=1}^{n} (X_i - \mu)^T (X_i - \mu)(Y_i^T Y_i)^4 \\
&+ n^{-6} \sum_{i \neq j} (X_i - \mu)^T (X_i - \mu)\}\{(Y_j^T Y_i)^4 + 6(Y_i^T Y_i)^2 (Y_j^T Y_i)^2 \\
&+ 4(Y_i^T Y_i)(Y_j^T Y_i)^3 + 4(Y_i^T Y_i)^3 (Y_j^T Y_i)\} \\
&+ 6n^{-3} \sum_{i \neq j \neq l} (Y_j^T Y_i)^2 (Y_l^T Y_i)^2 (X_i - \mu)^T (X_i - \mu).
\end{aligned}
\tag{3.42}
$$

As there are more terms in F_{n2}, we classify them by the number of distinct subscripts involved. In particular we assign F_{n2j}, $j = 2, 3, 4$, to be terms of F_{n2} which have j distinct subscripts. Then,

$$
\begin{aligned}
F_{n22} &= n^{-6} \sum_{i \neq j} (X_i - \mu)^T (X_j - \mu)\{2(Y_j^T Y_i)^4 + 4Y_i^T Y_i Y_i^T Y_j (Y_j^T Y_j)^2 + (Y_i^T Y_i)^2 (Y_j^T Y_j)^2 \\
&\qquad\qquad + (Y_i^T Y_i)(Y_j^T Y_i)^2 Y_j^T Y_j\},
\end{aligned}
$$

$$
\begin{aligned}
F_{n23} &= n^{-6} \sum_{i_1 \neq i_2 \neq j} (X_{i_1} - \mu)^T (X_{i_2} - \mu)\{(Y_j^T Y_{i_1})^2 (Y_j^T Y_{i_2})^2 + 4(Y_{i_1}^T Y_{i_1})(Y_{i_2}^T Y_{i_2})^2 Y_j^T Y_{i_1} Y_j^T Y_{i_2} \\
&\qquad\qquad + 8(Y_{i_2}^T Y_{i_1})^2 (Y_j^T Y_{i_1})^2 Y_j^T Y_{i_2} Y_j^T Y_{i_2}\}
\end{aligned}
$$

and

$$
\begin{aligned}
F_{n24} &= n^{-6} \sum_{i_1 \neq i_2 \neq j \neq l} (X_{i_1} - \mu)^T (X_{i_2} - \mu)\{2(Y_j^T Y_{i_1})^2 (Y_l^T Y_{i_2})^2 \\
&\qquad\qquad + 4(Y_{i_1}^T Y_{i_1})(Y_l^T Y_{i_1})(Y_j^T Y_{i_2})(Y_l^T Y_{i_2})\}.
\end{aligned}
$$

Here we only derive $E(F_{n24})$ as it has the largest number of terms $\{n(n-1)(n-2)(n-3)\}$ in the summation. Working out the expectation for the other terms is similar, and it can be shown that the order of magnitude of these expectations is at most $O\{\gamma_p p^2/n^2\}$.

Recall that $\Gamma^T\Gamma = (\nu_{ls})_{1\leq l,s\leq m}$ and $(X_{i_1} - \mu)^T(X_{i_2} - \mu) = \sum_{t_1,t_2} Z_{i_1 t_1} Z_{i_2 t_2} \nu_{t_1 t_2}$. Hence for four mutually different i_1, i_2, j and l,

$$E\{(X_{i_1} - \mu)^T(X_{i_2} - \mu)(Y_j^T Y_{i_1})^2(Y_l^T Y_{i_2})^2\}$$

$$= \sum_{k_1,\cdots,k_4,s_1,\cdots s_4,t_1,t_2}^m E(Z_{jk_1} Z_{jk_3} Z_{i_1 k_2} Z_{i_1 k_4} Z_{ls_1} Z_{ls_3} Z_{i_2 s_2} Z_{i_2 s_4} Z_{i_1 t_1} Z_{i_2 t_2}) \tilde{\nu}_{k_1 k_2} \tilde{\nu}_{k_3 k_4} \tilde{\nu}_{s_1 s_2} \tilde{\nu}_{s_3 s_4} \nu_{t_1 t_2}$$

$$= \sum_{k_1,k_2,s_1,s_2}^m E(Z_{jk_1} Z_{jk_1} Z_{i_1 k_2} Z_{i_1 k_2} Z_{ls_1} Z_{ls_1} Z_{i_2 s_2} Z_{i_2 s_2} Z_{i_1 k_2} Z_{i_2 s_2}) \tilde{\nu}_{k_1 k_2} \tilde{\nu}_{k_1 k_2} \tilde{\nu}_{s_1 s_2} \tilde{\nu}_{s_1 s_2} \nu_{k_2 s_2}$$

$$= m_3^2 \sum_{k_1,k_2,s_1,s_2}^m \tilde{\nu}_{k_1 k_2}^2 \tilde{\nu}_{s_1 s_2}^2 \nu_{k_2 s_2}.$$

We note here in the first equation above, k_1 and k_3, s_1 and s_3, k_2, k_4 and t_1, and s_2, s_4 and t_2 must be the same respectively to avoid zero means. As $|\nu_{k_2 s_2}| \leq \gamma_p(\Gamma^T\Gamma) = \gamma_p$,

$$|E\{(X_{i_1} - \mu)^T(X_{i_2} - \mu)(Y_j^T Y_{i_1})^2(Y_l^T Y_{i_2})^2\}| \leq m_3^2 \gamma_p (\sum_{k_1,k_2} k\tilde{\nu}_{k_1 k_2}^2)^2 = m_3^2 \gamma_p p^2. \quad (3.43)$$

The mean of the second term in F_{n24} is

$$E\{(X_{i_1} - \mu)^T(X_{i_2} - \mu)(Y_{i_1}^T Y_{i_1})(Y_l^T Y_{i_1})(Y_j^T Y_{i_2})(Y_l^T Y_{i_2})\}$$

$$= \sum_{k_1,\cdots,k_4,s_1,\cdots s_4,t_1,t_2}^m E(Z_{jk_1} Z_{jk_3} Z_{i_1 k_2} Z_{i_1 s_2} Z_{ls_1} Z_{ls_3} Z_{i_2 k_4} Z_{i_2 s_4} Z_{i_1 t_1} Z_{i_2 t_2}) \tilde{\nu}_{k_1 k_2} \tilde{\nu}_{k_3 k_4} \tilde{\nu}_{s_1 s_2} \tilde{\nu}_{s_3 s_4} \nu_{t_1 t_2}$$

$$= m_3^2 \sum_{t,s,t_1,t_2} \tilde{\nu}_{tt_1} \tilde{\nu}_{st_1} \tilde{\nu}_{tt_2} \tilde{\nu}_{st_2} \nu_{t_1 t_2} = m_3^2 \sum_{t_1,t_2} (\tilde{\nu}_{t_1 t_2}^{(2)})^2 \nu_{t_1 t_2}.$$

Hence $|E\{(X_{i_1} - \mu)^T(X_{i_2} - \mu)(Y_{i_1}^T Y_{i_1})(Y_l^T Y_{i_1})(Y_j^T Y_{i_2})(Y_l^T Y_{i_2})\}| \leq \gamma_p tr\{(\tilde{\Gamma}^T\tilde{\Gamma})^2\} = \gamma_p p$.

Thus, $E(F_{n24}) = O(\gamma_p p^2/n^2)$ and $E(\|\beta_{n2}\|^2) = O(\gamma_p p^2/n^2)$, which lead to the lemma.

Lemma 3.10 *Under the conditions of Theorem 3.3,*

$$n\bar{Y}^T D_n \bar{Y} = O_p(p/\sqrt{n}) \quad and \quad n\bar{Y}^T D_n^2 \bar{Y} = O(p^2/n).$$

Proof:

Write $A_n = \bar{Y}^T D_n \bar{Y}$. Recall that $\tilde{\Sigma} = \Gamma^T \Sigma^{-1} \Gamma = (\tilde{\nu}_{lk})$. As $\tilde{\Sigma}$ is idempotent and $tr(\tilde{\Sigma}) = p$, we are to use the following facts repeatedly throughout our derivations:

$$\sum_{s=1}^{m} \tilde{\nu}_{ss}^3 \le \gamma_p(\tilde{\Sigma})^2 tr(\tilde{\Sigma}) = p, \quad \sum_{s,l}^{m} \tilde{\nu}_{sl}^2 \tilde{\nu}_{ll} \le \gamma_p(\tilde{\Sigma}) tr(\tilde{\Sigma}^2) = p \quad \text{and}$$

$$|\sum_{s,l} \tilde{\nu}_{ss} \tilde{\nu}_{ls} \tilde{\nu}_{ll}| \le \sum_{s,l} \tilde{\nu}_{ss} \tilde{\nu}_{ll} = p^2. \tag{3.44}$$

Since

$$\begin{aligned}
A_n &= \bar{Z} \tilde{\Sigma} n^{-1} \sum Z_i Z_i^T \tilde{\Sigma} \bar{Z} - \bar{Z} \tilde{\Sigma} \bar{Z} \\
&= n^{-3} \sum_{s,l,l_1,l_2} \sum_{i_1,i_2,,i_3}^{n} Z_{i_1 s} Z_{i_2 l} Z_{i_3 l_1} Z_{i_3 l_2} \tilde{\nu}_{sl_1} \tilde{\nu}_{l_2 l} - n^{-2} \sum_{s,l} \sum_{i_1,i_2}^{n} Z_{i_1 s} Z_{i_2 l} \tilde{\nu}_{sl},
\end{aligned}$$

$$\begin{aligned}
E(A_n) &= n^{-3} \sum_{s,l,l_1,l_2}^{m} \{n E(Z_{is} Z_{il} Z_{il_1} Z_{il_2}) \tilde{\nu}_{sl_1} \tilde{\nu}_{l_2 l} + n(n-1) \delta_{sl} \delta_{l_1 l_2}\} - n^{-1} \sum_{s}^{m} \tilde{\nu}_{ss} \\
&= n^{-2} \{m_4 \sum_{s}^{m} \tilde{\nu}_{ss}^2 + \sum_{s \ne l} (2\tilde{\nu}_{sl}^2 + \tilde{\nu}_{ss} \tilde{\nu}_{ll})\} - n^{-2} p \\
&= n^{-2} \{(m_4 - 3) \sum_{s} \tilde{\nu}_{ss}^2 + p + p^2\} = O(p^2/n^2) \tag{3.45}
\end{aligned}$$

as $\sum_s \tilde{\nu}_{ss}^2 \le \gamma_p(\tilde{\Sigma}) tr(\tilde{\Sigma}) = p$.

As D_n is not necessarily non-negative definite, we have to derive $E(A_n^2)$, which can

be expressed as

$$
\begin{aligned}
E(A_n^2) &= n^{-6} E \left(\sum_{q,l,s,t}^{m} \sum_{i_1,i_2,,i_3}^{n} Z_{i_1 q} Z_{i_2 l} Z_{i_3 s} Z_{i_3 t} \tilde{\nu}_{qs} \tilde{\nu}_{tl} \right)^2 \\
&\quad - 2n^{-5} \sum_{k_1,l_1,l_2,l_3,k_2,q}^{m} \sum_{i_1,i_2,i_3,j_1,j_2}^{n} E \left(Z_{i_1 k_1} Z_{i_2 l_1} Z_{i_3 l_2} Z_{i_3 l_3} Z_{j_1 k_2} Z_{j_2 q} \right) \tilde{\nu}_{k_1 l_2} \tilde{\nu}_{l_3 l_1} \tilde{\nu}_{k_2 q} \\
&\quad + n^{-4} \sum_{k_1,k_2,l_1,l_2}^{m} \sum_{i_1,i_2,i_3,i_4}^{n} E(Z_{i_1 k_1} Z_{i_2 l_1} Z_{i_3 k_2} Z_{i_4 l_2}) \tilde{\nu}_{k_1 l_1} \tilde{\nu}_{k_2 l_2} \\
&=: I_1 - 2I_2 + I_3.
\end{aligned}
\tag{3.46}
$$

Now

$$
\begin{aligned}
I_3 &= n^{-4} \sum_{k_1,k_2,l_1,l_2}^{m} \tilde{\nu}_{k_1 l_1} \tilde{\nu}_{k_2 l_2} \Big\{ \sum_{i}^{n} E(Z_{ik_1} Z_{il_1} Z_{ik_2} Z_{il_2}) + \sum_{i_1 \neq i_2} (\delta_{k_1 l_1} \delta_{k_2 l_2} + \delta_{k_1 l_2} \delta_{k_2 l_1} + \delta_{k_1 k_2} \delta_{l_1 l_2}) \Big\} \\
&= n^{-3} \Big\{ m_4 \sum_{s=1}^{m} \tilde{\nu}_{ss}^2 + \sum_{s \neq l}(2\tilde{\nu}_{sl}^2 + \tilde{\nu}_{ss}\tilde{\nu}_{ll}) \Big\} + n^{-2}(1-n^{-1}) \sum_{s,l}^{m} \left(\tilde{\nu}_{ss}\tilde{\nu}_{ll} + 2\tilde{\nu}_{sl}^2 \right) \\
&= n^{-3}(m_3 - 3) \sum_{s=1}^{m} \tilde{\nu}_{ss}^2 + n^{-2}(1-n^{-1})(p^2 + p)
\end{aligned}
\tag{3.47}
$$

and

$$
\begin{aligned}
I_2 &= n^{-5} \sum_{k_1,l_1,l_2,l_3,k_2,q}^{m} \tilde{\nu}_{k_1 l_2} \tilde{\nu}_{l_3 l_1} \tilde{\nu}_{k_2 q} \Big\{ \sum_{i=1}^{n} E \left(Z_{ik_1} Z_{il_1} Z_{il_2} Z_{il_3} Z_{ik_2} Z_{iq} \right) \\
&\quad + \sum_{i \neq j}^{n} E(Z_{ik_1} Z_{il_1} Z_{il_2}) E(Z_{jl_3} Z_{jk_2} Z_{jq})[4; q; k_1, l_1, l_2] \\
&\quad + \sum_{i_1 \neq i_2 \neq i_3} \delta_{l_2 l_3}(\delta_{k_1 l_1} \delta_{k_2 q} + \delta_{k_1 k_2} \delta_{l_1 q} + \delta_{k_1 q} \delta_{k_2 l_1}) \Big\} \\
&= n^{-4} \Big\{ m_6 \sum_{s=1}^{m} \tilde{\nu}_{ss}^3 + m_4 \sum_{s \neq l}^{m} \left(12\tilde{\nu}_{sl}^2 \tilde{\nu}_{ll} + 3\tilde{\nu}_{ss}\tilde{\nu}_{ll}^2 \right) \\
&\quad + m_3^2 \sum_{s \neq l}(6\tilde{\nu}_{ss}\tilde{\nu}_{ls}\tilde{\nu}_{ll} + 4\tilde{\nu}_{sl}^3) + \sum_{s \neq l \neq q}(24\tilde{\nu}_{sl}^2 \tilde{\nu}_{qq} + 6\tilde{\nu}_{sl}\tilde{\nu}_{lq}\tilde{\nu}_{sq}) \Big\} \\
&\quad + 2n^{-3}(1-n^{-1}) \sum_{s \neq l}(\tilde{\nu}_{sl}^3 + \tilde{\nu}_{ss}\tilde{\nu}_{ls}\tilde{\nu}_{ll}) \\
&\quad + n^{-2}(1-n^{-1})(1-n^{-2}) \sum_{s,l,q}(\tilde{\nu}_{sl}^2 \tilde{\nu}_{qq} + 2\tilde{\nu}_{sl}\tilde{\nu}_{lq}\tilde{\nu}_{sq})
\end{aligned}
$$

where $[4; q; k_1, l_1, l_2]$ denotes 4 rotations of having q, k_1, l_1 and l_2 taking the current position of q. Using (3.44), we note that the terms starting with n^{-4} and n^{-3} in both I_3 and I_2 contribute only to the order of $O(p^2/n^{-3})$ or smaller. Hence, it can be readily shown that

$$I_3 = n^{-2}(p^2 + 2p) + O(n^{-3}p^2) \quad \text{and} \quad I_2 = n^{-2}(p^2 + 2p) + O(n^{-3}p^2). \tag{3.48}$$

