MULTIVARIATE ONE-SIDED TESTS FOR NONLINEAR MIXED-EFFECTS MODELS WITH INCOMPLETE DATA

YIXIN ZHANG

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS
YORK UNIVERSITY
TORONTO, ONTARIO

2023 January

Abstract

Nonlinear mixed-effects (NLME) models are widely used in the analysis of longitudinal studies. The parameters in an NLME model typically have meaningful scientific interpretations, and these parameters may have some natural order restrictions such as being strictly positive. The problems of testing parameters with order restrictions are known as multivariate one-sided hypothesis testing. However, multivariate one-sided testing problems in NLME models have not been discussed thoroughly.

In many longitudinal studies, the inter-individual variation can be partially explained by the time-varying covariates which, however, may be measured with substantial errors. Moreover, censoring and non-ignorable missingness in response are very common in practice. Standard testing procedures ignoring covariate measurement errors and/or response censoring/missingness may lead to biased results. We propose multiple imputation methods to address the foregoing data complication. The multiple imputation methods allow us to use existing "complete-data" hypoth-

esis testing procedures for parameters with order restrictions. In this thesis, we propose testing statistics for the multivariate one-sided testing problems in NLME models with: (i) mis-measured covariates, (ii) both mis-measured covariates and left-censored response, and (iii) both mis-measured covariates and non-ignorable missing response, which are discussed in Chapters 2-4, respectively. Some asymptotic null distributions of the proposed test statistics are derived. The proposed methods are illustrated by HIV data examples and evaluated by simulation studies under different scenarios. Simulation results have shown the power advantage of the proposed testing statistics over the commonly used ones.

Acknowledgements

I would like to first gratefully and sincerely thank my advisor, Dr. Wei Liu. After working with her for six years, I still find it difficult to believe that she is so knowledgeable in her field. I have been amazingly fortunate to have an advisor who gave me insightful guidance at each stage of my research, and at the same time taught me how to question thoughts and express ideas. It was truly a once-in-a-lifetime experience to receive her support and encouragement. I am so proud to be one of her students.

I wish to express my sincerest thanks to the members of my committee, Xin Gao and Jianhong Wu, for their insightful comments and suggestions on my work.

I would like to thank all people in the Department of Mathematics and Statistics at York University, especially those friends of mine for their valuable discussions and accessibility.

Finally, I am so grateful to my mom, Jie Lin, for her understanding, endless patience, and encouragement when it was most required. Her unwavering love helped

me through each step of my doctoral study. It was under her watchful eyes that I gained so much drive and an ability to tackle challenges.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Background

Some order constraints on the parameters of a statistical model arise frequently in many practical problems. In hypothesis testing, it is desirable to incorporate such constraints on the parameters of interest to have a more efficient hypothesis test, which is a so-called order-restricted hypotheses test. A one-sided hypothesis may be considered as a special case of order-restricted hypotheses.

We take the clinical trials as an example. A treatment is expected to be at least as effective as a control. More specifically, we start with a simple example of testing one parameter. Let μ represent the efficacy of a new drug or a new treatment. And μ could be $\geq 0, \leq 0$, or $= 0$, indicating that the drug is effective, toxic, or neutral, respectively. To evaluate the efficacy of the drug or the treatment, we may consider the following one-sided test,

$$H_0 : \mu \leq 0 \quad v.s. \quad H_1 : \mu > 0. \tag{1.1}$$

Apparently, the above one-sided test is more appropriate than the following two-sided test,

$$H_0 : \mu = 0 \quad v.s. \quad H_1 : \mu \neq 0. \tag{1.2}$$

Furthermore, we may even consider the following one-sided test if subject-area knowledge indicates that the drug is unlikely to be toxic,

$$H_0 : \mu = 0 \quad v.s. \quad H_1 : \mu > 0. \tag{1.3}$$

Now we briefly illustrate the ideas for testing one-parameter problems (1.1), (1.2) and (1.3). Suppose that $X_1, X_2, \cdots, X_n$ are n independent and identically distributed observations drawn from $N(\mu, \sigma^2)$. Given the realization of $(x_1, x_2, \cdots, x_n)$ of $(X_1, X_2, \cdots, X_n)$, the log-likelihood function is written as

$$l_n(\mu, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 . \tag{1.4}$$

Then we can derive the maximum likelihood estimate (MLE) of the mean parameter μ under different constraints:

- Without any constraint on μ, the MLE of μ is $\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.
- Under the one-sided constraint $\mu \leq 0$, the MLE of μ is $\hat{\mu}^- = \begin{cases} \bar{x}, & \bar{x} \leq 0 \\ 0, & \bar{x} > 0 \end{cases}$.
- Under the one-sided constraint $\mu \geq 0$, the MLE of μ is $\hat{\mu}^+ = \begin{cases} \bar{x}, & \bar{x} \geq 0 \\ 0, & \bar{x} < 0 \end{cases}$.

- Under the constraint $\mu = 0$, the MLE of μ is $\hat{\mu}^0 = 0$.

Based on the two-sided hypothesis (1.2), the Wald test statistic is given as

$$T_W = \frac{\bar{x}^2}{\widehat{\text{Var}}(\bar{x})} = \frac{n^2\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

which has a large sample approximate χ_1^2 null distribution (given in Appendix (1.1)). Rewrite the test statistic as

$$T_W = \frac{\bar{x}}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \Big/ \sqrt{n}}.$$

Recall that the Student's test statistics is

$$T = \frac{\bar{x}}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \Big/ \sqrt{n}} = \frac{\sqrt{n-1}}{\sqrt{n}} T_W,$$

which has a Student's t-distribution with $(n-1)$ degrees freedom under H_0. In this case, the Wald test is equivalent to the Student's t-test.

However, the methods cannot be extended to multi-parameter problems directly. In this case, the Hotelling's T-square statistic instead of the Student's t-test is used for the two-sided hypotheses. And the null distribution of the corresponding Wald test statistic is not the familiar χ^2 distribution anymore since the standard regularity conditions are violated for the likelihood theory when the parameters have constraints. In the next section, we briefly review the multivariate one-sided or order-restricted testing problems.

1.2 Multivariate One-sided Hypothesis Testing

For the multivariate one-sided hypothesis with complete data, various testing methods such as the likelihood ratio, Wald, and score tests have been proposed [Silvapulle and Sen, 2005].

In this section, we briefly review the basic results for the multivariate one-sided or constrained hypothesis testing under multivariate normal populations. Suppose $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are n independent and identically distributed observations drawn from the p-dimension multivariate normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{X}_i = (X_{i1}, \cdots, X_{ip})^T$, $i = 1, 2, \cdots, n$. Let $\bar{\boldsymbol{X}}$ and S respectively denote the sample mean and sample variance. We begin with the following two-sided hypothesis

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \qquad \text{versus} \qquad H_1 : \boldsymbol{\mu} \neq \mathbf{0}. \tag{1.5}$$

When the covariance matrix Σ is known, the Wald test statistic can be written as the following form

$$T_W = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu}),$$

which follows a χ^2_p distribution.

When the covariance matrix Σ is unknown, we use the sample covariance S instead and have the following Hotelling's T-square statistic

$$T_H = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^T S^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu}),$$

with $\dfrac{n-p}{p(n-1)}T_H \sim F_{p,n-p}$ under H_0.

However, in the case of multivariate one-sided hypotheses, the testing statistics often involve orthogonal projections of the sample mean onto the one-sided parameter space, which leads to difficulties in the computation of the test statistics and the derivation of their null distributions. Consider the following multivariate one-sided hypothesis

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \qquad \text{versus} \qquad H_1 : \boldsymbol{\mu} \geq \mathbf{0},\ \boldsymbol{\mu} \neq \mathbf{0}. \tag{1.6}$$

Here $\boldsymbol{\mu} \geq \mathbf{0}$ means $\mu_i \geq 0$ for all i and $\boldsymbol{\mu} \neq \mathbf{0}$ means $\mu_i \neq 0$ for some i. Figure (1.1) shows an example of the parameter spaces for hypothesis (1.6) when $p = 2$ with Q_i denoting the quadrant i in the two-dimension plane, $i = 1, 2, 3, 4$.

When the covariance matrix Σ is known, the Wald test statistic for the multivariate one-sided test (1.6) is written as

$$\begin{aligned} T_1 =& n\bar{\boldsymbol{X}}\Sigma^{-1}\bar{\boldsymbol{X}} - \min_{\boldsymbol{\mu}\in\mathbb{O}^+} n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right)^T \Sigma^{-1}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right) \\ =& \left\|\pi_{\Sigma_n}\left(\bar{\boldsymbol{X}}; \mathbb{O}^+\right)\right\|_{\Sigma_n}^2, \end{aligned}$$

where $\Sigma_n = \Sigma/n$, and $\mathbb{O}^+ = \{(\mu_1, \cdots, \mu_p)^T | \mu_i \geq 0, i = 1, \cdots, p\}$, and $\pi_{\Sigma_n}\left(\bar{\boldsymbol{X}}; \mathbb{O}^+\right)$ is the orthogonal projection of $\bar{\boldsymbol{X}}$ onto $\mathbb{O}^+$ with respect to Σ_n. $\|\boldsymbol{x}\|_A^2$ is the l_2-norm with respect to matrix A which is defined as $\boldsymbol{x}^T A^{-1}\boldsymbol{x}$. The larger value of T_1 is, the stronger evidence against H_0.

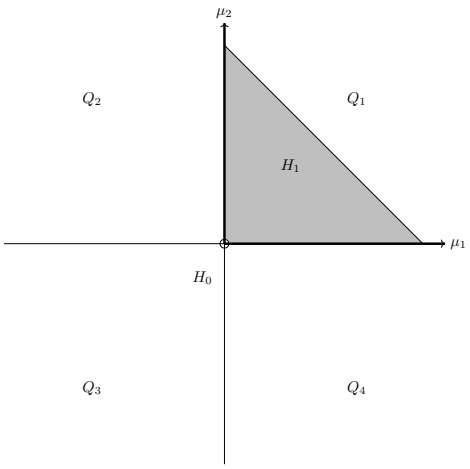


Figure 1.1: Null and alternative parameter spaces in hypothesis (1.6) for $p = 2$.

To determine critical values or p-values of the above test, we need to find the null distribution of the test statistic, i.e., the distribution of T_1 if the null hypothesis H_0 holds. Unlike the univariate case discussed in the previous section, the null distribution of T_1 in the multivariate case is different from that of the "two-sided" Hotelling's T-square test statistic, due to the shape of the null and alternative parameter spaces. In the general case, the null distribution of the Wald statistic T_1, when Σ is known, is a *chi-bar-square* ($\bar{\chi}^2$) distribution [Silvapulle and Sen, 2005]

$$\Pr(T_1 \leq c) = \sum_{i=0}^{p} \omega_i(p, \Sigma_n, \mathbb{O}^+) F_{\chi^2}(c; i), \text{ for any } c \geq 0, \tag{1.7}$$

where c is a constant, and $F_{\chi^2}(c; i)$ represents a probability that a χ_i^2 random

variable is less than or equal to c ($i = 1, \cdots, p$). χ_0^2 denotes a distribution of a random variable that takes value 0 with probability 1. The probability weight $\omega_i(p, \Sigma_n, \mathbb{O}^+)$ denotes the probability that $\pi_{\Sigma_n}(\boldsymbol{x}; \mathbb{O}^+) \in$ the i dimensional face of $\mathbb{O}^+$ with $\sum_{i=0}^{p} \omega_i(p, \Sigma_n, \mathbb{O}^+) = 1$.

A simple example for $p = 2$ and $\Sigma = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is given to illustrate the basic ideas. $\pi_{\Sigma_n}\left(\bar{\boldsymbol{X}}; \mathbb{O}^+\right)$ and T_1 take values in four possible regions, depending on which quadrant the value of $\bar{\boldsymbol{X}} = (\bar{X}_1, \bar{X}_2)^T$ falls (see Figure 1.1) as follows

$$\pi_{\Sigma_n}\left(\bar{\boldsymbol{X}}; \mathbb{O}^+\right) = \begin{cases} (\bar{X}_1, \bar{X}_2)^T & \text{if } \bar{\boldsymbol{X}} \in Q_1, \\ (0, \bar{X}_2)^T & \text{if } \bar{\boldsymbol{X}} \in Q_2, \\ (0, 0)^T & \text{if } \bar{\boldsymbol{X}} \in Q_3, \\ (\bar{X}_1, 0)^T & \text{if } \bar{\boldsymbol{X}} \in Q_4, \end{cases} \tag{1.8}$$

where Q_i is the quadrant i for $i = 1, 2, 3, 4$ in the two-dimension plane.

Correspondingly, we have

$$T_1 = \left\|\pi_{\Sigma_n}\left(\bar{\boldsymbol{X}}; \mathbb{O}^+\right)\right\|_{\Sigma_n}^2 = \begin{cases} n(\bar{X}_1^2 + \bar{X}_2^2) & \text{if } \bar{\boldsymbol{X}} \in Q_1, \\ n\bar{X}_2^2 & \text{if } \bar{\boldsymbol{X}} \in Q_2, \\ 0 & \text{if } \bar{\boldsymbol{X}} \in Q_3, \\ n\bar{X}_1^2 & \text{if } \bar{\boldsymbol{X}} \in Q_4. \end{cases} \tag{1.9}$$

Since Σ is a 2×2 identity matrix, $\bar{\boldsymbol{X}}$ are equally likely to fall in any of the four quadrants, i.e. $\Pr(\bar{\boldsymbol{X}} \in Q_i) = \frac{1}{4}$ for $i = 1, 2, 3, 4$. Thus, for any positive constant c,

we have

$$\Pr(T_1 \leq c) = \sum_{i=1}^{4} \Pr(T_1 \leq c|\bar{\boldsymbol{X}} \in Q_i) \Pr(\bar{\boldsymbol{X}} \in Q_i) = \frac{1}{4} \sum_{i=1}^{4} \Pr(T_1 \leq c|\bar{\boldsymbol{X}} \in Q_i).$$

Note that under H_0, $\bar{\boldsymbol{X}} = \left(\begin{smallmatrix} \bar{X}_1 \\ \bar{X}_2 \end{smallmatrix}\right) \sim N\left(\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1/n & 0 \\ 0 & 1/n \end{smallmatrix}\right)\right)$, then $\sqrt{n}\bar{X}_1$ and $\sqrt{n}\bar{X}_2$ independently follow the standard normal distribution and consequently $n(\bar{X}_1^2 + \bar{X}_2^2)$ has a χ_2^2 distribution. As a result, we have

$$\begin{aligned}
\Pr(T_1 \leq c|\bar{\boldsymbol{X}} \in Q_1) &= \Pr(n(\bar{X}_1^2 + \bar{X}_2^2) \leq c) = F_{\chi^2}(c; 2), \\
\Pr(T_1 \leq c|\bar{\boldsymbol{X}} \in Q_2) &= \Pr(n\bar{X}_2^2 \leq c) = F_{\chi^2}(c; 1), \\
\Pr(T_1 \leq c|\bar{\boldsymbol{X}} \in Q_3) &= \Pr(0 \leq c) = F_{\chi^2}(c; 0), \\
\Pr(T_1 \leq c|\bar{\boldsymbol{X}} \in Q_4) &= \Pr(n\bar{X}_1^2 \leq c) = F_{\chi^2}(c; 1),
\end{aligned}$$

where $F_{\chi^2}(c; i)$ denotes the probability that a random variable is less than or equal to c, following a χ^2-distribution with i degrees of freedom, $i = 0, 1, 2$. Thus we have the expression as in (1.7)

$$\Pr(T_1 \leq c) = \frac{1}{4} F_{\chi^2}(c; 0) + \frac{1}{2} F_{\chi^2}(c; 1) + \frac{1}{4} F_{\chi^2}(c; 2), \text{ for any } c \geq 0, \tag{1.10}$$

with $\omega_0(2, \Sigma_n, \mathbb{O}^+) = 1/4$, $\omega_1(2, \Sigma_n, \mathbb{O}^+) = 1/2$, and $\omega_2(2, \Sigma_n, \mathbb{O}^+) = 1/4$.

In general, the calculation of chi-bar-weights $\omega_i(p, \Sigma_n, \mathbb{O}^+)$ $(i = 0, 1, \cdots, p)$ could be difficult, so we can compute them by Monte Carlo simulations. Recall that $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are n independent and identically distributed observations from the

p-dimension multivariate normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$. Let $\bar{\boldsymbol{X}} = \sum_{i=1}^{n} \boldsymbol{X}_i/n$ and $\Sigma_n = \Sigma/n$. The steps are as follows:

(i) Generate $\bar{\boldsymbol{X}} \sim N(\boldsymbol{\mu}, \Sigma_n)$.

(ii) compute $\pi_{\Sigma_n}(\bar{\boldsymbol{X}}; \mathbb{O}^+) = \bar{\boldsymbol{X}} - \hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\mu}}$ is the solution for $\arg\min_{\boldsymbol{\mu} \in \mathbb{O}^+}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^T \Sigma_n^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$.

(iii) Repeat the foregoing two steps (i) and (ii) for (say) $N = 10000$ times.

Then, for $i = 0, 1, \cdots, p$, we have,

$$\omega_i(p, \Sigma_n, \mathbb{O}^+) \approx \frac{n_i}{N},$$

where n_i is the number of times that $\pi_{\Sigma_n}(\bar{\boldsymbol{X}}; \mathbb{O}^+)$ has exactly i positive components.

When Σ is unknown, a consistent estimate $\hat{\Sigma}$ should be found to replace Σ in T_1. By the Slutsky's theorem, the asymptotic null distributions of the new testing statistic obtained this way is the same as the null distributions of T_1 for the known Σ. This method is particularly useful in complicated cases where the exact null distributions of the test statistics are intractable. However, to obtain a good estimate of Σ, the sample size needs to be large enough. 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 regression models, and the asymptotic null distribution of the wald test is some *chi -bar-square* distribution under the

regularity conditions, as discussed in Silvapulle and Sen [2005].

1.3 Longitudinal Data and Mixed-effects Models

Compared with cross-sectional data, which have only one observation for a variable of interest for each independent individuals, longitudinal data are characterized by repeated observations or measurements collected at different time points for each independent individual. They arise frequently in practice, such as in clinical trials and epidemiology studies. When the variables are continuous and the number of the observations of each individual is the same, we may assume the longitudinal data have a certain multivariate normal structure. Longitudinal data combine elements of multivariate data and time series data.

Figure 1.2 and 1.3 show a longitudinal dataset in an anti-HIV treatment. In this study, 46 HIV infected patients received a potent anti-retro viral treatment. The viral load and some covariates such as CD4 cell count of each patient were repeatedly measured during a period of 48 weeks after the initiation of the treatment. We can see that although the trajectories vary a lot among patients, the general trend of the viral load (as shown in Figure 1.2) seems to decrease over time, which indicates that the anti-HIV treatment may be effective. Specifically, the overall viral load decreases rapidly at the beginning of treatment while some decreases slower and

some even rebounds later. Figure 1.3 shows the trajectories of the CD4 cell count are likely to have a quadratic trend during the anti-HIV treatment.

In longitudinal data analysis, we need to consider both the within-individual variation and the between-individual variation. Modeling the within-individual variation helps us to study the change of each individual over time, while modeling the between-individual variation allows us to understand the differences between individuals. In many longitudinal studies, the between-individual variation may be large and may be partially explained by covariates. There are two types of covariates: time-invariant covariates such as a patient's gender and time-varying covariates such as the blood pressure measured at different time points.

Mixed-effects models are one of the most commonly used approaches to analyze longitudinal data. A mixed-effects model assumes that the response is linked to a function of covariates with both fixed-effect coefficients and random-effect coefficients. In the following, we will review the linear and nonlinear mixed-effects models in details.

1.3.1 Linear Mixed-effects Models

Linear mixed-effects (LME) models assume the existence of the random effects and a linear relationship between the response variable and covariates and random

effects. Therefore, LME models can be constructed by adding appropriate random effects to linear regression models.

Suppose that there are n individuals in a study. Let y_{ij} be the response value for individual i $(i = 1, 2, \cdots, n)$ at time $t_{ij}, j = 1, \cdots, m_i$, and $\mathbf{y}_i = (y_{i1}, \cdots, y_{im_i})^T$ be the collection of response observations for individual i. A general LME model can be expressed as [Laird and Ware, 1982, Wu, 2009]

$$\begin{aligned} &\mathbf{y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \mathbf{e}_i, \qquad i = 1, \cdots, n, \\ &\mathbf{b}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, B), \qquad \mathbf{e}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, D_i), \end{aligned} \tag{1.11}$$

where X_i and Z_i are known as the design matrices with dimensions $m_i \times (p+1)$ and $m_i \times q$ respectively that may contain observation time and covariates, and Z_i is usually a submatrix of X_i, the vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^T$ denotes the population parameters (also called fixed-effects). In LME models, we assume the random-effects $\mathbf{b}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, B)$ with B being the unstructured variance-covariance matrix, and the within-individual error vector $\mathbf{e}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, D_i)$ with D_i being the $m_i \times m_i$ variance-covariance matrix. In practice , D_i is usually assumed to be equal to $\sigma^2 I_{m_i \times m_i}$, that is, the random errors are independent and have the same variance.

1.3.2 Generalized Linear Mixed-effects Models

Generalized linear mixed-effects (GLME) models extend linear mixed-effects models by allowing non-normal responses. Let $\mathbf{y}_i = (y_{i1}, \cdots, y_{im_i})$ be the vector of m_i repeated observations of the response for individual i, $i = 1, 2, \cdots, n$, which is assumed to follow a distribution from the exponential family including normal, binomial, Poisson, and some other distributions. Let $\mu_i = E(\mathbf{y}_i)$, and let $h(\cdot)$ be a known monotone link function. A general GLME model can be written as [Wu, 2009]

$$\begin{aligned} h(E(\mathbf{y}_i)) &= X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i, \qquad i = 1, \cdots, n. \\ \mathbf{b}_i &\overset{i.i.d.}{\sim} N(\mathbf{0}, B), \end{aligned} \tag{1.12}$$

where X_i and Z_i are known as the design matrices that may contain observation time and covariates, $\boldsymbol{\beta}$ denotes population parameters (also called fixed-effects), $\mathbf{b}_i$ denotes random-effects, and B is the variance-covariance matrix.

1.3.3 Nonlinear Mixed-effects (NLME) Models

LME models may not truly describe the underlying relationship between the response and covariates since they are usually empirical models. On the other hand, NLME models may provide more reliable predictions in the sense that they are usually based on scientific justifications or underlying data-generating mechanisms. NLME models can be constructed by introducing appropriate random effects to the

corresponding nonlinear regression models.

Recall the longitudinal data shown in Figure 1.2. A LME model may not be suitable here since the trajectory pattern is quite complicated. Based on the biological and statistical arguments in Wu [2002], the following bi-exponential NLME model has been suggested for the HIV viral dynamics

$$\begin{aligned} y_{ij} =& \log_{10}(e^{P_{1i}-\lambda_{1ij}t_{ij}} + e^{P_{2i}-\lambda_{2i}t_{ij}}) + e_{ij}, \qquad i = 1, \cdots, n, \quad j = 1, \cdots, m_i, \\ P_{1i} =& P_1 + b_{1i}, \quad \lambda_{1ij} = \lambda_1 + \xi_1 CB4_{ij} + b_{2i}, \\ P_{2i} =& P_2 + b_{3i}, \quad \lambda_{2i} = \lambda_2 + b_{4i}, \end{aligned} \tag{1.13}$$

where y_{ij} is the log10-transformation of viral load measurements for patient i at time t_{ij}, $CD4_{ij}$ is the CD4 count for patient i at time t_{ij} which can partially explain the between-patient variation. $\boldsymbol{\beta} = (P_1, P_2, \lambda_1, \lambda_2, \xi_1)^T$ are the fixed-effects, $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^T$ are random-effects, and e_{ij} is a within-individual random error. We assume that $\boldsymbol{b}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, B)$, $e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, and $\boldsymbol{b}_i$ is independent of e_{ij}'s. In this study, the fixed-effect parameters e^{P_1} and e^{P_2} respectively denote the baseline viral load in the productively infected cells and the long-lived or latently infected cells. λ_1 and λ_2 denote initial viral decay rate in in the productively infected cells and the long-lived or latently infected cells respectively. The initial viral decay rate λ_{1ij} is an important parameter since it reflects the efficacy of the treatment [Wu and Ding, 1999], which may vary substantially across individuals. Moreover, it is known that

parameter λ_1 is positive during an anti-HIV treatment since the viral load typically declines at the beginning of the treatment and there is usually a positive association between CD4 count and initial decay rates; i.e., the higher the CD4 count is, the faster the decay is.

To evaluate the efficacy of the anti-HIV treatment, Wang and Wu [2011] considered multivariate one-sided tests based on multiple imputations for nonlinear mixed-effects models. Zhou et al. [2016] and Zhou and Wu [2021] also considered multivariate one-sided tests using likelihood-based methods for nonlinear mixed-effects models with arbitrary missing mechanism and censoring, respectively.

A general NLME form of the above model (1.13) can be written as

$$\begin{aligned} &y_{ij} = g(t_{ij}, \boldsymbol{\beta}_{ij}) + e_{ij}, \quad e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2), \\ &\boldsymbol{\beta}_{ij} = \mathbf{d}(\boldsymbol{z}_{ij}, \boldsymbol{\beta}, \mathbf{b}_i), \quad \mathbf{b}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, B), \quad i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, m_i, \end{aligned} \tag{1.14}$$

where $g(\cdot)$ is a known nonlinear function, $\mathbf{d}(\cdot)$ is a known multivariate linear function, $\boldsymbol{\beta}_{ij}$ is a vector of individual-specific and time-varying parameters, $\boldsymbol{\beta}$ is a vector of fixed-effects, $\boldsymbol{z}_{ij}$ is the covariates value, $\boldsymbol{e}_i = (e_{i1}, \cdots, e_{im_i})^T$ is a vector of within-individual random errors, $\boldsymbol{b}_i$ is a vector of random-effects, σ^2 is a variance parameter, and B is the variance-covariance matrix of $\boldsymbol{b}_i$. We assume $\boldsymbol{e}_i$ and $\boldsymbol{b}_i$ are independent. Note that NLME models are more general and include LME modes as a special case.

