

Approximate Methods for Joint Models in Longitudinal Studies

by

Libo Lu

B.Sc., Shandong University, 2002

M.Sc., University of Science and Technology of China, 2005

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

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Statistics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2010

© Libo Lu 2010

Abstract

Longitudinal studies often contain several statistical issues, such as longitudinal process and time-to-event process, the association among which requires joint modeling strategy.

We firstly review the recent researches on the joint modeling topic. After that, four popular inference methods are introduced for jointly analyzing longitudinal data and time-to-event data based on a combination of typical parametric models. However, some of them may suffer from non-ignorable bias of the estimators. Others may be computationally intensive or even lead to convergence problems.

In this thesis, we propose an approximate likelihood-based simultaneous inference method for jointly modeling longitudinal process and time-to-event process with covariate measurement errors problem. By linearizing the joint model, we design a strategy for updating the random effects that connect the two processes, and propose two algorithm frameworks for different scenarios of joint likelihood function. Both frameworks approximate the multidimensional integral in the observed-data joint likelihood by analytic expressions, which greatly reduce the computational intensity of the complex joint modeling problem.

We apply this new method to a real dataset along with some available methods. The inference result provided by our new method agrees with those from other popular methods, and makes sensible biological interpretation. We also conduct a simulation study for comparing these methods. Our new method looks promising in terms of estimation precision, as well as computation efficiency, especially when

Abstract

more subjects are given. Conclusions and discussions for future research are listed in the end.

Table of Contents

Abstract	ii
Table of Contents	iv
List of Tables	vii
List of Figures	viii
Acknowledgments	ix
1 Introduction	1
1.1 Some Issues in Longitudinal Studies	1
1.1.1 Longitudinal Data	1
1.1.2 Incomplete Data	4
1.1.3 Time-to-Event Data	6
1.2 Joint Modeling Longitudinal Data and Time-to-Event Data	7
1.2.1 Motivation	7
1.2.2 Statistical Models	9
1.2.3 Inference Approaches	10
1.3 Motivating Example	12
1.3.1 Brief Biological Background	12
1.3.2 Statistical Question	13
1.4 Objective and Outline	14

Table of Contents

2	Statistical Models	16
2.1	Nonlinear Mixed Effects (NLME) Model for Longitudinal Process	16
2.2	Nonparametric Model for Covariate Measurement Errors Problem	19
2.3	AFT Model for Time-to-Event Data	22
2.4	Models for Nonignorable Missing Data	26
3	Simultaneous Inference	29
3.1	Two-Step Method	31
3.1.1	Naive Two-Step Method	31
3.1.2	Modified Two-Step Method	32
3.2	EM Algorithm	33
3.2.1	E-Step	34
3.2.2	M-Step	36
3.3	Laplace Approximation	37
3.4	Linearization Method	39
3.4.1	Framework I	43
3.4.2	Framework II	43
4	Data Analysis	45
4.1	Data Description	45
4.2	Model Specification	46
4.3	Analysis and Results	48
5	Simulation	53
5.1	Simulation Design	53
5.2	Simulation Results	54
5.2.1	Number of Subjects	54
5.2.2	Numbers of Repeated Measurements	56
5.2.3	Between-Subject Variance	57

Table of Contents

5.2.4 Within-Subject Variance	58
6 Conclusion and Discussion	61
Bibliography	64

List of Tables

- 4.1 Estimation Results 51

- 5.1 Simulation Results: $n = 60$ 54
- 5.2 Simulation Results: $n = 120$ 55
- 5.3 Simulation Results: $n_i = 12$ 56
- 5.4 Simulation Results: A_2, B_2 58
- 5.5 Simulation Results: $\mu_2, \lambda_2, \sigma_2$ 59

List of Figures

1.1	Longitudinal Data of 6 Subjects	2
1.2	Plots of Longitudinal Data of Subject 10	13
1.3	Plots of Longitudinal Data of Subject 18	14
1.4	Plots of Longitudinal Data of Subject 33	14
4.1	Longitudinal Data Trajectories	46
4.2	Estimate of Survivor Function	47
4.3	Residuals Plots I of NLME Model	49
4.4	Residuals Plots II of NLME Model	49
4.5	Residuals Plots I of Covariate Measurement Errors Model	50
4.6	Residuals Plots II of Covariate Measurement Errors Model	50

Acknowledgments

First and foremost, I would like to express my gratitude to my supervisors, Dr. Lang Wu and Dr. Jiahua Chen for their support, encouragement and patience throughout my study at the University of British Columbia. Their world-leading expertise and excellent guidance defined a role model for my career. This thesis would not have been completed without their inspiring comments and constructive suggestions.

I would also like to thank everyone in the Department of Statistics at the University of British Columbia. In the past two years, they have kindly left me invaluable knowledge, along with wonderful memories. I had a great time with their company.

Many many many thanks to my parents and Miss. Marie Li for their unconditional love and support, and to my lifelong friends for their lifelong loyalty. And finally, I must thank Buddha for always being with me no matter what happens, dispelling my fear, healing my pain, comforting my soul and guiding my life forward. It is lucky to be his follower.

Chapter 1

Introduction

1.1 Some Issues in Longitudinal Studies

With the development of modern statistical methodology, *longitudinal studies* play an increasingly important role in health science, medical research, social science and economics. This type of study obviously differs from the classical *cross-sectional studies* in terms of the times of measurement of the response. That is to say, in a longitudinal study, some variables are measured more than once at different time points for each subject. Longitudinal studies, which enable the researchers to separate the *cohort* and *time* effects, are valuable because the effect of time could be a serious confounding factor to other covariates of interest.

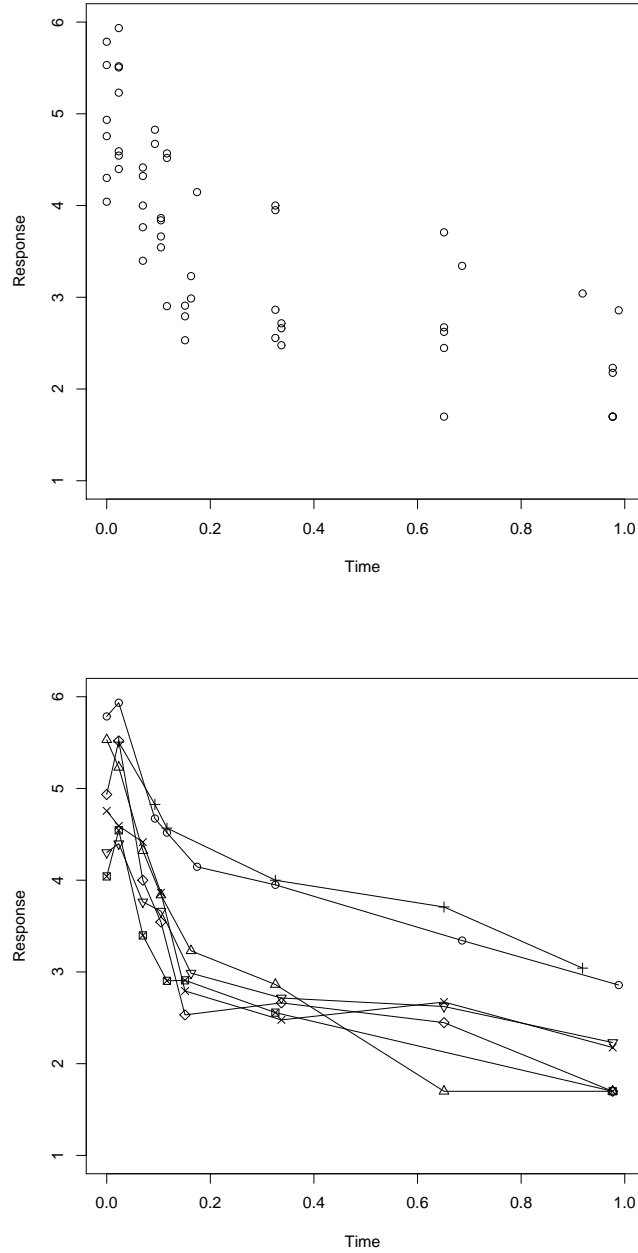
1.1.1 Longitudinal Data

Suppose n subjects are included a longitudinal study. The value of the response variable y for subject i at time t_{ij} is denoted as y_{ij} , where $i = 1, \dots, n, j = 1, \dots, n_i$. This type of data are called *longitudinal data*. The graphical display of longitudinal data requires attention. A sample plot of longitudinal data of 6 subjects is on the top of Figure 1.1, which is quite difficult for discovering growth patterns. It is often helpful to distinguish subjects with different types of points and connect the observations of each subject with lines, such as the bottom plot of Figure 1.1.

Comparing to the data obtained from a cross-sectional study, the data col-

1.1. Some Issues in Longitudinal Studies

Figure 1.1: Longitudinal Data of 6 Subjects



lected from a longitudinal study are naturally correlated. This feature requires a statistical methodology to take the correlation into account. For a longitudinal data analysis, a primary interest is to estimate the population mean and its relationship with covariates. A popular method for this purpose is called the *marginal model*, which separately models the mean structure and the variance-covariance structure of the response variable. Its underlying idea is close to the quasi-likelihood approach that only requires the first and second moment information of the response variable without specific distributional assumption. Another approach is the *transitional model*, which assumes a Markov structure for the longitudinal process to model the correlation among the repeated measurements. This approach is appealing if the current observation strongly depends on previous ones.

Marginal models and transitional models consider the between-subject variation and within-subject variation separately. A *mixed effects model* considers the two sources of variation simultaneously by incorporating *random effects* to represent the between-subject variation and within-subject correlations. Laird and Ware (1982) proposed a general *linear mixed effects* (LME) model:

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \mathbf{e}_i \quad (1.1)$$

$$\mathbf{b}_i \sim N(0, D), \quad \mathbf{e}_i|\mathbf{b}_i \sim N(0, R_i), \quad (1.2)$$

where $\boldsymbol{\beta}$ are the fixed effects (population parameters), $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})^T$ are the random effects (subject-specific parameters), X_i is a $n_i \times (p + 1)$ design matrix containing the covariates of subject i , Z_i is a $n_i \times q$ design matrix (Z_i is often a submatrix of X_i) for the random effects, $\mathbf{e}_i = (e_{i1}, \dots, e_{in_i})^T$ are the random errors of those within-subject measurements, R_i is a $n_i \times n_i$ matrix characterizing variance-covariance of within-subject measurements, and D is the variance-

covariance matrix of the random effects.

The inference procedure of mixed effects models is a natural extension of classic regression methods for cross-sectional models, and most ideas and results, such as the maximum likelihood estimation (MLE) and asymptotic normality, also apply to the estimation of LME models.

Diggle et al. (2002) provided a comprehensive overview of the above three methods. Verbeke and Molenberghs (2001) and Davidian and Giltinan (1995) gave more discussions on linear and nonlinear mixed effects models. In practice, the longitudinal trajectories could be so complex that parametric models are not flexible enough. For those cases, Wu and Zhang (2006) introduced some semiparametric or nonparametric models.

1.1.2 Incomplete Data

The data observed in longitudinal studies are very likely to be incomplete, such as measurement errors and missing data. In the following, we briefly introduce these problems.

Measurement Errors Problem

It is ideal that we can get precise values of the selected variables, such as gender and age. However, in practice, the measurements of some continuous variables, such as blood pressure and CD4, are often imprecise due to various reasons. We call this issue as *measurement error problem*, the distinguishing feature of which is that we can only observe a variable \mathbf{z} that are related to \mathbf{z}^* . A major goal of measurement error modeling is to obtain nearly unbiased estimates of the parameters in the main model by fitting a sub-model based on \mathbf{z} . This objective requires careful analysis, because simply substituting \mathbf{z} for \mathbf{z}^* without proper adjustments may lead to biased estimates in some cases. Carroll et al. (2006)

reviewed general methods for measurement errors problems.

Missing Data Problem

Missing data problems mainly encompass missing responses, missing covariates, and dropouts. A main reason is that not all subjects are available at every measurement time throughout the entire study, especially if the study lasts for a long period.

In the cases of missing responses, missing covariates, and dropouts, part of the data are completely missing. The missingness can be classified into three mechanisms:

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

For instance, suppose a student has not submitted his assignment in class on the due day, then his grade for that assignment was missing. If he left his assignment at home, the missing grade is MCAR. If he missed that class, then the missing grade is MAR, since the missingness is determined by the observed data “attendance”. If he has not finished his assignment, then the missing grade is MNAR.

There are enormous valuable literatures on this topic, since the missing data are often informative and nonignorable. Statistical inference without considering this issue might result in biased estimation. Little and Rubin (2002) provided a systematic introduction to major missing data methods. Schafer (1997) gave a

general introduction for incomplete multivariate data analysis. Wu (2009) offered an overview of analyzing incomplete data problems with mixed effects models.

1.1.3 Time-to-Event Data

In many longitudinal studies, the time to certain events happen, such as time of patient's death, are of great interest. It is conventional to talk about survival data and survival analysis, regardless of the nature of the events. Often, statistical models are built for studying the relationship between the event time and other covariates.

The distinguishing feature of time-to-event data is *censoring*, which means that the event time may not be observed exactly. For example, suppose that a subject is lost to follow-up after 5 years of observation. According to these complex reasons, censoring issues can be classified into three types:

- *Right-censoring*. The time of event is not observed because it happened to the *right* of the 5th year. This case is called right-censoring.
- *Left-censoring*. Suppose a subject had an event before the 5th year but the exact time of the event is unknown. This case is called left-censoring.
- *Interval-censoring*. Suppose a subject had an event between the 4th year and 5th year but the exact time of the event is unknown. This case is called interval-censoring.

Here are some possible reasons why censoring may occur:

- a subject does not experience the event before the study end;
- a person is lost to follow-up during the study period;
- a person withdraws from the study because of a death (if death is not the event of interest) or some other reasons.

The censoring feature, which is sometimes referred as a special missing data issue, determines that specific statistical methodology is necessary.

Common approaches in time-to-event data analysis include *Cox Proportional Hazards model* (Cox model), which assumes that the effect of a covariate is to multiply the hazard by some constant, and *Accelerated Failure Time model* (AFT model), which assumes that the effect of a covariate is to multiply the predicted event time by some constant. Time-to-event data have been extensively studied. Lawless (2003) provided comprehensive introduction to the relevant statistical models and methods.

1.2 Joint Modeling Longitudinal Data and Time-to-Event Data

1.2.1 Motivation

The statistical issues mentioned in section 1.1 often occur simultaneously but used to be analyzed separately in longitudinal studies. Previous relevant researches reported that the intrinsic relationships among longitudinal process, time-to-event process and incomplete data mechanism can be greatly influential to the statistical inference for diverse scientific objectives.