It remains to derive I_1. Note that

$$
\begin{aligned}
I_1 \;=\; & \sum_{k_1,k_2,l_1,l_2,s_1,s_2,t_1,t_2} \tilde{\nu}_{k_1 s_1} \tilde{\nu}_{t_1 l_1} \tilde{\nu}_{k_2 s_2} \tilde{\nu}_{t_2 l_2} \Big[n^{-5} E(Z_{ik_1} Z_{ik_2} Z_{il_1} Z_{il_2} Z_{is_1} Z_{is_2} Z_{it_1} Z_{it_2}) \\
& + n^{-6} \sum_{i \neq j} E(Z_{ik_1} Z_{il_1} Z_{is_1} Z_{it_1}) E(Z_{jk_2} Z_{jl_2} Z_{js_2} Z_{jt_2})[6; k_1, l_1; k_2, l_2] \\
& + n^{-6} \sum_{i_1 \neq i_3 \neq i_6} \delta_{k_1,k_2} E(Z_{i_3 l_1} Z_{i_3 s_1} Z_{i_3 t_1}) E(Z_{i_6 l_2} Z_{i_6 s_2} Z_{i_6 t_2})[12; l_1; l_2; k_1, k_2] \\
& + n^{-6} \sum_{i_1 \neq i_3 \neq i_6} \{ \delta_{s_1 t_1} \delta_{s_2 t_2} E(Z_{i_1 k_1} Z_{i_1 l_1} Z_{i_1 k_2} Z_{i_1 l_2}) \\
& \quad + E(Z_{i_3 s_1} Z_{i_3 t_1} Z_{i_3 s_2} Z_{i_3 t_2})(\delta_{k_1 l_1} \delta_{k_2 l_2} + \delta_{k_1 l_2} \delta_{k_2 l_1} + \delta_{k_1 k_2} \delta_{l_1 l_2}) \} \\
& + n^{-6} \sum_{i_1 \neq i_2 \neq i_3 \neq i_6} \delta_{s_1 t_1} \delta_{s_2 t_2} (\delta_{k_1 l_1} \delta_{k_2 l_2} + \delta_{k_1 l_2} \delta_{k_2 l_1} + \delta_{k_1 k_2} \delta_{l_1 l_2}) \Big].
\end{aligned}
$$

All the terms except the last term in the above equation are $O(n^{-3}p^2)$ and by working out the last term, we have

$$I_1 = n^{-2}(p^2 + 2p) + O(n^{-3}p^2). \tag{3.49}$$

Combine (3.46), (3.49) and (3.48),

$$E(\bar{Y}^T D_n \bar{Y})^2 = O(n^{-3}p^2) \tag{3.50}$$

which establishes the first part of the lemma.

On the second part of the lemma, let $B_n = \bar{Z}\tilde{\Sigma}n^{-1}\sum_{i=1}^n Z_i Z_i^T \tilde{\Sigma}n^{-1}\sum_{i=1}^n Z_i Z_i^T \tilde{\Sigma}$, then

$$\bar{Y}^T D_n^2 \bar{Y} = B_n - 2A_n - \bar{Y}^T \bar{Y}$$

where $A_n = \bar{Y}^T D_n \bar{Y}$. From (3.45),

$$E(-2A_n - \bar{Y}^T \bar{Y}) = -2n^{-2}\{(m_4 - 3)\sum_s \tilde{\nu}_{ss}^2 + p + p^2\} - n^{-1}p. \qquad (3.51)$$

It remains to derive $E(B_n)$. As

$$B_n = n^{-4} \sum_{k_1,l_1,l_2,s_1,s_2,t} \sum_{i_1,i_2,i_3,i_4} \tilde{\nu}_{k_1 l_1} \tilde{\nu}_{s_1 l_2} \tilde{\nu}_{s_2 t} Z_{i_1 k_1} Z_{i_2 l_1} Z_{i_2 s_1} Z_{i_3 s_2} Z_{i_3 l_2} Z_{i_4 t},$$

by carrying out derivations similar to, but slightly less involved than, those of $E(A_n^2)$, it can be shown that

$$E(B_n) = n^{-1}p + O(n^{-2}p^2).$$

This together with (3.51) means that

$$E(\bar{Y}^T D_n^2 \bar{Y}) = O(p^2/n^2)$$

which leads to the second conclusion of the lemma.

Proof of Theorem 3.3.

The key in our proof here is to update the rates given in (3.28) and (3.29) when we have more moments for Z_{il} under our disposal.

We first update (3.25) by noting that

$$
\begin{aligned}
||\beta_{n1}|| &\leq \max_{1 \leq i \leq n} ||X_i - \mu|| n^{-1} \sum_{i=1}^n \lambda^T (X_i - \mu)(X_i - \mu)^T \lambda = \max_{1 \leq i \leq n} ||X_i - \mu|| ||\lambda||^2 O_p(\gamma_p) \\
&= o_p(\gamma_p^{\frac{4k+1}{4k}} \{tr(\Sigma)\}^{-\frac{6k-1}{4k}} n^{-\frac{4k-1}{4k}}) + O_p(tr^{3/2}(\Sigma) n^{-1} \gamma_p). \qquad (3.52)
\end{aligned}
$$

Let $\lambda_0 = \Sigma^{-1}(\bar{X} - \mu)$, $\eta_1 = (S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu)$ and $\eta_2 = S_n^{-1}\beta_n$. Then we can write (3.26) as

$$
\begin{aligned}
\lambda &= \Sigma^{-1}(\bar{X} - \mu) + (S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) + S_n^{-1}\beta_n \\
&=: \lambda_0 + \eta_1 + \eta_2. \qquad (3.53)
\end{aligned}
$$

The order of $||\lambda_0||$ is of $\sqrt{tr(\Sigma^{-1})/n}$ which can be smaller than $\sqrt{tr(\Sigma)/n}$, the existing order for $||\lambda||$ given in Theorem 3.1. From Lemmas 3.7 and 3.10, and (**??**)

$$
\begin{aligned}
||\eta_1||^2 &= (\bar{X} - \mu)^T (S_n^{-1} - \Sigma^{-1})^2 (\bar{X} - \mu) = \bar{Y}^T (V_n^{-1} - I_p) \Sigma^{-1} (V_n^{-1} - I_p) \bar{Y} \\
&\leq \gamma_1^{-1} \bar{Y}^T (V_n^{-1} - I_p)^2 \bar{Y} = \gamma_1^{-1} \gamma_1^{-1} \bar{Y}^T \{-D_n + D_n^2 + D_n^2 (V_n^{-1} - I_p)\}^2 \bar{Y} \\
&= O_p(\bar{Y}^T D_n \bar{Y}) = O_p(p^2/n^2).
\end{aligned}
\tag{3.54}
$$

Hence $||\eta_1|| = o(||\lambda_0||)$ as $p = o[\{ntr(\Sigma^{-1})\}^{1/2}]$ is trivially true. Also $||\eta_2||^2 \leq ||\beta_n||^2 \gamma_1^{-2}(S_n)$. Hence from (3.52) and Lemma 3.4, $||\eta_{n2}|| = o_p(||\lambda_0||)$ if

$$
\{tr(\Sigma)\}^{\frac{6k-1}{2k}} tr^{-1}(\Sigma^{-1}) = O_p(\gamma_p^{-\frac{4k+1}{2k}} n^{\frac{2k-1}{2k}})
$$

which is implied by the assumption $p^3 \gamma_p^5 = o\{tr(\Sigma^{-1})n^{1 - \frac{1}{4k}}\}$.

With (3.53), the log EL ratio

$$
\begin{aligned}
w_n(\mu) &= 2n\lambda^T(\bar{X} - \mu) - n\lambda^T S_n \lambda + \tfrac{2}{3} R_n\{1 + o_p(1)\} \\
&= n(\bar{X} - \mu)^T \Sigma^{-1}(\bar{X} - \mu) + n(\bar{X} - \mu)^T (\Sigma^{-1} S_n \Sigma^{-1} - \Sigma^{-1})(\bar{X} - \mu) \\
&+ n(\bar{X} - \mu)^T (S_n^{-1} - \Sigma^{-1}) S_n (S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) \\
&+ n\beta_n^T(S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) + \beta_n^T S_n^{-1} \beta_n + \tfrac{2}{3} R_n\{1 + o_p(1)\}
\end{aligned}
\tag{3.55}
$$

where $R_n = \sum_{i=1}^n \{\lambda^T(X_i - \mu)\}^3$.

From Lemma 3.10,

$$
n(\bar{X} - \mu)^T(\Sigma^{-1} S_n \Sigma^{-1} - \Sigma^{-1})(\bar{X} - \mu) = n\bar{Y}^T D_n \bar{Y} = O_p(p/\sqrt{n}) = o_p(1).
$$

Since $V_n^{-1} - I_p = -D_n + D_n^2 + D_n^2(V_n^{-1} - I_p)$,

$$
\begin{aligned}
&n(\bar{X} - \mu)^T(S_n^{-1} - \Sigma^{-1}) S_n (S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) \\
&= n\bar{Y}^T(V_n^{-1} - I_p) V_n (V_n^{-1} - I_p)\bar{Y} = n\bar{Y}^T\{D_n^2 - D_n^3 - D_n^3(V_n^{-1} - I_p)\}\bar{Y}.
\end{aligned}
$$

As $|n\bar{Y}^T D_n^3(V_n^{-1} - I_p)\bar{Y}| \le |\gamma_p(V_n^{-1} - I_p)|n\bar{Y}^T D_n^3 \bar{Y}$ and $\gamma_p(V_n^{-1} - I_p) = o_p(1)$ as implied from Lemma 3.4, we have from (3.23) and Lemma 3.10 that

$$n(\bar{X} - \mu)^T(S_n^{-1} - \Sigma^{-1})S_n(S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) = O_p(p/\sqrt{n}) = o_p(1).$$

From (3.53), we can use λ_0 to replace λ in β_{n1} and R_n, which results in

$$\beta_n = \beta_{n2}\{1 + o_p(1)\} \quad \text{and} \quad \xi_n = n^{-1}\sum_{i=1}^n W_i^4 = \xi_{n1}\{1 + o_p(1)\}$$

where β_{n2} and ξ_{n1} are defined in (3.30). From Lemma 3.9

$$
\begin{aligned}
|n\beta_n S_n^{-1}\beta_n| &\le n||\beta_n||^2\gamma_p(S_n^{-1}) = n||\beta_{n2}||^2/\gamma_1\{1 + o_p(1)\} \\
&= O_p(\gamma_p p^2 n^{-1}) = o_p(\sqrt{p})
\end{aligned}
$$

as $p\gamma_p = o(\sqrt{n})$. From Lemma 3.8 and the assumption $p^3\gamma_p^5 = o\{tr(\Sigma^{-1})n^{1 - \frac{1}{4k}}\}$

$$
\begin{aligned}
R_n &\le \sum_{i=1}^n |W_i|^3\{1 + o_p(1)\} \le \sqrt{n\lambda^T S_n \lambda}\left\{\sum_{i=1}^n ||\lambda||^4||X_i - \mu||^4\right\}^{1/2} \\
&= O_p(\sqrt{tr(\Sigma)\gamma_p})O_p\{\sqrt{tr^2(\Sigma)n^{-2}np^2/n^2}\} \\
&= O_p\{tr^{3/2}(\Sigma)\gamma_p^{1/2}pn^{-3/2}\} = o_p(\sqrt{p}).
\end{aligned}
$$

At last, from (3.54) and Lemma 3.4,

$$|n\beta_n^T(S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu)| \le n||\beta_n||||\eta_{n1}|| = O_p(\sqrt{\gamma_p}p^2/n) = o_p(\sqrt{p}).$$

Repeating the last part of the proof of Theorem 3.2, the proof Theorem 3.3 is completed.

Figure 3.1 Normal Q-Q plots with the faster growth rate $p = c_1 n^{0.4}$ for the normal (upper panels) and the pareto (lower panel) innovations: $c_1 = 3$ (solid line), 4 (dotted lines) and 5 (dashed lines).

