2011

Topics on high dimensional statistical inference and ANOVA for longitudinal data

Pingshou Zhong
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/etd/12245

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
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
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table 2.1</th>
<th>Empirical size and power of the F-test, the EB test and the proposed test (new) for $H_0 : \beta = 0_{p \times 1}$ vs $H_1 : \beta \neq 0_{p \times 1}$ at significant level 5% for normal residual. LP represents the theoretical local power.</th>
<th>39</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.2</td>
<td>Empirical size and power of the F-test, the EB test and the proposed test (new) for $H_0 : \beta = 0_{p \times 1}$ vs $H_1 : \beta \neq 0_{p \times 1}$ at significant level 5% for centralized gamma residual. LP represents the theoretical local power.</td>
<td>40</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Empirical size and power of the EB test and the proposed test (new) for $H_0 : \beta = 0_{p \times 1}$ vs $H_1 : \beta \neq 0_{p \times 1}$ at significant level 5% for normal residual. LP represents the theoretical local power.</td>
<td>41</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>Empirical size and power of the EB test and the proposed test (new) for $H_0 : \beta = 0_{p \times 1}$ vs $H_1 : \beta \neq 0_{p \times 1}$ at significant level 5% for centralized gamma residual. LP represents the theoretical local power.</td>
<td>42</td>
</tr>
<tr>
<td>Table 2.5</td>
<td>Empirical size and power of the proposed test for $H_0 : \beta = 0_{p \times 1}$ in a 2 × 2 factorial design with $n_1 = 20$ and $n_2 = 30$ replicates in each cell.</td>
<td>44</td>
</tr>
<tr>
<td>Table 2.6</td>
<td>P-values of the GO terms which are significant under at least three designs using the proposed test, and their number of genes.</td>
<td>49</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>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.</td>
<td>87</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>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.</td>
<td>88</td>
</tr>
</tbody>
</table>
Table 4.1 Empirical size and power of the 5% ANOVA test for \( H_{0a} : \beta_{10} = \beta_{20} = \beta_{30} \).

Table 4.2 Empirical size and power of the 5% test for the existence of interaction
\( H_{0c} : \gamma_{20} = 0 \).

Table 4.3 Empirical size and power of the 5% ANOVA test for \( H_{0d} : \gamma_{10} = \gamma_{20} = \gamma_{30} \).

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 \( \Delta_{3n}(t) = 2 \sin(2\pi t) - 2 \sin(2\pi(t + V_n)) \).

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 \( H_{1b} : g_1(\cdot) = g_2(\cdot) + \Delta_{2n}(\cdot) \).

Table 4.6 Differences in the AIC and BIC scores among three models (M1-M3).

Table 4.7 Parameter estimates and their standard errors.

Table 4.8 P-values of ANOVA tests for \( \beta_j \)s.

Table 4.9 P-values of ANOVA tests on \( g_j(\cdot) \)s.

Table 4.10 Differences in the AIC and BIC scores among three models (M1-M3) for \( d = 1 \).

Table 4.11 Parameter estimates and their standard errors with \( d = 1, 2, 3 \).

Table 4.12 P-values of ANOVA tests on \( \beta \)s with \( d = 1, 2, 3 \).

Table 4.13 P-values of ANOVA tests on \( g_j(\cdot) \)s with \( d = 1, 2, 3 \).
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 \((\beta,r)\) 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 Silverman (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 $p^2/n \to 0$ and if $p^3/2n \to 0$, the likelihood ratio test statistic $\Lambda_n$ for a simple hypothesis has a chi-square distribution in the sense that $(-2 \log(\Lambda_n) - p) / \sqrt{2p} \to N(0,1)$. 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 $p/n \to c$ for a finite constant $c$. They found the sphericity test is robust under $p/n \to c$, 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 Srivastavaa 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 $p^5/n \to 0$.

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 $X_i = (X_{i1}, \ldots, X_{ip})'$ 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

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

(1.2.1)

where $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ 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),

$$\sqrt{\frac{n}{2p}} T_n \overset{d}{\to} N(0,1).$$

(1.2.2)

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

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

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

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

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 $Y = (Y_1, \cdots, Y_n)$, $\epsilon = (\epsilon_1, \cdots, \epsilon_n)$ and $X = (X_1, \cdots, X_n)'$ 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

\[
H_0 : \beta = 0 \quad \text{vs} \quad H_1 : \beta \neq 0. 
\] (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 \( \Psi \) is a given function satisfying some conditions. They showed that \( \hat{\beta} \) is still consistent if \( p \log(p)/n \to 0 \) and further if \( \{p \log(p)\}^{3/2}/n \to 0 \), \( \frac{d_\Psi}{\sigma_\Psi^2} (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \xrightarrow{d} \chi^2_p \), 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 \( \alpha \) if

\[
\frac{d_\Psi}{\sigma_\Psi^2} \hat{\beta}'X'X\hat{\beta} > \chi^2_{p, \alpha},
\]

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

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 \( \beta \) is a random vector with mean 0 and covariance \( \tau^2 I_p \). Then testing hypothesis of (1.2.4) is equivalent to testing

\[
H_0 : \tau^2 = 0 \quad \text{vs} \quad H_1 : \tau^2 \neq 0. \tag{1.2.5}
\]

Suppose that \( Y \sim N(X\beta, \sigma^2 I) \) and \( L(\beta; Y) \) be the likelihood of \( \beta \) given \( Y \). Then the marginal density of \( Y \) is given by

\[
\hat{f}(\tau^2; y) = E_{\beta|\tau^2}\{L(\beta; Y)\},
\]

where \( E_{\beta|\tau^2}\{\cdot\} \) is the expectation with respect to prior distribution of \( \beta \) given \( \tau^2 \). A score test (Goeman et al., 2006) can be used to test (1.2.5), which is

\[
S = \frac{\partial \log \{\hat{f}(0; Y)\}}{\partial \tau^2}.
\]

The test statistics is equivalent to \( S' = \frac{X'X'y}{y'y} \). 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 \( t_1, \cdots, t_p \) and \( X_i = (X_i(t_1), \cdots, X_i(t_p))' \), 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 \( t = t_1, \cdots, t_p \). 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 \( \beta \). 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 \( X_1, \cdots, X_n \) are IID \( p \)-dimensional vectors with mean \( \mu \). We consider testing for the high dimensional mean under sparsity i.e., testing

\[ H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0 \quad (1.2.6) \]

where \( \mu = (\mu_1, \cdots, \mu_p)' \) and we assume that only a small fraction \( p^{-\beta} (\beta > 1/2) \) of the \( p \) components of \( \mu \) 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 \((\beta, r)\) 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 of 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 $\frac{nT_{CQ}}{\sqrt{2\pi \Sigma}} \overset{d}{\rightarrow} N(0, 1)$ 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

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

where \( \beta_j \)'s are the treatment effects, \( \alpha_m \)'s are the time effects, \( (\alpha\beta)_{jm} \)'s are the interactions between time and treatment, \( \omega_{i(j)} \overset{iid}{\sim} N(0, \sigma^2_w) \) are the whole plot errors and \( \varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2) \) are split plot error, \( \omega_{i(j)} \) and \( \varepsilon_{ijk} \) 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,

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

where \( \{Y_{ji}(t), X_{ji}(t)\} \) 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 \( \beta_{j0} \)'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 $X_1, \ldots, X_n$ 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 \( \theta(F) \in \mathbb{R}^p \) which could be defined by \( q \) dimensional estimating equation \( g(X; \theta) \) such that \( E g(X; \theta_0) = 0 \), where \( \theta_0 \) is the true value of \( \theta \). For example, \( \theta(F) = E(X) = \int x dF(x) \) and \( g(X; \theta) = X - \theta \) such that \( E g(X; \theta_0) = 0 \). The just identifiable and over identified cases corresponding to \( q = p \) and \( q > p \) respectively. Assuming that \( 0 \) is in the convex hull of \( g(X_1; \theta), \ldots, g(X_n; \theta) \), the empirical likelihood ratio for \( \theta \) is defined as

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

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 \( \lambda \) satisfies

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

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

\[
\ell(\theta_0) - \ell(\hat{\theta}) \xrightarrow{d} \chi^2_p.
\]

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, \( \ell(\theta_0) \xrightarrow{d} \chi^2_p \).

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

\[
R(\theta_1, \ldots, \theta_k) = \sup \left\{ \prod_{j=1}^{k} \prod_{i=1}^{n_j} n_j p_{ji} : \sum_{i=1}^{n} p_{ji} \psi_j(X_{ji}; \theta_j) = 0, \sum_{i=1}^{n} p_{ji} = 1, p_{ji} \geq 0 \right\}.
\]
Then an empirical likelihood ratio test statistic for testing $H_0 : \theta_1 = \theta_2 = \cdots = \theta_k$ is

$$\ell_n : = -2 \log \left\{ \frac{\max_{\theta_1, \ldots, \theta_k} R(\theta_1, \ldots, \theta_k)}{\max_{\theta_1, \ldots, \theta_k} R(\theta_1, \ldots, \theta_k)} \right\}$$

$$= -2 \max_{\theta_1, \ldots, \theta_k} \log\{R(\theta_1, \ldots, \theta_k)\} + 2 \max_{\theta_1, \ldots, \theta_k} \log\{R(\theta_1, \ldots, \theta_k)\}$$

$$= 2 \min_{\theta_1, \ldots, \theta_k} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \log\{1 + \lambda_j' \psi_j(X_{ji}; \theta_j)\}$$

(1.4.8)

where $\lambda_j$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 $\max_{\theta_1, \ldots, \theta_k} \log\{R(\theta_1, \ldots, \theta_k)\} = 0$. It is shown in Chapter 4 that $\ell_n \rightarrow \chi^2_{(k-1)p}$ under the null hypothesis. We also consider the ANOVA test for time effect functions, i.e., comparing nonparametric functions, $H_0 : g_1(\cdot) = \cdots = g_k(\cdot)$. 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 $t$,

$$L_n\{g(t)\} : = 2 \min_{g_1(t) = \cdots = g_k(t)} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \log\{1 + \eta_j \varphi_j(X_{ji}; g_j(t))\}$$

(1.4.9)

where $\eta_j$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

$$T_n = \int_0^1 L_n(g(t)) \varpi(t) dt$$

(1.4.10)

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

The asymptotic distribution of $T_n$ 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 $p/n \to c \in (0, 1)$. 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


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.


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


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

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

(2.2.1)

for \( i = 1, \ldots, n \) where \( X_1, \ldots, X_n \) are independent and identically distributed \( p \)-dimensional covariates and \( Y_1, \ldots, Y_n \) are independent responses, \( \beta \) is the vector of regression coefficients, and \( \alpha \) 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 \((\alpha, \beta)\) in the linear regression model is defined as

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

To make \( \beta \) 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 \( \Sigma \) 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

\[ H_0 : \beta = \beta_0 \quad \text{vs} \quad H_1 : \beta \neq \beta_0 \]  

(2.2.2)

for a specific \( \beta_0 \in \mathbb{R}^p \). 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 \( p/n \to \rho \in (0, 1) \) when both \( p \) and \( n \) diverge to infinity.

Let \( U = (1, X) \) which is assumed to be of full rank and \( A = (0, I_p) \), where \( 1 \) 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

$$G_{n,p} = \frac{(\hat{\gamma} - \gamma_0)'(A(U'U)^{-1}A')^{-1}A(\hat{\gamma} - \gamma_0)/p}{Y'(I_n - P_U)Y/(n - p - 1)}$$

$$= \frac{(\hat{\beta} - \beta_0)'(A(U'U)^{-1}A')^{-1}(\hat{\beta} - \beta_0)/p}{Y'(I_n - P_U)Y/(n - p - 1)}$$

(2.2.3)

where $\hat{\gamma} = (\hat{\alpha}, \hat{\beta})' = (U'U)^{-1}U'Y$ is the least square estimator of $\gamma$ and $Y = (Y_1, \ldots, Y_n)'$.

Under $H_0$, $G_{n,p} \sim F_{p, n-p-1}$. Hence, an $\alpha$-level F-test rejects $H_0$ if $G_{n,p} > F_{p, n-p-1, \alpha}$, the upper $\alpha$ 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

There exists a $m$-variate random vector $Z_i = (Z_{i1}, \ldots, Z_{im})'$ for some $m \geq p$ so that $X_i = \Gamma Z_i + \mu$, where $\Gamma$ is a $p \times m$ matrix such that $\Gamma \Gamma' = \Sigma$, and $E(Z_i) = 0$, $\text{Var}(Z_i) = I_m$; each $Z_{il}$ has finite 8-th moment, $E(Z_{il}^8) = 3 + \Delta$ for some constant $\Delta$;

for any $\sum_{\nu=1}^{d} \ell_{\nu} \leq 8$ and $i_1 \neq \cdots \neq i_d$, $E(Z_{i_1}^{\ell_{i_1}} Z_{i_2}^{\ell_{i_2}} \cdots Z_{i_d}^{\ell_{i_d}}) = E(Z_{i_1}^{\ell_{i_1}})E(Z_{i_2}^{\ell_{i_2}}) \cdots E(Z_{i_d}^{\ell_{i_d}})$.

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 $Y_i|X_i \sim N(X_i'\Sigma, \sigma^2)$, 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

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

We notice that the denominator of the F statistic (2.2.3) estimates $\sigma^2$. When $p$ is closer to $n$, there are fewer degrees of freedom left to estimate $\sigma^2$. The impact of the dimensionality on
the F-test is revealed in Theorem 1 by $\sqrt{(1 - \rho)}/\rho$ being a decreasing function of $\rho$. Hence, the power is adversely impacted by an increased dimension even $p < n - 1$, reflecting a reduced degree of freedom in estimating $\sigma^2$ 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 $\beta$ and $\beta_0$. We note that when $\alpha = 0$, $||Y - X\beta_0||^2$ is a measure between $\beta$ and $\beta_0$, 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 $\alpha$, we consider a U-statistic

$$T_{n,p} = \frac{1}{P_{m}^{n}} \sum_{i_1, i_2, i_3, i_4}^{*} \phi(i_1, i_2, i_3, i_4),$$

where

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

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 $\{i_1 \neq i_2 \neq i_3 \neq i_4, \text{ for } i_1, i_2, i_3, i_4 \in \{1, \cdots, n\}\}$ and $P_{m}^{n} = n!/(n-m)!$. 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

$$p(n) \to \infty \text{ as } n \to \infty, \Sigma > 0 \text{ and } tr(\Sigma^4) = o(tr^2(\Sigma^2)).$$
These conditions do not impose any explicit relative growth rates between \( p \) and \( n \), and they are quite mild. Assuming \( \Sigma \) being positive definite assures the identification of the regression coefficient. We allow some eigenvalues of \( \Sigma \) diverge to infinity as \( p \to \infty \). 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 \( W_1, W_2, \ldots, W_n \) are independent and identically distributed observations from a distribution \( F \) on \( \mathbb{R}^q \), where \( q \) may diverge. Consider a U-statistic of \( s \)-th order for a fixed \( s < n \)

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

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

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

\[
M_{nc} = \sum_{1 \leq i_1 < \cdots < i_c \leq n} g_c(w_{i_1}, \ldots, w_{i_c}).
\]

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

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

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

(2.3.9)
and (iii)
\[
\text{Var}(U_{n,q}) = \binom{n}{s}^{-1} \sum_{c=1}^{s} \binom{s}{c} \binom{n-s}{s-c} \zeta_c. \tag{2.3.10}
\]

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 \( E\{h_{c+1}(w_1, \cdots, w_c, W_{c+1})\} = h_c(w_1, \cdots, w_c) \) and
\[
\zeta_{c+1} = E\{\text{Var}(h_{c+1}(W_1, \cdots, W_{c+1})|W_1, \cdots, W_c)\} + \zeta_c.
\]

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 \( \sigma \)-field generated by \( \{W_1, \ldots, W_c\} \) and \( \text{Var}(M_{nc}) = O(\zeta_c) \).

When \( q \to \infty \), unlike the fixed dimension cases, \( \zeta_c \) 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 \( \zeta_c \) 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, \( \zeta_c \) may diverge and a higher order projection \( M_{nc(c+1)} \) 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 \( \zeta_c \) is monotone non-decreasing, the following strategy may be applied to determine the dominant terms of \( U_{n,q} \). We can start evaluating \( \zeta_c \)s from the two ends, namely \( \zeta_1 \) and \( \zeta_s \). If \( \zeta_1 \) and \( \zeta_s \) 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 \( \zeta_s \) and \( \zeta_1 \) are not the same order, but \( \zeta_2 \) and \( \zeta_s \) 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 $\phi$ defined in $\text{(2.2.7)}$ by

$$h(W_i, W_j, W_k, W_l) = \frac{1}{4} \{ \phi(i, j, k, l) + \phi(i, k, j, l) + \phi(i, l, j, k) \}$$

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

$$T_{n,p} = \frac{1}{n} \sum_{i=1}^{n} h(W_i, W_j, W_k, W_l).$$

(2.4.11)

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

$$h_1(w_1) = \frac{1}{2} (\beta - \beta_0)' (x_1 x_1' + \Sigma) (\beta - \beta_0) + \frac{1}{2} \varepsilon_1 x_1' \Sigma (\beta - \beta_0),$$

$$h_2(w_1, w_2) = \frac{1}{6} \left[ (\beta - \beta_0)' (x_1 - x_2) (x_1 - x_2)' \Sigma (\beta - \beta_0) + (\varepsilon_1 - \varepsilon_2) (x_1 - x_2)' \Sigma (\beta - \beta_0) + ((\beta - \beta_0)' (x_1 x_1' + \Sigma) + \varepsilon_1 x_1') (\varepsilon_2 x_2 + (x_2 x_2' + \Sigma) (\beta - \beta_0)) \right]$$

and

$$h_3(w_1, w_2, w_3) = \frac{1}{12} \left[ ((x_1 - x_2)' (\beta - \beta_0) + (\varepsilon_1 - \varepsilon_2)) (x_1 - x_2)' ((x_3 x_3' + \Sigma) (\beta - \beta_0) + x_3 \varepsilon_3) \right. \right.$$  

$$\left. + \frac{1}{12} ((x_1 - x_3)' (\beta - \beta_0) + (\varepsilon_1 - \varepsilon_3)) (x_1 - x_3)' ((x_2 x_2' + \Sigma) (\beta - \beta_0) + x_2 \varepsilon_2) \right. \right.$$  

$$\left. + \frac{1}{12} ((x_2 - x_3)' (\beta - \beta_0) + (\varepsilon_2 - \varepsilon_3)) (x_2 - x_3)' ((x_1 x_1' + \Sigma) (\beta - \beta_0) + x_1 \varepsilon_1) \right].$$

Let $B_i = (\beta - \beta_0)' \Sigma (\beta - \beta_0)$ for $i = 1, 2, 3$, $A_0 = \Gamma \Sigma$, $A_1 = \Gamma' (\beta - \beta_0)' (\beta - \beta_0) \Sigma$, $A_2 = \Gamma' \Sigma (\beta - \beta_0)' (\beta - \beta_0) \Sigma \Gamma$ 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

$$\zeta_1^* = (B_1 + \sigma^2) B_3 + B_2^2 + \Delta tr(A_1 \circ A_2) \quad \text{and}$$

$$\zeta_2^* = \sigma^4 tr(\Sigma^2) + 21 B_2^2 + 22 B_1 B_3 + 22 \sigma^2 B_3 + B_1^2 tr(\Sigma^2) + 2 \sigma^2 tr(\Sigma^2) B_1$$

$$+ 2 \Delta (B_1 + \sigma^2) tr(A_1 \circ A_3) + 20 \Delta tr(A_1 \circ A_2) + \Delta^2 tr((A_0 \text{diag}(A_1))^2),$$

where $C \circ B = (c_{ij} b_{ij})$ for matrices $C = (c_{ij})$ and $B = (b_{ij})$, and $\text{diag}(A) = \text{diag}\{a_{11}, \cdots , a_{mm}\}$ 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 \( n \to \infty \),

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

(ii) \( T_{n,p} - \|\Sigma(\beta - \beta_0)\|^2 = \{4^2/n M_{n1} + 2/n^2 M_{n2}\} \{1 + o_p(1)\} \), where \( E(M_{n1}^2) = \zeta_1 \) 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,

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

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

\[
(\beta - \beta_0)' \Sigma(\beta - \beta_0) = o(1) \quad \text{and} \quad (\beta - \beta_0)' \Sigma^3(\beta - \beta_0) = o\{n^{-1} tr(\Sigma^2)\}. \tag{2.4.12}
\]

As this subclass prescribes a smaller difference between \( \beta \) and \( \beta_0 \), 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 \( n \to \infty \),

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

To formulate a test procedure based on \( T_{n,p} \), we need to estimate \( tr(\Sigma^2) \) and \( \sigma^2 \) appeared in the asymptotic variance. We will use the estimator of \( tr(\Sigma^2) \) proposed in Chen, Zhang and Zhong (2010). Specifically, let \( Y_{1n} = \frac{1}{P_n^2} \sum_i (X_{i1}' X_{i2})^2 \), \( Y_{2n} = \frac{1}{P_n} \sum_i X_{i1}' X_{i2} X_{i2}' X_{i3} \) and \( Y_{3n} = \frac{1}{P_n} \sum_i X_{i1}' X_{i2} X_{i3}' X_{i4} \). Then an unbiased and ratio consistent estimator of \( tr(\Sigma^2) \) is

\[
tr(\Sigma^2) = Y_{1n} - 2Y_{2n} + Y_{3n}.
\]

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

\[
\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - X_i \beta_0 - \bar{Y} + \bar{X} \beta_0)^2. \tag{2.4.14}
\]
Applying Theorem 3 and the Slutsky Theorem, the proposed test rejects \( H_0 \) at a significant level \( \alpha \) if

\[
n T_{n,p} \geq \sqrt{2 \text{tr}(\Sigma^2)\hat{\sigma}^2} z_\alpha,  
\]

where \( z_\alpha \) is the upper-\( \alpha \) 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

\[
\Omega_L(\|\beta - \beta_0\|) = \Phi \left( -z_\alpha + \frac{n\|\Sigma (\beta - \beta_0)\|^2}{\sqrt{2 \text{tr}(\Sigma^2)\sigma^2}} \right).  
\]

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 \( \alpha \) 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 \( \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p \) be the eigenvalues of \( \Sigma \). Then, a sufficient condition for the test to have a non-trivial power is \( \|\beta - \beta_0\| = O(n^{-1/2}S_\chi^{1/4}\lambda_1^{-1}) \) where \( S_\chi = \sum_{i=1}^p \lambda_i^2 \). 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 \( \delta_\beta \) 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

\[
(\beta - \beta_0)'\Sigma (\beta - \beta_0) \text{ is not } o(1)  
\]

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

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

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 \((2.4.17)\) is

\[
\frac{1}{n} (\beta - \beta_0)'\Sigma (\beta - \beta_0) \text{tr}(\Sigma^2) = o \left\{ (\beta - \beta_0)'\Sigma^3 (\beta - \beta_0) \right\}.  
\]
If \((\beta - \beta_0)\Sigma(\beta - \beta_0)\) is bounded, (2.4.19) implies \(\frac{1}{n}tr(\Sigma^2) = o\{(\beta - \beta_0)'\Sigma^3(\beta - \beta_0)\}\), which prescribes a larger discrepancies between \(\beta\) and \(\beta_0\). 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

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

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 \(\zeta_1\), 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)

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

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

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

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

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

\[
\Omega_{H_2}(||(\beta - \beta_0)||) \doteq \Phi \left( -\frac{\sqrt{2tr(\Sigma^2)\sigma^2z_\alpha}}{\sqrt{(n - 1)\sigma_{A_2}^2}} + \frac{\sqrt{n}\|\Sigma(\beta - \beta_0)\|^2}{\sigma_{A_2}} \right). \quad (2.4.22)
\]
As (2.4.19) implies \( \frac{1}{n} \text{tr}(\Sigma^2)/\sigma_{A_2}^2 = o(1) \), the proposed test is consistent as long as

\[
\sqrt{n} \| \Sigma(\beta - \beta_0) \|^2 / \sigma_{A_2}^2 \to \infty. (2.4.23)
\]

Even if \( \sqrt{n} \| \Sigma(\beta - \beta_0) \|^2 / \sigma_{A_2}^2 \) does not converge to \( \infty \), 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 \( c = 1, \cdots, IJ \), 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

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

where \( \gamma_i \) represent for the effect of A, \( \theta_j \) 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

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

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

\[
E(Y_{ijk}|X_{ijk}) = \mu_{ij} + X'_{ijk} \beta, \quad k = 1, \cdots, n_c. (2.5.26)
\]
Define $Y = (Y^1', \ldots, Y^{IJ}')'$, $X = (X^1', X^2', \ldots, X^{IJ}')'$ where

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

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

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

where $D = I_{IJ} \otimes 1_{nc}$ is the design matrix, $\alpha$ corresponding to the cell means parameters $\mu_{ij}$. Multiply $I - P_D$ on both sides of (2.5.27) where $P_D = D(D'D)^{-1}D' = I_{IJ} \otimes n_c^{-1}1_{nc}'1_{nc}$ 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 $\alpha$ in (2.5.27). So a natural generalization of $T_{n,p}$ to the factorial design is

$$T_{n,p} = \frac{1}{IJ} \sum_{c=1}^{IJ} (P_4^{1c})^{-1} \sum_{i,j,k,l} \phi(i,j,k,l), \quad (2.5.28)$$

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

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

where $\Gamma_c$ 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

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

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

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

$$(\beta - \beta_0)'\Sigma_c^3(\beta - \beta_0) = o\{n_c^{-1}tr(\Sigma_c^2)\}.$$
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 $\sigma_{fac,0}^2 = \frac{2\sigma^4}{IJ} \sum_{c=1}^{IJ} tr(\Sigma_c^2)/\{n_c(n_c - 1)\}$.

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

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

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

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

and the other is

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

Let

$$\sigma_{A_1,fac}^2 = (IJ)^{-2} \sum_{c=1}^{IJ} \left\{ 2(\sigma^4 + B_{1c}^2)\hat{tr}(\Sigma_c^2) + 4\sigma^2 \hat{tr}(\Sigma_c^2)B_{1c} \right. \right.$$

$$\left. + 4\Delta(\sigma^2 + B_{1c})\hat{tr}(A_{1c} \circ A_{3c}) + 2\Delta^2 \hat{tr}\left\{ (A_{0c} diag(A_{1c}))^2 \right\} \right\} / n_c^2$$

and

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

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 $\min C n_c \to \infty$, (i) under the fixed alternative (2.5.34),

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

(ii) under the fixed alternatives (2.5.35),

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

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 $\varepsilon_i$. 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

$$H_0 : \beta = \mathbf{0}_{p \times 1} \quad \text{vs} \quad H_1 : \beta \neq \mathbf{0}_{p \times 1}.$$
for some $T < p$. Here $Z_i = (Z_{i1}, \ldots, Z_{i(p+T-1)})'$ 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 $\{\rho_t\}_{t=1}^T$ 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.

![Auto-correlation for T=10](image1)

![Auto-correlation for T=20](image2)

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 $\beta$-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 $p >> n$. 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 $p >> n$, 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
\[
\frac{n\|\Sigma(\beta-\beta_0)\|^2}{\sqrt{2tr(\Sigma^2)\sigma^2}},
\]
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 \( n \to \infty \).

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) | $||\beta||^2$ | $T = 10$ | $T = 20$ |
|-------|---------------|----------|----------|
|       | SNR | F-test | EB | New | LP | SNR | F-test | EB | New |
|       |     |       |     |     |    |     |       |     |     |
| (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 |
| (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 $H_0 : \beta = 0_p \times 1$ vs $H_1 : \beta \neq 0_p \times 1$ at significant level 5% for centralized gamma residual. LP represents the theoretical local power.

<table>
<thead>
<tr>
<th>(n, p)</th>
<th>$|\beta|^2$</th>
<th>(T = 10)</th>
<th>(T = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SNR</td>
<td>F-test</td>
<td>EB</td>
</tr>
<tr>
<td>(40, 34)</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.96</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>1.92</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>2.89</td>
<td>0.47</td>
</tr>
<tr>
<td>(60, 54)</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>1.48</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>2.95</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>4.44</td>
<td>0.63</td>
</tr>
<tr>
<td>(80, 76)</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>1.25</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>2.51</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>3.76</td>
<td>0.51</td>
</tr>
</tbody>
</table>

(b) Sparse case

<table>
<thead>
<tr>
<th>(n, p)</th>
<th>$|\beta|^2$</th>
<th>(T = 10)</th>
<th>(T = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SNR</td>
<td>F-test</td>
<td>EB</td>
</tr>
<tr>
<td>(40, 34)</td>
<td>0.02</td>
<td>0.59</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>1.19</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>1.78</td>
<td>0.15</td>
</tr>
<tr>
<td>(60, 54)</td>
<td>0.02</td>
<td>0.81</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>1.63</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>2.44</td>
<td>0.15</td>
</tr>
<tr>
<td>(80, 76)</td>
<td>0.02</td>
<td>0.62</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>1.25</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>1.87</td>
<td>0.12</td>
</tr>
</tbody>
</table>
Table 2.3 Empirical size and power of the EB test and the proposed test (new) for $H_0: \beta = 0_{p \times 1}$ vs $H_1: \beta \neq 0_{p \times 1}$ at significant level 5% for normal residual. LP represents the theoretical local power.

<table>
<thead>
<tr>
<th>$(n, p)$</th>
<th>$|\beta|^2$</th>
<th>(T = 10)</th>
<th>(T = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\text{SNR})</td>
<td>(\text{EB})</td>
<td>(\text{New})</td>
</tr>
<tr>
<td>(40, 310)</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.30</td>
<td>0.01</td>
</tr>
<tr>
<td>(60, 400)</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>1.47</td>
<td>0.08</td>
</tr>
<tr>
<td>(80, 550)</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>1.66</td>
<td>0.13</td>
</tr>
</tbody>
</table>

(a) Non-sparse case

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

<table>
<thead>
<tr>
<th>(n, p)</th>
<th>$|\beta|^2$</th>
<th>SNR</th>
<th>$T = 10$</th>
<th>SNR</th>
<th>$T = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>EB</td>
<td>New</td>
<td>LP</td>
<td></td>
</tr>
<tr>
<td>(40, 310)</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.01</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.30</td>
<td>0.01</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.61</td>
<td>0.03</td>
<td>0.19</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.92</td>
<td>0.05</td>
<td>0.24</td>
<td>0.23</td>
</tr>
<tr>
<td>(60, 400)</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.01</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.49</td>
<td>0.02</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.98</td>
<td>0.05</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>1.47</td>
<td>0.10</td>
<td>0.36</td>
<td>-</td>
</tr>
<tr>
<td>(80, 550)</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.55</td>
<td>0.03</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>1.11</td>
<td>0.07</td>
<td>0.23</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>1.66</td>
<td>0.16</td>
<td>0.40</td>
<td>-</td>
</tr>
<tr>
<td>(a) Non-sparse case</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(40, 310)</td>
<td>0.02</td>
<td>0.16</td>
<td>0.01</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.32</td>
<td>0.01</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.48</td>
<td>0.02</td>
<td>0.14</td>
<td>0.12</td>
</tr>
<tr>
<td>(60, 400)</td>
<td>0.02</td>
<td>0.27</td>
<td>0.02</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.54</td>
<td>0.02</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.82</td>
<td>0.04</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>(80, 550)</td>
<td>0.02</td>
<td>0.35</td>
<td>0.01</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.70</td>
<td>0.03</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>1.05</td>
<td>0.06</td>
<td>0.25</td>
<td>0.28</td>
</tr>
</tbody>
</table>

| (a) Sparse case |

| (40, 310) | 0.02           | 0.16 | 0.01 | 0.08 | 0.07 | 0.58 | 0.05 | 0.16 | 0.14 |
| (60, 400) | 0.02           | 0.27 | 0.02 | 0.09 | 0.08 | 0.60 | 0.04 | 0.15 | 0.15 |
| (80, 550) | 0.02           | 0.35 | 0.01 | 0.10 | 0.10 | 1.05 | 0.10 | 0.24 | 0.28 |
| (a) Sparse case |
test under the factorial designs. We simulated a two-factor balanced design with two levels for each factor:

\[ Y_{ijk} = \alpha_{ij} + X_{ijk}'/\beta + \varepsilon_{ijk}, \quad k = 1, 2, \ldots, n_c \]  

(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 \( \beta \) 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 \( \Sigma \). 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
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 \(H_0: \beta = 0_{p \times 1}\) in a \(2 \times 2\) factorial design with \(n_1 = 20\) and \(n_2 = 30\) replicates in each cell.

<table>
<thead>
<tr>
<th>(p)</th>
<th>(|\beta|^2)</th>
<th>(\text{SNR}_f)</th>
<th>(n_1)</th>
<th>(\text{SNR}_f)</th>
<th>(n_2)</th>
<th>(\text{SNR}_f)</th>
<th>(n_1)</th>
<th>(\text{SNR}_f)</th>
<th>(n_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Normal residuals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>3.05</td>
<td>0.65</td>
<td>4.58</td>
<td>0.85</td>
<td>0.70</td>
<td>0.20</td>
<td>1.06</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>6.10</td>
<td>0.88</td>
<td>9.16</td>
<td>0.98</td>
<td>1.41</td>
<td>0.29</td>
<td>2.12</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>9.16</td>
<td>0.96</td>
<td>13.74</td>
<td>1.00</td>
<td>2.12</td>
<td>0.44</td>
<td>3.18</td>
<td>0.65</td>
</tr>
<tr>
<td>150</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>2.59</td>
<td>0.57</td>
<td>3.89</td>
<td>0.77</td>
<td>0.57</td>
<td>0.15</td>
<td>0.85</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>5.18</td>
<td>0.84</td>
<td>7.78</td>
<td>0.97</td>
<td>1.14</td>
<td>0.28</td>
<td>1.71</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>7.78</td>
<td>0.94</td>
<td>11.67</td>
<td>0.99</td>
<td>1.71</td>
<td>0.35</td>
<td>2.57</td>
<td>0.54</td>
</tr>
<tr>
<td>200</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>2.28</td>
<td>0.50</td>
<td>3.43</td>
<td>0.73</td>
<td>0.49</td>
<td>0.14</td>
<td>0.73</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>4.57</td>
<td>0.78</td>
<td>6.86</td>
<td>0.94</td>
<td>0.98</td>
<td>0.22</td>
<td>1.47</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>6.86</td>
<td>0.89</td>
<td>10.29</td>
<td>0.99</td>
<td>1.47</td>
<td>0.31</td>
<td>2.21</td>
<td>0.48</td>
</tr>
<tr>
<td>(b) Gamma residuals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>3.05</td>
<td>0.66</td>
<td>4.58</td>
<td>0.83</td>
<td>0.70</td>
<td>0.15</td>
<td>1.06</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>6.10</td>
<td>0.86</td>
<td>9.16</td>
<td>0.97</td>
<td>1.41</td>
<td>0.31</td>
<td>2.12</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>9.16</td>
<td>0.95</td>
<td>13.74</td>
<td>0.99</td>
<td>2.12</td>
<td>0.47</td>
<td>3.18</td>
<td>0.66</td>
</tr>
<tr>
<td>150</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>2.59</td>
<td>0.57</td>
<td>3.89</td>
<td>0.78</td>
<td>0.57</td>
<td>0.16</td>
<td>0.85</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>5.18</td>
<td>0.81</td>
<td>7.78</td>
<td>0.96</td>
<td>1.14</td>
<td>0.28</td>
<td>1.71</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>7.78</td>
<td>0.93</td>
<td>11.67</td>
<td>0.99</td>
<td>1.71</td>
<td>0.37</td>
<td>2.57</td>
<td>0.57</td>
</tr>
<tr>
<td>200</td>
<td>0.00 (size)</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>2.28</td>
<td>0.53</td>
<td>3.43</td>
<td>0.74</td>
<td>0.49</td>
<td>0.14</td>
<td>0.73</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>4.57</td>
<td>0.77</td>
<td>6.86</td>
<td>0.93</td>
<td>0.98</td>
<td>0.24</td>
<td>1.47</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>6.86</td>
<td>0.89</td>
<td>10.29</td>
<td>0.98</td>
<td>1.47</td>
<td>0.30</td>
<td>2.21</td>
<td>0.48</td>
</tr>
</tbody>
</table>
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 Yolkshire 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 $X_{gijk}^g$ 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:

Design I: \[ Y_k = \alpha + X_k^g \beta^g + \varepsilon_k^g, \quad k = 1, \cdots, 24; \]
Design II: \[ Y_{ik} = \alpha + \mu_i + X_{ik}^g \beta^g + \varepsilon_{ik}^g, \quad k = 1, \cdots, 12; \]
Design III: \[ Y_{jk} = \alpha + \tau_j + X_{jk}^g \beta^g + \varepsilon_{jk}^g, \quad k = 1, \cdots, 12; \]
Design IV: \[ Y_{ijk} = \alpha + \mu_i + \tau_j + \mu \tau_{ij} + X_{ijk}^g \beta^g + \varepsilon_{ijk}^g, \quad k = 1, \cdots, 6 \]

for $i = 1, 2$, $j = 1, 2$ and $g = 1, \cdots, G$ where $G = 6176$ is the total number of the GO terms, $\mu_i$ stand for diet treatment effects, $\tau_j$ 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.