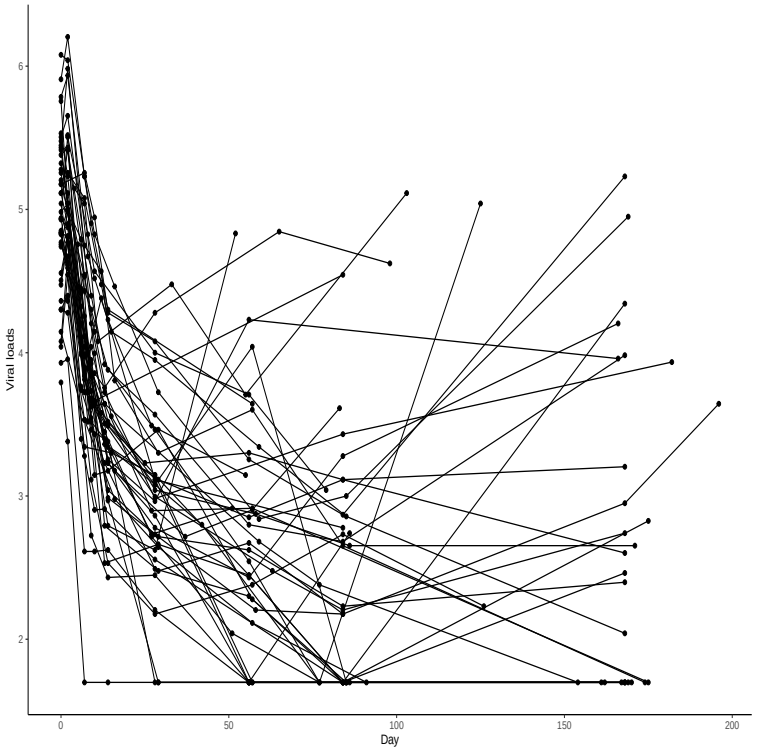


Figure 1.2: Trajectories of viral load in $\log_{10}$ scale in an anti-HIV treatment.

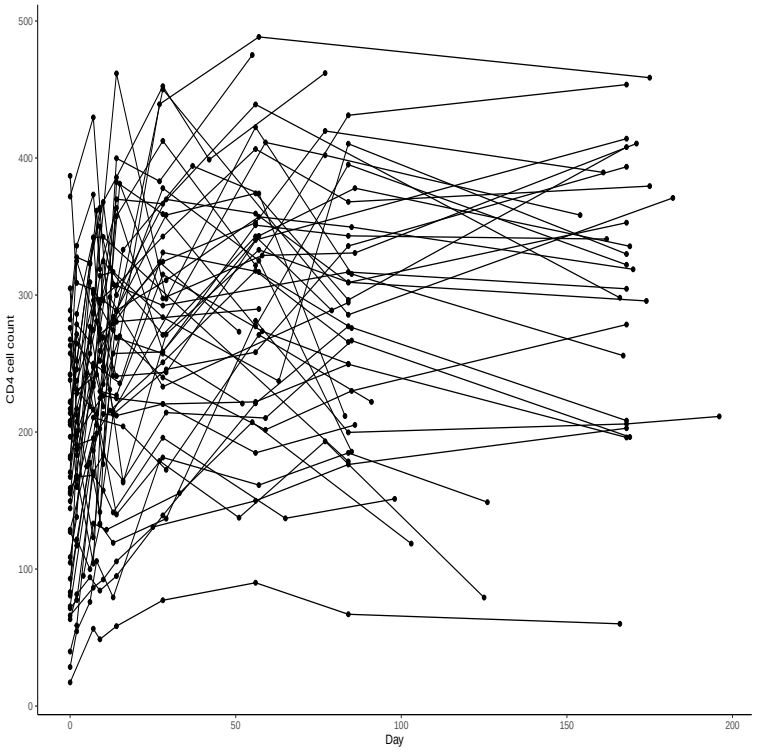


Figure 1.3: Trajectories of CD4 cell count in an anti-HIV treatment.

1.4 Complexity in Longitudinal Data and Multiple Imputation

In practice, longitudinal data often become complicated in the sense that i) longitudinal data may contain missing values; ii) some time-varying covariates may be measured with substantial errors; iii) some data may be censored due to a lower or upper detection limit.

1.4.1 Missingness

In the presence of missing data in longitudinal studies, it is important to understand the missing mechanism since the methods to handle missing data depend on the missing data mechanisms. In general, there are three possible missing data mechanisms [Rubin, 1976] as follows.

- missing complete at random (MCAR): missingness depends neither on the observed data nor on the unobserved data. For example, missing data are due to the broken of a device, which is irrelevant to the treatment.

- missing at random (MAR): missingness may depend on the observed data but not on the unobserved data. For example, a patient is too old to visit a clinic since the missingness only relates to the age which is observed.

- missing not at random (MNAR): missingness may depend on the observed data and on the unobserved data. For example, a patient misses a schedule due to drug intolerance or other drug side-effects during a treatment.

Generally, MCAR and MAR cases are known as the "ignorable" missing since the probability of observing a data item is independent of the value of that data item and the missing data mechanism can be safely ignored consequently, whereas missing data is non-ignorable in the MNAR cases since the missingness depends on the values of being missing and the missing data mechanism need to be included in the analysis.

1.4.2 Measurement Error in Time-varying Covariates

In many longitudinal studies, the within-individual variation may be large and this variation may be partially explained by time-varying covariates. Some covariates, however, may be measured with substantial errors. For example, it is known that the CD4 count is usually measured with errors in an HIV study. Ignoring measurement errors in covariates may lead to biased results [Higgins et al., 1997]. When a variable is measured with error, the true value of this variable is unknown or "missing". Thus, the true values of covariates with measurement errors can be treated as a special case of "missing data". Therefore, the measurement errors in covariates need

to be addressed to perform valid statistical inference.

1.4.3 Censoring Data

Censored or truncated data may arise when the devices used to measure a variable are subject to a lower and/or an upper detection limit. For example, in HIV/AIDS studies, HIV viral load typically has a lower detection limit such as 100 (copies/ml), i.e., viral load values less than 100 copies/ml cannot be measured. In this case, viral load observations may be left-censored. Although the true values are not observed, the censored values are known to be less or greater than a known number. Thus, the censoring is related to the unobserved values (the censored values). So censored data may be considered as another special case of missing data with missing not at random mechanism.

A naive way to deal with the left-censoring problem is to substitute them by the limit of detection(LOD) or some arbitrary values such as LOD/2 or LOD/$\sqrt{2}$. However, the naive method may lead to biased results when the proportion of censored data is not trivial [Hughes, 1999, Vaida and Liu, 2009, Wu, 2002, Qiu and Wu, 2010]. biased results [Hughes et al., 1999].

1.4.4 Multiple Imputation Approach for Incomplete Data

In the presence of missing data, naive methods ignoring incomplete observations may lead to biased results when the missing data are not MCAR [Little and Rubin, 2002]. After introduced by Rubin [1987], the multiple imputation(MI) technique has been extensively used to replace the missing value in different ways, including mismeasured data [Cole et al., 2006] and censoring data [Wei and Tanner, 1991, Grover and Gupta, 2015]. With the multiple imputations for incomplete data, the existing standard complete-data methods can be used in the subsequent statistical analysis.

In a MI method, the simulated values can be drawn from a predictive distribution of the missing data given the observed data. After m "complete datasets" are obtained, existing standard complete-data methods can be used to analyze each of the m "complete datasets". These analysis results are then combined to produce an overall inference, which takes the missing data uncertainty into account. Moreover, when the missing data rate is not high, the robustness against the assumed imputation models is also a main advantage of MI methods. In practice, $m = 5, 6$, would be a good choice for various practical concerns [Schafer, 1997, Wu and Wu, 2002]. The MI methods are also developed to address measurement errors[Cole et al., 2006] and censoring data [Wei and Tanner, 1991, Grover and Gupta, 2015]. With the multiple imputations for incomplete data, the existing standard complete-data methods can

be used in the subsequent statistical analysis.

In this thesis, we mainly focus on the multivariate one-sided hypothesis testing problems based on m multiple imputations for incomplete data including covariate measurement errors, left-censoring/non-ignorable missing response. As discussed in Li et al. [1991] and Meng and Rubin [1992], the statistical inference performs well with the "equal missing ratio" assumption which requires the fractional loss of information is the same for all components of the parameters of interest when m is small, while the multivariate two-sided test still performs well even when the assumption is violated. Therefore, we consider the MI method for order-constrained inferences with incomplete data since the closed-form parameter estimators are available for complete-data and the subsequent analysis can be conducted by the existing standard complete-data method. After generating m multiple imputations for incomplete data, we propose combined testing statistics for multivariate one-sided hypothesis in NLME models with various data complications.

1.5 Outline

Although some research work on the multivariate one-sided hypothesis testing problems in NLME models has been done, based on our best knowledge, there is very limited literature on these problems for complex longitudinal data including

measurement errors in time-varying covariates, censoring or non-ignorable missing response, and other complications.

In this thesis, we propose testing statistics for multivariate one-sided hypothesis in NLME models with: (i) mis-measured time-varying covariates, (ii) both mis-measured time-varying covariates and the left-censored response, and (iii) both mis-measured time-varying covariates and the non-ignorable missing response. The general approach to deal with the complex data is based on the multiple imputation (MI) method. We mainly focus on the Wald-type tests since these tests are commonly used in one-sided or order-restricted inference.

The rest of the thesis is organized as follows. In Chapter 2, we propose two new testing statistics for multivariate one-sided hypothesis in NLME models with measurement errors in time-varying covariates. In Chapter 3, we consider multivariate one-sided testing problems in NLME models with measurement errors in time-varying covariates and the left-censored response. In Chapter 4, we address multivariate one-sided testing problems in NLME models with the non-ignorable missing response and measurement errors in time-varying covariates. Chapter 5 presents conclusions with discussions and plans for the future work.

2 Multivariate One-sided Hypothesis Tests in NLME Models with Covariate Measurement Errors

2.1 Introduction

In the analysis of longitudinal data, nonlinear mixed-effects (NLME) models have received much attention in the literature, such as HIV viral dynamics and pharmacokinetics. NLME models are typically based on scientific justifications or underlying data-generating mechanisms, so they may provide more reliable predictions for missing or mis-measured data than empirical models such as linear mixed-effects (LME) models. Moreover, parameters in a NLME model may have meaningful scientific interpretations, and these parameters may also have some natural restrictions such as being strictly positive. Hypothesis testing for these parameters should incorporate the natural restrictions, which leads to more powerful tests than the correspond-

ing tests ignoring the restrictions [Silvapulle and Sen, 2005]. Since these restrictions may often be expressed as "one-sided" or "order-constrained", these tests are usually called one-sided tests or order-restricted tests. In practice, a common complication is that some data may be measured with errors, such as CD4 cell count and blood pressure. The measurement errors need to be addressed in the hypothesis tests in order to gain powers and avoid biased results [Fuller, 2009]. In this chapter, we consider the multivariate one-sided tests or order-restricted tests for NLME models with measurement errors in time-varying covariates.

Recall the motivating HIV viral dynamic model for the trajectories of viral load during an anti-HIV treatment ((1.13) in Chapter 1)

$$\begin{aligned} y_{ij} =& \log_{10}(e^{P_{1i}-\lambda_{1ij}t_{ij}} + e^{P_{2i}-\lambda_{2i}t_{ij}}) + e_{ij}, \qquad i = 1, \cdots, n, \quad j = 1, \cdots, m_i, \\ P_{1i} =& P_1 + b_{1i}, \quad \lambda_{1ij} = \lambda_1 + \xi_1 CD4_{ij} + b_{2i}, \quad P_{2i} = P_2 + b_{3i}, \quad \lambda_{2i} = \lambda_2 + b_{4i}, \end{aligned} \tag{2.1}$$

where y_{ij} is the log10-transformation of viral load measurements for patient i at time t_{ij}, $CD4_{ij}$ is the CD4 count for patient i at time t_{ij}, $\boldsymbol{\beta} = (P_1, P_2, \lambda_1, \lambda_2, \xi_1)^T$ are the fixed-effect parameters, $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^T$ are the random-effects, and e_{ij} is a within-individual random error. We assume that $\boldsymbol{b}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, B)$, $e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, and $\boldsymbol{b}_i$ is independent of e_{ij}'s.

The initial viral decay rate λ_{1ij} is an important parameter since it reflects the efficacy of the treatment [Wu and Ding, 1999], but it may vary substantially across

individuals. An important question is that if the large between-individual variation in the viral decay rate λ_{1ij} may be partially explained by the variation in CD4 count.

Based on many studies, it is known that parameter λ_1 is positive during an anti-HIV treatment since the viral load typically declines at the beginning of the treatment. Moreover, it is also known that there is usually a positive association between CD4 count and the initial decay rate; i.e., the higher the CD4 count is, the faster the decay is. In other words, these parameters should have the following restrictions: $\lambda_1 \geq 0,\ \xi_1 \geq 0$. When evaluating and comparing different treatments, we wish to test the following multivariate one-sided hypothesis

$$H_0 : \lambda_1 = 0,\ \xi_1 = 0 \qquad \text{versus} \qquad H_1 : \lambda_1 \geq 0,\ \xi_1 \geq 0, \tag{2.2}$$

where at least one inequality strictly holds in H_1.

In addition, it is known that CD4 count is typically measured with substantial errors. Ignoring these measurement errors in statistical inference may under-estimate the association between CD4 count and the initial decay rate. Therefore, the above testing problem is further complicated by the measurement errors in CD4 count.

To address measurement errors in time-varying covariates, we may treat the repeated measurements of the covariates as "replicates" and model the covariate process to estimate the magnitude of the measurement errors. Statistical inference can then be based on the joint covariate model and response model using (say) a Monte-

Carlo expectation maximization (EM) algorithm for simultaneous inference [Liu and Wu, 2007]. However, such an approach can be computationally very intensive in the current context. In this chapter, we propose to use a multiple imputation method to address covariate measurement errors in NLME models, treating measurement errors as "missing data". An advantage of using a multiple imputation method to address covariate measurement errors is that many existing methods for "complete-data" one-sided tests may be readily to use. This advantage is particularly important here since the multivariate one-sided tests for NLME models are already challenging, even without measurement errors.

Treating measurement errors as a "missing data" problem has appeared in the literature [Cole et al., 2006, Blackwell et al., 2017]. The intuition is that, when a variable is measured with errors, the true values of this variable are unknown or "missing". However, the error-prone observed values of this variable may be used to help "predicting" the (unobserved) true values, based on a model fitted to the observed data. The missing data here may be viewed as missing at random since the "missingness" may depend on the observed data but not the unobserved true values [Rubin, 1976].

To perform one-sided tests for NLME models with covariate measurement errors based on a multiple imputation method, a key issue is how to combine the results

from multiply imputed datasets. We propose two approaches: (i) combine individual sufficient statistics from imputed data; and (ii) combine individual test statistics from imputed data. In either case, it is not straightforward to combine these statistics to make an overall (combined) and valid inference, and it can be challenging to derive the null distributions of the overall test statistics. Following Li et al. [1991], Meng and Rubin [1992], and Wang and Wu [2011], we propose two methods to appropriately combine individual sufficient statistics and individual test statistics. These methods will be evaluated based on a simulation study.

The chapter is organized as follows. In Section 2.2, we propose a multiple imputation method for NLME models with covariate measurement errors. In Section 2.3, we propose the two approaches to combine individual sufficient statistics or test statistics based on imputed datasets. We illustrate the methods in a real dataset in Section 2.4. In Section 2.5, we conduct a simulation study to evaluate the proposed methods. We conclude the chapter in Section 2.6 with some discussions.

2.2 A Multiple Imputation Method for NLME Models with Covariate Measurement Errors

Let y_{ij} be the response value for individual i at time t_{ij}, and let z_{qij} be the q-th possibly error-prone and time-varying covariate for individual i at time t_{ij},

$q = 1, \cdots, s,\ i = 1, \cdots, n,\ j = 1, \cdots, m_i$. Some of the covariates may be error-free and thus suppressed. Let $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{im_i})^T$ and $\boldsymbol{z}_i = (\boldsymbol{z}_{i1}^T, \cdots, \boldsymbol{z}_{im_i}^T)^T$ with $\boldsymbol{z}_{ij} = (z_{1ij}, \cdots, z_{sij})^T$, where T denotes a transpose. We consider the following general NLME model

$$\begin{aligned} &y_{ij} = g(t_{ij}, \boldsymbol{\beta}_{ij}) + e_{ij}, \quad e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2), \quad i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, m_i, \\ &\boldsymbol{\beta}_{ij} = \mathbf{d}(\boldsymbol{z}_{ij}^*, \boldsymbol{\beta}, \boldsymbol{b}_i), \quad \boldsymbol{b}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, B), \end{aligned} \tag{2.3}$$

where $g(\cdot)$ is a known nonlinear function, $\mathbf{d}(\cdot)$ is a known multivariate linear function, $\boldsymbol{\beta}_{ij}$ is a vector of individual-specific and time-varying parameters and $\boldsymbol{\beta}$ is a vector of fixed-effect, $\boldsymbol{z}_{ij}^*$ is the (unobserved) true covariate value corresponding to the observed and possibly mis-measured $\boldsymbol{z}_{ij}$, $\boldsymbol{e}_i = (e_{i1}, \cdots, e_{im_i})^T$ is a vector of within-individual random errors, $\boldsymbol{b}_i$ is a vector of random-effect, σ^2 is a variance parameter, and B is the variance-covariance matrix of $\boldsymbol{b}_i$. We assume $\boldsymbol{e}_i$ and $\boldsymbol{b}_i$ are independent.

In NLME model (2.3), we assume that the individual-specific parameters $\boldsymbol{\beta}_{ij}$ depend on the true (unobservable) covariate $\boldsymbol{z}_{ij}^*$ rather than the observed but possibly contaminated covariate $\boldsymbol{z}_{ij}$. To address the measurement errors in time-varying covariates, we consider the following classical measurement error model [Carroll et al., 2006]

$$\boldsymbol{z}_{ij} = U_{ij}\boldsymbol{\alpha} + V_{ij}\boldsymbol{a}_i + \boldsymbol{\epsilon}_{ij} \ \left(\equiv \boldsymbol{z}_{ij}^* + \boldsymbol{\epsilon}_{ij}\right), \quad i = 1, \cdots, n, j = 1, \cdots, m_i, \tag{2.4}$$

where U_{ij} and V_{ij} are known design matrices including time, $\boldsymbol{\alpha}$ and $\boldsymbol{a}_i$ are the

fixed-effect vector and random-effect vector, respectively, and $\boldsymbol{\epsilon}_{ij}$ are the random measurement errors for individual i at time t_{ij}. Here $\boldsymbol{a}_i$ is assumed to i.i.d. (independently and identically distributed) follow $N(\mathbf{0}, A)$, and $\boldsymbol{\epsilon}_{ij} = (\epsilon_{1ij}, \cdots, \epsilon_{sij})^T$ is assumed to i.i.d. follow $N(\mathbf{0}, D)$. For example, if $\boldsymbol{z}_{ij}$ is the error-prone CD4 count for individual i at time t_{ij}, we may consider a measurement error model $z_{ij} = (\alpha_0 + a_{0i}) + (\alpha_1 + a_{1i})t_{ij} + (\alpha_2 + a_{2i})t_{ij}^2 + \epsilon_{ij} (\equiv z_{ij}^* + \epsilon_{ij})$, where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)^T$ are the fixed-effect parameters and $\boldsymbol{a}_i = (a_{0i}, a_{1i}, a_{2i})^T$ are the random-effects. That is, we assume that CD4 count follows a smooth quadratic trajectory and use the repeated measurements as "replicates" to address measurement errors. More generally, we may also assume a nonparametric mixed-effects model for $\boldsymbol{z}_{ij}$ and then approximate it by an LME model using a basis-based approach [Liu and Wu, 2007].

Since the true covariate $\boldsymbol{z}_i^*$ is not observed (or "missing"), we propose to use the observed data $\mathcal{D} = \{(\boldsymbol{z}_i, \boldsymbol{y}_i), i = 1, \cdots, n\}$ from the covariates and the response to create imputations for the unobserved $\boldsymbol{z}_i^*$, incorporating imputation uncertainty. Specifically, by conditioning on $\mathcal{D}$, we may create multiple imputations for $\boldsymbol{z}_i^*$ by generating random samples from an assumed predictive distribution $f(\boldsymbol{z}_i^*|\mathcal{D})$, where $f(\cdot|\cdot)$ denotes a generic conditional density function. To ensure that the multiple imputations are "proper" in the sense of Rubin [1976], we generate imputations under Bayesian framework.

Let $\Psi = (\boldsymbol{\alpha}, \Phi) \equiv (\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma, D, A, B)$ be the collection of all unknown parameters in the model (2.3) and (2.4), where $\Phi = (\boldsymbol{\beta}, \sigma, D, A, B)$. Denote the collection of the random-effects $\boldsymbol{a}_i$'s in the covariate model (2.4) and $\boldsymbol{b}_i$'s in the response model (2.3) by $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. Let $\pi(\cdot)$ denote a prior density function for Ψ. Note that, in the covariate model (2.4), the unobserved "true" covariate values $\boldsymbol{z}^*_{ij} = U_{ij}\boldsymbol{\alpha} + V_{ij}\boldsymbol{a}_i$ are completely determined by the mean parameters $\boldsymbol{\alpha}$ and the random-effects $\boldsymbol{a}_i$ since the design matrices U_{ij} and V_{ij} are known. Therefore, to generate multiple imputations for $\boldsymbol{z}^*_{ij}$'s, it is sufficient to generate multiple imputations for $(\boldsymbol{\alpha}, \boldsymbol{a})$ from the distribution $f(\boldsymbol{\alpha}, \boldsymbol{a}|\mathcal{D})$.

Under the Bayesian framework, we need to specify a prior distribution for the parameter Ψ. In practice, the prior information may be obtained from previous studies or reference literature. For some of the parameters with reliable prior information, we may use strong priors with small variances, while for the other parameters without enough prior information, non-informative prior distributions (with large variances) may be employed [Liu and Li, 2015, Huang et al., 2006]. We assume that the parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma, D, A$ and B are independent, i.e. $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma, D, A, B) = \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\sigma)\pi(D)\pi(A)\pi(B)$. Normal prior distributions are assigned to the mean parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, while the Inverse Gamma or Inverse Wishart prior distributions

are assigned to the variance-covariance parameters σ^2, D, A and B as follows.

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \Sigma_{\boldsymbol{\alpha}}), \qquad \boldsymbol{\beta} \sim N(\mathbf{0}, \Sigma_{\boldsymbol{\beta}}),$$

$$\sigma^2 \sim G^{-1}(\gamma, \delta), \qquad D \sim W^{-1}(\Sigma_D, \kappa),$$

$$A \sim W^{-1}(\Sigma_A, \rho), \qquad B \sim W^{-1}(\Sigma_B, \tau),$$

where G^{-1} and W^{-1} respectively denote the Inverse Gamma distribution and the Inverse Wishart distribution, and the hyper-parameters $\gamma, \delta, \kappa, \rho, \tau, \Sigma_{\boldsymbol{\alpha}}, \Sigma_{\boldsymbol{\beta}}, \Sigma_D, \Sigma_A$ and Σ_B are known.

With the prior distributions for the unknown model parameters, the multiple imputations for $(\boldsymbol{\alpha}, \boldsymbol{a})$ can be generated by using the data augmentation method and the following Gibbs sampling method. Specifically, we sample from the full conditional distributions in turn, at iteration t $(t = 1, 2, \cdots)$,

Step I: Draw a value $(\boldsymbol{\alpha}^{(t)}, \boldsymbol{a}^{(t)})$ of $(\boldsymbol{\alpha}, \boldsymbol{a})$ in the following order

i. Draw a value of $\boldsymbol{\alpha}^{(t)}$ of the mean parameter $\boldsymbol{\alpha}$ in the covariate model from

$$f(\boldsymbol{\alpha}|\mathcal{D}, \boldsymbol{a}^{(t-1)}, \boldsymbol{b}^{(t-1)}; \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}, R^{(t-1)})$$
$$\propto \prod_{i=1}^{n} \left[f(\boldsymbol{y}_i|\boldsymbol{a}_i^{(t-1)}, \boldsymbol{b}_i^{(t-1)}; \boldsymbol{\alpha}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \times f(\boldsymbol{z}_i|\boldsymbol{a}_i^{(t-1)}; \boldsymbol{\alpha}, R^{(t-1)}) \right] \cdot \pi(\boldsymbol{\alpha}).$$

ii. Draw a value $\boldsymbol{a}_i^{(t)}$ of the random-effect $\boldsymbol{a}_i$ in the covariate model from

$$f(\boldsymbol{a}_i|\boldsymbol{z}_i, \boldsymbol{y}_i, \boldsymbol{\alpha}^{(t)}, \boldsymbol{b}_i^{(t-1)}, \Phi^{(t-1)})$$
$$\propto f(\boldsymbol{y}_i|\boldsymbol{a}_i, \boldsymbol{b}_i^{(t-1)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \times f(\boldsymbol{z}_i|\boldsymbol{a}_i; \boldsymbol{\alpha}^{(t)}, R^{(t-1)}) f(\boldsymbol{a}_i; A^{(t-1)}), \ i = 1, \cdots, n.$$

Step II: Draw a value $(\boldsymbol{b}^{(t)}, \Phi^{(t)})$ of $(\boldsymbol{b}, \Phi)$ in the following order

i Draw a value $\boldsymbol{b}_i^{(t)}$ of the random-effect $\boldsymbol{b}_i$ in the response model from

$$f(\boldsymbol{b}_i|\boldsymbol{y}_i, \boldsymbol{a}_i^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}, D^{(t-1)})$$

$$\propto f(\boldsymbol{y}_i|\boldsymbol{a}_i^{(t)}, \boldsymbol{b}_i; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) f(\boldsymbol{b}_i; D^{(t-1)}), \quad i = 1, \cdots, n.$$

ii Draw a value $\boldsymbol{\beta}^{(t)}$ of the mean parameter $\boldsymbol{\beta}$ in the response model from

$$f(\boldsymbol{\beta}|\mathcal{D}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{a}^{(t)}, \boldsymbol{b}^{(t)}, \sigma^{2(t-1)}) \propto \prod_{i=1}^{n} f(\boldsymbol{y}_i|\boldsymbol{a}_i^{(t)}, \boldsymbol{b}_i^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}, \sigma^{2(t-1)}) \cdot \pi(\boldsymbol{\beta}).$$

iii Draw a value $(\sigma^{2(t)}, D^{(t)}, A^{(t)}, B^{(t)})$ of the variance-covariance parameters (σ^2, D, A, B) in the models from the corresponding full conditional distributions.

$$\sigma^2|\mathcal{D}, \boldsymbol{a}^{(t)}, \boldsymbol{b}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t)}, B^{(t-1)}, A^{(t-1)}, D^{(t-1)}$$

$$\sim G^{-1}\left(\gamma + \frac{1}{2}\sum_{i=1}^{n} m_i, \left[\frac{1}{\delta} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(y_{ij} - g(t_{ij}, \boldsymbol{\beta}_{ij}^{(t)})\right)^2\right]^{-1}\right),$$

$$D|\mathcal{D}, \boldsymbol{a}^{(t)}, \boldsymbol{b}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t)}, \sigma^{(t)}, A^{(t-1)}, B^{(t-1)}$$

$$\sim W^{-1}\left(\left[\Sigma_D^{-1} + \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\boldsymbol{z}_{ij} - U_{ij}\boldsymbol{\alpha}^{(t)} - V_{ij}\boldsymbol{a}_i^{(t)}\right)\left(\boldsymbol{z}_{ij} - U_{ij}\boldsymbol{\alpha}^{(t)} - V_{ij}\boldsymbol{a}_i^{(t)}\right)^T\right]^{-1}, \kappa + \sum_{i=1}^{n} m_i\right)$$

$$A|\mathcal{D}, \boldsymbol{a}^{(t)}, \boldsymbol{b}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t)}, \sigma^{(t)}, D^{(t)}, B^{(t-1)}$$

$$\sim W^{-1}\left(\left[\Sigma_A^{-1} + \sum_{i=1}^{n}\boldsymbol{a}_i^{(t)}\boldsymbol{a}_i^{(t)T}\right]^{-1}, \rho + n\right),$$

$$B|\mathcal{D}, \boldsymbol{a}^{(t)}, \boldsymbol{b}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t)}, \sigma^{(t)}, D^{(t)}, A^{(t)}$$

$$\sim W^{-1}\left(\left[\Sigma_B^{-1} + \sum_{i=1}^{n}\boldsymbol{b}_i^{(t)}\boldsymbol{b}_i^{(t)T}\right]^{-1}, \tau + n\right).$$

The starting values $\Psi^{(0)}$ can be obtained from the two-step method. Beginning with $\Psi^{(0)}$, we iterate the foregoing procedures for a burn-in period until the resulting Markov Chain converges to its stationary distribution. In order to determine the number of "burn-in" iterations and to check the convergence of the Markov Chain, we start multiple Markov Chains from dispersed initial values and graphically compare the within and between variation of the simulated samples. Note that a *proper* multiple imputation method needs independent imputations [Rubin, 1987]. We may save every Mth (M is a large integer) simulation samples after the burn-in period such that the retained samples are approximately independent. At convergence, we obtain m value of $(\boldsymbol{\alpha}, \boldsymbol{a})$ from $f(\boldsymbol{\alpha}, \boldsymbol{a}|\mathcal{D})$, and thus a value of $\boldsymbol{z}_i^*$ from $f(\boldsymbol{z}_i^*|\mathcal{D})$ since $\boldsymbol{z}_{ij}^* = U_{ij}\boldsymbol{\alpha} + V_{ij}\boldsymbol{a}_i$. Repeating the above procedure m times, we obtain m imputations of $\boldsymbol{z}_i^*$ from $f(\boldsymbol{z}_i^*|\mathcal{D})$. Thus, we obtain m "complete datasets" $\left\{\left(\boldsymbol{y}_i, \boldsymbol{z}_i^{*(l)}\right),\ i = 1, 2, \cdots, n\right\},\ l = 1, 2, \cdots, m$. Based on each of the m "complete datasets", we can conduct inference for the NLME model (2.3) with covariate measurement errors being addressed, and then we combine the m results to obtain an overall conclusion. A common choice for m is $m = 5$ or 6.