89
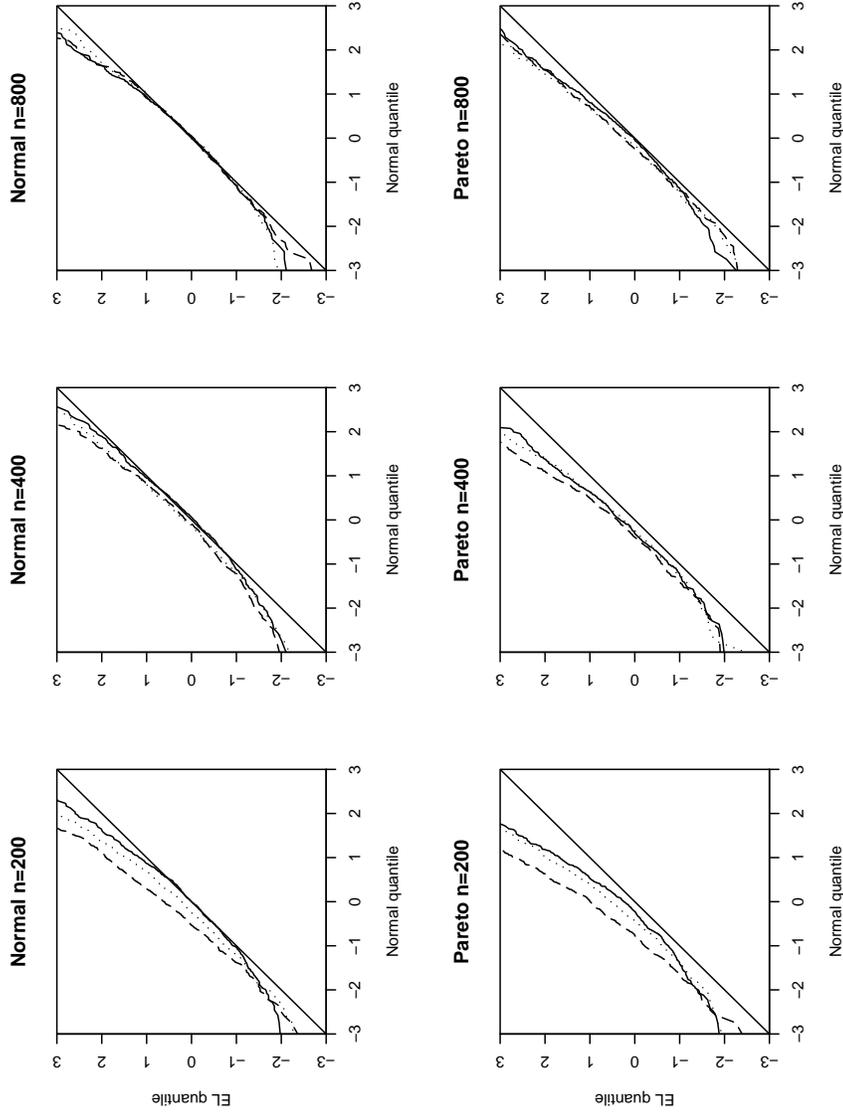
Figure 3.2   Normal Q-Q plots with the slower growth rate $p = c_1 n^{0.24}$ for the normal (upper panels) and the pareto (lower panel) innovations: $c_2 = 4$ (solid line), 6 (dotted lines) and 8 (dashed lines).
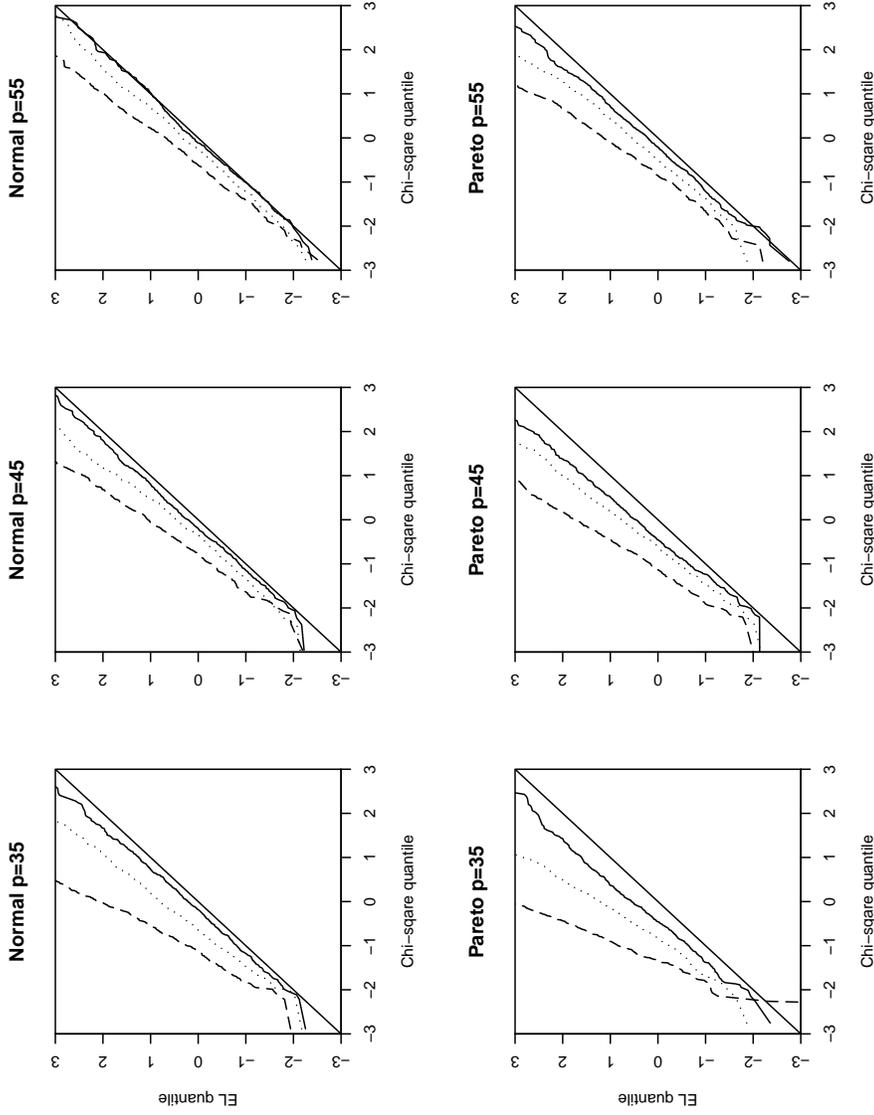
Figure 3.3 Chi-square Q-Q plots with the slower growth rate $n = (p/c_1)^{1/0.4}$ for the normal (upper panels) and the pareto (lower panel) innovations: $c_1 = 3$ (solid line), 4 (dotted lines) and 5 (dashed lines).
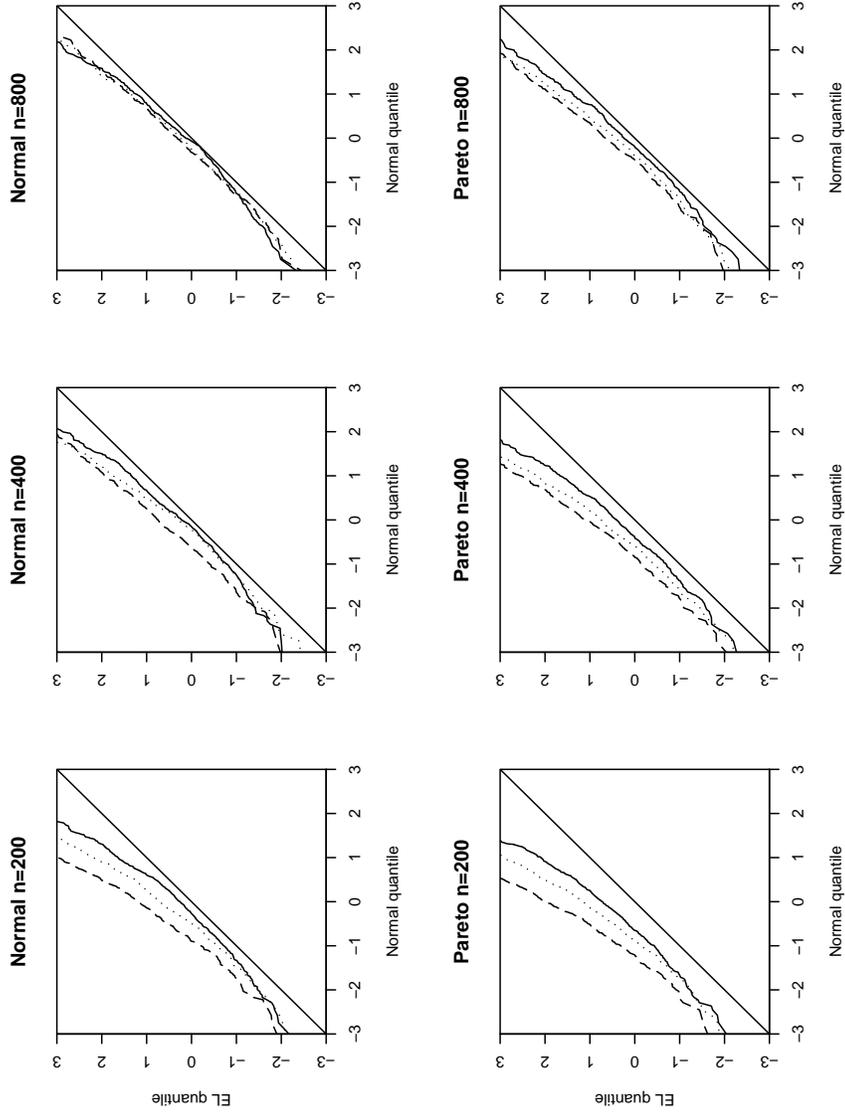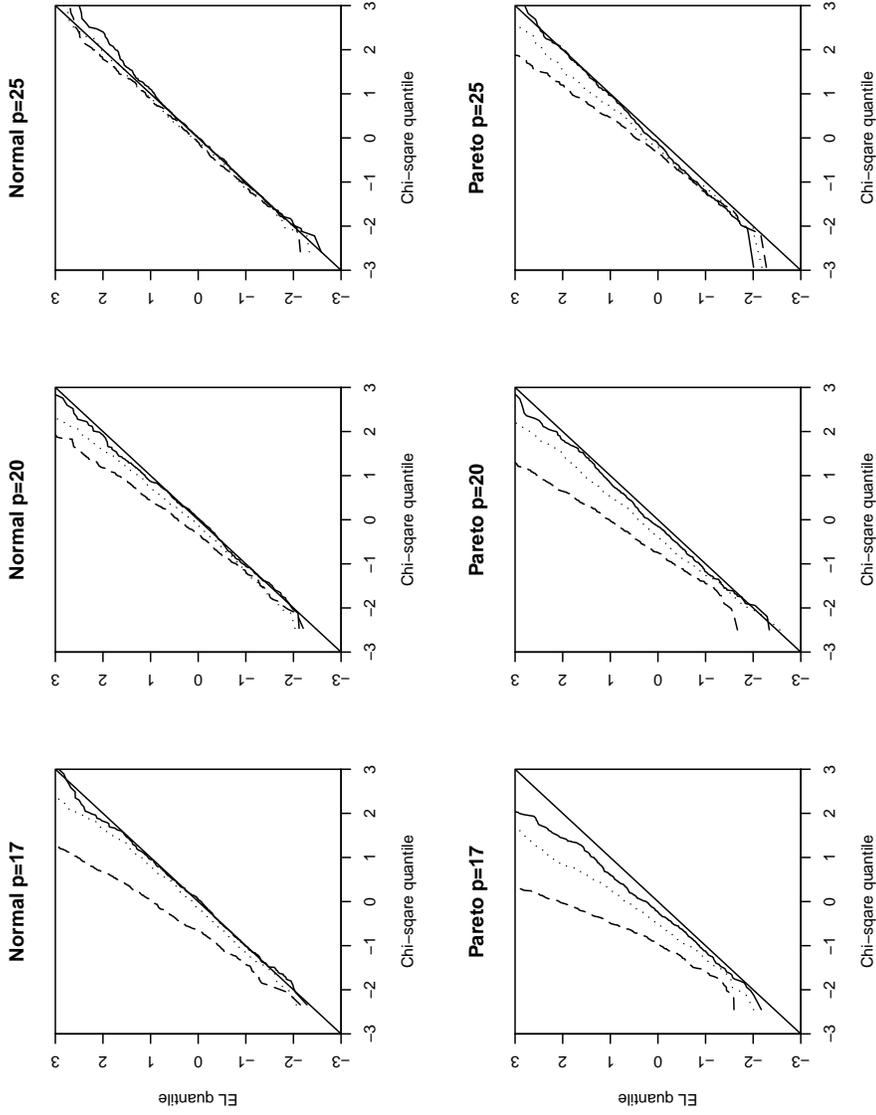
Figure 3.4 Chi-square Q-Q plots with the faster growth rate $n = (p/c_2)^{1/0.24}$ for the normal (upper panels) and the pareto (lower panel) innovations: $c_2 = 4$ (solid line), 5 (dotted lines) and 6 (dashed lines).

Table 3.1   Empirical likelihood tests for Dow Jones data

Sector	Dimension	Test Statistic	P-values
Basic Material	4	3.69	0.0001
Consumer Goods	4	2.11	0.02
Finance	5	-.54	0.69
Health Care	3	0.75	0.23
Industrial Goods	3	1.01	0.16
Services	4	1.74	0.04
Technology	7	1.35	0.09
Overall	30	1.56	0.056

CHAPTER 4.   Hypothesis Testing For High-Dimensional Distributions

4.1   Introduction

Let us consider two random samples $X_{i1}, X_{i2}, \cdots, X_{in_i} \in R^p$ for $i = 1$ and 2, with multivariate continuous distributions F_1 and F_2. The distinctions between two distributions are not limited to inequality of their mean vectors. We therefore consider in this chapter a test for equality of distributions and that is, the hypothesis of interest is

$$H_0 : F_1 = F_2 \text{ vs. } H_1 : F_1 \neq F_2.$$

For the univariate case, traditionally, people would use rank tests, Kolmogrorov-Smirnov and Cramér-von Mises tests and many others. See Darling (1957) for detailed discussions of "goodness of fit" tests and two-sample tests based on the empirical distribution functions. For multivariate Kolmogrorov-Smirnov tests, see Peacock (1983) and Fasano and Franceschini (1987). Friedman and Rafsky (1979) presented multivariate generalizations of the Wald-Wolfowitz runs statistic and the Smirnov maximum deviation statistic for the two-sample problems. Bickel (1969) considered a multivariate Smirnov test which is consistent against all alternatives given the conditional convergence of the empirical distribution functions. Ahmad (1996) provided some results on modification of two-sample univariate and multivariate Cramér-von Mises

test. All of these existing methods have been effective in testing for equal distributions for fixed dimensional data. However, the generalization to the case where $p \to \infty$ has not been explored.

For arbitrarily high-dimensional data, Baringhaus and Franz (2004) proposed a test statistic (BF test) based on between and within groups' pair-wise Euclidean distances. As the BF test converges in distribution to a Brownian bridge which depends on an unknown distribution, the authors suggested to use the bootstrap method to get critical values. Alba Fernández et al. (2008) constructed a similar test statistic as shown in (1.3), using an empirical characteristic function. Hall and Tajvidi (2002) proposed a permutation test (HT test). Its critical values are determined conditional on the pairwise distances between pooled data. There is an issue of computational burden in the HT test due to the distance ranking strategy. Motivated by the application of identifying differentially expressed gene-sets, a Multiresponse Permutation Procedure (MRPP) test was investigated in Nettleton et al. (2008). The authors also addressed the problem of controlling false discovery rate for multiple gene-sets testing.

In this chapter we propose a test the equality of two distributions in high-dimensional settings. In conjunction with the two-sample test for means we pursue the asymptotic distribution for the proposed distribution test. This chapter is organized as follows. In Section 4.2, we present the distribution test for high-dimensional data and explore its asymptotic distribution via two-sample U-statistic theory and the martingale central limit theorem. Section 4.3 contains simulation results. Technical proofs are provided in Section 4.4.

4.2 Main Results

The statistic T_{n_1,n_2} which is a distance measure between two continuous cumulative distribution functions (cdfs) F_1 and F_2, is

$$
\begin{aligned}
T_{n_1,n_2} \;=\; \int \Bigg\{ & \frac{1}{n_1(n_1-1)} \sum_{i_1 \neq i_2}^{n_1} I(X_{1i_1} \leq x)I(X_{1i_2} \leq x) \\
& -\frac{2}{n_1 n_2} \sum_{i=1}^{n_1} I(X_{1i} \leq x) \sum_{j=1}^{n_2} I(X_{2j} \leq x) \\
& +\frac{1}{n_2(n_2-1)} \sum_{j_1 \neq j_2}^{n_2} I(X_{2j_1} \leq x)I(X_{2j_2} \leq x) \Bigg\} w(x)dx,
\end{aligned}
$$

where $w(x)$ is a known density (weight) function on R^p, $I(\cdot)$ is the indicator function and $X_{ij} \leq x$ if and only if $X_{ijl} \leq x_l$, for $l = 1, \cdots, p$. For simplicity of notation, let us define

$$
\text{``} \int \text{''} = \int_{x_1} \int_{x_2} \cdots \int_{x_p} \quad \text{and} \quad \text{``}dx\text{''} = dx_p \cdots dx_1.
$$

It is shown in Section 4.4 that $E(T_{n_1,n_2}) = \int \left\{ F_1(x) - F_2(x) \right\}^2 w(x)dx$. Precisely, this formula explains that T_{n_1,n_2} is a weighted average of distance measure between two cdfs F_1 and F_2.

