

Topics on high dimensional statistical inference and ANOVA for longitudinal data

by

Pingshou Zhong

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

Peng Liu

William Q. Meeker

Dan Nettleton

Zhengyuan Zhu

Iowa State University

Ames, Iowa

2011

Copyright © Pingshou Zhong, 2011. All rights reserved.

DEDICATION

To my family

TABLE OF CONTENTS

LIST OF TABLES	v
LIST OF FIGURES	vii
ACKNOWLEDGEMENTS	ix
CHAPTER 1. Introduction	1
1.1 Introduction	1
1.2 High Dimensional Tests	2
1.2.1 High Dimensional Tests for Regression Coefficients	4
1.2.2 Threshold Test for High Dimensional Mean under Dependency	7
1.3 ANOVA Tests for Longitudinal Data	9
1.4 Empirical Likelihood	11
1.5 Thesis Organization	13
References	15
CHAPTER 2. Tests for High Dimensional Regression Coefficients with Factorial Designs	21
2.1 Introduction	21
2.2 Models and Test Statistics	24
2.2.1 F-test and Its Performances under High Dimensionality	24
2.2.2 A New Test Statistic	26
2.3 U-Statistics under High Dimensionality	27
2.4 Main Results	29
2.5 Generalization to Factorial Designs	33
2.6 Simulation Study	36

2.7	Association Test for Gene-sets	45
2.8	Appendix: Technical Details	48
	References	62
CHAPTER 3. Threshold Test for High Dimensional Mean under Dependency		66
3.1	Introduction	66
3.2	Large Deviation Approximation to the Mean and Variance	70
3.3	Asymptotic Distribution of T_n	73
3.4	Extension to the Maximum Test	76
3.5	Optimal Detection Boundary and the Best Power	77
3.6	Simulation Results	82
3.7	Appendix: Technical Details	86
	References	121
CHAPTER 4. ANOVA for Longitudinal Data with Missing Values		124
4.1	Introduction	124
4.2	Models, Hypotheses and Missing Values	127
4.3	ANOVA Test for Covariate Effects	130
4.4	ANOVA Test for Time Effects	134
4.5	Tests on Interactions	138
4.6	Bootstrap Calibration	139
4.7	Simulation Results	141
4.8	Analysis on HIV-CD4 Data	146
	4.8.1 Monotone Missingness	148
	4.8.2 Not-monotone Missingness	153
4.9	Appendix: Technical Details	156
	References	172
CHAPTER 5. Summary and General Discussion		177
	References	179

LIST OF TABLES

Table 2.1	Empirical size and power of the F-test, the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p \times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p \times 1}$ at significant level 5% for normal residual. LP represents the theoretical local power.	39
Table 2.2	Empirical size and power of the F-test, the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p \times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p \times 1}$ at significant level 5% for centralized gamma residual. LP represents the theoretical local power.	40
Table 2.3	Empirical size and power of the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p \times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p \times 1}$ at significant level 5% for normal residual. LP represents the theoretical local power.	41
Table 2.4	Empirical size and power of the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p \times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p \times 1}$ at significant level 5% for centralized gamma residual. LP represents the theoretical local power.	42
Table 2.5	Empirical size and power of the proposed test for $H_0 : \beta = \mathbf{0}_{p \times 1}$ in a 2×2 factorial design with $n_1 = 20$ and $n_2 = 30$ replicates in each cell.	44
Table 2.6	P-values of the GO terms which are significant under at least three designs using the proposed test, and their number of genes.	49
Table 3.1	Empirical sizes of the Oracle test, C-Q test, FDR, maximum test and threshold tests with different threshold levels $\lambda_n = 2s \log(p)$ for Gaussian process.	87
Table 3.2	Empirical sizes of the Oracle test, C-Q test, FDR, maximum test and threshold tests with different threshold levels $\lambda_n = 2s \log(p)$ for process with standardized Gamma(2,2) marginal distribution.	88

Table 4.1	Empirical size and power of the 5% ANOVA test for $H_{0a} : \beta_{10} = \beta_{20} = \beta_{30}$.	144
Table 4.2	Empirical size and power of the 5% test for the existence of interaction $H_{0c} : \gamma_{20} = 0$.	145
Table 4.3	Empirical size and power of the 5% ANOVA test for $H_{0d} : \gamma_{10} = \gamma_{20} = \gamma_{30}$.	146
Table 4.4	Empirical size and power of the 5% ANOVA test for $H_{0b} : g_1(\cdot) = g_2(\cdot) = g_3(\cdot)$ with $\Delta_{2n}(t) = U_n \sin(2\pi t)$ and $\Delta_{3n}(t) = 2 \sin(2\pi t) - 2 \sin(2\pi(t + V_n))$.	147
Table 4.5	The empirical sizes and powers of the proposed test (CZ) and the test (SZ) proposed by Scheike and Zhang (1998) for $H_{0b} : g_1(\cdot) = g_2(\cdot)$ vs $H_{1b} : g_1(\cdot) = g_2(\cdot) + \Delta_{2n}(\cdot)$.	148
Table 4.6	Differences in the AIC and BIC scores among three models (M1-M3)	149
Table 4.7	Parameter estimates and their standard errors	150
Table 4.8	P-values of ANOVA tests for β_j s.	151
Table 4.9	P-values of ANOVA tests on $g_j(\cdot)$ s.	152
Table 4.10	Differences in the AIC and BIC scores among three models (M1-M3) for $d = 1$.	153
Table 4.11	Parameter estimates and their standard errors with $d = 1, 2, 3$	154
Table 4.12	P-values of ANOVA tests on β s with $d = 1, 2, 3$	155
Table 4.13	P-values of ANOVA tests on $g_j(\cdot)$ s with $d = 1, 2, 3$	156

LIST OF FIGURES

Figure 2.1	The auto-correlation functions for series $\{X_{ij}\}_{j=1}^p$	37
Figure 2.2	The null distributions of standardized $T_{n,p}$	43
Figure 2.3	Histograms of the p-values on all GO terms using the proposed tests. .	46
Figure 2.4	Histograms of the p-values on all GO terms using Empirical Bayes (EB) tests.	47
Figure 2.5	Differences in the p-values among Designs I-IV.	48
Figure 3.1	The detectable region of the threshold and the maximum test in (β, r) plane, which is the union of I-IV areas in the plot.	80
Figure 3.2	The histograms for the simulated null distributions of standardized T_n using plug-in, theoretical variance (3.3.14) estimate and spectral smoothed variance estimate introduced in the Appendix. The (p, n) is (1000,20). Marginal distribution: Gaussian.	83
Figure 3.3	The histograms for the simulated null distributions of standardized T_n using plug-in, theoretical variance (3.3.14) estimate and spectral smoothed variance estimate introduced in Appendix. The (p, n) is (2000,30). Marginal distribution: Gaussian.	84
Figure 3.4	The histograms for the simulated null distributions of standardized T_n using plug-in, theoretical variance (3.3.14) estimate and spectral smoothed variance estimate introduced in Appendix. The (p, n) is (2500,40). Marginal distribution: Gaussian.	85

Figure 3.5	The ROC curves of the Oracle test, C-Q test, FDR test, Maximum test and the threshold test at different levels with Type I error between 0-0.2. From top to bottom, $r = 0.4, 0.6$ and 0.9 . From left to right panels $\beta = 0.6, 0.7, 0.8$. ($p = 1000, n = 20$)	89
Figure 3.6	The ROC curves of the Oracle test, C-Q test, FDR test, Maximum test and the threshold test at different levels with Type I error between 0-0.2. From top to bottom, $r = 0.4, 0.6$ and 0.9 . From left to right panels $\beta = 0.6, 0.7, 0.8$. ($p = 2000, n = 30$)	90
Figure 3.7	The ROC curves of the Oracle test, C-Q test, FDR test, Maximum test and the threshold test at different levels with Type I error between 0-0.2. From top to bottom, $r = 0.4, 0.6$ and 0.9 . From left to right panels $\beta = 0.6, 0.7, 0.8$. ($p = 2500, n = 40$)	91
Figure 4.1	(a) The raw data plots with the estimates of $g_j(t)$ ($j = 1, 2, 3, 4$). (b) The estimates of $g_j(t)$ in the same plot: Treatment I (solid line), Treatment II (short dashed line), Treatment III (dashed and dotted line) and Treatment IV (long dashed line).	151

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Song Xi Chen for his guidance, support, encouragement and tremendous effort to my research projects during my Ph.D studies. His enthusiasm and dedication toward the research will always inspire me. I am also grateful to Dr. Dan Nettleton and Dr. Peng Liu for many useful discussion and suggestions. I also appreciate my committee members for their efforts and advices to my work: Dr. Wayne Fuller, Dr. William Meeker and Dr. Zhengyuan Zhu. Sincerely thank Dr. Ken Koehler for various help. I also thank Dr. Long Qu for providing me the Yorkshire gilt data set and introducing the biological background. Thanks also goes to Heng Wang for her support and encouragement. Finally, thanks my family for consistent support.

CHAPTER 1. Introduction

1.1 Introduction

The first part of this thesis contains Chapters 2 and 3, where we investigate the statistical inference for high dimensional data. High dimensional data are more and more easily collected in scientific research, engineering, financial and medical areas. In genetic studies, high-throughput technologies such as microarray and next generation sequencing produce thousands of measurements in a single chip. For example, microarray allows researchers monitor the expression levels of thousands of genes in an experiment to study effects of treatments and diseases on gene expression. Next generation sequencing can examine DNA copy numbers at thousands sites of a genome by mapping tens of millions of short reads. For more examples and statistical challenges in high dimensional data, see Donoho (2000), Johnstone and Titterington (2009) and Fan and Lv (2010). Chapters 2 and 3 focus on high dimensional simultaneous tests. Chapter 2 considers the high dimensional tests for regression coefficients while Chapter 3 is on tests for high dimensional means under sparsity and dependency. We examine the limitation of some classical tests and propose new tests which are applicable in the high dimension and small sample size scenarios.

The second part of this thesis is on ANOVA test for longitudinal data. Longitudinal studies collect repeated measurements on each individual, which can be collected either prospectively or retrospectively. The benefit of a longitudinal study is the ability to distinguish the time effects (i.e., changes over time) and cohort effects (i.e., differences between subjects at baseline). For this reason, longitudinal studies are widely used in clinical trials and social science studies. Despite its advantages, the modelling of the correlation structure, missing values and unbalanced designs are challenging. Chapter 4 of this thesis proposes empirical likelihood based tests for

comparing the treatment effects including time effects and cohort effects in longitudinal studies with missing values.

In the rest of this chapter, we will review some important issues and developments in high dimensional simultaneous tests in Section 1.2. Section 1.3 provides an introduction to ANOVA test for longitudinal data. Section 1.4 reviews concepts of empirical likelihood. More literature reviews can be found in each chapter.

1.2 High Dimensional Tests

High dimensionality is one aspect of the massive data we encounter nowadays. The small sample size is another typical situation we face. Due to financial restrictions and the resources available to produce many replicates, we often have very limited sample sizes. For example, sample sizes for most microarray data are less than one hundred. High dimensionality itself poses a challenge to classical statistical inference, together with small sample size makes the situation more difficult. This is the so-called “large p , small n ” problem.

Classical asymptotic statistical inference typically assumes that the dimension p fixed but the sample size n goes to infinity. This “large n , fixed p ” setup is of course not suitable for high dimensional data where p is much larger than n . It is natural to consider p increasing to infinity as n goes to infinity. So it is important to evaluate how the classical methods perform when the data dimension increases with the sample size and what is the limit of the dimensionality in the classical methods. More importantly, we need new methods that are able to handle the high dimensionality.

The historic development can be traced back to Neyman and Scott (1948), who pointed out the inconsistency of some parameter estimators when the dimension increases with the sample size. Huber (1973), Yohai and Maronna (1979) and Portnoy (1984, 1985) treated the consistency and asymptotic normality of least square and M-estimators of the regression parameters when the number of parameters grows with the sample size. In analysis of spectral density of the large dimensional sample covariance matrix, the dimension of the covariance matrix is often assumed to increase with the sample size. The pioneering works in this area include Marčenko and Pastur (1967) and Pastur (1972, 1973). See Bai and Siliverman (2006)

for a review.

The literature on high dimensional tests is expanding quickly as other fields in high dimensional statistical inference. Portnoy (1988) considered the asymptotic distribution of the maximum likelihood estimator (MLE) for exponential families. He showed that the asymptotic normality of MLE would hold if $p^2/n \rightarrow 0$ and if $p^{3/2}/n \rightarrow 0$, the likelihood ratio test statistic Λ_n for a simple hypothesis has a chi-square distribution in the sense that $(-2 \log(\Lambda_n) - p)/\sqrt{2p} \rightarrow N(0,1)$. Bai and Saranadasa (1996) showed that the Hotelling T^2 test lose power when p is close to n and proposed a new method, which is further improved by Chen and Qin (2010). Ledoit and Wolf (2002) studied two likelihood ratio tests (John 1971, 1972 and Nagao 1973) for the sphericity and identity covariance matrix hypotheses for high-dimensional normally distributed random vectors when $p/n \rightarrow c$ for a finite constant c . They found the sphericity test is robust under $p/n \rightarrow c$, whereas the identity test is untenable. Some recent developments on testing the high dimensional covariance matrix including Chen, Zhang and Zhong (2010), Schott (2006), Srivastava (2005) and Srivastava and Yanagiharab (2010) among others. Fan and Peng (2004) studied the penalized likelihood ratio test with diverging number of parameters and they showed that Wilks' theorem holds if $p^5/n \rightarrow 0$.

To better illustrate the high dimensional asymptotics and the effect of high dimensionality, let us take a look at an example from Bai and Silverstein (2006):

Example: Let $X_i = (X_{i1}, \dots, X_{ip})'$ where X_{ij} are independent and identically distributed (IID) random variables with $N(0,1)$. Define the sample covariance $S_n = \frac{1}{n} \sum_{i=1}^n X_i X_i'$. An important test statistic in multivariate statistics is

$$T_n = \log\{\det(S_n)\} = \sum_{i=1}^p \log(\hat{\lambda}_i) \quad (1.2.1)$$

where $\hat{\lambda}_1, \dots, \hat{\lambda}_p$ are the eigenvalues of S_n . When p is fixed, we know that S_n is distributed as $W_p(n, \frac{1}{n}I_p)$, the Wishart distribution with n degrees of freedom and covariance matrix $\frac{1}{n}I_p$. Then by a result from multivariate statistics (see Muirhead, 1982, page 102),

$$\sqrt{\frac{n}{2p}} T_n \xrightarrow{d} N(0, 1). \quad (1.2.2)$$

This result holds for any fixed p and $n \rightarrow \infty$. However, this result is not true if p goes to infinity at the same rate as n . Let us consider $p/n \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$ and c is a constant. We can

write $\frac{1}{p}T_n$ as

$$\frac{1}{p}T_n = \frac{1}{p} \sum_{i=1}^p \log(\hat{\lambda}_i) = \int \log(x) dF^S(x)$$

where $F^S(x)$ is the empirical spectral distribution of S_n defined as

$$F^S(x) = \frac{1}{p} \#\{j \leq p : \hat{\lambda}_j \leq x\}.$$

By Theorem 3.5 in Bai and Silverstein (2005), with probability one,

$$\sup_x |F^S(x) - F(x)| \rightarrow 0,$$

where $F(x)$ is the so-called Marčenko-Pastur (MP) law that has a density

$$f_c(x) = \begin{cases} \frac{1}{2\pi xc} \sqrt{(b-x)(x-a)}, & \text{if } a \leq x \leq b; \\ 0, & \text{otherwise} \end{cases}$$

and has a point mass $1 - 1/c$ at 0 if $c > 1$, where $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$. Therefore, with probability one,

$$\frac{1}{p}T_n \rightarrow \int \log(x) dF(x) = \frac{c-1}{c} \log(1-c) - 1 < 0, \quad (1.2.3)$$

which implies that as $p/n \rightarrow c$ and $n \rightarrow \infty$, almost surely,

$$\sqrt{\frac{n}{2p}} T_n = \sqrt{\frac{np}{2}} \frac{1}{p} T_n \rightarrow -\infty.$$

This small example shows that using the asymptotic distribution given in (1.2.2) for high dimensional data is no longer adequate, and new asymptotics have to be developed for high dimensional inference.

1.2.1 High Dimensional Tests for Regression Coefficients

Let us consider a high dimensional linear model

$$Y = X\beta + \epsilon,$$

where $Y = (Y_1, \dots, Y_n)$, $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ and $X = (X_1, \dots, X_n)'$ is a $n \times p$ matrix of predictors. Assume that $p \rightarrow \infty$ as $n \rightarrow \infty$. In the variable selection context, much attention has been

focused on the consistency of the model selection and estimation efficiency in the “large p , small n ” scenario. See for example, Chen and Chen (2008), Greenshtein and Ritov (2004), Meinshausen and Bühlmann (2006) and Zhao and Yu (2006).

Our purpose in Chapter 2 is to test

$$H_0 : \beta = 0 \quad \text{vs} \quad H_1 : \beta \neq 0. \quad (1.2.4)$$

This was motivated by the latest need in biology to identify significant sets of genes, which are associated with certain clinical outcome, rather than identifying individual gene. As advocated by Subramanian *et al.* (2005), gene-set testing gives more consistent results across different studies and can detect much smaller effects than the single gene based methods. There are several resampling methods proposed to test the significance of a gene-set. The basic idea is to test if a given set is over represented or enriched by significant genes. Subramanian *et al.* (2005) proposed a gene-set enrichment score method (GSEA). Efron and Tibshirani (2007) further improved GSEA by using the maxmean statistic for summarizing gene-sets and restandardization for more accurate inferences. Newton *et al.* (2007) improved the GSEA by considering the size of the gene sets and the dependence among enrichment scores. However, all of these methods fail to account for the dependence among the genes in the gene sets and essentially use the marginal regression coefficients.

In the fixed dimensional case, the F-test is often used for testing regression coefficients simultaneously in a linear model. Under the conditional normality assumption, the F-test is the uniformly most powerful (UMP) test. However, if $p > n$, the F-test is not applicable. In a similar but essentially different problem, Meinshausen, Meier and Bühlmann (2008) discussed a multi-split method for assigning p-values to each regression coefficient instead of the regression coefficient vector in high dimensional regression context, which extends the single split method proposed by Wasserman and Roeder (2008).

More closely related studies were pioneered by Portnoy (1984, 1985), who studied the consistency and the asymptotic normality of the M-estimator $\hat{\beta}$ when p goes to infinity as n goes to infinity, where $\hat{\beta}$ is the solution of the following equation:

$$\sum_{i=1}^n X_i \Psi(Y_i - X_i' \beta) = 0$$

and Ψ is a given function satisfying some conditions. They showed that $\hat{\beta}$ is still consistent if $p \log(p)/n \rightarrow 0$ and further if $\{p \log(p)\}^{3/2}/n \rightarrow 0$, $\frac{d_{\Psi}^2}{\hat{\sigma}_{\Psi}^2}(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \xrightarrow{d} \chi_p^2$, where $d_{\Psi} = E(\Psi'(\epsilon_i))$ and $\sigma_{\Psi}^2 = E(\Psi^2(\epsilon_i))$. Thus we may formulate a test by rejecting the null hypothesis at level α if

$$\frac{d_{\Psi}^2}{\hat{\sigma}_{\Psi}^2} \hat{\beta}'X'X\hat{\beta} > \chi_{p; \alpha}^2,$$

where $\hat{\sigma}_{\Psi}^2$ is a consistent estimator of σ_{Ψ}^2 . But this method only works for very mild dimensionality, i.e. $\{p \log(p)\}^{3/2}/n \rightarrow 0$.

Goeman *et al.* (2006, 2009) proposed an Empirical Bayes (EB) method to test the regression coefficients, which can be applied to the high dimensional case. Basically, they assumed that β is a random vector with mean 0 and covariance $\tau^2 I_p$. Then testing hypothesis of (1.2.4) is equivalent to testing

$$H_0 : \tau^2 = 0 \quad \text{vs} \quad H_1 : \tau^2 \neq 0. \quad (1.2.5)$$

Suppose that $Y \sim N(X\beta, \sigma^2 I)$ and $L(\beta; Y)$ be the likelihood of β given Y . Then the marginal density of Y is given by $\bar{f}(\tau^2; y) = E_{\beta|\tau^2}\{L(\beta; Y)\}$, where $E_{\beta|\tau^2}\{\cdot\}$ is the expectation with respect to prior distribution of β given τ^2 . A score test (Goeman *et al.*, 2006) can be used to test (1.2.5), which is

$$S = \frac{\partial \log\{\bar{f}(0; Y)\}}{\partial \tau^2}.$$

The test statistics is equivalent to $S' = \frac{y'X X' y}{y' y}$. In Goeman *et al.* (2009), they derived an asymptotic distribution for S' . However, the asymptotic distribution was based on fixed p but sample size n goes to infinity.

Although the test for regression coefficients was motivated by gene-set testing, it can be also applied to many other possible situations. One possibility is for functional data. The functional data are similar to longitudinal data with repeated measures for each individual. The difference exists in the dimension of the repeated measurements for each individual. The functional data usually have a much higher dimension than the longitudinal data. If we assume that $X_i(t)$ were collected at time t_1, \dots, t_p and $X_i = (X_i(t_1), \dots, X_i(t_p))'$, then the corresponding model can be written as

$$Y_i = \sum_{j=1}^p X_i(t_j)\beta(t_j) + \epsilon_i.$$

Thus the hypothesis (1.2.4) is equivalent to

$$H_0 : \beta(t) = 0 \quad \text{vs} \quad H_1 : \beta(t) \neq 0$$

for $t = t_1, \dots, t_p$. The null hypothesis means that the functional data are not significant associated with the response Y .

In Chapter 2, we proposed new test statistics and derived the asymptotic distributions of the test statistics under the “large p , small n ” scenarios. We allow the dimension p to grow much faster than the sample size n . Our simulation showed that the test performs better than the F-test and EB test both in moderate and high dimensional cases.

1.2.2 Threshold Test for High Dimensional Mean under Dependency

In the last section, we discussed the test for high dimensional regression coefficients β . The test is suitable to test against both non-sparse and sparse alternatives. However, due to high dimensionality, the test potentially loses power in the sparse case. The purpose of this section is to discuss the high dimensional test under sparsity and dependency. The sparsity condition is the key to many variable selection procedures, which makes the high dimensional parameter estimation possible (Fan and Lv, 2010).

Suppose X_1, \dots, X_n are IID p -dimensional vectors with mean μ . We consider testing for the high dimensional mean under sparsity i.e., testing

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0 \tag{1.2.6}$$

where $\mu = (\mu_1, \dots, \mu_p)'$ and we assume that only a small fraction $p^{-\beta}$ ($\beta > 1/2$) of the p components of μ are not zero under the alternative, which is a very sparse case. The hypothesis has been considered by several other papers including Fan (1996), Donoho and Jin (2004) and Kim and Akritas (2010) among others. The Donoho and Jin (2004)’s study provided an elegant analysis on the asymptotic detection boundary of the higher criticism in the (β, r) plane for independent and normally distributed data with non-zero signals $\mu_i = \sqrt{2r \log(p)}$.

Multiple comparison is a way for simultaneously testing (1.2.6). The challenges in multiple testing include controlling the family wise error rate (FWER) or false discovery rate (FDR),

and accounting for the dependence among marginal test statistics. There is a huge amount literature on multiple comparison. See Benjamini and Hochberg (1995), Hommel (1988), Simes (1986) for controlling FDR or FWER procedures; Storey (2002, 2003), Storey *et al.* (2004) and Storey (2007) for controlling positive FDR and optimal discovery procedure. See Efron *et al.* (2001) for local FDR. For testing grouped hypothesis, Cai and Sun (2009) introduced a procedure minimizing the false nondiscovery rate while controlling the false discovery rate. However, these multiple comparison methods are not able to provide a p-value for indicating the significance of the grouped hypothesis (1.2.6) and most of methods are only justified for independent hypotheses. Some exceptions are Benjamini and Yekutieli (2001) and Hall and Wang (2010).

It is a common practice to apply t-tests or tests based on asymptotic normality in multiple comparison. Normal distribution or t-distribution can serve as good calibrations for the null distributions of the test statistics in low dimensional inference. However, are these calibrations accurate enough in “large p , small n ” scenarios when the underlying distribution is non-normal? An interesting study conducted by Fan, Hall and Yao (2007) showed that the level of the simultaneous test based on p-values calculated from normality approximation is accurate provided that $\log(p)$ increases at a strictly slower rate than $n^{1/3}$ as n diverges, see also Kosorov and Ma (2007) for more discussion.

Chen and Qin (2010) proposed a method for simultaneous testing high dimensional two sample means. Their method can be applied to the test above hypothesis (1.2.6). Basically, we can formulate a test statistic

$$T_{CQ} = \frac{1}{n(n-1)} \sum_{i \neq j} X_i' X_j.$$

As shown by Chen and Qin (2010), the test statistic $\frac{nT_{CQ}}{\sqrt{2\text{tr}(\Sigma^2)}} \xrightarrow{d} N(0, 1)$ under the null hypothesis. The advantage of this proposal is that the asymptotic distribution does not depend on the underlying distribution of X and it can adapt to a wide range of high dimensionality. However, this test statistic is not effective in testing sparse alternatives, see the simulation results in Chen and Qin (2010).

To better account for the sparsity and dependency, we propose a threshold test statistic for

testing (1.2.6) in Chapter 3. The test statistic is

$$T_n = \sum_{i=1}^p Y_{i,n} I\{Y_{i,n} > \lambda_n\} \quad (1.2.7)$$

where $Y_{i,n} = n\bar{X}_i^2$, $\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}/\sigma_i$ is the scaled sample mean of the i -th component of X and $\lambda_n = 2s \log(p)$ is the level of threshold where $0 < s < 1$. Fan (1996) showed the asymptotic distribution of T_n assuming $Z_i = \sqrt{n}\bar{X}_i$ s are normally distributed and the components of X are independent. However, some questions remain unanswered.

- (a) Is the threshold test able to attain the optimal detection boundary for dependent and non-normal data?
- (b) When can the normal calibration for Z_i be used? i.e., the robustness of the distribution assumption.
- (c) How does the dependence among the components of X_i affect the asymptotic distribution of T_n ?

The purpose of Chapter 3 is to answer these questions. We showed that the normal calibration to Z_i is valid if $\log(p) = o(n^{1/3})$ and the Cramér condition holds. The asymptotic normality of the threshold test statistic was established for strong mixing sequences. To avoid selection of threshold parameter, we also propose a maximum test, which maximizes the standardized threshold test statistic over a range of threshold values. It is shown that the maximum threshold test can attain the optimal detection boundary (Donoho and Jin, 2004).

1.3 ANOVA Tests for Longitudinal Data

The ANOVA test is one of the most common problems for many practitioners. The ANOVA table can be easily constructed for balanced data. We can treat balanced longitudinal data as a split plot design, each individual as a whole plot while each repeated measurements as a split plot. Thus the ANOVA for split plot designs can be used for longitudinal data. Linear mixed-effect models have a long history for modelling longitudinal data. Let Y_{jim} be the m -th measurement of the i -th individual in the j -th treatment. A mixed effect model corresponding

to split plot design is

$$Y_{jim} = \mu + \beta_j + \alpha_m + (\alpha\beta)_{jm} + \omega_{i(j)} + \varepsilon_{ijk}$$

where β_j s are the treatment effects, α_m s are the time effects, $(\alpha\beta)_{jm}$ s are the interactions between time and treatment, $\omega_{i(j)} \stackrel{iid}{\sim} N(0, \sigma_w^2)$ are the whole plot errors and $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$ are split plot error, $\omega_{i(j)}$ and ε_{ijk} are independent. An elegant variance decomposition can be done for balanced data and F-test can be applied to test treatment, time and interaction effects. However, it is only appropriate for balanced data and it has very restrictive assumptions on the covariance structure for the repeated measures on the same subject. See Diggle, Liang and Zeger (1994) for more detail.

Multivariate analysis of variance (MANOVA) is another alternative for ANOVA test for longitudinal data (see Muirhead, 1982). MANOVA is intended to test the effect of treatments on multivariate dependent random vectors. In contrast to univariate ANOVA, it can accommodate more general covariance structure. However, MANOVA requires balanced data and missing values are not allowed. Furthermore, both univariate and multivariate ANOVA are based on normality assumptions and are purely parametric, hence lacking robustness to model longitudinal data. There is also a line of ANOVA methods for longitudinal data based on derived variables. See Fitzmaurice *et al.* (2009) for a review on these methods for longitudinal data.

We considered modelling longitudinal data by semi-linear models, which includes nonparametric functions to increase the flexibility of modelling the time-effect and unbalanced data. More specifically,

$$Y_{ji}(t) = X_{ji}^T(t)\beta_{j0} + M^T(X_{ji}(t), t)\gamma_{j0} + g_{j0}(t) + \varepsilon_{ji}(t), \quad j = 1, 2, \dots, k$$

where $\{Y_{ji}(t), X_{ji}^T(t)\}$ are response and covariates measured at time t , M is a known interaction function and $g_{j0}(t)$ s represent time effects. The ANOVA tests we considered in Chapter 4 contain both ANOVA for β_{j0} s, i.e., the treatment effects with respect to covariates and ANOVA for nonparametric time-effect functions $g_{j0}(t)$ s.

Drop-outs (Missing values) are very common in longitudinal studies and may be due to many reasons. If the drop-out is related to treatment effects, we need to consider missing mechanism

to avoid possible bias in estimating the parameters and nonparametric time-effect functions. Rubin (1976) gave a general discussion on missing mechanism. An overview can be found in Little and Rubin (2002). Three commonly methods are used in handling the missing values in longitudinal data: multiple imputation (Rubin, 1978), likelihood based methods (Dempster, Laird and Rubin, 1977; Little, 1995) and propensity weighted methods (Robins *et al.*, 1995). We adjusted for missing value using the weighted propensity method and extended the monotone missing (Robins *et al.*, 1995) to a more general missing mechanism which can utilize a sequence of observations after some casual drop outs. More details can be found in Chapter 4.

We formulated ANOVA test statistics by using the empirical likelihood, a nonparametric likelihood introduced by Owen (1990). A brief review on empirical likelihood is presented in the next section. We extended the Wilk's theorem to multi-sample case for longitudinal data with missing values. The asymptotic distributions of the ANOVA test statistics for the nonparametric time-effect functions were also given in the paper. But the converge rate based on the asymptotic distribution may be slow. So we further generalized the wild Bootstrap (Wu, 1986) to implement the ANOVA tests for time-effect functions.

1.4 Empirical Likelihood

Empirical likelihood is a nonparametric likelihood introduced by Owen (1990, 2001). Unlike the parametric likelihood, empirical likelihood does not assume any specific underlying distributions, therefore it is more robust than parametric likelihood. Besides it's robustness, it also enjoys good properties as the parametric likelihood. The Wilks' theorem for empirical likelihood, which resembles the Wilks' theorem for the parametric likelihood, was obtained by Owen (1990). The empirical likelihood is also Bartlett correctable (DiDicco and Hall, 1990; Chen and Cui, 2006). For more developments of the empirical likelihood, see Hall and La Scala (1990) for methodology and algorithms, Qin and Lawless (1994) for empirical likelihood for general estimating equations, Kitamura (1997) for dependent data. See also Chen and van Keilegom (2009) for a review on empirical likelihood for regressions.

Let X_1, \dots, X_n be independent and identically distributed random variables with common

distribution F . The empirical likelihood is defined as

$$L_n(F) = \prod_{i=1}^n p_i$$

with $p_i \geq 0$ and $\sum_i p_i = 1$. Notice that $L_n(F)$ attains its maximum when $p_i = 1/n$, which corresponding to $F = F_n$ where F_n is the empirical cumulative distribution of X_i .

Suppose we are interested in making inference for a parameter $\theta(F) \in R^p$ which could be defined by q dimensional estimating equation $g(X; \theta)$ such that $Eg(X; \theta_0) = 0$, where θ_0 is the true value of θ . For example, $\theta(F) = E(X) = \int x dF(x)$ and $g(X; \theta) = X - \theta$ such that $Eg(X; \theta_0) = 0$. The just identifiable and over identified cases corresponding to $q = p$ and $q > p$ respectively. Assuming that 0 is in the convex hull of $g(X_1; \theta), \dots, g(X_n; \theta)$, the empirical likelihood ratio for θ is defined as

$$R(\theta) = \sup \left\{ \prod_{i=1}^n np_i : \sum_{i=1}^n p_i g(X_i; \theta) = 0, \sum_{i=1}^n p_i = 1, p_i \geq 0 \right\}.$$

By using the Langrange multiplier, it can be shown that

$$\ell(\theta) = -2 \log\{R(\theta)\} = 2 \sum_{i=1}^n \log\{1 + \lambda' g(X_i; \theta)\},$$

where λ satisfies

$$\sum_{i=1}^n \frac{g(X_i; \theta)}{1 + \lambda' g(X_i; \theta)} = 0.$$

The maximum empirical likelihood estimate of θ is defined as $\hat{\theta} = \arg \max_{\theta} R(\theta)$. Under some conditions, Qin and Lawless (1994) showed that

$$\ell(\theta_0) - \ell(\hat{\theta}) \xrightarrow{d} \chi_p^2.$$

It is easy to see that if $q = p$, $\hat{\theta}$ is the solution of the empirical version of the estimating equation, i.e., $\frac{1}{n} \sum_{i=1}^n g(X_i; \hat{\theta}) = 0$ which corresponding to $\ell(\hat{\theta}) = 0$. Thus, $\ell(\theta_0) \xrightarrow{d} \chi_p^2$.

In Chapter 4, we consider ANOVA test statistics formulated by empirical likelihood. Suppose we have $X_{j1}, \dots, X_{jn_j} \sim F_j$ for $j = 1, \dots, k$ and our interests is on $\theta_j := \theta(F_j) \in R^p$ which are defined by p dimensional estimating equations ψ_j such that $E\{\psi_j(X_{ji}; \theta_j)\} = 0$. The empirical likelihood ratio for $(\theta_1, \dots, \theta_k)$ is defined as

$$R(\theta_1, \dots, \theta_k) = \sup \left\{ \prod_{j=1}^k \prod_{i=1}^{n_j} n_j p_{ji} : \sum_{i=1}^n p_{ji} \psi_j(X_{ji}; \theta_j) = 0, \sum_{i=1}^n p_{ji} = 1, p_{ji} \geq 0 \right\}.$$

Then an empirical likelihood ratio test statistic for testing $H_0 : \theta_1 = \theta_2 = \dots = \theta_k$ is

$$\begin{aligned} \ell_n &:= -2 \log \left\{ \frac{\max_{\theta_1 = \dots = \theta_k} R(\theta_1, \dots, \theta_k)}{\max_{\theta_1, \dots, \theta_k} R(\theta_1, \dots, \theta_k)} \right\} \\ &= -2 \max_{\theta_1 = \dots = \theta_k} \log\{R(\theta_1, \dots, \theta_k)\} + 2 \max_{\theta_1, \dots, \theta_k} \log\{R(\theta_1, \dots, \theta_k)\} \\ &= 2 \min_{\theta_1 = \dots = \theta_k} \sum_{j=1}^k \sum_{i=1}^{n_j} \log\{1 + \lambda'_j \psi_j(X_{ji}; \theta_j)\} \end{aligned} \quad (1.4.8)$$

where λ_j s satisfy

$$\sum_{i=1}^{n_j} \frac{\psi_j(X_{ji}; \theta_j)}{1 + \lambda'_j \psi_j(X_{ji}; \theta_j)} = 0.$$

Notice that we used $\max_{\theta_1, \dots, \theta_k} \log\{R(\theta_1, \dots, \theta_k)\} = 0$. It is shown in Chapter 4 that $\ell_n \rightarrow \chi_{(k-1)p}^2$ under the null hypothesis. We also consider the ANOVA test for time effect functions, i.e., comparing nonparametric functions, $H_0 : g_1(\cdot) = \dots = g_k(\cdot)$. Suppose for each time t , $E\{\varphi_j(X_{ji}; g_j(t))\} = 0$. Then we can construct an empirical likelihood ratio statistic at each time t ,

$$\mathcal{L}_n\{g(t)\} := 2 \min_{g_1(t) = \dots = g_k(t)} \sum_{j=1}^k \sum_{i=1}^{n_j} \log\{1 + \eta_j \varphi_j(X_{ji}; g_j(t))\} \quad (1.4.9)$$

where η_j s satisfy

$$\sum_{i=1}^{n_j} \frac{\varphi_j(X_{ji}; g_j(t))}{1 + \eta_j \varphi_j(X_{ji}; g_j(t))} = 0.$$

We then use

$$\mathcal{T}_n = \int_0^1 \mathcal{L}_n(g(t)) \varpi(t) dt \quad (1.4.10)$$

as a test statistic for ANOVA test for nonparametric functions where $\varpi(t)$ is a weight function.

The asymptotic distribution of \mathcal{T}_n is also derived in Chapter 4.

1.5 Thesis Organization

Chapter 2 proposes a high dimensional simultaneous test for regression coefficients in linear model, a short version of which is published in the *Journal of American Statistical Association*. This test aims to test the significance of a large number of covariates simultaneously under the

so-called “large p , small n ” situations where the conventional F-test is no longer applicable. We derive the asymptotic distribution of the proposed test statistic under the high dimensional null hypothesis and various scenarios of the alternatives, which allow power evaluations. We further extend the result to linear model with factorial designs. We also evaluate the power of the F-test under very mild dimensionality, namely $p < n$ and $p/n \rightarrow c \in (0, 1)$. The proposed tests are employed to analyze a micro-array data on Yorkshire Gilts to find gene ontology terms which are significantly associated with the thyroid hormone after accounting for the designs of the experiment.

Chapter 3 considers a test for high dimensional means under sparsity and dependency. We propose a threshold test statistic, which is designed to detect sparse and faint signal. The asymptotic distribution is obtained for non normal and dependent data under the “large p , small n ” setting, where the data dimension can grow exponentially fast as the sample size grows. A maximum test, which maximizes the standardized threshold test statistic over a range of thresholds, is also proposed. It is shown that the maximum test can attain the optimal detection boundary, in the sense that asymptotically, all the tests would be powerless below the boundary.

Chapter 4 is a longer version of a paper published in the *Annals of Statistics*, which contains analysis of variance (ANOVA) tests for treatment effects in longitudinal studies with missing values. The treatment effects are modelled semiparametrically via a partially linear regression which is flexible in quantifying the time effects of treatments. The empirical likelihood is employed to formulate model-robust nonparametric ANOVA tests for treatment effects with respect to covariates, the nonparametric time-effect functions and interactions between covariates and time. The proposed tests can be readily modified for a variety of data and model combinations, that encompass parametric, semiparametric and nonparametric regression models; cross-sectional and longitudinal data, and with or without missing values.

References

- Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6, 311-329.
- Bai, Z. and Silverman, J.W. (2006). *Spectral analysis of large dimensional random matrices*. Springer, New York.
- 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*, 57, 289-300.
- Benjamini, Y. and Yekutieli, D. (2001). The control of false discovery rate under dependency. *The Annals of Statistics*, 29, 1165-1188.
- Cai, T. and Sun, W. (2009). Simultaneous testing of grouped hypotheses: finding needles in multiple haystacks. *Journal of the American Statistical Association*, 104, 1467-1481.
- Chen, J. and Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95, 759-771.
- Chen, S. X. and Cui, H.-J. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika*, 93, 215-220.
- Chen, S. X. and Qin, Y. (2010). A two sample test for high dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38, 808-835.
- Chen, S. X. and van Keilegom, I. (2009). A review on empirical likelihood for regressions. *Test*, 18, 415-447.

- Chen, S. X., Zhang, L.X. and Zhong, P.S. (2010). Tests for high dimensional covariance matrices. *Journal of the American Statistical Association*, 109, 810-819.
- Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.
- DiCiccio, T., Hall, P. and Romano, J. (1991). Empirical likelihood is Barterlett-correctable. *The Annals of Statistics*, 19, 1053-1061.
- Diggle, P., Liang, K.Y. and Zeger, S. (1994). *Analysis of longitudinal data*. Oxford University Press, New York.
- Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. Available at <http://www-stat.stanford.edu/~donoho/>.
- Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, 32, 962-994.
- Efron, B., Tibshirani, R., Storey, J. and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, 96, 1151-1160.
- Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association*, 91, 674-688.
- Fan, J. and Hall, P. and Yao, Q. (2007). To how many simultaneous hypothesis tests can normal student's t or bootstrap calibrations be applied. *Journal of the American Statistical Association*, 102, 1282-1288.
- Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20, 101-148.
- Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, 32, 928-961.
- Fitzmaurice, G., Laird, N. and Ware, J. (2004). *Applied longitudinal analysis*. John Wiley & Sons.

- Goeman, J., Sara A. van de G. and van Houwelingen, J. C. (2006). Testing against a high dimensional alternative. *Journal of Royal Statistical Society, Series B*, 68, 477-493.
- Goeman, J., Finos, L. and van Houwelingen, J. C. (2009). Testing against a high dimensional alternative in the generalized linear model: asymptotic alpha-level control. *Manuscript*.
- Greenshtein, E. and Ritov, Y. (2004). Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli*, 10, 971-988.
- Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review*, 58, 109-127.
- Hall, P. and Wang, Q. (2010). Strong approximations of level exceedences related to multiple hypothesis testing. *Bernoulli*, 16, 418-434.
- Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75, 383-386.
- Huber, P. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 1, 799-821.
- John, S. (1971). Some optimal multivariate tests. *Biometrika*, 59 123-127.
- John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika*, 59, 169-173.
- Johnstone, I. M. and Titterton, D. M. (2009). Statistical challenges of high-dimensional data. *Philosophical Transactions of the Royal Society, A*, 367, 4237-4253.
- Kim, M. and Akritas, M. (2010). Order thresholding. *The Annals of Statistics*, 38, 2314-2350.
- Kosorok, M. R. and Ma, S. (2007). Marginal asymptotics for the “large p, small n” paradigm: with applications to microarray data. *The Annals of Statistics*, 35, 1456-1486.
- Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compare to the sample size. *The Annals of Statistics*, 30, 1081-1102.

- Little, R. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90, 1112-1121.
- Little, R. and Rubin, B. (2002). *Statistical analysis with missing data*. Second edition. John Wiley and Sons.
- Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik (N.S.)*, 72, 507-536.
- Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the LASSO. *The Annals of Statistics*, 34, 1436-1462.
- Meinshausen, N., Meier, L. and Bühlmann, P. (2009). p-Values for high-dimensional regression. *Journal of the American Statistical Association*, 104, 1671-1681.
- Muirhead, R. (2005). *Aspects of multivariate statistical theory*. John Wiley and Sons, New York.
- Nagao, H. (1973). On some test criteria for covariance matrix. *The Annals of Statistics*, 1, 700-709.
- Newton, M., Quintana, F., Den Boon, J., 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, 85-106.
- Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16, 1-32.
- Pastur, L. A. (1972). On the spectrum of random matrices. *Theoretical and Mathematical Physics*, 10, 67-74.
- Pastur, L. A. (1973). Spectra of random self adjoint operators. *Russian Math. Survey*, 28, 1-67.
- Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18, 90-120.

- Owen, A. (2001). *Empirical likelihood*. Chapman and Hall/CRC.
- Portnoy, S. (1984). Asymptotic behavior of the M-estimators of p-regression parameters when p^2/n is large: Consistency. *The Annals of Statistics*, 12, 1298-1309.
- Portnoy, S. (1985). Asymptotic behavior of the M-estimators of p-regression parameters when p^2/n is large: Normal approximation. *The Annals of Statistics*, 13, 1403-1417.
- Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families with the number of parameters tends to infinity. *The Annals of Statistics*. 16, 356-366.
- Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*. 22, 300-325.
- Robins, J., Rotnitzky, A. and Zhao, L. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of American Statistical Association*, 90, 106-121.
- Rubin, B. (1976). Inference and missing data (with discussion). *Biometrika*, 63, 581-592.
- Schott, J.R. (2004). A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. *Journal of Multivariate Analysis*, 97, 827-843.
- Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73, 751-754.
- Storey, J. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, 64, 479-498.
- Storey, J. (2003). The positive false discovery rate: A Bayesian interpretation and q-value. *The Annals of Statistics*, 31, 2013-2035.
- Storey, J., Taylor J. 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*, 66, 187-205.

- Storey, J. (2007). The optimal discovery procedure: A new approach to simultaneous significance testing. *Journal of the Royal Statistical Society, Series B*, 69, 347-368.
- Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of Japan Statistical Society*, 35, 251-272.
- Srivastava, M. S. and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis*, 101, 1319-1329.
- 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. and Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102 15545-15550.
- Wasserman, L. and Roeder, K. (2009). High dimensional variable selection. *The Annals of Statistics*, 37, 2178-2201.
- Wu, C.-F.-J. (1986). Jackknife, Bootstrap and other resampling method in regression analysis. *The Annals of Statistics*, 14, 1261-1295.
- Yohai, V.J. and Marorum, R.A. (1979). Asymptotic behavior of M-estimates for the linear model. *The Annals of Statistics* 7, 258-268.
- Zhao, P. and Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7, 2541-2563.

CHAPTER 2. Tests for High Dimensional Regression Coefficients with Factorial Designs

A paper published in *The Journal of American Statistical Association*, 106, 260-274.

Pingshou Zhong and Song Xi Chen

Department of Statistics

Iowa State University

Ames, IA 50011, USA

Abstract

We propose simultaneous tests for coefficients in high dimensional linear regression models with factorial designs. The proposed tests are designed for the “large p , small n ” situations where the conventional F-test is no longer applicable. We derive the asymptotic distribution of the proposed test statistic under the high dimensional null hypothesis and various scenarios of the alternatives, which allow power evaluations. We also evaluate the power of the F-test under very mild dimensionality. The proposed tests are employed to analyze a micro-array data on Yorkshire Gilts to find significant gene ontology terms which are significantly associated with the thyroid hormone after accounting for the designs of the experiment.

KEY WORDS: Factorial Design; Gene-set test; High dimensional regression; Large p , small n ; U-statistics.

2.1 Introduction

The emergence of high-dimensional data, such as the gene expression values in microarray and the single nucleotide polymorphism (SNP) data, brings challenges to many traditional

statistical methods and theory. One important aspect of the high-dimensional data under the regression setting is that the number of covariates greatly exceeds the sample size. For example, in microarray data, the number of genes (p) is in the order of thousands whereas the sample size (n) is much less, usually less than fifty due to limitation to replicate. This is the so-called “large- p , small- n ” paradigm, which translates to a regime of asymptotics where $p \rightarrow \infty$ much faster than n . See Kosorok and Ma (2007), Fan, Hall and Yao (2007), Huang, Wang and Zhang (2007), Chen and Qin (2010) among others. Kosorok and Ma (2007) considered uniform convergence for a large number of marginal discrepancy measures targeted on univariate distributions, means and medians. Chen and Qin (2010) proposed a two sample test on high dimensional means. Both of these two aforementioned papers considered testing under “large- p , small- n ” without a regression structure, which is the focus of the present paper. Much earlier, for more moderate dimensions, Portnoy (1984, 1985) had considered consistency and asymptotic normality for the M-estimators of linear regression coefficients when the dimension p of the covariates grows to infinity faster than the square root of the sample size n . The rates for p that Portnoy considered were $p = o(n/\log(p))$ for consistency and $p = o(n^{2/3}/\log(p))$ for asymptotic normality of the M-estimators.

Covariate selection for high dimensional linear regression has attracted much attention and has been intensively considered in recent years. Penalizing methods are alternatives to the traditional least square estimator for simultaneous variable selection and shrinkage estimation. These include the LASSO (Tibshirani, 1996) with a L_1 -penalty, the bridge regression with a L_2 -penalty (Frank and Friedman, 1996), the SCAD penalty proposed by Fan and Li (2001) and Candes and Tao (2007)’s Dantzig selector; see also Fan and Lv (2008) and Wang (2009) for other methods of variable selection. There is also a line of works on ANOVA with diverging number of treatments while the number of replications (cell sample sizes) is small and can be regarded as fixed. This includes the rank based nonparametric tests proposed by Brownie and Boos (1994), Boos and Brownie (1995), Akritas and Arnold (2000), Bathke and Lankowski (2005), Bathke and Harrar (2008), Harrar and Bathke (2008), and Wang and Akritas (2009). The problem can be viewed as “large p , fixed n ” in contrast to the conventional “fixed p , large n ” setting and the “large p , small n ” paradigm we are considering.

This chapter is aimed at developing simultaneous tests on linear regression coefficients that can accommodate high dimensionality and factorial designs. The latter is often encountered in statistical experiments especially those in biology, and there is no exception for high dimensional data. Testing hypotheses on the regression coefficients is a necessity in determining the effects of covariates on certain outcome variable. Our interest here is on testing the significance of a large number of covariates simultaneously. This is motivated by the latest need in biology to identify significant sets of genes (Subramanian *et al.*, 2005; Efron and Tibshirani, 2007; Newton *et al.*, 2007), which are associated with certain clinical outcome, rather than identifying individual gene. As the dimension of a gene-set ranges from a few to thousands, and the gene-sets can overlap as they share common genes, there are both high dimensionality and multiplicity in gene-set testing. In order to test for the significance of a gene-set, the P-value associated with a hypothesis regarding the regression coefficient corresponding to the gene-set is needed. This calls for multivariate tests for regression coefficients that can accommodate both high dimensionality and dependence among the covariates.

We propose tests for high dimensional regression coefficients for both simple random or factorial designs. A feature of the tests is that they do not require explicit relationships between the growth rates of p and n , which makes the tests adaptable to a wide range of high dimensionality. The tests also account for a variety of dependence among the high dimensional covariate. These together with their accommodation to factorial designs makes the tests more applicable in applications. The F-test is the conventional test for regression coefficients simultaneously under the normality and $p < n - 1$. We take the opportunity to study the F-test and find that it is adversely affected by an increasing dimension.

The chapter is organized as follows. We first study the F-test and propose a new test statistic in Section 2.2 for simple random designs. Section 2.3 discusses some general properties of U-statistics under high dimensionality. Section 2.4 establishes the main properties of the proposed test. Extensions to factorial designs are made in Section 2.5. Section 2.6 reports results from simulation studies. Empirical analyses on a microarray dataset on Yorkshire Gilts with factorial designs are reported in Section 2.7. All technical details are relegated to the Appendix.

2.2 Models and Test Statistics

Consider a linear regression model

$$E(Y_i|X_i) = \alpha + X_i' \beta \quad \text{and} \quad \text{Var}(Y_i|X_i) = \sigma^2 \quad (2.2.1)$$

for $i = 1, \dots, n$ where X_1, \dots, X_n are independent and identically distributed p -dimensional covariates and Y_1, \dots, Y_n are independent responses, β is the vector of regression coefficients, and α is a nuisance intercept. We do not impose any specific distribution on Y_i given X_i except when studying the F-test in the next subsection.

The true parameter (α, β) in the linear regression model is defined as

$$(\alpha, \beta) = \arg \min_{\tilde{\alpha} \in R^1, \tilde{\beta} \in R^p} E(Y_i - \tilde{\alpha} - X_i' \tilde{\beta})^2.$$

To make β identifiable, we assume that $\Sigma = \text{Var}(X_i) > 0$. This is weaker than the sparse Riesz condition in Zhang and Huang (2008), which requires the eigenvalues of Σ are all bounded from below and above. The sparse Riesz condition is for the purpose of parameter estimation and variable selection, which are different from the agenda of this Chapter.

Our interest is in testing a high dimensional hypothesis

$$H_0 : \beta = \beta_0 \quad \text{vs} \quad H_1 : \beta \neq \beta_0 \quad (2.2.2)$$

for a specific $\beta_0 \in R^p$. For instance $\beta_0 = 0$ which arises in the context of gene-set testing with H_0 indicating a particular set of genes to be insignificant.

2.2.1 F-test and Its Performances under High Dimensionality

When the conditional distribution of Y_i given X_i is normally distributed, the conventional test for (2.2.2) is the F-test when $p < n - 1$. The F-statistic is a monotone function of the likelihood ratio statistic and is distributed as a non-central F distribution under the alternative (Anderson, 2003). It is interesting to know the power implication on the F-test when $p/n \rightarrow \rho \in (0, 1)$ when both p and n diverge to infinity.

Let $U = (\mathbf{1}, X)$ which is assumed to be of full rank and $A = (\mathbf{0}, I_p)$, where $\mathbf{1}$ denotes the n -dimensional vector of 1's. Let $\gamma^T = (\alpha, \beta^T)$ and $\gamma_0^T = (\alpha, \beta_0^T)$, then the null hypothesis in

(2.2.2) becomes $H_0 : A\gamma = A\gamma_0$. The F statistic for testing H_0 (Rao *et al.*, 2008, p51) is

$$\begin{aligned} G_{n,p} &= \frac{(\hat{\gamma} - \gamma_0)' A' (A(U'U)^{-1}A')^{-1} A (\hat{\gamma} - \gamma_0)/p}{Y'(I_n - P_U)Y/(n-p-1)} \\ &= \frac{(\hat{\beta} - \beta_0)' (A(U'U)^{-1}A')^{-1} (\hat{\beta} - \beta_0)/p}{Y'(I_n - P_U)Y/(n-p-1)} \end{aligned} \quad (2.2.3)$$

where $\hat{\gamma} = (\hat{\alpha}, \hat{\beta})' = (U'U)^{-1}U'Y$ is the least square estimator of γ and $Y = (Y_1, \dots, Y_n)'$. Under H_0 , $G_{n,p} \sim F_{p, n-p-1}$. Hence, an α -level F-test rejects H_0 if $G_{n,p} > F_{p, n-p-1; \alpha}$, the upper α quantile of the $F_{p, n-p-1}$ distribution.

In this chapter, we use I_m to denote the $m \times m$ identity matrix and $\Phi(\cdot)$ as the distribution function of $N(0,1)$. To facilitate our analysis, like Bai and Saranadasa (1996), we assume that

$$\begin{aligned} &\text{There exists a } m\text{-variate random vector } Z_i = (Z_{i1}, \dots, Z_{im})' \text{ for} \\ &\text{some } m \geq p \text{ so that } X_i = \Gamma Z_i + \mu, \text{ where } \Gamma \text{ is a } p \times m \text{ matrix such that } \Gamma\Gamma' = \Sigma, \text{ and } E(Z_i) = 0, \text{Var}(Z_i) = I_m; \text{ each } Z_{il} \\ &\text{has finite 8-th moment, } E(Z_{il}^4) = 3 + \Delta \text{ for some constant } \Delta; \\ &\text{for any } \sum_{\nu=1}^d \ell_\nu \leq 8 \text{ and } i_1 \neq \dots \neq i_d, E(Z_{1i_1}^{\ell_1} Z_{1i_2}^{\ell_2} \dots Z_{1i_d}^{\ell_d}) = \\ &E(Z_{1i_1}^{\ell_1})E(Z_{1i_2}^{\ell_2}) \dots E(Z_{1i_d}^{\ell_d}). \end{aligned} \quad (2.2.4)$$

Model (2.2.4) resembles a factor model where the p -variate X is linearly generated by a m -variate factor Z . However, unlike the factor model which assumes far less number of factors than p so as to achieve a dimension reduction, we assume here the number of factors m is at least as larger as p . Model (2.2.4) slightly differs from the one assumed in Bai and Saranadasa (1996) in relaxing their assumption of Z_i having independent components. We also require the existence of the 8-th moments for Z_i .

The power property of the F-test when $p/n \rightarrow \rho \in (0, 1)$ is depicted in the following theorem.

Theorem 1 *Assume $Y_i|X_i \sim N(X_i'\beta, \sigma^2)$, Model (2.2.4), $(\beta - \beta_0)'\Sigma(\beta - \beta_0) = o(1)$ and $\rho_n = p/n \rightarrow \rho \in (0, 1)$ as $n \rightarrow \infty$ then $\Omega_F(\|\beta - \beta_0\|)$, the power of the F-test, satisfies*

$$\Omega_F(\|\beta - \beta_0\|) - \Phi\left(-z_\alpha + \sqrt{\frac{(1-\rho)n}{2\rho}}(\beta - \beta_0)'\Sigma(\beta - \beta_0)\right) \rightarrow 0. \quad (2.2.5)$$

We notice that the denominator of the F statistic (2.2.3) estimates σ^2 . When p is closer to n , there are fewer degrees of freedom left to estimate σ^2 . The impact of the dimensionality on

the F-test is revealed in Theorem 1 by $\sqrt{(1-\rho)/\rho}$ being a decreasing function of ρ . Hence, the power is adversely impacted by an increased dimension even $p < n - 1$, reflecting a reduced degree of freedom in estimating σ^2 when the dimensionality is close to the sample size.

2.2.2 A New Test Statistic

We have seen two limitations with the F-test under mild dimensionality above. One is that p can not be larger than $n - 1$; and the other is the conditional normality assumption. To test for regression coefficients in the “large p, small n” paradigm without the normality assumption, we modify the F-statistic in two aspects. One is to remove the denominator as it is a major contributor to F-test’s fragile power performance under even mild dimensionality as shown in Theorem 1. Another is to renovate the numerator to make it more effective in measuring the discrepancy between β and β_0 . We note that when $\alpha = 0$, $\|Y - X\beta_0\|^2$ is a measure between β and β_0 , whose expectation is $(\beta - \beta_0)'E(X'X)(\beta - \beta_0) + n\sigma^2$. To avoid the $n\sigma^2$ term, we consider $(Y_i - X_i'\beta_0)(Y_j - X_j'\beta_0)$ for $i \neq j$ and a U-statistic with $X_i'X_j(Y_i - X_i'\beta_0)(Y_j - X_j'\beta_0)$ as the kernel. Our proposal here is similar to the effort made in improving the Wald type F-statistics as demonstrated in Brunner, Dette and Munk (1997) and Ahmad, Brunner and Werner (2008).

When the nuisance parameter $\alpha \neq 0$, to remove α , we consider a U-statistic

$$T_{n,p} = \frac{1}{P_n^4} \sum_n^* \phi(i_1, i_2, i_3, i_4), \quad (2.2.6)$$

$$\text{where } \phi(i_1, i_2, i_3, i_4) = \frac{1}{4}(X_{i_1} - X_{i_2})'(X_{i_3} - X_{i_4})\Delta_{i_1, i_2}\Delta_{i_3, i_4} \quad (2.2.7)$$

and $\Delta_{i,j} = Y_i - Y_j - (X_i - X_j)'\beta_0$. Through this Chapter, we use \sum^* to denote summations over distinct indices. For example, in (2.2.6), the summation is over the set $\{i_1 \neq i_2 \neq i_3 \neq i_4, \text{ for } i_1, i_2, i_3, i_4 \in \{1, \dots, n\}\}$ and $P_n^m = n!/(n-m)!$. As $T_{n,p}$ is invariant to location shifts in both X_i and Y_i . We assume, without loss of generality, that $\alpha = \mu = 0$ in the rest of this Chapter.

The set of conditions we use to regulate for the “large p, small n” is

$$p(n) \rightarrow \infty \text{ as } n \rightarrow \infty, \Sigma > 0 \text{ and } tr(\Sigma^4) = o\{tr^2(\Sigma^2)\}. \quad (2.2.8)$$

These conditions do not impose any explicit relative growth rates between p and n , and they are quite mild. Assuming Σ being positive definite assures the identification of the regression coefficient. We allow some eigenvalues of Σ diverge to infinity as $p \rightarrow \infty$. If all the eigen-values are bounded, the last part of (2.2.8) is trivially true for any p .

2.3 U-Statistics under High Dimensionality

As $T_{n,p}$ is a U-statistic, we devote this section to discuss U-statistics for high dimensional data. The theory of U-statistics for fixed dimensional data, as pioneered by Hoeffding (1948), has been well documented; see Serfling (1980) and Lee (1990) for summaries. We will demonstrate below that, while some results in the classical U-statistic remain valid, others may not be directly applicable if p diverges.

Suppose W_1, W_2, \dots, W_n are independent and identically distributed observations from a distribution F on R^q , where q may diverge. Consider a U-statistic of s -th order for a fixed $s < n$

$$U_{n,q} = \frac{1}{\binom{n}{s}} \sum_{C_{n,s}} h(W_{i_1}, \dots, W_{i_s}),$$

where $C_{n,s} = \{\text{all distinct combinations of } \{i_1, i_2, \dots, i_s\} \text{ from } \{1, \dots, n\}\}$. The kernel h is symmetric so that its value is invariant to the permutations of its arguments. Let $E\{h(W_1, \dots, W_s)\} = \theta(F)$, say. In our current testing problem, $q = p + 1$, $s = 4$ and $\theta(F) = \|\Sigma(\beta - \beta_0)\|^2$.

Let $h_c(w_1, \dots, w_c) = E\{h(w_1, \dots, w_c, W_{c+1}, \dots, W_s)\}$ be projections of h to lower dimensional sample spaces, $\tilde{h} = h - \theta(F)$ and $\tilde{h}_c = h_c - \theta(F)$ for $c = 1, \dots, s$. Let $g_c(w_1, \dots, w_c) = \tilde{h}_c - \sum_{j=1}^{c-1} \sum_{1 \leq i_1 < \dots < i_j \leq c} g_j(w_{i_1}, \dots, w_{i_j})$ where $g_1(w_1) = \tilde{h}_1(w_1)$, and

$$M_{nc} = \sum_{1 \leq i_1 < \dots < i_c \leq n} g_c(w_{i_1}, \dots, w_{i_c}).$$

The following theorem provides the Hoeffding decompositions (Hoeffding, 1948) for $U_{n,q}$ and its variance respectively, which are valid regardless of q being fixed or diverging.

Proposition 1 *Assume $E\{h^2(W_1, \dots, W_s)\}$ exist and let $\zeta_c = \text{Var}(h_c)$ for $c = 1, 2, \dots, s$.*

Then (i) $\zeta_{c+1} \geq \zeta_c$; (ii)

$$U_{n,q} - \theta(F) = \sum_{c=1}^s \binom{s}{c} \binom{n}{c}^{-1} M_{nc} \quad (2.3.9)$$

and (iii)

$$\text{Var}(U_{n,q}) = \binom{n}{s}^{-1} \sum_{c=1}^s \binom{s}{c} \binom{n-s}{s-c} \zeta_c. \quad (2.3.10)$$

The proof in Hoeffding (1948)(see also Serfling, 1980) is applicable even when q is increasing to infinity. Specifically, the result in (i) is implied by $E\{h_{c+1}(w_1, \dots, w_c, W_{c+1})\} = h_c(w_1, \dots, w_c)$ and

$$\zeta_{c+1} = E\{\text{Var}(h_{c+1}(W_1, \dots, W_{c+1})|W_1, \dots, W_c)\} + \zeta_c.$$

The variance decomposition for the variance in (2.3.10) reflects the decomposition of the U-statistic in (2.3.9) as $\{M_{nc}, \mathcal{F}_c\}_{c \geq 1}$ forms a forward martingale where \mathcal{F}_c denotes the σ -field generated by $\{W_1, \dots, W_c\}$ and $\text{Var}(M_{nc}) = O(\zeta_c)$.

When $q \rightarrow \infty$, unlike the fixed dimension cases, ζ_c may no longer be bounded and can diverge. This brings ambiguity in assessing the relative orders of terms in the decomposition (2.3.9). To appreciate this point, we note that if q is fixed, all ζ_c are bounded provided $\zeta_s < \infty$, hence the $(c+1)$ -th term in the variance decomposition (2.3.10) is a smaller order of the c -th term. This means that the asymptotic behavior of the U-statistic is determined by the c -th term where c is the smallest integer such that $\zeta_c \neq 0$. However, if q diverges, ζ_c may diverge and a higher order projection $M_{n(c+1)}$ may be at the same order or higher than M_{nc} . Hence, for high dimensional data, the leading order terms of the U-statistics may consist of multiple terms.

As ζ_c is monotone non-decreasing, the following strategy may be applied to determine the dominant terms of $U_{n,q}$. We can start evaluating ζ_c s from the two ends, namely ζ_1 and ζ_s . If ζ_1 and ζ_s are of the same order, then $U_{n,q}$ will be dominated by the first term so that

$$U_{n,q} - \theta(F) = \binom{s}{1} \binom{n}{1}^{-1} M_{n1} \{1 + o_p(1)\}.$$

If ζ_s and ζ_1 are not the same order, but ζ_2 and ζ_s are, then $U_{n,q}$ will be dominated by the first two terms so that

$$U_{n,q} - \theta(F) = \sum_{c=1}^2 \binom{s}{c} \binom{n}{c}^{-1} M_{nc} \{1 + o_p(1)\}.$$

This process can be continued until the dominating terms are found. We will employ this strategy on the proposed test statistic $T_{n,p}$ in the next section.

2.4 Main Results

We first symmetrize ϕ defined in (2.2.7) by

$$h(W_i, W_j, W_k, W_l) = \frac{1}{3}\{\phi(i, j, k, l) + \phi(i, k, j, l) + \phi(i, l, j, k)\}$$

where $W_i = (X_i^\tau, \varepsilon_i)^\tau$ and $\varepsilon_i = Y_i - X_i' \beta_0$. Then,

$$T_{n,p} = \frac{1}{\binom{n}{4}} \sum_{C_{n,4}} h(W_i, W_j, W_k, W_l). \quad (2.4.11)$$

It can be shown that the projections of h are, respectively,

$$\begin{aligned} h_1(w_1) &= \frac{1}{2}(\beta - \beta_0)'(x_1 x_1' + \Sigma)\Sigma(\beta - \beta_0) + \frac{1}{2}\varepsilon_1 x_1' \Sigma(\beta - \beta_0), \\ h_2(w_1, w_2) &= \frac{1}{6}\left\{(\beta - \beta_0)'(x_1 - x_2)(x_1 - x_2)' \Sigma(\beta - \beta_0) \right. \\ &\quad + (\varepsilon_1 - \varepsilon_2)(x_1 - x_2)' \Sigma(\beta - \beta_0) \\ &\quad \left. + ((\beta - \beta_0)'(x_1 x_1' + \Sigma) + \varepsilon_1 x_1')(\varepsilon_2 x_2 + (x_2 x_2' + \Sigma)(\beta - \beta_0))\right\} \end{aligned}$$

and

$$\begin{aligned} h_3(w_1, w_2, w_3) &= \frac{1}{12}\{(x_1 - x_2)'(\beta - \beta_0) + (\varepsilon_1 - \varepsilon_2)\}(x_1 - x_2)' \{(x_3 x_3' + \Sigma)(\beta - \beta_0) + x_3 \varepsilon_3\} \\ &\quad + \frac{1}{12}\{(x_1 - x_3)'(\beta - \beta_0) + (\varepsilon_1 - \varepsilon_3)\}(x_1 - x_3)' \{(x_2 x_2' + \Sigma)(\beta - \beta_0) + x_2 \varepsilon_2\} \\ &\quad + \frac{1}{12}\{(x_2 - x_3)'(\beta - \beta_0) + (\varepsilon_2 - \varepsilon_3)\}(x_2 - x_3)' \{(x_1 x_1' + \Sigma)(\beta - \beta_0) + x_1 \varepsilon_1\}. \end{aligned}$$

Let $B_i = (\beta - \beta_0)' \Sigma^i (\beta - \beta_0)$ for $i = 1, 2, 3$, $A_0 = \Gamma' \Gamma$, $A_1 = \Gamma'(\beta - \beta_0)(\beta - \beta_0)' \Gamma$, $A_2 = \Gamma' \Sigma(\beta - \beta_0)(\beta - \beta_0)' \Sigma \Gamma$ and $A_3 = \Gamma' \Sigma \Gamma$. Derivations given in the Appendix show that $\zeta_1 = \frac{1}{4} \zeta_1^*$ and $\zeta_2 = \frac{1}{36} \zeta_2^*$ where

$$\begin{aligned} \zeta_1^* &= (B_1 + \sigma^2)B_3 + B_2^2 + \Delta \text{tr}(A_1 \circ A_2) \quad \text{and} \\ \zeta_2^* &= \sigma^4 \text{tr}(\Sigma^2) + 21B_2^2 + 22B_1 B_3 + 22\sigma^2 B_3 + B_1^2 \text{tr}(\Sigma^2) + 2\sigma^2 \text{tr}(\Sigma^2) B_1 \\ &\quad + 2\Delta(B_1 + \sigma^2) \text{tr}(A_1 \circ A_3) + 20\Delta \text{tr}(A_1 \circ A_2) + \Delta^2 \text{tr}\{(A_0 \text{diag}(A_1))^2\}, \end{aligned}$$

where $C \circ B = (c_{ij} b_{ij})$ for matrices $C = (c_{ij})$ and $B = (b_{ij})$, and $\text{diag}(A) = \text{diag}\{a_{11}, \dots, a_{mm}\}$ for $A = (a_{ij})_{m \times m}$. The proof of the following theorem in the Appendix shows that $\{\zeta_c\}_{c=2}^4$ are of the same order. This means that the test statistic is dominated by the first two terms corresponding M_{n1} and M_{n2} .