- Wu and Carroll (1988) reported that analyzing longitudinal data without modeling informative censoring issue might lead to biased results. Their work was generalized by DeGruttola and Tu (1994). Little (1995) pointed out the possible estimation bias for analyzing longitudinal data due to ignoring the missing data problems. Wu (2002) and Liu and Wu (2007) revealed the effects of measurement errors problems and missing data problems on the inference of longitudinal process. These researches imply that classical statistical methods for longitudinal data analysis may need to be

adjusted for incorporating certain critical events in longitudinal studies, such as outcome-dependent drop-out.

- Faucett and Thomas (1996) and Wulfshon and Tsiatis (1997) modeled time-to-event data with time-varying covariates measured with errors, where the joint modeling approach and the separate modeling approach provided different estimates. These researches indicated that valid modeling of time-to-event data might be conditional on relevant longitudinal processes, for instance, the time-varying covariates. Their researches were generalized respectively by Xu and Zeger (2001a,b) and Song et al. (2002a,b).
- Henderson et al. (2000) jointly analyzed longitudinal data and time-to-event data with equal interests using the likelihood-based approaches. Wang and Taylor (2001) and Brown and Ibrahim (2003) further generalized the modeling assumptions on longitudinal process. Joint modeling of both longitudinal process and time-to-event process is particularly essential if the association between the event times and growth trends of the longitudinal process is of interest.

Generally speaking, joint modeling links the longitudinal process and the time-to-event process via a common set of random effects. Under the joint modeling framework, the main focus could be just on the longitudinal process or the time-to-event process, or it could be on both processes. Since this joint modeling approach contains at least two processes, the parameters to be estimated are usually more than a single inference. If only one specific process is of interest, then the other process would better be modeled as simple as possible to reduce the nuisance parameters.

Tsiatis and Davidian (2004) provided an excellent overview on the motivations of joint modeling and relevant preceding literature. The rest of this section briefly

reviews both statistical model and inference approach that have been adopted for recent researches.

1.2.2 Statistical Models

Models for Longitudinal Data

Many researches adopt mixed effects models for modeling longitudinal data. One typical model, adopted by Ratcliffe et al. (2004), Zeng and Cai (2005a,b), Li et al. (2009) and Huang et al. (2009), is parametric linear mixed effects models with covariates as predictors. Other researchers, such as Brown et al. (2005), Ding and Wang (2008), Ye et al. (2008) and Zhang et al. (2010), used nonparametric models for more flexible descriptions of the longitudinal processes. Vonesh et al. (2006) and Yu et al. (2008) adopted nonlinear mixed effects models according to the corresponding scientific backgrounds. Rizopoulos et al. (2008), Liu (2009) and Li et al. (2010) established their approaches based on generalized mixed effects models for analyzing categorical longitudinal data.

Models for Time-to-event Data

For modeling time-to-event data, some joint modeling studies, including Ratcliffe et al. (2004), Song and Huang (2006), Ding and Wang (2008), Li et al. (2009) and Zhang et al. (2010), were based on Cox proportional hazard model, partially because Cox model is widely used in practice. Recently, according to the scientific background of the data, more and more attention have been turned to accelerated failure time model, such as Tseng et al. (2005) and Rizopoulos et al. (2008). Particularly, Vonesh et al. (2006) and Rizopoulos et al. (2010) adopted Weibull model, which bears the characteristics of both Cox model and AFT model.

1.2.3 Inference Approaches

Two-stage Approach

The idea of two-stage approach is to firstly estimate the shared parameters \mathbf{b}_i from either the longitudinal process or the time-to-event process, such as Zhang et al. (2008) and Ye et al. (2008), with the corresponding parameters of the first model being estimated. Then the estimated shared parameters $\hat{\mathbf{b}}_i$ are used for the estimation of the parameters of the second model. Although two-stage approach is easy to implement, there are several drawbacks (Wulfshon and Tsiatis, 1997). For example, the estimated parameters of one process from the first step are regarded as fixed in the modeling of the other process in the second step, which does not propagate uncertainty from step one to step two.

Likelihood-Based Approach

A more unified alternative of statistical inference is based on the joint likelihood function given both longitudinal data and time-to-event data. The joint modeling approaches simultaneously estimate the parameters that describe the longitudinal process, as well as those that describe the time-to-event process. Besides enhancing the efficiency of the inference, joint modeling is also expected to provide more precise estimation of the relationship between the two processes.

Tseng et al. (2005) considered a LME model for longitudinal data along with a general AFT model for time-to-event model, and developed an EM algorithm to maximize the resulting log-likelihood, which involves intractable integrals over the distribution of fixed effects. The similar EM algorithms have been applied to many other researches based on Cox models, such as Ratcliffe et al. (2004), Wu et al. (2008), Rizopoulos et al. (2009) and Li et al. (2010). Vonesh et al. (2006) and Rizopoulos et al. (2009) effectively reduced the computational difficulties in

the E-step by using Laplace approximation.

Zeng and Cai (2005a) adopted a LME model and a Cox model for the longitudinal process and the time-to-event process, and discussed the asymptotic behavior of the MLE obtained from EM algorithm. Hsieh et al. (2006) reinforced the merit of the joint modeling approach in Wulfshon and Tsiatis (1997) by providing a theoretical explanation of the robustness features observed in the literature. They suggested that the likelihood-based procedure with normal random effects can be very efficient and robust as long as there is rich enough information available from the longitudinal data. However, if the longitudinal data are too sparse or carry too large measurement errors, the efficiency loss of joint modeling can be quite substantial. Hsieh et al. (2006) also recommended to use bootstrap method to get more reliable estimates of the standard errors of the MLEs.

Brown et al. (2005), Yu et al. (2008) and Zhang et al. (2010) took Bayesian approaches to joint model two processes and developed Markov Chain Monte Carlo (MCMC) implementations, which were also based on likelihood. Ibrahim et al. (2001b, chap. 7) provided a detailed discussion on joint modeling from a Bayesian perspective. Yu et al. (2004) compared the inference results for Bayesian approach and EM approach, both of which rely on the specification of likelihood functions.

Semiparametric Approach

Song and Huang (2006) and Song and Wang (2008) focused on estimating the fixed effects in the joint models of longitudinal process and time-to-event process by developing a set of unbiased estimating equations (conditional and corrected score approaches), which yield consistent and asymptotically normal estimators with no assumptions on the random effects. Their approaches reduce reliance on the parametric modeling assumptions.

1.3 Motivating Example

1.3.1 Brief Biological Background

Human Immunodeficiency Virus (HIV) is a virus that directly attacks certain human organs, such as the brain, heart, and kidneys, as well as the human immune system. Infection with HIV occurs by the transfer of blood, semen, vaginal fluid, pre-ejaculate, or breast milk. Within these bodily fluids, HIV is present as both free virus particles and virus within infected immune cells. The four major routes of transmission are unsafe sex, contaminated needles, breast milk, and transmission from an infected mother to her baby at birth (Vertical transmission). Screening of blood products for HIV has largely eliminated transmission through blood transfusions or infected blood products in the developed world.

The primary cells attacked by HIV are the **CD4** lymphocytes, which help direct immune function in the body. HIV infection leads to low levels of CD4 cells through three main mechanisms: direct viral killing of infected cells; increased rates of apoptosis in infected cells; and killing of infected CD4 cells by CD8 cytotoxic lymphocytes that recognize infected cells. Since CD4 cells are required for proper immune system function, when enough CD4 lymphocytes have been destroyed by HIV, the immune system barely works. Many problems experienced by people infected with HIV result from a failure of the immune system to protect them from certain opportunistic infections (OIs) and cancers.

Most people infected with HIV eventually develop *Acquired Immunodeficiency Syndrome (AIDS)*. These individuals mostly die from opportunistic infections or malignancies associated with the progressive failure of the immune system. HIV progresses to AIDS at a variable rate affected by viral, host, and environmental factors. Most individuals will progress to AIDS within 10 years of HIV infection, sooner or later. HIV treatment with anti-retrovirals increases the life expectancy

1.3. Motivating Example

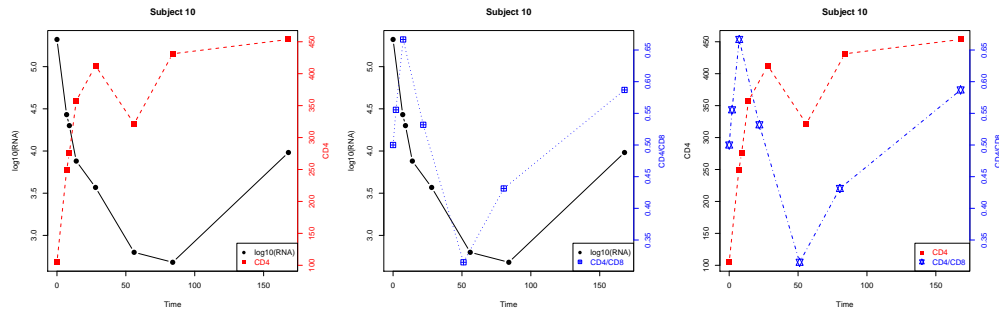
of people infected with HIV. Even after HIV has progressed to diagnosable AIDS, the average survival time with antiretroviral therapy was estimated to be more than 5 years as of 2005. Without antiretroviral therapy, someone who has AIDS typically dies within a year.

1.3.2 Statistical Question

According to the clinical experiences, HIV infection often occurs along with the variation of CD4 cells, CD8 cells and their ratio. Recently, the CD4/CD8 ratio has become a new biomarker for assessing the relative condition of HIV subjects, and further more, for predicting the progression from HIV to AIDS.

After having certain anti-HIV treatments, the viral loads of HIV-infected subjects are expected to decline, roughly with biphasic exponential decay. Meanwhile, we might observe a corresponding increase of CD4 and a decrease of CD8, and consequently, an increase of CD4/CD8 ratio. However, the viral load might rebound after a period of treatment due to various reasons. The relationship between HIV viral suppression and immune restoration is of great attention in AIDS research.

Figure 1.2: Plots of Longitudinal Data of Subject 10



Our research was motivated by an AIDS Clinical Trials Group (ACTG) study (Wu and Ding, 1999) for demonstrating that the initial viral decay rate reflects

1.4. Objective and Outline

Figure 1.3: Plots of Longitudinal Data of Subject 18

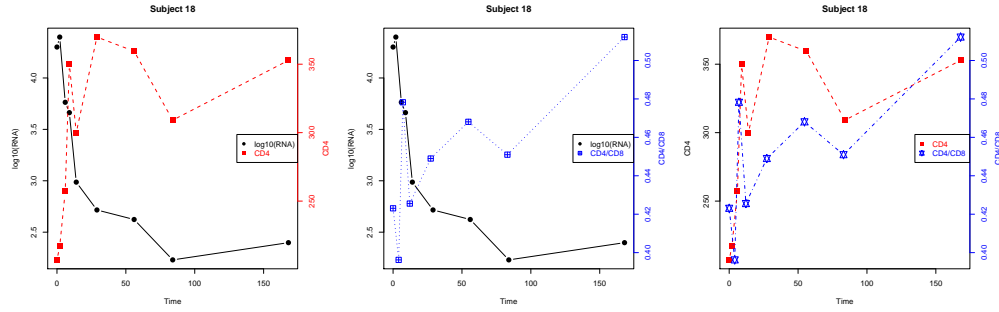
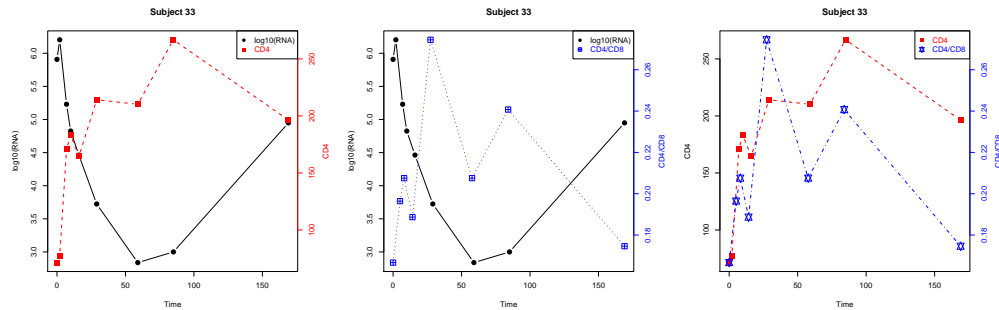


Figure 1.4: Plots of Longitudinal Data of Subject 33



the efficacy of an anti-HIV treatment. Figure 1.2, 1.3 and 1.4 are the trajectories of HIV viral load, CD4 count, and CD4/CD8 ratio of three randomly selected subjects. HIV viral load appears to be negatively correlated with both CD4 count and CD4/CD8 ratio. This article studies the potential underlying association between the viral load decay and the CD4/CD8 time trend by jointly modeling the HIV viral load dynamics and the time of the first decrease of CD4/CD8.

1.4 Objective and Outline

In this thesis, we develop a new approximate statistical inference method for jointly modeling longitudinal process and time-to-event process with covariate

1.4. Objective and Outline

measurement errors. Comparing to other existing inference methods, the methods that we proposed are easy to implement and much more computationally efficient.

This thesis is organized as follows. Chapter 2 introduces the statistical models that will be included in our joint modeling. Chapter 3 discusses currently available inference methods and proposes our new methods with procedure details. These methods are to be applied to a real dataset from the motivating example in Chapter 4 for data analysis. The simulation results are displayed in Chapter 5. We finally make the conclusion and future research discussion in Chapter 6.

Chapter 2

Statistical Models

The joint modeling of longitudinal process and time-to-event process is basically composed of several simpler models which represent respectively different specific processes. In this chapter, we introduce the statistical models for each process in our longitudinal study.

2.1 Nonlinear Mixed Effects (NLME) Model for Longitudinal Process

The general LME model (1.1) has achieved numerous successes in both theoretical and applied research for its simplicity in modeling, computation and interpretation. However, linear structure only provides local linear approximation to the growth trend of the response variable and thus limits our vision within the range of the observed data. In many longitudinal studies, the developments of response variables have certain deterministic mechanisms underlying, which can often be represented by nonlinear mathematical formulations of physically meaningful parameters. The nonlinear formulations, if being correctly specified, support more precise estimation for the statistical model, both inside and outside of the range of the observed data, with less parameters than linear models. Therefore, nonlinear mixed effects models (NLME) are becoming increasingly favorable in representing the longitudinal process with known developing mechanism.