4.2.1 Limiting distribution of T_{n_1,n_2} under the null H_0

The idea is to consider T_{n_1,n_2} or part of T_{n_1,n_2} as a sum of martingale differences. The adventure of exploring the limiting distribution of T_{n_1,n_2} is enlivened by the possibility of applying the martingale central limit theorem. Let us begin with partitioning

T_{n_1,n_2} into $T^{(1)}_{n_1,n_2}$ and $T^{(2)}_{n_1,n_2}$.

$$T^{(1)}_{n_1,n_2} = \int \left[\frac{1}{n_1(n_1-1)} \sum_{i_1 \neq i_2}^{n_1} \left\{ I(X_{1i_1} \leq x) - F_1(x) \right\} \left\{ I(X_{1i_2} \leq x) - F_1(x) \right\} \right.$$

$$- \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \left\{ I(X_{1i} \leq x) - F_1(x) \right\} \sum_{j=1}^{n_2} \left\{ I(X_{2j} \leq x) - F_2(x) \right\}$$

$$\left. + \frac{1}{n_2(n_2-1)} \sum_{j_1 \neq j_2}^{n_2} \left\{ I(X_{2j_1} \leq x) - F_2(x) \right\} \left\{ I(X_{2j_2} \leq x) - F_2(x) \right\} \right] w(x)dx,$$

$$T^{(2)}_{n_1,n_2} = \int \left[\left\{ 2F_1(x) - 2F_2(x) \right\} \left\{ \hat{F}_1(x) - \hat{F}_2(x) \right\} - \left\{ F_1(x) - F_2(x) \right\}^2 \right] w(x)dx.$$

where $\hat{F}_i(x)$ is the empirical cumulative distribution function for the ith sample and $\hat{F}_i(x) = \sum_{j=1}^{n_i} I(X_{ij} \leq x)/n_i$. Under $H_0 : F_1 = F_2$, let F be the common cdf. It is easy to check that $T^{(2)}_{n_1,n_2} = 0$ a.e., $E(T^{(1)}_{n_1,n_2}) = 0$ and

$$Var(T^{(1)}_{n_1,n_2}) = \int \left\{ \frac{4}{n_1 n_2} + \frac{2}{n_1(n_1-1)} + \frac{2}{n_2(n_2-1)} \right\} \left\{ F(x \wedge y) - F(x)F(y) \right\}^2$$

$$w(x)w(y)dxdy$$

$$=: \sigma_1^2.$$

Here $(x \wedge y) = (\min\{x_1, y_1\}, \cdots, \min\{x_p, y_p\})^T \in R^p$, for $x = (x_1, \cdots, x_p)^T$ and $y = (y_1, \cdots, y_p)^T$. Some notations are to be introduced before we continue to present the lemmas. Let $n = n_1 + n_2$, $Y_i = X_{1i}$ for $i = 1, \cdots, n_1$ and $Y_{j+n_1} = X_{2j}$ for $j = 1, \cdots, n_2$,

$$\phi_{ij}(x) = \lambda_{ij}\{I(Y_i < x) - F(x)\}\{I(Y_j < x) - F(x)\}, \text{ where}$$

$$\lambda_{ij} = \begin{cases} \frac{1}{n_1(n_1-1)} & \text{if } i \neq j \in \{1, 2, \cdots, n_1\}, \\ -\frac{1}{n_1 n_2} & \text{if } i \in \{1, 2, \cdots, n_1\} \text{ and } j \in \{n_1+1, \cdots, n_1+n_2\}, \\ \frac{1}{n_2(n_2-1)} & \text{if } i \neq j \in \{n_1+1, \cdots, n_1+n_2\}. \end{cases}$$

Define $V_{nj}(x) = \sum_{i=1}^{j-1} \int \phi_{ij}(x)w(x)dx$ for $j = 2, 3, \cdots, n$, $S_{nm}(x) = \sum_{j=2}^{m} V_{nj}(x)$ for $2 \leq m \leq n$ and $\mathcal{F}_{nm} = \sigma\{Y_1, Y_2, \cdots, Y_m\}$ which is the σ-algebra generated by

$\{Y_1, \cdots, Y_m\}$. We can therefore rewrite $T^{(1)}_{n_1,n_2}$ as

$$T^{(1)}_{n_1,n_2} = 2 \sum_{j=2}^{n_1+n_2} V_{nj}(x).$$

Lemma 4.1 *For each n, $\{S_{nm}, \mathcal{F}_{nm}\}$ is a sequence of zero mean, square integrable martingale.*

Lemma 4.1 can be proved by following the similar proof of Lemma 2.1.

The asymptotic normality of $T^{(1)}_{n_1,n_2}$ is to be established via the martingale central limit theorem under three assumptions:

(A1) $\lim_{\min\{n_1,n_2\}\to\infty} n/n_k = \rho_k^2 \in (1, +\infty)$ for $k = 1$ and 2.

(A2) $\Phi_2/n^5 = o(\sigma_1^4)$, where

$$
\begin{aligned}
\Phi_2 &= \int E\Big[\{I(Y_i < x) - F(x)\}\{I(Y_i < y) - F(y)\}\{I(Y_i < u) - F(u)\} \\
&\quad \times \{I(Y_i < v) - F(v)\}\Big]\{F(x \wedge y) - F(x)F(y)\}\{F(u \wedge v) - F(u)F(v)\} \\
&\quad w(x)w(y)w(u)w(v)dxdydudv.
\end{aligned}
$$

(A3) $\Phi_3/n^6 = o(\sigma_1^4)$, where

$$
\begin{aligned}
\Phi_3 &= \int E^2\Big[\{I(Y_i < x) - F(x)\}\{I(Y_i < y) - F(y)\}\{I(Y_i < u) - F(u)\} \\
&\quad \times \{I(Y_i < v) - F(v)\}\Big]w(x)w(y)w(u)w(v)dxdydudv.
\end{aligned}
$$

Lemma 4.2 *Under (A1) and (A2), then*

$$\frac{\sum_{j=2}^{n_1+n_2} E[V_{nj}^2|\mathcal{F}_{n,j-1}]}{\sigma_1^2} \xrightarrow{P} \frac{1}{4}, \quad \text{when } p \to \infty \text{ as } \min\{n_1, n_2\} \to \infty.$$

Lemma 4.3 *Under (A1), (A2) and (A3), then*

$$\sum_{j=2}^{n_1+n_2} \sigma_1^{-2} E\{V_{nj}^2 I(|V_{nj}| > \epsilon\sigma_1)|\mathcal{F}_{nj-1}\} \xrightarrow{p} 0, \quad \text{when } p \to \infty \text{ as } \min\{n_1, n_2\} \to \infty.$$

The proofs of Lemmas 4.2 and 4.3 are available in Section 4.4. The Lemmas 4.1, 4.2 and 4.3 verify the sufficient conditions required by the martingale central limit theorem. The assumption *(A1)* ensures that sample size n_1 and n_2 increase to $+\infty$ proportionally. Both *(A2)* and *(A3)* are apparently satisfied if p is fixed. When $p \to +\infty$, *(A2)* and *(A3)* address the relationships between p and n, presented via an integral consisting of F and w. To elaborate on these assumptions, consider the the case where F consists of independent marginals and w is its corresponding probability density function (pdf). For simplicity of notation, denote $w(x)w(y)w(u)w(v)dxdydudv$ as $w(\cdot)d\cdot$. Then

$$
\begin{aligned}
\frac{\Phi_2/n^5}{\sigma_1^4} &\leq \eta(n)\frac{\int\left\{F(x \wedge y) - F(x)F(y)\right\}^2\left\{F(u \wedge v) - F(u)F(v)\right\}w(\cdot)d\cdot}{\int\left\{F(x \wedge y) - F(x)F(y)\right\}^2\left\{F(u \wedge v) - F(u)F(v)\right\}^2 w(\cdot)d\cdot} \\
&= \eta(n)\frac{\int\left\{F(u \wedge v) - F(u)F(v)\right\}w(u)w(v)dudv}{\int\left\{F(u \wedge v) - F(u)F(v)\right\}^2 w(u)w(v)dudv} \\
&= \eta(n)\frac{(1/3^p - 1/4^p)}{1/6^p + 1/9^p - 2(2/15)^p} =: \eta(n)R_p.
\end{aligned}
\tag{4.1}
$$

where

$$
\eta(n) = \frac{1/n^5}{\left\{\frac{4}{n_1 n_2} + \frac{2}{n_1(n_1-1)} + \frac{2}{n_2(n_2-1)}\right\}^2} = O(1/n).
$$

We notice that

$$
\begin{aligned}
\int F(u \wedge v)w(u)w(v)dudv &= \left\{\int\int F_i(u_i \wedge v_i)w_i(u_i)w_i(v_i)du_idv_i\right\}^p \\
&= \left\{2\int_{F_i(u_i)=0}^{1}\int_{F_i(v_i)=0}^{F_i(u_i)} F_i(v_i)dF_i(u_i)dF_i(v_i)\right\}^p \\
&= (1/3)^p
\end{aligned}
$$

and

$$\int F(u)F(v)w(u)w(v)dudv = \left\{ \int\int F_i(u_i)F_i(v_i)w_i(u_i)w_i(v_i)du_idv_i \right\}^p$$

$$= \left\{ \int_{F_i(u_i)=0}^{1} F_i(u_i)dF_i(u_i) \right\}^{2p}$$

$$= (1/4)^p.$$

We therefore have $(1/3^p - 1/4^p)$ as the numerator in (4.1) and the denominator can be derived by following the same procedure. The arguments for *(A3)* are similar. Note that $R_{10} \approx 1204$ and R_p is monotonically increasing in p. Namely, *(A2)* ensures that n grows much faster than p. Notably, one important issue is how stringent the relationship between p and n should be in order to validate the asymptotic normality. The analytical solution to this problem is closely related to the F and w functions.

Theorem 4.1 *Under H_0 along with (A1), (A2) and (A3),*

$$\frac{T_{n_1,n_2}}{\sqrt{\sigma_1^2}} \xrightarrow{d} N(0,1), \ \ when \ p \to \infty \ as \ \min\{n_1,n_2\} \to \infty.$$

Proof: Under *(A1)* , *(A2)* and *(A3)*, by combining Lemmas 4.1, 4.2 and 4.3 and applying the martingale central limit theorem in Chapter 1, we finish the proof.

In order to carry out the test, a ratio consistent estimator for σ_1^2 needs to be constructed. In practice, we suggest using a bootstrap method. The bootstrap estimator $\widehat{\sigma}_1^2$ is described as follows.

Given an integer B and the statistic T_{n_1,n_2}, compute versions T_{n_1,n_2}^{1*}, $T_{n_1,n_2}^{2*} \cdots$, T_{n_1,n_2}^{B*} of T_{n_1,n_2} by randomly resampling n_1 observations, with replacement, from $\{X_{1i}\}_{i=1}^{n_1}$ and assigning them to sample 1, and randomly resampling another n_2 observations with replacement, from $\{X_{1i}\}_{i=1}^{n_1}$ and assigning them to sample 2; repeat this procedure independently for each calculation of T_{n_1,n_2}^*. One bootstrap estimator for σ_1^2 is

$$\widehat{\sigma}_1^{2(1)} = Var(T_{n_1,n_2}^*).$$

By repeating the above procedure for $\{X_{2j}\}_{j=1}^{n_2}$, we get the second bootstrap estimator $\widehat{\sigma}_1^{2(2)}$. Later, in the simulation studies, we employ both variance estimators along with a pooled bootstrap estimator $\widehat{\sigma}_1^2$.

$$\widehat{\sigma}_1^2 = \frac{n_1 \widehat{\sigma}_1^{2(1)} + n_2 \widehat{\sigma}_1^{2(2)}}{n_1 + n_2}.$$

The simulation results show that using the pooled variance estimator always lowers the empirical power.

Remark: From the definition of T_{n_1,n_2}, it is expected that w function will affect the power of the test. More precisely, the power will be significantly increased if w puts heavier weights at regions where F_1 and F_2 are further apart than at regions where they overlap and vice versa.

We establish the asymptotic normality to T_{n_1,n_2} under H_1 by applying two-sample U-statistic theory. Some basic definitions and theorems of two-sample U-statistics are presented next.

4.2.2 Two-sample U-statistic

Let $X_{i1}, X_{i2}, \cdots, X_{in_i} \in R^p$ be independent observations on a distribution F_i for $i = 1, 2$. Consider a parameter $\theta = \theta(F_1, F_2)$ for which there is an unbiased estimator and $\theta(F_1, F_2)$ can be written as

$$\theta(F_1, F_2) = Eh(X_{1i_1}, \cdots, X_{1i_{m_1}}; X_{2j_1}, \cdots, X_{2j_{m_2}})$$

for some symmetric function h, such that h is invariant to any permutation of its arguments within each sample.

Definition 4.1 *A two-sample U-statistic with m_1 and m_2 arguments for the first and second sample is*

$$U_{n_1,n_2} = \left[\binom{n_1}{m_1}\binom{n_2}{m_2}\right]^{-1} \sum_{l=1}^{2} \sum_{(i_1,i_2,\cdots,i_{m_l})\in C_{m_l}^{n_l}} h(X_{1i_1},\cdots,X_{1i_{m_1}};X_{2j_1},\cdots,X_{2j_{m_2}})$$

(4.2)

where h is a symmetric kernel with respect to arguments within each sample and its degrees are (m_1,m_2). The sum $\sum_{(i_1,i_2,\cdots,i_{m_l})\in C_{m_l}^{n_l}}$ is taken over all subsets $1 \le i_1 < i_2 \cdots < i_{ml} \le n_l$ of $\{1,\cdots,n_l\}$.

Definition 4.1 reveals that a two-sample U-statistic is a sum of identically distributed random variables and the summands are not independent except in the case where $m_1 = m_2 = 1$. Thus, the classic central limit theorem is not applicable if $m_i \ge 2$ for $i = 1$ or 2.

The following theorem (Lee, 1990) provides a variance decomposition for two-sample U-statistics.

Theorem 4.2 *Let U_{n_1,n_2} be a two-sample U-statistic based on a kennel function h of degrees (m_1,m_2). Then*

$$Var(U_{n_1,n_2}) = \sum_{c=0}^{m_1}\sum_{d=0}^{m_2} \frac{\binom{m_1}{c}\binom{m_2}{d}\binom{n_1-m_1}{m_1-c}\binom{n_2-m_2}{m_2-d}}{\binom{n_1}{m_1}\binom{n_2}{m_2}}\xi_{c,d}^2,$$

(4.3)

where $\xi_{c,d}^2 = Var\left\{h_{c,d}(X_{1i_1},\cdots,X_{1i_c};X_{2j_1},\cdots,X_{2j_d})\right\}$ and

$$h_{c,d}(x_{11},\cdots,x_{1c};x_{21},\cdots,x_{2d}) = E\Big\{h(x_{11},\cdots,x_{1c},X_{1c+1},\cdots,X_{1m_1};$$

$$x_{21},\cdots,x_{2d},X_{2d+1},\cdots,X_{2m_2})|X_{11}=x_{11},\cdots,X_{1c}=x_{1c};X_{21}=x_{21},\cdots,X_{2d}=x_{2d}\Big\}.$$

Theorem 4.2 holds for both fixed and growing dimensions. In the case when p is fixed, it is expected that $Var(U_{n_1,n_2})$ is dominated by both summands for $c = 1, d = 0$ and $c = 0, d = 1$.