Theorem 2 Under Model (2.2.4) and as $n \rightarrow \infty$,

$$(i) E(T_{n,p}) = \|\Sigma(\beta - \beta_0)\|^2 \text{ and } \text{Var}(T_{n,p}) = \left\{ \frac{4}{n}\zeta_1^* + \frac{2}{n(n-1)}\zeta_2^* \right\} \{1 + o(1)\};$$

$$(ii) T_{n,p} - \|\Sigma(\beta - \beta_0)\|^2 = \left\{ \frac{4^2}{n}M_{n1} + \frac{2 \times 6^2}{n(n-1)}M_{n2} \right\} \{1 + o_p(1)\}, \text{ where } E(M_{n1}^2) = \zeta_1 \text{ and } E(M_{n2}^2) = \zeta_2 - 2\zeta_1.$$

Under $H_0 : \beta = \beta_0$, $A_1 = A_2 = B_i = 0$ for $i = 1, 2, 3$. Thus, $\zeta_1 = 0$ and $T_{n,p}$ is a degenerate U-statistic dominated by M_{n2} . In this case,

$$\text{Var}(T_{n,p}) = \frac{2}{n(n-1)}\sigma^4 \text{tr}(\Sigma^2) \{1 + o(1)\}.$$

This form of the variance for $T_{n,p}$ is also valid under a subclass of H_1 specified by

$$(\beta - \beta_0)' \Sigma(\beta - \beta_0) = o(1) \quad \text{and} \quad (2.4.12)$$

$$(\beta - \beta_0)' \Sigma^3(\beta - \beta_0) = o\{n^{-1} \text{tr}(\Sigma^2)\}.$$

As this subclass prescribes a smaller difference between β and β_0 , we call it the local alternatives. Under the local alternatives, $\zeta_1 = o(n^{-1}\zeta_2)$ which means like the case under H_0 , M_{n2} is also the dominating term while M_{n1} is of smaller order.

Theorem 3 Assume Model (2.2.4) and Condition (2.2.8), then under either H_0 or the local alternatives (2.4.12), as $n \rightarrow \infty$,

$$\frac{n}{\sigma^2 \sqrt{2 \text{tr}(\Sigma^2)}} (T_{n,p} - \|\Sigma(\beta - \beta_0)\|^2) \xrightarrow{d} N(0, 1). \quad (2.4.13)$$

To formulate a test procedure based on $T_{n,p}$, we need to estimate $\text{tr}(\Sigma^2)$ and σ^2 appeared in the asymptotic variance. We will use the estimator of $\text{tr}(\Sigma^2)$ proposed in Chen, Zhang and Zhong (2010). Specifically, let $Y_{1n} = \frac{1}{P_n^2} \sum^* (X'_{i_1} X_{i_2})^2$, $Y_{2n} = \frac{1}{P_n^3} \sum^* X'_{i_1} X_{i_2} X'_{i_2} X_{i_3}$ and $Y_{3n} = \frac{1}{P_n^4} \sum^* X'_{i_1} X_{i_2} X'_{i_3} X_{i_4}$. Then an unbiased and ratio consistent estimator of $\text{tr}(\Sigma^2)$ is

$$\widehat{\text{tr}(\Sigma^2)} = Y_{1n} - 2Y_{2n} + Y_{3n}.$$

We note here that a closely related estimator, that only employs $Y_{1,n}$, has been proposed in Ahmad, Werner and Brunner (2008) for normally distributed X_i with zero mean. The estimator of σ^2 under H_0 is

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - X_i \beta_0 - \bar{Y} + \bar{X} \beta_0)^2. \quad (2.4.14)$$

Applying Theorem 3 and the Slutsky Theorem, the proposed test rejects H_0 at a significant level α if

$$nT_{n,p} \geq \sqrt{2\text{tr}(\Sigma^2)}\delta^2 z_\alpha, \quad (2.4.15)$$

where z_α is the upper- α quantile of $N(0, 1)$.

Theorem 3 also implies that $\Omega_L(\|\beta - \beta_0\|)$, the asymptotic power of the proposed test under the local alternatives is

$$\Omega_L(\|\beta - \beta_0\|) \doteq \Phi\left(-z_\alpha + \frac{n\|\Sigma(\beta - \beta_0)\|^2}{\sqrt{2\text{tr}(\Sigma^2)}\sigma^2}\right). \quad (2.4.16)$$

The power is largely impacted by $\eta_n(\beta - \beta_0, \Sigma, \sigma^2) = n\|\Sigma(\beta - \beta_0)\|^2/\{\sqrt{2\text{tr}(\Sigma^2)}\sigma^2\}$, which may be viewed as a signal to noise ratio (SNR). In particular, the power converges to α if $\eta_n(\beta - \beta_0, \Sigma, \sigma^2) = o(1)$ which means that the test can not distinguish H_0 from the local alternative in this case. If it is of a larger order of 1, the power converges to 1, indicating consistency of the test.

Let $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p$ be the eigenvalues of Σ . Then, a sufficient condition for the test to have a non-trivial power is $\|\beta - \beta_0\| = O(n^{-1/2}S_\lambda^{1/4}\lambda_1^{-1})$ where $S_\lambda = \sum_{i=1}^p \lambda_i^2$. Suppose all the eigenvalues are bounded from zero and infinity, let $\delta_\beta = \|\beta - \beta_0\|/\sqrt{p}$ define ‘‘signal strength’’, then the test has non-trivial power if δ_β is of order $n^{-1/2}p^{-1/4}$. This is a smaller order than $n^{-1/2}$, the corresponding ‘‘signal’’ strength for the fixed dimensional case.

We can also evaluate power of the proposed test under other scenarios of H_1 such that

$$(\beta - \beta_0)'\Sigma(\beta - \beta_0) \text{ is not } o(1) \quad (2.4.17)$$

violating the first part of (2.4.12) in the specification of the local alternatives. We consider two scenarios of fixed alternatives under (2.4.17). One is

$$(\beta - \beta_0)'\Sigma^3(\beta - \beta_0) = o\left\{\frac{1}{n}(\beta - \beta_0)'\Sigma(\beta - \beta_0)\text{tr}(\Sigma^2)\right\}, \quad (2.4.18)$$

which complements (2.4.12). If $(\beta - \beta_0)'\Sigma(\beta - \beta_0)$ is truly bounded, (2.4.18) implies $(\beta - \beta_0)'\Sigma^3(\beta - \beta_0) = o\left\{\frac{1}{n}\text{tr}(\Sigma^2)\right\}$ which mimics the second part of (2.4.12).

A complement to both (2.4.12) and (2.4.18) is

$$\frac{1}{n}(\beta - \beta_0)'\Sigma(\beta - \beta_0)\text{tr}(\Sigma^2) = o\left\{(\beta - \beta_0)'\Sigma^3(\beta - \beta_0)\right\}. \quad (2.4.19)$$

If $(\beta - \beta_0)' \Sigma (\beta - \beta_0)$ is bounded, (2.4.19) implies $\frac{1}{n} \text{tr}(\Sigma^2) = o\{(\beta - \beta_0)' \Sigma^3 (\beta - \beta_0)\}$, which prescribes a larger discrepancies between β and β_0 . Without causing much confusion, we call both (2.4.18) and (2.4.19) under (2.4.17) as fixed alternatives.

To quantify the asymptotic power, we define

$$\begin{aligned} \sigma_{A_1}^2 &= 2\sigma^4 \text{tr}(\Sigma^2) + 2B_1^2 \text{tr}(\Sigma^2) + 4\sigma^2 \text{tr}(\Sigma^2) B_1 + 4\Delta (B_1 + \sigma^2) \text{tr}(A_1 \circ A_3) \\ &\quad + 2\Delta^2 \text{tr}\{(A_0 \text{diag}(A_1))^2\} \quad \text{and} \\ \sigma_{A_2}^2 &= (B_1 + \sigma^2) B_3 + B_2^2 + \Delta \text{tr}(A_1 \circ A_2). \end{aligned}$$

We note that $\sigma_{A_1}^2$ is part of the variance of M_{n2} , where we only keep the leading order terms under (2.4.18) and $\sigma_{A_2}^2$ is the same as ζ_1 , the variance of M_{n1} up to a constant.

Theorem 4 *Assume Model (2.2.4), Conditions (2.2.8) and (2.4.17), then (i) under the first fixed alternatives (2.4.18)*

$$\frac{n}{\sigma_{A_1}} (T_{n,p} - \|\Sigma(\beta - \beta_0)\|^2) \xrightarrow{d} N(0, 1); \quad (2.4.20)$$

(ii) under the second fixed alternatives (2.4.19)

$$\frac{\sqrt{n}}{\sigma_{A_2}} (T_{n,p} - \|\Sigma(\beta - \beta_0)\|^2) \xrightarrow{d} N(0, 1). \quad (2.4.21)$$

The theorem implies that the asymptotic power of the test under the first fixed alternatives (2.4.18) is

$$\Omega_{H_1}(\|(\beta - \beta_0)\|) \doteq \Phi \left(-\frac{\sqrt{2\text{tr}(\Sigma^2)} \sigma^2 z_\alpha}{\sigma_{A_1}} + \frac{n \|\Sigma(\beta - \beta_0)\|^2}{\sigma_{A_1}} \right). \quad (2.4.22)$$

Since B_1 is not $o(1)$ and $\sigma_{A_1}^2 > 2B_1^2 \text{tr}(\Sigma^2)$, the first term $\sqrt{2\text{tr}(\Sigma^2)} \sigma^2 z_\alpha / \sigma_{A_1} < \sigma^2 z_\alpha / B_1$ is always bounded from infinity. In particular, if B_1 diverges to ∞ , the first term converges to 0. Hence, the test attains at least 50% power in this case. If $n \|\Sigma(\beta - \beta_0)\|^2 / \sigma_{A_1} \rightarrow \infty$, the power converges to 1.

The asymptotic power under the second fixed alternatives (2.4.19) is

$$\Omega_{H_2}(\|(\beta - \beta_0)\|) \doteq \Phi \left(-\frac{\sqrt{2\text{tr}(\Sigma^2)} \sigma^2 z_\alpha}{\sqrt{(n-1)\sigma_{A_2}^2}} + \frac{\sqrt{n} \|\Sigma(\beta - \beta_0)\|^2}{\sigma_{A_2}} \right).$$

As (2.4.19) implies $\frac{1}{n}tr(\Sigma^2)/\sigma_{A_2}^2 = o(1)$, the proposed test is consistent as long as

$$\sqrt{n}\|\Sigma(\beta - \beta_0)\|^2/\sigma_{A_2} \rightarrow \infty. \quad (2.4.23)$$

Even if $\sqrt{n}\|\Sigma(\beta - \beta_0)\|^2/\sigma_{A_2}$ does not converge to ∞ , the power is still at least 50% asymptotically. The power of the test under the fixed alternatives attains at least 50% power is assuring and it can be shown that the proposed test is more powerful under two fixed alternatives than the local alternative if all the eigenvalues are of the same order. It is also the reason that we call the two alternatives in (2.4.18) and (2.4.19) as fixed alternatives. It may be shown that a sufficient condition for (2.4.23) is $\lambda_p/\lambda_1 = o(n)$.

2.5 Generalization to Factorial Designs

So far we have assumed that $\{(X_i, Y_i)\}_{i=1}^n$ is a simple random sample. However, in many scientific studies, observations are obtained via certain designs of experiments. For example, a randomized factorial design was used in a micro-array study that we will analyze in the next section. In this section, we provide an extension of the proposed high dimensional regression test to accommodate factorial designs.

For ease of expedition, we will concentrate on two way factorial designs with two factors A and B, where A has I levels and B has J levels. Let c indicate a cell for $c = 1, \dots, IJ$, which has n_c observations in the cell. The observations (X'_{ijk}, Y_{ijk}) in the i -th level of A and j -th level of B satisfy a linear model

$$E(Y_{ijk}|X_{ijk}) = \alpha_0 + \gamma_i + \theta_j + \gamma\theta_{ij} + X'_{ijk}\beta, \quad k = 1, \dots, n_c, \quad (2.5.24)$$

where γ_i represent for the effect of A, θ_j for that of B, and $\gamma\theta_{ij}$ for their interactions. These effects could be either random effects or fixed effects. Our purpose in this section is to generalize the test given in Section 2.4 for

$$H_0 : \beta = \beta_0 \quad \text{vs} \quad H_1 : \beta \neq \beta_0 \quad (2.5.25)$$

for Model (2.5.24) while treating $(\alpha_0, \gamma_i, \theta_j, \gamma\theta_{ij})$ as nuisance parameters.

Let $\mu_{ij} = \alpha_0 + \gamma_i + \theta_j + \gamma\theta_{ij}$. Model (2.5.24) can be written as

$$E(Y_{ijk}|X_{ijk}) = \mu_{ij} + X'_{ijk}\beta, \quad k = 1, \dots, n_c. \quad (2.5.26)$$

Define $Y = (Y^{1'}, \dots, Y^{IJ'})'$, $X = (X^{1'}, X^{2'}, \dots, X^{IJ'})'$ where

$$X^c = (X_{ij1}, \dots, X_{ijn_c})' := (X_{c1}, \dots, X_{cn_c})'$$

and $Y^c = (Y_{ij1}, \dots, Y_{ijn_c})' := (Y_{c1}, \dots, Y_{cn_c})'$ for $c = (i-1)J + j$. Then,

$$E(Y|X) = D\alpha + X\beta, \quad (2.5.27)$$

where $D = \mathbf{I}_{IJ} \otimes \mathbf{1}_{n_c}$ is the design matrix, α corresponding to the cell means parameters μ_{ij} . Multiply $I - P_D$ on both sides of (2.5.27) where $P_D = D(D'D)^{-1}D' = \mathbf{I}_{IJ} \otimes n_c^{-1}\mathbf{1}_{n_c}\mathbf{1}'_{n_c}$ is the projection matrix of D , we have

$$E\{(I - P_D)Y|X\} = (I - P_D)X\beta,$$

where we eliminate the nuisance parameters α in (2.5.27). So a natural generalization of $T_{n,p}$ to the factorial design is

$$T_{n,p} = \frac{1}{IJ} \sum_{c=1}^{IJ} (P_{n_c}^A)^{-1} \sum^* \phi(i, j, k, l), \quad (2.5.28)$$

where $\phi(i, j, k, l) = \frac{1}{4}(X_{ci} - X_{cj})'(X_{ck} - X_{cl})\Delta(i, j)\Delta(k, l)$, $\Delta(i, j) = \{Y_{ci} - Y_{cj} - (X_{ci} - X_{cj})'\beta_0\}$, and the second summation is over distinct observations in the c -th cell.

As an extension to Model (2.2.4), we assume in each cell

$$X_{ci} = \Gamma_c Z_{ci} + \mu_c, \quad (2.5.29)$$

where Γ_c is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma_c \Gamma_c' = \Sigma_c = \text{Var}(X_{ijk})$ for $c = (i-1)J + j$, and Z_{ci} are independent and identically distributed random vectors having the same qualifications as in Model (2.2.4). An extension of Condition (2.2.8) is

$$p(n_c) \rightarrow \infty \text{ as } \min_c n_c \rightarrow \infty, \Sigma_c > 0 \text{ and } \text{tr}(\Sigma_c^4) = o\{\text{tr}^2(\Sigma_c^2)\}. \quad (2.5.30)$$

For $c = 1, \dots, IJ$, the factorial design version of the local alternative hypothesis (2.4.12) is

$$(\beta - \beta_0)' \Sigma_c (\beta - \beta_0) = o(1) \text{ and} \quad (2.5.31)$$

$$(\beta - \beta_0)' \Sigma_c^3 (\beta - \beta_0) = o\{n_c^{-1} \text{tr}(\Sigma_c^2)\}.$$

The following corollary can be readily established by modifying the proof of Theorem 3.

Corollary 1 Assume Model (2.5.29) and assumption (2.5.30), then under either H_0 or (2.5.31),

$$\sigma_{fac,0}^{-1} \left(T_{n,p} - \frac{1}{IJ} \sum_{c=1}^{IJ} \|\Sigma_c(\beta - \beta_0)\|^2 \right) \xrightarrow{d} N(0, 1), \quad (2.5.32)$$

where $\sigma_{fac,0}^2 = \frac{2\sigma^4}{(IJ)^2} \sum_{c=1}^{IJ} tr(\Sigma_c^2) / \{n_c(n_c - 1)\}$.

Let $\widehat{tr}(\Sigma_c^2)$ be the analog of the $tr(\Sigma^2)$ estimator given in (2.4.14) and $\hat{\sigma}^2 = \frac{1}{IJ} \sum_{i,j} \frac{1}{n_c - 1} \sum_{k=1}^{n_c} (Y_{ijk} - X'_{ijk}\beta_0 - \bar{Y}_{ij\cdot} + \bar{X}'_{ij\cdot}\beta_0)^2$, where $\bar{Y}_{ij\cdot} = \frac{1}{n_c} \sum_{k=1}^{n_c} Y_{ijk}$ and $\bar{X}_{ij\cdot} = \frac{1}{n_c} \sum_{k=1}^{n_c} X_{ijk}$. Then, an α -level test for the factorial design rejects H_0 if

$$T_{n,p} \geq \frac{\hat{\sigma}^2 z_\alpha}{(IJ)} \left\{ 2 \sum_{c=1}^{IJ} \widehat{tr}(\Sigma_c^2) / \{n_c(n_c - 1)\} \right\}^{1/2}.$$

Similar to our analysis in the last Section for the simple random design, we can also evaluate the power of the test for two fixed alternatives under

$$(\beta - \beta_0)' \Sigma_c (\beta - \beta_0) \text{ is not } o(1) \text{ for any } c. \quad (2.5.33)$$

One is

$$(\beta - \beta_0)' \Sigma_c^3 (\beta - \beta_0) = o \left\{ \frac{1}{n_c} (\beta - \beta_0)' \Sigma_c (\beta - \beta_0) tr(\Sigma_c^2) \right\} \quad (2.5.34)$$

and the other is

$$n_c^{-1} (\beta - \beta_0)' \Sigma_c (\beta - \beta_0) tr(\Sigma_c^2) = o \left\{ (\beta - \beta_0)' \Sigma_c^3 (\beta - \beta_0) \right\}. \quad (2.5.35)$$

Let

$$\begin{aligned} \sigma_{A_1, fac}^2 &= (IJ)^{-2} \sum_{c=1}^{IJ} \left\{ 2(\sigma^4 + B_{1c}^2) tr(\Sigma_c^2) + 4\sigma^2 tr(\Sigma_c^2) B_{1c} \right. \\ &\quad \left. + 4\Delta(\sigma^2 + B_{1c}) tr(A_{1c} \circ A_{3c}) + 2\Delta^2 tr\{(A_{0c} \text{diag}(A_{1c}))^2\} \right\} / P_{n_c}^2 \end{aligned}$$

and

$$\sigma_{A_2, fac}^2 = \frac{1}{(IJ)^2} \sum_{c=1}^{IJ} \left\{ (B_{1c} + \sigma^2) B_{3c} + B_{2c}^2 + \Delta tr(A_{1c} \circ A_{2c}) \right\} / n_c$$

where B_{ic} and A_{ic} are the factorial design versions of A_i and B_i respectively.

The following corollary establishes the asymptotic normality of $T_{n,p}$ under the two fixed alternatives, whose proof can be made by modifying that of Theorem 4 in Appendix.

Corollary 2 Assume Model (2.5.29) and (2.5.33), then as $\min_c n_c \rightarrow \infty$, (i) under the fixed alternative (2.5.34),

$$\sigma_{A_1, fac}^{-1} (T_{n,p} - \frac{1}{IJ} \sum_{c=1}^{IJ} \|\Sigma_c(\beta - \beta_0)\|^2) \xrightarrow{d} N(0, 1); \quad (2.5.36)$$

(ii) under the fixed alternatives (2.5.35),

$$\sigma_{A_2, fac}^{-1} (T_{n,p} - \frac{1}{IJ} \sum_{c=1}^{IJ} \|\Sigma_c(\beta - \beta_0)\|^2) \xrightarrow{d} N(0, 1). \quad (2.5.37)$$

The results in the above corollaries can be used to evaluate the power properties of the proposed test under factorial designs in a fashion similar to that in last Section.

2.6 Simulation Study

We conducted numerical simulations to evaluate the finite sample performance of the proposed tests under both simple random and factorial designs. For comparison purposes, we also carried out simulation for the F-test and an Empirical Bayes (EB) test proposed by Goeman *et al.* (2009). The empirical Bayes test is formulated via a score test on the hyper-parameter of a prior distribution assumed on the regression coefficients. As it allows $p > n$, it is applicable for high dimensional data.

The first set of simulations were designed to evaluate the performance of the test for the linear regression model with the simple random designs:

$$Y_i = \alpha + X_i' \beta + \varepsilon_i, \quad (2.6.38)$$

where $\text{Var}(\varepsilon_i) = \sigma^2 = 4$. Two distributions were experimented for ε_i . One was $N(0, 4)$; the other was a centralized gamma distribution with the shape parameter 1 and the scale parameter 0.5. The hypotheses to be tested were

$$H_0 : \beta = \mathbf{0}_{p \times 1} \quad vs \quad H_1 : \beta \neq \mathbf{0}_{p \times 1}.$$

Independent and identically distributed covariates X_1, \dots, X_n with $X_i = (X_{i1}, \dots, X_{ip})'$ were generated according to a moving average model

$$X_{ij} = \rho_1 Z_{ij} + \rho_2 Z_{i(j+1)} + \dots + \rho_T Z_{i(j+T-1)} + \mu_j, \quad j = 1, \dots, p; \quad (2.6.39)$$

for some $T < p$. Here $Z_i = (Z_{i1}, \dots, Z_{i(p+T-1)})'$ is a $(p+T-1)$ -dimensional $N(0, I_{p+T-1})$ random vector, $\{\mu_j\}_{j=1}^p$ were fixed constants generated from the Uniform (2,3) distribution. The coefficients $\{\rho_l\}_{l=1}^T$ were generated independently from the Uniform (0, 1) distribution and were kept fixed once generated. Model (2.6.39) implied that $\Sigma = \left(\sum_{k=1}^{T-|j-l|} \rho_k \rho_{k+|j-l|} I\{|j-l| < T\} \right)$. Hence the correlation among X_{ij} and X_{il} were determined by $|j-l|$ and T . We chose two values of T , 10 and 20, to generate different levels of dependence. The auto-correlation functions for model (2.6.39) are displayed in Figure 2.1.

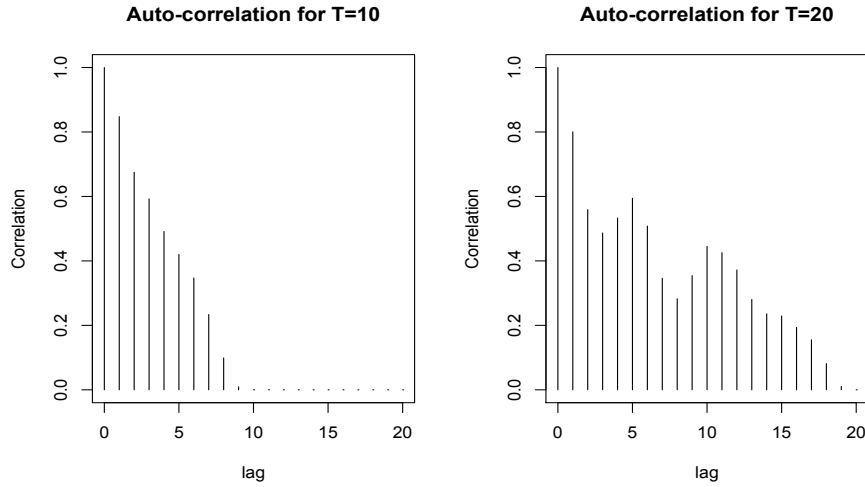


Figure 2.1 The auto-correlation functions for series $\{X_{ij}\}_{j=1}^p$.

Two configurations of the alternative hypothesis H_1 were experimented. One allocated half of the β -components of equal magnitude to be non-zeros, the so-called the “non-sparse case”. The other has only five non-zero components of equal magnitude, the so-called “sparse case”. In both cases, we fixed $\|\beta\|^2$ at three levels: 0.02, 0.04 and 0.06. To gain information on the performance of the proposed test, we consider two settings regarding p and n . One is $p < n$, which allowed F-test; and the other one is $p \gg n$. In the first setting, we set $\rho_n = p/n = (0.85, 0.90, 0.95)$, where $p = 34, 54, 76$ and $n = 40, 60, 80$ respectively. For the setting of $p \gg n$, we chose $p = 310, 400$ and 550 , which was increased exponentially, according to $p = \exp(n^{0.4}) + 230$ for $n = 40, 60, 80$ respectively. All the results were based on 1000

simulation replicates.

Tables 2.1 and 2.2 summarize the empirical sizes and powers of the proposed tests as well as those for the F-tests and EB tests with the normally and the centralized gamma distributed residuals for $p < n$. The empirical sizes of the proposed tests, EB tests and the F-tests were quite reasonably around 0.05. We find that the proposed tests consistently outperformed the EB and the F-tests for both normally and gamma distributed residuals, for different levels of dependence ($T=10$ or 20), and for both the sparse and the non-sparse settings. In particular, in the sparse setting, although there were some reduction of power for all three tests, the power reduction in the F-test was the most significant. The empirical power of the proposed test was quite responsive to the signal to the noise ratio (SNR), which is $n\|\Sigma(\beta-\beta_0)\|^2/\{\sqrt{2tr(\Sigma^2)}\sigma^2\}$, in all the settings. We also computed the theoretical power given in (2.4.16) derived from Theorem 3 under the so-called local alternatives. It was found that there was a good agreement between the empirical power and the theoretical power when the SNR was relatively small. This makes sense as a small SNR is much in tune with the local alternatives.

Table 2.3 and 2.4 report the empirical powers and sizes of the proposed tests and the EB tests when p were much larger than n , which makes F-test unapplicable. We observe that the sizes of the proposed tests became closer to the nominal level 0.05 than Table 2.1 and 2.2. This is also confirmed by the null distributions plots in Figure 2.2. The power of the proposed test were increased quite rapidly as the SNR was increased. In contrast, the EB test suffered from rather severe size distortion for all cases considered. At the meanwhile, the power of the EB test endured very low power when $T = 10$. This alarming performance may be due to the fact that its justification as in Goeman *et al.* (2009) was made for p being fixed while $n \rightarrow \infty$.

Considering that the proposed test is an asymptotic test, we plotted in Figure 2.2 the kernel density estimates for the standardized test statistics of proposed test under H_0 for $T = 10$ and compared them with the standard normal distribution. It shows that the null distribution was quite closer to that of $N(0, 1)$, which confirmed the asymptotic null distribution of the standardized test statistic given in Theorem 3. There was some right skewness when p is less than n . However, as p was increased, this skewness was largely reduced when p was increased.

The second set of the simulations were designed to understand performance of the proposed

Table 2.1 Empirical size and power of the F-test, the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p \times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p \times 1}$ at significant level 5% for normal residual. LP represents the theoretical local power.

(n, p)	$\ \beta\ ^2$	SNR	$T = 10$			LP	SNR	$T = 20$		
			F-test	EB	New			F-test	EB	New
(a) Non-sparse case										
(40, 34)	0.00 (size)	0.00	0.05	0.04	0.06	0.05	0.00	0.05	0.04	0.07
	0.02	0.96	0.16	0.19	0.26	0.25	4.31	0.19	0.65	0.71
	0.04	1.92	0.31	0.36	0.44	-	8.62	0.35	0.90	0.93
	0.06	2.89	0.41	0.48	0.57	-	12.94	0.51	0.97	0.98
(60, 54)	0.00 (size)	0.00	0.05	0.03	0.06	0.05	0.00	0.05	0.04	0.06
	0.02	1.48	0.21	0.26	0.34	-	8.19	0.28	0.92	0.95
	0.04	2.95	0.43	0.53	0.62	-	16.38	0.53	1.00	1.00
	0.06	4.44	0.62	0.70	0.80	-	24.57	0.72	1.00	1.00
(80, 76)	0.00 (size)	0.00	0.06	0.03	0.06	0.05	0.00	0.04	0.04	0.06
	0.02	1.25	0.19	0.24	0.33	0.35	6.19	0.25	0.87	0.91
	0.04	2.51	0.34	0.48	0.56	-	12.39	0.41	0.99	1.00
	0.06	3.76	0.52	0.68	0.77	-	18.58	0.56	1.00	1.00
(b) Sparse case										
(40, 34)	0.02	0.59	0.08	0.12	0.18	0.15	1.41	0.09	0.25	0.32
	0.04	1.19	0.12	0.19	0.27	0.32	2.82	0.15	0.43	0.52
	0.06	1.78	0.17	0.29	0.38	-	4.23	0.20	0.60	0.68
(60, 54)	0.02	0.81	0.09	0.14	0.22	0.20	2.22	0.09	0.42	0.50
	0.04	1.63	0.13	0.26	0.36	-	4.45	0.18	0.68	0.76
	0.06	2.44	0.18	0.40	0.50	-	6.68	0.22	0.85	0.90
(80, 76)	0.02	0.62	0.07	0.11	0.17	0.15	1.67	0.09	0.34	0.42
	0.04	1.25	0.10	0.22	0.33	0.35	3.35	0.11	0.57	0.67
	0.06	1.87	0.13	0.32	0.44	-	5.03	0.16	0.80	0.87

Table 2.2 Empirical size and power of the F-test, the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p \times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p \times 1}$ at significant level 5% for centralized gamma residual. LP represents the theoretical local power.

(n, p)	$\ \beta\ ^2$	SNR	$T = 10$				LP	SNR	$T = 20$		
			F-test	EB	New	F-test			EB	New	
(a) Non-sparse case											
(40, 34)	0.00 (size)	0.00	0.04	0.04	0.05	0.05	0.00	0.05	0.04	0.06	
	0.02	0.96	0.14	0.22	0.28	0.25	4.31	0.20	0.67	0.73	
	0.04	1.92	0.30	0.36	0.45	-	8.62	0.35	0.88	0.92	
	0.06	2.89	0.47	0.49	0.59	-	12.94	0.52	0.95	0.96	
(60, 54)	0.00 (size)	0.00	0.06	0.03	0.06	0.05	0.00	0.05	0.04	0.06	
	0.02	1.48	0.22	0.29	0.39	-	8.19	0.28	0.90	0.93	
	0.04	2.95	0.46	0.55	0.63	-	16.38	0.53	0.99	0.99	
	0.06	4.44	0.63	0.73	0.79	-	24.57	0.72	1.00	1.00	
(80, 76)	0.00 (size)	0.00	0.04	0.03	0.06	0.05	0.00	0.05	0.04	0.06	
	0.02	1.25	0.21	0.23	0.31	0.35	6.19	0.24	0.86	0.90	
	0.04	2.51	0.38	0.48	0.58	-	12.39	0.41	0.98	0.98	
	0.06	3.76	0.51	0.68	0.75	-	18.58	0.59	1.00	1.00	
(b) Sparse case											
(40, 34)	0.02	0.59	0.07	0.13	0.20	0.15	1.41	0.09	0.26	0.35	
	0.04	1.19	0.14	0.22	0.31	0.32	2.82	0.13	0.49	0.58	
	0.06	1.78	0.15	0.29	0.40	-	4.23	0.21	0.62	0.70	
(60, 54)	0.02	0.81	0.09	0.15	0.23	0.20	2.22	0.09	0.42	0.49	
	0.04	1.63	0.11	0.30	0.40	-	4.45	0.17	0.69	0.76	
	0.06	2.44	0.15	0.45	0.56	-	6.68	0.24	0.86	0.91	
(80, 76)	0.02	0.62	0.06	0.11	0.18	0.15	1.67	0.08	0.37	0.43	
	0.04	1.25	0.10	0.24	0.33	0.35	3.35	0.12	0.65	0.72	
	0.06	1.87	0.12	0.35	0.48	-	5.03	0.14	0.77	0.84	

Table 2.3 Empirical size and power of the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p \times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p \times 1}$ at significant level 5% for normal residual. LP represents the theoretical local power.

(n, p)	$\ \beta\ ^2$	SNR	$T = 10$			SNR	$T = 20$		
			EB	New	LP		EB	New	LP
(a) Non-sparse case									
(40, 310)	0.00 (size)	0.00	0.00	0.06	0.05	0.00	0.02	0.06	-
	0.02	0.30	0.01	0.09	0.09	1.99	0.26	0.46	-
	0.04	0.61	0.01	0.15	0.15	3.99	0.47	0.68	-
	0.06	0.92	0.05	0.21	0.23	5.98	0.62	0.81	-
(60, 400)	0.00 (size)	0.00	0.01	0.05	0.05	0.00	0.01	0.05	-
	0.02	0.49	0.02	0.14	0.12	2.51	0.30	0.54	-
	0.04	0.98	0.05	0.23	0.25	5.03	0.63	0.82	-
	0.06	1.47	0.08	0.31	-	7.54	0.83	0.93	-
(80, 550)	0.00 (size)	0.00	0.00	0.05	0.05	0.00	0.02	0.06	-
	0.02	0.55	0.02	0.15	0.14	4.02	0.63	0.79	-
	0.04	1.11	0.08	0.29	0.30	8.05	0.91	0.96	-
	0.06	1.66	0.13	0.37	-	12.08	0.98	0.99	-
(b) Sparse case									
(40, 310)	0.02	0.16	0.01	0.08	0.07	0.58	0.05	0.15	0.14
	0.04	0.32	0.01	0.12	0.09	1.17	0.09	0.23	0.32
	0.06	0.48	0.01	0.11	0.12	1.75	0.12	0.30	-
(60, 400)	0.02	0.27	0.01	0.08	0.08	0.60	0.05	0.16	0.15
	0.04	0.54	0.02	0.14	0.13	1.21	0.09	0.25	0.33
	0.06	0.82	0.04	0.18	0.20	1.82	0.14	0.35	-
(80, 550)	0.02	0.35	0.02	0.10	0.10	1.05	0.11	0.24	0.28
	0.04	0.70	0.03	0.16	0.17	2.11	0.25	0.46	-
	0.06	1.05	0.05	0.25	0.28	3.17	0.38	0.58	-

Table 2.4 Empirical size and power of the EB test and the proposed test (new) for $H_0 : \beta = \mathbf{0}_{p \times 1}$ vs $H_1 : \beta \neq \mathbf{0}_{p \times 1}$ at significant level 5% for centralized gamma residual. LP represents the theoretical local power.

(n, p)	$\ \beta\ ^2$	SNR	$T = 10$			SNR	$T = 20$		
			EB	New	LP		EB	New	LP
(a) Non-sparse case									
(40, 310)	0.00 (size)	0.00	0.01	0.06	0.05	0.00	0.01	0.06	-
	0.02	0.30	0.01	0.12	0.09	1.99	0.24	0.45	-
	0.04	0.61	0.03	0.19	0.15	3.99	0.52	0.70	-
	0.06	0.92	0.05	0.24	0.23	5.98	0.69	0.83	-
(60, 400)	0.00 (size)	0.00	0.01	0.04	0.05	0.00	0.01	0.04	-
	0.02	0.49	0.02	0.13	0.12	2.51	0.35	0.57	-
	0.04	0.98	0.05	0.24	0.25	5.03	0.65	0.82	-
	0.06	1.47	0.10	0.36	-	7.54	0.82	0.93	-
(80, 550)	0.00 (size)	0.00	0.01	0.05	0.05	0.00	0.02	0.05	-
	0.02	0.55	0.03	0.16	0.14	4.02	0.67	0.82	-
	0.04	1.11	0.07	0.23	0.30	8.05	0.91	0.97	-
	0.06	1.66	0.16	0.40	-	12.08	0.97	0.99	-
(a) Sparse case									
(40, 310)	0.02	0.16	0.01	0.08	0.07	0.58	0.05	0.16	0.14
	0.04	0.32	0.01	0.10	0.09	1.17	0.11	0.25	0.32
	0.06	0.48	0.02	0.14	0.12	1.75	0.14	0.33	-
(60, 400)	0.02	0.27	0.02	0.09	0.08	0.60	0.04	0.15	0.15
	0.04	0.54	0.02	0.12	0.13	1.21	0.10	0.25	0.33
	0.06	0.82	0.04	0.20	0.20	1.82	0.18	0.38	-
(80, 550)	0.02	0.35	0.01	0.10	0.10	1.05	0.10	0.24	0.28
	0.04	0.70	0.03	0.17	0.17	2.11	0.27	0.48	-
	0.06	1.05	0.06	0.25	0.28	3.17	0.39	0.60	-

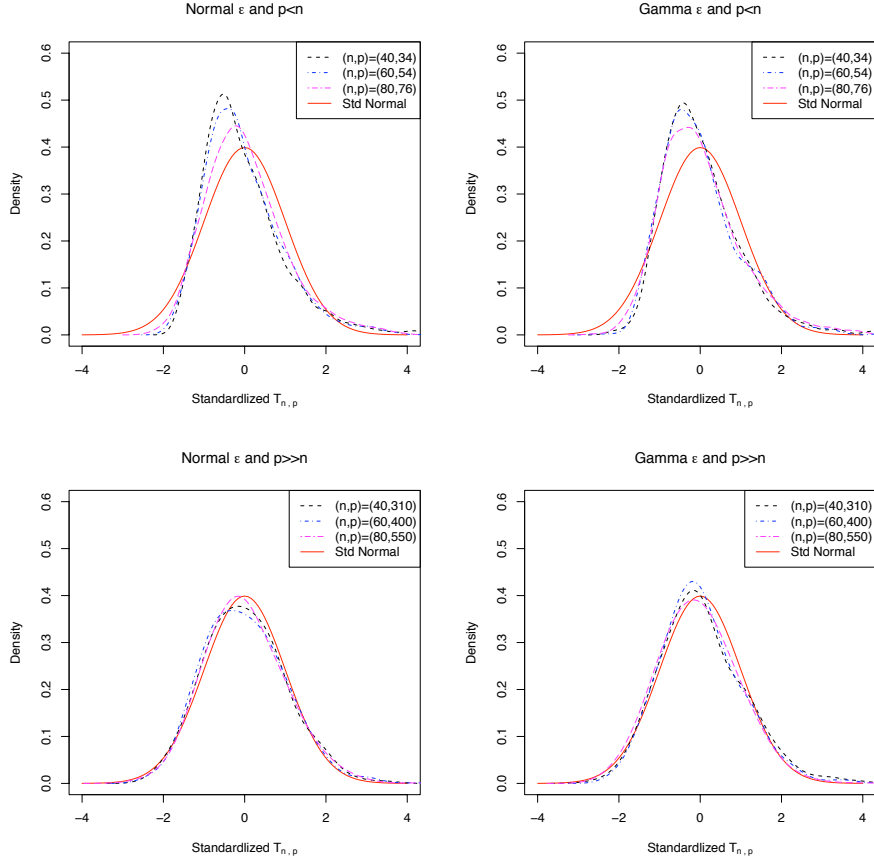


Figure 2.2 The null distributions of standardized $T_{n,p}$.

test under the factorial designs. We simulated a two-factor balanced design with two levels for each factor:

$$Y_{ijk} = \alpha_{ij} + X'_{ijk}\beta + \varepsilon_{ijk}, \quad k = 1, 2, \dots, n_c \quad (2.6.40)$$

where $c = 2(i - 1) + j$ and $i, j = 1, 2$, corresponding to (i, j) -th cell and the parameters $(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}) = (1, 3, 3, 4)$. The sparsity set-ups for β were the same to those for simple random designs used in (2.6.38). Within each cell, independent and identically distributed p -dimensional X_{ijk} were generated from the moving average model (2.6.39) with $T = T_c$, where T_c equals to 10, 15, 20 and 25 for $c = 1, 2, 3, 4$ respectively. Using the different T values was to generate different dependence structure in Σ . We assigned the $n_c = 20$ and 30 in all cells, and three values of p : 100, 150 and 200. The simulation results for the proposed test are summarized

in Table 2.5. We observe that the sizes of the proposed test were satisfactorily around 0.05. The power of the test increased as the SNR_f , the factorial design version of SNR, was increased. When the sample size was increased from 20 to 30, we observed significant increase in the power under all settings.

Table 2.5 Empirical size and power of the proposed test for $H_0 : \beta = \mathbf{0}_{p \times 1}$ in a 2×2 factorial design with $n_1 = 20$ and $n_2 = 30$ replicates in each cell.

p	$\ \beta\ ^2$	Non-sparse				Sparse			
		SNR_f	n_1	SNR_f	n_2	SNR_f	n_1	SNR_f	n_2
(a) Normal residuals									
100	0.00 (size)	0.00	0.06	0.00	0.06	0.00	0.07	0.00	0.05
	0.02	3.05	0.65	4.58	0.85	0.70	0.20	1.06	0.26
	0.04	6.10	0.88	9.16	0.98	1.41	0.29	2.12	0.48
	0.06	9.16	0.96	13.74	1.00	2.12	0.44	3.18	0.65
150	0.00 (size)	0.00	0.06	0.00	0.06	0.00	0.05	0.00	0.06
	0.02	2.59	0.57	3.89	0.77	0.57	0.15	0.85	0.21
	0.04	5.18	0.84	7.78	0.97	1.14	0.28	1.71	0.39
	0.06	7.78	0.94	11.67	0.99	1.71	0.35	2.57	0.54
200	0.00 (size)	0.00	0.07	0.00	0.06	0.00	0.07	0.00	0.06
	0.02	2.28	0.50	3.43	0.73	0.49	0.14	0.73	0.18
	0.04	4.57	0.78	6.86	0.94	0.98	0.22	1.47	0.35
	0.06	6.86	0.89	10.29	0.99	1.47	0.31	2.21	0.48
(b) Gamma residuals									
100	0.00 (size)	0.00	0.07	0.00	0.05	0.00	0.07	0.00	0.06
	0.02	3.05	0.66	4.58	0.83	0.70	0.15	1.06	0.28
	0.04	6.10	0.86	9.16	0.97	1.41	0.31	2.12	0.48
	0.06	9.16	0.95	13.74	0.99	2.12	0.47	3.18	0.66
150	0.00 (size)	0.00	0.07	0.00	0.05	0.00	0.04	0.00	0.06
	0.02	2.59	0.57	3.89	0.78	0.57	0.16	0.85	0.22
	0.04	5.18	0.81	7.78	0.96	1.14	0.28	1.71	0.39
	0.06	7.78	0.93	11.67	0.99	1.71	0.37	2.57	0.57
200	0.00 (size)	0.00	0.05	0.00	0.06	0.00	0.06	0.00	0.05
	0.02	2.28	0.53	3.43	0.74	0.49	0.14	0.73	0.18
	0.04	4.57	0.77	6.86	0.93	0.98	0.24	1.47	0.32
	0.06	6.86	0.89	10.29	0.98	1.47	0.30	2.21	0.48

2.7 Association Test for Gene-sets

We applied the proposed test for association between gene-sets and certain clinical outcomes in a randomized factorial design experiment applied to 24 six-month-old Yorkshire gilts. The gilts were genotyped according to the melanocortin-4 receptor gene, 12 of them with D298 and the other with N298. Two diet treatments were randomly assigned to the 12 gilts in each genotype. One treatment is ad libitum (no restrictions) in the amount of feed consumed; the other is fasting. More details of the experiment could be found at Lkhagvadorj *et al.*(2009). The genotypes and the diet treatments were the two factors in the factorial experiments. The purpose of our study was to identify associations between gene-sets and triiodothyronine (T_3) measurement, a vital thyroid hormone that increases the metabolic rate, protein synthesis and stimulates breakdown of cholesterol.

The gene expression values were obtained for 24,123 genes in liver tissues, as well as measurements of T_3 in the blood on each gilt. Gene sets are defined by Gene Ontology (GO term) (The Gene Ontology Consortium, 2000), which classifies genes into different sets according to their biological functions among three broad categories: cellular component, molecular function and biological process. The data-set contained 6176 GO terms. Our objective is to find the GO terms which are significantly correlated with T_3 after accounting for the design factors.

Let i, j, k be indices for treatment, genotype and observations, respectively. For instance, Y_{ijk} denote the T_3 measurement for the k -th gilt in the i -th treatment with j -th genotype, and X_{ijk}^g be the corresponding p_g -dimension gene expressions for the g -th GO term. We consider the following four models corresponding to four types of designs:

$$\text{Design I: } Y_k = \alpha + X_k^{g'}\beta^g + \varepsilon_k^g, \quad k = 1, \dots, 24;$$

$$\text{Design II: } Y_{ik} = \alpha + \mu_i + X_{ik}^{g'}\beta^g + \varepsilon_{ik}^g, \quad k = 1, \dots, 12;$$

$$\text{Design III: } Y_{jk} = \alpha + \tau_j + X_{jk}^{g'}\beta^g + \varepsilon_{jk}^g, \quad k = 1, \dots, 12;$$

$$\text{Design IV: } Y_{ijk} = \alpha + \mu_i + \tau_j + \mu\tau_{ij} + X_{ijk}^{g'}\beta^g + \varepsilon_{ijk}^g, \quad k = 1, \dots, 6$$

for $i = 1, 2$, $j = 1, 2$ and $g = 1, \dots, G$ where $G = 6176$ is the total number of the GO terms, μ_i stand for diet treatment effects, τ_j for genotype effects and $\mu\tau_{ij}$ represent the interaction

between treatment and genotype. For each GO term, we tested for

$$H_0 : \beta^g = 0 \quad vs \quad H_1 : \beta^g \neq 0.$$

Among the 6176 GO terms, the dimension p_g of the gene-sets ranged from 1 to 5158, and many of the gene-sets shared common genes. Hence, there were both high dimensionality and multiplicity. We applied the proposed high dimensional test for $p_g \geq 5$ and the F-test for $p_g < 5$. Without confusion, we call this combination of the proposed high dimensional test and F-test as the proposed test in this Section. For comparison purposes, the Empirical Bayes test was also carried out.

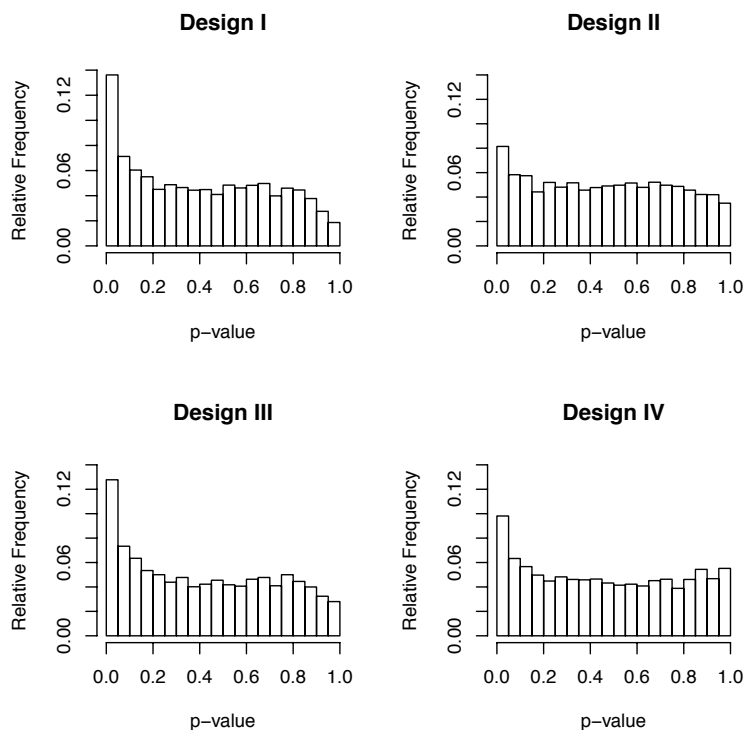


Figure 2.3 Histograms of the p-values on all GO terms using the proposed tests.

Figure 2.3 and 2.4 display histograms of p-values of the proposed tests and the EB tests under the four designs (I-IV) for all the gene-sets, respectively. Both Figures 2.3 and 2.4 show that the histograms for Designs I and III were very similar, so were the histograms of Designs II and IV. This was confirmed by Figure 2.5 where we plots the histograms for the differences

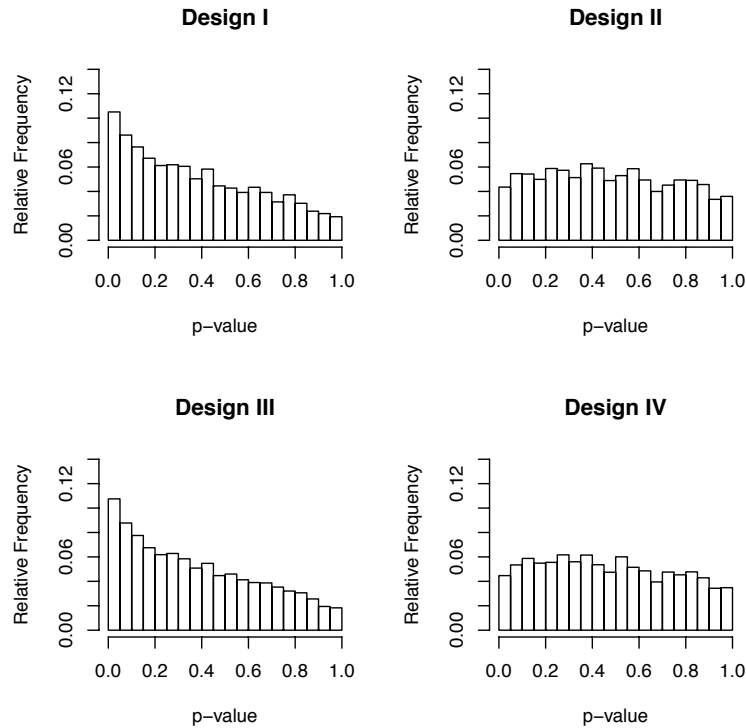


Figure 2.4 Histograms of the p-values on all GO terms using Empirical Bayes (EB) tests.

in the p-values from the proposed tests. We observed that the p-values from Design I and III had higher portion of small p-values than those under Design II and IV. These features show that the form of design is important and it is necessary to account for different designs into the analysis.

By controlling the false discover rate (FDR) for the p-values from the proposed tests at 5%, 129, 23, 51 and 40 GO terms were declared statistically significant under designs I-IV respectively. We list in Table 2.6 significant GO terms identified by the proposed tests under at least three designs, together with their p-values and dimensions. They include GO terms that significant under all four designs: GO:0005086, GO:0007528 and GO:0032012. GO:0005086 is related to the molecular function, which stimulates the exchange of guanyl nucleotides associated with the GTPase ARF. GO:0007528 belongs to the biological process category. GO:0032012 also belongs to the biological process, which was also found significant by the EB test.

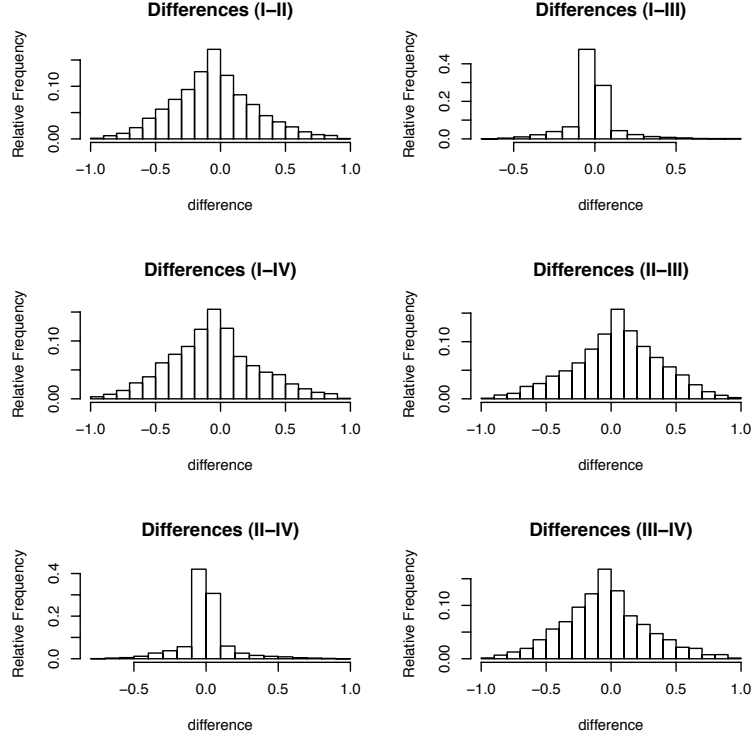


Figure 2.5 Differences in the p-values among Designs I-IV.

The EB tests detected one significant GO term for each design: GO:0032012 for Designs I and III, and GO:0004731 for Designs II and IV. They were all among the significant GO terms discovered by the proposed tests. That the EB test detected quite few gene-sets is not entirely unexpected as our simulation has shown it tends to have relative low power.

2.8 Appendix: Technical Details

In this appendix, we give technical proofs for the results we presented in Sections 2.2 and 2.4. We will use $\delta_\beta = \beta - \beta_0$ through the Appendix. Let $tr(A \circ B) = \sum_i A_{ii}B_{ii}$, and without loss of generality, assume A_0 is symmetric in $X'A_0X$ as we could use $A = (A_0 + A'_0)/2$ to replace A_0 .

Lemma 1 (i) For any symmetric matrix A and B , $tr^2(A \circ B) \leq tr(A \circ A)tr(B \circ B) \leq tr(A^2)tr(B^2)$.

Table 2.6 P-values of the GO terms which are significant under at least three designs using the proposed test, and their number of genes.

GO term	Design I	Design II	Design III	Design IV	No. of Genes
GO:0004115	3.253E-04	2.774E-06		1.992E-06	8
GO:0005086	2.345E-10	1.945E-05	7.220E-06	1.629E-05	14
GO:0005677	1.082E-04	3.102E-06		7.575E-05	5
GO:0006342	3.068E-04	3.444E-06		5.951E-05	5
GO:0007528	1.110E-16	7.922E-07	2.235E-08	3.203E-04	8
GO:0017136	1.082E-04	3.102E-06		7.575E-05	5
GO:0032012	0.000E-04	2.586E-06	2.746E-10	5.418E-06	12
GO:0050909	1.545E-09	3.842E-05	4.216E-05		5

(ii) For any symmetric non-negative matrix Q and vector a , $a'Qaa'Q^3a \geq (a'Q^2a)^2$.

Proof (i) By Cauchy-Schwarz inequality, we have

$$tr^2(A \circ B) = \left(\sum_i A_{ii}B_{ii} \right)^2 \leq \left(\sum_i A_{ii}^2 \right) \left(\sum_i B_{ii}^2 \right) = tr(A \circ A)tr(B \circ B).$$

Notice that $tr(A^2) = \sum_i A_{ii}^2 + \sum_{i \neq j} A_{ij}^2 \geq tr(A \circ A)$. The second inequality follows.

(ii) Since for any symmetric matrix, there exist an orthogonal matrix P such that $Q = Pdiag(\lambda_1, \dots, \lambda_p)P'$. Plugging this decomposition and applying the Cauchy-Schwarz inequality again, the inequality follows immediately. \square

Lemma 2 For random variables Z_1 and Z_2 from the model assumption (2.2.4), we have

(i) For any $m \times m$ symmetric matrices B_1 and B_2 ,

$$E\{(Z_1' B_1 Z_1)(Z_1' B_2 Z_1)\} = tr(B_1)tr(B_2) + 2tr(B_1 B_2) + \Delta tr(B_1 \circ B_2);$$

(ii) Let $A_0 = \Gamma' \Gamma$. Then

$$E\{(Z_1' A Z_2)^4\} = 3tr^2(\Sigma^2) + 6tr(\Sigma^4) + 6\Delta tr(A_0^2 \circ A_0^2) + \Delta^2 \sum_{i,k=1}^m (A_{0,ik})^4.$$

(iii) For any $m \times m$ symmetric matrices B_1 and B_2 then there exists a finite constant C such that

$$E\{(Z_1' B_1 Z_1 - tr(B_1))(Z_1' B_2 Z_1 - tr(B_2))\}^2 \leq C tr(B_1^2) tr(B_2^2)$$

(iv) For any $m \times m$ symmetric matrices A, B and D and suppose that $E(Z_{11}^6) = \tau_6$. Then

$$\begin{aligned} E\{Z_1'AZ_1Z_1'BZ_1Z_1'DZ_1\} &= 8tr(ABC) + 2tr(AB)tr(D) + 2tr(BD)tr(A) \\ &+ tr(A)tr(B)tr(D) + \Delta\{tr(A)tr(B \circ D) + tr(B)tr(A \circ D) + tr(D)tr(A \circ B) \\ &+ tr(AB \circ D) + tr(AD \circ B) + tr(BD \circ A)\} + (\tau_6 - 15 - 6\Delta)tr(A \circ B \circ D). \end{aligned}$$

Proof Let z_i denotes the i -th entry of Z_1 . To derive part (i), we denote the (j, k) -th entry of B_i by $B_{jk}^{(i)}$, for $i = 1, 2$. Then, by model assumption (2.2.4),

$$\begin{aligned} &E\{(Z_1'B_1Z_1)(Z_1'B_2Z_1)\} \\ &= \sum_{i_1, \dots, i_4} B_{i_1i_2}^{(1)} B_{i_3i_4}^{(2)} E(z_{i_1}z_{i_2}z_{i_3}z_{i_4}) \\ &= \sum_{i=1}^m b_{ii}^{(1)} B_{ii}^{(2)} E(z_i^4) + \sum_{i_1 \neq i_2} B_{i_1i_1}^{(1)} B_{i_2i_2}^{(2)} E(z_{i_1}^2)E(z_{i_2}^2) + 2 \sum_{i_1 \neq i_2} B_{i_1i_2}^{(1)} B_{i_1i_2}^{(2)} E(z_{i_1}^2)E(z_{i_2}^2) \\ &= tr(B_1)tr(B_2) + 2tr(B_1B_2) + \Delta tr(B_1 \circ B_2). \end{aligned}$$

For (ii), due to the independence between Z_1 and Z_2 , (ii) follows by applying twice the result in (i). In particular,

$$E\{(Z_1'AZ_2)^4\} = 3E\{(Z_1'A^2Z_1)^2\} + \Delta \sum_{i=1}^m E\{(Z_1'Ae_i e_i'AZ_1)^2\},$$

where e_i is an m -vector with all the components 0 except the i -th component being 1. Then applying (i) again, the asserted result follows.

To show (iii), notice that the left hand side of the inequality can be written as

$$\begin{aligned} &E\left\{\left(\sum_i B_{ii}^{(1)}(z_i^2 - 1) + \sum_{j \neq l} B_{jl}^{(1)} z_j z_l\right) \left(\sum_k B_{kk}^{(2)}(z_k^2 - 1) + \sum_{s \neq t} B_{st}^{(2)} z_s z_t\right)\right\}^2 \\ &\leq J_1 + J_2 + J_3 + J_4 \end{aligned}$$

where

$$J_1 = 4 \sum_{i, k, i_1, k_1} B_{ii}^{(1)} B_{kk}^{(2)} B_{i_1i_1}^{(1)} B_{k_1k_1}^{(2)} (z_i^2 - 1)(z_{i_1}^2 - 1)(z_k^2 - 1)(z_{k_1}^2 - 1),$$

$$\begin{aligned}
J_2 &= 4 \sum_{i, i_1, s \neq t, s_1 \neq t_1} B_{ii}^{(1)} B_{st}^{(2)} B_{i_1 i_1}^{(1)} B_{s_1 t_1}^{(2)} (z_i^2 - 1)(z_{i_1}^2 - 1) z_s z_t z_{s_1} z_{t_1}, \\
J_3 &= 4 \sum_{k, k_1, j \neq l, j_1 \neq l_1} B_{kk}^{(2)} B_{jl}^{(1)} B_{k_1 k_1}^{(2)} B_{j_1 l_1}^{(1)} (z_k^2 - 1)(z_{k_1}^2 - 1) z_j z_l z_{j_1} z_{l_1} \quad \text{and} \\
J_4 &= 4 \sum_{s \neq t, s_1 \neq t_1, j \neq l, j_1 \neq l_1} B_{jl}^{(1)} B_{st}^{(2)} B_{j_1 l_1}^{(1)} B_{s_1 t_1}^{(2)} z_j z_l z_s z_t z_{j_1} z_{l_1} z_{s_1} z_{t_1}.
\end{aligned}$$

In the rest of the appendix, we use C to denote finite positive constants whose values may change and $|B_i| = (|B_{kl}^{(i)}|)_{kl}$ for $i = 1, 2$. Then,

$$\begin{aligned}
|J_1| &\leq C \left(2 \sum_i (B_{ii}^{(1)} B_{ii}^{(2)})^2 + \sum_{i \neq k} (B_{ii}^{(1)} B_{kk}^{(2)})^2 + \sum_{i \neq k} |B_{ii}^{(1)}| |B_{ii}^{(2)}| |B_{kk}^{(1)}| |B_{kk}^{(2)}| \right) \\
&= C \left(\sum_{i, k} (B_{ii}^{(1)} B_{kk}^{(2)})^2 + \sum_{i, k} |B_{ii}^{(1)}| |B_{ii}^{(2)}| |B_{kk}^{(1)}| |B_{kk}^{(2)}| \right) \\
&= C \left(\text{tr}(B_1 \circ B_1) \text{tr}(B_2 \circ B_2) + \text{tr}(|B_1| \circ |B_2|) \text{tr}(|B_1| \circ |B_2|) \right) \leq C \text{tr}(B_1^2) \text{tr}(B_2^2).
\end{aligned}$$

Applying Cauchy-Schwarz inequality,

$$\begin{aligned}
|J_2| &\leq C \left(\sum_{i, s \neq t} (B_{ii}^{(1)} B_{st}^{(2)})^2 + \sum_{i \neq i_1} |B_{ii}^{(1)}| |B_{i_1 i_1}^{(1)}| (B_{ii}^{(2)})^2 \right) \\
&\leq C \left(\sum_i (B_{ii}^{(1)})^2 \sum_{s, t} (B_{st}^{(2)})^2 + \sum_i |B_{ii}^{(1)}| \left(\sum_{i_1} (B_{i_1 i_1}^{(1)})^2 \right)^{1/2} \left(\sum_{i_1} (B_{ii}^{(2)})^4 \right)^{1/2} \right) \\
&\leq C \left(\text{tr}(B_1 \circ B_1) \text{tr}(B_2^2) + \left(\sum_i (B_{ii}^{(1)})^2 \right) \left(\sum_i \sum_{i_1} (B_{ii}^{(2)})^4 \right)^{1/2} \right) \leq C \text{tr}(B_1^2) \text{tr}(B_2^2).
\end{aligned}$$

Similarly, we have $|J_3| \leq C \text{tr}(B_1^2) \text{tr}(B_2^2)$. Finally,

$$\begin{aligned}
|J_4| &\leq C \left(\sum_{j \neq l, s \neq t} (B_{jl}^{(1)} B_{st}^{(2)})^2 + \sum_{j \neq l, s \neq t} |B_{jl}^{(1)}| |B_{jl}^{(2)}| |B_{st}^{(1)}| |B_{st}^{(2)}| \right. \\
&\quad \left. + \sum_{\substack{j \neq l, t \neq l_1 \\ j \neq t, l \neq l_1}} |B_{lj}^{(1)}| |B_{jt}^{(2)}| |B_{tl_1}^{(2)}| |B_{l_1 l}^{(1)}| + \sum_{\substack{j \neq l, j \neq j_1 \\ j_1 \neq l_1, l \neq l_1}} |B_{lj}^{(1)}| |B_{jj_1}^{(2)}| |B_{j_1 l_1}^{(1)}| |B_{l_1 l}^{(2)}| \right) \\
&\leq C \left(\sum_{jl} (B_{jl}^{(1)})^2 \sum_{s, t} (B_{st}^{(2)})^2 + \sum_{j, l} |B_{jl}^{(1)}| |B_{jl}^{(2)}| \sum_{s, t} |B_{st}^{(1)}| |B_{st}^{(2)}| \right. \\
&\quad \left. + 2 \sum_{l, j} (B_{lj}^{(1)})^2 \sum_{t, l_1} (B_{tl_1}^{(2)})^2 \right) \\
&\leq C \text{tr}(B_1^2) \text{tr}(B_2^2).
\end{aligned}$$

Therefore, (iii) is true. The conclusion (iv) could be established based on the fact

$$\begin{aligned}
E\{Z_1'AZ_1Z_1'BZ_1Z_1'DZ_1\} &= \tau_6 \sum_i A_{ii}B_{ii}D_{ii} + (3 + \Delta) \left(\sum_{i \neq k} A_{ii}B_{kk}D_{kk} + 4 \sum_{i \neq j} A_{ij}B_{ij}D_{jj} \right. \\
&\quad \left. + 4 \sum_{i \neq j} A_{ij}B_{jj}D_{ij} + \sum_{i \neq k} A_{ii}B_{kk}D_{ii} + \sum_{i \neq k} A_{ii}B_{ii}D_{kk} + \sum_{i \neq k} A_{ii}B_{ki}D_{ki} \right) \\
&\quad + \sum_{i \neq k \neq s} A_{ii}B_{kk}D_{ss} + 2 \sum_{i \neq j \neq s} A_{ij}B_{ij}D_{ss} + 2 \sum_{i \neq j \neq k} A_{ij}B_{kk}D_{ij} \\
&\quad + 2 \sum_{i \neq k \neq l} A_{ii}B_{kl}D_{kl} + 8 \sum_{i \neq j \neq l} A_{ij}B_{il}D_{lj}.
\end{aligned}$$

□

Proof of Theorem 1 Let $\gamma_0 = (\alpha, \beta_0^\tau)^\tau$. By plugging in the least square estimate $\hat{\gamma}$, we could write the F-statistics in (2.2.3) as

$$G_{n,p} = \frac{(Y - U\gamma_0)'P_{Au}(Y - U\gamma_0)/p}{Y'(I_n - P_U)Y/(n - p - 1)}$$

where $P_{Au} = U(U'U)^{-1}A'(A(U'U)^{-1}A')^{-1}A(U'U)^{-1}U'$, $P_U = U(U'U)^{-1}U'$ and $P_{\mathbf{1}} = \mathbf{1}\mathbf{1}'/n$ be the projection matrices of $U(U'U)^{-1}A'$, U and $\mathbf{1}$ respectively. By applying the matrix inverse formula on $(U'U)^{-1}$, $U(U'U)^{-1}A' = (I - P_{\mathbf{1}})X\{X'(I - P_{\mathbf{1}})X\}^{-1}$. It then follows that $P_{Au} = (I - P_{\mathbf{1}})X(X'(I - P_{\mathbf{1}})X)^{-1}X'(I - P_{\mathbf{1}})$.

Since $P_{Au}(I - P_U) = 0$, the numerator and the denominator of $G_{n,p}$ are independent, and P_{Au} is an idempotent matrix with rank p . We may write

$$\frac{p}{n - p - 1}G_{n,p} \stackrel{d}{=} \frac{\{Q\varepsilon + Q(U(\gamma - \gamma_0))\}'diag(\mathbf{1}'_p, \mathbf{0}'_{n-p})\{Q\varepsilon + Q(U(\gamma - \gamma_0))\}}{\mathbf{z}_1'\mathbf{z}_1},$$

where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \sim N(0, I_n)$ and $\mathbf{z}_1 \sim N(0, I_{n-p-1})$ are independent random variables, and Q is an orthogonal matrix such that $P_{Au} = Q'diag(\mathbf{1}'_p, \mathbf{0}'_{n-p})Q$. Here $\stackrel{d}{=}$ means the two random vectors on either side have the same distribution. Write $Q = (Q_1, Q_2, \dots, Q_n)'$. Note that $Q\varepsilon \stackrel{d}{=} \varepsilon$. Furthermore, write $pG_{n,p}/(n - p - 1)$ as

$$\frac{p}{n - p - 1}G_{n,p} \stackrel{d}{=} \sum_{i=1}^p \{\varepsilon_i^2 + 2\varepsilon_i Q_i'X\delta_\beta\} / \mathbf{z}_1'\mathbf{z}_1 + \delta_\beta'X'P_{Au}X\delta_\beta / \mathbf{z}_1'\mathbf{z}_1 \quad (2.8.41)$$

where $X'P_{Au}X = X'(I - P_{\mathbf{1}})X = \Gamma Z'(I - P_{\mathbf{1}})Z\Gamma'$ and $Z = (Z_1, \dots, Z_n)'$.