![Histograms of p-values](image.png)

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
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.
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 $B$, $tr^2(A \circ B) \leq tr(A \circ A)tr(B \circ B) \leq tr(A^2)tr(B^2)$. 

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

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

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

**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 $Q = P diag(\lambda_1, \ldots, \lambda_p) P'$. 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 $B_2$,

$$E\{(Z_1'B_1Z_1)(Z_1'B_2Z_1)\} = tr(B_1)tr(B_2) + 2tr(B_1B_2) + \Delta tr(B_1 \circ B_2);$$

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

$$E\{Z_1'AZ_2^4\} = 3tr^2(\Sigma^2) + 6tr(\Sigma^4) + 6\Delta tr(A_0^2 \circ A_0^2) + \Delta^2 \sum_{i,k=1}^{m} (A_{0,ik})^4.$$

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

$$E\{(Z_1'B_1Z_1 - tr(B_1))(Z_1'B_2Z_1 - tr(B_2))\}^2 \leq C tr(B_1^2) tr(B_2^2)$$
(iv) For any $m \times m$ symmetric matrices $A, B$ and $D$ and suppose that $E(Z_{11}^2) = \tau_0$. Then

$$E\{Z_1'AZ_1Z_1'BZ_1Z_1'DZ_1\} = 8\text{tr}(ABC) + 2\text{tr}(AB)\text{tr}(D) + 2\text{tr}(BD)\text{tr}(A) + \text{tr}(A)\text{tr}(B)\text{tr}(D) + \Delta\{\text{tr}(A)\text{tr}(B \circ D) + \text{tr}(B)\text{tr}(A \circ D) + \text{tr}(D)\text{tr}(A \circ B) + \text{tr}(AB \circ D) + \text{tr}(AD \circ B) + \text{tr}(BD \circ A)\} + (\tau_0 - 15 - 6\Delta)\text{tr}(A \circ B \circ D).$$

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),

$$E\{(Z_1'B_1Z_1)(Z_1'B_2Z_1)\} = \sum_{i_1, \ldots, i_4} B_{i_1i_2}^{(1)} B_{i_3i_4}^{(2)} E(z_{i_1}z_{i_2}z_{i_3}z_{i_4})$$

$$= \sum_{i=1}^m b_{ii}^{(1)} B_{ii}^{(2)} E(z_i^4) + \sum_{i_1 \neq i_2} B_{i_1i_1}^{(1)} B_{i_2i_2}^{(2)} E(z_{i_1}^2)E(z_{i_2}^2) + 2 \sum_{i_1 \neq i_2} B_{i_1i_2}^{(1)} B_{i_1i_2}^{(2)} E(z_{i_1}^2)E(z_{i_2}^2)$$

$$= \text{tr}(B_1)\text{tr}(B_2) + 2\text{tr}(B_1B_2) + \Delta\text{tr}(B_1 \circ B_2).$$

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

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

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

$$E\left\{ \left( \sum_i B_{ii}^{(1)} (z_i^2 - 1) + \sum_{j \neq l} B_{jl}^{(1)} z_i z_l \right) \left( \sum_k B_{kk}^{(2)} (z_k^2 - 1) + \sum_{s \neq t} B_{st}^{(2)} z_s z_t \right) \right\}^2$$

$$\leq J_1 + J_2 + J_3 + J_4$$

where

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

\[ J_2 = 4 \sum_{i,i_1,j \neq t, t \neq t_1} B_{ii}^{(1)} B_{st}^{(2)} B_{i_1i_1}^{(1)} B_{s_1t_1}^{(2)} (z_i^2 - 1)(z_{i_1}^2 - 1)z_s z_t z_{i_1} z_{t_1}, \]

\[ J_3 = 4 \sum_{k,k_1,j \neq t, j \neq t_1} B_{kk}^{(2)} B_{ji}^{(1)} B_{k_1j_1}^{(2)} B_{i_1i_1}^{(1)} (z_k^2 - 1)(z_{k_1}^2 - 1)z_j z_{j_1} z_{i_1} \]

\[ J_4 = 4 \sum_{s \neq t, s \neq t_1, j \neq t_1} B_{jl}^{(1)} B_{st}^{(2)} B_{j_1i_1}^{(1)} B_{s_1t_1}^{(2)} z_j z_{j_1} z_t z_{i_1} z_{s_1} z_{t_1}. \]

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,

\[
|J_1| \leq C \left( 2 \sum_i (B_{ii}^{(1)} B_{ii}^{(2)})^2 + \sum_{i \neq k} (B_{ii}^{(1)} B_{kk}^{(2)})^2 + \sum_{i \neq k} |B_{ii}^{(1)}||B_{ii}^{(2)}||B_{kk}^{(1)}||B_{kk}^{(2)}| \right)
\]

\[
= C \left( \sum_{i,k} (B_{ii}^{(1)} B_{kk}^{(2)})^2 + \sum_{i,k} |B_{ii}^{(1)}||B_{ii}^{(2)}||B_{kk}^{(1)}||B_{kk}^{(2)}| \right)
\]

\[ = C \left( tr(B_1 \circ B_1) tr(B_2 \circ B_2) + tr(|B_1| \circ |B_2|) tr(|B_1| \circ |B_2|) \right) \leq C tr(B_1^2) tr(B_2^2). \]

Applying Cauchy-Schwarz inequality,

\[
|J_2| \leq C \left( \sum_{i,s \neq t} (B_{ii}^{(1)} B_{st}^{(2)})^2 + \sum_{i \neq i_1} |B_{ii}^{(1)}||B_{i_1i_1}^{(1)}|(B_{ii}^{(2)})^2 \right)
\]

\[ \leq C \left( \sum_i (B_{ii}^{(1)})^2 \sum_{s,t} (B_{st}^{(2)})^2 + \sum_i |B_{ii}^{(1)}| \left( \sum_{i_1} (B_{i_1i_1}^{(1)})^2 \right)^{1/2} \left( \sum_i (B_{ii}^{(2)})^4 \right)^{1/2} \right)
\]

\[ \leq C (tr(B_1 \circ B_1) tr(B_2^2) + \left( \sum_i (B_{ii}^{(1)})^2 \right) \left( \sum_i \sum_{i_1} (B_{i_1i_1}^{(2)})^4 \right)^{1/2} \) \leq C tr(B_1^2) tr(B_2^2). \]

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

\[
|J_4| \leq C \left( \sum_{j,l \neq t, l \neq t_1} (B_{jl}^{(1)} B_{st}^{(2)})^2 + \sum_{j,l \neq t, l \neq t_1} |B_{jl}^{(1)}||B_{jl}^{(2)}||B_{st}^{(1)}||B_{st}^{(2)}| \right)
\]

\[ \leq C \left( \sum_{j,l} (B_{jl}^{(1)})^2 \sum_{s,t} (B_{st}^{(2)})^2 + \sum_{j,l} |B_{jl}^{(1)}||B_{jl}^{(2)}| \sum_{s,t} |B_{st}^{(1)}||B_{st}^{(2)}| \right)
\]

\[ + 2 \sum_{j,l} (B_{jl}^{(1)})^2 (B_{st}^{(2)})^2 \]

\[ \leq C tr(B_1^2) tr(B_2^2). \]
Therefore, (iii) is true. The conclusion (iv) could be established based on the fact

\[
E\{Z_1'AZ_1'Z_1'BZ_1'Z_1'DZ_1\} = \tau_0 \sum_i A_{ii}B_{ii}D_{ii} + (3 + \Delta) \left( \sum_{i \neq k} A_{ii}B_{kk}D_{kk} + 4 \sum_{i \neq j} A_{ij}B_{ij}D_{jj} \right) \\
+ 4 \sum_{i \neq j} A_{ij}B_{ij}D_{ij} + \sum_{i \neq k} A_{ii}B_{kk}D_{ii} + \sum_{i \neq k} A_{ii}B_{ii}D_{kk} + \sum_{i \neq k} A_{ii}B_{ki}D_{ki} \\
+ \sum_{i \neq k \neq s} A_{ii}B_{kk}D_{ss} + 2 \sum_{i \neq j \neq s} A_{ij}B_{ij}D_{ss} + 2 \sum_{i \neq j \neq k} A_{ij}B_{kk}D_{ij} \\
+ 2 \sum_{i \neq k \neq l} A_{ii}B_{kl}D_{kl} + 8 \sum_{i \neq j \neq l} A_{ij}B_{il}D_{lj}.
\]

\[\square\]

**Proof of Theorem 1** Let \( \gamma_0 = (\alpha, \beta_0')' \). 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 \( P_1 = \mathbf{1}'/n \) be the projection matrices of \( U(U'U)^{-1}A' \), \( U \) and \( \mathbf{1} \) respectively. By applying the matrix inverse formula on \( (U'U)^{-1} \), \( U(U'U)^{-1}A' = (I - P_1)X\{X'(I - P_1)X\}^{-1} \). It then follows that \( P_{Au} = (I - P_1)X(X'(I - P_1)X)^{-1}X'(I - P_1) \).

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

\[
\frac{p}{n - p - 1}G_{n,p} \overset{d}{=} \left\{ Q\varepsilon + Q(U(\gamma - \gamma_0)) \right\}' \text{diag}(1_p', 0_{n-p}') \left\{ Q\varepsilon + Q(U(\gamma - \gamma_0)) \right\}, \]

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

\[
\frac{p}{n - p - 1}G_{n,p} \overset{d}{=} \sum_{i=1}^p\left\{ \varepsilon_i^2 + 2\varepsilon_iQ_i'X\delta_\beta \right\}/\mathbf{z}_1'\mathbf{z}_1 + \delta'X'P_{Au}X\delta_\beta/\mathbf{z}_1'\mathbf{z}_1 \quad (2.8.41)
\]

where \( X'P_{Au}X = X'(I - P_1)X = \Gamma Z'(I - P_1)Z \Gamma' \) and \( Z = (Z_1, \cdots, Z_n)' \).
For the numerator of the (2.8.41), we can show that under Model (2.2.4), \( E\{\delta'X'P_{Au}X\delta\} = (n - 1)\delta'\Sigma\delta \). It is easy to see that \( E\{\sum_{i=1}^{p} \varepsilon_iQ_i'X\delta\} = 0 \) and

\[
\text{Var}\{\sum_{i=1}^{p} \varepsilon_iQ_i'X\delta\} = (n - 1)\sigma^2\delta'\Sigma\delta. \tag{2.8.42}
\]

It can be shown that

\[
\text{Var}\{\delta'X'P_{Au}X\delta\} = 2(n - 1)(\delta'\Sigma\delta)^2 + (n + 2 + 1/n)\Delta tr(A_1 \circ A_1). \tag{2.8.43}
\]

Direct calculation shows that \( E(\frac{1}{z_1^2}) = 1/(n - p - 3) \) and \( E(\frac{1}{z_1^2}) = 1/\{(n - p - 3)(n - p - 5)\} \). (2.8.42) implies that \( \sum_{i=1}^{p} \varepsilon_iQ_i'X\delta\beta/z_1'z_1 = O_p(\frac{1}{\sqrt{n}}\delta'\Sigma\delta) \) and note that \( E(X'P_{Au}X) = (n - 1)\Sigma \). Then (2.8.43) yields

\[
\frac{\delta'X'P_{Au}X\delta}{z_1'z_1} = \frac{\delta'\Sigma\delta}{1 - \rho} + O_p(\frac{1}{\sqrt{n}}\delta'\Sigma\delta).
\]

If \( \delta'\Sigma\delta = o(1) \), then

\[
\frac{p}{n - p - 1}G_{n,p} \overset{d}{=} \sum_{i=1}^{p} \frac{\varepsilon_i^2}{z_1'z_1} + \frac{\delta'\Sigma\delta}{1 - \rho} + o_p(n^{-1/2}).
\]

From Bai and Saranadasa (1996),

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

where \( z_{\alpha} \) is the \( \alpha \) quantile of \( N(0,1) \) and it can be shown

\[
\sqrt{\frac{(1 - \rho)^3n}{2\rho}} \left( \sum_{i=1}^{p} \frac{\varepsilon_i^2}{z_1'z_1} - \frac{\rho_n}{1 - \rho_n} \right) \overset{d}{=} N(0,1).
\]

Therefore the power of the F-test is

\[
\Omega_F(\|\beta - \beta_0\|) = P\left( \frac{p}{n - p - 1}G_{n,p} > \frac{p}{n - p - 1}F_{p, n-p-1; \alpha} \right)
= P\left\{ \sqrt{\frac{(1 - \rho)^3n}{2\rho}} \left( \sum_{i=1}^{p} \frac{\varepsilon_i^2}{z_1'z_1} - \frac{\rho_n}{1 - \rho_n} \right) > z_{\alpha} - \sqrt{\frac{(1 - \rho)^3n}{2\rho}}\delta'\Sigma\delta + o_p(1) \right\}
= \Phi(-z_{\alpha} - \sqrt{\frac{(1 - \rho)n}{2\rho}}\delta'\Sigma\delta) + o(1).
\]

\( \square \)
Proof of Theorem 2: It is straightforward to show that $E(T_{n,p}) = ||\Sigma \delta||^2$. To derive 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 $A_0 = \Gamma' \Gamma$, $A_1 = \Gamma' \delta \delta' \Gamma$, $A_2 = \Gamma' \Sigma \delta \delta' \Sigma \Gamma$, $A_3 = \Gamma' \Sigma \Gamma$ and $B_i = \delta \delta' \Sigma \delta$. It can be shown that

$$\zeta_1 = \frac{1}{4} B_1 B_3 + \frac{1}{4} \sigma^2 B_3 + \frac{1}{4} B_2^2 + \frac{1}{4} \Delta tr(A_1 \circ A_2). \quad (2.8.44)$$

We can also show that

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

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

$$E[\delta_\beta(X_1 - X_2)(X_1 - X_2)' \Sigma \delta \beta] = 2 \Delta tr(A_1 \circ A_2) + 8 B_2^2 + 4 B_1 B_3,$$

$$E[\varepsilon_1 X_1(X_2 X_2' + \Sigma) \delta \beta] = 5 \sigma^2 B_3 + \sigma^2 tr(\Sigma^2) B_1 + \Delta \sigma^2 tr(A_1 \circ A_3),$$

$$E[\delta_\beta(X_1 X_1' + \Sigma)(X_2 X_2' + \Sigma) \delta \beta] = tr[5 \Sigma \delta \delta' \Sigma + \delta \delta' \Sigma \delta \Sigma + \Delta \Gamma diag(A_1) \Gamma']^2$$

$$= 25 B_2^2 + B_1^2 tr(\Sigma^2) + \Delta^2 tr\{ A_0 diag(A_1) \}^2 + 10 B_1 B_3 + 10 \Delta tr(A_1 \circ A_2)$$

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

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

$$= E[2 \delta_\beta(X_1 X_1' + X_2 X_2') \Sigma \delta \beta \delta_\beta(X_1 X_1' + \Sigma)(X_2 X_2' + \Sigma) \delta \beta]$$

$$= 8 E[\delta_\beta X_1 X_1' \Sigma \delta \beta \delta_\beta X_1 X_1' \Sigma \delta \beta] + 8 [\delta \beta \Sigma \delta \beta]^2 = 8 \Delta tr(A_1 \circ A_2) + 8 B_1 B_3 + 24 B_2^2.$$

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

$$E(h^2(W_1, W_2, W_3, W_4)) = \frac{1}{3} E\{ \phi(1, 2, 3, 4) + \phi(1, 3, 2, 4) + \phi(1, 4, 2, 3) \}^2$$

$$= \frac{1}{3} E \phi^2(1, 2, 3, 4) + \frac{2}{3} E \phi(1, 2, 3, 4) \phi(1, 3, 2, 4)$$

$$+ \frac{2}{3} E \phi(1, 2, 3, 4) \phi(1, 4, 2, 3) + \frac{2}{3} E \phi(1, 3, 2, 4) \phi(1, 4, 2, 3)$$

$$= \frac{1}{3} E \phi^2(1, 2, 3, 4) + \frac{2}{3} E \phi(1, 2, 3, 4) \phi(1, 3, 2, 4).$$
where
\[
\frac{1}{3}E\phi^2(1, 2, 3, 4) = \frac{1}{3}\sigma^4 tr(\Sigma^2) + \frac{4}{3}B_2^2 + \frac{4}{3}B_1B_3 + \frac{4}{3}\sigma^2 B_3 + \frac{1}{3}B_1^2 tr(\Sigma^2) + \frac{2}{3}\sigma^2 tr(\Sigma^2) B_1 + \frac{1}{3}(B_1 + \sigma^2) \Delta tr(A_1 \circ A_3) + \frac{2}{3} \Delta tr(A_1 \circ A_2) + \frac{1}{12}\Delta^2 tr\{A_0 diag(A_1)\}^2 \quad \text{and}
\]
\[
\frac{2}{3}E\phi(1, 2, 3, 4)\phi(1, 3, 2, 4) = \frac{1}{24}\left\{8\sigma^4 tr(\Sigma^2) + 32B_2^2 + 32B_1B_3 + 32\sigma^2 B_3 + 16\sigma^2 B_1 tr(\Sigma^2) + 8B_1^2 tr(\Sigma^2) + 16\Delta tr(A_1 \circ A_2) + 8\Delta^2 tr(A_1 \circ A_3) + 8\Delta B_1 tr(A_1 \circ A_3) + 2\Delta^2 tr\{A_0 diag(A_1)\}^2 \right\}.
\]

It then follows that
\[
\zeta_4 = \frac{1}{24}\left\{12\sigma^4 tr(\Sigma^2) + 45B_2^2 + 65B_1B_3 + 40\sigma^2 B_3 + 10B_1^2 tr(\Sigma^2) + 24\sigma^2 tr(\Sigma^2) B_1 + 12(B_1 + \sigma^2) \Delta tr(A_1 \circ A_3) + 37\Delta tr(A_1 \circ A_2) + 4\Delta^2 tr\{A_0 diag(A_1)\}^2 \right\}, \tag{2.8.46}
\]

Note that (2.8.45) and (2.8.46) show that \( \zeta_2 \) and \( \zeta_4 \) are both the linear combination of \( tr(\Sigma^2), B_2^2, B_1B_3, B_3, B_1^2 tr(\Sigma^2), B_1 tr(\Sigma^2), (B_1 + \sigma^2) tr(A_1 \circ A_3), tr(A_1 \circ A_2) \) and \( tr\{A_0 diag(A_1)\}^2 \). So it implies that \( \zeta_2 \) and \( \zeta_4 \) are of the same order. By Proposition 1, \( \zeta_2, \zeta_3 \) and \( \zeta_4 \) 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 \( \zeta_1 \) and \( \zeta_2 \), the results in Theorem 2 follow.

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
\[
\zeta_1 \leq \left\{\frac{1}{2} + \frac{1}{4}\Delta\right\}B_1 + \frac{1}{4}\sigma^2\right\} B_3, \tag{2.8.47}
\]
\[
\zeta_2 \leq \frac{1}{36}\left\{[\sigma^2 + (\Delta + 1)B_1]^2 tr(\Sigma^2) + [22\sigma^2 + (43 + 20\Delta)B_1]B_3\right\}. \tag{2.8.48}
\]

**Proof of Theorem 3**

Let
\[
\widehat{T}_{n,p} - \|\Sigma \delta\|^2 = \frac{12}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \tilde{h}_2(W_{i_1}, W_{i_2}) \tag{2.8.49}
\]
be the projection of \( T_{n,p} \). We can decompose \( T_{n,p} - \|\Sigma \delta\|^2 = \widehat{T}_{n,p} - \|\Sigma \delta\|^2 + (T_{n,p} - \widehat{T}_{n,p}) \)
where \( T_{n,p} - \widehat{T}_{n,p} \) can still be written as a U-statistics with kernel
\[
H(W_1, W_2, W_3, W_4) = \tilde{h}(W_1, W_2, W_3, W_4) - \sum_{1 \leq i_1 < i_2 \leq 4} \tilde{h}_2(W_{i_1}, W_{i_2}). \tag{2.8.50}
\]
The projections of $H$ are $H_1(w_1) = -2\hat{h}_1(w_1), H_2(w_1, w_2) = -2\sum_{i=1}^{2} \hat{h}_1(w_i)$ and $H_3(w_1, w_2, w_3) = \hat{h}_3(w_1, w_2, w_3) - \sum_{i=1}^{3} \hat{h}_1(w_i) - \sum_{1 \leq i < j \leq 3} \hat{h}_2(w_i, w_j)$. 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 $\zeta_2, \zeta_3$ and $\zeta_4$ are of the same order as we have shown in Theorem 2. Thus,

$$\frac{T_{n,p} - \|\Sigma\delta\|_2^2}{\sqrt{\text{Var}(\hat{T}_{n,p})}} \sim \frac{\hat{T}_{n,p} - \|\Sigma\delta\|_2^2}{\sqrt{\text{Var}(\hat{T}_{n,p})}} + o_p(1).$$

Hence we only need to show that

$$\frac{\hat{T}_{n,p} - \|\Sigma\delta\|_2^2}{\sqrt{\text{Var}(\hat{T}_{n,p})}} \xrightarrow{d} N(0, 1).$$

(2.8.51)

From (2.8.49), $\hat{T}_{n,p} - \|\Sigma\delta\|_2^2 = \hat{T}_{n,p}^{(1)} + \hat{T}_{n,p}^{(2)}$ where

$$\hat{T}_{n,p}^{(1)} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \left\{ [\delta_\beta(X_i - X_j) + (\varepsilon_i - \varepsilon_j)](X_i - X_j)'\Sigma\delta\beta + [\delta_\beta'(X_iX_i' + \Sigma) + \varepsilon_iX_i'X_j'] + \varepsilon_jX_jX_i' + \Sigma)\delta\beta + \varepsilon_jX_jX_i' + \Sigma)\delta\beta \right\} - 6\|\Sigma\delta\|_2^2$$

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

$$\frac{\hat{T}_{n,p}^{(2)}/\sqrt{\text{Var}(\hat{T}_{n,p}^{(2)})}}{\sqrt{\frac{n}{2}}} \xrightarrow{d} N(0, 1).$$

(2.8.52)

Now write $\tilde{T}_{nk} = \binom{n}{2}^{1/2} \hat{T}_{n,p}^{(2)} = \sum_{i=2}^{k} Z_{ni}$ and $\tilde{T}_{nn} = \tilde{T}_{n,p}$, where $Z_{ni} = \sum_{j=1}^{i-1} \varepsilon_i\varepsilon_jX_i'X_j/\binom{n}{2}$. Let $\mathcal{F}_i = \sigma\left\{ (X_{\varepsilon_i}, \ldots, (X_i)) \right\}$ be the $\sigma$–field generated by $\{(X_j, \varepsilon_j), j \leq i\}$. 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

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

(2.8.53)

and for any $\epsilon > 0$

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

(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

\[
\frac{V_n}{\text{Var}(T_{n,p})} = \frac{1}{\binom{n}{2}^2 tr(\Sigma^2) \sigma^4} \left\{ \sum_{j=1}^{n-1} j \varepsilon_j^2 X_j' \Sigma X_j + 2 \sum_{1 \leq j < k \leq n} \varepsilon_j \varepsilon_k X_j' \Sigma X_k \right\}
\]

\[
= C_{n1} + C_{n2}, \text{ say.}
\]

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

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

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

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

\[
= \frac{4}{\binom{n}{2}^4} \sum_{i=3}^{n-1} \left\{ \frac{n}{2} \right\} (n-i) \left( \frac{i}{2} \right)^2 tr(\Sigma^4).
\]

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

It remains to show (2.8.54). Since

\[
E\left\{ Z_{ni}^2 I(|Z_{ni}| > \epsilon \sigma^2 \sqrt{tr(\Sigma^2)}) | F_{i-1} \right\} \leq E(Z_{ni}^4 | F_{i-1})/(\epsilon^2 \sigma^4 tr(\Sigma^2)),
\]

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

\[
\sum_{i=1}^{n} E(Z_{ni}^4) = o\{\sigma^4 tr^2(\Sigma^2)\}.
\]

(2.8.55)

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

\[
\sum_{i=1}^{n} E(Z_{ni}^4) \leq \binom{n}{2}^{-1} \kappa_4^2 \left( 3tr^2(\Sigma^2) + (6 + 6 \Delta + \Delta^2) tr(\Sigma^4) \right)
\]

\[
+ \binom{n}{2}^{-2} \frac{1}{3} (n^3 - 3n^2 + 2n) \kappa_4 \sigma^4 \left( tr^2(\Sigma^2) + (2 + \Delta) tr(\Sigma^4) \right).
\]

Under the assumption that \( tr(\Sigma^4) = o\{tr^2(\Sigma^2)\} \), (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) \( E(\eta_1^t A\eta_1) = tr(\Sigma B_1 + \sigma^2) + \delta_\beta^t \Sigma A\Sigma\delta_\beta + \Delta tr(\Gamma' A\Gamma \circ \Gamma' \delta_\beta \delta_\beta' \Gamma) \).

(ii) \( E(\eta_1^t A\eta_2)^2 \leq Ctr(\Sigma B_1 + \sigma^2)^2 + C\{B_1 + \sigma^2\} \delta_\beta^t \Sigma A\Sigma A\Sigma\delta_\beta + \{\delta_\beta^t \Sigma A\Sigma\delta_\beta\}^2 \).

(iii) \( E(\eta_1^t A\eta_1)^2 \leq Ctr(\Sigma B_1 + \sigma^4) + Ctr(\Sigma B_1^2 + \sigma^4) + C\{\delta_\beta^t \Sigma A\Sigma\delta_\beta\}^2 \)
\[ + C\{B_1 + \sigma^2\} \delta_\beta^t \Sigma A\Sigma A\Sigma\delta_\beta + Ctr(\Sigma A\Sigma) tr(\Gamma' A\Gamma \circ \Gamma' \delta_\beta \delta_\beta' \Gamma). \]

**Proof** The first result is immediately obtained from (i) in Lemma 2. We only show (ii) and (iii). For (ii), notice that
\[ E(\eta_1^t A\eta_2) = (B_1 + \sigma^2) \Sigma + \Sigma \delta_\beta \delta_\beta' \Sigma + \text{diag}(\Gamma' \delta_\beta \delta_\beta' \Gamma) \text{Gamma} \text{ and } E(\eta_1^t A\eta_2)^2 = E[\eta_2^t A\Sigma A\eta_2] = E[\eta_2^t A\Sigma A\eta_2 + \eta_2 A\Sigma \delta_\beta \delta_\beta' \Sigma A\eta_2 + \Delta tr(\Gamma' A\eta_2^t A\Gamma \circ \Gamma' \delta_\beta \delta_\beta' \Gamma)])].

Applying the formula in (i) again and note that \( tr(\Gamma' A\Gamma \circ \Gamma' \delta_\beta \delta_\beta' \Gamma) \leq tr(\Sigma A\Sigma) B_1^2 \), and \( tr(\Gamma' A\Sigma A\Gamma \circ \Gamma' \delta_\beta \delta_\beta' \Gamma) \leq tr(\Sigma A\Sigma) B_1 \) the result in (ii) follows immediately. Using Cauchy-Schwarz inequality, we have
\[ E(\eta_1^t A\eta_1)^2 \leq 24 E[(X_1^t A X_1 - tr(\Sigma A)](X_1^t \delta_\beta \delta_\beta' X_1 - B_1)] + 24 B_1^2 tr(\Sigma A) \]
\[ + 24 tr(\Sigma A)]E[(X_1^t \delta_\beta \delta_\beta' X_1 - B_1)^2] + 24 B_1^2 E[(X_1^t A X_1 - tr(\Sigma A))]^2 \]
\[ + 24 E[X_1^t A\Sigma \delta_\beta \delta_\beta' X_1]^2 + 6(\delta_\beta^t \Sigma A\Sigma \delta_\beta)^2 + 8\sigma^2 E[(X_1^t A X_1)^2 X_1^t \delta_\beta \delta_\beta' X_1] \]
\[ + 8\delta_\beta^t \Sigma A\Sigma A\Sigma \delta_\beta + 2\sigma^4 E(X_1^t A X_1)^2. \] (2.8.56)

The first term on the right hand side of (2.8.56) is bounded by \( Ctr(\Sigma A\Sigma) B_1^2 \), 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. \(\Box\)

**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
\[ \tilde{T}_{n,p} - \|\Sigma \delta_\beta\|^2 = \tilde{T}_{n,p}^{(1)} + \tilde{T}_{n,p}^{(2)} \]
where
\[ \tilde{T}_{n,p}^{(1)} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \left\{ (\delta_\beta(X_i - X_j) + (\varepsilon_1 - \varepsilon_2))(X_i - X_j)^t \Sigma \delta_\beta \right\} - 2\|\Sigma \delta_\beta\|^2 \]
and \[ \tilde{T}_{n,p}^{(2)} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \{\eta_i^t \eta_j + 2\delta_\beta \Sigma \eta_j + 2\delta_\beta^t \Sigma \eta_i\}. \]
Utilizing the facts listed in the proof of Theorem 2, we see that \( \hat{T}_{np}^{(1)} \) is a smaller order term of \( \hat{T}_{np}^{(2)} \). Therefore, we only need to show the asymptotic normality of \( \hat{T}_{np}^{(2)} \). Using Proposition 1, \[
\text{Var}(\hat{T}_{np}^{(2)}) = \frac{4(n-2)}{n(n-1)} \zeta_1 + \frac{2}{n(n-1)} \zeta_2, \]
where \( \zeta_1 = E(2\delta_{\beta}^2 \Sigma \eta_1) \) and \( \zeta_2 = E(\eta_1 \eta_j + 2\delta_{\beta}^2 \Sigma \eta_j + 2\delta_{\beta}' \Sigma \eta_i) \). Under condition \((2.4.18)\), it is can be shown that \( \text{Var}(\hat{T}_{np}^{(2)}) = \frac{2}{n(n-1)} \zeta_2 \{1 + o(1)\} \).

Let \( \hat{T}_{np} = \sum_{i=2}^{n} Z_{ni} \), where \( Z_{ni} = \binom{n}{2} \frac{1}{2} \sum_{j=1}^{i-1} \{\eta_i \eta_j + 2\delta_{\beta}^2 \Sigma \eta_j + 2\delta_{\beta}' \Sigma \eta_i\} \). Then \( Z_{ni}^2 = Z_{ni}^{(2)} + Z_{ni}^{(2)} \), where \( Z_{ni}^{(2)} = \binom{n}{2} \sum_{j=1}^{i-1} \{\eta_i \eta_j + 2\delta_{\beta}^2 \Sigma \eta_j + 2\delta_{\beta}' \Sigma \eta_i\} \). Let \( Z_{ni}^{(2)} = \frac{1}{2} \sum_{1 \leq j < k \leq i-1} \{\eta_i \eta_j + 2\delta_{\beta}^2 \Sigma \eta_j + 2\delta_{\beta}' \Sigma \eta_i\} \).

It follows that

\[
V_{ni}^{(1)} = E(Z_{ni}^{(1)} | \mathcal{F}_{i-1}) = \binom{n}{2} \sum_{j=1}^{i-1} \{E[B_1 + \sigma^2] \eta_i \eta_j + \Delta tr(\Gamma' \eta_i \eta_j \Gamma \circ \Gamma' \delta_{\beta}^2 \Gamma) + 4B_2^2 \eta_i \eta_j + 4\Delta tr(\Gamma' \eta_i \eta_j \Gamma \circ \Gamma' \delta_{\beta} \Sigma \eta_j) + 4\Delta tr(\Gamma' \eta_i \eta_j \Gamma \circ \Gamma' \delta_{\beta} \Sigma \eta_j)\}.
\]

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

\[
\text{Var}(V_{n1}) \leq \binom{n}{2}^{-2} \sum_{j=1}^{n-1} j^2 \{6B_1 + \sigma^2 \}^2 E[\eta_j \Sigma \eta_j - E(\eta_j \Sigma \eta_j)]^2 + 6\Delta^2 E(\eta_j \Sigma \eta_j)^2 + 30E[\eta_j \Sigma \delta_{\beta} \delta_{\beta}' \Sigma \eta_j - E[\eta_j \Sigma \delta_{\beta} \delta_{\beta}' \Sigma \eta_j]]^2 + 24B_2^2 E[\eta_j \Sigma \delta_{\beta} \delta_{\beta}' \Sigma \eta_j] + 24\{B_1 + \sigma^2 \}^2 + \Delta^2 B_1^2 E[\eta_j \Sigma \delta_{\beta} \delta_{\beta}' \Sigma \eta_j] \}.
\]

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

\[
B_1^2 B_5 \geq B_2^2 B_1^2 \geq B_1^4.
\]

Hence, \( \text{Var}(V_{n1}) \leq C(n)^2 \sum_{j=1}^{n-1} j^2 \{B_1^4 \{tr(\Sigma^4) + tr(\Sigma^2)\} + B_2^2 B_1^2 \} \). As \( tr(\Sigma^4) = o\{tr(\Sigma^2)\} \) implies that \( \lambda_1^4 = o\{\sum_{i=1}^{n} \lambda_i^2 \} \), \( B_1^2 B_5 = o\{B_1^4 tr(\Sigma^2)\} \). Therefore, \( \text{Var}(V_{n1})/\zeta_2^2 \to 0 \) as \( n \to \infty \), which means \( V_{n1} \to \zeta_2 \).

\[
V_{ni}^{(2)} = E(Z_{ni}^{(2)} | \mathcal{F}_{i-1}) = \binom{n}{2}^{-1} \sum_{1 \leq j < k \leq i-1} I_{2jk},
\]
where \( I_{2jk} = I_{1jk} + I_{1k} + I_{1j} + C_0, \ C_0 = 4\delta^i_\beta \Sigma^3 \delta_\beta (B_1 + \sigma^2) + 4B_2^2 + \Delta tr(\Gamma^' \Sigma \delta_\beta \delta^i_\gamma \Sigma^\Gamma \circ \Gamma^' \delta_\beta \delta^i_\gamma \Gamma) \) and

\[
I_{1jk} = (B_1 + \sigma^2)\eta_j^i \Sigma \eta_k + 5\eta_j^i \Sigma \delta_\beta \delta^i_\gamma \eta_k + \Delta tr(\Gamma^' \eta_j^i \eta_k \Gamma \circ \Gamma^' \delta_\beta \delta^i_\gamma \Gamma) \quad \text{and} \\
I_{1k} = 2(B_1 + \sigma^2)\eta_j^i \Sigma^2 \delta_\beta + 2\eta_j^i \Sigma \delta_\beta \delta^i_\gamma \Sigma^2 \delta_\beta + 2\Delta tr(\Gamma^' \eta_j^i \Sigma \Gamma \circ \Gamma^' \delta_\beta \delta^i_\gamma \Gamma).
\]

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

\[
\text{Var}(V_{n2}) = \left( \frac{n}{2} \right)^2 \sum_{k=2}^{n-1} \sum_{j=1}^{k-1} (n-k)^2 E(I_{2jk} - E(I_{2jk}))^2 \\
+ \left( \frac{n}{2} \right)^2 \sum_{k=2}^{n-1} \sum_{j<i} (n-k)^2 E(I_{1k}^2) + \left( \frac{n}{2} \right)^2 \sum_{k<i} \sum_{j=1}^{k} (n-k)(n-k_i) E(I_{1j}^2).
\]

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

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

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

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 \( S_{ij} = \eta_i^j \eta_j^i + 2\delta_\beta \delta^i_\gamma \Sigma \eta_j + 2\delta^i_\beta \Sigma \eta_i \). Then

\[
J_{1i} = 2 \left( \frac{n}{2} \right)^2 \sum_{j,k}^{i-1} E(S_{ij}^4) + 2 \left( \frac{n}{2} \right)^2 \sum_{j,k}^{i-1} E(S_{ij}^2 S_{ik}^2) \leq 2 \left( \frac{n}{2} \right)^2 \sum_{j,k}^{i-1} E(S_{ij}^4).
\]

By Cauchy-Schwarz inequality, \( E(S_{ij}^4) \leq 8E(\eta_i^j \eta_j^i \eta_i^j \eta_j^i)^2 + 2^{11} E(\eta_i^j \Sigma \delta_\beta \delta^i_\gamma \Sigma \eta_j)^2 \). Again, by Lemma 3, \( E(\eta_i^j \eta_j^i \eta_i^j \eta_j^i)^2 \leq C(B_1^2 + \sigma^4)^2 \{tr(\Sigma^4) + tr^2(\Sigma^2)\} + CB_2^2 B_3^2 + CB_1^3 B_5 + CB_4^2 \) and \( E(\eta_i^j \Sigma \delta_\beta \delta^i_\gamma \Sigma \eta_j)^2 \leq CB_1^2 B_3^2 + CB_2^4 \).