In the next section, we consider the one-sided hypothesis testing for parameters in the NLME models based on multiply imputed datasets and show how to combine the results.

2.3 Multivariate One-sided Tests for NLME Models with Multiply Imputed Covariates

We first discuss the multivariate one-sided hypothesis tests concerning the parameters in general NLME models.

2.3.1 The Multivariate One-sided Tests in NLME Models

Let $\boldsymbol{\theta}$ denote the $k \times 1$ parameter vector of a NLME model and $\Omega = \{ \boldsymbol{\theta} : \boldsymbol{\theta} \in \mathbb{R}^k, R\boldsymbol{\theta} \geq \mathbf{0} \}$ denote the parameter space. The multivariate one-sided hypothesis of interest is given as

$$H_0 : R\boldsymbol{\theta} = \mathbf{0} \qquad \text{versus} \qquad H_1 : R\boldsymbol{\theta} \geq \mathbf{0},\ R\boldsymbol{\theta} \neq \mathbf{0}, \tag{2.5}$$

where R is a known $r \times k$ ($r \leq k$) full rank matrix with elements of 1's and 0's.

Suppose there are n observations. Let $l_i(\boldsymbol{\theta})$ denote the observed-data log-likelihood based on the ith observation ($i = 1, \cdots, n$) and $l_{obs}(\boldsymbol{\theta})$ denote the observed-data log-likelihood based on n observations. Then $l_{obs}(\boldsymbol{\theta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Omega} l_{obs}(\boldsymbol{\theta})$ denote the unconstrained MLE, $\tilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \{\boldsymbol{\theta} : R\boldsymbol{\theta} \geq \mathbf{0}\}} l_{obs}(\boldsymbol{\theta})$ denote the MLE under H_1 and $\bar{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \{\boldsymbol{\theta} : R\boldsymbol{\theta} = \mathbf{0}\}} l_{obs}(\boldsymbol{\theta})$ denote the MLE under H_0.

To develop the Wald test for hypothesis (2.5), we need the following two assump-

tions,

$$\mathbf{A1}: \qquad -\frac{1}{n}\nabla^2 l_{obs}(\boldsymbol{\theta}) \xrightarrow{a.s.} H_{\boldsymbol{\theta}},$$

$$\mathbf{A2}: \qquad n^{-1/2}\nabla l_{obs}(\boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, H_{\boldsymbol{\theta}}),$$

where $\nabla = \partial/\partial\boldsymbol{\theta}, \nabla^2 = \partial^2/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$, and $H_{\boldsymbol{\theta}}$ is the finite and positive definite Fisher information matrix. We assume the following regularity conditions hold.

C1. For each $\boldsymbol{\theta} \in \Omega, l_i(\boldsymbol{\theta}) = \log(f(\boldsymbol{y}_i|\boldsymbol{\theta}))$ is third-order differentiable with respect to $\boldsymbol{\theta}$.

C2. There exist real-valued functions $g_i(\boldsymbol{y}_i)$ such that $\int g_i(\boldsymbol{y}_i) < \infty$ and the absolute value of the first two derivatives of $l_i(\boldsymbol{\theta})$ are bounded by $g_i(\boldsymbol{y}_i)$, $i = 1, \cdots, n$.

C3. For each $\boldsymbol{\theta} \in \Omega, H_{\boldsymbol{\theta}}^{(i)} = (H_{\boldsymbol{\theta}}^{(i)}(j_1, j_2))_{k\times k} = E\{\nabla l_i(\boldsymbol{\theta}|\boldsymbol{y}_i)[\nabla l_i(\boldsymbol{\theta}|\boldsymbol{y}_i)]^T\}$ is positive definite and has finite entries.

C4. For each $\boldsymbol{\theta} \in \Omega$, $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n H_{\boldsymbol{\theta}}^{(i)} = H_{\boldsymbol{\theta}}$ exists and is positive definite.

C5. For each $\boldsymbol{\theta} \in \Omega$, there exists a $\delta > 0$ and a finite M such that for $g_i(\boldsymbol{y}_i)$ in **C2**, $E_{\boldsymbol{\theta}}\left[\left(g_i(\boldsymbol{y}_i) + H_{\boldsymbol{\theta}}^{(i)}(j_1, j_2)\right)^{1+\delta}\right] \leq M$ with $E_{\boldsymbol{\theta}}(\cdot)$ denoting an expectation evaluated under distributions corresponding to the parameter values $\boldsymbol{\theta}$.

C6. The Linderberg condition is satisfied if $\frac{1}{n}\sum_{i=1}^n E\{\|\nabla l_i(\boldsymbol{\theta}|\boldsymbol{y}_i)\|^2\}\mathbf{1}\{\|\nabla l_i(\boldsymbol{\theta}|\boldsymbol{y}_i)\| > \varepsilon\sqrt{n}\} \xrightarrow{p} 0$ for all $\epsilon > 0$, where $\mathbf{1}\{\cdot\}$ is the indicator function and $\|\cdot\|$ denotes the Euclidean norm.

C7. There exists a function $l(\boldsymbol{\theta})$ such that $\sup_{\boldsymbol{\theta}\in\Omega}\left|\frac{1}{n}l_{obs}(\boldsymbol{\theta}) - l(\boldsymbol{\theta})\right| \xrightarrow{p} 0$ as $n \to \infty$, and $\sup_{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\epsilon} l(\boldsymbol{\theta}) < l(\boldsymbol{\theta}_0)$ for any $\epsilon > 0$. Moreover, $l_{obs}(\boldsymbol{\theta})$ has a unique local maximizer. For $\boldsymbol{\theta}$ on the boundary of Ω, "local" is in a one-sided sense.

In the above regularity conditions, **C1** ensures that the observed-data log-likelihood function has a Taylor expansion as a function of $\boldsymbol{\theta}$. **C2** ensures that the observed-data log-likelihood function is differentiable with respect to $\boldsymbol{\theta}$ under the integral sign, which leads to $E\{\nabla l_i(\boldsymbol{\theta}|\boldsymbol{y}_i)\} = \mathbf{0}$ and $E\{\nabla^2 l_i(\boldsymbol{\theta}|\boldsymbol{y}_i)\} = -E\{\nabla l_i(\boldsymbol{\theta}|\boldsymbol{y}_i)[\nabla l_i(\boldsymbol{\theta}|\boldsymbol{y}_i)]^T\}$ for each $\boldsymbol{\theta} \in \Omega$.

By Markov's strong law of large numbers, if $\{x_1, x_2, \cdots\}$ is a sequence of independent random variables with $E(x_i) = \mu_i < \infty$ and if for some $\delta > 0, \sum_{i=1}^{\infty} E[|x_i - \mu_i|^{1+\delta}]/i^{1+\delta} < \infty$, then $\bar{x}_n - \bar{\mu}_n \xrightarrow{a.s.} 0$ where $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}\mu_i$. Therefore, when x_i is the (j_1, j_2) entry of $-\frac{1}{n}\nabla^2 l_i(\boldsymbol{\theta})$, for δ in **C5**, we have by **C2**

$$\left|x_i - H_{\boldsymbol{\theta}}^{(i)}(j_1, j_2)\right|^{1+\delta} \leq \left|g_i(\boldsymbol{y}_i) + H_{\boldsymbol{\theta}}^{(i)}(j_1, j_2)\right|^{1+\delta}.$$

As a result, $E_{\boldsymbol{\theta}}\left[\left|x_i - H_{\boldsymbol{\theta}}^{(i)}(j_1, j_2)\right|^{1+\delta}\right]$ is uniformly bounded by a finite M by **C5**, which leads to

$$\sum_{i=1}^{\infty} E_{\boldsymbol{\theta}}\left[\left|x_i - H_{\boldsymbol{\theta}}^{(i)}(j_1, j_2)\right|^{1+\delta} \Big/ i^{1+\delta}\right] \leq M\sum_{i=1}^{\infty} 1/i^{1+\delta} < \infty.$$

The (j_1, j_2) entry of $-\frac{1}{n}\nabla^2 l_{obs}(\boldsymbol{\theta})$ converges almost surely to the (j_1, j_2) entry of $H_{\boldsymbol{\theta}}$

consequently. Then, we have by **C4**

$$-\frac{1}{n}\nabla^2 l_{obs}(\boldsymbol{\theta}) \xrightarrow{a.s.} H_{\boldsymbol{\theta}}. \tag{2.6}$$

The Lindbergh-Feller central limit theorem states that for each n, let $\mathbf{Y}_{n,1}, \cdots, \mathbf{Y}_{n,k_n}$ be independent random vectors with finite variances such that $\sum_{i=i}^{k_n} E\|\mathbf{Y}_{n,i}\|^2 \mathbf{1}\{\|\mathbf{Y}_{n,i}\| > \varepsilon\} \to 0, i = 1, \cdots, k_n, \forall \varepsilon > 0$, and $\sum_{i=1}^{k_n} \text{Cov}\mathbf{Y}_{n,i} \to \Sigma$, then the sequence $\sum_{i=1}^{k_n}(\mathbf{Y}_{n,i} - E(\mathbf{Y}_{n,i})) \xrightarrow{d} N(0, \Sigma)$. With **C4**, **C6**, and equation (2.6), we have the following result

$$n^{-1/2}\nabla l_{obs}(\boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, H_{\boldsymbol{\theta}}). \tag{2.7}$$

Based on the above arguments, the Wald statistic for testing (2.5) is given by

$$T_W = n\left(R\tilde{\boldsymbol{\theta}}\right)^T \left(R\hat{H}_{\boldsymbol{\theta}}^{-1}R^T\right)^{-1} \left(R\tilde{\boldsymbol{\theta}}\right),$$

where $\hat{H}_{\boldsymbol{\theta}}$ is an estimator of $H_{\boldsymbol{\theta}}$ under $H_1 : R\boldsymbol{\theta} \geq \mathbf{0}$; thus, $\hat{H}_{\boldsymbol{\theta}}$ can be $\hat{H}_{\tilde{\boldsymbol{\theta}}}$ or $\hat{H}_{\hat{\boldsymbol{\theta}}}$. However, the explicit form of the Wald test usually does not exist due to the complexity of the observed-data likelihood function. We will derive the asymptotic distribution for the above Wald test.

Theorem 2.3.1 *With the regularity conditions* **C1-C7** *and the assumptions* **A1** *and* **A2**, *the asymptotic null distribution of the Wald test for one-sided hypothesis (2.5) is given by the following* $\bar{\chi}^2$*- distribution.*

$$\lim_{n\to\infty} \Pr(T_W \leq c | R\boldsymbol{\theta} = \mathbf{0}) = \sum_{i=0}^{r} \omega_i(r, RH_{\boldsymbol{\theta}_0}^{-1}R^T, \mathbb{O}^+) \Pr(\chi_i^2 \leq c), \quad \textit{for any} \;\; c \geq 0, \tag{2.8}$$

where $\boldsymbol{\theta}_0$ *denotes the true value of* $\boldsymbol{\theta}$, $\mathbb{O}^+ = \{R\boldsymbol{\theta} | R_i\boldsymbol{\theta} \geq 0, i = 1, \cdots, r\}$ *with* R_i *being the row elements of* R, $\omega_i(r, RH_{\boldsymbol{\theta}_0}^{-1}R^T, \mathbb{O}^+)$ *is the probability weight as given in section 1.2, i.e. the probability that* $\pi_{RH_{\boldsymbol{\theta}_0}^{-1}R^T}(\boldsymbol{x}; \mathbb{O}^+) \in$ *the* i *dimensional face of* $\mathbb{O}^+$ *with* $\sum_{i=0}^{r} \omega_i(r, \Sigma_n, \mathbb{O}^+) = 1$, χ_i^2 *is the chi-square distribution with* i *degree of freedom and* χ_0^2 *denotes the distribution that takes the value zero with probability one.*

The proof of Theorem 2.3.1 is given in Appendix (A.2.1).

Although we have derived the null distribution for the Wald test statistic, the limiting variance covariance matrix $H_{\boldsymbol{\theta}_0}$ in (2.8) is unknown. In order to actually implement the $\bar{\chi}^2$ test in practice, we can either substitute $H_{\boldsymbol{\theta}_0}$ by the observed information matrix $-\frac{1}{n}\nabla^2 l_{obs}(\boldsymbol{\theta})$ under H_0 and use a simulation method to obtain the cut-off values, or use the upper bound and the lower bound of the null probability to do a conservative test. These ideas are stated in the following theorem and the proof is given in Appendix (A.2.2).

Theorem 2.3.2 *With the regularity conditions* **C1-C7** *and the assumptions* **A1** *and* **A2**, *the asymptotic null distribution of* T_W *can be estimated by*

$$\widehat{\Pr}(T_W \leq c | R\boldsymbol{\theta} = \mathbf{0}) = \sum_{i=0}^{r} \omega_i(r, R\hat{H}_{\boldsymbol{\theta}}^{-1}R^T, \mathbb{O}^+) \Pr(\chi_i^2 \leq c), \quad \text{for any } \ c \geq 0,$$

where $\hat{H}_{\boldsymbol{\theta}} = -\frac{1}{n}\nabla^2 l_{obs}(\tilde{\boldsymbol{\theta}})$. *Moreover, we have the lower and upper bound*

$$\frac{\Pr(\chi_0^2 \geq c) + \Pr(\chi_1^2 \geq c)}{2} \leq \Pr(T_W \geq c | R\boldsymbol{\theta} = \mathbf{0}) \ \leq \frac{\Pr(\chi_{r-1}^2 \geq c) + \Pr(\chi_r^2 \geq c)}{2}.$$

In the following, we develop two Wald type tests for the hypothesis testing problem (2.9) in NLME models with measurement errors based on the above results. In the previous section, we create m imputations for the unobserved true covariate value $\boldsymbol{z}_i^*$ and obtain m "complete datasets" $\left\{\left(\boldsymbol{y}_i, \boldsymbol{z}_i^{*(l)}\right),\ i = 1, \cdots, n\right\}$, $l = 1, \cdots, m$. For each of the m datasets, we can conduct the one-sided hypothesis testing using existing complete-data methods and then we combine the results to obtain an overall conclusion. Here, a key question is how to combine the results from the m hypothesis tests. Following Meng and Rubin [1992] and Wang and Wu [2011], we propose two approaches: (i) combine the m *sufficient statistics* for the parameters of interest from the m "complete datasets", and then construct a single test statistic such as a Wald-type test statistic; (ii) combine the m *test statistics* based on the m "complete datasets" to obtain a single test statistic. For either case, another important question is how to derive the null distribution or p-value for the single test statistic. The details are discussed below.

Under the NLME model (2.3), the hypothesis (2.2) can be written in a more general form as the hypothesis (2.5), i.e.

$$H_0 : R\boldsymbol{\theta} = \mathbf{0} \qquad \text{versus} \qquad H_1 : R\boldsymbol{\theta} \geq \mathbf{0},\ R\boldsymbol{\theta} \neq \mathbf{0}, \tag{2.9}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma, \text{vec}(B))^T \subset \mathbb{R}^k$ denote the collection of all parameters in a NLME model with vec$(\cdot)$ being the vectorization of a matrix, k denotes the number of

distinct parameters and R is a known $r \times k$ ($r \leq k$) full rank matrix with elements of 1's and 0's. We consider the anti-HIV treatment (2.1) as a simple example to illustrate the hypothesis (2.9). Here B is assumed to be diagonal for simplicity, i.e., $\text{vec}(B) = diag(B_{11}, B_{22}, B_{33}, B_{44})$, then $\boldsymbol{\theta} = (P_1, P_2, \lambda_1, \lambda_2, \xi_1, \sigma, B_{11}, B_{22}, B_{33}, B_{44})^T$. So the matrix R is given by

$$R = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{2\times 10},$$

with $r = 2$.

Let $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n)^T$ be the response data, and let $\boldsymbol{z}^{*(l)} = (\boldsymbol{z}_1^{*(l)}, \boldsymbol{z}_2^{*(l)}, \cdots, \boldsymbol{z}_n^{*(l)})$ be the l-th imputed covariate data. Then we have m "complete datasets" $\{\left(\boldsymbol{y}, \boldsymbol{z}^{*(l)}\right),\ l = 1, \cdots, m\}$. For the l-th "complete dataset" $\{\boldsymbol{y}, \boldsymbol{z}^{*(l)}\}$, let $\tilde{\boldsymbol{\theta}}_{*l}$ be the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ under the parameter space $\{\boldsymbol{\theta} : R\boldsymbol{\theta} \geq \mathbf{0}\}$, $\hat{H}_{n*l}$ be the corresponding observed information matrix $-\frac{1}{n}\nabla^2 l_{obs}(\boldsymbol{\theta})$ evaluated at $\tilde{\boldsymbol{\theta}}$, hence $\hat{H}_{n*l}^{-1}$ is an estimate of variance covariance matrix H_n^{-1}. We consider the following combined statistics

$$\begin{aligned} \bar{\boldsymbol{\theta}}_m &= \frac{1}{m}\sum_{l=1}^{m} \tilde{\boldsymbol{\theta}}_{*l}, \\ \bar{H}_m &= \frac{1}{m}\sum_{l=1}^{m} \hat{H}_{n*l}, \\ \bar{B}_m &= \frac{1}{m-1}\sum_{l=1}^{m} (\tilde{\boldsymbol{\theta}}_{*l} - \bar{\boldsymbol{\theta}}_m)(\tilde{\boldsymbol{\theta}}_{*l} - \bar{\boldsymbol{\theta}}_m)^T, \end{aligned}$$

where $\bar{\boldsymbol{\theta}}_m$ is an estimator of $\boldsymbol{\theta}$, $\bar{H}_m^{-1}$ estimates the within-imputation variance-covariance matrix (denoted by H_n^{-1}), and $\bar{B}_m$ estimates the unknown between-imputation variance-covariance matrix B_n of $\tilde{\boldsymbol{\theta}}_{*l}$. Following Rubin [1987], when n is large, approximately we have

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

where $\Sigma_{n,m} = H_n^{-1} + (1 + 1/m)B_n$. Thus, approximately we have

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

where $\Sigma^*_{n,m} = RH_n^{-1}R^T + (1 + 1/m)RB_nR^T$. In the context of multivariate normal data with missing values, Li et al. [1991] and Meng and Rubin [1992] assumed equal ratios of missing to observed information for small m, i.e. $B_n \propto H_n^{-1}$. This assumption suggests that all eigenvalues λ_i of matrix H_nB_n are equal to $\bar{\lambda} = \frac{1}{k}\sum_{i=1}^{k}\lambda_i$, which leads to $B_n = \bar{\lambda}H_n^{-1}$. Thus, $\Sigma_{n,m} = H_n^{-1} + (1+1/m)\bar{\lambda}H_n^{-1}$. Following Johnson et al. [1995], we have $\sum_{i=1}^{k}\lambda_i = tr(H_nB_n)$. Therefore, $\Sigma_{n,m}$ can be estimated by $\widehat{\Sigma}_{n,m} = \bar{H}_m^{-1} + r_m\bar{H}_m^{-1}$, with $r_m = (1 + \frac{1}{m})\frac{1}{k}tr(\bar{H}_m\bar{B}_m)$.

For measurement errors in time-varying covariates, the assumption of equal ratios of missing to observed information is modified as the assumption of equal measurement errors at each time point, i.e., $RB_nR^T \propto RH_n^{-1}R^T$. This assumption seems

reasonable in the current context. The estimate of $\Sigma^*_{n,m}$ is correspondingly given by

$$\widehat{\Sigma}^*_{n,m} = R\bar{H}_m^{-1}R^T + r^*_m R\bar{H}_m^{-1}R^T,$$

where $r^*_m = (1+\frac{1}{m})\frac{1}{r}tr\left(R\bar{B}_m R^T\left(R\bar{H}_m^{-1}R^T\right)^{-1}\right)$.

Now we propose an overall Wald-type test statistic T^*_W for testing hypothesis (2.9) based on the combined "sufficient statistic" $\bar{\boldsymbol{\theta}}_m$ as follows.

$$\begin{aligned} T^*_W &= n(R\bar{\boldsymbol{\theta}}_m)^T\widehat{\Sigma}^{*-1}_{n,m}(R\bar{\boldsymbol{\theta}}_m) \\ &= n(1+r^*_m)^{-1}(R\bar{\boldsymbol{\theta}}_m)^T\left(R\bar{H}_m^{-1}R^T\right)^{-1}(R\bar{\boldsymbol{\theta}}_m). \end{aligned} \tag{2.10}$$

Note that the test statistic T^*_W measures the distance from the estimate $R\bar{\boldsymbol{\theta}}_m$ to the null parameter space $H_0 : R\boldsymbol{\theta} = 0$. Thus, the larger this distance, the stronger the evidence against H_0.

Alternatively, we also propose to obtain an overall test statistic by combining the individual test statistics based on each of the m "complete datasets". Specifically, for each of the m "complete datasets" $\left\{\left(\boldsymbol{y}, \boldsymbol{z}^{*(l)}\right), l = 1, \cdots, m\right\}$, the Wald-type test statistic for testing hypothesis (2.9) is given by

$$W_{*l} = n(R\tilde{\boldsymbol{\theta}}_{*l})^T\left(R\hat{H}^{-1}_{n*l}R^T\right)^{-1}(R\tilde{\boldsymbol{\theta}}_{*l}), \qquad l = 1, 2, \cdots, m. \tag{2.11}$$

We might tend to consider combining the above test statistics with a simple average $W_m = \sum_{l=1}^m W_{*l}/m$. However, this may lead to difficulties in computing the null distribution and thus difficulties in making inference [Wang and Wu, 2011]. We

propose to combine the individual test statistics as follows.

$$T^*_{CW} = \frac{W_m - \frac{(m-1)r}{m+1} r^*_m}{1 + r^*_m}.$$

The next step is to derive the asymptotic null distributions for the proposed test statistics T^*_W and T^*_{CW}. We have the following results, with the proofs given in the Appendix (A.2).

Theorem 2.3.3 *For the multivariate one-sided hypothesis (2.9) $H_0 : R\boldsymbol{\theta} = \mathbf{0}$ versus $H_1 : R\boldsymbol{\theta} \geq \mathbf{0}, R\boldsymbol{\theta} \neq \mathbf{0}$, the following result holds for the proposed test statistic T^*_W.*

$$\widehat{\Pr}(T^*_W \leq c | R\boldsymbol{\theta} = \mathbf{0}) = \sum_{i=0}^{r} \omega_i(r, \hat{\Sigma}^*_{n,m}, \mathbb{O}^+) \Pr(\chi^2_i \leq c), \quad \text{for any } \ c \geq 0. \tag{2.12}$$

where $\mathbb{O}^+ = \{R\boldsymbol{\theta} | R_i^T \boldsymbol{\theta} \geq 0, i = 1, \cdots, r\}$ with R_i being the ith row vector of R, $\omega_i(r, \hat{\Sigma}^*_{n,m}, \mathbb{O}^+)$ is the probability that $\pi_{\hat{\Sigma}^*_{n,m}}(\boldsymbol{x}; \mathbb{O}^+) \in$ the i dimensional face of $\mathbb{O}^+$ with $\sum_{i=0}^{r} \omega_i(r, \hat{\Sigma}^*_{n,m}, \mathbb{O}^+) = 1$, χ^2_i is the chi-square distribution with i degrees of freedom, and χ^2_0 denotes the distribution that takes the value zero with probability one.