For the numerator of the (2.8.41), we can show that under Model (2.2.4), $E\{\delta'_\beta X' P_{Au} X \delta_\beta\} = (n-1)\delta'_\beta \Sigma \delta_\beta$. It is easy to see that $E\{\sum_{i=1}^p \varepsilon_i Q'_i X \delta_\beta\} = 0$ and

$$\text{Var}\left\{\sum_{i=1}^p \varepsilon_i Q'_i X \delta_\beta\right\} = (n-1)\sigma^2 \delta'_\beta \Sigma \delta_\beta. \quad (2.8.42)$$

It can be shown that

$$\text{Var}\{\delta'_\beta X' P_{Au} X \delta_\beta\} = 2(n-1)(\delta'_\beta \Sigma \delta_\beta)^2 + (n+2+1/n)\Delta \text{tr}(A_1 \circ A_1). \quad (2.8.43)$$

Direct calculation shows that $E(\frac{1}{\mathbf{z}_1' \mathbf{z}_1}) = 1/(n-p-3)$ and $E(\frac{1}{\mathbf{z}_1' \mathbf{z}_1})^2 = 1/\{(n-p-3)(n-p-5)\}$. (2.8.42) implies that $\sum_{i=1}^p \varepsilon_i Q'_i X \delta_\beta / \mathbf{z}_1' \mathbf{z}_1 = O_p\{\frac{1}{\sqrt{n}} \sqrt{\delta'_\beta \Sigma \delta_\beta}\}$ and note that $E(X' P_{Au} X) = (n-1)\Sigma$. Then (2.8.43) yields

$$\frac{\delta'_\beta X' P_{Au} X \delta_\beta}{\mathbf{z}_1' \mathbf{z}_1} = \frac{\delta'_\beta \Sigma \delta_\beta}{1-\rho} + O_p\left\{\frac{1}{\sqrt{n}} \delta'_\beta \Sigma \delta_\beta\right\}.$$

If $\delta'_\beta \Sigma \delta_\beta = o(1)$, then

$$\frac{p}{n-p-1} G_{n,p} \stackrel{d}{=} \sum_{i=1}^p \frac{\varepsilon_i^2}{\mathbf{z}_1' \mathbf{z}_1} + \frac{\delta'_\beta \Sigma \delta_\beta}{1-\rho} + o_p(n^{-1/2}).$$

From Bai and Saranadasa (1996),

$$\frac{p}{n-p-1} F_{p, n-p-1; \alpha} = \frac{\rho_n}{1-\rho_n} + \sqrt{\frac{2\rho}{(1-\rho)^3 n}} z_\alpha + o(n^{-1/2}),$$

where z_α is the α quantile of $N(0,1)$ and it can be shown

$$\sqrt{\frac{(1-\rho)^3 n}{2\rho}} \left(\sum_{i=1}^p \frac{\varepsilon_i^2}{\mathbf{z}_1' \mathbf{z}_1} - \frac{\rho_n}{1-\rho_n} \right) \stackrel{d}{\rightarrow} N(0,1).$$

Therefore the power of the F-test is

$$\begin{aligned} \Omega_F(\|\beta - \beta_0\|) &= P\left(\frac{p}{n-p-1} G_{n,p} > \frac{p}{n-p-1} F_{p, n-p-1; \alpha}\right) \\ &= P\left\{\sqrt{\frac{(1-\rho)^3 n}{2\rho}} \left(\sum_{i=1}^p \frac{\varepsilon_i^2}{\mathbf{z}_1' \mathbf{z}_1} - \frac{\rho_n}{1-\rho_n}\right) > z_\alpha - \sqrt{\frac{(1-\rho)^3 n}{2\rho}} \frac{\delta'_\beta \Sigma \delta_\beta}{1-\rho} + o_p(1)\right\} \\ &= \Phi\left(-z_\alpha + \sqrt{\frac{(1-\rho)n}{2\rho}} \delta'_\beta \Sigma \delta_\beta\right) + o(1). \end{aligned}$$

□

Proof of Theorem 2 It is straightforward to show that $E(T_{n,p}) = \|\Sigma\delta_\beta\|^2$. To derive $\text{Var}(T_{n,p})$, we need to derive the variance of h_1, h_2, h_3 and h and then apply the variance decomposition given in (2.3.10).

Let $A_0 = \Gamma'\Gamma$, $A_1 = \Gamma'\delta_\beta\delta_\beta'\Gamma$, $A_2 = \Gamma'\Sigma\delta_\beta\delta_\beta'\Sigma\Gamma$, $A_3 = \Gamma'\Sigma\Gamma$ and $B_i = \delta_\beta'\Sigma^i\delta_\beta$. It can be shown that

$$\zeta_1 = \frac{1}{4}B_1B_3 + \frac{1}{4}\sigma^2B_3 + \frac{1}{4}B_2^2 + \frac{1}{4}\Delta\text{tr}(A_1 \circ A_2). \quad (2.8.44)$$

We can also show that

$$\begin{aligned} \zeta_2 = & \frac{1}{36} \left\{ \sigma^4\text{tr}(\Sigma^2) + 21B_2^2 + 22B_1B_3 + 22\sigma^2B_3 + B_1^2\text{tr}(\Sigma^2) + 2\sigma^2\text{tr}(\Sigma^2)B_1 \right. \\ & \left. + 2\Delta\{B_1 + \sigma^2\}\text{tr}(A_1 \circ A_3) + 20\Delta\text{tr}(A_1 \circ A_2) + \Delta^2\text{tr}\{A_0\text{diag}(A_1)\}^2 \right\}, \end{aligned} \quad (2.8.45)$$

where we need the following facts: $E[(\varepsilon_1 - \varepsilon_2)(X_1 - X_2)'\Sigma\delta_\beta]^2 = 4\sigma^2B_3$,

$$E[\delta_\beta'(X_1 - X_2)(X_1 - X_2)'\Sigma\delta_\beta]^2 = 2\Delta\text{tr}(A_1 \circ A_2) + 8B_2^2 + 4B_1B_3,$$

$$E[\varepsilon_1X_1'(X_2X_2' + \Sigma)\delta_\beta]^2 = 5\sigma^2B_3 + \sigma^2\text{tr}(\Sigma^2)B_1 + \Delta\sigma^2\text{tr}(A_1 \circ A_3),$$

$$E[\delta_\beta'(X_1X_1' + \Sigma)(X_2X_2' + \Sigma)\delta_\beta]^2 = \text{tr}[5\Sigma\delta_\beta\delta_\beta'\Sigma + \delta_\beta'\Sigma\delta_\beta\Sigma + \Delta\Gamma\text{diag}(A_1)\Gamma']^2$$

$$= 25B_2^2 + B_1^2\text{tr}(\Sigma^2) + \Delta^2\text{tr}\{A_0\text{diag}(A_1)\}^2 + 10B_1B_3 + 10\Delta\text{tr}(A_1 \circ A_2)$$

$$+ 2\Delta B_1\text{tr}(A_1 \circ A_3) \quad \text{and}$$

$$E[2\delta_\beta'(X_1 - X_2)(X_1 - X_2)'\Sigma\delta_\beta\delta_\beta'(X_1X_1' + \Sigma)(X_2X_2' + \Sigma)\delta_\beta]$$

$$= E[2\delta_\beta'(X_1X_1' + X_2X_2')\Sigma\delta_\beta\delta_\beta'(X_1X_1' + \Sigma)(X_2X_2' + \Sigma)\delta_\beta]$$

$$= 8E[\delta_\beta'X_1X_1'\Sigma\delta_\beta\delta_\beta'X_1X_1'\Sigma\delta_\beta] + 8[\delta_\beta'\Sigma^2\delta_\beta]^2 = 8\Delta\text{tr}(A_1 \circ A_2) + 8B_1B_3 + 24B_2^2.$$

As $\zeta_4 \geq \zeta_3$, we first derive ζ_4 . It may be shown that

$$\begin{aligned} E(h^2(W_1, W_2, W_3, W_4)) &= \frac{1}{9}E\{\phi(1, 2, 3, 4) + \phi(1, 3, 2, 4) + \phi(1, 4, 2, 3)\}^2 \\ &= \frac{1}{3}E\phi^2(1, 2, 3, 4) + \frac{2}{9}E\phi(1, 2, 3, 4)\phi(1, 3, 2, 4) \\ &\quad + \frac{2}{9}E\phi(1, 2, 3, 4)\phi(1, 4, 2, 3) + \frac{2}{9}E\phi(1, 3, 2, 4)\phi(1, 4, 2, 3) \\ &= \frac{1}{3}E\phi^2(1, 2, 3, 4) + \frac{2}{3}E\phi(1, 2, 3, 4)\phi(1, 3, 2, 4) \end{aligned}$$

where

$$\begin{aligned} \frac{1}{3}E\phi^2(1, 2, 3, 4) &= \frac{1}{3}\sigma^4tr(\Sigma^2) + \frac{4}{3}B_2^2 + \frac{4}{3}B_1B_3 + \frac{4}{3}\sigma^2B_3 + \frac{1}{3}B_1^2tr(\Sigma^2) + \frac{2}{3}\sigma^2tr(\Sigma^2)B_1 \\ &\quad + \frac{1}{3}\{B_1 + \sigma^2\}\Delta tr(A_1 \circ A_3) + \frac{2}{3}\Delta tr(A_1 \circ A_2) + \frac{1}{12}\Delta^2tr\{A_0diag(A_1)\}^2 \quad \text{and} \\ \frac{2}{3}E\phi(1, 2, 3, 4)\phi(1, 3, 2, 4) &= \frac{1}{24}\left\{8\sigma^4tr(\Sigma^2) + 32B_2^2 + 32B_1B_3 + 32\sigma^2B_3 + 16\sigma^2B_1tr(\Sigma^2) \right. \\ &\quad + 8B_1^2tr(\Sigma^2) + 16\Delta tr(A_1 \circ A_2) + 8\Delta\sigma^2tr(A_1 \circ A_3) + 8\Delta B_1tr(A_1 \circ A_3) \\ &\quad \left. + 2\Delta^2tr\{A_0diag(A_1)\}^2\right\}. \end{aligned}$$

It then follows that

$$\begin{aligned} \zeta_4 &= \frac{1}{24}\left\{12\sigma^4tr(\Sigma^2) + 45B_2^2 + 65B_1B_3 + 40\sigma^2B_3 + 10B_1^2tr(\Sigma^2) + 24\sigma^2tr(\Sigma^2)B_1 \right. \\ &\quad \left. + 12\{B_1 + \sigma^2\}\Delta tr(A_1 \circ A_3) + 37\Delta tr(A_1 \circ A_2) + 4\Delta^2tr\{A_0diag(A_1)\}^2\right\}. \end{aligned} \quad (2.8.46)$$

Note that (2.8.45) and (2.8.46) show that ζ_2 and ζ_4 are both the linear combination of $tr(\Sigma^2)$, B_2^2 , B_1B_3 , B_3 , $B_1^2tr(\Sigma^2)$, $B_1tr(\Sigma^2)$, $(B_1 + \sigma^2)tr(A_1 \circ A_3)$, $tr(A_1 \circ A_2)$ and $tr\{A_0diag(A_1)\}^2$.

So it implies that ζ_2 and ζ_4 are of the same order. By Proposition 1, ζ_2 , ζ_3 and ζ_4 are of the same order. Hence, the third and fourth term in the Hoeffding decomposition are all of smaller order. Thus $\text{Var}(T_{n,p}) = \left\{\frac{16}{n}\zeta_1 + \frac{72}{n(n-1)}\zeta_2\right\} \{1 + o(1)\}$. Substituting ζ_1 and ζ_2 , the results in Theorem 2 follow. \square

The following two inequalities will be useful in the proof of Theorem 3. By the Cauchy-Schwarz inequality together with (2.8.44) and (2.8.45), we have

$$\zeta_1 \leq \left\{\left(\frac{1}{2} + \frac{1}{4}\Delta\right)B_1 + \frac{1}{4}\sigma^2\right\}B_3, \quad (2.8.47)$$

$$\zeta_2 \leq \frac{1}{36}\left\{[\sigma^2 + (\Delta + 1)B_1]^2tr(\Sigma^2) + [22\sigma^2 + (43 + 20\Delta)B_1]B_3\right\}. \quad (2.8.48)$$

Proof of Theorem 3 Let

$$\widehat{T}_{n,p} - \|\Sigma\delta_\beta\|^2 = \frac{12}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \tilde{h}_2(W_{i_1}, W_{i_2}) \quad (2.8.49)$$

be the projection of $T_{n,p}$. We can decompose $T_{n,p} - \|\Sigma\delta_\beta\|^2 = \widehat{T}_{n,p} - \|\Sigma\delta_\beta\|^2 + (T_{n,p} - \widehat{T}_{n,p})$, where $T_{n,p} - \widehat{T}_{n,p}$ can still be written as a U-statistics with kernel

$$H(W_1, W_2, W_3, W_4) = \tilde{h}(W_1, W_2, W_3, W_4) - \sum_{1 \leq i_1 < i_2 \leq 4} \tilde{h}_2(W_{i_1}, W_{i_2}). \quad (2.8.50)$$

The projections of H are $H_1(w_1) = -2\tilde{h}_1(w_1)$, $H_2(w_1, w_2) = -2\sum_{i=1}^2\tilde{h}_1(w_i)$ and $H_3(w_1, w_2, w_3) = \tilde{h}_3(w_1, w_2, w_3) - \sum_{i=1}^3\tilde{h}_1(w_i) - \sum_{1\leq i<j\leq 3}\tilde{h}_2(w_i, w_j)$. Thus if the null hypothesis or the local alternatives conditions (2.4.12) hold, $\text{Var}(h_1) = o(n^{-1}\zeta_2)$. By Hoeffding's variance formula, $\text{Var}(\hat{T}_{n,p}) = O(n^{-2}\zeta_2)$ and $\text{Var}(T_{n,p} - \hat{T}_{n,p}) = o(n^{-2}\zeta_2)$. Here we used the fact that ζ_2, ζ_3 and ζ_4 are of the same order as we have shown in Theorem 2. Thus,

$$\frac{T_{n,p} - \|\Sigma\delta_\beta\|^2}{\sqrt{\text{Var}(\hat{T}_{n,p})}} = \frac{\hat{T}_{n,p} - \|\Sigma\delta_\beta\|^2}{\sqrt{\text{Var}(\hat{T}_{n,p})}} + o_p(1).$$

Hence we only need to show that

$$\frac{\hat{T}_{n,p} - \|\Sigma\delta_\beta\|^2}{\sqrt{\text{Var}(\hat{T}_{n,p})}} \xrightarrow{d} N(0, 1). \quad (2.8.51)$$

From (2.8.49), $\hat{T}_{n,p} - \|\Sigma\delta_\beta\|^2 = \hat{T}_{n,p}^{(1)} + \hat{T}_{n,p}^{(2)}$ where

$$\begin{aligned} \hat{T}_{n,p}^{(1)} &= \binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} \left\{ [\delta'_\beta(X_i - X_j) + (\varepsilon_i - \varepsilon_j)](X_i - X_j)' \Sigma \delta_\beta \right. \\ &\quad \left. + [\delta'_\beta(X_i X_i' + \Sigma) + \varepsilon_i X_i'](X_j X_j' + \Sigma) \delta_\beta + \varepsilon_j X_j'(X_i X_i' + \Sigma) \delta_\beta \right\} - 6\|\Sigma\delta_\beta\|^2 \end{aligned}$$

and $\hat{T}_{n,p}^{(2)} = \binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} \varepsilon_i \varepsilon_j X_i' X_j$. Under the assumptions of this theorem and following (2.8.47) and (2.8.48), $\text{Var}(\hat{T}_{n,p}) = \text{Var}(\hat{T}_{n,p}^{(2)})\{1 + o(1)\}$ and $\hat{T}_{n,p}^{(1)}/\sqrt{\text{Var}(\hat{T}_{n,p})} = o_p(1)$. To prove the theorem, we only need to show

$$\hat{T}_{n,p}^{(2)}/\sqrt{\text{Var}(\hat{T}_{n,p}^{(2)})} = \sqrt{\binom{n}{2}} \hat{T}_{n,p}^{(2)}/\sqrt{\sigma^4 \text{tr}(\Sigma^2)} \xrightarrow{d} N(0, 1). \quad (2.8.52)$$

Now write $\tilde{T}_{nk} = \sqrt{\binom{n}{2}} \hat{T}_{n,p}^{(2)} = \sum_{i=2}^k Z_{ni}$ and $\tilde{T}_{nn} = \tilde{T}_{n,p}$, where $Z_{ni} = \sum_{j=1}^{i-1} \varepsilon_i \varepsilon_j X_i' X_j / \sqrt{\binom{n}{2}}$. Let $\mathcal{F}_i = \sigma \left\{ \binom{X_1}{\varepsilon_1}, \dots, \binom{X_i}{\varepsilon_i} \right\}$ be the σ -field generated by $\{(X_j^\tau, \varepsilon_j), j \leq i\}$. It is easy to see that $E(Z_{ni} | \mathcal{F}_{i-1}) = 0$ and it follows that $\{\tilde{T}_{nk}, \mathcal{F}_k : 2 \leq k \leq n\}$ is a zero mean martingale. Let $v_{ni} = E(Z_{ni}^2 | \mathcal{F}_{i-1})$, $2 \leq i \leq n$ and $V_n = \sum_{i=2}^n v_{ni}$. The central limit theorem will hold (Hall and Heyde, 1980) if we can show

$$\frac{V_n}{\text{Var}(\tilde{T}_{n,p})} \xrightarrow{p} 1 \quad (2.8.53)$$

and for any $\epsilon > 0$

$$\sum_{i=1}^n \sigma^{-4} \text{tr}^{-1}(\Sigma^2) E\{Z_{ni}^2 I(|Z_{ni}| > \epsilon \sigma^2 \sqrt{\text{tr}(\Sigma^2)}) | \mathcal{F}_{i-1}\} \xrightarrow{p} 0. \quad (2.8.54)$$

It can be shown that $v_{ni} = \binom{n}{2}^{-1} \sigma^2 \left\{ \sum_{j=1}^{i-1} \varepsilon_j^2 X_j' \Sigma X_j + 2 \sum_{1 \leq j < k < i} \varepsilon_j \varepsilon_k X_j' \Sigma X_k \right\}$ and

$$\begin{aligned} \frac{V_n}{\text{Var}(\widetilde{T}_{n,p})} &= \frac{1}{\binom{n}{2}^2 \text{tr}(\Sigma^2) \sigma^2} \left\{ \sum_{j=1}^{n-1} j \varepsilon_j^2 X_j' \Sigma X_j + 2 \sum_{1 \leq j < k \leq n} \varepsilon_j \varepsilon_k X_j' \Sigma X_k \right\} \\ &= C_{n1} + C_{n2}, \text{ say.} \end{aligned}$$

We know that $E(C_{n1}) = 1$ and

$$\text{Var}(C_{n1}) = \frac{1}{\binom{n}{2}^4 \text{tr}^2(\Sigma^2) \sigma^4} E \left\{ \sum_{j=1}^{n-1} j^2 (\varepsilon_j^4 (X_j' \Sigma X_j)^2 - \text{tr}^2(\Sigma^2) \sigma^4) \right\}.$$

As $\text{tr}(\Sigma^4) = o\{\text{tr}^2(\Sigma^2)\}$ implies $E\{(X_j' \Sigma X_j)^2\} = o(n) \text{tr}^2(\Sigma^2)$. Hence, $\text{Var}(C_{n1}) \rightarrow 0$ and $C_{n1} \xrightarrow{p} 1$. Similarly, $E(C_{n2}) = 0$ and

$$\text{Var}(C_{n2}) = \frac{4}{\binom{n}{2}^4} \left\{ \sum_{i=3}^n \binom{i}{2} + \sum_{i=3}^{n-1} (n-i) \binom{i}{2} \right\} \frac{\text{tr}(\Sigma^4)}{\text{tr}^2(\Sigma^2)}.$$

Thus, $\text{tr}(\Sigma^4) = o\{\text{tr}^2(\Sigma^2)\}$ implies $C_{n2} \xrightarrow{p} 0$. In summary, (2.8.53) holds.

It remains to show (2.8.54). Since

$$E\{Z_{ni}^2 I(|Z_{ni}| > \epsilon \sigma^2 \sqrt{\text{tr}(\Sigma^2)}) | \mathcal{F}_{i-1}\} \leq E(Z_{ni}^4 | \mathcal{F}_{i-1}) / (\epsilon^2 \sigma^4 \text{tr}(\Sigma^2)),$$

by the law of large numbers, we only need to prove that

$$\sum_{i=1}^n E(Z_{ni}^4) = o\{\sigma^4 \text{tr}^2(\Sigma^2)\}. \quad (2.8.55)$$

Let $\kappa_4 = E(\varepsilon^4)$ which is assumed to be finite. Then

$$\begin{aligned} \sum_{i=1}^n E(Z_{ni}^4) &\leq \binom{n}{2}^{-1} \kappa_4^2 \left(3 \text{tr}^2(\Sigma^2) + (6 + 6\Delta + \Delta^2) \text{tr}(\Sigma^4) \right) \\ &\quad + \binom{n}{2}^{-2} \frac{1}{3} (n^3 - 3n^2 + 2n) \kappa_4 \sigma^4 \left(\text{tr}^2(\Sigma^2) + (2 + \Delta) \text{tr}(\Sigma^4) \right). \end{aligned}$$

Under the assumption that $\text{tr}(\Sigma^4) = o\{\text{tr}^2(\Sigma^2)\}$, (2.8.55) follows immediately. This completes the proof. \square

We need the following lemma in the proof of Theorem 4.

Lemma 3 *Let $\eta_1 = \varepsilon_1 X_1 + (X_1 X_1' - \Sigma) \delta_\beta$ and A be a symmetric matrix, where $\delta_\beta = \beta - \beta_0$. Assume $E(Z_{11}^8) < \infty$. Then for some finite positive constants C ,*

$$(i) \quad E(\eta'_1 A \eta_1) = \text{tr}(A \Sigma) \{B_1 + \sigma^2\} + \delta'_\beta \Sigma A \Sigma \delta_\beta + \Delta \text{tr}(\Gamma' A \Gamma \circ \Gamma' \delta_\beta \delta'_\beta \Gamma).$$

$$(ii) \quad E(\eta'_1 A \eta_2)^2 \leq \text{Ctr}(A \Sigma A \Sigma) \{B_1 + \sigma^2\}^2 + C \{B_1 + \sigma^2\} \delta'_\beta \Sigma A \Sigma A \Sigma \delta_\beta + \{\delta'_\beta \Sigma A \Sigma \delta_\beta\}^2.$$

$$(iii) \quad E(\eta'_1 A \eta_1)^2 \leq \text{Ctr}(A \Sigma A \Sigma) \{B_1^2 + \sigma^4\} + \text{Ctr}^2(A \Sigma) \{B_1^2 + \sigma^4\} + C \{\delta'_\beta \Sigma A \Sigma \delta_\beta\}^2 \\ + C \{B_1 + \sigma^2\} \delta'_\beta \Sigma A \Sigma A \Sigma \delta_\beta + \text{Ctr}(A \Sigma) \text{tr}(\Gamma' A \Gamma \circ \Gamma' \delta_\beta \delta'_\beta \Gamma).$$

Proof The first result is immediately obtained from (i) in Lemma 2. We only show (ii) and (iii). For (ii), notice that $E(\eta_2 \eta'_2) = (B_1 + \sigma^2) \Sigma + \Sigma \delta_\beta \delta'_\beta \Sigma + \Gamma \text{diag}(\Gamma' \delta_\beta \delta'_\beta \Gamma) \Gamma'$ and $E(\eta'_1 A \eta_2)^2 = E[E(\eta'_1 A \eta_2 \eta'_2 A \eta_1 | \eta_2)] = E[\eta'_2 A \Sigma A \eta_2 (B_1 + \sigma^2) + \eta'_2 A \Sigma \delta_\beta \delta'_\beta \Sigma A \eta_2 + \Delta \text{tr}(\Gamma' A \eta_2 \eta'_2 A \Gamma \circ \Gamma' \delta_\beta \delta'_\beta \Gamma)]$. Applying the formula in (i) again and note that $\text{tr}(\Gamma' A \Gamma \text{diag}(\Gamma' \delta_\beta \delta'_\beta \Gamma) \Gamma' A \Gamma \circ \Gamma' \delta_\beta \delta'_\beta \Gamma) \leq \text{tr}(A \Sigma A \Sigma) B_1^2$, and $\text{tr}(\Gamma' A \Sigma A \Gamma \circ \Gamma' \delta_\beta \delta'_\beta \Gamma) \leq \text{tr}(A \Sigma A \Sigma) B_1$ the result in (ii) follows immediately. Using Cauchy-Schwartz inequality, we have

$$E(\eta'_1 A \eta_1)^2 \leq 24E[(X'_1 A X_1 - \text{tr}(A \Sigma))(X'_1 \delta_\beta \delta'_\beta X_1 - B_1)] + 24B_1^2 \text{tr}^2(A \Sigma) \\ + 24\text{tr}^2(A \Sigma) E[(X'_1 \delta_\beta \delta'_\beta X_1 - B_1)]^2 + 24B_1^2 E[(X'_1 A X_1 - \text{tr}(A \Sigma))]^2 \\ + 24E[X'_1 A \Sigma \delta_\beta \delta'_\beta X_1]^2 + 6(\delta'_\beta \Sigma A \Sigma \delta_\beta)^2 + 8\sigma^2 E[(X'_1 A X_1)^2 X'_1 \delta_\beta \delta'_\beta X_1] \\ + 8\delta'_\beta \Sigma A \Sigma A \Sigma \delta_\beta + 2\sigma^4 E(X'_1 A X_1)^2. \quad (2.8.56)$$

The first term on the right hand side of (2.8.56) is bounded by $\text{Ctr}(A \Sigma A \Sigma) B_1^2$, which is an implication of (iii) in Lemma 2. The 7-th term on the right hand side of (2.8.56) can be calculated by the formula given in (iv) in Lemma 2. Applying (i) in Lemma 2, all the other expectations in (2.8.56) can be calculated. Then the result (iii) could be obtained by some algebras. \square

Proof of Theorem 4 We first show the conclusion in (i). Following the proof of Theorem 3, we only need to show that (2.8.51) hold under the fixed alternative condition (2.4.18). Write

$$\widehat{T}_{n,p} - \|\Sigma \delta_\beta\|^2 = \widehat{T}_{n,p}^{(1)} + \widehat{T}_{n,p}^{(2)} \text{ where}$$

$$\widehat{T}_{n,p}^{(1)} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \left\{ [\delta'_\beta (X_i - X_j) + (\varepsilon_1 - \varepsilon_2)] (X_i - X_j)' \Sigma \delta_\beta \right\} - 2\|\Sigma \delta_\beta\|^2$$

$$\text{and } \widehat{T}_{n,p}^{(2)} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \{ \eta'_i \eta_j + 2\delta'_\beta \Sigma \eta_j + 2\delta'_\beta \Sigma \eta_i \}.$$

Utilizing the facts listed in the proof of Theorem 2, we see that $\widehat{T}_{n,p}^{(1)}$ is a smaller order term of $\widehat{T}_{n,p}^{(2)}$. Therefore, we only need to show the asymptotic normality of $\widehat{T}_{n,p}^{(2)}$. Using Proposition 1, $\text{Var}(\widehat{T}_{n,p}^{(2)}) = \frac{4(n-2)}{n(n-1)}\zeta_1 + \frac{2}{n(n-1)}\zeta_2$, where $\zeta_1 = E(2\delta'_\beta \Sigma \eta_1)^2$ and $\zeta_2 = E(\eta'_i \eta_j + 2\delta'_\beta \Sigma \eta_j + 2\delta'_\beta \Sigma \eta_i)^2$. Under condition (2.4.18), it is can be shown that $\text{Var}(\widehat{T}_{n,p}^{(2)}) = \frac{2}{n(n-1)}\zeta_2\{1 + o(1)\}$.

Let $\tilde{T}_{n,p} = \sum_{i=2}^n Z_{ni}$, where $Z_{ni} = \binom{n}{2}^{-\frac{1}{2}} \sum_{j=1}^{i-1} \{\eta'_i \eta_j + 2\delta'_\beta \Sigma \eta_j + 2\delta'_\beta \Sigma \eta_i\}$. Then $Z_{ni}^2 = Z_{ni}^{2(1)} + Z_{ni}^{2(2)}$, where $Z_{ni}^{2(1)} = \binom{n}{2}^{-1} \sum_{j=1}^{i-1} \{\eta'_i \eta_j + 2\delta'_\beta \Sigma \eta_j + 2\delta'_\beta \Sigma \eta_i\}^2$ and

$$Z_{ni}^{2(2)} = \binom{n}{2}^{-1} \sum_{1 \leq j < k \leq i-1} \{\eta'_i \eta_j + 2\delta'_\beta \Sigma \eta_j + 2\delta'_\beta \Sigma \eta_i\} \{\eta'_i \eta_k + 2\delta'_\beta \Sigma \eta_k + 2\delta'_\beta \Sigma \eta_i\}.$$

It follows that

$$\begin{aligned} V_{ni}^{(1)} &= E(Z_{ni}^{2(1)} | \mathcal{F}_{i-1}) = \binom{n}{2}^{-1} \sum_{j=1}^{i-1} \left\{ (B_1 + \sigma^2) \eta'_j \Sigma \eta_j + \Delta \text{tr}(\Gamma' \eta_j \eta'_j \Gamma \circ \Gamma' \delta_\beta \delta'_\beta \Gamma) \right. \\ &\quad + 4(B_1 + \sigma^2) \delta'_\beta \Sigma^3 \delta_\beta + 5\eta'_j \Sigma \delta_\beta \delta'_\beta \Sigma \eta'_j + 4\Delta \text{tr}(\Gamma' \Sigma \delta_\beta \delta'_\beta \Sigma \Gamma \circ \Gamma' \delta_\beta \delta'_\beta \Gamma) + 4B_2^2 \\ &\quad \left. + 4(B_1 + \sigma^2) \delta'_\beta \Sigma^2 \eta_j + 4\delta'_\beta \Sigma^2 \delta_\beta \delta'_\beta \Sigma^2 \eta_j + 4\Delta \text{tr}(\Gamma' \eta_j \delta'_\beta \Sigma \Gamma \circ \Gamma' \delta_\beta \delta'_\beta \Gamma) \right\}. \end{aligned}$$

Let $V_{n1} = \sum_{i=2}^n V_{ni}^{(1)}$. It is easy to see that $E(V_{n1}) = \zeta_2$. Then

$$\begin{aligned} \text{Var}(V_{n1}) &\leq \binom{n}{2}^{-2} \sum_{j=1}^{n-1} j^2 \left\{ 6(B_1 + \sigma^2)^2 E[\eta'_j \Sigma \eta_j - E(\eta'_j \Sigma \eta_j)]^2 + 6\Delta^2 E(\eta'_j \Sigma \eta_j)^2 \right. \\ &\quad + 30E[\eta'_j \Sigma \delta_\beta \delta'_\beta \Sigma \eta_j - E(\eta'_j \Sigma \delta_\beta \delta'_\beta \Sigma \eta_j)]^2 + 24B_1^2 E[\eta'_j \Sigma \delta_\beta \delta'_\beta \Sigma \eta_j] \\ &\quad \left. + 24\{(B_1 + \sigma^2)^2 + \Delta^2 B_1^2\} E[\eta'_j \Sigma^2 \delta_\beta \delta'_\beta \Sigma^2 \eta_j] \right\}. \end{aligned}$$

Applying Cauchy-Schwarz inequality, the summand on the right hand side of the above inequality is smaller than $C \left\{ (\delta'_\beta \Sigma \delta_\beta)^4 [\text{tr}(\Sigma^4) + \text{tr}^2(\Sigma^2)] + B_3^2 B_1^2 + B_1^3 B_5 + B_2^4 \right\}$. Using Cauchy-Schwarz inequality, we can shown that

$$B_1^3 B_5 \geq B_3^2 B_1^2 \geq B_2^4. \quad (2.8.57)$$

Hence, $\text{Var}(V_{n1}) \leq C \binom{n}{2}^{-2} \sum_{j=1}^{n-1} j^2 \left\{ B_1^4 [\text{tr}(\Sigma^4) + \text{tr}^2(\Sigma^2)] + B_1^3 B_5 \right\}$. As $\text{tr}(\Sigma^4) = o\{\text{tr}^2(\Sigma^2)\}$ implies that $\lambda_p^4 = o\{(\sum_{i=1}^p \lambda_i^2)^2\}$, $B_1^3 B_5 = o\{B_1^4 \text{tr}^2(\Sigma^2)\}$. Therefore, $\text{Var}(V_{n1})/\zeta_2^2 \rightarrow 0$ as $n \rightarrow \infty$, which means $V_{n1} \xrightarrow{p} \zeta_2$.

$$V_{ni}^{(2)} = E(Z_{ni}^{(2)2} | \mathcal{F}_{i-1}) = \binom{n}{2}^{-1} \sum_{1 \leq j < k \leq i-1} I_{2jk},$$

where $I_{2jk} = I_{1jk} + I_{1k} + I_{1j} + C_0$, $C_0 = 4\delta'_\beta \Sigma^3 \delta_\beta (B_1 + \sigma^2) + 4B_2^2 + \Delta \text{tr}(\Gamma' \Sigma \delta_\beta \delta'_\beta \Sigma \Gamma \circ \Gamma' \delta_\beta \delta'_\beta \Gamma)$ and

$$\begin{aligned} I_{1jk} &= (B_1 + \sigma^2) \eta'_j \Sigma \eta_k + 5\eta_j \Sigma \delta_\beta \delta'_\beta \eta_k + \Delta \text{tr}(\Gamma' \eta_k \eta_j \Gamma \circ \Gamma' \delta_\beta \delta'_\beta \Gamma) \quad \text{and} \\ I_{1k} &= 2(B_1 + \sigma^2) \eta'_k \Sigma^2 \delta_\beta + 2\eta_k \Sigma \delta_\beta \delta'_\beta \Sigma^2 \delta_\beta + 2\Delta \text{tr}(\Gamma' \eta_k \delta'_\beta \Sigma \Gamma \circ \Gamma' \delta_\beta \delta'_\beta \Gamma). \end{aligned}$$

Let $V_{n2} = \sum_{i=2}^n V_{ni}^{(2)}$. Then $E(V_{n2}) = 0$ and

$$\begin{aligned} \text{Var}(V_{n2}) &= \binom{n}{2}^{-2} \sum_{k=2}^{n-1} \sum_{j=1}^{k-1} (n-k)^2 E\{I_{2jk} - E(I_{2jk})\}^2 \\ &+ \binom{n}{2}^{-2} \sum_{k=2}^{n-1} \sum_{j < j_1}^{k-1} (n-k)^2 E(I_{1k}^2) + \binom{n}{2}^{-2} \sum_{k < k_1}^{n-1} \sum_{j=1}^k (n-k)(n-k_1) E(I_{1j}^2). \end{aligned}$$

It can be shown that, by Lemma 3 and (2.8.57),

$$E\{I_{2jk} - E(I_{2jk})\}^2 \leq C \left\{ (B_1 + \sigma^2)^4 \text{tr}(\Sigma^4) + (B_1 + \sigma^2)^3 B_5 \right\}$$

and $E\{I_{1k}\}^2 \leq C(B_1 + \sigma^2)^3 B_5$. Thus, under the condition given in Theorem 4, as $n \rightarrow \infty$, $\text{Var}(V_{n2})/\zeta_2^2 \leq C E\{I_{2jk} - E(I_{2jk})\}^2/\zeta_2^2 \rightarrow 0$, which implies that $V_{n2}/\zeta_2 \xrightarrow{P} 0$. In summary, we have $E(\sum_{i=2}^n Z_{ni}^2 | \mathcal{F}_{i-1}) \xrightarrow{P} \zeta_2$.

It remains to establish the second condition (2.8.55). Note that

$$E(Z_{ni}^4) \leq 2E(Z_{ni}^{2(1)})^2 + 2E(Z_{ni}^{2(2)})^2 =: J_{1i} + J_{2i}.$$

Let $S_{ij} = \eta'_i \eta_j + 2\delta'_\beta \Sigma \eta_j + 2\delta'_\beta \Sigma \eta_i$. Then

$$J_{1i} = 2 \binom{n}{2}^{-2} \sum_{j=1}^{i-1} E\{S_{ij}^4\} + 2 \binom{n}{2}^{-2} \sum_{j \neq k}^{i-1} E\{S_{ij}^2 S_{ik}^2\} \leq (2i-2) \binom{n}{2}^{-2} \sum_{j=1}^{i-1} E\{S_{ij}^4\}.$$

By Cauchy-Schwarz inequality, $E(S_{ij}^4) \leq 8E(\eta'_i \eta_j \eta'_j \eta_i)^2 + 2^{11} E(\eta'_j \Sigma \delta_\beta \delta'_\beta \Sigma \eta_j)^2$. Again, by Lemma 3, $E(\eta'_i \eta_j \eta'_j \eta_i)^2 \leq C(B_1^2 + \sigma^4)^2 \{ \text{tr}(\Sigma^4) + \text{tr}^2(\Sigma^2) \} + C B_1^3 B_3^2 + C B_1^3 B_5 + C B_2^4$ and $E(\eta'_j \Sigma \delta_\beta \delta'_\beta \Sigma \eta_j)^2 \leq C B_1^2 B_3^2 + C B_2^4$.

Therefore, $\sum_{i=1}^n J_{1i}/\zeta_2^2 = \frac{4(2n^3 - 3n^2 + n)}{3n^2(n-1)^2} E(S_{ij}^4)/\zeta_2^2 \rightarrow 0$ as $n \rightarrow \infty$. Write

$$\begin{aligned} J_{2i} &= 4 \binom{n}{2}^{-2} \sum_{j \neq k}^{i-1} E\{S_{ij}^2 S_{ik}^2\} + 8 \binom{n}{2}^{-2} \sum_{j \neq k \neq l}^{i-1} E\{S_{ij}^2 S_{ik} S_{il}\} \\ &+ 2 \binom{n}{2}^{-2} \sum_{j \neq k \neq l \neq m}^{i-1} E\{S_{ij} S_{ik} S_{il} S_{im}\} =: J_{2i}^{(1)} + J_{2i}^{(2)} + J_{2i}^{(3)}, \end{aligned}$$

where

$$\begin{aligned}
J_{2i}^{(2)} &= 8 \binom{n}{2}^{-2} \sum_{j \neq k \neq l}^{i-1} E\{4(\delta'_\beta \Sigma \eta_i)^2 (\eta'_i \eta_j)^2 + 16(\delta'_\beta \Sigma \eta_i)^4 + 16(\delta'_\beta \Sigma \eta_i)^2 (\delta'_\beta \Sigma \eta_j)^2\} \\
&\leq C \frac{i(i-1)(i-2)}{n^2(n-1)^2} E(S_{ij}^4). \\
J_{2i}^{(3)} &= 32 \binom{n}{2}^{-2} \sum_{j \neq k \neq l \neq m}^{i-1} E(\delta'_\beta \Sigma \eta_i)^4 \leq C \frac{i(i-1)(i-2)(i-3)}{n^2(n-1)^2} B_1^2 B_3^2.
\end{aligned}$$

Under conditions given in Theorem 4, $\sum_{i=1}^n J_{1i}/\zeta_2^2 \rightarrow 0$ as $n \rightarrow \infty$. Hence, $\sum_{i=1}^n E(Z_{ni}^4) = o(\zeta_2^2)$.

By the Martingale Central Limit Theorem, the asymptotic normality holds for $T_{n,p}$, which completes the proof of part (i).

Now we turn to part (ii). Let $\widehat{T}_{n,p}^F - \|\Sigma \delta_\beta\|^2 = \frac{2}{n} \sum_{i=1}^n \tilde{h}_1(W_i)$ be the projection of $T_{n,p}$. Similar to the local alternatives cases, we may decompose $T_{n,p} - \|\Sigma \delta_\beta\|^2$ into two parts. One is $\widehat{T}_{n,p}^F - \|\Sigma \delta_\beta\|^2$, which is asymptotically normal and the other is $T_{n,p} - \widehat{T}_{n,p}^F$ which is a U-statistic with kernel $H^F(W_1, W_2, W_3, W_4) = \tilde{h}(W_1, W_2, W_3, W_4) - \sum_{i=1}^4 \tilde{h}_1(W_i)$. It is straightforward to show that the projections of H^F are $H_1^F(w_1) = 0$, $H_2^F(w_1, w_2) = \tilde{h}_2(w_1, w_2) - \sum_{i=1}^2 \tilde{h}_1(w_i)$ and $H_3^F(w_1, w_2, w_3) = \tilde{h}_3(w_1, w_2, w_3) - \sum_{i=1}^3 \tilde{h}_1(w_i)$. Suppose conditions (2.4.19) hold, $\text{Var}(h_2) = o\{\text{Var}(h_1)\}$. We observe that $\text{Var}(\widehat{T}_{n,p}^F) = O(n^{-1}\zeta_1)$ and $\text{Var}(T_{n,p} - \widehat{T}_{n,p}^F) = O(n^{-2}\zeta_1)$. Therefore the asymptotic normality can be obtained by the asymptotic normality of $\widehat{T}_{n,p}^F$. The latter is obtained by the conventional central limit theorem. This completes the proof of part (ii). \square

References

- Ahmad, M., Werner, C. and Brunner E. (2008). Analysis of high-dimensional repeated measures designs: The one sample case. *Computational Statistics and Data Analysis*, 53, 416-427.
- Akritis, M. and Arnold, S. (2000). Asymptotics for analysis of variance when the number of levels is large. *Journal of the American Statistical association*, 95, 212-226.
- Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. Wiley.
- Bathke, A. and Harrar, S. (2008). Nonparametric methods in multivariate factorial designs for large number of factor levels. *Journal of Statistical Planning and Inference*, 138, 588-610.
- Bathke, A. and Lankowski, D. (2005). Rank procedures for a large number of treatments. *Journal of Statistical Planning and Inference*, 133, 223-238.
- Brunner, E., Dette, H. and Munk, A. (1997). Box-Type Approximations in nonparametric factorial designs. *Journal of the American Statistical association*, 92, 1494-1502.
- Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6, 311-329.
- Boos, D. D. and Brownie, C. (1995). ANOVA and rank tests when the number of treatments is large. *Statistics and Probability Letters*, 23, 183-191.
- Chen, S. X. and Qin, Y. (2010). A two sample test for high dimensional data with applications to gene-set testing. *The Annals of Statistics*. 38, 808-835.
- Chen, S. X., Zhang, L.X. and Zhong, P.S. (2010). Tests for high dimensional covariance matrices. *Journal of the American Statistical Association*, 109, 810-819.

- Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *The Annals of Statistics*, 35, 2313-2351.
- Efron, B. and Tibshirani, R. (2007). On testing the significance of sets of genes. *The Annals of Applied Statistics*, 1, 107-129.
- Fan, J. and Hall, P. and Yao, Q. (2007). To how many simultaneous hypothesis tests can normal student's t or bootstrap calibrations be applied. *Journal of the American Statistical Association*, 102, 1282-1288.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348-1360.
- Fan, J. and Lv, J. (2008). Sure independence screening for ultra-high dimensional feature space (with discussion). *Journal of Royal Statistical Society, Series B*, 70, 849-911.
- Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools (with discussion). *Technometrics*, 35, 109-148.
- The Gene Ontology Consortium. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics*, 25, 25-29.
- Goeman, J., Finos, L. and van Houwelingen, J. C. (2009). Testing against a high dimensional alternative in the generalized linear model: asymptotic alpha-level control. *Manuscript*.
- Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its application*. New York : Academic Press.
- Harrar, S. and Bathke, A. (2008). Nonparametric methods for unbalanced multivariate data and many factor levels, *Journal of Multivariate Analysis*. 99, 1635-1664.
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19, 293-325.

- Huang, J., Wang, D. 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, 814-829.
- Kosorok, M. R. and Ma, S. (2007). Marginal asymptotics for the “large p, small n” paradigm: With applications to microarray data. *The Annals of Statistics*, 35, 1456-1486.
- Lee, A. J. (1990). *U-statistics: theory and practice*. Marcel Dekker.
- Lkhagvadorj, S., Qu, L., Cai, W., Couture, O., Barb, C., Hausman, G., Nettleton, D., Anderson, L., Dekkers, J. and Tuggle, C. (2009). Microarray gene expression profiles of fasting induced changes in liver and adipose tissues of pigs expressing the melanocortin-4 receptor D298N variant. *Physiol. Genomeics*, 38, 98-111, 2009.
- Newton, M., Quintana, F., Den Boon, J., 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, 85-106.
- Portnoy, S. (1984). Asymptotic behavior of the M-estimators of p-regression parameters when p^2/n is large: Consistency. *The Annals of Statistics*, 12, 1298-1309.
- Portnoy, S. (1985). Asymptotic behavior of the M-estimators of p-regression parameters when p^2/n is large: Normal approximation. *The Annals of Statistics*, 13, 1403-1417.
- Rao, C. R., Toubert, H., Shalabh and Heumann, C. (2008). *Linear models and generalizations*. Springer, New York.
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley and Sons.
- 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. and Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102, 15545-15550.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of Royal Statistical Society, Series B*, 58, 267-288.

- Wang H.S. (2009). Forward regression for ultra-high dimensional variable screening. *Journal of the American Statistical Association*, 104, 1512-1524.
- Wang, H. and Akritas, M. (2009). Rank tests in heteroscedastic multi-way HANOVA. *Journal of Nonparametric Statistics*, 21, 663-681
- Zhang, C.H. and Huang, J. (2008). The sparsity and bias of the Lasso selection in high-dimensional linear regression. *The Annals of Statistics*, 36, 1567-1594.

CHAPTER 3. Threshold Test for High Dimensional Mean under Dependency

Song Xi Chen and Pingshou Zhong

Department of Statistics

Iowa State University

Ames, IA 50011, USA

Abstract

We consider a test for high dimensional means under sparsity and dependency. A threshold test statistic is proposed, and the asymptotic distribution is obtained for dependent data under the “large p , small n ” paradigm without a specific distribution assumption. To avoid selection of threshold parameter, we also propose a maximum test, which maximizes the standardized threshold test statistic over a range of threshold values. It is shown that the maximum threshold test is able to attain the optimal detection boundary (Donoho and Jin, 2004). Our analysis provides conditions under which the threshold test based on an independence assumption can be justified for weakly dependent data.

KEY WORDS: Detection boundary; Large deviation; Large p , small n ; Sparsity; Strong mixing; Threshold test.

3.1 Introduction

Assume we have independent identically distributed p -variate random vectors $X_1, \dots, X_n \sim F$ and

$$X_j = W_j + \mu \quad j = 1, \dots, n, \tag{3.1.1}$$

where $\mu = (\mu_1, \dots, \mu_p)^T$ is a p -dimensional unknown vector, $W_j = (W_{1j}, \dots, W_{pj})^T$ and $\{W_{ij}\}_{i=1}^p$ is a sequence of dependent random variables with mean 0 and variance σ_i^2 . We are interested in testing the high dimensional mean vector

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0. \quad (3.1.2)$$

We assume the nonzero components in μ are sparse, namely most of μ_i s are 0 and only a small portion $p^{-\beta}$ of them are non-zero. Fan (1996), Fan and Lin (1998), Donoho and Jin (2004) and Hall and Jin (2008) elaborated various applications of above high dimensional test problem, including signal detection, analysis of variance for functional data and goodness-of-fit tests for distributions.

Many traditional multivariate tests suffer low power and may even be invalid in “large p , small n ” scenarios. In an important work of Bai and Saranadasa (1996), they raised such concern about Hotelling’s T^2 -test when $p/n \rightarrow c \in (0, 1)$ and proposed a test that is valid when $p/n \rightarrow c \in (0, \infty)$. Chen and Qin (2010) further improved the test proposed by Bai and Saranadasa (1996) by allowing p to grow much faster than n . See also Chen, Zhang and Zhong (2010), Ledoit and Wolf (2002), Goeman *et al.* (2006) and Zhong and Chen (2011) for high dimensional tests in testing covariances and regression coefficients. These proposals are designed to detect weak signals that are non-sparse, i.e., the signals exist in many dimensions. These tests could lose power in sparse settings (see the simulation results Zhong and Chen, 2011). Another way testing (3.1.2) could be based on multiple comparisons. There is a huge literature on multiple comparisons. See for example, Simes (1986), Benjamini and Hochberg (1995) and Storey (2002, 2003). Principle methods are used to control the family-wise error rate or the false discovery rate (FDR) based on the p-values of the marginal tests, which do not efficiently use the dependence among the components of X_i . Furthermore, these methods are not able to provide a p-value for indicating the significance of the grouped hypothesis (3.1.2). As shown in Donoho and Jin (2004), the FDR procedure is not able to attain the optimal detection boundary in moderately sparse case ($0.5 < \beta < 0.75$), in the sense that under the detection boundary all the tests are asymptotically powerless (Donoho and Jin, 2004; Chen and Xu, 2011).

This chapter aims at improving the above high dimensional tests when the signal is sparse and faint. We consider the following threshold statistic for testing hypothesis (3.1.2)

$$T_n(\lambda_n) = \sum_{i=1}^p Y_{i,n} I\{Y_{i,n} > \lambda_n\} \quad (3.1.3)$$

where $Y_{i,n} = n\bar{X}_i^2$, $\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}/\sigma_i$ is the scaled sample mean of the i -th component of X and $\lambda_n = 2s \log(p)$ is the level of threshold where $0 < s < 1$. Without causing confusion, we will suppress λ_n and write $T_n(\lambda_n)$ as T_n except in Section 3.4. For discussion simplicity, we shall assume that σ_i^2 is known. When σ_i^2 is unknown, we may use the sample variance to estimate σ_i^2 which results in the t-statistic. This estimate would put less restrictive conditions on the validity of our test procedure, because the large deviation results for t-statistic, a self-normalized statistic, can be established under mild moment conditions (see Shao, 1997; Wang and Hall, 2009 and Delaigle *et al.*, 2011). However, to keep the essential, we will continue assuming that σ_i^2 is known and $\sigma_i^2 = 1$. We also extend the threshold test to a maximum test, which maximizes the standardized $T(\lambda_n)$ in searching for a good threshold λ_n . We show that the maximum threshold test is able to achieve the optimal detection boundary. In some part of the detectable region, the maximum test could achieve the same power as an oracle test up to an order $\log(p)$, where the oracle test is constructed by assuming that the exact locations of non-zero signals were known in advance. Donoho and Jin (2004) used such an idea in the so-called “higher criticism” for detecting sparse and weak signals by maximizing the normalized indicator function $I\{Y_i > \sqrt{2s \log(p)}\}$ over a range of $s \in (0, 1)$ where Y_i is the test statistic with standard normal distribution for each component.

Using a threshold statistic to test (3.1.2) is particularly appealing when the signal is sparse (the location of signal is unknown) and the components of X_j are dependent to each other. Because of the high dimensionality and sparse signal, most of the components of X_j contribute to noise. The threshold is designed to remove the noisy components and retain the components that contribute to the signal. The idea of threshold was used in selecting significant wavelet coefficients in Donoho and Johnstone (1994). Moreover, the formulation (3.1.3) allows us to account for dependence explicitly. In this Chapter, we consider the asymptotic distributions of T_n by assuming that $\{X_{ij}\}_{i=1}^p$ is a strong mixing dependent sequence. Although the dependence

exists in the original data, we show that the effect of dependence on the threshold test statistics is small and asymptotically negligible under some conditions. In other words, the threshold test building on the independence assumption may be applicable to weakly dependent data without causing large error. This feature mimics the asymptotic tail independence of two random variables that are jointly bivariate normally distributed with correlation less than one (Sibuya, 1960).

In a closely related work, Fan (1996) considered testing (3.1.2) by assuming the components of X are independent identically normally distributed with mean μ_i and unit variance. Despite the common use of the normality assumption in many works, in reality, data are rarely exactly normal. The robustness of the distribution assumption is a concern in practice, especially for high dimension data, because the level accuracy of the high dimensional simultaneous test often depends on the accuracy of the distribution approximation. Fan, Hall and Yao (2006) showed that if $\log(p) = o(n^{1/3})$, using the p-values calculated from the normal approximation for multiple comparison is still able to control the false discovery rate (FDR) or false family-wise error rate (FWER) at the given level. Delaigle, Hall and Jin (2011) showed that the standard higher criticism (Donoho and Jin, 2004) based on normality assumption could perform poorly when the underlying data deviate from the normal distribution. These works motivate us to consider the robustness of the distribution assumption in the threshold test. We study the threshold test for a class of distributions satisfying the Cramér condition. We show that within this class, the normal approximation to marginal distribution of X is justified.

The rest of the chapter is organized as follows. In Section 3.2, we approximate the mean and variance of the threshold test statistic using the large deviation method. The asymptotic distribution of T_n under the “large p , small n ” is provided in Section 3.3. We extend the threshold test statistic to a maximum test in Section 3.4. The detectable region of the maximum test and the best power are discussed in Section 3.5. Section 3.6 summarizes the simulation results. All the technical details are relegated to the Appendix.

3.2 Large Deviation Approximation to the Mean and Variance

We begin by deriving the mean and variance of the threshold test statistic (3.1.3). Because the underlying distribution of X is unknown and purely nonparametric, direct calculation is impossible. However, we could approximate the mean and variance with enough accuracy. Due to the threshold, the moments of the test statistic highly depend on the tail of the distribution. Notice that

$$E\{YI(Y > \lambda_n)\} = \lambda_n P(Y \geq \lambda_n) + \int_{\lambda_n}^{\infty} P(Y \geq z) dz.$$

Since $\lambda_n \rightarrow \infty$ as $p \rightarrow \infty$, we approximate the tail probability $P(Y \geq z)$ using the large deviation results. The large deviation results for random variable can be found in Petrov (1995). To study the variance of T_n , we further establish a similar large deviation result for a bivariate random vector, which is given in the Appendix.

Let $\eta_{ni}^- = \sqrt{\lambda_n} - \sqrt{n}\mu_i$, $\eta_{ni}^+ = \sqrt{\lambda_n} + \sqrt{n}\mu_i$ and $Z_{i,n}(\lambda_n) = Y_{i,n}I(Y_{i,n} > \lambda_n)$. Throughout this chapter we use $L_p = C \log^b(p)$ to denote slow varying functions for some constants b and positive C , $\phi(\cdot)$ and $\bar{\Phi}(\cdot)$ are the density and survival functions of the standard normal distribution, respectively. Let ρ_k be the correlation coefficient between X_{1i} and $X_{(k+1)i}$ and write $\rho_1 = \rho$ for simplicity. Let

$$Q(a_{n1}, a_{n2}; \rho) = \phi(a_{n1})\bar{\Phi}\left(\frac{a_{n2} - \rho a_{n1}}{\sqrt{1 - \rho^2}}\right), \quad q(a_{n1}, a_{n2}; \rho) = \phi(a_{n1})\phi\left(\frac{a_{n2} - \rho a_{n1}}{\sqrt{1 - \rho^2}}\right) \quad \text{and}$$

$$U(a, b; \rho) = \{2\pi(1 - \rho^2)^{\frac{1}{2}}\}^{-1} \int_a^{\infty} \int_b^{\infty} \exp\left\{-\frac{y_1^2 + y_2^2 - 2\rho y_1 y_2}{2(1 - \rho^2)}\right\} dy_1 dy_2.$$

We write $\mu_{T_n,0}$ and $\sigma_{T_n,0}^2$ as the mean and variance of T_n under the null hypothesis and $\mu_{T_n,1}$ and $\sigma_{T_n,1}^2$ as the corresponding mean and variance under the alternative. The following theorem summarizes the mean and variance of the test statistics T_n , whose proof is given in the Appendix. This theorem suggests that under some conditions, we could approximate the mean and variance of the test statistic T_n as though $\sqrt{n}\bar{X}_i$ are normally distributed.

Theorem 1 *Assume $\lambda_n = o(n^{1/3})$, $n\mu_i^2 = O(\log p)$ and there exists a positive H such that $E(e^{h'(X_{ki}^d, X_{li}^d)}) < \infty$ for $h \in [-H, H] \times [-H, H]$, $1 < d \leq 2$ and any $k \neq l \in \{1, \dots, p\}$. Then $E(T_n) = \sum_{i=1}^p G_{T_n,i}$, where*

$$G_{T_n,i} = \left\{ (\eta_{ni}^+) \phi(\eta_{ni}^-) + (\eta_{ni}^-) \phi(\eta_{ni}^+) + (n\mu_i^2 + 1)(\bar{\Phi}(\eta_{ni}^-) + \bar{\Phi}(\eta_{ni}^+)) \right\} \{1 + o(1)\}$$

and

$$\text{Var}(T_n) = \left\{ \sum_{i=1}^p \nu(\mu_i, \lambda_n) + 2 \sum_{i=1}^p \sum_{j=i+1}^p \gamma_{j-i}(\sqrt{n}\mu_i, \sqrt{n}\mu_j, \lambda_n) \right\} \{1 + o(1)\} \quad (3.2.4)$$

where

$$\begin{aligned} \nu(\mu_i, \lambda_n) &= [(\sqrt{\lambda_n})^3 + (\sqrt{n}\mu_i)^3 + \sqrt{n}\mu_i\lambda_n + n\sqrt{\lambda_n}\mu_i^2 + 5\sqrt{n}\mu_i + 3\sqrt{\lambda_n}]\phi(\eta_{ni}^-) \\ &\quad + [(\sqrt{\lambda_n})^3 - (\sqrt{n}\mu_i)^3 - \sqrt{n}\mu_i\lambda_n + n\sqrt{\lambda_n}\mu_i^2 - 5\sqrt{n}\mu_i + 3\sqrt{\lambda_n}]\phi(\eta_{ni}^+) \\ &\quad + [n^2\mu_i^4 + 6n\mu_i^2 + 3](\bar{\Phi}(\eta_{ni}^-) + \bar{\Phi}(\eta_{ni}^+)) - G_{T_n, i}^2 \quad \text{and} \end{aligned} \quad (3.2.5)$$

$$\gamma_{j-i}(\sqrt{n}\mu_i, \sqrt{n}\mu_j, \lambda_n) = \tilde{E}\{Z_{i,n}(\lambda_n)Z_{j,n}(\lambda_n)\} - G_{T_n, i}G_{T_n, j} \quad (3.2.6)$$

with the expression of $\tilde{E}\{Z_{i,n}(\lambda_n)Z_{j,n}(\lambda_n)\}$ given in the Appendix.

Using Theorem 1, we now evaluate the order of the mean and variance of test statistics T_n under model (3.1.1) when all the non-zero signal μ_i are on the same level $\sqrt{2r \log(p)}/n$ in corollary 1.

Corollary 1 *Assume the conditions in Theorem 1 and $\sum_k |\rho_k| < \infty$. If $p^{1-\beta}$ non-zero components of X with mean $\sqrt{2r \log(p)}/n$, then*

$$\begin{aligned} E(T_n) &= L_p^{(1)} \max\{p^{1-\beta}, p^{1-s}\}I(r > s) + L_p^{(2)} \max\{p^{1-(\sqrt{s}-\sqrt{r})^2-\beta}, p^{1-s}\}I(r < s) \quad \text{and} \\ \text{Var}(T_n) &= L_p^{(1)} \max\{p^{1-\beta}, p^{1-s}\}I(r > s) + L_p^{(2)} \max\{p^{1-(\sqrt{s}-\sqrt{r})^2-\beta}, p^{1-s}\}I(r < s) \end{aligned}$$

where $L_p^{(1)}$ and $L_p^{(2)}$ may be different slow varying functions in each appearance.

It is interesting to notice that the mean and variance given in Corollary 1 are in the same order as that of T_n for X having independent components and have the same slow varying functions $L_p^{(1)}$ and $L_p^{(2)}$ in the leading order. This implies that the threshold test building on the independent assumption is valid for weakly dependent data in the leading order asymptotically.

To appreciate this, let S_0 be the set of indices where $\mu_i = 0$ and S_1 be the set of indices where $\mu_i \neq 0$.

$$T_n = \sum_{i \in S_0} n\bar{X}_i^2 I\{n\bar{X}_i^2 > \lambda_n\} + \sum_{i \in S_1} n\bar{X}_i^2 I\{n\bar{X}_i^2 > \lambda_n\} := T_n(S_0) + T_n(S_1).$$

Then

$$\begin{aligned}
E(T_n) &= E\{T_n(S_0)\} + E\{T_n(S_1)\} = 2\sqrt{2s \log p} \phi(\sqrt{2s \log p})(1 - p^{-\beta})p \\
&\quad + \left[2 \log p (\sqrt{r} + \sqrt{s}) \phi(2(\sqrt{s} - \sqrt{r})\sqrt{\log p}) \right. \\
&\quad \left. + (2r \log p + 1) \bar{\Phi}(2(\sqrt{s} - \sqrt{r})\sqrt{\log p}) \right] p^{1-\beta} \\
&= L_p p^{1-s} + [L_p p^{1-\beta} I(r > s) + L_p p^{1-(\sqrt{s}-\sqrt{r})^2-\beta} I(r < s)] \\
&= L_p^{(1)} \max\{p^{1-\beta}, p^{1-s}\} I(r > s) + L_p^{(2)} \max\{p^{1-(\sqrt{s}-\sqrt{r})^2-\beta}, p^{1-s}\} I(r < s). \quad (3.2.7)
\end{aligned}$$

To know the order of the variance, note that

$$\text{Var}(T_n) = \text{Var}\{T_n(S_0)\} + \text{Var}\{T_n(S_1)\} + 2\text{Cov}\{T_n(S_0), T_n(S_1)\}$$

and

$$\begin{aligned}
\text{Var}(T_n) &= \sum_{j=1}^p \text{Var}\{Z_j(\lambda_n)\} + \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in S_0}} \text{Cov}\{Z_{j_1}(\lambda_n), Z_{j_2}(\lambda_n)\} + \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in S_1}} \text{Cov}\{Z_{j_1}(\lambda_n), Z_{j_2}(\lambda_n)\} \\
&\quad + \sum_{\substack{j_1 \neq j_2 \\ j_1 \in S_1, j_2 \in S_0}} \text{Cov}\{Z_{j_1}(\lambda_n), Z_{j_2}(\lambda_n)\} \\
&:= I_{(1)} + I_{(2)} + I_{(3)} + I_{(4)}. \quad (3.2.8)
\end{aligned}$$

It can be shown that

$$I_{(1)} = L_p^{(1)} \max\{p^{1-\beta}, p^{1-s}\} I(r > s) + L_p^{(2)} \max\{p^{1-(\sqrt{s}-\sqrt{r})^2-\beta}, p^{1-s}\} I(r < s),$$

which is the same as the variance of T_n if X have independent components. Since $\sum_k |\rho_k| < \infty$, only finite number of $|\rho_k| > 1 - \epsilon$ for some $\epsilon > 0$. From the derivation given in the Appendix, we know that,

$$\begin{aligned}
|I_{(2)}| &\leq \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in S_0}} |\gamma_{|j_1-j_2|}(0, 0, \lambda_n)| = L_p \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in S_0}} |\rho_{|j_1-j_2|}| p^{-\frac{2s}{1+|\rho_{|j_1-j_2|}|}} \\
&\leq L_p \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in S_0}} |\rho_{|j_1-j_2|}| p^{-2s/(2-\epsilon)} I\{|\rho_{|j_1-j_2|}| < 1 - \epsilon\} \\
&\quad + L_p \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in S_0}} |\rho_{|j_1-j_2|}| p^{-s} I\{|\rho_{|j_1-j_2|}| > 1 - \epsilon\} \leq L_p p^{1-2s/(2-\epsilon)} + L_p p^{-s} \\
&\sim L_p p^{1-2s/(2-\epsilon)}. \quad (3.2.9)
\end{aligned}$$

and for $s < r$, $I_{(3)} = 4r \log(p) \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in S_1}} \rho_{|j_1-j_2|} U(\eta_{n1}^-, \eta_{n1}^-; \rho_{|j_1-j_2|}) \{1 + o(1)\}$ which is a smaller order of $O(\log^2(p)p^{1-\beta})$, the order of $L_p^{(1)}p^{1-\beta}$ in $I_{(1)}$ and for $s > r$

$$\begin{aligned} |I_{(3)}| &\leq \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in S_1}} |\rho_{|j_1-j_2|}| |\rho_{|j_1-j_2|}^{-1}| \gamma_{|j_1-j_2|}(\sqrt{n}\mu_{j_1}, \sqrt{n}\mu_{j_2}, \lambda_n) |I\{\rho_{|j_1-j_2|} \neq 0\}| \\ &\leq L_p \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in S_1}} |\rho_{|j_1-j_2|}| p^{-\frac{2(\sqrt{s}-\sqrt{r})^2}{1+|\rho_{|j_1-j_2|}|}} \leq L_p p^{1-\beta-2(\sqrt{s}-\sqrt{r})^2/(2-\epsilon)}. \end{aligned} \quad (3.2.10)$$

Also,

$$|I_{(4)}| \leq \sum_{\substack{j_1 \neq j_2 \\ j_1 \in S_1, j_2 \in S_0}} |\gamma_{|j_1-j_2|}(\sqrt{n}\mu_{j_1}, 0, \lambda_n)| \leq L_p \sum_{\substack{j_1 \neq j_2 \\ j_1 \in S_1, j_2 \in S_0}} |\rho_{|j_1-j_2|}| p^{-s} \leq L_p p^{1-s-\beta}. \quad (3.2.11)$$

In summary, $I_{(2)}$, $I_{(3)}$ and $I_{(4)}$ are smaller order than $I_{(1)}$. By (3.2.8), we have

$$\text{Var}(T_n) = L_p^{(1)} \max\{p^{1-\beta}, p^{1-s}\} I(r > s) + L_p^{(2)} \max\{p^{1-(\sqrt{s}-\sqrt{r})^2-\beta}, p^{1-s}\} I(r < s). \quad (3.2.12)$$

The difference in the variance of T_n between independent and weakly dependent data exists in the terms $I_{(2)}$, $I_{(3)}$ and $I_{(4)}$ in (3.2.8). As we shown in (3.2.9), (3.2.10) and (3.2.11), these terms are all negligible comparing to $I_{(1)}$.

3.3 Asymptotic Distribution of T_n

In this section, we will discuss the asymptotic distribution of the threshold statistics T_n . The asymptotic distribution is derived for α -mixing (strong mixing) sequences $\{X_{ij}\}_{i=1}^p$. Let us firstly recall two concepts that will be used in the following.

Definition 1. Assume that a sequence of random variables $\{V_i\}_{i=1}^p$ is defined on the same probability space (Ω, \mathcal{F}, P) and $\mathcal{F}_a^b = \sigma\{V_i : i \in (a, b)\}$ is the σ -algebras generated by $\{V_i\}_{i=a}^b$. Then the sequence of $\{V_i\}_{i=1}^p$ (not necessary stationary) is said to be an α -mixing array if $\lim_{k \rightarrow \infty} \alpha_V(k) = 0$ where

$$\alpha_V(k) = \sup_{i \in \mathbb{Z}} \alpha(\mathcal{F}_{-\infty}^i, \mathcal{F}_{i+k}^\infty),$$

$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A)P(B)|$ and \mathcal{A}, \mathcal{B} are σ -algebras in Ω .

Definition 2. A stochastic process W_t is called **weak stationary** if $E\{W_t\} = E\{W_{t+\delta}\}$ for any δ and covariance $\text{Cov}\{W_t, W_{t+\delta}\}$ exists and only depends on δ .

Assume the sequence $\{X_{ij}\}_{i=1}^p$ satisfy the α -mixing condition for each replicate j . Because $\{X_{ij}\}_{j=1}^n$ is a sequence of IID random variables and $Z_{i,n}$ is a Borel function of $\{X_{i1}, \dots, X_{in}\}$, the α -mixing coefficient $\alpha_Z(k)$ for $Z_{i,n}$ satisfy (Refer to Theorem 5.2 in Bradley, 2005)

$$\alpha_Z(k) \leq \sum_{j=1}^n \alpha_{X_j}(k) = n\alpha_X(k), \text{ for each } k = 1, \dots, p,$$

where $\alpha_X(k) = \alpha_{X_j}(k)$ for $j = 1, \dots, n$ (since X_j are IID random vectors). Suppose k depends on n and denote it by $k(n)$. If $\alpha_X\{k(n)\} \rightarrow 0$ as $k(n) \rightarrow \infty$ and $\alpha_X\{k(n)\} = o(n^{-1})$, then $\alpha_Z\{k(n)\} \rightarrow 0$ as $k(n) \rightarrow \infty$, which means that $\{Z_{i,n}\}_{i=1}^p$ is also an α -mixing sequence. If $\sum_{k(n)=1}^{\infty} \alpha_X\{k(n)\} < \infty$, then $\alpha_X\{k(n)\} = o\{k(n)^{-1}\}$, which is a smaller order of n^{-1} if $k(n)$ is a higher order than n , in this case, we can take $k(n)$ large enough than n such that $\alpha_Z(k) \rightarrow 0$.

Recall that $\gamma_k(\mu_1, \mu_{k+1}, \lambda_n) = \text{Cov}\{Z_{1,n}(\lambda_n), Z_{k+1,n}(\lambda_n)\}$ and define

$$\sigma_0^2(p; \lambda_n) = \text{Var}\{Z_{1,n}(\lambda_n)\} + 2 \sum_{k=1}^{p-1} \gamma_k(0, 0, \lambda_n).$$

The following theorem provides the asymptotic distributions of the test statistic T_n under the null hypothesis and alternatives. The proof is provided in the Appendix.

Theorem 2 *Suppose $\{X_{ij}\}_{i=1}^p$ is an α -mixing sequence for each sample j and $\{W_{ij}\}_{i=1}^p$ in model (3.1.1) is a weak stationary sequence. Let $\mu_{T_n,0}^{(i)}$ and $\mu_{T_n,1}^{(i)}$ be the mean of $Z_{i,n}(\lambda_n)$ under the null and alternative, respectively. Under the conditions in Theorem 1 and assume $\sum_{k=1}^{\infty} k\alpha_X^{\delta/(4+\delta)}(k) < \infty$ for some $\delta > 0$ and $p^{6a/5}n^{4/5}\alpha_X^{4/5}(p^c) \rightarrow 0$ for some a and $c \in (0, 1-a)$.*

(i) *If there exists $a \in ((4+2\delta)s/(4+\delta), 1)$ such that $n^{\delta/(4+\delta)}p^{(4+2\delta)s/(4+\delta)-a} \rightarrow 0$, then under the null hypothesis in (3.1.2),*

$$p^{-1/2}S_n \xrightarrow{d} N(0, 1),$$

where $S_n = \sigma_0^{-1}(p; \lambda_n) \sum_{i=1}^p \{Z_{i,n}(\lambda_n) - \mu_{T_n,0}^{(i)}\}$.