Therefore, \( \sum_{i=1}^{n} J_{1i}/\zeta_2^2 = \frac{4(2n^3 - 3n^2 + n)}{5n^3(n-1)^2} E(S_{ij}^4)/\zeta_2^2 \to 0 \) as \( n \to \infty \). Write

\[
J_{2i} = 4 \left( \frac{n}{2} \right)^2 \sum_{j,k}^{i-1} E(S_{ij}^2 S_{ik}^2) + 8 \left( \frac{n}{2} \right)^2 \sum_{j,k \neq l}^{i-1} E(S_{ij}^2 S_{ik} S_{il} S_{ij}) \\
+ 2 \left( \frac{n}{2} \right)^2 \sum_{j,k \neq l \neq m}^{i-1} E(S_{ij} S_{ik} S_{il} S_{im}) =: J_{2i}^{(1)} + J_{2i}^{(2)} + J_{2i}^{(3)}.
\]
where

\[ J_{2i}^{(2)} = 8 \left( \frac{n}{2} \right)^{-2} \sum_{j \neq k \neq l} E\{ 4(\delta_\beta \Sigma \eta_i)^2(\eta'_j \eta_j)^2 + 16(\delta_\beta \Sigma \eta_i)^4 + 16(\delta_\beta \Sigma \eta_i)^2(\delta_\beta \Sigma \eta_j)^2 \} \]

\[ \leq C \left( \frac{i(i-1)(i-2)}{n^2(n-1)^2} \right) E(S_{ij}^4). \]

\[ J_{2i}^{(3)} = 32 \left( \frac{n}{2} \right)^{-2} \sum_{j \neq k \neq l \neq m} E(\delta_\beta \Sigma \eta_i)^4 \leq C \left( \frac{i(i-1)(i-2)(i-3)}{n^2(n-1)^2} \right) B_1^2 B_3^2. \]

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

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


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.


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 $X_1, \ldots, X_n \sim F$ and

$$X_j = W_j + \mu \quad j = 1, \ldots, n,$$

(3.1.1)
where \( \mu = (\mu_1, \cdots, \mu_p)^T \) is a \( p \)-dimensional unknown vector, \( W_j = (W_{1j}, \cdots, W_{pj})^T \) and \( \{W_{ij}\}_{i=1}^p \) is a sequence of dependent random variables with mean 0 and variance \( \sigma_i^2 \). We are interested in testing the high dimensional mean vector

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

We assume the nonzero components in \( \mu \) are sparse, namely most of \( \mu_i \)'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)

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

(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 $\lambda_n$ and write $T_n(\lambda_n)$ as $T_n$ except in Section 3.4. For discussion simplicity, we shall assume that $\sigma_i^2$ is known. When $\sigma_i^2$ is unknown, we may use the sample variance to estimate $\sigma_i^2$ 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 $\sigma_i^2$ 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 $\lambda_n$. 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 $\mu_i$ 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 $\lambda_n \to \infty$ as $p \to \infty$, 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 $\Phi(\cdot)$ are the density and survival functions of the standard normal distribution, respectively. Let $\rho_k$ be the correlation coefficient between $X_{1i}$ and $X_{(k+1)i}$, and write $\rho_1 = \rho$ for simplicity. Let

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

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 $k \neq l \in \{1, \ldots, p\}$. Then

$$E(T_n) = \sum_{i=1}^{p} G_{T_{n,i}}, \text{ where}$$

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

and
and

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

where

\[
\nu(\mu_i, \lambda_n) = [(\sqrt{n}\lambda_n)^3 + (\sqrt{n}\mu_i)^3 + \sqrt{n}\mu_i\lambda_n + n\sqrt{\lambda_n}\mu_i^2 + 5\sqrt{n}\mu_i + 3\sqrt{\lambda_n}\phi(n^{-i})
\]
\[
+ [(\sqrt{n}\lambda_n)^3 - (\sqrt{n}\mu_i)^3 - \sqrt{n}\mu_i\lambda_n + n\sqrt{\lambda_n}\mu_i^2 - 5\sqrt{n}\mu_i + 3\sqrt{\lambda_n}\phi(n^{+i})
\]
\[
+ [n^2\mu_i^4 + 6n\mu_i^2 + 3] \left( \bar{\Phi}(\eta^{-i}_n) + \Phi(\eta^{+i}_n) \right) - G^2_{T_n, i} \quad \text{and}
\]

\[
\gamma_{j-i}(\sqrt{n}\mu_i, \sqrt{n}\mu_j, \lambda_n) = \bar{E}\{Z_{i,n}(\lambda_n)Z_{j,n}(\lambda_n)\} - G_{T_n, i}G_{T_n, j} \tag{3.2.6}
\]

with the expression of \(\bar{E}\{Z_{i,n}(\lambda_n)Z_{j,n}(\lambda_n)\}\) 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 \(\mu_i\) 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

\[
E(T_n) = L_p^{(1)} \max\{p^{1-\beta},p^{1-s}\} I(r > s) + L_p^{(2)} \max\{p^{1-(\sqrt{r}-\sqrt{\tau})^2-\beta},p^{1-s}\} I(r < s) \quad \text{and}
\]

\[
\Var(T_n) = L_p^{(1)} \max\{p^{1-\beta},p^{1-s}\} I(r > s) + L_p^{(2)} \max\{p^{1-(\sqrt{r}-\sqrt{\tau})^2-\beta},p^{1-s}\} I(r < s)
\]

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

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

(3.2.7)

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

\[
\text{Var}(T_n) = \sum_{j=1}^{p} \text{Var}\{Z_j(\lambda_n)\} + \sum_{j_1 \neq j_2} \text{Cov}\{Z_{j_1}(\lambda_n), Z_{j_2}(\lambda_n)\}
+ \sum_{j_1 \neq j_2} \text{Cov}\{Z_{j_1}(\lambda_n), Z_{j_2}(\lambda_n)\}
:= I_{(1)} + I_{(2)} + I_{(3)} + I_{(4)}.
\]

(3.2.8)

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,

\[
|I_{(2)}| \leq \sum_{j_1 \neq j_2} |\gamma_{j_1-j_2}|(0, 0, \lambda_n)| = L_p \sum_{j_1 \neq j_2} |\rho_{j_1-j_2}| p^{-1+2s|\rho_{j_1-j_2}|} |\rho_{j_1-j_2}| p^{-1+2s|\rho_{j_1-j_2}|} |\rho_{j_1-j_2}| p^{-1+2s|\rho_{j_1-j_2}|} \\
\leq L_p \sum_{j_1 \neq j_2} |\rho_{j_1-j_2}| p^{-2s/(2-\epsilon)} I\{|\rho_{j_1-j_2}| < 1 - \epsilon\}
+ L_p \sum_{j_1 \neq j_2} |\rho_{j_1-j_2}| p^{-2s/(2-\epsilon)} I\{|\rho_{j_1-j_2}| > 1 - \epsilon\} \leq L_pp^{1-2s/(2-\epsilon)} + L_pp^{-s}
\sim L_pp^{1-2s/(2-\epsilon)}.
\]

(3.2.9)
and for \( s < r \), \( I(3) = 4r \log(p) \sum_{j_1 \neq j_2} \rho_{j_1-j_2} U(\eta_n, \eta_n; \rho_{j_1-j_2}) \{1 + o(1)\} \) 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 \( s > r \)

\[
|I(3)| \leq \sum_{j_1 \neq j_2, j_1, j_2 \in S_1} |\rho_{j_1-j_2}| |\rho_{j_1-j_2}^{-1}| \gamma_{j_1-j_2} (\sqrt{n} \mu_{j_1}, \sqrt{n} \mu_{j_2}, \lambda_n) |I\{\rho_{j_1-j_2} \neq 0\}
\]

\[
\leq L_p \sum_{j_1 \neq j_2, j_1, j_2 \in S_1} |\rho_{j_1-j_2}| p^{-2(\sqrt{s}-\sqrt{r})^2/(2-\epsilon)} \leq L_p p^{1-\beta - 2(\sqrt{s}-\sqrt{r})^2/(2-\epsilon)}. \tag{3.2.10}
\]

Also,

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

In summary, \( I(2), I(3) \) and \( I(4) \) are smaller order than \( I(1) \). By (3.2.8), we have

\[
\Var(T_n) = L_p^{(1)} \max\{p^{1-\beta}, p^{1-s}\} |I(r > s) + L_p^{(2)} \max\{p^{1-(\sqrt{s}-\sqrt{r})^2-\beta}, p^{1-s}\} |I(r < s). \tag{3.2.12}
\]

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 \( \alpha \)-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 \((\Omega, \mathcal{F}, P)\) and \( \mathcal{F}_a^b = \sigma\{V_i : i \in (a, b)\} \) is the \( \sigma \)-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 \( \alpha \)-mixing array if

\[
\lim_{k \to \infty} \alpha_V(k) = 0
\]

where

\[
\alpha_V(k) = \sup_{i \in \mathbb{Z}} \alpha(\mathcal{F}_i, \mathcal{F}_{i+k}),
\]

\[
\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A)P(B)| \quad \text{and} \quad \mathcal{A}, \mathcal{B} \text{ are } \sigma \text{-algebras in } \Omega.
\]

**Definition 2.** A stochastic process \( W_t \) is called weak stationary if \( E\{W_t\} = E\{W_{t+\delta}\} \) for any \( \delta \) and covariance \( \text{Cov}\{W_t, W_{t+\delta}\} \) exists and only depends on \( \delta \).
Assume the sequence \( \{X_{ij}\}_{i=1}^{p} \) satisfy the \( \alpha \)-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 \( \{X_{i1}, \ldots, X_{in}\} \), the \( \alpha \)-mixing coefficient \( \alpha_Z(k) \) for \( Z_{i,n} \) satisfy (Refer to Theorem 5.2 in Bradley, 2005)
\[
\alpha_Z(k) \leq \sum_{j=1}^{n} \alpha_{X_j}(k) = n \alpha_X(k), \quad \text{for each } k = 1, \ldots, p,
\]
where \( \alpha_X(k) = \alpha_{X_j}(k) \) for \( j = 1, \ldots, n \) (since \( X_j \) are IID random vectors). Suppose \( k \) depends on \( n \) and denote it by \( k(n) \). If \( \alpha_X\{k(n)\} \to 0 \) as \( k(n) \to \infty \) and \( \alpha_X\{k(n)\} = o(n^{-1}) \), then \( \alpha_Z\{k(n)\} \to 0 \) as \( k(n) \to \infty \), which means that \( \{Z_{i,n}\}_{i=1}^{p} \) is also an \( \alpha \)-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 \( \alpha_Z(k) \to 0 \).

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
\[
\sigma^2_0(p; \lambda_n) = \text{Var}\{Z_{1,n}(\lambda_n)\} + 2 \sum_{k=1}^{p-1} \gamma_k(0, 0, \lambda_n).
\]
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 \( \alpha \)-mixing sequence for each sample \( j \) and \( \{W_{ij}\}_{i=1}^{p} \) in model (3.1.1) is a weak stationary sequence. Let \( \mu_{Tn,0}^{(i)} \) and \( \mu_{Tn,1}^{(i)} \) be the mean of \( Z_{i,n}(\lambda_n) \) under the null and alternative, respectively. Under the conditions in Theorem 1 and assume
\[
\sum_{k=1}^{\infty} k \alpha_X^{\delta/(4+\delta)}(k) < \infty \text{ for some } \delta > 0 \text{ and } p^{6a/5} n^{4/5} \alpha_X^{4/5}(p^c) \to 0 \text{ for some } a \text{ and } c \in (0, 1-a).
\]
(i) If there exists \( a \in ((4+2\delta)s/(4+\delta), 1) \) such that \( n^{\delta/(4+\delta)} p^{(4+2\delta)s/(4+\delta)-a} \to 0 \), then under the null hypothesis in (3.1.2),
\[
p^{-1/2} S_n \overset{d}{\to} N(0, 1),
\]
where \( S_n = \sigma_0^{-1}(p; \lambda_n) \sum_{i=1}^{p} \{Z_{i,n}(\lambda_n) - \mu_{Tn,0}^{(i)}\} \).

(ii) Let \( \mu_i = \sqrt{2r_i \log(p)/n} \). Suppose that there exist \( \sigma^2_1(p; \lambda_n) \) such that
\[
H_l = \sup_{l \geq 1} \left| l^{-1} \text{Var} \left\{ \sum_{i=j}^{j+l-1} Z_{i,n}(\lambda_n) \right\} - \sigma^2_1(l; \lambda_n) \right| \to 0 \text{ as } l \to \infty, \quad (3.3.13)
\]
and \( \sigma^2_1(p; \lambda_n) = L_n p^{-h_1} \) (0 \leq h_1 \leq s). Assume (a) \( \max_i r_i > s \) and exists \( a \in (2h_1, 1) \) such that \( n^{\delta/(4+\delta)} p^{2h_1-a} \to 0 \); (b) \( \max_i r_i < s \) and exists \( a \in (2h_1 - \frac{4s}{4+5}, 1) \) such that
\[ n^{\delta/(4+\delta)}p^{2\beta_1 - 4s^*/4s^* - a} \to 0 \text{ where } s^* = (\sqrt{s} - \max_i \sqrt{r_i})^2. \text{ 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_{i,T_n,1}^{(i)}\} \overset{d}{\to} N(0, 1). \\

\textbf{Remark 1.} Under the null hypothesis, } \{Z_{i,n}(\lambda_n)\}_{i=1}^p \text{ is a weak stationary sequence. The condition } \sum_{k=1}^\infty k\alpha_X^{\delta/(4+\delta)}(k) = O(1) \text{ assures that} \\
\lim_{p \to \infty} \text{Var}\left[p^{-1/2} \sum_{i=1}^p Z_{i,n}(\lambda_n)\right] = \sigma_0^2(p; \lambda_n) < \infty.

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

\text{Under the conditions in Theorem 1, } \mu_{i,T_n,0}^{(i)} = G_{T_n}\{1 + o(1)\} \text{ where} \\
G_{T_n} = 2\sqrt{\lambda_n} \phi(\sqrt{\lambda_n}) + 2\Phi(\sqrt{\lambda_n})

\text{and } \sigma_0^2(p; \lambda_n) = \tilde{\sigma}_0^2(p; \lambda_n)\{1 + o(1)\} \text{ where} \\
\tilde{\sigma}_0^2(p; \lambda_n) = 2[(\sqrt{\lambda_n})^3 + 3\sqrt{\lambda_n}] \phi(\sqrt{\lambda_n}) + 6\Phi(\sqrt{\lambda_n}) + 2 \sum_{k=1}^{p-1} \gamma_k(0, 0, \lambda_n) - G_{T_n}^2, \tag{3.3.14}

\text{where} \\
\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})[\Phi(\theta\lambda_n^{1/2}) + \Phi(-1\lambda_n^{1/2})] \\
+ 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}) \\
+ 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}) \\
+ 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 \tag{3.3.15}

\text{and } \theta = \sqrt{(1 - \rho_k)/(1 + \rho_k).}
From (3.2.9), we know $G_{T_n} = L_p p^{-s}$ and $\bar{\sigma}_0(p; \lambda_n) = L_p p^{-s/2}$. So the approximation error is $o(G_{T_n}/\bar{\sigma}_0(p; \lambda_n)) = o(L_p p^{-s/2}) \to 0$. Therefore, by Theorem 2, we can construct an asymptotic $\alpha$ level test which rejects the null hypothesis if

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

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

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

(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 $\bar{\sigma}_0^2(p; \lambda_n)$ can be obtained though formula (3.3.14). Another method for estimating $\bar{\sigma}_0^2(p; \lambda_n)$ is to use the smoothed spectral density $g(w)$ of $\{Z_{i,n}(\lambda_n) := n\tilde{W}_i^2 I(n\tilde{W}_i^2 > \lambda_n)\}_{i=1}^p$, 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 $\lambda_n$. The basic idea is to maximize the threshold test 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

$$T_\lambda(\lambda_n) = \frac{T_n(\lambda_n) - \mu_{T_n,0}(\lambda_n)}{\sigma_{T_n,0}(\lambda_n)}.$$  

Under the null hypothesis, $T_\lambda(\lambda_n)$ is a zero mean stochastic process with respect to $\lambda_n$. For any different thresholds $\lambda_n$ and $\eta_n$, the correlation between $T_\lambda(\lambda_n)$ and $T_\lambda(\eta_n)$ is

$$\Omega(\lambda_n, \eta_n) := \text{Corr}\{T_\lambda(\lambda_n), T_\lambda(\eta_n)\} = \frac{\text{Cov}\{T_n(\lambda_n), T_n(\eta_n)\}}{\sigma_{T_n,0}(\lambda_n)\sigma_{T_n,0}(\eta_n)}$$  

(3.4.17)

where $\sigma_{T_n,0}^2(\lambda_n) = p\bar{\sigma}_0^2(p; \lambda_n)$. It can be shown that, under the null hypothesis,

$$p^{-1}\text{Cov}\{T_n(\lambda_n), T_n(\eta_n)\} = \left\{2\left[(\sqrt{\lambda_n})^3 + 3\sqrt{\lambda_n}\phi(\sqrt{\lambda_n}) + 6\Phi(\sqrt{\lambda_n})\right]^3$$

$$+ 2\sum_{k=1}^{p-1} \tilde{\gamma}_k(\sqrt{\lambda_n}, \sqrt{\eta_n}) - pG_{T_n}(\lambda_n)G_{T_n}(\eta_n)\right\}\{1 + o(1)\},$$

where $\tilde{\gamma}_k(\lambda_n, \eta_n)$ is given in the Appendix.
Let will also show that the maximum test is able to attain the optimal detection boundary (Donoho give the best power that could be attained by the threshold test in the detectable region. We can be estimated by the quantile of which converges to \( B \). Then \( T_S(\lambda_n) := (T_S(\lambda_{n1}), \cdots, T_S(\lambda_{nd}))' \) 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 \( T_S(t) \) 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

\[
\mathcal{M}_n = \max_{\lambda_n \in \Lambda} T_S(\lambda_n),
\]

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

### 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 
\[ T_{oc,n} = \sum_{i \in S_1} n \hat{X}_i^2. \]

As shown in Theorem 1, we can approximate the mean and variance by assuming that \( \sqrt{n} \hat{X}_i \) is normally distributed. And similar to Theorem 2, we could also show the asymptotic normality of \( T_{oc,n} \).

Under the null hypothesis, 
\[ E(T_{oc,n}) = p_1 \quad \text{and} \quad \text{Var}(T_{oc,n}) = 2p_1 + 2 \sum_{i \neq j, i,j \in S_1} \rho_{[i-j]}^2 = 2p_1 \{1 + o(1)\}. \]

An \( \alpha \) level oracle test will reject the null hypothesis if 
\[ \{ T_{oc,n} > \sqrt{2p_1 z_\alpha + p_1}\}. \]

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 \( \beta \). Recall that \( \mu_{Tn,0} \) and \( \sigma_{Tn,0}^2 \) are the mean and variance of \( T_n \) under the null and \( \mu_{Tn,1} \) and \( \sigma_{Tn,1}^2 \) are the corresponding mean and variance under the alternative. We will reject the null hypothesis if \( T_n > \mu_{Tn,0} + \sigma_{Tn,0} z_\alpha \) at \( \alpha \)-level, where \( z_\alpha \) is the upper \( \alpha \) quantile of the standard normal distribution. So the power of the threshold test will be
\[ \beta(\mu) = \Phi\left(-\frac{\sigma_{Tn,0} z_\alpha + (\mu_{Tn,1} - \mu_{Tn,0})}{\sigma_{Tn,1}}\right). \]

From (3.2.7) and (3.2.12), we have
\[ \mu_{Tn,1} - \mu_{Tn,0} = L_p^{(1)} p^{1-\beta} I\{r > s\} + L_p^{(2)} p^{1-(\sqrt{r} - \sqrt{s})^2 - \beta} I(r \leq s) \]
and
\[
\sigma_{Tn,1} = L_p^{(1)} \max\{p^{(1-\beta)/2}, p^{(1-s)/2}\} I\{r > s\}
+ L_p^{(2)} \max\{p^{(1-(\sqrt{s}-\sqrt{r})^2 - \beta)/2}, p^{(1-s)/2}\} I\{r \leq s\}.
\] (3.5.21)

We should discuss the best power in the detectable region on \((\beta, r)\) 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 \(\alpha_n \to 0\) as \(n \to \infty\) be the significant level and \(\{T_n \geq \mu_{Tn,0} + \sigma_{Tn,0} z_{\alpha_n}\}\) be the rejection region of the threshold test with type I error \(\alpha_n\). Then the sum of type I error and type II error is
\[
\text{Err}_{\alpha_n} := \alpha_n + \Phi\left(-\frac{\sigma_{Tn,0}}{\sigma_{Tn,1}} z_{\alpha_n} + \frac{\mu_{Tn,1} - \mu_{Tn,0}}{\sigma_{Tn,1}}\right).
\]

An alternative is asymptotically detectable if \(\text{Err}_{\alpha_n} \to 0\) as \(n \to \infty\) and undetectable if \(\text{Err}_{\alpha_n} \to 1\). Because \(\sigma_{Tn,1} \geq \sigma_{Tn,0}\) asymptotically, a necessary and sufficient condition for detectable is that
\[
\frac{\mu_{Tn,1} - \mu_{Tn,0}}{\sigma_{Tn,1}} \to \infty.
\]

We verified this condition for any point in the detectable region of \((\beta, r)\) 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 \(L_p p^{1/2} - \frac{1}{2} (r + \beta^2) / (4r)\). In the
Figure 3.1 The detectable region of the threshold and the maximum test in $(\beta, r)$ 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 \( L_p \beta - 1/2 \). In the region I where \( r > \beta \), the best power is of order \( L_p (1 - \beta)/2 \), 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 \( \lambda_n \) 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 \((\beta, r)\) defined in Figure 3.1 as the significant level \( \alpha_n \) goes to 0. Because of the maximum test is an asymptotic \( \alpha_n \) 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 \( \alpha_n \) level rejection region for the maximum test is \( R_{\alpha_n} = \{ M_n > B_{\alpha_n} \} \). The tightness of \( T(\lambda_n) \) for \( \lambda_n \in \Lambda \) ensures the existence of \( M < \infty \) such that \( B_{\alpha_n} < M \) with probability one. Then it is enough to show

\[
P(M_n \rightarrow \infty) \rightarrow 1, \quad \text{as } n \rightarrow \infty \quad (3.5.22)
\]

at every point \((\beta, r)\) in the detectable region. Notice that \( M_n \geq T(\lambda_n) \) 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 \( \lambda_n \) such that \( P(T(\lambda_n) \rightarrow \infty) \rightarrow 1 \). For any fixed \( M < \infty \), we want to show

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

Because of \( \sigma_{T_n,0} \leq \sigma_{T_n,1} \) and finiteness of \( M \), (3.5.23) is true as long as \( \frac{\mu_{T_n,1} - \mu_{T_n,0}}{\sigma_{T_n,1}} \rightarrow \infty \). As we have shown in the Appendix, at every point in the detectable region, there exist a \( \lambda_n \) such that \( \frac{\mu_{T_n,1} - \mu_{T_n,0}}{\sigma_{T_n,1}} \rightarrow \infty \). This concludes that \( T(\lambda_n) \rightarrow \infty \) with probability one for some \( \lambda_n \) and hence the \( M_n \) 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.,

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

where $W_j = (W_{1j}, \cdots, W_{pj})'$ 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 $\rho$ 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 $\rho$ 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 $\rho_k$ 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
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 $\lambda_n = \{2s \log(p) : s = 0.50, 0.55, 0.60, \cdots, 0.90\}$. 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 $\tilde{E}(Z_{1,n}^2, n \lambda_n)Z_{2,n}^2(\lambda_n)$ 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.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Oracle</th>
<th>C-Q</th>
<th>FDR</th>
<th>MAX</th>
<th>0.50</th>
<th>0.55</th>
<th>0.60</th>
<th>0.65</th>
<th>0.70</th>
<th>0.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(p,n)=(1000,20)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>0.050</td>
<td>0.033</td>
<td>0.016</td>
<td>0.048</td>
<td>0.031</td>
<td>0.039</td>
<td>0.041</td>
<td>0.042</td>
<td>0.049</td>
<td>0.046</td>
</tr>
<tr>
<td>0.050</td>
<td>0.066</td>
<td>0.062</td>
<td>0.044</td>
<td>0.064</td>
<td>0.056</td>
<td>0.059</td>
<td>0.056</td>
<td>0.069</td>
<td>0.072</td>
<td>0.058</td>
</tr>
<tr>
<td>0.075</td>
<td>0.078</td>
<td>0.080</td>
<td>0.068</td>
<td>0.074</td>
<td>0.089</td>
<td>0.069</td>
<td>0.078</td>
<td>0.087</td>
<td>0.086</td>
<td>0.078</td>
</tr>
<tr>
<td>0.100</td>
<td>0.087</td>
<td>0.109</td>
<td>0.096</td>
<td>0.084</td>
<td>0.100</td>
<td>0.098</td>
<td>0.098</td>
<td>0.105</td>
<td>0.113</td>
<td>0.110</td>
</tr>
<tr>
<td>0.125</td>
<td>0.100</td>
<td>0.129</td>
<td>0.129</td>
<td>0.093</td>
<td>0.127</td>
<td>0.122</td>
<td>0.116</td>
<td>0.127</td>
<td>0.137</td>
<td>0.144</td>
</tr>
<tr>
<td>0.150</td>
<td>0.112</td>
<td>0.138</td>
<td>0.155</td>
<td>0.105</td>
<td>0.145</td>
<td>0.139</td>
<td>0.131</td>
<td>0.144</td>
<td>0.146</td>
<td>0.158</td>
</tr>
<tr>
<td>0.175</td>
<td>0.120</td>
<td>0.175</td>
<td>0.183</td>
<td>0.113</td>
<td>0.162</td>
<td>0.156</td>
<td>0.157</td>
<td>0.164</td>
<td>0.158</td>
<td>0.164</td>
</tr>
<tr>
<td>0.200</td>
<td>0.132</td>
<td>0.199</td>
<td>0.211</td>
<td>0.130</td>
<td>0.188</td>
<td>0.182</td>
<td>0.178</td>
<td>0.185</td>
<td>0.169</td>
<td>0.167</td>
</tr>
<tr>
<td>$(p,n)=(2000,30)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>0.035</td>
<td>0.023</td>
<td>0.036</td>
<td>0.040</td>
<td>0.034</td>
<td>0.039</td>
<td>0.042</td>
<td>0.043</td>
<td>0.044</td>
<td>0.037</td>
</tr>
<tr>
<td>0.050</td>
<td>0.048</td>
<td>0.042</td>
<td>0.063</td>
<td>0.058</td>
<td>0.057</td>
<td>0.064</td>
<td>0.069</td>
<td>0.071</td>
<td>0.064</td>
<td>0.078</td>
</tr>
<tr>
<td>0.075</td>
<td>0.063</td>
<td>0.060</td>
<td>0.080</td>
<td>0.079</td>
<td>0.088</td>
<td>0.087</td>
<td>0.093</td>
<td>0.092</td>
<td>0.093</td>
<td>0.087</td>
</tr>
<tr>
<td>0.100</td>
<td>0.077</td>
<td>0.081</td>
<td>0.098</td>
<td>0.091</td>
<td>0.107</td>
<td>0.106</td>
<td>0.111</td>
<td>0.113</td>
<td>0.108</td>
<td>0.098</td>
</tr>
<tr>
<td>0.125</td>
<td>0.093</td>
<td>0.094</td>
<td>0.120</td>
<td>0.100</td>
<td>0.130</td>
<td>0.129</td>
<td>0.132</td>
<td>0.136</td>
<td>0.124</td>
<td>0.127</td>
</tr>
<tr>
<td>0.150</td>
<td>0.105</td>
<td>0.124</td>
<td>0.138</td>
<td>0.111</td>
<td>0.143</td>
<td>0.150</td>
<td>0.158</td>
<td>0.155</td>
<td>0.149</td>
<td>0.160</td>
</tr>
<tr>
<td>0.175</td>
<td>0.116</td>
<td>0.150</td>
<td>0.163</td>
<td>0.123</td>
<td>0.167</td>
<td>0.168</td>
<td>0.172</td>
<td>0.164</td>
<td>0.171</td>
<td>0.188</td>
</tr>
<tr>
<td>0.200</td>
<td>0.130</td>
<td>0.177</td>
<td>0.194</td>
<td>0.138</td>
<td>0.190</td>
<td>0.184</td>
<td>0.193</td>
<td>0.178</td>
<td>0.193</td>
<td>0.195</td>
</tr>
<tr>
<td>$(p,n)=(2500,40)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>0.035</td>
<td>0.022</td>
<td>0.022</td>
<td>0.046</td>
<td>0.026</td>
<td>0.031</td>
<td>0.039</td>
<td>0.036</td>
<td>0.046</td>
<td>0.039</td>
</tr>
<tr>
<td>0.050</td>
<td>0.048</td>
<td>0.041</td>
<td>0.046</td>
<td>0.061</td>
<td>0.047</td>
<td>0.056</td>
<td>0.057</td>
<td>0.061</td>
<td>0.068</td>
<td>0.069</td>
</tr>
<tr>
<td>0.075</td>
<td>0.063</td>
<td>0.073</td>
<td>0.070</td>
<td>0.081</td>
<td>0.071</td>
<td>0.073</td>
<td>0.076</td>
<td>0.090</td>
<td>0.089</td>
<td>0.085</td>
</tr>
<tr>
<td>0.100</td>
<td>0.077</td>
<td>0.095</td>
<td>0.093</td>
<td>0.089</td>
<td>0.099</td>
<td>0.096</td>
<td>0.098</td>
<td>0.103</td>
<td>0.108</td>
<td>0.103</td>
</tr>
<tr>
<td>0.125</td>
<td>0.093</td>
<td>0.119</td>
<td>0.118</td>
<td>0.101</td>
<td>0.117</td>
<td>0.126</td>
<td>0.118</td>
<td>0.119</td>
<td>0.124</td>
<td>0.126</td>
</tr>
<tr>
<td>0.150</td>
<td>0.105</td>
<td>0.137</td>
<td>0.145</td>
<td>0.116</td>
<td>0.133</td>
<td>0.150</td>
<td>0.139</td>
<td>0.139</td>
<td>0.145</td>
<td>0.151</td>
</tr>
<tr>
<td>0.175</td>
<td>0.116</td>
<td>0.166</td>
<td>0.167</td>
<td>0.132</td>
<td>0.152</td>
<td>0.169</td>
<td>0.159</td>
<td>0.164</td>
<td>0.173</td>
<td>0.196</td>
</tr>
<tr>
<td>0.200</td>
<td>0.130</td>
<td>0.179</td>
<td>0.184</td>
<td>0.140</td>
<td>0.183</td>
<td>0.184</td>
<td>0.181</td>
<td>0.188</td>
<td>0.202</td>
<td>0.224</td>
</tr>
</tbody>
</table>
Table 3.2 Empirical sizes of the Oracle test, C-Q test, FDR, maximum test and threshold tests with different threshold levels $\lambda_\alpha = 2s\log(p)$ for process with standardized Gamma(2,2) marginal distribution.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Oracle</th>
<th>C-Q</th>
<th>FDR</th>
<th>MAX</th>
<th>0.50</th>
<th>0.55</th>
<th>0.60</th>
<th>0.65</th>
<th>0.70</th>
<th>0.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p,n)=(1000,20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>0.052</td>
<td>0.034</td>
<td>0.021</td>
<td>0.041</td>
<td>0.035</td>
<td>0.029</td>
<td>0.034</td>
<td>0.032</td>
<td>0.036</td>
<td>0.039</td>
</tr>
<tr>
<td>0.050</td>
<td>0.067</td>
<td>0.061</td>
<td>0.039</td>
<td>0.055</td>
<td>0.051</td>
<td>0.052</td>
<td>0.054</td>
<td>0.052</td>
<td>0.057</td>
<td>0.057</td>
</tr>
<tr>
<td>0.075</td>
<td>0.076</td>
<td>0.091</td>
<td>0.069</td>
<td>0.067</td>
<td>0.075</td>
<td>0.076</td>
<td>0.072</td>
<td>0.077</td>
<td>0.072</td>
<td>0.079</td>
</tr>
<tr>
<td>0.100</td>
<td>0.085</td>
<td>0.114</td>
<td>0.094</td>
<td>0.073</td>
<td>0.093</td>
<td>0.097</td>
<td>0.088</td>
<td>0.095</td>
<td>0.094</td>
<td>0.107</td>
</tr>
<tr>
<td>0.125</td>
<td>0.097</td>
<td>0.142</td>
<td>0.120</td>
<td>0.085</td>
<td>0.115</td>
<td>0.117</td>
<td>0.109</td>
<td>0.115</td>
<td>0.121</td>
<td>0.132</td>
</tr>
<tr>
<td>0.150</td>
<td>0.109</td>
<td>0.165</td>
<td>0.144</td>
<td>0.092</td>
<td>0.138</td>
<td>0.140</td>
<td>0.125</td>
<td>0.142</td>
<td>0.142</td>
<td>0.153</td>
</tr>
<tr>
<td>0.175</td>
<td>0.121</td>
<td>0.184</td>
<td>0.160</td>
<td>0.108</td>
<td>0.157</td>
<td>0.157</td>
<td>0.152</td>
<td>0.162</td>
<td>0.158</td>
<td>0.159</td>
</tr>
<tr>
<td>0.200</td>
<td>0.133</td>
<td>0.203</td>
<td>0.186</td>
<td>0.118</td>
<td>0.182</td>
<td>0.180</td>
<td>0.172</td>
<td>0.171</td>
<td>0.177</td>
<td>0.161</td>
</tr>
<tr>
<td>(p,n)=(2000,30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>0.057</td>
<td>0.026</td>
<td>0.023</td>
<td>0.033</td>
<td>0.031</td>
<td>0.030</td>
<td>0.040</td>
<td>0.039</td>
<td>0.037</td>
<td>0.034</td>
</tr>
<tr>
<td>0.050</td>
<td>0.075</td>
<td>0.054</td>
<td>0.042</td>
<td>0.050</td>
<td>0.056</td>
<td>0.056</td>
<td>0.066</td>
<td>0.061</td>
<td>0.059</td>
<td>0.059</td>
</tr>
<tr>
<td>0.075</td>
<td>0.082</td>
<td>0.077</td>
<td>0.062</td>
<td>0.063</td>
<td>0.078</td>
<td>0.080</td>
<td>0.093</td>
<td>0.086</td>
<td>0.080</td>
<td>0.064</td>
</tr>
<tr>
<td>0.100</td>
<td>0.097</td>
<td>0.104</td>
<td>0.081</td>
<td>0.074</td>
<td>0.108</td>
<td>0.107</td>
<td>0.108</td>
<td>0.102</td>
<td>0.099</td>
<td>0.081</td>
</tr>
<tr>
<td>0.125</td>
<td>0.109</td>
<td>0.133</td>
<td>0.103</td>
<td>0.088</td>
<td>0.128</td>
<td>0.128</td>
<td>0.127</td>
<td>0.131</td>
<td>0.123</td>
<td>0.110</td>
</tr>
<tr>
<td>0.150</td>
<td>0.120</td>
<td>0.150</td>
<td>0.122</td>
<td>0.103</td>
<td>0.154</td>
<td>0.147</td>
<td>0.150</td>
<td>0.147</td>
<td>0.146</td>
<td>0.140</td>
</tr>
<tr>
<td>0.175</td>
<td>0.134</td>
<td>0.172</td>
<td>0.150</td>
<td>0.111</td>
<td>0.175</td>
<td>0.164</td>
<td>0.177</td>
<td>0.161</td>
<td>0.179</td>
<td>0.168</td>
</tr>
<tr>
<td>0.200</td>
<td>0.150</td>
<td>0.199</td>
<td>0.175</td>
<td>0.122</td>
<td>0.190</td>
<td>0.191</td>
<td>0.193</td>
<td>0.181</td>
<td>0.200</td>
<td>0.182</td>
</tr>
<tr>
<td>(p,n)=(2500,40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>0.057</td>
<td>0.033</td>
<td>0.022</td>
<td>0.045</td>
<td>0.027</td>
<td>0.032</td>
<td>0.039</td>
<td>0.039</td>
<td>0.044</td>
<td>0.042</td>
</tr>
<tr>
<td>0.050</td>
<td>0.075</td>
<td>0.056</td>
<td>0.052</td>
<td>0.066</td>
<td>0.059</td>
<td>0.056</td>
<td>0.063</td>
<td>0.053</td>
<td>0.069</td>
<td>0.080</td>
</tr>
<tr>
<td>0.075</td>
<td>0.082</td>
<td>0.080</td>
<td>0.076</td>
<td>0.080</td>
<td>0.085</td>
<td>0.088</td>
<td>0.085</td>
<td>0.079</td>
<td>0.097</td>
<td>0.095</td>
</tr>
<tr>
<td>0.100</td>
<td>0.097</td>
<td>0.100</td>
<td>0.092</td>
<td>0.087</td>
<td>0.108</td>
<td>0.110</td>
<td>0.106</td>
<td>0.098</td>
<td>0.115</td>
<td>0.104</td>
</tr>
<tr>
<td>0.125</td>
<td>0.109</td>
<td>0.116</td>
<td>0.123</td>
<td>0.098</td>
<td>0.123</td>
<td>0.129</td>
<td>0.126</td>
<td>0.120</td>
<td>0.123</td>
<td>0.119</td>
</tr>
<tr>
<td>0.150</td>
<td>0.120</td>
<td>0.147</td>
<td>0.144</td>
<td>0.103</td>
<td>0.140</td>
<td>0.146</td>
<td>0.137</td>
<td>0.145</td>
<td>0.140</td>
<td>0.150</td>
</tr>
<tr>
<td>0.175</td>
<td>0.134</td>
<td>0.162</td>
<td>0.173</td>
<td>0.120</td>
<td>0.161</td>
<td>0.165</td>
<td>0.161</td>
<td>0.171</td>
<td>0.168</td>
<td>0.181</td>
</tr>
<tr>
<td>0.200</td>
<td>0.150</td>
<td>0.184</td>
<td>0.201</td>
<td>0.132</td>
<td>0.174</td>
<td>0.181</td>
<td>0.174</td>
<td>0.184</td>
<td>0.188</td>
<td>0.198</td>
</tr>
</tbody>
</table>
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 $\beta = 0.6, 0.7, 0.8$. ($p = 2000, n = 30$)
\[ \beta = 0.6 \quad \beta = 0.7 \quad \beta = 0.8 \]

