

**MULTIVARIATE ONE-SIDED TESTS FOR
MULTIVARIATE NORMAL AND NONLINEAR
MIXED EFFECTS MODELS WITH COMPLETE
AND INCOMPLETE DATA**

by

Tao Wang

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Statistics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

March 2011

©Tao Wang, 2011

ABSTRACT

Multivariate one-sided hypotheses testing problems arise frequently in practice. Various tests have been developed for multivariate normal data. However only limited literatures are available for multivariate one-sided testing problems in regression models. In particular, one-sided tests for nonlinear mixed effects (NLME) models, which are popular in many longitudinal studies, have not been studied yet, even in the cases of complete data. In practice, there are often missing values in multivariate data and longitudinal data. In this case, standard testing procedures based on complete data may not be applicable or may perform poorly if the observations that contain missing data are discarded. In this thesis, we propose testing methods for multivariate one-sided testing problems in multivariate normal distributions with missing data and for NLME models with complete and incomplete data. In the missing data case, testing methods are based on multiple imputations. Some theoretical results are presented. The proposed methods are evaluated using simulations. Real data examples are presented to illustrate the methods.

PREFACE

Chapter 2, Chapter 3 and Chapter 4 are works collaborated with Prof. Lang Wu, Department of Statistics, The University of British Columbia. A version of Chapter 2 has been accepted for publication. Versions of Chapter 3 and Chapter 4 will be submitted for publications. Check the first pages of these chapters to see footnotes with similar information. In these chapters, I identified and designed the research topics, performed the research, analyzed the data, did the simulation studies, and wrote multiple versions of the manuscripts. Prof. Lang Wu guided the identification and design of the research topics, the research, data analyses and simulation studies, and edited multiple versions of the manuscripts.

TABLE OF CONTENTS

Abstract	ii
Preface	iii
Table of Contents	iv
List of Tables	vii
List of Figures	ix
Acknowledgements	x
Dedication	xi
1 Introduction	1
1.1 Background and Motivation	1
1.2 The Likelihood Ratio, Wald, and Score Tests	4
1.3 The Multivariate Normal Distribution and the Hotelling's T^2 Test	6
1.4 Multivariate One-sided Hypothesis Testing	8
1.5 Longitudinal Data and Mixed Effects Models	14
1.6 Inference with Missing Data	18
1.7 Outline of Thesis	22
2 Multiple Imputation Methods for Multivariate One-sided Tests with Missing Data	27
2.1 Introduction	27
2.2 Multiple Imputation Method	30
2.3 One-Sided Tests based on Multiple Imputations	34

2.3.1	Tests Based on Combined Parameter Estimates	34
2.3.2	Tests Based on Combining Test Statistics	38
2.4	Real Data Example	41
2.5	A Simulation Study	43
2.6	Discussion	46
3	Multivariate One-sided Hypothesis Tests for Nonlinear Mixed-Effects	
	Models	58
3.1	Introduction	58
3.2	Multivariate One-sided Tests for NLME Models	62
3.2.1	The LRT	62
3.2.2	Wald and Score Tests	65
3.3	Computing Methods for NLME Models	66
3.4	Real Data Example	68
3.5	A Simulation Study	69
3.6	Discussion	71
4	Multivariate One-sided Hypothesis Testing in NLME Models with	
	Missing Covariates	75
4.1	Introduction	75
4.2	A Motivating NLME Model and Multivariate One-sided Hypothesis . .	77
4.3	Multiple Imputation for NLME Model with Missing Data	79
4.4	Multivariate One-sided Tests based on Multiple Imputations for NLME	
	Models with Missing Covariates	85

4.5	Real Data Example	89
4.6	Simulation Study	90
4.7	Discussion	91
5	Conclusion and Discussion	94
	References	99
	Appendix	103
A.1	Proof of Theorem 2.1	103
A.2	Proof of Theorem 2.2	104
A.3	Proof of Theorem 3.1	106
A.4	Proof of Theorem 3.2	107
A.5	Proof of Theorem 3.3	108
A.6	Lindeberg Condition	111

LIST OF TABLES

Table 2.1	P-values of the tests for the mental distress data	48
Table 2.2	Simulation results for $p=2$ based on the bound method (equal missing rate = 20%)	49
Table 2.3	Simulation results for $p=2$ based on the substitution method (equal missing rate = 20%)	50
Table 2.4	Simulation results for $p=2$ under unequal missing rates ($n=200$, $\rho = 0.6$)	51
Table 2.5	Simulation results for $p=2$ under MAR ($n=200$, equal missing rate=20%)	52
Table 2.6	Simulation results for $p=5$ based on the bound method ($n=50$, equal missing rate = 20%)	53
Table 2.7	Simulation results for $p=5$ based on the substitution method ($n=50$, equal missing rate = 20%)	54
Table 2.8	Simulation results for $p=5$ based on AR(1) correlations ($n=50$, equal missing rate = 20%)	55
Table 3.1	Simulation results for $H_0 : \boldsymbol{\lambda} = \mathbf{0}$ vs $H_1 : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \neq \mathbf{0}$: bound method ($n=46$)	73
Table 3.2	Simulation results for $H_0 : \boldsymbol{\lambda} = \mathbf{0}$ vs $H_1 : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \neq \mathbf{0}$: substitution method ($n=46$)	74

Table 4.1	Simulation results: bound method (n=48, missing rate = 20% on each covariate)	93
Table 4.2	Simulation results: substitution method (n=48, missing rate = 20% on each covariate)	93

LIST OF FIGURES

Figure 1.1	Null and alternative parameter spaces for one-sided hypotheses.	24
Figure 1.2	Projection of \mathbf{x} onto \mathbb{O}^- .	25
Figure 1.3	Trajectories for viral loads in an anti-HIV treatment study.	26
Figure 2.1	Six randomly selected subjects and the total average scores.	56
Figure 2.2	Time series plot for draws of parameters.	57

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor, Prof. Lang Wu, for his continuous support, patience, motivation, and encouragement. His insightful guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Harry Joe and Prof. Matias Salibian-Barrera, for their encouragement, perceptive comments and penetrating questions.

I offer my enduring gratitude to the faculty, staff and my fellow students at the Department of Statistics, UBC, who have inspired and supported me to continue my work.

Special thanks are owed to my parents: Huiming Wang and Lihua Zhang, and my wife Jin Xia, for their unconditional support and deepest love to me. To them I dedicate this thesis.

DEDICATION

To my wife and my parents

1 Introduction

1.1 Background and Motivation

Multivariate one-sided hypotheses arise frequently in practice. In many research problems one-sided (inequality) restrictions are often imposed on model parameters due to various scientific and practical reasons. Intuitively, we expect that statistical methods incorporating these constraints would be more efficient than methods that simply ignore such information. Incorporating these restrictions into hypotheses testing leads to the so-called order-restricted hypotheses testing problems. Note that a one-sided hypothesis may be considered as a special case of order-restricted hypotheses.

For example, in clinical trials treatment effects may be measured by both efficacy and toxicity. Researchers may be interested in testing whether the treatment is superior to the control in at least one of the multiple endpoints. Let μ_1 and μ_2 denote the mean differences (say, treatment – control) of efficacy and toxicity respectively. The testing problem can then be written as

$$H_0 : \mu_1 \leq 0 \text{ and } \mu_2 \leq 0 \quad \text{versus} \quad H_1 : \mu_1 > 0 \text{ or } \mu_2 > 0. \quad (1.1)$$

Alternatively, one may test $\mu_1 = 0$ and $\mu_2 = 0$ versus $H_1 : \mu_1 > 0$ or $\mu_2 > 0$, if one believes that the treatment should be at least as efficient as the control. Figure 1.1 shows the null and alternative parameter spaces for hypotheses (1.1). For the same research question, the conventional multivariate two-sided methods such as the Hotelling's T^2 test neglects the inequality constraints and test the following two-sided

hypotheses

$$H_0 : \mu_1 = 0 \text{ and } \mu_2 = 0 \quad \text{versus} \quad H_1 : \mu_1 \neq 0 \text{ or } \mu_2 \neq 0.$$

Simply ignoring the one-sided constraints may result in loss of information and may even lead to misleading inferences (Silvapulle and Sen, 2004).

Another example is from an HIV/AIDS study, in which a biphasic viral decay model is proposed to model the viral load trajectories during an anti-HIV treatment (Wu and Ding, 1999). Let λ_1 and λ_2 denote the two-phase viral decay rates in the HIV viral dynamic model respectively. In this type of AIDS study, the viral decay rates are known to be non-negative, and a positive decay rate indicates a decrease in HIV viral loads and thus indicates an efficacy of the anti-HIV treatment. Therefore, the following one-sided testing problem is important for the evaluation of the treatment

$$H_0 : \boldsymbol{\lambda} = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \neq \mathbf{0}, \quad (1.2)$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$. More details of this problem will be provided in Section 1.3.

Order-restricted hypotheses testing has been an active research field for several decades, although the focus has mostly been on testing the mean vectors in multivariate normal distributions for continuous data or multinomial distributions for categorical data. When the data are completely observed, various testing methods have been proposed for multivariate one-sided hypotheses. Silvapulle and Sen (2004) provide an excellent review. However, only limited literatures are available for multivariate one-sided testing problems in regression models. In particular, one-sided or order-restricted inferences for nonlinear mixed effects (NLME) models, which are popular in

many longitudinal studies, have not been studied yet, even in the cases of complete data.

For multivariate data and longitudinal data in practice, missing values are very common, when it is difficult to observe each value for each variable or for each time point. Existing complete-data methods may not be applicable when the data contain missing values. In the presence of missing data, the commonly used naive complete-case method, which discards all incomplete observations, may be very inefficient if the missing rate is not low and may even give biased results if the missing data are not missing completely at random (Little and Rubin, 2002).

To our knowledge, little research has been done on the order restricted inference in the presence of missing data. Our work in this thesis is the first attempt to fill this important gap by using multiple imputation (MI) methods (Rubin, 1987). In other words, order-restricted testing methods for multivariate normal models with missing data and for NLME models with complete and incomplete data will be proposed in this thesis. Our proposed tests for missing data are based on MI methods.

In the following sections, brief reviews will be given on commonly used testing methods, the multivariate normal distribution, multivariate one-sided hypothesis testing problems, longitudinal data and NLME models, and missing data problems and multiple imputation methods. At the end of this chapter, an outline of the thesis will be provided.

1.2 The Likelihood Ratio, Wald, and Score Tests

The likelihood ratio test (LRT), Wald test, and score test are three commonly used hypotheses testing methods. They will be used throughout this thesis for order-restricted hypotheses with missing data. In this section, we provide a brief review of the standard theory of these three methods.

Suppose that $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ is a random sample from a probability distribution with density function $f(x|\theta), \theta \in \Theta$. For simplicity, we consider the simple case where the parameter space is one dimension. A general hypothesis about θ has the following form

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1,$$

where Θ_0 and Θ_1 are subspaces of the parameter space Θ . The LRT is based on the likelihood function $L(\theta|\mathbf{x})$ or the log-likelihood function $l(\theta|\mathbf{x}) = \log(L(\theta|\mathbf{x}))$. Let $\hat{\theta}_0$ and $\hat{\theta}$ denote the maximum likelihood estimates (MLEs) of θ under H_0 and $H_0 \cup H_1$ respectively, i.e.

$$L(\hat{\theta}_0|\mathbf{x}) = \max_{\theta \in \Theta_0} L(\theta|\mathbf{x}) \quad \text{and} \quad L(\hat{\theta}|\mathbf{x}) = \max_{\theta \in \Theta_0 \cup \Theta_1} L(\theta|\mathbf{x}).$$

The LRT statistic is then defined as the ratio of the above two maximum likelihoods:

$$T_{LR} = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}.$$

The smaller the LRT statistic T_{LR} , the stronger the evidence against H_0 . Inference is often based on the “null distribution”: the distribution of the test statistic given that H_0 is true. When H_0 consists of more than one point of the parameter values, the

“least favourable null distribution” is used as follows to calculate the p-value

$$p = \sup_{\boldsymbol{\theta} \in \Theta_0} \Pr(T_{LR} \leq t),$$

where t is the observed value of T_{LR} based on a given sample. Usually, the following alternative form of the LRT statistic is used

$$T_{LR}^* = -2[l(\widehat{\theta}_0|\mathbf{x}) - l(\widehat{\theta}|\mathbf{x})].$$

In the usual case of testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, T_{LR}^* has an asymptotic χ^2 distribution, under some regularity conditions. This asymptotic result is also true with the multi-dimensional $\boldsymbol{\theta}$ for two-sided or equality constrained hypotheses testing problems. However, this result does not hold for order-restricted or one-sided hypotheses.

The Wald test is named after Abraham Wald and is also widely used. Consider the following two-sided test

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0. \quad (1.3)$$

The Wald test statistic measures the (squared) distance between an estimate (say, MLE) $\widehat{\theta}$ and the null parameter θ_0 with respect to (w.r.t.) the variance of $\widehat{\theta}$, i.e.,

$$T_W = \frac{(\widehat{\theta} - \theta_0)^2}{\text{Var}(\widehat{\theta})}.$$

The asymptotic null distribution of T_W is the same as the LRT statistic T_{LR} for the above two-sided hypothesis. Many computer programs often produce the value of the Wald statistic by default. The extension to more general cases is straightforward and will be described later.

The score test is another commonly used large-sample testing method. The score test statistic is based on the “score” $S(\theta) = dl(\theta|\mathbf{x})/d\theta$ and the Fisher information $I_n(\theta) = E[dl(\theta|\mathbf{x})/d\theta]^2$. So the score test statistic for (1.3) is given by

$$\frac{S(\theta_0)^2}{I_n(\theta_0)}. \tag{1.4}$$

It has the same asymptotic null distribution as the Wald and LRT test statistics for the hypothesis (1.3). It has an advantage of easy computation.

The foregoing three tests are often called large-sample tests since they have the same asymptotic null distribution for the two-sided hypothesis (1.3). In other words, when the sample size $n \rightarrow \infty$, the asymptotic null distributions for the above LRT, Wald and Score testing statistics for the two-sided hypothesis (1.3) are the same and follow a χ^2 distribution under certain regularity conditions. However, the small-sample or finite sample performance of the three tests may differ, and no test is uniformly superior to the others. More details can be found in Casella and Berger (2001), among others.

1.3 The Multivariate Normal Distribution and the Hotelling’s

T^2 Test

The multivariate normal distribution is a generalized version of the univariate normal distribution to higher dimensions. It is the most popular distribution for multivariate continuous data. A random vector is called *multivariate normally distributed* if every linear combination of its components has a univariate normal distribution. Al-

ternatively, a p -dimensional random vector $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$ is said to follow a multivariate normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$ (Σ is positive definite) if its probability density function (pdf) can be expressed as

$$\frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right), \quad (1.5)$$

where $\boldsymbol{\mu}$ and Σ are the expectation and variance-covariance matrix of \mathbf{x} , respectively. When $p = 1$, the multivariate normal distribution reduces to the familiar univariate normal distribution.

In a multivariate normal distribution, researchers are often interested in testing the following multivariate two-sided hypotheses

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0.$$

Suppose n i.i.d. (independently and identically distributed) observations $\mathbf{x}_1, \dots, \mathbf{x}_n$ were drawn from $N_p(\boldsymbol{\mu}, \Sigma)$. Let $\bar{\mathbf{x}}$ and V denote the sample mean and sample covariance matrix respectively. When the covariance matrix Σ is known, both the LRT and Wald test statistics have the following form and follow a χ_p^2 distribution.

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \quad (1.6)$$

When Σ is unknown, both LRT and Wald test statistics are again the same, leading to the well-known Hotelling's T-square (T^2) testing statistic, which is a generalization of the Student's t statistic in the univariate case. Specifically, replacing the unknown Σ in (1.6) with the sample covariance matrix V , the *Hotelling's T-square statistic* is defined as

$$T_H = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{V}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0). \quad (1.7)$$

Under the null hypothesis H_0 , we have the following result

$$\frac{n-p}{p(n-1)}T_H \sim F_{p, n-p}, \quad (1.8)$$

which can be used to compute a p-value or critical value. More details can be found in Anderson (1984), among others.

In the case of multivariate one-sided hypotheses, such as (1.1) and (1.2), the testing statistics often involve orthogonal projections of the sample mean onto the one-sided parameter space. Thus, it is more difficult to compute the test statistics and derive their null distributions. More details will be provided in the next section.

1.4 Multivariate One-sided Hypothesis Testing

The two examples in Section 1.1 represent two common multivariate one-sided hypotheses in practice. For these one-sided hypotheses, various testing methods such as the likelihood ratio, Wald, and score tests have been proposed when there are no missing data. See Silvapulle and Sen (2004) for a review. Literatures on multivariate one-sided hypotheses testing for regression models are somewhat limited. Constrained inference for linear mixed-effects (LME) models is discussed in Stram and Lee (1994) and Verbeke and Molenberghs (2000). However, little has been done for multivariate one-sided hypothesis testing for NLME models. In this section, we provide a brief review of the standard results for multivariate one-sided or constrained hypothesis testing for multivariate normal data. Since many readers may be unfamiliar with these results, we use a similar example as in Silvapulle and Sen (2004) to illustrate the essential ideas.

Without loss of generality, consider an observation \mathbf{x} from $N_p(\boldsymbol{\mu}, \Sigma)$ and consider the following two common multivariate one-sided or constrained hypotheses about $\boldsymbol{\mu}$:

$$H_0 : \boldsymbol{\mu} \in \mathbb{O}^- \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \in \mathbb{R}^p \setminus \mathbb{O}^-, \quad (1.9)$$

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \geq \mathbf{0}, \boldsymbol{\mu} \neq \mathbf{0}, \quad (1.10)$$

where $\mathbb{O}^- = \{(\mu_1, \dots, \mu_p) \mid \max_{1 \leq j \leq p} \mu_j \leq 0\}$ is the nonpositive orthant in the p -dimensional space \mathbb{R}^p . When Σ is known, the LRT statistic for multivariate one-sided hypothesis (1.9) is given by

$$T_1 = \min_{\boldsymbol{\mu} \in \mathbb{O}^-} (\mathbf{x} - \boldsymbol{\mu}) \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \|\mathbf{x} - \pi_{\Sigma}(\mathbf{x}; \mathbb{O}^-)\|_{\Sigma}^2,$$

where $\pi_{\Sigma}(\mathbf{x}; \mathbb{O}^-) = \arg \min_{\boldsymbol{\mu} \in \mathbb{O}^-} (\mathbf{x} - \boldsymbol{\mu}) \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$ represents the point in \mathbb{O}^- that is Σ -closest to \mathbf{x} . Such a point is called the orthogonal projection of \mathbf{x} onto \mathbb{O}^- with respect to (w.r.t.) Σ . Similar to the univariate case, here the LRT statistic is also a Wald-type statistic and it measures the squared “distance” from \mathbf{x} to the null parameter space \mathbb{O}^- . The larger the “distance”, the stronger the evidence against H_0 .

Note that the expression of the projection $\pi_{\Sigma}(\mathbf{x}; \mathbb{O}^-)$ varies depending on the location where the sample point \mathbf{x} falls. For illustration, let $p = 2$ and let Q_1, Q_2, Q_3, Q_4 denote the four quadrants in the two-dimensional plane. As shown in Figure 1.2, the projection $\pi_{\Sigma}(\mathbf{x}; \mathbb{O}^-)$ when $\Sigma = I$ (identity matrix) is given by

$$\pi_I(\mathbf{x}; \mathbb{O}^-) = \begin{cases} (0, 0), & \text{if } \mathbf{x} \in Q_1 \\ (x_1, 0), & \text{if } \mathbf{x} \in Q_2 \\ (x_1, x_2), & \text{if } \mathbf{x} \in Q_3 \\ (0, x_2), & \text{if } \mathbf{x} \in Q_4 \end{cases}$$

Correspondingly, T_1 has the following expression

$$T_1 = \begin{cases} x_1^2 + x_2^2, & \text{given } \mathbf{x} \in Q_1 \\ x_2^2, & \text{given } \mathbf{x} \in Q_2 \\ 0, & \text{given } \mathbf{x} \in Q_3 \\ x_1^2, & \text{given } \mathbf{x} \in Q_4 \end{cases}$$

Therefore, we have the following conditional distributions given $\boldsymbol{\mu} = \mathbf{0}$,

$$T_1 \sim \begin{cases} \chi_2^2, & \text{given } \mathbf{x} \in Q_1 \\ \chi_1^2, & \text{given } \mathbf{x} \in Q_2 \\ \chi_0^2, & \text{given } \mathbf{x} \in Q_3 \\ \chi_1^2, & \text{given } \mathbf{x} \in Q_4 \end{cases}$$

where χ_0^2 indicates a distribution of a random variable that takes value 0 with probability 1. Consequently, the null distribution of T_1 at $\boldsymbol{\mu} = \mathbf{0}$ can be expressed as follows,

$$\begin{aligned} \Pr(T_1 \leq c) &= \sum_{i=1}^4 \Pr(T_1 \leq c, \mathbf{x} \in Q_i) \\ &= \sum_{i=1}^4 \Pr(T_1 \leq c | \mathbf{x} \in Q_i) \Pr(\mathbf{x} \in Q_i) \\ &= \frac{1}{4} \Pr(W_0 \leq c) + \frac{1}{2} \Pr(W_1 \leq c) + \frac{1}{4} \Pr(W_2 \leq c) \\ &= \sum_{j=0}^2 \omega_{2-j}(2, I, \mathbb{O}^-) \Pr(W_j \leq c), \end{aligned}$$

where c is a constant and $W_j \sim \chi_j^2$ ($j = 0, 1, 2$) and χ_j^2 indicates a χ^2 distribution with degree of freedom (d.f.) of j . The weight $\omega_{2-j}(2, I, \mathbb{O}^-)$ indicates the following probability

$$\omega_{2-j}(2, I, \mathbb{O}^-) = \Pr\left(\pi_I(\mathbf{x}; \mathbb{O}^-) \in \text{the } (2-j) \text{ dimensional face of } \mathbb{O}^-\right),$$

where the 2 dimensional face of \mathbb{O}^- indicates the origin point, the 1 dimensional face of \mathbb{O}^- indicates the negative sides of the two axis lines, and the 0 dimensional face of \mathbb{O}^- indicates the inside region of \mathbb{O}^- (the points within \mathbb{O}^- except the 2 and 1 dimensional faces). This distribution consists of a weighted sum of chi-square distributions, therefore it is named as a *chi-bar-square* ($\bar{\chi}^2$) *distribution*. Note that H_0 for hypothesis (1.9) consists of more than one parameter point. The least favourable null distribution is obtained at point $\boldsymbol{\mu} = \mathbf{0}$ (Silvapulle and Sen, 2004), i.e.

$$\sup_{\boldsymbol{\mu} \in H_0} \Pr(T_1 \geq c) = \Pr(T_1 \geq c | \boldsymbol{\mu} = \mathbf{0}).$$

In the general case, the null distribution of the LRT statistic T_1 at $\boldsymbol{\mu} = \mathbf{0}$, when Σ is known, is a $\bar{\chi}^2$ distribution given as follows

$$\Pr(T_1 \leq c | \boldsymbol{\mu} = \mathbf{0}) = \sum_{i=0}^p \omega_{p-i}(p, \Sigma, \mathbb{O}^-) \Pr(W_i \leq c), \quad \text{for any } c \geq 0,$$

where W_i represents a χ_i^2 distribution ($i = 1, 2, \dots, p$). As illustrated in the above $p = 2$ case, the weights

$$\omega_{p-i}(p, \Sigma, \mathbb{O}^-) = \Pr\left(\pi_{\Sigma}(\mathbf{x}; \mathbb{O}^-) \in \text{the } (p-i) \text{ dimensional face of } \mathbb{O}^-\right),$$

are called *probability weights*. Note that

$$\sum_{i=0}^p \omega_i(p, \Sigma, \mathbb{O}^-) = 1.$$

The computation of the weights ω_i can be difficult, but we may often compute the cut-off values of a $\bar{\chi}^2$ distribution using simulations. e.g. we may generate 10,000 random samples of \mathbf{x} from a multivariate normal distribution under the null hypothesis and then calculate the corresponding values T_1 to obtain the simulated cut-off values.