(ii) *Let $\mu_i = \sqrt{2r_i \log(p)/n}$. Suppose that there exist $\sigma_1^2(p; \lambda_n)$ such that*

$$H_l = \sup_{j \geq 1} \left| l^{-1} \text{Var} \left\{ \sum_{i=j}^{j+l-1} Z_{i,n}(\lambda_n) \right\} - \sigma_1^2(l; \lambda_n) \right| \rightarrow 0 \text{ as } l \rightarrow \infty, \quad (3.3.13)$$

and $\sigma_1^2(p; \lambda_n) = L_p p^{-h_1}$ ($0 \leq h_1 \leq s$). Assume (a) $\max_i r_i > s$ and exists $a \in (2h_1, 1)$ such that $n^{\delta/(4+\delta)}p^{2h_1-a} \rightarrow 0$; (b) $\max_i r_i < s$ and exists $a \in (2h_1 - \frac{4s^*}{4+\delta}, 1)$ such that

$n^{\delta/(4+\delta)} p^{2h_1 - \frac{4s^*}{4+\delta} - a} \rightarrow 0$ where $s^* = (\sqrt{s} - \max_i \sqrt{r_i})^2$. Then either under (a) or (b),

$$p^{-1/2} \sigma_1^{-1}(p; \lambda_n) \sum_{i=1}^p \{Z_{i,n}(\lambda_n) - \mu_{T_n,1}^{(i)}\} \xrightarrow{d} N(0, 1).$$

Remark 1. Under the null hypothesis, $\{Z_{i,n}(\lambda_n)\}_{i=1}^p$ is a weak stationary sequence. The condition $\sum_{k=1}^{\infty} k \alpha_X^{\delta/(4+\delta)}(k) = O(1)$ assures that

$$\lim_{p \rightarrow \infty} \text{Var} \left\{ p^{-1/2} \sum_{i=1}^p Z_{i,n}(\lambda_n) \right\} = \sigma_0^2(p; \lambda_n) < \infty.$$

Remark 2. Assume $p = \kappa n$ for some constant $\kappa \in (0, \infty)$ and $\alpha_X(p) = p^{-4(4+\delta)/\delta}$. Let $\delta = 1$ and $a = 6s/5 + \Delta$ with some $\Delta > 1/5$ and $s < 2/3$. Then $n^{\delta/(4+\delta)} p^{(4+2\delta)s/(4+\delta)-a} = \kappa^{-\Delta} n^{1/5-\Delta} \rightarrow 0$ and $p^{6a/5} n^{4/5} \alpha_X^{4/5}(p^c) = \kappa^{(36s+30\Delta)/25-16c} n^{(20+36s+30\Delta)/25-16c} \rightarrow 0$ for $c \in ((10 + 18s + 15\Delta)/200, 1 - 6s/5 - \Delta)$ and $s < 0.57$. Thus, in this case, a sufficient condition such that the asymptotic normality holds under the null hypothesis is $s < 0.57$. The conditions $p^{6a/5} n^{4/5} \alpha_X^{4/5}(p^c) \rightarrow 0$ and $n^{\delta/(4+\delta)} p^{(4+2\delta)s/(4+\delta)-a} \rightarrow 0$ are more easily to be satisfied if p increases at a faster rate than n . These conditions can be relaxed to $p^{6a/5} \alpha_X^{4/5}(p^c) \rightarrow 0$ and $p^{(4+2\delta)s/(4+\delta)-a} \rightarrow 0$ if $\alpha_Z(k) \leq C \alpha_X(k)$ for some constant C .

Under the conditions in Theorem 1, $\mu_{T_n,0}^{(i)} = G_{T_n} \{1 + o(1)\}$ where

$$G_{T_n} = 2\sqrt{\lambda_n} \phi(\sqrt{\lambda_n}) + 2\bar{\Phi}(\sqrt{\lambda_n})$$

and $\sigma_0^2(p; \lambda_n) = \tilde{\sigma}_0^2(p; \lambda_n) \{1 + o(1)\}$ where

$$\tilde{\sigma}_0^2(p; \lambda_n) = 2[(\sqrt{\lambda_n})^3 + 3\sqrt{\lambda_n}] \phi(\sqrt{\lambda_n}) + 6\bar{\Phi}(\sqrt{\lambda_n}) + 2 \sum_{k=1}^{p-1} \gamma_k(0, 0, \lambda_n) - G_{T_n}^2 \quad (3.3.14)$$

where

$$\begin{aligned} \gamma_k(0, 0, \lambda_n) &= 4(\lambda_n^{3/2} \rho_k^2 + (1 + 2\rho_k^2) \lambda_n^{1/2}) \phi(\lambda_n^{1/2}) [\bar{\Phi}(\theta \lambda_n^{1/2}) + \bar{\Phi}(\theta^{-1} \lambda_n^{1/2})] \\ &\quad + 2\sqrt{1 - \rho_k^2} ((1 + 2\rho_k) \lambda_n + 3\rho_k) \phi(\theta \lambda_n^{1/2}) \phi(\lambda_n^{1/2}) \\ &\quad + 2\sqrt{1 - \rho_k^2} ((1 - 2\rho_k) \lambda_n - 3\rho_k) \phi(\theta^{-1} \lambda_n^{1/2}) \phi(\lambda_n^{1/2}) \\ &\quad + 2(1 + 2\rho_k^2) [U(\lambda_n^{1/2}, \lambda_n^{1/2}; \rho_k) + U(\lambda_n^{1/2}, \lambda_n^{1/2}; -\rho_k)] - G_{T_n}^2 \end{aligned} \quad (3.3.15)$$

and $\theta = \sqrt{(1 - \rho_k)/(1 + \rho_k)}$.

From (3.2.9), we know $G_{T_n} = L_p p^{-s}$ and $\tilde{\sigma}_0(p; \lambda_n) = L_p p^{-s/2}$. So the approximation error is $o\{G_{T_n}/\tilde{\sigma}_0(p; \lambda_n)\} = o(L_p p^{-s/2}) \rightarrow 0$. Therefore, by Theorem 2, we can construct an asymptotic α level test which rejects the null hypothesis if

$$\{p^{-1/2}(T_n - pG_{T_n}) > z_\alpha \tilde{\sigma}_0(p; \lambda_n)\}.$$

An estimate of $\tilde{\sigma}_0^2(p; \lambda_n)$ is necessary for implement the above test. To estimate $\gamma_k(0, 0, \lambda_n)$, we need an estimate of ρ_k . An estimate of the ρ_k for any fixed $k < p$ is

$$\hat{\rho}_k = \frac{1}{(n-1)(p-k)} \sum_{i=1}^{p-k} \sum_{j=1}^n (X_{ij} - \bar{X}_i)(X_{(i+k)j} - \bar{X}_{(i+k)}) \quad (3.3.16)$$

where \bar{X}_i is the sample mean of the i -th components of X . For each fixed k , $\hat{\rho}_k$ is a \sqrt{np} -consistent estimator. Hence, a consistent estimate of $\tilde{\sigma}_0^2(p; \lambda_n)$ can be obtained though formula (3.3.14). Another method for estimating $\tilde{\sigma}_0^2(p; \lambda_n)$ is to use the smoothed spectral density $g(w)$ of $\{Z_{i,n}^0(\lambda_n) := n\bar{W}_i^2 I(n\bar{W}_i^2 > \lambda_n)\}_{i=1}^p$, which is given in the Appendix.

3.4 Extension to the Maximum Test

We now extend the threshold test to a maximum test, which avoids the choice of λ_n . The basic idea is to maximize the threshold statistic in a reasonable range of threshold values. To emphasize the dependence on threshold, we write $\mu_{T_n,0}$ and $\sigma_{T_n,0}$ as $\mu_{T_n,0}(\lambda_n)$ and $\sigma_{T_n,0}(\lambda_n)$. Define the standardized version of $T_n(\lambda_n)$ as

$$\mathcal{T}_S(\lambda_n) = \frac{T_n(\lambda_n) - \mu_{T_n,0}(\lambda_n)}{\sigma_{T_n,0}(\lambda_n)}.$$

Under the null hypothesis, $\mathcal{T}_S(\lambda_n)$ is a zero mean stochastic process with respect to λ_n . For any different thresholds λ_n and η_n , the correlation between $\mathcal{T}_S(\lambda_n)$ and $\mathcal{T}_S(\eta_n)$ is

$$\Omega(\lambda_n, \eta_n) := \text{Corr}\{\mathcal{T}_S(\lambda_n), \mathcal{T}_S(\eta_n)\} = \frac{\text{Cov}\{T_n(\lambda_n), T_n(\eta_n)\}}{\sigma_{T_n,0}(\lambda_n)\sigma_{T_n,0}(\eta_n)} \quad (3.4.17)$$

where $\sigma_{T_n,0}^2(\lambda_n) = p\sigma_0^2(p; \lambda_n)$. It can be shown that, under the null hypothesis,

$$\begin{aligned} p^{-1}\text{Cov}\{T_n(\lambda_n), T_n(\eta_n)\} &= \left\{ 2[(\sqrt{\lambda_n^*})^3 + 3\sqrt{\lambda_n^*}]\phi(\sqrt{\lambda_n^*}) + 6\bar{\Phi}(\sqrt{\lambda_n^*}) \right. \\ &\quad \left. + 2 \sum_{k=1}^{p-1} \tilde{\gamma}_k(\sqrt{\lambda_n}, \sqrt{\eta_n}) - pG_{T_n}(\lambda_n)G_{T_n}(\eta_n) \right\} \{1 + o(1)\}, \end{aligned}$$

where $\lambda_n^* = \max(\lambda_n, \eta_n)$, $G_{T_n}(\lambda_n) = 2\sqrt{\lambda_n}\phi(\sqrt{\lambda_n}) + 2\bar{\Phi}(\sqrt{\lambda_n})$ and

$$\begin{aligned} \tilde{\gamma}_k(\lambda_n, \eta_n) &= 2(1 + 2\rho_k^2)[U(\lambda_n, \eta_n; \rho_k) + U(\lambda_n, \eta_n; -\rho_k)] \\ &\quad + 2(\lambda_n^3\rho_k^2 + (1 + 2\rho_k^2)\lambda_n)[Q(\lambda_n, \eta_n; \rho_k) + Q(\lambda_n, \eta_n; -\rho_k)] \\ &\quad + 2(\eta_n^3\rho_k^2 + (1 + 2\rho_k^2)\eta_n)[Q(\eta_n, \lambda_n; \rho_k) + Q(\eta_n, \lambda_n; -\rho_k)] \\ &\quad + 2\sqrt{1 - \rho_k^2}(\rho_k\eta_n^2 + \lambda_n\eta_n + \rho_k\lambda_n^2 + 3\rho_k)q(\eta_n, \lambda_n; \rho_k) \\ &\quad - 2\sqrt{1 - \rho_k^2}(\rho_k\eta_n^2 - \lambda_n\eta_n + \rho_k\lambda_n^2 + 3\rho_k)q(\eta_n, \lambda_n; -\rho_k). \end{aligned} \tag{3.4.18}$$

Theorem 3 *Let $\underline{\lambda}_n = (\lambda_{n1}, \dots, \lambda_{nd})' = 2\log(p)(s_1, \dots, s_d)' \in \Lambda$ and for any $\lambda_{nk} \in \Lambda$, $\mathcal{T}_S(\lambda_{nk})$ is asymptotically normally distributed, namely the conditions given in Theorem 2(i) hold for λ_{nk} . Then $\mathcal{T}_S(\underline{\lambda}_n) := (\mathcal{T}_S(\lambda_{n1}), \dots, \mathcal{T}_S(\lambda_{nd}))'$ is asymptotically multivariate normally distributed with mean 0 and covariance $(\Omega(\lambda_{ns}, \lambda_{nt}))_{s,t}$ under the null hypothesis.*

It might be further shown that $\mathcal{T}_S(t)$ converges uniformly to a zero mean Gaussian process $\mathcal{N}(t)$ with covariance $\text{Cov}\{\mathcal{N}(s), \mathcal{N}(t)\} = \Omega(s, t)$ under the null hypothesis. We define the maximum test as

$$\mathcal{M}_n = \max_{\lambda_n \in \Lambda} \mathcal{T}_S(\lambda_n),$$

which converges to $\mathcal{B} := \max_{t \in \Lambda} \mathcal{N}(t)$ under the null hypothesis. Therefore, an asymptotic α level rejection region is $\{\mathcal{M}_n > \mathcal{B}_\alpha\}$ where \mathcal{B}_α is the α quantile of \mathcal{B} . The distribution of \mathcal{B} can be obtained by simulation. For a set of $\underline{\lambda}_n = (\lambda_{n1}, \dots, \lambda_{nd})' \in \Lambda$, simulating a large number $N_{\mathcal{B}}$ of normally distributed d -variate random vectors $\xi_i = (\xi_{i1}, \dots, \xi_{id})'$ ($i = 1, \dots, N_{\mathcal{B}}$) with mean 0 and covariance $(\Omega(\lambda_{ns}, \lambda_{nt}))_{s,t}$. For each ξ_i , we can get $\mathcal{B}_i = \max_{j \in \{1, \dots, d\}} \xi_{ij}$. Then \mathcal{B}_α can be estimated by the α quantile of $\{\mathcal{B}_i\}_{i=1}^{N_{\mathcal{B}}}$.

3.5 Optimal Detection Boundary and the Best Power

In this section, we would like to compare the proposed threshold test with an oracle test and give the best power that could be attained by the threshold test in the detectable region. We will also show that the maximum test is able to attain the optimal detection boundary (Donoho

and Jin, 2004). For discussion simplicity, we shall assume that all the non-zero signals are equal to $\mu_i := r_n = \sqrt{2r \log p/n} = o(1)$.

The oracle test assumes that there were an oracle who knows the non-zero components' locations in advance. An oracle test can be constructed through the non-zero components without involving with a large amount of noises, which should perform the best. The oracle test statistic can be defined as $T_{oc,n} = \sum_{i \in S_1} n \bar{X}_i^2$. As shown in Theorem 1, we can approximate the mean and variance by assuming that $\sqrt{n} \bar{X}_i$ is normally distributed. And similar to Theorem 2, we could also show the asymptotic normality of $T_{oc,n}$. Under the null hypothesis,

$$E(T_{oc,n}) = p_1 \quad \text{and} \quad \text{Var}(T_{oc,n}) = 2p_1 + 2 \sum_{\substack{i \neq j \\ i, j \in S_1}} \rho_{|i-j|}^2 = 2p_1 \{1 + o(1)\}$$

where $p_1 = p^{1-\beta}$ is the number of non-zero components in S_1 . Under the alternative, $E(T_{oc,n}) = (nr_n^2 + 1)p_1$ and

$$\text{Var}(T_{oc,n}) = 2(2nr_n^2 + 1)p_1 + \sum_{\substack{i \neq j \\ i, j \in S_1}} (2\rho_{|i-j|}^2 + 4\rho_{|i-j|}nr_n^2) = 2(2nr_n^2 + 1)p_1 \{1 + o(1)\}.$$

An α level oracle test will reject the null hypothesis if $\{T_{oc,n} > \sqrt{2p_1}z_\alpha + p_1\}$. It can be shown that the asymptotic power of the oracle test is $\Phi\left(-z_\alpha + \sqrt{\frac{p_1}{2}nr_n^2}\right)$.

Now comparing the oracle test with the proposed threshold test. Our purpose here is to see what is the best power the threshold test can achieve under cases regarding r and β . Recall that $\mu_{T_n,0}$ and $\sigma_{T_n,0}^2$ are the mean and variance of T_n under the null and $\mu_{T_n,1}$ and $\sigma_{T_n,1}^2$ are the corresponding mean and variance under the alternative. We will reject the null hypothesis if $T_n > \mu_{T_n,0} + \sigma_{T_n,0}z_\alpha$ at α -level, where z_α is the upper α quantile of the standard normal distribution. So the power of the threshold test will be

$$\beta(\mu) = \Phi\left(-\frac{\sigma_{T_n,0}}{\sigma_{T_n,1}}z_\alpha + \frac{\mu_{T_n,1} - \mu_{T_n,0}}{\sigma_{T_n,1}}\right). \quad (3.5.19)$$

From (3.2.7) and (3.2.12), we have

$$\mu_{T_n,1} - \mu_{T_n,0} = L_p^{(1)}p^{1-\beta}I\{r > s\} + L_p^{(2)}p^{1-(\sqrt{s}-\sqrt{r})^2-\beta}I\{r \leq s\} \quad (3.5.20)$$

and

$$\begin{aligned} \sigma_{T_{n,1}} &= L_p^{(1)} \max\{p^{(1-\beta)/2}, p^{(1-s)/2}\} I\{r > s\} \\ &\quad + L_p^{(2)} \max\{p^{(1-(\sqrt{s}-\sqrt{r})^2-\beta)/2}, p^{(1-s)/2}\} I\{r \leq s\}. \end{aligned} \quad (3.5.21)$$

We should discuss the best power in the detectable region on (β, r) plane. An alternative H_1 is said to be asymptotically detectable if H_0 and H_1 can be separated asymptotically and otherwise it is not detectable. Let $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$ be the significant level and $\{T_n \geq \mu_{T_{n,0}} + \sigma_{T_{n,0}} z_{\alpha_n}\}$ be the rejection region of the threshold test with type I error α_n . Then the sum of type I error and type II error is

$$Err_{\alpha_n} := \alpha_n + \bar{\Phi} \left(-\frac{\sigma_{T_{n,0}}}{\sigma_{T_{n,1}}} z_{\alpha_n} + \frac{\mu_{T_{n,1}} - \mu_{T_{n,0}}}{\sigma_{T_{n,1}}} \right).$$

An alternative is asymptotically detectable if $Err_{\alpha_n} \rightarrow 0$ as $n \rightarrow \infty$ and undetectable if $Err_{\alpha_n} \rightarrow 1$. Because $\sigma_{T_{n,1}} \geq \sigma_{T_{n,0}}$ asymptotically, a necessary and sufficient condition for detectable is that

$$\frac{\mu_{T_{n,1}} - \mu_{T_{n,0}}}{\sigma_{T_{n,1}}} \rightarrow \infty.$$

We verified this condition for any point in the detectable region of (β, r) plane, i.e., there exists a $\lambda_n = 2s \log(p)$ such that the above condition holds for any point in the detectable region. The detail of the verification is provided in the Appendix. The detectable region is summarized in Figure 3.1, which is the union of regions marked by I-IV. The boundary of this region is given by

$$\varrho(\beta) = \begin{cases} \beta - 1/2, & 1/2 \leq \beta \leq 3/4; \\ (1 - \sqrt{1 - \beta})^2, & 3/4 < \beta < 1. \end{cases}$$

This boundary is optimal, in the sense if $r < \varrho(\beta)$, all the tests are powerless (Donoho and Jin, 2004).

It turns out the best rates can be classified by three different areas in above detectable region. We summarize the best power that the threshold test may achieve in Figure 3.1, whose derivation is given in Appendix. In the regions II and IV with blue color (encircled by $r = \beta$, $r = \beta/3$ and $r = (1 - \sqrt{1 - \beta})^2$), the best power is of order $L_p p^{\frac{1}{2}} - \frac{1}{2}(r + \beta)^2/(4r)$. In the

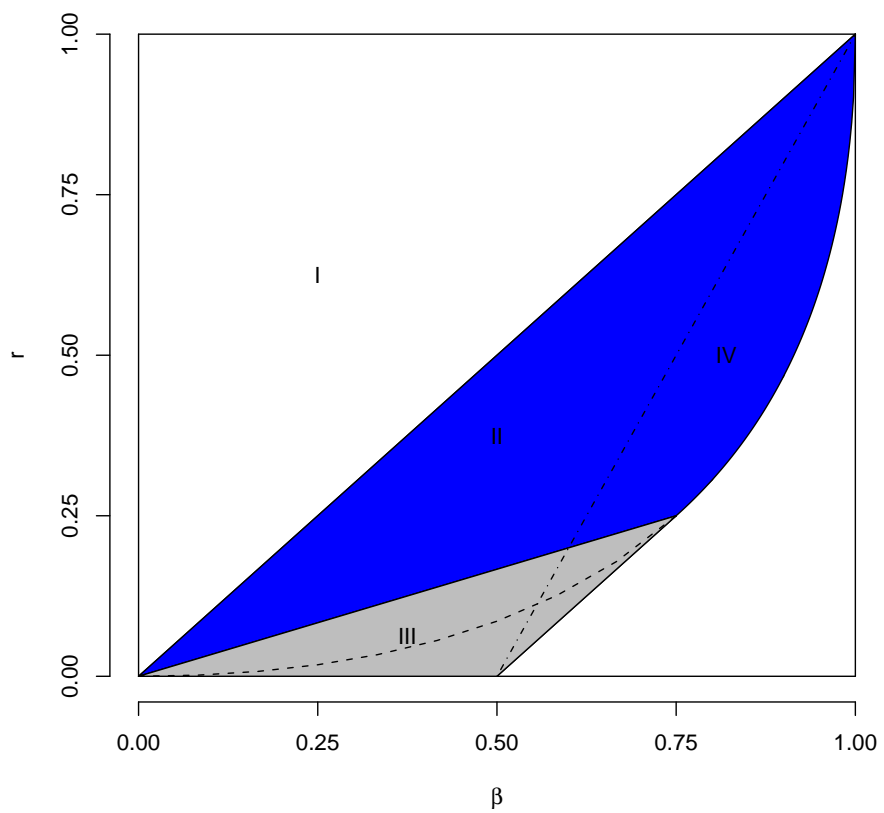


Figure 3.1 The detectable region of the threshold and the maximum test in (β, r) plane, which is the union of I-IV areas in the plot.

region III with gray color (encircled by $r = \beta/3$, $r = \beta - \frac{1}{2}$ and $r = 0$), the best power is of order $L_p p^{\frac{1}{2} - \beta + r}$. In the region I where $r > \beta$, the best power is of order $L_p p^{(1-\beta)/2}$, which has the power the same order as the oracle test up to a slow varying function L_p . This means that in the region I, the power of the threshold test and the oracle test are only different by a slow varying function L_p .

In practice, we have to choose the threshold λ_n to implement the threshold test. Therefore, the optimal detection boundary may not be attained for the threshold test. However, we would like to show that the optimal detection boundary can be achieved by the maximum test given in Section 3.4.

To this end, we only need to show that the sum of type I and II errors of the maximum test goes to 0 in the detectable region of (β, r) defined in Figure 3.1 as the significant level α_n goes to 0. Because of the maximum test is an asymptotic α_n level test, the type I error less than $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$. Thus, it suffices to show that the power of the maximum test goes to 1 in the detectable region as n goes to infinity.

To appreciate this point, recall that the α_n level rejection region for the maximum test is $R_{\alpha_n} = \{\mathcal{M}_n > \mathcal{B}_{\alpha_n}\}$. The tightness of $\mathcal{T}(\lambda_n)$ for $\lambda_n \in \Lambda$ ensures the existence of $M < \infty$ such that $\mathcal{B}_{\alpha_n} < M$ with probability one. Then it is enough to show

$$P(\mathcal{M}_n \rightarrow \infty) \rightarrow 1, \quad \text{as } n \rightarrow \infty \quad (3.5.22)$$

at every point (β, r) in the detectable region. Notice that $\mathcal{M}_n \geq \mathcal{T}(\lambda_n)$ everywhere for any fixed $\lambda_n \in \Lambda$. Therefore, (3.5.22) is true if for any point in the detectable region, there exists a λ_n such that $P(\mathcal{T}(\lambda_n) \rightarrow \infty) \rightarrow 1$. For any fixed $M < \infty$, we want to show

$$P(\mathcal{T}(\lambda_n) > M) = \Phi\left(-\frac{\sigma_{Tn,0}}{\sigma_{Tn,1}}M + \frac{\mu_{Tn,1} - \mu_{Tn,0}}{\sigma_{Tn,1}}\right) \rightarrow 1. \quad (3.5.23)$$

Because of $\sigma_{Tn,0} \leq \sigma_{Tn,1}$ and finiteness of M , (3.5.23) is true as long as $\frac{\mu_{Tn,1} - \mu_{Tn,0}}{\sigma_{Tn,1}} \rightarrow \infty$. As we have shown in the Appendix, at every point in the detectable region, there exist a λ_n such that $\frac{\mu_{Tn,1} - \mu_{Tn,0}}{\sigma_{Tn,1}} \rightarrow \infty$. This concludes that $\mathcal{T}(\lambda_n) \rightarrow \infty$ with probability one for some λ_n and hence the \mathcal{M}_n at every point in the detectable region given in Figure 3.1, which means that the maximum test can attain the optimal detection boundary.

3.6 Simulation Results

The simulation was designed to understand the performance of the threshold test, the maximum test and compare them with false discovery rate (Benjamini and Hochberg, 1995), oracle test and the test (C-Q test) proposed by Chen and Qin (2010).

Independent and identically distributed random vectors X_j were generated by model (3.1.1) i.e.,

$$X_j = W_j + \mu_j, \quad j = 1, \dots, n$$

where $W_j = (W_{1j}, \dots, W_{pj})'$ is a random vector and the sequence $\{W_{ij}\}_{i=1}^p$ was a realization of a stochastic process. Two processes were considered in the simulation, one is the Gaussian process, which is simulated according to the method proposed by Wood and Chan (1994). The other process was generated exactly the same as the Gaussian process except the marginal distribution used was the standardized Gamma(2,2). We set an autoregressive correlation structure. That is $\text{Cov}(W_{i1}, W_{j1}) = \rho^{|i-j|}$ and ρ was set to be 0.6 in the simulation. The sparsity was controlled by setting $\beta = 0.6, 0.7$ and 0.8 . The non-zero signals $\mu_i = \sqrt{2r \log(p)/n}$ were set at the same levels and r were fixed at 0.4, 0.6 and 0.9. The dimension p of X_j was chosen to be 1000, 2000 and 2500, and the sample size was chosen to be 20, 30 and 40 respectively in the simulation. All the simulation results were based on 1000 replicates.

The null distributions of the standardized threshold test statistics T_n were plotted in Figure 3.2, 3.3 and 3.4 for $\{X_{ij}\}_{i=1}^p$ generated from Gaussian process. We also obtained null distribution plots for processes with marginal centralized Gamma(2,2) and Gamma(1,2). Because there were largely the same as the plots for Gaussian process, we only present the null distributions of T_n from Gaussian process. In each plot, T_n was standardized by theoretical variance (3.3.14) with $\sum_{k=1}^{p-1} \gamma_k(0, 0, \lambda_n)$ replaced by $\sum_{k=1}^{p-1} (1 - \frac{k}{p}) \gamma_k(0, 0, \lambda_n)$ using the true value of ρ and known covariance structure (“Theoretical” in the plot legend), estimated variance (“Plug-in” in the plot legend) using the same variance expression as “Theoretical” but with plug-in estimates of ρ_k provided in (3.3.16) and the spectral density based kernel smooth estimate of the variance (“Spectral” in the plot legend) of T_n given in the Appendix. We plotted histograms of the null distributions of the standardized T_n with thresholds at $\lambda_n = 2s \log(p)$ for

$s = 0.50, 0.55, 0.60, 0.65$. It can be seen that the null distributions which are standardized by plug-in variance are closer to the standard normal than that by spectral smoothed variance. Thus we only report the simulation results utilizing plug-in variance in the following.

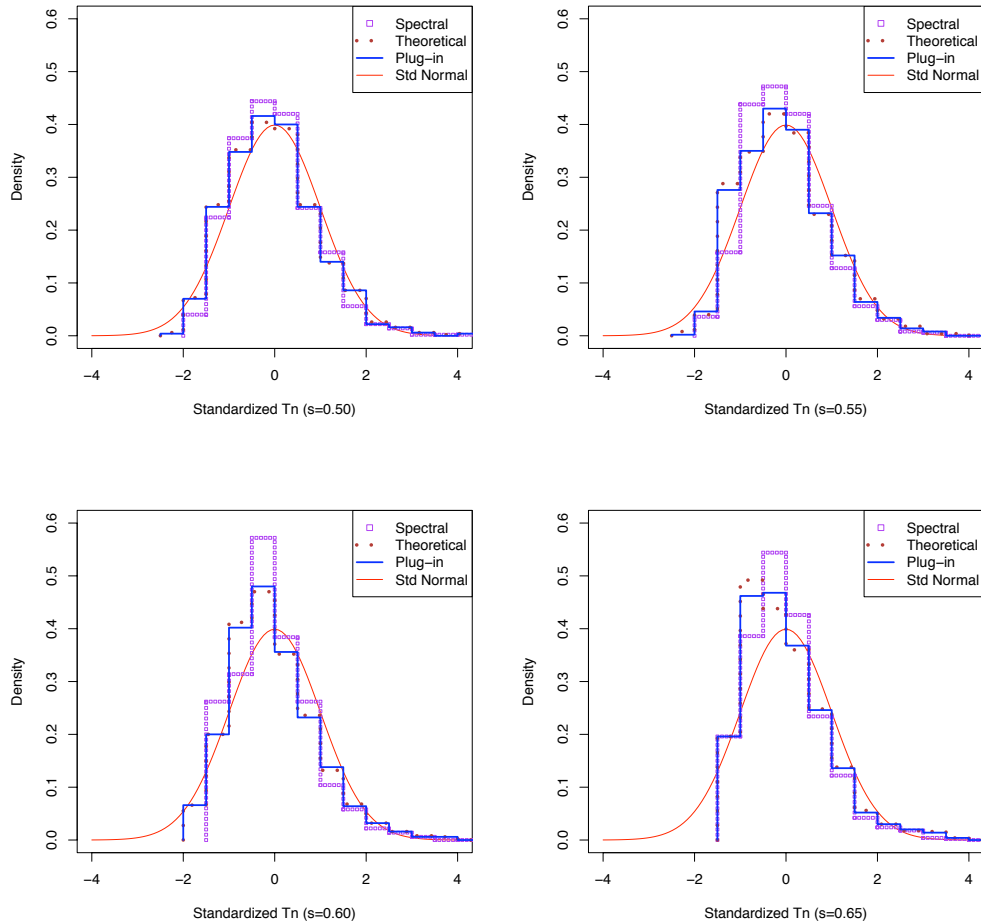


Figure 3.2 The histograms for the simulated null distributions of standardized T_n using plug-in, theoretical variance (3.3.14) estimate and spectral smoothed variance estimate introduced in the Appendix. The (p, n) is $(1000, 20)$. Marginal distribution: Gaussian.

Table 3.1 and 3.2 summarize the sizes of the oracle test, FDR, maximum test and threshold test for different threshold levels for Gaussian process and process with marginal standardized Gamma distributions. The maximum test statistic maximizes the standardized threshold test statistic at threshold levels with $\underline{\lambda}_n = \{2s \log(p) : s = 0.50, 0.55, 0.60, \dots, 0.90\}$. The cutoff

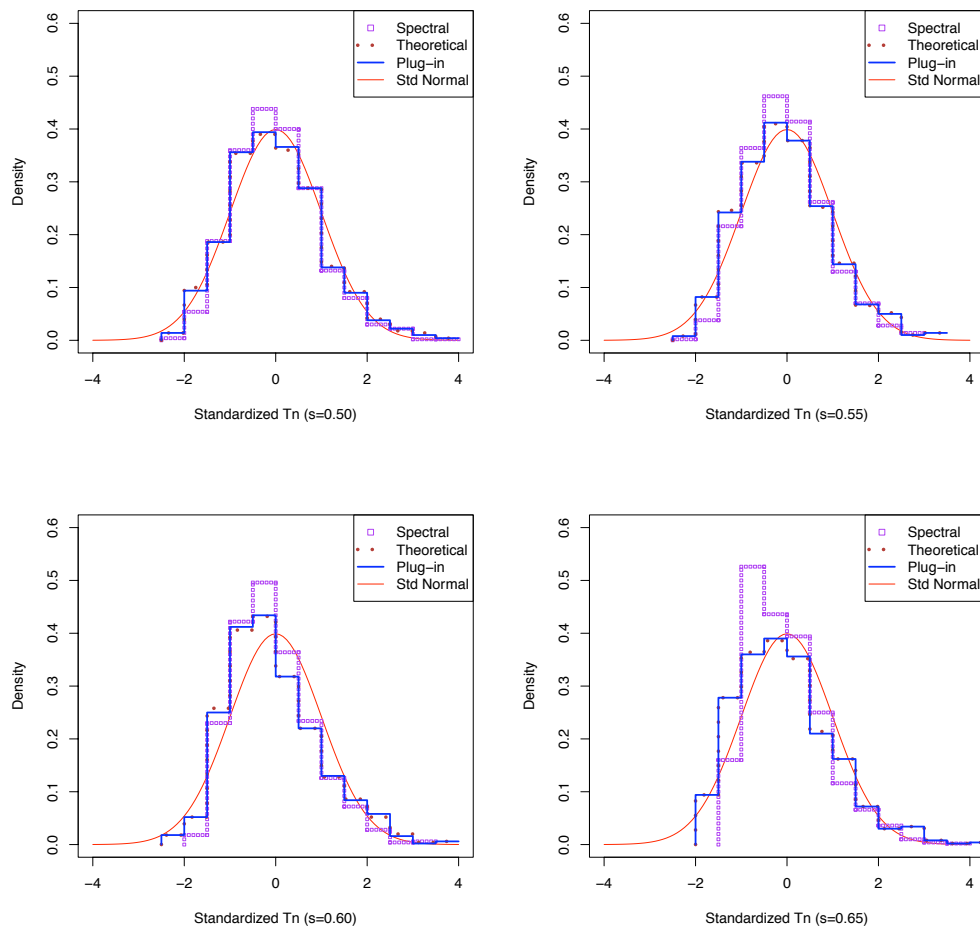


Figure 3.3 The histograms for the simulated null distributions of standardized T_n using plug-in, theoretical variance (3.3.14) estimate and spectral smoothed variance estimate introduced in Appendix. The (p, n) is $(2000, 30)$. Marginal distribution: Gaussian.

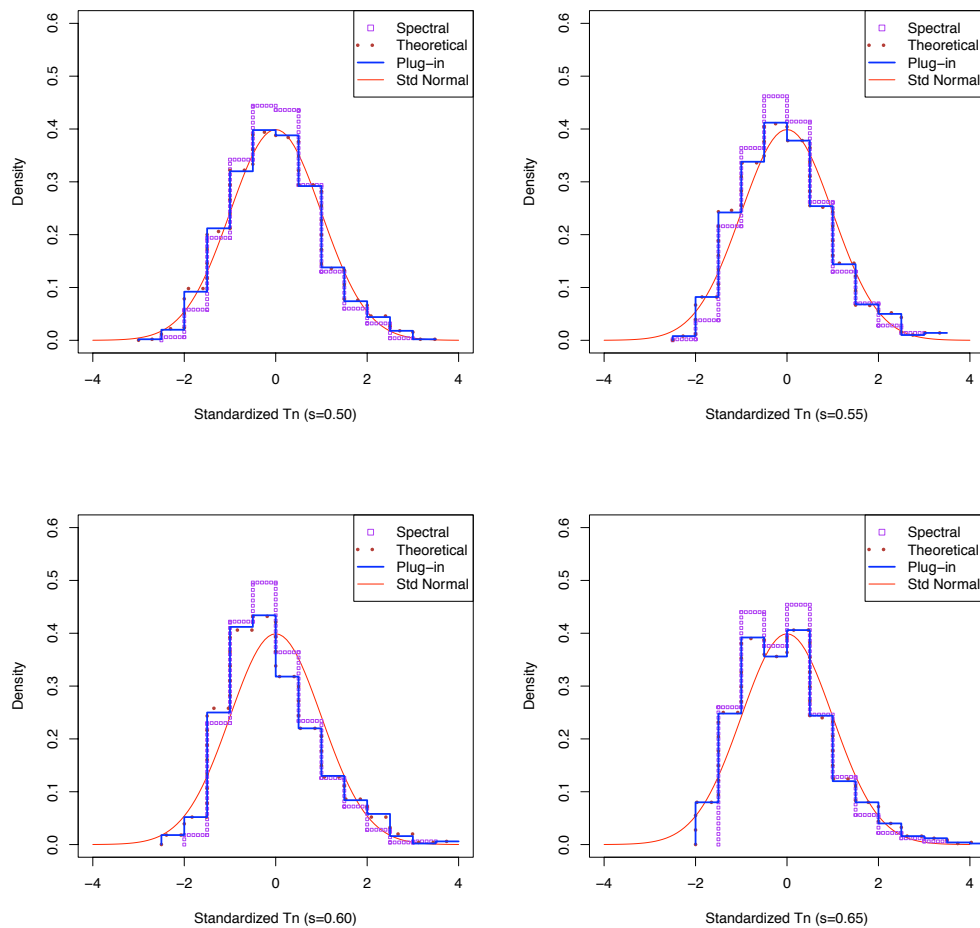


Figure 3.4 The histograms for the simulated null distributions of standardized T_n using plug-in, theoretical variance (3.3.14) estimate and spectral smoothed variance estimate introduced in Appendix. The (p, n) is $(2500, 40)$. Marginal distribution: Gaussian.

points were based on the simulated distributions of \mathcal{B} provided in Section 3.4. It is observed that the threshold tests preserved sizes reasonably at all levels in Table 3.1 and 3.2. There are some slightly larger sizes at level 0.025 but it is getting better when the levels increase. From Table 3.2, we see that the non-normal marginal distributions had little impact on the sizes of the tests. The maximum test was conservative at high significance level for $\alpha > 0.10$.

Figure 3.5, 3.6 and 3.7 present the receiver operating characteristic (ROC) curves with type I error between 0 and 0.2 for oracle tests, FDR, C-Q tests, maximum tests and threshold tests at several levels. We observed that (i) in all the cases, the threshold tests were more powerful than the C-Q tests and FDR. (ii) the powers of the threshold tests were not responsive to the threshold levels for sparsity parameter $\beta = 0.6$ and 0.7 . The threshold tests can have significant improvement in power for proper choices of threshold levels when $\beta = 0.8$. This is a quite sparse case, notice that $[1000^{0.2}] = 3$, $[2000^{0.2}] = 4$ and $[2500^{0.2}] = 4$, so there are only a few locations with signals. Thus, high threshold levels are preferable in the threshold test. (iii) the C-Q tests is suitable for data with non-sparsity signals while the FDR is good to use when the signal is very sparse. (iv) when the signal level r is high, the threshold tests could achieve the power of oracle tests. But it is hard to achieve in most cases. (v) The maximum test almost attained the best power among all the tests (except the oracle test) even though it has relative smaller sizes than the other tests on high significant levels.

3.7 Appendix: Technical Details

In this Appendix, we give the expression of $\tilde{E}(Z_{1,n}(\lambda_n)Z_{2,n}(\lambda_n))$ in Theorem 1, present the derivation of $I_{(2)} - I_{(4)}$ in Section 3.2, verify the detection boundary of the threshold test and outline the proofs of the Theorems presented in Sections 3.2, 3.3 and 3.4.

Table 3.1 Empirical sizes of the Oracle test, C-Q test, FDR, maximum test and threshold tests with different threshold levels $\lambda_n = 2s \log(p)$ for Gaussian process.

α	Oracle	C-Q	FDR	MAX	s					
					0.50	0.55	0.60	0.65	0.70	0.75
(p,n)=(1000,20)										
0.025	0.050	0.033	0.016	0.048	0.031	0.039	0.041	0.042	0.049	0.046
0.050	0.066	0.062	0.044	0.064	0.056	0.059	0.056	0.069	0.072	0.058
0.075	0.078	0.080	0.068	0.074	0.089	0.069	0.078	0.087	0.086	0.078
0.100	0.087	0.109	0.096	0.084	0.100	0.098	0.098	0.105	0.113	0.110
0.125	0.100	0.129	0.129	0.093	0.127	0.122	0.116	0.127	0.137	0.144
0.150	0.112	0.138	0.155	0.105	0.145	0.139	0.131	0.144	0.146	0.158
0.175	0.120	0.175	0.183	0.113	0.162	0.156	0.157	0.164	0.158	0.164
0.200	0.132	0.199	0.211	0.130	0.188	0.182	0.178	0.185	0.169	0.167
(p,n)=(2000,30)										
0.025	0.035	0.023	0.036	0.040	0.034	0.039	0.042	0.043	0.044	0.037
0.050	0.048	0.042	0.063	0.058	0.057	0.064	0.069	0.071	0.064	0.078
0.075	0.063	0.060	0.080	0.079	0.088	0.087	0.093	0.092	0.093	0.087
0.100	0.077	0.081	0.098	0.091	0.107	0.106	0.111	0.113	0.108	0.098
0.125	0.093	0.094	0.120	0.100	0.130	0.129	0.132	0.136	0.124	0.127
0.150	0.105	0.124	0.138	0.111	0.143	0.150	0.158	0.155	0.149	0.160
0.175	0.116	0.150	0.163	0.123	0.167	0.168	0.172	0.164	0.171	0.188
0.200	0.130	0.177	0.194	0.138	0.190	0.184	0.193	0.178	0.193	0.195
(p,n)=(2500,40)										
0.025	0.035	0.022	0.022	0.046	0.026	0.031	0.039	0.036	0.046	0.039
0.050	0.048	0.041	0.046	0.061	0.047	0.056	0.057	0.061	0.068	0.069
0.075	0.063	0.073	0.070	0.081	0.071	0.073	0.076	0.090	0.089	0.085
0.100	0.077	0.095	0.093	0.089	0.099	0.096	0.098	0.103	0.108	0.103
0.125	0.093	0.119	0.118	0.101	0.117	0.126	0.118	0.119	0.124	0.126
0.150	0.105	0.137	0.145	0.116	0.133	0.150	0.139	0.139	0.145	0.151
0.175	0.116	0.166	0.167	0.132	0.152	0.169	0.159	0.164	0.173	0.196
0.200	0.130	0.179	0.184	0.140	0.183	0.184	0.181	0.188	0.202	0.224

Table 3.2 Empirical sizes of the Oracle test, C-Q test, FDR, maximum test and threshold tests with different threshold levels $\lambda_n = 2s \log(p)$ for process with standardized Gamma(2,2) marginal distribution.

α	Oracle	C-Q	FDR	MAX	s					
					0.50	0.55	0.60	0.65	0.70	0.75
(p,n)=(1000,20)										
0.025	0.052	0.034	0.021	0.041	0.035	0.029	0.034	0.032	0.036	0.039
0.050	0.067	0.061	0.039	0.055	0.051	0.052	0.054	0.052	0.057	0.057
0.075	0.076	0.091	0.069	0.067	0.075	0.076	0.072	0.077	0.072	0.079
0.100	0.085	0.114	0.094	0.073	0.093	0.097	0.088	0.095	0.094	0.107
0.125	0.097	0.142	0.120	0.085	0.115	0.117	0.109	0.115	0.121	0.132
0.150	0.109	0.165	0.144	0.092	0.138	0.140	0.125	0.142	0.142	0.153
0.175	0.121	0.184	0.160	0.108	0.157	0.157	0.152	0.162	0.158	0.159
0.200	0.133	0.203	0.186	0.118	0.182	0.180	0.172	0.171	0.177	0.161
(p,n)=(2000,30)										
0.025	0.057	0.026	0.023	0.033	0.031	0.030	0.040	0.039	0.037	0.034
0.050	0.075	0.054	0.042	0.050	0.056	0.056	0.066	0.061	0.059	0.059
0.075	0.082	0.077	0.062	0.063	0.078	0.080	0.093	0.086	0.080	0.064
0.100	0.097	0.104	0.081	0.074	0.108	0.107	0.108	0.102	0.099	0.081
0.125	0.109	0.133	0.103	0.088	0.128	0.128	0.127	0.131	0.123	0.110
0.150	0.120	0.150	0.122	0.103	0.154	0.147	0.150	0.147	0.146	0.140
0.175	0.134	0.172	0.150	0.111	0.175	0.164	0.177	0.161	0.179	0.168
0.200	0.150	0.199	0.175	0.122	0.190	0.191	0.193	0.181	0.200	0.182
(p,n)=(2500,40)										
0.025	0.057	0.033	0.022	0.045	0.027	0.032	0.039	0.039	0.044	0.042
0.050	0.075	0.056	0.052	0.066	0.059	0.056	0.063	0.053	0.069	0.080
0.075	0.082	0.080	0.076	0.080	0.085	0.088	0.085	0.079	0.097	0.095
0.100	0.097	0.100	0.092	0.087	0.108	0.110	0.106	0.098	0.115	0.104
0.125	0.109	0.116	0.123	0.098	0.123	0.129	0.126	0.120	0.123	0.119
0.150	0.120	0.147	0.144	0.103	0.140	0.146	0.137	0.145	0.140	0.150
0.175	0.134	0.162	0.173	0.120	0.161	0.165	0.161	0.171	0.168	0.181
0.200	0.150	0.184	0.201	0.132	0.174	0.181	0.174	0.184	0.188	0.198

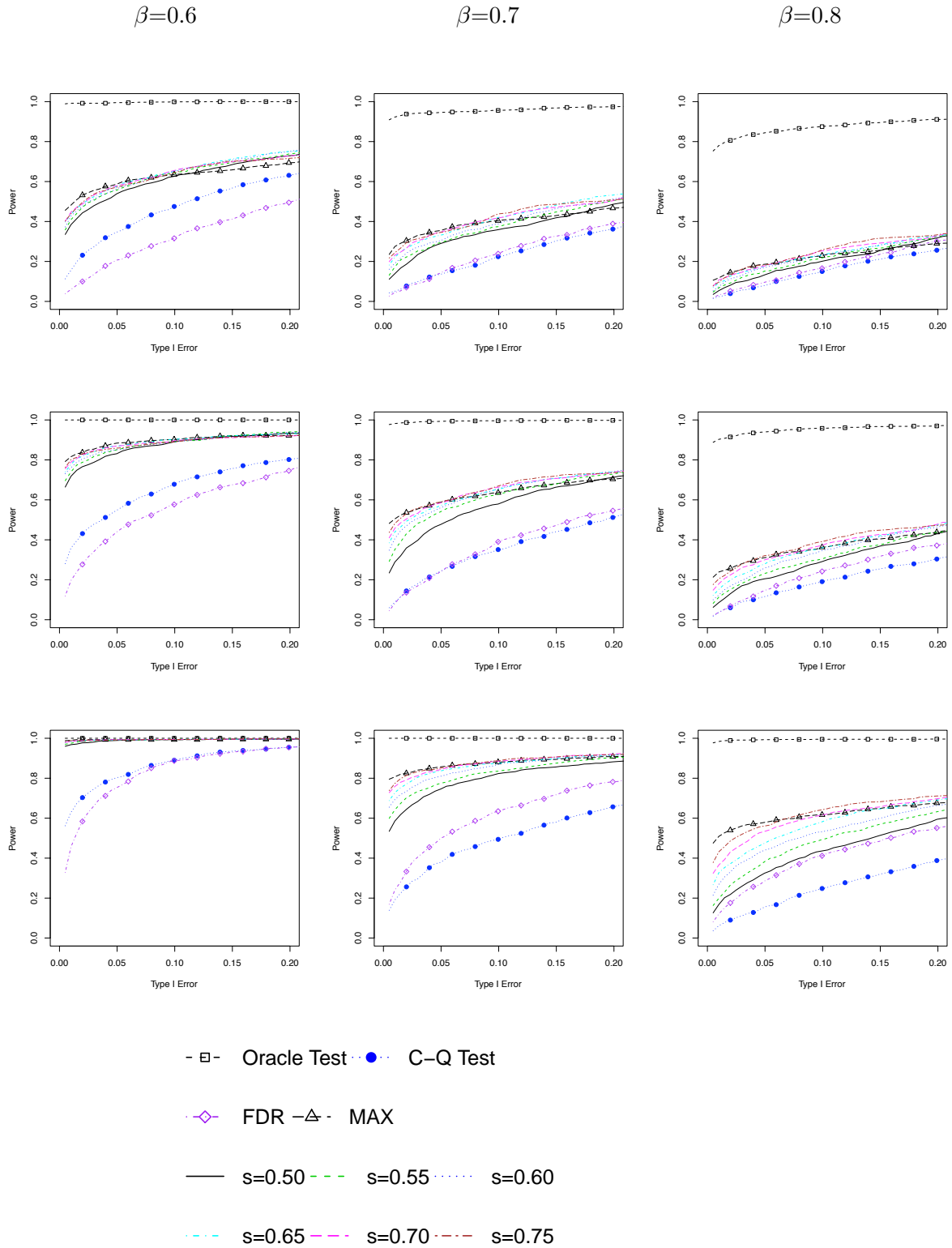


Figure 3.5 The ROC curves of the Oracle test, C-Q test, FDR test, Maximum test and the threshold test at different levels with Type I error between 0-0.2. From top to bottom, $r = 0.4, 0.6$ and 0.9 . From left to right panels $\beta = 0.6, 0.7, 0.8$. ($p = 1000, n = 20$)

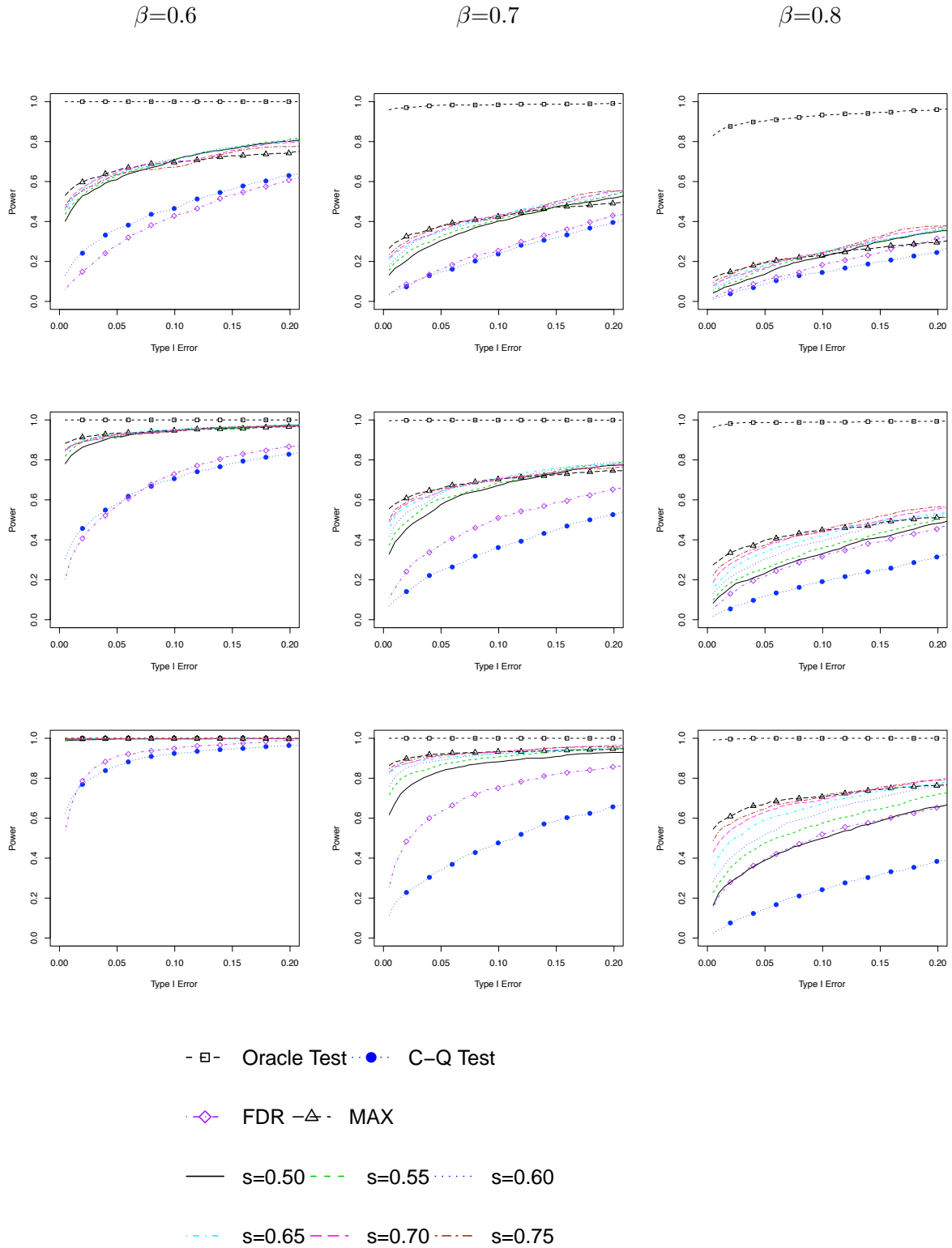


Figure 3.6 The ROC curves of the Oracle test, C-Q test, FDR test, Maximum test and the threshold test at different levels with Type I error between 0-0.2. From top to bottom, $r = 0.4, 0.6$ and 0.9 . From left to right panels $\beta = 0.6, 0.7, 0.8$. ($p = 2000, n = 30$)

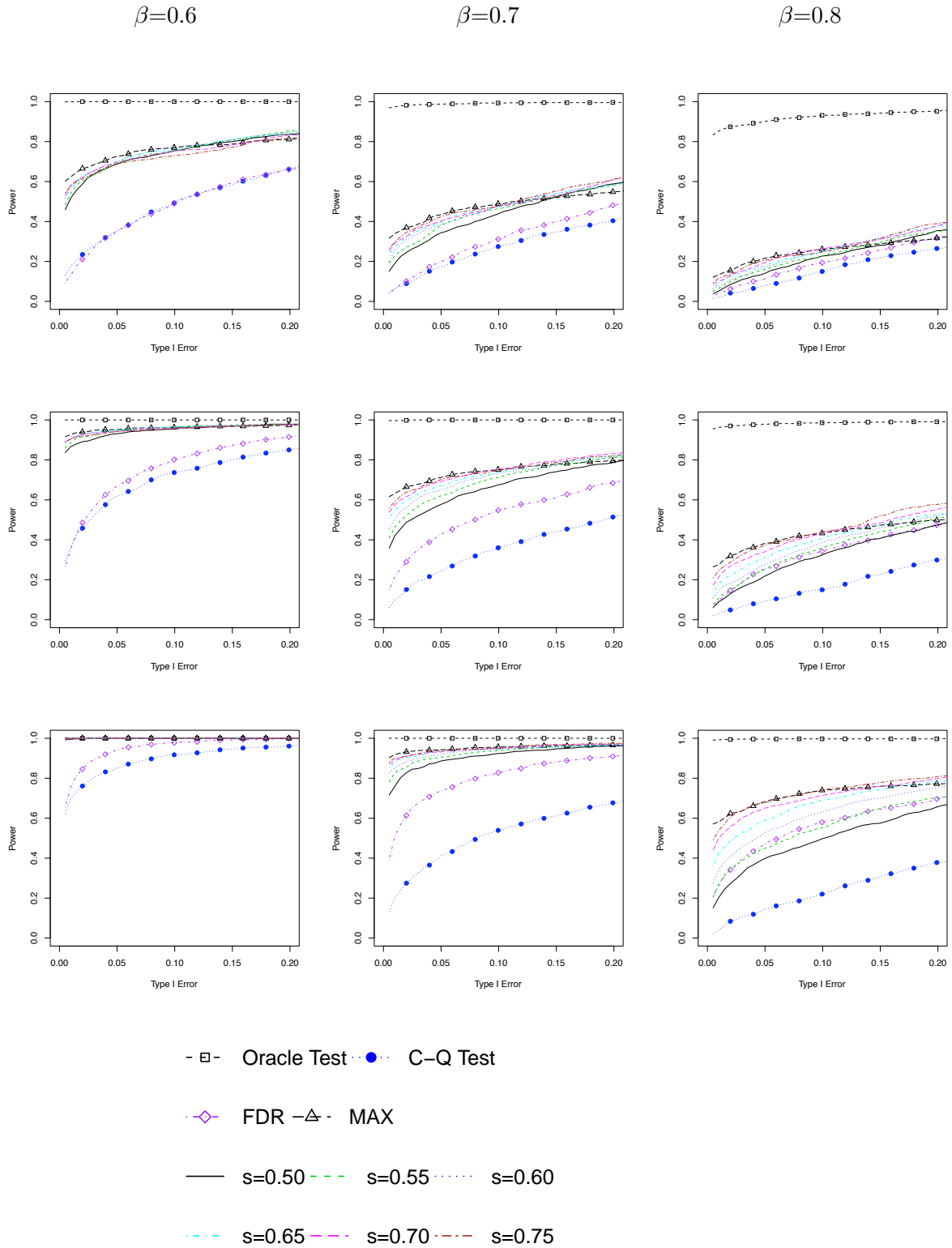


Figure 3.7 The ROC curves of the Oracle test, C-Q test, FDR test, Maximum test and the threshold test at different levels with Type I error between 0-0.2. From top to bottom, $r = 0.4, 0.6$ and 0.9 . From left to right panels $\beta = 0.6, 0.7, 0.8$. ($p = 2500, n = 40$)

Expression of $\tilde{E}\{Z_{1,n}(\lambda_n)Z_{2,n}(\lambda_n)\}$ in Theorem 1:

$$\begin{aligned}
& \tilde{E}\{Z_{1,n}(\lambda_n)Z_{2,n}(\lambda_n)\} \tag{3.7.24} \\
&= A(\eta_{n1}^-, \sqrt{n}\mu_1, \sqrt{n}\mu_2; \rho)Q(\eta_{n1}^-, \eta_{n2}^-; \rho) + A(\eta_{n2}^-, \sqrt{n}\mu_2, \sqrt{n}\mu_1; \rho)Q(\eta_{n2}^-, \eta_{n1}^-; \rho) \\
&+ B(\eta_{n1}^-, \eta_{n2}^-, \sqrt{n}\mu_1, \sqrt{n}\mu_2; \rho)q(\eta_{n1}^-, \eta_{n2}^-; \rho) + C(\sqrt{n}\mu_1, \sqrt{n}\mu_2; \rho)U(\eta_{n1}^-, \eta_{n2}^-; \rho) \\
&+ A(\eta_{n1}^+, -\sqrt{n}\mu_1, -\sqrt{n}\mu_2; \rho)Q(\eta_{n1}^+, \eta_{n2}^+; \rho) + A(\eta_{n2}^+, -\sqrt{n}\mu_2, -\sqrt{n}\mu_1; \rho)Q(\eta_{n2}^+, \eta_{n1}^+; \rho) \\
&+ B(\eta_{n1}^+, \eta_{n2}^+, -\sqrt{n}\mu_1, -\sqrt{n}\mu_2; \rho)q(\eta_{n1}^+, \eta_{n2}^+; \rho) + C(-\sqrt{n}\mu_1, -\sqrt{n}\mu_2; \rho)U(\eta_{n1}^+, \eta_{n2}^+; \rho) \\
&+ A(\eta_{n1}^-, \sqrt{n}\mu_1, -\sqrt{n}\mu_2; -\rho)Q(\eta_{n1}^-, \eta_{n2}^+; -\rho) + A(\eta_{n2}^+, -\sqrt{n}\mu_2, \sqrt{n}\mu_1; -\rho)Q(\eta_{n2}^+, \eta_{n1}^-; -\rho) \\
&+ B(\eta_{n1}^-, \eta_{n2}^+, \sqrt{n}\mu_1, -\sqrt{n}\mu_2; -\rho)q(\eta_{n1}^-, \eta_{n2}^+; -\rho) + C(\sqrt{n}\mu_1, -\sqrt{n}\mu_2; -\rho)U(\eta_{n1}^-, \eta_{n2}^+; -\rho) \\
&+ A(\eta_{n1}^+, -\sqrt{n}\mu_1, \sqrt{n}\mu_2; -\rho)Q(\eta_{n1}^+, \eta_{n2}^-; -\rho) + A(\eta_{n2}^-, \sqrt{n}\mu_2, -\sqrt{n}\mu_1; -\rho)Q(\eta_{n2}^-, \eta_{n1}^+; -\rho) \\
&+ B(\eta_{n1}^+, \eta_{n2}^-, -\sqrt{n}\mu_1, \sqrt{n}\mu_2; -\rho)q(\eta_{n1}^+, \eta_{n2}^-; -\rho) + C(-\sqrt{n}\mu_1, \sqrt{n}\mu_2; -\rho)U(\eta_{n1}^+, \eta_{n2}^-; -\rho)
\end{aligned}$$

where

$$\begin{aligned}
A(\eta, \mu_1, \mu_2; \rho) &= \eta^3 \rho^2 + (1 + 2\rho^2)\eta + 2\mu_2(\eta^2 \rho + 2\rho) + 2\mu_1(1 + \rho^2 + \rho^2 \eta^2) \\
&\quad + 4\mu_1 \mu_2 \rho \eta + \mu_2^2 \eta + \mu_1^2 \rho^2 \eta + 2\mu_1 \mu_2^2 + 2\rho \mu_1^2 \mu_2; \\
B(\eta_1, \eta_2, \mu_1, \mu_2; \rho) &= \sqrt{1 - \rho^2}(\rho \eta_1^2 + \eta_1 \eta_2 + \rho \eta_2^2 + 3\rho) \\
&\quad + 2\mu_2(\eta_1 + \rho \eta_2) + 2\mu_1(\eta_2 + \rho \eta_1) + 4\mu_1 \mu_2 + \rho(\mu_1^2 + \mu_2^2); \\
C(\mu_1, \mu_2; \rho) &= 1 + 2\rho^2 + 4\mu_1 \mu_2 \rho + \mu_1^2 + \mu_2^2 + \mu_1^2 \mu_2^2.
\end{aligned}$$

□

Derivation of $I_{(2)} - I_{(4)}$ in Section 2: To obtain the order of $I_{(2)}$, notice that in set S_0 , we have $\mu_1 = \mu_{k+1} = 0$ and $\lambda_n = 2s \log(p)$.

$$\begin{aligned}
\gamma_k(0, 0, \lambda_n) &= 4\rho_k^2(\lambda_n^{3/2} + 2\lambda_n^{1/2})\phi(\lambda_n^{1/2})[\bar{\Phi}(\theta\lambda_n^{1/2}) + \bar{\Phi}(\theta^{-1}\lambda_n^{1/2})] \\
&\quad + 4\lambda_n^{1/2}\phi(\lambda_n^{1/2})[\bar{\Phi}(\theta\lambda_n^{1/2}) + \bar{\Phi}(\theta^{-1}\lambda_n^{1/2}) - 2\bar{\Phi}(\lambda_n^{1/2})] \\
&\quad + 2\lambda_n\phi(\lambda_n^{1/2})[\sqrt{1 - \rho_k^2}\phi(\theta\lambda_n^{1/2}) + \sqrt{1 - \rho_k^2}\phi(\theta^{-1}\lambda_n^{1/2}) - 2\phi(\lambda_n^{1/2})] \\
&\quad + 2\rho_k\sqrt{1 - \rho_k^2}(2\lambda_n + 3)\phi(\theta\lambda_n^{1/2})\phi(\lambda_n^{1/2}) - 2\rho_k\sqrt{1 - \rho_k^2}(2\lambda_n + 3)\phi(\theta^{-1}\lambda_n^{1/2})\phi(\lambda_n^{1/2}) \\
&\quad + 2[U(\lambda_n^{1/2}, \lambda_n^{1/2}; \rho_k) + U(\lambda_n^{1/2}, \lambda_n^{1/2}; -\rho_k) - 2\bar{\Phi}^2(\theta\lambda_n^{1/2})] \\
&\quad + 4\rho_k^2[U(\lambda_n^{1/2}, \lambda_n^{1/2}; \rho_k) + U(\lambda_n^{1/2}, \lambda_n^{1/2}; -\rho_k)] \tag{3.7.25}
\end{aligned}$$

where $\theta = \sqrt{(1 - \rho_k)/(1 + \rho_k)}$. From the above expression of $\gamma_k(0, 0, \lambda_n)$, we notice that the value of $\gamma_k(0, 0, \lambda_n)$ is the same if we replace ρ_k by $-\rho_k$. Thus, without loss of generality, assume $\rho_k \geq 0$.

Because $\phi(\lambda_n^{1/2}) = \frac{1}{\sqrt{2\pi}}p^{-s}$ and $\phi(\theta\lambda_n^{1/2}) = \frac{1}{\sqrt{2\pi}}p^{-\theta^2 s}$, $\phi(\theta\lambda_n^{1/2})\phi(\lambda_n^{1/2}) \sim p^{-\frac{2s}{1+\rho_k}}$, where “ \sim ” represents that both sides are at the same order. It can be shown that if $x \rightarrow \infty$ and $x > 0$, $\bar{\Phi}(x) \sim \frac{1}{x}\phi(x)$. By an inequality given in Willink (2004), we have for $\rho_k \geq 0$,

$$\bar{\Phi}(\theta\lambda_n^{1/2})\bar{\Phi}(\lambda_n^{1/2}) \leq U(\lambda_n^{1/2}, \lambda_n^{1/2}; \rho_k) \leq \bar{\Phi}(\theta\lambda_n^{1/2})\bar{\Phi}(\lambda_n^{1/2})(1 + \rho_k). \quad (3.7.26)$$

By the Mean Value theorem, there exist a $\theta_0 \in (\theta, 1)$ such that $\bar{\Phi}(\theta x) - \bar{\Phi}(x) = \phi(\theta_0 x)(1 - \theta)x$ and $(1 - \theta)/\rho_k \rightarrow 1$ as $\rho_k \rightarrow 0$. Therefore,

$$\frac{\bar{\Phi}(\theta\lambda_n^{1/2}) - \bar{\Phi}(\lambda_n^{1/2})}{\rho_k \lambda_n^{1/2} \phi(\theta_0 \lambda_n^{1/2})} \rightarrow 1. \quad (3.7.27)$$

Similarly, there exists $\theta'_0 \in (\theta, 1)$ such that

$$\begin{aligned} \sqrt{1 - \rho_k^2} \phi(\theta x) - \phi(x) &= \sqrt{1 - \rho_k^2} \{\phi(\theta x) - \phi(x)\} + (\sqrt{1 - \rho_k^2} - 1)\phi(x) \\ &= \sqrt{1 - \rho_k^2} \phi(\theta'_0 x)(\theta'_0 x)(1 - \theta)x + (\sqrt{1 - \rho_k^2} - 1)\phi(x) \end{aligned}$$

and $(\sqrt{1 - \rho_k^2} - 1)/\rho_k^2 \rightarrow 1$ as $\rho_k \rightarrow 0$. Thus

$$\frac{\sqrt{1 - \rho_k^2} \phi(\theta\lambda_n^{1/2}) - \phi(\lambda_n^{1/2})}{\rho_k \lambda_n \phi(\theta'_0 \lambda_n^{1/2}) + \rho_k^2 \phi(\lambda_n^{1/2})} \rightarrow 1. \quad (3.7.28)$$

By (3.7.26) and (3.7.27), we conclude that

$$\frac{U(\lambda_n^{1/2}, \lambda_n^{1/2}; \rho_k) - \bar{\Phi}^2(\lambda_n^{1/2})}{\rho_k \lambda_n^{1/2} \phi(\theta_0 \lambda_n^{1/2}) \bar{\Phi}(\lambda_n^{1/2})} \rightarrow 1. \quad (3.7.29)$$

Replacing θ with θ^{-1} in (3.7.27), (3.7.28) and (3.7.29) show that $\bar{\Phi}(\theta^{-1}\lambda_n^{1/2}) - \bar{\Phi}(\lambda_n^{1/2})$, $\sqrt{1 - \rho_k^2} \phi(\theta^{-1}\lambda_n^{1/2}) - \phi(\lambda_n^{1/2})$ and $U(\lambda_n^{1/2}, \lambda_n^{1/2}; -\rho_k) - \bar{\Phi}^2(\lambda_n^{1/2})$ are all higher order terms. Thus, in summary, $\gamma_k(0, 0, \lambda_n) \sim \rho_k L_p p^{-\frac{2s}{1+\rho_k}}$.

To evaluate $I_{(3)}$, we consider $\mu_1 = \mu_{k+1} = \sqrt{2r \log(p)/n}$ and $\lambda_n = 2s \log(p)$. Assuming $\rho_k \geq 0$, it can be shown that the leading order terms of (3.7.24) are the first fourth terms.