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

$$\mathcal{E}\{Z_{1,n}(\lambda_n)Z_{2,n}(\lambda_n)\} = A(\eta_{n1}, \sqrt{n}\mu_1, \sqrt{n}\mu_2; \rho)Q(\eta_{n1}, \eta_{n2}; \rho) + A(\eta_{n2}, \sqrt{n}\mu_1, \sqrt{n}\mu_2; \rho)Q(\eta_{n2}, \eta_{n1}; \rho)$$

$$+ B(\eta_{n1}, \eta_{n2}, \sqrt{n}\mu_1, \sqrt{n}\mu_2; \rho)q(\eta_{n1}, \eta_{n2}; \rho) + C(\sqrt{n}\mu_1, \sqrt{n}\mu_2; \rho)U(\eta_{n1}, \eta_{n2}; \rho)$$

$$+ A(\eta_{n1}^+, -\sqrt{n}\mu_1, -\sqrt{n}\mu_2; \rho)Q(\eta_{n1}^+; \eta_{n1}^-; \rho) + A(\eta_{n2}^+, -\sqrt{n}\mu_2, -\sqrt{n}\mu_1; \rho)Q(\eta_{n2}^+; \eta_{n1}^-; \rho)$$

$$+ B(\eta_{n1}^+, \eta_{n2}^+, -\sqrt{n}\mu_1, -\sqrt{n}\mu_2; \rho)q(\eta_{n1}^+; \eta_{n2}^+/; \rho) + C(-\sqrt{n}\mu_1, -\sqrt{n}\mu_2; \rho)U(\eta_{n1}^+, \eta_{n2}^+; \rho)$$

$$+ A(\eta_{n1}^+, -\sqrt{n}\mu_1, -\sqrt{n}\mu_2; -\rho)Q(\eta_{n1}^+, \eta_{n2}^-; -\rho) + A(\eta_{n2}^+, -\sqrt{n}\mu_2, -\sqrt{n}\mu_1; -\rho)Q(\eta_{n2}^+, \eta_{n1}^-; -\rho)$$

$$+ B(\eta_{n1}^+, \eta_{n2}^+, -\sqrt{n}\mu_1, -\sqrt{n}\mu_2; -\rho)q(\eta_{n1}^+; \eta_{n2}^+/; -\rho) + C(-\sqrt{n}\mu_1, -\sqrt{n}\mu_2; -\rho)U(\eta_{n1}^+, \eta_{n2}^+; -\rho)$$

where

$$A(\eta, \mu_1, \mu_2; \rho) = \eta^3 \rho^2 + (1 + 2\rho^2)\eta + 2\mu_2(\eta^3 + 2\rho) + 2\mu_1(1 + \rho^2 + \rho^3 \eta^2)$$

$$+ 4\mu_1\mu_2\rho\eta + \mu_2^2 \eta + \mu_1^2 \rho^2\eta + 2\mu_1\mu_2^2 + 2\rho\mu_1^2 \mu_2;$$

$$B(\eta_1, \eta_2, \mu_1, \mu_2; \rho) = \sqrt{1 - \rho^3}(\rho\eta_1^2 + \eta_1\eta_2 + \rho\eta_2^2 + 3\rho)$$

$$+ 2\mu_2(\eta_1 + \rho\eta_2) + 2\mu_1(\eta_2 + \rho\eta_1) + 4\mu_1\mu_2 + \rho(\mu_1^2 + \mu_2^2));$$

$$C(\mu_1, \mu_2; \rho) = 1 + 2\rho^2 + 4\mu_1\mu_2\rho + \mu_1^2 + \mu_2^2 + \mu_1^2 \mu_2.$$

□

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 $\lambda_n = 2s \log(p)$.

$$\gamma_k(0, 0, \lambda_n) = 4\rho_k^2(\lambda_n^{3/2} + 2\lambda_n^{1/2})\phi(\lambda_n^{1/2})[\Phi (\theta \lambda_n^{1/2}) + \Phi (-\theta^{-1} \lambda_n^{1/2})]$$

$$+ 4\lambda_n^{1/2} \phi(\lambda_n^{1/2})[\Phi (\theta \lambda_n^{1/2}) + \Phi (-\theta^{-1} \lambda_n^{1/2}) - 2\Phi (\lambda_n^{1/2})]$$

$$+ 2\lambda_n \phi(\lambda_n^{1/2}) \sqrt{1 - \rho_k^2 \phi(\theta \lambda_n^{1/2}) + \sqrt{1 - \rho_k^2 \phi(\theta^{-1} \lambda_n^{1/2}) - 2 \phi(\lambda_n^{1/2})}}$$

$$+ 2\rho_k \sqrt{1 - \rho_k^2 (2\lambda_n + 3) \phi(\theta \lambda_n^{1/2}) \phi(\lambda_n^{1/2}) - 2\rho_k \sqrt{1 - \rho_k^2 (2\lambda_n + 3) \phi(\theta^{-1} \lambda_n^{1/2}) \phi(\lambda_n^{1/2})}}$$

$$+ 2[U(\lambda_n^{1/2}, \lambda_n^{1/2}; \rho_k) + U(\lambda_n^{1/2}, \lambda_n^{1/2}; -\rho_k) - 2\Phi^2 (\theta \lambda_n^{1/2})]$$

$$+ 4\rho_k^2 [U(\lambda_n^{1/2}, \lambda_n^{1/2}; \rho_k) + U(\lambda_n^{1/2}, \lambda_n^{1/2}; -\rho_k)]$$

(3.7.25)
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 \( \rho_k \) 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 \( x \to \infty \) and \( x > 0 \), \( \Phi(x) \sim \frac{1}{\sqrt{2\pi}} \), By an inequality given in Willink (2004), we have for \( \rho_k \geq 0 \),

\[
\Phi(\theta_{\lambda_{n}^{1/2}}) \Phi(\lambda_{n}^{1/2}) \leq U(\lambda_{n}^{1/2}, \lambda_{n}^{1/2}; \rho_k) \leq \Phi(\theta_{\lambda_{n}^{1/2}}) \Phi(\lambda_{n}^{1/2})(1 + \rho_k). \tag{3.7.26}
\]

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

\[
\frac{\Phi(\theta_{\lambda_{n}^{1/2}}) - \Phi(\lambda_{n}^{1/2})}{\rho_k \lambda_{n}^{1/2} \phi(\theta_0, \lambda_{n}^{1/2})} \to 1. \tag{3.7.27}
\]

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

\[
\sqrt{1 - \rho_k^2} \phi(\theta x) - \phi(x) = \sqrt{1 - \rho_k^2} \{\phi(\theta x) - \phi(x)\} + (\sqrt{1 - \rho_k^2} - 1)\phi(x) = \sqrt{1 - \rho_k^2} \phi(\theta_0 x)(\theta'_0 x)(1 - \theta)x + (\sqrt{1 - \rho_k^2} - 1)\phi(x)
\]

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

\[
\frac{\sqrt{1 - \rho_k^2} \phi(\theta_{\lambda_{n}^{1/2}}) - \phi(\lambda_{n}^{1/2})}{\rho_k \lambda_{n}^{1/2} \phi(\theta_0, \lambda_{n}^{1/2}) + \rho_k^2 \phi(\lambda_{n}^{1/2})} \to 1. \tag{3.7.28}
\]

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

\[
\frac{U(\lambda_{n}^{1/2}, \lambda_{n}^{1/2}; \rho_k) - \Phi^2(\lambda_{n}^{1/2})}{\rho_k \lambda_{n}^{1/2} \phi(\theta_0, \lambda_{n}^{1/2}) \Phi(\lambda_{n}^{1/2})} \to 1. \tag{3.7.29}
\]

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

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,

\[
\gamma_k(\sqrt{n}_1, \sqrt{n}_{k+1}, \lambda_n) = \left\{ (\eta_n^+ \phi(\eta_n^-) + (n\mu_1^2 + 1)(\Phi(\eta_n^-))) \right\}^2
\]

\[
= 2A(\eta_n^-, \sqrt{n}_1, \sqrt{n}_{k+1}; \rho_k) Q(\eta_n^-, \eta_{n(k+1)}^-; \rho_k) - 2(\eta_n^+)(n\mu_1^2 + 1)\phi(\eta_n^-)\Phi(\eta_n^-)
\]

\[
+ B(\eta_n^-, \eta_{n(k+1)}^-; \sqrt{n}_1, \sqrt{n}_{k+1}; \rho_k) q(\eta_n^-, \eta_{n(k+1)}^-; \rho_k) - (\eta_n^+)^2\phi^2(\eta_n^-)
\]

\[\]

\[
+ C(\sqrt{n}_1, \sqrt{n}_{k+1}; \rho_k) U(\eta_n^-, \eta_{n(k+1)}^-; \rho_k) - (n\mu_1^2 + 1)^2\Phi^2(\eta_n^-)
\]

\[
(3.7.30)
\]

where

\[
A(\eta_n^-, \sqrt{n}_1, \sqrt{n}_{k+1}; \rho_k) Q(\eta_n^-, \eta_{n(k+1)}^-; \rho_k) - (\eta_n^+)(n\mu_1^2 + 1)\phi(\eta_n^-)\Phi(\eta_n^-)
\]

\[
= (\eta_n^+)(n\mu_1^2 + 1)\phi(\eta_n^-)\{\Phi(\theta \eta_n^-) - \Phi(\eta_n^-)\} + \rho_k\{\eta_n^-3 \rho_k + 2\rho_k \eta_n^- + 2\sqrt{n}_1(n_{n_1}^- + 2)
\]

\[
+ 2\sqrt{n}_1(\rho_k + \eta_n^-2) + 4n\mu_1\mu_{k+1}\eta_n^- + n\mu_1^2\rho_k \eta_n^- + 2n^{3/2}\mu_1^2\mu_{k+1}Q(\eta_{n1}, \eta_{n(k+1)}^-; \rho_k);
\]

\[
(3.7.31)
\]

\[
B(\eta_n^-, \eta_{n(k+1)}^-; \sqrt{n}_1, \sqrt{n}_{k+1}; \rho_k) q(\eta_n^-, \eta_{n(k+1)}^-; \rho_k) - (\eta_n^+)^2\phi^2(\eta_n^-)
\]

\[
= \sqrt{1 - \rho_k^2 - \phi(\theta \eta_n^-) \phi(\eta_n^-)}(\eta_n^+)^2
\]

\[
+ \sqrt{1 - \rho_k^2 \rho_k}\{\eta_n^- + \sqrt{n}_1(\eta_n^- + \sqrt{n}_{k+1})^2 + 3\} q(\eta_n^-, \eta_{n(k+1)}^-; \rho_k)
\]

\[
(3.7.32)
\]

and

\[
C(\sqrt{n}_1, \sqrt{n}_{k+1}; \rho_k) U(\eta_n^-, \eta_{n(k+1)}^-; \rho_k) - (n\mu_1^2 + 1)^2\Phi^2(\eta_n^-)
\]

\[
= \{U(\eta_n^-, \eta_{n(k+1)}^-; \rho_k) - \Phi^2(\eta_n^-)\}(n\mu_1^2 + 1)^2 + 2\rho_k(\rho_k + 2n\mu_1\mu_{k+1}) U(\eta_{n1}, \eta_{n(k+1)}^-; \rho_k)
\]

\[
= \Phi(\eta_n^-)\{c\Phi(\theta \eta_n^-) - \Phi(\eta_n^-)\}(n\mu_1^2 + 1)^2 + 2\rho_k(\rho_k + 2n\mu_1\mu_{k+1}) U(\eta_{n1}, \eta_{n(k+1)}^-; \rho_k)
\]

\[
(3.7.33)
\]

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

Notice that \(\Phi(-\theta \eta_n^-) - \Phi(-\eta_n^-) = \Phi(\eta_n^-) - \Phi(\theta \eta_n^-)\) and \(U(-\eta_n^-, -\eta_{n2}; \rho_k) - \Phi^2(-\eta_n^-) = U(\eta_{n1}, \eta_{n2}; \rho_k) - \Phi^2(\eta_n^-)\). So no matter \(\eta_n^- > 0\) or \(\eta_n^- < 0\), the above differences maintain the same order. Hence, without loss of generality, we can assume \(\eta_n^- > 0\). Similar to (3.7.27), (3.7.28) and (3.7.29), we have \(\{\rho_k \eta_n^- \phi(\theta \eta_n^-)\}^{-1}(\Phi(\theta \eta_n^-) - \Phi(\eta_n^-)) \to 1\),

\[
\frac{\sqrt{1 - \rho_k^2 \phi(\theta \eta_n^-) \phi(\eta_n^-)}}{\rho_k \eta_n^- \phi(\theta \eta_n^-) + \rho_k^2 \phi(\eta_n^-)} \to 1\quad \text{and} \quad \frac{U(\eta_n^-, \eta_{n2}; \rho_k) - \Phi^2(\eta_n^-)}{\rho_k \eta_n^- \phi(\theta \eta_n^-) \Phi(\eta_n^-)} \to 1.
\]
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, $q(\eta_{n1}, \eta_{n2}; \rho_k) \sim p^{-\frac{2}{1+\rho_k} (\sqrt{s} - \sqrt{r})^2}$, $Q(\eta_{n1}, \eta_{n2}; \rho_k) \sim I(s > r) L_p p^{-\frac{2}{1+\rho_k} (\sqrt{s} - \sqrt{r})^2} + I(s < r) L_p p^{-(\sqrt{s} - \sqrt{r})^2}$ and $U(\eta_{n1}, \eta_{n2}; \rho_k) \sim I(s > r) L_p p^{-\frac{2}{1+\rho_k} (\sqrt{s} - \sqrt{r})^2} + I(s < r) 4r \log(p) U(\eta_{n1}, \eta_{n2}; \rho_k)$.

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

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

(3.7.34)

For $\rho_k < 0$, note that for any $\eta > 0$, $\hat{\Phi}(\theta \eta) \bar{\Phi}(\eta) (1 + \rho_k) \leq U(\eta, \eta; \rho_k) \leq \hat{\Phi}(\theta \eta) \Phi(\eta)$. 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

$$J_1 := 2 A(n_{n1}, \sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k) Q(\eta_{n1}, \eta_{n(k+1)}; \rho_k) - 2(\eta_{n1}^*) \Phi(\eta_{n1}^*)(n\mu_{k+1}^2 + 1) \Phi(\eta_{n(k+1)}^-)$$

$$+ B(\eta_{n1}^*, \eta_{n(k+1)}, \sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k) Q(\eta_{n1}, \eta_{n(k+1)}^-; \rho_k) - (\eta_{n1}^*)^2 \Phi(\eta_{n1}^-)$$

$$+ C(\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; \rho_k) U(\eta_{n1}, \eta_{n(k+1)}^-; \rho_k) - (n\mu_{k+1}^2 + 1)^2 \Phi(\eta_{n(k+1)}^-)^2$$

and

$$J_2 := A(\eta_{n(k+1)}^+, n\mu_{k+1}, \sqrt{n}\mu_1; -\rho_k) Q(\eta_{n(k+1)}^+, \eta_{n1}; -\rho_k) - (\eta_{n(k+1)}^+)^* \Phi(\eta_{n(k+1)}^+)(n\mu_{k+1}^2 + 1) \Phi(\eta_{n1}^-)$$

$$+ A(\eta_{n1}^+, \sqrt{n}\mu_1, n\mu_{k+1}; -\rho_k) Q(\eta_{n1}^+, \eta_{n(k+1)}^-; -\rho_k) - (\eta_{n1}^+)^* \Phi(\eta_{n1}^+)(n\mu_{k+1}^2 + 1) \Phi(\eta_{n(k+1)}^-)$$

$$+ B(\eta_{n1}^+, \eta_{n(k+1)}^-, \sqrt{n}\mu_1, -\sqrt{n}\mu_{k+1}; \rho_k) Q(\eta_{n1}^+, \eta_{n(k+1)}^-; -\rho_k) - \eta_{n(k+1)}^+ \eta_{n(k+1)}^- \eta_{n1}^+ \Phi(\eta_{n1}) \Phi(\eta_{n(k+1)})$$

$$+ B(\eta_{n1}^+, \eta_{n(k+1)}^-, \sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; -\rho_k) Q(\eta_{n1}^+, \eta_{n(k+1)}^-; -\rho_k) - \eta_{n(k+1)}^+ \eta_{n(k+1)}^- \eta_{n1}^+ \Phi(\eta_{n1}) \Phi(\eta_{n(k+1)}^-)$$

$$+ 2 C(\sqrt{n}\mu_1, \sqrt{n}\mu_{k+1}; -\rho) U(\eta_{n1}^+, \eta_{n(k+1)}^+; -\rho) - (n\mu_{k+1}^2 + 1)^2 \Phi(\eta_{n(k+1)}^+)^2 \Phi(\eta_{n(k+1)}^+)$$.
In a similar fashion as that for $\rho_k \geq 0$, we will have for $\rho_k < 0$,

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

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

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

and the “∼” 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

$$
\gamma_k(\mu_1, 0, \lambda_n) = \left\{ A(\eta_{n1}, \sqrt{n\mu_1}, 0; \rho_k) Q(\eta_{n1}, \eta_{n_{(k+1)}}, \rho_k) - (\eta_{n1}^+ \phi(\eta_{n1}) \Phi(\eta_{n_{(k+1)}})) I(s > r) + \left\{ A(\eta_{n1}, \sqrt{n\mu_1}, 0; -\rho_k) Q(\eta_{n1}, \eta_{n_{(k+1)}}, -\rho_k) - (\eta_{n1}^+ \phi(\eta_{n1}) \Phi(\eta_{n_{(k+1)}})) I(s < r) + A(\eta_{n_{(k+1)}}, 0, \sqrt{n\mu_1}; \rho_k) Q(\eta_{n_{(k+1)}}, \eta_{n1}; \rho_k) - (\eta_{n_{(k+1)}}^+ \eta_{n1}^- \phi(\eta_{n_{(k+1)}}) \Phi(\eta_{n1})) + B(\eta_{n1}, \eta_{n_{(k+1)}}, \sqrt{n\mu_1}, 0; \rho_k) q(\eta_{n1}, \eta_{n_{(k+1)}}; \rho_k) - (\eta_{n1}^+ \eta_{n_{(k+1)}}) \phi(\eta_{n1}) \Phi(\eta_{n_{(k+1)}}) + C(\sqrt{n\mu_1}, 0; \rho_k) U(\eta_{n1}, \eta_{n_{(k+1)}}, \rho_k) - (\eta_{n1}^2 + 1) \Phi(\eta_{n1}) \Phi(\eta_{n_{(k+1)}}) \right\} \right. \tag{3.7.37}
$$

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

$$
P_I = \rho_k \left\{ \eta_{n1}^{-3} \rho_k + 2 \rho_k \eta_{n1}^{-2} + 2 \sqrt{n\mu_1} \rho_k \eta_{n1}^{-2} + n \mu_1^2 \rho_k \eta_{n1}^{-2} \right\} Q(\eta_{n1}, \eta_{n_{(k+1)}}, \rho_k) + (\eta_{n1}^+ \phi(\eta_{n1}) \Phi(\eta_{n_{(k+1)}})) \right\} \right. \tag{3.7.37}
$$

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

$$
\Phi \left( \frac{\eta_{n_{(k+1)}} - \rho_k \eta_{n1}}{1-\rho_k^2} \right) - \Phi(\eta_{n_{(k+1)}}) \sim \rho_k \left\{ I(s > \frac{1}{4}(1+\theta^{-1})^2) L_p p^{-\frac{(1-\rho_k) \sqrt{s-\sqrt{r}} + r}{1-\rho_k^2}} - I(r < s < \frac{1}{4}(1+\theta^{-1})^2) L_p p^{-s} \right\}. \tag{3.7.37}
$$
Thus \( P_I \sim \rho_k \left\{ I(s > \frac{1}{4}(1 + \theta^{-1})^2 r) L_p p^{-\xi} - I(r < s < \frac{1}{4}(1 + \theta^{-1})^2 r) L_p p^{-s-(\sqrt{\tau} - \sqrt{\tau})^2} \right\}. \)

For \( P_{II} \), we only need consider \( s < r \),

\[
P_{II} = \rho_k \left\{ n_{n_1}^{-3} \rho_k + 2 \rho_k n_{n_1}^{-2} + 2 \sqrt{n} \mu_1 (\rho_k - n_{n_1}^{-2}) + n \mu_1^2 \rho_k n_{n_1}^{-1} \right\} Q(n_{n_1}^{-}, n_{n(k+1)}^{+}; \rho_k)
\]

\[
+ (n_{n_1}^{+}) \phi(n_{n_1}^{-}) \left\{ \bar{\Phi} \left( \frac{n_{n(k+1)}^{+} + \rho_k n_{n_1}^{-}}{\sqrt{1 - \rho_k^2}} \right) - \Phi(n_{n(k+1)}^{-}) \right\}
\]

where

\[
Q(n_{n_1}^{-}, n_{n(k+1)}^{+}; \rho_k) \sim I(s < \frac{\rho_k^2 r}{(1 + \rho_k)^2}) L_p p^{-(\sqrt{\tau} - \sqrt{\tau})^2}
\]

\[
+ I(r > s > \frac{\rho_k^2 r}{(1 + \rho_k)^2}) L_p p^{-\xi} - \frac{4 \rho_k}{1 - \rho_k^2} (s - \sqrt{s^2 r}) \text{ and }
\]

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

\[
+ I(\frac{\rho_k^2 r}{(1 + \rho_k)^2} < s < r) L_p p^{-(\frac{1 + \rho_k \sqrt{\tau} - \rho_k \sqrt{\tau}}{1 - \rho_k^2})^2} \right\}.
\]

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

Now let us consider \( P_{III} \),

\[
P_{III} = \rho_k \left\{ n_{n_1}^{-3} \rho_k + 2 \rho_k n_{n_1}^{-2} + 2 \sqrt{n} \mu_1 (2 + n_{n_1}^{-2}) \right\} Q(n_{n_1}^{-}, n_{n(k+1)}^{-}; \rho_k)
\]

\[
+ (n_{n_1}^{+}) (n \mu_1^2 + 1) \phi(n_{n_1}^{-}) \left\{ \bar{\Phi} \left( \frac{n_{n_1}^{-} - \rho_k n_{n_1}^{-}}{\sqrt{1 - \rho_k^2}} \right) - \Phi(n_{n_1}^{-}) \right\}
\]

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

\[
\bar{\Phi} \left( \frac{n_{n_1}^{-} - \rho_k n_{n_1}^{-}}{\sqrt{1 - \rho_k^2}} \right) - \Phi(n_{n_1}^{-}) \sim \rho_k \left\{ I(s > \frac{r}{(1 - \rho_k)^2}) L_p p^{-(\frac{1 + \rho_k \sqrt{\tau} - \rho_k \sqrt{\tau}}{1 - \rho_k^2})^2}
\]

\[
+ I(r < s < \frac{r}{(1 - \rho_k)^2}) C + I(s < r) L_p p^{-(\sqrt{\tau} - \sqrt{\tau})^2} \right\}.
\]

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 \( P_{IV} \)

\[
P_{IV} = \rho_k \sqrt{1 - \rho_k^2} (\sqrt{\bar{\Phi}(\eta_{n_1}^{-2} + \eta_{n_{n_1}}^{-2} + 3 + 2 \sqrt{n} \mu_1 \eta_{n_1}^{-2} + n \mu_1^2) \eta_{n_1}^{-1} \eta_{n_{n(k+1)}}^{-1}; \rho_k})
\]

\[
+ (n_{n_1}^{+}) (n_{n_1}^{-2}) \left\{ \sqrt{1 - \rho_k^2} \eta_{n_1}^{-2} \eta_{n_{n(k+1)}}^{-1}; \rho_k) - \phi(n_{n_1}^{-}) \phi(n_{n_{n(k+1)}}^{-1}) \right\}\]
where $q(\eta_1, \eta_{n+1}; \rho_k) \sim p^{-\xi}$ and $\sqrt{1 - \rho_k^2} q(\eta_1, \eta_n; \rho_k) - \phi(\eta_1) \phi(\eta_n) \sim \rho_k p^{-\xi}$. Thus $P_{IV} \sim \rho_k L_p p^{-\xi}$. Finally, taking a look at $P_V$

$$P_V = 2 \rho_k^2 U(\eta_1, \eta_2; \rho_k) + (n \mu_1^2 + 1) \{ U(\eta_1, \eta_2; \rho_k) - \Phi(\eta_1) \Phi(\eta_2) \}.$$  

Because $U(\eta_1, \eta_2; \rho_k) = z(\rho_k) \Phi(\eta_2) \Phi(\eta_1) \frac{\eta_1 - \rho \eta_2}{\sqrt{1 - \rho_k^2}}$ for some $z(\rho_k) \to 1$ as $\rho_k \to 0$, $P_V$ has the same order as $P_{III}$ up to a slow varying function.

Thus, in summary,

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

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

**Spectral Density Based Estimation of $\sigma_0^2(p; \lambda_n)$:** Let $\iota^2 = -1$ be the imaginary number. The spectral density of $\{Z_{t,n}^0(\lambda_n)\}_{i=1}^p$ is

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

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

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

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

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

where $\hat{\gamma}(\tau) = \frac{1}{p} \sum_{t=1}^{p-\tau} (Z_{t,n}^0(\lambda_n) - Z_0)(Z_{t+n, n}^0(\lambda_n) - Z_0)$ and $Z_0 = \frac{1}{p} \sum_{t=1}^{p} Z_{t,n}^0$. 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
\[
\hat{I}_p(w_k) = \sum_{\tau=-(p-1)}^{p-1} \hat{\gamma}_{\tau}(0,0,\lambda_n)e^{-i\tau w_k}.
\]

By Brillinger (1981) for \( k \in T \),
\[
I_p(w_k) = 2\pi g(w_k)E_k + R_k
\]
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 \( \varepsilon_k \) are IID random variables with mean 0 and variance \( \pi^2/6 \). Hence we can estimate \( m(w) \) by
\[
\hat{m}_h(w) = \frac{\sum_{k \in T} K_h(w - w_k)W_k}{\sum_{k \in T} K_h(w - w_k)},
\]
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 \( \hat{\sigma}_0^2(p;\lambda_n) = 2\pi \hat{g}(0) \).

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, \( \mu_{Tn,1} - \mu_{Tn,0} = L_p p^{1-\beta} \) and \( \sigma_{Tn,1} = \sigma_{Tn,0} = L_p p^{(1-s)/2} \). Hence
\[
\frac{\mu_{Tn,1} - \mu_{Tn,0}}{\sigma_{Tn,1}} = L_p p^{(1+s-2\beta)/2}.
\]
So to make the test detectable, i.e. such that \( (\mu_{Tn,1} - \mu_{Tn,0})/\sigma_{Tn,1} \to \infty \), \( s > 2\beta - 1 \). It follows that the detectable region in \((\beta, r)\) plane for this case is \( r \geq 2\beta - 1 \) (Chen and Xu,
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, $\mu_{T_{n,1}} - \mu_{T_{n,0}} = L_p p^{1-\beta}$, $\sigma_{T_{n,1}} = L_p p^{(1-\beta)/2}$, and $\sigma_{T_{n,0}} = L_p p^{(1-s)/2}$. Then

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

So the detectable region in $(\beta, r)$ 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, $\mu_{T_{n,1}} - \mu_{T_{n,0}} = L_p p^{1-(\sqrt{s} - \sqrt{r})^2 - \beta}$, $\sigma_{T_{n,1}} = \sigma_{T_{n,0}} = L_p p^{(1-s)/2}$. Then

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

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 $(\beta, r)$ 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 $L_p p^2 (1 - \beta + r)$. If $r > \beta/3$, the best power is of order $L_p p^2 (1 - (r + \beta)^2/(4r))$, 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, $\mu_{T_{n,1}} - \mu_{T_{n,0}} = L_p p^{1-(\sqrt{s} - \sqrt{r})^2 - \beta}$, $\sigma_{T_{n,1}} = L_p p^{(1-(\sqrt{s} - \sqrt{r})^2 - \beta)/2}$. Then

$$\frac{\mu_{T_{n,1}} - \mu_{T_{n,0}}}{\sigma_{T_{n,1}}} = L_p p^{(1-(\sqrt{s} - \sqrt{r})^2 - \beta)/2}. \quad (3.7.42)$$
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 \( L_p p^{2 - \frac{1}{2}(r + \beta)^2/(4r)} \). \( \square \)

**Lemma 1** Let \( X_i = (X_{i1}, X_{i2})' \) be IID random vectors with mean zero and covariance \( \Sigma \). 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

\[ 1 - F_n(t_1, t_2) = \exp \left\{ \frac{\theta(t_1, t_2)}{\sqrt{n}} \right\} \Phi_0,\Sigma(t) \{1 + O(n^{-1/2})\} \] (3.7.43)

hold uniformly for \( t_1, t_2 = o(n^{1/2}) \) where \( \Phi_0,\Sigma(t) = (2\pi|\Sigma|^{1/2})^{-1} \int_t^\infty \exp \left(-\frac{1}{2}y/\Sigma^{-1}y\right) dy \) and

\[ \theta(t_1, t_2) = b_{30} t_1^3 + b_{12} t_1^2 t_2 + b_{21} t_1 t_2^2 + b_{03} t_2^3 \]

with

\[ b_{30} = \frac{1}{2} \kappa(2,1) c_2^2 + \frac{1}{2} \kappa(1,2) c_1 c_2 + \frac{1}{3!} \kappa(3,0) c_1^3 + \frac{1}{3!} \kappa(0,3) c_2^3; \]

\[ b_{03} = \frac{1}{2} \kappa(2,1) c_1 c_2^2 + \frac{1}{2} \kappa(1,2) c_2^2 + \frac{1}{3!} \kappa(3,0) c_2^3 + \frac{1}{3!} \kappa(0,3) c_1^3; \]

\[ b_{12} = \frac{1}{2} \kappa(2,1)(c_1^2 + 2c_1 c_2^3) + \frac{1}{2} \kappa(1,2)(c_2^2 + 2c_1^2) + \frac{1}{2} \kappa(3,0) c_1^2 c_2 + \frac{1}{2} \kappa(0,3) c_1 c_2^2 \]

and

\[ b_{21} = \frac{1}{2} \kappa(2,1)(c_2^3 + 2c_1^2 c_2) + \frac{1}{2} \kappa(1,2)(c_1^3 + 2c_1 c_2^2) + \frac{1}{2} \kappa(3,0) c_1 c_2^2 + \frac{1}{2} \kappa(0,3) c_1^2 c_2, \]

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

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

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

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

\[
\tilde{v}(t) = \int \int \exp(it'\tilde{\Sigma}^{-1/2} x) d\tilde{V}(x) = \frac{1}{R(h)} \int \int \exp\{i(h - it')\tilde{\Sigma}^{-1/2} x\} d\tilde{V}(x)
= \frac{1}{R(h)} v(\tilde{\Sigma}^{-1/2} t - ih).
\]

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

\[
W_n(x) = \int_{-\infty}^{\sqrt{n}x} \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \exp(-it'y)\tilde{w}_n(t) dtdy
= \int_{-\infty}^{\sqrt{n}x} \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \exp(-it'y)\tilde{w}_n(\tilde{\Sigma}^{1/2}(t + ih)) dtdy
= \int_{-\infty}^{\infty} R^n e^{-h'y|\tilde{\Sigma}|^{-1/2}} \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \exp(-iz'\tilde{\Sigma}^{-1/2} y)\tilde{w}_n(z) dzdy
= \int_{-\infty}^{\infty} R^n e^{-h'y|\tilde{\Sigma}|^{-1/2}} \int_{-\infty}^{\sqrt{n}x} \tilde{f}_{\tilde{S}_n}(\tilde{\Sigma}^{-1/2} y) dy
= \int_{-\infty}^{\sqrt{n}x} R^n e^{-h'y|\tilde{\Sigma}|^{-1/2}} d\tilde{W}_n(\tilde{\Sigma}^{-1/2} y).
\]

So it follows that

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

where \( B = \{t : \Sigma^{1/2} t > x - \sqrt{n}\tilde{m}\}. \)
Now let \( x = \sqrt{n\tilde{m}} \), we get

\[
1 - F_n(\sqrt{n\tilde{m}}) = \exp\{n \log R - n\tilde{m}\} \int_{\{t: \Sigma^2 t > 0\}} e^{-\sqrt{n\tilde{m}}} t \, d\tilde{F}_n(t) \\
= \exp\{n \log R - n\tilde{m}\} \int_{\{t: \Sigma^2 t > 0\}} e^{-\sqrt{n\tilde{m}}} t \, d\Phi_0(t) \\
+ \exp\{n \log R - n\tilde{m}\} \int_{\{t: \Sigma^2 t > 0\}} e^{-\sqrt{n\tilde{m}}} t \, d(\tilde{F}_n(t) - \Phi_0(t)) \\
:= \exp\{n \log R - n\tilde{m}\} \{I + II\} = P_I + P_{II}.
\]

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

\[
\exp\{n \log R - n\tilde{m}\} \int_{\{t: \Sigma^2 t > 0\}} e^{-\sqrt{n\tilde{m}}} t \, d\Phi_0(t) \\
= \exp\{n \log R - n\tilde{m}\} \frac{1}{2\pi |\Sigma|^{\frac{1}{2}}} \int_0^\infty \exp\{-\sqrt{n\tilde{m}}\xi - \frac{\xi'(\Sigma^{-1}\xi)}{2}\} d\xi \\
= \exp\{n \log R - n\tilde{m} + \frac{n\tilde{m}\Sigma h}{2}\} \frac{1}{2\pi |\Sigma|^{\frac{1}{2}}} \int_0^\infty \exp\{-\frac{\left(\xi + \sqrt{n\tilde{m}}\sqrt{\Sigma}h\right)\Sigma^{-1}\left(\xi + \sqrt{n\tilde{m}}\sqrt{\Sigma}h\right)}{2}\} d\xi.
\]

For sufficient small \( h \),

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

where \( k_\nu \) is the cumulant of order \( \nu = (\nu_1, \nu_2) \). \( k_0 = 0 \) if \( |\nu| = 1 \). \( k_\nu = \text{Cov}(X_{i1}, X_{i2}) \) 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

\[
\frac{\partial \log R(h)}{\partial h_1} = k_{(1,1)} h_2 + k_{(2,0)} h_1 + \frac{1}{2} k_{(3,0)} h_2^2 + k_{(2,1)} h_1 h_2 + \frac{1}{2} k_{(1,2)} h_2^2 \\
\frac{\partial \log R(h)}{\partial h_2} = k_{(1,1)} h_1 + k_{(0,2)} h_2 + \frac{1}{2} k_{(3,0)} h_2^2 + k_{(1,2)} h_1 h_2 + \frac{1}{2} k_{(2,1)} h_2^2.
\]

Let \( a = \tilde{m} \) and

\[
\begin{pmatrix}
    a_1 \\
    a_2
\end{pmatrix} = \begin{pmatrix}
    k_{(1,1)} h_2 + k_{(2,0)} h_1 + \frac{1}{2} k_{(3,0)} h_2^2 + k_{(2,1)} h_1 h_2 + \frac{1}{2} k_{(1,2)} h_2^2 \\
    k_{(1,1)} h_1 + k_{(0,2)} h_2 + \frac{1}{2} k_{(3,0)} h_2^2 + k_{(1,2)} h_1 h_2 + \frac{1}{2} k_{(2,1)} h_2^2
\end{pmatrix}.
\]
Then the approximate solution of above equations are:

\[
\begin{pmatrix}
    h_1 \\
    h_2
\end{pmatrix} = \begin{pmatrix}
    c_1a_1 + c_2a_2 \\
    d_1a_1 + d_2a_2
\end{pmatrix}
\]