Moreover, the asymptotic null distribution of T^*_W has the lower and upper bound:

$$\begin{aligned} \frac{\Pr(\chi^2_0 \geq c) + \Pr(\chi^2_1 \geq c)}{2} &\leq \Pr(T^*_W \geq c | R\boldsymbol{\theta} = \mathbf{0}) \\ &\leq \frac{\Pr(\chi^2_{r-1} \geq c) + \Pr(\chi^2_r \geq c)}{2}. \end{aligned} \tag{2.13}$$

Theorem 2.3.4 *For the multivariate one-sided hypothesis (2.9)* $H_0 : R\boldsymbol{\theta} = \mathbf{0}$ *versus* $H_1 : R\boldsymbol{\theta} \geq \mathbf{0}, R\boldsymbol{\theta} \neq \mathbf{0}$*, the proposed overall test statistics* T_W^* *and* T_{CW}^* *are asymptotically equivalent*

$$T_{CW}^* = T_W^* + o_p(1).$$

To calculate the p-values of proposed test statistics, we may consider two approaches: (a) *substitution method*: as shown in (2.12), substitute $\Sigma_{n,m}^*$ by its estimate $\hat{\Sigma}_{n,m}^*$, and compute the weights $\omega_i(r, \hat{\Sigma}_{n,m}^*, \mathbb{O}^+)$ by a simulation-based method; (b) *bound method:* we can use the lower and the upper bound in (2.13) to compute conservative p-values.

2.4 A Real Data Example

In this section, we apply the proposed tests T_W^* and T_{CW}^* to a real dataset for testing hypothesis (2.9) in NLME model (2.1) based on the MI method. For comparison purpose, we also consider a naive method which ignores covariate measurement errors, as well as a simple two-step method which first estimates the "true" covariate from the assumed covariate model and then fit the model (2.1) as if the estimated covariates were true values. And a two-sided alternative hypothesis $H_1 : \lambda_1 \neq 0$, or $\xi_1 \neq 0$ is considered based on the naive method and the two-step method.

The dataset is from an HIV/AIDS study which evaluates an anti-HIV treatment.

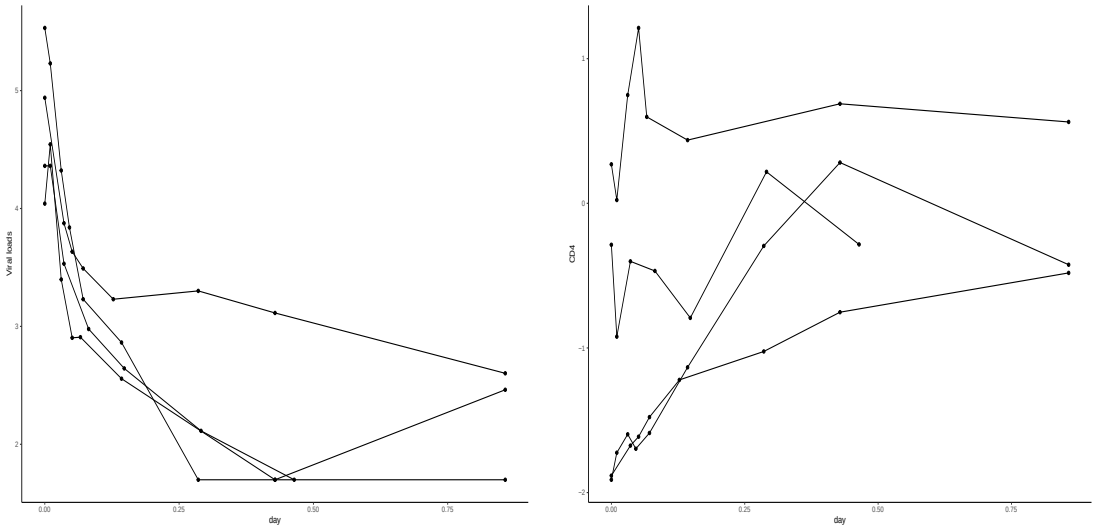
In this study, viral load of 46 HIV infected patients were repeatedly measured on days 0, 2, 7, 10, 14, 21, 28 and weeks 8, 12, 24, and 48 after the initiation of the treatment. The number of repeated measurements for each individual varies from 4 to 10. Various covariates such as CD4 count were also recorded throughout the study. Some values of viral load were below a detection limit of 100 copies/ml, and these values were imputed by half the detection limit for simplicity. Viral load values were $\log_{10}$-transformed to make their values more normally distributed and more stabilized variances over time. To avoid very small estimates, which may be unstable, we also standardize the CD4 count values and re-scale the original time t (in days) so that the new time scale is between 0 and 1. Figure 2.1 shows the trajectories of viral load and CD4 count of four randomly selected patients. We can see that, in most cases, the viral load declines during the treatment, while CD4 count shows a quadratic trend.

We fit the NLME model (2.1) to the viral load data, with CD4 count being an error-prone time-varying covariate. To address the measurement errors in CD4 count, we consider several empirical polynomial LME models for the CD4 process, and choose the best model based on AIC and BIC values. Specifically, we consider the covariate model (2.4) with $U_{ij} = V_{ij} = (1, t_{ij}, \ldots, t_{ij}^{d-1})$, where t_{ij}'s are CD4 measurement times, and focus on linear ($d = 2$), quadratic ($d = 3$), and cubic

($d = 4$) polynomials, whose AIC (BIC) values are 796.17 (819.50), 703.19 (761.52), and 742.12 (781.01), respectively. So the following quadratic polynomial LME model fits the observed CD4 data reasonably well

$$CD4_{ij} = (\alpha_1 + a_{i1}) + (\alpha_2 + a_{i2})t_{ij} + (\alpha_3 + a_{i3})t_{ij}^2 + \epsilon_{ij} \equiv CD4_{ij}^* + \epsilon_{ij}, \qquad (2.14)$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ are the fixed-effects and $\boldsymbol{a} = (a_1, a_2, a_3)$ are the random-effects. As discussed in Section 2.1, we are interested in testing the multivariate one-sided hypothesis $H_0 : \boldsymbol{\lambda} = 0$ versus $H_1 : \boldsymbol{\lambda} \geq 0, \boldsymbol{\lambda} \neq 0$, where $\boldsymbol{\lambda} = (\lambda_1, \xi_1)$.



(a) Viral Load (b) CD4 Cell Count

Figure 2.1: Trajectories of viral load in $\log_{10}$ scale and CD4 count for four randomly selected patients.

We consider $m = 5$ multiple imputations to address the measurement errors

Table 2.1: Parameter estimates of the viral load model (2.3) based on $m = 5$ multiple imputations

Response model						
Fixed-effect		Covariance matrix				
P_1	11.71			B		σ
λ_1	66.36	$\begin{pmatrix} 1.10 & 0.63 & 1.49 & -2.38 \\ & 69.71 & 0.84 & -1.21 \\ & & 2.01 & -4.96 \\ & & & 22.93 \end{pmatrix}$				
ξ_1	1.51					0.35
P_2	6.85					
λ_2	-2.81					

in CD4 cell count. For sensitivity analysis, we take two different values of hyper-parameters representing noninformative and strong priors and find out that the traces of the Markov chains with the different values of hyper-parameters almost overlap each other, which indicates that the Markov chain is not sensitive to the values of hyper-parameters. Here, we use noninformative prior distributions for model parameters as follows:

$$\begin{aligned}
\boldsymbol{\beta} &\sim N(\mathbf{0}, diag(100, 100, 100, 100, 100)), \\
\boldsymbol{\alpha} &\sim N(\mathbf{0}, diag(100, 100, 100)), \\
\sigma^2 &\sim G^{-1}(0.01, 100), \qquad D \sim W^{-1}(1, 1), \\
A &\sim W^{-1}(I, 3), \qquad B \sim W^{-1}(I, 4),
\end{aligned}$$

where the I's are appropriate identity matrices and $W(1, 1)$ (i.e., the chi-square distribution with degree of freedom 1) is a prior Wishart distribution for the scale parameter D.

In order to determine the number of "burn-in" iterations and to check the convergence of the Markov Chain, we start three Markov Chains from dispersed initial values. Based on these Markov Chains, we suggest that, after an initial number of 2000 burn-in iterations, every 500th MCMC sample is retained from the next 2500 samples for $m = 5$ independent multiple imputations. The Markov sampling is conducted by the software *WinBUGS* through R package *R2WinBUGS*.

The parameter estimates in model (2.1) and model (2.14) are listed in Table 1. The 5 estimates of λ_1 and ξ_1 based on the 5 "complete datasets" are (66.57, 1.53), (66.70, 1.55), (66.13, 1.50), (65.87, 1.50), (66.52, 1.47) respectively. Thus, we have $\bar{\boldsymbol{\lambda}}_m = (66.36, 1.51)$. The overall test statistics are calculated as $T_W^* = 494.27$ and $T_{CW}^* = 488.54$. Both the *substitution method* and the *bound method* are used to calculate the p-values, and both p-values are close to 0. Therefore, we have strong evidence to reject H_0 and conclude that the viral load declines significantly during the anti-HIV treatment and the CD4 count has a significantly positive association with the initial viral decay rate. That is, the higher the CD4 count, the faster the viral load declines during the initial period of the treatment. The testing statistics based on the naive and the two-step method are 306.49 and 359.45, respectively, which are substantially smaller than T_W^* and T_{CW}^*, and thus they provide weaker evidence against H_0. In other words, the proposed new tests provide stronger evidence about the efficacy of the treatment and its association with CD4 count than those tests based on the naive and the two-step method.

2.5 Simulation Study

In this section, we perform a simulation study to evaluate the performance of the proposed tests and compare them with those two-sided tests based on the naive

and the two-step method. We evaluate the methods in terms of both type I error probabilities and powers. The significance level $\alpha = 0.05$ is chosen. The simulation design uses the same NLME response model and the same covariate model as those in the real data example, but with different choices of the true parameters of λ_1 and ξ_1 values in order to evaluate the tests under different settings. The different true values of the parameters λ_1 and ξ_1 are listed in Tables 2.2 –2.3. The true values of the other fixed-effect parameters are chosen to be similar to those estimated from the real data example in the previous section. Specifically, $(P_1, P_2, \lambda_2) = (12, 7, -3)$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2) = (-0.5, 4, -4)$, $Diag(A) = (0.5, 2, 1), D = 0.5, Diag(B) = (1, 10, 2, 5)$, and $\sigma = 0.2$, where the covariance matrices A and B are chosen to be diagonal matrices.

We generate 1000 datasets with the two different sample sizes $n = 46$ and $n = 100$, respectively. The number of the within-individual repeated measurements are with the same measurement times $m_i = 8$. Both the *substitution method* and the *bound method* are used to calculate the approximate p-values, together with the naive and the two-step method described in the previous section.

The simulation results are reported in Table 2.2–2.3. We see that all tests approximately attain the nominal significant level $\alpha = 5\%$ in most of the cases, except for the *substitution method* when sample size $n = 46$. This might be due to the

fact that the *substitution method* may not estimate the covariance matrix $\Sigma^*_{n,m}$ well for the small sample size ($n = 46$). This problem disappears for large sample sizes ($n = 100$) since in this scenario the covariance matrix can be better estimated. We can see that our proposed new tests T^*_W and T^*_{CW} have somewhat lower type I errors than the naive and the two-step method in most cases, and in the meantime they are more powerful in all scenarios. These results clearly show the better performance of the proposed methods than the naive and the two-step method. Note that the *bound method* is more conservative because it uses the upper and lower bound of the null probabilities to compute the p-values, i.e., the *substitution method* produces higher powers than the *bound method* for both $n = 46$ and $n = 100$. However, the *substitution method* relies on a good estimation of the covariance matrix $\Sigma^*_{n,m}$, so it might not perform very well when the sample size is small.

Table 2.2: Type I error probabilities and powers when n=46.

(λ_1, ξ_1)	$T_W^{*(b)+}$	$T_{CW}^{*(b)+}$	$T_W^{*(s)+}$	$T_{CW}^{*(s)+}$	Naive	Two-step
	Type I error probability					
(0,0)	3.4	3.7	4.4	5.6	4.4	4.2
	Power					
(0.008,0)	10.9	12.0	13.4	14.5	6.1	7.6
(0,0.02)	38.7	40.1	43.3	44.5	11.0	13.3
(0.05,0.01)	53.5	53.3	55.7	55.6	35.0	38.1

$^{+}$ $T^{*(b)}$ and $T^{*(s)}$ denote test statistics based on the *bound* and the *substitution method*, respectively.

Table 2.3: Type I error probabilities and powers when n=100.

(λ_1, ξ_1)	$T_W^{*(b)}$	$T_{CW}^{*(b)}$	$T_W^{*(s)}$	$T_{CW}^{*(s)}$	Naive	Two-step
	Type I error probability					
(0,0)	3.1	3.6	4.0	4.8	3.3	4.1
	Power					
(0.008,0)	12.3	13.9	15.2	15.3	6.7	8.3
(0,0.02)	59.3	61.5	65.2	65.7	27.9	30.4
(0.05,0.01)	77.8	77.2	80.1	80.0	55.5	56.8

2.6 Discussions

In this chapter, we have proposed a multiple imputation method for time-varying covariate measurement errors in NLME models in the context of multivariate one-sided tests. A main advantage of multiple imputation methods is that existing complete-data methods for multivariate one-sided tests can be used with the imputed datasets. We have focused on the error-prone time-varying covariates in NLME models. The repeated measurements of the covariates allow us to model the covariate process to estimate measurement errors. Simulation results show that the proposed methods perform better than the naive and the two-step method in the sense that

they offer higher powers while maintaining nominal type I errors.

Multiple imputation methods are typically used to address missing data problems. Here we use them for measurement error problems. A good imputation model is important to generate "good imputations" in the sense that the imputed values are close to the unobserved true values while incorporating prediction uncertainty. To obtain a good imputation, we should use as much observed data as possible to "predict" the unobserved true values while maintaining the underlying model structures. This can be challenging for NLME models with "missing" covariates. In this chapter, we use both the observed covariate data and the response data to help generating the imputations. Another advantage of the proposed method is that the imputations are generated in the individual levels, so the method is very useful when there are large variations between individuals in the dataset. Simulation results show that the proposed approach works well.

For the problem under consideration, an alternative approach would be based on the likelihood method. A (joint) likelihood can be specified based on the assumed models. An (Monte Carlo) EM algorithm may then be used for parameter estimation and inference. While the likelihood method may be conceptually straightforward, such an approach may be computationally very intensive in the current model settings, especially convergence issues. Moreover, the existing methods for the mul-

tivariate one-sided hypothesis testing may not be readily used. In some sense, the likelihood method may be viewed as a multiple imputation method with infinite many imputations. On the other hand, it is known that a multiple imputation method can produce good results with only 5 or 6 imputations, saving much computation time [Rubin, 1987, Schafer, 1997].

There is a large literature on multivariate one-sided hypothesis testing, though most of the tests appear to be only of theoretical interest. It is well known that such tests are more powerful than those ignoring the restrictions on the parameter space. We hope that such tests can be more commonly used to analyze real datasets. In this chapter, we consider the multivariate one-sided tests in NLME models with time-varying covariate measurement errors. The proposed methods can also be used for missing covariates in NLME models when the missingness is missing at random. Moreover, the proposed methods may also be extended to generalized linear mixed-effects models or semi-parametric mixed-effects models with measurement errors in time-varying covariates.

3 Multivariate One-sided Hypothesis Tests for NLME Models with Covariate Measurement Errors and Response Left-censoring

3.1 Introduction

HIV dynamic models have been developed to help study HIV parthenogenesis and treatment strategies for AIDS patients. In practice, the viral load in a HIV study is are usually left-censored at the lower detection limit due to the technology constraints. The "fill-in" method is commonly used to deal with the left-censoring problem which substitutes the censoring data by the limit of detection(LOD) or some arbitrary values such as LOD/2 or LOD/$\sqrt{2}$. However, the "fill-in" method may lead to biased results [Wu, 2002, Vaida and Liu, 2009].

Note that the censoring is related to the unobserved values (the censored values) since the censored values are known to be less or greater than a known number. So

censored data can be considered as a special case of missing data with missing not at random mechanism. There are two main formal methods to deal with missing data problems, i.e. multiple imputation (MI) methods and expectation maximization (EM) algorithms. The EM algorithm needs to compute the expectation of the complete log-likelihood function with respect to missing data distribution at current estimates in the E-step, then maximize the expected log-likelihood function to update unknown parameters in the M-step. Compared with EM algorithm, the multiple imputation methods are more computationally efficient and the existing standard complete methods can be used in the subsequent statistical analysis.

In this chapter, we consider the MI method to address both covariate measurement errors and response left-censoring in NLME models. We proposed two Wald-type test statistics for the multivariate one-sided hypothesis for testing parameters in NLME models. This chapter is organized as follows. In Section 3.2, we describe a multiple imputation method to address both the left-censored response and the mis-measured time-varying covariates. In Section 3.3, we propose two Wald-type testing statistics for the multivariate one-sided hypothesis of parameters in NLME models based on multiple imputations of mis-measured covariates and left-censored response. In Sections 3.4, we illustrate the proposed approaches in a real data example. In section 3.5, a simulation study is conducted to evaluate the performance of

the proposed tests. We conclude this chapter in Section 3.6 with some discussions.

3.2 A Multiple Imputation for NLME Models with Response Left-censoring and Covariate Measurement Errors

Let y_{ij} be the response value for individual i at time t_{ij} subject to the left-censoring due to a known limit d of detection, $i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, m_i$. In the presence of left-censored response, we denote $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{im_i}) = (\boldsymbol{y}_i^c, \boldsymbol{y}_i^o)$ be the collection of the response values for individual i, where $\boldsymbol{y}_i^c$ and $\boldsymbol{y}_i^o$ are the censored and the observed component of $\boldsymbol{y}_i$, respectively. So for individual i, some of y_{ij}'s are observed while the other may be left-censored. To indicate the response left-censoring for individual i, we denote the censoring indicator vector by $\boldsymbol{c}_i = (c_{i1}, \cdots, c_{im_i})$ such that $c_{ij} = 1$ if y_{ij} is left-censored, i.e. $y_{ij} \leq d$ and 0 otherwise. Let z_{qij} and z_{qij}^* be the observed and the corresponding unobservable "true" value of the qth error-prone covariate for individual i at time t_{ij}, $q = 1, \cdots, s, i = 1, \cdots, n, j = 1, \cdots, m_i$, then we have $\boldsymbol{z}_i = (\boldsymbol{z}_{i1}^T, \cdots, \boldsymbol{z}_{im_i}^T)^T$, $\boldsymbol{z}_{ij} = (z_{1ij}, \cdots, z_{sij})^T$, where T denotes a transpose. Then the observed data can be denoted by $\mathcal{D}_1 = (\boldsymbol{y}_i^o, \boldsymbol{c}_i, \boldsymbol{z}_i), i = 1, \cdots, n$.

For completeness, we describe the general response NLME models with covariate

measurement errors as follows.

$$\begin{aligned} y_{ij} &= g(t_{ij}, \boldsymbol{\beta}_{ij}) + e_{ij}, \quad e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2), \\ \boldsymbol{\beta}_{ij} &= \mathbf{d}(\boldsymbol{z}^*_{ij}, \boldsymbol{\beta}, \mathbf{b}_i), \quad \mathbf{b}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, B), \qquad i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, m_i \end{aligned} \tag{3.1}$$

where $g(\cdot)$ is a known nonlinear function, $\mathbf{d}(\cdot)$ is a known multivariate linear function, $\boldsymbol{\beta}_{ij}$ is a vector of individual-specific and time-varying parameters and $\boldsymbol{\beta}$ is a vector of fixed-effect, $\boldsymbol{z}^*_{ij}$ is the (unobserved) true covariate value corresponding to the observed and possibly mis-measured $\boldsymbol{z}_{ij}$, $\boldsymbol{e}_i = (e_{i1}, \cdots, e_{im_i})^T$ is a vector of within-individual random errors, $\mathbf{b}_i$ is a vector of random-effects, σ^2 is a variance parameter, and B is the variance-covariance matrix of $\boldsymbol{b}_i$. We assume $\boldsymbol{e}_i$ and $\boldsymbol{b}_i$ are independent.

In the presence of measurement errors in time-varying covariates, the following classical measurement error model is considered to model the true (unobservable) covariate $\boldsymbol{z}^*_{ij}$ in NLME model (3.1).

$$\boldsymbol{z}_{ij} = U_{ij}\boldsymbol{\alpha} + V_{ij}\boldsymbol{a}_i + \boldsymbol{\epsilon}_{ij} \ \left(\equiv \boldsymbol{z}^*_{ij} + \boldsymbol{\epsilon}_{ij}\right), \quad i = 1, \cdots, n, j = 1, \cdots, m_i, \tag{3.2}$$

where U_{ij} and V_{ij} are known design matrices including time, and $\boldsymbol{\alpha}$ and $\boldsymbol{a}_i$ are the fixed-effect vector and random-effect, vector respectively, $\boldsymbol{\epsilon}_{ij}$ are the random w terrors for individual i at time t_{ij}.

We propose a MI method to address the left-censored response and the mis-measured covariate in NLME models. Specifically, we can generate m proper multiple imputations for $\{(\boldsymbol{y}^c_i, \boldsymbol{z}^*_i), i = 1, \cdots, n\}$ under the Bayesian framework. As

discussed in section 2.2, the unobserved "true" covariate values $\boldsymbol{z}_{ij}^{*} = U_{ij}\boldsymbol{\alpha} + V_{ij}\boldsymbol{a}_i$ are completely determined by its mean parameters $\boldsymbol{\alpha}$ and its random-effects $\boldsymbol{a}_i$ since the design matrices U_{ij} and V_{ij} are known. Therefore, it is sufficient to generate multiple imputations for $(\boldsymbol{y}^c, \boldsymbol{\alpha}, \boldsymbol{a})$ from the predictive distribution $f(\boldsymbol{y}^c, \boldsymbol{\alpha}, \boldsymbol{a}|\mathcal{D}_1)$, where $\boldsymbol{y}^c = \{\boldsymbol{y}_i^c, i = 1, \cdots, n\}$ and $\boldsymbol{a} = \{\boldsymbol{a}_i, i = 1, \cdots, n\}$.

Let $\Psi = (\boldsymbol{\alpha}, \Phi) \equiv (\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, D, A, B)$ be the collection of all unknown parameters in models (3.1) and (3.2), where $\Phi = (\boldsymbol{\beta}, \sigma^2, D, A, B)$, $\boldsymbol{b}$ denotes the collection of the random-effects $\boldsymbol{b}_i$'s in the response model (3.1). Under the Bayesian framework, we need to specify prior distributions for the parameter Ψ. We assume that the parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma, D, A$ and B are independent, i.e. $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, D, A, B) = \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(D)\pi(A)\pi(B)$. Normal prior distributions are assigned to the mean parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, while the Inverse Gamma or Inverse Wishart prior distributions are assigned to the variance-covariance parameters σ^2, D, A and B.

However, the analytical computation convenience associated with the conjugate prior does not hold for the censored data. Thus, the data augmentation method is applied to handle the censoring problems within the Gibbs sampling framework [Chib, 1992, Wei and Tanner, 1990]. In particular, for the censored response $\boldsymbol{y}_i^c$, we will sample the imputations successively from the conditional distribution $f(\boldsymbol{y}_i^c|\mathcal{D}_1)$, which is a truncated multivariate normal distribution. Based on the model (3.1) and

(3.2), we have $y_{ij}|\boldsymbol{a}_i, \boldsymbol{b}_i \overset{i.i.d.}{\sim} N(g(t_{ij}, d(U_{ij}\boldsymbol{\alpha} + V_{ij}\boldsymbol{a}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \sigma^2)$. Then the multiple imputations for $(\boldsymbol{y}^c, \boldsymbol{\alpha}, \boldsymbol{a})$ can be generated through sampling the full conditionals in turn at iteration t $(t = 1, 2, \cdots)$,

Step I: Draw a value $(\boldsymbol{\alpha}^{(t)}, \boldsymbol{a}^{(t)})$ of $(\boldsymbol{\alpha}, \boldsymbol{a})$ in the following order

i Draw a value of $\boldsymbol{\alpha}^{(t)}$ of the mean parameter $\boldsymbol{\alpha}$ in the covariate model (3.2) from

$$
\begin{aligned}
&f(\boldsymbol{\alpha}|\boldsymbol{y}^{c(t-1)}, \mathcal{D}_1, \boldsymbol{a}^{(t-1)}, \boldsymbol{b}^{(t-1)}; \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}, D^{(t-1)}) \\
&\propto \prod_{i=1}^{n} \Big[f(\boldsymbol{y}_i^{c(t-1)}, \boldsymbol{y}_i^o|\boldsymbol{a}_i^{(t-1)}, \boldsymbol{b}_i^{(t-1)}; \boldsymbol{\alpha}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \\
&\qquad \times f(\boldsymbol{z}_i|\boldsymbol{a}_i^{(t-1)}; \boldsymbol{\alpha}, D^{(t-1)}) \Big] \cdot \pi(\boldsymbol{\alpha}).
\end{aligned}
$$

ii Draw a value $\boldsymbol{a}_i^{(t)}$ of the random-effect $\boldsymbol{a}_i$ in the covariate model (3.2) from

$$
\begin{aligned}
&f(\boldsymbol{a}_i|\boldsymbol{y}_i^{c(t-1)}, \boldsymbol{y}_i^o, \boldsymbol{z}_i, \boldsymbol{b}_i^{(t-1)}; \boldsymbol{\alpha}^{(t)}, \Phi^{(t-1)}) \\
&\propto f(\boldsymbol{y}_i^{c(t-1)}, \boldsymbol{y}_i^o|\boldsymbol{a}_i, \boldsymbol{b}_i^{(t-1)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \\
&\qquad \times f(\boldsymbol{z}_i|\boldsymbol{a}_i; \boldsymbol{\alpha}^{(t)}, D^{(t-1)}) f(\boldsymbol{a}_i|A^{(t-1)}), \qquad i = 1, \cdots, n.
\end{aligned}
$$

Step II: Draw a value $(\boldsymbol{b}^{(t)}, \Phi^{(t)})$ of $(\boldsymbol{b}, \Phi)$ in the following order

i Draw a value $\boldsymbol{b}_i^{(t)}$ of the random-effect $\boldsymbol{b}_i$ in the response model (3.1) from

$$
\begin{aligned}
&f(\boldsymbol{b}_i|\boldsymbol{y}_i^{c(t-1)}, \boldsymbol{y}_i^o, \boldsymbol{a}_i^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}, B^{(t-1)}) \\
&\propto f(\boldsymbol{y}_i^{c(t-1)}, \boldsymbol{y}_i^o|\boldsymbol{a}_i^{(t)}, \boldsymbol{b}_i; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \\
&\qquad \times f(\boldsymbol{b}_i|B^{(t-1)}), \quad i = 1, \cdots, n.
\end{aligned}
$$

ii Draw a value $\boldsymbol{\beta}^{(t)}$ of the mean parameter $\boldsymbol{\beta}$ in the response model (3.1) from

$$f(\boldsymbol{\beta}|\boldsymbol{y}_i^{c(t-1)}, \mathcal{D}_1, \boldsymbol{a}^{(t)}, \boldsymbol{b}^{(t)}; \boldsymbol{\alpha}^{(t)}, \sigma^{2(t-1)})$$
$$\propto \prod_{i=1}^{n} \left[f(\boldsymbol{y}_i^{c(t-1)}, \boldsymbol{y}_i^o|\boldsymbol{a}_i^{(t)}, \boldsymbol{b}_i^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}, \sigma^{2(t-1)}) \right] \cdot \pi(\boldsymbol{\beta}).$$

Step III: If $c_{ij} = 1, i = 1, \cdots, n, j = 1, \cdots, m_i$, draw a value $y_{ij}^{c(t)}$ of the censored response y_{ij}^c successively from

$$y_{ij}^c|c_{ij} = 1, \boldsymbol{a}_i^{(t)}, \boldsymbol{b}_i^{(t)}; \boldsymbol{\alpha}^{(t)}, \sigma^{2(t)} \sim TruncNormal_{(-\infty,d]}\left(\mu_{ij}^{(t)}, \sigma^{2(t)}\right), j = 1, \cdots, m_i,$$

with $\mu_{ij}^{(t)} = g(t_{ij}, d(U_{ij}\boldsymbol{\alpha}^{(t)} + V_{ij}\boldsymbol{a}_i^{(t)}, \boldsymbol{\beta}^{(t)}, \boldsymbol{b}_i^{(t)})$, $\text{truncNormal}_{(a,b)}(\mu, \sigma^2)$ denoting the normal distribution density $N(\mu, \sigma^2)$ truncated on the interval (a, b) as illustrated in Gelfand and Smith [1990], Lee et al. [2012].