Hence,

$$\begin{aligned}
\gamma_k(\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}, \lambda_n) &= E\{Z_{1,n}(\lambda_n)Z_{k+1,n}(\lambda_n)\} - \left\{(\eta_{n_1}^+) \phi(\eta_{n_1}^-) + (n\mu_1^2 + 1)(\bar{\Phi}(\eta_{n_1}^-))\right\}^2 \\
&= 2A(\eta_{n_1}^-, \sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k)Q(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) - 2(\eta_{n_1}^+)(n\mu_1^2 + 1)\phi(\eta_{n_1}^-)\bar{\Phi}(\eta_{n_1}^-) \\
&\quad + B(\eta_{n_1}^-, \eta_{n(k+1)}^-; \sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k)q(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) - (\eta_{n_1}^+)^2\phi^2(\eta_{n_1}^-) \\
&\quad + C(\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k)U(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) - (n\mu_1^2 + 1)^2\bar{\Phi}^2(\eta_{n_1}^-) \tag{3.7.30}
\end{aligned}$$

where

$$\begin{aligned}
&A(\eta_{n_1}^-, \sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k)Q(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) - (\eta_{n_1}^+)(n\mu_1^2 + 1)\phi(\eta_{n_1}^-)\bar{\Phi}(\eta_{n_1}^-) \\
&= (\eta_{n_1}^+)(n\mu_1^2 + 1)\phi(\eta_{n_1}^-)\{\bar{\Phi}(\theta\eta_{n_1}^-) - \bar{\Phi}(\eta_{n_1}^-)\} + \rho_k\{\eta_{n_1}^-^3\rho_k + 2\rho_k\eta_{n_1}^- + 2\sqrt{n}\mu_{k+1}(\eta_{n_1}^-^2 + 2) \\
&\quad + 2\sqrt{n}\mu_1(\rho_k + \eta_{n_1}^-^2) + 4n\mu_1\mu_{k+1}\eta_{n_1}^- + n\mu_1^2\rho_k\eta_{n_1}^- + 2n^{3/2}\mu_1^2\mu_{k+1}\}Q(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k); \tag{3.7.31}
\end{aligned}$$

$$\begin{aligned}
&B(\eta_{n_1}^-, \eta_{n(k+1)}^-; \sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k)q(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) - (\eta_{n_1}^+)^2\phi^2(\eta_{n_1}^-) \\
&= \{\sqrt{1 - \rho_k^2}\phi(\theta\eta_{n_1}^-) - \phi(\eta_{n_1}^-)\}\phi(\eta_{n_1}^-)(\eta_{n_1}^+)^2 \\
&\quad + \sqrt{1 - \rho_k^2}\rho_k\{(\eta_{n_1}^- + \sqrt{n}\mu_1)^2 + (\eta_{n(k+1)}^- + \sqrt{n}\mu_{k+1})^2 + 3\}q(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) \tag{3.7.32}
\end{aligned}$$

and

$$\begin{aligned}
&C(\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k)U(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) - (n\mu_1^2 + 1)^2\bar{\Phi}^2(\eta_{n_1}^-) \\
&= \{U(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) - \bar{\Phi}^2(\eta_{n_1}^-)\}(n\mu_1^2 + 1)^2 + 2\rho_k(\rho_k + 2n\mu_1\mu_{k+1})U(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) \\
&= \bar{\Phi}(\eta_{n_1}^-)\{c\bar{\Phi}(\theta\eta_{n_1}^-) - \bar{\Phi}(\eta_{n_1}^-)\}(n\mu_1^2 + 1)^2 + 2\rho_k(\rho_k + 2n\mu_1\mu_{k+1})U(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) \tag{3.7.33}
\end{aligned}$$

for some $c \in [1, 1 + \rho_k]$ and $\rho_k \geq 0$.

Notice that $\bar{\Phi}(-\theta\eta_{n_1}^-) - \bar{\Phi}(-\eta_{n_1}^-) = \bar{\Phi}(\eta_{n_1}^-) - \bar{\Phi}(\theta\eta_{n_1}^-)$ and $U(-\eta_{n_1}^-, -\eta_{n_2}^-; \rho_k) - \bar{\Phi}^2(-\eta_{n_1}^-) = U(\eta_{n_1}^-, \eta_{n_2}^-; \rho_k) - \bar{\Phi}^2(\eta_{n_1}^-)$. So no matter $\eta_{n_1}^- > 0$ or < 0 , the above differences maintain the same order. Hence, without loss of generality, we can assume $\eta_{n_1}^- > 0$. Similar to (3.7.27), (3.7.28) and (3.7.29), we have $\{\rho_k\eta_{n_1}^-\phi(\theta\eta_{n_1}^-)\}^{-1}(\bar{\Phi}(\theta\eta_{n_1}^-) - \bar{\Phi}(\eta_{n_1}^-)) \rightarrow 1$,

$$\frac{\sqrt{1 - \rho_k^2}\phi(\theta\eta_{n_1}^-) - \phi(\eta_{n_1}^-)}{\rho_k\eta_{n_1}^-^2\phi(\theta\eta_{n_1}^-) + \rho_k^2\phi(\eta_{n_1}^-)} \rightarrow 1 \quad \text{and} \quad \frac{U(\eta_{n_1}^-, \eta_{n_2}^-; \rho_k) - \bar{\Phi}^2(\eta_{n_1}^-)}{\rho_k\eta_{n_1}^-\phi(\theta\eta_{n_1}^-)\bar{\Phi}(\eta_{n_1}^-)} \rightarrow 1.$$

Thus it can be shown that the orders of the first term in (3.7.31), (3.7.32) and (3.7.33) are small order of $\rho_k L_p p^{-\frac{2}{1+\rho_k}(\sqrt{s}-\sqrt{r})^2}$. It can be also shown that, $q(\eta_{n1}^-, \eta_{n2}^-, \rho_k) \sim p^{-\frac{2}{1+\rho_k}(\sqrt{s}-\sqrt{r})^2}$,

$$\begin{aligned} Q(\eta_{n1}^-, \eta_{n2}^-, \rho_k) &\sim I(s > r) L_p p^{-\frac{2}{1+\rho_k}(\sqrt{s}-\sqrt{r})^2} + I(s < r) L_p p^{-(\sqrt{s}-\sqrt{r})^2} \quad \text{and} \\ U(\eta_{n1}^-, \eta_{n2}^-, \rho_k) &\sim I(s > r) L_p p^{-\frac{2}{1+\rho_k}(\sqrt{s}-\sqrt{r})^2} + I(s < r) 4r \log(p) U(\eta_{n1}^-, \eta_{n2}^-, \rho_k). \end{aligned}$$

In summary, we have for $\rho_k \geq 0$

$$\gamma_k(\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}, \lambda_n) \sim I(s > r) \rho_k L_p p^{-\frac{2}{1+\rho_k}(\sqrt{s}-\sqrt{r})^2} + I(s < r) \rho_k 4r \log(p) U(\eta_{n1}^-, \eta_{n2}^-, \rho_k). \quad (3.7.34)$$

For $\rho_k < 0$, note that for any $\eta > 0$, $\bar{\Phi}(\theta\eta)\bar{\Phi}(\eta)(1+\rho_k) \leq U(\eta, \eta; \rho_k) \leq \bar{\Phi}(\theta\eta)\bar{\Phi}(\eta)$. Then it can be shown that the leading order terms of $\gamma_k(\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}, \lambda_n)$ are

$$\gamma_k(\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}, \lambda_n) = J_1 I(\rho_k^2 s < (\rho_k - 1)^2 r) + J_2 I(\rho_k^2 s > (\rho_k - 1)^2 r)$$

where

$$\begin{aligned} J_1 &:= 2A(\eta_{n1}^-, \sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k) Q(\eta_{n1}^-, \eta_{n(k+1)}^-; \rho_k) - 2(\eta_{n1}^+) \phi(\eta_{n1}^-) (n\mu_{k+1}^2 + 1) \bar{\Phi}(\eta_{n(k+1)}^-) \\ &\quad + B(\eta_{n1}^-, \eta_{n(k+1)}^-, \sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k) q(\eta_{n1}^-, \eta_{n(k+1)}^-; \rho_k) - (\eta_{n1}^+)^2 \phi^2(\eta_{n1}^-) \\ &\quad + C(\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k) U(\eta_{n1}^-, \eta_{n(k+1)}^-; \rho_k) - (n\mu_{k+1}^2 + 1)^2 \bar{\Phi}^2(\eta_{n(k+1)}^-) \end{aligned}$$

and

$$\begin{aligned} J_2 &:= A(\eta_{n(k+1)}^+, -\sqrt{n}\mu_{k+1}, \sqrt{n}\mu_1; -\rho_k) Q(\eta_{n(k+1)}^+, \eta_{n1}^-; -\rho_k) - (\eta_{n(k+1)}^-) \phi(\eta_{n(k+1)}^+) (n\mu_1^2 + 1) \bar{\Phi}(\eta_{n1}^-) \\ &\quad + A(\eta_{n1}^+, -\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; -\rho_k) Q(\eta_{n1}^+, \eta_{n(k+1)}^-; -\rho_k) - (\eta_{n1}^-) \phi(\eta_{n1}^+) (n\mu_{k+1}^2 + 1) \bar{\Phi}(\eta_{n(k+1)}^-) \\ &\quad + B(\eta_{n1}^-, \eta_{n(k+1)}^-, \sqrt{n}\mu_1, -\sqrt{n}\mu_{k+1}; \rho_k) q(\eta_{n1}^-, \eta_{n(k+1)}^+; -\rho_k) - \eta_{n(k+1)}^+ \eta_{n1}^- \phi(\eta_{n1}^-) \phi(\eta_{n(k+1)}^+) \\ &\quad + B(\eta_{n1}^-, \eta_{n(k+1)}^-, -\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k) q(\eta_{n(k+1)}^+, \eta_{n1}^-; -\rho_k) - \eta_{n(k+1)}^+ \eta_{n1}^- \phi(\eta_{n1}^-) \phi(\eta_{n(k+1)}^+) \\ &\quad + 2C(\sqrt{n}\mu_1, -\sqrt{n}\mu_{k+1}; -\rho) U(\eta_{n1}^-, \eta_{n(k+1)}^+; -\rho) - (n\mu_{k+1}^2 + 1)^2 \bar{\Phi}(\eta_{n(k+1)}^-) \bar{\Phi}(\eta_{n(k+1)}^+). \end{aligned}$$

In a similar fashion as that for $\rho_k \geq 0$, we will have for $\rho_k < 0$,

$$\begin{aligned}
\rho_k^{-1} \gamma_k(\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}, \lambda_n) &\sim I\left(r < \min\left\{\frac{\rho_k^2}{(\rho_k - 1)^2}, \frac{(\rho_k + 1)^2}{(\rho_k - 1)^2}\right\} s\right) L_p p^{-2\left(\frac{s}{1-\rho_k} + \frac{r}{1+\rho_k}\right)} \\
&+ I\left(s > r > \min\left\{\frac{\rho_k^2}{(\rho_k - 1)^2}, \frac{(\rho_k + 1)^2}{(\rho_k - 1)^2}\right\} s\right) L_p \max\left\{p^{-\frac{2}{1+\rho_k}(\sqrt{s}-\sqrt{r})^2}, p^{-(\sqrt{s}+\sqrt{r})^2}\right\} \\
&+ I(s < r) 4r \log(p) U(\eta_{n1}^-, \eta_{n2}^-; \rho_k) \\
&\leq I(s > r) L_p p^{-\frac{2}{1-\rho_k}(\sqrt{s}-\sqrt{r})^2} + I(s < r) 4r \log(p) U(\eta_{n1}^-, \eta_{n2}^-; \rho_k).
\end{aligned} \tag{3.7.35}$$

Thus, together with (3.7.34), we have for $\rho_k \neq 0$,

$$\rho_k^{-1} \gamma_k(\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}, \lambda_n) \leq I(s > r) L_p p^{-\frac{2}{1+|\rho_k|}(\sqrt{s}-\sqrt{r})^2} + I(s < r) 4r \log(p) U(\eta_{n1}^-, \eta_{n2}^-; \rho_k). \tag{3.7.36}$$

and the “ \sim ” relation holds if $r > s$. $I_{(3)}$ follows from the above expression.

For $I_{(4)}$, assume $\mu_1 = \sqrt{2r \log(p)/n}$, $\mu_{k+1} = 0$. Let $\gamma_k(\mu_1, 0, \lambda_n; \rho_k)$ be the covariance between $Z_{1,n}(\lambda_n)$ and $Z_{k+1,n}(\lambda_n)$. Notice that $\gamma_k(\mu_1, 0, \lambda_n; \rho_k) = \gamma_k(\mu_1, 0, \lambda_n; -\rho_k)$. Thus, we only need to show $I_{(4)}$ for $\rho_k \geq 0$. It can be shown that the leading order terms of (3.7.24) are

$$\begin{aligned}
\gamma_k(\mu_1, 0, \lambda_n) &= \{A(\eta_{n1}^-, \sqrt{n}\mu_1, 0; \rho_k) Q(\eta_{n1}^-, \eta_{n(k+1)}^-; \rho_k) - (\eta_{n1}^+) \phi(\eta_{n1}^-) \bar{\Phi}(\eta_{n(k+1)}^-)\} I(s > r) \\
&+ \{A(\eta_{n1}^-, \sqrt{n}\mu_1, 0; -\rho_k) Q(\eta_{n1}^-, \eta_{n(k+1)}^+; -\rho_k) - (\eta_{n1}^+) \phi(\eta_{n1}^-) \bar{\Phi}(\eta_{n(k+1)}^+)\} I(s < r) \\
&+ A(\eta_{n(k+1)}^-, 0, \sqrt{n}\mu_1; \rho_k) Q(\eta_{n(k+1)}^-, \eta_{n1}^-; \rho_k) - (\eta_{n(k+1)}^+) (n\mu_1^2 + 1) \phi(\eta_{n(k+1)}^-) \bar{\Phi}(\eta_{n1}^-) \\
&+ B(\eta_{n1}^-, \eta_{n(k+1)}^-, \sqrt{n}\mu_1, 0; \rho_k) q(\eta_{n1}^-, \eta_{n(k+1)}^-; \rho_k) - (\eta_{n1}^+) (\eta_{n(k+1)}^+) \phi(\eta_{n1}^-) \phi(\eta_{n(k+1)}^-) \\
&+ C(\sqrt{n}\mu_1, 0; \rho_k) U(\eta_{n1}^-, \eta_{n(k+1)}^-; \rho_k) - (n\mu_1^2 + 1) \bar{\Phi}(\eta_{n1}^-) \bar{\Phi}(\eta_{n(k+1)}^-) \\
&:= P_I + P_{II} + P_{III} + P_{IV} + P_V.
\end{aligned} \tag{3.7.37}$$

Let $\xi = \frac{2(1-\rho_k)\sqrt{s}(\sqrt{s}-\sqrt{r})+r}{1-\rho_k^2}$. Firstly, for P_I , we only need consider $s > r$,

$$\begin{aligned}
P_I &= \rho_k \{ \eta_{n1}^-^3 \rho_k + 2\rho_k \eta_{n1}^- + 2\sqrt{n}\mu_1(\rho_k + \eta_{n1}^-^2) + n\mu_1^2 \rho_k \eta_{n1}^- \} Q(\eta_{n1}^-, \eta_{n(k+1)}^-; \rho_k) \\
&+ (\eta_{n1}^+) \phi(\eta_{n1}^-) \left\{ \bar{\Phi}\left(\frac{\eta_{n(k+1)}^- - \rho_k \eta_{n1}^-}{\sqrt{1-\rho_k^2}}\right) - \bar{\Phi}(\eta_{n(k+1)}^-) \right\};
\end{aligned}$$

where $Q(\eta_{n1}^-, \eta_{n(k+1)}^-; \rho_k) \sim L_p p^{-\xi}$ and

$$\begin{aligned}
\bar{\Phi}\left(\frac{\eta_{n(k+1)}^- - \rho_k \eta_{n1}^-}{\sqrt{1-\rho_k^2}}\right) - \bar{\Phi}(\eta_{n(k+1)}^-) &\sim \rho_k \left\{ I\left(s > \frac{1}{4}(1+\theta^{-1})^2 r\right) L_p p^{-\frac{[(1-\rho_k)\sqrt{s}+\rho_k\sqrt{r}]^2}{1-\rho_k^2}} \right. \\
&\left. - I\left(r < s < \frac{1}{4}(1+\theta^{-1})^2 r\right) L_p p^{-s} \right\}.
\end{aligned}$$

Thus $P_I \sim \rho_k \left\{ I(s > \frac{1}{4}(1 + \theta^{-1})^2 r) L_p p^{-\xi} - I(r < s < \frac{1}{4}(1 + \theta^{-1})^2 r) L_p p^{-s - (\sqrt{s} - \sqrt{r})^2} \right\}$.

For P_{II} , we only need consider $s < r$,

$$P_{II} = \rho_k \{ \eta_{n_1}^{-3} \rho_k + 2\rho_k \eta_{n_1}^- + 2\sqrt{n} \mu_1 (\rho_k - \eta_{n_1}^{-2}) + n \mu_1^2 \rho_k \eta_{n_1}^- \} Q(\eta_{n_1}^-, \eta_{n(k+1)}^+; \rho_k) \\ + (\eta_{n_1}^+) \phi(\eta_{n_1}^-) \{ \bar{\Phi} \left(\frac{\eta_{n(k+1)}^+ + \rho_k \eta_{n_1}^-}{\sqrt{1 - \rho_k^2}} \right) - \bar{\Phi}(\eta_{n(k+1)}^-) \}$$

where

$$Q(\eta_{n_1}^-, \eta_{n(k+1)}^+; -\rho_k) \sim I(s < \rho_k^2 r / (1 + \rho_k)^2) L_p p^{-(\sqrt{s} - \sqrt{r})^2} \\ + I(r > s > \rho_k^2 r / (1 + \rho_k)^2) L_p p^{-\xi - \frac{4\rho_k}{1 - \rho_k^2} (s - \sqrt{sr})} \quad \text{and} \\ \bar{\Phi} \left(\frac{\eta_{n(k+1)}^+ + \rho_k \eta_{n_1}^-}{\sqrt{1 - \rho_k^2}} \right) - \bar{\Phi}(\eta_{n(k+1)}^-) \sim \rho_k \left\{ I(s < \rho_k^2 r / (1 + \rho_k)^2) C \right. \\ \left. + I(\rho_k^2 r / (1 + \rho_k)^2 < s < r) L_p p^{-\frac{[(1 + \rho_k)\sqrt{s} - \rho_k \sqrt{r}]^2}{1 - \rho_k^2}} \right\}.$$

Thus $P_{II} \sim \rho_k \left\{ I(s < \frac{\rho_k^2}{(1 + \rho_k)^2} r) L_p p^{-(\sqrt{s} - \sqrt{r})^2} + I(\frac{\rho_k^2}{(1 + \rho_k)^2} r < s < r) L_p p^{-\xi - \frac{4\rho_k}{1 - \rho_k^2} (s - \sqrt{sr})} \right\}$.

Now let us consider P_{III} ,

$$P_{III} = \rho_k \{ \eta_{n(k+1)}^{-3} \rho_k + 2\rho_k \eta_{n(k+1)}^- + 2\sqrt{n} \mu_1 (2 + \eta_{n(k+1)}^{-2}) \} Q(\eta_{n(k+1)}^-, \eta_{n_1}^-; \rho_k) \\ + (\eta_{n(k+1)}^+) (n \mu_1^2 + 1) \phi(\eta_{n(k+1)}^-) \{ \bar{\Phi} \left(\frac{\eta_{n_1}^- - \rho_k \eta_{n(k+1)}^-}{\sqrt{1 - \rho_k^2}} \right) - \bar{\Phi}(\eta_{n_1}^-) \}$$

where $Q(\eta_{n(k+1)}^-, \eta_{n_1}^-; \rho_k) \sim I(s > r / (1 - \rho_k)^2) L_p p^{-\xi} + I(s < r / (1 - \rho_k)^2) L_p p^{-s}$ and

$$\bar{\Phi} \left(\frac{\eta_{n_1}^- - \rho_k \eta_{n(k+1)}^-}{\sqrt{1 - \rho_k^2}} \right) - \bar{\Phi}(\eta_{n_1}^-) \sim \rho_k \left\{ I(s > \frac{r}{(1 - \rho_k)^2}) L_p p^{-\frac{[(1 - \rho_k)\sqrt{s} - \sqrt{r}]^2}{1 - \rho_k^2}} \right. \\ \left. + I(r < s < \frac{r}{(1 - \rho_k)^2}) C + I(s < r) L_p p^{-(\sqrt{s} - \sqrt{r})^2} \right\}.$$

Thus $P_{III} \sim \rho_k \left\{ I(s > r / (1 - \rho_k)^2) L_p p^{-\xi} + I(s < r / (1 - \rho_k)^2) L_p p^{-s} \right\}$.

Next, we evaluate P_{IV}

$$P_{IV} = \rho_k \sqrt{1 - \rho_k^2} (\eta_{n_1}^2 + \eta_{n(k+1)}^2 + 3 + 2\sqrt{n} \mu_1 \eta_{n_1} + n \mu_1^2) q(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) \\ + (\eta_{n_1}^+) (\eta_{n(k+1)}^+) \{ \sqrt{1 - \rho_k^2} q(\eta_{n_1}^-, \eta_{n(k+1)}^-; \rho_k) - \phi(\eta_{n_1}^-) \phi(\eta_{n(k+1)}^-) \}$$

where $q(\eta_{n1}^-, \eta_{n(k+1)}^-; \rho_k) \sim p^{-\xi}$ and $\sqrt{1 - \rho_k^2} q(\eta_{n1}^-, \eta_{n(k+1)}^-; \rho_k) - \phi(\eta_{n1}^-) \phi(\eta_{n(k+1)}^-) \sim \rho_k p^{-\xi}$. Thus $P_{IV} \sim \rho_k L_p p^{-\xi}$. Finally, taking a look at P_V

$$P_V = 2\rho_k^2 U(\eta_{n1}^-, \eta_{n2}^-; \rho_k) + (n\mu_1^2 + 1) \{U(\eta_{n1}^-, \eta_{n2}^-; \rho_k) - \bar{\Phi}(\eta_{n1}^-) \bar{\Phi}(\eta_{n2}^-)\}.$$

Because $U(\eta_{n1}^-, \eta_{n2}^-; \rho_k) = z(\rho_k) \bar{\Phi}(\eta_{n2}^-) \bar{\Phi}\left(\frac{\eta_{n1}^- - \rho_k \eta_{n2}^-}{\sqrt{1 - \rho_k^2}}\right)$ for some $z(\rho_k) \rightarrow 1$ as $\rho_k \rightarrow 0$, P_V has the same order as P_{III} up to a slow varying function.

Thus, in summary,

$$\begin{aligned} \gamma_k(\sqrt{n}\mu_1, 0, \lambda_n) \sim \rho_k \left\{ I(s < \rho_k^2 r / (1 + \rho_k)^2) L_p p^{-(\sqrt{s} - \sqrt{r})^2} \right. \\ \left. + I(\rho_k^2 r / (1 + \rho_k)^2 < s < r / (1 - \rho_k)^2) L_p p^{-s} + I(s > r / (1 - \rho_k)^2) L_p p^{-\xi} \right\}. \end{aligned}$$

Since $s < \rho_k^2 r / (1 + \rho_k)^2$, we have $\sqrt{r} \geq (1/\rho_k + 1)\sqrt{s} \geq 2\sqrt{s}$. It follows that $(\sqrt{s} - \sqrt{r})^2 - s = \sqrt{r}(\sqrt{r} - 2\sqrt{s}) \geq 0$. It is easy to see $\xi - s = ((1 - \rho_k)\sqrt{s} - \sqrt{r})^2 \geq 0$. It follows that $|\gamma_k(\sqrt{n}\mu_1, 0, \lambda_n)| \leq |\rho_k| L_p p^{-s}$. \square

Spectral Density Based Estimation of $\sigma_0^2(p; \lambda_n)$: Let $\iota^2 = -1$ be the imaginary number.

The spectral density of $\{Z_{i,n}^0(\lambda_n)\}_{i=1}^p$ is

$$g(w) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_k(0, 0, \lambda_n) \exp(-\iota k w) \text{ for } w \in [-\pi, \pi].$$

According to Brockwell and Davis (2009), $\lim_{p \rightarrow \infty} \sigma_0^2(p; \lambda_n) = 2\pi g(0)$. Hence we can estimate $\sigma_0^2(p; \lambda_n)$ by estimating $g(0)$. To estimate $g(0)$, we will introduce periodogram of $Z_{i,n}^0(\lambda_n)$ which is defined as

$$I_p(w_k) = \frac{1}{p} \left| \sum_{t=1}^p Z_{t,n}^0(\lambda_n) e^{-\iota t w_k} \right|^2, \quad k = -[(p-1)/2], \dots, -1, 0, 1, \dots, [p/2],$$

where $w_k = 2\pi k/p$. By Theorem 2.13 in Fan and Yao (2005), we know that for $k \in \mathcal{T} = \{ -[(p-1)/2], \dots, -1, 1, \dots, [p/2] \}$,

$$I_p(w_k) = \sum_{\tau=-(p-1)}^{p-1} \hat{\gamma}(\tau) e^{-\iota \tau w_k},$$

where $\hat{\gamma}(\tau) = \frac{1}{p} \sum_{t=1}^{p-\tau} (Z_{t,n}^0(\lambda_n) - \bar{Z}^0)(Z_{t+\tau,n}^0(\lambda_n) - \bar{Z}^0)$ and $\bar{Z}^0 = \frac{1}{p} \sum_{t=1}^p Z_{t,n}^0$. Notice that $Z_{t,n}^0(\lambda_n)$ is unobservable when the data coming from the alternative, hence we can not use it

directly. However, because $\hat{\gamma}(\tau)$ is a consistent estimate of $\gamma_\tau(0, 0, \lambda_n)$, we can replace $\hat{\gamma}(\tau)$ by $\hat{\gamma}_\tau(0, 0, \lambda_n)$ with $\hat{\rho}_k$ being estimated and plugging in to (3.3.15). Hence we have

$$\hat{I}_p(w_k) = \sum_{\tau=-(p-1)}^{p-1} \hat{\gamma}_\tau(0, 0, \lambda_n) e^{-\iota\tau w_k}.$$

By Brillinger (1981) for $k \in \mathcal{T}$,

$$I_p(w_k) = 2\pi g(w_k) E_k + R_k \quad (3.7.38)$$

where $\{E_k\}$ are independent standard exponential random variables and $\{R_k\}$ are asymptotically negligible terms. As commonly used method for estimating the spectral density (see Fan and Yao, 2005; Fuller, 1996; Chen and Tang, 2005), let $W_k = \log\{\hat{I}_p(w_k)/(2\pi)\} + 0.57721$, $\varepsilon_k = \log(E_k) + 0.57721$ and $m(w) = \log\{g(w)\}$. Then (3.7.38) can be approximated as

$$W_k = m(w_k) + \varepsilon_k,$$

where ε_k are IID random variables with mean 0 and variance $\pi^2/6$. Hence we can estimate $m(w)$ by

$$\hat{m}_h(w) = \frac{\sum_{k \in \mathcal{T}} K_h(w - w_k) W_k}{\sum_{k \in \mathcal{T}} K_h(w - w_k)},$$

where $K_h(t) = K(t/h)/h$, $K(\cdot)$ is the kernel function and h is the bandwidth. Then an estimator of $g(0)$ is $\hat{g}(0) = \exp\{\hat{m}_h(0)\}$ and $\hat{\sigma}_0^2(p; \lambda_n) = 2\pi\hat{g}(0)$. \square

Detection Boundary of the Threshold Test: The analyses are similar to the analyses given by Chen and Xu (2011) for the normal and independent random variables. We will discuss the detection boundary of the threshold test by four cases. For each case, we find the corresponding detectable region and the union of the four detectable regions is the detectable region of threshold test.

Case 1: $s \leq r$ and $s \leq \beta$. In this case, $\mu_{T_{n,1}} - \mu_{T_{n,0}} = L_p p^{1-\beta}$ and $\sigma_{T_{n,1}} = \sigma_{T_{n,0}} = L_p p^{(1-s)/2}$. Hence

$$\frac{\mu_{T_{n,1}} - \mu_{T_{n,0}}}{\sigma_{T_{n,1}}} = L_p p^{(1+s-2\beta)/2}. \quad (3.7.39)$$

So to make the test detectable, i.e. such that $(\mu_{T_{n,1}} - \mu_{T_{n,0}})/\sigma_{T_{n,1}} \rightarrow \infty$, $s > 2\beta - 1$. It follows that the detectable region in (β, r) plane for this case is $r \geq 2\beta - 1$ (Chen and Xu,

2011). If we could select $s = \min\{r, \beta\}$, then the best power of the threshold test is of order $L_p p^{(1+\min\{r, \beta\}-2\beta)/2}$.

Case 2: $s \leq r$ and $s > \beta$. In this case, $\mu_{Tn,1} - \mu_{Tn,0} = L_p p^{1-\beta}$, $\sigma_{Tn,1} = L_p p^{(1-\beta)/2}$, and $\sigma_{Tn,0} = L_p p^{(1-s)/2}$. Then

$$\frac{\mu_{Tn,1} - \mu_{Tn,0}}{\sigma_{Tn,1}} = L_p p^{(1-\beta)/2}. \quad (3.7.40)$$

So the detectable region in (β, r) plane is $r > \beta$. In this detection region, the best power is of order $L_p p^{(1-\beta)/2}$.

Case 3: $s > r$ and $s \leq (\sqrt{s} - \sqrt{r})^2 + \beta$. The case is equivalent to $\sqrt{r} < \sqrt{s} \leq (r + \beta)/(2\sqrt{r})$. In this case, $\mu_{Tn,1} - \mu_{Tn,0} = L_p p^{1-(\sqrt{s}-\sqrt{r})^2-\beta}$, $\sigma_{Tn,1} = \sigma_{Tn,0} = L_p p^{(1-s)/2}$. Then

$$\frac{\mu_{Tn,1} - \mu_{Tn,0}}{\sigma_{Tn,1}} = L_p p^{\frac{1}{2} - \beta + r - (\sqrt{s} - 2\sqrt{r})^2/2}. \quad (3.7.41)$$

So to make the test detectable, we want (3.7.41) goes to infinity. It follows that we need

$$2\sqrt{r} - \sqrt{1 - 2\beta + 2r} < \sqrt{s} < 2\sqrt{r} + \sqrt{1 - 2\beta + 2r}.$$

Thus the detectable region in (β, r) plane must satisfy

$$\sqrt{r} < (r + \beta)/(2\sqrt{r}), 1 - 2\beta + 2r > 0 \text{ and } 2\sqrt{r} - \sqrt{1 - 2\beta + 2r} \leq (r + \beta)/(2\sqrt{r}).$$

It is corresponding to the detectable region

$$r < \beta, r > \beta - \frac{1}{2} \text{ and } \{r \leq \beta/3 \text{ or } (r > \beta/3 \text{ and } r \geq (1 - \sqrt{1 - \beta})^2)\}.$$

Now in the above detectable region, if $2\sqrt{r} \leq (r + \beta)/(2\sqrt{r})$, i.e., $r \leq \beta/3$ then we can take $\sqrt{s} = 2\sqrt{r}$. So the best power under area $r \leq \beta/3$ in the detectable region is of order $L_p p^{\frac{1}{2} - \beta + r}$. If $r > \beta/3$, the best power is of order $L_p p^{\frac{1}{2} - \frac{1}{2}(r + \beta)^2/(4r)}$, which is attained at $\sqrt{s} = (r + \beta)/(2\sqrt{r})$.

Case 4: $s > r$ and $s > (\sqrt{s} - \sqrt{r})^2 + \beta$. This is equivalent to $\sqrt{s} > \max\{(r + \beta)/(2\sqrt{r}), \sqrt{r}\}$. In this case, $\mu_{Tn,1} - \mu_{Tn,0} = L_p p^{1-(\sqrt{s}-\sqrt{r})^2-\beta}$, $\sigma_{Tn,1} = L_p p^{(1-(\sqrt{s}-\sqrt{r})^2-\beta)/2}$. Then

$$\frac{\mu_{Tn,1} - \mu_{Tn,0}}{\sigma_{Tn,1}} = L_p p^{(1-(\sqrt{s}-\sqrt{r})^2-\beta)/2}. \quad (3.7.42)$$

The detectable condition requires that

$$\sqrt{r} - \sqrt{1 - \beta} < \sqrt{s} < \sqrt{r} + \sqrt{1 - \beta}.$$

In order to find an s such that we could implement the test, we need $\sqrt{r} + \sqrt{1 - \beta} > \max\{(r + \beta)/(2\sqrt{r}), \sqrt{r}\}$. If $\sqrt{r} > (r + \beta)/(2\sqrt{r})$, i.e. $r > \beta$, then the above inequality is obviously true. If $r \leq \beta$, then $\sqrt{r} + \sqrt{1 - \beta} > (r + \beta)/(2\sqrt{r})$ is equivalent to $r > (1 - \sqrt{1 - \beta})^2$. So the detectable region is $r > (1 - \sqrt{1 - \beta})^2$.

When $r > \beta$, we can take $s = r$ such that (3.7.42) is of order $L_p p^{(1-\beta)/2}$. If $(1 - \sqrt{1 - \beta})^2 < r \leq \beta$, then the best rate of (3.7.42) is attained at $\sqrt{s} = (r + \beta)/(2\sqrt{r})$, where the best rate is $L_p p^{\frac{1}{2}} - \frac{1}{2}(r + \beta)^2/(4r)$. \square

Lemma 1 *Let $X_i = (X_{i1}, X_{i2})'$ be IID random vectors with mean zero and covariance Σ . Suppose that there exist a positive H such that $E(e^{h'X_i}) < \infty$ for $h \in [-H, H] \times [-H, H]$. Then for $t = (t_1, t_2)' > 0$, we have*

$$1 - F_n(t_1, t_2) = \exp\left\{\frac{\theta(t_1, t_2)}{\sqrt{n}}\right\} \bar{\Phi}_{0, \Sigma}(t) \{1 + O(n^{-1/2}t)\} \quad (3.7.43)$$

hold uniformly for $t_1, t_2 = o(n^{1/2})$ where $\bar{\Phi}_{0, \Sigma}(t) = (2\pi|\Sigma|^{1/2})^{-1} \int_t^\infty \exp\left(-\frac{1}{2}y'\Sigma^{-1}y\right) dy$ and $\theta(t_1, t_2) = b_{30}t_1^3 + b_{12}t_1^2t_2 + b_{21}t_1t_2^2 + b_{03}t_2^3$ with

$$\begin{aligned} b_{30} &= \frac{1}{2}\kappa_{(2,1)}c_1^2c_2 + \frac{1}{2}\kappa_{(1,2)}c_1c_2^2 + \frac{1}{3!}\kappa_{(3,0)}c_1^3 + \frac{1}{3!}\kappa_{(0,3)}c_2^3, \\ b_{03} &= \frac{1}{2}\kappa_{(2,1)}c_1c_2^2 + \frac{1}{2}\kappa_{(1,2)}c_1^2c_2 + \frac{1}{3!}\kappa_{(3,0)}c_2^3 + \frac{1}{3!}\kappa_{(0,3)}c_1^3, \\ b_{12} &= \frac{1}{2}\kappa_{(2,1)}(c_1^3 + 2c_1c_2^2) + \frac{1}{2}\kappa_{(1,2)}(c_2^3 + 2c_1^2c_2) + \frac{1}{2}\kappa_{(3,0)}c_1^2c_2 + \frac{1}{2}\kappa_{(0,3)}c_1c_2^2 \text{ and} \\ b_{21} &= \frac{1}{2}\kappa_{(2,1)}(c_2^3 + 2c_1^2c_2) + \frac{1}{2}\kappa_{(1,2)}(c_1^3 + 2c_1c_2^2) + \frac{1}{2}\kappa_{(3,0)}c_1c_2^2 + \frac{1}{2}\kappa_{(0,3)}c_1^2c_2, \end{aligned}$$

where $c_1 = -(\kappa_{(1,1)}^2 - \kappa_{(0,2)}\kappa_{(2,0)})^{-1}\kappa_{(0,2)}$ and $c_2 = (\kappa_{(1,1)}^2 - \kappa_{(0,2)}\kappa_{(2,0)})^{-1}\kappa_{(1,1)}$.

Proof Let $V(x_1, x_2)$ be the distribution function of X_i . Introduce a conjugate random vector \tilde{X}_i with a common distribution function

$$\tilde{V}(x_1, x_2) = \frac{1}{R(h)} \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} e^{h'y} dV(y)$$

where $R(h) = E(e^{h'X_1}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{h'y} dV(y)$.

Define $v(t_1, t_2) = E\{\exp(it'X_1)\}$, $\tilde{v}(t_1, t_2) = E\{\exp(it'\tilde{\Sigma}^{-\frac{1}{2}}\tilde{X}_1)\}$, $\tilde{m} = E(\tilde{X}_1)$, $\tilde{\Sigma} = \text{Var}(\tilde{X}_1)$, $S_n = \sum_{j=1}^n X_j$ and $\tilde{S}_n = \sum_{j=1}^n \tilde{\Sigma}^{-\frac{1}{2}}\tilde{X}_j$. Let x be a two-dimensional vector. $W_n(x) = P(S_n < x)$, $\tilde{W}_n(x) = P(\tilde{S}_n < x)$, $w_n(t) = E(e^{it'S_n})$, $\tilde{w}_n(t) = E(e^{it'\tilde{S}_n})$.

$$F_n(x) = P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i < x\right), \quad \text{and} \quad \tilde{F}_n(x) = P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\Sigma}^{-\frac{1}{2}}(\tilde{X}_i - \tilde{m}) < x\right).$$

Then $F_n(x) = W_n(\sqrt{n}x)$, $\tilde{F}_n(x) = \tilde{W}_n(\sqrt{n}x + n\tilde{\Sigma}^{-\frac{1}{2}}\tilde{m})$, $w_n(t) = v^n(t)$ and $\tilde{w}_n(t) = \tilde{v}^n(t)$.

$$\begin{aligned} \tilde{v}(t) &= \int \int \exp\{it'\tilde{\Sigma}^{-\frac{1}{2}}x\} d\tilde{V}(x) = \frac{1}{R(h)} \int \int \exp\{(h - it'\tilde{\Sigma}^{-\frac{1}{2}})x\} dV(x) \\ &= \frac{1}{R(h)} v(\tilde{\Sigma}^{-\frac{1}{2}}t - ih). \end{aligned}$$

So, $v(t) = R\tilde{v}(\tilde{\Sigma}^{\frac{1}{2}}(t + ih))$ and $w_n(t) = R^n\tilde{w}_n(\tilde{\Sigma}^{\frac{1}{2}}(t + ih))$. Hence,

$$\begin{aligned} W_n(x) &= \int_{-\infty}^x \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \exp(-it'y) w_n(t) dt dy \\ &= \int_{-\infty}^x \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \exp(-it'y) \tilde{w}_n(\tilde{\Sigma}^{\frac{1}{2}}(t + ih)) dt dy \\ &= \int_{-\infty}^x R^n e^{-h'y} |\tilde{\Sigma}|^{-\frac{1}{2}} \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \exp(-iz'\tilde{\Sigma}^{-\frac{1}{2}}y) \tilde{w}_n(z) dz dy \\ &= \int_{-\infty}^x R^n e^{-h'y} |\tilde{\Sigma}|^{-\frac{1}{2}} f_{\tilde{S}_n}(\tilde{\Sigma}^{-\frac{1}{2}}y) dy = \int_{-\infty}^x R^n e^{-h'y} d\tilde{W}_n(\tilde{\Sigma}^{-\frac{1}{2}}y). \end{aligned}$$

So it follows that

$$\begin{aligned} 1 - F_n(x) &= 1 - W_n(\sqrt{n}x) = \int_{\sqrt{n}x}^{\infty} R^n e^{-h'y} d\tilde{W}_n(\tilde{\Sigma}^{-\frac{1}{2}}y) \\ &= \int_{\sqrt{n}x}^{\infty} R^n e^{-h'y} d\tilde{F}_n\left(\frac{1}{\sqrt{n}}\tilde{\Sigma}^{-\frac{1}{2}}y - \sqrt{n}\tilde{\Sigma}^{-\frac{1}{2}}\tilde{m}\right) \\ &= \int_B R^n e^{-h'(\sqrt{n}\tilde{\Sigma}^{\frac{1}{2}}t + n\tilde{m})} d\tilde{F}_n(t) \\ &= \exp\{n \log R - nh\tilde{m}\} \int_B e^{-\sqrt{nh}\tilde{\Sigma}^{\frac{1}{2}}t} d\tilde{F}_n(t), \end{aligned}$$

where $B = \{t : \tilde{\Sigma}^{\frac{1}{2}}t > x - \sqrt{n}\tilde{m}\}$.

Now let $x = \sqrt{n\tilde{m}}$, we get

$$\begin{aligned} 1 - F_n(\sqrt{n\tilde{m}}) &= \exp\{n \log R - nh\tilde{m}\} \int_{\{t: \tilde{\Sigma}^{\frac{1}{2}} t > 0\}} e^{-\sqrt{nh'} \tilde{\Sigma}^{\frac{1}{2}} t} d\tilde{F}_n(t) \\ &= \exp\{n \log R - nh\tilde{m}\} \int_{\{t: \tilde{\Sigma}^{\frac{1}{2}} t > 0\}} e^{-\sqrt{nh'} \tilde{\Sigma}^{\frac{1}{2}} t} d\Phi_{0, I_2}(t) \end{aligned} \quad (3.7.44)$$

$$+ \exp\{n \log R - nh\tilde{m}\} \int_{\{t: \tilde{\Sigma}^{\frac{1}{2}} t > 0\}} e^{-\sqrt{nh'} \tilde{\Sigma}^{\frac{1}{2}} t} d(\tilde{F}_n(t) - \Phi_{0, I_2}(t)) \quad (3.7.45)$$

$$:= \exp\{n \log R - nh\tilde{m}\} \{I + II\} = P_I + P_{II}.$$

Now let us firstly take a look at (3.7.44). Then

$$\begin{aligned} &\exp\{n \log R - nh\tilde{m}\} \int_{\{t: \tilde{\Sigma}^{\frac{1}{2}} t > 0\}} e^{-\sqrt{nh'} \tilde{\Sigma}^{\frac{1}{2}} t} d\Phi_{0, I_2}(t) \\ &= \exp\{n \log R - nh\tilde{m}\} \frac{1}{2\pi} |\tilde{\Sigma}|^{-\frac{1}{2}} \int_0^\infty \exp\{-\sqrt{nh'} \xi - \frac{\xi' \tilde{\Sigma}^{-1} \xi}{2}\} d\xi \\ &= \exp\{n \log R - nh\tilde{m} + \frac{nh' \tilde{\Sigma} h}{2}\} \frac{1}{2\pi} |\tilde{\Sigma}|^{-\frac{1}{2}} \int_0^\infty \exp\{-\frac{(\xi + \sqrt{n\tilde{\Sigma}h})' \tilde{\Sigma}^{-1} (\xi + \sqrt{n\tilde{\Sigma}h})}{2}\} d\xi. \end{aligned} \quad (3.7.46)$$

For sufficient small h ,

$$\log R(h) = \sum_{|\nu|=1}^{\infty} \frac{k_\nu}{\nu!} h^\nu$$

where k_ν is the cumulant of order $\nu = (\nu_1, \nu_2)$. $k_\nu = 0$ if $|\nu| = 1$. $k_\nu = \text{Cov}(X_{i_1}, X_{i_2})$ if $\nu = (1, 1)$.

$k_\nu = \mu_{(2,0)} - \mu_{(1,0)}^2$ if $\nu = (2, 0)$. $k_\nu = \mu_{(0,2)} - \mu_{(0,1)}^2$ if $\nu = (0, 2)$.

$$\tilde{m} = E(\tilde{X}_1) = \frac{1}{R(h)} \int y \exp(h'y) dV(y) = \frac{\partial \log R(h)}{\partial h}.$$

Then

$$\begin{aligned} \frac{\partial \log R(h)}{\partial h_1} &= k_{(1,1)} h_2 + k_{(2,0)} h_1 + \frac{1}{2} k_{(3,0)} h_1^2 + k_{(2,1)} h_1 h_2 + \frac{1}{2} k_{(1,2)} h_2^2 \\ \frac{\partial \log R(h)}{\partial h_2} &= k_{(1,1)} h_1 + k_{(0,2)} h_2 + \frac{1}{2} k_{(3,0)} h_2^2 + k_{(1,2)} h_1 h_2 + \frac{1}{2} k_{(2,1)} h_1^2. \end{aligned}$$

Let $a = \tilde{m}$ and

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} k_{(1,1)} h_2 + k_{(2,0)} h_1 + \frac{1}{2} k_{(3,0)} h_1^2 + k_{(2,1)} h_1 h_2 + \frac{1}{2} k_{(1,2)} h_2^2 \\ k_{(1,1)} h_1 + k_{(0,2)} h_2 + \frac{1}{2} k_{(3,0)} h_2^2 + k_{(1,2)} h_1 h_2 + \frac{1}{2} k_{(2,1)} h_1^2 \end{pmatrix}.$$

Then the approximate solution of above equations are:

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} c_1 a_1 + c_2 a_2 \\ d_1 a_1 + d_2 a_2 \end{pmatrix} \quad (3.7.47)$$

where $d_1 = c_2$ and $d_2 = c_1$. We can also know that

$$\begin{aligned} \tilde{\Sigma} &= \frac{\partial^2 \log R(h)}{\partial h \partial h} = \begin{pmatrix} k_{(2,0)} + k_{(3,0)}h_1 + k_{(2,1)}h_2 & k_{(1,1)} + k_{(2,1)}h_1 + k_{(1,2)}h_2 \\ k_{(1,1)} + k_{(2,1)}h_1 + k_{(1,2)}h_2 & k_{(0,2)} + k_{(0,3)}h_2 + k_{(1,2)}h_1 \end{pmatrix} \\ &\xrightarrow{h_1, h_2 \rightarrow 0} \begin{pmatrix} k_{(2,0)} & k_{(1,1)} \\ k_{(1,1)} & k_{(0,2)} \end{pmatrix} = \Sigma. \end{aligned}$$

Thus $\tilde{\Sigma}h = \tilde{m}\{1 + O(h)\}$ and $\tilde{\Sigma} = \Sigma\{1 + O(h)\}$. It follows that

$$\log R - h\tilde{m} + \frac{h'\tilde{\Sigma}h}{2} = \frac{1}{2}k_{(2,1)}h_1^2h_2 + \frac{1}{2}k_{(1,2)}h_1h_2^2 + \frac{1}{3!}k_{(3,0)}h_1^3 + \frac{1}{3!}k_{(0,3)}h_2^3 + O(h^4). \quad (3.7.48)$$

Plugging (3.7.47) into (3.7.48), we get

$$\log R - h\tilde{m} + \frac{h'\tilde{\Sigma}h}{2} = \theta(a_1, a_2) + O(a^4)$$

where $\theta(a_1, a_2)$ is given in Lemma 1. Now combing it with (3.7.46) and replacing \tilde{m} by a , we get

$$P_I = \exp\left\{n\theta(a_1, a_2)\right\} \frac{1}{2\pi} |\Sigma|^{-\frac{1}{2}} \int_0^\infty \exp\left\{-\frac{(\xi + \sqrt{na})'\Sigma^{-1}(\xi + \sqrt{na})}{2}\right\} d\xi. \quad (3.7.49)$$

If P_{II} is a smaller order of P_I in (3.7.44), Then it can be seen that, $1 - F_n(\sqrt{na}) = P_I\{1 + o(1)\}$.

Let $t = \sqrt{na}$. we obtain the result (3.7.43).

It remains to show that P_{II} is a smaller order of P_I . We only need to show II is a small order of I . Choosing $h = (h_1, h_2)$ such that $h_1, h_2 \rightarrow 0$ as $n \rightarrow \infty$ and $nh_1^2, nh_2^2 \leq C_0$ for sufficient large n . Following from (3.7.46), we know

$$\begin{aligned} I &= \frac{1}{2\pi} |\tilde{\Sigma}|^{-\frac{1}{2}} \int_0^\infty \exp\left\{-\sqrt{nh}'\xi - \frac{\xi'\tilde{\Sigma}^{-1}\xi}{2}\right\} d\xi \\ &= \exp\left\{\frac{nh'\tilde{\Sigma}h}{2}\right\} \frac{1}{2\pi} |\tilde{\Sigma}|^{-\frac{1}{2}} \int_0^\infty \exp\left\{-\frac{(\xi + \sqrt{n\tilde{\Sigma}h})'\tilde{\Sigma}^{-1}(\xi + \sqrt{n\tilde{\Sigma}h})}{2}\right\} d\xi. \end{aligned}$$

Assume $\tilde{\Sigma} = \begin{pmatrix} a^2 & ac\tilde{\rho} \\ ac\tilde{\rho} & c^2 \end{pmatrix}$ and $L = \begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix}$. Let $\xi = L\xi^*$. Then

$$\begin{aligned} & \frac{1}{2\pi} |\tilde{\Sigma}|^{-\frac{1}{2}} \int_0^\infty \exp\left\{-\frac{(\xi + \sqrt{n}\tilde{\Sigma}h)' \tilde{\Sigma}^{-1} (\xi + \sqrt{n}\tilde{\Sigma}h)}{2}\right\} d\xi \\ &= \frac{1}{2\pi} |L| |\tilde{\Sigma}|^{-\frac{1}{2}} \int_0^\infty \exp\left\{-\frac{(\xi^* + L^{-1}\tilde{\Sigma}\sqrt{nh})' (L^{-1}\tilde{\Sigma}L^{-1})^{-1} (\xi^* + L^{-1}\tilde{\Sigma}\sqrt{nh})}{2}\right\} d\xi^* \\ &= \frac{1}{2\pi} |\tilde{\Sigma}^*|^{-\frac{1}{2}} \int_0^\infty \exp\left\{-\frac{(\xi^* + L^{-1}\tilde{\Sigma}\sqrt{nh})' \tilde{\Sigma}^{*-1} (\xi^* + L^{-1}\tilde{\Sigma}\sqrt{nh})}{2}\right\} d\xi^* \\ &= U(\sqrt{n}(ah_1 - c\tilde{\rho}h_2), \sqrt{n}(a\tilde{\rho}h_1 - ch_2); \tilde{\rho}) \end{aligned}$$

where $\tilde{\Sigma}^* = \begin{pmatrix} 1 & \tilde{\rho} \\ \tilde{\rho} & 1 \end{pmatrix}$. It follows that

$$I = \exp\left\{\frac{nh'\tilde{\Sigma}h}{2}\right\} U(\sqrt{n}(ah_1 - c\tilde{\rho}h_2), \sqrt{n}(a\tilde{\rho}h_1 - ch_2); \tilde{\rho}).$$

According to the choice of h , $\sqrt{n}(ah_1 - c\tilde{\rho}h_2)$ and $\sqrt{n}(a\tilde{\rho}h_1 - ch_2)$ are bounded constants. Thus, I is bounded away from 0 and infinity for sufficient large n .

Now let us show that $II = O(n^{-\frac{1}{2}})$. Let $Q(t) = \tilde{F}_n(t) - \Phi_{0,I_2}(t)$ and $B_0 = \{t : \tilde{\Sigma}^{\frac{1}{2}}t > 0\}$. Let $A_1 = \{t : Q(t) > 0\}$ and $A_2 = \{t : Q(t) < 0\}$ be the sets where Q take positive and negative values. Writing $Q^+(t)$ and $Q^-(t)$ to represent the positive and negative parts of the function $Q(t)$. Then $Q(t) = Q^+(t) - Q^-(t)$. Denote $\tilde{F}_{n,Q^+}(A) = \tilde{F}_n(A \cap A_1)$ and $\tilde{F}_{n,Q^-}(A) = \tilde{F}_n(A \cap A_2)$. Similarly, we could define $\Phi_{0,I_2,Q^+}(A)$ and $\Phi_{0,I_2,Q^-}(A)$. Then

$$Q^+(A) = \tilde{F}_{n,Q^+}(A) - \Phi_{0,I_2,Q^+}(A) \text{ and } Q^-(A) = \tilde{F}_{n,Q^-}(A) - \Phi_{0,I_2,Q^-}(A).$$

It follows that

$$\int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} dQ^+(t) = \int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} d\tilde{F}_{n,Q^+}(t) - \int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} d\Phi_{0,I_2,Q^+}(t).$$

Then $\int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} d\tilde{F}_{n,Q^+}(t) = \int_{B_0 \cap A_1} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} d\tilde{F}_n(t)$. Since $e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t}$ is positive and $\int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} d\tilde{F}_n(t) < \infty$, we have $\int_{B_0 \cap A_1} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} d\tilde{F}_n(t) \leq \int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} d\tilde{F}_n(t) < \infty$, which means $\int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} d\tilde{F}_{n,Q^+}(t) < \infty$. Similarly, we can show $\int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} d\Phi_{0,I_2,Q^+}(t) < \infty$. Thus

$$\int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} dQ^+(t) < \infty.$$

Fix $\epsilon > 0$, there exist a finite $C > 0$ such that $\int_{B_0 \cap D_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} dQ^+(t) < \epsilon/2$, where $D_0 = \{e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} > C\}$. Then

$$\begin{aligned} \int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} dQ^+(t) &= \int_{B_0 \cap D_0^c} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} dQ^+(t) + \int_{B_0 \cap D_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} dQ^+(t) \\ &\leq CQ^+(B_0 \cap D_0^c) + \epsilon/2 = CQ(B_0 \cap D_0^c \cap A_1) + \epsilon/2. \end{aligned} \quad (3.7.50)$$

By the Berry-Essen bounds given by Bhattacharya (1968), we have

$$\sup_B |Q(B)| = \sup_B |\tilde{F}_n(B) - \Phi_{0, I_2}(B)| \leq cn^{-\frac{1}{2}}\theta_{3+\delta}^{3(1+\delta)/(3+\delta)},$$

where $\theta_s = \sum_{i=1}^2 E|\tilde{X}_{1i}|^s$. As shown on page 181 in Petrov (1995), if the Cramér's condition hold, then $\theta_{3+\delta} < \infty$. Therefore, $Q(B_0 \cap D_0^c \cap A_1) \leq Cn^{-\frac{1}{2}}$. Because of the arbitrary of ϵ and from (3.7.50), we know $\int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} dQ^+(t) = O(n^{-\frac{1}{2}})$. Similarly, we can also show that $\int_{B_0} e^{-\sqrt{nh'}\tilde{\Sigma}^{\frac{1}{2}}t} dQ^-(t) = O(n^{-\frac{1}{2}})$. Hence, $II = O(n^{-\frac{1}{2}})$ and II is a smaller order of I . Therefore, P_{II} is a smaller order of P_I . This completes the proof of Lemma 1. \square

Lemma 2 *Let (X, Y) be bivariate normal random vector*

$$(X, Y)' \sim N \left((0, 0)', \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

Using the definition of $U(\lambda, \eta; \rho)$, $Q(\lambda, \eta; \rho)$ and $q(\lambda, \eta; \rho)$ given in Section 2, we have

$$\begin{aligned} (i) \quad E\{X^2Y^2I(X > \eta)I(Y > \lambda)\} &= (1 + 2\rho^2)U(\lambda, \eta; \rho) + (\lambda^3\rho^2 + (1 + 2\rho^2)\lambda)Q(\lambda, \eta; \rho) \\ &\quad + (\eta^3\rho^2 + (1 + 2\rho^2)\eta)Q(\eta, \lambda; \rho) + \sqrt{1 - \rho^2}(\rho\eta^2 + \lambda\eta + \rho\lambda^2 + 3\rho)q(\eta, \lambda; \rho); \\ (ii) \quad E\{X^2YI(X > \eta)I(Y > \lambda)\} &= (2 + \eta^2)\rho Q(\eta, \lambda; \rho) + (1 + \lambda^2\rho^2 + \rho^2)Q(\lambda, \eta; \rho) \\ &\quad + \sqrt{1 - \rho^2}\eta q(\eta, \lambda; \rho) + \rho\sqrt{1 - \rho^2}\lambda q(\lambda, \eta; \rho); \\ (iii) \quad E\{XYI(X > \eta)I(Y > \lambda)\} &= \rho\eta Q(\eta, \lambda; \rho) + \rho\lambda Q(\lambda, \eta; \rho) + \sqrt{1 - \rho^2}q(\eta, \lambda; \rho); \end{aligned}$$

$$(iv) \quad E\{XI(X > \eta)I(Y > \lambda)\} = Q(\eta, \lambda; \rho) + \rho Q(\lambda, \eta; \rho);$$

(v) Let $\eta'_1 = \eta - \mu_1$ and $\lambda'_2 = \lambda - \mu_2$. Then

$$\begin{aligned} & E\{(X + \mu_1)^2(Y + \mu_2)^2I((X + \mu_1) > \eta)I((Y + \mu_2) > \lambda)\} \\ &= E\{X^2Y^2I(X > \eta'_1)I(Y > \lambda'_2)\} + 2\mu_2E\{X^2YI(X > \eta'_1)I(Y > \lambda'_2)\} \\ & \quad + 2\mu_1E\{XY^2I(X > \eta'_1)I(Y > \lambda'_2)\} + \mu_2^2E\{X^2I(X > \eta'_1)I(Y > \lambda'_2)\} \\ & \quad + \mu_1^2E\{Y^2I(X > \eta'_1)I(Y > \lambda'_2)\} + 4\mu_1\mu_2E\{XYI(X > \eta'_1)I(Y > \lambda'_2)\}. \end{aligned}$$

The proof of Lemma 2 is straightforward but tedious, hence we omit it here.

Proof of Theorem 1 We first calculate the mean of the test statistic. By Fubini's theorem,

$$\begin{aligned} E(T_n) &= \sum_{i=1}^p E\{Y_{i,n}I(Y_{i,n} \geq \lambda_n)\} = \sum_{i=1}^p \int_0^\infty \int_{z \vee \lambda_n}^\infty dF(y)dz \\ &= \sum_{i=1}^p \lambda_n P(Y_{i,n} \geq \lambda_n) + \sum_{i=1}^p \int_{\lambda_n}^\infty P(Y_{i,n} \geq z)dz \\ &= \sum_{i=1}^p \lambda_n P(\sqrt{n}|\bar{X}_i| \geq \sqrt{\lambda_n}) + 2 \sum_{i=1}^p \int_{\sqrt{\lambda_n}}^\infty z P(\sqrt{n}|\bar{X}_i| \geq z)dz. \end{aligned} \quad (3.7.51)$$

Without loss of generality, we would assume $\mu_i = E(X_i) \geq 0$ as we can replace X_i with $X_i^* = -X_i$ in (3.7.51). This replacement won't change the value in (3.7.51), so the following analysis hold exactly the same for X_i^* .

Followed by (3.7.51), we have

$$\begin{aligned} E(T_n) &= \sum_{i=1}^p \lambda_n \{P(\sqrt{n}(\bar{X}_i - \mu_i) \geq \eta_{ni}^-) + P(\sqrt{n}(\bar{X}_i - \mu_i) \leq -\eta_{ni}^+)\} \\ & \quad + 2 \sum_{i=1}^p \int_{\sqrt{\lambda_n}}^\infty z \{P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) + P(\sqrt{n}(\bar{X}_i - \mu_i) \leq -z - \sqrt{n}\mu_i)\} dz. \end{aligned} \quad (3.7.52)$$

By the result given on page 183 in Petrov (1995), we have

$$\begin{aligned} P(\sqrt{n}(\bar{X}_i - \mu_i) \geq y) &= \bar{\Phi}(y) \left\{ \exp\left(\frac{\kappa_3 y^3}{6\sigma^3 n^{1/2}}\right) \right\} \left[1 + o\left(\frac{1+y}{n^{1/2}}\right) \right] \quad \text{and} \\ P(\sqrt{n}(\bar{X}_i - \mu_i) \leq -y) &= \bar{\Phi}(y) \left\{ \exp\left(-\frac{\kappa_3 y^3}{6\sigma^3 n^{1/2}}\right) \right\} \left[1 + o\left(\frac{1+y}{n^{1/2}}\right) \right] \end{aligned}$$

hold uniformly for any $y = o(n^{1/6})$. Based on this result we know that for the integral in (3.7.52), we can use the normal approximation if $z + \sqrt{n}\mu_i = o(n^{1/6})$ while for $z + \sqrt{n}\mu_i$ is high

order than $n^{1/6}$ we can not use normal approximation. However, the Cramér's condition in the theorem implies that (Lemma 2.2 in Petrov (1995))

$$P(|X_{ij} - \mu_i| > x) \leq be^{-x^d}. \quad (3.7.53)$$

Hence, under the exponential tail assumption (3.7.53), the integral above high order $n^{1/6}$ is actually a smaller order than the first term in (3.7.52).