(3.7.47)

where \(d_1 = c_2\) and \(d_2 = c_1\). We can also know that

\[
\tilde{\Sigma} = \frac{\partial^2 \log R(h)}{\partial h \partial h} = \begin{pmatrix}
    k_{(2,0)} + k_{(3,0)}h_1 + k_{(2,1)}h_2 & k_{(1,1)} + k_{(2,1)}h_1 + k_{(1,2)}h_2 \\
    (k_{(1,1)} + k_{(2,1)}h_1 + k_{(1,2)}h_2 & k_{(0,2)} + k_{(0,3)}h_2 + k_{(1,2)}h_1
\end{pmatrix}
\]

\[
h_{1,2} \rightarrow 0 \left( \begin{array}{cc}
    k_{(2,0)} & k_{(1,1)} \\
    k_{(1,1)} & k_{(0,2)}
\end{array} \right) = \Sigma.
\]

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

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

(3.7.48)

Plugging (3.7.47) into (3.7.48), we get

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

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

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

(3.7.49)

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

Let \(t = \sqrt{n}a\). 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

\[
I = \frac{1}{2\pi} |\Sigma|^{-\frac{1}{2}} \int_0^\infty \exp\{-\sqrt{n}h'\xi - \frac{\xi'\Sigma^{-1}\xi}{2}\} d\xi
\]

\[
= \exp\left\{-\frac{nh'\Sigma h}{2}\right\} \frac{1}{2\pi} |\Sigma|^{-\frac{1}{2}} \int_0^\infty \exp\{-\left(\xi + \sqrt{n}\Sigma h\right)\Sigma^{-1}\left(\xi + \sqrt{n}\Sigma h\right)\} d\xi.
\]

(3.7.49)
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

$$
\frac{1}{2\pi} \sqrt{\tilde{\Sigma}}^{-\frac{1}{2}} \int_0^\infty \exp\left\{ -\frac{(\xi + \sqrt{n\tilde{\Sigma}}h)'\tilde{\Sigma}^{-1}(\xi + \sqrt{n\tilde{\Sigma}}h)}{2} \right\} d\xi
$$

$$
= \frac{1}{2\pi} |L|L^{-\frac{1}{2}} \int_0^\infty \exp\left\{ -\frac{(\xi^* + L^{-1}\tilde{\Sigma}\sqrt{n}h)'(L^{-1}\tilde{\Sigma}L^{-1})^{-1}(\xi^* + L^{-1}\tilde{\Sigma}\sqrt{n}h)}{2} \right\} d\xi^*
$$

$$
= \frac{1}{2\pi} \tilde{\Sigma}^*^{-\frac{1}{2}} \int_0^\infty \exp\left\{ -\frac{(\xi^* + L^{-1}\tilde{\Sigma}\sqrt{n}h)'\tilde{\Sigma}^*^{-1}(\xi^* + L^{-1}\tilde{\Sigma}\sqrt{n}h)}{2} \right\} d\xi^*
$$

$$
= U(\sqrt{n}(ah_1 - c\tilde{\rho}h_2), \sqrt{n}(a\tilde{\rho}h_1 - ch_2); \tilde{\rho})
$$

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

$$
\int_{B_0} e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t} dQ^+(t) = \int_{B_0} e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t} d\tilde{F}_{n,Q^+}(t) - \int_{B_0} e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t} d\Phi_{0,I_2,Q^+}(t).
$$

Then $\int_{B_0} e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t} d\tilde{F}_{n,Q^+}(t) = \int_{B_0 \cap A_1} e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t} d\tilde{F}_n(t)$. Since $e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t}$ is positive and $\int_{B_0} e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t} d\tilde{F}_n(t) < \infty$, we have $\int_{B_0 \cap A_1} e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t} d\tilde{F}_n(t) \leq \int_{B_0} e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t} d\tilde{F}_n(t) < \infty$, which means $\int_{B_0} e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t} d\tilde{F}_{n,Q^+}(t) < \infty$. Similarly, we can show $\int_{B_0} e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t} d\Phi_{0,I_2,Q^+}(t) < \infty$. Thus

$$
\int_{B_0} e^{-\sqrt{n\tilde{\Sigma}}\frac{1}{2}t} dQ^+(t) < \infty.
$$
Fix $\epsilon > 0$, there exist a finite $C > 0$ such that $\int_{B_0 \cap D_0} e^{-\sqrt{n}\Sigma_{2}^{1} t} dQ^{+}(t) < \epsilon/2$, where $D_0 = \{e^{-\sqrt{n}\Sigma_{2}^{1} t} > C\}$. Then
\[
\int_{B_0} e^{-\sqrt{n}\Sigma_{2}^{1} t} dQ^{+}(t) = \int_{B_0 \cap D_{0}^{c}} e^{-\sqrt{n}\Sigma_{2}^{1} t} dQ^{+}(t) + \int_{B_0 \cap D_{0}} e^{-\sqrt{n}\Sigma_{2}^{1} t} dQ^{+}(t) \leq CQ^{+}(B_0 \cap D_{0}^{c}) + \epsilon/2 = CQ(B_0 \cap D_{0}^{c} \cap A_1) + \epsilon/2. \quad (3.7.50)
\]

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 $\theta_{k} = \sum_{i=1}^{2} E|\bar{X}_i|^k$. 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 $\epsilon$ and from (3.7.50), we know $\int_{B_0} e^{-\sqrt{n}h^{1/2} t} dQ^{+}(t) = O(n^{-\frac{1}{2}})$. Similarly, we can also show that $\int_{B_0} e^{-\sqrt{n}h^{1/2} t} dQ^{-}(t) = O(n^{-\frac{1}{2}})$. Hence, $I_2 = O(n^{-\frac{1}{2}})$ 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. \qed

**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
\[
(i) \quad E\{X^2Y^2 I(X > \eta)I(Y > \lambda)\} = (1 + 2\rho^2)U(\lambda, \eta; \rho) + (\lambda^2\rho^2 + (1 + 2\rho^2)\lambda)Q(\lambda, \eta; \rho) + (\eta^2\rho^2 + (1 + 2\rho^2)\eta)Q(\eta, \lambda; \rho) + \sqrt{1 - \rho^2}(\rho\eta^2 + \lambda\eta + \rho\lambda^2 + 3\rho)q(\eta, \lambda; \rho); \\
(ii) \quad E\{X^2Y I(X > \eta)I(Y > \lambda)\} = (2 + \eta^2)\rho Q(\eta, \lambda; \rho) + (1 + \lambda^2\rho^2 + \rho^2)Q(\lambda, \eta; \rho) + \sqrt{1 - \rho^2}(\rho\eta q(\eta, \lambda; \rho) + \rho\lambda q(\lambda, \eta; \rho); \\
(iii) \quad E\{XY I(X > \eta)I(Y > \lambda)\} = \rho\eta Q(\eta, \lambda; \rho) + \rho\lambda Q(\lambda, \eta; \rho) + \sqrt{1 - \rho^2}q(\eta, \lambda; \rho); 
\]
(iv) \( E \{ XI(X > \eta)I(Y > \lambda) \} = Q(\eta, \lambda; \rho) + \rho Q(\lambda, \eta; \rho); \)

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

\[
E \{(X + \mu_1)^2(Y + \mu_2)^2I((X + \mu_1) > \eta)I((Y + \mu_2) > \lambda)\}
\]

\( = E \{X^2Y^2I(X > \eta'_1)I(Y > \lambda'_2)\} + 2\mu_2E \{X^2YI(X > \eta'_1)I(Y > \lambda'_2)\}
\]

\[+ 2\mu_1E \{XY^2I(X > \eta'_1)I(Y > \lambda'_2)\} + \mu_2^2E \{X^2I(X > \eta'_1)I(Y > \lambda'_2)\}
\]

\[+ \mu_1^2E \{Y^2I(X > \eta'_1)I(Y > \lambda'_2)\} + 4\mu_1\mu_2E \{XYI(X > \eta'_1)I(Y > \lambda'_2)\}.\]

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,

\[
E(T_n) = \sum_{i=1}^{p} E \{Y_{i,n}I(Y_{i,n} \geq \lambda_n)\} = \sum_{i=1}^{p} \int_{0}^{\infty} \int_{z \geq \lambda_n} dF(y)dz
\]

\[= \sum_{i=1}^{p} \lambda_nP(Y_{i,n} \geq \lambda_n) + \sum_{i=1}^{p} \int_{\lambda_n}^{\infty} P(Y_{i,n} \geq z)dz
\]

\[= \sum_{i=1}^{p} \lambda_nP(\sqrt{n}|\bar{X}_i| \geq \sqrt{n}\lambda_n) + 2\sum_{i=1}^{p} \int_{\sqrt{n}\lambda_n}^{\infty} zP(\sqrt{n}|\bar{X}_i| \geq z)dz. \tag{3.7.51}\]

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

\[
E(T_n) = \sum_{i=1}^{p} \lambda_n \{P(\sqrt{n}|\bar{X}_i - \mu_i| \geq \eta_{ni}) + P(\sqrt{n}|\bar{X}_i - \mu_i| \leq -\eta_{ni})\}
\]

\[+ 2\sum_{i=1}^{p} \int_{\sqrt{n}\lambda_n}^{\infty} z\{P(\sqrt{n}|\bar{X}_i - \mu_i| \geq z - \sqrt{n}\mu_i) + P(\sqrt{n}|\bar{X}_i - \mu_i| \leq -z - \sqrt{n}\mu_i)\}dz. \tag{3.7.52}\]

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

\[
P(\sqrt{n}|\bar{X}_i - \mu_i| \geq y) = \Phi(y) \left\{ \exp \left( \frac{k_3y^3}{6\sigma^3n^{1/2}} \right) \right\} \left[ 1 + o \left( \frac{1 + y}{n^{1/2}} \right) \right] \quad \text{and}
\]

\[
P(\sqrt{n}|\bar{X}_i - \mu_i| \leq -y) = \Phi(y) \left\{ \exp \left( -\frac{k_3y^3}{6\sigma^3n^{1/2}} \right) \right\} \left[ 1 + o \left( \frac{1 + y}{n^{1/2}} \right) \right]
\]

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))

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

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 $y > 0$,

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

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

$$P(\sqrt{n}(X_i - \mu_i) \geq \eta_{ni}^{-}) + P(\sqrt{n}(X_i - \mu_i) \leq -\eta_{ni}^{+}) = \{\Phi(\eta_{ni}^{-}) + \Phi(\eta_{ni}^{+})\} \{1 + o(1)\}$$

$$= \{(\sqrt{2\pi}(\eta_{ni}^{-}))^{-1} \exp(-\frac{1}{2}(\eta_{ni}^{-})^2) + (\sqrt{2\pi}(\eta_{ni}^{+}))^{-1} \exp(-\frac{1}{2}(\eta_{ni}^{+})^2)\} \{1 + o(1)\}. \quad (3.7.54)$$

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

$$P(\sqrt{n}(X_i - \mu_i) \geq \eta_{ni}^{-}) + P(\sqrt{n}(X_i - \mu_i) \leq -\eta_{ni}^{+})$$

$$= 1 - P(\sqrt{n}(X_i - \mu_i) \leq \eta_{ni}^{-}) + P(\sqrt{n}(X_i - \mu_i) \leq -\eta_{ni}^{+})$$

$$= \{1 - \Phi(-\eta_{ni}^{-}) + \Phi(\eta_{ni}^{+})\} \{1 + o(1)\}$$

$$= \{1 - (\sqrt{2\pi}(\eta_{ni}^{-}))^{-1} \exp(-\frac{1}{2}(\eta_{ni}^{-})^2) + (\sqrt{2\pi}(\eta_{ni}^{+}))^{-1} \exp(-\frac{1}{2}(\eta_{ni}^{+})^2)\} \{1 + o(1)\}. \quad (3.7.55)$$

Let $\epsilon > 0$ be an arbitrary small number and $\xi_n = O(n^{1/6-\epsilon})$ such that $\xi_n$ 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

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

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

and

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

$$+ \int_{C\xi_n}^{-\infty} zP(\sqrt{n}(X_i - \mu_i) \leq -z - \sqrt{n\mu_i})dz. \quad (3.7.59)$$
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 $\xi_n - \sqrt{n} \mu_i \to \infty$. Then the sum of (3.7.57) and (3.7.59) can be bounded by

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

$$\leq 2 \int_{C\xi_n}^{\infty} zP(\sqrt{n}|\bar{X}_i - \mu| \geq z - \sqrt{n}\mu_i)dz$$

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

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

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

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

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,

$$E(T_n) = \left\{ \sum_{i=1}^{p} \lambda_i \{ P(\sqrt{n}(\bar{X}_i - \mu_i) \geq \eta_{n_i}^+ - \eta_{n_i}^-) \} + 2 \sum_{i=1}^{p} \int_{C\xi_n}^{\infty} zP(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i)dz$$

$$+ P(\sqrt{n}(\bar{X}_i - \mu_i) \leq -z - \sqrt{n}\mu_i)dz \right\} \{1 + o(1)\}.$$

It follows that

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

$$+ \int_{\sqrt{\lambda_n}}^{C\xi_n} z\{ P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) - \Phi(z - \sqrt{n}\mu_i) \}dz$$  (3.7.65)
where (3.7.65) equals to

\[
| \int_{\sqrt{\lambda_n}}^{C\xi_n} z \Phi(z - \sqrt{n}\mu_i) P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) - \Phi(z - \sqrt{n}\mu_i) \, dz |
\]

\[
\leq \int_{\sqrt{\lambda_n}}^{C\xi_n} z \Phi(z - \sqrt{n}\mu_i) \frac{P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) - \Phi(z - \sqrt{n}\mu_i)}{\Phi(z - \sqrt{n}\mu_i)} \, dz
\]

\[
\leq \sup_{z = o(n^{1/6})} \left| \frac{P(\sqrt{n}(\bar{X}_i - \mu_i) \geq z - \sqrt{n}\mu_i) - \Phi(z - \sqrt{n}\mu_i)}{\Phi(z - \sqrt{n}\mu_i)} \int_{\sqrt{\lambda_n}}^{C\xi_n} z \Phi(z - \sqrt{n}\mu_i) \, dz \right|
\]

\[
= o(1) \times \int_{\sqrt{\lambda_n}}^{C\xi_n} z \Phi(z - \sqrt{n}\mu_i) \, dz.
\]

So, we have

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

Thus,

\[
E(T_n) = \left[ \sum_{i=1}^{p} \lambda_n \left( \Phi(n\mu_i) + \Phi(n\mu_i) \right) + 2 \sum_{i=1}^{p} \int_{\sqrt{\lambda_n}}^{C\xi_n} z \Phi(z - \sqrt{n}\mu_i) + \Phi(z + \sqrt{n}\mu_i) \, dz \right] \{1 + o(1)\}
\]

\[
= \left[ \sum_{i=1}^{p} \lambda_n \left( \Phi(n\mu_i) + \Phi(n\mu_i) \right) + 2 \sum_{i=1}^{p} \int_{\sqrt{\lambda_n}}^{\infty} z \Phi(z - \sqrt{n}\mu_i) + \Phi(z + \sqrt{n}\mu_i) \, dz \right] \{1 + o(1)\}.
\]

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

\[
E(n\bar{X}_i^2 I\{n\bar{X}_i^2 > \lambda_n\}) = \left\{ \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\lambda_n}}^{\infty} x^2 \exp\left(-\frac{(x - \sqrt{n}\mu_i)^2}{2}\right) dx \right\} \{1 + o(1)\}
\]

\[
+ \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{\lambda_n}}^{\infty} x^2 \exp\left(-\frac{(x - \sqrt{n}\mu_i)^2}{2}\right) dx \{1 + o(1)\}
\]

\[
= \left\{ (\eta_{ni}^+) \phi(\eta_{ni}^-) + (\eta_{ni}^-) \phi(\eta_{ni}^+) + (n\mu_i^2 + 1)(\Phi(\eta_{ni}^-) + \Phi(\eta_{ni}^+)) \right\} \{1 + o(1)\}.
\]

Hence the expectation of \( T_n \) is

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

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 

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

\[ E(n^2 \bar{X}_i^4 I\{n \bar{X}_i^2 > \lambda_n\}) = \lambda_n^2 \mathbb{P}(\sqrt{n} \bar{X}_i \geq \sqrt{\lambda_n}) + 4 \int_{\sqrt{\lambda_n}}^{\infty} z^3 \mathbb{P}(\sqrt{n} \bar{X}_i \geq z) dz. \]

Similar to the calculation of the expectation for \( T_n \), we want to show that for some \( \xi_n = o(n^{1/6}) \) and \( \xi_n \) is higher order than \( \sqrt{n} \mu_i \),

\[ \int_{\xi_n}^{\infty} z^3 \mathbb{P}(\sqrt{n} \bar{X}_i \geq z) dz = o\{\lambda_n^2 \mathbb{P}(\sqrt{n} \bar{X}_i \geq \sqrt{\lambda_n})\}. \] (3.7.68)

We have

\[ \int_{\xi_n}^{\infty} z^3 \mathbb{P}(\sqrt{n} (\bar{X}_i - \mu_i) \geq z - \sqrt{n} \mu_i) dz + \int_{\xi_n}^{\infty} z^3 \mathbb{P}(\sqrt{n} (\bar{X}_i - \mu_i) \leq z - \sqrt{n} \mu_i) dz. \]

Then followed by the exponential tail assumption,

\[ \int_{\xi_n}^{\infty} z^3 \mathbb{P}(\sqrt{n} (\bar{X}_i - \mu_i) \geq z - \sqrt{n} \mu_i) dz \leq \int_{\xi_n - \sqrt{n} \mu_i}^{\infty} (z + \sqrt{n} \mu_i)^3 \exp(-z^d/M) dz = \int_{\xi_n - \sqrt{n} \mu_i}^{\infty} (z^3 + 3z^2 \sqrt{n} \mu_i + 3zn \mu_i^2 + n^{3/2} \mu_i^3) \exp(-z^d/M) dz. \]

It is straightforward to show that

\[ \int_{\xi_n - \sqrt{n} \mu_i}^{\infty} z^a \exp(-z^d/M) dz = \frac{1}{d} \int_{\xi_n - \sqrt{n} \mu_i}^{\infty} u^{a+1/d-1} \exp(-u/M) du \leq (\xi_n - \sqrt{n} \mu_i)^{(a+1)(1-d)} M^{a+1} \int_{\xi_n - \sqrt{n} \mu_i}^{\infty} u^a \exp(-u/M) du = M^2 (\xi_n - \sqrt{n} \mu_i)^{(a+1-2d)} \exp(- (\xi_n - \sqrt{n} \mu_i)^{d/M}\{1 + o(1)\}). \]

From (3.7.54) and (3.7.55), we could take \( \xi_n \) to be \( (M \max(\lambda_n, n \mu_i^2))^{1/d} \) such that (3.7.68) holds. It follows that

\[ E(n^2 \bar{X}_i^4 I\{n \bar{X}_i^2 > \lambda_n\}) = \left\{ \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\lambda_n}}^{\infty} x^4 \exp(-\frac{(x - \sqrt{n} \mu_i)^2}{2}) dx + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{\lambda_n}} x^4 \exp(-\frac{(x - \sqrt{n} \mu_i)^2}{2}) dx \right\} \{1 + o(1)\}. \]
By integrate by parts, we have

\[
E(n^2 \bar{X}_i I\{n \bar{X}_i > \lambda\}) = \left\{ [(\eta_{ni}^-)^3 - 4\sqrt{n}\mu_i(\eta_{ni}^-) + 3(1 + 2n\mu_i^2)(\eta_{ni}^-) + 4\sqrt{n}\mu_i(2 + n\mu_i^2)]\phi(\eta_{ni}^-) + (\eta_{ni}^+)^3 - 4\sqrt{n}\mu_i(\eta_{ni}^+) + 3(1 + 2n\mu_i^2)(\eta_{ni}^+) - 4\sqrt{n}\mu_i(2 + n\mu_i^2)]\phi(\eta_{ni}^+) + [\mu^4_i + 6n\mu_i^2 + 3]\Phi(\eta_{ni}^-) + [\mu^4_i + 6n\mu_i^2 + 3]\Phi(\eta_{ni}^+) \right\} \{1 + o(1)\}
\]

Therefore

\[
V(\mu, \lambda) = \left\{ [(\sqrt{n}\mu)^3 + \sqrt{n}\mu\lambda + n\sqrt{\lambda}\mu^2 + 5\sqrt{n}\mu_i + 3\sqrt{\lambda}\phi(\eta_{ni}^-) + [(\sqrt{n}\mu)^3 - \sqrt{n}\mu\lambda + n\sqrt{\lambda}\mu^2 - 5\sqrt{n}\mu_i + 3\sqrt{\lambda}\phi(\eta_{ni}^+) + [\mu^4_i + 6n\mu_i^2 + 3]\Phi(\eta_{ni}^-) + \Phi(\eta_{ni}^+)
\]

- [(\eta_{ni}^+\phi(\eta_{ni}^-) + (\eta_{ni}^-)\phi(\eta_{ni}^+) + (\mu_i^2 + 1)(\Phi(\eta_{ni}^-) + \Phi(\eta_{ni}^+))] \right\} \{1 + o(1)\}. (3.7.69)
\]

Recall that \(Y_{i,n} = n\bar{X}_i^2\) and by Fubini’s Theorem, we have

\[
E\{Y_{1,n}Y_{2,n}I(Y_{1,n} > \lambda_n)I(Y_{2,n} > \lambda_n)\} = \lambda_n^2 P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n)
\]

+ \(\lambda_n \int_{\lambda_n}^{\infty} P(Y_{1,n} > z_1, Y_{2,n} > \lambda_n)dz_1\)

+ \(\lambda_n \int_{\lambda_n}^{\infty} P(Y_{1,n} > \lambda_n, Y_{2,n} > z_2)dz_2\)

+ \(\int_{\lambda_n}^{\infty} \int_{\lambda_n}^{\infty} P(Y_{1,n} > z_1, Y_{2,n} > z_2)dz_1dz_2.\) (3.7.70)

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 \(\sqrt{n}(\tilde{X}_{1,n}, \tilde{X}_{2,n})^T\) is a bivariate normally distributed
vector. Without loss of generality, suppose $\rho > 0$. Because

$$P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n) = P(\sqrt{n}(\bar{X}_{1,n} - \mu_1) > \eta_{n1}, \sqrt{n}(\bar{X}_{2,n} - \mu_2) > \eta_{n2})$$

$$+ P(\sqrt{n}(X^*_{1,n} + \mu_1) > \eta_{n1}^+, \sqrt{n}(\bar{X}_{2,n} - \mu_2) > \eta_{n2})$$

$$+ P(\sqrt{n}(X_{1,n} - \mu_1) > \eta_{n1}, \sqrt{n}(X^*_{2,n} + \mu_2) > \eta_{n2}^+)$$

$$+ P(\sqrt{n}(X^*_{1,n} + \mu_1) > \eta_{n1}^+, \sqrt{n}(X^*_{2,n} + \mu_2) > \eta_{n2}^+)$$

and we observe that

$$P(\sqrt{n}(\bar{X}_{1,n} - \mu_1) > \eta_{n1}, \sqrt{n}(\bar{X}_{2,n} - \mu_2) > \eta_{n2}) \geq P(\sqrt{n}(\bar{X}_{1,n} - \mu_1) > \eta_{n1}^+, \sqrt{n}(\bar{X}_{2,n} - \mu_2) > \eta_{n2})$$

$$\geq \Phi(\eta_{n1}^+) \Phi\left(-\rho\eta_{n1}^+ + \eta_{n2}^-\right) \geq \Phi(\eta_{n1}^+) \Phi\left(\frac{\rho\eta_{n1}^+ + \eta_{n2}^-}{\sqrt{1-\rho^2}}\right)$$

$$\geq P(\sqrt{n}(X^*_{1,n} + \mu_1) > \eta_{n1}^+, \sqrt{n}(\bar{X}_{2,n} - \mu_2) > \eta_{n2})$$

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

$$P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n) \geq \Phi(\eta_{n2}^-) \Phi\left(-\rho\eta_{n2}^- + \eta_{n1}^-\right).$$

When $\eta_{n1} > 0$ such that $\eta_{n1}^- - \rho\eta_{n2} > 0$,

$$P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n) \geq \frac{\sqrt{1-\rho^2}}{\eta_{n2}(\eta_{n1}^- - \rho\eta_{n2})} \exp\left\{ -\frac{\eta_{n2}^2 + \eta_{n2}^2 - 2\rho\eta_{n2}\eta_{n1}^-}{2(1-\rho^2)} \right\}$$

$$\sim C_1\lambda_n^{-1} \exp\{-C_2\lambda_n\}.$$ 

Otherwise, $P(Y_{1,n} > \lambda_n, Y_{2,n} > \lambda_n) \geq \Phi(\eta_{n2}^-) \left(1 - \Phi\left(\frac{\rho\eta_{n2}^- - \eta_{n1}^-}{\sqrt{1-\rho^2}}\right)\right) \sim C_3\lambda_n^{-1/2} \exp\{-C_4\sqrt{\lambda_n}\}.$

Let $\xi_n = o(n^{1/6})$ such that $\xi_n$ is higher order than $\sqrt{n}\mu_i$

$$\lambda_n \int_{\xi_n}^\infty P(Y_{1,n} > z_1, Y_{2,n} > \lambda_n)dz_1$$

$$\leq 2\lambda_n \int_{\xi_n}^{\infty} z_1 P(\sqrt{n}|X_{1,n} - \mu_1| > z_1 - \sqrt{n}\mu_1, \sqrt{n}|X_{2,n} - \mu_2| > \eta_{n2})dz_1$$

$$\leq 2\lambda_n \int_{\xi_n}^{\infty} z_1 P(\sqrt{n}|X_{1,n} - \mu_1| > z_1 - \sqrt{n}\mu_1)dz_1.$$
Thus, like (3.7.63), we can find $\xi_n$ 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

$$\int_{\xi_n}^{\infty} \int_{\xi_n}^{\infty} P(Y_{1,n} > z_1, Y_{2,n} > z_2)dz_1dz_2$$

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

$$E\{Y_{1,n}Y_{2,n}I(Y_{1,n} > \lambda_n)I(Y_{2,n} > \lambda_n)\} = E\{Y_{1,n}Y_{2,n}I(\sqrt{n}X_{1,n} > \sqrt{\lambda_n})I(\sqrt{n}X_{2,n} > \sqrt{\lambda_n})\}$$

$$+ E\{Y_{1,n}Y_{2,n}^*I(\sqrt{n}X_{1,n} > \sqrt{\lambda_n})I(\sqrt{n}X_{2,n}^* > \sqrt{\lambda_n})\}$$

$$+ E\{Y_{1,n}^*Y_{2,n}I(\sqrt{n}X_{1,n}^* > \sqrt{\lambda_n})I(\sqrt{n}X_{2,n} > \sqrt{\lambda_n})\}$$

$$+ E\{Y_{1,n}^*Y_{2,n}^*I(\sqrt{n}X_{1,n}^* > \sqrt{\lambda_n})I(\sqrt{n}X_{2,n}^* > \sqrt{\lambda_n})\}.$$ 

Applying the formula given in Lemma 2, we can approximate above expectations by (3.7.24).

It follows that

$$\gamma_1(\mu_1, \mu_2, \lambda_n) := Cov\{Y_{1,n}I(Y_{1,n} > \lambda_n), Y_{2,n}I(Y_{2,n} > \lambda_n)\}$$

$$= E\{Y_{1,n}Y_{2,n}I(Y_{1,n} > \lambda_n)I(Y_{2,n} > \lambda_n)\}$$

$$- \left\{ (\eta_{n1}^+ \phi(\eta_{n1}^+) + (\eta_{n1}^-) \phi(\eta_{n1}^+) + (n\mu_1^2 + 1)(\Phi(\eta_{n1}^-) + \Phi(\eta_{n1}^+)) \right\}$$

$$\times \left\{ (\eta_{n2}^+ \phi(\eta_{n2}^+) + (\eta_{n2}^-) \phi(\eta_{n2}^+) + (n\mu_2^2 + 1)(\Phi(\eta_{n2}^-) + \Phi(\eta_{n2}^+)) \right\}. \quad (3.7.74)$$

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. \hfill \Box

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

$$\sum_{i=1}^{\infty} t^{-1} \alpha(i)^{\delta/(2r+\delta)} < \infty,$$
then
\[ E\left( \sum_{i=1}^{p} Z_i \right)^{2r} \leq C p^r \left[ M_{2r}^{2r} + M_{2r+\delta}^{2r} \sum_{i=1}^{\infty} i^{r-1} \alpha(i) \delta/(2r+\delta) \right] \] (3.7.75)

where \( C \) is a finite constant only depending on \( r \).

**Proof of Theorem 2** (i) Let \( k' \to \infty, k'/k \to 0 \) and \( k/p \to 0 \) as \( n \to \infty \). 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
\[ \tilde{\zeta}_j^r = \sigma_0^{-1}(p; \lambda_n) \sum_{i=1}^{k'} \tilde{T}_{(j-1)b+k+i,n} \]
for \( j = 1, \cdots, r \). Also the residual block \( \delta_n = \sigma_0^{-1}(p; \lambda_n) \sum_{i=rb+1}^{p} \tilde{T}_{i,n} \). Then
\[ p^{-1/2} S_n = p^{-1/2} \sum_{j=1}^{r} \tilde{\zeta}_{j,n} + p^{-1/2} \sum_{j=1}^{r} \tilde{\zeta}_j^r + p^{-1/2} \delta_n := S_{n,1} + S_{n,2} + S_{n,3}. \]

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 \( q > 2 \)
\[ |\rho_k| \leq \frac{2q}{q-2} \left( 2\alpha_X(k) \right)^{1-2/q} ||X_{ij}||_q^2 = C \alpha_X^{1-2/q}(k) \]
and from (3.7.37), \( |\gamma_k(0,0,\lambda_n)| = C|\rho_k|L_p p^{-\frac{2\delta}{1+|\rho_k|}} \leq C \alpha_X^{1-2/q}(k)p^{-\delta}. \) 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 \( \sum_{k=1}^{\infty} |\gamma_k(0,0,\lambda_n)| = O(p^{-\delta}). \)

Because for \( j \geq 1 \),
\[ |\text{Cov}(\tilde{\zeta}_{1,n}, \tilde{\zeta}_{j+1,n})| = \left| \sum_{i=1}^{k'} \sum_{i'=1}^{k'} \text{Cov}(Z_{k+i,n}(\lambda_n), Z_{j+b+k+i',n}(\lambda_n)) \right| \]
\[ = \left| \sum_{i=1}^{k'} \sum_{i'=1}^{k'} \gamma_{j+b+\delta(i-i')}(0,0,\lambda_n) \right| \leq \sum_{i=1}^{k'} \left| \gamma_{j+b+\delta(i-k')} (0,0,\lambda_n) \right|. \]
we have
\[
\text{Var}(S_{n,2}) = p^{-1} \sigma_0^{-2}(p; \lambda_n) \left\{ r k' \sigma_0^2(k'; \lambda_n) + 2 \sum_{j=1}^{r-1} (r-j) \text{Cov}(\tilde{\zeta}_{j,n}', \tilde{\zeta}_{j+1,n}') \right\} \\
\leq p^{-1} \sigma_0^{-2}(p; \lambda_n) \left\{ r k' \sigma_0^2(k'; \lambda_n) + 2 \sum_{j=1}^{r-1} (r-j) \sum_{i=1}^{k'} |\gamma_{j+(i-k')(0,0,\lambda_n)}| \right\} \\
\leq p^{-1} \sigma_0^{-2}(p; \lambda_n) \left\{ r k' \sigma_0^2(k'; \lambda_n) + 2r \sum_{k=1}^{\infty} |\gamma_k(0,0,\lambda_n)| \right\} = O(rk'/p) \rightarrow 0,
\]
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,
\[
p^{-1/2} S_n = p^{-1/2} \sum_{j=1}^{r} \tilde{\zeta}_{j,n} + o_p(1). \tag{3.7.76}
\]
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 \( \epsilon > 0 \),
\[
P(\tilde{\zeta}_{j,n} - W_{j,n} | \leq \epsilon p^{1/2}/r) \leq 11(\epsilon p^{1/2}/r)^{-2/5}(E(\tilde{\zeta}^2))^{1/5} \alpha_5^{4/5}(k') \leq C \epsilon^{-2/5} r^{1/5} \alpha_5^{4/5}(k').
\]
Let \( \Delta_n = S_{n,1} - p^{-1/2} \sum_{j=1}^{r} W_{j,n} \). Then
\[
P(|\Delta_n| > \epsilon) \leq \sum_{j=1}^{r} P(|\tilde{\zeta}_{j,n} - W_{j,n}| \leq \epsilon p^{1/2}/r) \leq C_1 \epsilon^{-2/5} r^{6/5} \alpha_5^{4/5}(k').
\]
Choosing \( r = p^a \) for \( a \in (0,1) \), \( b = p^{1-a} \) and \( k' = p^c \) for \( c \in (0,1-a) \). Because
\[
p^{6a/5} \alpha_5^{4/5}(p^c) \leq p^{6a/5} n^{4/5} \alpha_5^{4/5}(p^c) \rightarrow 0, \quad \text{as } p \rightarrow \infty,
\]
then \( \Delta_n = o_p(1) \). It follows that
\[
p^{-1/2} S_n = p^{-1/2} \sum_{j=1}^{r} W_{j,n} + o_p(1)
\]
It suffices to show the Lyapounov’s condition, which is

$$\lim_{r \to \infty} r^{-2} \sum_{j=1}^{r} E|k^{-1/2}W_{j,n}^*|^4 = 0. \quad (3.7.77)$$

By Lemma 3, we can get that

$$E|k^{-1/2}W_{j,n}^*|^4 = k^{-2}\sigma_0^{-4}(k; \lambda_n)E\left\{\sum_{i=1}^{k}(Z_{i,n}(\lambda_n) - \mu_{T_n,0}^{(i)})\right\}^4 \leq C\sigma_0^{-4}(k; \lambda_n)\left[M_4^4 + M_{4+\delta}^4 \sum_{i=1}^{\infty} i\alpha Z(i)^{\delta/(4+\delta)}\right].$$

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 $M_{4+\delta}^4 = O((\log p)^{4+14} p^{-4+4s/4+\delta})$. Because

$$\sum_{i=1}^{\infty} i\alpha X(i)^{\delta/(4+\delta)} \leq n^{\delta/(4+\delta)} \sum_{i=1}^{\infty} i\alpha_X(i)^{\delta/(4+\delta)} \leq Cn^{\delta/(4+\delta)},$$

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 $n^{\delta/(4+\delta)} p^{(4+2\delta)s/(4+\delta)} - a \to 0$, 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 $\tilde{\zeta}_{j,n}, \tilde{\gamma}_{j,n}$ and $\rho_n$ 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 $\mu_1$ and $\mu_{k+1}$ satisfy the conditions in Theorem 1,

$$\left|\gamma_k(\mu_1, \mu_{k+1}, \lambda_n)\right| \leq CL_p |\rho_k| \leq C L_p \alpha_{X_1}^{1/2}(k). \quad (3.7.78)$$

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

$$\tilde{W}_{j,n} \overset{d}{=} \tilde{\gamma}_{j,n} := \sigma_1^{-1}(p; \lambda_n) \sum_{i=1}^{k}(Z_{i,n}(\lambda_n) - \mu_{T_n,1}^{(i)})$$

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:

\[
\left\{ \sum_{j=1}^{r} E(\tilde{W}_{j,n}^2) \right\}^{-2} \sum_{j=1}^{r} E(\tilde{W}_{j,n}^4) \to 0.
\] (3.7.79)

Note that by condition (3.3.13),

\[
\left| \frac{1}{p} \sum_{j=1}^{r} E(\tilde{W}_{j,n}^2) - 1 \right| = \left| \frac{1}{p} \sum_{j=1}^{r} E(\tilde{\xi}_{j,n}^2) - 1 \right| \leq \frac{1}{p} \sum_{j=1}^{r} |E(\tilde{\xi}_{j,n}^2) - k| + p^{-1}|rk - p|
\]

\[
\leq rkH_k/p + p^{-1}|rk - p| \to 0.
\]

Now we will use Lemma 3 to show that

\[
\frac{1}{p^2} \sum_{j=1}^{r} E(\tilde{W}_{j,n}^4) = \frac{(rk)^2}{p^2} \frac{1}{r^2} \sum_{j=1}^{r} E((k^{-1/2}\tilde{W}_{j,n})^4) \to 0.
\] (3.7.80)

Similar to part (i),

\[
E|k^{-1/2}\tilde{W}_{j,n}|^4 = k^{-2}\sigma_1^{-4}(k; \lambda_n)E \left\{ \sum_{i=1}^{k} (Z_{i,n}(\lambda_n) - \mu^{(i)}_{T_{n,1}}) \right\}^4
\]

\[
\leq C\sigma_1^{-4}(k; \lambda_n) \left[ M_4^4 + M_{4+\delta}^4 \sum_{i=1}^{\infty} i\alpha Z(i)^{\delta/(4+\delta)} \right].
\]

where \( M_4^4 = \max_i E(Z_{i,n}(\lambda_n) - \mu^{(i)}_{T_{n,1}})^4 \) and \( M_{4+\delta}^4 = \max_i E(Z_{i,n}(\lambda_n) - \mu^{(i)}_{T_{n,1}})^{4+\delta} i^{\delta/2} \).