Step VI: Draw a value $(\sigma^{2(t)}, D^{(t)}, A^{(t)}, B^{(t)})$ of the precision parameters σ^2, D, A, and B in the models (3.1) and (3.2) from the corresponding full conditional distributions in similar ways as discussed in Chapter 2.

Beginning with the starting value $\Psi^{(0)}$ obtained from the two-step method, we iterate the foregoing procedures for a burn-in period until the resulting Markov Chain converges to its stationary distribution. At convergence, we obtain a value of $(\boldsymbol{y}_i^c, \boldsymbol{\alpha}, \boldsymbol{a}_i)$ from $f(\boldsymbol{y}_i^c, \boldsymbol{\alpha}, \boldsymbol{a}_i|\mathcal{D}_1)$. Repeating the above procedure m times, we obtain m imputations of $\left(\boldsymbol{y}_i^{c(l)}, \boldsymbol{z}_i^{*(l)}\right)$ from $f(\boldsymbol{y}^c, \boldsymbol{z}^*|\mathcal{D}_1)$. Thus, we obtain m "complete datasets" $\left\{\left(\boldsymbol{y}_i^{(l)}, \boldsymbol{z}_i^{*(l)}\right), i = 1, 2, \cdots, n\right\}$, $l = 1, 2, \cdots, m$ with $\boldsymbol{y}_i^{(l)} = \left(\boldsymbol{y}_i^{c(l)}, \boldsymbol{y}_i^o\right)$.

Based on each of the m "complete datasets", we can conduct inference for the NLME model (3.1) with both response left-censoring and covariate measurement errors being addressed, and then we combine the m results to obtain an overall conclusion.

In the next section, we consider the multivariate one-sided hypothesis testing for parameters in the NLME models based on multiply imputed datasets and show how to combine the results.

3.3 Multivariate One-sided Tests Based on Multiple Imputations in NLME Models with Response Censoring and Covariate Measurement Errors

In the previous section, we create m imputations for the left-censored response $\boldsymbol{y}_i^c$ and the unobserved true covariate value $\boldsymbol{z}_i^*$ and obtain m "complete datasets" $\left\{\left(\boldsymbol{y}_i^{(l)}, \boldsymbol{z}_i^{*(l)}\right),\ i = 1, 2, \cdots, m\right\}$, $l = 1, 2, \cdots, m$ with $\boldsymbol{y}_i^{(l)} = \left(\boldsymbol{y}_i^{c(l)}, \boldsymbol{y}_i^o\right)$ being the l-th imputed response data for individual i. For each of the m "complete datasets", we can conduct the multivariate one-sided hypothesis testing for parameters in the NLME model (3.1) using existing complete-data methods and then we combine the results to draw an overall conclusion.

As in Section 2.3, we propose two approaches to combine the results from the m hypothesis tests, (i) combine the m "complete data" parameter estimates and then

construct an overall single test statistic; (ii) combine the m sufficient statistics based on each of the m "complete datasets" to obtain an overall test statistic [Meng and Rubin, 1992, Wang and Wu, 2011]. For either case, it is important to derive the null distribution or calculate the p-value for the test statistics.

We consider the following multivariate one-sided hypothesis test of interest

$$H_0 : R\boldsymbol{\theta} = \mathbf{0} \qquad \text{versus} \qquad H_1 : R\boldsymbol{\theta} \geq \mathbf{0}, R\boldsymbol{\theta} \neq \mathbf{0}. \tag{3.3}$$

where $\boldsymbol{\theta}$ is a k-dimension parameter vector in NLME models and R is a $r \times k$ ($r \leq k$) full rank matrix with elements of 1's and 0's indicating the parameters of interest. For the l-th "complete dataset" $\left\{\boldsymbol{y}_i^{(l)}, \boldsymbol{z}_i^{*(l)}, i = 1, \cdots, n\right\}$, let $\tilde{\boldsymbol{\theta}}_{*l}$ be the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ under the parameter space $\{\boldsymbol{\theta} : R\boldsymbol{\theta} \geq \mathbf{0}\}$, $\hat{H}_{n*l}$ be the corresponding observed information matrix $-\frac{1}{n}\nabla^2 l_{obs}(\boldsymbol{\theta})$ evaluated at $\tilde{\boldsymbol{\theta}}$, which is an estimate of variance covariance matrix H_n^{-1}. Consider the following combined statistics

$$\begin{aligned}
\bar{\boldsymbol{\theta}}_m &= \frac{1}{m}\sum_{l=1}^{m} \tilde{\boldsymbol{\theta}}_{*l}, \\
\bar{H}_m &= \frac{1}{m}\sum_{l=1}^{m} \hat{H}_{n*l}, \\
\bar{B}_m &= \frac{1}{m-1}\sum_{l=1}^{m} (\tilde{\boldsymbol{\theta}}_{*l} - \bar{\boldsymbol{\theta}}_m)(\tilde{\boldsymbol{\theta}}_{*l} - \bar{\boldsymbol{\theta}}_m)^T,
\end{aligned}$$

where $\bar{\boldsymbol{\theta}}_m$ is an estimator of $\boldsymbol{\theta}$, $\bar{H}_m^{-1}$ estimates the within-imputation variance-covariance matrix of $\boldsymbol{\theta}$ (denoted by H_n^{-1}), and $\bar{B}_m$ estimates the between-imputation

variance-covariance B_n of $\tilde{\boldsymbol{\theta}}_{*l}$. When n is large, we have $\bar{\boldsymbol{\theta}}_m \overset{approx.}{\sim} N(\boldsymbol{\theta}, \Sigma_{n,m})$ with $\Sigma_{n,m} = H_n^{-1} + (1 + 1/m)B_n$ [Rubin, 1987]. Therefore, we have $R\bar{\boldsymbol{\theta}}_m \overset{approx.}{\sim} N(R\boldsymbol{\theta}, \Sigma^*_{n,m})$ with $\Sigma^*_{n,m} = RH_n^{-1}R^T + (1 + 1/m)RB_nR^T$.

B_n can not be well estimated by B_m unless m is large enough, thereby leading to substantially intensive computations. Alternatively, we have B_n proportional to H_n^{-1} under the assumption of equal missing fractions, that is, the fractional loss of information is the same for all components of $\boldsymbol{\theta}$. Specifically, all eigenvalues λ_i of matrix H_nB_n are equal to $\bar{\lambda} = \frac{1}{k}\sum_{i=1}^{k}\lambda_i$ and $B_n = \bar{\lambda}H_n^{-1}$. Following Johnson et al. [1995], we have

$$\widehat{\Sigma}_{n,m} = \bar{H}_m^{-1} + r_m\bar{H}_m^{-1}, \quad \text{with } r_m = (1 + \frac{1}{m})\frac{1}{k}tr(\bar{H}_m\bar{B}_m).$$

Besides, even though the equal missing assumption is violated when $m \geq 3$, $\widehat{\Sigma}_{n,m}$ would be asymptotically calibrated and suffer only modest loss of power [Li et al., 1991]. Therefore, for the case with left-censored response and measurement errors in time-varying covariates, the estimate of $\Sigma^*_{n,m}$ can be correspondingly given by

$$\widehat{\Sigma}^*_{n,m} = R\bar{H}_m^{-1}R^T + r^*_mR\bar{H}_m^{-1}R^T,$$

where $r^*_m = (1 + \frac{1}{m})\frac{1}{r}tr\left(\left(R\bar{H}_m^{-1}R^T\right)^{-1}R\bar{B}_mR^T\right)$.

Then we can propose an overall Wald-type test statistic for hypothesis (3.3) based

on the combined parameter estimates $\bar{\boldsymbol{\theta}}_m$ as follows,

$$T_W^* = n(R\bar{\boldsymbol{\theta}}_m)^T \widehat{\Sigma}_{n,m}^{*-1}(R\bar{\boldsymbol{\theta}}_m). \tag{3.4}$$

Alternatively, we also propose to obtain an overall test statistic by combining the individual test statistics based on each of the m "complete datasets". The Wald-type test statistic for each of the m "complete datasets" $\left\{\boldsymbol{y}_i^{(l)}, \boldsymbol{z}_i^{*(l)}, i = 1, \cdots, n\right\}, l = 1, \cdots, m$ of hypothesis (3.3) is given by

$$W_{*l} = n(R\tilde{\boldsymbol{\theta}}_{*l})^T \left(R\hat{H}_{*l}^{-1}R^T\right)^{-1}(R\tilde{\boldsymbol{\theta}}_{*l}), \qquad l = 1, 2, \cdots, m. \tag{3.5}$$

We propose to combine the individual test statistics as follows.

$$T_{CW}^* = \frac{W_m - \frac{(m-1)r}{m+1} r_m^*}{1 + r_m^*}. \tag{3.6}$$

The asymptotic null distributions for the proposed test statistics T_W^* and T_{CW}^* are given in the following theorem, which can be proved in a same way as in Theorem 2.3.3 and Theorem 2.3.4.

Theorem 3.3.1 *For the multivariate one-sided hypothesis (3.3) $H_0 : R\boldsymbol{\theta} = \mathbf{0}$ versus $H_1 : R\boldsymbol{\theta} \geq \mathbf{0}, R\boldsymbol{\theta} \neq \mathbf{0}$, the following results hold for the proposed test statistics T_W^* and T_{CW}^*.*

(i) *The asymptotic null distribution of T_W^* is estimated by the following $\bar{\chi}^2$ distribu-*

tion.

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

where $\mathbb{O}^+ = \{R\boldsymbol{\theta} | R_i^T \boldsymbol{\theta} \geq 0, i = 1, \cdots, r\}$ *with* R_i *being the* i*th row vector of* R, $\omega_i(r, \hat{\Sigma}_{n,m}^*, \mathbb{O}^+)$ *is the probability that* $\pi_{\hat{\Sigma}_{n,m}^*}(\boldsymbol{x}; \mathbb{O}^+) \in$ *the* i *dimensional face of* $\mathbb{O}^+$ *with* $\sum_{i=0}^{r} \omega_i(r, \hat{\Sigma}_{n,m}^*, \mathbb{O}^+) = 1$, χ_i^2 *is the chi-square distribution with* i *degrees of freedom, and* χ_0^2 *denotes the distribution that takes the value zero with probability one.*

(ii) *The asymptotic null distribution of* T_W^* *has the lower and the upper bound:*

$$\begin{aligned} \frac{\Pr(\chi_0^2 \geq c) + \Pr(\chi_1^2 \geq c)}{2} &\leq \Pr(T_W^* \geq c | R\boldsymbol{\theta} = \mathbf{0}) \\ &\leq \frac{\Pr(\chi_{r-1}^2 \geq c) + \Pr(\chi_r^2 \geq c)}{2}. \end{aligned}$$

(iii) *The proposed overall test statistics* T_W^* *and* T_{CW}^* *are asymptotically equivalent*

$$T_{CW}^* = T_W^* + o_p(1).$$

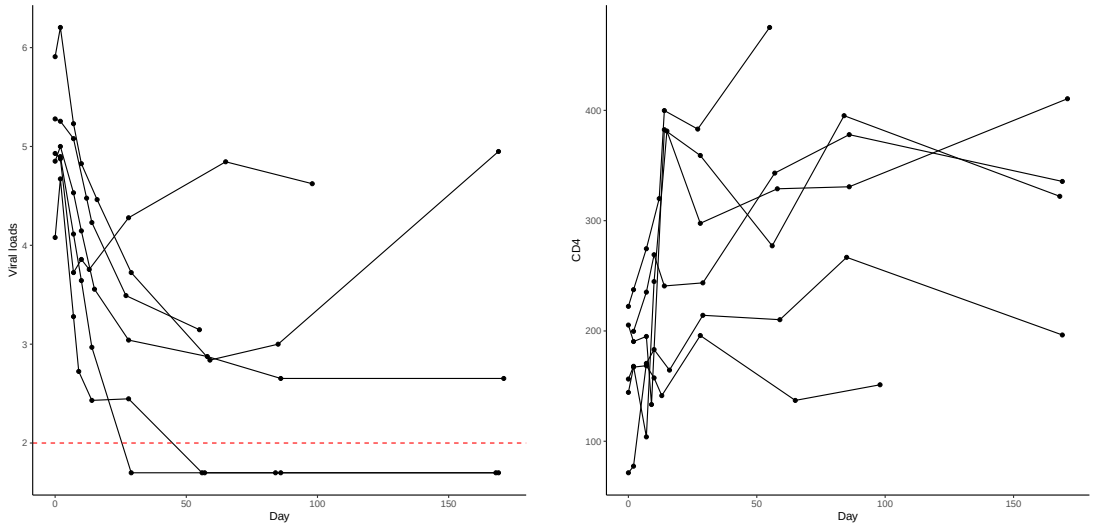
Similar to Chapter 2, we may either use the *substitution method* in Theorem 3.3.1 (i) or the *bound method* in Theorem 3.3.1 (ii) to approximate the null distribution. The *substitution method* would substitute $\Sigma_{n,m}^*$ by its estimate $\hat{\Sigma}_{n,m}^*$, and compute the weights $\omega_i(p, \hat{\Sigma}_{n,m}^*, \mathbb{O}^+)$ by a simulation-based method [Silvapulle and Sen, 2005]. The *bound method* uses the lower and upper bound in Theorem 3.3.1 to compute p-values which is computationally simple, but somehow conservative.

3.4 A Real Data Example

In this section, we apply the proposed tests T_W^* and T_{CW}^* to a real data example for the multivariate one-sided hypothesis test (3.3). For comparison purpose, we also consider a two-sided alternative hypothesis $H_1 : \lambda_1 \neq 0$, or $\xi_1 \neq 0$ based on the commonly-used estimation methods: the naive method and the two-step method. The naive method ignores the covariate measurement errors and simply imputes the left-censoring response values with half of the detection limit, while the two-step method first estimates the "true" mis-measured covariate values based on an assumed covariate model and then incorporates the estimated "true" covariate values into the response model with the censored response values imputed by half of the detection limit.

We use the same dataset as in Chapter 2. The data are from an anti-HIV treatment study which includes the viral load, CD4 count and other variables for 46 HIV infected patients measured over a period of 48 weeks after the initiation of the treatment. 40 out of 361 viral load observations are below a detection limit of 100, and they are simply imputed by half of the detection limit. Figure 3.1 (a) shows the trajectories of viral load in $\log_{10}$ scale for six randomly selected patients with the red dashed line representing the censoring threshold. We can see that the long-term viral load trajectories can be very complex after the initial phase viral decay. It may

continue to decay, fluctuate or even start to rise(rebound). And the left-censored viral load is demonstrated by a solid lines below the red dashed line whose value is equal to the half of the detection limit 100. The CD4 count trajectories for six randomly selected patients plotted in Fig 3.1 (b) show large inter-individual variation but a quadratic trend overall, suggesting a mixed-effects model. To avoid very small estimates, which may be unstable, we standardize the CD4 count values and re-scale the original time t (in days) so that the new time scale is between 0 and 1.



(a) Viral Load

(b) CD4 Cell Count

Figure 3.1: Trajectories of viral load in $\log_{10}$ scale and CD4 count for six randomly selected patients with the dashed red horizontal line indicating the limit of detection 100 units.

Based on biological arguments, we consider the following NLME model [Wu, 2002]

$$\begin{aligned} y_{ij} &= \log_{10}(e^{P_{1i}-\lambda_{1ij}t_{ij}} + e^{P_{2i}-\lambda_{2i}t_{ij}}) + e_{ij}, \\ P_{1i} &= P_1 + b_{1i}, \quad \lambda_{1ij} = \lambda_1 + \xi_1 CD4^*_{ij} + b_{2i}, \\ P_{2i} &= P_2 + b_{3i}, \quad \lambda_{2i} = \lambda_2 + b_{4i}, \\ & i = 1, 2, \cdots, n, \quad j = 1, 2, \cdots, m_i. \end{aligned} \tag{3.7}$$

where y_{ij} is the log10-transformation of viral load measurements for patient i at time t_{ij}, $CD4_{ij}$ is the CD4 count for patient i at time t_{ij}, P_{1i} and P_{2i} are baseline viral load, λ_{1ij} and λ_{2i} are the first (initial) and the second phases of viral decay rates, respectively. $\boldsymbol{\beta} = (P_1, P_2, \lambda_1, \lambda_2, \xi_1)^T$ are the fixed-effect parameters, $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^T$ are random-effects, and e_{ij} is a within-individual random error. We assume that $\boldsymbol{b}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, B)$, $e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, and $\boldsymbol{b}_i$ is independent of e_{ij}'s.

To address the measurement errors, we consider a linear mixed-effects model for CD4 count [Wu, 2002]

$$CD4_{ij} = (\alpha_1 + a_{i1}) + (\alpha_2 + a_{i2})t_{ij} + (\alpha_3 + a_{i3})t_{ij}^2 + \epsilon_{ij} \tag{3.8}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T$ are the fixed-effects and $\boldsymbol{a} = (a_1, a_2, a_3)^T$ are the random-effects. We assume the unobservable true $CD4^*_{ij} = (\alpha_1 + a_{i1}) + (\alpha_2 + a_{i2})t_{ij} + (\alpha_3 + a_{i3})t_{ij}^2$.

We use the same noninformative prior distributions as we used in Chapter 2. After an initial number of 2000 burn-in iterations, every 500th MCMC sample is retained

Table 3.1: Parameter estimates of the viral load model (3.7) based on $m = 5$ multiple imputations

Response model						
Fixed effect		Covariance matrix				
P_1	11.32	B				σ
λ_1	67.08	$\begin{pmatrix} 1.09 & 0.61 & 1.49 & -2.38 \\ & 70.62 & 0.85 & -1.20 \\ & & 2.00 & -4.93 \\ & & & 19.81 \end{pmatrix}$				0.35
ξ_1	1.48					
P_2	6.88					
λ_2	-1.98					

from the next 2500 samples for $m = 5$ independent multiple imputations. The estimates of λ_1 and ξ_1 based on the 5 generated " complete datasets" are (67.43, 1.42), (67.03, 1.44), (66.82, 1.49), (66.97, 1.50), and (67.16, 1.57) respectively. Therefore, $\bar{\boldsymbol{\lambda}}_m = (67.08, 1.48)$. The other parameter estimates of the model (3.7) based on the multiple imputations are listed in Table 3.4. To evaluate the efficacy of the treatment, we are interested in testing the one-sided hypothesis $H_0 : \boldsymbol{\lambda} = 0$ v.s. $H_1 : \boldsymbol{\lambda} \geq 0, \boldsymbol{\lambda} \neq 0$, where $\boldsymbol{\lambda} = (\lambda_1, \xi_1)$. The statistics T_W^* and T_{CW}^* for testing H_0 versus H_1 using the *substitution method* are calculated as $T_W^* = 532.19$ and $T_{CW}^* = 507.25$. The *bound method* is also used to compute the p-values. All the p-values of these tests

are close to 0. Note that the statistics in Chapter 2 are calculated as $T_W^* = 494.27$ and $T_{CW}^* = 488.54$, which is smaller than their counterparts in Chapter 3, since we only address the covariate measurement errors in Chapter 2. Thus, it is important to take the response left-censoring into consideration.

Therefore, we have strong evidence to reject H_0, and we may conclude that the CD4 count increases significantly and the viral load declines significantly in the initial period of the anti-HIV treatment. The two-sided testing statistics based on the naive and the two-step method are 307.01 and 360.14, respectively. They reject H_0 too but with weaker evidence than the newly proposed testing statistics T_W^* and T_{CW}^*.

3.5 Simulation Study

In this section, we conduct a simulation study to evaluate the performance of the proposed test statistics T_W^* and T_{CW}^* for the multivariate one-sided hypothesis testing problem in NLME models with response left-censoring and covariate measurement errors simultaneously under different scenarios. The corresponding two-sided hypothesis tests based on the naive and the two step method are considered for the comparison purpose. The significance level $\alpha = 0.05$ is chosen.

The simulation design uses the same NLME response model (3.7) and the same covariate model (3.8) as that in the real data example, but with different choices of

true parameter values in order to evaluate the tests under different settings. The sample size n are set to be 46 and 100 respectively, and the number of repeated measurements is $m_i = 8$.

The true values of the parameters λ_1 and ξ_1 are listed in Table 3.2. The true values of the other fixed-effect parameters are: $(P_1, P_2, \lambda_2) = (10, 5, -1)$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2) = (-0.5, 4, -4)$, $Diag(A) = (0.5, 2, 1), D = 0.5, Diag(B) = (1, 10, 2, 5), \sigma = 0.5$, where the covariance matrices A amd B are set to be diagonal matrices. Besides, the censoring threshold for response is set to be $d_{100} = log_{10}(100) = 2$ which is the detection limits of commonly used HIV viral load measurement devices [Wu, 2009]. All the simulations are repeated 1000 times. And the censoring rate is the average censoring rate over the 1000 simulated datasets.

The multivariate one-sided hypothesis test of interest in the simulation is $H_0 : \lambda_1 = \xi_1 = 0$ versus $H_1 : \lambda_1 \geq 0, \xi_1 \geq 0$ (with at least one strict inequality). As a comparison, the two-sided hypothesis $H_0 : \lambda_1 = \xi_1 = 0$ versus $H_1 : \lambda_1 \neq 0$ or $\xi_1 \neq 0$ using the naive method and two-step method would be considered.

The simulation results are shown in Table 3.2. We can see that all the tests attain the significant level $\alpha = 5\%$ except for the *substitution method* when sample size $n = 46$. The exception may be due to that the *substitution method* cannot estimate the covariance matrix $\Sigma^*_{n,m}$ well for the small sample size ($n = 46$). Overall,

the type I error rates of each testing procedures are close to the nominal level 5%.

Power is an important performance measure of the hypothesis tests. The proposed multivariate one-sided tests consistently show higher powers than the two-sided tests based on the commonly-used naive and two-step method, especially for the scenarios with the smaller sample size ($n = 46$) and the higher censoring rates. We can see that the proposed new tests T_W^* and T_{CW}^* have higher powers than those based on the naive and the two-step method all scenarios while maintaining the nominal level, which demonstrate the better ability to detect the true difference.

Regarding the two p-value calculation methods, the *bound method* has smaller type I errors and lower powers in both $n = 46$ and $n = 100$ cases since it uses the upper and lower bound of the null probabilities and produce conservative results. The *substitution method* relies on the estimation of the covariance matrix $\Sigma_{n,m}^*$. It may not perform well when the sample size is small. In summary, the simulation results show that the proposed multivariate one-sided tests based on multiple imputations for NLME models with left-censored response and mis-measured covariates are more powerful than the usual two-sided tests based on the naive or the two-step method for testing parameters with natural restrictions since the multivariate one-sided tests incorporate the restrictions in the hypothesis.

Table 3.2: Type I error probabilities and powers with detection limit = 100 at 5% significance level

Sample size	Censoring rate	(λ_1, ξ_1)	$T_W^{*(b)}$	$T_W^{*(s)}$	$T_{CW}^{*(b)}$	$T_{CW}^{*(s)}$	Naive	Two-step
					Type I error in %			
46	13%	(0,0)	2.9	5.2	3.4	5.6	4.0	4.3
100	13%	(0,0)	2.6	4.0	3.1	4.4	3.6	3.8
					Power comparison in %			
46	18%	(0.08,0)	30.3	38.5	31.2	38.6	16.2	18.3
100	18%	(0.08,0)	39.9	45.7	40.0	46.2	20.4	22.2
46	17%	(0,0.02)	43.9	50.0	43.9	51.7	14.1	15.4
100	17%	(0,0.02)	60.8	65.1	64.5	66.2	25.8	27.9
46	23%	(0.05,0.01)	75.4	79.3	75.6	79.6	63.5	68.2
100	23%	(0.05,0.01)	87.8	89.5	88.2	90.0	65.6	70.4

3.6 Discussions

In this chapter, we have proposed the two Wald-type multivariate one-sided tests for testing parameters in NLME models with measurement errors in time-varying covariates and left-censored response. Before we conduct multivariate one-sided hypothesis tests, a multiple imputation method is applied to deal with the left-censored response and mis-measured covariates since both of them can be treated as a special case of missing data. We use both the observed covariate data and the observed response data to help generating the multiple imputations for the true but unobserved covariate values and the left-censored responses. Then we propose the two Wald-type test statistics by either combining the m "complete data" parameter estimates and then construct an overall single test statistic or combing the m sufficient statistics based on each of the m "complete datasets" to obtain an overall test statistic.