Suppose n subjects are included in a longitudinal study. Let y_{ij} be the response

value of subject i at time t_{ij} , and let z_{ij} be the corresponding covariate value. We can write the NLME model in the following two-stage hierarchical nonlinear model:

$$y_{ij} = g(t_{ij}, \boldsymbol{\beta}_i) + e_{ij} \quad (2.1)$$

$$\boldsymbol{\beta}_i = h(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i) \quad (2.2)$$

$$\mathbf{b}_i \sim N(0, B), \quad \mathbf{e}_i | \mathbf{b}_i \sim N(0, R_i), \quad (2.3)$$

where $g(\cdot)$ and $h(\cdot)$ are given nonlinear functions, $\boldsymbol{\beta}_i$ are the subject-specific parameters and $\boldsymbol{\beta}$ are the fixed effects (population parameters), \mathbf{z}_i contain the covariates of subject i , $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})^T$ are the random effects distinguishing subjects, $\mathbf{e}_i = (e_{i1}, \dots, e_{in_i})^T$ are the random errors of within-subject measurements, B is the variance-covariance matrix of the random effects, and R_i is a $n_i \times n_i$ matrix characterizing the variance-covariance of within-subject measurements.

In practice, the function $g(\cdot)$ is determined by the scientific problem and the data background. The function $h(\cdot)$ is often a linear combination of the fixed effects and the random effects, such as

$$\boldsymbol{\beta}_i = A_i \boldsymbol{\beta} + B_i \mathbf{b}_i, \quad (2.4)$$

where A_i is a design matrix depending on the elements of \mathbf{z}_i , and B_i is a design matrix, which typically involve only zeros and ones, allowing some elements of $\boldsymbol{\beta}_i$ to have no associated random effects. The random effects \mathbf{b}_i represent between-subject variance that is not explained by covariate \mathbf{z}_i . Thus, the variance-covariance matrix B is usually unstructured. The error term \mathbf{e}_i represent the lack of fit of the models and possible measurement errors of the data. The choice of R_i should be guided under practical background, on which Davidian and Giltinan (2003) gave detailed discussions.

A widely applicable approach for estimating parameters of NLME is based on likelihood. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\kappa}, B)$ denotes all parameters, where $\boldsymbol{\kappa}$ is the collection of parameters in $R_i, i = 1, \dots, n$. The marginal distribution of the response \mathbf{y}_i is given by

$$f(\mathbf{y}_i|\boldsymbol{\theta}) = \int f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\theta}, \mathbf{b}_i)f(\mathbf{b}_i|B)d\mathbf{b}_i, \quad (2.5)$$

and the likelihood is

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^n L(\boldsymbol{\theta}|\mathbf{y}_i) = \prod_{i=1}^n \int f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\theta}, \mathbf{b}_i)f(\mathbf{b}_i|B)d\mathbf{b}_i, \quad (2.6)$$

The statistical inference is then based on maximum likelihood estimation (MLE) approach, which means that the classic large-sample asymptotic result is promising to hold in this situation.

In general, the marginal distribution (2.5) or the likelihood function (2.6) has no closed-form expression, which greatly distinguishes NLME model from LME model. This characteristic implies that a major challenge of likelihood-based approach for NLME model is the maximization of the likelihood function (2.6). Commonly applicable methods are:

- Numerical method or Monte Carlo method, which uses different integration techniques to approximate the integral of (2.6) and then maximizes the approximate likelihood.
- EM algorithm, which iteratively maximizes the likelihood function (2.6).
- Laplace approximation method, which approximates the integral of (2.6) by a simple form and then maximizes the approximate likelihood.
- Taylor expansion method, which linearizes the NLME into a sequence of LMEs and approximate the maximum point of the likelihood function (2.6).

Davidian and Giltinan (1995) and Wu (2009) provided comprehensive discussions on the asymptotic results and implementational details.

2.2 Nonparametric Model for Covariate Measurement Errors Problem

In longitudinal studies, the covariates are usually recorded and analyzed to partially explain the between-subject variance. However, the covariates may be measured with substantial errors. It is well known that ignoring covariates measurement errors might lead to severe bias in statistical inference (see, e.g., Carroll et al. (2006)). Therefore, the covariate measurement errors problem should be incorporated into the analysis of longitudinal data. Generally, the choice of appropriate analysis method for measurement errors problem depends on the following two considerations.

- The choice of statistical model depends on the type and amount of data available for analysis. A sophisticated error model may be less reliable without sufficient data available.
- The measurement error models can mainly be classified into *functional models* and *structural models*. Comparatively speaking, functional models treat \mathbf{z}^* as a sequence of unknown fixed constants or a random variable with minimal distributional assumption. It leads to estimation procedures that are robust to mis-specification of the distribution of \mathbf{z}^* . On the other hand, structural models assume \mathbf{z}^* follow specific parametric distributions, the estimates based on which are usually more efficient if the parametric distributions are correctly specified.

From the viewpoint of functional data analysis, measurement errors problems in longitudinal studies are closely related to fitting several smooth curves. Thanks

to the magnificent developments of functional data analysis and nonparametric statistical analysis in recent years, the researchers now have more flexible techniques to handle complex data with high precision. Rice and Wu (2001) proposed a linear mixed effects regression spline model for making subject-specific inference for the covariates with measurement errors and distinguishing the between-subject and within-subject variance. Liu and Muller (2009) developed a general and model-free dynamic equation for obtaining the derivatives of the growth curves for sparse data. This empirical differential equation has potential to provide valuable insights into the time-dynamics of random trajectories.

In our study, we adopt the approach of Rice and Wu (2001) for modeling the time-varying covariate process to incorporate the measurement errors. Since the covariate process could be quite complex, a flexible empirical nonparametric linear mixed effects model is considered.

$$z_i(t) = r(t) + h_i(t) + \xi_i(t) \equiv z_i^*(t) + \xi_i(t), \quad i = 1, \dots, n. \quad (2.7)$$

where $z_i^*(t_{ik}) = r(t_{ik}) + h_i(t_{ik})$ are the true but unobserved covariate values, $r(t_{ik})$ and $h_i(t_{ik})$ are unknown nonparametric smooth functions representing the population means and the subject-specific variation respectively, and $\xi_i(t_{ik})$ are the error terms following $N(0, \lambda^2)$. The random smooth function $h_i(\cdot)$ is introduced to incorporate the large between-subject variation in the covariate process, while the fixed smooth function $r(\cdot)$ represents population average of the covariate process. We assume that $h_i(\cdot)$ is the realization of a zero-mean stochastic process.

As in Rice and Wu (2001), we can approximate the nonparametric functions $r(t)$ and $h_i(t)$ by linear combinations of some basis functions $\Psi_p(t) =$

2.2. Nonparametric Model for Covariate Measurement Errors Problem

$[\psi_0(t), \psi_1(t), \dots, \psi_{p-1}(t)]^T$ and $\Phi_q(t) = [\phi_0(t), \phi_1(t), \dots, \phi_{q-1}(t)]^T$ as follows:

$$r(t) \approx r_p(t) = \sum_{k=0}^{p-1} \alpha_k \psi_k(t) = \Psi_p(t)^T \boldsymbol{\alpha}, \quad (2.8)$$

$$h_i(t) \approx h_{iq}(t) = \sum_{k=0}^{q-1} a_{ik} \phi_k(t) = \Phi_q(t)^T \mathbf{a}_i, \quad i = 1, \dots, n. \quad (2.9)$$

where $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_{p-1})^T$ is a $p \times 1$ vector of fixed effects and $\mathbf{a}_i = (a_{i0}, \dots, a_{i,q-1})^T$ is a $q \times 1$ vector of random effects, and we assume \mathbf{a}_i are i.i.d. $\sim N(0, A)$. The functions $\Psi(\cdot)$ and $\Phi(\cdot)$ can theoretically be taken as any basis functions, including various types of splines, such as B-splines (de Boor, 1978), smoothing splines (Chambers, 1991) and P-splines (Ramsay and Silverman, 2005), or local polynomials (Fan and Gijbels, 1996). The above approximations of $r(t)$ and $h_i(t)$ can be arbitrarily accurate as the values of p and q increase. However, too large values of p and q may lead to overfitting problem. Appropriate values of p and q can be determined by standard model selection criteria such as AIC or BIC values (Rice and Wu, 2001). By approximating $r(t)$ and $h_i(t)$ with $r_p(t)$ and $h_{iq}(t)$ respectively, the covariate model (2.7) can be approximated by the following LME model

$$\mathbf{z}_i = \mathbf{z}_i^* + \boldsymbol{\xi}_i = \Psi_p^T \boldsymbol{\alpha} + \Phi_q^T \mathbf{a}_i + \boldsymbol{\xi}_i \quad (2.10)$$

$$\mathbf{a}_i \sim N(0, A), \quad \boldsymbol{\xi}_i | \mathbf{a}_i \sim N(0, \lambda^2 I_i). \quad (2.11)$$

Model (2.10) can be used to model covariate measurement errors problems (Carroll et al., 2006). For multiple covariates, we can model each covariate separately using (2.10). However, a more reasonable approach is to consider a multivariate version of model (2.10), such as a multivariate LME model (Shah et al., 1997).

2.3 AFT Model for Time-to-Event Data

In the analysis of time-to-event data, the *survivor function* and the *hazard function*, which depend on time, are of particular interest. Denote $f(t)$ as the probability density function (p.d.f.) of event time T . The survivor function $S(t)$ is defined as the probability of surviving to time t

$$S(t) = Pr(T \geq t) = \int_t^{\infty} f(x)dx. \quad (2.12)$$

The graph of $S(t)$ against t is called the *survival curve*. The hazard function $h(t)$ is defined as the event rate at time t conditional on no event occurs until time t or later.

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t). \end{aligned} \quad (2.13)$$

Many statistical analysis methods for analyzing time-to-event data, no matter they are parametric, semiparametric or nonparametric, are closely related to the likelihood function given the data, which is determined by the existence and specific type of censoring issue. Let's consider a simple right-censoring case for example. Let T_i be the event time for subject $i, i = 1, \dots, N$. Let C_i be the censoring time for subject i . Obviously, if no censoring occurs, T_i is observed and less than the potential censoring time ($T_i \leq C_i$). If censoring occurs, we only know that $T_i > C_i$. In general, we denote

$$\rho_i = \min(T_i, C_i), \quad \delta_i = I(T_i \leq C_i). \quad (2.14)$$

Both ρ_i and δ_i are random variables and their joint p.d.f. is

$$f(\rho_i)^{\delta_i} S(\rho_i)^{1-\delta_i}. \quad (2.15)$$

Suppose T_1, \dots, T_n are independent, we may obtain the likelihood function given the right-censored data as

$$L = \prod_{i=1}^N f(\rho_i)^{\delta_i} S(\rho_i)^{1-\delta_i}. \quad (2.16)$$

The *Kaplan-Meier method*, the nonparametric MLE of the survivor function $S(t)$, can be used to estimate this curve from the observed survival times without assumption on distribution. However, in many cases, the survival function $S(t)$ is assumed following certain distributions. For instance, if $S(t)$ can be written as

$$S(T|u, \sigma) = S_0 \left(\frac{\log T - u}{\sigma} \right), \quad (2.17)$$

where $u \in \mathbb{R}$ is a location parameter and $\sigma > 0$ is a scale parameter, we say T has a *log-location-scale* distribution, which is a widely used parametric distribution for time-to-event data. If we generalize u to $u(\mathbf{z})$, we may have more covariates \mathbf{z} into account, that is

$$S(T|\mathbf{z}, u, \sigma) = S_0 \left(\frac{\log T - u(\mathbf{z})}{\sigma} \right). \quad (2.18)$$

Suppose ϵ is a random variable with survivor function $S_0(\epsilon)$ and is independent from \mathbf{z} , then (2.18) can be rewritten as

$$\log T = u(\mathbf{z}) + \sigma\epsilon. \quad (2.19)$$

Here the survivor function is represented in a regression setting.

2.3. AFT Model for Time-to-Event Data

Formula (2.19) is a general semiparametric AFT model. In practice, $u(\cdot)$ can be taken as a linear function and we have the following parametric AFT model

$$\log T = \gamma_0 + \gamma_1^T \mathbf{z} + \sigma \epsilon, \quad (2.20)$$

where ϵ is independent from \mathbf{z} . In this thesis, we consider the random effects \mathbf{a}_i and \mathbf{b}_i as covariates and build the following AFT frailty model for time-to-event data.

$$\log T_i = \gamma_0 + \gamma_1^T \mathbf{a}_i + \gamma_2^T \mathbf{b}_i + \sigma \epsilon_i, \quad i = 1, \dots, n. \quad (2.21)$$

where $\gamma = (\gamma_0, \gamma_1, \gamma_2)$ are the unknown parameters and ϵ_i are i.i.d. random variables. Note that the NLME longitudinal model (2.1) and the time-to-event model (2.21) are connected through the random effects \mathbf{b}_i .

There are three major parametric AFT models based on different distributions of ϵ .

- *Weibull AFT model.* Assume the event times follow a Weibull distribution with scale parameter $\exp(-(\gamma_0 + \gamma_1^T \mathbf{a}_i + \gamma_2^T \mathbf{b}_i)/\sigma)$ and shape parameter $1/\sigma$, the survivor function is

$$S(t) = \exp \left\{ - \exp \left(\frac{\log t - \gamma_0 - \gamma_1^T \mathbf{a}_i - \gamma_2^T \mathbf{b}_i}{\sigma} \right) \right\}, \quad (2.22)$$

Because the Weibull distribution has both the proportional hazards property and the accelerated failure time property, this model is particularly attractive.