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

\[
E\{Z_{i,n}(\lambda_n) - \mu^{(i)}_{T_{n,1}}\}^d = \sum_{i=0}^{2d} \binom{2d}{i} (2r_i \log p)^i \frac{1}{\sqrt{2\pi}} \int_{2(\sqrt{s} - \sqrt{r_i})\sqrt{\log p}} y^{2d-i} \exp(-y^2/2)dy.
\] (3.7.81)

When \( i \) is an even number,

\[
\int_{2(\sqrt{s} - \sqrt{r_i})\sqrt{\log p}} y^{2d-i} \exp(-y^2/2)dy = \{2(\sqrt{s} - \sqrt{r_i})\sqrt{\log p}\}^{2d-i-1} \phi(2(\sqrt{s} - \sqrt{r_i})\sqrt{\log p}) \{1 + o(1)\}
\]

\[
+ \Phi(2(\sqrt{s} - \sqrt{r_i})\sqrt{\log p}).
\]

When \( i \) is an odd number,

\[
\int_{2(\sqrt{s} - \sqrt{r_i})\sqrt{\log p}} y^{2d-i} \exp(-y^2/2)dy = \{2(\sqrt{s} - \sqrt{r_i})\sqrt{\log p}\}^{2d-i-1} \phi(2(\sqrt{s} - \sqrt{r_i})\sqrt{\log p}) \{1 + o(1)\}.
\]
Therefore, \( E\{Z_{i,n}(\lambda_n) - \mu^{(i)}_{T_{n,0}}\}^d = L_p^{(1)} I(r_i > s) + L_p^{(2)} p^{-(\sqrt{s} - \sqrt{r})^2} I(r_i < s) \). Then it can be seen that \( M^4_p = L_p^{(1)} I(\max_i r_i > s) + L_p^{(2)} p^{-(\sqrt{s} - \max_i \sqrt{r_i})^2} I(\max_i r_i < s) \). Suppose \( \sigma_1(p; \lambda_n) \sim L_p p^{-h_1} \). If \( \max_i r_i > s \), then \( E|k^{-1/2} \tilde{W}_{j,n}|^4 \leq L_p p^{\delta/(4+\delta)} p^{2h_1} \). Now if we take \( r = p^a \) with \( a > 2h_1 \) such that \( n^{\delta/(4+\delta)} p^{2h_1-a} \to 0 \), then the Lyapunov’s condition holds.

If \( \max_i r_i < s \), then \( E|k^{-1/2} \tilde{W}_{j,n}|^4 \leq L_p p^{\delta/(4+\delta)} p^{2h_1 - \frac{4s^*}{4+s}} \) where \( s^* = (\sqrt{s} - \max_i \sqrt{r_i})^2 \). In this case, if we take \( a > 2h_1 - \frac{4s^*}{4+s} \) such that \( n^{\delta/(4+\delta)} p^{2h_1 - \frac{4s^*}{4+s} - a} \to 0 \), the Lyapunov’s condition holds. \( \square \)

**Proof of Theorem 3** Let \( \Delta_n = (\lambda_{n1}, \cdots, \lambda_{nd})' = 2 \log(p)(s_1, \cdots, s_d)' \). We want to show that \((T_S(\lambda_{n1}), \cdots, T_S(\lambda_{nd}))\) are multivariate normally distributed. Applying the Cramér-wold device, we only need to show that for any \( c_k \) such that \( \sum_k c_k^2 = 1 \), \( T_{S^*} := \sum_{k=1}^d c_k T_S(\lambda_{nk}) \) is normally distributed. Let

\[
T_{S^*} = \sum_{k=1}^d c_k \frac{Z_{i,n}(\lambda_{nk}) - \mu^{(i)}_{T_{n,0}}(\lambda_{nk})}{\sigma_0(p; \lambda_{nk})}.
\]

By the definition of \( T_S(\lambda_{nk}) \), we have \( T_{S^*} = \frac{1}{\sqrt{p}} \sum_{i=1}^p T_{S^* i,n} \). Let \( k^* / p \to 0 \) as \( n \to \infty \). As in the proof of Theorem 2, define \( \tilde{\zeta}_{j,n} = \sum_{i=1}^k T_{S^* (j-1)b+i,n} \), \( \tilde{\zeta}'_{j,n} = \sum_{i=1}^{k'} T_{S^* (j-1)b+i,n} \) and \( \delta_n^* \) 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 \( s_k > s_{k'} \),

\[
\tilde{r}_k (\sqrt{\lambda_{nk}}, \sqrt{\lambda_{nk'}}) - G_{T_n}(\lambda_{nk}) G_{T_n}(\lambda_{nk'}) = L_p \rho_j p^{-\frac{s_k - s_{k'}}{2}} I(s_{k'} \leq \rho_j^2 s_k) + \frac{L_p \rho_j p^{-\frac{s_k - s_{k'}}{2}}}{1-\rho_j^2} (\sqrt{\lambda_{nk}} - \sqrt{\lambda_{nk'}})^2 I(s_{k'} > \rho_j^2 s_k).
\]

and \( 2[(\sqrt{\lambda_{nk}})^3 + 3\sqrt{\lambda_{nk}}] \phi(\sqrt{\lambda_{nk}}) = L_p p^{-\frac{s_k - s_{k'}}{2}} \sigma_0(p; \lambda_{nk}) \sigma_0(p; \lambda_{nk'}) \) where \( \lambda_{nk}^* \) is max \( \lambda_{nk}, \lambda_{nk'} \). Thus, if \( \sum_k |\rho_k| < \infty \) and \( \lambda_{nk} \neq \lambda_{nk'} \), \( \text{Cov}(T_S(\lambda_{nk}), T_S(\lambda_{nk'})) = o(1) \). It follows that \( \text{Var}(T_{S^*}) = (\sum_{k=1}^d c_k^2)^2 \{1 + o(1)\} = 1 + o(1) \). We can also show that \( \text{Var}(\tilde{\zeta}'_{j,n}) = k^* \{1 + o(1)\} \) and

\[
\text{Cov}(T_{S^* i,n}, T_{S^* j+1,n}) = \sum_{k=1}^d \sum_{k'=1}^d c_k c_k' \left\{ L_p \rho_j p^{-\frac{s_k - s_{k'}}{2}} I(s_{k'} \leq \rho_j^2 s_k) + L_p \rho_j p^{-\frac{s_k - s_{k'}}{2}} (\sqrt{\lambda_{nk}} - \sqrt{\lambda_{nk'}})^2 I(s_{k'} > \rho_j^2 s_k) \right\}.
\]
By the same arguments in the proof of Theorem 2, we obtain
\[
\text{Var}(S_{n,2}) \leq p^{-1}\{r\text{Var}(\tilde{\zeta}_{j,n}) + 2r \sum_{k=1}^{\infty} \text{Cov}(TS_{1,n}^*, TS_{j+1,n}^*)}\} = O(rk^*/p) \to 0.
\]
Similarly, we could show that \( \text{Var}(S_{n,3}) = o(1) \). Therefore, \( TS^* = S_{n,1} + o_p(1) \). Following the proof of Theorem 2, we only need to verify the Lyapounov’s condition:
\[
\lim_{r \to \infty} r^{-2} \sum_{j=1}^{r} E|k^*-\frac{1}{2}\tilde{\zeta}_{j,n}^*|^4 = 0.
\]
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
\[
M_4^4 = E(TS_{i,n}^*)^4 \leq \sum_{k=1}^{d} d E \left( c_k \frac{Z_{i,n}(\lambda_{nk}) - \mu_{T,0}(\lambda_{nk})}{\sigma_0(p, \lambda_{nk})} \right)^4 = L_p d \sum_{k=1}^{d} c_k^4 p s_k = O(L_p^p \max s_k).
\]
and similarly, we have \( M_{4+\delta} = O(L_p^p (4+2\delta) \max s_k/(4+\delta)) \). It follows that
\[
E|k^*\tilde{\zeta}_{j,n}^*|^4 \leq L_p n^{\delta/(4+\delta)} p^{(4+2\delta) \max s_k/(4+\delta)}.\]

If we take \( r = p^a \) with \( a > (4+2\delta) \max s_k/(4+\delta) \) such that \( n^{\delta/(4+\delta)} p^{(4+2\delta) \max s_k/(4+\delta) - a} \to 0 \), then the Lyapounov’s condition holds. If all \( TS(\lambda_{nk}) \) are normally distributed, by Theorem 2(i), the above condition holds. Thus, \( TS^* \) is normally distributed with mean 0 and variance 1. Hence, the \( (TS(\lambda_{n1}), \cdots, TS(\lambda_{nd})) \) is normally distributed with mean 0 and covariance \( \Omega(\lambda_{ns}, \lambda_{nt})_{st} \).

\[\square\]
References


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.


CHAPTER 4. ANOVA for Longitudinal Data with Missing Values


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 trials 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:

$$\{(Y_{1i}(t), X_{1i}(t))\}_{i=1}^{n_1}, \ldots, \{(Y_{ki}(t), X_{ki}(t))\}_{i=1}^{n_k}$$

where the response variable $Y_{ji}(t)$ and the covariate $X_{ji}(t)$ are supposed to be measured at time points $t = t_{j1}, \ldots, t_{jT_j}$. Here $T_j$ is the fixed number of scheduled observations for the $j$-th treatment. However, $\{Y_{ji}(t), X_{ji}(t)\}$ 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

$$Y_{ji}(t) = X_{ji}(t)^T \beta_j + M^T(X_{ji}(t), t) \gamma_j + g_j(t) + \epsilon_{ji}(t), \quad j = 1, 2, \ldots, k \tag{4.1.1}$$

where $M(X_{ji}(t), t)$ are known functions of $X_{ji}(t)$ and time $t$ representing interactions between the covariates and the time, $\beta_j$ and $\gamma_j$ are $p$- and $q$-dimensional parameters respectively, $g_j(t)$ are unknown smooth functions representing the time effect, and $\{\epsilon_{ji}(t)\}$ 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 $\beta_{j0}$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

\[
Y_{ji}(t_{jim}) = X^\tau_{ji}(t_{jim})\beta_{j0} + M^\tau(X_{ji}(t_{jim}), t_{jim})\gamma_{j0} + g_{j0}(t_{jim}) + \varepsilon_{ji}(t_{jim}),
\]

for \( j = 1, \cdots, k \), \( i = 1, \cdots, n_j \), \( m = 1, \ldots, T_j \). Here \( \beta_{j0} \) and \( \gamma_{j0} \) 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 \( (Y_{jim}, X^\tau_{jim}, M^\tau_{jim}) \) to denote \( (Y_{ji}(t_{jim}), X^\tau_{ji}(t_{jim}), M^\tau(X_{ji}(t_{jim}), t_{jim})) \). Also, we will use \( X^\tau_{jim} = (X^\tau_{jim}, M^\tau_{jim}) \) and \( \xi^\tau_j = (\beta^\tau_j, \gamma^\tau_j) \). For each individual, the residuals \( \{\varepsilon_{ji}(t)\} \) satisfy

\[
E\{\varepsilon_{ji}(t)|X_{ji}(t)\} = 0, \quad \text{Var}\{\varepsilon_{ji}(t)|X_{ji}(t)\} = \sigma_j^2(t) \quad \text{and} \quad \text{Cov}\{\varepsilon_{ji}(t), \varepsilon_{ji}(s)|X_{ji}(t), X_{ji}(s)\} = \rho_j(s, t)\sigma_j(t)\sigma_j(s)
\]

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 \( \beta_{j0}, \gamma_{j0} \) and \( g_{j0}(t) \), we assume

\[
(\beta_{j0}, \gamma_{j0}, g_{j0}) = \arg \min_{(\beta_j, \gamma_j, g_j)} \frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E\{Y_{jim} - X_{jim}^T \beta_j - M_{jim} \gamma_j - g_j(t_{jim})\}^2.
\]

We also require that \( \frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E(\tilde{X}_{jim} \tilde{X}_{jim}^T) > 0 \), where \( \tilde{X}_{jim} = X_{jim} - E(X_{jim}|t_{jim}) \).

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 \( X_{jim} \):

\[
X_{jim} = h_j(t_{jim}) + u_{jim}, \quad j = 1, 2, \ldots, k, \quad i = 1, \ldots, n_j, \quad m = 1, \ldots, T_j,
\]

where \( h_j(\cdot) \)s are \( p \)-dimensional smooth functions with continuous second derivatives, the residual \( u_{jim} = (u_{jim}^1, \ldots, u_{jim}^p)^T \) satisfy \( E(u_{jim}) = 0 \) and \( u_{jl} \) and \( u_{jk} \) are independent for \( l \neq k \), where \( u_{jl} = (u_{jl1}, \ldots, u_{jlT_j}) \). 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:

\[
H_{0a}: \beta_{10} = \beta_{20} = \ldots = \beta_{k0} \quad \text{vs} \quad H_{1a}: \beta_{i0} \neq \beta_{j0} \text{ for some } i \neq j.
\]

The second one is regarding the time effect functions:

\[
H_{0b}: g_{10}(\cdot) = \ldots = g_{k0}(\cdot) \quad \text{vs} \quad H_{1b}: g_{i0}(\cdot) \neq g_{j0}(\cdot) \text{ for some } i \neq j.
\]

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

\[
H_{0d}: \gamma_{10} = \gamma_{20} = \ldots = \gamma_{k0} \quad \text{vs} \quad H_{1d}: \gamma_{i0} \neq \gamma_{j0} \text{ for some } i \neq j.
\]

Let \( X_{ji} = \{X_{jio}, \ldots, X_{jiT_j}\} \) and \( Y_{ji} = \{Y_{jio}, \ldots, Y_{jiT_j}\} \) 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
$\tilde{Y}_{ji}(t-d) = \{Y_{ji(t-d)}, \ldots, Y_{ji(t-1)}\}$ and $\tilde{X}_{ji}(t-d) = \{X_{ji(t-d)}, \ldots, X_{ji(t-1)}\}$ 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 $\delta_{ji(t-1)} = 1$ if $(X_{ji(t-1)}, Y_{ji})$ is observed and $\delta_{ji(t-1)} = 0$ if $(X_{ji(t-1)}, Y_{ji})$ is missing. Here, we assume $X_{ji(t)}$ and $Y_{ji(t)}$ are either both observed or both missing. This simultaneous missingness of $X_{ji(t)}$ and $Y_{ji(t)}$ 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 $\delta_{ji(t)} = 0$. 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 $\delta_{ji(t-d)} = \prod_{l=1}^{d} \delta_{ji(t-l)}$. We assume the missingness of $(X_{ji(t)}, Y_{ji(t)})$ is missing at random (MAR) (Rubin, 1976) given its immediate past $d$ complete observations, namely

$$P(\delta_{ji(t)} = 1 | \delta_{ji(t-d)} = 1, X_{ji}, Y_{ji}) = P(\delta_{ji(t)} = 1 | \delta_{ji(t-d)} = 1, \tilde{X}_{ji(t-d)}, \tilde{Y}_{ji(t-d)}) = p_{ji}(\tilde{X}_{ji(t-d)}, \tilde{Y}_{ji(t-d)}; \theta_{j0}).$$

(4.2.4)

Here the missing propensity $p_{ji}$ is known up to a parameter $\theta_{j0}$. To allow derivation of a binary likelihood function, we need to set $\delta_{ji(t)} = 0$ if $\delta_{ji(t-d)} = 0$ when there is some drop-outs among the past $d$ visits, which is only temporarily if $\delta_{ji(t)} = 1$. This set-up ensures

$$P(\delta_{ji(t)} = 0 | \delta_{ji(t-d)} = 0, \tilde{X}_{ji(t-d)}, \tilde{Y}_{ji(t-d)}) = 1.$$  

(4.2.5)

Now the conditional binary likelihood for $\{\delta_{ji(t)}\}_{t=1}^{T_j}$ given $X_{ji}$ and $Y_{ji}$ is

$$P(\delta_{j0}, \ldots, \delta_{ji(T_j)} | X_{ji}, Y_{ji}) = \prod_{m=1}^{T_j} P(\delta_{jim} | \delta_{jim(m-1)}, \ldots, \delta_{j0}, X_{ji}, Y_{ji})$$

$$= \prod_{m=1}^{T_j} P(\delta_{jim} | \delta_{jim,d} = 1, \tilde{X}_{jim,d}, \tilde{Y}_{jim,d})$$

$$= \prod_{m=1}^{T_j} [p_{ji}(\tilde{X}_{jim,d}, \tilde{Y}_{jim,d}; \theta_{ji})^{\delta_{jim}} (1 - p_{ji}(\tilde{X}_{jim,d}, \tilde{Y}_{jim,d}; \theta_{ji}))^{(1-\delta_{jim})}]^{\delta_{jim,d}}.$$
In the second equation above we use both the MAR in (4.2.4) and (4.2.5). Hence, the parameters \( \theta_{j0} \) can be estimated by maximizing the binary likelihood

\[
\mathcal{L}_{B_j}(\theta_j) = \prod_{i=1}^{n_j} \prod_{t=1}^{T_j} \left[ p_j(\tilde{X}_{jit,d}, \tilde{Y}_{jit,d}; \theta_j)^{\delta_{jit}} \left( 1 - p_j(\tilde{X}_{jit,d}, \tilde{Y}_{jit,d}; \theta_j) \right)^{(1-\delta_{jit})} \right]^{\delta_{jit,d}}. \tag{4.2.6}
\]

Under some regular conditions, the binary maximum likelihood estimator \( \hat{\theta}_j \) is \( \sqrt{n} \)-consistent estimator of \( \theta_{j0} \); 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 \( H_{0a} : \beta_{10} = \beta_{20} = \ldots = \beta_{k0} \) with respect to the covariates. Let \( \pi_{jim}(\theta_j) = \prod_{l=m-d}^{m} p_j(\tilde{X}_{jil,d}, \tilde{Y}_{jil,d}; \theta_j) \) 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 \( \beta_{j0} \) and \( \gamma_{j0} \) were known, \( g_{j0}(t) \) would be estimated by

\[
\hat{g}_j(t; \beta_{j0}) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t)(Y_{jim} - X_{jim}^T \beta_{j0} - M_{jim}^T \gamma_{j0}), \tag{4.3.7}
\]

where

\[
w_{jim,h_j}(t) = \frac{(\delta_{jim}/\pi_{jim}(\hat{\theta}_j))K_{h_j}(t_{jim} - t)}{\sum_{s=1}^{n_j} \sum_{l=1}^{T_j} (\delta_{jls}/\pi_{jls}(\hat{\theta}_j))K_{h_j}(t_{jls} - t)} \tag{4.3.8}
\]

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 \( \pi_{jst}(\hat{\theta}_j) \) may be inconsistent if the missingness depends on the responses \( Y_{jit} \), which can be the case for missing covariates.

Let \( A_{jim} \) denote any of \( X_{jim}, Y_{jim} \) and \( M_{jim} \) and define

\[
\tilde{A}_{jim} = A_{jim} - \sum_{t_{1}=1}^{n_j} \sum_{m_{1}=1}^{T_j} w_{jim_{1},h_j}(t_{jim}) A_{jim_{1}m_{1}} \tag{4.3.9}
\]
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

$$Z_{ji}(\beta_j) = \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{X}_{jim}(\tilde{Y}_{jim} - \tilde{X}_{jim}'\beta_j - \tilde{M}_{jim}'\tilde{\gamma}_j),$$

where $\tilde{\gamma}_j$ is the solution of

$$\sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{M}_{jim}(\tilde{Y}_{jim} - \tilde{X}_{jim}'\beta_j - \tilde{M}_{jim}'\tilde{\gamma}_j) = 0$$

at given $\beta_j$. Note that $E\{Z_{ji}(\beta_j_0)\} = o(1)$. Although it is not exactly zero, $Z_{ji}(\beta_j_0)$ can still be used as an approximate zero mean estimating function to formulate an empirical likelihood for $\beta_j$ as follows.

Let $\{p_{ji}\}_{i=1}^{n_j}$ be non-negative weights allocated to $\{(X_{ji}'_r, Y_{ji})\}_{i=1}^{n_j}$. The empirical likelihood for $\beta_j$ 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 $\lambda_j$ to solve the above optimization problem and following the standard derivation in empirical likelihood (Owen, 1990), it can be shown that

$$L_{n_j}(\beta_j) = \prod_{i=1}^{n_j} \left\{ \frac{1}{n_j} + \frac{1}{1 + \lambda_j Z_{ji}(\beta_j)} \right\}, \quad (4.3.11)$$

where $\lambda_j$ satisfies

$$\sum_{i=1}^{n_j} Z_{ji}(\beta_j) = 0. \quad (4.3.12)$$

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 $n = \sum_{i=1}^{k} n_j, n_j/n \to \rho_j$ for some non-zero $\rho_j$ as $n \to \infty$ such that $\sum_{i=1}^{k} \rho_j = 1$. As the $k$ samples are independent, the joint empirical likelihood for $(\beta_1, \beta_2, \ldots, \beta_k)$ is

$$L_n(\beta_1, \beta_2, \ldots, \beta_k) = \prod_{j=1}^{k} L_{n_j}(\beta_j).$$
The log likelihood ratio statistic for $H_{0a}$ is

$$\ell_n : = -2 \max_{\beta} \log L_n(\beta, \beta, \ldots, \beta) + \sum_{j=1}^{k} n_j \log n_j$$

$$= 2 \min_{\beta} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \log \{1 + \lambda_j^2 Z_{ji}(\beta)\}. \quad (4.3.13)$$

Using a Taylor expansion and the Lagrange multiplier to carry out the minimization in (4.3.13), the optimal solution to $\beta$ is

$$\left( \sum_{j=1}^{k} \Omega_{xj} B_j^{-1} \Omega_{xj} \right)^{-1} \left( \sum_{j=1}^{k} \Omega_{xj} B_j^{-1} \Omega_{xj} y_j \right) + o_p(1), \quad (4.3.14)$$

where $B_j = \lim_{n_j \to \infty} (n_j T_j)^{-1} \sum_{i=1}^{n_j} E\{Z_{ji}(\beta_{j0})Z_{ji}(\beta_{j0})^\tau\}$,

$$\Omega_{xj} = \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E \left\{ \frac{\delta_{jim}}{\pi_{jim}(\theta_j)} \tilde{X}_{jim} \tilde{X}_{jim}^\tau \right\}$$

and

$$\Omega_{xj} y_j = \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\theta_j)} \tilde{X}_{jim} (\tilde{Y}_{jim} - M_{jim} \tilde{\gamma}_j).$$

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 $\ell_n$.

**Theorem 1** If Conditions (A1-A4) given in the Appendix hold, then under $H_{0a}$, $\ell_n \overset{d}{\to} \chi^2_{(k-1)p}$ as $n \to \infty$.

The theorem suggests an empirical likelihood ANOVA test that rejects $H_{0a}$ if $\ell_n > \chi^2_{(k-1)p, \alpha}$ where $\alpha$ is the significant level and $\chi^2_{(k-1)p, \alpha}$ is the upper $\alpha$ quantile of the $\chi^2_{(k-1)p}$ distribution.

We next evaluate the power of the empirical likelihood ANOVA test under a series of local alternative hypotheses:

$$H_{1a} : \beta_{j0} = \beta_{10} + c_n n_j^{-1/2} \quad \text{for} \quad 2 \leq j \leq k$$

where $\{c_n\}$ is a sequence of bounded constants. Define $\Delta_\beta = (\beta_{10} - \beta_{20}, \beta_{10} - \beta_{30}, \ldots, \beta_{10} - \beta_{k0})^\tau$, $D_{1j} = \Omega_{x1}^{-1} \Omega_{x1y1} - \Omega_{xj}^{-1} \Omega_{xjyj}$ for $2 \leq j \leq k$ and $D = (D_{12}, D_{13}, \ldots, D_{1k})^\tau$. Let $\Sigma_D = \text{Var}(D)$ and $\gamma^2 = \Delta_\beta^\tau \Sigma_D^{-1} \Delta_\beta$. Theorem 2 gives the asymptotic distribution of $\ell_n$ under the local alternatives.
Theorem 2 Suppose Conditions (A1-A4) in the Appendix hold, then under \( H_{1a} \), \( \ell_n \xrightarrow{d} \chi^2_{(k-1)p}(\gamma^2) \) as \( n \to \infty \).

It can be shown that

\[
\Sigma_D = \Omega_{x_1}^{-1} B_1 \Omega_{x_1}^{-1} 1_{(k-1)} \otimes 1_{(k-1)} + \text{diag}\{ \Omega_{x_2}^{-1} B_2 \Omega_{x_2}^{-1}, \ldots, \Omega_{x_k}^{-1} B_k \Omega_{x_k}^{-1} \}. \tag{4.3.15}
\]

As each \( \Omega_{x_j}^{-1} \) is \( O(n^{1/2}) \), the non-central component \( \gamma^2 \) is non-zero and bounded. The power of the \( \alpha \) level empirical likelihood ANOVA test is \( \beta(\gamma) = P\{\chi^2_{(k-1)p}(\gamma^2) > \chi^2_{(k-1)p, \alpha} \} \). 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 \( \theta_j \). 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 \( Y_{jit,d} \). 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 \( \eta_j \). When there is absence of interaction, we may formulate the empirical likelihood for \( (\beta_j, \eta_j) \in R^{p+s} \) using

\[
Z_{ji}(\beta_j; \eta_j) = \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \left( X_{jim}^\tau, \frac{\partial g_{j0}^\tau(t_{jim}; \eta_j)}{\partial \eta_j} \right) \tau \left\{ Y_{jim} - X_{jim}^\tau \beta_j - g_{j0}(t_{jim}; \eta_j) \right\}
\]

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 \( \eta_j \).

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

\[
H_{0b} : g_{10}(\cdot) = \ldots = g_{k0}(\cdot).
\]

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: \( X_{jim}^\tau = (X_{jim}^\tau, M_{jim}^\tau) \) and \( \xi_j^\tau = (\beta_j^\tau, \gamma_j^\tau) \).

Plugging-in the estimator \( \hat{\xi}_j \) to (4.3.7), we have

\[
\tilde{g}_j(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t) (Y_{jim} - X_{jim}^\tau \hat{\xi}_j) \quad (4.4.16)
\]

It follows that, for any \( t \in [0, 1] \),

\[
\tilde{g}_j(t) - g_{j0}(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t) \left\{ \varepsilon_{ji}(t_{jim}) + X_{jim}^\tau (\xi_j - \hat{\xi}_j) + g_{j0}(t_{jim}) - g_{j0}(t) \right\} \quad (4.4.17)
\]

However, there is a bias of order \( h_j^2 \) in the kernel estimation since

\[
\sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t) \left\{ g_{j0}(t_{jim}) - g_{j0}(t) \right\} = \frac{1}{2} \{ \int z^2 K(z) dz \} g_{j0}(t)^2 h_j^2 + o_p(h_j^2).
\]

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

\[
\hat{g}_j(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t) \{ Y_{jim} - X_{jim}^\tau \hat{\xi}_j - (\tilde{g}_j(t_{jim}) - \tilde{g}_j(t)) \}.
\]

Based on this modified estimator \( \hat{g}_j(t) \), we define the auxiliary variable

\[
R_j \{ g_j(t) \} = \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\theta_j)} K \left( \frac{t_{jim} - t}{h_j} \right) \left\{ Y_{jim} - X_{jim}^\tau \hat{\xi}_j - g_j(t) - (\tilde{g}_j(t_{jim}) - \tilde{g}_j(t)) \right\}
\]
for empirical likelihood formulation. At true function \( g_{j0}(t) \), \( E[R_{ji}\{g_{j0}(t)\}] = o(1) \).

Using a similar procedure to \( L_{nj}(\beta_j) \) as given in (4.3.11) and (4.3.12), the empirical likelihood for \( g_{j0}(t) \) is

\[
L_{nj}\{g_{j0}(t)\} = \max \left\{ \prod_{i=1}^{n_j} p_{ji} \right\}
\]

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

\[
L_{nj}\{g_{j0}(t)\} = \prod_{i=1}^{n_j} \left\{ \frac{1}{n_j 1 + \eta_j(t) R_{ji}\{g_{j0}(t)\}} \right\}^n,
\]

where \( \eta_j(t) \) is a Lagrange multiplier that satisfies

\[
\sum_{i=1}^{n_j} \frac{R_{ji}\{g_{j0}(t)\}}{1 + \eta_j(t) R_{ji}\{g_{j0}(t)\}} = 0.
\]

(4.4.18)

The log empirical likelihood ratio for \( g_{10}(t) = \ldots = g_{k0}(t) := g(t) \), say, is

\[
L_n(t) = 2 \min_{g(t)} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \log(1 + \eta_j(t) R_{ji}\{g(t)\}),
\]

which is analogues of \( \ell_n \) in (4.3.13). As shown in the proof of Theorem 3 given in the Appendix, the leading order term of the \( L_n(t) \) is a studentized version of the distance

\[
(\hat{g}_1(t) - \hat{g}_2(t), \hat{g}_1(t) - \hat{g}_3(t), \ldots, \hat{g}_1(t) - \hat{g}_k(t)),
\]

namely between \( \hat{g}_1(t) \) and the other \( \hat{g}_j(t) \) \((j \neq 1)\). This motivates us to propose using

\[
T_n = \int_0^1 L_n(t) \varpi(t) dt
\]

(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 \( T_n \), we assume without loss of generality that for each \( h_j \) and \( T_j \), \( j = 1, \ldots, k \), there exist fixed finite positive constants \( \alpha_j \) and \( b_j \) such that \( \alpha_j T_j = T \) and \( b_j h_j = h \) for some \( T \) and \( h \) as \( h \to 0 \). Effectively, \( T \) is the smallest common multiple of \( T_1, \ldots, T_k \). 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 \( \{t_{jm}\} \). Let \( a_j = \rho^{-1}_j \alpha_j \),

\[
W_j(t) = \frac{f_j(t) / \{a_j b_j \sigma_{xj}^2\}}{\sum_{l=1}^{k} f_l(t) / \{a_l b_l \sigma_{xl}^2\}}
\]
and \( V_j(t) = K^{(2)}(0)\sigma_{f_j}^2 f_j(t) \) where \( \sigma_{f_j}^2 = \frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E\{\frac{\varepsilon_{j,m}}{\pi_{j,m}(\theta_j)}\} \). Furthermore, we define

\[
\Lambda(t) = \sum_{j=1}^{k} b_j^{-1} K^{(4)}(0)(1 - W_j(t))^2 + \sum_{j \neq j_1} (b_j b_{j_1})^{-1/2} K^{(4)}(0) W_j(t) W_{j_1}(t) \quad \text{and}
\]

\[
\mu_1 = \int_0^1 \left[ \sum_{j=1}^{k} b_j^{-2} V_j^{-1}(t) f_j^2(t) \Delta_{j_2}^2(t) - \left( \sum_{s=1}^{k} b_s^{-1} V_s^{-1/2}(t) W_s^2(t) f_s(t) \Delta_{s_2}(t) \right)^2 \right] \varpi(t) dt.
\]

We consider a sequence of local alternative hypotheses:

\[
g_{j_0}(t) = g_{10}(t) + C_{jn} \Delta_{jn}(t), \tag{4.4.21}
\]

where \( C_{jn} = (n_j T_j)^{-1/2} h_j^{-1/4} \) for \( j = 2, \ldots, k \) 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),

\[
h^{-1/2}(T_n - \mu_0) \overset{d}{\to} N(0, \sigma_0^2),
\]

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 \( H_{0\theta} : g_{10}(\cdot) = \ldots = g_{k0}(\cdot), \Delta_{jn}(t) = 0 \) which yields \( \mu_1 = 0 \) and

\[
h^{-1/2}\{T_n - (k - 1)\} \overset{d}{\to} N(0, \sigma_0^2).
\]

This may lead to an asymptotic test at a nominal significance level \( \alpha \) that rejects \( H_{0\theta} \) if

\[
T_n \geq h^{1/2} \tilde{\sigma}_0 z_\alpha + (k - 1), \tag{4.4.22}
\]

where \( z_\alpha \) is the upper \( \alpha \) quantile of \( N(0,1) \) and \( \tilde{\sigma}_0 \) is a consistent estimator of \( \sigma_0 \). 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 \( h_j \) (\( j = 1, \ldots, k \)) 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 $H_{0k}$ 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 $T_n$ is no longer available, and estimation of $\sigma^2_0$ 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 $T_n$ 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 $L_n(t)$. In this case, the model (4.2.2) takes the form

$$Y_{ji}(t) = g_j(X_{ji}(t), t) + \varepsilon_{ji}(t),$$

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 $\gamma_{j0}$, 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 $\beta_{j0}$. For this purpose, we define estimating equations for $\gamma_{j0}$

$$
\phi_{ji}(\gamma_{j0}) = \sum_{m=1}^{T_j} \frac{\delta_{jm}}{\pi_{jim}(\theta_j)} \tilde{M}_{jim}(\bar{Y}_{jim} - \bar{X}_{jim}^r \tilde{\beta}_j - \tilde{M}_{jim}^r \gamma_{j0}),
$$

(4.5.24)

where

$$
\tilde{\beta}_j = \left\{ \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jm}}{\pi_{jim}(\theta_j)} \bar{X}_{jim} \bar{X}_{jim}^r \right\}^{-1} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jm}}{\pi_{jim}(\theta_j)} \bar{X}_{jim}(\bar{Y}_{jim} - \bar{M}_{jim}^r \gamma_{j0})
$$

(4.5.25)

is the “estimator” of $\beta_j$ at the true $\gamma_{j0}$. Similar to establishing $\ell_{nj}(\beta_j)$ in Section 4.3, the log-empirical likelihood for $\gamma_{j0}$ can be written as

$$
\ell_{nj}^r(\gamma_j) = 2 \sum_{i=1}^{n_j} \log \left\{ 1 + \Lambda_j^r \phi_{ji}(\gamma_j) \right\},
$$

where the Lagrange multipliers $\Lambda_j$ satisfies

$$
\sum_{i=1}^{n_j} \frac{\phi_{ji}(\gamma_j)}{1 + \Lambda_j^r \phi_{ji}(\gamma_j)} = 0.
$$

(4.5.26)

To test for $H_{0d} : \gamma_{10} = \gamma_{20} = \cdots = \gamma_{k0} \ vs \ H_{1d} : \gamma_{i0} \neq \gamma_{j0}$ for some $i \neq j$, we construct the joint empirical likelihood ratio

$$
\ell^r_n : = 2 \min_{\gamma} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \log \left\{ 1 + \Lambda_j^r \phi_{ji}(\gamma) \right\},
$$

(4.5.27)

where $\Lambda_j$ satisfy (4.5.26).

The asymptotic distributions of the empirical likelihood ratios $\ell_{nj}^r(0)$ and $\ell^r_n$ 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 $\beta_{j0}$ with $M_{jim}$ and $\gamma_{j0}$ respectively.
Theorem 4  Under Conditions (A1-A4) given in the Appendix, then (i) under $H_{0c}$, $\ell^c_n(0) \xrightarrow{d} \chi^2_q$ as $n \to \infty$; (ii) under $H_{0d}$, $\ell_n \xrightarrow{d} \chi^2_{(k-1)q}$ as $n \to \infty$.