We illustrate the proposed tests using a real HIV study dataset and evaluate their performances via a simulation study. The simulation results show that the proposed tests based on multiple imputations perform better than the two-sided tests based on the naive and the two-step method in the sense that they offer higher powers while maintaining nominal type I errors.

4 Multivariate One-sided Hypothesis Tests in NLME Models with Covariate Measurement Errors and Non-ignorable Missing Response

4.1 Introduction

In longitudinal studies, the between-individual variation may be large and can be partially explained by covariates. However, some covariates may be measured with substantial errors and may be measured at different time points from the response measurement schedule, which leads to missing data in the covariates and the missingness in covariates is usually missing at random. Moreover, some individuals may drop out of the study or miss scheduled visits due to drug intolerance and other problems and they may possibly return at a later time, which leads to the intermittent missingness in response. Here the missingness in response is non-ignorable, i.e. the missingness depends on the values of being missing, so the missing data mechanism

needs to be included in the analysis.

In this chapter, we consider the multivariate one-sided hypothesis testing problem of the parameters in NLME models simultaneously with non-ignorable missing response and measurement errors in time-varying covariates. We propose to use a multiple imputation (MI) method to address both non-ignorable response missingness and measurement errors and missingness in time-varying covariates. Based on multiple imputations, we proposed two Wald-type test statistics for the multivariate one-sided hypothesis parameters in NLME models.

The remainder of chapter is organized as follows. In section 4.2, the non-ignorable missing mechanism, the covariate process, and the response process are modeled respectively. Then a multiple imputation method for NLME models with non-ignorable missing response and covariate measurement errors is described in section 4.3. In section 4.4, two approaches are proposed to combine individual sufficient statistics or test statistics based on multiple imputations. In sections 4.5, we illustrate the proposed approaches in a real data example. In section 4.6, a simulation study is conducted to evaluate the performance of the proposed tests. We conclude this chapter in section 4.7 with some discussions.

4.2 NLME Models with Non-ignorable Missing Response and Covariate Measurement Errors

Let y_{ij} be the response value for individual i at time t_{ij}, subject to non-ignorable missingness, $i = 1, 2, \cdots, n,\ \ j = 1, 2, \cdots, m_i$. In the presence of non-ignorable missing response, we denote $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{im_i}) = (\boldsymbol{y}_{mis,i}, \boldsymbol{y}_{obs,i})$ be the collection of the response values for individual i, where $\boldsymbol{y}_{mis,i}$ and $\boldsymbol{y}_{obs,i}$ are the missing and the observed component of $\boldsymbol{y}_i$, respectively. Let $\boldsymbol{r}_i = (r_{i1}, \cdots, r_{im_i})$ be the missing indicator vector such that $r_{ij} = 1$ if y_{ij} is missing and 0 otherwise. Note that $r_{ij} = 1$ does not necessarily imply that $r_{i,j+1} = 1$, i.e. intermittent missing. Let z_{qit} and z^*_{qit} be the observed and the corresponding unobservable "true" value of the qth error-prone covariate for individual i at time u_{it}, $q = 1, \cdots, s, i = 1, \cdots, n, t = 1, \cdots, n_i$. Here the covariate measurement time u_{it} may differ from the response measurement time t_{ij}, that is, missing data may exist in the covariates. Let $\boldsymbol{z}_i = (\boldsymbol{z}_{i1}^T, \cdots, \boldsymbol{z}_{in_i}^T)^T$, where $\boldsymbol{z}_{it} = (z_{1it}, \cdots, z_{sit})^T,\ t = 1, \cdots, n_i$, and T denotes a transpose. Thus, we have the observed data $\mathcal{D}_2 = \{(\boldsymbol{y}_{obs,i}, \boldsymbol{z}_i, \boldsymbol{r}_i), i = 1, 2, \cdots, n\}$.

4.2.1 NLME Response Models with Covariate Measurement Errors

For the response process, we consider the following general NLME model to incorporate possibly mis-measured time-varying covariates

$$\begin{aligned} &y_{ij} = g(t_{ij}, \boldsymbol{\beta}_{ij}) + e_{ij}, \quad e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2), \quad i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, m_i, \\ &\boldsymbol{\beta}_{ij} = \mathbf{d}(\boldsymbol{z}^*_{ij}, \boldsymbol{\beta}, \mathbf{b}_i), \quad \mathbf{b}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, B), \end{aligned} \tag{4.1}$$

where $g(\cdot)$ is a known nonlinear function, $\mathbf{d}(\cdot)$ is a known multivariate linear function, $\boldsymbol{\beta}_{ij}$ is a vector of individual-specific and time-varying parameters and $\boldsymbol{\beta}$ is a vector of fixed-effects, $\boldsymbol{z}^*_{ij}$ is the (unobserved) true covariate value corresponding to the observed and possibly mis-measured $\boldsymbol{z}_{ij}$, $\boldsymbol{e}_i = (e_{i1}, \cdots, e_{im_i})^T$ is a vector of within-individual random errors, $\mathbf{b}_i$ is a vector of random-effects, σ^2 is a variance parameter, and B is the variance-covariance matrix of $\mathbf{b}_i$. We assume $\boldsymbol{e}_i$ and $\boldsymbol{b}_i$ are independent.

In NLME model (4.1), we assume that the individual-specific parameters $\boldsymbol{\beta}_{ij}$ depend on the true (unobservable) covariate $\boldsymbol{z}^*_{ij}$ rather than the observed but possibly contaminated covariate $\boldsymbol{z}_{ij}$.

4.2.2 Models for Time-varying Covariates with Missingness and Measurement Errors

We consider a classical measurement error model to address both missingness and measurement errors in the time-varying covariates. To incorporate possible difference

between the response measurement time t_{ij} and the covariate measurement time u_{it}, we consider the following continuous form which is similar to the model (2.4)

$$\boldsymbol{z}_i(t) = U_i(t)\boldsymbol{\alpha} + V_i(t)\boldsymbol{a}_i + \boldsymbol{\epsilon}_i(t), \quad i = 1, \cdots, n, \tag{4.2}$$

where $\boldsymbol{z}_i(t), U_i(t), V_i(t)$ and $\boldsymbol{\epsilon}_i(t)$ are the covariate value, design matrices, and measurement error at time t, respectively. Then the unobserved "true" covariate values at time t_{ij} is $\boldsymbol{z}^*_{ij} = \boldsymbol{z}_i(t_{ij}) = U_i(t_{ij})\boldsymbol{\alpha} + V_i(t_{ij})\boldsymbol{a}_i$. Note that the missing data in time-varying covariates are ignorable since the missingness is led by the different measurement schedules for the response and the covariates.

4.2.3 Models for Non-ignorable Missingness

To address the non-ignorable missingness in response, we need to model the non-ignorable missing response mechanism, that is, we need to assume a distribution for the missing indicator vector $\boldsymbol{r}_i$. In general, the non-ignorable missingness may depend on the responses values, and the covariate value $\boldsymbol{z}_i$ or the individual random-effects $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$. Little [1995] defined two types of the non-ignorable missing response mechanism:

- outcome-based non-ignorable missing if the distribution for $\boldsymbol{r}_i$ depends on the response $\boldsymbol{y}_i$ and the covariate $\boldsymbol{z}_i$.

- random-effect-based non-ignorable missing if the distribution for $\boldsymbol{r}_i$ depends on the random-effects $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$.

We consider the *outcome-based* missing mechanism in this chapter. Since r_{ij} is binary, we can use a logistic regression model for r_{ij}:

$$\text{logit}\left[\Pr\left(r_{ij}=1|\boldsymbol{y}_i,\boldsymbol{z}_i;\boldsymbol{\eta}\right)\right]=\log\frac{\Pr\left(r_{ij}=1|\boldsymbol{y}_i,\boldsymbol{z}_i;\boldsymbol{\eta}\right)}{1-\Pr\left(r_{ij}=1|\boldsymbol{y}_i,\boldsymbol{z}_i;\boldsymbol{\eta}\right)}=h(\boldsymbol{y}_i,\boldsymbol{z}_i;\boldsymbol{\eta}), \tag{4.3}$$

where $\boldsymbol{\eta}$ are the unknown nuisance parameters, $h(\cdot)$ is usually chosen to be a linear function of $\boldsymbol{y}_i,\boldsymbol{z}_i$. Following Little [1995] and Liu and Wu [2007], we may assume that, for example, r_{ij}'s are independent with $\text{logit}[\Pr(r_{ij}=1|\boldsymbol{y}_i,\boldsymbol{z}_i;\boldsymbol{\eta})]=\eta_1+\eta_2 z_{1ij}+\cdots+\eta_{s+1}z_{sij}+\eta_{s+2}y_{ij}$. Other missing data models can be specified in a similar way. It is important to carry out the sensitivity analysis based on different plausible missing data models since the assumed missing data models are not testable based on the observed data. If the parameter estimates are not sensitive to the assumed missing data models, the statistical inference is reliable.

4.3 A Multiple Imputation Method for NLME Models with Non-ignorable Response Missingness and Covariate Measurement Errors

We consider a MI method for the "missing" values $\boldsymbol{y}_{mis,i}$'s and $\boldsymbol{z}_i^*$'s based on the model (4.1), model (4.2) and model (4.3). Under the Bayesian framework, we can generate m proper multiple imputations for $\{(\boldsymbol{y}_{mis,i}, \boldsymbol{z}_i^*), i = 1, \cdots, n\}$ to obtain m complete datasets to which the existing hypothesis testing approaches are applicable. Specifically, these multiple imputations for $(\boldsymbol{y}_{mis,i}, \boldsymbol{z}_i^*)$ can be generated from the assumed predictive distribution $f(\boldsymbol{y}_{mis,i}, \boldsymbol{z}_i^* | \mathcal{D}_2)$, $i = 1, \cdots, n$.

Let $\Psi = (\boldsymbol{\alpha}, \Phi) \equiv (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \sigma, D, A, B)$, $\boldsymbol{a} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$ and $\boldsymbol{b} = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n)$ be the collection of all unknown parameters and random-effects in models (4.1), (4.2) and (4.3) respectively, where $\Phi = (\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma, D, A, B)$. Let $\pi(\cdot)$ denote a prior density function for Ψ, and we assume the parameters are independent, i.e., $\pi(\Psi) = \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\eta})\pi(\sigma)\pi(D)\pi(A)\pi(B)$. Here the prior distribution is assumed as follows.

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \Sigma_{\boldsymbol{\alpha}}), \qquad \boldsymbol{\beta} \sim N(\mathbf{0}, \Sigma_{\boldsymbol{\beta}}), \qquad \boldsymbol{\eta} \sim N(\mathbf{0}, \Sigma_{\boldsymbol{\eta}}),$$

$$\sigma^2 \sim G^{-1}(\gamma, \delta), \qquad D \sim W^{-1}(\Sigma_D, \kappa),$$

$$A \sim W^{-1}(\Sigma_A, \rho), \qquad B \sim W^{-1}(\Sigma_B, \tau),$$

where G^{-1} and W^{-1} respectively denote the Inverse Gamma distribution and the In-

verse Wishart distribution, and the hyper-parameters $\gamma, \delta, \kappa, \rho, \tau$, $\Sigma_{\boldsymbol{\alpha}}, \Sigma_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\eta}}$, Σ_D, Σ_A and Σ_B are known. Note that, in the covariate model (4.2), the unobserved "true" covariate values $\boldsymbol{z}_{ij}^* = U_i(t_{ij})\boldsymbol{\alpha} + V_i(t_{ij})\boldsymbol{a}_i$ is completely determined by its mean parameters $\boldsymbol{\alpha}$ and its random-effects $\boldsymbol{a}_i$ since the design matrices $U_i(t_{ij})$ and $V_i(t_{ij})$ are known. Therefore, to generate multiple imputations for $\boldsymbol{z}_{ij}^*$, it is sufficient to generate multiple imputations for $(\boldsymbol{\alpha}, \boldsymbol{a})$. Thus, we need to generate multiple imputations for $(\boldsymbol{y}_{mis}, \boldsymbol{\alpha}, \boldsymbol{a})$ from the distribution $f(\boldsymbol{y}_{mis}, \boldsymbol{\alpha}, \boldsymbol{a}|\mathcal{D}_2)$ by using the data augmentation method and the Gibbs sampling method. The MI procedure can be accomplished by the following Gibbs sampling through the full conditionals in turn: at iteration t $(t = 1, 2, \cdots)$,

Step I: Draw a value $\boldsymbol{y}_{mis,i}^{(t)}$ of $\boldsymbol{y}_{mis,i}$ from

$$\begin{aligned} & f(\boldsymbol{y}_{mis,i}|\boldsymbol{y}_{obs,i}, \boldsymbol{z}_i, \boldsymbol{r}_i, \boldsymbol{a}_i^{(t-1)}, \boldsymbol{b}_i^{(t-1)}; \Psi^{(t)}) \\ & \quad \propto f(\boldsymbol{y}_i|\boldsymbol{a}_i^{(t-1)}, \boldsymbol{b}_i^{(t-1)}; \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \\ & \quad \times f(\boldsymbol{r}_i|\boldsymbol{y}_i, \boldsymbol{z}_i; \boldsymbol{\eta}^{(t-1)}), \quad i = 1, \cdots, n. \end{aligned}$$

Step II: Draw a value $(\boldsymbol{\alpha}^{(t)}, \boldsymbol{a}^{(t)})$ of $(\boldsymbol{\alpha}, \boldsymbol{a})$ in the following order

i Draw a value of $\boldsymbol{\alpha}^{(t)}$ of the mean parameter $\boldsymbol{\alpha}$ in the covariate model from

$$
\begin{aligned}
&f(\boldsymbol{\alpha}|\boldsymbol{y}_{mis}^{(t)}, \mathcal{D}_2, \boldsymbol{a}^{(t-1)}, \boldsymbol{b}^{(t-1)}; \Phi^{(t-1)}) \\
&\quad \propto \prod_{i=1}^{n} \Big[f(\boldsymbol{y}_{obs,i}|\boldsymbol{a}_i^{(t-1)}, \boldsymbol{b}_i^{(t-1)}; \boldsymbol{\alpha}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \\
&\qquad \times f(\boldsymbol{y}_{mis,}|\boldsymbol{a}_i^{(t-1)}, \boldsymbol{b}_i^{(t-1)}; \boldsymbol{\alpha}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \\
&\qquad \times f(\boldsymbol{z}_i|\boldsymbol{a}_i^{(t-1)}; \boldsymbol{\alpha}, D^{(t-1)}) \Big] \cdot \pi(\boldsymbol{\alpha}).
\end{aligned}
$$

ii Draw a value $\boldsymbol{a}_i^{(t)}$ of the random-effect $\boldsymbol{a}_i$ in the covariate model from

$$
\begin{aligned}
&f(\boldsymbol{a}_i|\boldsymbol{y}_{mis,i}^{(t)}, \mathcal{D}_2, \boldsymbol{\alpha}^{(t)}, \boldsymbol{b}_i^{(t-1)}; \Phi^{(t-1)}) \\
&\quad \propto f(\boldsymbol{y}_{obs,i}|\boldsymbol{a}_i, \boldsymbol{b}_i^{(t-1)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \\
&\qquad \times f(\boldsymbol{y}_{mis,i}|\boldsymbol{a}_i, \boldsymbol{b}_i^{(t-1)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \\
&\qquad \times f(\boldsymbol{z}_i|\boldsymbol{a}_i; \boldsymbol{\alpha}^{(t)}, D^{(t-1)}) f(\boldsymbol{a}_i|A^{(t-1)}), \quad i = 1, \cdots, n.
\end{aligned}
$$

Step III: Draw a value $(\boldsymbol{b}^{(t)}, \Phi^{(t)})$ of $(\boldsymbol{b}, \Phi)$ in the following order

i Draw a value $\boldsymbol{b}_i^{(t)}$ of the random-effect $\boldsymbol{b}_i$ in the response model from

$$
\begin{aligned}
&f(\boldsymbol{b}_i|\boldsymbol{y}_{mis,i}^{(t)}, \mathcal{D}_2, \boldsymbol{a}_i^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}, B^{(t-1)}) \\
&\quad \propto f(\boldsymbol{y}_{obs,i}|\boldsymbol{a}_i^{(t)}, \boldsymbol{b}_i; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \\
&\qquad \times f(\boldsymbol{y}_{mis,i}|\boldsymbol{a}_i^{(t)}, \boldsymbol{b}_i; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t-1)}, \sigma^{2(t-1)}) \\
&\qquad \times f(\boldsymbol{b}_i|B^{(t-1)}), \ i = 1, \cdots, n.
\end{aligned}
$$

ii Draw a value $\boldsymbol{\beta}^{(t)}$ of the mean parameters $\boldsymbol{\beta}$ in the response model from

$$f(\boldsymbol{\beta}|\mathcal{D}_2, \boldsymbol{a}_i^{(t)}, \boldsymbol{b}_i^{(t)}; \boldsymbol{\alpha}^{(t)}, \sigma^{2(t-1)}) \propto \prod_{i=1}^{n} f(\boldsymbol{y}_{obs,i}|\boldsymbol{a}_i^{(t)}, \boldsymbol{b}_i^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}, \sigma^{2(t-1)})$$
$$\times f(\boldsymbol{y}_{mis,i}^{(t)}|\boldsymbol{a}_i^{(t)}, \boldsymbol{b}_i^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}, \sigma^{2(t-1)}) \cdot \pi(\boldsymbol{\beta})$$

iii Draw a value $(\boldsymbol{\eta}^{(t)}, \sigma^{2(t)}, D^{(t)}, A^{(t)}, B^{(t)})$ of the variance-covariance parameters $\boldsymbol{\eta}, \sigma^2, D, A$, and B in the models from the corresponding full conditional distributions in similar ways as discussed in Chapter 2.

In the above proposed multiple imputation method under a Bayesian framework, the starting values are obtained as follows. $(\boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)}, \sigma^{2(0)}, D^{(0)}, A^{(0)}, B^{(0)})$ are calculated from two-step method. $\boldsymbol{\eta}^{(0)}$ is calculated from the non-ignorable missing model with $\boldsymbol{y}_{mis,i}$ equal to the average on $\boldsymbol{y}_{obs,i}$. And $(\boldsymbol{a}_i^{(0)}, \boldsymbol{b}_i^{(0)})$ is set to be $(\mathbf{0}, \mathbf{0})$. Then we iterate the foregoing procedures for a burn-in period until the resulting Markov Chain converges to its stationary distribution. At convergence, we obtain a value of $(\boldsymbol{y}_{mis,i}, \boldsymbol{\alpha}, \boldsymbol{a})$ from $f(\boldsymbol{y}_{mis,i}, \boldsymbol{\alpha}, \boldsymbol{a}|\mathcal{D}_2)$. Repeating the above procedure m times, we obtain m "complete datasets" $\{(\boldsymbol{y}_i^{(l)}, \boldsymbol{z}_i^{*(l)}), i = 1, \cdots, n\}$, $l = 1, 2, \cdots, m$ with $\boldsymbol{y}_i^{(l)} = \left(\boldsymbol{y}_{mis,i}^{(l)}, \boldsymbol{y}_{obs,i}\right)$ and $\boldsymbol{z}_i^{*(l)} = \left(\boldsymbol{z}_{i1}^{(l)}, \cdots, \boldsymbol{z}_{im_i}^{(l)}\right)$. Based on each of the m "complete datasets", we can conduct statistical inference for the NLME model (4.1) with non-ignorable missingness in response and measurement errors in time-varying covariates being addressed simultaneously, and then we combine the m results to obtain an overall conclusion.

In the next section, we consider the multivariate one-sided hypothesis testing for parameters in the NLME model (4.1) and propose the test statistics based on multiply imputed datasets.

4.4 Multivariate One-sided Tests for NLME Models with Multiply Imputed Response and Covariates

In the previous section, we create m imputations for the non-ignorable missing response $\boldsymbol{y}_{mis,i}$ and the unobserved true covariate value $\boldsymbol{z}_i^*$ to obtain m "complete datasets" $\left\{\left(\boldsymbol{y}_i^{(l)}, \boldsymbol{z}_i^{*(l)}\right), i = 1, 2, \cdots, n\right\}$, $l = 1, 2, \cdots, m$. For each of the m imputed datasets, we can conduct the multivariate one-sided hypothesis testing using existing complete-data methods and then we combine the results to obtain an overall conclusion.

We consider the following hypothesis

$$H_0 : R\boldsymbol{\theta} = \mathbf{0} \text{ versus } H_1 : R\boldsymbol{\theta} \geq \mathbf{0}, R\boldsymbol{\theta} \neq \mathbf{0}, \tag{4.4}$$

where $\boldsymbol{\theta} \in \mathbb{R}^k$ is the collection of parameters in the NLME model (4.1), and R is a $r \times k$ ($r \leq k$) full rank matrix of 0's and 1's, indicating the parameters of interest to be tested. For the l-th "complete dataset" $\{\boldsymbol{y}_i^{(l)}, \boldsymbol{z}_i^{*(l)}, i = 1, \cdots, n\}$, $l = 1, \cdots, m$, let $\tilde{\boldsymbol{\theta}}_{*l}$ be the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ under the parameter space

$\{\boldsymbol{\theta} : R\boldsymbol{\theta} \geq \mathbf{0}\}$, $\hat{H}_{n*l}$ be the corresponding observed information matrix $-\frac{1}{n}\nabla^2 l_{obs}(\boldsymbol{\theta})$ evaluated at $\tilde{\boldsymbol{\theta}}$, hence $\hat{H}_{n*l}^{-1}$ is an estimate of variance covariance matrix H_n^{-1}. Then we define

$$\bar{\boldsymbol{\theta}}_m = \frac{1}{m}\sum_{l=1}^{m} \tilde{\boldsymbol{\theta}}_{*l},$$

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

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

where $\bar{\boldsymbol{\theta}}_m$ is an estimator of $\boldsymbol{\theta}$, $\bar{H}_m^{-1}$ estimates the within-imputation variance-covariance matrix H_n^{-1} of $\boldsymbol{\theta}$, and $\bar{B}_m$ estimates the between-imputation variance-covariance B_n of $\tilde{\boldsymbol{\theta}}_{*l}$. When n is large enough, $\bar{\boldsymbol{\theta}}_m \sim N(\boldsymbol{\theta}, \Sigma_{n,m})$ with $\Sigma_{n,m} = H_n^{-1} + (1 + 1/m)B_n$ [Rubin, 1976].

Note that $\bar{B}_m$ may be an inefficient estimator of B_n when m is small since the dimension of $\boldsymbol{\theta}$ may be larger than the number of the multiple imputations. With the equal ratios of missing to observed information assumption for small m [Li et al., 1991, Meng and Rubin, 1992], we have $B_n \propto H_n^{-1}$. This assumption suggests that all eigenvalues λ_i of matrix $H_n B_n$ are equal to $\bar{\lambda} = \frac{1}{k}\sum_{i=1}^{k} \lambda_i$, which leads to $B_n = \bar{\lambda} H_n^{-1}$. Following Johnson et al. [1995], we have $\sum_{i=1}^{k} \lambda_i = tr(H_n B_n)$. Therefore, $\Sigma_{n,m}$ can be estimated by $\widehat{\Sigma}_{n,m} = \bar{H}_m^{-1} + r_m \bar{H}_m^{-1}$, with $r_m = (1 + \frac{1}{m})\frac{1}{k} tr(\bar{H}_m \bar{B}_m)$. In fact, the estimation of B_n based on $\bar{H}_m$ would be asymptotically calibrated and suffers only modest loss of power even though the equal missing assumption is violated when

$m \geq 3$ [Li et al., 1991]. Therefore, for the parameter $R\boldsymbol{\theta}$ of interest, we have

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

where $\Sigma^*_{n,m} = RH_n^{-1}R^T + (1+1/m)RB_nR^T$. Then we utilize $RB_nR^T \propto RH_n^{-1}R^T$ to have an estimate of $\Sigma^*_{n,m}$

$$\widehat{\Sigma}^*_{n,m} = R\bar{H}_m^{-1}R^T + r^*_m R\bar{H}_m^{-1}R^T,$$

where $r^*_m = (1+\frac{1}{m})\frac{1}{r}tr\left(\left(R\bar{H}_m^{-1}R^T\right)^{-1}R\bar{B}_mR^T\right)$.

Now we propose an overall Wald-type test statistic for hypothesis (4.4) based on the combined "sufficient statistic" $\bar{\boldsymbol{\theta}}_m$ as follows

$$T^*_W = n(R\bar{\boldsymbol{\theta}}_m)^T\widehat{\Sigma}^{*-1}_{n,m}(R\bar{\boldsymbol{\theta}}_m). \tag{4.5}$$

We also propose an overall test statistic by combining the individual test statistics based on each of the m "complete datasets" rather than the parameter estimates. Specifically, for each of the m "complete datasets" $\left\{\left(\boldsymbol{y}_i^{(l)}, \boldsymbol{z}_i^{*(l)}\right),\ i = 1, \cdots, n\right\}$, $l = 1, \cdots, m$, the individual Wald-type test statistic for testing hypothesis (4.4) is given by

$$W_{*l} = n(R\tilde{\boldsymbol{\theta}}_{*l})^T\left(R\widehat{H}^{-1}_{n*l}R^T\right)^{-1}(R\tilde{\boldsymbol{\theta}}_{*l}), \qquad l = 1, 2, \cdots, m. \tag{4.6}$$

The simple average $W_m = \sum_{i=1}^m W_{*l}/m$ of the above test statistics in (4.6) may lead to difficulties in computing the null distribution and thus difficulties in making

inference [Wang and Wu, 2011]. Following Wang and Wu [2011], we combine the individual test statistics as follows.

$$T^*_{CW} = \frac{W_m - \frac{(m-1)r}{m+1} r^*_m}{1 + r^*_m}.$$

We derive the asymptotic null distributions for the proposed test statistics T^*_W and T^*_{CW} in the following Theorem 4.4.1 whose proof is similar to those of Theorem 2.3.3 and Theorem 2.3.4.