Whereas the classic central limit theorem does not apply to U_{n_1,n_2} when $m_i \geq 2$ for $i = 1$ or 2, it does apply to the first order projection of U_{n_1,n_2}. Further, its first order projection and the U-statistic itself share the same limiting distribution if the difference between them is negligible.

Definition 4.2 *The first order projection of the two-sample U-statistic is defined as*

$$\hat{U}_{n_1,n_2} = \sum_{i=1}^{2} \sum_{j=1}^{n_i} E(U_{n_1,n_2}|X_{ij}) - (n_1 + n_2 - 1)\theta, \tag{4.4}$$

where $E(U_{n_1,n_2}|X_{ij})$ is the conditional expectation of U_{n_1,n_2} given X_{ij}.

Thus $\hat{U}_{n_1,n_2}$ can be written as a sum of independent random variables $\widetilde{h}_1^{(i)}(X_{ij})$, that is

$$\hat{U}_{n_1,n_2} - \theta = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \widetilde{h}_1^{(i)}(X_{ij}),$$

where

$$\widetilde{h}_1^{(i)}(x_{i1}) = E\Big\{h(X_{11}, \cdots, X_{1m_1}; X_{21}, \cdots, X_{2m_2})|X_{i1} = x_{i1}\Big\} - \theta.$$

Notably, the difference $U_{n_1,n_2} - \hat{U}_{n_1,n_2}$ itself is also a two-sample U-statistic with zero mean and can be expressed as

$$U_{n_1,n_2} - \hat{U}_{n_1,n_2} = \left[\binom{n_1}{m_1}\binom{n_2}{m_2}\right]^{-1} \sum_{l=1}^{2} \sum_{(i_1,i_2,\cdots,i_{m_l})\in C_{m_l}^{n_l}} \phi(X_{1i_1}, \cdots, X_{1i_{m_1}}; X_{2j_1}, \cdots, X_{2j_{m_2}})$$

$$\tag{4.5}$$

where

$$\phi(X_{1i_1}, \cdots, X_{1i_{m_1}}; X_{2j_1}, \cdots, X_{2j_{m_2}}) = h(X_{1i_1}, \cdots, X_{1i_{m_1}}; X_{2j_1}, \cdots, X_{2j_{m_2}})$$

$$- \sum_{i=1}^{2} \sum_{j=1}^{n_i} \widetilde{h}_1^{(i)}(X_{ij}) - \theta.$$

4.2.3 Limiting distribution of T_{n_1,n_2} under the alternative H_1

Based on Definition 4.1, it can be argued that T_{n_1,n_2} is indeed a two-sample U-statistic with $m_1 = m_2 = 2$ and the kernel function

$$
\begin{aligned}
h(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2}) &= \int \Big[I(X_{1i_1} \leq x)I(X_{1i_2} \leq x) + I(X_{2j_1} \leq x)I(X_{2j_2} \leq x) \\
&\quad - \frac{1}{2}I(X_{1i_1} \leq x)I(X_{2j_1} \leq x) - \frac{1}{2}I(X_{1i_1} \leq x)I(X_{2j_2} \leq x) \\
&\quad - \frac{1}{2}I(X_{1i_2} \leq x)I(X_{2j_1} \leq x) - \frac{1}{2}I(X_{1i_2} \leq x)I(X_{2j_2} \leq x) \Big] \\
&\quad w(x)dx.
\end{aligned}
$$

We now note that $h(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2})$ is symmetric within each sample. In particular, $h(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2})$, $h(X_{1i_2}, X_{1i_1}; X_{2j_1}, X_{2j_2})$, $h(X_{1i_1}, X_{1i_2}; X_{2j_2}, X_{2j_1})$ and $h(X_{1i_2}, X_{1i_1}; X_{2j_2}, X_{2j_1})$ are all equal. Let $\theta = \int \left\{ F_1(x) - F_2(x) \right\}^2 w(x)dx$, then it is shown in Section 4.4 that T_{n_1,n_2} is an unbiased estimator for θ.

From (4.4), the projection of T_{n_1,n_2} can be expressed as

$$
\hat{T}_{n_1,n_2} - \theta = \frac{2}{n_1} \sum_{i=1}^{n_1} \tilde{h}_1^{(1)}(X_{1i}) + \frac{2}{n_2} \sum_{j=1}^{n_2} \tilde{h}_1^{(2)}(X_{2j}),
$$

where

$$
\begin{aligned}
\tilde{h}_1^{(1)}(X_{1i}) &= \int \Big[I(X_{1i} \leq x)F_1(x) + F_2^2(x) - I(X_{11} \leq x)F_2(x) - F_1(x)F_2(x) \Big] w(x)dx - \theta \\
&= \int \Big[\{I(X_{1i} \leq x) - F_2(x)\}\{F_1(x) - F_2(x)\} \Big] w(x)dx - \theta, \\
\tilde{h}_1^{(2)}(X_{2j}) &= \int \Big[\{F_1(x) - I(X_{2j} \leq x)\}\{F_1(x) - F_2(x)\} \Big] w(x)dx - \theta.
\end{aligned}
$$

We proceed to $Var(\tilde{h}_1^{(i)}(X_{ij}))$, for $i = 1$ and 2. It is easy to check that

$$
\begin{aligned}
Var(\tilde{h}_1^{(1)}(X_{1i})) &= \int \Big[\{F_1(x) - F_2(x)\}\{F_1(y) - F_2(y)\}\{F_1(x \wedge y) - F_1(x)F_1(y)\} \Big] \\
&\quad w(x)w(y)dxdy, \\
Var(\tilde{h}_1^{(2)}(X_{2j})) &= \int \Big[\{F_1(x) - F_2(x)\}\{F_1(y) - F_2(y)\}\{F_2(x \wedge y) - F_2(x)F_2(y)\} \Big] \\
&\quad w(x)w(y)dxdy.
\end{aligned}
$$

Under the alternative $H_1 : F_1 \neq F_2$, both $Var\{\tilde{h}_1^{(1)}(X_{11})\}$ and $Var\{\tilde{h}_1^{(2)}(X_{21})\}$ result positive. Then the following theorem presents the limiting distribution of $\hat{T}_{n_1,n_2}$ under H_1.

Theorem 4.3 *Let*

$$\delta_1^2 = 4\rho_1^2 Var\{\tilde{h}_1^{(1)}(X_{1i})\} + 4\rho_2^2 Var\{\tilde{h}_1^{(2)}(X_{2j})\}.$$

Assume that $\lim_{n_1,n_2 \to \infty} \frac{n}{n_k} = \rho_k^2 < \infty$ *for* $k = 1, 2$ *and*

$$\lim_{p \to \infty} \frac{Var(\tilde{h}_1^{(1)}(X_{1i}))}{Var(\tilde{h}_1^{(2)}(X_{2j}))} = \gamma \in (0, \infty). \tag{4.6}$$

Then,

$$\sqrt{n}(\hat{T}_{n_1,n_2} - \theta)/\delta_1 \xrightarrow{d} N(0,1) \text{ when } p \to \infty \text{ as } \min\{n_1, n_2\} \to \infty. \tag{4.7}$$

Proof:

We partition $\sqrt{n}(\hat{T}_{n_1,n_2} - \theta)$ into two independent sums of i.i.d. random variables as

$$
\begin{aligned}
\sqrt{n}(\hat{T}_{n_1,n_2} - \theta) &= 2\frac{\sqrt{n}}{\sqrt{n_1}}\frac{\sqrt{n_1}}{n_1}\sum_{i=1}^{n_1}\tilde{h}_1^{(1)}(X_{1i}) + 2\frac{\sqrt{n}}{\sqrt{n_2}}\frac{\sqrt{n_2}}{n_2}\sum_{j=1}^{n_2}\tilde{h}_1^{(2)}(X_{2j}) \\
&=: S_{n1} + S_{n2}, \text{ say.}
\end{aligned}
$$

By applying the classic central limit theorem,

$$\frac{S_{nk}}{\sqrt{4\rho_k^2 Var\{\tilde{h}_1^{(k)}(X_{ki})\}}} \xrightarrow{d} N(0,1), \text{ for } k = 1, 2.$$

As a result,

$$\sqrt{n}(\hat{T}_{n_1,n_2} - \theta)/\delta_1 = \frac{S_{n1}}{\sqrt{4\rho_1^2 Var\{\tilde{h}_1^{(1)}(X_{1i})\}}}\sqrt{\frac{\gamma}{1+\gamma}} + \frac{S_{n2}}{\sqrt{4\rho_2^2 Var\{\tilde{h}_1^{(2)}(X_{2i})\}}}\sqrt{\frac{1}{1+\gamma}}.$$

Then (4.6) follows. This finishes the proof.

We note that when p is fixed, the assumption (4.5) becomes apparently true, when $p \to +\infty$, it requires $\{F_1(x \wedge y) - F_1(x)F_1(y)\}$ and $\{F_2(x \wedge y) - F_2(x)F_2(y)\}$ to be comparable such that $Var\{\tilde{h}_1^{(1)}(X_{1i})\}$ and $Var\{\tilde{h}_1^{(2)}(X_{2j})\}$ are of the same magnitude.

It is expected that T_{n_1,n_2} and $\hat{T}_{n_1,n_2}$ share the same limiting distribution if $e_n =: T_{n_1,n_2} - \hat{T}_{n_1,n_2}$ is negligible. More precisely, based on (4.6) and the Slutsky Theorem we can conclude that

$$\sqrt{n}(T_{n_1,n_2} - \theta)/\delta_1 \xrightarrow{d} N(0,1)$$

if $\sqrt{n}e_n/\delta_1 = o_p(1)$. For simplicity of notation, let us define

$$\Delta_i = \int \frac{2}{n_i(n_i-1)}\Big\{F_i(x \wedge y) - F_i(x)F_i(y)\Big\}^2 w(x)w(y)dxdy, \text{ for } i = 1,2.$$

$$\Delta_{12} = \int \frac{4}{n_1 n_2}\Big\{F_1(x \wedge y) - F_1(x)F_1(y)\Big\}\Big\{F_2(x \wedge y) - F_2(x)F_2(y)\Big\}w(x)w(y)dxdy.$$

(A4) Assume for $i, j = 1$ and 2,

$$\int \Big\{F_i(x \wedge y) - F_i(x)F_i(y)\Big\}^2 w(x)w(y)dxdy = o\Big[nVar\{h_1^{(j)}(X_{j1})\}\Big].$$

(A5) Assume Δ_1, Δ_2 and Δ_{12} are of the same order.

Theorem 4.4 *Under (A4) and (A5),*

$$\sqrt{n}e_n/\delta_1 = o_p(1).$$

The proof of Theorem 4.4 is available in Section 4.4. The assumption *(A5)* is once again ensures $\{F_1(x \wedge y) - F_1(x)F_1(y)\}$ and $\{F_2(x \wedge y) - F_2(x)F_2(y)\}$ are comparable as in (4.5). Moreover, *(A4)* further restricts the relationship between p and n. We do not offer analytical solutions to this restriction which is determined by cdfs F_1, F_2 and w, but leave it as a future research problem.

To obtain the limiting distribution of T_{n_1,n_2} under H_1, we combine the Slutsky Theorem, Theorem 4.3 and 4.4. Thus

$$\sqrt{n}(T_{n_1,n_2} - \theta)/\delta_1 \xrightarrow{d} N(0,1) \text{ when } p \to \infty \text{ as } \min\{n_1, n_2\} \to \infty.$$

Under H_0, we find $\delta_1^2 = 0$. Applying the classic central limit theorem to the first order projection $\hat{T}_{n_1,n_2}$ is not suitable any more. As in Chapter 2, we attempt, in Section 4.2.1, to establish the asymptotic normality by using the martingale central limit theorem.

Remark: One possibility of pursuing the limiting distribution for T_{n_1,n_2} under H_0 is to consider T_{n_1,n_2} again as a two-sample U-statistic. We find the following theorem (Koroljuk and Borovskich, 1989) for univariate case might be possibly extended to high-dimensional settings. Yet the way of extending it to where $p \to \infty$ is outside the scope of existing two-sample U-statistic theory, it needs further investigation.

Theorem 4.5 *If $\xi_{1,0}^2 = \xi_{0,1}^2 = 0$, $\xi_{2,0}^2 \neq 0$, $\xi_{0,2}^2 \neq 0$ and $\xi_{1,1}^2 \neq 0$. The two-sample U-statistic U_{n_1,n_2} is asymptotically distributed as the sum*

$$n_1^{-1} z_{20} + n_1^{-1/2} n_2^{-1/2} z_{11} + n_2^{-1} z_{02}, \tag{4.8}$$

where $z_{20} = \sum_{i=1}^{\infty} \lambda_i(\tau_i^2 - 1)$, $z_{02} = \sum_{i=1}^{\infty} \lambda_i(\zeta_i^2 - 1)$ and $z_{11} = \sum_{i=1}^{\infty} \kappa_i \tau_i \zeta_i$, τ_i and ζ_i are all independent standard normal random variables, and λ_i and κ_i are eigenvalues of $h_{2,0}(\cdot, \cdot)(h_{0,2}(\cdot, \cdot))$ and $h_{1,1}(\cdot, \cdot)$ respectively.

Theorem 4.5 indicates that $n_1^{1/2} n_2^{1/2} U_{n_1,n_2} \xrightarrow{d} \sqrt{\rho_1^2 - 1} z_{20} + z_{11} + \sqrt{\rho_2^2 - 1} z_{02}$. When $p \to +\infty$, the orthonormal eigenfunctions and eigenvalues in connection with the symmetric kernel function h may hinge on p. Upon assuming their existence, the eigenvalues λ_i and κ_i are still hard to find. In practice, we recommend a bootstrap method to estimate the limiting distribution given in (4.8). We have described two schemes for establishing asymptotic normality for the test statistic. Our concerns, after all, emerge in related assumptions made towards the relationship between p and n, and the orthonormal decomposition for symmetric kernel function h when $p \to +\infty$.

4.3 Numerical Results

We report numerical results from three simulation studies in which we compared the proposed distribution test with the MRPP test and the two-sample mean test proposed in Chapter 2. For the MRPP test, we set the number of permutations to be 1,000. Our target p-variate observations $\{X_{ij}\}_{j=1}^{n_i}$ were generated from the following multivariate model,

$$X_{ij} = \Gamma_i U_{ij} + \mu_i, \text{ for } i = 1, 2, j = 1, 2, \cdots, n_i \tag{4.9}$$

where $U_{ij} \in R^{p+1}$. Every standardized U_{ijk} (with zero mean and unit variance), for $k = 1, \cdots, p+1$, was generated independently from a candidate distribution C_i. Two candidate distributions we considered were $N(0,1)$ and $\chi^2(6)$.

To better examine the performance of the proposed test, we considered three types of alternatives which are listed in Table 4.1 as Case 1, Case 2 and Case 3. Note that Case 4 leads to the size of the test. We chose the significance level $\alpha = .05$.