Based on Theorem 4, an $\alpha$-level empirical likelihood ratio test for the presence of the interaction in the $j$-th sample rejects $H_{0c}$ if $\ell^c_n(0) > \chi^2_{q,\alpha}$, and the ANOVA test for the equivalence of the interactive effects rejects $H_{0d}$ if $\ell_n > \chi^2_{(k-1)q,\alpha}$. The ANOVA test for $H_{0d}$ has a similar local power performance as that described after Theorem 2 for the ANOVA test regarding $\beta_{j0}$ 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 $\sigma_0^2$ in Theorem 3 and to speed up the convergence of $T_n$, 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 $\tilde{t}^f_j$ and $\tilde{t}^m_j$ 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 $\hat{X}_{ji}(t)$, $t \in \tilde{t}^m_j$, so that for any $t \in \tilde{t}^m_j$

$$\hat{X}_{ji}(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t)X_{jim}, \quad (4.6.28)$$

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

$$\hat{\epsilon}_{jim} = Y_{jim} - X^T_{jim}\hat{\xi}_j - \hat{g}_j(t_{jim}).$$
An estimator of $\sigma^2_j(t)$, the variance of $\varepsilon_{ji}(t)$, is $\hat{\sigma}^2_{ji}(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{ji,m}(t) \varepsilon^2_{ji}$ and an estimator of $\rho_{ji}(s,t)$, the correlation coefficient between $\varepsilon_{ji}(t)$ and $\varepsilon_{ji}(s)$ for $s \neq t$, is

$$\hat{\rho}_{ji}(s,t) = \sum_{i=1}^{n_j} \sum_{m \neq m'} H_{ji,m'}(s,t) \hat{e}_{ji,m'} \hat{e}_{ji,m'},$$

where $\hat{e}_{ji,m} = \hat{e}_{ji} / \hat{\sigma}_{ji}(t_{ji})$,

$$H_{ji,m'}(s,t) = \frac{\delta_{ji,m'}K_{b_j}(s-t_{ji})K_{b_j}(t-t_{jim'})/\pi_{ji,m'}(\hat{\theta}_{ji})}{\sum_{i=1}^{n_j} \sum_{m \neq m'} \delta_{ji,m'}K_{b_j}(s-t_{ji})K_{b_j}(t-t_{jim'})/\pi_{ji,m'}(\hat{\theta}_{ji})},$$

and $\pi_{ji,m'}(\hat{\theta}_{ji}) = \pi_{ji}(\hat{\theta}_{ji})\pi_{ji,m'}(\hat{\theta}_{ji})$ if $|m-m'| > d$; $\pi_{ji,m'}(\hat{\theta}_{ji}) = \pi_{ji,m}(\hat{\theta}_{ji})$ 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 $T_n$ (Fan, Huang and Li 2007). Then, the covariance $\Sigma_{ij}$ of $\varepsilon_{ji} = (\varepsilon_{ji1}, \cdots, \varepsilon_{jiT_j})^\tau$ is estimated by $\hat{\Sigma}_{ij}$ which has $\hat{\sigma}^2_{ji}(t_{ji})$ as its $m$-th diagonal element and $\hat{\rho}_{ji}(t_{ji,k}, t_{ji})$ for $k = 1, \ldots, T_j$ 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, $X_{ji} = (X_{ji1}(t_{ji1}), \cdots, X_{ji(T_j)}(t_{ji}))^\tau$ and $g_{ji}(t_{si}) = (g_{ji0}(t_{si1}), \cdots, g_{ji0}(t_{siT_j}))^\tau$, where $s$ may be different from $j$. Let $X^c_{ji} = \{X^o_{ji}, \hat{X}^m_{ji}\}$, where $X^o_{ji}$ contains observed $X_{ji}(t)$ for $t_j \in \bar{t}^o$ and $\hat{X}^m_{ji}$ collects the imputed $X_{ji}(t)$ for $t \in \bar{t}^m$ according to (4.6.28). Plugging the value of $X^c_{ji}$, we get $M^c_{ji} = \{M^o_{ji}, M^m_{ji}\}$, the observed and the imputed interactions for $(j, i)$-th subject, and then $X^c_{ji}$.

The proposed bootstrap procedure consists of the following steps:

**Step 1.** Generate a bootstrap re-sample $\{Y^*_ji, X^c_{ji}, \delta^*_ji, t_{ji}\}$ for the $(j, i)$-th subject by

$$Y^*_ji = X^c_{ji} \hat{\xi}_j + \hat{g}_1(t_{ji}) + \hat{\Sigma}_{ji} e^*_ji,$$

where $e^*_ji$'s are i.i.d. random vectors simulated from a distribution satisfying $E(e^*_ji) = 0$ and $\text{Var}(e^*_ji) = I_{T_j}$, $\delta^*_ji \sim \text{Bernoulli}(\pi_{ji}(\hat{\theta}_{ji}))$ where $\hat{\theta}_{ji}$ 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 $\xi_j$, $\theta_j$ and $g_j(t)$ based on the re-sample $\{Y^*_ji, X^c_{ji}, \delta^*_ji, t_{ji}\}$ 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

$$R^*_ji(\hat{g}_1(t)) = \sum_{m=1}^{T_j} \frac{\delta^*_ji}{\pi_{ji,m}(\theta_j)} K\left(\frac{t_{ji,m} - t}{h_j}\right) \{Y^*_ji - X^c_{ji} \hat{\xi}_j^* - \hat{g}_1(t_{ji,m}) - \hat{g}_j^*(t_{ji,m}) - \hat{g}_j^*(t)\}.$$
and use it to substitute \( R_{ji}\{g_j(t)\} \) in the formulation of \( L_n(t) \), we obtain \( L^*_n(t) \) and then
\[
T^*_n = \int L^*_n(t)\varpi(t)dt.
\]

**Step 3.** Repeat the above two steps \( B \) times for a large integer \( B \) and obtain \( B \) bootstrap values \( \{T^*_{nb}\}_{b=1}^B \). Let \( \hat{t}_\alpha \) be the \( 1 - \alpha \) quantile of \( \{T^*_{nb}\}_{b=1}^B \), which is a bootstrap estimate of the \( 1 - \alpha \) quantile of \( T_n \). Then, we reject the null hypothesis \( H_0 \) if \( T_n > \hat{t}_\alpha \).

The following theorem justifies the bootstrap procedure.

**Theorem 5** Assume Conditions (A1-A4) in the Appendix hold and \( h = O(n^{-1/5}) \). Let \( X_n \) denote the original sample, \( h \) and \( \sigma_0^2 \) be defined as in Theorem 3. The conditional distribution of \( h^{-1/2}(T^*_n - \mu_0) \) given \( X_n \) converges to \( N(0, \sigma_0^2) \) almost surely, namely,
\[
h^{-1/2}\{T^*_n - (k-1)\}|X_n \xrightarrow{d} N(0, \sigma_0^2) \quad \text{a.s.}
\]

### 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 \( j = \{1, 2, 3\} \), \( i = 1, \ldots, n_j \) and \( m = 1, \ldots, T_j \). This structure used to generate \( \{e_{jim}\}_{m=1}^{T_j} \) 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

$$\text{logit}\{P(\delta_{jim} = 1|\delta_{jim,m-1} = 1, X_{ji}, Y_{ji})\} = \theta_j X_{ji(m-1)} \text{ for } m > 1,$$

which is not dependent on the response $Y$, with $\theta_1 = 3, \theta_2 = 2$ and $\theta_3 = 2$. In the second mechanism (II),

$$\text{logit}\{P(\delta_{jim} = 1|\delta_{jim,m-1} = 1, X_{ji}, Y_{ji})\} = \begin{cases} 
\theta_j X_{ji(m-1)} + \theta_j Y_{ji(m-1)} - Y_{ji(m-2)} & \text{if } m > 2, \\
\theta_j X_{ji(m-1)} & \text{if } m = 2;
\end{cases}$$

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

$$\sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{n_{jim}(\hat{\theta}_j)} \left( Y_{jim} - X_{jim}^\tau \hat{\beta}_j^{(i)} - M_{jim}^\tau \hat{\gamma}_j^{(-i)} - \hat{g}_j^{(-i)}(t_{jim}) \right)^2,$$

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 $H_0: \beta_{10} = \beta_{20} = \beta_{30}$ 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 $H_0$ 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 $\beta_{0j}$ 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 $H_{0a}: \beta_{10} = \beta_{20} = \beta_{30}$.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Missingness</th>
<th>Missingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>$n_2$</td>
<td>$n_3$</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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).
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 \(T(1-a)/\sqrt{\text{Var}\{T(1-a)\}}\) 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 \(\Delta_{2n}(\cdot)\), 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 $H_0: \gamma_{10} = \gamma_{20} = \gamma_{30}$.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>D_{2n}</th>
<th>D_{3n}</th>
<th>T</th>
<th>Missingness</th>
<th>Missingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_{1}</td>
<td>n_{2}</td>
<td>n_{3}</td>
<td>0.0</td>
<td>0.0 (size)</td>
<td>5</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
<td>0.2</td>
<td>0.0</td>
<td>0.134</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0</td>
<td>0.358</td>
<td>0.486</td>
<td>0.510</td>
<td>0.622</td>
</tr>
<tr>
<td>0.0</td>
<td>0.2</td>
<td>0.136</td>
<td>0.166</td>
<td>0.230</td>
<td>0.218</td>
</tr>
<tr>
<td>0.0</td>
<td>0.3</td>
<td>0.356</td>
<td>0.414</td>
<td>0.466</td>
<td>0.474</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.170</td>
<td>0.208</td>
<td>0.286</td>
<td>0.276</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
<td>0.292</td>
<td>0.328</td>
<td>0.462</td>
<td>0.428</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
<td>0.266</td>
<td>0.356</td>
<td>0.498</td>
<td>0.474</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>0.392</td>
<td>0.476</td>
<td>0.578</td>
<td>0.588</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
<td>0.0</td>
<td>0.0 (size)</td>
<td>5</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0</td>
<td>0.262</td>
<td>0.366</td>
<td>0.354</td>
<td>0.432</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0</td>
<td>0.654</td>
<td>0.744</td>
<td>0.744</td>
<td>0.820</td>
</tr>
<tr>
<td>0.0</td>
<td>0.2</td>
<td>0.272</td>
<td>0.330</td>
<td>0.340</td>
<td>0.334</td>
</tr>
<tr>
<td>0.0</td>
<td>0.3</td>
<td>0.590</td>
<td>0.676</td>
<td>0.722</td>
<td>0.672</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.282</td>
<td>0.332</td>
<td>0.412</td>
<td>0.410</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
<td>0.528</td>
<td>0.582</td>
<td>0.716</td>
<td>0.640</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
<td>0.502</td>
<td>0.580</td>
<td>0.680</td>
<td>0.728</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>0.672</td>
<td>0.674</td>
<td>0.814</td>
<td>0.808</td>
</tr>
</tbody>
</table>

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 $H_0: g_1(\cdot) = g_2(\cdot) = g_3(\cdot)$ with $
abla 2n(t) = \nu_n \sin(2\pi t)$ and $\Delta 3n(t) = 2 \sin(2\pi t) - 2 \sin(2\pi(t + \nu_n))$.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$U_n$</th>
<th>$V_n$</th>
<th>$T$</th>
<th>Missingness</th>
<th>Missingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>$n_2$</td>
<td>$n_3$</td>
<td></td>
<td>$T$</td>
<td>I</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
<td>0.00</td>
<td>0.00</td>
<td>0.040</td>
</tr>
<tr>
<td>0.30</td>
<td>0.00</td>
<td>0.186</td>
<td>0.232</td>
<td>0.282</td>
<td>0.256</td>
</tr>
<tr>
<td>0.50</td>
<td>0.00</td>
<td>0.666</td>
<td>0.718</td>
<td>0.828</td>
<td>0.840</td>
</tr>
<tr>
<td>0.00</td>
<td>0.05</td>
<td>0.664</td>
<td>0.726</td>
<td>0.848</td>
<td>0.842</td>
</tr>
<tr>
<td>0.00</td>
<td>0.10</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
<td>0.00</td>
<td>0.00</td>
<td>0.032</td>
</tr>
<tr>
<td>0.30</td>
<td>0.00</td>
<td>0.434</td>
<td>0.518</td>
<td>0.526</td>
<td>0.540</td>
</tr>
<tr>
<td>0.50</td>
<td>0.00</td>
<td>0.938</td>
<td>0.980</td>
<td>0.992</td>
<td>0.998</td>
</tr>
<tr>
<td>0.00</td>
<td>0.05</td>
<td>0.916</td>
<td>0.974</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.00</td>
<td>0.10</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

(PreCD4), and gender (Gender) on $Y = \text{log(CD4 counts +1)}$. The semiparametric model regression is, for $j = 1, 2, 3$ and 4,

$$Y_{ji}(t) = \beta_1 Age_{ji} + \beta_2 PreCD4_{ji} + \beta_3 Gender_{ji} + g_j(t) + \varepsilon_{ji}(t), \quad (4.8.32)$$

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
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 \( p_j(\vec{X}_{jit,3}, \vec{Y}_{jit,3}; \theta_j0) \) 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>
<thead>
<tr>
<th>Models</th>
<th>Treatment I</th>
<th>Treatment II</th>
<th>Treatment III</th>
<th>Treatment VI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC</td>
<td>BIC</td>
<td>AIC</td>
<td>BIC</td>
</tr>
<tr>
<td>M2-M3</td>
<td>-2.47</td>
<td>-11.47</td>
<td>0.93</td>
<td>-8.12</td>
</tr>
</tbody>
</table>

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 \( \beta_j \) 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 \( \beta_j \) 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.

We then formally tested \( H_{0a} : \beta_1 = \beta_2 = \beta_3 = \beta_4 \). The empirical likelihood ratio statistic \( \ell_n \) was 8.1348, which was smaller than \( \chi^2_{9,0.95} = 16.9190 \), 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 \( \beta_j \)'s as well as for equality of any three \( \beta_j \)'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>
<thead>
<tr>
<th>Coefficients</th>
<th>Treatment I</th>
<th>Treatment II</th>
<th>Treatment III</th>
<th>Treatment VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.0063(0.0039)</td>
<td>0.0050(0.0040)</td>
<td>0.0047(0.0058)</td>
<td>0.0056(0.0046)</td>
</tr>
<tr>
<td>PreCD4</td>
<td>0.7308(0.0462)</td>
<td>0.7724(0.0378)</td>
<td>0.7587(0.0523)</td>
<td>0.8431(0.0425)</td>
</tr>
<tr>
<td>Gender</td>
<td>0.1009(0.0925)</td>
<td>0.1045(0.0920)</td>
<td>-0.3300(0.1510)</td>
<td>-0.3055(0.1136)</td>
</tr>
</tbody>
</table>
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.

Table 4.8 P-values of ANOVA tests for $\beta_j$'s.

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>p-value</th>
<th>$H_0$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1 = \beta_2$</td>
<td>0.9661</td>
<td>$\beta_1 = \beta_2 = \beta_3$</td>
<td>0.7399</td>
</tr>
<tr>
<td>$\beta_1 = \beta_3$</td>
<td>0.4488</td>
<td>$\beta_1 = \beta_2 = \beta_4$</td>
<td>0.4011</td>
</tr>
<tr>
<td>$\beta_1 = \beta_4$</td>
<td>0.1642</td>
<td>$\beta_1 = \beta_3 = \beta_4$</td>
<td>0.3846</td>
</tr>
<tr>
<td>$\beta_2 = \beta_3$</td>
<td>0.4332</td>
<td>$\beta_2 = \beta_3 = \beta_4$</td>
<td>0.4904</td>
</tr>
<tr>
<td>$\beta_2 = \beta_4$</td>
<td>0.2523</td>
<td>$\beta_1 = \beta_2 = \beta_3 = \beta_4$</td>
<td>0.5206</td>
</tr>
<tr>
<td>$\beta_3 = \beta_4$</td>
<td>0.8450</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 doted 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 $T_n$ 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 $T_n$ 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.

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>p-value</th>
<th>$H_0$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_1(\cdot) = g_2(\cdot)$</td>
<td>0.894</td>
<td>$g_1(\cdot) = g_2(\cdot) = g_3(\cdot)$</td>
<td>0.046</td>
</tr>
<tr>
<td>$g_1(\cdot) = g_3(\cdot)$</td>
<td>0.018</td>
<td>$g_1(\cdot) = g_2(\cdot) = g_4(\cdot)$</td>
<td>0.010</td>
</tr>
<tr>
<td>$g_1(\cdot) = g_4(\cdot)$</td>
<td>0.004</td>
<td>$g_1(\cdot) = g_3(\cdot) = g_4(\cdot)$</td>
<td>0.000</td>
</tr>
<tr>
<td>$g_2(\cdot) = g_3(\cdot)$</td>
<td>0.020</td>
<td>$g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$</td>
<td>0.014</td>
</tr>
<tr>
<td>$g_2(\cdot) = g_4(\cdot)$</td>
<td>0.006</td>
<td>$g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$</td>
<td>0.004</td>
</tr>
<tr>
<td>$g_3(\cdot) = g_4(\cdot)$</td>
<td>0.860</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Models</th>
<th>Treatment I</th>
<th>Treatment II</th>
<th>Treatment III</th>
<th>Treatment VI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC</td>
<td>BIC</td>
<td>AIC</td>
<td>BIC</td>
</tr>
<tr>
<td>M1-M2</td>
<td>-7.885</td>
<td>-2.992</td>
<td>-3.870</td>
<td>1.039</td>
</tr>
<tr>
<td>M2-M3</td>
<td>6.506</td>
<td>-3.281</td>
<td>2.125</td>
<td>-7.693</td>
</tr>
</tbody>
</table>

Table 4.10 Differences in the AIC and BIC scores among three models (M1-M3) for $d = 1$.

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>
<thead>
<tr>
<th>Coefficients</th>
<th>Treatment I</th>
<th>Treatment II</th>
<th>Treatment III</th>
<th>Treatment VI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>( \beta_2 )</td>
<td>( \beta_3 )</td>
<td>( \beta_4 )</td>
</tr>
<tr>
<td>( d = 1 ) Age</td>
<td>0.0036(0.0031)</td>
<td>0.0061(0.0036)</td>
<td>0.0039(0.0043)</td>
<td>0.0059(0.0037)</td>
</tr>
<tr>
<td>PreCD4</td>
<td>0.7590(0.0389)</td>
<td>0.7440(0.0339)</td>
<td>0.7735(0.0441)</td>
<td>0.8441(0.0334)</td>
</tr>
<tr>
<td>Gender</td>
<td>0.0650(0.0874)</td>
<td>0.0343(0.1082)</td>
<td>-0.1941(0.1208)</td>
<td>-0.1892(0.0790)</td>
</tr>
<tr>
<td>( d = 2 ) Age</td>
<td>0.0065(0.0037)</td>
<td>0.0059(0.0042)</td>
<td>0.0002(0.0053)</td>
<td>0.0050(0.0041)</td>
</tr>
<tr>
<td>PreCD4</td>
<td>0.7538(0.0429)</td>
<td>0.7282(0.0360)</td>
<td>0.7574(0.0443)</td>
<td>0.8409(0.0392)</td>
</tr>
<tr>
<td>Gender</td>
<td>0.0309(0.0895)</td>
<td>0.0318(0.1075)</td>
<td>-0.2134(0.1282)</td>
<td>-0.3019(0.0910)</td>
</tr>
<tr>
<td>( d = 3 ) Age</td>
<td>0.0054(0.0036)</td>
<td>0.0049(0.0040)</td>
<td>0.0056(0.0053)</td>
<td>0.0044(0.0042)</td>
</tr>
<tr>
<td>PreCD4</td>
<td>0.7540(0.0443)</td>
<td>0.7666(0.0368)</td>
<td>0.7607(0.0482)</td>
<td>0.8476(0.0406)</td>
</tr>
<tr>
<td>Gender</td>
<td>0.0716(0.0955)</td>
<td>0.0942(0.0930)</td>
<td>-0.2776(0.1294)</td>
<td>-0.2527(0.1081)</td>
</tr>
</tbody>
</table>

Next, we summarize the ANOVA test results on \( \beta \)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 \( \beta_1, \beta_2 \) and \( \beta_4 \) 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 βs with $d = 1, 2, 3$

<table>
<thead>
<tr>
<th></th>
<th>$H_{0a}$</th>
<th>p-value</th>
<th>$H_{0a}$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d = 1$</td>
<td>$\beta_1 = \beta_2$</td>
<td>0.9810</td>
<td>$\beta_1 = \beta_2 = \beta_3$</td>
<td>0.9513</td>
</tr>
<tr>
<td></td>
<td>$\beta_1 = \beta_3$</td>
<td>0.7123</td>
<td>$\beta_1 = \beta_2 = \beta_4$</td>
<td>0.6240</td>
</tr>
<tr>
<td></td>
<td>$\beta_1 = \beta_4$</td>
<td>0.3855</td>
<td>$\beta_1 = \beta_3 = \beta_4$</td>
<td>0.7032</td>
</tr>
<tr>
<td></td>
<td>$\beta_2 = \beta_3$</td>
<td>0.7776</td>
<td>$\beta_2 = \beta_3 = \beta_4$</td>
<td>0.7006</td>
</tr>
<tr>
<td></td>
<td>$\beta_2 = \beta_4$</td>
<td>0.3322</td>
<td>$\beta_1 = \beta_2 = \beta_3 = \beta_4$</td>
<td>0.8208</td>
</tr>
<tr>
<td></td>
<td>$\beta_3 = \beta_4$</td>
<td>0.8362</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d = 2$</td>
<td>$\beta_1 = \beta_2$</td>
<td>0.9910</td>
<td>$\beta_1 = \beta_2 = \beta_3$</td>
<td>0.9404</td>
</tr>
<tr>
<td></td>
<td>$\beta_1 = \beta_3$</td>
<td>0.6640</td>
<td>$\beta_1 = \beta_2 = \beta_4$</td>
<td>0.4341</td>
</tr>
<tr>
<td></td>
<td>$\beta_1 = \beta_4$</td>
<td>0.2654</td>
<td>$\beta_1 = \beta_3 = \beta_4$</td>
<td>0.5296</td>
</tr>
<tr>
<td></td>
<td>$\beta_2 = \beta_3$</td>
<td>0.7375</td>
<td>$\beta_2 = \beta_3 = \beta_4$</td>
<td>0.5465</td>
</tr>
<tr>
<td></td>
<td>$\beta_2 = \beta_4$</td>
<td>0.2102</td>
<td>$\beta_1 = \beta_2 = \beta_3 = \beta_4$</td>
<td>0.6614</td>
</tr>
<tr>
<td></td>
<td>$\beta_3 = \beta_4$</td>
<td>0.7282</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d = 3$</td>
<td>$\beta_1 = \beta_2$</td>
<td>0.9967</td>
<td>$\beta_1 = \beta_2 = \beta_3$</td>
<td>0.8562</td>
</tr>
<tr>
<td></td>
<td>$\beta_1 = \beta_3$</td>
<td>0.5803</td>
<td>$\beta_1 = \beta_2 = \beta_4$</td>
<td>0.5230</td>
</tr>
<tr>
<td></td>
<td>$\beta_1 = \beta_4$</td>
<td>0.2916</td>
<td>$\beta_1 = \beta_3 = \beta_4$</td>
<td>0.5927</td>
</tr>
<tr>
<td></td>
<td>$\beta_2 = \beta_3$</td>
<td>0.5399</td>
<td>$\beta_2 = \beta_3 = \beta_4$</td>
<td>0.5723</td>
</tr>
<tr>
<td></td>
<td>$\beta_2 = \beta_4$</td>
<td>0.2702</td>
<td>$\beta_1 = \beta_2 = \beta_3 = \beta_4$</td>
<td>0.6891</td>
</tr>
<tr>
<td></td>
<td>$\beta_3 = \beta_4$</td>
<td>0.8404</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the p-values when \( d = 3 \) are quite similar to the monotone case.

<table>
<thead>
<tr>
<th>( H_{0b} )</th>
<th>p-value</th>
<th>( H_{0b} )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d = 1 )</td>
<td>( g_1(\cdot) = g_2(\cdot) )</td>
<td>0.750</td>
<td>( g_1(\cdot) = g_2(\cdot) = g_3(\cdot) )</td>
</tr>
<tr>
<td>( g_1(\cdot) = g_3(\cdot) )</td>
<td>0.068</td>
<td>( g_1(\cdot) = g_2(\cdot) = g_4(\cdot) )</td>
<td>0.070</td>
</tr>
<tr>
<td>( g_1(\cdot) = g_4(\cdot) )</td>
<td>0.026</td>
<td>( g_1(\cdot) = g_3(\cdot) = g_4(\cdot) )</td>
<td>0.038</td>
</tr>
<tr>
<td>( g_2(\cdot) = g_3(\cdot) )</td>
<td>0.110</td>
<td>( g_2(\cdot) = g_3(\cdot) = g_4(\cdot) )</td>
<td>0.058</td>
</tr>
<tr>
<td>( g_2(\cdot) = g_4(\cdot) )</td>
<td>0.016</td>
<td>( g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot) )</td>
<td>0.056</td>
</tr>
<tr>
<td>( g_3(\cdot) = g_4(\cdot) )</td>
<td>0.550</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( d = 2 )</td>
<td>( g_1(\cdot) = g_2(\cdot) )</td>
<td>0.896</td>
<td>( g_1(\cdot) = g_2(\cdot) = g_3(\cdot) )</td>
</tr>
<tr>
<td>( g_1(\cdot) = g_3(\cdot) )</td>
<td>0.154</td>
<td>( g_1(\cdot) = g_2(\cdot) = g_4(\cdot) )</td>
<td>0.036</td>
</tr>
<tr>
<td>( g_1(\cdot) = g_4(\cdot) )</td>
<td>0.016</td>
<td>( g_1(\cdot) = g_3(\cdot) = g_4(\cdot) )</td>
<td>0.054</td>
</tr>
<tr>
<td>( g_2(\cdot) = g_3(\cdot) )</td>
<td>0.216</td>
<td>( g_2(\cdot) = g_3(\cdot) = g_4(\cdot) )</td>
<td>0.106</td>
</tr>
<tr>
<td>( g_2(\cdot) = g_4(\cdot) )</td>
<td>0.048</td>
<td>( g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot) )</td>
<td>0.046</td>
</tr>
<tr>
<td>( g_3(\cdot) = g_4(\cdot) )</td>
<td>0.446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( d = 3 )</td>
<td>( g_1(\cdot) = g_2(\cdot) )</td>
<td>0.886</td>
<td>( g_1(\cdot) = g_2(\cdot) = g_3(\cdot) )</td>
</tr>
<tr>
<td>( g_1(\cdot) = g_3(\cdot) )</td>
<td>0.016</td>
<td>( g_1(\cdot) = g_2(\cdot) = g_4(\cdot) )</td>
<td>0.010</td>
</tr>
<tr>
<td>( g_1(\cdot) = g_4(\cdot) )</td>
<td>0.002</td>
<td>( g_1(\cdot) = g_3(\cdot) = g_4(\cdot) )</td>
<td>0.004</td>
</tr>
<tr>
<td>( g_2(\cdot) = g_3(\cdot) )</td>
<td>0.042</td>
<td>( g_2(\cdot) = g_3(\cdot) = g_4(\cdot) )</td>
<td>0.026</td>
</tr>
<tr>
<td>( g_2(\cdot) = g_4(\cdot) )</td>
<td>0.014</td>
<td>( g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot) )</td>
<td>0.004</td>
</tr>
<tr>
<td>( g_3(\cdot) = g_4(\cdot) )</td>
<td>0.812</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4.9 Appendix: Technical Details

We provide the conditions used for Theorems 1-5 and some remark 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 \( \mathcal{L}_{\theta_j}(\theta_j) \) for a q-dimensional parameter \( \theta_j \) defined in (4.2.6), and \( \theta_{j0} \) is in the interior of compact \( \Theta_j \). We assume \( E\{S(\theta_j)\} \neq 0 \) if \( \theta_j \neq \theta_{j0} \), \( \operatorname{Var}(S(\theta_{j0})) \) 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 $n_j h_j^2 / \log^2 n_j \to \infty$, $n_j^{1/2} h_j^4 \to 0$ and $h_j \to 0$ as $n_j \to \infty$.

(ii) For each treatment $j$ ($j = 1, \cdots, k$), 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 $T_j$, $j = 1, \cdots, k$, there exist finite positive constants $\alpha_j$, $b_j$ and $T_j$ such that $\alpha_j T_j = T$ and $b_j h_j = h$ for some $h$ as $h \to 0$. Let $n = \sum_{i=1}^{k} n_j$, $n_j/n \to \rho_j$ for some non-zero $\rho_j$ as $n \to \infty$ 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; $\max_{1 \leq i \leq n_j} \|u_{jim}\| = o_p\{n_j^{2(4+r)}(\log n_j)^{-1}\}$, $\max_{1 \leq i \leq n_j} E|\varepsilon_{jim}|^{4+r} < \infty$, for some $r > 0$; And assume that

$$
\lim_{n_j \to \infty} (n_j T_j)^{-1} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E\{\tilde{X}_{jim}\tilde{X}_{jim}^r\} = \Sigma_x > 0,
$$

where $\tilde{X}_{jim} = X_{jim} - E(X_{jim}|t_{jim})$.

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 $\Sigma_x$ 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 $\Sigma_x$.

**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

$$\lambda_j = \left(\sum_{i=1}^{n_j} Z_{ji}(\beta)Z_{ji}(\beta)^\tau\right)^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) + o_p(n_j^{-1/2}), \ j = 1, 2, \ldots, k.$$  

Then we can write

$$\ell_n = 2 \min_{\beta} \frac{1}{2} \sum_{j=1}^k \left\{ \sum_{i=1}^{n_j} Z_{ji}^\tau(\beta) \left( \sum_{i=1}^{n_j} Z_{ji}(\beta) Z_{ji}(\beta)^\tau \right)^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) \right\} + o_p \left\{ (\min n_j)^{-1/2} \right\}$$

$$= \min_{\beta} \sum_{j=1}^k \frac{1}{n_j T_j} \left\{ \sum_{i=1}^{n_j} Z_{ji}^\tau(\beta) B_j^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) \right\} + o_p \left\{ (\min n_j)^{-1/2} \right\}$$

where $B_j := \lim_{n_j \to \infty} \frac{1}{n_j T_j} \sum_{i=1}^{n_j} E\{ Z_{ji}(\beta_0) Z_{ji}(\beta_0)^\tau \}$, which is not related with $\beta$ for any $\beta = \beta_0 + \Delta_j n$ and $\Delta_j n = O(n_j^{-1/2})$.

Using the Lagrange method to carry out the minimizations in (4.9.34), we want to minimize

$$Q = \frac{1}{2} \sum_{j=1}^k \frac{1}{n_j T_j} \left( \sum_{i=1}^{n_j} Z_{ji}^\tau(\beta_j) B_j^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta_j) \right) - \sum_{j=2}^k \eta_j (\beta_1 - \beta_j),$$

where $\eta_1, \ldots, \eta_k$ are lagrange multipliers. Then

$$\frac{\partial Q}{\partial \beta_1} = \frac{1}{n_1 T_1} \sum_{i=1}^{n_1} Z_{i1}^\tau(\beta_1) B_1^{-1} \sum_{i=1}^{n_1} \sum_{m=1}^{T_1} \frac{\delta_{1im}}{\pi_{1im}(\theta)} \hat{X}_{1im} \hat{X}_{1im}^\tau - \sum_{j=2}^k \eta_j,$$

and

$$\frac{\partial Q}{\partial \beta_j} = \frac{1}{n_j T_j} \sum_{i=1}^{n_j} Z_{ji}^\tau(\beta_j) B_j^{-1} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\theta)} \hat{X}_{jim} \hat{X}_{jim}^\tau + \eta_j, \ j = 2, \ldots, k,$$

Setting $\beta_1 = \beta_2 = \cdots = \beta_k = \beta$, then the minima $\beta$ satisfies

$$\sum_{j=1}^k \frac{1}{\sqrt{n_j T_j}} \Omega_{xj} B_j^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) = o_p(1).$$

(4.9.35)

Inverting (4.9.35) for $\beta$, we have

$$\beta = \left( \sum_{j=1}^k \Omega_{xj} B_j^{-1} \Omega_{xj} \right)^{-1} \left( \sum_{j=1}^k \Omega_{xj} B_j^{-1} \Omega_{xjy_j} \right) + o_p(1).$$
Lemma 1 Suppose \((e_{i1}, \ldots, e_{iT})_{i=1}^T\) 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
\[
\max_{1 \leq j \leq n} \sum_{i=1}^n \sum_{m=1}^T a_{jim} e_{im} = O\{\max(n^{1/\delta} d_n, d_n^{1/2}) \log n\} \quad \text{a.s.}
\]

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: \(\beta_{l0} = \beta_{g0}\),
\[
\{(\Omega_{x_j}^{-1} B_j \Omega_{x_j}^{-1}) + (\Omega_{x_g}^{-1} B_g \Omega_{x_g}^{-1})\}^{-1/2} (\Omega_{x_j}^{-1} \Omega_{x_jy_j} - \Omega_{x_g}^{-1} \Omega_{x_gy_g}) \overset{d}{\to} N(0, I_p).
\]

Proof Since we know that
\[
\Omega_{x_j}^{-1} \Omega_{x_jy_j} = \frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\theta_j)} \tilde{X}_{jim} (\tilde{Y}_{jim} - \tilde{M}_{jim} \tilde{\gamma}_j)
\]
\[
= \frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\theta_j)} \tilde{X}_{jim} (\tilde{Y}_{jim} - \tilde{M}_{jim} \beta_j + \tilde{\gamma}_j) + \beta_j
\]
\[
= \frac{1}{n_j T_j} \sum_{i=1}^{n_j} Z_{ji}(\beta_j) + \beta_j
\]
and because samples \(l\) and \(g\) are mutually independent, we need to show that, for \(j = l, g\),
\[
(\Omega_{x_j}^{-1} B_j \Omega_{x_j}^{-1})^{-1/2} \Omega_{x_jy_j} \overset{d}{\to} N(0, I_p),
\]
which is equivalent to show that
\[
\frac{1}{n_j T_j} \sum_{i=1}^{n_j} Z_{ji}(\beta_j) \overset{d}{\to} N(0, B_j).
\]
Recall that \(\tilde{Y}_{jim} = \tilde{X}_{jim} \beta_j + \tilde{M}_{jim} \gamma_j + \tilde{\gamma}_j + \tilde{\varepsilon}_{jim}\), where \(\tilde{\gamma}_j = g_j(t_{jim}) - g_j(t_{jim}) - \tilde{\gamma}_j(t_{jim})\),
\(\tilde{\varepsilon}_{jim} = \varepsilon_{jim} - \varepsilon_{jim}\) and
\[
\overline{A}(t_{jim}) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,ni}(t_{jim}) A(t_{jim}).
\]
Then, it follows that
\[
\sum_{i=1}^{n_j} Z_{ji}(\beta_j) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\theta_j)} \tilde{X}_{jim} \left\{ \tilde{g}_j(t_{jim}) + \tilde{M}_{jim} (\gamma_j - \tilde{\gamma}_j) + \tilde{\varepsilon}_{jim} \right\}
\]
\[
= \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\theta_j)} \tilde{X}_{jim} \left\{ \tilde{g}_j(t_{jim}) + \tilde{M}_{jim} (\gamma_j - \tilde{\gamma}_j) + \tilde{\varepsilon}_{jim} \right\} \{1 + o_p(1)\}. 
\]
The last equality is true, since \( \hat{\theta}_j - \theta_{j0} = O_p(n_j^{-1/2}) \). At the same time, we can decompose
\[
\bar{X}_{jim} \left\{ \tilde{g}_{j0}(t_{jim}) + \bar{M}_{jim}(\gamma_{j0} - \bar{\gamma}_j) + \tilde{\varepsilon}_{jim} \right\}
\]
\[
= \left\{ \hat{h}(t_{jim}) + u_{jim} - \bar{u}(t_{jim}) \right\} \left\{ \tilde{g}_{j0}(t_{jim}) + \varepsilon_{jim} - \bar{\varepsilon}_{jim} \right\}
\]
\[
= u_{jim} \varepsilon_{jim} + \left\{ (\hat{h}(t_{jim}) - \bar{u}(t_{jim})) \varepsilon_{jim} + (\tilde{g}_{j0}(t_{jim}) - \bar{\varepsilon}_{jim}) u_{jim} \right\}
\]
\[
+ \left\{ (\tilde{h}(t_{jim}) - \bar{u}(t_{jim})) (\tilde{g}_{j0}(t_{jim}) - \bar{\varepsilon}_{jim}) \right\} + \bar{X}_{jim} \bar{M}_{jim}(\gamma_{j0} - \bar{\gamma}_j)
\]
\[
:= I_1 + I_2 + I_3 + I_4 , \text{ say.}
\]

From assumptions in A2(i) and the facts that
\[
\max_{1 \leq i,i' \leq n_j, 1 \leq m, m_1 \leq T_j} w_{jim1, m}(t_{jim}) = O\{(n_j h_j)^{-1}\}
\]
and \( \sum_{i=1}^{n_j} \sum_{m=1}^{h_{m1}} w_{jim1, m}(t_{jim}) = 1 \), we have, by applying Lemma 1
\[
\max_{1 \leq i \leq n_j} \| \hat{h}(t_{jim}) - \bar{u}(t_{jim}) \| = o(1) \text{ a.s., } \max_{1 \leq i \leq n_j} | \tilde{g}_{j0}(t_{jim}) - \bar{\varepsilon}_{jim} | = o(1) \text{ a.s.,} \quad (4.9.36)
\]
\[
\max_{1 \leq i \leq n_j} \| (\tilde{h}(t_{jim}) - \bar{u}(t_{jim}))(\tilde{g}_{j0}(t_{jim}) - \bar{\varepsilon}_{jim}) \| = o(n_j^{-1/2}) \text{ a.s.}
\]

Therefore,
\[
\frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} Z_{ji}(\beta_{j0}) = \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \left\{ \delta_{jim} \pi^{-1}_{jim}(\theta_{j0}) \right\} (I_1 + I_2 + I_3 + I_4)
\]
\[
:= J_1 + J_2 + J_3 + J_4 , \text{ say.}
\]

It is easy to see \( J_3 = o_p(1), J_4 = o_p(1) \) and from (4.9.36),
\[
|J_2| \leq o_p(1) \times \left\| \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \left\{ \delta_{jim} \pi^{-1}_{jim}(\theta_{j0}) \right\} u_{jim} \right\|
\]
\[
+ o_p(1) \times \left\| \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \left\{ \delta_{jim} \pi^{-1}_{jim}(\theta_{j0}) \right\} \varepsilon_{jim} \right\| = o_p(1).
\]

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 \( L_n := \sum_{j=1}^{n_j} \text{Var}(\sum_{m=1}^{T_j} \alpha' u_{jim} \nu_{jim} \varepsilon_{jim}) = O(n_j) \), where \\
\[ \Lambda_n(\epsilon) = \frac{1}{L_n} \sum_{i=1}^{n_j} E \left[ I \left\{ \sum_{m=1}^{T_j} \alpha' u_{jim} \nu_{jim} \varepsilon_{jim} \geq \epsilon \sqrt{L_n} \right\} \left\{ \sum_{m=1}^{T_j} \alpha' u_{jim} \nu_{jim} \varepsilon_{jim} \right\}^2 \right] \]
\[ \leq C \frac{C}{L_n} \sum_{i=1}^{n_j} E \left[ \sum_{m=1}^{T_j} \varepsilon_{jim} \left| 4+r \alpha \left\{ n_j^{r+2} (\log n_j)^{(4+r)} \right\} \right\| \alpha \right|^{4+r} \right] \]
\[ = \frac{C}{\epsilon^{r+2}} \alpha \left\{ \log^{-4+r}(n_j) \right\} \to 0, \]

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. \( \beta_{10} = \beta_{20}, \ell_n \to \chi^2_p \).