Theorem 4.4.1 *For the multivariate one-sided hypothesis test $H_0 : R\boldsymbol{\theta} = \mathbf{0}$ versus $H_1 : R\boldsymbol{\theta} \geq \mathbf{0}, R\boldsymbol{\theta} \neq \mathbf{0}$, the following results hold for the proposed test statistics T^*_W and T^*_{CW}.*

(i) *The asymptotic null distribution of T^*_W is estimated by the following $\bar{\chi}^2$ distribution.*

$$\widehat{\Pr}(T^*_W \leq c | R\boldsymbol{\theta} = \mathbf{0}) = \sum_{i=0}^{p} \omega_i(r, \hat{\Sigma}^*_{n,m}, \mathbb{O}^+) \Pr(\chi^2_i \leq c), \qquad \text{for any } \ c \geq 0,$$

where $\mathbb{O}^+ = \{R\boldsymbol{\theta} | R_i^T \boldsymbol{\theta} \geq 0, i = 1, \cdots, r\}$ with R_i being the ith row vector of R, $\omega_i(r, \hat{\Sigma}^*_{n,m}, \mathbb{O}^+)$ is the probability that $\pi_{\hat{\Sigma}^*_{n,m}}(\boldsymbol{x}; \mathbb{O}^+) \in$ the i dimensional face of $\mathbb{O}^+$ with $\sum_{i=0}^{r} \omega_i(r, \hat{\Sigma}^*_{n,m}, \mathbb{O}^+) = 1$, χ^2_i is the chi-square distribution with i degrees of freedom, and χ^2_0 denotes the distribution that takes the value zero with probability one.

(ii) *The above null distribution has thelower and upper bound:*

$$\frac{\Pr(\chi_0^2 \geq c) + \Pr(\chi_1^2 \geq c)}{2} \leq \Pr(T_W^* \geq c | R\boldsymbol{\theta} = \mathbf{0}) \leq \frac{\Pr(\chi_{r-1}^2 \geq c) + \Pr(\chi_r^2 \geq c)}{2}.$$

(iii) *The proposed test statistics T_W^* and T_{CW}^* are asymptotically equivalent*

$$T_{CW}^* = T_W^* + o_p(1).$$

As discussed in Chapter 2 and Chapter 3, we may either use the *substitution method* in Theorem 4.4.1 (i) or the *bound method* in Theorem 4.4.1 (ii) to approximate the null distribution. The *substitution method* would substitute $\Sigma_{n,m}^*$ by its estimate $\hat{\Sigma}_{n,m}^*$, and compute the weights $\omega_i(p, \hat{\Sigma}_{n,m}^*, \mathbb{O}^+)$ by a simulation-based method [Silvapulle and Sen, 2005]. The *bound method* uses the lower and upper bound in Theorem 4.4.1 to compute p-values.

4.5 A Real Data Example

In this section, we illustrate the proposed tests T_W^* and T_{CW}^* to an HIV/AIDS study where the viral load, the CD4 count, and other variables for 48 HIV infected patients were repeatedly measured over a period of 48 weeks after the initiation of the treatment. The number of repeated measurements for each individual varies from 6 to 10. There were 16 patients with missing viral load at scheduled time points and

37 out of 403 viral load observations missing at scheduled time points. There also existed left-censoring in the viral load values due to the detection limit 100 units of the device and these left-censored values were imputed by half of the detection limit for simplicity in this dataset. After the treatment, the viral load of each patient would decay in the initial phase, and the decay rates may reveal the efficacy of the treatment. However, the viral load may continue to decay, fluctuate, or even start to rise(rebound) during the time course, which were likely to be contaminated by long-term clinical factors. Besides, the CD4 count was often measured with substantial errors and measured at time points different from the viral load measurement schedule, which led to missing values in CD4 cell count. This type of missingness may be ignorable (or missing at random) in the sense of Little and Rubin [2002].

Therefore, a bi-exponential NLME model has been suggested to fit the viral load trajectories in this HIV study

$$\begin{aligned} y_{ij} =& \log_{10}(e^{P_{1i}-\lambda_{1ij}t_{ij}} + e^{P_{2i}-\lambda_{2i}t_{ij}}) + e_{ij}, \qquad i = 1, \cdots, n, \quad j = 1, \cdots, m_i, \\ P_{1i} =& P_1 + b_{1i}, \quad \lambda_{1ij} = \lambda_1 + \xi_1 CD4^*_{ij} + b_{2i}, \quad P_{2i} = P_2 + b_{3i}, \quad \lambda_{2i} = \lambda_2 + b_{4i}, \end{aligned} \tag{4.7}$$

where y_{ij} is the log10-transformation of viral load measurements for patient i at time t_{ij}, $CD4_{ij}$ is the CD4 count for patient i at time t_{ij}, P_{1i} and P_{2i} are the baseline viral load of the first and second phase, λ_{1ij} and λ_{2i} are the first (initial) and the second phase of viral decay rates, respectively. $\boldsymbol{\beta} = (P_1, P_2, \lambda_1, \lambda_2, \xi_1)^T$ are the

fixed-effect parameters, $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^T$ are random-effects, and e_{ij} is a within-individual random error. We assume that $\boldsymbol{b}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, B)$, $e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, and $\boldsymbol{b}_i$ is independent of e_{ij}'s.

Moreover, we also consider a LME model to fit the CD4 process in order to address measurement errors and missing data.

$$CD4_{ij} = (\alpha_1 + a_{i1}) + (\alpha_2 + a_{i2})t_{ij} + (\alpha_3 + a_{i3})t_{ij}^2 + \epsilon_{ij} \equiv CD4_{ij}^* + \epsilon_{ij}, \tag{4.8}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T$ are the fixed-effects and $\boldsymbol{a} = (a_1, a_2, a_3)^T$ are the random-effects.

Note that the missing response is usually non-ignorable in such studies since some patients may drop out of the study or miss scheduled visits due to drug-resistance intolerance and other problems [Liu and Wu, 2007]. Subject-area knowledge suggests that the non-ignorable missingness may be related to current or previous viral load and CD4 count measurements. Therefore, we consider the following plausible non-ignorable missing mechanism models [Fitzmaurice et al., 1996],

$$\textit{Model I:} \quad \text{logit}[\Pr(r_{ij} = 1 | \boldsymbol{y}_i, \boldsymbol{z}_i; \boldsymbol{\eta})] = \eta_1 + \eta_2 CD4_{ij} + \eta_3 y_{ij},$$

$$\textit{Model II:} \quad \text{logit}[\Pr(r_{ij} = 1 | \boldsymbol{y}_i, \boldsymbol{z}_i; \boldsymbol{\eta})] = \eta_1 + \eta_2 CD4_{ij},$$

$$\textit{Model III:} \quad \text{logit}[\Pr(r_{ij} = 1 | \boldsymbol{y}_i, \boldsymbol{z}_i; \boldsymbol{\eta})] = \eta_1 + \eta_2 y_{ij}.$$

However, the assumed missing mechanism models are not testable based on the observed data. It is critical to carry out sensitivity analysis based on different missing

mechanism models.

As discussed in Section 4.1, we are interested in testing the multivariate one-sided hypothesis, $H_0 : \boldsymbol{\lambda} = 0$ versus $H_1 : \boldsymbol{\lambda} \geq 0, \boldsymbol{\lambda} \neq 0$, where $\boldsymbol{\lambda} = (\lambda_1, \xi_1)$. Based on these models, we conduct MI method for viral load missing data and mis-measured CD4 count under Bayesian framework with $m = 5$. There are similar parameter estimates based on the three non-ignorable missing data models which indicates that the parameter estimation may be robust under different assumed non-ignorable missing data models. The average of the parameter estimates of $m = 5$ multiple imputations for model (4.7) based on *Model I* are listed in Table 4.1.

Here, we use the following noninformative prior distributions for model parameters,

$$\begin{aligned}
\boldsymbol{\beta} &\sim N(\mathbf{0}, diag(100, 100, 100, 100, 100)), \\
\boldsymbol{\alpha} &\sim N(\mathbf{0}, diag(100, 100, 100)), \qquad \boldsymbol{\eta} \sim N(\mathbf{0}, diag(100, 100)), \\
\sigma^2 &\sim G^{-1}(0.01, 100), \qquad D \sim W^{-1}(1, 1), \\
A &\sim W^{-1}(I, 3), \qquad B \sim W^{-1}(I, 4),
\end{aligned}$$

where the I's are appropriate identity matrices and $W(1, 1)$ (i.e., the chi-square distribution with degree of freedom 1) is a prior Wishart distribution for the scale parameter D. The number of "burn-in" iterations is 2000. After an initial number of 2000 burn-in iterations, every 500th MCMC sample is retained from the next 2500

samples for $m = 5$ independent multiple imputations. The 5 estimates of λ_1 and ξ_1 based on the 5 imputed datasets are (66.93, 1.40), (67.01, 1.50), (66.97, 1.51), (66.94, 1.48), (67.05, 1.49) respectively. Thus, we have $\bar{\boldsymbol{\lambda}}_m = (66.98, 1.48)$. The proposed overall test statistics are calculated as $T_W^* = 540.16$ and $T_{CW}^* = 538.41$. Both the *substitution method* and the *bound method* are used to calculate the p-values and both p-values are close to 0. For comparison purpose, we also consider a two-sided test based on multiple imputation method which doesn't incorporate the constraints, as well as a a naive method which ignores response missingness and covariate measurement errors. Here, the two-sided alternative hypothesis $H_1 : \lambda_1 \neq 0$, or $\xi_1 \neq 0$, which ignores the one-sided constraints, is considered to investigate whether there is gain in power when the constraints are incorporated in hypothesis testing based on the imputed datasets. Moreover, the Wald testing statistics based on the unconstrained two-sided test and the native method are 483.08 and 301.15 respectively, which are substantially smaller than T_W^* and T_{CW}^*, and thus they provide weaker evidence against H_0. In other words, the proposed new tests provide stronger evidence about the efficacy of the treatment and its association with CD4 than existing tests.

Therefore, we have strong evidence to reject H_0 and conclude that the viral load declines significantly during the anti-HIV treatment and the CD4 count has a

significantly positive association with the initial viral decay rate. That is, the higher the CD4 count is, the faster the viral load declines during the initial period of the treatment.

Table 4.1: Parameter estimates of the viral load model (4.7) based on $m = 5$ multiple imputations

Response model						
Fixed effect		Covariance matrix				
P_1	11.52	B				σ
λ_1	66.98	$\begin{pmatrix} 1.13 & 0.63 & 1.49 & -2.30 \\ & 67.19 & 0.77 & -1.02 \\ & & 2.00 & -4.88 \\ & & & 20.86 \end{pmatrix}$				0.36
ξ_1	1.48					
P_2	6.92					
λ_2	-2.63					

4.6 Simulation Study

In this section, we conduct a simulation study to evaluate the performance of the proposed test procedures for the multivariate one-sided hypothesis testing and compare them with the counterpart two-sided tests based on multiple imputation method and the naive method respectively. Both the *substitution method* and the

bound method are used to calculate the approximate p-values. All the simulation settings are repeated for 1000 times. We evaluate the performance of the tests in terms of both the type I error probability and the test power. The significance level $\alpha = 0.05$ is set.

In the simulation study, we use the NLME response model (4.7), the covariate model (4.8), and the missing mechanism *model I*, with different choices of the true parameters of λ_1 and ξ_1 values in order to evaluate the tests under different settings. The true values of the parameters λ_1 and ξ_1 are listed in Table 4.2 – 4.4. The true values of the other fixed-effect parameters are $(P_1, P_2, \lambda_2) = (10, 5, -3)$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2) = (-0.5, 4, -4)$, $Diag(A) = (0.5, 2, 1)$, $D = 0.5$, $Diag(B) = (1, 10, 2, 5)$, $\sigma = 0.5$, where the covariance matrices A and B are set to be diagonal matrices. We consider two different sample sizes $n = 50$ and $n = 100$ and three different average missing rate 10%, 20% and 40%. The number of response measurement times m_i and covariate measurement times n_i are both set to be 8.

The simulation results are reported in Table 4.2 – 4.4. We can see that the proposed multivariate one-sided Wald-type test T_W^* and combined Wald-type test T_{CW}^* using both the *substitution* and the *bound method* consistently provide the lower type I error rates which are close to the nominal level 5%, while the two-sided test Wald test gives the higher type I error rates, especially for the scenarios with the smaller

sample size and the higher response missing rates. The naive method also approximately attains the nominal significant level $\alpha = 5\%$ in most of the cases since it may underestimate the parameters. Note that the type I errors of the proposed tests for the smaller sampler size ($n = 50$) based on the *substitution method* are slightly higher than the nominal level. This may be due to the fact that the *substitution method* may not estimate the covariance matrix $\Sigma^*_{n,m}$ well under the small sample size ($n = 50$), and this problem disappears for large sample sizes ($n = 100$).

As for the power of these tests, we can see that the proposed tests T^*_W and T^*_{CW} are more powerful in all the scenarios than the two-sided counterparts. These simulation results clearly show the better performance of the proposed testing approaches than the two-sided counterparts, indicating the gain of power when the constraints are incorporated in the hypothesis tests. Note that the *bound method* is more conservative than the *substitution method* because it uses the upper and the lower bound of the null distributions to compute the p-values, i.e., the *substitution method* produces higher powers in both $n = 50$ and $n = 100$ cases. However, the *substitution method* relies on good estimation of the covariance matrix $\Sigma^*_{n,m}$, so it might not perform well when the sample size is small.

Table 4.2: Type I error probabilities and powers with 10% response missing rate.

n	(λ_1, ξ_1)	$T_W^{*(b)}$ †	$T_{CW}^{*(b)}$†	$T_W^{*(s)}$†	$T_{CW}^{*(s)}$ †	$T_W^{(u)}$ †	Naive
		Type I error probability in %					
50	(0,0)	3.1	4.0	5.3	6.1	7.2	4.5
100	(0,0)	2.2	2.7	4.2	4.4	6.0	3.9
		Power in %					
50	(0,0.02)	22.3	24.6	23.4	24.5	20.4	13.5
100	(0,0.02)	45.3	46.1	46.8	47.1	40.6	28.1
50	(0.02,0)	29.2	30.4	32.3	33.5	25.8	18.3
100	(0.02,0)	50.4	53.3	51.6	52.7	45.8	38.3
50	(0.05,0.01)	64.1	64.7	66.6	67.2	60.3	48.5
100	(0.05,0.01)	88.5	88.7	89.1	90.6	84.3	66.2

† $T^{*(b)}$ and $T^{*(s)}$ denote the constrained Wald-type test statistics based on the *bound* and the *substitution method*, respectively. $T_W^{(u)}$ denotes the unconstrained Wald test statistic.

Table 4.3: Type I error probabilities and powers with 20% response missing rate.

n	(λ_1, ξ_1)	$T_W^{*(b)}$	$T_{CW}^{*(b)}$	$T_W^{*(s)}$	$T_{CW}^{*(s)}$	$T_W^{(u)}$	Naive
		Type I error probability					
50	(0,0)	3.5	3.4	4.8	5.1	7.8	4.2
100	(0,0)	2.5	2.2	3.1	3.3	6.5	3.0
		Power					
50	(0,0.02)	22.0	23.4	24.3	24.5	19.5	11.3
100	(0,0.02)	44.6	45.3	45.6	46.7	40.5	23.8
50	(0.02, 0)	29.0	30.4	33.1	33.5	25.1	13.8
100	(0.02,0)	49.5	52.7	55.6	56.9	45.3	28.0
50	(0.05,0.01)	64.0	65.3	66.5	66.9	61.4	48.1
100	(0.05,0.01)	85.1	85.3	86.5	86.9	83.5	60.4

Table 4.4: Type I error probabilities and powers with 40% response missing rate.

n	(λ_1, ξ_1)	$T_W^{*(b)}$	$T_{CW}^{*(b)}$	$T_W^{*(s)}$	$T_{CW}^{*(s)}$	$T_W^{(u)}$	Naive
		Type I error probability					
50	(0,0)	3.9	4.0	6.7	6.3	8.9	6.5
100	(0,0)	3.1	3.0	4.6	4.7	7.4	5.0
		Power					
50	(0,0.02)	19.8	20.3	24.3	24.5	18.1	11.0
100	(0,0.02)	42.4	43.	44.5	45.6	39.3	20.8
50	(0.02,0)	27.1	28.4	29.5	31.0	24.9	14.7
100	(0.02,0)	46.4	47.1	49.2	50.6	42.7	37.6
50	(0.05,0.01)	63.7	64.2	65.6	66.2	60.3	58.2
100	(0.05,0.01)	82.9	83.3	84.9	86.0	78.7	59.5

4.7 Discussions

In this chapter, we have proposed a multiple imputation method to address both non-ignorable missing response and covariate measurement errors in NLME models. Based on multiple imputation for missing response and covariate measurement error

and missing data, we have proposed two Wald-type test statistics for the multivariate one-sided hypothesis testing, and we have derived their null distributions. Simulation results show that the proposed multivariate one-sided tests perform better than the two-sided tests in the sense that they offer higher powers while maintaining nominal type I errors.

We assume the non-ignorable missing data model depends on the observed response or unobserved covariates, which is called outcome-based missing mechanism. However, the missing mechanism could also be related to the random-effects in response and covariate process. The methods in this chapter may be extended to the random-effect based missing mechanism models. And the sensitivity analysis is needed based on both outcome-based missing mechanism models and random-effects-based missing mechanism models. Besides, the proposed methods may also be extended to generalized linear mixed-effects models or semi-parametric mixed-effects models. The proposed methods can also be used in the case where there are measurement errors, censoring and missing are all present simultaneously.

5 Conclusions and Future Work

5.1 Conclusions

In this thesis, we have proposed testing statistics for the multivariate one-sided testing problems in NLME models with: (i) mis-measured time-varying covariates, (ii) both mis-measured time-varying covariates and left-censored response, and (iii) both mis-measured time-varying covariates and non-ignorable missing response. Multiple imputation methods are used throughout this thesis to address the mentioned data complications. For each project, we have developed the two Wald-type tests for the multivariate one-sided hypothesis on the parameters of NLME models based on multiple imputations for incomplete data. Specifically, we generate the multiple "complete-data" sets for NLME models, then the proposed test statistics can be obtained by either combining the m "complete data" parameter estimates and then constructing an overall single test statistic or combing the m sufficient statistics based on each of the m "complete datasets" to obtain an overall test statistic.

For the both test statistics, T_W^* and T_{CW}^*, we have derived their asymptotic null distributions. For the calculation of p-values, we have proposed two methods: (i) the *substitution method* which substitutes the unknown parameters in the null distributions with their estimates; (ii) the *bound method* which uses the upper and the lower bound of the null probabilities as conservative p-values. Simulation results have shown that in terms of the test powers, the two proposed one-sided tests based on MI method perform better than the commonly used two-sided tests based on the naive and the two-step method. In particular, the *substitution method* performs well in most cases except when the sample size is small, it may produces liberal type I errors. This exception may be due to the fact that the *substitution method* may not estimate the covariance matrix $\Sigma_{n,m}^*$ well under the small sample size, and this problem disappears for large sample size. The *bound method* also performs well in all cases, although it may be conservative to some extent since it uses the upper and the lower bound of the null probabilities to calculate the p-values.

The scientific contribution of this thesis is that, by jointly addressing covariate measurement errors and censoring/non-ignorable missing response for NLME models, more reliable one-sided hypothesis tests based on multiple imputation method are proposed, which fill the research gap.

5.2 Future Work

We discuss possible future work relevant to this thesis as follows.

1 In Chapter 3 and Chapter 4, we addressed the covariate measurement errors with the left-censoring response as well as covariate measurement errors and non-ignorable missing response in NLME models, respectively. However, in practice, there may exist covariate measurement errors, left-censoring, and non-ignorable missing response simultaneously in a study. A new project can be conducted to develop statistical methodologies to address these three issues simultaneously.

2 In Chapter 4, we have proposed the test statistics for multivariate one-sided hypothesis testing for the NLME models with both covariate measurement errors and outcome-based missing mechanism. Another plausible missing data model is random-effect-based missing mechanism which assumes the missingness is related to the unobserved individual-related information. In the future, we may consider the random-effect-based non-ignorable missing mechanism, and both the outcome-based missing mechanism models and the random-effect-based missing mechanism models will be used for the sensitivity analysis.

3 In this thesis, the multivariate one-sided test is based on Wald-type statistics. An

intuitive extension is that we may develop Likelihood Ratio type statistics and Score type statistics. Theories and performances for the testing approaches based on Likelihood Ratio type and Score type statistics can be investigated and reported in the future.

4 The methods proposed in this thesis may be extended to other mixed-effects models, such as generalized linear mixed-effects models, and semi-parametric or non-parametric mixed-effects models. The extensions are conceptually straightforward, but the technical details may be complicated in some cases and the performances of the multivariate one-sided tests under different models need to be evaluated separately.

A Appendix

A.1 Some results for Section 1.1

Recall that, in Section (1.1), the Wald test statistic for testing the one parameter two-sided hypothesis $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$ is given by

$$T_W = \frac{\bar{X}^2}{\widehat{\text{Var}}(\bar{X})} = \frac{n^2 \bar{X}^2}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}$$

where $X_i \overset{i.i.d.}{\sim} N(\mu, \sigma^2), i = 1, \ldots, n$, and $\bar{X} = \sum_{i=1}^{n} X_i/n$ is the sample mean. Let $\xrightarrow{d}$ and $\xrightarrow{p}$ denote convergence in distribution and in probability as $n \to n$ respectively. We will show that $T_W \xrightarrow{d} \chi_1^2$ holds under H_0.

Proof. Given $\mu = 0$, by the central limit theorem (CLT), we have

$$\frac{\sqrt{n}\bar{X}}{\sigma} \xrightarrow{d} N(0, 1),$$

which is equivalent to

$$\frac{n\bar{X}^2}{\sigma^2} \xrightarrow{d} \chi_1^2.$$

By the law of large numbers, we have

$$\frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{n} \xrightarrow{p} \sigma^2,$$

$$\bar{X} \xrightarrow{p} \mu.$$

Therefore, we have the following results from the Slutsky's theorem,

$$\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{n} - \left(\bar{X}-\mu\right)^2 \xrightarrow{p} \sigma^2,$$

$$\frac{\sigma^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2/n} \xrightarrow{p} 1,$$

Therefore, we have

$$T_W = \frac{n^2\bar{X}^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} = \frac{n\bar{X}^2}{\sigma^2} \times \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2/n} \xrightarrow{d} \chi_1^2 \times 1 = \chi_1^2.$$

■

The Wald test for testing the one one-sided hypothesis $H_0 : \mu = 0$ versus $H_1 : \mu > 0$ is given by

$$T_W = \begin{cases} 0, & \text{if } \bar{X} \le 0, \\ \dfrac{n^2\bar{X}^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}, & \text{if } \bar{X} > 0. \end{cases}$$

We will show that as $n \to \infty$, $\Pr(T_W \le c) \to \frac{1}{2} + \frac{1}{2}F_{\chi^2}(c;1)$ under H_0.

Proof. First, when $\mu = 0$, we have

$$\begin{aligned}\Pr(T_W \le c) &= \Pr(T_W \le c|\bar{X} \le 0)\Pr(|\bar{X} \le 0) + \Pr(T_W \le c|\bar{X} > 0)\Pr(|\bar{X} > 0) \\ &= \frac{1}{2} + \frac{1}{2}\Pr\left[\frac{n^2\bar{X}^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} \le c\right] \to \frac{1}{2} + \frac{1}{2}F_{\chi^2}(c;1)\end{aligned}$$

since that $\dfrac{n^2\bar{X}^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} \xrightarrow{d} \chi_1^2$ shown in the previous proof. ■

A.2 Proofs of Theorems

A.2.1 Proof of Theorem 2.3.1

Proof. We begin to show the $\sqrt{n}$-consistency of $\tilde{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$. Let A denote a subset of Ω containing $\boldsymbol{\theta}_0$ which is the true value of $\boldsymbol{\theta}$. Specifically, $A = \{\boldsymbol{\theta} : R\boldsymbol{\theta} \geq \mathbf{0}, R\boldsymbol{\theta} \neq \mathbf{0}\}$ when H_0 does not hold and $A = \{\boldsymbol{\theta} : R\boldsymbol{\theta} = \mathbf{0}\}$ when H_0 holds. Let $\hat{\boldsymbol{\theta}}_A = \arg\max_{\boldsymbol{\theta} \in A} l_{obs}(\boldsymbol{\theta})$. For any sufficiently small $\epsilon > 0$, the intersection of the closure of a ϵ-neighborhood of $\boldsymbol{\theta}_0$ and the closure of A is closed. The local maximum of the continuous function $l_{obs}(\boldsymbol{\theta})$ would be attained in this intersection set. By **C7**, $\hat{\boldsymbol{\theta}}_A$ is the local maximum. We first show that $\forall \epsilon > 0$, $\exists \delta > 0$, when n is greater than some N_δ depending on δ, for any $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \epsilon$,

$$\Pr\left[l_{obs}(\boldsymbol{\theta}) - l_{obs}(\boldsymbol{\theta}_0) < 0\right] > 1 - \delta. \tag{A.1}$$

Then the consistency of $\hat{\boldsymbol{\theta}}_A$ holds by definition.

To show (A.1), we consider the following Taylor expansion of $\frac{1}{n} l_{obs}(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

$$\begin{aligned} \frac{1}{n} l_{obs}(\boldsymbol{\theta}) - \frac{1}{n} l_{obs}(\boldsymbol{\theta}_0) =& (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \frac{1}{n} \nabla l_{obs}(\boldsymbol{\theta}_0) \\ &+ \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \frac{1}{n} \nabla^2 l_{obs}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ &+ O_p(1)\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^3. \end{aligned} \tag{A.2}$$

Note that the first term can be arbitrarily small for all large n since $\frac{1}{n} \nabla l_{obs}(\boldsymbol{\theta}_0) =$

$\frac{1}{\sqrt{n}} \times \frac{1}{\sqrt{n}} \nabla l_{obs}(\boldsymbol{\theta}_0) \xrightarrow{p} \mathbf{0}$ by (2.7), and the third term is smaller in norm than the second term for all sufficiently small ϵ. Therefore, the sign of the right side of (A.2) be the determined by the second term for all large n. Note that the second term has a negative limit (in probability) $-(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T H_{\boldsymbol{\theta}_0} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ by (2.6). Then we have $\frac{1}{n} l_{obs}(\boldsymbol{\theta}) - \frac{1}{n} l_{obs}(\boldsymbol{\theta}_0) < 0$ as $n \to \infty$, and (A.1) is justified accordingly.