- *The log-logistic AFT model.* Assume the event times follow a log-logistic distribution with parameters $(\gamma_0 + \gamma_1^T \mathbf{a}_i + \gamma_2^T \mathbf{b}_i)/\sigma$ and $1/\sigma$, the survivor

function is

$$S(t) = \left\{ 1 + \exp \left(\frac{\log t - \gamma_0 - \gamma_1^T \mathbf{a}_i - \gamma_2^T \mathbf{b}_i}{\sigma} \right) \right\}^{-1}. \quad (2.23)$$

- *The log-normal AFT model.* Assume the event times follow a log-normal distribution with parameters $\gamma_0 + \gamma_1^T \mathbf{a}_i + \gamma_2^T \mathbf{b}_i$ and σ , the survivor function is

$$S(t) = 1 - \Phi \left(\frac{\log t - \gamma_0 - \gamma_1^T \mathbf{a}_i - \gamma_2^T \mathbf{b}_i}{\sigma} \right). \quad (2.24)$$

Once the distribution is assumed, one may find the corresponding likelihood function by specifying the $f(t)$ and $S(t)$ in (2.16). Suppose the survival time ϵ_i in AFT model follows Weibull distribution, the pdf and survivor function of ϵ_i is

$$f(\epsilon) = \exp(\epsilon - e^\epsilon), \quad S(\epsilon) = \exp(-e^\epsilon), \quad (2.25)$$

that is

$$f(t) = \exp \left[\sigma^{-1} (\log t - \gamma_0 - \gamma_1^T \mathbf{a}_i - \gamma_2^T \mathbf{b}_i) \right] \quad (2.26)$$

$$- \exp \left(\sigma^{-1} (\log t - \gamma_0 - \gamma_1^T \mathbf{a}_i - \gamma_2^T \mathbf{b}_i) \right), \quad (2.27)$$

and

$$S(t) = \exp \left[- \exp \left(\sigma^{-1} (\log t - \gamma_0 - \gamma_1^T \mathbf{a}_i - \gamma_2^T \mathbf{b}_i) \right) \right]. \quad (2.28)$$

The log-likelihood function of the AFT model for subject i is

$$\begin{aligned} \ell(\boldsymbol{\gamma} | \mathbf{a}_i, \mathbf{b}_i) &= \sum_{i=1}^n (\delta_i \epsilon_i - e^{\epsilon_i}) \\ &= \sum_{i=1}^n \left[\delta_i \sigma^{-1} (\log T_i - \gamma_0 - \gamma_1^T \mathbf{a}_i - \gamma_2^T \mathbf{b}_i) \right. \\ &\quad \left. - \exp \left(\sigma^{-1} (\log T_i - \gamma_0 - \gamma_1^T \mathbf{a}_i - \gamma_2^T \mathbf{b}_i) \right) \right]. \end{aligned}$$

By integrating out the random effects $\mathbf{a}_i, \mathbf{b}_i$, we can obtain the log-likelihood function of γ .

$$\begin{aligned}
 \ell(\gamma) &= \sum_{i=1}^n \iint \ell(\gamma|\mathbf{a}_i, \mathbf{b}_i) f(\mathbf{a}_i|A) f(\mathbf{b}_i|B) d\mathbf{a}_i d\mathbf{b}_i & (2.29) \\
 &= \sum_{i=1}^n \iint (\delta_i \sigma^{-1} (\log T_i - \gamma_0 - \gamma_1^T \mathbf{a}_i - \gamma_2^T \mathbf{b}_i)) f(\mathbf{a}_i|A) f(\mathbf{b}_i|B) d\mathbf{a}_i d\mathbf{b}_i \\
 &\quad - \sum_{i=1}^n \iint [\exp(\sigma^{-1} (\log T_i - \gamma_0 - \gamma_1^T \mathbf{a}_i - \gamma_2^T \mathbf{b}_i))] f(\mathbf{a}_i|A) f(\mathbf{b}_i|B) d\mathbf{a}_i d\mathbf{b}_i \\
 &= \sigma^{-1} \sum_{i=1}^n \delta_i [\log T_i - \gamma_0 - \gamma_1^T E(\mathbf{a}_i) - \gamma_2^T E(\mathbf{b}_i)] \\
 &\quad - \sum_{i=1}^n \exp(\sigma^{-1} (\log T_i - \gamma_0)) \int \exp(-\sigma^{-1} \gamma_1^T \mathbf{a}_i) f(\mathbf{a}_i|A) d\mathbf{a}_i \\
 &\quad \times \int \exp(-\sigma^{-1} \gamma_2^T \mathbf{b}_i) f(\mathbf{b}_i|B) d\mathbf{b}_i \\
 &= \sigma^{-1} \sum_{i=1}^n \delta_i [\log T_i - \gamma_0 - \gamma_1^T E(\mathbf{a}_i) - \gamma_2^T E(\mathbf{b}_i)] - \sum_{i=1}^n \left[\exp\left(\frac{\log T_i - \gamma_0}{\sigma}\right) \right. \\
 &\quad \left. \times \exp\left(-\frac{\gamma_1^T E(\mathbf{a}_i)}{\sigma} + \frac{\gamma_1^T \Sigma(\mathbf{a}_i) \gamma_1}{\sigma^2} - \frac{\gamma_2^T E(\mathbf{b}_i)}{\sigma} + \frac{\gamma_2^T \Sigma(\mathbf{b}_i) \gamma_2}{\sigma^2}\right) \right]
 \end{aligned}$$

Then the statistical inference could be carried out based on the log-likelihood function (2.29) and the MLE theory applies.

2.4 Models for Nonignorable Missing Data

Lots of generally applicable statistical inference methods for incomplete data issues are likelihood-based, the idea of which is illustrated by a missing response case in the following. Suppose some values of the response variable \mathbf{y} are missing. Let $r_i = 1$ if y_i is missing and $r_i = 0$ if y_i is observed. Denote the latent variable that carries the information of the missingness as ω , and the joint distribution of response \mathbf{y} and missingness indicator \mathbf{r} , condition on covariates \mathbf{x} , is given by

$Pr(\mathbf{y}, \mathbf{r}|\mathbf{x}, \omega)$. Since this distribution can be rewritten as

$$Pr(\mathbf{y}, \mathbf{r}|\mathbf{x}, \omega) = \mathbf{Pr}(\mathbf{y}|\mathbf{x}, \theta(\omega))\mathbf{Pr}(\mathbf{r}|\mathbf{y}, \mathbf{x}, \psi(\omega)), \quad (2.30)$$

where $\theta(\omega), \psi(\omega)$ are subsets or transformation of ω , the likelihood function can be obtained if the full-data response model $Pr(\mathbf{y}|\mathbf{x}, \theta(\omega))$ and the missing mechanism $Pr(\mathbf{r}|\mathbf{y}, \mathbf{x}, \psi(\omega))$ are specified. The statistical inference can therefore be carried out based on by maximum likelihood theory.

Little (1992) reviewed the statistical methods of estimation in regression models with missing covariates. Six methods dealing with missing covariates are compared: complete-case methods, available-case methods, least squares on imputed data, maximum likelihood methods, Bayesian methods and multiple imputation. He suggested that the maximum likelihood method, Bayesian methods, and multiple imputation method perform well, and the maximum likelihood method is preferred in a large samples scenario and Bayesian methods or multiple imputation method are preferred in a small samples scenario.

In the specification of NLME model (2.1), all covariate values should be available, either observed or estimated, at the response measurement times t_{ij} . However, in practice this may not be the case since covariates may be measured at different time points or may be missing at t_{ij} . Model (2.10) automatically incorporates missing covariates when the missing data mechanism is ignorable (e.g., MAR or MCAR), since model (2.10) is fitted to the observed covariates which may be measured at different time points than the responses. When the missing covariates are nonignorable (i.e., the missingness may be related to the missing values), the missing data mechanism should be modeled.

Little (1995) gave an overview on modeling dropout mechanism in longitudinal studies. For example, we may assume that the missingness is related to the true

but unobserved covariate value and the previous missing status. That is, we may assume that

$$\text{logit}(Pr(r_{ij} = 1)) = \eta_0 + \eta_1 z_i^*(t_{ij}) + \eta_2 r_{i,j-1}, \quad (2.31)$$

where $z_i^*(t_{ij}) = \Psi_p(t_{ij})^T \boldsymbol{\alpha} + \Phi_q(t_{ij})^T \mathbf{a}_i$. We can then incorporate the missing data model into the joint likelihood for inference. Alternative missing data models may also be considered.

The statistical analysis in this thesis is mainly based on likelihood. When the missing covariates are nonignorable, we only need to add an additional model for the missing mechanism to the joint likelihood and then proceed the analysis. For simplicity of presentation, we only focus on ignorable missing covariates in the following sections.

Chapter 3

Simultaneous Inference

Having introduced several models in previous chapter for the statistical problems discussed in Section 1.1, the attentions are turned to how to analyze the data based on these models. As we emphasized in Section 1.2.1, the longitudinal data and the survival data are associated and the models share the same random effects. Thus, separate analysis of the longitudinal data and survival data may lead to biased results. This essential feature requires a simultaneous inference for those statistical models. In this chapter, we firstly introduce a simple method that can be implemented easily by standard softwares. Then we discuss two likelihood-based simultaneous inference methods, EM algorithm and Laplace approximation, which both have their own advantages and are widely applicable. Finally, we linearize the joint models and propose a new likelihood-based approximate simultaneous inference approach. The purpose of Laplace approximation and the linearization method is to reduce computation burden since the EM algorithm is computationally intensive.

A briefly review below is to remind the readers of the statistical models that are adopted in this thesis. Suppose n subjects are included a longitudinal study. Let y_{ij} be the response value for subject i at time t_{ij} , z_{ij} be the corresponding covariate, and T_i be the event time of subject i , $i = 1, \dots, n$. The statistical models that we introduced in Chapter 2 are reviewed as follows:

- Parametric NLME model for the longitudinal process:

$$y_{ij} = g(t_{ij}, \boldsymbol{\beta}_i) + e_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, n_i, \quad (3.1)$$

$$\boldsymbol{\beta}_i = h(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i) \quad (3.2)$$

$$\mathbf{b}_i \sim N(0, B), \quad \mathbf{e}_i | \mathbf{b}_i \sim N(0, \mu^2 I_i), \quad (3.3)$$

where $g(\cdot)$ and $h(\cdot)$ are given nonlinear functions, $\boldsymbol{\beta}_i$ are the subject-specific parameters and $\boldsymbol{\beta}$ are the fixed effects (population parameters), \mathbf{z}_i contain the covariates for subject i , $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})^T$ are the random effects between subjects, $\mathbf{e}_i = (e_{i1}, \dots, e_{in_i})^T$ are the random errors of within-subject measurements, B is the variance-covariance matrix of the random effects, and μ characterizes variance of within-subject measurements.

- Semiparametric LME model for the covariate measurement errors problem:

$$\mathbf{z}_i = \mathbf{z}_i^* + \boldsymbol{\xi}_i = \Psi_p^T \boldsymbol{\alpha} + \Phi_q^T \mathbf{a}_i + \boldsymbol{\xi}_i, \quad i = 1, \dots, n, \quad (3.4)$$

$$\mathbf{a}_i \sim N(0, A), \quad \boldsymbol{\xi}_i | \mathbf{a}_i \sim N(0, \lambda^2 I_i), \quad (3.5)$$

where $\Psi(\cdot)$ and $\Phi(\cdot)$ are some basis functions, $\boldsymbol{\alpha}$ are the fixed effects and \mathbf{a}_i are the random effects between subjects, $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{in_i})^T$ are the random errors of within-subject measurements, A is the variance-covariance matrix of the random effects, and λ characterizes variance of within-subject measurements.

- Parametric AFT model for the time-to-event process:

$$\log T_i = \gamma_0 + \boldsymbol{\gamma}_1^T \mathbf{a}_i + \boldsymbol{\gamma}_2^T \mathbf{b}_i + \sigma \epsilon_i, \quad i = 1, \dots, n. \quad (3.6)$$

where ϵ_i follows a standard extreme value distribution.

3.1 Two-Step Method

Most joint modelings of longitudinal process and time-to-event process are based on the assumption that two processes are linked with some shared parameters, which are usually the random effects representing variations between subjects. A simple statistical inference method is to make independent inference for one process in the first step to estimate the shared parameters, and then incorporate the estimated parameters into the inference for the other process in the second step. We call this method *Two-Step Method*.

3.1.1 Naive Two-Step Method

Since both the longitudinal process and the time-to-event process depend on the covariate \mathbf{z} , which were measured with errors, we need to solve the measurement errors problem (3.4) at the beginning. Laird and Ware (1982) gave the LME model formulation and proposed an EM algorithm to fit the model by treating the random effects as missing data. Lindstrom and Bates (1988) employed the linear feature of LME model and obtained the first two marginal moments for \mathbf{y}_i . The fixed effects β can be estimated by generalized least squares, and the variance components can be estimated by normal-theory pseudo-likelihood or restricted maximum pseudo-likelihood using a general framework based on Newton-Raphson method. Pinheiro and Bates (2002) described the covariance parametrizations of A , which could be more diverse and complex than our case. By fitting the LME model (3.4), we have the “true” covariate \mathbf{z}_i^* for modeling the longitudinal process and the estimate of random effects \mathbf{a}_i for modeling the time-to-event process.

For the NLME model (3.1), Lindstrom and Bates (1990) alternated between a penalized nonlinear least squares (PNLS) step and a linear mixed effects (LME) step. In the PNLS step, the current estimate of precision factor is held fixed, and the conditional estimates of the random effects \mathbf{b}_i and the fixed effects β are

obtained by minimizing a penalized nonlinear least squares objective function. The LME step then updates the estimate of the precision factor based on the first-order Taylor expansion of the likelihood function around the current estimates of β and \mathbf{b}_i respectively. Using this method to fit the longitudinal model (3.1), we can obtain the estimate of the fixed effects β and the random effects \mathbf{b}_i for fitting the time-to-event model.

The time-to-event process is analyzed by the AFT regression model. Having the random effects \mathbf{a}_i and \mathbf{b}_i estimated from above two models, we can fit the AFT regression model by standard statistical packages.

Thus, the naive two-step approach can be summarized as below.

- Covariate model inference. For a time-varying covariate, assume a flexible and robust model to estimate the true covariate values and the random effects between subjects.
- Nonlinear mixed effects model inference. Estimate the fixed effects, the random effects and the variance components with the covariate values measured with errors replaced by the true values.
- Time-to-event model inference. Fit the time-to-event model based on the estimated random effects from above two estimations.

3.1.2 Modified Two-Step Method

The naive two-step method is easily understandable and executable using standard statistical software. However, as pointed out by Ye et al. (2008) and Albert and Shih (2009), the naive two-step method may lead to two types of bias.

- The covariate trajectories may be different according to the censoring status of the event times. Thus, applying a single covariate model to all covariate data may lead to biased estimation for the first model.

- Inference for the second model that ignores the estimation uncertainty in the first model may also lead to biased results (e.g., under-estimating the standard errors).

The first bias, called bias from informative dropouts, may depend on the strength of the association between the longitudinal process and the time-to-event process. The second bias may depend on the magnitude of measurement errors in covariate. Therefore, we need to modify the naive two-step method to reduce these biases.

To incorporate the estimation uncertainty in the first step, we could adjust the standard errors of the parameters' estimates using bootstrap method by re-sampling subjects and keeping the repeated measurements of resampled subjects. After repeating above procedure B times, the standard errors for the estimates of the fixed parameters can be estimated by the sample variance-covariance matrix across the B bootstrap datasets, which is expected to be more reliable than the naive two-step method results, if the assumed models are correct.

However, the modified two-step method can not completely reduce the biases in the naive two-step method, because it still treats the two processes separately. In order to make reliable inference, simultaneous inference based on a joint model may be more appropriate. In order to provide an entirely decent inference method, making sure that the method considers two processes simultaneously is the first and most fundamental step to do. In the following sections, we consider some unified approaches based on the joint likelihood function given all observed longitudinal data and time-to-event data.