The LRT statistic for multivariate one-sided hypotheses (1.10), when Σ is known, is given by

$$\begin{aligned} T_2 &= \|\mathbf{x}\|_{\Sigma}^2 - \|\mathbf{x} - \pi_{\Sigma}(\mathbf{x}; \mathbb{O}^+)\|_{\Sigma}^2 \\ &= \mathbf{x}^T \Sigma^{-1} \mathbf{x} - \min_{\boldsymbol{\mu} \in \mathbb{O}^+} \{(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\}, \end{aligned} \quad (1.11)$$

where $\mathbb{O}^+ = \{\boldsymbol{\mu} \mid \mu_j \geq 0, 1 \leq j \leq p\}$ is the positive orthant in \mathbb{R}^p and $\pi_{\Sigma}(\mathbf{x}; \mathbb{O}^+)$ is the orthogonal projection of \mathbf{x} onto \mathbb{O}^+ w.r.t. Σ . So T_2 actually measures the difference of the “distance” from \mathbf{x} to the null parameter space $H_0 : \boldsymbol{\mu} = \mathbf{0}$ and the “distance” from \mathbf{x} to the union of the null and the alternative parameter spaces \mathbb{O}^+ w.r.t. Σ . The larger the difference, the stronger the evidence against H_0 . The null distribution of T_2 is given by the following $\bar{\chi}^2$ distribution,

$$\Pr(T_2 \leq c \mid \boldsymbol{\mu} = \mathbf{0}) = \sum_{i=1}^p \omega_i(p, \Sigma, \mathbb{O}^+) \Pr(W_i \leq c).$$

When Σ is unknown, Perlman (1969) developed classical results for the LRT, which are reviewed as follows. Suppose n i.i.d. observations $\mathbf{x}_1, \dots, \mathbf{x}_n$ are drawn from $N_p(\boldsymbol{\mu}, \Sigma)$. Let $\bar{\mathbf{x}}$ and V denote the sample mean and sample covariance matrix respectively. Also let $V^* = (n-1)V$. The LRT statistic for (1.10) is then given by

$$T_3 = \|\pi_{V^*}(\bar{\mathbf{x}}; \mathbb{O}^+)\|_{V^*}^2 \left[1 + \|\bar{\mathbf{x}} - \pi_{V^*}(\bar{\mathbf{x}}; \mathbb{O}^+)\|_{V^*}^2 \right], \quad (1.12)$$

and the null distribution for T_3 is given by

$$\Pr(T_3 \geq c \mid H_0) = \sum_{i=1}^p \omega_i(p, \Sigma, \mathbb{O}^+) \Pr(G_{i, d+1-p} \geq c), \quad (1.13)$$

where $G_{a,b} = W_a/W_b$ with $W_a \sim \chi_a^2$ and $W_b \sim \chi_b^2$, and $d = n - 1$. This result relies heavily on the fact that

$$V^* \sim \text{Wishart}(\Sigma, n - 1), \quad (1.14)$$

where $\text{Wishart}(\Sigma, n - 1)$ denotes a Wishart distribution with d.f. $n - 1$. Note that the null distribution of T_3 depends on the unknown nuisance parameters in Σ . So the supremum of the null probabilities over the nuisance parameter space is often used to calculate the p-value (Perlman, 1969)

$$\sup_{\Sigma} \Pr(T_3 \geq c) = \frac{1}{2} \Pr(G_{p-1, d+1-p} \geq c) + \frac{1}{2} \Pr(G_{p, d+1-p} \geq c). \quad (1.15)$$

Because a supremum is used, this test can be conservative.

An alternative method with Σ unknown is to find a consistent estimate $\widehat{\Sigma}$, and then replace Σ with $\widehat{\Sigma}$ in T_1 or T_2 . The asymptotic null distributions for the new testing statistics obtained this way are often the same as the null distributions for the Σ known case when the Slutsky's theorem can be applied. This method is particularly useful in complicated cases where the exact distributions of the test statistics are intractable. However, to obtain a good estimate $\widehat{\Sigma}$, the sample size needs to be reasonably large. Moreover, this is an approximate method so its performance needs to be evaluated in a case-by-case basis.

The foregoing results can be extended to the normal regression cases in which the mean vector involves covariates. The extension is conceptually straightforward but can be technically complicated. Silvapulle and Sen (2004) discuss these extensions. The foregoing results can also be extended to multivariate discrete data in which a

multinomial distribution is often assumed.

1.5 Longitudinal Data and Mixed Effects Models

Longitudinal data arise frequently in practice, such as in clinical trials and epidemiology studies. In longitudinal studies, multiple measurements for the variables are obtained for each subject in the study over a period of time. This allows us to study changes of variables over time. A key characteristic of longitudinal data is that measurements for the same subjects are often assumed to be correlated, while the measurements between different subjects are often assumed to be independent. When the variables are continuous, a multivariate normal model may be assumed for the longitudinal data with a certain structure for the mean vector. In fact, longitudinal data combine elements of multivariate data and time series data.

Figure 1.3 shows a longitudinal dataset in an anti-HIV treatment study. In this study, 53 HIV-1 infected patients were treated with a potent antiretroviral treatment. Five of them discontinued the study because of intolerance to the drug and other problems. The viral load (plasma HIV-1 RNA) was measured at baseline and between day 2 to day 90 after initiation of the treatment. The number of measurements for each individual varies from 3 to 8. Figure 1.3 shows the viral loads. We can see that the trajectories across different patients vary and the general trend of the viral loads seems to decrease over time, which indicates that the anti-HIV treatment may be effective.

In practice, regression models are widely used in both cross-sectional data and longitudinal data. In a regression model, covariates are used to partially explain the sys-

tematic variation in the response. The remaining unexplained variation in the response is treated as random errors. For longitudinal data, one should take into account the correlations of the repeated response measurements within a subject (cluster). There are three common approaches for analyzing longitudinal data: mixed effects models, copula or dependence models, and transitional models. In a mixed effects model, the correlation within a subject (cluster) is accounted because the observations within each subject (cluster) share the same random effects. Copula or dependence models separate the modelling of univariate models and the joint dependence structure. A transitional model assumes that the current response value depends on the previous response values, which suggests a Markov process. This may be specified by incorporating previous response values as additional “covariates”.

In this thesis, we will consider mixed effects models for longitudinal data. Mixed effects models are widely used in practice because

- they are natural extensions of the corresponding regression models for cross-sectional data by introducing random effects;
- subject-specific inference is available;
- in the presence of missing data or measurement errors, which arise frequently in practice, likelihood-based inference for mixed effects models is conceptually straightforward.

Moreover, for normal mixed effects models, the response distribution is closely related to the multivariate normal distribution. For example, in a linear mixed effects (LME)

model, the response is multivariate normal with special mean and covariance structures. More discussions can be found in Wu (2009). In the following, linear and nonlinear mixed effects models will be reviewed in more detail.

Linear mixed effects (LME) models first assume the existence of the random effects and then assume a linear relationship between the response variable and covariates and random effects. Specifically, let $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^T$ indicate the m_i response values for subject i ($i = 1, 2, \dots, n$) at time $t_{i1}, t_{i2}, \dots, t_{im_i}$. Also let X_i and Z_i indicate the known design matrices (with dimensions $m_i \times p$ and $m_i \times q$ respectively) that may contain covariates. A general LME model can be expressed as

$$\begin{aligned}\mathbf{y}_i &= X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \mathbf{e}_i, & i = 1, 2, \dots, n \\ \mathbf{e}_i &\sim N(\mathbf{0}, R_i), & \mathbf{b}_i \sim N(\mathbf{0}, D),\end{aligned}$$

where $\boldsymbol{\beta}$ is a p -dimensional vector of fixed effects that is the same for all subjects, \mathbf{b}_i is a q -dimensional vector of random effects that varies across subjects with the unknown covariance matrix D , and \mathbf{e}_i is a vector of random errors with the unknown covariance matrix R_i . Although there are various options for the structure of R_i , in practice researchers often assume that the random errors are independent and share the same variance, i.e. $R_i = \sigma^2 I$.

LME models, like other linear models, are empirical models which may not describe the underlying true relationship between the response and covariates. Nonlinear mixed effects (NLME) models, or hierarchical nonlinear models, are widely used in many applications such as HIV viral dynamics, pharmacokinetic analysis, and studies of

growth and decay. As an example, the following biphasic HIV viral dynamics NLME model has been widely used in HIV/AIDS studies to evaluate anti-HIV treatments (Wu and Ding, 1999; Wu, 2004):

$$\begin{aligned}
y_{ij} &= \log_{10}(e^{P_{1i}-\lambda_{1i}t_{ij}} + e^{P_{2i}-\lambda_{2i}t_{ij}}) + e_{ij}, \\
P_{1i} &= P_1 + b_{1i}, \quad \lambda_{1ij} = \lambda_1 + a_1CD4_{ij} + a_2CD8_{ij} + b_{2i}, \\
P_{2i} &= P_2 + b_{3i}, \quad \lambda_{2i} = \lambda_2 + b_{4i}, \\
j &= 1, \dots, m_i, \quad i = 1, \dots, n,
\end{aligned}$$

where $y_{ij} = \log_{10}(V(t_{ij}))$ is the \log_{10} -transformation of viral load measurements for patient i at time t_{ij} (the \log_{10} -transformation is used to stabilize the variance and make the data more normally distributed), $CD4_{ij}$ and $CD8_{ij}$ are CD4 and CD8 cells counts at time t_{ij} , $\boldsymbol{\beta} = (P_1, P_2, \lambda_1, \lambda_2, a_1, a_2)^T$ are the fixed effects, $\mathbf{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^T$ are random effects, and e_{ij} is a random error. We assume that e_{ij} i.i.d. $\sim N(0, \sigma^2)$ and $\mathbf{b}_i \sim N(0, D)$, and that the random error and random effects are independent with each other. In this HIV dynamic model, e^{P_1} and e^{P_2} denote baseline viral load values, and parameters λ_1 and λ_2 denote the viral decay rates, for the corresponding cells compartments (see detailed interpretations in Section 3.1). This model can be generalized to the following general form of a NLME model:

$$\begin{aligned}
y_{ij} &= g(\mathbf{t}_{ij}, \boldsymbol{\beta}_{ij}) + e_{ij}, \quad \mathbf{e}_{ij} \sim N(0, \sigma^2), \\
\boldsymbol{\beta}_{ij} &= \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{b}_i, \quad \mathbf{b}_i \sim N(\mathbf{0}, D), \\
j &= 1, \dots, m_i, \quad i = 1, \dots, n,
\end{aligned}$$

where $g(\cdot)$ is a known real-valued nonlinear function, and \mathbf{x}_{ij} is a vector of covariates. In our HIV dynamic model $\beta_{ij} = (P_{1i}, P_{2i}, \lambda_{1ij}, \lambda_{2i})^T$. In this thesis, testing methods for multivariate one-sided hypothesis in NLME models will be proposed in both complete and incomplete data cases.

1.6 Inference with Missing Data

Missing data problems arise frequently in practice, especially in multivariate data and longitudinal data. At the presence of missing data, it is important to first understand how the data were missing because the methods to handle missing data depend on the missing data mechanisms.

Rubin (1976) defines the following three missing data mechanisms:

- missing completely at random (MCAR): missingness depends neither on the observed data nor on the unobserved data;
- missing at random (MAR): missingness may depend on the observed data but not on the unobserved data;
- missing not at random (MNAR): missingness may depend on the observed data and on the unobserved data.

By assuming MCAR, the observed data can be treated as a random subsample from the population, so the missing data mechanism can be ignored. However, it is a very strong assumption, which may not be reasonable in many cases. In practice, MAR

may be a more reasonable assumption, especially if more observed information is incorporated into the analysis. In MAR cases, the missing data mechanism can be safely ignored from the likelihood or multiple imputation based analysis if we also assume that the parameters in the missing mechanism model and the inference model are distinct. These two assumptions (MAR and distinctness) combined together establish the “*ignorability*” assumption (Rubin 1976 and Rubin 1987), which we will assume throughout this thesis. More details will be provided in Chapter 2. In many cases, MNAR is also a possibility. In such cases, the missing data mechanism needs to be included in the analysis.

When there are missing data, some naive methods are commonly used in practice. These include the complete-case (CC) method, which discards all incomplete observations from the analysis. However, when the missing data are not MCAR, the CC method may give biased results (Little and Rubin, 2002). In longitudinal studies (especially in pharmaceutical studies), the last-value-carried-forward (LVCF) method is widely used. It imputes a missing value in a covariate or response by its observed value immediately before the time point when the missing value occurs. This method may also be misleading because it fails to incorporate the missing data uncertainty (Wu, 2009). Therefore, more formal statistical methods are needed to appropriately handle missing data. The most commonly used formal statistical methods for missing data are the observed-data likelihood method and the multiple imputation (MI) method (Schafer, 1997).

The observed-data likelihood method is conceptually straightforward: it makes in-

ferences based on the observed-data likelihood. This way, not only the complete cases but also the observed data contained in the incomplete cases are included into the analysis, with the possibility to include the information of the missing data mechanism into the likelihood function if one has such knowledge. Because the missing data patterns can be arbitrary, the parameter estimates based on the observed-data likelihoods often do not have closed-form expressions. Therefore, the computation often involves numeric methods such as the Quasi-Newton and the expectation-maximization (EM) algorithms.

The MI method, is often easy to implement for complex problems since it uses complete-data methods once the missing data are multiply imputed. In an MI method, several (say m) simulated values for *each* missing value are imputed. The simulated values are drawn from a predictive distribution of the missing data given the observed data. This way, “proper” multiple imputations can be generated in the sense that they can provide valid frequentist inferences (Rubin and Little, 2002). MCMC (Markov Chain Monte Carlo) methods such as the Gibbs sampler are often used to generate multiple imputations. Now that once m “complete datasets” are obtained, each of the m “complete datasets” may be analyzed by standard complete-data methods. These analysis results are then combined to produce an overall inference, which takes into account the missing data uncertainty. Specifically, the main advantages of MI methods include:

- the missing data uncertainty is incorporated;

- standard complete-data methods can be used;
- the MI methods are relatively robust against the assumed imputation models when the missing data rate is not high;
- once the multiply imputed datasets are obtained, they can be used for different analyses by different analysts.

In practice, people may prefer to have a small number of multiply imputed datasets, say 3-5, due to various practical concerns about time, finance, computing and human resources etc, especially for large datasets such as in sample survey and public health. Therefore, our study in this thesis mainly focus on the multivariate one-sided hypothesis testing problems based on multiple imputations with small m (say 3-5). Li (1985), Rubin (1987) and Raghunathan (1987) have shown that when m is small the statistical inference methods under the “*equal ratios of missing to observed information*” assumption (see Chapter 2 for details) usually perform well. Therefore this assumption will be assumed in our study. In practice this theoretical assumption is often difficult to check, however, even when the assumption is violated the multivariate two-sided test still perform well, as in Li et al. (1991), among others. We consider similar missing rates in all variables to approximate the situation where the assumption is satisfied. Also, simulation studies will be done to check the situations when the equal-ratio assumption is violated for our methods. General theories and implementations of MI methods can be found in Rubin (1987).

MI methods are particularly attractive for order-restricted inferences for multi-

variate normal data because closed-form estimators are available for complete data. So we only need to generate the multiple imputations and then find appropriate ways to combine the complete-data analysis results. For multivariate two-sided hypothesis testing, the methods to properly combine analysis results from m “complete datasets” are discussed by Li et al. (1991) and Meng and Rubin (1992). In this thesis, testing methods based on MI methods will be proposed for multivariate one-sided hypothesis for multivariate normal data and for NLME models with missing values.

1.7 Outline of Thesis

We have provided a brief review of multivariate one-sided tests for complete multivariate normal data. However, multivariate one-sided tests for incomplete multivariate normal data have not been studied yet, even though missing values are common for multivariate data. Moreover, multivariate one-sided tests for NLME models, with either complete data or incomplete data, have not been studied yet. In this thesis, we attempt to fill in this important gap and develop innovative methods for multivariate one-sided tests for incomplete multivariate normal data, NLME models with complete data, and NLME models with incomplete data. We will use MI methods for the missing values.

The thesis will be organized as follows. In Chapter 2, MI based tests for multivariate normal data with missing values will be proposed. These include a Wald-type test by combining individual sufficient statistics, and two tests by combining individual Wald and LRT statistics respectively. In Chapter 3, LRT, Wald, and Score tests will

be proposed for NLME models with complete data, with applications in HIV viral dynamics studies. In Chapter 4, MI based tests for NLME models with missing data will be proposed. Real data examples and simulation studies will be provided for Chapters 2 - 4. In Chapter 5, conclusions and discussions will be given.

Note that, due to the manuscript-based format of this thesis, throughout the thesis some notation such as T_W , T_{LR} , X , etc., will be repeatedly used in different chapters but may define somewhat different but related variables. For example, the statistic T_W may mean a general Wald test statistic, or a Wald-type test statistic based on multiple imputations from NLME models. In other words, the same notation will mean the same variable within a chapter, but may mean different variables at different chapters. The context in each chapter should make this clear and not cause much confusion. Also note that the methods described in this thesis focus on the “*balanced-data*” case. i.e. if the multivariate data were completely observed, the numbers of observations for all variables in the dataset are the same. Therefore, the term “*complete-data methods*” in this thesis refers to the methods used for completely observed balanced data.

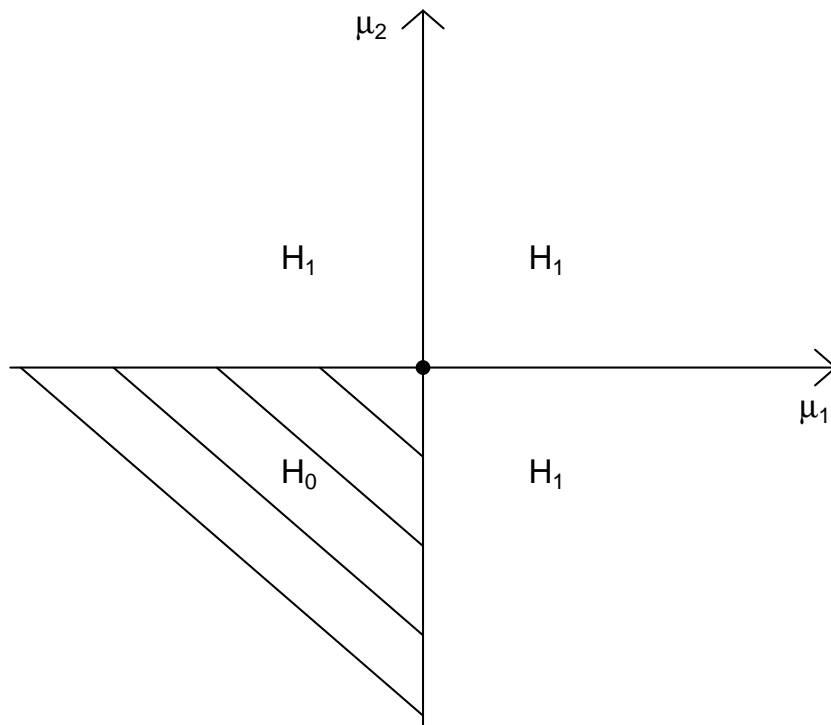


Figure 1.1: Null and alternative parameter spaces for one-sided hypotheses.

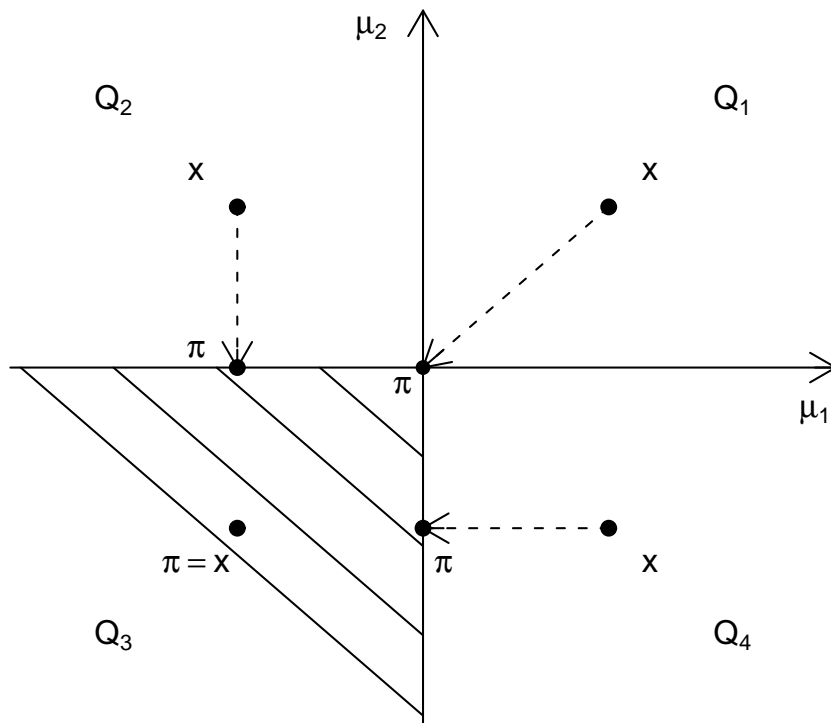


Figure 1.2: Projection of \mathbf{x} onto \mathbb{O}^- .

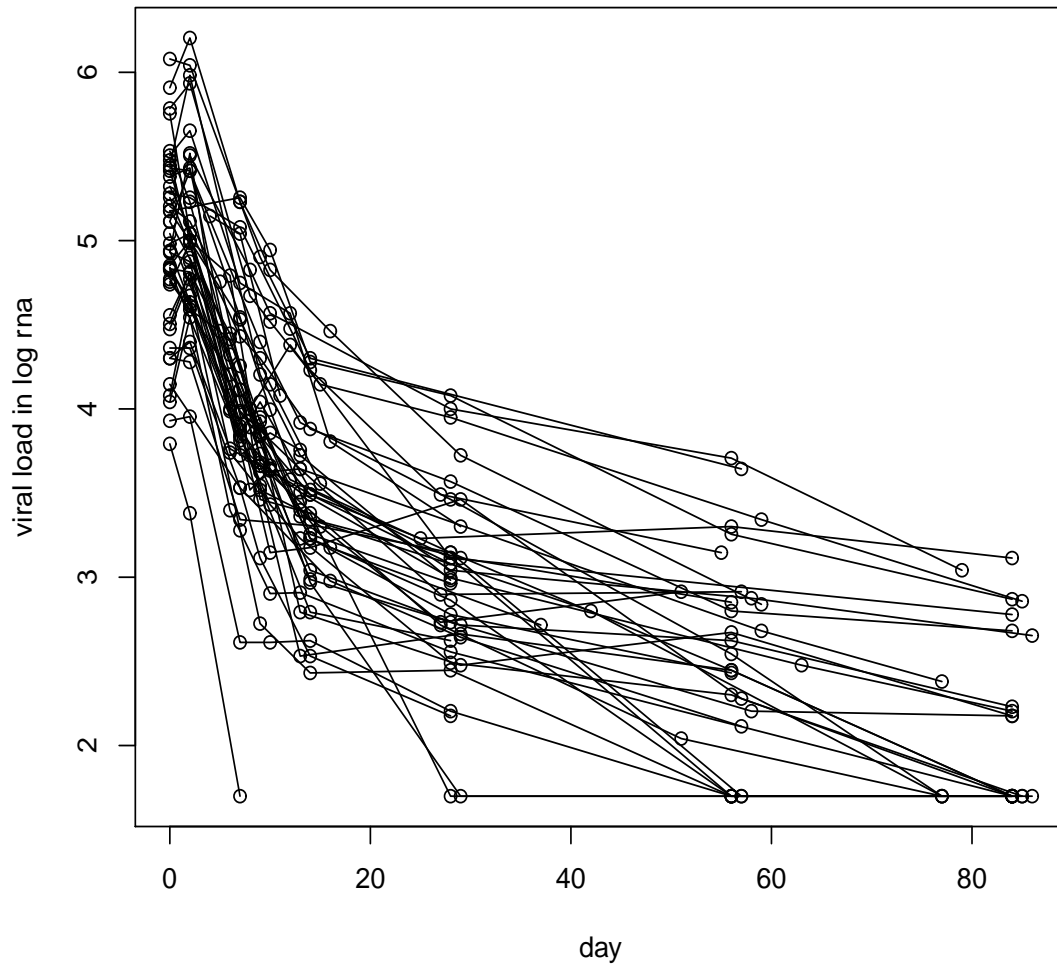


Figure 1.3: Trajectories for viral loads in an anti-HIV treatment study.