To calculate $T_{n_1,n_2}/\sqrt{\sigma_1^2}$, we set w to be the uniform distribution on $[-3,3]^p$. The Monte Carlo approximation of T_{n_1,n_2} was calculated based on 10,000 p-variate data points independently sampled from w and additional 1,000 p-variate data points $\{Z_i\}_{i=1}^{1000}$ where $Z_i = (Z_{i1}, \cdots, Z_{ip})^T$. In particular, for the kth dimension, $k = 1, 2, \cdots, p$, let $(Z_{1k} < Z_{2k} < \cdots < Z_{1000,k})$ equally partition the interval $[-3,3]$. Three types of alternatives and a null hypothesis were constructed as follows.

For case 1, let $\mu_1 = \mu_2 = (0, \cdots, 0)^T \in R^p$, $\Gamma_1 = \Gamma_2 = \Gamma(.5)$ where

$$\Gamma(\rho) = \begin{pmatrix} 1 & \rho & 0 & 0 & 0 & 0 & \cdots & \cdots \\ 0 & 1 & \rho & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 0 & 1 & \rho \end{pmatrix}_{p \times (p+1)}$$

Table 4.1   Various alternatives and the null

	$\mu_1 = \mu_2$	$\Gamma_1 = \Gamma_2$	$C_1 = C_2$
Case 1:	true	true	*false*
Case 2:	*false*	true	true
Case 3:	true	*false*	true
Case 4:	true	true	true

and chose $C_1 = N(0,1)$ and $C_2 = \chi^2(6)$. Therefore, both $\{X_{1i}\}_{i=1}^{n_1}$ and $\{X_{2j}\}_{j=1}^{n_2}$ generated from the multivariate model (4.9) had p-variate zero mean vectors and same covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$ where $\sigma_{ii} = 1 + .5^2$, $\sigma_{ij} = .5$ if $|i - j| = 1$ and $\sigma_{ij} = 0$ if $|i - j| > 1$. For case 2, let $\Gamma_1 = \Gamma_2 = \Gamma(.5)$, $C_1 = C_2$, $\mu_1 = (0, \cdots, 0)^T \in R^p$ and $\mu_2 = (\delta_\mu, \cdots, \delta_\mu)^T \in R^p$ where δ_μ took positive values. The difference between F_1 and F_2 was apparently caused by different mean vectors, μ_1 and μ_2. For case 3, let $\mu_1 = \mu_2 = (0, \cdots, 0)^T \in R^p$, $C_1 = C_2$, $\Gamma_1 = \Gamma(\rho)$ and $\Gamma_2 = \Gamma(-\rho)$. As a result, the covariance matrix of $\{X_{1i}\}_{i=1}^{n_1}$ and the covariance matrix of $\{X_{2j}\}_{j=1}^{n_2}$ differed in their off-diagonal elements. Finally, for case 4, let $\mu_1 = \mu_2 = (0, \cdots, 0)^T \in R^p$, $\Gamma_1 = \Gamma_2 = \Gamma(.5)$ and $C_1 = C_2$. As a result, we had $F_1 = F_2$.

Throughout our simulation studies, we set both n_1 and n_2 to be equal. Sample size n_1 (or n_2) was 25 and 50 and dimension p was 10, 20 and 50. We carried out 500 simulations for each combination of sample size and dimension in Case 1, 2, 3 and 4.

A permutation test using the proposed T_{n_1, n_2}

In the first simulation study, a permutation test based on the proposed T_{n_1, n_2} was undertaken. Consider the pooled sample

$$\{Z_1, Z_2, \cdots, Z_{n_1 + n_2}\} = \{X_{11}, \cdots, X_{1n_1}\} \bigcup \{X_{21}, \cdots, X_{2n_2}\}.$$

Under the null hypothesis, $Z_1, \cdots, Z_{n_1+n_2}$ are i.i.d. with common cdf F. Given an integer M and T_{n_1,n_2}, compute versions $T_{n_1,n_2}^{1*}, T_{n_1,n_2}^{2*}, \cdots, T_{n_1,n_2}^{M*}$ of T_{n_1,n_2} by randomly resampling n_1 observations without replacement from $\{Z_1, Z_2, \cdots, Z_{n_1+n_2}\}$ and assigning them to sample 1, and including all the remaining observations into sample 2; repeat this procedure independently for each calculation of T_{n_1,n_2}^*. The permutation p-value is

$$\frac{1}{M} \sum_{i=1}^{M} I\left(T_{n_1,n_2}^{i*} \geq T_{n_1,n_2}^{(obs)}\right)$$

where $I(\cdot)$ is the indicator function and $T_{n_1,n_2}^{(obs)}$ is the value of T_{n_1,n_2} for original two samples $\{X_{11}, \cdots, X_{1n_1}\}$ and $\{X_{21}, \cdots, X_{2n_2}\}$. We chose $M = 1000$ in our simulation.

For Case 2, we set $\delta_\mu = .1$ and $.05$, $C_1 = C_2 = N(0,1)$ and $\chi^2(6)$. For Case 3, we set $C_1 = C_2 = N(0,1)$ and $\chi^2(6)$, set $\rho = .5$ and 1. For Case 4, we set $C_1 = C_2 = N(0,1)$ and $\chi^2(6)$. The empirical sizes of the three tests shown in Table 4.2, are all close to the significance level $\alpha = .05$. Table 4.3 provides the empirical power in Case 1 where the two candidate distributions $(C_1 \neq C_2)$ are different. The empirical power of the distribution test increased as sample size or dimension p increased, and reached $.955$ when $n_1 = n_2 = 50$, $p = 50$. However, the power of the two-sample mean and MRPP tests remained around the significance level. For Case 2, we provided Table 4.4 and Table 4.5 for $\delta_\mu = .05$ and $\delta_\mu = .1$, respectively. From these two tables, we found that the two-sample mean test outperformed the other two tests and the empirical power of MRPP and the distribution tests were very close when $p = 10$ and 20 while the latter fell behind when $p = 50$. Overall, empirical power increased when δ_μ, sample size or dimension became larger. For Case 3 in which F_1 and F_2 differed only due to their covariance matrices, Table 4.6 presents the empirical power when $\rho = .5$ and Table 4.7 shows results for $\rho = 1$. The power of the distribution test and MRPP test both increased when sample size increased, but powers decreased as

dimension p increased. However, the two-sample mean test was only able to maintain the significance level. As the simulation results showed that the two-sample mean test should not be used to detect any differences between F_1 and F_2 caused by either different candidate distributions or different covariance matrices. The MRPP test and the distribution test were able to detect different covariance structures in the case where dimension p was not too high. When the marginal distributions differed in their higher (than second) order moments, the distribution test performed the best.

A simulation study based on asymptotic normality

In the previous simulation study, for Case 2, Case 3 and Case 4 three tests performed very similarly for $C_1 = C_2 = N(0,1)$ and $C_1 = C_2 = \chi^2(6)$. We therefore only considered $N(0,1)$ here to implement the distribution test based on its asymptotic normality by adopting three variance estimators described earlier. When calculating the bootstrap estimators for σ_1^2, we chose B to be 500. For Case 2, set $\delta_\mu = .1$ and for Case 3, set $\rho = 1$. The simulation results shown in Table 4.8 are the empirical power and size for the distribution test. Dist(BTV-S1) represents the distribution test using the bootstrap variance estimator (BTV) from the first sample (S1). Dist(BTV-S2) used BTV from the second sample (S2) and Dist (BTV-pooled) used the pooled BTV. Overall, the empirical powers of distribution tests using the individual bootstrap variance estimators were consistently better than using the pooled one but the sizes of the tests were not as well maintained with the individual estimators.

A simulation study based on bootstrapping T_{n_1,n_2} from only one sample

We also conducted a simulation study based on bootstrapping T_{n_1,n_2} from only one sample. The procedure can be described as follows.

Consider either of two samples, say the first sample $\{X_{11}, \cdots, X_{1n_1}\}$. Given an integer M and T_{n_1,n_2}, compute versions $T^{1*}_{n_1,n_2}, T^{2*}_{n_1,n_2}, \cdots, T^{M*}_{n_1,n_2}$ of T_{n_1,n_2} by randomly resampling n_1 observations with replacement from $\{X_{11}, \cdots, X_{1n_1}\}$ and assigning them to sample 1, and randomly resampling another n_2 observations with replacement from $\{X_{11}, \cdots, X_{1n_1}\}$ and assigning them to sample 2; repeat this procedure independently for each calculation of $T^*_{n_1,n_2}$. The permutation p-value based on bootstrapping T_{n_1,n_2} from the first sample is

$$\frac{1}{M} \sum_{i=1}^{M} I\left(T^{i*}_{n_1,n_2} \geq T^{(obs)}_{n_1,n_2}\right).$$

By repeating the above procedure for $\{X_{21}, \cdots, X_{2n_2}\}$, we could have another p-value. Both sets of p-values were provided in Table 4.9. The permutation test based on T_{n_1,n_2} in the first simulation study can be viewed as bootstrapping T_{n_1,n_2} from pooled sample without replacement. Comparing with the permutation test and the distribution tests using bootstrap variance estimators, the distribution test based on bootstrapping T_{n_1,n_2} from only one sample lost a little power but successfully controlled the size around the significance level.

In summary, we found that the empirical powers of the distribution tests were rather close under three testing schemes: permutation based, asymptotic normality based and bootstrapping T_{n_1,n_2} from only one sample based distribution tests.

4.4   Technical Proofs

Derivation of $Var(T^{(1)}_{n_1,n_2})$:

We are to work out $Var(T_{n1})$ as following.

$$
\begin{aligned}
Var(T^{(1)}_{n_1,n_2}) &= E(T^{(1)2}_{n_1,n_2}) \\
&= \int E\Bigg[\frac{1}{n_1^2(n_1-1)^2}\sum_{i_1\neq i_2}^{n_1}\Big\{I(X_{1i_1}\leq x)-F(x)\Big\}\Big\{I(X_{1i_2}\leq x)-F(x)\Big\}\times \\
&\quad \sum_{i_3\neq i_4}^{n_1}\Big\{I(X_{1i_3}\leq y)-F(y)\Big\}\Big\{I(X_{1i_4}\leq y)-F(y)\Big\} \\
&\quad +\frac{4}{n_1^2 n_2^2}\sum_{i=1}^{n_1}\Big\{I(X_{1i}\leq x)-F(x)\Big\}\sum_{j=1}^{n_2}\Big\{I(X_{2j}\leq x)-F(x)\Big\}\times \\
&\quad \sum_{l=1}^{n_1}\Big\{I(X_{1l}\leq y)-F(y)\Big\}\sum_{k=1}^{n_2}\Big\{I(X_{2k}\leq y)-F(y)\Big\} \\
&\quad +\frac{1}{n_2^2(n_2-1)^2}\sum_{j_1\neq j_2}^{n_2}\Big\{I(X_{2j_1}\leq x)-F(x)\Big\}\Big\{I(X_{2j_2}\leq x)-F(x)\Big\}\times \\
&\quad \sum_{j_3\neq j_4}^{n_2}\Big\{I(X_{2j_3}\leq y)-F(y)\Big\}\Big\{I(X_{2j_4}\leq y)-F(y)\Big\}\Bigg]w(x)w(y)dxdy \\
&= \int\Big\{\frac{4}{n_1 n_2}+\frac{2}{n_1(n_1-1)}+\frac{2}{n_2(n_2-1)}\Big\}\Big\{F(x\wedge y)-F(x)F(y)\Big\}^2 \\
&\quad w(x)w(y)dxdy.
\end{aligned}
$$

Proof of T_{n_1,n_2} being an unbiased estimator for θ:

As T_{n_1,n_2} is a two-sample U-statistic and

$$
T_{n_1,n_2} = \left[\binom{n_1}{m_1}\binom{n_2}{m_2}\right]^{-1}\sum_{l=1}^{2}\sum_{(i_1,i_2,\cdots,i_{m_l})\in C_{m_l}^{n_l}}h(X_{1i_1},\cdots,X_{1i_{m_1}};X_{2j_1},\cdots,X_{2j_{m_2}}),
$$

where

$$
\begin{aligned}
h(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2}) \;=\; & \int \Big[I(X_{1i_1} \le x)I(X_{1i_2} \le x) + I(X_{2j_1} \le x)I(X_{2j_2} \le x) \\
& - \frac{1}{2}I(X_{1i_1} \le x)I(X_{2j_1} \le x) - \frac{1}{2}I(X_{1i_1} \le x)I(X_{2j_2} \le x) \\
& - \frac{1}{2}I(X_{1i_2} \le x)I(X_{2j_1} \le x) - \frac{1}{2}I(X_{1i_2} \le x)I(X_{2j_2} \le x) \Big] \\
& w(x)dx.
\end{aligned}
$$

Note that T_{n_1,n_2} is a sum of identically distributed random variables and

$$
\left[\binom{n_1}{m_1} \binom{n_2}{m_2} \right]^{-1} \sum_{l=1}^{2} \sum_{(i_1,i_2,\cdots,i_{m_l}) \in C_{m_l}^{n_l}} 1 = 1.
$$

Due to the independence assumption,

$$
\begin{aligned}
E\Big\{ h(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2}) \Big\} \;=\; & \int \Big\{ F_1^2(x) + F_2^2(x) - 4(\frac{1}{2})F_1(x)F_2(x) \Big\} w(x)dx \\
=\; & \int \Big\{ F_1(x) - F_2(x) \Big\}^2 w(x)dx = \theta.
\end{aligned}
$$

Therefore, $E(T_{n_1,n_2}) = \theta$.