Furthermore, under **C7**, we have $l(\boldsymbol{\theta}_0) = \frac{1}{n} l_{obs}(\boldsymbol{\theta}_0) + o_p(1)$ and thus,

$$\begin{aligned} l(\boldsymbol{\theta_0}) - l(\hat{\boldsymbol{\theta}}_A) &= \frac{1}{n} l_{obs}(\boldsymbol{\theta_0}) - l(\hat{\boldsymbol{\theta}}_A) + o_p(1) \\ &\leq \frac{1}{n} l_{obs}(\hat{\boldsymbol{\theta}}_A) - l(\hat{\boldsymbol{\theta}}_A) + o_p(1) \\ &\leq \sup_{\boldsymbol{\theta} \in \Omega} \left| \frac{1}{n} l_{obs}(\boldsymbol{\theta}) - l(\boldsymbol{\theta}) \right| + o_p(1) \xrightarrow{p} 0. \end{aligned} \tag{A.3}$$

For any $\epsilon > 0$, we have $l(\boldsymbol{\theta_0}) - \sup_{\{\boldsymbol{\theta}: \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \epsilon\}} l(\boldsymbol{\theta}) > 0$ by **C7** and $\exists 0 < \eta_\epsilon < l(\boldsymbol{\theta_0}) - \sup_{\{\boldsymbol{\theta}: \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \epsilon\}} l(\boldsymbol{\theta})$, then $l(\boldsymbol{\theta}) < l(\boldsymbol{\theta}_0) - \eta_\epsilon$ as long as $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \epsilon$. That is, the event $\left\{ \left\| \hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0 \right\| \geq \epsilon \right\}$ is contained in the event $\left\{ l(\hat{\boldsymbol{\theta}}_A) < l(\boldsymbol{\theta}_0) - \eta_\epsilon \right\}$ whose probability converges to 0 as $n \to \infty$ by (A.3). Then we have $\hat{\boldsymbol{\theta}}_A \xrightarrow{p} \boldsymbol{\theta}_0$ by definition.

Next we show the $\sqrt{n}$-*consistency* of $\hat{\boldsymbol{\theta}}_A$ in a similar way as the proof of Lemma 1 in Chernoff [1954]. $\forall \epsilon > 0$, $\exists$ a positive sequence $\delta_{n\epsilon} \to 0$ and a positive K_ϵ depending

on ϵ such that

$$\Pr\left[\left\|\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0\right\| < \delta_{n\epsilon}\right] > 1-\epsilon \text{ (by the consistency of } \hat{\boldsymbol{\theta}}_A),$$
$$\Pr\left[\left\|\frac{1}{\sqrt{n}}\nabla l_{obs}(\boldsymbol{\theta}_0)\right\| < K_\epsilon\right] > 1-\epsilon \text{ (by (2.6))},$$
$$\Pr\left[\left\|\frac{1}{n}\nabla^2 l_{obs}(\boldsymbol{\theta}_0) + H_{\boldsymbol{\theta}_0}\right\| < \delta_{n\epsilon}\right] > 1-\epsilon \text{ (by (2.7))}.$$

and $O_p(1)\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^3 < K_\epsilon\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^3$. With $l_{obs}(\hat{\boldsymbol{\theta}}_A) \geq l_{obs}(\boldsymbol{\theta}_0)$,

$$\begin{aligned}
0 \leq & \frac{1}{n}l_{obs}(\hat{\boldsymbol{\theta}}_A) - \frac{1}{n}l_{obs}(\boldsymbol{\theta}_0) \\
&= (\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0)^T\frac{1}{n}\nabla l_{obs}(\boldsymbol{\theta}_0) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0)^T\frac{1}{n}\nabla^2 l_{obs}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0) + O_p(1)\left\|\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0\right\|^3 \\
&= T_1 + T_2 + T_3.
\end{aligned} \tag{A.4}$$

With the first term $T_1 = (\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0)^T\frac{1}{n}\nabla l_{obs}(\boldsymbol{\theta}_0)$, the second term $T_2 = \frac{1}{2}(\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0)^T\frac{1}{n}\nabla^2 l_{obs}$, and the third term $T_3 = O_p(1)\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^3$. Based on the above inequalities, there exist K_ϵ^* such that

$$T_1+T_2+T_3 < -\frac{1}{2}(\hat{\boldsymbol{\theta}}_A-\boldsymbol{\theta}_0)^T H_{\boldsymbol{\theta}_0}(\hat{\boldsymbol{\theta}}_A-\boldsymbol{\theta}_0) + K_\epsilon^* \cdot \left(\frac{\left\|\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0\right\|}{\sqrt{n}} + \delta_{n\epsilon}\cdot\left\|\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0\right\|^2\right). \tag{A.5}$$

As $n \to \infty$ (A.5) holds implies that there exists a $K_\epsilon^{**} > 0$ such that $\left\|\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0\right\| < \frac{K_\epsilon^{**}}{\sqrt{n}}$ occurs with probability $1-\epsilon$ for all sufficiently large n. Then the $\sqrt{n}$-consistency of $\hat{\boldsymbol{\theta}}_A$ holds by definition. Therefore, we have $\tilde{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$ is $\sqrt{n}$-consistent.

Finally, we derive the asymptotic null distribution of the Wald statistic as follows. We consider the following quadratic approximation of $l_{obs}(\boldsymbol{\theta})$ [Silvapulle and Sen,

2005]

$$l_{obs}(\boldsymbol{\theta}) = l_{obs}(\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\boldsymbol{Z}_n - \boldsymbol{u})^T H_{\boldsymbol{\theta}_0}(\boldsymbol{Z}_n - \boldsymbol{u}) + r_n(\boldsymbol{u}),$$

where $\boldsymbol{Z}_n = \sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$, $\hat{\boldsymbol{\theta}}$ is $\sqrt{n}$-consistent, $\boldsymbol{u} = \sqrt{n}\,(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ and the reminder $r_n(\boldsymbol{u})$ satisfies $\sup_{\|\boldsymbol{u}\|<K} |r_n(\boldsymbol{u})| = o_p(1)$ for $K > 0$. Moreover, we have

$$\inf_{\boldsymbol{\theta}=\mathbf{0}} (\boldsymbol{Z}_n - \boldsymbol{u})^T H_{\boldsymbol{\theta}_0}(\boldsymbol{Z}_n - \boldsymbol{u}) = \min\left\{ \|\boldsymbol{Z}_n - \boldsymbol{\mu}\|^2_{H_{\boldsymbol{\theta}_0}} : \boldsymbol{\theta} = \mathbf{0} \right\}$$

and

$$\inf_{\boldsymbol{\theta}\geq\mathbf{0}} (\boldsymbol{Z}_n - \boldsymbol{u})^T H_{\boldsymbol{\theta}_0}(\boldsymbol{Z}_n - \boldsymbol{u}) = \min\left\{ \|\boldsymbol{Z}_n - \boldsymbol{\mu}\|^2_{H_{\boldsymbol{\theta}_0}} : \boldsymbol{\theta} \geq \mathbf{0} \right\} \quad ,$$

Also we have the first-order Taylor expansion of $\nabla l_{obs}(\hat{\boldsymbol{\theta}})$ at $\boldsymbol{\theta}_0$

$$\nabla l_{obs}(\hat{\boldsymbol{\theta}}) = \nabla l_{obs}(\boldsymbol{\theta}_0) + \nabla^2 l_{obs}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(\frac{1}{\sqrt{n}})$$

since $\hat{\boldsymbol{\theta}}$ is $\sqrt{n}$-consistent. With $\nabla l_{obs}(\hat{\boldsymbol{\theta}}) = 0$, we have

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = \left(-\frac{1}{n}\nabla^2 l_{obs}(\boldsymbol{\theta}_0)\right)^{-1} \frac{1}{\sqrt{n}} \nabla l_{obs}(\boldsymbol{\theta}_0) + o_p(1). \tag{A.6}$$

By the Slutsky theorem, we have

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, H^{-1}_{\boldsymbol{\theta}_0}\right).$$

As given in Silvapulle and Sen [2005], with $\boldsymbol{Z}_n \xrightarrow{d} N\left(\mathbf{0}, H^{-1}_{\boldsymbol{\theta}_0}\right)$, the asymptotic null distribution of $D(\boldsymbol{Z}_n) = \inf_{\boldsymbol{\theta}=\mathbf{0}} (\boldsymbol{Z}_n - \boldsymbol{u})^T H_{\boldsymbol{\theta}_0}(\boldsymbol{Z}_n - \boldsymbol{u}) - \inf_{\boldsymbol{\theta}\geq\mathbf{0}} (\boldsymbol{Z}_n - \boldsymbol{u})^T H_{\boldsymbol{\theta}_0}(\boldsymbol{Z}_n - \boldsymbol{u})$

is the χ^2-bar distribution (denoted by $\bar{\chi}^2$) which is given by $\lim_{n\to\infty} \Pr(\boldsymbol{Z}_n \le c) = \sum_{i=0}^{k} \omega_i(k, H_{\boldsymbol{\theta}_0}^{-1}, \mathbb{O}^+) \Pr(\chi_i^2 \le c)$ for any $c \ge 0$.

Let $\boldsymbol{Z}_{nR} = \sqrt{n}\left(R\hat{\boldsymbol{\theta}} - R\boldsymbol{\theta}_0\right)$, $\boldsymbol{u}_R = \sqrt{n}\left(R\boldsymbol{\theta} - R\boldsymbol{\theta}_0\right)$, we have $\sqrt{n}\left(R\hat{\boldsymbol{\theta}} - R\boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, RH_{\boldsymbol{\theta}_0}^{-1}R^T\right)$. Note that under $H_0 : R\boldsymbol{\theta} = \mathbf{0}$, we also have $R\bar{\boldsymbol{\theta}} = \mathbf{0}$.

$$
\begin{aligned}
D(\boldsymbol{Z}_{nR}) =& \inf_{R\boldsymbol{\theta}=\mathbf{0}} (\boldsymbol{Z}_{nR} - \boldsymbol{u}_R)^T \left(RH_{\boldsymbol{\theta}_0}^{-1}R^T\right)^{-1} (\boldsymbol{Z}_{nR} - \boldsymbol{u}_R) \\
& - \inf_{R\boldsymbol{\theta}\ge\mathbf{0}} (\boldsymbol{Z}_{nR} - \boldsymbol{u}_R)^T \left(RH_{\boldsymbol{\theta}_0}^{-1}R^T\right)^{-1} (\boldsymbol{Z}_{nR} - \boldsymbol{u}_R) \\
=& \inf_{R\boldsymbol{\theta}=\mathbf{0}} \left(\sqrt{n}R\hat{\boldsymbol{\theta}} - \sqrt{n}R\boldsymbol{\theta}\right)^T \left(RH_{\boldsymbol{\theta}_0}^{-1}R^T\right)^{-1} \left(\sqrt{n}R\hat{\boldsymbol{\theta}} - \sqrt{n}R\boldsymbol{\theta}\right) \\
& - \inf_{R\boldsymbol{\theta}\ge\mathbf{0}} \left(\sqrt{n}R\hat{\boldsymbol{\theta}} - \sqrt{n}R\boldsymbol{\theta}\right)^T \left(RH_{\boldsymbol{\theta}_0}^{-1}R^T\right)^{-1} \left(\sqrt{n}R\hat{\boldsymbol{\theta}} - \sqrt{n}R\boldsymbol{\theta}\right) + o_p(1) \\
=& \min\left\{ \left\| \sqrt{n}R\hat{\boldsymbol{\theta}} - \sqrt{n}R\boldsymbol{\theta} \right\|_{RH_{\boldsymbol{\theta}_0}R^T} : R\boldsymbol{\theta} = \mathbf{0} \right\} - \min\left\{ \left\| \sqrt{n}R\hat{\boldsymbol{\theta}} - \sqrt{n}R\boldsymbol{\theta} \right\|_{RH_{\boldsymbol{\theta}_0}R^T} : R\boldsymbol{\theta} \ge \mathbf{0} \right\} \\
=& \left\| \sqrt{n}R\hat{\boldsymbol{\theta}} - \sqrt{n}R\boldsymbol{\theta} \right\|_{RH_{\boldsymbol{\theta}_0}R^T} - \min\left\{ \left\| \sqrt{n}R\hat{\boldsymbol{\theta}} - \sqrt{n}R\boldsymbol{\theta} \right\|_{RH_{\boldsymbol{\theta}_0}R^T} : R\boldsymbol{\theta} \ge \mathbf{0} \right\} \\
=& \left\| \sqrt{n}R\tilde{\boldsymbol{\theta}} \right\|_{RH_{\boldsymbol{\theta}_0}R^T} \\
=& n\left(R\tilde{\boldsymbol{\theta}}\right)^T \left(R\hat{H}_{\boldsymbol{\theta}}^{-1}R^T\right)^{-1} \left(R\tilde{\boldsymbol{\theta}}\right) + o_p(1) \\
=& T_W + o_p(1)
\end{aligned}
$$

Note that $D(\boldsymbol{Z}_{nR}) \xrightarrow{d} \bar{\chi}^2$, which is given as $\sum_{i=0}^{r} \omega_i(r, RH_{\boldsymbol{\theta}_0}^{-1}R^T, \mathbb{O}^+) F_{\chi^2}(c; i)$ [Silvapulle and Sen, 2005], so is the T_W. ■

A.2.2 Proof of Theorem 2.3.2

Proof. It is straightforward to estimate the asymptotic null distribution of T_W by $\widehat{\Pr}(T_W \leq c|R\boldsymbol{\theta} = \mathbf{0}) = \sum_{i=0}^{r} \omega_i(r, R\hat{H}_{\boldsymbol{\theta}} R^T, \mathbb{O}^+) \Pr(\chi_i^2 \leq c)$, for any $c \geq 0$, since the observed information matrix $-\frac{1}{n}\nabla^2 l_{obs}(\tilde{\boldsymbol{\theta}})$ is a consistent estimator of $H_{\boldsymbol{\theta}_0}$.

Note that $\Pr(\chi_i^2 \geq c) \leq \Pr(\chi_j^2 \geq c)$ for $i < j$, $\sum_{i=0}^{r} \omega_i(r, H_{\boldsymbol{\theta}_0}, \mathbb{O}^+) = 1$ and $0 \leq \omega_i(r, H_{\boldsymbol{\theta}_0}, \mathbb{O}^+) \leq \frac{1}{2}$ for any positive definite matrix $H_{\boldsymbol{\theta}_0}$ [Silvapulle and Sen, 2005].

We have

$$\begin{aligned}\sum_{i=0}^{r} \omega_i(r, R\hat{H}_{\tilde{\boldsymbol{\theta}}}^{-1} R^T, \mathbb{O}^+) \Pr(\chi_i \geq c) &\geq \sum_{i=1}^{r} \omega_i \Pr(\chi_1^2 \geq c) + \omega_0 \Pr(\chi_0^2 \geq c) \\ &= (1 - \omega_0) \Pr(\chi_1^2 \geq c) + \omega_0 \Pr(\chi_0^2 \geq c) \\ &\geq \frac{\Pr(\chi_1^2 \geq c) + \Pr(\chi_0^2 \geq c)}{2}\end{aligned}$$

Similarly, we have

$$\sum_{i=0}^{r} \omega_i(r, R\hat{H}_{\tilde{\boldsymbol{\theta}}}^{-1} R^T, \mathbb{O}^+) \Pr(\chi_i^2 \geq c) \leq \frac{\Pr(\chi_{r-1}^2 \geq c) + \Pr(\chi_r^2 \geq c)}{2}.$$

The proof is completed. ■

A.2.3 Proof of Theorem 2.3.3

Proof. (i)The $\sqrt{n}$-consistency of $\bar{\boldsymbol{\theta}}_m$ can be shown in a similar way as in the proof of Theorem 2.3.1. Moreover, we have that $R\bar{\boldsymbol{\theta}}_m \sim N(R\boldsymbol{\theta}, \Sigma_{n,m}^*)$, where $\Sigma_{n,m}^* =$

$RH_nR^T+(1+1/m)RH_nR^T$. By substituting $\hat{\Sigma}^*_{n,m}$ for $\Sigma^*_{n,m}$, we have the approximate distribution $R\bar{\boldsymbol{\theta}}_m \sim N(R\boldsymbol{\theta}, \hat{\Sigma}^*_{n,m})$. Therefore, the asymptotic null distribution of T^*_W can be estimated by

$$\widehat{\Pr}(T^*_W \le c) = \sum_{i=0}^{r} \omega_i(r, \hat{\Sigma}^*_{n,m}, \mathbb{O}^+) \Pr(\chi_i \le c), \text{ for any } c \ge 0.$$

(ii) Following the proof of theorem 2.3.2, we have

$$\begin{aligned}\sum_{i=0}^{r} \omega_i \Pr(\chi_i \ge c) &\ge \sum_{i=1}^{r} \omega_i \Pr(\chi^2_1 \ge c) + \omega_0 \Pr(\chi^2_0 \ge c) \\ &= (1-\omega_0)\Pr(\chi^2_1 \ge c) + \omega_0 \Pr(\chi^2_0 \ge c) \\ &\ge \frac{\Pr(\chi^2_1 \ge c) + \Pr(\chi^2_0 \ge c)}{2},\end{aligned}$$

where $\omega_i = \omega_i(r, \hat{\Sigma}^*_{n,m}, \mathbb{O}^+)$.

Similarly, we have

$$\sum_{i=0}^{r} \omega_i \Pr(\chi^2_i \ge c) \le \frac{\Pr(\chi^2_{r-1} \ge c) + \Pr(\chi^2_r \ge c)}{2}.$$

Therefore, the tail probability of the above null distribution of T^*_W has the lower and upper bound:

$$\begin{aligned}\frac{\Pr(\chi^2_0 \ge c) + \Pr(\chi^2_1 \ge c)}{2} &\le \Pr(T^*_W \ge c | R\boldsymbol{\theta} = \mathbf{0}) \\ &\le \frac{\Pr(\chi^2_{r-1} \ge c) + \Pr(\chi^2_r \ge c)}{2}.\end{aligned}$$

■

A.2.4 Proof of Theorem 2.3.4

Proof. Following Meng and Rubin [1992], when the sample size n goes to infinity, we have $\hat{H}_{n*l} = H_n + o_p(1)$ and $\bar{H}_m = H_n + o_p(1)$. By the Cholesky decomposition, the positive definite matrix $\left(RH_n^{-1}R^T\right)^{-1}$ can be factorized as $\left(RH_n^{-1}R^T\right)^{-1} = C^TC$

. Then W_m is given by

$$
\begin{aligned}
W_m =& \frac{1}{m}\sum_{l=1}^{m} W_{*l} \\
=& \frac{1}{m}\sum_{l=1}^{m} n(R\tilde{\boldsymbol{\theta}}_{*l})^T \left(R\widehat{H}_{n*l}^{-1}R^T\right)^{-1}(R\tilde{\boldsymbol{\theta}}_{*l}) \\
=& \frac{n}{m}\sum_{l=1}^{m}(R\tilde{\boldsymbol{\theta}}_{*l})^T C^T C(R\tilde{\boldsymbol{\theta}}_{*l}) + o_p(1) \\
=& \frac{n}{m}\sum_{l=1}^{m}(CR\tilde{\boldsymbol{\theta}}_{*l})^T (CR\tilde{\boldsymbol{\theta}}_{*l}) + o_p(1) \qquad \text{(A.7)} \\
=& n\left[\frac{1}{m}\sum_{l=1}^{m}\left(CR\tilde{\boldsymbol{\theta}}_{*l} - CR\bar{\boldsymbol{\theta}}_m\right)^T \left(CR\tilde{\boldsymbol{\theta}}_{*l} - CR\bar{\boldsymbol{\theta}}_m\right) + \left(CR\bar{\boldsymbol{\theta}}_m\right)^T \left(CR\bar{\boldsymbol{\theta}}_m\right)\right] + o_p(1) \qquad \text{(A.8)} \\
=& \frac{n}{m}\sum_{l=1}^{m}\left(R\hat{\boldsymbol{\theta}}_{*l} - R\bar{\boldsymbol{\theta}}_m\right)^T \left(RH_n^{-1}R^T\right)^{-1}\left(R\tilde{\boldsymbol{\theta}}_{*l} - R\bar{\boldsymbol{\theta}}_m\right) \\
& + n\left(R\bar{\boldsymbol{\theta}}_m\right)^T \left(R\bar{H}_m^{-1}R^T\right)^{-1}\left(R\bar{\boldsymbol{\theta}}_m\right) + o_p(1) \qquad \text{(A.9)} \\
=& n\left(R\bar{\boldsymbol{\theta}}_m\right)^T \left(R\bar{H}_m R^T\right)^{-1}\left(R\bar{\boldsymbol{\theta}}_m\right) + n\frac{m-1}{m}\mathrm{tr}\left[RB_m R^T\left(RH_n^{-1}R^T\right)^{-1}\right] + o_p(1) \qquad \text{(A.10)} \\
=& n\left(R\bar{\boldsymbol{\theta}}_m\right)^T \left(R\bar{H}_m R^T\right)^{-1}\left(R\bar{\boldsymbol{\theta}}_m\right) + n\frac{m-1}{m}\mathrm{tr}\left[RB_m R^T\left(R\bar{H}_m^{-1}R^T\right)^{-1}\right] + o_p(1) \\
=& n(R\bar{\boldsymbol{\theta}}_m)^T \left(R\bar{H}_m^{-1}R^T\right)^{-1}(R\bar{\boldsymbol{\theta}}_m) + n\frac{(m-1)r}{m+1}r_m^* + o_p(1) \\
=& (1+r_m^*)T_W^* + \frac{(m-1)r}{m+1}r_m^* + o_p(1).
\end{aligned}
$$

In the above arguments, the equivalence between (A.7) and (A.8) holds by letting $Q_l = CR\hat{\boldsymbol{\theta}}_{*l}$ in the following equation

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

where $\bar{Q}_m = \frac{1}{m}\sum_{l=1}^{m} Q_l$. And the equivalence between (A.9) and (A.10) holds by noting $B_m = \frac{1}{m-1}\sum_{l=1}^{m}(\tilde{\boldsymbol{\theta}}_{*l} - \bar{\boldsymbol{\theta}}_m)(\tilde{\boldsymbol{\theta}}_{*l} - \bar{\boldsymbol{\theta}}_m)^T$.

Now we have

$$T_{CW}^* = \frac{W_m - \frac{(m-1)r}{m+1} r_m^*}{1 + r_m^*} = T_W^* + o_p(1).$$

Therefore, The proposed overall test statistics T_W^* and T_{CW}^* are asymptotically equal when n is large. This completes the proof. ■

Bibliography

M. Blackwell, J. Honaker, and G. King. A unified approach to measurement error and missing data: Overview and applications. *Sociological Methods & Research*, 46(3):303–341, 2017.

R. J. Carroll, D. Midthune, L. S. Freedman, and V. Kipnis. Seemingly unrelated measurement error models, with application to nutritional epidemiology. *Biometrics*, 62(1):75–84, 2006.

H. Chernoff. On the Distribution of the Likelihood Ratio. *The Annals of Mathematical Statistics*, 25(3):573–578, 1954.

S. Chib. Bayes inference in the tobit censored regression model. *Journal of Econometrics*, 51(1):79–99, 1992.

S. R. Cole, H. Chu, and S. Greenland. Multiple-imputation for measurement-error correction. *International Journal of Epidemiology*, 35(4):1074–1081, 2006.

G. M. Fitzmaurice, N. M. Laird, and G. E. P. Zahner. Multivariate logistic models for incomplete binary responses. *Journal of the American Statistical Association*, 91(433):99–108, 1996.

W. A. Fuller. *Measurement error models*, volume 305. John Wiley & Sons, 2009.

A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

G. Grover and V. K. Gupta. Multiple imputation of censored survival data in the presence of missing covariates using restricted mean survival time. *Journal of Applied Statistics*, 42(4):817–827, 2015.

K. M. Higgins, M. Davidian, and D. M. Giltinan. A two-step approach to measurement error in time-dependent covariates in nonlinear mixed-effects models, with application to igf-i pharmacokinetics. *Journal of the American Statistical Association*, 92(438):436–448, 1997.

Y. Huang, D. Liu, and H. Wu. Hierarchical bayesian methods for estimation of parameters in a longitudinal hiv dynamic system. *Biometrics*, 62(2):413–423, 2006.

J. P. Hughes. Mixed effects models with censored data with application to hiv rna levels. *Biometrics*, 55(2):625–629, 1999.

J. P. Hughes, P. Guttorp, and S. P. Charles. A non-homogeneous hidden markov model for precipitation occurrence. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 48(1):15–30, 1999.

N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions, volume 2*, volume 289. John Wiley & Sons, 1995.

N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.

M. Lee, L. Kong, and L. Weissfeld. Multiple imputation for left-censored biomarker data based on gibbs sampling method. *Statistics in Medicine*, 31(17):1838–1848, 2012.

K. H. Li, T. E. Raghunathan, and D. B. Rubin. Large-sample significance levels from multiply imputed data using moment-based statistics and an f reference distribution. *Journal of the American Statistical Association*, 86(416):1065–1073, 1991.

R. J. A. Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90(431):1112–1121, 1995.

R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics. Wiley, 2nd edition, 2002.

W. Liu and S. Li. A multiple imputation approach to nonlinear mixed-effects models with covariate measurement errors and missing values. *Journal of Applied Statistics*, 42(3):463–476, 2015.

W. Liu and L. Wu. Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. *Biometrics*, 63 (2):342–350, 2007.

X. Meng and D. B. Rubin. Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, 79(1):103–111, 1992.

W. Qiu and L. Wu. Hiv viral dynamic models with censoring and informative dropouts. *Statistics in Biopharmaceutical Research*, 2(2):220–228, 2010.

D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

D. B. Rubin. *Multiple imputation for nonresponse in surveys*. Wiley Series in Probability and Statistics. John Wiley & Sons, 1987.

J. L. Schafer. *Analysis of Incomplete Multivariate Data*. Taylor & Francis, 1st edition, 1997.

M. J. Silvapulle and P. K. Sen. *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. John Wiley & Sons, 2005.

F. Vaida and L. Liu. Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics*, 18(4):797–817, 2009.

T. Wang and L. Wu. Multiple imputation methods for multivariate one-sided tests with missing data. *Biometrics*, 67(4):1452–1460, 2011.

G. C. Wei and M. A. Tanner. Applications of multiple imputation to the analysis of censored regression data. *Biometrics*, 47(4):1297–1309, 1991.

G. C. G. Wei and M. A. Tanner. A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699–704, 1990.

H. Wu and A. A. Ding. Population hiv-1 dynamics in vivo: applicable models and inferential tools for virological data from aids clinical trials. *Biometrics*, 55(2): 410–418, 1999.

L. Wu. A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to aids studies. *Journal of the American Statistical Association*, 97(460):955–964, 2002.

L. Wu. *Mixed Effects Models for Complex Data*. Taylor & Francis, 1st edition, 2009.

L. Wu and H. Wu. Missing time-dependent covariates in human immunodeficiency virus dynamic models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 51(3):297–318, 2002.

G. Zhou and L. Wu. Multiparameter one-sided tests for nonlinear mixed effects models with censored responses. *Statistics in Medicine*, 40, 04 2021.

G. Zhou, L. Wu, R. Brant, and M. Ansermino. A likelihood-based approach for multivariate one-sided tests with missing data. *Journal of Applied Statistics*, 44: 1–17, 2016.