2 Multiple Imputation Methods for Multivariate One-sided Tests with Missing Data¹

2.1 Introduction

Multivariate one-sided hypotheses testing problems arise frequently in practice. For example, in clinical trials treatment effects may be measured by both efficacy and toxicity. Researchers may be interested in testing the null hypothesis H_0 that the treatment is not superior to the control in both endpoints versus the alternative hypothesis H_1 that the treatment is superior to the control in at least one of the endpoints. After appropriate transformation (Perlman and Wu, 2005), such a testing problem can be written as

$$H_0 : \mu_1 \leq 0 \text{ and } \mu_2 \leq 0 \quad \text{versus} \quad H_1 : \mu_1 > 0 \text{ or } \mu_2 > 0,$$

which is called a multivariate one-sided test. Various testing methods for multivariate one-sided testing problems have been proposed. Silvapulle and Sen (2004) provide an excellent review.

In practice, missing data are very common in multivariate datasets. In the presence

¹A version of this chapter has been accepted for publication. Wang, T. and Wu, L. (2011) Multiple Imputation Methods for Multivariate One-sided Tests with Missing Data.

of the missing data, the commonly used complete-case method, which discards all incomplete observations, may be very inefficient if the missing rate is not low and may even give biased results if the missing data are not missing completely at random (Little and Rubin, 2002). To our knowledge, no formal methods for multivariate one-sided testing problems are available in the presence of missing data. In this chapter we propose several testing methods for multivariate one-sided hypotheses with missing data based on multiple imputations. We assume that the missing data are MAR (missing at random) and the parameters in the missing mechanism model and the inference model are distinct. Under such “*ignorability*” assumption, the missing data mechanism can be safely ignored (Rubin 1987). We also allow the missing data patterns to be arbitrary, but with a similar missing rate on each variable. Note that we will focus on the balanced-data case: if the data were completely observed, all variables will have the same number of observations, as discussed at the end of Chapter 1. In other words, the following general $n \times p$ dataset with missing values will be considered, where NA denotes a missing value, p denotes the number of variables, and n denotes the sample size. More details will be provided in Section 2.2 and Section 2.3.

Variable	X_1	X_2	\cdots	X_p
Data	x_{11}	x_{12}	\cdots	NA
	x_{21}	x_{22}	\cdots	x_{2p}
	NA	x_{32}	\cdots	NA
			
	x_{n1}	NA	\cdots	x_{np}

Multiple Imputation (MI) methods have been widely used for missing data problems. A main advantage of MI methods is that complete-data methods can be used. In an MI method, each missing value is imputed by several (say m) simulated versions, where the simulations are drawn from a predictive distribution of the missing data given the observed data. This leads to m “complete datasets”. Each of the m “complete datasets” is analyzed by complete-data methods. These m “complete-data” analyses are then combined to produce an overall inference, which takes into account the uncertainty due to missing values. General theories and implementations of MI methods can be found in Rubin (1987).

For hypothesis testing problems with missing data, a testing procedure based on a MI method can either use individual sufficient statistics or use individual test statistics computed from the “complete datasets”. In either cases, it is not straightforward to combine these sufficient statistics or test statistics to produce an overall inference, and it can be challenging to derive the null distributions of the overall test statistic. Following

Li et.al. (1991) and Meng and Rubin (1992), in this chapter we propose several methods to appropriately combine individual sufficient statistics and individual test statistics for multivariate one-sided testing problems. We mainly focus on multivariate normal populations.

In Section 2.2, we give a brief review of MI methods. In Section 2.3, we propose three overall testing procedures based on MI: a Wald-type test by combining individual sufficient statistics and two tests by combining individual Wald and likelihood ratio (LR) test statistics, respectively. In Section 2.4 and 2.5, a real data example and a simulation study are presented. We conclude the article in Section 2.6 with some discussions.

2.2 Multiple Imputation Method

Consider a multivariate normal random variable $\mathbf{x} = (x_1, \dots, x_p)^T \sim N_p(\boldsymbol{\mu}, \Sigma)$, with $\boldsymbol{\theta} = (\boldsymbol{\mu}, \text{vec}(\Sigma))^T$ where $\text{vec}(\cdot)$ represents vectorization. Suppose that we have n such i.i.d. multivariate observations $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ with missing data. Let $\text{vec}(X) = (\mathbf{x}_{obs}, \mathbf{x}_{mis})^T$, where \mathbf{x}_{obs} and \mathbf{x}_{mis} denote the observed and missing components in X , respectively. We are interested in the multivariate one-sided hypothesis $H_0 : \boldsymbol{\mu} \in \mathbb{O}^-$ versus $H_1 : \boldsymbol{\mu} \in \mathbb{R}^p \setminus \mathbb{O}^-$ (i.e., the testing problem (1.9)). Figure 1.1 illustrates the null and alternative parameter spaces when $p = 2$.

In the presence of missing data, the complete-data tests may not be directly applicable. However, based on a multiple imputation method, we can simulate several plausible values of each missing value to obtain several “complete datasets”. Then,

we can conduct complete-data tests based on known methods and then combine the results in an appropriate way to incorporate the missing data uncertainty. For such an approach, there are two major issues: (i) how to generate the missing data so that the imputations are *proper* in the sense of Rubin (1987); and (ii) how to combine the complete-data analyses appropriately. Note that a simple approach to generate multiple imputations is to simulate \mathbf{x}_{mis} from the predictive distribution $f(\mathbf{x}_{mis}|\mathbf{x}_{obs}, \hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ is obtained from (say) a complete-case method or some other simple methods. However, such a simple approach generates *improper* multiple imputations because the uncertainty of the estimate $\hat{\boldsymbol{\theta}}$ is not incorporated.

Proper imputations can be generated from the predictive distribution of missing data given the observed data $f(\mathbf{x}_{mis}|\mathbf{x}_{obs})$, which can be implemented by the following Bayesian framework

$$f(\mathbf{x}_{mis}|\mathbf{x}_{obs}) = \int f(\mathbf{x}_{mis}|\mathbf{x}_{obs}, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta},$$

where $\pi(\boldsymbol{\theta})$ is an assumed prior for $\boldsymbol{\theta}$. Note that the missing data mechanism has been ignored from the above expression. This is due to the “ignorability” assumption: MAR and distinct parameters. Specifically, let $M_{n \times p}$ be the indicator matrix of whether a value is observed or missing and let $g(M|\mathbf{x}_{mis}, \mathbf{x}_{obs}, \boldsymbol{\phi})$ be the probability density function of M given the data and the model parameter(s) $\boldsymbol{\phi}$. A predictive distribution that incorporates the missing data mechanism M has the following form.

$$\begin{aligned} & f(\mathbf{x}_{mis}|\mathbf{x}_{obs}, M) \\ &= \frac{\int \int f(\mathbf{x}_{mis}, \mathbf{x}_{obs}|\boldsymbol{\theta})g(M|\mathbf{x}_{mis}, \mathbf{x}_{obs}, \boldsymbol{\phi})\pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi})d\boldsymbol{\theta}d\boldsymbol{\phi}}{\int \int \int f(\mathbf{x}'_{mis}, \mathbf{x}_{obs}|\boldsymbol{\theta})g(M|\mathbf{x}'_{mis}, \mathbf{x}_{obs}, \boldsymbol{\phi})\pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi})d\boldsymbol{\theta}d\boldsymbol{\phi}d\mathbf{x}'_{mis}} \end{aligned} \quad (2.1)$$

where $\pi(\boldsymbol{\phi})$ is an assumed prior for $\boldsymbol{\phi}$. By noting the MAR assumption $g(M|\mathbf{x}_{mis}, \mathbf{x}_{obs}, \boldsymbol{\phi}) = g(M|\mathbf{x}_{obs}, \boldsymbol{\phi})$, and the assumption that $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are distinct, equation (2.1) becomes

$$\begin{aligned} & \frac{\int f(\mathbf{x}_{mis}, \mathbf{x}_{obs}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \cdot \int g(M|\mathbf{x}_{obs}, \boldsymbol{\phi})\pi(\boldsymbol{\phi})d\boldsymbol{\phi}}{\int \int f(\mathbf{x}'_{mis}, \mathbf{x}_{obs}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}d\mathbf{x}'_{mis} \cdot \int g(M|\mathbf{x}_{obs}, \boldsymbol{\phi})\pi(\boldsymbol{\phi})d\boldsymbol{\phi}} \\ &= f(\mathbf{x}_{mis}|\mathbf{x}_{obs}). \end{aligned}$$

To generate proper multiple imputations from $f(\mathbf{x}_{mis}|\mathbf{x}_{obs})$, the implementation can be based on MCMC methods such as a Gibbs sampler as follows. Beginning with a plausible starting value $\boldsymbol{\theta}^{(0)}$, we repeat the following two steps (i) and (ii) for a burn-in period (a large number of times, say 1000). Specifically, at iteration $t + 1$ ($t = 0, 1, 2, \dots$), given the values of parameters $\boldsymbol{\theta}^{(t)}$ and missing data $\mathbf{x}_{mis}^{(t)}$ from the previous iteration t , we proceed with the following two steps:

- (i) Draw new values of the missing data $\mathbf{x}_{mis}^{(t+1)}$ from the posterior $f(\mathbf{x}_{mis}|\boldsymbol{\theta}^{(t)}, \mathbf{x}_{obs})$;
- (ii) Draw new values of the parameters $\boldsymbol{\theta}^{(t+1)} = (\boldsymbol{\mu}^{(t+1)}, \boldsymbol{\Sigma}^{(t+1)})^T$ from the posterior $f(\boldsymbol{\theta}|\mathbf{x}_{mis}^{(t+1)}, \mathbf{x}_{obs})$.

This yields a stochastic chain (sequence) $\{(\boldsymbol{\theta}^{(t)}, \mathbf{x}_{mis}^{(t)}) : t = 0, 1, 2, \dots\}$, whose stationary distribution, after the burn-in period, is $f(\boldsymbol{\theta}, \mathbf{x}_{mis}|\mathbf{x}_{obs})$. The stationary distribution for the subsequence $\{\mathbf{x}_{mis}^{(t)} : t = 0, 1, 2, \dots\}$ is the desired $f(\mathbf{x}_{mis}|\mathbf{x}_{obs})$. The adequacy of the length of the burn-in period can be diagnosed by the time series plot of the draws of the parameters over iteration times. If the plot shows no obvious patterns or trends, the convergence is likely to be achieved after the burn-in period. Note that the multiple

imputations can be obtained either from multiple chains or from multiple draws of a single chain separated by as many iterations as a burn-in period. To implement the above Gibbs sampler in our study, first we assume a non-informative prior distribution for $\boldsymbol{\theta}$ as follows:

$$\boldsymbol{\theta} \sim |\Sigma|^{-\frac{p+1}{2}},$$

Based on Bayes theorem we obtain the posterior distributions in Step (i) as follows:

$$\begin{aligned} \Sigma & \mid \mathbf{x}_{mis}, \mathbf{x}_{obs} \sim W^{-1}(n-1, ((n-1)S)^{-1}), \\ \boldsymbol{\mu} & \mid \Sigma, \mathbf{x}_{mis}, \mathbf{x}_{obs} \sim N(\bar{X}, n^{-1}\Sigma), \end{aligned}$$

where S is the sample covariance matrix based on X . To simulate the missing data, the posterior distribution in Step (ii) is given by the standard conditional distribution results for the multivariate normal distribution. This Gibbs sampler can be implemented by using the “missing data” library in Splus.

After generating multiple imputations for the missing values, we have several “complete datasets”. For each “complete dataset”, standard complete-data methods can be used, and the results are then combined to obtain an overall inference. For hypothesis testing problems, there are different ways to combine complete-data analysis results:

- one may either *combine complete-data parameter estimates* (or combine sufficient statistics) and then propose a single test statistic based on the combined estimates, or
- one may *combine complete-data test statistics* and then propose an overall test.

These two approaches are discussed in details in the following sections.

2.3 One-Sided Tests based on Multiple Imputations

2.3.1 Tests Based on Combined Parameter Estimates

In the presence of missing data, we can generate m proper multiple imputations from $f(\mathbf{x}_{mis}|\mathbf{x}_{obs})$ to obtain m “complete datasets”. From each of these m “complete datasets”, we obtain estimates of the mean vector and the associated variance-covariance matrix. Then we *combine* these m parameter estimates to obtain an overall estimate of the mean vector and associated variance-covariance matrix, respectively. Finally, we propose a single testing procedure based on the combined estimates and incorporate the missing data uncertainty. We consider the commonly used unbiased estimates for a multivariate normal distribution in this section. This approach is described as follows.

Still consider n i.i.d. observations $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ from the multivariate normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$. Let $\hat{\boldsymbol{\mu}}$ be the unbiased estimate of $\boldsymbol{\mu}$, let U_n be the variance-covariance matrix of $\hat{\boldsymbol{\mu}}$, and let \hat{U}_n denote the unbiased estimate of U_n . When X is completely observed, we have

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}, \tag{2.2}$$

$$\begin{aligned} U_n &= \text{Var}(\hat{\boldsymbol{\mu}}) = \frac{1}{n}\Sigma, \\ \hat{U}_n &= \frac{1}{n}S = \frac{1}{n} \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T. \end{aligned} \tag{2.3}$$

In order to make valid statistical inference based on proper multiple imputations, fol-

lowing Rubin (1987), Li et al. (1991), Meng and Rubin (1992) and Rubin (1996), the complete-data estimates $\hat{\boldsymbol{\mu}}$ and \hat{U}_n should be asymptotically unbiased for $\boldsymbol{\mu}$ and U_n . Also, when n is large the following normal approximation should hold

$$\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, U_n).$$

We can see that the above conditions are satisfied by the estimates (2.2) and (2.3), based on the standard results of multivariate normal distributions. For more complicated models, maximum likelihood estimates (MLE) in most of the cases satisfy the above large-sample conditions.

When X contains missing values, let $\{X_{*l}, l = 1, 2, \dots, m\}$ be the m multiply imputed datasets. Correspondingly we have “complete-data” estimates (2.2) and (2.3) for each of these “complete datasets”, denoted by $\hat{\boldsymbol{\mu}}_{*l}$ and \hat{U}_{n*l} . We combine these m sets of estimates in a way similar to Li et al. (1991) as follows

$$\bar{\boldsymbol{\mu}}_m = \frac{1}{m} \sum_{l=1}^m \hat{\boldsymbol{\mu}}_{*l}, \quad (2.4)$$

$$\bar{U}_m = \frac{1}{m} \sum_{l=1}^m \hat{U}_{n*l},$$

$$\bar{B}_m = \frac{1}{m-1} \sum_{l=1}^m (\hat{\boldsymbol{\mu}}_{*l} - \bar{\boldsymbol{\mu}}_m)(\hat{\boldsymbol{\mu}}_{*l} - \bar{\boldsymbol{\mu}}_m)^T, \quad (2.5)$$

where $\bar{\boldsymbol{\mu}}_m$ is an unbiased estimator of $\boldsymbol{\mu}$, \bar{U}_m estimates the within-imputation covariance U_n , and \bar{B}_m estimates B_n , with B_n denoting the unknown between-imputation variation of $\hat{\boldsymbol{\mu}}_{*l}$, due to the *missing data uncertainty*. Following Rubin (1987), Li et al. (1991) and Meng and Rubin (1992), we have that when n is large

$$\bar{\boldsymbol{\mu}}_m \sim N(\boldsymbol{\mu}, \Sigma_{n,m}), \quad (2.6)$$

where

$$\Sigma_{n,m} = U_n + (1 + m^{-1})B_n. \quad (2.7)$$

Based on the combined statistics (2.4) – (2.5) for the mean vector and the covariance matrix, we may construct a single Wald-type test statistic. The main challenge is to efficiently estimate the covariance matrix $\Sigma_{n,m}$ in (2.7). Based on Rubin (1987) among others, \bar{U}_m is a consistent estimator for U_n . However, although \bar{B}_m may look like a natural choice to estimate B_n , it can be very inefficient with small m (say 3 - 5). In fact, the estimation methods and two-sided hypothesis tests based on \bar{B}_m have been studied by Li (1985), Rubin (1987) and Raghunathan (1987), where it has been shown that for small m these methods are inferior to those based on the assumption of equal ratios of missing to observed information, i.e. $B_n \propto U_n$. Specifically, the ratios are given by the eigenvalues of matrix $U_n^{-1}B_n$, which are denoted by $(\lambda_1, \dots, \lambda_p) \in [0, \infty)^p$. Then the assumption suggests that all $\lambda_i = \bar{\lambda} = \frac{1}{p} \sum_{i=1}^p \lambda_i$, or equivalently $B_n = \bar{\lambda}U_n$. Thus, $\Sigma_{n,m} = U_n + (1 + m^{-1})\bar{\lambda}U_n$. Following Johnson and Kotz (1973), among others, we know that $\sum_{i=1}^p \lambda_i = \text{tr}(U_n^{-1}B_n)$. Therefore, the following estimate of $\Sigma_{n,m}$ is proposed (Li et al. 1991 and Meng and Rubin 1992)

$$\hat{\Sigma}_{n,m} = \bar{U}_m + r_m \bar{U}_m, \quad (2.8)$$

where $r_m = (1 + m^{-1})\text{tr}(\bar{B}_m \bar{U}_m^{-1})/p$. Note that $\hat{\Sigma}_{n,m}$ incorporates the missing data uncertainty. In practice, this theoretical assumption is often difficult to be checked, because B_n is usually unknown and \bar{B}_m can be very inefficient with small m (as discussed above). For the simulation study in Section 2.5, we consider the missing rates

being similar in all variables that have missing values (as shown by the example in Section 2.1) and the correlation structure being exchangeable to approximate the situation where the assumption is satisfied. In summary, this equal ratios assumption is made mainly for obtaining an estimate of B_n that can efficiently extract information from the imputation datasets when m is small, and we will show in the simulation study that our methods are robust against modest departures from the assumption.

Let $\pi_{\hat{\Sigma}_{n,m}}(\bar{\boldsymbol{\mu}}_m; \mathbb{O}^-)$ be the orthogonal projection of the overall estimate $\bar{\boldsymbol{\mu}}_m$ onto the null parameter space \mathbb{O}^- with respect to (w.r.t.) $\hat{\Sigma}_{n,m}$. We propose the following single (overall) Wald-type test statistic

$$T_W \equiv \left\| \bar{\boldsymbol{\mu}}_m - \pi_{\hat{\Sigma}_{n,m}}(\bar{\boldsymbol{\mu}}_m; \mathbb{O}^-) \right\|_{\hat{\Sigma}_{n,m}}^2 = \min_{\boldsymbol{\mu} \in \mathbb{O}^-} \{(\bar{\boldsymbol{\mu}}_m - \boldsymbol{\mu})^T \hat{\Sigma}_{n,m}^{-1} (\bar{\boldsymbol{\mu}}_m - \boldsymbol{\mu})\} \quad (2.9)$$

which measures the “distance” from the estimate $\bar{\boldsymbol{\mu}}_m$ to the null parameter space \mathbb{O}^- w.r.t. the estimated covariance matrix $\hat{\Sigma}_{n,m}$. The larger the “distance”, the stronger the evidence against the null hypothesis $H_0 : \boldsymbol{\mu} \in \mathbb{O}^-$. Note that the overall test statistic T_W incorporates the missing data uncertainty through the estimated variance-covariance matrix $\hat{\Sigma}_{n,m}$. The null distribution of T_W is given by the following theorem.

Theorem 2.1 Consider the multivariate one-sided hypotheses (1.9) and multivariate normally distributed dataset X with missing values. For the proposed overall test statistic T_W given in (2.9), assuming equal ratios of missing to observed information, i.e. $B_n \propto U_n$, the following results hold:

(i) The least favorable null distribution of T_W can be approximated by the following

$\bar{\chi}^2$ distribution at $\boldsymbol{\mu} = \mathbf{0}$,

$$\Pr(T_W \leq c | \boldsymbol{\mu} = \mathbf{0}) \approx \sum_{i=0}^p \omega_{p-i}(p, \hat{\Sigma}_{n,m}, \mathbb{O}^-) \Pr(\chi_i^2 \leq c).$$

(ii) The tail probability of the above null distribution has the following lower and upper bounds:

$$\begin{aligned} \frac{\Pr(\chi_0^2 \geq c) + \Pr(\chi_1^2 \geq c)}{2} &\leq \sum_{i=0}^p \omega_{p-i}(p, \hat{\Sigma}_{n,m}, \mathbb{O}^-) \Pr(\chi_i^2 \leq c) \\ &\leq \frac{\Pr(\chi_{p-1}^2 \geq c) + \Pr(\chi_p^2 \geq c)}{2}, \end{aligned} \quad (2.10)$$

where χ_0^2 denotes the distribution that takes the value zero with probability one.

Proof of Theorem 2.1 is given in the Appendix.

Note that the lower and upper bounds in (2.10) are general bounds for all $\bar{\chi}^2$ distributions (Silvapulle and Sen, 2004), so tests based on these bounds may be conservative, but they are computationally simple.

Note also that, throughout this chapter, we assume the assumption of equal ratios of missing to observed information. Li et al. (1991) and Meng and Rubin (1992) have shown that the two-sided testing results are robust against modest departure of this assumption. We will also do some simulation studies to evaluate the influence of violations of this assumption to our one-sided test in Section 2.5.

2.3.2 Tests Based on Combining Test Statistics

The overall test statistic in Section 2.3.1 is obtained by combining parameter estimates of the mean vector $\boldsymbol{\mu}$ and the associated covariance matrix U_n from the multiply im-

puted “complete datasets”. In this section we propose overall tests which combine individual complete-data test statistics rather than parameter estimates. We first consider combining Wald-type test statistics and then we consider combining LRT statistics.

Combining Wald-type test statistics

Rubin (1987) considered combining individual test statistics for the two-sided hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$. However, those methods may not be easily extended to our one-sided hypotheses (1.9), because in our case the null parameter space is not just a single point, and both the null and alternative parameter spaces are constrained. Meng and Rubin (1992) proposed to use the combined estimates rather than the individual estimates in the expressions of the individual test statistics, attempting to obtain more efficient procedures than those in Rubin (1987). Motivated by Meng and Rubin (1992), we propose to combine the following individual Wald-type test statistics based on the l -th “complete dataset”

$$W_{*l} = \left(\hat{\boldsymbol{\mu}}_{*l} - \pi_{\hat{\Sigma}_{n,m}}(\bar{\boldsymbol{\mu}}_m; \mathbb{O}^-) \right)^T \hat{U}_{n*l}^{-1} \left(\hat{\boldsymbol{\mu}}_{*l} - \pi_{\hat{\Sigma}_{n,m}}(\bar{\boldsymbol{\mu}}_m; \mathbb{O}^-) \right), \quad l = 1, 2, \dots, m.$$

To combine these individual test statistics W_{*l} 's, a simple average without any adjustments, such as $\bar{W}_m = \sum_{i=1}^m W_{*i}/m$, may not properly incorporate the variability due to missing data. Moreover, the null distribution of the simple average \bar{W}_m may not be easily obtained. Motivated by Meng and Rubin (1992), here we propose to combine the individual test statistics W_{*l} as follows

$$T_{CW} = \frac{\bar{W}_m - \frac{(m-1)p}{m+1} r_m}{1 + r_m}, \quad (2.11)$$

where r_m is the estimated ratio of missing to observed information, as defined in (2.8). T_{CW} takes into account the missing data uncertainty, measured by the between-imputation variance B_n , by incorporating the quantity r_m , which is the estimator for the ratio of B_n and U_n .