3.2 EM Algorithm

The two-step methods introduced in the previous section, naive or modified, fit the models separately with a simple “plug-in” strategy. EM Algorithm can be

3.2. EM Algorithm

used to estimate all parameters in the longitudinal model, the covariate model and the time-to-event model simultaneously based on the joint likelihood with the bias of the naive two-step method avoided. The joint likelihood approach is quite general and can be extended to statistical inference for more than two related models. Such a joint likelihood method provides efficient estimation if the assumed models are correct.

Specifically, let $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mu, \lambda, \sigma, A, B)$ be the parameters' collection of above three models, and let $f(\cdot)$ denotes the generic density functions and $F(\cdot)$ be the corresponding cumulative distribution functions. The joint log-likelihood given all the observed data can be written as:

$$\begin{aligned} \ell_o(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}, \boldsymbol{\rho}, \boldsymbol{\delta}) &= \sum_{i=1}^n \ell_o^{(i)}(\boldsymbol{\theta}|\mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i) \\ &= \sum_{i=1}^n \iint \log[f(\mathbf{y}_i|\mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mu) f(\mathbf{b}_i|B) f(\mathbf{z}_i|\mathbf{a}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) f(\mathbf{a}_i|A) \\ &\quad \times f(\rho_i|\mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \sigma)^{\delta_i} (1 - F(\rho_i|\mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \sigma))^{1-\delta_i}] d\mathbf{a}_i d\mathbf{b}_i. \end{aligned} \quad (3.7)$$

The EM algorithm (Dempster et al., 1977) is a powerful iterative algorithm to compute MLEs in a wide variety of situations, including missing data problems and random effects models. To get the MLE of $\boldsymbol{\theta}$ in our case, we may consider an EM algorithm by treating the unobservable random effects $(\mathbf{a}_i, \mathbf{b}_i)$ as “missing data”. EM algorithm contains an E-step, which computes the conditional expectation of the “complete data” log-likelihood given the observed data and parameter estimates, and an M-step, which maximizes the conditional expectation in the E-step to update the parameters' estimates.

3.2.1 E-Step

The “complete data” are $\{\mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i, \mathbf{a}_i, \mathbf{b}_i\}$, $i = 1, 2, \dots, n$, and the parameter estimates at the t th iteration is denoted as $\boldsymbol{\theta}^{(t)}$. The E-step at the $(t + 1)$ th

3.2. EM Algorithm

iteration can be written as:

$$\begin{aligned}
 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= \sum_{i=1}^n \iint [\log f(\mathbf{b}_i|B) + \log f(\mathbf{y}_i|\mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mu) + \log f(\mathbf{a}_i|A) \\
 &\quad + \log f(\mathbf{z}_i|\mathbf{a}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) + \log(1 - F(\rho_i|\mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \sigma))^{1-\delta_i} \\
 &\quad + \log f(\rho_i|\mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \sigma)^{\delta_i}] f(\mathbf{a}_i, \mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i; \boldsymbol{\theta}^{(t)}) d\mathbf{a}_i d\mathbf{b}_i, \quad (3.8)
 \end{aligned}$$

The computation of the conditional expectation $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ in the E-step is usually non-trivial for high-dimensional integrals. However, since the integral is an expectation with respect to $f(\mathbf{a}_i, \mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i; \boldsymbol{\theta}^{(t)})$, it can be evaluated based on the Monte Carlo EM of Wei and Tanner (1990) and Ibrahim et al. (2001a). That is to generate m_t samples from $f(\mathbf{a}_i, \mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i; \boldsymbol{\theta}^{(t)})$ and approximate the expectation $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ by its empirical mean, with the random effects $\mathbf{a}_i, \mathbf{b}_i$ replaced by the simulated values $\mathbf{a}_i^{(j)}, \mathbf{b}_i^{(j)}, j = 1, \dots, m_t$.

$$\begin{aligned}
 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= \frac{1}{m_t} \sum_{j=1}^{m_t} \sum_{i=1}^n [\log f(\mathbf{y}_i|\mathbf{a}_i^{(j)}, \mathbf{b}_i^{(j)}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mu) + \log f(\mathbf{b}_i^{(j)}|B) \\
 &\quad + \log f(\mathbf{z}_i|\mathbf{a}_i^{(j)}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) + \log f(\mathbf{a}_i^{(j)}|A) + \delta_i \log f(\rho_i|\mathbf{a}_i^{(j)}, \mathbf{b}_i^{(j)}; \boldsymbol{\gamma}, \sigma) \\
 &\quad + (1 - \delta_i) \log(1 - F(\rho_i|\mathbf{a}_i^{(j)}, \mathbf{b}_i^{(j)}; \boldsymbol{\gamma}, \sigma))] \\
 &= Q^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mu) + Q^{(2)}(B) + Q^{(3)}(\boldsymbol{\alpha}, \lambda) + Q^{(4)}(A) + Q^{(5)}(\boldsymbol{\gamma}, \sigma). \quad (3.9)
 \end{aligned}$$

We may choose m_0 as a large number and $m_t = m_{t-1} + m_{t-1}/k, (k \geq 1)$ in the t th iteration. Increasing m_t at each iteration may speed up the algorithm's convergence (Booth and Hobert, 1999).

To generate independent samples from $f(\mathbf{a}_i, \mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i; \boldsymbol{\theta}^{(t)})$, we may use Gibbs sampler (Gelfand and Smith, 1990) by sampling from the following full

conditionals iteratively:

$$\begin{aligned}
 & f(\mathbf{a}_i | \mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) \\
 \propto & f(\mathbf{a}_i; \boldsymbol{\theta}^{(t)}) f(\mathbf{y}_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) f(\mathbf{z}_i | \mathbf{a}_i; \boldsymbol{\theta}^{(t)}) f^*(\rho_i, \delta_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}), \\
 & f(\mathbf{b}_i | \mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i, \mathbf{a}_i; \boldsymbol{\theta}^{(t)}) \\
 \propto & f(\mathbf{b}_i; \boldsymbol{\theta}^{(t)}) f(\mathbf{y}_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) f^*(\rho_i, \delta_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}),
 \end{aligned}$$

where $f^*(\rho_i, \delta_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) = f(\rho_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)})^{\delta_i} (1 - F(\rho_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}))^{1-\delta_i}$. Thus, we can use the Gibbs sampling method to iteratively sample from the above full conditionals. After a burn-in period, we can obtain a sample from the conditional distribution $f(\mathbf{a}_i, \mathbf{b}_i | \mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i; \boldsymbol{\theta}^{(t)})$.

3.2.2 M-Step

The parameter $\boldsymbol{\theta}$ at the $(t + 1)$ th iteration can be estimated by maximizing $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)})$. Generally, $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)})$ is a nonlinear function and there is no closed-form expression for the estimate $\hat{\boldsymbol{\theta}}^{(t+1)}$. The maximizers could be obtained via standard numerical optimization procedures for complete data, such as the Newton-Raphson method or more advanced algorithms (Fletcher, 1987; Nocedal and Wright, 2006).

EM algorithm iterates between the two steps until convergence and the final estimates of $\boldsymbol{\theta}$ are regarded as a local maximum. The asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ can be obtained using well-known complete-data formulas (Louis, 1982), where the expectations in those formulas can be approximated by Monte-Carlo means. Specifically, note that the observed information matrix equals the expected complete information minus the missing information

$$I_{obs}(\hat{\boldsymbol{\theta}}) = I_{com}(\hat{\boldsymbol{\theta}}) - I_{mis}(\hat{\boldsymbol{\theta}})$$

Define

$$\begin{aligned}\dot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \sum_{i=1}^n \dot{Q}_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{j=1}^{m_t} \sum_{i=1}^n \frac{1}{m_t} S_{ij}(\boldsymbol{\theta}) \\ \ddot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \sum_{j=1}^{m_t} \sum_{i=1}^n \frac{1}{m_t} \frac{\partial S_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.\end{aligned}$$

Since the parameters are all distinct, above matrices are block diagonal. Then the asymptotic observed information matrix is

$$I_{obs}(\hat{\boldsymbol{\theta}}) = -\ddot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) - \left[\sum_{j=1}^{m_t} \sum_{i=1}^n \frac{1}{m_t} S_{ij}(\hat{\boldsymbol{\theta}}) S_{ij}^T(\hat{\boldsymbol{\theta}}) - \sum_{i=1}^n \frac{1}{m_t} \dot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) \dot{Q}^T(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) \right]. \quad (3.10)$$

The asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ can be approximated by

$$\text{Cov}(\hat{\boldsymbol{\theta}}) = I_{obs}^{-1}(\hat{\boldsymbol{\theta}}).$$

3.3 Laplace Approximation

The convergence of the foregoing Monte-Carlo EM algorithm depends on the dimension of the random effects $(\mathbf{a}_i, \mathbf{b}_i)$. If the dimension of the random effects $(\mathbf{a}_i, \mathbf{b}_i)$ is not small, for instance larger than 2, the computation of Monte-Carlo EM method can be extremely intensive since the method involves simulating large samples at each iteration. Moreover, the algorithm may not converge in some cases. Therefore, it is valuable to approximate the observed-data joint log-likelihood $\ell_o(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}, \rho, \delta)$ using the first-order Laplace approximation, which completely avoids integration and fastens the entire computation speed significantly.

3.3. Laplace Approximation

The “complete data” log-likelihood can be written as:

$$\begin{aligned}
 \ell_c(\boldsymbol{\theta}|\mathbf{a}, \mathbf{b}) &= \sum_{i=1}^n \ell_c^{(i)}(\boldsymbol{\theta}|\mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i) \\
 &= \sum_{i=1}^n \log[f(\mathbf{y}_i|\mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mu) f(\mathbf{b}_i|B) f(\mathbf{z}_i|\mathbf{a}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) f(\mathbf{a}_i|A) \\
 &\quad \times f(\rho_i|\mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \sigma)^{\delta_i} (1 - F(\rho_i|\mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \sigma))^{1-\delta_i}] \quad (3.11)
 \end{aligned}$$

Vonesh et al. (2002) and Lee et al. (2006) showed that the observed-data log-likelihood $\ell_o(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}, \rho, \delta)$ can be approximated by its first-order Laplace approximation

$$\tilde{\ell}_o(\boldsymbol{\theta}|\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) = \ell_c(\boldsymbol{\theta}|\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) - \frac{1}{2} \log \left| -\frac{1}{2\pi} \frac{\partial^2 \ell_c(\boldsymbol{\theta}|\mathbf{a}, \mathbf{b})}{\partial(\mathbf{a}, \mathbf{b})^2} \right|_{(\mathbf{a}, \mathbf{b}=\tilde{\mathbf{a}}, \tilde{\mathbf{b}})}, \quad (3.12)$$

where $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) = \{(\tilde{\mathbf{a}}_i, \tilde{\mathbf{b}}_i), i = 1, 2, \dots, n\}$ solve the equations

$$\partial \ell_c^{(i)}(\boldsymbol{\theta}|\mathbf{a}_i, \mathbf{b}_i) / \partial(\mathbf{a}_i, \mathbf{b}_i) = \mathbf{0}, \quad i = 1, 2, \dots, n. \quad (3.13)$$

Thus, an approximate estimate of $\boldsymbol{\theta}$ can be obtained by solving the equation

$$\partial \tilde{\ell}_o(\boldsymbol{\theta}|\tilde{\mathbf{a}}_i, \tilde{\mathbf{b}}_i) / \partial \boldsymbol{\theta} = \mathbf{0}. \quad (3.14)$$

Given starting values $\boldsymbol{\theta}^{(0)}$ and $(\mathbf{a}^{(0)}, \mathbf{b}^{(0)})$, we iterate the above procedures until convergence and obtain the approximate estimate $\hat{\boldsymbol{\theta}}$ of the MLE. To be specific, we carry out the following two steps at the t th iteration.

- Estimate the random effects $(\tilde{\mathbf{a}}^{(t)}, \tilde{\mathbf{b}}^{(t)})$ by solving equations (3.13),
- Update the fixed effect $\boldsymbol{\theta}^{(t+1)}$ by solving equation (3.14).

Laplace approximation method completely avoids integration, and thus is computationally attractive. The estimated random effects obtained this way can be

interpreted as the empirical Bayes estimates. The standard errors of the approximate estimates can be calculated based on the approximate formula for their variance-covariance matrix, which can be obtained in a similar way by eliminating the mean parameters and the random effects using a similar Laplace approximation. Specifically,

$$\text{Cov}(\hat{\boldsymbol{\theta}}) = - \left[\frac{\partial^2 \ell_o(\boldsymbol{\theta} | \tilde{\mathbf{a}}^*, \tilde{\mathbf{b}}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}^{-1} \quad (3.15)$$

where $\hat{\boldsymbol{\theta}}$ are the estimated fixed effects, and $\tilde{\mathbf{a}}^*, \tilde{\mathbf{b}}^*$ are the estimated random effects at the final iteration.

Vonesh et al. (2002) showed that the approximate estimate $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normally distributed when both the sample size and the number of measurements within each individual go to infinity

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_p \left[\max \left\{ (n)^{-\frac{1}{2}}, (\min n_i)^{-1} \right\} \right], \quad (3.16)$$

and

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow N(\mathbf{0}, \bar{I}(\boldsymbol{\theta})^{-1}) \quad (3.17)$$

where $\bar{I}(\boldsymbol{\theta}) = \lim_n \sum_i I_i(\boldsymbol{\theta})/n$ and $I_i(\boldsymbol{\theta})$ is the information matrix of subject i .

3.4 Linearization Method

Inspired by a popular approximate method for solving NLME model, we propose a linearization method for our joint models. Following the similar idea of Lindstrom and Bates (1990), we firstly linearize the NLME (3.1). Given the estimates of parameters $\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{a}}_i, \tilde{\mathbf{b}}_i$, we have the first-order Taylor expansion of $g_{ij} = g(t_{ij}, \boldsymbol{\beta}_i)$, which is a LME model

$$\tilde{\mathbf{y}}_i = W_i \boldsymbol{\alpha} + X_i \boldsymbol{\beta} + U_i \mathbf{a}_i + V_i \mathbf{b}_i + \mathbf{e}_i, \quad (3.18)$$

3.4. Linearization Method

where

$$\begin{aligned}
 W_i &= (\mathbf{w}_{i1}, \dots, \mathbf{w}_{in_i})^T, \quad \mathbf{w}_{ij} = \partial g_{ij} / \partial \boldsymbol{\alpha}, \\
 X_i &= (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})^T, \quad \mathbf{x}_{ij} = \partial g_{ij} / \partial \boldsymbol{\beta}, \\
 U_i &= (\mathbf{u}_{i1}, \dots, \mathbf{u}_{in_i})^T, \quad \mathbf{u}_{ij} = \partial g_{ij} / \partial \mathbf{a}_i, \\
 V_i &= (\mathbf{v}_{i1}, \dots, \mathbf{v}_{in_i})^T, \quad \mathbf{v}_{ij} = \partial g_{ij} / \partial \mathbf{b}_i, \\
 \tilde{\mathbf{y}}_i &= \mathbf{y}_i - \mathbf{g}_i(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{a}}_i, \tilde{\mathbf{b}}_i) + W_i \tilde{\boldsymbol{\alpha}} + X_i \tilde{\boldsymbol{\beta}} + U_i \tilde{\mathbf{a}}_i + V_i \tilde{\mathbf{b}}_i,
 \end{aligned}$$

with all the partial derivatives being evaluated at $(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{a}}_i, \tilde{\mathbf{b}}_i)$. Now we have two LMEs and a time-to-event model, and we adopt the strategy of EM algorithm to estimate the parameters.