Proof of Lemma 4.2:

For simplicity of notation, we denote $\sum_{j=2}^{n_1+n_2} E[V_{nj}^2|\mathcal{F}_{n,j-1}]$ as Q_n. To prove Lemma 4.2, we only need to show $E(Q_n) = \sigma_1^2\{1+o(1)\}/4$ and $Var(Q_n) = o(\sigma_1^4)$. We further partition Q_n into $Q_{n1} + Q_{n2}$.

$$
\begin{aligned}
Q_n \;=\; & \sum_{j=2}^{n_1+n_2} E[V_{nj}^2|\mathcal{F}_{n,j-1}] \\
=\; & \sum_{j=2}^{n_1+n_2} \int \Big(\sum_{i_1,i_2=1}^{j-1} \phi_{i_1 i_2} \Big) \frac{\{F(x \wedge y) - F(x)F(y)\}}{\tilde{n}_j(\tilde{n}_j - 1)} w(x)w(y)dxdy \\
=\; & : Q_{n1} + Q_{n2},
\end{aligned}
$$

where $\tilde{n}_j = n_1$ if $j \in \{1, \cdots, n_1\}$, $\tilde{n}_j = n_2$ if $j \in \{n_1 + 1, \cdots, n_1 + n2\}$ and

$$Q_{n1} = \sum_{j=2}^{n_1+n_2} \int \Big(\sum_{i=1}^{j-1} \phi_{ii} \Big) \frac{\{F(x \wedge y) - F(x)F(y)\}}{\tilde{n}_j(\tilde{n}_j - 1)} w(x)w(y)dxdy,$$

$$Q_{n2} = \sum_{j=2}^{n_1+n_2} \int \Big(\sum_{i_1 \neq i_2}^{j-1} \phi_{i_1,i_2} \Big) \frac{\{F(x \wedge y) - F(x)F(y)\}}{\tilde{n}_j(\tilde{n}_j - 1)} w(x)w(y)dxdy.$$

Note that $E(Q_{n2}) = 0$ and $Cov(Q_{n1}, Q_{n2}) = 0$, to prove the lemma, we only need to show $E(Q_{n1}) = \sigma_1^2 \{1 + o(1)\}/4$, $Var(Q_{n1}) = o(\sigma_1^4)$ and $Var(Q_{n2}) = o(\sigma_1^4)$.

$$
\begin{aligned}
E(Q_{n1}) &= \sum_{j=2}^{n_1+n_2} \int \Big(\sum_{i=1}^{j-1} \phi_{ii} \Big) \frac{\{F(x \wedge y) - F(x)F(y)\}}{\tilde{n}_j(\tilde{n}_j - 1)} w(x)w(y)dxdy \\
&= \Big(\sum_{j=2}^{n_1} + \sum_{j=n_1+1}^{n_1+n_2} \Big) \int \Big(\sum_{i=1}^{j-1} \phi_{ii} \Big) \frac{\{F(x \wedge y) - F(x)F(y)\}}{\tilde{n}_j(\tilde{n}_j - 1)} w(x)w(y)dxdy \\
&= \sum_{j=2}^{n_1} \sum_{i=1}^{j-1} \int \frac{\{F(x \wedge y) - F(x)F(y)\}^2}{n_1^2(n_1 - 1)^2} w(x)w(y)dxdy \\
&\quad + \sum_{j=n_1+1}^{n_1+n_2} \sum_{i=1}^{n_1} \int \frac{\{F(x \wedge y) - F(x)F(y)\}^2}{n_1(n_1 - 1)n_2(n_2 - 1)} w(x)w(y)dxdy \\
&\quad + \sum_{j=n_1+1}^{n_1+n_2} \sum_{i=n_1+1}^{j-1} \int \frac{\{F(x \wedge y) - F(x)F(y)\}^2}{n_2^2(n_2 - 1)^2} w(x)w(y)dxdy \\
&= \sigma_1^2 \{1 + o(1)\}/4.
\end{aligned}
$$

We focus on $Var(Q_{n1})$ only as $Var(Q_{n2}) = o(\sigma_1^4)$ can be proved similarly.

$$
\begin{aligned}
E(Q_{n1}^2) &= E\Big[\sum_{j=2}^{n_1+n_2} \int \sum_{i=1}^{j-1} \frac{\{I(Y_i < x) - F(x)\}\{I(Y_i < y) - F(y)\}}{\tilde{n}_i(\tilde{n}_i - 1)} \times \\
&\quad \frac{\{F(x \wedge y) - F(x)F(y)\}}{\tilde{n}_j(\tilde{n}_j - 1)} w(x)w(y)dxdy \Big]^2 \\
&= \sum_{j_1=2}^{n_1+n_2} \sum_{j_2=2}^{n_1+n_2} \int E\Big[\sum_{i_1=1}^{j_1-1} \sum_{i_2=1}^{j_2-1} \frac{\{I(Y_{i_1} < x) - F(x)\}\{I(Y_{i_1} < y) - F(y)\}}{\tilde{n}_{i_1}(\tilde{n}_{i1} - 1)} \times \\
&\quad \frac{\{F(x \wedge y) - F(x)F(y)\}}{\tilde{n}_{j1}(\tilde{n}_{j1} - 1)} \frac{\{I(Y_{i_2} < u) - F(u)\}\{I(Y_{i_2} < v) - F(v)\}}{\tilde{n}_{i_2}(\tilde{n}_{i2} - 1)} \times \\
&\quad \frac{\{F(u \wedge v) - F(u)F(v)\}}{\tilde{n}_{j2}(\tilde{n}_{j2} - 1)} \Big] w(\cdot)d\cdot = E^2(Q_{n1})\{1 + 0(1)\} + \Phi_2/n^5,
\end{aligned}
$$

where

$$
\begin{aligned}
\Phi_2 \;=\; & \int E\Big[\{I(Y_i < x) - F(x)\}\{I(Y_i < y) - F(y)\}\{I(Y_i < u) - F(u)\} \\
& \times \{I(Y_i < v) - F(v)\}\Big]\{F(x \wedge y) - F(x)F(y)\}\{F(u \wedge v) - F(u)F(v)\} \\
& w(x)w(y)w(u)w(v)dxdydudv.
\end{aligned}
$$

Under *(A1)* and *(A2)*, $Var(Q_{n1}) = o(\sigma_1^4)$. This finishes the proof.

Proof of Lemma 4.3:

Since

$$
\sum_{j=2}^{n_1+n_2} \sigma_1^{-2} E\{V_{nj}^2 I(|V_{nj}| > \epsilon\sigma_1)|F_{nj-1}\} \leq \sigma_1^{-q}\epsilon^{2-q} \sum_{j=1}^{n_1+n_2} E(V_{nj}^q|F_{nj-1}) \text{ for some } q > 2.
$$

We choose $q = 4$. Then the conclusion of the lemma is true if we can show

$$
E\Big\{ \sum_{j=2}^{n_1+n_2} E(V_{nj}^4|F_{nj-1})\Big\} = \sum_{j=2}^{n_1+n_2} E(V_{nj}^4) = o(\sigma_1^4).
$$

Note that

$$
\begin{aligned}
\sum_{j=2}^{n_1+n_2} E(V_{nj}^4) \;=\; & \sum_{j=2}^{n_1+n_2} E\Big\{ \sum_{i=1}^{j-1} \int \phi_{ij}(x)w_i(x)dx\Big\}^4 \\
=\; & \sum_{j=2}^{n_1+n_2} E\Big\{ \sum_{i_1,i_2,i_3,i_4=1}^{j-1} \int \phi_{i_1j}(x)\phi_{i_2j}(y)\phi_{i_3j}(u)\phi_{i_4j}(v)w(\cdot)d \cdot \Big\} \\
=\; & \sum_{j=2}^{n_1+n_2} \sum_{i_1,i_2,i_3,i_4=1}^{j-1} E\Big\{ \int \phi_{i_1j}(x)\phi_{i_2j}(y)\phi_{i_3j}(u)\phi_{i_4j}(v)w(\cdot)d \cdot \Big\}
\end{aligned}
$$

The last term can be decomposed as $3Q + P$ where

$$
Q = \sum_{j=2}^{n_1+n_2} \sum_{i_1 \neq i_2}^{j-1} E\Big\{ \int \phi_{i_1j}(x)\phi_{i_2j}(y)\phi_{i_1j}(u)\phi_{i_2j}(v)w(\cdot)d \cdot \Big\}
$$

and

$$
P = \sum_{j=2}^{n_1+n_2} \sum_{i=1}^{j-1} E\Big\{ \int \phi_{ij}(x)\phi_{ij}(y)\phi_{ij}(u)\phi_{ij}(v)w(\cdot)d \cdot \Big\}
$$

Notice that

$$
\begin{aligned}
P &= \left(\sum_{j=2}^{n_1} + \sum_{j=n_1+1}^{n_1+n_2} \right) \sum_{i=1}^{j-1} E\left\{ \int \phi_{ij}(x)\phi_{ij}(y)\phi_{ij}(u)\phi_{ij}(v)w(\cdot)d\cdot \right\} \\
&= \left(\sum_{j=2}^{n_1} + \sum_{j=n_1+1}^{n_1+n_2} \right) \sum_{i=1}^{j-1} O(n^{-8})\Phi_3 \\
&= O(n^{-6})\Phi_3,
\end{aligned}
$$

where

$$
\begin{aligned}
\Phi_3 &= \int E^2\Big[\{I(Y_i < x) - F(x)\}\{I(Y_i < y) - F(y)\}\{I(Y_i < u) - F(u)\} \\
&\quad \times \{I(Y_i < v) - F(v)\}\Big] w(x)w(y)w(u)w(v)dxdydudv
\end{aligned}
$$

and

$$
\begin{aligned}
Q &= \sum_{j=2}^{n_1+n_2} \sum_{i_1 \neq i_2}^{j-1} E\left\{ \int \phi_{i_1 j}(x)\phi_{i_2 j}(y)\phi_{i_1 j}(u)\phi_{i_2 j}(v)w(\cdot)d\cdot \right\} \\
&= \left(\sum_{j=2}^{n_1} + \sum_{j=n_1+1}^{n_1+n_2} \right) \sum_{i_1 \neq i_2}^{j-1} E\left\{ \int \phi_{i_1 j}(x)\phi_{i_2 j}(y)\phi_{i_1 j}(u)\phi_{i_2 j}(v)w(\cdot)d\cdot \right\} \\
&= \left(\sum_{j=2}^{n_1} + \sum_{j=n_1+1}^{n_1+n_2} \right) \sum_{i_1 \neq i_2}^{j-1} O(n^{-8})\Phi_2 \\
&= O(n^{-5})\Phi_2.
\end{aligned}
$$

Under *(A1)*, *(A2)* and *(A3)*, we have $3Q + P = o(\sigma_1^4)$. This finishes the proof.

Proof of Theorem 4.4:

From (4.5), e_n can be written as

$$
e_n = \left[\binom{n_1}{2}\binom{n_2}{2} \right]^{-1} \sum_{i_1,i_2 \in C_{n_1}^2} \sum_{j_1,j_2 \in C_{n_2}^2} \phi(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2}).
$$

The kernel function ϕ is defined as

$$
\begin{aligned}
\phi(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2}) &= h(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2}) - \tilde{h}_1^{(1)}(X_{1i_1}) - \tilde{h}_1^{(1)}(X_{1i_2}) \\
&\quad - \tilde{h}_1^{(2)}(X_{2j_1}) - \tilde{h}_1^{(2)}(X_{2j_2}) - \theta.
\end{aligned}
$$

It can be shown that

$$E\left\{\phi(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2})\right\} = 0,$$

$$E\left\{\phi(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2})|X_{1s} = x_{1s}\right\} = 0, a.e. \text{ for } s = i_1 \text{ or } i_2$$

$$E\left\{\phi(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2})|X_{2t} = x_{2t}\right\} = 0, a.e. \text{ for } t = j_1 \text{ or } j_2.$$

Therefore the variances of these three terms are all equal to 0, that is $\xi_{0,0}^2 = 0; \xi_{1,0}^2 = 0; \xi_{0,1}^2 = 0$. Then the variance of e_n can be simplified as

$$
\begin{aligned}
Var(e_n) &= \frac{2}{n_1(n_1-1)}\{1+o(1)\}\xi_{0,2}^2 + \frac{2}{n_2(n_2-1)}\{1+o(1)\}\xi_{2,0}^2 \\
&+ \frac{4}{n_1 n_2}\{1+o(1)\}\xi_{1,1}^2 + \frac{8}{n_1 n_2(n_2-1)}\{1+o(1)\}\xi_{1,2}^2 \\
&+ \frac{8}{n_2 n_1(n_1-1)}\{1+o(1)\}\xi_{2,1}^2 + \frac{4}{n_1 n_2(n_1-1)(n_2-1)}\xi_{2,2}^2.
\end{aligned}
$$

where

$$
\begin{aligned}
\xi_{2,0}^2 &= Var\left\{E(\phi(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2})|X_{1i_1} = x_{1i_1}, X_{1i_2} = x_{1i_2})\right\} \\
&= Var\left(\int \left[I(X_{1i_1} \le x)I(X_{1i_2} \le x) - \{I(X_{1i_1} \le x) + I(X_{1i_2} \le x)\}F_1(x)\right]w(x)dx\right) \\
&= \int \left\{F_1(x \wedge y) - F_1(x)F_1(y)\right\}^2 w(x)w(y)dxdy,
\end{aligned}
$$

Similarly, $\xi_{0,2}^2 = \int \left\{F_2(x \wedge y) - F_2(x)F_2(y)\right\}^2 w(x)w(y)dxdy.$

$$
\begin{aligned}
\xi_{1,1}^2 &= Var\left\{E(\phi(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2})|X_{1i_1} = x_{1i_1}, X_{2j_1} = x_{2j_1})\right\} \\
&= Var\left(\int \frac{1}{2}\left[I(X_{1i_1} \le x)F_2(x) + I(I_{2j_1} \le x)F_1(x) - I(X_{1i_1} \le x)I(I_{2j_1} \le x)\right]w(x)dx\right) \\
&= \frac{1}{4}\int \left\{F_1(x \wedge y) - F_1(x)F_1(y)\right\}\left\{F_2(x \wedge y) - F_2(x)F_2(y)\right\}w(x)w(y)dxdy,
\end{aligned}
$$

$$\xi_{1,2}^2 = Var\Big\{E(\phi(X_{1i_1}, X_{1i_2}, X_{2j_1}, X_{2j_2})|X_{1i_1} = x_{1i_1}, X_{2j_1} = x_{2j_1}, X_{2j_2} = x_{2j_2})\Big\}$$

$$= \int\Big[\big\{F_2(x \wedge y) - F_2(x)F_2(y)\big\}^2$$

$$+ \frac{1}{2}\big\{F_1(x \wedge y) - F_1(x)F_1(y)\big\}\big\{F_2(x \wedge y) - F_2(x)F_2(y)\big\}\Big]\{1 + o(1)\}w(x)w(y)dxdy$$

$$= (\xi_{0,2}^2 + 2\xi_{1,1}^2)\{1 + o(1)\},$$

$$\xi_{2,1}^2 = \int\Big[\big\{F_1(x \wedge y) - F_1(x)F_1(y)\big\}^2$$

$$+ \frac{1}{2}\big\{F_1(x \wedge y) - F_1(x)F_1(y)\big\}\big\{F_2(x \wedge y) - F_2(x)F_2(y)\big\}\Big]\{1 + o(1)\}w(x)w(y)dxdy$$

$$= (\xi_{2,0}^2 + 2\xi_{1,1}^2)\{1 + o(1)\},$$

$$\xi_{2,2}^2 = Var\Big\{\phi(X_{1i_1}, X_{1i_2}; X_{2j_1}, X_{2j_2})\Big\}$$

$$= \int\Big[\big\{F_1(x \wedge y) - F_1(x)F_1(y)\big\}^2 + \big\{F_2(x \wedge y) - F_2(x)F_2(y)\big\}^2$$

$$+ \big\{F_1(x \wedge y) - F_1(x)F_1(y)\big\}\big\{F_2(x \wedge y) - F_2(x)F_2(y)\big\}\Big]\{1 + o(1)\}w(x)w(y)dxdy$$

$$= (\xi_{2,0}^2 + \xi_{0,2}^2 + 4\xi_{1,1}^2)\{1 + o(1)\},$$

All those $o(1)$ terms vanish when $p \to \infty$ and $\min\{n_1, n_2\} \to \infty$. In summary,

$$Var(e_n) = \Big[\frac{2}{n_1(n_1 - 1)}\xi_{2,0}^2 + \frac{2}{n_2(n_2 - 1)}\xi_{0,2}^2 + \frac{4}{n_1 n_2}\xi_{1,1}^2\Big]\{1 + o(1)\}.$$

Under *(A1)*, *(A4)* and *(A5)*, we have $nVar(e_n)/\delta_1^2 = o(1)$. This finishes the proof.

Table 4.2   Empirical size of MRPP test, two-sample mean test and a permutation distribution test using T_{n_1,n_2} in Case 4: $F_1 = F_2$.