To find the null distribution of T_{CW} , we have the following result.

Theorem 2.2 Following the same assumptions as in Theorem 2.1, we have $T_W = T_{CW} + o_p(1)$.

By Theorem 2.2, we know that the asymptotic null distribution of T_{CW} is the same as that of T_W in (2.9), and thus is determined by equations in Theorem 2.1. The proof for Theorem 2.2 is given in the Appendix.

Combining LR-type test statistics

Similarly, motivated by Meng and Rubin (1992), we can combine individual LR (likelihood ratio) type statistics to obtain an overall test. For a multivariate normal distribution and the hypothesis testing problem (1.9), the commonly used LRT has the same form as the Wald-type test statistic based on Perlman (1969). Therefore, the procedure for combining the individual LR statistics is the same as the one for combining individual Wald-type statistics. It follows in our study that the individual $R_{*l} = W_{*l}$ and the combined test statistic $T_{LR} = T_{CW}$. Therefore, the asymptotic null distribution of T_{LR} is the same as that of T_{CW} and T_W , and is determined by equations in Theorem 2.1. However, this property is not likely to remain true in more complex cases. Therefore after we first develop T_{LR} for the multivariate normal model,

we can further investigate its properties and performances in more complex models in the future.

2.4 Real Data Example

We consider a longitudinal study on the change of mental distress of parents whose children died by accident, suicide or homicide (Murphy et al. 2003). The mental distress was measured in terms of depression (DEP), anxiety (ANX), hostility (HOS), somatization (SOM) and global severity index (GSI) for parents at baseline, the end of 3 months, 6 months and 18 months after their children's deaths. We consider only the parents in the treatment group and obtain a sample of size $n = 156$. Let y_1, y_2, y_3 and y_4 denote the distress measurements at month 0 (baseline), 3, 6 and 18, respectively. We are interested in testing whether the changing rate (defined by change per month) of distress measurements consistently decreases during the three time spans (0-3 month, 3-6 month, 6-18 month).

Specifically, let $\delta_i = E(y_i)$, $i = 1, 2, 3, 4$. The hypotheses of interest are $H_0 : (\delta_2 - \delta_1)/3 \geq (\delta_3 - \delta_2)/3 \geq (\delta_4 - \delta_3)/12$ v.s. $H_1 : \text{not } H_0$. Let $x_1 = (y_3 - y_2) - (y_2 - y_1)$, $x_2 = (y_4 - y_3)/4 - (y_3 - y_2)$, and $\boldsymbol{\mu} = (\mu_1, \mu_2)^T = ((\delta_3 - \delta_2) - (\delta_2 - \delta_1), (\delta_4 - \delta_3)/4 - (\delta_3 - \delta_2))^T$. Then the testing problem becomes $H_0 : \boldsymbol{\mu} \in \mathbb{O}^-$ v.s. $H_1 : \boldsymbol{\mu} \in \mathbb{R}^p \setminus \mathbb{O}^-$ with $p = 2$. Figure 2.1 shows the trajectories of five variables for six randomly selected subjects, and the last graph shows the mean scores of the whole sample. We can see that in most cases the distress scores dropped from baseline to 18-month after the accidents. The decreasing trends during the three time spans seem to get flatter, which is the

opposite of H_0 .

Since there are missing values in the data (see Table 2.1), we analyze the data using the proposed multiple imputation methods, and compare the results with the naive complete-case (CC) method which discards all incomplete observations. We consider the Wald-type test T_W in Section 2.3.1 and the combined test T_{CW} in Section 2.3.2. To obtain null distributions, Theorem 2.1 shows that one can either use the “*substitution method*” to approximate the null distribution or the “*bound method*” (i.e., use the upper and lower bounds of the null probabilities). We consider both methods.

The analysis results are shown in Table 2.1. We see that our new tests produce smaller p-values than the CC method in most of the cases, which indicates the better ability to detect the true difference. Specifically, while the new tests rejected H_0 in the following cases, the CC method failed to reject H_0 for SOM (substitution method) at significance level 0.01, for GSI and SOM (bound method) at significance level 0.05, and for HOS (substitution method) at significance level 0.1. According to the results from our new tests, we can conclude that the changing trends are not consistently moving towards the decreasing direction in all 5 outcomes at different significance levels, which evidenced our observation earlier from the graphs.

We also see that the p-values of the substitution method are smaller than the corresponding p-values of the bound method. This is because the bound test is conservative due to the use of upper and lower bounds in computing the p-values. Since T_W and T_{CW} are approximately equal, they give similar results for this dataset as the sample size is reasonably large.

Note that in this dataset, the equal ratios of missing to observed information assumption is not likely to hold, as the missing rates are very different between variables (around 18% for x_1 and 45% for x_2). In next section, we will conduct a simulation study to evaluate the performances of these methods, as well as the influence of violations of the above equal ratios assumption.

2.5 A Simulation Study

In this section we perform a simulation study to compare the new tests T_W and T_{CW} with the naive CC method for the one-sided hypotheses testing problem (1.9). To show the difference of our proposed tests and the MI testing method proposed by Li et al. (1991), which is used for two-sided hypotheses ($H_0 : \boldsymbol{\mu} = \mathbf{0}$ v.s. $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$) and uses the following test statistic $D_m = \overline{\boldsymbol{\mu}}_m^T \overline{U}_m^{-1} \overline{\boldsymbol{\mu}}_m / [p(1 + r_m)]$, we include it in the simulation for $p = 2$ case for comparison purpose. Since under the normality assumption the combined LR test and the combined Wald test are very similar, the combined LR test is excluded from comparison. For the one-sided tests under consideration, both the bound method and the substitution method are applied. The missing data are mainly generated by the MCAR mechanism, but some simulations by the MAR mechanism are performed as well. Also note that the major part of our simulation study is under the MCAR and equal ratios assumption of missing to observed information. In order to check whether our tests are robust against violations of this assumption, we conduct a small simulation study for $p = 2$ case with unequal missing rates at the end of this section.

The data are simulated from a p -variate normal population $N(\boldsymbol{\mu}, \mathbf{V})$, with $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^T$ and \mathbf{V} being the variance-covariance matrix whose diagonal elements are all 1's and off-diagonal elements are all ρ . We consider the cases of $p=2$ and 5, sample sizes of $n = 50$ and 200, nominal significance level $\alpha = 0.05$, same missing data rate 20% for all variables (to approximate the situation where the equal ratios assumption of missing to observed information is true), and various choices of the mean vector $\boldsymbol{\mu}$ (under null and alternative parameter spaces) and the value of ρ . Simulated complete data are randomly deleted to create “missing values”, so the missing data are missing completely at random. For the two new tests based on multiple imputations, the number of imputations is $m = 5$. The imputations are created using the method described in Section 2.2, via a Gibbs Sampler with noninformative priors. The burn-in period for the Gibbs sampler consists of 1000 iterations. Figure 2.2 shows the time series plot of the 1000 draws of the parameters $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_5)^T$ when $p = 5$. The five graphs show that the convergence is achieved by 1000 iterations.

Tables 2.2, 2.3, 2.6 and 2.7 show the simulation results based on 1000 replications. We see that all three tests attain the nominal level $\alpha = 5\%$ in most of the cases. However, for the substitution method, which replaces the unknown covariance matrix $\Sigma_{n,m}$ by its estimate $\hat{\Sigma}_{n,m}$, the one-sided tests become somewhat liberal when the sample size is not large ($n = 50$) since the sizes of the tests are slightly larger than the nominal level. This is probably due to the fact that the covariance matrix $\Sigma_{n,m}$ may not be estimated well for small sample sizes. This problem disappears for large sample sizes ($n = 200$) since in this case the covariance matrix can be better estimated. The

two new tests T_W and T_{CW} have smaller Type I error rates and are more powerful than the naive CC method, especially when $p = 5$. The performances of T_W and T_{CW} appear to be very similar. The naive CC method, which discards all incomplete observations, is still valid since the missing data are missing completely at random but it loses power, especially when $p = 5$. Also note that in Table 2.2, Li et al.'s D_m test gives very liberal type I errors. This is because it does not take into account the whole one-sided null parameter space (i.e., it only uses one point in the null space: $\boldsymbol{\mu} = \mathbf{0}$). Simulations on multivariate normal distributions with unexchangeable correlation structures are also done for $p = 5$. This corresponds to the situation where the assumption of equal ratios of missing to observed information is likely to be violated even though the missing rates are the same across variables. Specifically, an AR(1) correlation structure is considered with $\rho = 0.5$. The results in Table 2.8 show similar patterns with those in Table 2.6 and 2.7 with the exchangeable correlation structure. In conclusion, the proposed two new tests T_W and T_{CW} perform well and are better than the CC method in the sense that they have smaller Type I errors and are more powerful (or smaller Type II errors). The new tests are also more appropriate for the multivariate one-sided hypothesis testing problems than Li et al.'s D_m test which ignores the one-sided null parameter space.

Table 2.4 shows the simulation results based on 1000 replications when the assumption of equal ratios of missing to observed information is violated. We consider $p = 2$, $\rho = 0.6$ and the sample size $n = 200$. We first choose missing rates for x_1 and x_2 to be modestly different, i.e. 10% and 25%. We also simulate a case with similar missing rates as in the mental distress example in Section 2.4, i.e. 20% and 40%. At

last, we consider a more extreme case where the missing rates are 10% and 40%. We see that all tests attain the nominal significance level 0.05 for all missing rates, and our new tests T_W and T_{CW} have smaller type I error rates and are more powerful than the naive CC methods. Moreover, we see that the powers under the modest departure from the assumption (10% and 25%) are similar with those in Tables 2.2 and 2.3 where the assumption is satisfied. Even for the more extreme missing rates 10% and 40%, there are no dramatic decreases in power rates. In conclusion, the proposed new tests are robust against modest violations to the equal ratios assumption.

Since all of the above missing data are MCAR, some simulations based on MAR mechanism are also done for $p = 2$. Specifically, x_1 is still MCAR with 20% missing rate, but the missingness of x_2 depends on the observed values of x_1 . With the observed value of x_1 in different regions, x_2 will be missing at different rates. The regions are carefully chosen so that the expected missing rate of x_2 is also 20%. Table 2.5 shows the results based on 1000 iterations. We can see that the proposed two new tests T_W and T_{CW} have smaller Type I errors and are more powerful than the CC method, which is consistent with the MCAR results.

2.6 Discussion

Another commonly used approach for missing data is based on the likelihood method. Multivariate one-sided testing methods can be developed based on the observed-data likelihood. This method is under investigation and the results will be reported separately. A main advantage of multiple imputation methods is that existing complete-

data methods can be used with the imputed datasets, and various complete-data hypothesis testing and data analyses can be performed on the imputed datasets.

When the missing data are informative or non-ignorable, one must assume a model for the missing data mechanism and incorporate the missing data model in the imputation models or in the likelihood. We will study this problem in the near future.

Although we have focused on testing the one-sided hypotheses where the null parameter space is the negative orthant, the methods proposed in this chapter are quite general and can be applied to other constrained testing problems. The methods can also be extended to regression problems in which the mean parameters may be constrained such as positiveness. Such an extension is currently under investigation.

Table 2.1: P-values of the tests for the mental distress data

Variable	T_W	T_{CW}	CC
Bound method			
GSI	.03	.03	.052
HOS	.12	.12	.27
ANX	.03	.03	.04
SOM	.007	.007	.055
DEP	.002	.002	.001
Substitution method			
GSI	.03	.03	.04
HOS	.092	.098	.22
ANX	.02	.02	.04
SOM	.005	.004	.04
DEP	.002	.002	.001

Table 2.2: Simulation results for $p=2$ based on the bound method (equal missing rate = 20%)

μ	$\Sigma_1 (\rho = 0)$			$\Sigma_2 (\rho = 0.2)$			$\Sigma_3 (\rho = 0.6)$			D_m
	T_W	T_{CW}	CC	T_W	T_{CW}	CC	T_W	T_{CW}	CC	
Type I Error Rates (in %), $n = 50$										
(0,0)	3.6	3.8	3.6	4.4	4.5	4.8	2.7	3.2	3.6	6.9
(-2,0)	1.3	1.4	1.4	1.6	1.6	2.1	1.2	1.3	1.8	100
Type I Error Rates (in %), $n = 200$										
(0,0)	3.3	3.3	4.0	2.9	2.9	3.5	2.3	2.3	3.0	5.9
(-2,0)	0.9	1.0	1.4	1.7	1.7	1.4	1.3	1.3	1.4	100
Power Comparisons (in %), $n = 50$										
(0.2,0.2)	45	45	40	42	42	37	29	29	23	24
(-1,0.2)	19	19	15	21	21	15	16	17	15	100
Power Comparisons (in %), $n = 200$										
(0.2,0.2)	92	92	87	87	87	81	77	77	69	69
(-1,0.2)	58	58	51	59	59	51	59	59	52	100

Table 2.3: Simulation results for p=2 based on the substitution method (equal missing rate = 20%)

μ	$\Sigma_1 (\rho = 0)$			$\Sigma_2 (\rho = 0.2)$			$\Sigma_3 (\rho = 0.6)$		
	T_W	T_{CW}	CC	T_W	T_{CW}	CC	T_W	T_{CW}	CC
Type I Error Rates (in %), n=50									
(0,0)	5.8	5.9	6.4	6.4	6.4	6.8	5.5	5.4	5.8
(-2,0)	1.9	2.0	2.9	2.8	2.8	2.9	2.5	2.4	2.8
Type I Error Rates (in %), n=200									
(0,0)	5.5	5.5	5.6	5.4	5.4	5.0	4.8	4.9	5.4
(-2,0)	2.1	2.1	1.7	2.6	2.7	2.2	2.9	2.9	2.6
Power Comparisons (in %), n=50									
(0.2,0.2)	56	56	48	52	52	47	41	41	35
(-1,0.2)	26	26	21	26	27	23	25	26	23
Power Comparisons (in %), n=200									
(0.2,0.2)	94	94	91	92	91	87	86	86	79
(-1,0.2)	66	66	57	68	68	61	70	71	64

Table 2.4: Simulation results for $p=2$ under unequal missing rates ($n=200$, $\rho = 0.6$)

μ	Missing rates	Bound method			Substitution method		
		T_W	T_{CW}	CC	T_W	T_{CW}	CC
Type I Error Rates (in %)							
(0,0)	(10%, 25%)	2.1	2.1	2.9	4.1	4.2	5.0
	(20%, 40%)	1.8	1.9	2.5	4.4	4.4	5.1
	(10%, 40%)	2.5	2.6	3.3	5.0	5.0	5.9
Power Comparisons (in %)							
(-1,0.2)	(10%, 25%)	60	60	54	72	72	67
	(20%, 40%)	48	48	38	63	64	52
	(10%, 40%)	52	53	44	67	67	58

Note: (10%, 25%) denotes the missing rates for x_1 and x_2 , respectively.

Table 2.5: Simulation results for p=2 under MAR (n=200, equal missing rate=20%)

$\boldsymbol{\mu}$	Σ	Bound method			Substitution method		
		T_W	T_{CW}	CC	T_W	T_{CW}	CC
Type I Error Rates (in %)							
(0,0)	$\Sigma_1(\rho = 0)$	3.7	3.7	3.5	5.4	5.5	5.5
	$\Sigma_2(\rho = 0.6)$	1.9	1.9	2.4	4.6	4.6	5.7
(-2,0)	$\Sigma_1(\rho = 0)$	1.3	1.5	1.2	2.7	2.8	2.1
	$\Sigma_2(\rho = 0.6)$	1.2	1.2	1.5	2.8	2.8	2.9
Power Comparisons (in %)							
(-1,0.2)	$\Sigma_1(\rho = 0)$	60	60	53	69	69	61
	$\Sigma_2(\rho = 0.6)$	60	60	51	73	73	65
(0.2,0.2)	$\Sigma_1(\rho = 0)$	92	92	87	95	95	91
	$\Sigma_2(\rho = 0.6)$	78	78	68	88	88	79

Table 2.6: Simulation results for $p=5$ based on the bound method ($n=50$, equal missing rate = 20%)

μ	$\Sigma_1 (\rho = 0)$			$\Sigma_2 (\rho = 0.2)$			$\Sigma_3 (\rho = 0.6)$		
	T_W	T_{CW}	CC	T_W	T_{CW}	CC	T_W	T_{CW}	CC
Type I Error Rates (in %)									
(0^5)	2.7	3.2	5.7	1.5	1.9	3.2	0.5	0.7	1.2
$(-1, 0^4)$	1.7	1.9	3.3	0.9	1.6	2.9	0.3	0.4	1.0
$(-1^2, 0^3)$	0.7	1.1	1.7	0.6	0.8	1.8	0.1	0.1	1.1
$(-1^3, 0^2)$	0.4	0.5	0.5	0.4	0.4	0.7	0.1	0.1	0.8
$(-1^4, 0)$	0.2	0.2	0.3	0.1	0.1	0.4	0.0	0.0	0.3
Power Comparisons (in %)									
(0.1^5)	24	27	19	11	12	11	4	4	6
$(-1, .3^4)$	89	90	60	70	72	38	28	30	18
$(-1^2, .3^3)$	72	74	38	52	54	27	22	24	13
$(-1^3, .3^2)$	41	45	18	33	35	17	15	17	11
$(-1^4, .3)$	12	14	5	13	14	6	12	13	6

Note: $(-1^2, .3^3)$ denotes $(-1, -1, 0.3, 0.3, 0.3)$, etc.

Table 2.7: Simulation results for $p=5$ based on the substitution method ($n=50$, equal missing rate = 20%)

μ	$\Sigma_1 (\rho = 0)$			$\Sigma_2 (\rho = 0.2)$			$\Sigma_3 (\rho = 0.6)$		
	T_W	T_{CW}	CC	T_W	T_{CW}	CC	T_W	T_{CW}	CC
Type I Error Rates (in %)									
(0^5)	6.1	7.1	11	6.4	7.1	10	4.1	4.7	8
$(-1, 0^4)$	4.5	5.0	7.8	4.8	5.3	7.6	3.3	3.9	6.9
$(-1^2, 0^3)$	2.6	2.8	5.2	3.5	3.8	5.1	2.8	3.5	6.4
$(-1^3, 0^2)$	1.8	1.8	2.0	1.8	2.1	3.2	1.3	1.6	4.8
$(-1^4, 0)$	0.5	0.5	0.6	0.7	1.0	1.8	0.1	0.1	0.7
Power Comparisons (in %)									
(0.1^5)	41	43	32	29	30	24	17	19	18
$(-1, .3^4)$	97	97	76	90	91	63	67	68	44
$(-1^2, .3^3)$	86	88	56	78	79	49	60	63	39
$(-1^3, .3^2)$	61	64	33	57	60	33	50	53	31
$(-1^4, .3)$	22	24	10	28	29	15	15	17	9

Note: $(-1^2, .3^3)$ denotes $(-1, -1, 0.3, 0.3, 0.3)$, etc.

Table 2.8: Simulation results for $p=5$ based on AR(1) correlations ($n=50$, equal missing rate = 20%)

	Bound method			Substitution method		
$\boldsymbol{\mu}$	T_W	T_{CW}	CC	T_W	T_{CW}	CC
Type I Error Rates (in %)						
(0^5)	5.3	6.1	11	0.9	1.0	3.3
$(-1^4, 0)$	0.4	0.4	1.8	0	0	0.5
Power Comparisons (in %)						
(0.1^5)	23	25	24	9.2	10	10
$(-1^4, .3)$	28	31	17	8.3	11	5.5

Note: $(-1^4, .3)$ denotes $(-1, -1, -1, -1, 0.3)$, etc.