**Proof** Let \( S_1 := \sum_{j=1}^{k} \Omega_{x_j} B_j^{-1} \Omega_{x_j} \) and \( S_2 := \sum_{j=1}^{k} \Omega_{x_j} B_j^{-1} \Omega_{x_j} \). In this Lemma, \( k = 2 \). Then,
\[ \ell_n = (\Omega_{x_{1y1}} - S_2 S_1^{-1} \Omega_{x_{1y1}}) B_1^{-1} (\Omega_{x_{1y1}} - \Omega_{x_1} S_1^{-1} S_2) \]
\[ + (\Omega_{x_{2y2}} - S_2 S_1^{-1} \Omega_{x_{2y2}}) B_2^{-1} (\Omega_{x_{2y2}} - \Omega_{x_2} S_1^{-1} S_2) + o_p(1) \]
\[ = (\Omega_{x_{1y1}} - S_2 S_1^{-1} \Omega_{x_{1y1}}) B_1^{-1} \Omega_{x_1} S_1^{-1} (S_1 \Omega_{x_1} - S_2) \]
\[ + (\Omega_{x_{2y2}} - S_2 S_1^{-1} \Omega_{x_{2y2}}) B_2^{-1} \Omega_{x_2} S_1^{-1} (S_1 \Omega_{x_2} - S_2) + o_p(1) \]

It is easy to show that
\[ S_1 \Omega_{x_1} S_1^{-1} = S_2 = \Omega_{x_2} B_2^{-1} (\Omega_{x_1} - \Omega_{x_2} \Omega_{x_2}) \]
\[ S_1 \Omega_{x_2} S_1^{-1} = S_2 = \Omega_{x_1} B_1^{-1} (\Omega_{x_2} - \Omega_{x_1} \Omega_{x_1}) \]

Then
\[ \ell_n = (\Omega_{x_{1y1}} - \Omega_{x_{2y2}} \Omega_{x_2}) V (\Omega_{x_1} - \Omega_{x_2} \Omega_{x_2}) + o_p(1), \]

where
\[ V = (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) \]
\[ + (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) \]
\[ := P_1 + P_2, \text{say.} \]
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 \( V (\Omega_{x_1} B_1 \Omega_{x_1}^{-1} + V' (\Omega_{x_2} B_2 \Omega_{x_2}^{-1}) = I \). Thus, to prove the theorem, we just need to show that

\[ (\Omega_{x_1}^{-1} \Omega_{x_1} y_1 - \Omega_{x_2}^{-1} \Omega_{x_2} y_2)^\tau \left\{ (\Omega_{x_1}^{-1} B_1 \Omega_{x_1}^{-1} + (\Omega_{x_2}^{-1} B_2 \Omega_{x_2}^{-1}) \right\}^{-1} (\Omega_{x_1}^{-1} \Omega_{x_1} y_1 - \Omega_{x_2}^{-1} \Omega_{x_2} y_2) \xrightarrow{d} \chi_p^2, \]

which is true as Lemma 2 implies

\[ \{ (\Omega_{x_1}^{-1} B_1 \Omega_{x_1}^{-1} + (\Omega_{x_2}^{-1} B_2 \Omega_{x_2}^{-1}) \right\}^{-1/2} (\Omega_{x_1}^{-1} \Omega_{x_1} y_1 - \Omega_{x_2}^{-1} \Omega_{x_2} y_2) \xrightarrow{d} N(0, I_p). \]

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 \( \ell_n \),

\[ \ell_n = \sum_{j=1}^k \left( \Omega_{x_j}^\tau y_j - S_2 \Omega_{x_j}^{-1} \Omega_{x_j} S_1^{-1} \Omega_{x_j} y_j \right) B_j^{-1} (\Omega_{x_j} y_j - \Omega_{x_j} S_1^{-1} S_2) + o_p(1) \]

\[ = \sum_{j=1}^k \left( \Omega_{x_j}^\tau y_j \Omega_{x_j}^{-1} S_1 - S_2 \right) S_1^{-1} \Omega_{x_j} B_j^{-1} \Omega_{x_j} S_1^{-1} (S_1 \Omega_{x_j}^{-1} \Omega_{x_j} y_j - S_2) + o_p(1). \]

It can be shown that (similar to the proof of Lemma 3),

\[ \ell_n = \begin{pmatrix} \Omega_{x_1}^{-1} \Omega_{x_1} y_1 - \Omega_{x_2}^{-1} \Omega_{x_2} y_2 \\ \vdots \\ \Omega_{x_k}^{-1} \Omega_{x_k} y_k - \Omega_{x_k}^{-1} \Omega_{x_k} y_k \end{pmatrix} \Sigma_0 \begin{pmatrix} \Omega_{x_1}^{-1} \Omega_{x_1} y_1 - \Omega_{x_2}^{-1} \Omega_{x_2} y_2 \\ \vdots \\ \Omega_{x_k}^{-1} \Omega_{x_k} y_k - \Omega_{x_k}^{-1} \Omega_{x_k} y_k \end{pmatrix} + o_p(1), \] (4.937)

where \( \Sigma_0 \) is a \((k-1)p \times (k-1)p \) matrix with \((j-1)\)th \((j = 2, \cdots, k)\) 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 \((p, q = 2, \cdots, k)\) 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

$$\ell_n = \left( \frac{\Omega^{-1}_{x_1} \Omega_{x_1 y_1} - \Omega^{-1}_{x_2} \Omega_{x_2 y_2}}{\Omega^{-1}_{x_1} \Omega_{x_1 y_1} - \Omega^{-1}_{x_3} \Omega_{x_3 y_3}} \right)^{\tau} \Sigma_0 \left( \frac{\Omega^{-1}_{x_1} \Omega_{x_1 y_1} - \Omega^{-1}_{x_2} \Omega_{x_2 y_2}}{\Omega^{-1}_{x_1} \Omega_{x_1 y_1} - \Omega^{-1}_{x_3} \Omega_{x_3 y_3}} \right) + o_p(1),$$

(4.9.38)

where $\Sigma_0 = \begin{pmatrix} A & C \\ C^\tau & B \end{pmatrix}$,

$$A = \Omega_{x_2} B_2^{-1} \Omega_{x_2} - (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2}),$$

$$B = \Omega_{x_3} B_3^{-1} \Omega_{x_3} - (\Omega_{x_3} B_3^{-1} \Omega_{x_3}) S_1^{-1} (\Omega_{x_3} B_3^{-1} \Omega_{x_3}) \text{ and}$$

$$C = -(\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_3} B_3^{-1} \Omega_{x_3}).$$

From the proof of Lemma 2, we know that

$$\Sigma_1 = \text{Var} \left( \frac{\Omega^{-1}_{x_1} \Omega_{x_1 y_1} - \Omega^{-1}_{x_2} \Omega_{x_2 y_2}}{\Omega^{-1}_{x_1} \Omega_{x_1 y_1} - \Omega^{-1}_{x_3} \Omega_{x_3 y_3}} \right)$$

$$= \begin{pmatrix} \Omega^{-1}_{x_1} B_1 \Omega^{-1}_{x_1} + \Omega^{-1}_{x_2} B_2 \Omega^{-1}_{x_2} & \Omega^{-1}_{x_1} B_1 \Omega^{-1}_{x_1} \\ \Omega^{-1}_{x_1} B_1 \Omega^{-1}_{x_1} & \Omega^{-1}_{x_1} B_1 \Omega^{-1}_{x_1} + \Omega^{-1}_{x_3} B_3 \Omega^{-1}_{x_3} \end{pmatrix}. $$

As $\Sigma_0 = \Sigma_1^{-1}$, from (4.9.38) $\ell_n \overset{d}{\to} \chi^2_p$. This completes the proof.

□

**Proof of Theorem 2** We note that $\Sigma_D^{-1/2} D \overset{d}{\to} N_{(k-1)p}(\gamma, I_{(k-1)p})$, where $D$ and $\Sigma_D$ 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 $\ell_n \overset{d}{\to} \chi^2_{(k-1)p}(\gamma^2)$, which completes the proof of the theorem.

□

**Proof of Theorem 3** Let $v_j(t, h_j) = \sum_{i=1}^{n_j} R_{ji}^2 \{g(t)\}$ and

$$d_j(t, h_j) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \delta_{jm} K \left( \frac{t_{jm} - t}{h_j} \right).$$

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 $L_n(t)$ 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 $L_n(t)$ will be

$$
\left( \frac{\hat{g}_1(t) - \hat{g}_2(t)}{\hat{g}_1(t) - \hat{g}_3(t)} \right) H_n \left( \frac{\hat{g}_1(t) - \hat{g}_2(t)}{\hat{g}_1(t) - \hat{g}_3(t)} \right) \quad (4.9.39)
$$

where

$$
H_n = \left\{ \sum_{j=1}^{3} \zeta_j^{-1} \right\}^{-1} \times \left( \begin{array}{ccc}
\zeta^{-1} \left( \zeta_1^{-1} + \zeta_3^{-1} \right) & -\zeta_2^{-1} \zeta_3^{-1} \\
-\zeta_2^{-1} \zeta_3^{-1} & \zeta_1^{-1} \left( \zeta_1^{-1} + \zeta_2^{-1} \right)
\end{array} \right).
$$

Under the local alternative $g_{\hat{\theta}_0}(t) = g_0(t) + C_{n s} \Delta_{n s}(t)$ for $s = 2, \cdots, k$, the test statistic $T_n$ can be written as

$$
T_n = \int_0^1 \sum_{j=1}^{k} v_j^{-1} \left\{ B_n^2(t) + A_n^2(t) + 2A_n(t)B_n(t) \right\} \varpi(t) dt + o_p(1)
$$

$$
:= T_{n1} + T_{n2} + T_{n3} + o_p(1), \quad (4.9.40)
$$

where $A_n(t) = d_{j} \left\{ C_{n j} \Delta_{n j}(t) - \left( \sum_{s=1}^{k} v_s^{-1} d_s^2 \right)^{-1} \sum_{s=1}^{k} v_s^{-1} d_s^2 C_{n s} \Delta_{n s}(t) \right\}$ 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 $\sigma_{\hat{\epsilon}_j}^2 = \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E \left\{ \frac{\epsilon_{jim}^2}{\pi_{jm}(\hat{\theta}_0)} \right\}, \quad R(K) = \int K(t)^2 dt$ and $V_j(t) = R(K) \sigma_{\hat{\epsilon}_j}^2 f_j(t).

We first show that $(n_j h_j T_j)^{-1} v_j(t, h_j) \xrightarrow{p} V_j(t)$. 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

$$
\frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} R_{ji}^2 \{ g(t) \} = \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left( \sum_{m=1}^{T_j} K \left( \frac{t_{jim} - t}{h_j} \right) \nu_{jim} \hat{\epsilon}_{jim}(t_{jim}) \right)^2
$$

$$
+ \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left( \sum_{m=1}^{T_j} K \left( \frac{t_{jim} - t}{h_j} \right) \nu_{jim} \left\{ (g_{j0}(t_{jim}) - g_{j0}(t)) - (\hat{g}_{j}(t_{jim}) - \hat{g}_{j}(t)) \right\} \right)^2
$$

$$
+ \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left( \sum_{m=1}^{T_j} K \left( \frac{t_{jim} - t}{h_j} \right) \nu_{jim} X_{jim}^2 (\beta_{j0} - \hat{\beta}_j) \right)^2 + o_p(1)
$$

$$
:= A_1(t) + A_2(t) + A_3(t) + o_p(1).
$$

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

\[
A_2(t) = \frac{1}{n_jh_jT_j} \sum_{i=1}^{n_j} \left( \sum_{m=1}^{T_j} \nu_{jim} K \left( \frac{t_{jim} - t}{h_j} \right) (g_{j0}'(t) - \hat{g}_j'(t))(t - t_{jim}) + O_p(h_j^2) \right)^2 \\
\leq \frac{1}{n_jh_jT_j} \sum_{i=1}^{n_j} \left( \sum_{m=1}^{T_j} \nu_{jim} K \left( \frac{t_{jim} - t}{h_j} \right) \right)^2 \left| g_{j0}'(t) - \hat{g}_j'(t) \right|^2 h_j^2 + o_p(h_j^2) = o_p(h_j^2),
\]

since \( g_{j0}'(t) - \hat{g}_j'(t) = o_p(1) \). Note that \( A_1(t) \) can be written as

\[
A_1(t) = \frac{1}{n_jh_jT_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \sum_{m_1=1}^{T_j} K \left( \frac{t_{jim} - t}{h_j} \right) K \left( \frac{t_{jim_1} - t}{h_j} \right) \nu_{jim} \nu_{jim_1} \varepsilon_{ji}(t_{jim}) \varepsilon_{ji}(t_{jim_1})
\]

\[
= \frac{1}{n_jh_jT_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} K^2 \left( \frac{t_{jim} - t}{h_j} \right) \nu_{jim}^2 \varepsilon_{ji}^2(t_{jim})
\]

\[
+ \frac{1}{n_jh_jT_j} \sum_{i=1}^{n_j} \sum_{m \neq m_1}^{T_j} K \left( \frac{t_{jim} - t}{h_j} \right) K \left( \frac{t_{jim_1} - t}{h_j} \right) \nu_{jim} \nu_{jim_1} \varepsilon_{ji}(t_{jim}) \varepsilon_{ji}(t_{jim_1})
\]

\[
:= A_{11}(t) + A_{12}(t).
\]

Then

\[
E\{A_{12}(t)\} = \frac{T_j - 1}{h_j} \int_0^1 \int_0^1 K \left( \frac{x - t}{h_j} \right) K \left( \frac{y - t}{h_j} \right) \rho_j(x,y)\sigma_j(x)\sigma_j(y)f_j(x)f_j(y)\,dx\,dy
\]

\[
= h_j(T_j - 1)\sigma_{\varepsilon j}^2(t)f_j^2(t)\{1 + o(1)\} = O(h_j),
\]

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 \( n_jT_jh_j^2 \). However, the denominator is \( n_jT_jh_j \) 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 \( A_{11}(t) \stackrel{p}{\rightarrow} V_j(t) \).
Let us first consider the first term in $T_n$ given in (4.9.40),

$$T_{n1} = \int_0^1 \sum_{j=1}^k v_j^{-1} B_s^2(t) \varpi(t) dt$$

$$= \int_0^1 \sum_{j=1}^k v_j^{-1} \left[ \sum_{i=1}^{n_j} R_{ji} \{g_{j0}(t)\} \right]^2 - \int_0^1 \left( \sum_{s=1}^k v_s^{-1} d_s \right)^{-1} \left[ \sum_{s=1}^k v_s^{-1} d_s \sum_{i=1}^{n_s} R_{si} \{g_{s0}(t)\} \right]^2 \varpi(t) dt$$

$$= \int_0^1 \sum_{j=1}^k \left( 1 - \left[ \sum_{s=1}^k v_s^{-1} d_s \right]^{-1} v_j^{-1} d_j \right) v_j^{-1} \left[ \sum_{i=1}^{n_j} R_{ji} \{g_{j0}(t)\} \right]^2 \varpi(t) dt$$

$$- \int_0^1 \sum_{j \neq j_1} \left[ \sum_{s=1}^k v_s^{-1} d_s \right]^{-1} v_j^{-1} d_j v_j^{-1} d_j \left[ \sum_{i=1}^{n_j} R_{ji} \{g_{j0}(t)\} \right] \left[ \sum_{i=1}^{n_{j_1}} R_{j_1 i} \{g_{j_10}(t)\} \right] \varpi(t) dt$$

$$:= T_{n1}^{(1)} - T_{n1}^{(2)}, \quad \text{say}, \quad (4.9.41)$$

Let

$$S_{j1}^2(t) = \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \sum_{m_1=1}^{T_j} \frac{\mathcal{O}(\delta_{jim} \delta_{jim_1} \varepsilon_{jim} \varepsilon_{jim_1} \pi_{jim}^{(\theta_j)} \pi_{jim_1}^{(\theta_{j1})})}{m_1!} K \left( \frac{t_{jim} - t}{h_j} \right) K \left( \frac{t_{jim_1} - t}{h_j} \right)$$

and

$$S_{j2}^2(t) = \frac{2}{n_j h_j T_j} \sum_{i \leq i_1} \left\{ \sum_{m=1}^{T_j} \frac{\delta_{jim} \varepsilon_{jim} \pi_{jim}^{(\theta_j)}}{m!} K \left( \frac{t_{jim} - t}{h_j} \right) \right\} \left\{ \sum_{m_1=1}^{T_j} \frac{\delta_{jim_1} \varepsilon_{jim_1} \pi_{jim_1}^{(\theta_{j1})}}{m_1!} K \left( \frac{t_{jim_1} - t}{h_j} \right) \right\}.$$

We observe that

$$T_{n1}^{(1)} = \left[ \int_0^1 \sum_{j=1}^k \{1 - W_j(t)\} V_j^{-1}(t) S_{j1}^2(t) \varpi(t) dt \right] + \int_0^1 \sum_{j=1}^k \{1 - W_j(t)\} V_j^{-1}(t) S_{j2}^2(t) \varpi(t) dt \{1 + o_p(1)\}$$

$$:= (T_{n1}^{(11)} + T_{n1}^{(12)}) \{1 + o_p(1)\}. $$

Since $E\{S_{j1}^2(t)\} = V_j(t) + O(h)$, $E(T_{n1}^{(11)}) = \sum_{j=1}^k \int_0^1 (1 - W_j(t)) \varpi(t) dt = k - 1 + O(h)$ and the variance of $T_{n1}^{(11)}$ is $O(n^{-1} h) = o(h)$, under the condition $h = O(n^{-1/5}).$ Thus

$$h^{-1/2} \{T_{n1}^{(11)} - (k - 1)\} \overset{p}{\rightarrow} 0. \quad (4.9.42)$$

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

$$T_{n1}^{(12)} = \sum_{j=1}^k \sum_{i \neq i_1}^{n_j} v_j^{-1} \int_0^1 \{1 - W_j(t)\} V_j^{-1}(t) \xi_{ji}(t) \xi_{ji}(t) \varpi(t) dt + o_p(h^{1/2}). \quad (4.9.43)$$
Let \( N = \sum_{j=1}^{k} n_j \). We stack \( \xi_{ji} \ (j = 1, \ldots, k, i = 1, \ldots, n_j) \) to form a sequence \( \phi_s, \ s = 1, \ldots, N \). Let \( G_j \) be the collection of the subscripts of \( \phi \) whose corresponding \( \xi \) are in Treatment \( j \). Define

\[
C_{ps}(t) = \frac{1}{n(p, s)} \left( \sum_{j=1}^{k} I(p \in G_j, s \in G_j)V_j^{-1}(t) - \sum_{j=1}^{k} \sum_{l=1}^{k} \left( \frac{W_j(t)V_j(t)}{V_l(t)V_l(t)} \right)^{1/2} I(p \in G_j, s \in G_l) \right),
\]

(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

\[
U_N := T_n^{(12)} - T_n^{(2)} = 2 \sum_{p=1}^{N} \sum_{s<p} \psi(\phi_p, \phi_s),
\]

(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

\[
\{\text{Var}(U_N)\}^{-1} \max_{1 \leq p \leq N} \sum_{s=1}^{N} \sigma_{ps}^2 \to 0 \quad \text{and}
\]

\[
\{\text{Var}(U_N)\}^{-2} \text{EU}_N^4 \to 3,
\]

(4.9.47, 4.9.48)

then (4.9.46) is asymptotically normally distributed with mean 0 and variance

\[
\text{Var}(U_N) = \text{Var}(T_n^{(12)}) + \text{Var}(T_n^{(2)}) - 2\text{Cov}(T_n^{(12)}, T_n^{(2)}).
\]

(4.9.49)

Let us first derive \( \text{Var}(U_N) \). We note that \( \text{Var}(T_n^{(12)}) = \sum_{j=1}^{k} \frac{4}{n_j} \sum_{i < i_1} \sigma_{j,ii_1}^2 \), where

\[
\sigma_{j,ii_1}^2 = E_i E_{i_1} \left[ \int_0^1 \int_0^1 \frac{\left(1 - W_j(t)\right)\{1 - W_j(u)\}}{V_j(t)V_j(u)} \xi_{ji}(t)\xi_{ji_1}(t)\xi_{ji}(u)\xi_{ji_1}(u) \varpi(t) \varpi(u) dt du \right]
\]

\[
= \frac{1}{T_j^2} \sum_{m, m_1} \frac{\left(1 - W_j(t_{jim})\right)^2}{V_j^2(t_{jim})} \sigma_{j,ji}(t_{jim}) \sigma_{j,ji_1}(t_{jim})
\]

\[
\times \varpi^2(t_{jim}) \left( K(2) \left( \frac{t_{jim} - t_{ji_1m_1}}{h_j} \right) \right)^2 \left\{1 + o(1)\right\},
\]

and

\[
T_n^{(2)} = \sum_{j \neq j_1} \sum_{i=1}^{n_j} \sum_{l=1}^{n_{ji}} (n_j n_{ji})^{-1/2} \int_0^1 \left( \frac{W_j(t)W_{j_1}(t)}{V_j(t)V_{j_1}(t)} \right)^{1/2} \xi_{ji}(t)\xi_{ji_1}(t) \varpi(t) dt + o(h^{1/2}).
\]

(4.9.44)
where $\sigma_{\epsilon ji}(t_{jim}) = E\{\epsilon_{\epsilon ji}/\pi_{\epsilon ji}(\theta_{0})\}$. Since $\{t_{jim}\}$ are fixed design points generated from a density $f_j(t)$, via a Taylor expansion and by Assumption A2(ii),

$$\text{Var}(T_{n1}^{(12)}) = 2hR(K)^{-2}K_1^{(4)}(0) \sum_{j=1}^k b_j^{-1} \int_0^1 (1 - W_j(t))^2 \varpi^2(t)dt \{1 + o(1)\}. \quad (4.9.50)$$

Similar to our derivation for the variance of $T_{n1}^{(12)}$, it may be shown that

$$\text{Var}(T_{n1}^{(2)}) = 2hR(K)^{-2} \sum_{j \neq j_1}^k K_{b_j/\bar{b}_j}^{(4)}(0)(b_j \bar{b}_j)^{-1/2} \{ \int_0^1 W_j(t)W_j(t) \varpi^2(t)dt \} \{1 + o(1)\}. \quad (4.9.51)$$

From (4.9.49), we also need to calculate the covariance between $T_{n1}^{(12)}$ and $T_{n1}^{(2)}$. Using the same method for calculating variance for $T_{n1}^{(12)}$ and $T_{n1}^{(2)}$, we may show that

$$\text{Cov}(T_{n1}^{(12)}, T_{n1}^{(2)}) = O(h^2), \quad (4.9.52)$$

In summary of (4.9.50), (4.9.51) and (4.9.52),

$$\text{Var}(U_N) := h\sigma_0^2 = 2hR(0)^{-2} \int_0^1 \Lambda(t)\varpi^2(t)dt \{1 + o(1)\}, \quad (4.9.53)$$

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

$$\{\text{Var}(U_N)\}^{-1} \max_{1 \leq p \leq N} \sum_{s=1}^N \sigma_{ps}^2 = (h\sigma_0^2)^{-1} \max_{1 \leq j \leq k} \left\{ \frac{1}{n_j^2} \sum_{i=1}^{n_j} \sigma_{1,jii_1}^2 + \sum_{j=1}^k \frac{1}{n_jn_{j_1}} \sum_{i=1}^{n_j} \sigma_{2,jij_1i_1}^2 \right\} \leq (h\sigma_0^2)^{-1} \left[ \max_{1 \leq j \leq j_1} \left\{ \frac{1}{n_j^2} \sum_{i=1}^{n_j} \sigma_{1,jji_1}^2 \right\} + \max_{1 \leq j \leq k} \left\{ \sum_{j_1=1}^{n_j} \frac{1}{n_jn_{j_1}} \sum_{i=1}^{n_j} \sigma_{2,jij_1i_1}^2 \right\} \right].$$

From conditions (A2) and (A3),

$$\max_{1 \leq j \leq k} \frac{1}{n_j^2} \sum_{i=1}^{n_j} \sigma_{1,jii_1}^2 = \max_{1 \leq j \leq k} \frac{1}{n_jT_j} \sum_{m_{j_1=1}^{n_j}} \{1 - W_j(t_{jim})\}^2 \sigma_{2,jjim}^2 (t_{jim}) \varpi^2(t_{jim})$$

$$\times \left\{ \frac{1}{n_jT_j} \sum_{i_1=1}^{n_j} \sum_{m_{j_1=1}^{n_j}} \sigma_{2,jji_1}^2 (t_{jii_1m_1}) \left( K^{(2)}(t_{jim} - \frac{t_{jim}m_1}{h_j}) \right) \right\} ^2$$

$$= \max_{1 \leq j \leq k} \left\{ \frac{1}{n_jT_j} \sum_{i_1=1}^{n_j} \sum_{m_{j_1=1}^{n_j}} \{1 - W_j(t_{jim})\}^2 \sigma_{2,jji_1}^2 (t_{jim}) \sigma_{\epsilon j}^2 (t_{jim}) \frac{1}{f_j(t_{jim})} (t_{jim}) \varpi^2(t_{jim}) \right\}$$

$$\times \{R(K)^{-2}K_1^{(4)}(0)\} h_j = O(n^{-1}h).$$

And similarly, max $\{ \sum_{j_1=1}^{k} \frac{1}{n_jn_{j_1}} \sum_{i_1=1}^{n_{j_1}} \sigma_{2,jij_1i_1}^2 \} = O(n^{-1}h)$. These imply (4.9.47).
It is remain to check \((4.9.48)\). By \((4.9.46)\), we have

\[
E(U_N^4) = E(T_{n1}^{(12)})^4 - 4E\{(T_{n1}^{(12)})^3T_{n1}^{(2)}\} + 6E\{(T_{n1}^{(12)})^2(T_{n1}^{(2)})^2\} - 4E\{T_{n1}^{(12)}(T_{n1}^{(2)})^3\} + E(T_{n1}^{(2)})^4. 
\]

(4.9.54)

It can be seen that \(E\{(T_{n1}^{(12)})^3T_{n1}^{(2)}\} = E\{T_{n1}^{(12)}(T_{n1}^{(2)})^3\} = 0\). At the same time, we observed that

\[
E(T_{n1}^{(12)})^4 = E\left\{\sum_{j=1}^{k} \sum_{i \neq i_1} n_j^{-4} \left[ \int_0^1 \{1 - W_j(t)\} V_j^{-1}(t)\xi_j(t)\xi_j(t, t) \varpi(t)dt\right]^4 \right\} 
\]

\[
+ 3E\left\{\sum_{j=1}^{k} \sum_{i \neq i_1} n_j^{-2} \left[ \int_0^1 \{1 - W_j(t)\} V_j^{-1}(t)\xi_j(t, t) \varpi(t)dt\right]^2 \right\} 
\]

\[
x \sum_{j=1}^{k} \sum_{i \neq i_1} n_j^{-2} \left[ \int_0^1 \{1 - W_j(t)\} V_j^{-1}(t)\xi_j(t, t) \varpi(t)dt\right]^2 + o(h). 
\]

(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 \(3\{\text{Var}(T_{n1}^{(12)})\}^2\). Similarly, we can show that \(E(T_{n1}^{(12)})^4 \rightarrow 3\{\text{Var}(T_{n1}^{(2)})\}^2\) and \(6E\{(T_{n1}^{(12)})^2(T_{n1}^{(2)})^2\} \rightarrow 6\text{Var}(T_{n1}^{(12)})\text{Var}(T_{n1}^{(2)}).\) From \((4.9.54)\),

\[
\lim_{n \rightarrow \infty} \{\text{Var}(U_N)\}^{-2} E(U_N^4) = \lim_{n \rightarrow \infty} 3\{\text{Var}(U_N)\}^{-2}\{\text{Var}(T_{n1}^{(12)}) + \text{Var}(T_{n1}^{(2)})\}^2 = 3. 
\]

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 \((4.9.53)\),

\[
h^{-1/2}\{T_{n1} - (k - 1)\} \overset{d}{\rightarrow} N(0, \sigma_0^2). 
\]

(4.9.56)

Let us consider \(T_{n2} = \int_0^1 \sum_{j=1}^{k} b_j^{-1} A_n(t) \varpi(t)dt\). 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 \(C_{jn} = a_j^{-1/2} b_j^{-1/4} (n T)^{-1/2} h^{-1/4}\). and Then it can be shown that \(h^{-1/2}(T_{n2} - \mu_1) = o_p(1)\) and where

\[
\mu_1 = \int_0^1 \left[ \sum_{j=1}^{k} b_j^{-2} V_j^{-1}(t) f_j^2(t) \Delta_{nj}(t) \right] - \left( \sum_{s=1}^{k} b_s^{-2} V_s(t) \Delta_{ns}(t) \right)^2 \varpi(t)dt. 
\]
It remains to consider \( T_{n3} = 2 \int_0^1 \sum_{j=1}^k v_j^{-1} A_n(t)B_n(t)\varpi(t)dt \). Using the expression of \( A_n(t) \) and \( B_n(t) \), we can decompose \( T_{n3} \) as

\[
T_{n3} = 2 \int_0^1 \sum_{j=1}^k v_j^{-1} d_j C_{nj}(t) \sum_{i=1}^{n_j} R_{ji}(g_{j0}(t)) \varpi(t)dt
- 2 \int_0^1 \left( \sum_{j=1}^k v_j^{-1} d_j^2 \right)^{-1} \left( \sum_{j=1}^k v_j^{-1} d_j^2 C_{nj}(t) \right) \left( \sum_{s=1}^k v_s^{-1} d_s \sum_{i=1}^{n_s} R_{si}(g_{s0}(t)) \right) \varpi(t)dt
:= T_{n3}^{(1)} - T_{n3}^{(2)}, \quad \text{say.}
\]

We know that

\[
T_{n3}^{(1)} = 2 \int_0^1 \sum_{j=1}^k V_j(t)^{-1} f_j(t) C_{nj}(t) \sum_{i=1}^{n_j} R_{ji}(g_{j0}(t)) \varpi(t)dt
= 2 \sum_{j=1}^k h_j^{3/4} T_{n3}^{(1j)} \{1 + o_p(1)\}
\]

Thus we can say \( T_{n3}^{(1j)} \) is \( O_p(h^{3/4}) \) if we can show \( T_{n3}^{(1j)} = O_p(1) \). It is sufficient to show that \( \text{Var}(T_{n3}^{(1j)}) = O(1) \). Indeed, after some algebra, we get

\[
\text{Var}(T_{n3}^{(1j)}) = R(K)^{-2} \int_0^1 \sigma_{v_j}(y) \Delta_j^2(y)dy \int K^{(2)}(z)dz \{1 + o(1)\} = O(1).
\]

Therefore \( T_{n3}^{(1)} = O_p(h^{3/4}) \). The second term of \( T_{n3}^{(2)} \) can be written in a similar form as \( T_{n3}^{(1)} \), which is also \( O_p(h^{3/4}) \). Thus \( T_{n3} = O_p(h^{3/4}) \). In summary of these and (4.9.56), \( h^{-1/2} T_{n3} = o_p(1) \), we have via Slutsky Theorem, \( h^{-1/2}(T_n - (k - 1) - \mu_1) \xrightarrow{d} N(0, \sigma_0^2) \). 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

\[
(n_j h_j T_j)^{-1} v_j^*(t, h_j) - V^*(t) = o_p(1), \quad \text{as } n_j \to \infty.
\]
where \( V^*(t) = R(K)\hat{\sigma}_{ij}^2 f_j(t) \). This can be seen from the following decomposition,

\[
\frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} R_{ji}^2(\hat{g}_1(t)) = \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left( \sum_{m=1}^{T_j} \nu_{jim}^* K \left( \frac{t_{jim} - t}{h_j} \right) \hat{\xi}_{jim} \right)^2 \\
+ \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left( \sum_{m=1}^{T_j} \nu_{jim}^* K \left( \frac{t_{jim} - t}{h_j} \right) \chi_{jim}(\hat{\xi}_j - \hat{\xi}_j^*) \right)^2 \\
+ \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left( \sum_{m=1}^{T_j} \nu_{jim}^* K \left( \frac{t_{jim} - t}{h_j} \right) \{ (\hat{g}_1(t_{jim}) - \hat{g}_1(t)) - (\hat{g}_j(t_{jim}) - \hat{g}_j(t)) \} \right)^2 + o_p(1)
\]

\[
:= A_1^* + A_2^* + A_3^* + o_p(1),
\]

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 \( T_n \), the leading term of \( T_n^* \) is

\[
\int_0^1 \sum_{j=1}^{k} v_j^{*-1} \left[ \sum_{s=1}^{n_j} R_{sj}^* \{ \hat{g}_1(t) \} - \sum_{s=1}^{n_j} v_j^{*-1} d_j^* \sum_{s=1}^{n_s} R_{sj}^* \{ \hat{g}_1(t) \} \right]^2 \varpi(t) dt \\
= \int_0^1 \sum_{j=1}^{k} \{ 1 - W_j^*(t) \} V_j^{*-1}(t) S_j^2(t) \varpi(t) dt + \left\{ \int_0^1 \sum_{j=1}^{k} \{ 1 - W_j^*(t) \} V_j^{*-1}(t) S_j^2(t) \varpi(t) dt \\
- \int_0^1 \sum_{j \neq 1} \left[ \sum_{s=1}^{k} v_j^{*-1} d_j^* \right] v_j^{*-1} d_j^* \sum_{s=1}^{n_j} R_{sj}^* \{ \hat{g}_1(t) \} \sum_{i=1}^{n_j} R_{si}^* \{ \hat{g}_1(t) \} \varpi(t) dt \right\}
\]

\[
:= B_1^* + B_2^*,
\]

where \( W_j^*(t) = \frac{f_j(t)/(a_{ij} h_j \hat{\sigma}_{ij}^2)}{\sum_{l=1}^k f_l(t)/(a_{il} \hat{\sigma}_{il}^2)} \), \( S_j^2(t) \) and \( S_j^2(t) \) 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 \( T_{n1} \) in (4.9.41) in the proof of Theorem 3. We may show that

\[
h^{-1/2} \{ B_1^* - (k - 1) \} = o_p(1) \quad \text{and} \quad h^{-1/2} B_2^* | \mathcal{X}_n \overset{d}{\rightarrow} N(0, \sigma_0^2) \quad \text{a.s.}
\]

Hence, Theorem 5 is established. \( \square \)
References


Rice, J. and Silverman, B. (1991). Estimating the mean and covariance structure nonparamet-

for repeated outcomes in the presence of missing data. *Journal of the American Statistical
Association*, 90, 106-121.


Statistical Association*, 91, 1278-1288.

tivariate Analysis*, 72, 132-148.

Society, Series B*, 50, 413-436.


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 \((p >> n)\) 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 \((p^5/n \to 0)\). Tang and Leng (2010) proposed a penalized empirical likelihood ratio for making inference \((p^2/n \to 0)\). 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.