To calculate (3.8), we need to estimate the random effects $\mathbf{a}_i, \mathbf{b}_i$ at first. The general EM approach samples $\mathbf{a}_i, \mathbf{b}_i$ from the conditional distribution of the random effects $f(\mathbf{a}_i, \mathbf{b}_i | \mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i; \boldsymbol{\theta}^{(t)})$ by iteratively updating their full conditionals:

$$\begin{aligned}
 & f(\mathbf{a}_i | \mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) \\
 \propto & f(\mathbf{a}_i; \boldsymbol{\theta}^{(t)}) f(\mathbf{y}_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) f(\mathbf{z}_i | \mathbf{a}_i; \boldsymbol{\theta}^{(t)}) f^*(\rho_i, \delta_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}), \\
 & f(\mathbf{b}_i | \mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i, \mathbf{a}_i; \boldsymbol{\theta}^{(t)}) \\
 \propto & f(\mathbf{b}_i; \boldsymbol{\theta}^{(t)}) f(\mathbf{y}_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) f^*(\rho_i, \delta_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}),
 \end{aligned}$$

where $f^*(\rho_i, \delta_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) = f(\rho_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)})^{\delta_i} (1 - F(\rho_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}))^{1-\delta_i}$.

Since we have

$$\begin{aligned}
f(\mathbf{a}_i; \boldsymbol{\theta}^{(t)}) &\propto |\boldsymbol{\Sigma}_{\mathbf{a}_{it}}|^{-\frac{n_i}{2}} \exp \left[-\frac{1}{2} (\mathbf{a}_i - \boldsymbol{\nu}_{\mathbf{a}_{it}})^T \boldsymbol{\Sigma}_{\mathbf{a}_{it}}^{-1} (\mathbf{a}_i - \boldsymbol{\nu}_{\mathbf{a}_{it}}) \right] \\
&\propto |\boldsymbol{\Sigma}_{\mathbf{a}_{it}}|^{-\frac{n_i}{2}} \exp \left[-\frac{1}{2} \mathbf{a}_i^T \boldsymbol{\Sigma}_{\mathbf{a}_{it}}^{-1} \mathbf{a}_i + \boldsymbol{\nu}_{\mathbf{a}_{it}}^T \boldsymbol{\Sigma}_{\mathbf{a}_{it}}^{-1} \mathbf{a}_i - \frac{1}{2} \boldsymbol{\nu}_{\mathbf{a}_{it}}^T \boldsymbol{\Sigma}_{\mathbf{a}_{it}}^{-1} \boldsymbol{\nu}_{\mathbf{a}_{it}} \right], \\
f(\mathbf{b}_i; \boldsymbol{\theta}^{(t)}) &\propto |\boldsymbol{\Sigma}_{\mathbf{b}_{it}}|^{-\frac{n_i}{2}} \exp \left[-\frac{1}{2} (\mathbf{b}_i - \boldsymbol{\nu}_{\mathbf{b}_{it}})^T \boldsymbol{\Sigma}_{\mathbf{b}_{it}}^{-1} (\mathbf{b}_i - \boldsymbol{\nu}_{\mathbf{b}_{it}}) \right] \\
&\propto |\boldsymbol{\Sigma}_{\mathbf{b}_{it}}|^{-\frac{n_i}{2}} \exp \left[-\frac{1}{2} \mathbf{b}_i^T \boldsymbol{\Sigma}_{\mathbf{b}_{it}}^{-1} \mathbf{b}_i + \boldsymbol{\nu}_{\mathbf{b}_{it}}^T \boldsymbol{\Sigma}_{\mathbf{b}_{it}}^{-1} \mathbf{b}_i - \frac{1}{2} \boldsymbol{\nu}_{\mathbf{b}_{it}}^T \boldsymbol{\Sigma}_{\mathbf{b}_{it}}^{-1} \boldsymbol{\nu}_{\mathbf{b}_{it}} \right], \\
f(\mathbf{z}_i | \mathbf{a}_i, \boldsymbol{\theta}^{(t)}) &\propto \lambda_{it}^{-n_i} \exp \left[-\frac{1}{2\lambda_{it}^2} (\mathbf{z}_i - \boldsymbol{\psi}_i^T \boldsymbol{\alpha}_t - \boldsymbol{\phi}_i^T \mathbf{a}_i)^T (\mathbf{z}_i - \boldsymbol{\psi}_i^T \boldsymbol{\alpha}_t - \boldsymbol{\phi}_i^T \mathbf{a}_i) \right] \\
&\propto \lambda_{it}^{-n_i} \exp \left[-\frac{1}{2\lambda_{it}^2} \mathbf{a}_i^T \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T \mathbf{a}_i + \frac{1}{\lambda_{it}^2} (\mathbf{z}_i - \boldsymbol{\psi}_i^T \boldsymbol{\alpha}_t)^T \boldsymbol{\phi}_i^T \mathbf{a}_i \right. \\
&\quad \left. - \frac{1}{2\lambda_{it}^2} (\mathbf{z}_i - \boldsymbol{\psi}_i^T \boldsymbol{\alpha}_t)^T (\mathbf{z}_i - \boldsymbol{\psi}_i^T \boldsymbol{\alpha}_t) \right], \\
f(\mathbf{y}_i | \mathbf{a}_i, \mathbf{b}_i, \boldsymbol{\theta}^{(t)}) &\propto \mu_{it}^{-n_i} \exp \left[-\frac{1}{2\mu_{it}^2} (\tilde{\mathbf{y}}_{it} - W_{it} \boldsymbol{\alpha}_t - X_{it} \boldsymbol{\beta}_t - U_{it} \mathbf{a}_i - V_{it} \mathbf{b}_i)^T \right. \\
&\quad \left. \times (\tilde{\mathbf{y}}_{it} - W_{it} \boldsymbol{\alpha}_t - X_{it} \boldsymbol{\beta}_t - U_{it} \mathbf{a}_i - V_{it} \mathbf{b}_i) \right] \\
&\propto \mu_{it}^{-n_i} \exp \left[-\frac{1}{2\mu_{it}^2} \mathbf{a}_i^T U_{it}^T U_{it} \mathbf{a}_i - \frac{1}{2\mu_{it}^2} \mathbf{b}_i^T V_{it}^T V_{it} \mathbf{b}_i \right. \\
&\quad + \frac{1}{\mu_{it}^2} (\tilde{\mathbf{y}}_{it} - W_{it} \boldsymbol{\alpha}_t - X_{it} \boldsymbol{\beta}_t - V_{it} \mathbf{b}_i)^T U_{it} \mathbf{a}_i \\
&\quad + \frac{1}{\mu_{it}^2} (\tilde{\mathbf{y}}_{it} - W_{it} \boldsymbol{\alpha}_t - X_{it} \boldsymbol{\beta}_t - U_{it} \mathbf{a}_i)^T V_{it} \mathbf{b}_i \\
&\quad \left. - \frac{1}{2\mu_{it}^2} (\tilde{\mathbf{y}}_{it} - W_{it} \boldsymbol{\alpha}_t - X_{it} \boldsymbol{\beta}_t)^T (\tilde{\mathbf{y}}_{it} - W_{it} \boldsymbol{\alpha}_t - X_{it} \boldsymbol{\beta}_t) \right],
\end{aligned}$$

The likelihood function given the event time data is linearized as follows:

$$\begin{aligned}
f^*(\rho_i, \delta_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) &\cong \propto \exp \left[-\frac{1}{2\sigma_t^2} \mathbf{a}_i^T (\gamma_{1t} \gamma_{1t}^T) \mathbf{a}_i + \frac{1}{\sigma_t^2} (\log T_i - \gamma_{0t} + \delta_i) \gamma_{1t}^T \mathbf{a}_i \right. \\
&\quad - \frac{1}{2\sigma_t^2} \mathbf{b}_i^T (\gamma_{2t} \gamma_{2t}^T) \mathbf{b}_i + \frac{1}{\sigma_t^2} (\log T_i - \gamma_{0t} + \delta_i) \gamma_{2t}^T \mathbf{b}_i \\
&\quad \left. - \frac{1}{2\sigma_t^2} (\log T_i - \gamma_{0t} + \delta_i)^2 \right].
\end{aligned}$$

Then we may obtain the approximate explicit form of the distribution of the ran-

3.4. Linearization Method

dom effects, which follows a multivariate normal distribution. Therefore, the full conditionals of the random effects can be approximated by the following multivariate normal distributions:

$$\begin{aligned}
& f(\mathbf{a}_i | \mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) \\
& \propto f(\mathbf{a}_i; \boldsymbol{\theta}^{(t)}) f(\mathbf{y}_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) f(\mathbf{z}_i | \mathbf{a}_i; \boldsymbol{\theta}^{(t)}) f^*(\rho_i, \delta_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) \\
& \propto \exp \left[-\frac{1}{2} (\mathbf{a}_i - \boldsymbol{\nu}_{\mathbf{a}_i(t+1)})^T \boldsymbol{\Sigma}_{\mathbf{a}_i(t+1)}^{-1} (\mathbf{a}_i - \boldsymbol{\nu}_{\mathbf{a}_i(t+1)}) \right] \\
& \sim \text{MVN}(\boldsymbol{\nu}_{\mathbf{a}_i(t+1)}, \boldsymbol{\Sigma}_{\mathbf{a}_i(t+1)}), \tag{3.19}
\end{aligned}$$

$$\begin{aligned}
& f(\mathbf{b}_i | \mathbf{y}_i, \mathbf{z}_i, \rho_i, \delta_i, \mathbf{a}_i; \boldsymbol{\theta}^{(t)}) \\
& \propto f(\mathbf{b}_i; \boldsymbol{\theta}^{(t)}) f(\mathbf{y}_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) f^*(\rho_i, \delta_i | \mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(t)}) \\
& \propto \exp \left[-\frac{1}{2} (\mathbf{b}_i - \boldsymbol{\nu}_{\mathbf{b}_i(t+1)})^T \boldsymbol{\Sigma}_{\mathbf{b}_i(t+1)}^{-1} (\mathbf{b}_i - \boldsymbol{\nu}_{\mathbf{b}_i(t+1)}) \right] \\
& \sim \text{MVN}(\boldsymbol{\nu}_{\mathbf{b}_i(t+1)}, \boldsymbol{\Sigma}_{\mathbf{b}_i(t+1)}), \tag{3.20}
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{a}_i(t+1)}^{-1} &= \boldsymbol{\Sigma}_{\mathbf{a}_i t}^{-1} + \frac{\boldsymbol{\phi}_i \boldsymbol{\phi}_i^T}{\lambda_{it}^2} + \frac{U_{it}^T U_{it}}{\mu_{it}^2} + \frac{\gamma_{1t} \gamma_{1t}^T}{\sigma_t^2}, \\
\boldsymbol{\nu}_{\mathbf{a}_i(t+1)}^T \boldsymbol{\Sigma}_{\mathbf{a}_i(t+1)}^{-1} &= \boldsymbol{\nu}_{\mathbf{a}_i t}^T \boldsymbol{\Sigma}_{\mathbf{a}_i t}^{-1} + \frac{1}{\lambda_{it}^2} (\mathbf{z}_i - \boldsymbol{\psi}_i^T \boldsymbol{\alpha}_t)^T \boldsymbol{\phi}_i^T + \frac{1}{\sigma_t^2} (\log T_i - \gamma_{0t} + \delta_i) \gamma_{1t}^T \\
&\quad + \frac{1}{\mu_{it}^2} (\tilde{\mathbf{y}}_{it} - W_{it} \boldsymbol{\alpha}_t - X_{it} \boldsymbol{\beta}_t - V_{it} \boldsymbol{\nu}_{\mathbf{b}_i t})^T U_{it}, \\
\boldsymbol{\Sigma}_{\mathbf{b}_i(t+1)}^{-1} &= \boldsymbol{\Sigma}_{\mathbf{b}_i t}^{-1} + \frac{V_{it}^T V_{it}}{\mu_{it}^2} + \frac{\gamma_{2t} \gamma_{2t}^T}{\sigma_t^2}, \\
\boldsymbol{\nu}_{\mathbf{b}_i(t+1)}^T \boldsymbol{\Sigma}_{\mathbf{b}_i(t+1)}^{-1} &= \boldsymbol{\nu}_{\mathbf{b}_i t}^T \boldsymbol{\Sigma}_{\mathbf{b}_i t}^{-1} + \frac{1}{\sigma_t^2} (\log T_i - \gamma_{0t} + \delta_i) \gamma_{2t}^T \\
&\quad + \frac{1}{\mu_{it}^2} (\tilde{\mathbf{y}}_{it} - W_{it} \boldsymbol{\alpha}_t - X_{it} \boldsymbol{\beta}_t - U_{it} \boldsymbol{\nu}_{\mathbf{a}_i t})^T V_{it}.
\end{aligned}$$

Thus, we may have an estimate of the distribution of the random effects \mathbf{a}_i and \mathbf{b}_i at the $(t+1)$ th iteration.

3.4.1 Framework I

With the estimated distributions of \mathbf{a}_i and \mathbf{b}_i , we may return to Section 3.2.1 and work on the log-likelihood function (3.8). If the random effects \mathbf{a}_i and \mathbf{b}_i can be integrated out, we may simply adopt the EM framework to complete the linearization method. To be specific, at the t th iteration, we need to:

- Linearize the joint model at the current steps $\tilde{\mathbf{a}}_i^{(t)}, \tilde{\mathbf{b}}_i^{(t)}$ and $\hat{\boldsymbol{\theta}}^{(t)}$;
- Estimate the random effects $(\tilde{\mathbf{a}}_i^{(t+1)}, \tilde{\mathbf{b}}_i^{(t+1)})$ as (3.19);
- Obtain the expectation of the joint log-likelihood function;
- Update $\boldsymbol{\theta}^{(t+1)}$ by maximizing the expectation of the observed data joint log-likelihood.