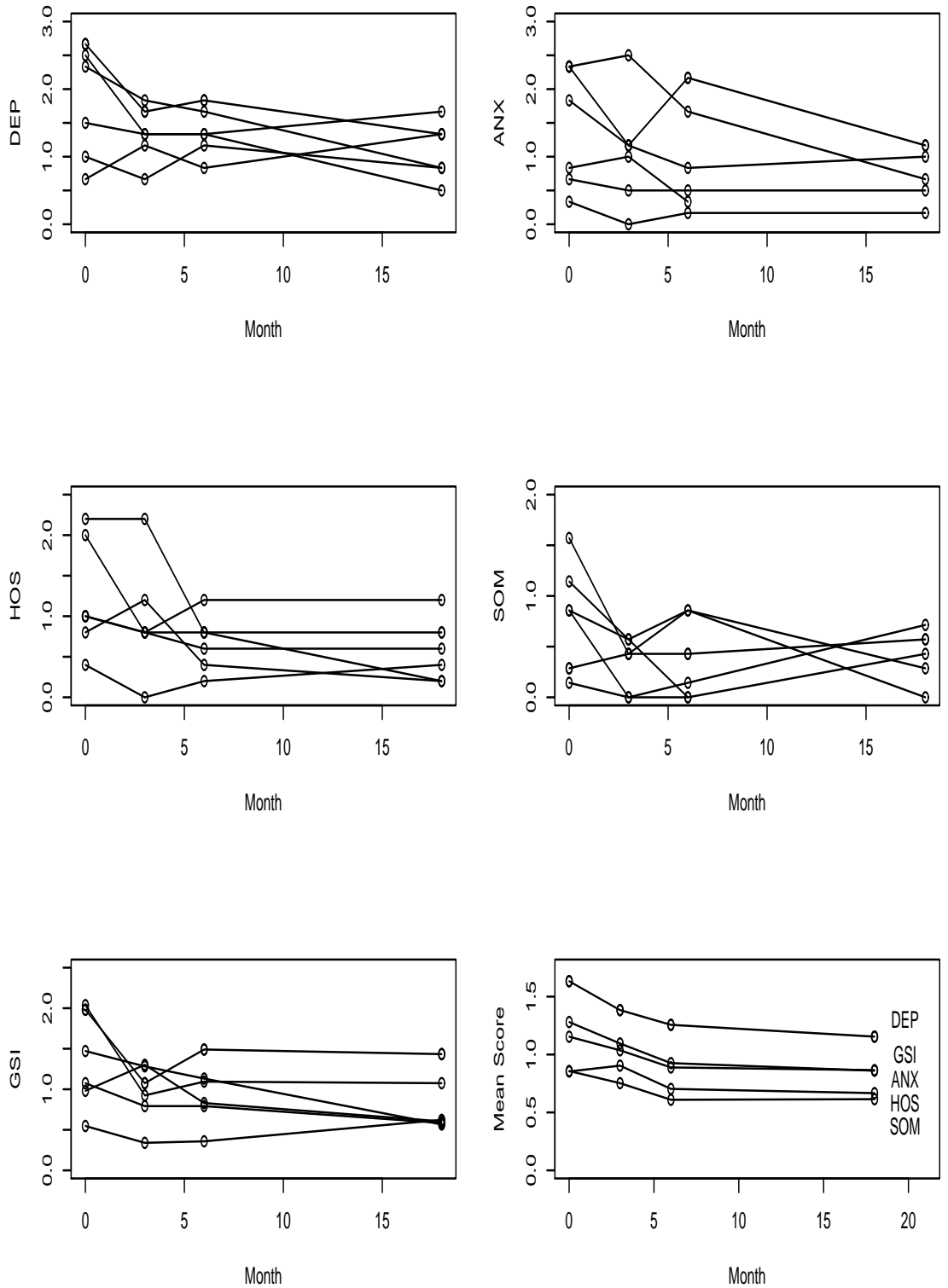


Figure 2.1: Six randomly selected subjects and the total average scores. 56

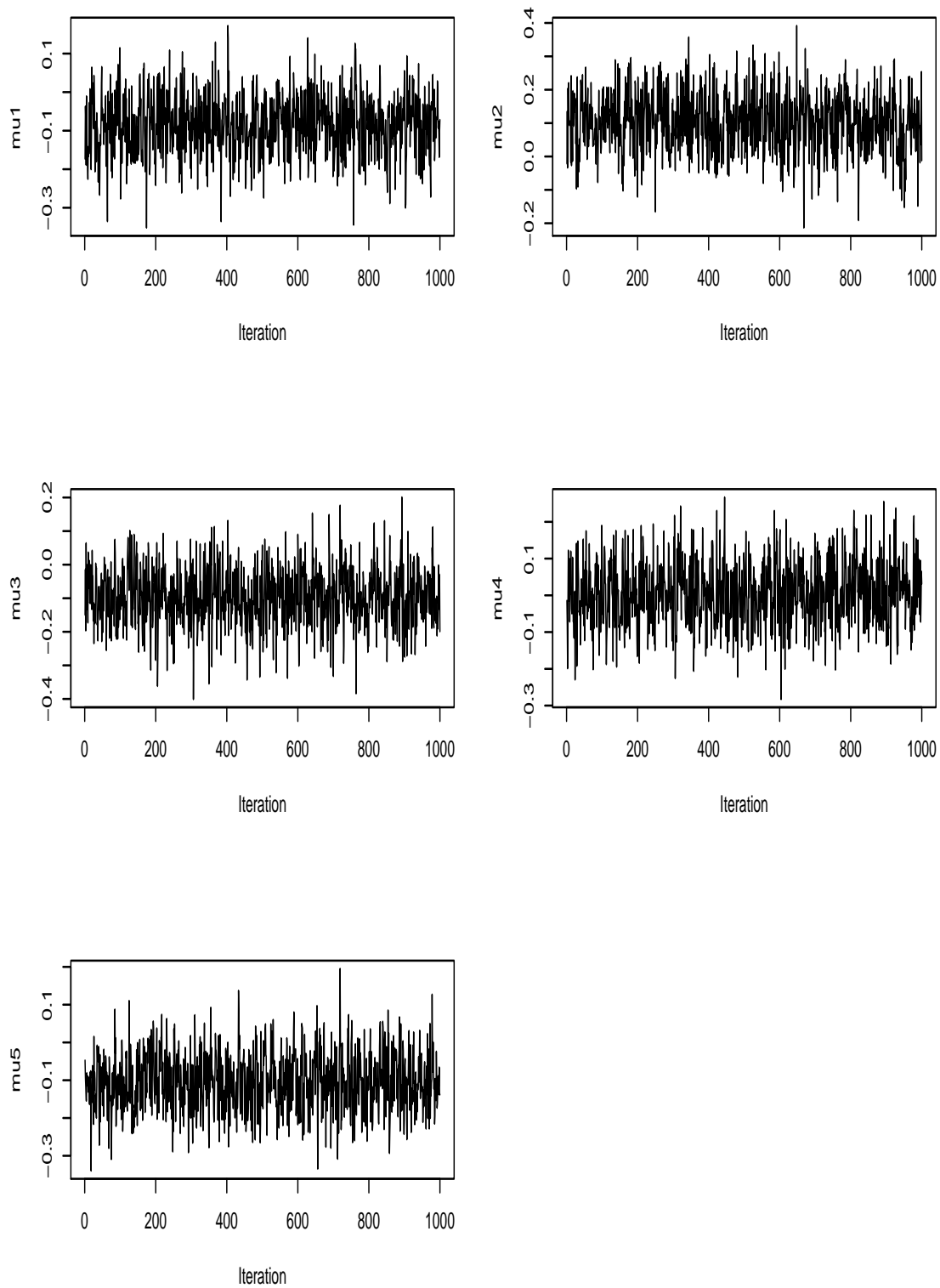


Figure 2.2: Time series plot for draws of parameters.

3 Multivariate One-sided Hypothesis Tests for Non-linear Mixed-Effects Models²

3.1 Introduction

Nonlinear mixed-effects (NLME) models, or hierarchical nonlinear models, have been widely used in longitudinal studies, such as human immunodeficiency (HIV) viral dynamics, pharmacokinetic analyses, and studies of growth and decay. These models are typically based on the underlying mechanisms that generate the data, so the model parameters often have meaningful physical interpretations, which implies natural restrictions or constraints on these parameters. For example, parameters such as the viral decay rates during an anti-HIV treatment or variances of random effects should be non-negative. Therefore, hypothesis testing for these parameters should incorporate the constraints, leading to one-sided or order-restricted tests. Incorporating natural parameter constraints in hypothesis testing should lead to more powerful tests than the corresponding unconstrained tests.

Constrained statistical inference has received much attention in the last few decades.

²A version of this chapter will be submitted for publication. Wang, T. and Wu, L. (2010) Multivariate One-sided Hypothesis Tests for Nonlinear Mixed-Effects Models.

See Silvapulle and Sen (2004) for a comprehensive overview. Most of the literature, however, focuses on multivariate normal distributions for continuous data or multinomial distributions for categorical data. The developments on regression models have been somewhat limited. Constrained inference for linear mixed-effects (LME) models are discussed in Stram and Lee (1994) and Verbeke and Molenberghs (2000). However, order-restricted inference for NLME models have not been studied. In this article, we consider one-sided hypothesis tests for parameters in NLME models.

Our research is motivated from HIV viral dynamic models, which model viral load trajectories during anti-HIV treatments. Based on biological and clinical arguments and preliminary analysis, Wu and Ding (1999), and Wu (2004) proposed the following exponential biphasic viral decay model

$$\begin{aligned}
 y_{ij} &= \log_{10}(e^{P_{1i}-\lambda_{1ij}t_{ij}} + e^{P_{2i}-\lambda_{2i}t_{ij}}) + e_{ij}, \\
 P_{1i} &= P_1 + b_{1i}, \quad \lambda_{1ij} = \lambda_1 + a_1 CD4_{ij} + b_{2i}, \\
 P_{2i} &= P_2 + b_{3i}, \quad \lambda_{2i} = \lambda_2 + b_{4i}, \\
 j &= 1, \dots, m_i, \quad i = 1, \dots, n,
 \end{aligned} \tag{3.1}$$

where $y_{ij} = \log_{10}(V(t_{ij}))$ is the \log_{10} -transformation of viral load measurements for patient i at time t_{ij} (the \log_{10} -transformation is used to stabilize the variance and make the data more normally distributed), $CD4_{ij}$ is CD4 counts at time t_{ij} , $\boldsymbol{\beta} = (P_1, P_2, \lambda_1, \lambda_2, a_1)^T$ are the fixed effects, $\mathbf{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^T$ are random effects, and e_{ij} is a random error. We assume that e_{ij} i.i.d. $\sim N(0, \sigma^2)$ and $\mathbf{b}_i \sim N(0, D)$, and that the random error and random effects are independent with each other. In this HIV dy-

dynamic model, e^{P_1} and e^{P_2} denote baseline viral load values in the productively infected cells, and long-lived infected and latently infected cells respectively. Parameters λ_1 and λ_2 denote the viral decay rates in the above two cells compartments correspondingly. Biological arguments and previous studies show that the rapid exponential decay phase mainly reflects the decay of productively infected cells. After several weeks the decay becomes slower, which may reflect the decay of long-lived infected and latently infected cells (Wu and Ding, 1999). Therefore the parameter constraints $P_1 > P_2$ and $\lambda_1 > \lambda_2$ may be considered when establishing the model or choosing the starting values for model fitting, in order to make the model identifiable.

Researchers are often interested in whether the true (population) decay rates are strictly positive or not, which indicates efficacy of the treatments. Such a testing problem can be written as follows

$$H_0 : \boldsymbol{\lambda} = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \neq \mathbf{0} \quad (3.2)$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$. This is a constrained or one-sided testing problem.

Let $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^T$, where y_{ij} is the response value for individual i at time t_{ij} , $i = 1, \dots, n$, $j = 1, \dots, m_i$. A general NLME model can be written as

$$\begin{aligned} y_{ij} &= g(\mathbf{t}_{ij}, \boldsymbol{\beta}_{ij}) + e_{ij}, & \mathbf{e}_{ij} &\sim N(0, \sigma^2), \\ \boldsymbol{\beta}_{ij} &= \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{b}_i, & \mathbf{b}_i &\sim N(\mathbf{0}, D), \\ & & j &= 1, \dots, m_i, \quad i = 1, \dots, n, \end{aligned}$$

where $g(\cdot)$ is a known real-valued nonlinear function, and \mathbf{x}_{ij} is a vector of covariates that may include both time-dependent or time-independent variables. In our HIV

dynamic model $\beta_{ij} = (P_{1i}, P_{2i}, \lambda_{1ij}, \lambda_{2i})^T$. All other notations are the same as in the HIV dynamic model example.

Let $\theta = (\beta, \sigma^2, \text{vec}(D))^T \in \mathbb{R}^k$ denote the collection of all parameters, where k denotes the total number of distinct parameters. A general one-sided multivariate hypotheses can be written as

$$H_0 : R\theta = \mathbf{0} \quad \text{versus} \quad H_1 : R\theta \geq \mathbf{0}, \quad R\theta \neq \mathbf{0} \quad (3.3)$$

where R is a given matrix with 0's or 1's, and the rank of R is denoted as r . In this chapter, we develop a likelihood ratio test, a Wald test, and a score test for testing (3.3). To illustrate the forms of R , let us consider the above testing problem (3.2) in AIDS study as a simple example. If D is assumed to be diagonal for simplicity, then we have $\theta = (P_1, P_2, \lambda_1, \lambda_2, \sigma^2, d_{11}, d_{22}, d_{33}, d_{44})^T$. So the matrix R should have the following form

$$R = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (3.4)$$

and r is equal to 2.

In Section 3.2, we develop LRT, Wald and Score tests for multivariate one-sided hypotheses (3.3) for NLME models. In Section 3.3, computing methods for NLME models will be discussed. In Section 3.4, a real data example and a simulation study are presented. We conclude the chapter in Section 3.5 with some discussions.

3.2 Multivariate One-sided Tests for NLME Models

In this section, we develop several testing methods for the general multivariate one-sided test problems (3.3) in NLME models, as introduced in Section 3.1.

3.2.1 The LRT

To develop the LRT for (3.3), we need the following two assumptions:

$$\mathbf{A1} : -n^{-1}\nabla^2 l_n(\boldsymbol{\theta}) \xrightarrow{p} V_{\boldsymbol{\theta}}, \quad (3.5)$$

$$\mathbf{A2} : n^{-\frac{1}{2}}\nabla l_n(\boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, V_{\boldsymbol{\theta}}), \quad \text{as } n \rightarrow \infty. \quad (3.6)$$

where $\nabla = \partial/\partial\boldsymbol{\theta}$, $\nabla^2 = \partial^2/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$ and $V_{\boldsymbol{\theta}}$ is a positive definite matrix. Assumptions **A1** and **A2** can be ensured by the following regularity conditions. Specifically, for $i = 1, \dots, n$,

C1: The first two partial derivatives of the log-likelihood function $l_i(\boldsymbol{\theta}|\mathbf{y}_i) = \log(f(\mathbf{y}_i|\boldsymbol{\theta}))$ with respect to $\boldsymbol{\theta}$ exist almost everywhere in the parameter space Θ ;

C2: For each $\boldsymbol{\theta}$ in the neighborhood of true value $\boldsymbol{\theta}_0$ there exist functions $G_i(\mathbf{y}_i)$ and $H_i(\mathbf{y}_i)$ such that $\int G_i(\mathbf{y}_i)d\mathbf{y}_i < \infty$, $\int H_i(\mathbf{y}_i)d\mathbf{y}_i < \infty$ and the absolute values of the first two partial derivatives of $\log(f(\mathbf{y}_i|\boldsymbol{\theta}))$ are bounded by $G_i(\mathbf{y}_i)$ and $H_i(\mathbf{y}_i)$ respectively;

C3: For each $\boldsymbol{\theta} \in \Theta$, $V_i = (V_i^{hl}) = E\{\nabla l_i(\boldsymbol{\theta}|\mathbf{y}_i)[\nabla l_i(\boldsymbol{\theta}|\mathbf{y}_i)]^T\}$ is finite and positive definite for $h = 1, \dots, k$ and $l = 1, \dots, k$.

C4: There exist positive numbers c_{hl} such that $0 < V_i^{hl} < c_{hl} < \infty$ for $h = 1, \dots, k$ and $l = 1, \dots, k$.

In the above regularity conditions, **C1** ensures that the log-likelihood function has a

Taylor expansion as a function of $\boldsymbol{\theta}$. **C2** ensures that the log-likelihood function is differentiable with respect to $\boldsymbol{\theta}$ under the integral sign, which consequently implies that $E\{\nabla l_i(\boldsymbol{\theta}|\mathbf{y}_i)[\nabla l_i(\boldsymbol{\theta}|\mathbf{y}_i)]^T\} = -E\{\nabla^2 l_i(\boldsymbol{\theta}|\mathbf{y}_i)\}$. **C4** and **C2** are used to ensure the convergence of the second derivative of the log-likelihood function $l_n(\boldsymbol{\theta}|Y) = \log(L_n(\boldsymbol{\theta}|Y))$, following Markov's Law of Large Numbers for non-i.i.d. observations, as discussed in Greene (2003). **A1** together with **C3** and the Lindeberg condition (see Appendix) can ensure **A2**, based on Lindeberg-Feller Central Limit Theorem, as discussed in Serfling (1980) and Van der Vaart (1998).

Let $\tilde{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}$ denote the maximizers of $l_n(\boldsymbol{\theta})$ over the parameter spaces $\{\boldsymbol{\theta} : R\boldsymbol{\theta} \geq \mathbf{0}\}$ and $\{\boldsymbol{\theta} : R\boldsymbol{\theta} = \mathbf{0}\}$ respectively. For testing (3.3), the LRT statistic is given by

$$T_{LR} = 2[l_n(\tilde{\boldsymbol{\theta}}) - l_n(\bar{\boldsymbol{\theta}})]. \quad (3.7)$$

Due to the complexity of the likelihood function as we discussed in the last paragraph, the exact null distribution of T_{LR} often does not have an explicit form. However, we can derive the asymptotic null distribution under the above assumptions as stated in the following theorem.

Theorem 3.1 Assuming the regularity conditions **C1** - **C4** and the assumptions **A1** and **A2**, in NLME models, the asymptotic null distribution of the LRT for testing $H_0 : R\boldsymbol{\theta} = \mathbf{0}$ vs $H_1 : R\boldsymbol{\theta} \geq \mathbf{0}, R\boldsymbol{\theta} \neq \mathbf{0}$ is given by the following $\bar{\chi}^2$ -distribution

$$\Pr(T_{LR} \leq c|H_0) \rightarrow \sum_{i=0}^r \omega_i(r, RV_{\boldsymbol{\theta}_0}^{-1}R^T) \Pr(\chi_i^2 \leq c), \quad \text{as } n \rightarrow \infty, \quad (3.8)$$

where n is the number of individuals, r is the rank of R , $\boldsymbol{\theta}_0$ denotes the true value of

$\boldsymbol{\theta}$ in H_0 , $\omega_i(r, RV_{\boldsymbol{\theta}_0}^{-1}R^T)$ is the probability weight (see details in Section 1.4), and χ_0^2 denotes the distribution that takes the value zero with probability one.

The proof is given in the Appendix.

Although we have derived the null distribution for LRT, the limiting variance-covariance matrix $V_{\boldsymbol{\theta}_0}$ in (3.8) is usually unknown. In order to actually implement the $\bar{\chi}^2$ test in practice, we can either substitute the observed information matrix $-n^{-1}\nabla^2 l_n(\bar{\boldsymbol{\theta}})$ estimated under H_0 for $V_{\boldsymbol{\theta}_0}$ and use simulation to obtain the cut-off values, or use the upper bound of the null probability to do a conservative test. Specifically, these ideas are stated in the following theorem.

Theorem 3.2 Assuming the regularity conditions **C1** - **C4** and the assumptions **A1** and **A2**, in NLME models, when the limiting variance-covariance matrix $V_{\boldsymbol{\theta}_0}$ is unknown, the asymptotic null distribution of T_{LR} in (3.8) can be estimated by

$$\widehat{\Pr}(T_{LR} \leq c | H_0) = \sum_{i=0}^r \omega_i(r, R\hat{V}^{-1}(\bar{\boldsymbol{\theta}})R^T) \Pr(\chi_i^2 \leq c), \quad (3.9)$$

where $\hat{V}(\bar{\boldsymbol{\theta}}) = -n^{-1}\nabla^2 l_n(\bar{\boldsymbol{\theta}})$. Alternatively, the asymptotic null distribution of T_{LR} in (3.8) has the following lower and upper bounds:

$$\frac{\Pr(\chi_0^2 \geq c) + \Pr(\chi_1^2 \geq c)}{2} \leq \Pr(T_{LR} \geq c | H_0) \leq \frac{\Pr(\chi_{r-1}^2 \geq c) + \Pr(\chi_r^2 \geq c)}{2}. \quad (3.10)$$

The proof of (3.10) is given in the Appendix.

The approximation in estimation (3.9) should be reasonably good when the sample size n is large. The test based on the bounds in (3.10) may be conservative, but it is computationally simpler. We will evaluate these two tests by simulation later.

3.2.2 Wald and Score Tests

Wald and score tests are commonly used in two-sided hypotheses testing problems and have the same asymptotic null distributions as the LRT. In this section we will develop these tests for the multivariate one-sided hypotheses for NLME models considered in the previous section, following Silvapulle and Sen (2004).

The Wald and score test statistics are given respectively by

$$\begin{aligned} T_W &= n(R\tilde{\boldsymbol{\theta}})^T [R\hat{V}^{-1}(\tilde{\boldsymbol{\theta}})R^T]^{-1}(R\tilde{\boldsymbol{\theta}}), \\ T_S &= \frac{1}{n}\mathbf{u}^T [R\hat{V}^{-1}(\tilde{\boldsymbol{\theta}})R^T]^{-1}\mathbf{u}, \end{aligned}$$

where

$$\begin{aligned} \hat{V}(\tilde{\boldsymbol{\theta}}) &= -n^{-1}\nabla^2 l_n(\tilde{\boldsymbol{\theta}}) \\ \mathbf{u} &= R\hat{V}^{-1}(\tilde{\boldsymbol{\theta}})(\nabla l_n(\tilde{\boldsymbol{\theta}}) - \nabla l_n(\bar{\boldsymbol{\theta}})), \end{aligned}$$

and $\tilde{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}$ denote the maximizers of $l_n(\boldsymbol{\theta})$ over the alternative parameter space $\{\boldsymbol{\theta} : R\boldsymbol{\theta} \geq \mathbf{0}\}$ and the null parameter space $\{\boldsymbol{\theta} : R\boldsymbol{\theta} = \mathbf{0}\}$, respectively. The Wald testing statistic T_W measures the norm of $R\tilde{\boldsymbol{\theta}}$. When $H_0 : R\boldsymbol{\theta} = \mathbf{0}$ is true, T_W is expected to be small. Similarly, the score testing statistic T_S measures the norm of $\nabla l_n(\tilde{\boldsymbol{\theta}}) - \nabla l_n(\bar{\boldsymbol{\theta}})$. When the null hypothesis holds, $\tilde{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}$ are expected to be close, so the difference of the two estimated scores is expected to be close to 0. Their asymptotic null distributions are given in the following theorem.

Theorem 3.3 Assuming the regularity conditions **C1** - **C4**, in NLME models, the Wald test statistic T_W and the score test statistic T_S have the same asymptotic distribution

as that of the LRT statistic T_{LR} in Theorem 3.1, i.e.

$$T_{LR} = T_W + o_p(1) = T_S + o_p(1).$$

The proof of Theorem 3.3 is very different from the two-sided hypothesis case and is mainly based on the Kühn-Tücker multipliers. We will give the proof in the Appendix.

3.3 Computing Methods for NLME Models

For general NLME models, the marginal joint distribution of $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{im_i})$ is

$$f(\mathbf{y}_i|\boldsymbol{\theta}) = \int f(\mathbf{y}_i|x_i, \boldsymbol{\beta}, \sigma^2, \mathbf{b}_i)f(\mathbf{b}_i|D)d\mathbf{b}_i, \quad (3.11)$$

for $i = 1, \dots, n$, and the likelihood is given by

$$L_n(\boldsymbol{\theta}|Y) = \prod_{i=1}^n \int f(\mathbf{y}_i|x_i, \boldsymbol{\beta}, \sigma^2, \mathbf{b}_i)f(\mathbf{b}_i|D)d\mathbf{b}_i. \quad (3.12)$$

The marginal distribution $f(\mathbf{y}_i|\boldsymbol{\theta})$ in (3.11) often does not have an explicit form, because NLME models are nonlinear in \mathbf{b}_i . So the likelihoods for NLME models typically do not have analytic or closed form expressions. In practice, the commonly used approaches include maximum likelihood estimation via Gauss-Hermite quadrature method and approximation methods. The Gauss-Hermite quadrature (GHQ) method approximates the desired integrals by a weighted summation evaluated at suitably chosen quadrature points (grid points). The accuracy of the numerical approximation can be improved by increasing the number of quadrature points. The “approximate” methods use Taylor expansions (Lindstrom and Bates, 1990) or Laplace approximations (Wolfinger, 1993; Joe, 2008) to approximate the model or likelihood. They avoid

numerical integrals in the likelihoods so are computationally very efficient and widely implemented in standard software such as “nlme” in R/Splus and “nlmixed” in SAS. Note that the Laplace method is also an option in “nlmixed”. More details and other methods can be found in Wu (2009).

For our study, one of the commonly used approximate methods: linearization method by first-order Taylor expansion is utilized. It is actually equivalent to iteratively solving linear mixed effects models, which avoids integrations with respect to the unobservable random effects so greatly simplifies computation. Although we have implemented the numerically fast linearization method in our study, all of the numerical results need to be validated with a method that gives more reliable estimates. Pinheiro and Bates have studied and compared the performance of the linearization method with Laplace approximation and adaptive GHQ method for some non-linear models (Pinheiro and Bates, 1995), but their conclusions do not cover all non-linear models. We did a limited simulation study on the HIV viral dynamic model (3.1) to evaluate the performance of the linearization method, using the GHQ method as a gold standard. The results show that the linearization method might have inaccuracy in the second significant digit in the estimates, and it might give liberal standard errors in estimating standard deviations of the random effects. But there are confounding factors in the parametrization and the number of random effects. Therefore, the GHQ or the Laplace method should probably be considered to have more accurate estimates, and the linearization method can be used to get a good starting point, as suggested by Pinheiro and Bates (1995).

3.4 Real Data Example

Consider the HIV viral dynamics example (3.1) described in Section 3.1. The biphasic NLME model approximates the viral load trajectories during anti-HIV treatments, as discussed in Section 3.1. In this study, 53 HIV-1 infected patients were treated with a potent antiretroviral treatment. Five patients discontinued the study because of intolerance to the drug and other problems. The viral load (plasma HIV-1 RNA) and CD4 cell counts were measured on day 0, semi-weekly in the first two weeks, weekly in the 3rd and 4th weeks, and monthly in the 2nd and 3rd months after treatment. There are several measurements after three months as well, but we consider only the first three months' data before viral rebound, as suggested by Wu and Ding (1999). Figure 1.3 shows the viral load data. 46 patients have at least one observation of CD4 cell counts. The total number of completely observed observations is 304. The number of measurements for each individual varies from 3 to 8. It is interesting to test if the viral loads actually decline during the anti-HIV treatment. Since it is unlikely that viral loads increase during the treatment (i.e., the decay rates should be non-negative), we are interested in testing the one-sided hypotheses (3.2) in Section 3.1, $H_0 : \boldsymbol{\lambda} = \mathbf{0}$ vs $H_1 : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \neq \mathbf{0}$, where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$. Here for simplicity and for illustration purpose, we have assumed that the matrix D is diagonal.