Type of		$C_1 = C_2 = N(0,1)$			$C_1 = C_2 = \chi^2(6)$	
Test	p	$n_1 = n_2 = 25$	$n_1 = n_2 = 50$	$n_1 = n_2 = 25$	$n_1 = n_2 = 50$	
Dist	10	.061	.047	.056	.050
	20	.052	.053	.053	.054
	50	.049	.053	.044	.057
Mean	10	.072	.060	.074	.090
	20	.070	.077	.073	.062
	50	.076	.067	.060	.060
MRPP	10	.055	.044	.049	.062
	20	.051	.055	.061	.047
	50	.060	.055	.055	.048

Table 4.3 Empirical power of MRPP test, two-sample mean test and a permutation distribution test using T_{n_1,n_2} in Case 1 where $C_1 = N(0,1)$ and $C_2 = \chi^2(6)$.

$n_1 = n_2$	p	Dist	Mean	MRPP
25	10	.108	.074	.068
	20	.314	.065	.047
	50	.711	.072	.068
50	10	.221	.081	.078
	20	.577	.060	.057
	50	.955	.081	.066

Table 4.4   Empirical power of MRPP test, two-sample mean test and a permutation distribution test using T_{n_1,n_2} in Case 2 where two-sample mean vectors are different: $\mu_1 = (0, \cdots, 0)^T$ and $\mu_2 = (.05, \cdots, .05)^T$.

Type of		$C_1 = C_2 = N(0,1)$		$C_1 = C_2 = \chi^2(6)$	
Test	p	$n_1 = n_2 = 25$	$n_1 = n_2 = 50$	$n_1 = n_2 = 25$	$n_1 = n_2 = 50$
Dist	10	.062	.061	.062	.063
	20	.060	.081	.057	.063
	50	.064	.076	.054	.061
Mean	10	.076	.077	.082	.101
	20	.085	.101	.080	.082
	50	.098	.105	.078	.099
MRPP	10	.051	.059	.055	.079
	20	.066	.083	.059	.073
	50	.076	.088	.068	.081

Table 4.5   Empirical power of MRPP test, two-sample mean test and a permutation distribution test using T_{n_1,n_2} in Case 2 where two-sample mean vectors are different: $\mu_1 = (0,\cdots,0)^T$ and $\mu_2 = (.10,\cdots,.10)^T$.

Type of		$C_1 = C_2 = N(0,1)$		$C_1 = C_2 = \chi^2(6)$	
Test	p	$n_1 = n_2 = 25$	$n_1 = n_2 = 50$	$n_1 = n_2 = 25$	$n_1 = n_2 = 50$
Dist	10	.085	.115	.068	.076
	20	.088	.142	.070	.079
	50	.100	.144	.062	.085
Mean	10	.112	.138	.119	.152
	20	.126	.205	.123	.188
	50	.164	.258	.130	.234
MRPP	10	.076	.113	.079	.112
	20	.097	.159	.097	.152
	50	.140	.227	.116	.210

Table 4.6 Empirical power of MRPP test, two-sample mean test and a permutation distribution test using T_{n_1,n_2} in Case 3 where two-sample covariance matrices are different: $\Gamma_1 = \Gamma(.5)$ and $\Gamma_2 = \Gamma(-.5)$.

Type of		$C_1 = C_2 = N(0,1)$		$C_1 = C_2 = \chi^2(6)$	
Test	p	$n_1 = n_2 = 25$	$n_1 = n_2 = 50$	$n_1 = n_2 = 25$	$n_1 = n_2 = 50$
Dist	10	.136	.206	.102	.155
	20	.077	.131	.083	.105
	50	.062	.084	.050	.062
Mean	10	.072	.060	.070	.065
	20	.067	.064	.078	.063
	50	.067	.069	.052	.059
MRPP	10	.130	.289	.136	.318
	20	.119	.224	.124	.256
	50	.099	.168	.081	.160

Table 4.7 Empirical power of MRPP test, two-sample mean test and a permutation distribution test using T_{n_1,n_2} in Case 3 where two-sample covariance matrices are different: $\Gamma_1 = \Gamma(1)$ and $\Gamma_2 = \Gamma(-1)$.

Type of Test	p	$C_1 = C_2 = N(0,1)$		$C_1 = C_2 = \chi^2(6)$	
		$n_1 = n_2 = 25$	$n_1 = n_2 = 50$	$n_1 = n_2 = 25$	$n_1 = n_2 = 50$
Dist	10	.201	.353	.139	.276
	20	.140	.237	.076	.095
	50	.112	.161	.052	.077
Mean	10	.067	.065	.066	.064
	20	.073	.058	.068	.080
	50	.068	.055	.059	.061
MRPP	10	.206	.592	.225	.662
	20	.202	.451	.204	.501
	50	.137	.278	.124	.290

Table 4.8 Empirical power of the distribution test using three types of bootstrap variance estimators: Case 1: $C_1 \neq C_2$; Cases 2: $\mu_1 \neq \mu_2$; Case 3: $\Gamma_1 \neq \Gamma_2$; Case 4: $F_1 = F_2$.

Type		$n_1 = n_2 = 25$			$n_1 = n_2 = 50$		
of test	Case	$p = 10$	$p = 20$	$p = 50$	$p = 10$	$p = 20$	$p = 50$
Dist	Case 1	.116	.310	.714	.206	.574	.958
(BTV-pooled)	Case 2	.074	.086	.114	.098	.134	.160
	Case 3	.180	.126	.094	.320	.226	.154
	Case 4	.050	.034	.052	.048	.046	.070
Dist	Case 1	.200	.390	.774	.336	.674	.972
(BTV-S1)	Case 2	.092	.106	.138	.122	.182	.184
	Case 3	.192	.134	.098	.340	.234	.158
	Case 4	.062	.050	.072	.070	.074	.092
Dist	Case 1	.118	.318	.754	.222	.592	.968
(BTV-S2)	Case 2	.094	.114	.148	.128	.174	.198
	Case 3	.254	.186	.182	.420	.276	.230
	Case 4	.064	.050	.078	.082	.068	.086

Table 4.9 Empirical power of the distribution test based on bootstrapping T_{n_1,n_2} from only one sample: Case 1: $C_1 \neq C_2$; Cases 2: $\mu_1 \neq \mu_2$; Case 3: $\Gamma_1 \neq \Gamma_2$; Case 4: $F_1 = F_2$.

Type of test	Case	$n_1 = n_2 = 25$			$n_1 = n_2 = 50$		
		$p = 10$	$p = 20$	$p = 50$	$p = 10$	$p = 20$	$p = 50$
Dist (BT-S1)	Case 1	.176	.338	.718	.286	.632	.962
	Case 2	.078	.084	.122	.100	.142	.148
	Case 3	.164	.110	.072	.294	.200	.122
	Case 4	.052	.038	.046	.048	.052	.074
Dist (BT-S2)	Case 1	.094	.272	.714	.178	.542	.954
	Case 2	.070	.078	.114	.098	.134	.160
	Case 3	.222	.156	.168	.360	.254	.200
	Case 4	.054	.038	.062	.048	.048	.068

BIBLIOGRAPHY

Abramovich, F., Benjamini, Y., Donoho, D.L., and Johnstone, I.M. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *The Annals of Statistics*, 34(2):584–653.

Ahmad, I.A. (1996). Modification of some goodness of fit statistics II: two-sample and symmetry testing. *Sankhyā: The Indian Journal of Statistics, Series A*, 80(2):464–472.

Alba Fernández, V., Jiménez Gamero, M.D., and Muñoz García, J. (2008). A test for the two-sample problem based on empirical characteristic functions. *Computational Statistics and Data Analysis*, 52(7):3730–3748.

Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2):193–212.

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29.

Bai, Z.D. and Saranadasa, H. (1996). Effect of high-dimension: by an example of a two-sample problem. *Statistica Sinica*, 6:311–329.

Bai, Z.D. , Silverstein, J.W., and Yin, Y.Q. (1988). A note on the largest eigenvalue of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis*, 26(2):166–168.

Bai, Z.D. and Yin, Y.Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *The Annals of Probability*, pages 1275–1294.

Baringhaus, L. and Franz, C. (2004). On a new multivariate two-sample test. *Journal of Multivariate Analysis*, 88(1):190–206.

Barry, W.T., Nobel, A.B., and Wright, F.A. (2005). Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinformatics*, 21(9):1943–1949.

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 (Methodology)*, 57(1):289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188.

Bickel, P.J. (1969). A distribution free version of the Smirnov two-sample test in the p-variate case. *The Annals of Mathematical Statistics*, 40(1):1–23.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Brown, B.M. and Chen, S.X. (1998). Combined and least squares empirical likelihood. *Annals of the Institute of Statistical Mathematics*, 50(4):697–714.

Chen, S.X. (1993). On the accuracy of empirical likelihood confidence regions for linear regression model. *Annals of the Institute of Statistical Mathematics*, 45(4):621–637.

Chen, S.X. (1994a). Comparing empirical likelihood and bootstrap hypothesis tests. *Journal of Multivariate Analysis*, 51(2):277–293.

Chen, S.X. (1994b). Empirical likelihood confidence intervals for linear regression coefficients. *Journal of Multivariate Analysis*, 49(1):24–40.

Chen, S.X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika*, 83(2):329–341.

Chen, S.X. and Cui, H.J. (2003). An extended empirical likelihood for generalized linear models. *Statistica Sinica*, 13:69–82.

Chen, S.X. and Cui, H.J. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika*, 93(1):215–220.

Chen, S.X., Liang, P. and Qin, Y.L. (2009). Effects of data dimension on empirical likelihood. *Biometrika*, 96(3):711722

Chen, S.X. and Qin, Y.L. (2009). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, to appear.

Chen, S.X. and Qin, Y.S. (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika*, 87(4):946–953.

Chiaretti, S., Li, X.C., 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.

Darling, D.A. (1957). The Kolmogorov-Smirnov, Cramer-Von Mises tests. *The Annals of Mathematical Statistics*, 28(4):823–838.

DiCiccio, T.J., Hall, P., and Romano, J.P. (1989). Comparison of parametric and empirical likelihood functions. *Biometrika*, 76(3):465–476.

DiCiccio, T.J., Hall, P., and Romano, J.P. (1991). Empirical likelihood is Bartlett-correctable. *The Annals of Statistics*, pages 1053–1061.

Dudoit, S., Keles, S., and van der Laan, M.J. (2008). Multiple tests of association with biological annotation metadata. *Institute of Mathematical Statistics Collections*, 2:153–218.

Efron, B. and Tibshirani, R. (2007). On testing the significance of sets of genes. *The Annals of Applied Statistics*, 1(1):107–129.

Fan, J.Q, Hall, P., and Yao, Q.W. (2007). To how many simultaneous hypothesis tests can normal, Student's t or bootstrap calibration be applied? *Journal of the American Statistical Association*, 102(480):1282–1288.

Fan, J.Q. , Peng, H., and Huang, T. (2005). Semilinear high-dimensional model for normalization of Microarray Data. *Journal of the American Statistical Association*, 100(471):781–796.

Fang, K.T. and Zhang, Y.T. (1990). *Generalized multivariate analysis*. Science Press.

Fasano, G. and Franceschini, A. (1987). A multidimensional version of the Kolmogorov-Smirnov test. *Royal Astronomical Society, Monthly Notices*, 225:155–170.

Friedman, J.H. and Rafsky, L.C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *The Annals of Statistics*, 4:697–717.

Gentleman, R., Carey, V., Huber, W., Irizarry, R., and Dudoit, S. (2005). *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer Verlag.

Hall, P. (1990). Pseudo-likelihood theory for empirical likelihood. *The Annals of Statistics*, 18(1):121–140.

Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and application*. Academic Press, New York

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review*, 58(2):109–127.

Hall, P. and Owen, A.B. (1993). Empirical likelihood confidence bands in density estimation. *Journal of Computational and Graphical Statistics*, 2(3):273–289.

Hall, P. and Tajvidi, N. (2002). Permutation tests for equality of distributions in high-dimensional settings. *Biometrika*, 89(2):359–374.

Hjort, N.L., McKeague, I.W., and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *The Annals of Statistics*, 37(3):1079–1111.

Huang, J., Wang, D.L., and Zhang, C.H. (2005). A two-way semilinear model for normalization and analysis of cDNA microarray data. *Journal of the American Statistical Association*, 100(471):814–829.

Jeffery, I.B., Higgins, D.G., and Culhane, A.C. (2006). Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinformatics*, 7:359.

Korolyuk, V.S. and Borovskikh, Y.V. (1989). *The theory of U-statistics.* Kluwer Academic Pulbishers.

Kosorok, M.R. and Ma, S.G. (2007). Marginal asymptotics for the "large p, small n" paradigm: with applications to microarray data. *The Annals of Statistics*, 35(4):1456–1486.

Lee, A. J. (1990). *U-statistics: Theory and Practice.* CRC.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30(4):1081–1102.

Liu, R. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 27(3):252–260.

Nettleton, D., Recknor, J., and Reecy, J.M. (2008). Identification of differentially expressed gene categories in microarray studies using nonparametric multivariate analysis. *Bioinformatics*, 24(2):192–201.

Newton, M.A., Quintana, F.A., den Boon, J.A., Sengupta, S., and Ahlquist, P. (2007). Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *The Annals of Applied Statistics*, 1(1):85–106.

Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Owen, A.B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120.

Owen, A.B. (2001). *Empirical likelihood.* Chapman & Hall Press.

Peacock, J.A. (1983). Two-dimensional goodness-of-fit testing in astronomy. *Royal Astronomical Society, Monthly Notices*, 202:615–627.

Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *The Annals of Statistics*, 12(4):1298–1309.

Portnoy, S. (1985). Asymptotic Behavior of M Estimators of p Regression Parameters when p^2/n is Large; II. Normal Approximation. *The Annals of Statistics*, 13(4):1403–1417.

Portnoy, S. (1986). On the central limit theorem in R^p when $p \to \infty$. *Probability Theory and Related Fields*, 73(4):571–583.

Qin, G.S. and Jing, B.Y. (2001). Empirical likelihood for Cox regression model under random censorship. *Communications in Statistics-Simulation and Computation*, 30(1):79–90.

Qin, J. (1993). Empirical likelihood in biased sample problems. *The Annals of Statistics*, 21(3):1182–1196.

Qin, J. (1999). Empirical likelihood ratio based confidence intervals for mixture proportions. *The Annals of Statistics*, 27(4):1368–1384.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, pages 300–325.

Qin, J. and Zhang, B. (2007). Empirical-likelihood-based inference in missing response problems and its application in observational studies. *Journal of the Royal Statistical Society: Series B (Methodology)*, 69(1):101–122.

Schott, J.R. (2005). Testing for complete independence in high dimensions. *Biometrika*, 92(4):951–956.

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

Srivastava, M.S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis*, 100(3):518–532.

Storey, J.D., Taylor, J.E., and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society. Series B (Methodology)*, 66(1):187–205.

Storey, J.D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550.

Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *The Annals of Statistics*, 32(3):1215–1221.

Van der Laan, M.J. and Bryan, J. (2001). Gene expression analysis with the parametric bootstrap. *Biostatistics*, 2(4):445–461.

Yin, Y.Q., Bai, Z.D., and Krishnaiah, P.R. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields*, 78(4):509–521.

Zuo, Y.J. and He, X.M. (2006). On the limiting distributions of multivariate depthbased rank sum statistics and related tests. *The Annals of Statistics*, 34(6):2879.