We give an evaluation of the first term of (3.7.52). From the following inequality, for $y > 0$,

$$\frac{y}{\sqrt{2\pi}(1+y^2)}e^{-\frac{y^2}{2}} \leq \bar{\Phi}(y) \leq \frac{1}{\sqrt{2\pi}y}e^{-\frac{y^2}{2}}$$

and hence for a sufficient large y , $\bar{\Phi}(y) \sim \frac{1}{\sqrt{2\pi}y}e^{-\frac{y^2}{2}}$. It follows that if $\sqrt{\lambda_n} > \sqrt{n}\mu_i$, then

$$\begin{aligned} & P(\sqrt{n}(\bar{X}_i - \mu_i) \geq \eta_{ni}^-) + P(\sqrt{n}(\bar{X}_i - \mu_i) \leq -\eta_{ni}^+) = \{\bar{\Phi}(\eta_{ni}^-) + \bar{\Phi}(\eta_{ni}^+)\}\{1 + o(1)\} \\ & = \{(\sqrt{2\pi}(\eta_{ni}^-))^{-1} \exp(-\frac{1}{2}(\eta_{ni}^-)^2) + (\sqrt{2\pi}(\eta_{ni}^+))^{-1} \exp(-\frac{1}{2}(\eta_{ni}^+)^2)\}\{1 + o(1)\}. \end{aligned} \quad (3.7.54)$$

If $\sqrt{\lambda_n} < \sqrt{n}\mu_i$,

$$\begin{aligned} & P(\sqrt{n}(\bar{X}_i - \mu_i) \geq \eta_{ni}^-) + P(\sqrt{n}(\bar{X}_i - \mu_i) \leq -\eta_{ni}^+) \\ & = 1 - P(\sqrt{n}(\bar{X}_i - \mu_i) \leq \eta_{ni}^-) + P(\sqrt{n}(\bar{X}_i - \mu_i) \leq -\eta_{ni}^+) \\ & = \{1 - \bar{\Phi}(-\eta_{ni}^-) + \bar{\Phi}(\eta_{ni}^+)\}\{1 + o(1)\} \\ & = \{1 - (\sqrt{2\pi}(-\eta_{ni}^-))^{-1} \exp(-\frac{1}{2}(\eta_{ni}^-)^2) + (\sqrt{2\pi}(\eta_{ni}^+))^{-1} \exp(-\frac{1}{2}(\eta_{ni}^+)^2)\}\{1 + o(1)\}. \end{aligned} \quad (3.7.55)$$

Let $\epsilon > 0$ be an arbitrary small number and $\xi_n = O(n^{1/6-\epsilon})$ such that ξ_n is a large order than $\lambda_n^{1/2}$ and $\sqrt{n}\mu_i$. Consider the third and fourth terms of the integral in (3.7.52), which can be written as

$$\int_{\sqrt{\lambda_n}}^{\infty} zP(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i)dz = \int_{\sqrt{\lambda_n}}^{C\xi_n} zP(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i)dz \quad (3.7.56)$$

$$+ \int_{C\xi_n}^{\infty} zP(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i)dz. \quad (3.7.57)$$

and

$$\int_{\sqrt{\lambda_n}}^{\infty} zP(\sqrt{n}(\bar{X}_i - \mu_i) \leq -z - \sqrt{n}\mu_i)dz = \int_{\sqrt{\lambda_n}}^{C\xi_n} zP(\sqrt{n}(\bar{X}_i - \mu_i) \leq -z - \sqrt{n}\mu_i)dz \quad (3.7.58)$$

$$+ \int_{C\xi_n}^{\infty} zP(\sqrt{n}(\bar{X}_i - \mu_i) \leq -z - \sqrt{n}\mu_i)dz. \quad (3.7.59)$$

We want to show that both (3.7.57) and (3.7.59) are smaller order than the first term in (3.7.52). To obtain the bound for (3.7.57) and (3.7.59), we need the following inequality. Assume $1 < d \leq 2$ in (3.7.53) and $\sum_{j=1}^n a_j^2 = 1$, we have (See, Huang *et al.*, 2008)

$$\max_{a_j} P \left(\left| \sum_{j=1}^n a_j X_{ij} - \mu_i \right| > t \right) \leq \exp(-t^d/M),$$

where M is a constant. Notice that $\xi_n - \sqrt{n}\mu_i \rightarrow \infty$. Then the sum of (3.7.57) and (3.7.59) can be bounded by

$$\begin{aligned} & \int_{C\xi_n}^{\infty} z P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) dz + \int_{C\xi_n}^{\infty} z P(\sqrt{n}(\bar{X}_i - \mu_i) \leq -z - \sqrt{n}\mu_i) dz \\ & \leq 2 \int_{C\xi_n}^{\infty} z P(\sqrt{n}|\bar{X}_i - \mu_i| \geq z - \sqrt{n}\mu_i) dz \end{aligned} \quad (3.7.60)$$

$$\leq 2 \int_{C\xi_n - \sqrt{n}\mu_i}^{\infty} z \exp(-z^d/M) dz + 2\sqrt{n}\mu_i \int_{C\xi_n - \sqrt{n}\mu_i}^{\infty} \exp(-z^d/M) dz \quad (3.7.61)$$

$$\leq M^2 (C\xi_n - \sqrt{n}\mu_i)^{2-2d} \exp(-(C\xi_n - \sqrt{n}\mu_i)^2/M)/d. \quad (3.7.62)$$

If $C\xi_n = (\lambda_n M)^{1/d}$ and $\sqrt{\lambda_n} > \sqrt{n}\mu_i$, then we know

$$\frac{(C\xi_n - \sqrt{n}\mu_i)^{2-2d} \exp(-(C\xi_n - \sqrt{n}\mu_i)^2/M)}{\lambda_n (\sqrt{\lambda_n} - \sqrt{n}\mu_i)^{-1} \exp(-(\sqrt{\lambda_n} - \sqrt{n}\mu_i)^2/2)} \sim \lambda_n^{\frac{2}{d}-\frac{3}{2}} \exp(-\lambda_n/2) \rightarrow 0, \quad (3.7.63)$$

which means that (3.7.57) and (3.7.59) are smaller order than the first term if $\sqrt{\lambda_n} > \sqrt{n}\mu_i$. If $\sqrt{\lambda_n} < \sqrt{n}\mu_i$, the first term in (3.7.52) is even larger order than the case when $\sqrt{\lambda_n} > \sqrt{n}\mu_i$. Hence, (3.7.57) and (3.7.59) are smaller order of the first term in (3.7.52). Therefore,

$$\begin{aligned} E(T_n) &= \left\{ \sum_{i=1}^p \lambda_n \{ P(\sqrt{n}(\bar{X}_i - \mu_i) \geq \eta_{ni}^-) + P(\sqrt{n}(\bar{X}_i - \mu_i) \leq -\eta_{ni}^+) \} \right. \\ &\quad + 2 \sum_{i=1}^p \int_{\sqrt{\lambda_n}}^{C\xi_n} z \{ P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) \\ &\quad \left. + P(\sqrt{n}(\bar{X}_i - \mu_i) \leq -z - \sqrt{n}\mu_i) \} dz \right\} \{1 + o(1)\}. \end{aligned} \quad (3.7.64)$$

It follows that

$$\begin{aligned} & \int_{\sqrt{\lambda_n}}^{C\xi_n} z P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) = \int_{\sqrt{\lambda_n}}^{C\xi_n} z \bar{\Phi}(z - \sqrt{n}\mu_i) dz \\ & + \int_{\sqrt{\lambda_n}}^{C\xi_n} z \{ P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) - \bar{\Phi}(z - \sqrt{n}\mu_i) \} dz \end{aligned} \quad (3.7.65)$$

where (3.7.65) equals to

$$\begin{aligned}
& \left| \int_{\sqrt{\lambda_n}}^{C\xi_n} z \bar{\Phi}(z - \sqrt{n}\mu_i) \frac{P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) - \bar{\Phi}(z - \sqrt{n}\mu_i)}{\bar{\Phi}(z - \sqrt{n}\mu_i)} dz \right| \\
& \leq \int_{\sqrt{\lambda_n}}^{C\xi_n} z \bar{\Phi}(z - \sqrt{n}\mu_i) \left| \frac{P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) - \bar{\Phi}(z - \sqrt{n}\mu_i)}{\bar{\Phi}(z - \sqrt{n}\mu_i)} \right| dz \\
& \leq \sup_{z=o(n^{1/6})} \left| \frac{P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) - \bar{\Phi}(z - \sqrt{n}\mu_i)}{\bar{\Phi}(z - \sqrt{n}\mu_i)} \right| \int_{\sqrt{\lambda_n}}^{C\xi_n} z \bar{\Phi}(z - \sqrt{n}\mu_i) dz \\
& = o(1) \times \int_{\sqrt{\lambda_n}}^{C\xi_n} z \bar{\Phi}(z - \sqrt{n}\mu_i) dz.
\end{aligned}$$

So, we have

$$\int_{\sqrt{\lambda_n}}^{C\xi_n} z P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) = \int_{\sqrt{\lambda_n}}^{C\xi_n} z \bar{\Phi}(z - \sqrt{n}\mu_i) dz \{1 + o(1)\}. \quad (3.7.66)$$

Thus,

$$\begin{aligned}
E(T_n) &= \left[\sum_{i=1}^p \lambda_n (\bar{\Phi}(\eta_{ni}^-) + \bar{\Phi}(\eta_{ni}^+)) + 2 \sum_{i=1}^p \int_{\sqrt{\lambda_n}}^{C\xi_n} z \{\bar{\Phi}(z - \sqrt{n}\mu_i) + \bar{\Phi}(z + \sqrt{n}\mu_i)\} dz \right] \{1 + o(1)\} \\
&= \left[\sum_{i=1}^p \lambda_n (\bar{\Phi}(\eta_{ni}^-) + \bar{\Phi}(\eta_{ni}^+)) + 2 \sum_{i=1}^p \int_{\sqrt{\lambda_n}}^{\infty} z \{\bar{\Phi}(z - \sqrt{n}\mu_i) + \bar{\Phi}(z + \sqrt{n}\mu_i)\} dz \right] \{1 + o(1)\}.
\end{aligned}$$

Equivalently, we can assume $\sqrt{n}\bar{X}_i \sim N(\sqrt{n}\mu_i, 1)$. Then we get

$$\begin{aligned}
E(n\bar{X}_i^2 I\{n\bar{X}_i^2 > \lambda_n\}) &= \left\{ \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\lambda_n}}^{\infty} x^2 \exp\left(-\frac{(x - \sqrt{n}\mu_i)^2}{2}\right) dx \right. \\
&\quad \left. + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{\lambda_n}} x^2 \exp\left(-\frac{(x - \sqrt{n}\mu_i)^2}{2}\right) dx \right\} \{1 + o(1)\} \\
&= \left\{ (\eta_{ni}^+) \phi(\eta_{ni}^-) + (\eta_{ni}^-) \phi(\eta_{ni}^+) + (n\mu_i^2 + 1)(\bar{\Phi}(\eta_{ni}^-) + \bar{\Phi}(\eta_{ni}^+)) \right\} \{1 + o(1)\}.
\end{aligned}$$

Hence the expectation of T_n is

$$E(T_n) = \sum_{i=1}^p \left\{ (\eta_{ni}^+) \phi(\eta_{ni}^-) + (\eta_{ni}^-) \phi(\eta_{ni}^+) + (n\mu_i^2 + 1)(\bar{\Phi}(\eta_{ni}^-) + \bar{\Phi}(\eta_{ni}^+)) \right\} \{1 + o(1)\}. \quad (3.7.67)$$

Next, we want to calculate the variance of the test statistics T_n . The variance of the test statistics is

$$\text{Var}(T_n) = \sum_{i=1}^p \text{Var}(n\bar{X}_i^2 I\{n\bar{X}_i^2 > \lambda_n\}) + \sum_{i \neq j}^p \text{Cov}(n\bar{X}_i^2 I\{n\bar{X}_i^2 > \lambda_n\}, n\bar{X}_j^2 I\{n\bar{X}_j^2 > \lambda_n\}).$$

Each term in the summation of first part in the variance of T_n can be written as

$$\text{Var}(n\bar{X}_i^2 I\{n\bar{X}_i^2 > \lambda_n\}) = E(n^2 \bar{X}_i^4 I\{n\bar{X}_i^2 > \lambda_n\}) - (E(n\bar{X}_i^2 I\{n\bar{X}_i^2 > \lambda_n\}))^2.$$

Let us consider the first term, which can be expressed as

$$E(n^2 \bar{X}_i^4 I\{n \bar{X}_i^2 > \lambda_n\}) = \lambda_n^2 P(\sqrt{n}|\bar{X}_i| \geq \sqrt{\lambda_n}) + 4 \int_{\sqrt{\lambda_n}}^{\infty} z^3 P(\sqrt{n}|\bar{X}_i| \geq z) dz.$$

Similar to the calculation of the expectation for T_n , we want to show that for some $\xi_n = o(n^{1/6})$ and ξ_n is higher order than $\sqrt{n}\mu_i$,

$$\int_{\xi_n}^{\infty} z^3 P(\sqrt{n}|\bar{X}_i| \geq z) dz = o\{\lambda_n^2 P(\sqrt{n}|\bar{X}_i| \geq \sqrt{\lambda_n})\}. \quad (3.7.68)$$

We have

$$\begin{aligned} \int_{\xi_n}^{\infty} z^3 P(\sqrt{n}|\bar{X}_i| \geq z) dz &= \int_{\xi_n}^{\infty} z^3 P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) dz \\ &\quad + \int_{\xi_n}^{\infty} z^3 P(\sqrt{n}(\bar{X}_i - \mu_i) \leq -z - \sqrt{n}\mu_i) dz. \end{aligned}$$

Then followed by the exponential tail assumption,

$$\begin{aligned} \int_{\xi_n}^{\infty} z^3 P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) dz &\leq \int_{\xi_n - \sqrt{n}\mu_i}^{\infty} (z + \sqrt{n}\mu_i)^3 \exp(-z^d/M) dz \\ &= \int_{\xi_n - \sqrt{n}\mu_i}^{\infty} (z^3 + 3z^2\sqrt{n}\mu_i + 3zn\mu_i^2 + n^{3/2}\mu_i^3) \exp(-z^d/M) dz. \end{aligned}$$

It is straightforward to show that

$$\begin{aligned} \int_{\xi_n - \sqrt{n}\mu_i}^{\infty} z^a \exp(-z^d/M) dz &= \frac{1}{d} \int_{(\xi_n - \sqrt{n}\mu_i)^d}^{\infty} u^{\frac{a+1}{d}-1} \exp(-u/M) du \\ &\leq (\xi_n - \sqrt{n}\mu_i)^{(a+1)(1-d)} M^{a+1} \int_{(\xi_n - \sqrt{n}\mu_i)^d/M}^{\infty} u^a \exp(-u) du \\ &= M^2 (\xi_n - \sqrt{n}\mu_i)^{(a+1-2d)} \exp(-(\xi_n - \sqrt{n}\mu_i)^d/M) \{1 + o(1)\}. \end{aligned}$$

From (3.7.54) and (3.7.55), we could take ξ_n to be $(M \max(\lambda_n, n\mu_i^2))^{1/d}$ such that (3.7.68)

holds. It follows that

$$\begin{aligned} E(n^2 \bar{X}_i^4 I\{n \bar{X}_i^2 > \lambda_n\}) &= \left\{ \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\lambda_n}}^{\infty} x^4 \exp\left(-\frac{(x - \sqrt{n}\mu_i)^2}{2}\right) dx \right. \\ &\quad \left. + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{\lambda_n}} x^4 \exp\left(-\frac{(x - \sqrt{n}\mu_i)^2}{2}\right) dx \right\} \{1 + o(1)\}. \end{aligned}$$

By integrate by parts, we have

$$\begin{aligned}
E(n^2 \bar{X}_i^4 I\{n \bar{X}_i^2 > \lambda_n\}) &= \left\{ [(\eta_{ni}^-)^3 + 4\sqrt{n}\mu_i(\eta_{ni}^-)^2 + 3(1 + 2n\mu_i^2)(\eta_{ni}^-) + 4\sqrt{n}\mu_i(2 + n\mu_i^2)]\phi(\eta_{ni}^-) \right. \\
&\quad + [(\eta_{ni}^+)^3 - 4\sqrt{n}\mu_i(\eta_{ni}^+)^2 + 3(1 + 2n\mu_i^2)(\eta_{ni}^+) - 4\sqrt{n}\mu_i(2 + n\mu_i^2)]\phi(\eta_{ni}^+) \\
&\quad \left. + [n^2\mu_i^4 + 6n\mu_i^2 + 3]\bar{\Phi}(\eta_{ni}^-) + [n^2\mu_i^4 + 6n\mu_i^2 + 3]\bar{\Phi}(\eta_{ni}^+) \right\} \{1 + o(1)\} \\
&= \left\{ [(\sqrt{\lambda_n})^3 + (\sqrt{n}\mu_i)^3 + \sqrt{n}\mu_i\lambda_n + n\sqrt{\lambda_n}\mu_i^2 + 5\sqrt{n}\mu_i + 3\sqrt{\lambda_n}]\phi(\eta_{ni}^-) \right. \\
&\quad + [(\sqrt{\lambda_n})^3 - (\sqrt{n}\mu_i)^3 - \sqrt{n}\mu_i\lambda_n + n\sqrt{\lambda_n}\mu_i^2 - 5\sqrt{n}\mu_i + 3\sqrt{\lambda_n}]\phi(\eta_{ni}^+) \\
&\quad \left. + [n^2\mu_i^4 + 6n\mu_i^2 + 3](\bar{\Phi}(\eta_{ni}^-) + \bar{\Phi}(\eta_{ni}^+)) \right\} \{1 + o(1)\}.
\end{aligned}$$

Therefore

$$\begin{aligned}
V(\mu_i, \lambda_n) &= \left\{ [(\sqrt{\lambda_n})^3 + (\sqrt{n}\mu_i)^3 + \sqrt{n}\mu_i\lambda_n + n\sqrt{\lambda_n}\mu_i^2 + 5\sqrt{n}\mu_i + 3\sqrt{\lambda_n}]\phi(\eta_{ni}^-) \right. \\
&\quad + [(\sqrt{\lambda_n})^3 - (\sqrt{n}\mu_i)^3 - \sqrt{n}\mu_i\lambda_n + n\sqrt{\lambda_n}\mu_i^2 - 5\sqrt{n}\mu_i + 3\sqrt{\lambda_n}]\phi(\eta_{ni}^+) \\
&\quad + [n^2\mu_i^4 + 6n\mu_i^2 + 3](\bar{\Phi}(\eta_{ni}^-) + \bar{\Phi}(\eta_{ni}^+)) \\
&\quad \left. - [(\eta_{ni}^+)\phi(\eta_{ni}^-) + (\eta_{ni}^-)\phi(\eta_{ni}^+) + (n\mu_i^2 + 1)(\bar{\Phi}(\eta_{ni}^-) + \bar{\Phi}(\eta_{ni}^+))]^2 \right\} \{1 + o(1)\}. \quad (3.7.69)
\end{aligned}$$

Recall that $Y_{i,n} = n\bar{X}_i^2$ and by Fubini's Theorem, we have

$$E\{Y_{1,n}Y_{2,n}I(Y_{1,n} > \lambda_n)I(Y_{2,n} > \lambda_n)\} = \lambda_n^2 P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n) \quad (3.7.70)$$

$$+ \lambda_n \int_{\lambda_n}^{\infty} P(Y_{1,n} > z_1, Y_{2,n} > \lambda_n) dz_1 \quad (3.7.71)$$

$$+ \lambda_n \int_{\lambda_n}^{\infty} P(Y_{1,n} > \lambda_n, Y_{2,n} > z_2) dz_2 \quad (3.7.72)$$

$$+ \int_{\lambda_n}^{\infty} \int_{\lambda_n}^{\infty} P(Y_{1,n} > z_1, Y_{2,n} > z_2) dz_1 dz_2. \quad (3.7.73)$$

For the purpose of evaluation of (3.7.70), (3.7.71) (3.7.72) and (3.7.73), we need use Lemma 1 regarding the large deviation results for bivariate random vectors. Based on this lemma, we could approximate (3.7.70) by assuming $\sqrt{n}(\bar{X}_{1,n}, \bar{X}_{2,n})'$ is a bivariate normally distributed

vector. Without loss of generality, suppose $\rho > 0$. Because

$$\begin{aligned} P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n) &= P(\sqrt{n}(\bar{X}_{1,n} - \mu_1) > \eta_{n1}^-, \sqrt{n}(\bar{X}_{2,n} - \mu_2) > \eta_{n2}^-) \\ &\quad + P(\sqrt{n}(\bar{X}_{1,n}^* + \mu_1) > \eta_{n1}^+, \sqrt{n}(\bar{X}_{2,n} - \mu_2) > \eta_{n2}^-) \\ &\quad + P(\sqrt{n}(\bar{X}_{1,n} - \mu_1) > \eta_{n1}^-, \sqrt{n}(\bar{X}_{2,n}^* + \mu_2) > \eta_{n2}^+) \\ &\quad + P(\sqrt{n}(\bar{X}_{1,n}^* + \mu_1) > \eta_{n1}^+, \sqrt{n}(\bar{X}_{2,n}^* + \mu_2) > \eta_{n2}^+) \end{aligned}$$

and we observe that

$$\begin{aligned} P(\sqrt{n}(\bar{X}_{1,n} - \mu_1) > \eta_{n1}^-, \sqrt{n}(\bar{X}_{2,n} - \mu_2) > \eta_{n2}^-) &\geq P(\sqrt{n}(\bar{X}_{1,n} - \mu_1) > \eta_{n1}^+, \sqrt{n}(\bar{X}_{2,n} - \mu_2) > \eta_{n2}^-) \\ &\geq \bar{\Phi}(\eta_{n1}^+) \bar{\Phi}\left(\frac{-\rho\eta_{n1}^+ + \eta_{n2}^-}{\sqrt{1-\rho^2}}\right) \geq \bar{\Phi}(\eta_{n1}^+) \bar{\Phi}\left(\frac{\rho\eta_{n1}^+ + \eta_{n2}^-}{\sqrt{1-\rho^2}}\right) \\ &\geq P(\sqrt{n}(\bar{X}_{1,n}^* + \mu_1) > \eta_{n1}^+, \sqrt{n}(\bar{X}_{2,n} - \mu_2) > \eta_{n2}^-), \end{aligned}$$

it is easy to see that

$$P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n) \sim P(\sqrt{n}(\bar{X}_{1,n} - \mu_1) > \eta_{n1}^-, \sqrt{n}(\bar{X}_{2,n} - \mu_2) > \eta_{n2}^-).$$

Assume $\eta_{n1}^- > 0$ or $\eta_{n2}^- > 0$. Without loss of generality, let $\eta_{n2}^- > 0$. It can be seen that

$$P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n) \geq \bar{\Phi}(\eta_{n2}^-) \bar{\Phi}\left(\frac{-\rho\eta_{n2}^- + \eta_{n1}^-}{\sqrt{1-\rho^2}}\right).$$

When $\eta_{n1}^- > 0$ such that $\eta_{n1}^- - \rho\eta_{n2}^- > 0$,

$$\begin{aligned} P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n) &\geq \frac{\sqrt{1-\rho^2}}{\eta_{n2}^-(\eta_{n1}^- - \rho\eta_{n2}^-)} \exp\left\{-\frac{\eta_{n2}^{-2} + \eta_{n1}^{-2} - 2\rho\eta_{n2}^-\eta_{n1}^-}{2(1-\rho^2)}\right\} \\ &\sim C_1 \lambda_n^{-1} \exp\{-C_2 \lambda_n\}. \end{aligned}$$

Otherwise, $P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n) \geq \bar{\Phi}(\eta_{n2}^-) \left(1 - \bar{\Phi}\left(\frac{\rho\eta_{n2}^- - \eta_{n1}^-}{\sqrt{1-\rho^2}}\right)\right) \sim C_3 \lambda_n^{-1/2} \exp\{-C_4 \sqrt{\lambda_n}\}$.

Let $\xi_n = o(n^{1/6})$ such that ξ_n is higher order than $\sqrt{n}\mu_i$

$$\begin{aligned} &\lambda_n \int_{\xi_n}^{\infty} P(Y_{1,n} > z_1, Y_{2,n} > \lambda_n) dz_1 \\ &\leq 2\lambda_n \int_{\xi_n}^{\infty} z_1 P(\sqrt{n}|\bar{X}_{1,n} - \mu_1| > z_1 - \sqrt{n}\mu_1, \sqrt{n}|\bar{X}_{2,n} - \mu_2| > \eta_{n2}^-) dz_1 \\ &\leq 2\lambda_n \int_{\xi_n}^{\infty} z_1 P(\sqrt{n}|\bar{X}_{1,n} - \mu_1| > z_1 - \sqrt{n}\mu_1) dz_1. \end{aligned}$$

Thus, like (3.7.63), we can find ξ_n such that $\lambda_n \int_{\xi_n}^{\infty} P(Y_{1,n} > z_1, Y_{2,n} > \lambda_n) dz_1$ is a higher order of $\lambda_n^2 P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n)$. If $\eta_{n1}^- \leq 0$ and $\eta_{n2}^- \leq 0$, then it is easy to see that $\lambda_n \int_{\xi_n}^{\infty} P(Y_{1,n} > z_1, Y_{2,n} > \lambda_n) dz_1$ is a higher order of $\lambda_n^2 P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n)$. Similarly, $\lambda_n \int_{\xi_n}^{\infty} P(Y_{1,n} > \lambda_n, Y_{2,n} > z_2) dz_2$ are smaller order of $\lambda_n^2 P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n)$. We may also show that

$$\int_{\xi_n}^{\infty} \int_{\xi_n}^{\infty} P(Y_{1,n} > z_1, Y_{2,n} > z_2) dz_1 dz_2$$

is a smaller order than $\lambda_n^2 P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n)$. So we can approximate $E\{Y_{1,n} Y_{2,n} I(Y_{1,n} > \lambda_n) I(Y_{2,n} > \lambda_n)\}$ by assuming normal assumption $X_i = (X_{i1}, X_{i2})' \sim N(\sqrt{n}(\mu_1, \mu_2)', \Sigma)$.

Let $X^* = -X$ and $Y^* = -Y$. Then

$$\begin{aligned} E\{Y_{1,n} Y_{2,n} I(Y_{1,n} > \lambda_n) I(Y_{2,n} > \lambda_n)\} &= E\{Y_{1,n} Y_{2,n} I(\sqrt{n} \bar{X}_{1,n} > \sqrt{\lambda_n}) I(\sqrt{n} \bar{X}_{2,n} > \sqrt{\lambda_n})\} \\ &\quad + E\{Y_{1,n} Y_{2,n}^* I(\sqrt{n} \bar{X}_{1,n} > \sqrt{\lambda_n}) I(\sqrt{n} \bar{X}_{2,n}^* > \sqrt{\lambda_n})\} \\ &\quad + E\{Y_{1,n}^* Y_{2,n} I(\sqrt{n} \bar{X}_{1,n}^* > \sqrt{\lambda_n}) I(\sqrt{n} \bar{X}_{2,n} > \sqrt{\lambda_n})\} \\ &\quad + E\{Y_{1,n}^* Y_{2,n}^* I(\sqrt{n} \bar{X}_{1,n}^* > \sqrt{\lambda_n}) I(\sqrt{n} \bar{X}_{2,n}^* > \sqrt{\lambda_n})\}. \end{aligned}$$

Applying the formula given in Lemma 2, we can approximate above expectations by (3.7.24).

It follows that

$$\begin{aligned} \gamma_1(\mu_1, \mu_2, \lambda_n) &:= Cov\{Y_{1,n} I(Y_{1,n} > \lambda_n), Y_{2,n} I(Y_{2,n} > \lambda_n)\} \\ &= E\{Y_{1,n} Y_{2,n} I(Y_{1,n} > \lambda_n) I(Y_{2,n} > \lambda_n)\} \\ &\quad - \left\{ (\eta_{n1}^+) \phi(\eta_{n1}^-) + (\eta_{n1}^-) \phi(\eta_{n1}^+) + (n\mu_1^2 + 1)(\bar{\Phi}(\eta_{n1}^-) + \bar{\Phi}(\eta_{n1}^+)) \right\} \\ &\quad \times \left\{ (\eta_{n2}^+) \phi(\eta_{n2}^-) + (\eta_{n2}^-) \phi(\eta_{n2}^+) + (n\mu_2^2 + 1)(\bar{\Phi}(\eta_{n2}^-) + \bar{\Phi}(\eta_{n2}^+)) \right\}. \quad (3.7.74) \end{aligned}$$

So the variance of T_n is $\text{Var}(T_n) = \sum_{i=1}^p V(\mu_i, \lambda_n) + 2 \sum_{i=1}^p \sum_{j=i+1}^p \gamma_{j-i}(\sqrt{n}\mu_i, \sqrt{n}\mu_j, \lambda_n)$.

This completes the proof of Theorem 1. \square

The following lemma is from Kim (1994), which will be useful for the proofs of Theorem 2 and 3.

Lemma 3 *Suppose $\{Z_i\}_{i=1}^p$ is a sequence of dependent zero mean random variables satisfying $M_{2r+\delta} = \sup_i \|Z_i\|_{2r+\delta} < \infty$ for some $\delta > 0$ and $r \geq 1$. If*

$$\sum_{i=1}^{\infty} i^{r-1} \alpha(i)^{\delta/(2r+\delta)} < \infty,$$

then

$$E\left(\sum_{i=1}^p Z_i\right)^{2r} \leq Cp^r \left[M_{2r}^{2r} + M_{2r+\delta}^{2r} \sum_{i=1}^{\infty} i^{r-1} \alpha(i)^{\delta/(2r+\delta)} \right] \quad (3.7.75)$$

where C is a finite constant only depending on r .

Proof of Theorem 2 (i) Let $k' \rightarrow \infty$, $k'/k \rightarrow 0$ and $k/p \rightarrow 0$ as $n \rightarrow \infty$. Define $b = k + k'$. Set integer r such that $rb \leq p \leq (r+1)b$.

$$\tilde{T}_{i,n} = Z_{i,n}(\lambda_n) - \mu_{T_n,0}^{(i)}.$$

We will use Bernstein's blocking method to show the central limit theorem. For this purpose, define large blocks

$$\tilde{\zeta}_{j,n} = \sigma_0^{-1}(p; \lambda_n) \sum_{i=1}^k \tilde{T}_{(j-1)b+i,n}$$

and small blocks

$$\tilde{\zeta}'_{j,n} = \sigma_0^{-1}(p; \lambda_n) \sum_{i=1}^{k'} \tilde{T}_{(j-1)b+k+i,n}$$

for $j = 1, \dots, r$. Also the residual block $\delta_n = \sigma_0^{-1}(p; \lambda_n) \sum_{i=rb+1}^p \tilde{T}_{i,n}$. Then

$$p^{-1/2} S_n = p^{-1/2} \sum_{j=1}^r \tilde{\zeta}_{j,n} + p^{-1/2} \sum_{j=1}^r \tilde{\zeta}'_{j,n} + p^{-1/2} \delta_n := S_{n,1} + S_{n,2} + S_{n,3}.$$

It follows that $E(S_{n,1}) = E(S_{n,2}) = E(S_{n,3}) = 0$.

By the Davydov's inequality (Bosq, 1998), for some $q > 2$

$$|\rho_k| \leq \frac{2q}{q-2} (2\alpha_X(k))^{1-2/q} \|X_{ij}\|_q^2 = C\alpha_X^{1-2/q}(k)$$

and from (3.7.37), $|\gamma_k(0, 0, \lambda_n)| = C|\rho_k| L_p p^{-\frac{2s}{1+|\rho_k|}} \leq C\alpha_X^{1-2/q}(k) p^{-s}$. Since $\sum_{k=1}^{\infty} k\alpha_X^{\delta/(4+\delta)}(k) < \infty$, taking $q > 2 + \delta/2$, then $\sum_{k=1}^{\infty} \alpha_X^{1-2/q}(k) < \sum_{k=1}^{\infty} \alpha_X^{\delta/(4+\delta)}(k) < \infty$. It follows that $\sum_{k=1}^{\infty} |\gamma_k(0, 0, \lambda_n)| = O(p^{-s})$.

Because for $j \geq 1$,

$$\begin{aligned} |Cov(\tilde{\zeta}'_{1,n}, \tilde{\zeta}'_{j+1,n})| &= \left| \sum_{i=1}^{k'} \sum_{i'=1}^{k'} Cov(Z_{k+i,n}(\lambda_n), Z_{j+1+k+i',n}(\lambda_n)) \right| \\ &= \left| \sum_{i=1}^{k'} \sum_{i'=1}^{k'} \gamma_{j+1+(i-i')}(0, 0, \lambda_n) \right| \leq \sum_{i=1}^{k'} |\gamma_{j+1+(i-k')}(0, 0, \lambda_n)|, \end{aligned}$$

we have

$$\begin{aligned}
\text{Var}(S_{n,2}) &= p^{-1}\sigma_0^{-2}(p; \lambda_n) \left\{ rk' \sigma_0^2(k'; \lambda_n) + 2 \sum_{j=1}^{r-1} (r-j) \text{Cov}(\tilde{\zeta}'_{1,n}, \tilde{\zeta}'_{j+1,n}) \right\} \\
&\leq p^{-1}\sigma_0^{-2}(p; \lambda_n) \left\{ rk' \sigma_0^2(k'; \lambda_n) + 2 \sum_{j=1}^{r-1} (r-j) \sum_{i=1}^{k'} |\gamma_{jb+(i-k')}(0, 0, \lambda_n)| \right\} \\
&\leq p^{-1}\sigma_0^{-2}(p; \lambda_n) \left\{ rk' \sigma_0^2(k'; \lambda_n) + 2r \sum_{k=1}^{\infty} |\gamma_k(0, 0, \lambda_n)| \right\} = O(rk'/p) \rightarrow 0,
\end{aligned}$$

where we use $\sigma_0^2(p; \lambda_n) = \sigma_0^2(k'; \lambda_n) = O(L_p p^{-s})$ and $\sum_{k=1}^{\infty} |\gamma_k(0, 0, \lambda_n)| = O(p^{-s})$. Similarly, we have $\text{Var}(S_{n,3}) \rightarrow 0$. Hence,

$$p^{-1/2} S_n = p^{-1/2} \sum_{j=1}^r \tilde{\zeta}_{j,n} + o_p(1). \quad (3.7.76)$$

By Bradley's lemma, there exist independent random variables $W_{j,n}$ such that $W_{j,n}$ and $\tilde{\zeta}_{j,n}$ are identically distributed and for any $\epsilon > 0$,

$$P(|\tilde{\zeta}_{j,n} - W_{j,n}| \leq \epsilon p^{1/2}/r) \leq 11(\epsilon p^{1/2}/r)^{-2/5} (E(\tilde{\zeta}^2))^{1/5} \alpha_Z^{4/5}(k') \leq C \epsilon^{-2/5} r^{1/5} \alpha_Z^{4/5}(k').$$

Let $\Delta_n = S_{n,1} - p^{-1/2} \sum_{j=1}^r W_{j,n}$. Then

$$P(|\Delta_n| > \epsilon) \leq \sum_{j=1}^r P(|\tilde{\zeta}_{j,n} - W_{j,n}| \leq \epsilon p^{1/2}/r) \leq C_1 \epsilon^{-2/5} r^{6/5} \alpha_Z^{4/5}(k').$$

Choosing $r = p^a$ for $a \in (0, 1)$, $b = p^{1-a}$ and $k' = p^c$ for $c \in (0, 1-a)$. Because

$$p^{6a/5} \alpha_Z^{4/5}(p^c) \leq p^{6a/5} n^{4/5} \alpha_X^{4/5}(p^c) \rightarrow 0, \text{ as } p \rightarrow \infty,$$

then $\Delta_n = o_p(1)$. It follows that

$$\begin{aligned}
p^{-1/2} S_n &= p^{-1/2} \sum_{j=1}^r W_{j,n} + o_p(1) \\
&= \left(\frac{rk}{p}\right)^{1/2} \frac{\sigma_0(k; \lambda_n)}{\sigma_0(p; \lambda_n)} (rk)^{-1/2} \sum_{j=1}^r W_{j,n}^* + o_p(1),
\end{aligned}$$

where $W_{j,n}^* = \frac{\sigma_0(p; \lambda_n)}{\sigma_0(k; \lambda_n)} W_{j,n}$ and $W_{j,n}^*$ are independent random variables with mean 0 and variance k . We also know that $\left(\frac{rk}{p}\right)^{1/2} \frac{\sigma_0(k; \lambda_n)}{\sigma_0(p; \lambda_n)} \rightarrow 1$. Therefore, we only need to show

$$(rk)^{-1/2} \sum_{j=1}^r W_{j,n}^* \xrightarrow{d} N(0, 1).$$

It suffices to show the Lyapounov's condition, which is

$$\lim_{r \rightarrow \infty} r^{-2} \sum_{j=1}^r E|k^{-1/2}W_{j,n}^*|^4 = 0. \quad (3.7.77)$$

By Lemma 3, we can get that

$$\begin{aligned} E|k^{-1/2}W_{j,n}^*|^4 &= k^{-2}\sigma_0^{-4}(k; \lambda_n) E \left\{ \sum_{i=1}^k (Z_{i,n}(\lambda_n) - \mu_{T_n,0}^{(i)}) \right\}^4 \\ &\leq C\sigma_0^{-4}(k; \lambda_n) \left[M_4^4 + M_{4+\delta}^4 \sum_{i=1}^{\infty} i\alpha_Z(i)^{\delta/(4+\delta)} \right]. \end{aligned}$$

where $M_4^4 = E(Z_{i,n}(\lambda_n) - \mu_{T_n,0}^{(i)})^4$. From (3.3.14), we know that $\sigma_0^2(k; \lambda_n) \sim L_p p^{-s}$. We can also show that $M_4^4 = O\{(\log p)^{7/2} p^{-s}\}$ and $M_{4+\delta}^4 = O\{(\log p)^{\frac{4\delta+14}{4+\delta}} p^{-\frac{4s}{4+\delta}}\}$. Because

$$\sum_{i=1}^{\infty} i\alpha_Z(i)^{\delta/(4+\delta)} \leq n^{\delta/(4+\delta)} \sum_{i=1}^{\infty} i\alpha_X(i)^{\delta/(4+\delta)} \leq Cn^{\delta/(4+\delta)},$$

we have $E|k^{-1/2}W_{j,n}^*|^4 \leq L_p n^{\delta/(4+\delta)} p^{(4+2\delta)s/(4+\delta)}$. If we take $r = p^a$ with $a > (4 + 2\delta)s/(4 + \delta)$ such that $n^{\delta/(4+\delta)} p^{(4+2\delta)s/(4+\delta)-a} \rightarrow 0$, then (3.7.77) holds. Hence, the central limit theorem holds for $p^{-1/2}S_{n,1}$ and by Slutsky's theorem, it is also hold for $p^{-1/2}S_n$. The proof of (i) is completed.

(ii) The proof of (ii) is similar to the proof of (i). We can also define $\tilde{\zeta}_{j,n}, \tilde{\zeta}'_{j,n}$ and δ_n by replacing $\sigma_0(p; \lambda_n)$ with $\sigma_1(p; \lambda_n)$, replacing $\mu_{T_n,0}^{(i)}$ with $\mu_{T_n,1}^{(i)}$ in the proof of (i). Similarly, we can define $S_{n,1}, S_{n,2}$ and $S_{n,3}$. Notice that, by Davydov's inequality, for any μ_1 and μ_{k+1} satisfy the conditions in Theorem 1,

$$|\gamma_k(\mu_1, \mu_{k+1}, \lambda_n)| \leq CL_p |\rho_k| \leq CL_p \alpha_X^{1-2/q}(k). \quad (3.7.78)$$

Hence, if $\sum_{k=1}^{\infty} k\alpha_X^{\delta/(4+\delta)}(k) < \infty$, then it is similar as part (i) to show that $p^{-1/2}S_{n,2}$ and $p^{-1/2}S_{n,3}$ are $o_p(1)$. Then by applying the Bradley's lemma, there exist independent random variables $\tilde{W}_{j,n}$ such that

$$\tilde{W}_{j,n} \stackrel{d}{=} \tilde{\zeta}_{j,n} := \sigma_1^{-1}(p; \lambda_n) \sum_{i=1}^k \{Z_{i,n}(\lambda_n) - \mu_{T_n,1}^{(i)}\}$$

and

$$p^{-1/2}S_n = p^{-1/2} \sum_{j=1}^r \tilde{W}_{j,n} + o_p(1).$$

We are left to show the Lyapounov's condition:

$$\left\{ \sum_{j=1}^r E(\tilde{W}_{j,n}^2) \right\}^{-2} \sum_{j=1}^r E(\tilde{W}_{j,n}^4) \rightarrow 0. \quad (3.7.79)$$

Note that by condition (3.3.13),

$$\begin{aligned} \left| \frac{1}{p} \sum_{j=1}^r E(\tilde{W}_{j,n}^2) - 1 \right| &= \left| \frac{1}{p} \sum_{j=1}^r E(\tilde{\zeta}_{j,n}^2) - 1 \right| \leq \frac{1}{p} \sum_{j=1}^r |E(\tilde{\zeta}_{j,n}^2) - k| + p^{-1}|rk - p| \\ &\leq rkH_k/p + p^{-1}|rk - p| \rightarrow 0. \end{aligned}$$

Now we will use Lemma 3 to show that

$$\frac{1}{p^2} \sum_{j=1}^r E(\tilde{W}_{j,n}^4) = \frac{(rk)^2}{p^2} \frac{1}{r^2} \sum_{j=1}^r E\{(k^{-1/2}\tilde{W}_{j,n})^4\} \rightarrow 0. \quad (3.7.80)$$

Similar to part (i),

$$\begin{aligned} E|k^{-1/2}\tilde{W}_{j,n}|^4 &= k^{-2}\sigma_1^{-4}(k; \lambda_n) E \left\{ \sum_{i=1}^k (Z_{i,n}(\lambda_n) - \mu_{T_n,1}^{(i)}) \right\}^4 \\ &\leq C\sigma_1^{-4}(k; \lambda_n) \left[M_4^4 + M_{4+\delta}^4 \sum_{i=1}^{\infty} i\alpha_Z(i)^{\delta/(4+\delta)} \right]. \end{aligned}$$

where $M_4^4 = \max_i E(Z_{i,n}(\lambda_n) - \mu_{T_n,1}^{(i)})^4$ and $M_{4+\delta}^4 = \max_i \{E(Z_{i,n}(\lambda_n) - \mu_{T_n,1}^{(i)})^{4+\delta}\}^{\frac{4}{4+\delta}}$.

Suppose the mean of the i -th component of X is $\sqrt{2r_i \log p/n}$. Similar to Theorem 1, we can approximate the moments by assuming normality assumption. It follows that

$$E\{Z_{i,n}(\lambda_n) - \mu_{T_n,1}^{(i)}\}^d = \sum_{i=0}^{2d} \binom{2d}{i} (\sqrt{2r_i \log p})^i \frac{1}{\sqrt{2\pi}} \int_{2(\sqrt{s}-\sqrt{r_i})\sqrt{\log p}}^{\infty} y^{2d-i} \exp(-y^2/2) dy. \quad (3.7.81)$$

When i is an even number,

$$\begin{aligned} \int_{2(\sqrt{s}-\sqrt{r_i})\sqrt{\log p}}^{\infty} y^{2d-i} \exp(-y^2/2) dy &= \{2(\sqrt{s}-\sqrt{r_i})\sqrt{\log p}\}^{2d-i-1} \phi(2(\sqrt{s}-\sqrt{r_i})\sqrt{\log p}) \{1 + o(1)\} \\ &\quad + \Phi(2(\sqrt{s}-\sqrt{r_i})\sqrt{\log p}). \end{aligned}$$

When i is an odd number,

$$\int_{2(\sqrt{s}-\sqrt{r_i})\sqrt{\log p}}^{\infty} y^{2d-i} \exp(-y^2/2) dy = \{2(\sqrt{s}-\sqrt{r_i})\sqrt{\log p}\}^{2d-i-1} \phi(2(\sqrt{s}-\sqrt{r_i})\sqrt{\log p}) \{1 + o(1)\}.$$

Therefore, $E\{Z_{i,n}(\lambda_n) - \mu_{T_n,1}^{(i)}\}^d = L_p^{(1)}I(r_i > s) + L_p^{(2)}p^{-(\sqrt{s}-\sqrt{r_i})^2}I(r_i < s)$. Then it can be seen that $M_4^4 = L_p^{(1)}I(\max_i r_i > s) + L_p^{(2)}p^{-(\sqrt{s}-\max_i \sqrt{r_i})^2}I(\max_i r_i < s)$. Suppose $\sigma_1(p; \lambda_n) \sim L_p p^{-h_1}$. If $\max_i r_i > s$, then $E|k^{-1/2}\tilde{W}_{j,n}|^4 \leq L_p n^{\delta/(4+\delta)} p^{2h_1}$. Now if we take $r = p^a$ with $a > 2h_1$ such that $n^{\delta/(4+\delta)} p^{2h_1-a} \rightarrow 0$, then the Lyapounov's condition holds.

If $\max_i r_i < s$, then $E|k^{-1/2}\tilde{W}_{j,n}|^4 \leq L_p n^{\delta/(4+\delta)} p^{2h_1 - \frac{4s^*}{4+\delta}}$ where $s^* = (\sqrt{s} - \max_i \sqrt{r_i})^2$. In this case, if we take $a > 2h_1 - \frac{4s^*}{4+\delta}$ such that $n^{\delta/(4+\delta)} p^{2h_1 - \frac{4s^*}{4+\delta} - a} \rightarrow 0$, the Lyapounov's condition holds. \square

Proof of Theorem 3 Let $\underline{\lambda}_n = (\lambda_{n1}, \dots, \lambda_{nd})' = 2 \log(p)(s_1, \dots, s_d)'$. We want to show that $(\mathcal{T}_S(\lambda_{n1}), \dots, \mathcal{T}_S(\lambda_{nd}))$ are multivariate normally distributed. Applying the Cramér-wold device, we only need to show that for any c_k such that $\sum_k c_k^2 = 1$, $\mathcal{T}_S^* := \sum_{k=1}^d c_k \mathcal{T}_S(\lambda_{nk})$ is normally distributed. Let

$$\mathcal{T}_{S_{i,n}}^* = \sum_{k=1}^d c_k \frac{Z_{i,n}(\lambda_{nk}) - \mu_{T_n,0}^{(i)}(\lambda_{nk})}{\sigma_0(p; \lambda_{nk})}.$$

By the definition of $\mathcal{T}_S(\lambda_{nk})$, we have $\mathcal{T}_S^* = \frac{1}{\sqrt{p}} \sum_{i=1}^p \mathcal{T}_{S_{i,n}}^*$. Let $k^{*'} \rightarrow \infty, k^{*'} / k^* \rightarrow 0$ and $k^* / p \rightarrow 0$ as $n \rightarrow \infty$. As in the proof of Theorem 2, define $\tilde{\zeta}_{j,n}^* = \sum_{i=1}^{k^*} \mathcal{T}_{S_{(j-1)b+i,n}}^*$, $\tilde{\zeta}_{j,n}^{*' } = \sum_{i=1}^{k^{*' }} \mathcal{T}_{S_{(j-1)b+i,n}}^*$ and δ_n^* to be the large, small and residual blocks, and their corresponding partial sums $S_{n,1} = p^{-1/2} \sum_{j=1}^r \tilde{\zeta}_{j,n}^*$, $S_{n,2} = p^{-1/2} \sum_{j=1}^r \tilde{\zeta}_{j,n}^{*' }$ and $S_{n,3}$.

Using the expression (3.4.18), it can be shown that, if $s_k > s_{k'}$,

$$\begin{aligned} \frac{\tilde{\gamma}_k(\sqrt{\lambda_{nk}}, \sqrt{\lambda_{nk'}}) - G_{T_n}(\lambda_{nk})G_{T_n}(\lambda_{nk'})}{\sigma_0(p; \lambda_{nk})\sigma_0(p; \lambda_{nk'})} &= L_p \rho_j p^{-\frac{s_k - s_{k'}}{2}} I(s_{k'} \leq \rho_j^2 s_k) \\ &\quad + L_p \rho_j p^{-\frac{\rho_j^2}{1-\rho_j^2}(\sqrt{s_{k'}} - \sqrt{s_k})^2} I(s_{k'} > \rho_j^2 s_k). \end{aligned}$$

and $2[(\sqrt{\lambda_n^*})^3 + 3\sqrt{\lambda_n^*}]\phi(\sqrt{\lambda_n^*}) = L_p p^{-\frac{s_k - s_{k'}}{2}} \sigma_0(p; \lambda_{nk})\sigma_0(p; \lambda_{nk'})$ where $\lambda_n^* = \max(\lambda_{nk}, \lambda_{nk'})$.

Thus, if $\sum_k |\rho_k| < \infty$ and $\lambda_{nk} \neq \lambda_{nk'}$, $\text{Cov}(\mathcal{T}_S(\lambda_{nk}), \mathcal{T}_S(\lambda_{nk'})) = o(1)$. It follows that $\text{Var}(\mathcal{T}_S^*) =$

$(\sum_{k=1}^d c_k^2)\{1 + o(1)\} = 1 + o(1)$. We can also show that $\text{Var}(\tilde{\zeta}_{j,n}^{*' }) = k^{*' }\{1 + o(1)\}$ and

$$\begin{aligned} \text{Cov}(\mathcal{T}_{S_{1,n}}^*, \mathcal{T}_{S_{j+1,n}}^*) &= \sum_{k=1}^d \sum_{k'=1}^d c_k c_{k'} \left\{ L_p \rho_j p^{\frac{s_{k'} - s_k}{2}} I(s_{k'} \leq \rho_j^2 s_k) \right. \\ &\quad \left. + L_p \rho_j p^{-\frac{\rho_j^2}{1-\rho_j^2}(\sqrt{s_{k'}} - \sqrt{s_k})^2} I(s_{k'} > \rho_j^2 s_k) \right\}. \end{aligned}$$

By the same arguments in the proof of Theorem 2, we obtain

$$\text{Var}(S_{n,2}) \leq p^{-1} \{r \text{Var}(\tilde{\zeta}_{j,n}^{*'}) + 2r \sum_{k=1}^{\infty} |\text{Cov}(\mathcal{T}_{S_{1,n}^*}, \mathcal{T}_{S_{j+1,n}^*})|\} = O(rk^{*'} / p) \rightarrow 0.$$

Similarly, we could show that $\text{Var}(S_{n,3}) = o(1)$. Therefore, $\mathcal{T}_S^* = S_{n,1} + o_p(1)$. Following the proof of Theorem 2, we only need to verify the Lyapounov's condition:

$$\lim_{r \rightarrow \infty} r^{-2} \sum_{j=1}^r E |k^{*-\frac{1}{2}} \tilde{\zeta}_{j,n}^*|^4 = 0.$$

Again, by Lemma 3, $E |k^* \tilde{\zeta}_{j,n}^*|^4 \leq C[M_4^4 + M_{4+\delta}^4 \sum_{i=1}^{\infty} i \alpha_Z(i)^{\delta/(4+\delta)}]$ where

$$M_4^4 = E(\mathcal{T}_{S_{i,n}^*})^4 \leq d^3 \sum_{k=1}^d E \left(c_k \frac{Z_{i,n}(\lambda_{nk}) - \mu_{T_n,0}^{(i)}(\lambda_{nk})}{\sigma_0(p; \lambda_{nk})} \right)^4 = L_p d^3 \sum_{k=1}^d c_k^4 p^{s_k} = O(L_p p^{\max_k s_k}).$$

and similarly, we have $M_{4+\delta}^4 = O(L_p p^{(4+2\delta) \max_k s_k / (4+\delta)})$. It follows that

$$E |k^{*-\frac{1}{2}} \tilde{\zeta}_{j,n}^*|^4 \leq L_p n^{\delta/(4+\delta)} p^{(4+2\delta) \max_k s_k / (4+\delta)}.$$

If we take $r = p^a$ with $a > (4+2\delta) \max_k s_k / (4+\delta)$ such that $n^{\delta/(4+\delta)} p^{(4+2\delta) \max_k s_k / (4+\delta) - a} \rightarrow 0$, then the Lyapounov's condition holds. If all $\mathcal{T}_S(\lambda_{nk})$ are normally distributed, by Theorem 2(i), the above condition holds. Thus, \mathcal{T}_S^* is normally distributed with mean 0 and variance 1. Hence, the $(\mathcal{T}_S(\lambda_{n1}), \dots, \mathcal{T}_S(\lambda_{nd}))$ is normally distributed with mean 0 and covariance $(\Omega(\lambda_{ns}, \lambda_{nt}))_{st}$. \square

References

- Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6, 311-329.
- 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*, 57, 289-300.
- Bhattacharya, R. (1968). Berry-Esseen bounds for the multi-dimensional central limit theorem. *Bulletin of American Mathematical Society*, 74, 285-287.
- Bosq, D. (1998). *Nonparametric statistics for stochastic processes*. Springer, New York.
- Bradley, R. (2005). Basic properties of strong mixing conditions: A survey and some open questions. *Probability Surveys*, 2, 107-144.
- Brillinger, D. R. (1981). *Time series analysis: Data analysis and theory (2nd ed)*. Holt, Rinehart & Winston, New York.
- Brockwell, P. and Davis, R. (2009). *Time series: Theory and methods*. Springer, New York.
- Chen, S. X. and Qin, Y. (2010). A two sample test for high dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38, 808-835.
- Chen, S. X. and Tang, C. Y. (2005). Nonparametric inference of value at risk for dependent financial returns. *Journal of Financial Econometrics*, 3, 227-255.
- Chen, S. X. and Xu, M. (2011). Manuscript.
- Chen, S. X., Zhang, L.X. and Zhong, P.S. (2010). Tests for high dimensional covariance matrices. *Journal of the American Statistical Association*, 109, 810-819.

- Delaigle, A., Hall, P. and Jin, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Students t-statistic. *Journal of Royal Statistical Society, Series B*, 73, 283-301.
- Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81, 425-455.
- Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, 32, 962-994.
- Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association*, 91, 674-688.
- Fan, J. and Lin, S. (1998). Test of significance when data are curves. *Journal of the American Statistical Association*, 93, 1007-1021.
- Fan, J. and Hall, P. and Yao, Q. (2007). To how many simultaneous hypothesis tests can normal student's t or bootstrap calibrations be applied. *Journal of the American Statistical Association*, 102, 1282-1288.
- Fan, J. and Yao, Q. (2005). *Nonlinear time series: Nonparametric and parametric methods*. Springer, New York.
- Fuller, W. (1996). *Introduction to statistical time series (second edition)*. Wiley and sons, Hoboken, NJ.
- Goeman, J., Sara A. van de G. and van Houwelingen, J. C. (2006). Testing against a high dimensional alternative. *Journal of Royal Statistical Society, Series B*, 68, 477-493.
- Hall, P. and Jin, J. (2008). Properties of higher criticism under strong dependence. *The Annals of Statistics*, 38, 1686-1732.
- Huang, J., Ma, S. and Zhang C.H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica*, 18, 1603-1618.

- Kim, T.Y. (1994). Moment bounds for non-stationary dependent sequences. *Journal of Applied Probability*, 31, 731-742.
- Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compare to the sample size. *Annals of Statistics*, 30, 1081-1102.
- Sibuya, M. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11, 195-210.
- Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73, 751-754.
- Shao, Q. M. (1997). Self-normalized large deviations. *Annals of Probability*, 25, 285-328.
- Storey, J. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, 64, 479-498.
- Storey, J. (2003). The positive false discovery rate: A Bayesian interpretation and q-value. *The Annals of Statistics*, 31, 2013-2035.
- Petrov, V. V. (1995). *Limit theorems of probability theory: sequences of independent random variables*. Clarendon Press, London.
- Wood, T. A. and Chan, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics*. 3, 409-432.
- Wang, Q. and Hall, P. (2009). Relative errors in central limit theorems for Student's t statistic, with application. *Statistica Sinica*. 19, 343-354.
- Willink, R. (2005). Bounds on the bivariate normal distribution function. *Communications in Statistics-Theory and Methods*, 33, 2281-2297.
- Zhong, P. S. and Chen, S. X. (2011). Tests for high dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association*, 106, 260-274.

CHAPTER 4. ANOVA for Longitudinal Data with Missing Values

A paper published in *the Annals of Statistics*, 38, 3630-3659.

Song Xi Chen and Pingshou Zhong

Department of Statistics

Iowa State University

Ames, IA 50011, USA

Abstract

We carry out ANOVA comparisons of multiple treatments for longitudinal studies with missing values. The treatment effects are modelled semiparametrically via a partially linear regression which is flexible in quantifying the time effects of treatments. The empirical likelihood is employed to formulate model-robust nonparametric ANOVA tests for treatment effects with respect to covariates, the nonparametric time-effect functions and interactions between covariates and time. The proposed tests can be readily modified for a variety of data and model combinations, that encompass parametric, semiparametric and nonparametric regression models; cross-sectional and longitudinal data, and with or without missing values.

KEY WORDS: Analysis of Variance; Empirical likelihood; Kernel smoothing; Missing at random; Semiparametric model; Treatment effects.

4.1 Introduction

Randomized clinical trials and observational studies are often used to evaluate treatment effects. While the treatment versus control studies are popular, multi-treatment comparisons beyond two samples are commonly practised in clinical trails and observational studies. In

addition to evaluate overall treatment effects, investigators are also interested in intra-individual changes over time by collecting repeated measurements on each individual over time. Although most longitudinal studies are desired to have all subjects measured at the same set of time points, such “balanced” data may not be available in practice due to missing values. Missing values arise when scheduled measurements are not made, which make the data “unbalanced”. There is a good body of literature on parametric, nonparametric and semiparametric estimation for longitudinal data with or without missing values. This includes Liang and Zeger (1986), Laird and Ware (1982), Wu (1998, 2000), Fitzmaurice *et al.* (2004) for methods developed for longitudinal data without missing values; and Little and Rubin (2002), Little (1995), Laird (2004), Robins, Rotnitzky and Zhao (1995) for missing values.

The aim of this chapter is to develop ANOVA tests for multi-treatment comparisons in longitudinal studies with or without missing values. Suppose that at time t , corresponding to k treatments there are k mutually independent samples:

$$\{(Y_{1i}(t), X_{1i}^\tau(t))\}_{i=1}^{n_1}, \quad \dots, \quad \{(Y_{ki}(t), X_{ki}^\tau(t))\}_{i=1}^{n_k}$$

where the response variable $Y_{ji}(t)$ and the covariate $X_{ji}(t)$ are supposed to be measured at time points $t = t_{ji1}, \dots, t_{jiT_j}$. Here T_j is the fixed number of scheduled observations for the j -th treatment. However, $\{Y_{ji}(t), X_{ji}^\tau(t)\}$ may not be observed at some times, resulting in missing values in either the response $Y_{ji}(t)$ or the covariates $X_{ji}(t)$.

We consider a semiparametric regression model for the longitudinal data

$$Y_{ji}(t) = X_{ji}^\tau(t)\beta_{j0} + M^\tau(X_{ji}(t), t)\gamma_{j0} + g_{j0}(t) + \varepsilon_{ji}(t), \quad j = 1, 2, \dots, k \quad (4.1.1)$$

where $M(X_{ji}(t), t)$ are known functions of $X_{ji}(t)$ and time t representing interactions between the covariates and the time, β_{j0} and γ_{j0} are p - and q -dimensional parameters respectively, $g_{j0}(t)$ are unknown smooth functions representing the time effect, and $\{\varepsilon_{ji}(t)\}$ are residual time series. Such a semiparametric model may be viewed as an extended partially linear model. The partially linear model has been used for longitudinal data analysis; see Zeger and Diggle (1994), Zhang, Lin, Raz and Sowers (1998), Lin and Ying (2001), Wang, Carroll and Lin (2005). Wu *et al.* (1998) and Wu and Chiang (2000) proposed estimation and confidence

regions for a semiparametric varying coefficient regression model. Despite a body of works on estimation for longitudinal data, analysis of variance for longitudinal data have attracted much less attention. A few exceptions include Forcina (1992) who proposed an ANOVA test in a fully parametric setting; and Scheike and Zhang (1998) who considered a two sample test in a fully nonparametric setting.

In this chapter, we propose ANOVA tests for differences among the β_{j0} s and the baseline time functions g_{j0} s respectively in the presence of the interactions. The ANOVA statistics are formulated based on the empirical likelihood (Owen, 1988 and 2001), which can be viewed as a nonparametric counterpart of the conventional parametric likelihood. Despite its not requiring a fully parametric model, the empirical likelihood enjoys two key properties of a conventional likelihood, the Wilks' theorem (Owen 1990, Qin and Lawless 1994, Fan and Zhang 2004) and Bartlett correction (DiCiccio, Hall and Romano 1991; Chen and Cui 2006); see Chen and Van Keilegom (2009) for an overview on the empirical likelihood for regression. This resemblance to the parametric likelihood ratio motivates us to consider using empirical likelihood to formulate ANOVA test for longitudinal data in nonparametric situations. This will introduce a much needed model-robustness in the ANOVA testing.

Empirical likelihood has been used in studies for either missing or longitudinal data. Wang *et al.* (2002, 2004) considered an empirical likelihood inference with a kernel regression imputation for missing responses. Liang and Qin (2008) treated estimation for the partially linear model with missing covariates. For longitudinal data, Xue and Zhu (2007a, 2007b) proposed a bias correction method to make the empirical likelihood statistic asymptotically pivotal in a one sample partially linear model; see also You, Chen and Zhou (2007) and Huang, Qin and Follman (2008).

In this chapter, we propose three empirical likelihood based ANOVA tests for the equivalence of the treatment effects with respect to (i) the covariate X_{ji} ; (ii) the interactions $M(X_{ji}(t), t)$ and (iii) the time effect functions $g_{j0}(\cdot)$ s, by formulating empirical likelihood ratio test statistics. It is shown that for the proposed ANOVA tests for the covariates effects and the interactions, the empirical likelihood ratio statistics are asymptotically chi-squared distributed, which resembles the conventional ANOVA statistics based on parametric likelihood ratios. This is achieved

without parametric model assumptions for the residuals and in the presence of the nonparametric time effect functions and missing values. Hence the empirical likelihood ANOVA tests have the needed model-robustness. Another attraction of the proposed ANOVA tests is that they encompass a set of ANOVA tests for a variety of data and model combinations. Specifically, they imply specific ANOVA tests for both cross-sectional and longitudinal data; for parametric, semiparametric and nonparametric regression models; and with or without missing values.

The chapter is organized as below. In Section 4.2, we describe the model and the missing value mechanism. Section 4.3 outlines the ANOVA test for comparing treatment effects due to the covariates; whereas the tests regarding interaction are proposed in Section 4.4. Section 4.5 considers ANOVA test for the nonparametric time effects. The bootstrap calibration to the ANOVA test on the nonparametric part is outlined in Section 4.6. Section 4.7 reports simulation results. We applied the proposed ANOVA tests in Section 4.8 to analyze an HIV-CD4 data set. Technical assumptions and all the technical proofs to the theorems are reported in the Appendix.

4.2 Models, Hypotheses and Missing Values

For the i -th individual of the j -th treatment, the measurements taken at time t_{jim} follow a semiparametric model

$$Y_{ji}(t_{jim}) = X_{ji}^T(t_{jim})\beta_{j0} + M^T(X_{ji}(t_{jim}), t_{jim})\gamma_{j0} + g_{j0}(t_{jim}) + \varepsilon_{ji}(t_{jim}), \quad (4.2.2)$$

for $j = 1, \dots, k$, $i = 1, \dots, n_j$, $m = 1, \dots, T_j$. Here β_{j0} and γ_{j0} are unknown p - and q -dimensional parameters and $g_{j0}(t)$ are unknown functions representing the time effects of the treatments. The time points $\{t_{jim}\}_{m=1}^{T_j}$ are known design points. For the ease of notation, we write $(Y_{jim}, X_{jim}^T, M_{jim}^T)$ to denote $(Y_{ji}(t_{jim}), X_{ji}^T(t_{jim}), M^T(X_{ji}(t_{jim}), t_{jim}))$. Also, we will use $\mathbb{X}_{jim}^T = (X_{jim}^T, M_{jim}^T)$ and $\xi_j^T = (\beta_j^T, \gamma_j^T)$. For each individual, the residuals $\{\varepsilon_{ji}(t)\}$ satisfy $E\{\varepsilon_{ji}(t)|X_{ji}(t)\} = 0$, $\text{Var}\{\varepsilon_{ji}(t)|X_{ji}(t)\} = \sigma_j^2(t)$ and

$$\text{Cov}\{\varepsilon_{ji}(t), \varepsilon_{ji}(s)|X_{ji}(t), X_{ji}(s)\} = \rho_j(s, t)\sigma_j(t)\sigma_j(s)$$

where $\rho_j(s, t)$ is the conditional correlation coefficient between two residuals at two different times. And the residual time series $\{\varepsilon_{ji}(t)\}$ from different subjects and different treatments are

independent. Without loss of generality, we assume $t, s \in [0, 1]$. For the purpose of identifying β_{j0} , γ_{j0} and $g_{j0}(t)$, we assume

$$(\beta_{j0}, \gamma_{j0}, g_{j0}) = \arg \min_{(\beta_j, \gamma_j, g_j)} \frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E\{Y_{jim} - X_{jim}^\tau \beta_j - M_{jim}^\tau \gamma_j - g_j(t_{jim})\}^2.$$

We also require that $\frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E(\tilde{\mathbb{X}}_{jim} \tilde{\mathbb{X}}_{jim}^\tau) > 0$, where $\tilde{\mathbb{X}}_{jim} = \mathbb{X}_{jim} - E(\mathbb{X}_{jim} | t_{jim})$. This condition also rules out $M(X_{ji}(t), t)$ being a pure function of t , and hence it has to be genuine interaction. For the same reason, the intercept in model (4.2.2) is absorbed into the nonparametric part $g_{j0}(t)$.

As commonly exercised in the partially linear model (Speckman 1988; Linton 1995), there is a secondary model for the covariate X_{jim} :

$$X_{jim} = h_j(t_{jim}) + u_{jim}, \quad j = 1, 2, \dots, k, \quad i = 1, \dots, n_j, \quad m = 1, \dots, T_j, \quad (4.2.3)$$

where $h_j(\cdot)$ s are p -dimensional smooth functions with continuous second derivatives, the residual $u_{jim} = (u_{jim}^1, \dots, u_{jim}^p)^\tau$ satisfy $E(u_{jim}) = 0$ and u_{jl} and u_{jk} are independent for $l \neq k$, where $u_{jl} = (u_{jl1}, \dots, u_{jlT_j})$. By the identification condition given above, the covariance matrix of u_{jim} is assumed to be finite and positive definite.

We are interested in testing three ANOVA hypotheses. The first one is on the treatment effects with respect to the covariates:

$$H_{0a} : \beta_{10} = \beta_{20} = \dots = \beta_{k0} \quad \text{vs} \quad H_{1a} : \beta_{i0} \neq \beta_{j0} \text{ for some } i \neq j.$$

The second one is regarding the time effect functions:

$$H_{0b} : g_{10}(\cdot) = \dots = g_{k0}(\cdot) \quad \text{vs} \quad H_{1b} : g_{i0}(\cdot) \neq g_{j0}(\cdot) \text{ for some } i \neq j.$$

The third one is on the existence of the interaction $H_{0c} : \gamma_{j0} = 0$ and $H_{1c} : \gamma_{j0} \neq 0$. And the last one is the ANOVA test for

$$H_{0d} : \gamma_{10} = \gamma_{20} = \dots = \gamma_{k0} \quad \text{vs} \quad H_{1d} : \gamma_{i0} \neq \gamma_{j0} \text{ for some } i \neq j.$$

Let $X_{ji} = \{X_{ji0}, \dots, X_{jiT_j}\}$ and $Y_{ji} = \{Y_{ji0}, \dots, Y_{jiT_j}\}$ be the complete time series of the covariates and responses of the (j, i) -th subject (the i -th subject in the j -th treatment), and

$\overleftarrow{Y}_{jit,d} = \{Y_{ji(t-d)}, \dots, Y_{ji(t-1)}\}$ and $\overleftarrow{X}_{jit,d} = \{X_{ji(t-d)}, \dots, X_{ji(t-1)}\}$ be the past d observations at time t for a positive integer $d \leq \min_j\{T_j\}$. For $t < d$, we set $d = t - 1$.

Define the missing value indicator $\delta_{jit} = 1$ if (X_{jit}^T, Y_{jit}) is observed and $\delta_{jit} = 0$ if (X_{jit}^T, Y_{jit}) is missing. Here, we assume X_{jit} and Y_{jit} are either both observed or both missing. This simultaneous missingness of X_{jit} and Y_{jit} is for the ease of mathematical exposition. We also assume that $\delta_{ji0} = 1$, namely the first visit of each subject is always made.

Monotone missingness is a common assumption in the analysis of longitudinal data (Robins *et al.*, 1995). It assumes that if $\delta_{ji(t-1)} = 0$ then $\delta_{jit} = 0$. However, in practice after missing some scheduled appointments people may re-join the study. This kind of casual drop-out appears quite often in empirical studies. To allow more data being included in the analysis, we relax the monotone missingness to allow segments of consecutive d visits being used. Let $\delta_{jit,d} = \prod_{l=1}^d \delta_{ji(t-l)}$. We assume the missingness of (X_{jit}^T, Y_{jit}) is missing at random (MAR) (Rubin, 1976) given its immediate past d complete observations, namely

$$\begin{aligned} P(\delta_{jit} = 1 | \delta_{jit,d} = 1, X_{ji}, Y_{ji}) &= P(\delta_{jit} = 1 | \delta_{jit,d} = 1, \overleftarrow{X}_{jit,d}, \overleftarrow{Y}_{jit,d}) \\ &= p_j(\overleftarrow{X}_{jit,d}, \overleftarrow{Y}_{jit,d}; \theta_{j0}). \end{aligned} \quad (4.2.4)$$

Here the missing propensity p_j is known up to a parameter θ_{j0} . To allow derivation of a binary likelihood function, we need to set $\delta_{jit} = 0$ if $\delta_{jit,d} = 0$ when there is some drop-outs among the past d visits, which is only temporarily if $\delta_{jit} = 1$. This set-up ensures

$$P(\delta_{jit} = 0 | \delta_{jit,d} = 0, \overleftarrow{X}_{jit,d}, \overleftarrow{Y}_{jit,d}) = 1. \quad (4.2.5)$$

Now the conditional binary likelihood for $\{\delta_{jit}\}_{t=1}^{T_j}$ given X_{ji} and Y_{ji} is

$$\begin{aligned} P(\delta_{ji0}, \dots, \delta_{jiT_j} | X_{ji}, Y_{ji}) &= \prod_{m=1}^{T_j} P(\delta_{jim} | \delta_{ji(m-1)}, \dots, \delta_{ji0}, X_{ji}, Y_{ji}) \\ &= \prod_{m=1}^{T_j} P(\delta_{jim} | \delta_{jim,d} = 1, \overleftarrow{X}_{jim,d}, \overleftarrow{Y}_{jim,d}) \\ &= \prod_{m=1}^{T_j} [p_j(\overleftarrow{X}_{jim,d}, \overleftarrow{Y}_{jim,d}; \theta_j)^{\delta_{jim}} \{1 - p_j(\overleftarrow{X}_{jim,d}, \overleftarrow{Y}_{jim,d}; \theta_j)\}^{(1-\delta_{jim})}]^{\delta_{jim,d}}. \end{aligned}$$

In the second equation above we use both the MAR in (4.2.4) and (4.2.5). Hence, the parameters θ_{j0} can be estimated by maximizing the binary likelihood

$$\mathcal{L}_{B_j}(\theta_j) = \prod_{i=1}^{n_j} \prod_{t=1}^{T_j} [p_j(\bar{X}_{jit,d}, \bar{Y}_{jit,d}; \theta_j)^{\delta_{jit}} \{1 - p_j(\bar{X}_{jit,d}, \bar{Y}_{jit,d}; \theta_j)\}^{(1-\delta_{jit})}]^{\delta_{jit,d}}. \quad (4.2.6)$$

Under some regular conditions, the binary maximum likelihood estimator $\hat{\theta}_j$ is \sqrt{n} -consistent estimator of θ_{j0} ; see Chen *et al.* (2008) for results on a related situation. Some guidelines on how to choose models for the missing propensity are given in Section 4.8 in the context of the empirical study. The robustness of the ANOVA tests with respect to the missing propensity model are discussed in Sections 4.3 and 4.4.

4.3 ANOVA Test for Covariate Effects

We consider testing for $H_{0a} : \beta_{10} = \beta_{20} = \dots = \beta_{k0}$ with respect to the covariates. Let $\pi_{jim}(\theta_j) = \prod_{l=m-d}^m p_j(\bar{X}_{jil,d}, \bar{Y}_{jil,d}; \theta_j)$ be the overall missing propensity for the (j, i) -th subject up to time t_{jim} . To remove the nonparametric part in (4.2.2), we first estimate the nonparametric function $g_{j0}(t)$. If β_{j0} and γ_{j0} were known, $g_{j0}(t)$ would be estimated by

$$\hat{g}_j(t; \beta_{j0}) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h}(t) (Y_{jim} - X_{jim}^\tau \beta_{j0} - M_{jim}^\tau \gamma_{j0}), \quad (4.3.7)$$

where

$$w_{jim,h_j}(t) = \frac{(\delta_{jim}/\pi_{jim}(\hat{\theta}_j))K_{h_j}(t_{jim} - t)}{\sum_{s=1}^{n_j} \sum_{l=1}^{T_j} (\delta_{jsl}/\pi_{jsl}(\hat{\theta}_j))K_{h_j}(t_{jsl} - t)} \quad (4.3.8)$$

is a kernel weight that has been inversely weighted by the propensity $\pi_{jim}(\hat{\theta}_j)$ to correct for selection bias due to the missing values. In (4.3.8), K is a univariate kernel function which is a symmetric probability density, $K_{h_j}(t) = K(t/h_j)/h_j$ and h_j is a smoothing bandwidth. The conventional kernel estimation of $g_{j0}(t)$ without weighting by $\pi_{jsl}(\hat{\theta}_j)$ may be inconsistent if the missingness depends on the responses Y_{jil} , which can be the case for missing covariates.

Let A_{jim} denote any of X_{jim} , Y_{jim} and M_{jim} and define

$$\tilde{A}_{jim} = A_{jim} - \sum_{i_1=1}^{n_j} \sum_{m_1=1}^{T_j} w_{ji_1m_1,h_j}(t_{jim}) A_{ji_1m_1} \quad (4.3.9)$$

to be the centering of A_{jim} by the kernel conditional mean estimate, as is commonly exercised in the partially linear regression (Härdle, Liang and Gao, 2000). An estimating function for the (j, i) -th subject is

$$Z_{ji}(\beta_j) = \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{X}_{jim} (\tilde{Y}_{jim} - \tilde{X}_{jim}^\tau \beta_j - \tilde{M}_{jim}^\tau \tilde{\gamma}_j),$$

where $\tilde{\gamma}_j$ is the solution of

$$\sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{M}_{jim} (\tilde{Y}_{jim} - \tilde{X}_{jim}^\tau \beta_j - \tilde{M}_{jim}^\tau \tilde{\gamma}_j) = 0$$

at given β_j . Note that $E\{Z_{ji}(\beta_{j0})\} = o(1)$. Although it is not exactly zero, $Z_{ji}(\beta_{j0})$ can still be used as an approximate zero mean estimating function to formulate an empirical likelihood for β_j as follows.

Let $\{p_{ji}\}_{i=1}^{n_j}$ be non-negative weights allocated to $\{(X_{ji}^\tau, Y_{ji})\}_{i=1}^{n_j}$. The empirical likelihood for β_j is

$$L_{n_j}(\beta_j) = \max \left\{ \prod_{i=1}^{n_j} p_{ji} \right\}, \quad (4.3.10)$$

subject to $\sum_{i=1}^{n_j} p_{ji} = 1$ and $\sum_{i=1}^{n_j} p_{ji} Z_{ji}(\beta_j) = 0$.

By introducing a Lagrange multiplier λ_j to solve the above optimization problem and following the standard derivation in empirical likelihood (Owen, 1990), it can be shown that

$$L_{n_j}(\beta_j) = \prod_{i=1}^{n_j} \left\{ \frac{1}{n_j} \frac{1}{1 + \lambda_j^\tau Z_{ji}(\beta_j)} \right\}, \quad (4.3.11)$$

where λ_j satisfies

$$\sum_{i=1}^{n_j} \frac{Z_{ji}(\beta_j)}{1 + \lambda_j^\tau Z_{ji}(\beta_j)} = 0. \quad (4.3.12)$$

The maximum of $L_{n_j}(\beta_j)$ is $\prod_{i=1}^{n_j} \frac{1}{n_j}$, achieved at $\beta_j = \hat{\beta}_j$ and $\lambda_j = 0$, where $\hat{\beta}_j$ solves $\sum_{i=1}^{n_j} Z_{ji}(\hat{\beta}_j) = 0$.