We repeat the above steps until the algorithm converges. The standard errors of the MLE of $\boldsymbol{\theta}$ can be calculated based on the asymptotic observed information matrix (3.10).

3.4.2 Framework II

If we can not obtain an explicit form of the integral (3.8), we may simply consider the means of the estimated distributions of the random effects \mathbf{a}_i and \mathbf{b}_i , $\boldsymbol{\nu}_{\mathbf{a}_i(t+1)}$ and $\boldsymbol{\nu}_{\mathbf{b}_i(t+1)}$, as their empirical Bayesian estimates. Then the fixed effects can be updated in the same way as Laplace approximation method. That is to say, we carry out the following steps at the t th iteration.

- Linearize the joint model at current iterate points $\tilde{\mathbf{a}}_i^{(t)}, \tilde{\mathbf{b}}_i^{(t)}$ and $\hat{\boldsymbol{\theta}}^{(t)}$;
- Estimate the random effects $(\tilde{\mathbf{a}}_i^{(t+1)}, \tilde{\mathbf{b}}_i^{(t+1)})$ by (3.19),
- Update $\boldsymbol{\theta}^{(t+1)}$ by solving equations (3.14) with $\boldsymbol{\nu}_{\mathbf{a}_i(t+1)}$ and $\boldsymbol{\nu}_{\mathbf{b}_i(t+1)}$ being estimates of $\tilde{\mathbf{a}}_i^{(t)}$ and $\tilde{\mathbf{b}}_i^{(t)}$.

3.4. Linearization Method

Similar to Laplace approximation method, the standard errors of the approximate MLE of $\boldsymbol{\theta}$ can be approximated by (3.15).

Chapter 4

Data Analysis

4.1 Data Description

We consider an AIDS clinical trial, which included 46 HIV infected patients being treated with an anti-HIV treatment. The viral loads were measured on days 0, 2, 7, 10, 14, 21, 28 and weeks 8, 12, 24, and 48 after initiation of the treatment. The CD4 and CD8 cell counts and other variables were also recorded throughout the study. The lower bound of the detectable limit of the viral load is 100 copies/ml, and 40 out of all 361 viral load measurements (11%) are below the detectable limit. For simplicity, we imputed the viral loads below the assay's lower detection limit 100 copies/ml with a half of the limit. The number of viral load measurements for each individual varies from 4 to 10. Figure 4.1 shows the trajectories of viral load, the CD4 and the CD4/CD8 of six randomly selected patients.

One of the primary objectives of this study is to assess the antiviral activity of the treatment in the first 12 weeks. Meanwhile, the association between the viral load and the time to the first decline of the ratio of CD4 and CD8 is of interest.

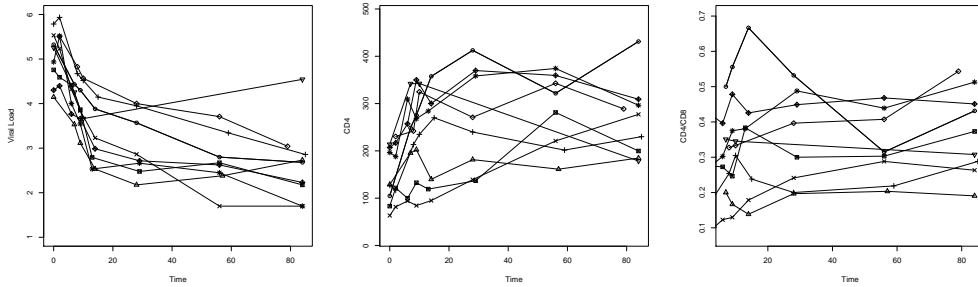
4.2 Model Specification

Based on previous researches (Wu and Ding, 1999; Wu, 2002; Wu et al., 2010), we consider the following HIV viral load dynamic model

$$\begin{aligned}
 y_{ij} &= \log_{10}(P_{1i}e^{-\lambda_{1i}t_{ij}} + P_{2i}e^{-\lambda_{2ij}t_{ij}}) + e_{ij} & (4.1) \\
 \log(P_{1i}) &= \beta_1 + b_{i1}, \quad \lambda_{1i} = \beta_2 + b_{i2}, \\
 \log(P_{2i}) &= \beta_3 + b_{i3}, \quad \lambda_{2ij} = \beta_4 + \beta_5 z_{ij}^* + b_{i4}.
 \end{aligned}$$

where y_{ij} is the log10-transformation of the viral load of patient i at time t_{ij} , P_{1i} and P_{2i} are subject-specific viral load baselines, λ_{1i} and λ_{2ij} subject-specific first-phase and second-phase viral load decay rates, z_{ij}^* is the “true” (but unobserved) CD4 count. It is known that CD4 counts are usually measured with substantial errors. Thus we assume that λ_{2ij} is related to the true CD4 value z_{ij}^* rather than the observed CD4 value z_{ij} . To avoid very small (large) estimates, which may be unstable, we standardize the CD4 counts and rescale the original time t (in days) so that the new time scale is between 0 and 1.

Figure 4.1: Longitudinal Data Trajectories



As we can see from Figure 4.1, the CD4 trajectories are complicated and the between-patient variations in CD4 trajectories are quite large, thus we consider a

4.2. Model Specification

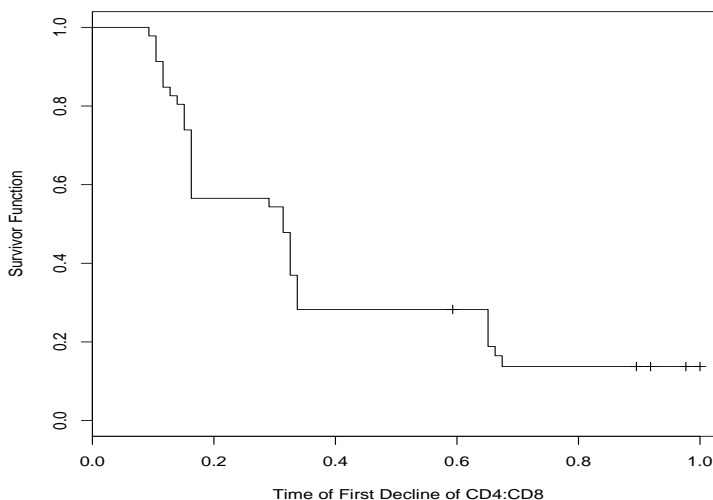
nonparametric linear mixed effects model to empirically describe the CD4 trajectories. We use linear combinations of natural cubic splines with percentile-based knots to approximate $r(t)$ and $h_i(t)$. We set $\psi_0(t) = \phi_0(t) = 1$ and take the same natural cubic splines in the approximations with $q \leq p$ in order to decrease the dimension of the random effects. AIC and BIC criteria are used to determine that $p = q = 2$, which leads to the following model for the CD4 process.

$$z_{ij} = (\alpha_1 + a_{i1})\phi_0(t_{ij}) + (\alpha_2 + a_{i2})\phi_1(t_{ij}) + (\alpha_3 + a_{i3})\phi_2(t_{ij}) + \xi_{ij}. \quad (4.2)$$

where z_{ij} is the observed CD4 value at time t_{ij} , $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are the basis functions, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T$ are the fixed parameters, and $\mathbf{a}_i = (a_{i1}, a_{i2}, a_{i3})^T$ are the random effects.

Let T_i be the time to the first decline of CD4/CD8 ratio of subject i . In AIDS

Figure 4.2: Estimate of Survivor Function



studies, it is of great interest to see if the time of the first decline of CD4/CD8 is related to the subject characteristics of the viral load and the CD4 trajectories,

represented by the random effects in the viral load model and the CD4 model. Therefore, we consider the following time-to-event model for T_i

$$\log T_i = \gamma_0 + \gamma_1 a_{i1} + \gamma_2 a_{i2} + \gamma_3 b_{i2} + \gamma_4 b_{i4} + \sigma \epsilon_i. \quad (4.3)$$

where ϵ_i follows standard extreme value distribution. The random effects b_{i2} and b_{i4} , which represent between-subject variations of viral decay rates, might be predictive for event times. The random effects a_{i1} and a_{i2} , which bear the main features of between-subject variations of CD4 trajectories, are also included in the parametric AFT model (4.3).

4.3 Analysis and Results

We estimate the models' parameters using two-step method (TS), modified two-step method (MTS), Laplace approximation method (LAP) and linearization method (LIN). TS and MTS are based on standard R algorithms in `library(nlme)` and `library(survival)`. The estimation results of TS and MTS are listed along with those of LAP and LIN in Table 4.1. Figure 4.3 and 4.4 check the normality assumptions of the residual and random effects of the NLME model. Figure 4.5 and 4.6 check the normality assumptions of the residual and random effects of the measurement error model. It seems that no assumption is severely violated.

LAP and LIN are iterative algorithms with convergence being achieved when the maximum percentage change of all estimates is less than 5% in two consecutive iterations. The parameters' estimates of TS method are taken as the initial values for two iterative algorithms LAP and LIN.

Based on the analysis results listed in Table 4.1, we have the following observations.

- Four inference methods provide similar parameter estimates for the NLME

4.3. Analysis and Results

Figure 4.3: Residuals Plots I of NLME Model

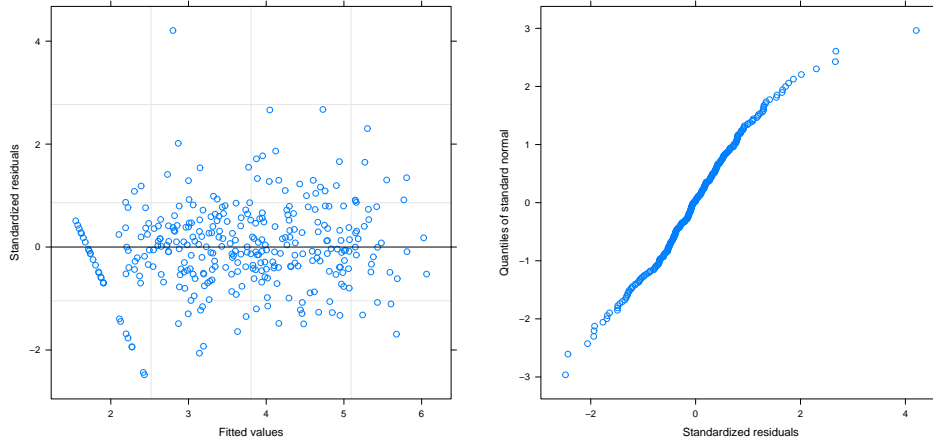
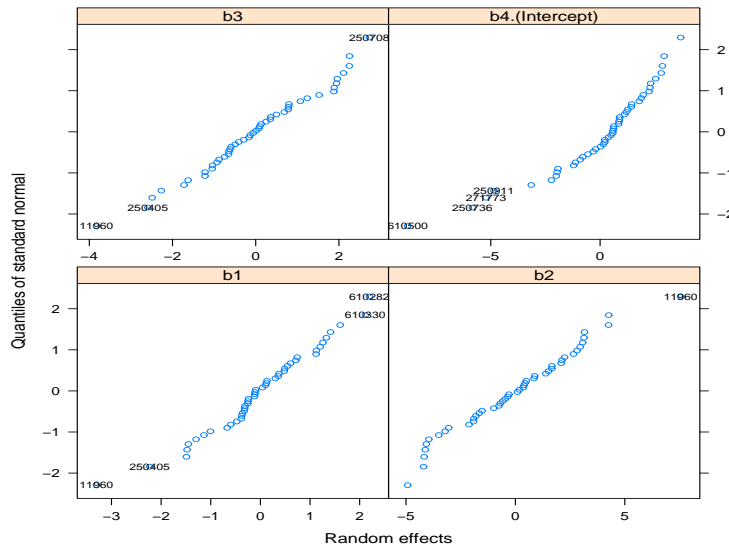


Figure 4.4: Residuals Plots II of NLME Model



model of the viral load data. LAP and LIN agree that all parameters of β are significant to the viral load, while TS and MTS suggest that β_5 might be insignificant.

4.3. Analysis and Results

Figure 4.5: Residuals Plots I of Covariate Measurement Errors Model

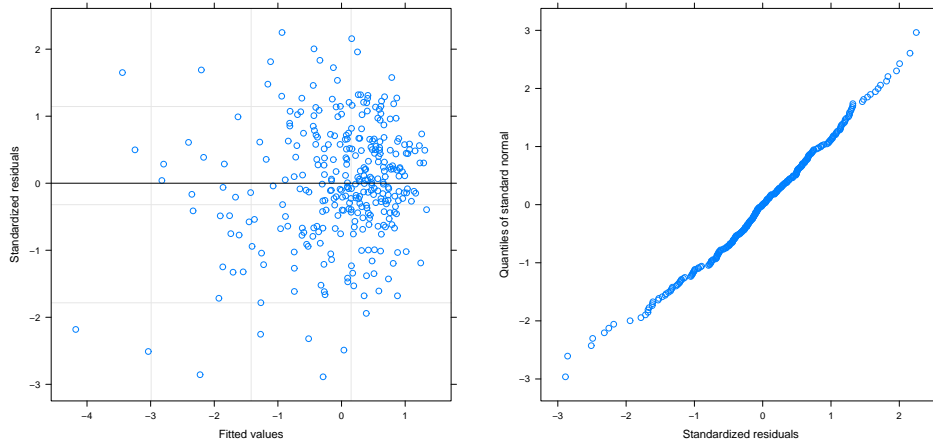
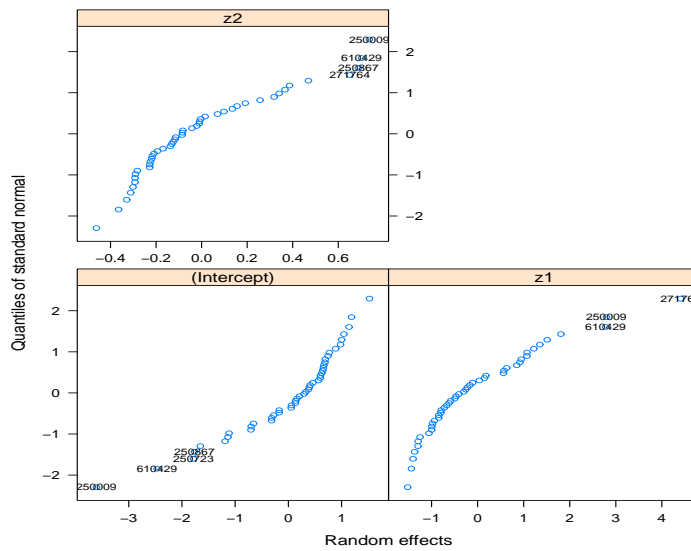


Figure 4.6: Residuals Plots II of Covariate Measurement Errors Model



- The estimation results of the time-to-event process obtained from four methods differ from each others. Generally speaking, TS and MTS suggest that none random effect is significant to the event time, while two simultaneous

4.3. Analysis and Results

Table 4.1: Estimation Results

Parameter	TS		MTS		LAP		LIN	
	Est	S.E.	Est	S.E.	Est	S.E.	Est	S.E.
β_1	11.75	0.19	11.76	0.19	11.78	0.09	11.80	0.06
β_2	29.84	1.47	30.22	1.69	29.25	1.31	30.03	0.87
β_3	7.48	0.29	7.43	0.32	7.22	0.15	7.40	0.10
β_4	1.34	0.51	1.31	0.64	1.43	0.23	1.18	0.15
β_5	0.63	0.46	0.52	0.70	0.42	0.16	0.55	0.10
μ	0.29	NA	0.29	NA	0.29	NA	0.22	NA
γ_0	-0.81	0.12	-0.82	0.01	-0.97	0.11	-0.71	0.20
γ_1	0.19	0.24	0.16	0.52	0.55	0.37	0.22	0.19
γ_2	-0.03	0.21	-0.01	0.52	-1.03	0.52	0.01	0.13
γ_3	-0.03	0.05	-0.02	0.11	-0.05	0.02	-0.03	0.04
γ_4	0.08	0.05	0.04	0.10	0.12	0.03	0.05	0.02
σ	0.75	NA	0.76	NA	0.31	NA	0.75	NA

approaches indicate that b_{i4} might be significant. This discovery distinguished the simultaneous approaches from the TS based methods.