We implemented the tests using two methods based on Theorem 3.2. One is called *substitution method*, which substitutes the unknown $V(\boldsymbol{\theta})$ by $\hat{V}(\tilde{\boldsymbol{\theta}})$ in Theorem 3.2. The other method is called *bound method*, which uses the bounds in Theorem 3.2 to

compute p-values. We focus on two tests: the LRT and the Wald test.

We calculate the LRT and Wald testing statistics for testing H_0 versus H_1 as follows. The LRT statistics is given by $T_{LR} = 2[-146.53 - (-230.21)] = 167.35$, where -146.53 and -230.21 are log-likelihoods under H_1 and H_0 , respectively. The Wald test statistic is given by $T_W = 46(0.34, 0.032)^T [R\hat{V}^{-1}(\tilde{\theta})R^T]^{-1}(0.34, 0.032) = 372.19$, where 0.34 and 0.032 are estimates of λ_1 and λ_2 under H_1 respectively,

$$R = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and $\theta = (P_1, P_2, \lambda_1, \lambda_2, a_1, \sigma^2, d_{11}, d_{22}, d_{33}, d_{44})^T$. The 95% critical values for the bound and substitution methods are both less than 10. The p-values for both tests are nearly 0 by either bound or substitution methods. We thus have strong evidence against H_0 , and we conclude that the viral loads decline significantly during the anti-HIV treatment.

3.5 A Simulation Study

In this section we perform a simulation study to evaluate the proposed LRT and the Wald test for multivariate one-sided testing problems (3.2). For comparison, we also consider the corresponding usual two-sided χ^2 tests, which ignore the one-sided nature of the hypotheses. We evaluate the methods by comparing both the sizes (type I errors) and the powers of the tests. The significance level $\alpha = 0.05$ is chosen.

The simulation study is still based on the biphasic HIV dynamics NLME model (3.1) in Section 3.1. The sample size of subjects is $n = 46$, and each individual has

the same number of repeated observations m_i as in the real data example in the last section. The measurement times and covariate values are also the same as those in the real data example. The viral loads y_{ij} are simulated from the model based on different settings of the parameters. The values of the parameters to be tested (λ) are listed in Tables 3.1 - 3.2. The values of the parameters that are not to be tested (including P_1, P_2, a_1, σ^2) take the values of the estimations from the real data example in Section 3.4. All simulations are repeated for 1000 times. Both the bound method and the substitution methods are used to calculate p-values.

Tables 3.1 and 3.2 show simulation results for the substitution method and the bound method, respectively. We see that all the tests attain the 0.05 nominal significance level. However the proposed one-sided LRT and Wald test have higher powers than the usual two-sided tests. This is expected since the proposed one-sided tests incorporate the order restrictions in the hypotheses while the two-sided tests ignore such restrictions. The tests based on the substitution method appear to be more powerful than the tests based on the bound method. This is again expected since the bound method is conservative.

In summary, the above simulation results show that the proposed one-sided tests are more powerful than the usual two-sided tests for testing one-sided or order-restricted hypotheses. This is because the one-sided tests incorporate the restrictions in the hypotheses. Although the proposed tests are large-sample tests, they appear to perform well for small/moderate samples. The bound method is usually conservative since it is based on the bounds of the null probabilities. The substitution method relies on

good estimation of the covariance matrix. If the sample size is small or the number of parameter is large, the methods may not perform well.

3.6 Discussion

In this chapter we have proposed several tests for multivariate one-sided hypotheses in NLME models. These tests are asymptotically equivalent but their finite-sample performances may differ. The simulation results show that they perform well in finite samples and they are more powerful than the usual two-sided tests which ignore the order restrictions in the hypotheses.

We have considered a simple case of testing the variance components by assuming that the covariance matrix of the random effects is diagonal. This is for simplicity and for illustration purpose only. In practice, the covariance matrices may not be diagonal. In this case, the testing procedure is more complicated since one needs to ensure the positive definiteness of the covariance matrices. Further research for this problem is required and is of great interest.

In practice, missing data often exist in longitudinal studies. Thus, a natural extension will be to test the multivariate one-sided hypotheses in NLME models with missing data. We are considering a multiple imputation method for this problem and will report the results separately.

The methods proposed in this chapter can be extended to other mixed effects models, such as generalized linear mixed models, frailty models, and semiparametric or nonparametric mixed effects models. The extensions are conceptually straightforward,

but the performances of the order-restricted tests under different models need to be evaluated separately.

Table 3.1: Simulation results for $H_0 : \boldsymbol{\lambda} = \mathbf{0}$ vs $H_1 : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \neq \mathbf{0}$: bound method
(n=46)

$\boldsymbol{\lambda}$	One-sided LRT	One-sided Wald	Two-sided LRT	Two-sided Wald
Type I Error Rates (in %)				
(0, 0)	3.1	3.1	2.0	2.0
Power Comparisons (in %)				
(0, 0.0008)	5.5	6.0	4.6	4.7
(0, 0.005)	62	65	55	58
(0.001, 0.001)	7.4	8.1	5.2	5.5
(0.001, 0.005)	62	65	55	59

Table 3.2: Simulation results for $H_0 : \boldsymbol{\lambda} = \mathbf{0}$ vs $H_1 : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \neq \mathbf{0}$: substitution method (n=46)

$\boldsymbol{\lambda}$	One-sided LRT	One-sided Wald	Two-sided LRT	Two-sided Wald
Type I Error Rates (in %)				
(0, 0)	5.2	5.0	2.0	2.0
Power Comparisons (in %)				
(0, 0.0008)	9.2	9.3	4.6	4.7
(0, 0.005)	71	72	55	58
(0.001, 0.001)	11	11	5.2	5.5
(0.001, 0.005)	71	73	55	59

4 Multivariate One-sided Hypothesis Testing in NLME Models with Missing Covariates ³

4.1 Introduction

Nonlinear mixed-effects (NLME) models, or hierarchical nonlinear models, have various applications in longitudinal studies such as pharmacokinetic analysis, studies of growth and decay, and human immunodeficiency virus (HIV) dynamics. These models are typically based on the underlying mechanisms which generate the data, so the model parameters often have meaningful physical interpretations, which implies natural restrictions or constraints on these parameters. For example, parameters such as decay rates or variances of random effects should be non-negative. Therefore, multivariate one-sided hypotheses arise frequently in these studies. When the data are complete, Silvapulle and Sen (2004) provide an excellent review on one-sided hypothesis testing methods for normal and general regression model cases. Recently, we, as in Chapter 3 in this thesis, developed the likelihood ratio (LR) and Wald tests for one-sided hypotheses in NLME models and showed that they are more powerful than

³A version of this chapter will be submitted for publication. Wang, T. and Wu, L. (2010) Multivariate One-sided Hypothesis Testing in NLME Models with Missing Covariates.

the conventional two-sided tests.

In practice, data are often missing at either responses or covariates. To our knowledge, very few of the existing multivariate one-sided testing methods can handle missing data in NLME model properly. The commonly used *complete-case* (CC) method, which discards all incomplete observations, may give inefficient results. When the missing data are not MCAR, the CC method may give biased results (Little and Rubin, 2002). Another widely used approach in longitudinal studies, the *last value carried forward method* (LVCF), which imputes a missing value in a covariate or response by the observed covariate or response value immediately before the time point when the missing value presents, may produce misleading results because it fails to incorporate the missing data uncertainty (Wu, 2009). In this chapter, we propose several testing methods for one-sided hypotheses in NLME models based on multiple imputations. The idea is motivated by Chapter 2 in this thesis where one-sided testing methods based on multiple imputations for a multivariate normal population were developed. We assume that the missing data are missing at random in the sense that the missingness may be related to observed data but not on missing data (Rubin 1987). The missing data pattern can be arbitrary.

Multiple imputation (MI) methods were first introduced by Rubin in early 1970's. In recent years they have been widely used for missing data problems in practice due to the availability of modern computers and software (e.g. SAS and Splus). The standard complete-data methods may be used once the missing data are multiply imputed. The resulting, say m , “complete-data” analyses are then combined to produce an overall

inference, which takes into account the uncertainty due to missing values. General theories and implementations of MI methods can be found in Rubin (1987), Li et.al. (1991), Meng and Rubin (1992), and Schafer (1997).

In Section 4.2, we introduce the motivating NLME model and multivariate one-sided hypotheses. In Section 4.3, we discuss how to generate proper multiple imputations for NLME models with missing time-varying covariates. In Section 4.4, based on multiple imputations we develop several testing statistics by combining individual analysis results for NLME models. In Sections 4.5 and 4.6, a real data example and a simulation study are presented. We conclude the chapter in Section 4.7 with some discussions.

4.2 A Motivating NLME Model and Multivariate One-sided Hypothesis

Our research is motivated by HIV viral dynamic models (Wu and Ding 1999; Wu and Wu 2002; Wu 2004), which model viral load trajectories during anti-HIV treatments. Based on biological and clinical arguments, the following exponential biphasic viral decay model was proposed (Wu and Wu, 2002)

$$\begin{aligned}
 y_{ij} &= \log_{10}(e^{P_{1i}-\lambda_{1ij}t_{ij}} + e^{P_{2i}-\lambda_{2ij}t_{ij}}) + e_{ij}, & (4.1) \\
 P_{1i} &= P_1 + b_{1i}, & \lambda_{1ij} = \lambda_1 + a_1CD4_{ij} + a_2CD8_{ij} + b_{2i}, \\
 P_{2i} &= P_2 + b_{3i}, & \lambda_{2i} = \lambda_2 + b_{4i},
 \end{aligned}$$

where $y_{ij} = \log_{10}(V(t_{ij}))$ is the \log_{10} -transformation of viral load measurements for patient i at time t_{ij} (the \log_{10} -transformation is used to stabilize the variance and make the data more normally distributed), $CD4_{ij}$ and $CD8_{ij}$ are CD4 and CD8 cells counts at time t_{ij} , $\boldsymbol{\beta} = (P_1, P_2, \lambda_1, \lambda_2, a_1, a_2)$ are the fixed effects, $\mathbf{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^T$ are random effects, and e_{ij} is a random error. We assume that e_{ij} i.i.d. $\sim N(0, \sigma^2)$ and $\mathbf{b}_i \sim N(0, D)$. In this HIV dynamic model, e^{P_1} and e^{P_2} denote baseline viral load values, and parameters λ_1 and λ_2 denote the viral decay rates, for the corresponding cells compartments (see detailed interpretations in Section 3.1).

In this AIDS study, researchers are often interested in whether the true (population) decay rates are positive or zero, which indicates efficacy of the treatments. Such a testing problem can be written as follows

$$H_0 : \boldsymbol{\lambda} = \mathbf{0} \quad vs \quad H_1 : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \neq \mathbf{0} \quad (4.2)$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$. This is a constrained or one-sided testing problem, since the decay rates during an anti-HIV treatment cannot be negative.

The above NLME model has the following general form:

$$\begin{aligned} y_{ij} &= g(\mathbf{t}_{ij}, \boldsymbol{\beta}_{ij}) + e_{ij}, & \mathbf{e}_{ij} &\sim N(0, \sigma^2), \\ \boldsymbol{\beta}_{ij} &= \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{b}_i, & \mathbf{b}_i &\sim N(\mathbf{0}, D), \\ & j = 1, \dots, m_i, & i &= 1, \dots, n, \end{aligned}$$

where $g(\cdot)$ is a nonlinear function, and \mathbf{x}_{ij} is a vector that contains both time-dependent and time-independent covariates. All other notations and settings are the same as in

the example. Let $\boldsymbol{\theta}$ denote the collection of all parameters in the above general NLME model. The general form of the multivariate one-sided hypotheses (4.2) can be written as

$$H_0 : R\boldsymbol{\theta} = \mathbf{0} \quad vs \quad H_1 : R\boldsymbol{\theta} \geq \mathbf{0}, \quad R\boldsymbol{\theta} \neq \mathbf{0}. \quad (4.3)$$

where R is a design matrix consisting of 1 and 0. Specifically for our example (4.2), $\boldsymbol{\theta} = (P_1, P_2, \lambda_1, \lambda_2, a_1, a_2, \sigma, d_{11}, d_{22}, d_{33}, d_{44})^T$ (assuming D is a diagonal matrix for simplicity) and

$$R = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & 0 & \dots & 0 \end{bmatrix}_{2 \times 11},$$

where the rank of the matrix R is $r = 2$. In the remaining of this chapter we will discuss how to test one-sided hypotheses (4.3) based on multiple imputations when missing data are present in practice.

4.3 Multiple Imputation for NLME Model with Missing Data

In longitudinal studies, the time-varying covariates are often not completely observed due to various reasons. For example, in our motivational HIV viral dynamics study, some of the CD4 and CD8 cell counts (covariates) were not observed at the same times as the viral loads (responses), leading to missing covariates.

At the presence of missing data, the covariates matrix X , whose columns contain data for individual covariates, can be expressed as $X = [X_{inc}, X_{com}]$, where “inc” and “com” denote the covariate columns that are incompletely and completely observed, respectively. As we discussed in Section 4.1, the complete-data methods may not be

directly applicable in such cases. However, if we generate several “complete datasets” by simulating several possible values for each missing value, then we may use complete-data methods and combine the results for an overall inference by properly incorporating the missing data uncertainty. In the following we discuss how to generate *proper* multiple imputations in the sense of Rubin (1987).

Write $\text{vec}(X_{inc}) = (\mathbf{x}_{obs}, \mathbf{x}_{mis})$, where \mathbf{x}_{obs} and \mathbf{x}_{mis} denote the observed and missing components in X , respectively, and $\text{vec}(\cdot)$ represents vectorization. As discussed in Rubin (1987), proper multiple imputations may be generated from the predictive distribution of the missing data given the observed data. Specifically, multiple imputations can be generated from the predictive distribution $f(\mathbf{x}_{mis} | \mathbf{x}_{obs}, X_{com}, \mathbf{y})$. Note that inferences based on multiple imputations are often robust in the sense that the imputation model may somewhat differ from the analytic model (Little and Rubin 2002), when the missing rate is not high. This is because imputations are only needed for the missing values, while the observed values remain unchanged.

We propose the following two-stage imputation procedure using a multivariate linear mixed effects (LME) model, similar to Wu and Wu (2002). The proposed imputation procedure first fits a NLME model without any covariates to obtain initial individual-level estimates $\hat{\beta}_i^{(0)}$. Then, we incorporate $\hat{\beta}_i^{(0)}$ together with the completely observed covariates into the imputation model for the covariates with missing values. Specifically, the proposed two-stage imputation procedure can be described as follows:

Stage 1: obtain the initial estimates $\widehat{\boldsymbol{\beta}}_i^{(0)}$ from the following NLME model without any covariates

$$\begin{aligned} y_{ij} &= g(\mathbf{t}_{ij}, \boldsymbol{\beta}_i^{(0)}) + e_{ij}, & \boldsymbol{\beta}_i^{(0)} &= \boldsymbol{\beta}^{(0)} + \mathbf{b}_i^{(0)}, \\ j &= 1, \dots, m_i, & i &= 1, \dots, n \end{aligned}$$

where e_{ij} i.i.d. $\sim N(0, \sigma^2)$ and $\mathbf{b}_i \sim N(0, D)$.

Stage 2: generate imputations based on the following multivariate LME model

$$\begin{aligned} X_{inc,i} &= (\mathbf{1} \ X_{com,i} \ \widehat{\boldsymbol{\beta}}_i^{(0)})_{m_i \times k} B_i + e_i, & e_i &\sim N_q(\mathbf{0}, \Sigma) \\ B_i &= B + V_i, & \text{vec}(V_i) &\sim N_{kq}(\mathbf{0}, \Psi) \\ i &= 1, \dots, n, \end{aligned} \tag{4.4}$$

where “inc” and “com” denote the covariate columns that are incompletely and completely observed, respectively. Note that the predicted value of $\mathbf{b}_i^{(0)}$ is used to obtain the individual level $\widehat{\boldsymbol{\beta}}_i^{(0)}$. Note also that q and k are the numbers of columns in X_{inc} and $(\mathbf{1} \ X_{com,i} \ \widehat{\boldsymbol{\beta}}_i^{(0)})_{m_i \times k}$, respectively, B and V_i are the $k \times q$ matrices of fixed and random effects, respectively, and $\text{vec}(V_i)$ denotes the vector by stacking columns of matrix V_i . Note that the notation $\widehat{\boldsymbol{\beta}}_i^{(0)}$ is being abused in the covariates matrix $(\mathbf{1} \ X_{com,i} \ \widehat{\boldsymbol{\beta}}_i^{(0)})_{m_i \times k}$, where $\widehat{\boldsymbol{\beta}}_i^{(0)}$ actually indicates a matrix with all rows being $\widehat{\boldsymbol{\beta}}_i^{(0)T}$. To simplify the notation, we write $Z_i = (\mathbf{1} \ X_{com,i} \ \widehat{\boldsymbol{\beta}}_i^{(0)})_{m_i \times k}$ in the remaining of the chapter.

After generating, say m ($m > 1$), multiple imputations from *Stage 2* (see below for details), we obtain m complete datasets. We can then fit the original NLME model (4.1) (with covariates) to each of the complete datasets and obtain an updated estimate

$\widehat{\beta}_i^{(1)}$ of β_i , which may be the average of the m estimates from the original NLME model (4.1). We can then go back to *Stage 2*, replace the initial estimate $\widehat{\beta}_i^{(0)}$ with the updated estimate $\widehat{\beta}_i^{(1)}$, and then generate updated multiple imputations. One may iterate this process until convergence is achieved in $\widehat{\beta}$. However, this iterative approach may require substantially intensive computations. Moreover, preliminary simulations done by Wu and Wu (2002) show that this iterative procedure may provide little improvement, possibly due to the reason we mentioned earlier that multiple imputations methods are generally robust to small or moderate mis-specifications of the imputation models. A reasonable method is simply to iterate a few times (say 2 or 3 times) to wash out uncertainty in the estimation.

In the following we discuss in details how to generate proper multiple imputations based on the proposed imputation model (4.4). Imputations can be generated from the predictive distribution $f(\mathbf{x}_{mis}|\mathbf{x}_{obs}, Z)$ via a Bayesian framework

$$f(\mathbf{x}_{mis}|\mathbf{x}_{obs}, Z) = \int f(\mathbf{x}_{mis}|\mathbf{x}_{obs}, Z, B, \Sigma, V) f(V|\Psi) f(\Phi) dV_1 \dots dV_n d\Phi,$$

where $\Phi = (B, \Sigma, \Psi)$ contains all unknown parameters, $f(\Phi)$ is an assumed prior for Φ , and $V = (\text{vec}(V_1), \dots, \text{vec}(V_n))^T$ are the matrix of random effects in the imputation model. In practice, the above integral often does not have an explicit form, but we may generate the imputations through Markov Chain Monte Carlo (MCMC) methods such as the Gibbs sampler. Note that besides the missing data \mathbf{x}_{mis} , the model parameters Φ and random effects V_i ($i=1, \dots, n$) are also unknown.

Specifically, imputations for missing values can be generated by the Gibbs sampler

as follows. At iteration $t+1$ ($t = 0, 1, 2, \dots$), given the parameter draws $\Phi^{(t)}$, simulated random effects $V^{(t)}$, and simulated missing values $\mathbf{x}_{mis}^{(t)}$ from the previous iteration t , we proceed with the following three steps to obtain the new simulated values $\Phi^{(t+1)}$, $V^{(t+1)}$ and $\mathbf{x}_{mis}^{(t+1)}$:

Step 1: Draw new random effects $V_1^{(t+1)}, \dots, V_n^{(t+1)}$ from the posterior $f(V_1, \dots, V_n \mid \mathbf{x}_{mis}^{(t)}, \Phi^{(t)}, \mathbf{x}_{obs}, Z)$.

Step 2: Draw new parameters $\Phi^{(t+1)} = (B^{(t+1)}, \Sigma^{(t+1)}, \Psi^{(t+1)})$ from the posterior $f(\Phi \mid V^{(t+1)}, \mathbf{x}_{mis}^{(t)}, \mathbf{x}_{obs}, Z)$.

Step 3: Simulate new values of the missing data $\mathbf{x}_{mis}^{(t+1)}$ from the posterior $f(\mathbf{x}_{mis} \mid \Phi^{(t+1)}, V^{(t+1)}, \mathbf{x}_{obs}, Z)$.

To implement the above MCMC approach in our study, first we assume a non-informative prior distribution for B : $P(B) \propto \text{constant}$, and assume proper inverted Wishart prior distributions for Σ and Ψ as follows:

$$\Sigma \sim W^{-1}(\gamma_1, \Lambda_1), \quad \gamma_1 \geq q,$$

$$\Psi \sim W^{-1}(\gamma_2, \Lambda_2), \quad \gamma_2 \geq kq,$$

where $\gamma_1^{-1}\Lambda_1^{-1}$ and $\gamma_2^{-1}\Lambda_2^{-1}$ may be considered as the prior guesses for Σ and Ψ , respectively. Based on Bayes theorem we obtain the posterior distributions in *Step 2* as

follows:

$$\begin{aligned} \Sigma & \mid V, \mathbf{x}_{mis}, \mathbf{x}_{obs}, Z \sim W^{-1} \left(\gamma_1 + \sum_{i=1}^n m_i - (k + q) + 1, \quad (\Lambda_1^{-1} + \sum_{i=1}^n \hat{e}_i^T \hat{e}_i)^{-1} \right), \\ \Psi & \mid V, \mathbf{x}_{mis}, \mathbf{x}_{obs}, Z \sim W^{-1} \left(\gamma_2 + n, \quad (\Lambda_2^{-1} + V^T V)^{-1} \right), \\ B & \mid V, \mathbf{x}_{mis}, \mathbf{x}_{obs}, Z \sim N \left(\hat{B}, \quad \Sigma \otimes \left(\sum_{i=1}^n Z_i^T Z_i \right)^{-1} \right), \end{aligned}$$

where

$$\begin{aligned} \hat{e}_i & = X_i^{inc} - Z_i \hat{B} - Z_i V_i, \\ \hat{B} & = \left(\sum_{i=1}^n Z_i^T Z_i \right)^{-1} \sum_{i=1}^n Z_i^T (X_i^{inc} - Z_i V_i). \end{aligned}$$

The posterior distribution in *Step 1* is given by

$$\text{vec}(V_i) \mid \Phi, \mathbf{x}_{mis}, \mathbf{x}_{obs}, Z \sim N(\text{vec}(\hat{V}_i), A_i),$$

where

$$\begin{aligned} \text{vec}(\hat{V}_i) & = A_i (\Sigma^{-1} \otimes Z_i^T) \text{vec}(X_i^{inc} - Z_i B), \\ A_i & = (\Psi^{-1} + (\Sigma^{-1} \otimes Z_i^T Z_i))^{-1}. \end{aligned}$$

To simulate the missing data, the posterior distribution in *Step 3* is given by

$$\begin{aligned} X_i^{inc} & \mid \Phi, V, Z \sim N(Z_i(B + V_i), \mathbf{1}_{m_i} \otimes \Sigma), \\ f(\mathbf{x}_{mis,i} \mid \mathbf{x}_{obs,i}, \Phi, V, Z) & = \prod_{j=1}^{m_i} f(\mathbf{x}_{mis,i,j} \mid \mathbf{x}_{obs,i,j}, \Phi, V, Z). \end{aligned}$$

Beginning with plausible starting values, we iterate the above three steps for a burn-in period, say 1000 iterations, and then we can obtain simulated \mathbf{x}_{mis} from the desired predictive distribution $f(\mathbf{x}_{mis} \mid \mathbf{x}_{obs}, Z)$ at the convergence. The adequacy of

the length of the burn-in period can be diagnosed by the time series plot of the draws of the parameters over iteration times. If the plot shows no obvious patterns or trends, the convergence is likely to be achieved after the burn-in period. This MCMC approach can be implemented by using the “pan” package in R. Repeating this MCMC approach for several times results in several “complete datasets”. For each “complete dataset”, standard complete-data methods may be used, and the results are then combined to obtain an overall inference. We will discuss the specific testing methods based on multiple imputations in the next section.