Let $n = \sum_{i=1}^k n_j$, $n_j/n \rightarrow \rho_j$ for some non-zero ρ_j as $n \rightarrow \infty$ such that $\sum_{i=1}^k \rho_j = 1$. As the k samples are independent, the joint empirical likelihood for $(\beta_1, \beta_2, \dots, \beta_k)$ is

$$L_n(\beta_1, \beta_2, \dots, \beta_k) = \prod_{j=1}^k L_{n_j}(\beta_j).$$

The log likelihood ratio statistic for H_{0a} is

$$\begin{aligned}\ell_n : &= -2 \max_{\beta} \log L_n(\beta, \beta, \dots, \beta) + \sum_{j=1}^k n_j \log n_j \\ &= 2 \min_{\beta} \sum_{j=1}^k \sum_{i=1}^{n_j} \log\{1 + \lambda_j^{\tau} Z_{ji}(\beta)\}.\end{aligned}\quad (4.3.13)$$

Using a Taylor expansion and the Lagrange multiplier to carry out the minimization in (4.3.13), the optimal solution to β is

$$\left(\sum_{j=1}^k \Omega_{x_j} B_j^{-1} \Omega_{x_j} \right)^{-1} \left(\sum_{j=1}^k \Omega_{x_j} B_j^{-1} \Omega_{x_j y_j} \right) + o_p(1), \quad (4.3.14)$$

where $B_j = \lim_{n_j \rightarrow \infty} (n_j T_j)^{-1} \sum_{i=1}^{n_j} E\{Z_{ji}(\beta_{j0}) Z_{ji}(\beta_{j0})^{\tau}\}$,

$$\Omega_{x_j} = \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E \left\{ \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{X}_{jim} \tilde{X}_{jim}^{\tau} \right\}$$

and

$$\Omega_{x_j y_j} = \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{X}_{jim} (\tilde{Y}_{jim} - M_{jim}^{\tau} \tilde{\gamma}_j).$$

The ANOVA test statistic (4.3.13) can be viewed as a nonparametric counterpart of the conventional parametric likelihood ratio ANOVA test statistic, for instance that considered in Forcina (1992). Like its parametric counterpart, the Wilks' theorem is maintained for ℓ_n .

Theorem 1 *If Conditions (A1-A4) given in the Appendix hold, then under H_{0a} , $\ell_n \xrightarrow{d} \chi_{(k-1)p}^2$ as $n \rightarrow \infty$.*

The theorem suggests an empirical likelihood ANOVA test that rejects H_{0a} if $\ell_n > \chi_{(k-1)p, \alpha}^2$ where α is the significant level and $\chi_{(k-1)p, \alpha}^2$ is the upper α quantile of the $\chi_{(k-1)p}^2$ distribution.

We next evaluate the power of the empirical likelihood ANOVA test under a series of local alternative hypotheses:

$$H_{1a} : \beta_{j0} = \beta_{10} + c_n n_j^{-1/2} \quad \text{for } 2 \leq j \leq k$$

where $\{c_n\}$ is a sequence of bounded constants. Define $\Delta_{\beta} = (\beta_{10}^{\tau} - \beta_{20}^{\tau}, \beta_{10}^{\tau} - \beta_{30}^{\tau}, \dots, \beta_{10}^{\tau} - \beta_{k0}^{\tau})^{\tau}$, $D_{1j} = \Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_j}^{-1} \Omega_{x_j y_j}$ for $2 \leq j \leq k$ and $D = (D_{12}^{\tau}, D_{13}^{\tau}, \dots, D_{1k}^{\tau})^{\tau}$. Let $\Sigma_D = \text{Var}(D)$ and $\gamma^2 = \Delta_{\beta}^{\tau} \Sigma_D^{-1} \Delta_{\beta}$. Theorem 2 gives the asymptotic distribution of ℓ_n under the local alternatives.

Theorem 2 *Suppose Conditions (A1-A4) in the Appendix hold, then under H_{1a} , $\ell_n \xrightarrow{d} \chi_{(k-1)p}^2(\gamma^2)$ as $n \rightarrow \infty$.*

It can be shown that

$$\Sigma_D = \Omega_{x_1}^{-1} B_1 \Omega_{x_1}^{-1} \mathbf{1}_{(k-1)} \otimes \mathbf{1}_{(k-1)} + \text{diag}\{\Omega_{x_2}^{-1} B_2 \Omega_{x_2}^{-1}, \dots, \Omega_{x_k}^{-1} B_k \Omega_{x_k}^{-1}\}. \quad (4.3.15)$$

As each $\Omega_{x_j}^{-1}$ is $O(n^{1/2})$, the non-central component γ^2 is non-zero and bounded. The power of the α level empirical likelihood ANOVA test is $\beta(\gamma) = P\{\chi_{(k-1)p}^2(\gamma^2) > \chi_{(k-1)p, \alpha}^2\}$. This indicates that the test is able to detect local departures of size $O(n^{-1/2})$ from H_{0a} , which is the best rate we can achieve under the local alternative set-up. This is attained despite the fact that nonparametric kernel estimation is involved in the formulation, which has a slower rate of convergence than \sqrt{n} , as the centering in (4.3.9) essentially eliminates the effects of the nonparametric estimation.

Remark 1. When there is no missing values, namely all $\delta_{jim} = 1$, we will assign all $\pi_{jim}(\hat{\theta}_j) = 1$ and there is no need to estimate each θ_j . In this case, Theorems 1 and 2 remain valid. It is a different matter for estimation as estimation efficiency with missing values will be less than that without missing values.

Remark 2. The above ANOVA test is robust against mis-specifying the missing propensity $p_j(\cdot; \theta_{j0})$ provided the missingness does not depend on the responses $\overleftarrow{Y}_{jit,d}$. This is because despite the mis-specification, the mean of $Z_{ji}(\beta)$ is still approximately zero and the empirical likelihood formulation remains valid, as well as Theorems 1 and 2. However, if the missingness depends on the responses and if the model is mis-specified, Theorems 1 and 2 will be affected.

Remark 3. The empirical likelihood test can be readily modified for ANOVA testing on pure parametric regressions with some parametric time effects $g_{j0}(t; \eta_j)$ with parameters η_j . When there is absence of interaction, we may formulate the empirical likelihood for $(\beta_j, \eta_j) \in R^{p+s}$ using

$$Z_{ji}(\beta_j; \eta_j) = \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \left(X_{jim}^\tau, \frac{\partial g_j^\tau(t_{jim}; \eta_j)}{\partial \eta_j} \right)^\tau \left\{ Y_{jim} - X_{jim}^\tau \beta_j - g_{j0}(t_{jim}; \eta_j) \right\}$$

as the estimating function for the (j, i) -th subject. The ANOVA test can be formulated following the same procedures from (4.3.11) to (4.3.13), and both Theorems 1 and 2 remaining valid after

updating p with $p + s$ where s is the dimension of η_j .

In our formulation for the ANOVA test here and in the next section, we rely on the Nadaraya-Watson type kernel estimator. The local linear kernel estimator may be employed when the boundary bias may be an issue. However, as we are interested in ANOVA tests instead of estimation, the boundary bias does not have a leading order effect.

4.4 ANOVA Test for Time Effects

In this section, we consider the ANOVA test for the nonparametric part

$$H_{0b} : g_{10}(\cdot) = \dots = g_{k0}(\cdot).$$

We will first formulate an empirical likelihood for $g_{j0}(t)$ at each t , which then lead to an overall likelihood ratio for H_{0b} . We need an estimator of $g_{j0}(t)$ that is less biased than the one in (4.3.7). Recall the notation defined in Section 4.2: $\mathbb{X}_{jim}^\tau = (X_{jim}^\tau, M_{jim}^\tau)$ and $\xi_j^\tau = (\beta_j^\tau, \gamma_j^\tau)$. Plugging-in the estimator $\hat{\xi}_j$ to (4.3.7), we have

$$\tilde{g}_j(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim, h_j}(t) (Y_{jim} - \mathbb{X}_{jim}^\tau \hat{\xi}_j). \quad (4.4.16)$$

It follows that, for any $t \in [0, 1]$,

$$\tilde{g}_j(t) - g_{j0}(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim, h_j}(t) \left\{ \varepsilon_{ji}(t_{jim}) + \mathbb{X}_{jim}^\tau (\xi_j - \hat{\xi}_j) + g_{j0}(t_{jim}) - g_{j0}(t) \right\}. \quad (4.4.17)$$

However, there is a bias of order h_j^2 in the kernel estimation since

$$\sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim, h_j}(t) \{g_{j0}(t_{jim}) - g_{j0}(t)\} = \frac{1}{2} \left\{ \int z^2 K(z) dz \right\} g_{j0}''(t) h_j^2 + o_p(h_j^2).$$

If we formulated the empirical likelihood based on $\tilde{g}_j(t)$, the bias will contribute to the asymptotic distribution of the ANOVA test statistic. To avoid that, we use the bias-correction method proposed in Xue and Zhu (2007a) so that the estimator of g_{j0} is

$$\hat{g}_j(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim, h_j}(t) \{Y_{jim} - \mathbb{X}_{jim}^\tau \hat{\xi}_j - (\tilde{g}_j(t_{jim}) - \tilde{g}_j(t))\}.$$

Based on this modified estimator $\hat{g}_j(t)$, we define the auxiliary variable

$$R_{ji}\{g_j(t)\} = \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} K\left(\frac{t_{jim} - t}{h_j}\right) \left\{ Y_{jim} - \mathbb{X}_{jim}^\tau \hat{\xi}_j - g_j(t) - (\tilde{g}_j(t_{jim}) - \tilde{g}_j(t)) \right\}$$

for empirical likelihood formulation. At true function $g_{j0}(t)$, $E[R_{ji}\{g_{j0}(t)\}] = o(1)$.

Using a similar procedure to $L_{n_j}(\beta_j)$ as given in (4.3.11) and (4.3.12), the empirical likelihood for $g_{j0}(t)$ is

$$L_{n_j}\{g_{j0}(t)\} = \max \left\{ \prod_{i=1}^{n_j} p_{ji} \right\}$$

subject to $\sum_{i=1}^{n_j} p_{ji} = 1$ and $\sum_{i=1}^{n_j} p_{ji} R_{ji}\{g_j(t)\} = 0$. The latter is obtained in a similar fashion as we obtain (4.3.11) by introducing Lagrange multipliers so that

$$L_{n_j}\{g_{j0}(t)\} = \prod_{i=1}^{n_j} \left\{ \frac{1}{n_j} \frac{1}{1 + \eta_j(t) R_{ji}\{g_{j0}(t)\}} \right\},$$

where $\eta_j(t)$ is a Lagrange multiplier that satisfies

$$\sum_{i=1}^{n_j} \frac{R_{ji}\{g_{j0}(t)\}}{1 + \eta_j(t) R_{ji}\{g_{j0}(t)\}} = 0. \quad (4.4.18)$$

The log empirical likelihood ratio for $g_{10}(t) = \dots = g_{k0}(t) := g(t)$, say, is

$$\mathcal{L}_n(t) = 2 \min_{g(t)} \sum_{j=1}^k \sum_{i=1}^{n_j} \log(1 + \eta_j(t) R_{ji}\{g(t)\}), \quad (4.4.19)$$

which is analogues of ℓ_n in (4.3.13). As shown in the proof of Theorem 3 given in the Appendix, the leading order term of the $\mathcal{L}_n(t)$ is a studentized version of the distance

$$(\hat{g}_1(t) - \hat{g}_2(t), \hat{g}_1(t) - \hat{g}_3(t), \dots, \hat{g}_1(t) - \hat{g}_k(t)),$$

namely between $\hat{g}_1(t)$ and the other $\hat{g}_j(t)$ ($j \neq 1$). This motivates us to propose using

$$\mathcal{T}_n = \int_0^1 \mathcal{L}_n(t) \varpi(t) dt \quad (4.4.20)$$

to test for the equivalence of $\{g_{j0}(\cdot)\}_{j=1}^k$, where $\varpi(t)$ is a probability weight function over $[0, 1]$.

To define the asymptotic distribution of \mathcal{T}_n , we assume without loss of generality that for each h_j and T_j , $j = 1, \dots, k$, there exist fixed finite positive constants α_j and b_j such that $\alpha_j T_j = T$ and $b_j h_j = h$ for some T and h as $h \rightarrow 0$. Effectively, T is the smallest common multiple of T_1, \dots, T_k . Let $K_c^{(2)}(t) = \int K(w) K(t-cw) dw$ and $K_c^{(4)}(0) = \int K_c^{(2)}(w\sqrt{c}) K_{1/c}^{(2)}(w/\sqrt{c}) dw$. For $c = 1$, we resort to the standard notations of $K^{(2)}(t)$ and $K^{(4)}(0)$ for $K_1^{(2)}(t)$ and $K_1^{(4)}(0)$, respectively. For each treatment j , let f_j be the super-population density of the design points $\{t_{jim}\}$. Let $a_j = \rho_j^{-1} \alpha_j$,

$$W_j(t) = \frac{f_j(t)/\{a_j b_j \sigma_{\varepsilon_j}^2\}}{\sum_{l=1}^k f_l(t)/\{a_l b_l \sigma_{\varepsilon_l}^2\}}$$

and $V_j(t) = K^{(2)}(0)\sigma_{\varepsilon_j}^2 f_j(t)$ where $\sigma_{\varepsilon_j}^2 = \frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E\left\{\frac{\varepsilon_{jim}^2}{\pi_{jim}(\theta_{j0})}\right\}$. Furthermore, we define

$$\begin{aligned} \Lambda(t) &= \sum_{j=1}^k b_j^{-1} K^{(4)}(0)(1 - W_j(t))^2 + \sum_{j \neq j_1}^k (b_j b_{j_1})^{-\frac{1}{2}} K_{b_j/b_{j_1}}^{(4)}(0) W_j(t) W_{j_1}(t) \quad \text{and} \\ \mu_1 &= \int_0^1 \left[\sum_{j=1}^k b_j^{-\frac{1}{2}} V_j^{-1}(t) f_j^2(t) \Delta_{n_j}^2(t) - \left(\sum_{s=1}^k b_s^{-\frac{1}{4}} V_s^{-\frac{1}{2}}(t) W_s^{\frac{1}{2}}(t) f_s(t) \Delta_{n_s}(t) \right)^2 \right] \varpi(t) dt. \end{aligned}$$

We consider a sequence of local alternative hypotheses:

$$g_{j0}(t) = g_{10}(t) + C_{jn} \Delta_{jn}(t), \quad (4.4.21)$$

where $C_{jn} = (n_j T_j)^{-1/2} h_j^{-1/4}$ for $j = 2, \dots, k$ and $\{\Delta_{jn}(t)\}_{n \geq 1}$ is a sequence of uniformly bounded functions.

Theorem 3 *Assume Conditions (A1-A4) in the Appendix and $h = O(n^{-1/5})$, then under (4.4.21),*

$$h^{-1/2}(\mathcal{T}_n - \mu_0) \xrightarrow{d} N(0, \sigma_0^2),$$

where $\mu_0 = (k-1) + h^{1/2} \mu_1$ and $\sigma_0^2 = 2K^{(2)}(0)^{-2} \int_0^1 \Lambda(t) \varpi^2(t) dt$.

We note that under $H_{0b} : g_{10}(\cdot) = \dots = g_{k0}(\cdot)$, $\Delta_{jn}(t) = 0$ which yields $\mu_1 = 0$ and

$$h^{-1/2}\{\mathcal{T}_n - (k-1)\} \xrightarrow{d} N(0, \sigma_0^2).$$

This may lead to an asymptotic test at a nominal significance level α that rejects H_{0b} if

$$\mathcal{T}_n \geq h^{1/2} \hat{\sigma}_0 z_\alpha + (k-1), \quad (4.4.22)$$

where z_α is the upper α quantile of $N(0, 1)$ and $\hat{\sigma}_0$ is a consistent estimator of σ_0 . The asymptotic power of the test under the local alternatives is $1 - \Phi(z_\alpha - \frac{\mu_1}{\sigma_0})$, where $\Phi(\cdot)$ is the standard normal distribution function. This indicates that the test is powerful in differentiating null hypothesis and its local alternative at the convergence rate $O(n_j^{-1/2} h_j^{-1/4})$ for C_{jn} . The rate is the best when a single bandwidth is used (Härdle and Mammen, 1993).

If all the h_j ($j = 1, \dots, k$) are the same, the asymptotic variance

$$\sigma_0^2 = 2(k-1)K^{(2)}(0)^{-2}K^{(4)}(0) \int_0^1 \varpi^2(t) dt,$$

which means that the test statistic under H_{0b} is asymptotic pivotal. However, when the bandwidths are not the same, which is most likely as different treatments may require different amount of smoothness in the estimation of $g_{j0}(\cdot)$, the asymptotical pivotalness of \mathcal{T}_n is no longer available, and estimation of σ_0^2 is needed for conducting the asymptotic test in (4.4.22). We will propose a test based on a bootstrap calibration to the distribution of \mathcal{T}_n in Section 4.6.

Remark 4. Similar to **Remarks 1 and 2** made on the ANOVA tests for the covariate effects, the proposed ANOVA test for the nonparametric baseline functions (Theorem 3) remains valid in the absence of missing values or if the missing propensity is mis-specified as long as the responses do not contribute to the missingness.

Remark 5. We note that the proposed test is not affected by the within-subject dependent structure (the longitudinal aspect) due to the fact that the formulation of the empirical likelihood is made for each subject. This is clearly shown in the construction of $R_{ji}\{g_j(t)\}$ and by the fact that the nonparametric functions can be separated from the covariate effects in the semiparametric model. Again this would be changed if we are interested in estimation as the correlation structure in the longitudinal data will affect the estimation efficiency. However, the test will be dependent on the choice of the weight function $\varpi(\cdot)$, and $\{\alpha_j\}$, $\{\rho_j\}$ and $\{b_j\}$, the relative ratios among $\{T_j\}$, $\{n_j\}$ and $\{h_j\}$.

Remark 6. The ANOVA test statistics for the time effects for the semiparametric model can be readily modified to obtain ANOVA test for purely nonparametric regression by simply setting $\hat{\xi}_j = 0$ in the formulation of the test statistic $\mathcal{L}_n(t)$. In this case, the model (4.2.2) takes the form

$$Y_{ji}(t) = g_j(X_{ji}(t), t) + \varepsilon_{ji}(t), \quad (4.4.23)$$

where $g_j(\cdot)$ is the unknown nonparametric function of $X_{ji}(t)$ and t . The proposed ANOVA test can be viewed as generalization of the tests considered in Mund and Detter (1998), Pardo-Fernández, Van Keilegom and González-Manteiga (2007) and Wang, Akritas and Van Keilegom (2008) by considering both the longitudinal and missing aspects. See also Cao and Van Keilegom (2006) for a two sample test for the equivalence of two probability densities.

4.5 Tests on Interactions

Model (4.1.1) contains an interactive term $M(X_{jim}, t)$ that is flexible in prescribing the interact between X_{jim} and the time, as long as the positive definite condition in Condition A3 is satisfied. In this section we propose tests for the presence of the interaction in the j -th treatment and the ANOVA hypothesis on the equivalence of the interactions among the treatments.

We firstly consider testing $H_{0c} : \gamma_{j0} = 0$ vs $H_{1c} : \gamma_{j0} \neq 0$ for a fixed j . In the formulation of the empirical likelihood for γ_{j0} , we treat $M_{jim} = M(X_{jim}, t)$ as a covariates with the same role like X_{jim} in the previous section when we constructed empirical likelihood for β_{j0} . For this purpose, we define estimating equations for γ_{j0}

$$\phi_{ji}(\gamma_{j0}) = \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{M}_{jim} (\tilde{Y}_{jim} - \tilde{X}_{jim}^{\tau} \tilde{\beta}_j - \tilde{M}_{jim}^{\tau} \gamma_{j0}), \quad (4.5.24)$$

where

$$\tilde{\beta}_j = \left\{ \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{X}_{jim} \tilde{X}_{jim}^{\tau} \right\}^{-1} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{X}_{jim} (\tilde{Y}_{jim} - \tilde{M}_{jim}^{\tau} \gamma_{j0}) \quad (4.5.25)$$

is the “estimator” of β_j at the true γ_{j0} . Similar to establishing $\ell_{n_j}(\beta_j)$ in Section 4.3, the log-empirical likelihood for γ_{j0} can be written as

$$\ell_{n_j}^{\gamma}(\gamma_j) = 2 \sum_{i=1}^{n_j} \log\{1 + \Lambda_j' \phi_{ji}(\gamma_j)\},$$

where the Lagrange multipliers Λ_j satisfies

$$\sum_{i=1}^{n_j} \frac{\phi_{ji}(\gamma_j)}{1 + \Lambda_j' \phi_{ji}(\gamma_j)} = 0. \quad (4.5.26)$$

To test for $H_{0d} : \gamma_{10} = \gamma_{20} = \dots = \gamma_{k0}$ vs $H_{1d} : \gamma_{i0} \neq \gamma_{j0}$ for some $i \neq j$, we construct the joint empirical likelihood ratio

$$\ell_n^{\gamma} : = 2 \min_{\gamma} \sum_{j=1}^k \sum_{i=1}^{n_j} \log\{1 + \Lambda_j^{\tau} \phi_{ji}(\gamma)\}, \quad (4.5.27)$$

where Λ_j satisfy (4.5.26).

The asymptotic distributions of the empirical likelihood ratios $\ell_{n_j}^{\gamma}(0)$ and ℓ_n^{γ} under the null hypotheses as given in the next theorem whose proofs will not be given as they follow the same routes in the proof of Theorem 1 by exchanging X_{jim} and β_{j0} with M_{jim} and γ_{j0} respectively.

Theorem 4 *Under Conditions (A1-A4) given in the Appendix, then (i) under H_{0c} , $\ell_{n_j}^\gamma(\mathbf{0}) \xrightarrow{d} \chi_q^2$ as $n_j \rightarrow \infty$; (ii) under H_{0d} , $\ell_n^\gamma \xrightarrow{d} \chi_{(k-1)q}^2$ as $n \rightarrow \infty$.*

Based on Theorem 4, an α -level empirical likelihood ratio test for the presence of the interaction in the j -th sample rejects H_{0c} if $\ell_{n_j}^\gamma(\mathbf{0}) > \chi_{q,\alpha}^2$, and the ANOVA test for the equivalence of the interactive effects rejects H_{0d} if $\ell_n^\gamma > \chi_{(k-1)q,\alpha}^2$. The ANOVA test for H_{0d} has a similar local power performance as that described after Theorem 2 for the ANOVA test regarding β_{j0} in Section 4.3. The power properties of the test for H_{0c} can be established using a much easier method.

We have assumed parametric models for the interaction in model (4.1.1). A semiparametric model would be employed to model the interaction given that the model for the time effect is nonparametric. The parametric interaction is a simplification and avoids some of the involved technicalities associated with a semiparametric model.

4.6 Bootstrap Calibration

To avoid direct estimation of σ_0^2 in Theorem 3 and to speed up the convergence of \mathcal{T}_n , we resort to the bootstrap. While the wild bootstrap (Wu 1986, Liu 1988 and Härdle and Mammen 1993) originally proposed for parametric regression without missing values has been modified by Shao and Sitter (1996) to take into account missing values, we extend it further to suit the longitudinal feature.

Let \vec{t}_j^o and \vec{t}_j^m be the sets of the time points with full and missing observations, respectively. According to model (4.2.3), we impute a missing $X_{ji}(t)$ from $\hat{X}_{ji}(t)$, $t \in \vec{t}_j^o$, so that for any $t \in \vec{t}_j^m$

$$\hat{X}_{ji}(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t) X_{jim}, \quad (4.6.28)$$

where $w_{jim,h_j}(t)$ is the kernel weight defined in (4.3.8).

To mimic the heteroscedastic and correlation structure in the longitudinal data, we estimate the covariance matrix for each subject in each treatment. Let

$$\hat{\varepsilon}_{jim} = Y_{jim} - \mathbb{X}_{jim}^\tau \hat{\xi}_j - \hat{g}_j(t_{jim}).$$

An estimator of $\sigma_j^2(t)$, the variance of $\varepsilon_{ji}(t)$, is $\hat{\sigma}_j^2(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t) \hat{\varepsilon}_{jim}^2$ and an estimator of $\rho_j(s, t)$, the correlation coefficient between $\varepsilon_{ji}(t)$ and $\varepsilon_{ji}(s)$ for $s \neq t$, is

$$\hat{\rho}_j(s, t) = \sum_{i=1}^{n_j} \sum_{m \neq m'}^{T_j} H_{jim,m'}(s, t) \hat{e}_{jim} \hat{e}_{jim'},$$

where $\hat{e}_{jim} = \hat{\varepsilon}_{jim} / \hat{\sigma}_j(t_{jim})$,

$$H_{jim,m'}(s, t) = \frac{\delta_{jim} \delta_{jim'} K_{b_j}(s - t_{jim}) K_{b_j}(t - t_{jim'}) / \pi_{jim,m'}(\hat{\theta}_j)}{\sum_{i=1}^{n_j} \sum_{m \neq m'} \delta_{jim} \delta_{jim'} K_{b_j}(s - t_{jim}) K_{b_j}(t - t_{jim'}) / \pi_{jim,m'}(\hat{\theta}_j)},$$

and $\pi_{jim,m'}(\hat{\theta}_j) = \pi_{jim}(\hat{\theta}_j) \pi_{jim'}(\hat{\theta}_j)$ if $|m - m'| > d$; $\pi_{jim,m'}(\hat{\theta}_j) = \pi_{jim_{m_b}}(\hat{\theta}_j)$ if $|m - m'| \leq d$ where $m_b = \max(m, m')$. Here b_j is a smoothing bandwidth which may be different from the bandwidth h_j for calculating the test statistics \mathcal{T}_n (Fan, Huang and Li 2007). Then, the covariance Σ_{ji} of $\varepsilon_{ji} = (\varepsilon_{ji1}, \dots, \varepsilon_{jiT_j})^\tau$ is estimated by $\hat{\Sigma}_{ji}$ which has $\hat{\sigma}_j^2(t_{jim})$ as its m -th diagonal element and $\hat{\rho}_j(t_{jik}, t_{jil}) \hat{\sigma}_j(t_{jik}) \hat{\sigma}_j(t_{jil})$ as its (k, l) -th element for $k \neq l$.

Let $Y_{ji}, \delta_{ji}, t_{ji}$ be the vector of random variables of the (j, i) -th subject, $X_{ji} = (X_{ji}(t_{ji1}), \dots, X_{ji}(t_{jiT_j}))^\tau$ and $g_{j0}(t_{sl}) = (g_{j0}(t_{sl1}), \dots, g_{j0}(t_{slT_k}))^\tau$, where s may be different from j . Let $X_{ji}^c = \{X_{ji}^o, \hat{X}_{ji}^m\}$, where X_{ji}^o contains observed $X_{ji}(t)$ for $t_j \in \bar{t}^o$ and \hat{X}_{ji}^m collects the imputed $X_{ji}(t)$ for $t \in \bar{t}_j^m$ according to (4.6.28). Plugging the value of X_{ji}^c , we get $M_{ji}^c = \{M_{ji}^o, \hat{M}_{ji}^m\}$, the observed and the imputed interactions for (j, i) -th subject, and then \mathbb{X}_{ji}^c .

The proposed bootstrap procedure consists of the following steps:

Step 1. Generate a bootstrap re-sample $\{Y_{ji}^*, \mathbb{X}_{ji}^c, \delta_{ji}^*, t_{ji}\}$ for the (j, i) -th subject by

$$Y_{ji}^* = \mathbb{X}_{ji}^c \tau \hat{\xi}_j + \hat{g}_1(t_{ji}) + \hat{\Sigma}_{ji} e_{ji}^*,$$

where e_{ji}^* 's are i.i.d. random vectors simulated from a distribution satisfying $E(e_{ji}^*) = 0$ and $\text{Var}(e_{ji}^*) = I_{T_j}$, $\delta_{jim}^* \sim \text{Bernoulli}(\pi_{jim}(\hat{\theta}_j))$ where $\hat{\theta}_j$ is estimated based on the original sample as given in (4.2.6). Here, $\hat{g}_1(t_{ji})$ is used as the common nonparametric time effect to mimic the null hypothesis H_{0b} .

Step 2. For each treatment j , we re-estimate ξ_j , θ_j and $g_j(t)$ based on the re-sample $\{Y_{ji}^*, \mathbb{X}_{ji}^c, \delta_{ji}^*, t_{ji}\}$ and denote them as $\hat{\xi}_j^*$, $\hat{\theta}_j^*$ and $\hat{g}_j^*(t)$. The bootstrap version of $R_{ji}\{g_1(t)\}$ is

$$R_{ji}^*\{\hat{g}_1(t)\} = \sum_{m=1}^{T_j} \frac{\delta_{jim}^*}{\pi_{jim}(\hat{\theta}_j^*)} K\left(\frac{t_{jim} - t}{h_j}\right) \left\{ Y_{jim}^* - \mathbb{X}_{jim}^{\tau} \hat{\xi}_j^* - \hat{g}_1(t) - \{\hat{g}_j^*(t_{jim}) - \hat{g}_j^*(t)\} \right\}$$

and use it to substitute $R_{ji}\{g_j(t)\}$ in the formulation of $\mathcal{L}_n(t)$, we obtain $\mathcal{L}_n^*(t)$ and then $\mathcal{T}_n^* = \int \mathcal{L}_n^*(t)\varpi(t)dt$.

Step 3. Repeat the above two steps B times for a large integer B and obtain B bootstrap values $\{\mathcal{T}_{nb}^*\}_{b=1}^B$. Let \hat{t}_α be the $1 - \alpha$ quantile of $\{\mathcal{T}_{nb}^*\}_{b=1}^B$, which is a bootstrap estimate of the $1 - \alpha$ quantile of \mathcal{T}_n . Then, we reject the null hypothesis H_{0b} if $\mathcal{T}_n > \hat{t}_\alpha$.

The following theorem justifies the bootstrap procedure.

Theorem 5 *Assume Conditions (A1-A4) in the Appendix hold and $h = O(n^{-1/5})$. Let \mathcal{X}_n denote the original sample, h and σ_0^2 be defined as in Theorem 3. The conditional distribution of $h^{-1/2}(\mathcal{T}_n^* - \mu_0)$ given \mathcal{X}_n converges to $N(0, \sigma_0^2)$ almost surely, namely,*

$$h^{-1/2}\{\mathcal{T}_n^* - (k - 1)\}|\mathcal{X}_n \xrightarrow{d} N(0, \sigma_0^2) \quad a.s.$$

4.7 Simulation Results

In this section, we report results from simulation studies which were designed to confirm the proposed ANOVA tests proposed in the previous sections. We simulated data from the following three-treatment model:

$$Y_{jim} = X_{jim}\beta_j + M_{jim}\gamma_j + g_j(t_{jim}) + \varepsilon_{jim} \quad \text{and} \quad X_{jim} = 2 - 1.5t_{jim} + u_{jim}, \quad (4.7.29)$$

where $M_{jim} = t_{jim} \times (X_{jim} - 1.5)^2$, $\varepsilon_{jim} = e_{ji} + \nu_{jim}$, $u_{jim} \sim N(0, \sigma_{a_j}^2)$, $e_{ji} \sim N(0, \sigma_{b_j}^2)$ and $\nu_{jim} \sim N(0, \sigma_{c_j}^2)$ for $j = \{1, 2, 3\}$, $i = 1, \dots, n_j$ and $m = 1, \dots, T_j$. This structure used to generate $\{\varepsilon_{jim}\}_{m=1}^{T_j}$ ensures dependence among the repeated measurements $\{Y_{jim}\}$ for each subject i . The correlation between Y_{jim} and Y_{jil} for any $m \neq l$ is $\sigma_{b_j}^2/(\sigma_{b_j}^2 + \sigma_{c_j}^2)$. The time points $\{t_{jim}\}_{m=1}^{T_j}$ were obtained by first independently generating uniform $[0, 1]$ random variables and then sorted in the ascending order. We set the number of repeated measures T_j to be the same, say T , for all three treatments; and chose $T = 5$ and 10 respectively. The standard deviation parameters in (4.7.29) were $\sigma_{a_1} = 0.5, \sigma_{b_1} = 0.5, \sigma_{c_1} = 0.2$ for the first treatment, $\sigma_{a_2} = 0.5, \sigma_{b_2} = 0.5, \sigma_{c_2} = 0.2$ for the second and $\sigma_{a_3} = 0.6, \sigma_{b_3} = 0.6, \sigma_{c_3} = 0.3$ for the third.

The parameters and the time effects for the three treatments were

Treatment 1: $\beta_1 = 2$, $\gamma_1 = 1$, $g_1(t) = 2 \sin(2\pi t) - \Delta_{1n}(t)$;

Treatment 2: $\beta_2 = 2 + D_{2n}$, $\gamma_2 = 1 + D_{2n}$, $g_2(t) = 2 \sin(2\pi t) - \Delta_{2n}(t)$;

Treatment 3: $\beta_3 = 2 + D_{3n}$, $\gamma_3 = 1 + D_{3n}$, $g_3(t) = 2 \sin(2\pi t) - \Delta_{3n}(t)$.

We designated different values of $D_{1n}, D_{2n}, D_{3n}, \Delta_{1n}(t), \Delta_{2n}(t)$, and $\Delta_{3n}(t)$ in the evaluation of the size and the power, whose details will be reported shortly.

We considered two missing data mechanisms. In the first mechanism (I), the missing propensity was

$$\text{logit}\{P(\delta_{jim} = 1 | \delta_{jim,m-1} = 1, X_{ji}, Y_{ji})\} = \theta_j X_{ji(m-1)} \quad \text{for } m > 1, \quad (4.7.30)$$

which is not dependent on the response Y , with $\theta_1 = 3, \theta_2 = 2$ and $\theta_3 = 2$. In the second mechanism (II),

$$\begin{aligned} & \text{logit}\{P(\delta_{jim} = 1 | \delta_{jim,m-1} = 1, X_{ji}, Y_{ji})\} \\ &= \begin{cases} \theta_{j1} X_{ji(m-1)} + \theta_{j2} \{Y_{ji(m-1)} - Y_{ji(m-2)}\}, & \text{if } m > 2, \\ \theta_{j1} X_{ji(m-1)}, & \text{if } m = 2; \end{cases} \end{aligned} \quad (4.7.31)$$

which is influenced by both covariate and response, with $\theta_1 = (\theta_{11}, \theta_{12})^\tau = (2, -1)^\tau, \theta_2 = (\theta_{21}, \theta_{22})^\tau = (2, -1.5)^\tau$ and $\theta_3 = (\theta_{31}, \theta_{32})^\tau = (2, -1.5)^\tau$. In both mechanisms, the first observation ($m = 1$) for each subject was always observed as we have assumed earlier.

We used the Epanechnikov kernel $K(u) = 0.75(1 - u^2)_+$ throughout the simulation where $(\cdot)_+$ stands for the positive part of a function. The bandwidths were chosen by the ‘leave-one-subject’ out cross-validation. Specifically, we chose the bandwidth h_j that minimized the cross-validation score functions

$$\sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} (Y_{jim} - X_{jim}^\tau \hat{\beta}_j^{(-i)} - M_{jim}^\tau \hat{\gamma}_j^{(-i)} - \hat{g}_j^{(-i)}(t_{jim}))^2,$$

where $\hat{\beta}_j^{(-i)}, \hat{\gamma}_j^{(-i)}$ and $\hat{g}_j^{(-i)}(t_{jim})$ were the corresponding estimates without using observations of the i -th subject. The cross-validation was used to choose an optimal bandwidth for representative data sets and fixed the chosen bandwidths in the simulations with the same sample size. We fixed the number of simulations to be 500.

The average missing percentages based on 500 simulations for the missing mechanism I were 8%, 15% and 17% for Treatments 1-3 respectively when $T = 5$, and were 16%, 28% and 31% when $T = 10$. In the missing mechanism II, the average missing percentages were 10%, 8% and 15% for $T = 5$, and 23%, 20% and 36% for $T = 10$, respectively.

For the ANOVA test for $H_{0a} : \beta_{10} = \beta_{20} = \beta_{30}$ with respect to the covariate effects, three values of D_{2n} and D_{3n} : 0, 0.2 and 0.3, were used respectively, while $\Delta_{1n}(t) = \Delta_{2n}(t) = \Delta_{3n}(t) = 0$. Table 4.1 summarizes the empirical size and power of the proposed EL ANOVA test with 5% nominal significant level for H_{0a} for 9 combinations of (D_{2n}, D_{3n}) , where the sizes corresponding to $D_{2n} = 0$ and $D_{3n} = 0$. We observed that the size of the ANOVA tests improved as the sample sizes and the observational length T increased, and the overall level of size were close to the nominal 5%. This is quite re-assuring considering the ANOVA test is based on the asymptotic chi-square distribution. We also observed that the power of the test increased as sample sizes and T were increased, and as the distance among the three β_{j0} was increased. For example, when $D_{2n} = 0.0$ and $D_{3n} = 0.3$, the L_2 distance was $\sqrt{0.3^2 + 0.3^2} = 0.424$, which is larger than $\sqrt{0.1^2 + 0.2^2 + 0.3^2} = 0.374$ for $D_{2n} = 0.2$ and $D_{3n} = 0.3$. This explains why the ANOVA test was more powerful for $D_{2n} = 0.0$ and $D_{3n} = 0.3$ than $D_{2n} = 0.2$ and $D_{3n} = 0.3$. At the same time, we see similar power performance between the two missing mechanisms.

To gain information on the empirical performance of the test on the existence of interaction, we carried out a test for $H_{0c} : \gamma_{20} = 0$. In the simulation, we chose $\gamma_{20} = 0, 0.2, 0.3, 0.4$, $\beta_{20} = 2 + \gamma_{20}$ and fixed $\Delta_{2n}(t) = 0$ respectively. Table 4.2 summarizes the sizes and the powers of the test. Table 4.3 reports the simulation results of the ANOVA test on the interaction effect $H_{0d} : \gamma_{10} = \gamma_{20} = \gamma_{30}$ with a similar configurations as those used as the ANOVA tests for the covarites effects reported in Table 4.1. We observe satisfactory performance of these two tests in terms of both the accurate of the size approximation and the empirical power. In particular, the performance of the ANOVA tests were very much similar to that conveyed in Table 4.1.

We then evaluate the power and size of the proposed ANOVA test regarding the nonparametric components. To study the local power of the test, we set $\Delta_{2n}(t) = U_n \sin(2\pi t)$ and $\Delta_{3n}(t) = 2 \sin(2\pi t) - 2 \sin(2\pi(t + V_n))$, and fixed $D_{2n} = 0$ and $D_{3n} = 0.2$. Here U_n and V_n were designed to adjust the amplitude and phase of the sine function. The same kernel and

Table 4.1 Empirical size and power of the 5% ANOVA test for $H_{0a} : \beta_{10} = \beta_{20} = \beta_{30}$.

Sample Size						Missingness		Missingness		
n_1	n_2	n_3	D_{2n}	D_{3n}	T	I	II	T	I	II
60	65	55	0.0	0.0 (size)	5	0.042	0.050	10	0.046	0.044
			0.2	0.0		0.192	0.254		0.408	0.434
			0.3	0.0		0.548	0.630		0.810	0.864
			0.0	0.2		0.236	0.214		0.344	0.354
			0.0	0.3		0.508	0.546		0.714	0.722
			0.2	0.2		0.208	0.262		0.446	0.458
			0.2	0.3		0.412	0.440		0.680	0.698
			0.3	0.2		0.426	0.490		0.728	0.728
			0.3	0.3		0.594	0.620		0.836	0.818
100	110	105	0.0	0.0 (size)	5	0.052	0.054	10	0.042	0.038
			0.2	0.0		0.426	0.470		0.686	0.718
			0.3	0.0		0.854	0.854		0.964	0.974
			0.0	0.2		0.406	0.444		0.612	0.568
			0.0	0.3		0.816	0.836		0.936	0.910
			0.2	0.2		0.404	0.480		0.674	0.686
			0.2	0.3		0.744	0.694		0.944	0.882
			0.3	0.2		0.712	0.768		0.922	0.920
			0.3	0.3		0.824	0.814		0.972	0.970

bandwidths chosen by the cross-validation as outlined earlier in the parametric ANOVA test were used in the test for the nonparametric time effects. We calculated the test statistic \mathcal{T}_n with $\varpi(t)$ being the kernel density estimate based on all the time points in all treatments. We applied the wild bootstrap proposed in Section 4.5 with $B = 100$ to obtain $\hat{t}_{0.05}$, the bootstrap estimator of the 5% critical value. The simulation results of the nonparametric ANOVA test for the time effects are given in Table 4.4.

The sizes of the nonparametric ANOVA test were obtained when $U_n = 0$ and $V_n = 0$, which were quite close to the nominal 5%. We observe that the power of the test increased when the distance among $g_1(\cdot)$, $g_2(\cdot)$ and $g_3(\cdot)$ were becoming larger, and when the sample size or repeated measurement T were increased. We noticed that the power was more sensitive to change in V_n , the initial phase of the sine function, than U_n .

We then compared the proposed tests with a test proposed by Scheike and Zhang (1998).

Table 4.2 Empirical size and power of the 5% test for the existence of interaction $H_{0c} : \gamma_{20} = 0$.

Sample Size			γ_{20}	T	Missingness		T	Missingness	
n_1	n_2	n_3			I	II		I	II
60	65	55	0.0 (size)	5	0.052	0.048	10	0.048	0.052
			0.2		0.428	0.456		0.568	0.636
			0.3		0.722	0.788		0.848	0.882
			0.4		0.928	0.952		0.948	0.968
100	110	105	0.0 (size)	5	0.054	0.046	10	0.056	0.042
			0.2		0.608	0.718		0.694	0.812
			0.3		0.940	0.938		0.940	0.958
			0.4		0.986	0.994		0.952	0.966

Scheike and Zhang's test was comparing two treatments for the nonparametric regression model (4.4.23) for longitudinal data without missing values. Their test was based on a cumulative statistic

$$T(z) = \int_a^z (\hat{g}_1(t) - \hat{g}_2(t)) dt,$$

where a, z are in a common time interval $[0, 1]$. They showed that $\sqrt{n_1 + n_2}T(z)$ converges to a Gaussian Martingale with mean 0 and variance function $\rho_1^{-1}h_1(z) + \rho_2^{-1}h_2(z)$, where $h_j(z) = \int_a^z \sigma_j^2(y) f_j^{-1}(y) dy$. Hence, the test statistic $T(1-a)/\sqrt{\widehat{\text{Var}}\{T(1-a)\}}$ is used for two group time-effect functions comparison.

To make the proposed test and the test of Scheike and Zhang (1998) comparable, we conducted simulation in a set-up that mimics the setting of model (4.7.29) but with only the first two treatments, no missing values and only the nonparametric part in the regression by setting $\beta_j = \gamma_j = 0$. Specifically, we test for $H_0 : g_1(\cdot) = g_2(\cdot)$ vs $H_1 : g_1(\cdot) = g_2(\cdot) + \Delta_{2n}(\cdot)$ for three cases of the alternative shift function $\Delta_{2n}(\cdot)$ functions which are spelt out in Table 4.5 and set $a = 0$ in the test of Scheike and Zhang. The simulation results are summarized in Table 4.5. We found that in the first two cases (I and II) of the alternative shift function Δ_{2n} , the test of Scheike and Zhang had little power. It was only in the third case (III), the test started to pick up some power although it was still not as powerful as the proposed test.

Table 4.3 Empirical size and power of the 5% ANOVA test for $H_{0d} : \gamma_{10} = \gamma_{20} = \gamma_{30}$.

Sample Size					T	Missingness		T	Missingness	
n_1	n_2	n_3	D_{2n}	D_{3n}		I	II		I	II
60	65	55	0.0	0.0 (size)	5	0.058	0.058	10	0.068	0.036
			0.2	0.0		0.134	0.188		0.232	0.254
			0.3	0.0		0.358	0.486		0.510	0.622
			0.0	0.2		0.136	0.166		0.230	0.218
			0.0	0.3		0.356	0.414		0.466	0.474
			0.2	0.2		0.170	0.208		0.286	0.276
			0.2	0.3		0.292	0.328		0.462	0.428
			0.3	0.2		0.266	0.356		0.498	0.474
			0.3	0.3		0.392	0.476		0.578	0.588
100	110	105	0.0	0.0 (size)	5	0.068	0.040	10	0.054	0.046
			0.2	0.0		0.262	0.366		0.354	0.432
			0.3	0.0		0.654	0.744		0.744	0.820
			0.0	0.2		0.272	0.330		0.340	0.334
			0.0	0.3		0.590	0.676		0.722	0.672
			0.2	0.2		0.282	0.332		0.412	0.410
			0.2	0.3		0.528	0.582		0.716	0.640
			0.3	0.2		0.502	0.580		0.680	0.728
			0.3	0.3		0.672	0.674		0.814	0.808

4.8 Analysis on HIV-CD4 Data

In this section, we analyzed a longitudinal data set from AIDS Clinical Trial Group 193A Study (Henry *et al.* 1998), which was a randomized, double-blind study of HIV-AIDS patients with advanced immune suppression. The study was carried out in 1993 with 1309 patients who were randomized to four treatments with regard to HIV-1 reverse transcriptase inhibitors. Patients were randomly assigned to one of four daily treatment regimes: 600mg of zidovudine alternating monthly with 400mg didanosine (Treatment I); 600mg of zidovudine plus 2.25mg of zalcitabine (Treatment II); 600mg of zidovudine plus 400mg of didanosine (Treatment III); or 600mg of zidovudine plus 400mg of didanosine plus 400mg of nevirapine (Treatment VI). The four treatments had 325, 324, 330 and 330 patients respectively.

The aim of our analysis was to compare the effects of age (Age), baseline CD4 counts

Table 4.4 Empirical size and power of the 5% ANOVA test for $H_{0b} : g_1(\cdot) = g_2(\cdot) = g_3(\cdot)$ with $\Delta_{2n}(t) = U_n \sin(2\pi t)$ and $\Delta_{3n}(t) = 2 \sin(2\pi t) - 2 \sin(2\pi(t + V_n))$

Sample Size					T	Missingness		T	Missingness	
n_1	n_2	n_3	U_n	V_n		I	II		I	II
60	65	55	0.00	0.00(size)	5	0.040	0.050	10	0.054	0.060
			0.30	0.00		0.186	0.232		0.282	0.256
			0.50	0.00		0.666	0.718		0.828	0.840
			0.00	0.05		0.664	0.726		0.848	0.842
			0.00	0.10		1.000	1.000		1.000	1.000
100	110	105	0.00	0.00(size)	5	0.032	0.062	10	0.050	0.036
			0.30	0.00		0.434	0.518		0.526	0.540
			0.50	0.00		0.938	0.980		0.992	0.998
			0.00	0.05		0.916	0.974		1.000	1.000
			0.00	0.10		1.000	1.000		1.000	1.000

(PreCD4), and gender (Gender) on $Y = \log(\text{CD4 counts} + 1)$. The semiparametric model regression is, for $j = 1, 2, 3$ and 4,

$$Y_{ji}(t) = \beta_{j1}\text{Age}_{ji} + \beta_{j2}\text{PreCD4}_{ji} + \beta_{j3}\text{Gender}_{ji} + g_j(t) + \varepsilon_{ji}(t), \quad (4.8.32)$$

with the intercepts absorbed in the nonparametric $g_j(\cdot)$ functions, and $\beta_j = (\beta_{j1}, \beta_{j2}, \beta_{j3})^T$ is the regression coefficients to the covariates (Age, PreCD4, Gender).

To make $g_j(t)$ more interpretable, we centralized Age and PreCD4 so that their sample means in each treatment were 0, respectively. As a result, $g_j(t)$ can be interpreted as the baseline evolution of Y for a female (Gender=0) with the average PreCD4 counts and the average age in Treatment j . This kind of normalization is used in Wu and Chiang (2000) in their analyzes for another CD4 data set. Our objectives were to detect any difference in the treatments with respect to (i) the covariates; and (ii) the nonparametric baseline functions.

Measurements of CD4 counts were scheduled at the start time 1 and at a 8-week intervals during the follow-up. However, the data were unbalanced due to variations from the planned measurement time and missing values resulted from skipped visits and dropouts. The number of CD4 measurements for patients during the first 40 weeks of follow-up varied from 1 to 9, with a median of 4. There were 5036 complete measurements of CD4, and 2826 scheduled

Table 4.5 The empirical sizes and powers of the proposed test (CZ) and the test (SZ) proposed by Scheike and Zhang (1998) for $H_{0b} : g_1(\cdot) = g_2(\cdot)$ vs $H_{1b} : g_1(\cdot) = g_2(\cdot) + \Delta_{2n}(\cdot)$.

Sample Size			U_n	T	Tests		T	Tests	
n_1	n_2	n_3			CZ	SZ		CZ	SZ
60	65	55	Case I: $\Delta_{2n}(t) = U_n \sin(2\pi t)$						
			0.00(size)	5	0.060	0.032	10	0.056	0.028
			0.30		0.736	0.046		0.844	0.028
			0.50		1.000	0.048		1.000	0.026
			Case II: $\Delta_{2n}(t) = 2 \sin(2\pi t) - 2 \sin(2\pi(t + U_n))$						
			0.05		1.000	0.026		1.000	0.042
			0.10		1.000	0.024		1.000	0.044
			Case III: $\Delta_{2n}(t) = -U_n$						
			0.10		0.196	0.162		0.206	0.144
			0.20		0.562	0.514		0.616	0.532
100	110	105	Case I: $\Delta_{2n}(t) = U_n \sin(2\pi t)$						
			0.00(size)	5	0.056	0.028	10	0.042	0.018
			0.30		0.982	0.038		0.994	0.040
			0.50		1.000	0.054		1.000	0.028
			Case II: $\Delta_{2n}(t) = 2 \sin(2\pi t) - 2 \sin(2\pi(t + U_n))$						
			0.05		1.000	0.022		1.000	0.030
			0.10		1.000	0.026		1.000	0.030
			Case III: $\Delta_{2n}(t) = -U_n$						
			0.10		0.290	0.260		0.294	0.218
			0.20		0.780	0.774		0.760	0.730

measurements were missing. Hence, considering missing values is very important in this analysis. Most of the missing values follow the monotone pattern. Therefore, we firstly model the missing mechanism under the monotone assumption in next subsection.

4.8.1 Monotone Missingness

We considered three logistic regression models for the missing propensities and used the AIC and BIC criteria to select the one that was the mostly supported by data. The first model (M1) was a logistic regression model for $p_j(\overleftarrow{X}_{jit,3}, \overleftarrow{Y}_{jit,3}; \theta_{j0})$ that effectively depends on X_{jit} (the PreCD4) and $(Y_{ji(t-1)}, Y_{ji(t-2)}, Y_{ji(t-3)})$ if $t > 3$. For $t < 3$, it relies on all Y_{jit} observed before t . In the second model (M2), we replace the X_{jit} in the first model with an intercept.

In the third model (M3), we added to the second logistic model with covariates representing the square of $Y_{ji(t-1)}$ and the interactions between $Y_{ji(t-1)}$ and $Y_{ji(t-2)}$. In the formulation of the AIC and BIC criteria we used the binary conditional likelihood given in (4.2.6) with the respective penalties. The difference of AIC and BIC values among these models for four treatment groups is given in Table 4.6. Under the BIC criterion, M2 was the best model for all four treatments. For Treatments II and III, M3 had smaller AIC values than M2, but the differences were very small. For Treatments I and VI, M2 had smaller AIC than M3. As the AIC tends to select more explanatory variables, we chose M2 as the model for the parametric missing propensity.

Table 4.6 Differences in the AIC and BIC scores among three models (M1-M3)

Models	Treatment I		Treatment II		Treatment III		Treatment VI	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
M1-M2	3.85	3.85	14.90	14.90	17.91	17.91	10.35	10.35
M2-M3	-2.47	-11.47	0.93	-8.12	0.30	-8.75	-3.15	-12.27

Model (4.8.32) does not have interactions. It is interesting to check if there is an interaction between gender and time. Then the model becomes

$$Y_{ji}(t) = \beta_{j1}\text{Age}_{ji} + \beta_{j2}\text{PreCD4}_{ji} + \beta_{j3}\text{Gender}_{ji} + \gamma_{j4}\text{Gender}_{ji} \times t + g_j(t) + \varepsilon_{ji}(t), \quad (4.8.33)$$

We applied the proposed test in Section 4.5 for $H_{0c} : \gamma_{j4} = 0$ for $j = 1, 2, 3$ and 4 respectively. The p-values were 0.9234, 0.9885, 0.9862 and 0.5558 respectively, which means that the interaction was not significant. Therefore, in the following analyzes, we would not include the interaction term and continue to use Model (4.8.32).

Table 4.7 reports the parameter estimates $\hat{\beta}_j$ of β_j based on the estimating function $Z_{ji}(\beta_j)$ given in Section 4.3. It contains the standard errors of the estimates, which were obtained from the length of the EL confidence intervals based on the marginal empirical likelihood ratio for each β_j as proposed in Chen and Hall (1994). In getting these estimates, we use the ‘leave-one-subject’ cross-validation (Rice and Silverman 1991) to select the smoothing bandwidths

$\{h_j\}_{j=1}^4$ for the four treatments, which were 12.90, 7.61, 8.27 and 16.20 respectively. We see that the estimates of the coefficients for the Age and PreCD4 were similar among all four treatments with comparable standard errors, respectively. In particular, the estimates of the Age coefficients endured large variations while the estimates of the PreCD4 coefficients were quite accurate. However, estimates of the Gender coefficients had different signs among the treatments. We may also notice that the confidence intervals from treatments I-IV for each coefficient were overlap.

Table 4.7 Parameter estimates and their standard errors

Coefficients	Treatment I	Treatment II	Treatment III	Treatment VI
	β_1	β_2	β_3	β_4
Age	0.0063(0.0039)	0.0050(0.0040)	0.0047(0.0058)	0.0056(0.0046)
PreCD4	0.7308(0.0462)	0.7724(0.0378)	0.7587(0.0523)	0.8431(0.0425)
Gender	0.1009(0.0925)	0.1045(0.0920)	-0.3300(0.1510)	-0.3055(0.1136)

We then formally tested $H_{0a} : \beta_1 = \beta_2 = \beta_3 = \beta_4$. The empirical likelihood ratio statistic ℓ_n was 8.1348, which was smaller than $\chi_{9,0.95}^2 = 16.9190$, which produced a p-value of 0.5206. So we do not have enough evidence to reject H_{0a} at a significant level 5 %. The parameter estimates reported in Table 4.7 suggested similar covariate effects between Treatments I and II, and between Treatments III and IV, respectively; but different effects between the first two treatments and the last two treatments. To verify this suggestion, we carry out formal ANOVA test for pair-wise equality among the β_j 's as well as for equality of any three β_j 's. The p-values of these ANOVA test are reported in Table 4.8. Indeed, the difference between the first two treatments and between the last two treatments were insignificant. However, the differences between the first three (I, II and III) treatments and the last treatment were also not significant.

We then tested for the nonparametric baseline time effects. The kernel estimates $\hat{g}_j(t)$ are displayed in Figure 4.1, which shows that Treatments I and II and Treatments III and IV had similar baselines evolution overtime, respectively. However, a big difference existed between

Table 4.8 P-values of ANOVA tests for β_j s.

H_{0a}	p-value	H_{0a}	p-value
$\beta_1 = \beta_2$	0.9661	$\beta_1 = \beta_2 = \beta_3$	0.7399
$\beta_1 = \beta_3$	0.4488	$\beta_1 = \beta_2 = \beta_4$	0.4011
$\beta_1 = \beta_4$	0.1642	$\beta_1 = \beta_3 = \beta_4$	0.3846
$\beta_2 = \beta_3$	0.4332	$\beta_2 = \beta_3 = \beta_4$	0.4904
$\beta_2 = \beta_4$	0.2523	$\beta_1 = \beta_2 = \beta_3 = \beta_4$	0.5206
$\beta_3 = \beta_4$	0.8450		

the first two treatments and the last two treatments. Treatment IV decreased more slowly than that of the other three treatments, which seemed to be the most effective in slowing down the decline of CD4. We also found that during the first 16 weeks the CD4 counts decrease slowly and then the decline became faster after 16 weeks for Treatments I, II and III.

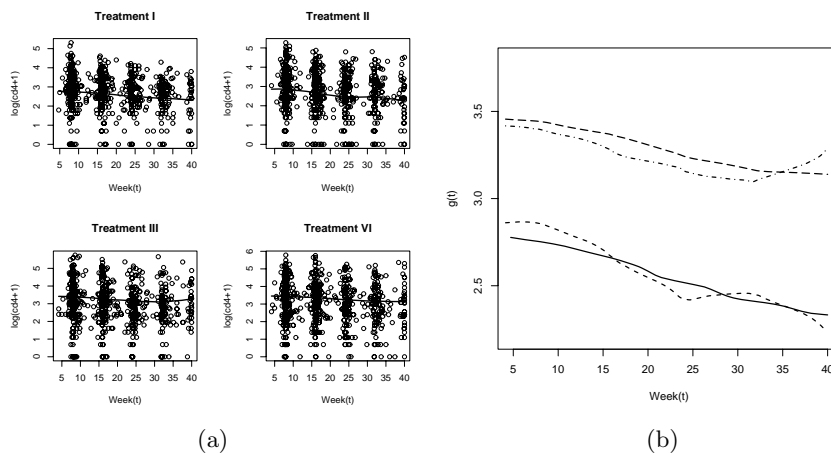


Figure 4.1 (a) The raw data plots with the estimates of $g_j(t)$ ($j = 1, 2, 3, 4$). (b) The estimates of $g_j(t)$ in the same plot: Treatment I (solid line), Treatment II (short dashed line), Treatment III (dashed and dotted line) and Treatment IV (long dashed line).

The p-value for testing $H_{0b} : g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$ is shown in Table 4.9. The entries were based on 500 bootstrapped resamples according to the procedure introduced in Section 4.6. The statistics \mathcal{T}_n for testing $H_{0b} : g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$ was 3965.00, where we take $\varpi(t) = 1$ over the range of t . The p-value of the test was 0.004. Thus, there

existed significant difference in the baseline time effects $g_j(\cdot)$'s among Treatments I-IV. At the same time, we also calculate the test statistics \mathcal{T}_n for testing $g_1(\cdot) = g_2(\cdot)$ and $g_3(\cdot) = g_4(\cdot)$. The statistics values were 19.26 and 26.22, with p-values 0.894 and 0.860, respectively. These p-values are much bigger than 0.05. We conclude that treatment I and II has similar baseline time effects, but they are significantly distinct from the baseline time effects of treatment III and IV, respectively. P-values of testing other combinations on equalities of $g_1(\cdot), g_2(\cdot), g_3(\cdot)$ and $g_4(\cdot)$ are also reported in Table 4.9.

Table 4.9 P-values of ANOVA tests on $g_j(\cdot)$ s.

H_{0b}	p-value	H_{0b}	p-value
$g_1(\cdot) = g_2(\cdot)$	0.894	$g_1(\cdot) = g_2(\cdot) = g_3(\cdot)$	0.046
$g_1(\cdot) = g_3(\cdot)$	0.018	$g_1(\cdot) = g_2(\cdot) = g_4(\cdot)$	0.010
$g_1(\cdot) = g_4(\cdot)$	0.004	$g_1(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.000
$g_2(\cdot) = g_3(\cdot)$	0.020	$g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.014
$g_2(\cdot) = g_4(\cdot)$	0.006	$g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.004
$g_3(\cdot) = g_4(\cdot)$	0.860		

This data set has been analyzed by Fitzmaurice, Laird and Ware (2004) using a random effects model that applied the Restricted Maximum Likelihood (REML) method. They conducted a two sample comparison test via parameters in the model for the difference between the dual therapy (Treatment I-III) versus triple therapy (Treatment VI) without considering the missing values. More specifically, they denoted Group = 1 if subject in the triple therapy treatment and Group = 0 if subject in the dual therapy treatment, and the linear mixed effect was

$$E(Y|b) = \beta_1 + \beta_2 t + \beta_3(t - 16)_+ + \beta_4 \text{Group} \times t \\ + \beta_5 \text{Group} \times (t - 16)_+ + b_1 + b_2 t + b_3(t - 16)_+,$$

where $b = (b_1, b_2, b_3)$ are random effects. They tested $H_0 : \beta_4 = \beta_5 = 0$. This is equivalent to test the null hypothesis of no treatment group difference in the changes in log CD4 counts between therapy and dual treatments. Both Wald test and likelihood ratio test rejected the

null hypothesis, indicating the difference between dual and triple therapy in the change of log CD4 counts. Their results are consistent with the result we illustrated in Table 4.9.

4.8.2 Not-monotone Missingness

We also analyzed the data without assuming monotone missingness for the missing values in this subsection. Instead of monotone assumption, we assume the missing propensity depends on the past $d(t)$ observations for a given time t as we described at Section 4.2. Recall that from Section 4.2, if we assume small d for the missing propensity function, more data could be used for analysis than monotone assumption. We presented the results for $d = 1, 2, 3$ in this subsection.

For $d = 1$, three logistic models were used to model the missing propensity functions. In the first model (M1) we include intercept, PreCD4, and $Y_{ji(t-1)}$ as covariates. In the second model (M2), only intercept and $Y_{ji(t-1)}$ are included. In the third model, we used a nonlinear model with intercept, $Y_{ji(t-1)}$, $Y_{ji(t-1)}^2$, and PreCD4 as covariates. As we did in previous monotone case, AIC and BIC value differences among M1-M3 are reported in the following Table 4.10. We observed that model M1 had the smallest AIC for four treatments among M1-M3. M1 also had the smaller BIC values than M3, for Treatment II-IV, M2 had slightly smaller BIC values. So, overall we would choose M1 to model the missing propensity. For $d = 2$ and $d = 3$, we chose the missing propensity function in a similar way, but we do not report the AIC and BIC values here for saving space.

Table 4.10 Differences in the AIC and BIC scores among three models (M1-M3) for $d = 1$.

Models	Treatment I		Treatment II		Treatment III		Treatment VI	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
M1-M2	-7.885	-2.992	-3.870	1.039	-3.519	1.365	-0.065	5.020
M2-M3	6.506	-3.281	2.125	-7.693	2.278	-7.491	-1.368	-11.155

Table 4.11 reports the parameter estimates and their corresponding standard errors. The estimates for the coefficient of PrdCD4 are very much similar for $d = 1, 2, 3$, but the estimates

for the coefficient of Age and Gender seem more variable among $d = 1, 2, 3$. Nevertheless, all of the estimates at one d value are in the 95% confidence interval of the estimates at another d value. For example, the 95% confidence interval for PreCD4 in Treatment I with $d = 1$ is (0.6812, 0.8368) and the corresponding estimates with $d = 2, 3$ are in this confidence interval. Basically, we may say the estimates at $d = 1, 2, 3$ are not significantly different among four treatments.

Table 4.11 Parameter estimates and their standard errors with $d = 1, 2, 3$

Coefficients		Treatment I	Treatment II	Treatment III	Treatment VI
		β_1	β_2	β_3	β_4
$d = 1$	Age	0.0036(0.0031)	0.0061(0.0036)	0.0039(0.0043)	0.0059(0.0037)
	PreCD4	0.7590(0.0389)	0.7440(0.0339)	0.7735(0.0441)	0.8441(0.0334)
	Gender	0.0650(0.0874)	0.0343(0.1082)	-0.1941(0.1208)	-0.1892(0.0790)
$d = 2$	Age	0.0065(0.0037)	0.0059(0.0042)	0.0002(0.0053)	0.0050(0.0041)
	PreCD4	0.7538(0.0429)	0.7282(0.0360)	0.7574(0.0443)	0.8409(0.0392)
	Gender	0.0309(0.0895)	0.0318(0.1075)	-0.2134(0.1282)	-0.3019(0.0910)
$d = 3$	Age	0.0054(0.0036)	0.0049(0.0040)	0.0056(0.0053)	0.0044(0.0042)
	PreCD4	0.7540(0.0443)	0.7666(0.0368)	0.7607(0.0482)	0.8476(0.0406)
	Gender	0.0716(0.0955)	0.0942(0.0930)	-0.2776(0.1294)	-0.2527(0.1081)

Next, we summarize the ANOVA test results on β s with $d = 1, 2, 3$ at Table 4.12. The p-values are consistent in the sense that the order of the p-values at different d values were almost the same. For instance, the test for $\beta_2 = \beta_4$ always had the smallest p-value among all the p-values with same d . The tests among β_1, β_2 and β_4 had smaller p-values than the other tests. All the test results showed some similarity treatment effects due to covariates among Treatments I-III (dual therapy treatments) and difference comparing to Treatment IV (triple therapy treatments), but there were not significant.

Finally, Table 4.13 illustrate the ANOVA test for the nonparametric baseline time effect functions. The p-values were obtained from the bootstrap calibration test we introduced in Section 4.5. Each p-value were based on 500 times resampling. The bandwidth selection method and the weight function $\varpi(t)$ are the same with the monotone case. We found that

Table 4.12 P-values of ANOVA tests on β s with $d = 1, 2, 3$

	$H_{0\alpha}$	p-value	$H_{0\alpha}$	p-value
$d = 1$	$\beta_1 = \beta_2$	0.9810	$\beta_1 = \beta_2 = \beta_3$	0.9513
	$\beta_1 = \beta_3$	0.7123	$\beta_1 = \beta_2 = \beta_4$	0.6240
	$\beta_1 = \beta_4$	0.3855	$\beta_1 = \beta_3 = \beta_4$	0.7032
	$\beta_2 = \beta_3$	0.7776	$\beta_2 = \beta_3 = \beta_4$	0.7006
	$\beta_2 = \beta_4$	0.3322	$\beta_1 = \beta_2 = \beta_3 = \beta_4$	0.8208
	$\beta_3 = \beta_4$	0.8362		
$d = 2$	$\beta_1 = \beta_2$	0.9910	$\beta_1 = \beta_2 = \beta_3$	0.9404
	$\beta_1 = \beta_3$	0.6640	$\beta_1 = \beta_2 = \beta_4$	0.4341
	$\beta_1 = \beta_4$	0.2654	$\beta_1 = \beta_3 = \beta_4$	0.5296
	$\beta_2 = \beta_3$	0.7375	$\beta_2 = \beta_3 = \beta_4$	0.5465
	$\beta_2 = \beta_4$	0.2102	$\beta_1 = \beta_2 = \beta_3 = \beta_4$	0.6614
	$\beta_3 = \beta_4$	0.7282		
$d = 3$	$\beta_1 = \beta_2$	0.9967	$\beta_1 = \beta_2 = \beta_3$	0.8562
	$\beta_1 = \beta_3$	0.5803	$\beta_1 = \beta_2 = \beta_4$	0.5230
	$\beta_1 = \beta_4$	0.2916	$\beta_1 = \beta_3 = \beta_4$	0.5927
	$\beta_2 = \beta_3$	0.5399	$\beta_2 = \beta_3 = \beta_4$	0.5723
	$\beta_2 = \beta_4$	0.2762	$\beta_1 = \beta_2 = \beta_3 = \beta_4$	0.6891
	$\beta_3 = \beta_4$	0.8404		

the p-values when $d = 3$ are quite similar to the monotone case.

Table 4.13 P-values of ANOVA tests on $g_j(\cdot)$ s with $d = 1, 2, 3$

	H_{0b}	p-value	H_{0b}	p-value
$d = 1$	$g_1(\cdot) = g_2(\cdot)$	0.750	$g_1(\cdot) = g_2(\cdot) = g_3(\cdot)$	0.224
	$g_1(\cdot) = g_3(\cdot)$	0.068	$g_1(\cdot) = g_2(\cdot) = g_4(\cdot)$	0.070
	$g_1(\cdot) = g_4(\cdot)$	0.026	$g_1(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.038
	$g_2(\cdot) = g_3(\cdot)$	0.110	$g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.058
	$g_2(\cdot) = g_4(\cdot)$	0.016	$g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.056
	$g_3(\cdot) = g_4(\cdot)$	0.550		
$d = 2$	$g_1(\cdot) = g_2(\cdot)$	0.896	$g_1(\cdot) = g_2(\cdot) = g_3(\cdot)$	0.358
	$g_1(\cdot) = g_3(\cdot)$	0.154	$g_1(\cdot) = g_2(\cdot) = g_4(\cdot)$	0.036
	$g_1(\cdot) = g_4(\cdot)$	0.016	$g_1(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.054
	$g_2(\cdot) = g_3(\cdot)$	0.216	$g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.106
	$g_2(\cdot) = g_4(\cdot)$	0.048	$g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.046
	$g_3(\cdot) = g_4(\cdot)$	0.446		
$d = 3$	$g_1(\cdot) = g_2(\cdot)$	0.886	$g_1(\cdot) = g_2(\cdot) = g_3(\cdot)$	0.044
	$g_1(\cdot) = g_3(\cdot)$	0.016	$g_1(\cdot) = g_2(\cdot) = g_4(\cdot)$	0.010
	$g_1(\cdot) = g_4(\cdot)$	0.002	$g_1(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.004
	$g_2(\cdot) = g_3(\cdot)$	0.042	$g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.026
	$g_2(\cdot) = g_4(\cdot)$	0.014	$g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$	0.004
	$g_3(\cdot) = g_4(\cdot)$	0.812		

4.9 Appendix: Technical Details

We provide the conditions used for Theorems 1-5 and some remarks in this Appendix. The proofs for Theorems 1, 2, 3 and 5 are also contained in this Appendix. The proof for Theorem 4 is largely similar to that of Theorem 1 and hence is omitted.

The following assumptions are made in this Chapter:

- A1. Let $S(\theta_j)$ be the score function of the partial likelihood $\mathcal{L}_{B_j}(\theta_j)$ for a q -dimensional parameter θ_j defined in (4.2.6), and θ_{j0} is in the interior of compact Θ_j . We assume $E\{S(\theta_j)\} \neq 0$ if $\theta_j \neq \theta_{j0}$, $\text{Var}(S(\theta_{j0}))$ is finite and positive definite, and $E\left(\frac{\partial S(\theta_{j0})}{\partial \theta_{j0}}\right)$ exists and is invertible. The missing propensity $\pi_{jim}(\theta_{j0}) > b_0 > 0$ for all j, i, m .

A2. (i) The kernel function K is a symmetric probability density which is differentiable of Lipschitz order 1 on its support $[-1,1]$. The bandwidths satisfy $n_j h_j^2 / \log^2 n_j \rightarrow \infty$, $n_j^{1/2} h_j^4 \rightarrow 0$ and $h_j \rightarrow 0$ as $n_j \rightarrow \infty$.

(ii) For each treatment j ($j = 1, \dots, k$), the design points $\{t_{jim}\}$ are thought to be independent and identically distributed from a super-population with density $f_j(t)$. There exist constants b_l and b_u such that $0 < b_l \leq \sup_{t \in S} f_j(t) \leq b_u < \infty$.

(iii) For each h_j and T_j , $j = 1, \dots, k$, there exist finite positive constants α_j , b_j and T such that $\alpha_j T_j = T$ and $b_j h_j = h$ for some h as $h \rightarrow 0$. Let $n = \sum_{i=1}^k n_j$, $n_j/n \rightarrow \rho_j$ for some non-zero ρ_j as $n \rightarrow \infty$ such that $\sum_{i=1}^k \rho_j = 1$.

A3. The residuals $\{\varepsilon_{ji}\}$ and $\{u_{ji}\}$ are independent of each other and each of $\{\varepsilon_{ji}\}$ and $\{u_{ji}\}$ are mutually independent among different j or i , respectively; $\max_{1 \leq i \leq n_j} \|u_{jim}\| = o_p\{n_j^{\frac{2+r}{2(4+r)}} (\log n_j)^{-1}\}$, $\max_{1 \leq i \leq n_j} E|\varepsilon_{jim}|^{4+r} < \infty$, for some $r > 0$; And assume that

$$\lim_{n_j \rightarrow \infty} (n_j T_j)^{-1} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E\{\tilde{\mathbb{X}}_{jim} \tilde{\mathbb{X}}_{jim}^T\} = \Sigma_x > 0,$$

where $\tilde{\mathbb{X}}_{jim} = \mathbb{X}_{jim} - E(\mathbb{X}_{jim} | t_{jim})$.

A4. The functions $g_{j0}(t)$ and $h_j(t)$ are, respectively, 1-dimensional and p -dimensional smooth functions with continuously second derivatives on $S = [0, 1]$.

Remark: Condition A1 are the regular conditions for the consistency of the binary MLE for the parameters in the missing propensity. Condition A2(i) are the usual conditions for the kernel and bandwidths in nonparametric curve estimation. Note that the optimal rate for the bandwidth $h_j = O(n_j^{-1/5})$ satisfies A2(i). The requirement of design points $\{t_{jim}\}$ in (A2)(ii) is a common assumption similar to the ones in Müller (1987). Condition A2(iii) is a mild assumption on the relationship between bandwidths and sample sizes among different samples. In A3, we do not require the residuals $\{\varepsilon_{ji}\}$ and $\{u_{ji}\}$ being respectively identically distributed for each fixed j . This allows extra heterogeneity among individuals for a treatment. The positive definite of Σ_x in Condition A3 is used to identify the “parameters” $(\beta_{j0}, \gamma_{j0}, g_{j0})$ uniquely, which is a generalization of the identification condition used in Härdle, Liang and

Gao (2000) to longitudinal data. This condition can be checked empirically by constructing consistent estimate of Σ_x .

Derivation of (4.3.14) : To appreciate this, we note from (4.3.12) that via standard derivations in empirical likelihood (Owen, 1990) that $\|\lambda_j\| = O_p(n_j^{-1/2})$, and

$$\lambda_j = \left(\sum_{i=1}^{n_j} Z_{ji}(\beta) Z_{ji}(\beta)^\tau \right)^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) + o_p(n_j^{-1/2}), \quad j = 1, 2, \dots, k.$$

Then we can write

$$\begin{aligned} \ell_n &= 2 \min_{\beta} \frac{1}{2} \sum_{j=1}^k \left\{ \sum_{i=1}^{n_j} Z_{ji}^\tau(\beta) \left(\sum_{i=1}^{n_j} Z_{ji}(\beta) Z_{ji}(\beta)^\tau \right)^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) \right\} + o_p \left\{ (\min_j n_j)^{-1/2} \right\} \\ &= \min_{\beta} \sum_{j=1}^k \frac{1}{n_j T_j} \left\{ \sum_{i=1}^{n_j} Z_{ji}^\tau(\beta) B_j^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) \right\} + o_p \left\{ (\min_j n_j)^{-1/2} \right\} \end{aligned} \quad (4.9.34)$$

where $B_j := \lim_{n_j \rightarrow \infty} \frac{1}{n_j T_j} \sum_{i=1}^{n_j} E\{Z_{ji}(\beta_{j0}) Z_{ji}(\beta_{j0})^\tau\}$, which is not related with β for any $\beta = \beta_{j0} + \Delta_{jn}$ and $\Delta_{jn} = O(n_j^{-1/2})$.

Using the Lagrange method to carry out the minimizations in (4.9.34), we want to minimize

$$Q = \frac{1}{2} \sum_{j=1}^k \frac{1}{n_j T_j} \left(\sum_{i=1}^{n_j} Z_{ji}^\tau(\beta_j) B_j^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta_j) \right) - \sum_{j=2}^k \eta_j^\tau (\beta_1 - \beta_j),$$

where η_1, \dots, η_k are lagrange multipliers. Then

$$\frac{\partial Q}{\partial \beta_1} = \frac{1}{n_1 T_1} \sum_{i=1}^{n_1} Z_{1i}^\tau(\beta_1) B_1^{-1} \sum_{i=1}^{n_1} \sum_{m=1}^{T_1} \frac{\delta_{1im}}{\pi_{1im}(\hat{\theta})} \tilde{X}_{1im} \tilde{X}_{1im}^\tau - \sum_{j=2}^k \eta_j,$$

and

$$\frac{\partial Q}{\partial \beta_j} = \frac{1}{n_j T_j} \sum_{i=1}^{n_j} Z_{ji}^\tau(\beta_j) B_j^{-1} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta})} \tilde{X}_{jim} \tilde{X}_{jim}^\tau + \eta_j, \quad j = 2, \dots, k,$$

Setting $\beta_1 = \beta_2 = \dots = \beta_k = \beta$, then the minima β satisfies

$$\sum_{j=1}^k \frac{1}{\sqrt{n_j T_j}} \Omega_{x_j} B_j^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) = o_p(1). \quad (4.9.35)$$

Inverting (4.9.35) for β , we have

$$\beta = \left(\sum_{j=1}^k \Omega_{x_j} B_j^{-1} \Omega_{x_j} \right)^{-1} \left(\sum_{j=1}^k \Omega_{x_j} B_j^{-1} \Omega_{x_j y_j} \right) + o_p(1).$$

□

Lemma 1 Suppose $(e_{i1}, \dots, e_{iT})_{i=1}^T$ is a sequence of T -dimensional independent random vectors and T is a fixed finite number, and $\max_{1 \leq i \leq n} E(|e_{im}|^\delta) < \infty$ for some $\delta > 1$ and all m . Let $\{a_{jim}, 1 \leq i, j \leq n, 1 \leq m \leq T\}$ be a collection of real numbers such that $\max_{1 \leq j \leq n} \sum_{i=1}^n \sum_{m=1}^T |a_{jim}| < \infty$. Let $d_n = \max_{1 \leq i, j \leq n, 1 \leq m \leq T} |a_{jim}|$, then

$$\max_{1 \leq j \leq n} \left| \sum_{i=1}^n \sum_{m=1}^T a_{jim} e_{im} \right| = O\{\max(n^{1/\delta} d_n, d_n^{1/2}) \log n\} \quad a.s.$$

Proof This can be proved in a similar way as Lemma 1 of Shi and Lau (2000).

Lemma 2 Under assumptions A1, A2(i), A3 and A4, for any $1 \leq l \neq g \leq k$, under the hypothesis: $\beta_{l0} = \beta_{g0}$,

$$\{(\Omega_{x_l}^{-1} B_l \Omega_{x_l}^{-1}) + (\Omega_{x_g}^{-1} B_g \Omega_{x_g}^{-1})\}^{-1/2} (\Omega_{x_l}^{-1} \Omega_{x_l y_l} - \Omega_{x_g}^{-1} \Omega_{x_g y_g}) \xrightarrow{d} N(0, I_p).$$

Proof Since we know that

$$\begin{aligned} \Omega_{x_j}^{-1} \Omega_{x_j y_j} &= \frac{1}{\sqrt{n_j T_j}} \Omega_{x_j}^{-1} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{X}_{jim} (\tilde{Y}_{jim} - \tilde{M}_{jim}^\tau \tilde{\gamma}_j) \\ &= \frac{1}{\sqrt{n_j T_j}} \Omega_{x_j}^{-1} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{X}_{jim} (\tilde{Y}_{jim} - \tilde{X}_{jim}^\tau \beta_{j0} - \tilde{M}_{jim}^\tau \tilde{\gamma}_j) + \beta_{j0} \\ &= \frac{1}{\sqrt{n_j T_j}} \Omega_{x_j}^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta_{j0}) + \beta_{j0}, \end{aligned}$$

and because samples l and g are mutually independent, we need to show that, for $j = l, g$, $(\Omega_{x_j}^{-1} B_j \Omega_{x_j}^{-1})^{-1/2} \Omega_{x_j}^{-1} \Omega_{x_j y_j} \xrightarrow{d} N(0, I_p)$, which is equivalent to show that

$$\frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} Z_{ji}(\beta_{j0}) \xrightarrow{d} N(0, B_j).$$

Recall that $\tilde{Y}_{jim} = \tilde{X}_{jim}^\tau \beta_{j0} + \tilde{M}_{jim}^\tau \gamma_{j0} + \tilde{g}_{j0}(t_{jim}) + \tilde{\varepsilon}_{jim}$, where $\tilde{g}_{j0}(t_{jim}) = g_{j0}(t_{jim}) - \bar{g}_{j0}(t_{jim})$, $\tilde{\varepsilon}_{jim} = \varepsilon_{jim} - \bar{\varepsilon}_{jim}$ and

$$\bar{A}(t_{jim}) = \sum_{i_1=1}^{n_j} \sum_{m_1=1}^{T_j} w_{j i_1 m_1, h}(t_{jim}) A(t_{j i_1 m_1}).$$

Then, it follows that

$$\begin{aligned} \sum_{i=1}^{n_j} Z_{ji}(\beta_{j0}) &= \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{X}_{jim} \left\{ \tilde{g}_{j0}(t_{jim}) + \tilde{M}_{jim}^\tau (\gamma_{j0} - \tilde{\gamma}_j) + \tilde{\varepsilon}_{jim} \right\} \\ &= \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\theta_{j0})} \tilde{X}_{jim} \left\{ \tilde{g}_{j0}(t_{jim}) + \tilde{M}_{jim}^\tau (\gamma_{j0} - \tilde{\gamma}_j) + \tilde{\varepsilon}_{jim} \right\} \{1 + o_p(1)\}. \end{aligned}$$

The last equality is true, since $\hat{\theta}_j - \theta_{j0} = O_p(n_j^{-1/2})$. At the same time, we can decompose

$$\begin{aligned}
& \tilde{X}_{jim} \left\{ \tilde{g}_{j0}(t_{jim}) + \tilde{M}_{jim}^\tau(\gamma_{j0} - \tilde{\gamma}_j) + \tilde{\varepsilon}_{jim} \right\} \\
&= \left\{ \tilde{h}(t_{jim}) + u_{jim} - \bar{u}(t_{jim}) \right\} \left\{ \tilde{g}_{j0}(t_{jim}) + \varepsilon_{jim} - \bar{\varepsilon}_{jim} \right\} \\
&= u_{jim} \varepsilon_{jim} + \left\{ (\tilde{h}(t_{jim}) - \bar{u}(t_{jim})) \varepsilon_{jim} + (\tilde{g}_{j0}(t_{jim}) - \bar{\varepsilon}_{jim}) u_{jim} \right\} \\
&\quad + \left\{ (\tilde{h}(t_{jim}) - \bar{u}(t_{jim})) (\tilde{g}_{j0}(t_{jim}) - \bar{\varepsilon}_{jim}) \right\} + \tilde{X}_{jim} \tilde{M}_{jim}^\tau(\gamma_{j0} - \tilde{\gamma}_j) \\
&:= I_1 + I_2 + I_3 + I_4, \text{ say.}
\end{aligned}$$

From assumptions in A2(i) and the facts that

$$\max_{1 \leq i, i_1 \leq n_j, 1 \leq m, m_1 \leq T_j} w_{ji_1 m_1, h}(t_{jim}) = O\{(n_j h_j)^{-1}\}$$

and $\sum_{i=1}^{n_j} \sum_{m=1}^{k_{ji}} w_{ji_1 m_1, h}(t_{jim}) = 1$, we have, by applying Lemma 1

$$\max_{1 \leq i \leq n_j} \|\tilde{h}(t_{jim}) - \bar{u}(t_{jim})\| = o(1) \text{ a.s.}, \quad \max_{1 \leq i \leq n_j} |\tilde{g}_{j0}(t_{jim}) - \bar{\varepsilon}_{jim}| = o(1) \text{ a.s.}, \quad (4.9.36)$$

$$\max_{1 \leq i \leq n_j} \|(\tilde{h}(t_{jim}) - \bar{u}(t_{jim}))(\tilde{g}_{j0}(t_{jim}) - \bar{\varepsilon}_{jim})\| = o(n_j^{-1/2}) \text{ a.s.}$$

Therefore,

$$\begin{aligned}
\frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} Z_{ji}(\beta_{j0}) &= \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \left\{ \delta_{jim} \pi_{jim}^{-1}(\theta_{j0}) \right\} (I_1 + I_2 + I_3 + I_4) \\
&:= J_1 + J_2 + J_3 + J_4, \quad \text{say.}
\end{aligned}$$

It is easy to see $J_3 = o_p(1)$, $J_4 = o_p(1)$ and from (4.9.36),

$$\begin{aligned}
|J_2| &\leq o_p(1) \times \left\| \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \left\{ \delta_{jim} \pi_{jim}^{-1}(\theta_{j0}) \right\} u_{jim} \right\| \\
&\quad + o_p(1) \times \left\| \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \left\{ \delta_{jim} \pi_{jim}^{-1}(\theta_{j0}) \right\} \varepsilon_{jim} \right\| = o_p(1).
\end{aligned}$$

We note that $\text{Var}(J_1) = B_j$ and J_1 is a sum of independent random variables. Therefore, we will complete our proof by verifying the Linderberg-Feller condition for $\alpha' J_1$, for any $\alpha \in R^p$. Let

$\nu_{jim} = \delta_{jim} \pi_{jim}^{-1}(\theta_{j0})$. Then for any $\epsilon > 0$, let $L_n := \sum_{i=1}^{n_j} \text{Var}\left(\sum_{m=1}^{T_j} \alpha' u_{jim} \nu_{jim} \varepsilon_{jim}\right) = O(n_j)$,

$$\begin{aligned} \Lambda_n(\epsilon) &= \frac{1}{L_n} \sum_{i=1}^{n_j} E \left[I \left\{ \sum_{m=1}^{T_j} \alpha' u_{jim} \nu_{jim} \varepsilon_{jim} \geq \epsilon \sqrt{L_n} \right\} \left\{ \sum_{m=1}^{T_j} \alpha' u_{jim} \nu_{jim} \varepsilon_{jim} \right\}^2 \right] \\ &\leq \frac{C}{L_n} \sum_{i=1}^{n_j} \frac{E \left| \sum_{m=1}^{T_j} \varepsilon_{jim} \right|^{4+r} o \left\{ n_j^{(r+2)/2} (\log n_j)^{-(4+r)} \right\} \|\alpha\|^{4+r}}{(\epsilon \sqrt{L_n})^{r+2}} \\ &= \frac{C \|\alpha\|^{4+r}}{\epsilon^{r+2}} o \left\{ \log^{-(4+r)}(n_j) \right\} \rightarrow 0, \end{aligned}$$

where C is a finite positive constant. This completes the proof of this Lemma. \square

Lemma 3 *If the conditions A1, A2, A3 and A4 hold, then under null hypothesis H_{0a} , i.e. $\beta_{10} = \beta_{20}$, $\ell_n \xrightarrow{d} \chi_p^2$.*

Proof Let $S_1 := \sum_{j=1}^k \Omega_{x_j} B_j^{-1} \Omega_{x_j}$ and $S_2 := \sum_{j=1}^k \Omega_{x_j} B_j^{-1} \Omega_{x_j y_j}$. In this Lemma, $k = 2$. Then,

$$\begin{aligned} \ell_n &= (\Omega_{x_1 y_1}^\tau - S_2^\tau S_1^{-1} \Omega_{x_1}) B_1^{-1} (\Omega_{x_1 y_1} - \Omega_{x_1} S_1^{-1} S_2) \\ &\quad + (\Omega_{x_2 y_2}^\tau - S_2^\tau S_1^{-1} \Omega_{x_2}) B_2^{-1} (\Omega_{x_2 y_2} - \Omega_{x_2} S_1^{-1} S_2) + o_p(1) \\ &= (\Omega_{x_1 y_1}^\tau \Omega_{x_1}^{-1} S_1 - S_2^\tau) S_1^{-1} \Omega_{x_1} B_1^{-1} \Omega_{x_1} S_1^{-1} (S_1 \Omega_{x_1}^{-1} \Omega_{x_1 y_1} - S_2) \\ &\quad + (\Omega_{x_2 y_2}^\tau \Omega_{x_2}^{-1} S_1 - S_2^\tau) S_1^{-1} \Omega_{x_2} B_2^{-1} \Omega_{x_2} S_1^{-1} (S_1 \Omega_{x_2}^{-1} \Omega_{x_2 y_2} - S_2) + o_p(1) \end{aligned}$$

It is easy to show that

$$\begin{aligned} S_1 \Omega_{x_1}^{-1} \Omega_{x_1 y_1} - S_2 &= \Omega_{x_2} B_2^{-1} \Omega_{x_2} (\Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_2}^{-1} \Omega_{x_2 y_2}), \\ S_1 \Omega_{x_2}^{-1} \Omega_{x_2 y_2} - S_2 &= \Omega_{x_1} B_1^{-1} \Omega_{x_1} (\Omega_{x_2}^{-1} \Omega_{x_2 y_2} - \Omega_{x_1}^{-1} \Omega_{x_1 y_1}). \end{aligned}$$

Then

$$\ell_n = (\Omega_{x_1 y_1}^\tau \Omega_{x_1}^{-1} - \Omega_{x_2 y_2}^\tau \Omega_{x_2}^{-1}) V (\Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_2}^{-1} \Omega_{x_2 y_2}) + o_p(1),$$

where

$$\begin{aligned} V &= (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) \\ &\quad + (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) \\ &:= P_1 + P_2, \text{ say.} \end{aligned}$$

We note that

$$P_1 = (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) - (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1})$$

and

$$P_2 = (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) - (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1}).$$

It follows that $V = (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1})$. Notice that V is symmetric and $V(\Omega_{x_1}^{-1} B_1 \Omega_{x_1}^{-1}) + V'(\Omega_{x_2}^{-1} B_2 \Omega_{x_2}^{-1}) = I$. Thus, to prove the theorem, we just need to show that

$$(\Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_2}^{-1} \Omega_{x_2 y_2})^\tau \{(\Omega_{x_1}^{-1} B_1 \Omega_{x_1}^{-1}) + (\Omega_{x_2}^{-1} B_2 \Omega_{x_2}^{-1})\}^{-1} (\Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_2}^{-1} \Omega_{x_2 y_2}) \xrightarrow{d} \chi_p^2,$$

which is true as Lemma 2 implies

$$\{(\Omega_{x_1}^{-1} B_1 \Omega_{x_1}^{-1}) + (\Omega_{x_2}^{-1} B_2 \Omega_{x_2}^{-1})\}^{-1/2} (\Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_2}^{-1} \Omega_{x_2 y_2}) \xrightarrow{d} N(0, I_p).$$

This completes the proof of Lemma 3. \square

Proof of Theorems 1 Let $S_1 := \sum_{j=1}^k \Omega_{x_j} B_j^{-1} \Omega_{x_j}$ and $S_2 := \sum_{j=1}^k \Omega_{x_j} B_j^{-1} \Omega_{x_j y_j}$. From the definition of ℓ_n ,

$$\begin{aligned} \ell_n &= \sum_{j=1}^k (\Omega_{x_j y_j}^\tau - S_2^\tau S_1^{-1} \Omega_{x_j}) B_j^{-1} (\Omega_{x_j y_j} - \Omega_{x_j} S_1^{-1} S_2) + o_p(1) \\ &= \sum_{j=1}^k (\Omega_{x_j y_j}^\tau \Omega_{x_j}^{-1} S_1 - S_2^\tau) S_1^{-1} \Omega_{x_j} B_j^{-1} \Omega_{x_j} S_1^{-1} (S_1 \Omega_{x_j}^{-1} \Omega_{x_j y_j} - S_2) + o_p(1). \end{aligned}$$

It can be shown that (similar to the proof of Lemma 3),

$$\ell_n = \begin{pmatrix} \Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_2}^{-1} \Omega_{x_2 y_2} \\ \Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_3}^{-1} \Omega_{x_3 y_3} \\ \vdots \\ \Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_k}^{-1} \Omega_{x_k y_k} \end{pmatrix}^\tau \Sigma_0 \begin{pmatrix} \Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_2}^{-1} \Omega_{x_2 y_2} \\ \Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_3}^{-1} \Omega_{x_3 y_3} \\ \vdots \\ \Omega_{x_1}^{-1} \Omega_{x_1 y_1} - \Omega_{x_k}^{-1} \Omega_{x_k y_k} \end{pmatrix} + o_p(1), \quad (4.9.37)$$

where Σ_0 is a $(k-1)p \times (k-1)p$ matrix with $(j-1)$ -th ($j = 2, \dots, k$) diagonal matrix component as $(\Omega_{x_j} B_j^{-1} \Omega_{x_j}) - (\Omega_{x_j} B_j^{-1} \Omega_{x_j}) S_1^{-1} (\Omega_{x_j} B_j^{-1} \Omega_{x_j})$ and $(p-1, q-1)$ -th ($p, q = 2, \dots, k$) matrix component is $-(\Omega_{x_p} B_p^{-1} \Omega_{x_p}) S_1^{-1} (\Omega_{x_q} B_q^{-1} \Omega_{x_q})$.

To make the derivation easily presentable, we only present the detail proof for $k = 3$, as the general case can be done similarly except more tedious. From (4.9.37), we have

$$\ell_n = \begin{pmatrix} \Omega_{x_1}^{-1}\Omega_{x_1y_1} - \Omega_{x_2}^{-1}\Omega_{x_2y_2} \\ \Omega_{x_1}^{-1}\Omega_{x_1y_1} - \Omega_{x_3}^{-1}\Omega_{x_3y_3} \end{pmatrix}^\tau \Sigma_0 \begin{pmatrix} \Omega_{x_1}^{-1}\Omega_{x_1y_1} - \Omega_{x_2}^{-1}\Omega_{x_2y_2} \\ \Omega_{x_1}^{-1}\Omega_{x_1y_1} - \Omega_{x_3}^{-1}\Omega_{x_3y_3} \end{pmatrix} + o_p(1), \quad (4.9.38)$$

where $\Sigma_0 = \begin{pmatrix} A & C \\ C^\tau & B \end{pmatrix}$,

$$A = \Omega_{x_2} B_2^{-1} \Omega_{x_2} - (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2}),$$

$$B = \Omega_{x_3} B_3^{-1} \Omega_{x_3} - (\Omega_{x_3} B_3^{-1} \Omega_{x_3}) S_1^{-1} (\Omega_{x_3} B_3^{-1} \Omega_{x_3}) \quad \text{and}$$

$$C = -(\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_3} B_3^{-1} \Omega_{x_3}).$$

From the proof of Lemma 2, we know that

$$\begin{aligned} \Sigma_1 &= \text{Var} \begin{pmatrix} \Omega_{x_1}^{-1}\Omega_{x_1y_1} - \Omega_{x_2}^{-1}\Omega_{x_2y_2} \\ \Omega_{x_1}^{-1}\Omega_{x_1y_1} - \Omega_{x_3}^{-1}\Omega_{x_3y_3} \end{pmatrix} \\ &= \begin{pmatrix} \Omega_{x_1}^{-1}B_1\Omega_{x_1}^{-1} + \Omega_{x_2}^{-1}B_2\Omega_{x_2}^{-1} & \Omega_{x_1}^{-1}B_1\Omega_{x_1}^{-1} \\ \Omega_{x_1}^{-1}B_1\Omega_{x_1}^{-1} & \Omega_{x_1}^{-1}B_1\Omega_{x_1}^{-1} + \Omega_{x_3}^{-1}B_3\Omega_{x_3}^{-1} \end{pmatrix}. \end{aligned}$$

As $\Sigma_0 = \Sigma_1^{-1}$, from (4.9.38) $\ell_n \xrightarrow{d} \chi_{2p}^2$. This completes the proof. \square

Proof of Theorem 2 We note that $\Sigma_D^{-1/2} D \xrightarrow{d} N_{(k-1)p}(\gamma, I_{(k-1)p})$, where D and Σ_D are defined before Theorem 2 and (4.3.15) respectively, and $\gamma = \Sigma_D^{-1/2} D$. From (4.9.38), $\ell_n = D^\tau \Sigma_D^{-1} D + o_p(1)$, therefore $\ell_n \rightarrow \chi_{(k-1)p}^2(\gamma^2)$, which completes the proof of the theorem. \square

Proof of Theorem 3 Let $v_j(t, h_j) = \sum_{i=1}^{n_j} R_{ji}^2 \{g(t)\}$ and

$$d_j(t, h_j) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} K\left(\frac{t_{jim} - t}{h_j}\right).$$

To simplify notation, we sometimes hide the arguments of $v_j(t, h_j)$ and $d_j(t, h_j)$. After plugging in the leading term of $\mathcal{L}_n(t)$ into \mathcal{T}_n , we have the leading term of \mathcal{T}_n , which is

$$\int \sum_{j=1}^k v_j^{-1} \left[\sum_{i=1}^{n_j} R_{ji} \{0\} - d_j \left(\sum_{s=1}^k v_s^{-1} d_s^2 \right)^{-1} \sum_{s=1}^k v_s^{-1} d_s \sum_{i=1}^{n_s} R_{si} \{0\} \right]^2 \varpi(t) dt.$$

Note that the leading order term of $\mathcal{L}_n(t)$ is $\left\{ \sum_{j=1}^2 \zeta_j \right\}^{-1} \left\{ \hat{g}_1(t) - \hat{g}_2(t) \right\}^2$ with $\zeta_j = v_j(t, h_j)/d_j^2(t, h_j)$ for $k = 2$; If $k = 3$, the leading term of $\mathcal{L}_n(t)$ will be

$$\begin{pmatrix} \hat{g}_1(t) - \hat{g}_2(t) \\ \hat{g}_1(t) - \hat{g}_3(t) \end{pmatrix}' H_n \begin{pmatrix} \hat{g}_1(t) - \hat{g}_2(t) \\ \hat{g}_1(t) - \hat{g}_3(t) \end{pmatrix} \quad (4.9.39)$$

where

$$H_n = \left\{ \sum_{j=1}^3 \zeta_j^{-1} \right\}^{-1} \times \begin{pmatrix} \zeta_2^{-1}(\zeta_1^{-1} + \zeta_3^{-1}) & -\zeta_2^{-1}\zeta_3^{-1} \\ -\zeta_2^{-1}\zeta_3^{-1} & \zeta_3^{-1}(\zeta_1^{-1} + \zeta_2^{-1}) \end{pmatrix}.$$

Under the local alternative $g_{s0}(t) = g_{10}(t) + C_{ns}\Delta_{ns}(t)$ for $s = 2, \dots, k$, the test statistic \mathcal{T}_n can be written as

$$\begin{aligned} \mathcal{T}_n &= \int_0^1 \sum_{j=1}^k v_j^{-1} \{ B_n^2(t) + A_n^2(t) + 2A_n(t)B_n(t) \} \varpi(t) dt + o_p(1) \\ &:= \mathcal{T}_{n1} + \mathcal{T}_{n2} + \mathcal{T}_{n3} + o_p(1), \end{aligned} \quad (4.9.40)$$

where $A_n(t) = d_j \left\{ C_{nj}\Delta_{nj}(t) - \left(\sum_{s=1}^k v_s^{-1} d_s^2 \right)^{-1} \sum_{s=1}^k v_s^{-1} d_s^2 C_{ns}\Delta_{ns}(t) \right\}$ and

$$B_n(t) = \sum_{i=1}^{n_j} R_{ji} \{ g_{j0}(t) \} - d_j \left(\sum_{s=1}^k v_s^{-1} d_s^2 \right)^{-1} \sum_{s=1}^k v_s^{-1} d_s \sum_{i=1}^{n_s} R_{si} \{ g_{s0}(t) \}.$$

Define $\sigma_{\varepsilon_j}^2 = \frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E \left\{ \frac{\varepsilon_{jim}^2}{\pi_{jim}(\theta_{j0})} \right\}$, $R(K) = \int K^2(t) dt$ and $V_j(t) = R(K) \sigma_{\varepsilon_j}^2 f_j(t)$.

We first show that $(n_j T_j h_j)^{-1} v_j(t, h_j) \xrightarrow{p} V_j(t)$. According to the definition of $v_j(t, h_j)$, $R_{ji} \{ g(t) \}$ and $g(t) = g_{j0}(t) + O\{(n_j h_j)^{-1/2}\}$, we get

$$\begin{aligned} \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} R_{ji}^2 \{ g(t) \} &= \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left(\sum_{m=1}^{T_j} K \left(\frac{t_{jim} - t}{h_j} \right) \nu_{jim} \varepsilon_{ji}(t_{jim}) \right)^2 \\ &+ \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left(\sum_{m=1}^{T_j} K \left(\frac{t_{jim} - t}{h_j} \right) \nu_{jim} \{ (g_{j0}(t_{jim}) - g_{j0}(t)) - (\hat{g}_j(t_{jim}) - \hat{g}_j(t)) \} \right)^2 \\ &+ \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left(\sum_{m=1}^{T_j} K \left(\frac{t_{jim} - t}{h_j} \right) \nu_{jim} X_{jim}^\tau (\beta_{j0} - \hat{\beta}_j) \right)^2 + o_p(1) \\ &:= A_1(t) + A_2(t) + A_3(t) + o_p(1). \end{aligned}$$

It is easy to see that $A_3(t) = O_p(n_j^{-1})$, since $\beta_{j0} - \hat{\beta}_j = O_p(n^{-1/2})$. For $A_2(t)$, we note that the kernel $K(t)$ has support on $[-1, 1]$ and is Lipchitz continuous from assumption A2(i). Then

a Taylor expansion yields

$$\begin{aligned} A_2(t) &= \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left(\sum_{m=1}^{T_j} \nu_{jim} K \left(\frac{t_{jim} - t}{h_j} \right) (g'_{j0}(t) - \hat{g}'_j(t))(t - t_{jim}) + O_p(h_j^2) \right)^2 \\ &\leq \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left(\sum_{m=1}^{T_j} \nu_{jim} K \left(\frac{t_{jim} - t}{h_j} \right) \right)^2 |g'_{j0}(t) - \hat{g}'_j(t)|^2 h_j^2 + o_p(h_j^2) = o_p(h_j^2), \end{aligned}$$

since $g'_{j0}(t) - \hat{g}'_j(t) = o_p(1)$. Note that $A_1(t)$ can be written as

$$\begin{aligned} A_1(t) &= \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \sum_{m_1=1}^{T_j} K \left(\frac{t_{jim} - t}{h_j} \right) K \left(\frac{t_{jim_1} - t}{h_j} \right) \nu_{jim} \nu_{jim_1} \varepsilon_{ji}(t_{jim}) \varepsilon_{ji}(t_{jim_1}) \\ &= \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} K^2 \left(\frac{t_{jim} - t}{h_j} \right) \nu_{jim}^2 \varepsilon_{ji}^2(t_{jim}) \\ &\quad + \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \sum_{m \neq m_1}^{T_j} K \left(\frac{t_{jim} - t}{h_j} \right) K \left(\frac{t_{jim_1} - t}{h_j} \right) \nu_{jim} \nu_{jim_1} \varepsilon_{ji}(t_{jim}) \varepsilon_{ji}(t_{jim_1}) \\ &:= A_{11}(t) + A_{12}(t). \end{aligned}$$

Then

$$\begin{aligned} E\{A_{12}(t)\} &= \frac{T_j - 1}{h_j} \int_0^1 \int_0^1 K \left(\frac{x - t}{h_j} \right) K \left(\frac{y - t}{h_j} \right) \rho_j(x, y) \sigma_j(x) \sigma_j(y) f_j(x) f_j(y) dx dy \\ &= h_j (T_j - 1) \sigma_{\varepsilon_j}^2(t) f_j^2(t) \{1 + o(1)\} = O(h_j), \end{aligned}$$

which is the case since T_j is finite. Note here, when $m \neq m_1$, $A_{12}(t)$ is similar to the kernel estimator for a bivariate function. Whereas in two dimensional kernel estimator are divided by $n_j T_j h_j^2$. However, the denominator is $n_j T_j h_j$ in $A_{12}(t)$, so this term is a smaller order term comparing to $A_{11}(t)$. From assumptions A2(ii) and A3(i), we know $A_{11}(t) \xrightarrow{p} V_j(t)$.

Let us first consider the first term in \mathcal{T}_n given in (4.9.40),

$$\begin{aligned}
\mathcal{T}_{n1} &= \int_0^1 \sum_{j=1}^k v_j^{-1} B_n^2(t) \varpi(t) dt \\
&= \int_0^1 \sum_{j=1}^k v_j^{-1} \left[\sum_{i=1}^{n_j} R_{ji} \{g_{j0}(t)\} \right]^2 - \int_0^1 \left(\sum_{s=1}^k v_s^{-1} d_s^2 \right)^{-1} \left[\sum_{s=1}^k v_s^{-1} d_s \sum_{i=1}^{n_s} R_{si} \{g_{s0}(t)\} \right]^2 \varpi(t) dt \\
&= \int_0^1 \sum_{j=1}^k \left(1 - \left[\sum_{s=1}^k v_s^{-1} d_s^2 \right]^{-1} v_j^{-1} d_j^2 \right) v_j^{-1} \left[\sum_{i=1}^{n_j} R_{ji} \{g_{j0}(t)\} \right]^2 \varpi(t) dt \\
&\quad - \int_0^1 \sum_{j \neq j_1} \left[\sum_{s=1}^k v_s^{-1} d_s^2 \right]^{-1} v_j^{-1} d_j v_{j_1}^{-1} d_{j_1} \left[\sum_{i=1}^{n_j} R_{ji} \{g_{j0}(t)\} \right] \left[\sum_{i=1}^{n_{j_1}} R_{j_1 i} \{g_{j_1 0}(t)\} \right] \varpi(t) dt \\
&:= \mathcal{T}_{n1}^{(1)} - \mathcal{T}_{n1}^{(2)}, \quad \text{say,} \tag{4.9.41}
\end{aligned}$$

Let

$$\begin{aligned}
S_{j1}^2(t) &= \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \sum_{m_1=1}^{T_j} \frac{\delta_{jim} \delta_{jim_1} \varepsilon_{jim} \varepsilon_{jim_1}}{\pi_{jim}(\theta_{j0}) \pi_{jim_1}(\theta_{j0})} K\left(\frac{t_{jim} - t}{h_j}\right) K\left(\frac{t_{jim_1} - t}{h_j}\right) \quad \text{and} \\
S_{j2}^2(t) &= \frac{2}{n_j h_j T_j} \sum_{i < i_1}^{n_j} \left\{ \sum_{m=1}^{T_j} \frac{\delta_{jim} \varepsilon_{jim}}{\pi_{jim}(\theta_{j0})} K\left(\frac{t_{jim} - t}{h_j}\right) \right\} \left\{ \sum_{m_1=1}^{T_j} \frac{\delta_{ji_1 m_1} \varepsilon_{ji_1 m_1}}{\pi_{ji_1 m_1}(\theta_{j0})} K\left(\frac{t_{ji_1 m_1} - t}{h_j}\right) \right\}.
\end{aligned}$$

We observe that

$$\begin{aligned}
\mathcal{T}_{n1}^{(1)} &= \left[\int_0^1 \sum_{j=1}^k \{1 - W_j(t)\} V_j^{-1}(t) S_{j1}^2(t) \varpi(t) dt \right. \\
&\quad \left. + \int_0^1 \sum_{j=1}^k \{1 - W_j(t)\} V_j^{-1}(t) S_{j2}^2(t) \varpi(t) dt \right] \{1 + o_p(1)\} \\
&:= (\mathcal{T}_{n1}^{(11)} + \mathcal{T}_{n1}^{(12)}) \{1 + o_p(1)\}.
\end{aligned}$$

Since $E\{S_{j1}^2(t)\} = V_j(t) + O(h)$, $E(\mathcal{T}_{n1}^{(11)}) = \sum_{j=1}^k \int_0^1 (1 - W_j(t)) \varpi(t) dt = k - 1 + O(h)$ and the variance of $\mathcal{T}_{n1}^{(11)}$ is $O(n^{-1}h) = o(h)$, under the condition $h = O(n^{-1/5})$. Thus

$$h^{-1/2} \{ \mathcal{T}_{n1}^{(11)} - (k - 1) \} \xrightarrow{p} 0. \tag{4.9.42}$$

Define $\xi_{ji}(t) = \frac{1}{\sqrt{h_j T_j}} \sum_{m=1}^{T_j} K\left(\frac{t_{jim} - t}{h_j}\right) \nu_{jim} \varepsilon_{jim}$, then we have

$$\mathcal{T}_{n1}^{(12)} = \sum_{j=1}^k \sum_{i \neq i_1}^{n_j} n_j^{-1} \int_0^1 \{1 - W_j(t)\} V_j^{-1}(t) \xi_{ji}(t) \xi_{ji_1}(t) \varpi(t) dt + o_p(h^{1/2}). \tag{4.9.43}$$

and

$$\mathcal{T}_{n1}^{(2)} = \sum_{j \neq j_1} \sum_{i=1}^{n_j} \sum_{l=1}^{n_{j_1}} (n_j n_{j_1})^{-\frac{1}{2}} \int_0^1 \left(\frac{W_j(t)W_{j_1}(t)}{V_j(t)V_{j_1}(t)} \right)^{1/2} \xi_{ji}(t)\xi_{j_1l}(t)\varpi(t)dt + o_p(h^{1/2}). \quad (4.9.44)$$

Let $N = \sum_{j=1}^k n_j$. We stack ξ_{ji} ($j = 1, \dots, k, i = 1, \dots, n_j$) to form a sequence ϕ_s , $s = 1, \dots, N$. Let G_j be the collection of the subscripts of ϕ whose corresponding ξ are in Treatment j . Define

$$C_{ps}(t) = \frac{1}{n(p,s)} \left\{ \sum_{j=1}^k I(p \in G_j, s \in G_j) V_j^{-1}(t) - \sum_{j=1}^k \sum_{l=1}^k \left(\frac{W_j(t)W_l(t)}{V_j(t)V_l(t)} \right)^{1/2} I(p \in G_j, s \in G_l) \right\}, \quad (4.9.45)$$

where $n(p,s) = \sum_{j=1}^k \sum_{l=1}^k (n_j n_l)^{1/2} I(p \in G_j, s \in G_l)$ and $I(p \in G_j, s \in G_l)$ is the usual indicator function. Using these notations, we may write

$$U_N := \mathcal{T}_{n1}^{(12)} - \mathcal{T}_{n1}^{(2)} = 2 \sum_{p=1}^N \sum_{s < p} \psi(\phi_p, \phi_s), \quad (4.9.46)$$

where $\psi(\phi_p, \phi_s) = \int_0^1 C_{ps}(t)\phi_p(t)\phi_s(t)\varpi(t)dt$. Then (4.9.46) is a quadratic form with kernel $\psi(\phi_p, \phi_s)$. Let $\sigma_{ps}^2 = \text{Var}\{\psi(\phi_p, \phi_s)\}$. Using results for generalized quadratic form with independent but not identically distributed random variables (de Jong, 1987) if

$$\{\text{Var}(U_N)\}^{-1} \max_{1 \leq p \leq N} \sum_{s=1}^N \sigma_{ps}^2 \rightarrow 0 \quad \text{and} \quad (4.9.47)$$

$$\{\text{Var}(U_N)\}^{-2} E U_N^4 \rightarrow 3, \quad (4.9.48)$$

then (4.9.46) is asymptotically normally distributed with mean 0 and variance

$$\text{Var}(U_N) = \text{Var}(\mathcal{T}_{n1}^{(12)}) + \text{Var}(\mathcal{T}_{n1}^{(2)}) - 2\text{Cov}(\mathcal{T}_{n1}^{(12)}, \mathcal{T}_{n1}^{(2)}). \quad (4.9.49)$$

Let us first derive $\text{Var}(U_N)$. We note that $\text{Var}(\mathcal{T}_{n1}^{(12)}) = \sum_{j=1}^k \frac{4}{n_j^2} \sum_{i < i_1} \sigma_{1,jii_1}^2$ where

$$\begin{aligned} \sigma_{1,jii_1}^2 &= E_i E_{i_1} \left[\int_0^1 \int_0^1 \frac{\{1 - W_j(t)\}\{1 - W_j(u)\}}{V_j(t)V_j(u)} \xi_{ji}(t)\xi_{j_1i_1}(t)\xi_{ji}(u)\xi_{j_1i_1}(u)\varpi(t)\varpi(u)dtdu \right] \\ &= \frac{1}{T_j^2} \sum_{m, m_1} \frac{\{1 - W_j(t_{jim})\}^2}{V_j^2(t_{jim})} \sigma_{\varepsilon_{ji}}^2(t_{jim}) \sigma_{\varepsilon_{j_1i_1}}^2(t_{j_1i_1m_1}) \\ &\quad \times \varpi^2(t_{jim}) \left(K^{(2)} \left(\frac{t_{jim} - t_{j_1i_1m_1}}{h_j} \right) \right)^2 \{1 + o(1)\}, \end{aligned}$$

where $\sigma_{\varepsilon_{ji}}^2(t_{jim}) = E\{\varepsilon_{jim}^2/\pi_{jim}(\theta_{j0})\}$. Since $\{t_{jim}\}$ are fixed design points generated from a density $f_j(t)$, via a Taylor expansion and by Assumption A2(ii),

$$\text{Var}(\mathcal{T}_{n1}^{(12)}) = 2hR(K)^{-2}K_1^{(4)}(0) \sum_{j=1}^k b_j^{-1} \int_0^1 (1 - W_j(t))^2 \varpi^2(t) dt \{1 + o(1)\}. \quad (4.9.50)$$

Similar to our derivation for the variance of $\mathcal{T}_{n1}^{(12)}$, it may be shown that

$$\text{Var}(\mathcal{T}_{n1}^{(2)}) = 2hR(K)^{-2} \sum_{j \neq j_1}^k K_{b_j/b_{j_1}}^{(4)}(0) (b_j b_{j_1})^{-1/2} \left\{ \int_0^1 W_j(t) W_{j_1}(t) \varpi^2(t) dt \right\} \{1 + o(1)\}. \quad (4.9.51)$$

From (4.9.49), we also need to calculate the covariance between $\mathcal{T}_{n1}^{(12)}$ and $\mathcal{T}_{n1}^{(2)}$. Using the same method for calculating variance for $\mathcal{T}_{n1}^{(12)}$ and $\mathcal{T}_{n1}^{(2)}$, we may show that

$$\text{Cov}(\mathcal{T}_{n1}^{(12)}, \mathcal{T}_{n1}^{(2)}) = O(h^2), \quad (4.9.52)$$

In summary of (4.9.50), (4.9.51) and (4.9.52),

$$\text{Var}(U_N) := h\sigma_0^2 = 2hK^{(2)}(0)^{-2} \int_0^1 \Lambda(t) \varpi^2(t) dt \{1 + o(1)\}, \quad (4.9.53)$$

where $\Lambda(t)$ is defined just before Theorem 3.

Next we need to establish the conditions (4.9.47) and (4.9.48). For (4.9.47), we have

$$\begin{aligned} \{\text{Var}(U_N)\}^{-1} \max_{1 \leq p \leq N} \sum_{s=1}^N \sigma_{ps}^2 &= (h\sigma_0^2)^{-1} \max_{\substack{1 \leq j \leq k \\ 1 \leq i \leq n_j}} \left\{ \frac{1}{n_j^2} \sum_{i_1=1}^{n_j} \sigma_{1,jii_1}^2 + \sum_{j_1=1}^k \frac{1}{n_j n_{j_1}} \sum_{i_1=1}^{n_{j_1}} \sigma_{2,jj_1i_1}^2 \right\} \\ &\leq (h\sigma_0^2)^{-1} \left[\max_{\substack{1 \leq j \leq k \\ 1 \leq i \leq n_j}} \left\{ \frac{1}{n_j^2} \sum_{i_1=1}^{n_j} \sigma_{1,jii_1}^2 \right\} + \max_{\substack{1 \leq j \leq k \\ 1 \leq i \leq n_j}} \left\{ \sum_{j_1=1}^k \frac{1}{n_j n_{j_1}} \sum_{i_1=1}^{n_{j_1}} \sigma_{2,jj_1i_1}^2 \right\} \right]. \end{aligned}$$

From conditions (A2) and (A3),

$$\begin{aligned} \max_{\substack{1 \leq j \leq k \\ 1 \leq i \leq n_j}} \frac{1}{n_j^2} \sum_{i_1=1}^{n_j} \sigma_{1,jii_1}^2 &= \max_{\substack{1 \leq j \leq k \\ 1 \leq i \leq n_j}} \frac{1}{n_j T_j} \sum_m \frac{\{1 - W_j(t_{jim})\}^2}{V_j^2(t_{jim})} \sigma_{\varepsilon_{ji}}^2(t_{jim}) \varpi^2(t_{jim}) \\ &\quad \times \left\{ \frac{1}{n_j T_j} \sum_{i_1=1}^{n_j} \sum_{m_1=1}^{T_j} \sigma_{\varepsilon_{ji_1}}^2(t_{ji_1 m_1}) \left(K^{(2)} \left(\frac{t_{jim} - t_{ji_1 m_1}}{h_j} \right) \right)^2 \right\} \\ &= \max_{\substack{1 \leq j \leq k \\ 1 \leq i \leq n_j}} \left\{ \frac{1}{n_j T_j} \sum_m \{1 - W_j(t_{jim})\}^2 \sigma_{\varepsilon_{ji}}^2(t_{jim}) \sigma_{\varepsilon_j}^{-2}(t_{jim}) f_j^{-1}(t_{jim}) \varpi^2(t_{jim}) \right\} \\ &\quad \times \{R(K)^{-2} K_1^{(4)}(0)\} h_j = O(n^{-1}h). \end{aligned}$$

And similarly, $\max_{\substack{1 \leq j \leq k \\ 1 \leq i \leq n_j}} \left\{ \sum_{j_1=1}^k \frac{1}{n_j n_{j_1}} \sum_{i_1=1}^{n_{j_1}} \sigma_{2,jj_1i_1}^2 \right\} = O(n^{-1}h)$. These imply (4.9.47).

It is remain to check (4.9.48). By (4.9.46), we have

$$E(U_N^4) = E(\mathcal{T}_{n1}^{(12)})^4 - 4E\{(\mathcal{T}_{n1}^{(12)})^3\mathcal{T}_{n1}^{(2)}\} + 6E\{(\mathcal{T}_{n1}^{(12)})^2(\mathcal{T}_{n1}^{(2)})^2\} - 4E\{\mathcal{T}_{n1}^{(12)}(\mathcal{T}_{n1}^{(2)})^3\} + E(\mathcal{T}_{n1}^{(2)})^4. \quad (4.9.54)$$

It can be seen that $E\{(\mathcal{T}_{n1}^{(12)})^3\mathcal{T}_{n1}^{(2)}\} = E\{\mathcal{T}_{n1}^{(12)}(\mathcal{T}_{n1}^{(2)})^3\} = 0$. At the same time, we observed that

$$\begin{aligned} E(\mathcal{T}_{n1}^{(12)})^4 &= E\left\{\sum_{j=1}^k \sum_{i \neq i_1}^{n_j} n_j^{-4} \left[\int_0^1 \{1 - W_j(t)\} V_j^{-1}(t) \xi_{ji}(t) \xi_{ji_1}(t) \varpi(t) dt \right]^4\right\} \\ &\quad + 3E\left\{\sum_{j=1}^k \sum_{i \neq i_1}^{n_j} n_j^{-2} \left[\int_0^1 \{1 - W_j(t)\} V_j^{-1}(t) \xi_{ji}(t) \xi_{ji_1}(t) \varpi(t) dt \right]^2\right. \\ &\quad \times \left. \sum_{j_1=1}^k \sum_{\substack{i_2 \neq i_3 \\ i \neq i_1, i_1 \neq i_3}}^{n_{j_1}} n_{j_1}^{-2} \left[\int_0^1 \{1 - W_{j_1}(t)\} V_{j_1}^{-1}(t) \xi_{j_1 i_2}(t) \xi_{j_1 i_3}(t) \varpi(t) dt \right]^2\right\} + o(h). \end{aligned} \quad (4.9.55)$$

The term marked by (4.9.55) is $O(n^{-2})$, hence is negligible; and the second term on the right hand side converges to $3\{\text{Var}(\mathcal{T}_{n1}^{(12)})\}^2$. Similarly, we can show that $E(\mathcal{T}_{n1}^{(2)})^4 \rightarrow 3\{\text{Var}(\mathcal{T}_{n1}^{(2)})\}^2$ and $6E\{(\mathcal{T}_{n1}^{(12)})^2(\mathcal{T}_{n1}^{(2)})^2\} \rightarrow 6\text{Var}(\mathcal{T}_{n1}^{(12)})\text{Var}(\mathcal{T}_{n1}^{(2)})$. From (4.9.54),

$$\lim_{n \rightarrow \infty} \{\text{Var}(U_N)\}^{-2} E(U_N^4) = \lim_{n \rightarrow \infty} 3\{\text{Var}(U_N)\}^{-2} \{\text{Var}(\mathcal{T}_{n1}^{(12)}) + \text{Var}(\mathcal{T}_{n1}^{(2)})\}^2 = 3.$$

Therefore, (4.9.48) is verified and then we have the asymptotic normality of U_N .

In summary of (4.9.42), (4.9.46) and (4.9.53),

$$h^{-1/2}\{\mathcal{T}_{n1} - (k-1)\} \xrightarrow{d} N(0, \sigma_0^2). \quad (4.9.56)$$

Let us consider $\mathcal{T}_{n2} = \int_0^1 \sum_{j=1}^k v_j^{-1} A_n^2(t) \varpi(t) dt$. Recall the definition of $A_n(t)$ in (4.9.40). From Assumption A2(iii) that there exist finite number a_j and b_j such that $n_j T_j h_j = (a_j b_j)^{-1} n T h$ and $C_{jn} = a_j^{-1/2} b_j^{-1/4} (nT)^{-1/2} h^{-1/4}$. and Then it can be shown that $h^{-1/2}(\mathcal{T}_{n2} - \mu_1) = o_p(1)$ and where

$$\mu_1 = \int_0^1 \left[\sum_{j=1}^k b_j^{-\frac{1}{2}} V_j^{-1}(t) f_j^2(t) \Delta_{n_j}^2(t) - \left(\sum_{s=1}^k b_s^{-\frac{1}{4}} V_s(t)^{-\frac{1}{2}} W_s^{\frac{1}{2}}(t) f_s(t) \Delta_{n_s}^2(t) \right)^2 \right] \varpi(t) dt.$$

It remains to consider $\mathcal{T}_{n3} = 2 \int_0^1 \sum_{j=1}^k v_j^{-1} A_n(t) B_n(t) \varpi(t) dt$. Using the expression of $A_n(t)$ and $B_n(t)$, we can decompose \mathcal{T}_{n3} as

$$\begin{aligned} \mathcal{T}_{n3} &= 2 \int_0^1 \sum_{j=1}^k v_j^{-1} d_j C_{nj} \Delta_{nj}(t) \sum_{i=1}^{n_j} R_{ji} \{g_{j0}(t)\} \varpi(t) dt \\ &\quad - 2 \int_0^1 \left(\sum_{j=1}^k v_j^{-1} d_j^2 \right)^{-1} \left(\sum_{j=1}^k v_j^{-1} d_j^2 C_{nj} \Delta_{nj}(t) \right) \left(\sum_{s=1}^k v_s^{-1} d_s \sum_{i=1}^{n_s} R_{si} \{g_{s0}(t)\} \right) \varpi(t) dt \\ &:= \mathcal{T}_{n3}^{(1)} - \mathcal{T}_{n3}^{(2)}, \quad \text{say.} \end{aligned}$$

We know that

$$\begin{aligned} \mathcal{T}_{n3}^{(1)} &= 2 \int_0^1 \sum_{j=1}^k V_j(t)^{-1} f_j(t) C_{nj} \Delta_{nj}(t) \sum_{i=1}^{n_j} R_{ji} \{g_{j0}(t)\} \varpi(t) dt \\ &= 2 \sum_{j=1}^k h_j^{3/4} \mathcal{T}_{n3}^{(1j)} \{1 + o_p(1)\} \end{aligned}$$

Thus we can say $\mathcal{T}_{n3}^{(1)}$ is $O_p(h^{3/4})$ if we can show $\mathcal{T}_{n3}^{(1j)} = O_p(1)$. It is sufficient to show that $\text{Var}(\mathcal{T}_{n3}^{(1j)}) = O(1)$. Indeed, after some algebra, we get

$$\text{Var}(\mathcal{T}_{n3}^{(1j)}) = R(K)^{-2} \int_0^1 \sigma_{\varepsilon_j}(y) \Delta_j^2(y) dy \int K^{(2)}(z) dz \{1 + o(1)\} = O(1).$$

Therefore $\mathcal{T}_{n3}^{(1)} = O_p(h^{3/4})$. The second term of \mathcal{T}_{n3} , $\mathcal{T}_{n3}^{(2)}$ can be written in a similar form as $\mathcal{T}_{n3}^{(1)}$, which is also $O_p(h^{3/4})$. Thus $\mathcal{T}_{n3} = O_p(h^{3/4})$. In summary of these and (4.9.56), $h^{-1/2} \mathcal{T}_{n3} = o_p(1)$, we have via Slutsky Theorem, $h^{-1/2}(\mathcal{T}_n - (k-1) - \mu_1) \xrightarrow{d} N(0, \sigma_0^2)$. Thus the proof is completed. \square

Proof of Theorem 5 We want to establish the bootstrap version of Theorem 3. To avoid repetition, we only outline some important steps in proving this theorem.

We use $v_j^*(t, h_j)$ and $d^*(t, h_j)$ to denote the bootstrap counterparts of $v_j(t, h_j)$ and $d(t, h_j)$ respectively. Let $o_p^*(1)$ and $O_p^*(1)$ be the stochastic order with respect to the conditional probability measure given the original samples.

We want to show first that

$$(n_j h_j T_j)^{-1} v_j^*(t, h_j) - V^*(t) = o_p^*(1), \quad \text{as } n_j \rightarrow \infty. \quad (4.9.57)$$

where $V^*(t) = R(K)\hat{\sigma}_{\varepsilon_j}^2 f_j(t)$. This can be seen from the following decomposition,

$$\begin{aligned} \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} R_{ji}^{*2} \{\hat{g}_1(t)\} &= \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left(\sum_{m=1}^{T_j} \nu_{jim}^* K \left(\frac{t_{jim} - t}{h_j} \right) \hat{\varepsilon}_{jim} \right)^2 \\ &+ \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left(\sum_{m=1}^{T_j} \nu_{jim}^* K \left(\frac{t_{jim} - t}{h_j} \right) \mathbb{X}_{jim}^\tau (\hat{\xi}_j - \hat{\xi}_j^*) \right)^2 \\ &+ \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left(\sum_{m=1}^{T_j} \nu_{jim}^* K \left(\frac{t_{jim} - t}{h_j} \right) \{(\hat{g}_1(t_{jim}) - \hat{g}_1(t)) - (\hat{g}_j^*(t_{jim}) - \hat{g}_j^*(t))\} \right)^2 + o_p^*(1) \\ &:= A_1^* + A_2^* + A_3^* + o_p^*(1), \end{aligned}$$

where $\nu_{jim}^* = \frac{\delta_{jim}^*}{\pi_{jim}(\hat{\theta}_j^*)} = \frac{\delta_{jim}^*}{\pi_{jim}(\hat{\theta}_j)} \left(1 - \frac{\pi_{jim}(\hat{\theta}_j^*) - \pi_{jim}(\hat{\theta}_j)}{\pi_{jim}(\hat{\theta}_j)} \right)$. Then we can apply $\pi_{jim}(\hat{\theta}_j) - \pi_{jim}(\hat{\theta}_j^*) = O_p^*(n_j^{-1/2})$, $\hat{\xi}_j - \hat{\xi}_j^* = O_p^*(n_j^{-1/2})$ and $\hat{g}_j'(t) - \hat{g}_j^{*'}(t) = o_p^*(1)$ to A_2^* and A_3^* . By the similar procedure as we derive expression for $v_j(t, h_j)$ in the proof of Theorem 3, we can get (4.9.57).

Corresponding to the leading term of \mathcal{T}_n , the leading term of \mathcal{T}_n^* is

$$\begin{aligned} &\int_0^1 \sum_{j=1}^k v_j^{*-1} \left[\sum_{i=1}^{n_j} R_{ji}^* \{\hat{g}_1(t)\} - d_j^* \left(\sum_{s=1}^k v_s^{*-1} d_s^{*2} \right)^{-1} \sum_{s=1}^k v_s^{*-1} d_s^* \sum_{i=1}^{n_s} R_{si}^* \{\hat{g}_1(t)\} \right]^2 \varpi(t) dt \\ &= \int_0^1 \sum_{j=1}^k \{1 - W_j^*(t)\} V_j^{*-1}(t) S_{j1}^{*2}(t) \varpi(t) dt + \left\{ \int_0^1 \sum_{j=1}^k \{1 - W_j^*(t)\} V_j^{*-1}(t) S_{j2}^{*2}(t) \varpi(t) dt \right. \\ &\quad \left. - \int_0^1 \sum_{j \neq j_1} \left[\sum_{s=1}^k v_s^{*-1} d_s^{*2} \right]^{-1} v_j^{*-1} d_j^* v_{j_1}^{*-1} d_{j_1}^* \left[\sum_{i=1}^{n_j} R_{ji}^* \{\hat{g}_1(t)\} \right] \left[\sum_{i=1}^{n_s} R_{si}^* \{\hat{g}_1(t)\} \right] \varpi(t) dt \right\} \\ &:= B_1^* + B_2^*, \end{aligned}$$

where $W_j^*(t) = \frac{f_j(t)/\{a_j b_j \hat{\sigma}_{\varepsilon_j}^2\}}{\sum_{i=1}^k f_i(t)/\{a_i b_i \hat{\sigma}_{\varepsilon_i}^2\}}$, $S_{j1}^{*2}(t)$ and $S_{j2}^{*2}(t)$ are the bootstrap version of $S_{j1}^2(t)$ and $S_{j2}^2(t)$ defined in the proof of Theorem 3. Then, using a similar approach to the one used in establishing the asymptotic normality of \mathcal{T}_{n1} in (4.9.41) in the proof of Theorem 3. We may show that

$$h^{-1/2} \{B_1^* - (k-1)\} = o_p^*(1) \quad \text{and} \quad h^{-1/2} B_2^* | \mathcal{X}_n \xrightarrow{d} N(0, \sigma_0^2) \quad a.s.$$

Hence, Theorem 5 is established. \square

References

- Cao, R. and van Keilegom, I. (2006). Empirical likelihood tests for two-sample problems via nonparametric density estimation. *The Canadian Journal of Statistics*, 34, 61-77.
- Chen, S.-X. and Cui, H.-J. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika*, 93, 215-220.
- Chen, S.-X. and Hall, P. (1994). On the calculation of standard error for quotation in confidence statements. *Statistics & Probability Letter*, 19, 147-151.
- Chen, S.-X., Leung, D. and Qin, J. (2008). Improved semiparametric estimation using surrogate data. *Journal of Royal Statistical Society, Series B*, 70, 803-823.
- Chen, S.-X. and van Keilegom, I. (2009). A review on empirical likelihood for regressions. *Test*, 18, 415-447.
- de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory Related Fields*, 75, 261-277.
- DiCiccio, T., Hall, P. and Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *The Annals of Statistics*, 19, 1053-1061.
- Fan, J., Huang, T., and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*, 102, 632-641.
- Fan, J. and Zhang, J. (2004). Sieve empirical likelihood ratio tests for nonparametric functions. *The Annals of Statistics*, 32, 1858-1907.

- Fitzmaurice, G., Laird, N. and Ware, J. (2004). *Applied longitudinal analysis*. John Wiley & Sons.
- Forcina, A. (1992). Modelling balanced longitudinal data: Maximum likelihood estimation and analysis of variance. *Biometrics*, 48, 743-750.
- Härdle, W., Liang, H. and Gao, J. (2000). *Partially linear models*. Physica-Verlag Heidelberg.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21, 1926-1947.
- Henry, K., Erice, A., Tierney, C., Balfour, H., Fischl, M., Kmack, A., Liou, S., Kenton, A., Hirsch, M., Phair, J., Martinez, A. and Kahn, J. for the AIDS Clinical Trial Group 193A Study Team (1998). A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-Drug, two-Drug, and alternating-Drug) for the treatment of advanced AIDS. *Journal of Acquired Immune Deficiency Syndromes And Human Retrovirology*, 19, 339-349.
- Huang, C. Y., Qin, J. and Follmann, D. (2008). Empirical likelihood-based estimation of the treatment effect in a pretest-posttest study. *Journal of the American Statistical Association*, 103, 1270-1280.
- Laird, N. (2004). *Analysis of longitudinal and cluster-correlated data*. Region conference series in probability and statistics.
- Laird, N. and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.
- Liang, H. and Qin, Y.-S. (2008). Empirical likelihood-based inferences for partially linear models with missing covariates. *The Australian & New Zealand Journal of Statistics*, 50, 347-359.
- Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

- Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 96, 103-126.
- Linton, O. and Nielsen J. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82, 93-100.
- Liu, R. (1988). Bootstrap procedures under some non-I.I.D. models. *The Annals of Statistics*, 16, 1696-1708.
- Little, R. and Rubin, B. (2002). *Statistical analysis with missing data*. Second edition. John Wiley and Sons.
- Little, R. (1995). Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association*, 90, 1112-1121.
- Müller, H. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *Journal of the American Statistical Association*, 82, 231-238.
- Mund, A. and Dettle, H. (1998). Nonparametric comparison of several regression functions: exact and asymptotic theory, *The Annals of Statistics*, 26, 2339-2368.
- Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75, 237-249.
- Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18, 90-120.
- Owen, A. (2001). *Empirical likelihood*. Chapman and Hall/CRC.
- Pardo-Fernández, J., Van Keilegom, I. and González-Manteiga, W. (2007). Testing for the equality of k regression curves, *Statistica Sinica*, 17, 1115-1137.
- Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22, 300-325.

- Rice, J. and Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of Royal Statistical Society, Series B*, 53, 233-243.
- Robins, J., Rotnitzky, A. and Zhao, L. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90, 106-121.
- Rubin, B. (1976). Inference and missing data (with discussion). *Biometrika*, 63, 581-592.
- Scheike, T.H. and Zhang, M.J. (1998). Cumulative regression function tests for regression models for longitudinal data. *The Annals of Statistics*, 26, 1328-1355.
- Shao, J. and Sitter, R. (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association*, 91, 1278-1288.
- Shi, J. and Lau, T.S. (2000). Empirical likelihood for partially linear models. *Journal of Multivariate Analysis*, 72, 132-148.
- Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of Royal Statistical Society, Series B*, 50, 413-436.
- Wang, L., Akritas, M., and Van Keilegom, I. (2008). An ANOVA-type nonparametric diagnostic test for heteroscedastic regression Models. *Journal of Nonparametric Statistics*, 20, 365-382.
- Wang, N., Carroll, R. and Lin, X. (2005). Efficient marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association*, 100, 147-157.
- Wang, Q.-H., Linton, O. and Härdle, W. (2004). Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association*, 99, 334-345.
- Wang, Q.-H. and Rao, J. (2002). Empirical likelihood-based inference under imputation for missing response data. *The Annals of Statistics*, 30, 896-924.
- Wu, C.-F.-J. (1986). Jackknife, Bootstrap and other resampling method in regression analysis. *The Annals of Statistics*, 14, 1261-1295.

- Wu, C., Chiang, C. and Hoover, D. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association*, 93, 1388-1402.
- Wu, C. and Chiang, C. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statistica Sinica*, 10, 433-456.
- Xue, L. and Zhu, L. (2007a). Empirical likelihood semiparametric regression analysis for longitudinal data. *Biometrika*, 94, 921-937.
- Xue, L. and Zhu, L. (2007b). Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of the American Statistical Association*, 102, 642-654.
- You, J., Chen, G. and Zhou, Y. (2006). Block empirical likelihood for longitudinal partially linear regression models. *The Canadian Journal of Statistics*, 34, 79-96.
- Zeger, S. and Diggle, P. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, 50, 689-699.
- Zhang, D., Lin, X., Raz, J. and Sowers, J. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, 93, 710-719.

CHAPTER 5. Summary and General Discussion

In Chapter 2 and 3 of this thesis, we discussed the simultaneous tests for high dimensional data. In Chapter 2, we considered the simultaneous test for regression coefficients, while the high dimensional test for high dimensional mean under sparsity and dependency was proposed in Chapter 3. The test proposed for regression coefficients is very powerful against the non-sparse alternative, i.e., a large number of predictors are associated with the response and most of which account for only a small effect. The test proposed in Chapter 3 was designed to detect the sparse alternatives where we have a prior information that only a small portion of the alternatives are different from the null hypothesis.

The test we developed in Chapter 3 is for the high dimensional means. One of the future directions is to develop a high dimensional test for regression coefficients under sparsity assumption. The difficulty in high dimensional regression context ($p \gg n$) comparing to that in means is that the explicitly consistent estimate of coefficients is not available and has to be found by minimizing a penalized likelihood. Therefore, the method based on threshold may not be directly applicable except relatively strong conditions are assumed for the design matrix (Arias-Castro *et al.*, 2010). A more appealing method for constructing test statistic in such case may base on the likelihood ratio statistics. It would be interesting to investigate the distribution or asymptotic distribution of the penalized likelihood ratio statistics under the high dimensional null hypothesis and alternative hypothesis. For the mild dimension case, Fan and Peng (2004) proposed a parametric penalized likelihood ratio ($p^5/n \rightarrow 0$). Tang and Leng (2010) proposed a penalized empirical likelihood ratio for making inference ($p^2/n \rightarrow 0$). But their methods are only applicable when the data dimension is smaller than the sample size. It is worthwhile to explore how to extend these methods to “large p , small n ” cases, where we have dimension higher than sample size.

Sparsity is often assumed in the current high dimensional statistical inference literature, especially in the variable selection context. The assumption serves as a simplification of the possible complex structure in high dimensional problems. In the regression context, the sparsity increases the estimation accuracy and makes the consistent variable selection possible. In the dimension reduction context, the sparsity means effective representation lying in a low dimension. However, in practice, it is often not clear if a sparsity or non-sparsity model is more appropriate. So it would be interesting to develop a test which can perform well under both sparse and non-sparse scenarios in the future.

In Chapter 4, we proposed empirical likelihood ratio based test statistics for comparing treatment effects in longitudinal data, including ANOVA tests for cohort effects and time effects. In our proposal, the number of repeated measurement for each individual is assumed fixed but the sample size growing. In contrast to our setting, some papers in the literature considered the growing number of repeated measurements but keep the sample size fixed (for example, Fan and Zhang, 2000). Under their setting, the longitudinal data can be regarded as functional data (Ramsey and Silverman, 1997). Therefore the functional ANOVA could be employed to analyze such data. However, most of the longitudinal data only have very limited repeated measurements. It would be interesting to investigate how to apply the functional ANOVA technique to longitudinal data with a small number repeated measurements. Recently developments in applying functional data analysis method to sparsely sampled longitudinal data can be found in Hall, Müller and Wang (2006) and Yao, Müller and Wang (2005) among others.

References

- Arias-Castro, E., Cadès, E. and Plan, Y. (2010). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Manuscript*.
- Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, 32, 928-961.
- Fan, J. and Zhang, J. (2000). Two-step estimation of functional linear models with application to longitudinal data. *Journal of Royal Statistical Society, Series B*, 62, 303-322.
- Hall, P., Müller, H. and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, 34, 1493-1517.
- Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*. Springer-Verlag, New York.
- Tang, C.Y. and Leng, C. (2010). Penalized high dimensional empirical likelihood. *Biometrika*, 97, 905-920.
- Yao, F., Müller, H. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577-590.