4.4 Multivariate One-sided Tests based on Multiple Imputations for NLME Models with Missing Covariates

After generating proper multiple imputations for NLME models as discussed in Section 4.3, we obtain, say m , “complete datasets”. Statistical inference such as hypothesis testing for NLME models can then be based on these m “complete datasets” as if there were no missing data, and the m results are then combined to produce an overall inference. For hypothesis testing problems, there are two approaches to combine several complete-data analysis results: we may either combine complete-data parameter estimates and then propose a single test statistic based on the combined estimates, such as a Wald-type test statistic, or we could combine several complete-data test statistics, say LRT statistics, and then propose an overall test. These methods were proposed and discussed in details in Chapter 2 in this thesis for multivariate normal populations. In

the following, we extend the results in Chapter 2 to the NLME models (as discussed in Section 4.2) with missing covariates for the multivariate one-sided hypothesis testing problem (4.3), i.e. $H_0 : R\boldsymbol{\theta} = \mathbf{0}$ v.s. $H_1 : R\boldsymbol{\theta} \geq \mathbf{0}$, $R\boldsymbol{\theta} \neq \mathbf{0}$.

Combining MLEs

Let $\{\mathbf{y}, X_*^{(l)}\}$ be the l -th “complete dataset” obtained from the l -th imputation, and let $\{\{\mathbf{y}, X_*^{(l)}\}, l = 1, 2, \dots, m\}$ be the m “complete datasets”. Suppose that $\hat{\boldsymbol{\theta}}_{*l}$ and \hat{U}_{n*l} are complete-data maximum likelihood estimates for $\boldsymbol{\theta}$ and its associated variance covariance matrix U_n respectively based on the l -th “complete dataset” $\{\mathbf{y}, X_*^{(l)}\}$, where \hat{U}_{n*l} is often obtained from the inverse of the negative second derivative of the log-likelihood function $(-\nabla^2 l_n(\boldsymbol{\theta}))^{-1}$ evaluated at $\hat{\boldsymbol{\theta}}_{*l}$. We can then combine the m complete-data estimates in a way similar to Li et al. (1991) as follows

$$\bar{\boldsymbol{\theta}}_m = \frac{1}{m} \sum_{l=1}^m \hat{\boldsymbol{\theta}}_{*l}, \quad (4.5)$$

$$\bar{U}_m = \frac{1}{m} \sum_{l=1}^m \hat{U}_{n*l},$$

$$\bar{B}_m = \frac{1}{m-1} \sum_{l=1}^m (\hat{\boldsymbol{\theta}}_{*l} - \bar{\boldsymbol{\theta}}_m)(\hat{\boldsymbol{\theta}}_{*l} - \bar{\boldsymbol{\theta}}_m)^T, \quad (4.6)$$

where $\bar{\boldsymbol{\theta}}_m$ is an unbiased estimator of $\boldsymbol{\theta}$, \bar{U}_m estimates the within-imputation covariance U_n , and \bar{B}_m estimates B_n , with B_n denoting the unknown between-imputation variation of $\hat{\boldsymbol{\theta}}_{*l}$, due to the *missing data uncertainty*. Following Rubin (1987), as discussed in Chapter 2, we have that when n is large

$$\bar{\boldsymbol{\theta}}_m \sim N(\boldsymbol{\theta}, \Sigma_{n,m}),$$

where

$$\Sigma_{n,m} = U_n + (1 + m^{-1})B_n,$$

where B_n reflects the missing data uncertainty for the data X with missing values.

Note that in our multivariate one-sided hypotheses testing problem (4.3), the parameters of interest are $R\boldsymbol{\theta}$. Therefore we modify the above results as follows. When n is large

$$R\bar{\boldsymbol{\theta}}_m \sim N(R\boldsymbol{\theta}, \Sigma_{n,m}^*),$$

where $\Sigma_{n,m}^* = RU_nR^T + (1 + m^{-1})RB_nR^T$. Therefore the usual assumption of equal ratios of missing to observed information (Meng and Rubin, 1992), as discussed in Chapter 2, is correspondingly modified to $RB_nR^T \propto RU_nR^T$. Similar with that in Chapter 2, the estimate of $\Sigma_{n,m}^*$ is given by

$$\widehat{\Sigma}_{n,m}^* = R\bar{U}_mR^T + r_mR\bar{U}_mR^T, \quad (4.7)$$

where $r_m = (1 + m^{-1})tr(R\bar{B}_mR^T(R\bar{U}_mR^T)^{-1})/p$ is the estimated ratio of missing to observed information. The estimate $\widehat{\Sigma}_{n,m}^*$ incorporates the missing data uncertainty.

Based on the combined statistics (4.5) – (4.6) and (4.7), we can construct an overall Wald-type test statistic for one-sided hypothesis (4.3) as follows,

$$\begin{aligned} T_W &= (R\bar{\boldsymbol{\theta}}_m)^T \widehat{\Sigma}_{n,m}^{*-1} (R\bar{\boldsymbol{\theta}}_m) - \min_{R\boldsymbol{\theta} \in \mathbb{O}^+} \{ (R\bar{\boldsymbol{\theta}}_m - R\boldsymbol{\theta})^T \widehat{\Sigma}_{n,m}^{*-1} (R\bar{\boldsymbol{\theta}}_m - R\boldsymbol{\theta}) \} \\ &= \left\| R\bar{\boldsymbol{\theta}}_m \right\|_{\widehat{\Sigma}_{n,m}^*}^2 - \left\| R\bar{\boldsymbol{\theta}}_m - \pi_{\widehat{\Sigma}_{n,m}^*} (R\bar{\boldsymbol{\theta}}_m; \mathbb{O}^+) \right\|_{\widehat{\Sigma}_{n,m}^*}^2, \end{aligned}$$

where $\pi_{\widehat{\Sigma}_{n,m}^*} (R\bar{\boldsymbol{\theta}}_m; \mathbb{O}^+)$ is the orthogonal projection of the overall estimate $R\bar{\boldsymbol{\theta}}_m$ onto the union of the null and the alternative parameter spaces with respect to (w.r.t.)

$\widehat{\Sigma}_{n,m}^*$. Similar to the statistic T_2 in (1.11), the statistic T_W measures the difference of the distance from $R\bar{\theta}_m$ to the null parameter space $H_0 : R\theta = \mathbf{0}$ and the distance from $R\bar{\theta}_m$ to space \mathbb{O}^+ . The larger the difference, the stronger the evidence against H_0 .

Combining individual Wald-type statistics

Another way of conducting an overall test based on m imputed datasets, $\{\mathbf{y}, X_*^{(l)}\}$, $l = 1, 2, \dots, m$, is to combine the following m Wald-type test statistics

$$W_{*l} = \left\| R\widehat{\theta}_{*l} \right\|_{R\widehat{U}_{n*l}R^T}^2 - \left\| R\widehat{\theta}_{*l} - \pi_{\widehat{\Sigma}_{n,m}^*}(R\bar{\theta}_m; \mathbb{O}^+) \right\|_{R\widehat{U}_{n*l}R^T}^2, \quad l = 1, 2, \dots, m.$$

Note that a simple average, such as $\bar{W}_m = \sum_{i=1}^m W_{*l}/m$, may lead to difficulties in computing the null distribution and making inference. Motivated by Meng and Rubin (1992) and Chapter 2 in this thesis, the individual test statistics may be combined as follows

$$T_{CW} = \frac{\bar{W}_m}{1 + r_m}.$$

Its null distribution will be discussed below.

The asymptotic null distributions for the proposed test statistics T_W and T_{CW} are given in the following theorem, which may be proved in a similar way as for Theorem 2.1 and Theorem 2.2.

Theorem 4.1. Consider the multivariate one-sided hypotheses (4.3). Assuming equal ratios of missing to observed information: $RB_nR^T \propto RU_nR^T$, the following results hold:

- (i) The proposed overall test statistics are asymptotically equal, i.e. $T_W = T_{CW} + o_p(1)$.

(ii) The null distribution of T_W can be approximated by the following $\bar{\chi}^2$ distribution

$$\Pr(T_W \leq c | R\boldsymbol{\theta} = \mathbf{0}) \approx \sum_{i=0}^r \omega_i(r, \widehat{\Sigma}_{n,m}^*, \mathbb{O}^+) \Pr(\chi_i^2 \leq c).$$

where χ_0^2 denotes the distribution that takes the value zero with probability one.

(iii) The tail probability of the above null distribution has the following lower and upper bounds:

$$\frac{\Pr(\chi_0^2 \geq c) + \Pr(\chi_1^2 \geq c)}{2} \leq \sum_{i=0}^r \omega_i(r, \widehat{\Sigma}_{n,m}^*, \mathbb{O}^+) \Pr(\chi_i^2 \leq c) \leq \frac{\Pr(\chi_{r-1}^2 \geq c) + \Pr(\chi_r^2 \geq c)}{2}.$$

4.5 Real Data Example

In this section, we apply the proposed multiple imputation tests T_W and T_{CW} to a real data example for the multivariate one-sided hypotheses (4.2). For comparison purpose, we also apply the naive CC and LVCF test. The data are from the anti-HIV treatment study as discussed in Chapter 1, Chapter 3, and Section 4.2. Figure 1.3 shows the detailed scatter plot of the data, where the observations for the same patients are combined by solid lines. In this study the HIV viral loads for 48 patients were measured semi-weekly in the first two weeks, weekly in the 3rd and 4th weeks, and monthly in the 2nd and 3rd months after treatment. To better investigate the viral load trend after treatment, baseline data before treatment are excluded, as in Wu and Wu (2002). Patients each have 2 - 8 measurements and there are 317 observations in total. Missing data appear in the two time-varying covariates with around 20% missing rates in both CD4 and CD8. In fact, quite a few missing cases were measured at different days from the HIV viral loads. The *missing at random* assumption is likely

to be satisfied because the missingness is unlikely to depend on the missing values.

Based on the computing scheme described in Section 4.3, we first generate $m(= 5)$ multiple imputation datasets for CD4 and CD8. The MLE's for the two parameters of interest: λ_1 and λ_2 based on 5 “complete datasets” are (.480, .04000), (.479, .03986), (.479, .03990), (.480, .03991), (.485, .04001) respectively. We calculated the combined MLE statistics as described in Section 4.4: $\bar{\theta}_m = (.481, .03992)$. The two testing statics are also obtained: $T_W = 326.9$ and $T_{CW} = 327.1$. Therefore we have strong evidence to reject H_0 , and we may conclude that the viral loads decline significantly during the anti-HIV treatment. The testing statistics based on CC and LVCF methods are 131.3 and 183.6, respectively. They reject H_0 too but with weaker evidence than the newly proposed tests T_W and T_{CW} .

4.6 Simulation Study

In this section we perform a simulation study to evaluate the proposed tests T_W and T_{CW} based on multiple imputations for multivariate one-sided testing problem (4.2) comparing to the CC method and LVCF method. We evaluate the methods by comparing both the sizes (type I errors) and the powers of the tests. The significance level $\alpha = 0.05$ is chosen.

The simulation study is based on the biphasic HIV dynamics NLME model (4.1). The sample size of subjects is 48, and each individual has the same number of repeated observations as in the real data example in the last section. The measurement times and covariate values are also the same as those in the real data example. The missing

values in CD4 and CD8 in the real data example are imputed by a randomly selected imputation dataset from last section. This way we obtain our “completely observed original” CD4 and CD8. The viral loads are then simulated from model (4.1) based on different settings of λ_1 and λ_2 . The true values for the other parameters are chosen to be similar to those estimated from the real data example in the previous section. For each of the above simulated dataset, we randomly generate missing values in CD4 and CD8 independently with missing rate 20% for each. Missing values are then imputed by $m = 5$ sets of imputations from the MCMC approach described in Section 4.3. All simulations are repeated for 500 times. Both the bound method and the substitution methods in Theorem 4.1 are used to calculate p-values.

Table 4.1 and 4.2 show the simulation results. We can see that our proposed new tests T_W and T_{CW} have lower type I error rates and are more powerful than the naive CC and LVCF methods. T_{CW} test tends to be the most powerful one among the four. Regarding the two p-value calculation methods, the bound method is more conservative because it takes the maximum of the null probability. Note that the substitution method for T_{CW} gives a type I error rate that is a bit over 0.05. This may be because that the covariance matrix $\Sigma_{n,m}$ is not well estimated under the small sample size ($n = 48$).

4.7 Discussion

We have shown that the new testing methods based on multiple imputations performed better than the naive CC and LVCF methods by a simulation study. However, due to

the limited scope of the simulation, more simulation studies based on different missing rates and different sample sizes are needed to provide further evidences. These are already under investigation and will be reported in the near future.

The methods proposed in this thesis can also be extended to other mixed effects models, such as generalized linear mixed models, frailty models, and semiparametric or nonparametric mixed effects models. Further researches are of great interest.

Table 4.1: Simulation results: bound method (n=48, missing rate = 20% on each covariate)

λ	T_W	T_{CW}	CC	LVCF
Type I Error Rates (in %)				
(0, 0)	2.7	4.5	3.4	4.3
Power Comparisons (in %)				
(0, 0.005)	7.6	13	3.7	3.2
(0.05, 0)	41	44	5.1	17
(0.02, 0.05)	38	45	9	10

Table 4.2: Simulation results: substitution method (n=48, missing rate = 20% on each covariate)

λ	T_W	T_{CW}	CC	LVCF
Type I Error Rates (in %)				
(0, 0)	4.2	6.3	4.5	5.1
Power Comparisons (in %)				
(0, 0.005)	11	15	6.1	4.6
(0.05, 0)	45	51	11	24
(0.02, 0.05)	39	49	15	15

5 Conclusion and Discussion

In previous chapters, we have proposed new testing methods for multivariate one-sided hypotheses for multivariate normal data and for mixed effects models, in the presence of missing data. Multiple imputation methods are used to impute the missing data, existing complete-data methods are used for the imputed datasets, and the results are appropriately combined to produce overall inferences. These proposed methods are evaluated by simulations and are shown to perform better than naive methods.

Specifically, in Chapter 2, for multivariate normal data with missing values, three testing methods for a multivariate one-sided hypothesis on the mean vector are developed. Based on multiply imputed datasets, the three methods combine sufficient statistics, Wald statistics, and LRT statistics, respectively. The simulation study shows that the new tests have smaller type I error rates and are more powerful than the naive CC method.

In Chapter 3, for multivariate one-sided hypothesis for NLME models with complete data, we develop LRT, Wald test, and score test for one-sided hypotheses on the regression parameters as well as on variance components. The proposed methods are then applied to a biphasic HIV viral dynamic model to test the two-phase viral decay rates and to test variance components of the random effects. The simulation study shows that the proposed tests are more powerful than the conventional two-sided tests.

In Chapter 4, for multivariate one-sided hypothesis for NLME models with incomplete data, we develop three testing methods for one-sided hypothesis. The testing

statistics are based on multiple imputations and obtained by combining MLE's, Wald test statistics and LRT statistics. The simulation study again confirms that the new proposed tests attain the nominal significance level, have smaller type I error rates, and are more powerful than the naive CC and LVCF methods.

In the foregoing multivariate one-sided tests, to calculate the null probabilities, we use both the bound method, which is based on the upper bound of null probabilities over all nuisance parameters (for testing the mean vector, the nuisance parameters are the variance-covariance parameters), and substitution method, which replaces the nuisance parameters by their consistent estimates. The bound method generally performs well in all cases, although it may be somewhat conservative in some cases. The substitution method also perform reasonably well in most cases, but it may give somewhat liberal rejection rates when the sample size is not large, possibly due to the fact that the covariance matrix may not be estimated well for small sample sizes.

The scope of the simulation studies in this thesis is somewhat limited due to computational resources and time limitations. Currently, for NLME models as in Chapter 4, simulations based on different missing rates, and different missing data mechanisms, and more extensive simulations based on violations of the equal ratios of missing-to-observed information assumption are under investigation and will be reported in the near future.

In this thesis, we focus on multiple imputation methods for small m (say 3-5) and assume the equal ratios of missing to observed information assumption. It is interesting to study large m cases too. In such cases, intuitively we will not need the equal ratios

assumption, because the between imputation variation B_n can be efficiently estimated by B_m directly with large m . Theories and simulations for testing methods based on B_m are under investigation and will be reported in the near future.

In this thesis, for the case when the covariance matrix is unknown, the $\bar{\chi}^2$ distribution is developed based on asymptotic null distribution which requires large samples. Another option is to find the exact null distribution similar to (1.12) following Perlman (1969). However, the key assumption in Perlman (1969) that the estimated covariance matrix should follow a Wishart distribution can be hard to be incorporated in regression models and multiply imputed datasets. We have done some simulation studies on the exact null distributions by assuming the Wishart distribution of the estimated covariance matrix and approximating its d.f. based on the relevant two-sided results in Li et al. (1991). Preliminary simulation results show that this may be a promising approach to approximate the exact null distribution. Another alternative approach is to use bootstrap methods to approximate the complex null distributions using computer simulations. We did some preliminary studies on such approaches but substantial more work is required to provide a complete picture of these approaches.

In the process of writing this thesis, we have also considered testing the variance components in our NLME model. For mixed effects models, the variances of the random effects indicate the magnitudes of the between-individual variation. In model building, it is important to decide whether some random effects are really needed in the mixed effects models, since too many random effects may lead to too many parameters and computational difficulties. The problem is equivalent to test whether the variances

of some random effects are strictly positive or not. Let $\mathbf{d} = (d_{11}, d_{22}, \dots, d_{qq})^T$ be the squared roots of diagonal elements of the random effects covariance matrix D . In general, we may test the following multivariate one-sided hypotheses

$$H_0^* : \mathbf{d}_s = \mathbf{0} \quad \text{versus} \quad H_1^* : \mathbf{d}_s \geq \mathbf{0}, \mathbf{d}_s \neq \mathbf{0},$$

where \mathbf{d}_s is a subset of \mathbf{d} . If we reject H_0^* , we should keep the corresponding random effects in the model; otherwise, no random effects are needed for the corresponding parameters. However the underlying theories do not directly follow from the theories in Chapter 3, because the true parameter values may be on the boundary of the parameter space. Even though by directly applying the results in Chapter 3, we obtained some fairly good results, the theories need to be formalized. Also, as we discussed in Chapter 3, the GHQ method probably should be used instead of the linearization method for the analyses of the random effects.

In this thesis, the MI method is used to handle the missing data problem. One of its main advantages is that existing complete-data methods on multivariate one-sided tests can be used, although the way to combine the results may not be straightforward. Another commonly used approach for missing data is based on the observed-data likelihood method. That is, multivariate one-sided testing methods can be developed based on the observed-data likelihood, typically implemented by an EM algorithm. This approach can be computationally intensive, and much effort is required to develop the theoretical results of the testing methods. This will be one of our future research topics.

The methods proposed in this thesis can be extended to other mixed effects mod-

els, such as generalized linear mixed models, frailty models, and semiparametric or nonparametric mixed effects models. The extensions are conceptually straightforward, but the technical details may be complicated in some cases and the performances of the order-restricted tests under different models need to be evaluated separately.

The proposed methods can also be extended to the case when the missing data are nonignorable. In this case, one needs to incorporate the missing data mechanism when generating multiple imputations. The remaining developments are then similar. This is another possible future research topic.

Another extension is to develop multivariate one-sided tests for multivariate categorical data. In this case, a multinomial model may be assumed for the data and contingency tables are used. Although some general ideas for multivariate normal data can also be used in this case, the extension is not necessarily straightforward in some cases. Agresti and Coull have done much work in this area, see for example Agresti and Coull (2002) and Agresti (2002).

References

- Agresti, A. (2002). *Categorical Data Analysis, 2nd edition*. Wiley.
- Agresti, A. and Coull, B. A. (2002). The analysis of contingency tables under inequality constraints. *Journal of Statistical Planning and Inference*, 107, 4573.
- Anderson, T.W. (1984). *An introduction to multivariate statistical analysis*. Wiley.
- Cassela, G. and Berger, R. L. (2001). *Statistical Inference*. Duxbury Press.
- Greene, W. H. (2003). *Econometric Analysis*. Prentice Hall.
- Gourieroux, C. and Monfort, A. (1995). *Statistics and Econometric Models*. Cambridge University Press.
- Joe, H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis*, 52, 5066-74.
- Johnson, N.L. and Kotz, S. (1973). *Continuous Univariate Distributions-2*. New York: John Wiley.
- Li, K.H. (1985). Hypotheses testing in multiple imputation-with emphasis on mixed-up frequencies in contingency tables. Ph.D. thesis, University of Chicago, Dept. of Statistics.
- Li, K.H., Raghunathan, T.E. and Rubin, D.B. (1991). Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *Journal of the American Statistical Association*, 86, 1065-73.

- Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, Second Edition, Wiley-Interscience.
- Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46, 673-687.
- Meng, X.L. and Rubin, D.B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, 79, 103-111.
- Murphy, S.A., Johnson, L.C., Wu, L., Fan, J.J. and Lohan, J. (2003). Bereaved parents' outcomes 4 to 60 months after their children's deaths by accident, suicide, or homicide: a comparative study demonstrating differences. *Death Studies*, 27, 39-61.
- Perlman, M.D. (1969). One-sided testing problems in multivariate analysis. *The Annals of Mathematical Statistics*, 40, 549-567.
- Perlman, M.D. and Wu, L. (2006). Some improved tests for multivariate one-sided hypotheses. *Metrika*, 64, 23-39.
- Pinheiro, J.C. and Bates, D.M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4, 12-35.
- Raghunathan, T.E. (1987). Large sample significance levels from multiply-imputed data. Ph.D. thesis, Harvard University, Dept. of Statistics.

- Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581-92.
- Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley.
- Rubin, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91, 473-389.
- Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. Chapman & Hall/CRC.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.
- Silvapulle, M.J. and Sen, P.K. (2004). *Constrained Statistical Inference : Inequality, Order, and Shape Restrictions*. Wiley.
- Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50, 1171-1177.
- Van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer.
- Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, 80, 791-795.
- Wu, H. and Ding, A. (1999). Population HIV-1 dynamics in Vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics*, 55, 410-418.

- Wu, L. (2004). Exact and approximate inferences for nonlinear mixed-effects models with missing covariates. *Journal of the American Statistical Association*, 99, 700-709.
- Wu, L. (2009). *Mixed Effects Models for Complex Data*. Chapman and Hall/CRC.
- Wu, L. and Wu, H. (2002). Missing time-dependent covariates in human immunodeficiency virus dynamic models. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 51, 297-318.

Appendix

A.1 Proof of Theorem 2.1

Consider testing the hypotheses $H_0 : \boldsymbol{\mu} \in \mathbb{O}^-$ versus $H_1 : \boldsymbol{\mu} \in \mathbb{R}^p \setminus \mathbb{O}^-$ (i.e., the testing problem (1.9)). In Section 2.3.1 we have that $\bar{\boldsymbol{\mu}}_m \sim N(\boldsymbol{\mu}, \Sigma_{n,m})$ (see (2.6)), where $\Sigma_{n,m} = U_n + (1 + m^{-1})B_n$. By substituting $\hat{\Sigma}_{n,m}$ (2.8) for $\Sigma_{n,m}$, we have the approximated distribution $\bar{\boldsymbol{\mu}}_m \sim N(\boldsymbol{\mu}, \hat{\Sigma}_{n,m})$. Therefore the least favorable null distribution of $T_W = \min_{\boldsymbol{\mu} \in \mathbb{O}^-} \{(\bar{\boldsymbol{\mu}}_m - \boldsymbol{\mu})^T \hat{\Sigma}_{n,m}^{-1} (\bar{\boldsymbol{\mu}}_m - \boldsymbol{\mu})\}$ can be approximated by the following $\bar{\chi}^2$ distribution based on the standard results reviewed in Section 1.4

$$\Pr(T_W \leq c | \boldsymbol{\theta} = \mathbf{0}) = \sum_{i=0}^p \omega_{p-i}(p, \hat{\Sigma}_{n,m}, \mathbb{O}^-) \Pr(\chi_i^2 \leq c).$$

As discussed in Section 1.4, the probability weights are sometimes hard to calculate, so we may obtain the cut-off values by simulation methods. Alternatively, we can find the upper and lower bounds of the null probabilities. Specifically, since $\Pr(\chi_i^2 \geq c) \leq \Pr(\chi_j^2 \geq c)$ for $i < j$, $\sum_{i=0}^p \omega_i(p, U_n, \mathbb{O}^-) = 1$ and $0 \leq \omega_i(p, U_n, \mathbb{O}^-) \leq \frac{1}{2}$ for any variance-covariance matrix U_n (Silvapulle and Sen, 2004). Thus,

$$\begin{aligned} \sum_{i=0}^p \omega_{p-i} \Pr(\chi_i^2 \geq c) &\geq \sum_{i=1}^p \omega_{p-i} \Pr(\chi_1^2 \geq c) + \omega_p \Pr(\chi_0^2 \geq c) \\ &= (1 - \omega_p) \Pr(\chi_1^2 \geq c) + \omega_p \Pr(\chi_0^2 \geq c) \\ &\geq \frac{1}{2} \Pr(\chi_1^2 \geq c) + \frac{1}{2} \Pr(\chi_0^2 \geq c), \end{aligned}$$

where $\omega_i = \omega_i(p, U_n, \mathbb{O}^-)$.

Similarly, we can show that

$$\sum_{i=0}^p \omega_{p-i} \Pr(\chi_i^2 \geq c) \leq \frac{1}{2} \Pr(\chi_{p-1}^2 \geq c) + \frac{1}{2} \Pr(\chi_p^2 \geq c).$$

This completes the proof. ■

A.2 Proof of Theorem 2.2

Following Rubin (1987), Li et al. (1991), Meng and Rubin (1992) and Rubin (1996), \widehat{U}_{n*l} is asymptotically unbiased for U_n and it has lower-order point-wise variances than U_n . Noting that all components of U_n ($= \frac{1}{n}\Sigma$) converge to 0, it follows that the point-wise variances of \widehat{U}_{n*l} converge to 0, which implies

$$\widehat{U}_{n*l} = U_n + o_p(1) \quad \text{and} \quad \bar{U}_m = U_n + o_p(1),$$

where the notation $o_p(1)$ is being abused because it actually indicates a matrix with all entries being $o_p(1)$.

By the Cholesky decomposition, U_n^{-1} can be factorized as $U_n^{-1} = \mathbf{A}^T \mathbf{A}$, where

$\mathbf{A}U_n^{-1}\mathbf{A}^T = \mathbf{I}$. So \overline{W}_m in (2.11) is given by

$$\begin{aligned}\overline{W}_m &= \frac{1}{m} \sum_{l=1}^m (\hat{\boldsymbol{\mu}}_{*l} - \pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-))^T \hat{U}_{n*l}^{-1} (\hat{\boldsymbol{\mu}}_{*l} - \pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-)) \\ &= \frac{1}{m} \sum_{l=1}^m (\mathbf{A}\hat{\boldsymbol{\mu}}_{*l} - \mathbf{A}\pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-))^T (\mathbf{A}\hat{\boldsymbol{\mu}}_{*l} - \mathbf{A}\pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-)) + o_p(1)\end{aligned}\quad (5.1)$$

$$\begin{aligned}&= \frac{1}{m} \sum_{l=1}^m (\mathbf{A}\hat{\boldsymbol{\mu}}_{*l} - \mathbf{A}\pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-))^T \cdot \frac{1}{m} \sum_{l=1}^m (\mathbf{A}\hat{\boldsymbol{\mu}}_{*l} - \mathbf{A}\pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-)) \\ &\quad + \frac{1}{m} \sum_{l=1}^m (\mathbf{A}\hat{\boldsymbol{\mu}}_{*l} - \mathbf{A}\overline{\boldsymbol{\mu}}_m)^T (\mathbf{A}\hat{\boldsymbol{\mu}}_{*l} - \mathbf{A}\overline{\boldsymbol{\mu}}_m) + o_p(1)\end{aligned}\quad (5.2)$$

$$\begin{aligned}&= (\overline{\boldsymbol{\mu}}_m - \pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-))^T U_n^{-1} (\overline{\boldsymbol{\mu}}_m - \pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-)) \\ &\quad + \frac{1}{m} \sum_{l=1}^m (\hat{\boldsymbol{\mu}}_{*l} - \overline{\boldsymbol{\mu}}_m)^T U_n^{-1} (\hat{\boldsymbol{\mu}}_{*l} - \overline{\boldsymbol{\mu}}_m) + o_p(1)\end{aligned}\quad (5.3)$$

$$\begin{aligned}&= (\overline{\boldsymbol{\mu}}_m - \pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-))^T U_n^{-1} (\overline{\boldsymbol{\mu}}_m - \pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-)) \\ &\quad + \frac{m-1}{m} \text{tr}(B_m U_n^{-1}) + o_p(1)\end{aligned}\quad (5.4)$$

$$= (\overline{\boldsymbol{\mu}}_m - \pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-))^T \overline{U}_m^{-1} (\overline{\boldsymbol{\mu}}_m - \pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-)) + \frac{m-1}{m} \text{tr}(B_m \overline{U}_m^{-1}) + o_p(1)$$

$$= (\overline{\boldsymbol{\mu}}_m - \pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-))^T \overline{U}_m^{-1} (\overline{\boldsymbol{\mu}}_m - \pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-)) + \frac{(m-1)p}{m+1} r_m + o_p(1)$$

$$= (1 + r_m) T_W + \frac{(m-1)p}{m+1} r_m + o_p(1).\quad (5.5)$$

In the above arguments, the equivalence between (5.1) and (5.2) holds by letting

$Q_l = \mathbf{A}\hat{\boldsymbol{\mu}}_{*l} - \mathbf{A}\pi_{\hat{\Sigma}_{n,m}}(\overline{\boldsymbol{\mu}}_m; \mathbb{O}^-)$ in the following equation

$$\frac{1}{m} \sum_{l=1}^m (Q_l - \overline{Q}_m)^T (Q_l - \overline{Q}_m) = \frac{1}{m} \sum_{l=1}^m Q_l^T Q_l - \overline{Q}_m^T \overline{Q}_m,$$

where $\overline{Q}_m = \frac{1}{m} \sum_{l=1}^m Q_l$. The equivalence between (5.3) and (5.4) holds by noting $B_m =$

$\frac{1}{m-1} \sum_{l=1}^m (\hat{\boldsymbol{\mu}}_{*l} - \overline{\boldsymbol{\mu}}_m)(\hat{\boldsymbol{\mu}}_{*l} - \overline{\boldsymbol{\mu}}_m)^T$ in (2.5) and some matrix algebra.

Now from (5.5) we have $T_{CW} = \frac{\overline{W}_m - \frac{(m-1)p}{m+1} r_m}{1+r_m} = T_W + o_p(1)$. So the null distribution

of T_{CW} is approximately equal to that of T_W when n is large. This completes the proof.

■

A.3 Proof of Theorem 3.1

Proof: First we derive a quadratic approximation of $l_n(\boldsymbol{\theta})$ by a second-order Taylor series expansion about $\boldsymbol{\theta}_0$:

$$\begin{aligned}
l_n(\boldsymbol{\theta}) &= l_n(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \nabla l_n(\boldsymbol{\theta}_0) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \nabla^2 l_n(\boldsymbol{\theta}_0) (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\
&\quad + n \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^3 O_p(1) \\
&\stackrel{\text{by (3.5)}}{=} l_n(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \nabla l_n(\boldsymbol{\theta}_0) - \frac{1}{2} n (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T V_{\boldsymbol{\theta}_0} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\
&\quad + n \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^3 O_p(1) + o_p(1) \\
&= l_n(\boldsymbol{\theta}_0) + \mathbf{u}^T V_{\boldsymbol{\theta}_0} Z_n - \frac{1}{2} \mathbf{u}^T V_{\boldsymbol{\theta}_0} \mathbf{u} + n^{-1/2} \|\mathbf{u}\|^3 O_p(1) + o_p(1) \\
&= l_n(\boldsymbol{\theta}_0) + \mathbf{u}^T V_{\boldsymbol{\theta}_0} Z_n - \frac{1}{2} \mathbf{u}^T V_{\boldsymbol{\theta}_0} \mathbf{u} - \frac{1}{2} Z_n^T V_{\boldsymbol{\theta}_0} Z_n + \frac{1}{2} Z_n^T V_{\boldsymbol{\theta}_0} Z_n + o_p(1) \\
&= l_n(\boldsymbol{\theta}_0) - \frac{1}{2} (Z_n - \mathbf{u})^T V_{\boldsymbol{\theta}_0} (Z_n - \mathbf{u}) + \frac{1}{2} Z_n^T V_{\boldsymbol{\theta}_0} Z_n + o_p(1) \tag{5.6}
\end{aligned}$$

where $Z_n = n^{-1/2} V_{\boldsymbol{\theta}_0}^{-1} \nabla l_n(\boldsymbol{\theta}_0)$ and $\mathbf{u} = n^{1/2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$. Note that the supreme of the remainder term converges to 0 in probability, i.e., $\sup_{\|\mathbf{u}\| < K} |n^{-1/2} \|\mathbf{u}\|^3 O_p(1)| = o_p(1)$ for any given $K > 0$.

The quadratic approximation (5.6) is essential in developing the null distribution of LRT for the multivariate one-sided test (3.3):

$$H_0 : R\boldsymbol{\theta} = \mathbf{0} \quad \text{vs} \quad H_1 : R\boldsymbol{\theta} \geq \mathbf{0}, R\boldsymbol{\theta} \neq \mathbf{0}.$$

According to the expression (5.6), following Silvapulle and Sen (2004), the LRT (3.7)

can be written as:

$$\begin{aligned}
T_{LR} &= 2[l_n(\tilde{\boldsymbol{\theta}}) - l_n(\bar{\boldsymbol{\theta}})] = 2[\sup_{R\boldsymbol{\theta} \geq \mathbf{0}} l_n(\boldsymbol{\theta}) - \sup_{R\boldsymbol{\theta} = \mathbf{0}} l_n(\boldsymbol{\theta})] \\
&= 2\{[l_n(\boldsymbol{\theta}_0) - \inf_{R\boldsymbol{\theta} \geq \mathbf{0}} \frac{1}{2}(Z_n - \mathbf{u})^T V_{\boldsymbol{\theta}_0} (Z_n - \mathbf{u}) + \frac{1}{2} Z_n^T V_{\boldsymbol{\theta}_0} Z_n + o_p(1)] \\
&\quad - [l_n(\boldsymbol{\theta}_0) - \inf_{R\boldsymbol{\theta} = \mathbf{0}} \frac{1}{2}(Z_n - \mathbf{u})^T V_{\boldsymbol{\theta}_0} (Z_n - \mathbf{u}) + \frac{1}{2} Z_n^T V_{\boldsymbol{\theta}_0} Z_n + o_p(1)]\} \\
&= \inf_{R\mathbf{u} = \mathbf{0}} (Z_n - \mathbf{u})^T V_{\boldsymbol{\theta}_0} (Z_n - \mathbf{u}) - \inf_{R\mathbf{u} \geq \mathbf{0}} (Z_n - \mathbf{u})^T V_{\boldsymbol{\theta}_0} (Z_n - \mathbf{u}) + o_p(1). \quad (5.7)
\end{aligned}$$

Note that in the last equation (5.7), $\mathbf{u} = n^{1/2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ and $R\boldsymbol{\theta}_0 = \mathbf{0}$. Therefore, $R\mathbf{u} = \mathbf{0}$ if and only if $R\boldsymbol{\theta} = \mathbf{0}$, and $R\mathbf{u} \geq \mathbf{0}$ if and only if $R\boldsymbol{\theta} \geq \mathbf{0}$.

Recall condition (3.6) that $n^{-1/2} \nabla l_n(\boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, V_{\boldsymbol{\theta}})$, so by the Slutsky theorem we have

$$Z_n \equiv n^{-1/2} V_{\boldsymbol{\theta}_0}^{-1} \nabla l_n(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, V_{\boldsymbol{\theta}_0}^{-1}), \quad n \rightarrow \infty.$$

Based on the results for multivariate normal random variables, as reviewed in Section 1.4, and given the fact that (5.7) is continuous in Z_n , it follows that $T_{LR} \xrightarrow{d} \bar{\chi}^2$ as in (3.8). ■

A.4 Proof of Theorem 3.2

Proof of (3.10): Following Silvapulle and Sen (2004), we have

$$\Pr(\chi_i^2 \geq c) \leq \Pr(\chi_j^2 \geq c), \quad \text{for } i < j,$$

$\sum_{i=0}^r \omega_i(r, V_{\theta_0}) = 1$, and $0 \leq \omega_i(r, V_{\theta_0}) \leq \frac{1}{2}$ for any positive definite matrix V_{θ_0} . For notational convenience, we denote $\omega_i(r, V_{\theta_0})$ by ω_i in the following proof. We have

$$\begin{aligned} \sum_{i=0}^r \omega_i \Pr(\chi_i^2 \geq c) &\geq \sum_{i=1}^r \omega_i \Pr(\chi_1^2 \geq c) + \omega_0 \Pr(\chi_0^2 \geq c) \\ &= (1 - \omega_0) \Pr(\chi_1^2 \geq c) + \omega_0 \Pr(\chi_0^2 \geq c) \\ &\geq \frac{1}{2} \Pr(\chi_1^2 \geq c) + \frac{1}{2} \Pr(\chi_0^2 \geq c). \end{aligned}$$

Similarly, we have

$$\begin{aligned} \sum_{i=0}^r \omega_i \Pr(\chi_i^2 \geq c) &\leq \sum_{i=0}^{r-1} \omega_i \Pr(\chi_{r-1}^2 \geq c) + \omega_r \Pr(\chi_r^2 \geq c) \\ &= (1 - \omega_r) \Pr(\chi_{r-1}^2 \geq c) + \omega_r \Pr(\chi_r^2 \geq c) \\ &\leq \frac{1}{2} \Pr(\chi_{r-1}^2 \geq c) + \frac{1}{2} \Pr(\chi_r^2 \geq c). \end{aligned}$$

These proves Theorem 3.2. ■

A.5 Proof of Theorem 3.3

First, let us briefly review the Lagrange multiplier $\bar{\lambda}$ and the Kühn-Tücker multiplier $\tilde{\lambda}$ for calculating the MLE's $\bar{\theta}$ and $\tilde{\theta}$ under the null and alternative hypothesis parameter constraints respectively. The Lagrangian function is given by

$$L_n(\theta) = l_n(\theta) + (R\theta)^T \lambda.$$

Under the null constraint $R\theta = \mathbf{0}$, the first-order conditions for $\bar{\lambda}$ and $\bar{\theta}$ are respectively

$$\nabla l_n(\bar{\theta}) + R^T \bar{\lambda} = \mathbf{0},$$

$$R\bar{\theta} = \mathbf{0}.$$

Under the alternative constraint $R\boldsymbol{\theta} \geq \mathbf{0}$, the first-order conditions for $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\boldsymbol{\theta}}$ are respectively

$$\nabla l_n(\tilde{\boldsymbol{\theta}}) + R^T \tilde{\boldsymbol{\lambda}} = \mathbf{0}, \quad (5.8)$$

$$\text{sign constraints : } R\tilde{\boldsymbol{\theta}} \geq \mathbf{0}, \tilde{\boldsymbol{\lambda}} \geq \mathbf{0}$$

$$\text{exclusion constraints : } (R\tilde{\boldsymbol{\theta}})^T \tilde{\boldsymbol{\lambda}} = \mathbf{0}. \quad (5.9)$$

The second-order Taylor series expansion about $\boldsymbol{\theta}_0$ gives the following expression for the LRT statistic T_{LR} :

$$\begin{aligned} T_{LR} &= 2[l_n(\tilde{\boldsymbol{\theta}}) - l_n(\bar{\boldsymbol{\theta}})] \\ &= 2[(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \nabla l_n(\boldsymbol{\theta}_0) - (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \nabla l_n(\boldsymbol{\theta}_0) + \frac{1}{2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \nabla^2 l_n(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &\quad - \frac{1}{2}(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \nabla^2 l_n(\boldsymbol{\theta}_0)(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)] + o_p(1). \end{aligned} \quad (5.10)$$

Also the first-order Taylor series expansion about $\boldsymbol{\theta}_0$ gives the following expression for (5.8):

$$\begin{aligned} \nabla l_n(\boldsymbol{\theta}_0) + \nabla^2 l_n(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + R^T \tilde{\boldsymbol{\lambda}} + o_p(1) &= 0, \\ \nabla l_n(\boldsymbol{\theta}_0) &= -\nabla^2 l_n(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - R^T \tilde{\boldsymbol{\lambda}} + o_p(1). \end{aligned} \quad (5.11)$$

Similarly, we have the following for $\bar{\boldsymbol{\theta}}$

$$\nabla l_n(\boldsymbol{\theta}_0) = -\nabla^2 l_n(\boldsymbol{\theta}_0)(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - R^T \bar{\boldsymbol{\lambda}} + o_p(1). \quad (5.12)$$

Combining (5.10), (5.11) and (5.12), we have

$$\begin{aligned} T_{LR} &= -(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \nabla^2 l_n(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \nabla^2 l_n(\boldsymbol{\theta}_0)(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &\quad + (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T R^T \bar{\boldsymbol{\lambda}} - (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T R^T \tilde{\boldsymbol{\lambda}}. \end{aligned} \quad (5.13)$$

In (5.13), we have

$$(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T R^T \bar{\boldsymbol{\lambda}} = 0$$

immediately from the constraints: $R\boldsymbol{\theta}_0 = \mathbf{0}$ and $R\bar{\boldsymbol{\theta}} = \mathbf{0}$. Also under $H_0 : R\boldsymbol{\theta} = \mathbf{0}$, the first-order Taylor series expansion about $\boldsymbol{\theta}_0$ gives the following expression for (5.9):

$$(R\boldsymbol{\theta}_0)^T \tilde{\boldsymbol{\lambda}} + (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T R^T \tilde{\boldsymbol{\lambda}} + o_p(1) = 0,$$

$$(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T R^T \tilde{\boldsymbol{\lambda}} + o_p(1) = 0.$$

Thus, we have the simplified expression for T_{LR} :

$$T_{LR} = -(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \nabla^2 l_n(\boldsymbol{\theta}_0) (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \nabla^2 l_n(\boldsymbol{\theta}_0) (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(1).$$

Now, replacing $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ by $\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}} + \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$, we have

$$T_{LR} = -(\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^T \nabla^2 l_n(\boldsymbol{\theta}_0) (\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}) - 2(\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^T \nabla^2 l_n(\boldsymbol{\theta}_0) (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(1).$$

On the other side, equations (5.11) and (5.12) give us the following

$$\nabla^2 l_n(\boldsymbol{\theta}_0) (\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}) = -R^T (\tilde{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}) + o_p(1). \quad (5.14)$$

So we have

$$\begin{aligned} T_{LR} &= -(\tilde{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}})^T R (\nabla^2 l_n(\boldsymbol{\theta}_0))^{-1} R^T (\tilde{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}) - 2(\tilde{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}})^T R (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(1) \\ &= -(\tilde{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}})^T R (\nabla^2 l_n(\boldsymbol{\theta}_0))^{-1} R^T (\tilde{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}) + o_p(1) \end{aligned} \quad (5.15)$$

where (5.15) is obtained by noting $R\boldsymbol{\theta}_0 = \mathbf{0}$ and $R\bar{\boldsymbol{\theta}} = \mathbf{0}$. Also (5.14) gives

$$\begin{aligned} R\tilde{\boldsymbol{\theta}} &= R(\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}) \\ &= -R(\nabla^2 l_n(\boldsymbol{\theta}_0))^{-1} R^T (\tilde{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}), \\ \text{so } (\tilde{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}) &= -[R(\nabla^2 l_n(\boldsymbol{\theta}_0))^{-1} R^T]^{-1} R\tilde{\boldsymbol{\theta}}. \end{aligned} \quad (5.16)$$

Substituting (5.16) into (5.15), we finally have the following desired expression for T_{LR} .

$$\begin{aligned}
T_{LR} &= -(R\tilde{\boldsymbol{\theta}})^T [R(\nabla^2 l_n(\boldsymbol{\theta}_0))^{-1} R^T]^{-1} (R\tilde{\boldsymbol{\theta}}) + o_p(1) \\
&= n(R\tilde{\boldsymbol{\theta}})^T [R\hat{V}^{-1}(\tilde{\boldsymbol{\theta}})R^T]^{-1} (R\tilde{\boldsymbol{\theta}}) + o_p(1) \\
&= T_W + o_p(1)
\end{aligned}$$

where $\hat{V}(\tilde{\boldsymbol{\theta}}) = -n^{-1}\nabla^2 l_n(\tilde{\boldsymbol{\theta}})$.

Next, let us establish the asymptotic equivalence between T_{LR} and T_S . By noting that the rank of R is r , $\tilde{\lambda}$ in (5.8) can be solved by the following expression:

$$\tilde{\lambda} = -[R\hat{V}^{-1}(\tilde{\boldsymbol{\theta}})R^T]^{-1} R\hat{V}^{-1}(\tilde{\boldsymbol{\theta}})\nabla l_n(\tilde{\boldsymbol{\theta}}). \quad (5.17)$$

Similarly we have the following expression for $\bar{\lambda}$:

$$\bar{\lambda} = -[R\hat{V}^{-1}(\bar{\boldsymbol{\theta}})R^T]^{-1} R\hat{V}^{-1}(\bar{\boldsymbol{\theta}})\nabla l_n(\bar{\boldsymbol{\theta}}). \quad (5.18)$$

On the other hand, by noting $\hat{V}(\tilde{\boldsymbol{\theta}}) = -n^{-1}\nabla^2 l_n(\tilde{\boldsymbol{\theta}})$ and (5.15) and (5.14), T_{LR} can be expressed as

$$\begin{aligned}
T_{LR} &= n(\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^T \hat{V}^{-1}(\tilde{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}) + o_p(1) \\
&= \frac{1}{n}(\tilde{\lambda} - \bar{\lambda})^T R^T \hat{V}^{-1}(\tilde{\boldsymbol{\theta}})R(\tilde{\lambda} - \bar{\lambda}) + o_p(1).
\end{aligned} \quad (5.19)$$

Finally, substituting (5.17) and (5.18) into (5.19) yields the desired result:

$$\begin{aligned}
T_{LR} &= \frac{1}{n}(\nabla l_n(\tilde{\boldsymbol{\theta}}) - \nabla l_n(\bar{\boldsymbol{\theta}}))^T \hat{V}^{-1}(\tilde{\boldsymbol{\theta}})R^T [R\hat{V}^{-1}(\tilde{\boldsymbol{\theta}})R^T]^{-1} R\hat{V}^{-1}(\tilde{\boldsymbol{\theta}})(\nabla l_n(\tilde{\boldsymbol{\theta}}) \\
&\quad - \nabla l_n(\bar{\boldsymbol{\theta}})) + o_p(1) \\
&= T_S + o_p(1).
\end{aligned} \quad (5.20)$$

This completes the proof. ■

A.6 Lindeberg Condition

Let the settings be the same as those in Section 3.2. The Lindeberg condition is as follows

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}\{\|\nabla l_i(\boldsymbol{\theta}|\mathbf{y}_i)\|^2 \mathbf{1}\{\|\nabla l_i(\boldsymbol{\theta}|\mathbf{y}_i)\| > \varepsilon\sqrt{n}\}\} \xrightarrow{p} 0,$$

as $n \rightarrow \infty$ for each $\varepsilon > 0$.