- MTS provides very close parameter estimates as TS for both the NLME model and the measurement errors model, but with greater standard errors. However, by inferring those three models simultaneously, LAP and LIN provided parameter estimates with smaller standard errors.

Based on this data analysis result, we may conclude that:

- For the analysis of the longitudinal process based on NLME model, the results obtained from two simultaneous inference methods agree with previous researches (Ding and Wu, 2001; Wu et al., 2008, 2010) in terms of the significance conclusion of the parameters. However, some estimated values of our analysis are different from previous researches, since we only considered the observations in the first three months in our study.
- Two simultaneous inference methods suggest that CD4 is significant to the

viral load, while TS and MTS miss this reasonable and important property. According to the conclusions from LAP and LIN, the higher the CD4 is, the lower viral load the subject might have.

- Two simultaneous inference methods also indicate that the viral decay rate is significant to the first decline time of the CD4/CD8 ratio. The lower the viral load is, the later the CD4/CD8 may firstly decline. TS based methods fail again in detecting this important feature.

The discrepancies between these data analysis results obtained from different methods are important in terms of both biological interpretation and statistical methodology development. To study if the performance of the simultaneous inference methods exceed the TS based methods where longitudinal process and time-to-event process are truly connected, we need a simulation study to compare these methods.

Chapter 5

Simulation

A simulation study is conducted to evaluate the performance of the three methods: two-step method (TS), Laplace approximation method (LAP) and linearization method (LIN). Modified two-step method (MTS) is omitted since MTS performs quite similar to TS only expect that MTS generally yields in greater variances of the estimates.

5.1 Simulation Design

We simulate the longitudinal data and covariate data based on model (4.1) and (4.2). The event times are simulated based on (4.3) with 20% randomly censored subject. While having the parameter β , α and γ fixed, we compare these inference methods based on their estimation performances in four scenarios with different settings of

- Number of subjects,
- Numbers of repeated measurements,
- Between-subject variance,
- Within-subject variance.

The simulations were implemented on Dell R900 server with 16 CPUs and 16-32Gb RAM. The estimation performance includes the standard errors of the parameters' estimates and the corresponding bias and MSE, along with the average running

5.2. Simulation Results

time. Since we are more interested in the NLME model and the time-to-event model estimation, the estimation results of the covariate model are omitted in the following section.

5.2 Simulation Results

5.2.1 Number of Subjects

To examine how these estimation methods perform with different numbers of subjects, we simulate datasets with $n = 60$ and $n = 120$. The estimation results based on $n = 60$ are reported in Table 5.1. The estimation results based on $n = 120$ are reported in Table 5.2.

Table 5.1: Simulation Results: $n = 60$

Parameter	TS			LAP			LIN		
	Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
$\beta_1 = 12$	0.02	0.18	0.03	-0.04	0.20	0.04	-0.09	0.06	0.01
$\beta_2 = 30$	-1.13	2.16	5.95	-1.39	1.69	4.79	-0.93	1.25	2.41
$\beta_3 = 7$	-0.02	0.20	0.04	0.12	0.23	0.07	-0.14	0.17	0.05
$\beta_4 = -2$	-0.05	0.26	0.07	0.15	0.37	0.16	-0.01	0.23	0.05
$\beta_5 = 3$	-0.00	0.13	0.02	-0.30	0.21	0.13	-0.48	0.08	0.23
$\mu = 0.3$	-0.00	0.01	0.00	0.12	0.04	0.02	-0.01	0.02	0.00
$\gamma_0 = -1$	2.19	0.40	4.95	0.20	0.38	0.19	0.15	0.10	0.03
$\gamma_1 = 1$	0.01	0.43	0.18	-0.80	1.53	2.97	-0.91	0.18	0.87
$\gamma_2 = 1$	-0.01	0.48	0.23	0.52	1.10	1.47	-0.88	0.24	0.84
$\gamma_3 = 1$	-1.00	0.28	1.08	0.14	0.36	0.15	-0.07	0.21	0.05
$\gamma_4 = 1$	-0.02	0.40	0.16	-0.08	0.36	0.14	-0.10	0.21	0.05
$\sigma = 0.2$	2.19	0.30	4.88	0.06	0.23	0.06	0.08	0.03	0.01
Time (Sec)	680.42			180.47			13.39		

For Table 5.1, we have the following observations.

- Three methods provided generally decent estimations for β , except that the estimate of β_5 provided by LIN seems to be biased.

5.2. Simulation Results

- Though TS estimated $\hat{\gamma}_1, \hat{\gamma}_2$ and $\hat{\gamma}_4$ with less biases, its performance in the estimation of γ_0, γ_3 and σ is poor. Comparatively, LAP and LIN are more likely to provide better estimation of γ , at least for γ_0 .

The results in Table 5.1 are regarded as the baseline for the following comparisons.

Table 5.2: Simulation Results: $n = 120$

Parameter	TS			LAP			LIN		
	Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
$\beta_1 = 12$	0.05	0.14	0.02	0.00	0.11	0.01	-0.10	0.04	0.01
$\beta_2 = 30$	-0.94	1.55	3.28	-1.14	0.96	2.20	-1.09	0.92	2.03
$\beta_3 = 7$	-0.01	0.18	0.03	0.14	0.14	0.04	-0.12	0.16	0.04
$\beta_4 = -2$	-0.04	0.19	0.04	0.11	0.26	0.08	0.00	0.22	0.05
$\beta_5 = 3$	0.00	0.11	0.01	-0.19	0.13	0.05	-0.36	0.07	0.13
$\mu = 0.3$	0.00	0.01	0.00	0.09	0.03	0.01	-0.00	0.01	0.00
$\gamma_0 = -1$	2.01	0.29	4.13	-0.07	0.06	0.01	-0.05	0.09	0.01
$\gamma_1 = 1$	0.04	0.37	0.14	0.00	0.14	0.02	-0.01	0.19	0.04
$\gamma_2 = 1$	0.00	0.42	0.18	-0.00	0.11	0.01	-0.06	0.25	0.07
$\gamma_3 = 1$	-0.97	0.31	1.03	0.03	0.04	0.00	-0.15	0.24	0.08
$\gamma_4 = 1$	-0.01	0.30	0.09	-0.01	0.04	0.00	-0.15	0.23	0.08
$\sigma = 0.2$	2.26	0.25	5.18	-0.06	0.15	0.03	0.02	0.03	0.00
Time (Sec)	1146.46			360.46			28.83		

Similarly, the observations for Table 5.2, which are based on more subjects, are listed below.

- Overall, if more subjects are available, the estimation results of three methods can be effectively improved in terms of smaller SE, bias and MSE.
- The estimation result offered by LAP is overall better than the results from two methods.
- LIN gave satisfactory estimation for β and γ , except for β_5 . This method might be even more attractive for its extraordinarily fast speed.

- TS still provided severely biased estimates for γ_0, γ_3 and σ .

5.2.2 Numbers of Repeated Measurements

To examine how these estimation methods perform with different numbers of repeated measurements, we simulate datasets with each subject being repeatedly measured for 6 times and 12 times. Table 5.1 contains the simulation results based on $n_i = 6$. Here we just list the estimation results of $n_i = 12$ in Table 5.3.

Table 5.3: Simulation Results: $n_i = 12$

Parameter	TS			LAP			LIN		
	Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
$\beta_1 = 12$	0.02	0.20	0.04	-0.02	0.15	0.02	-0.15	0.06	0.03
$\beta_2 = 30$	-0.52	1.38	2.18	-1.76	1.07	4.24	-0.94	1.23	2.39
$\beta_3 = 7$	-0.01	0.19	0.04	0.18	0.19	0.07	-0.08	0.13	0.02
$\beta_4 = -2$	0.01	0.24	0.06	0.53	0.33	0.38	0.05	0.17	0.03
$\beta_5 = 3$	-0.00	0.10	0.01	-0.45	0.22	0.25	-0.28	0.07	0.08
$\mu = 0.3$	-0.00	0.01	0.00	0.09	0.03	0.01	0.01	0.01	0.00
$\gamma_0 = -1$	2.16	0.43	4.87	-0.05	0.20	0.04	-0.06	0.12	0.02
$\gamma_1 = 1$	0.00	0.47	0.22	0.16	0.58	0.36	-0.96	0.19	0.95
$\gamma_2 = 1$	0.03	0.47	0.22	-0.16	0.53	0.31	-0.94	0.23	0.93
$\gamma_3 = 1$	-0.85	0.46	0.93	0.12	0.21	0.06	-0.05	0.21	0.05
$\gamma_4 = 1$	0.04	0.45	0.20	0.01	0.16	0.03	-0.09	0.21	0.05
$\sigma = 0.2$	2.20	0.31	4.92	-0.02	0.22	0.05	0.08	0.05	0.01
Time (Sec)	183.73			187.58			15.60		

The observations for the results based on more measurements, Table 5.3, are

- Overall, it might be hard to reduce the bias or SE of the estimates by having more measurements for each subject.
- Specifically, LAP led the performance of estimating β and γ , though the bias of the estimates are greater than Table 5.2.
- LIN also provided good estimation for β and γ except for γ_1 and γ_2 , which are associated with the random effects of the measurement error model.

- With more repeated measurements ensured, the computation time of TS is greatly reduce. However, TS still provided severely biased estimates for γ_0, γ_3 and σ .

5.2.3 Between-Subject Variance

To examine how these estimation methods perform with different variance-covariance matrices, we simulate datasets with different settings of A and B follows:

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, B_1 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix},$$

and

$$A_2 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} B_2 = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 6 \end{pmatrix}.$$

Table 5.2 is based on A_1 and B_1 . Table 5.4 is based on A_2 and B_2 .

For Table 5.4,

- Surprisingly, the estimation result that LAP offered in the greater between-subject variance scenario is better than the less between-subject variance scenario (Table 5.1). It is almost as good as the more sample size scenario (Table 5.2). However, LIN is not that lucky, at least in terms of the bias of estimates of β_5 and γ_2 .
- The SE of estimates are reasonably greater than previous setting of A_1 and B_1 .

5.2. Simulation Results

Table 5.4: Simulation Results: A_2, B_2

Parameter	TS			LAP			LIN		
	Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
$\beta_1 = 12$	0.10	0.25	0.07	-0.03	0.16	0.03	-0.15	0.07	0.03
$\beta_2 = 30$	-0.87	2.18	5.52	-1.17	0.93	2.23	-0.32	2.14	4.69
$\beta_3 = 7$	-0.07	0.29	0.09	0.10	0.23	0.06	-0.24	0.23	0.11
$\beta_4 = -2$	-0.02	0.34	0.11	0.15	0.39	0.18	-0.12	0.31	0.11
$\beta_5 = 3$	0.00	0.11	0.01	-0.17	0.10	0.04	-0.56	0.10	0.32
$\mu = 0.3$	-0.00	0.01	0.00	0.14	0.04	0.02	0.03	0.05	0.00
$\gamma_0 = -1$	2.61	0.51	7.10	-0.02	0.08	0.01	-0.14	0.37	0.15
$\gamma_1 = 1$	0.06	0.36	0.14	-0.05	0.25	0.06	-0.03	0.32	0.10
$\gamma_2 = 1$	-0.01	0.38	0.15	0.04	0.21	0.05	-0.22	0.41	0.22
$\gamma_3 = 1$	-0.91	0.34	0.94	0.01	0.02	0.00	-0.09	0.36	0.13
$\gamma_4 = 1$	0.02	0.31	0.09	0.00	0.03	0.00	-0.09	0.35	0.13
$\sigma = 0.2$	2.84	0.37	8.19	-0.02	0.18	0.03	0.80	5.52	31.07
Time (Sec)	908.11			172.65			14.22		

- TS still provided severely biased estimates for γ_0, γ_3 and σ , while taking much longer time on computation.

5.2.4 Within-Subject Variance

To examine how these estimation methods perform with different within-subject variance, we simulate datasets with different settings of μ, λ and σ as follows:

$$\mu_1 = 0.3, \lambda_1 = 0.2, \sigma_1 = 0.2, \quad \text{and} \quad \mu_2 = 0.6, \lambda_2 = 0.5, \sigma_1 = 0.4.$$

Table 5.1 is based on μ_1, λ_1 and σ_1 . Table 5.5 is based on μ_2, λ_2 and σ_2 .

- When within-subject variances become greater, the estimation results become worse in terms of both SE and bias.
- No method provides satisfactory estimation for the simulated time-to-event data with greater within-subject variance, indicating both the TS based

5.2. Simulation Results

Table 5.5: Simulation Results: $\mu_2, \lambda_2, \sigma_2$

Parameter	TS			LAP			LIN		
	Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
$\beta_1 = 12$	0.05	0.20	0.04	0.04	0.23	0.06	-0.05	0.11	0.01
$\beta_2 = 30$	-2.68	2.43	13.09	-1.97	2.36	9.43	-1.31	2.15	6.31
$\beta_3 = 7$	-0.12	0.24	0.07	0.25	0.28	0.14	-0.19	0.34	0.16
$\beta_4 = -2$	-0.08	0.31	0.10	0.69	0.47	0.70	-0.11	0.51	0.28
$\beta_5 = 3$	0.02	0.16	0.03	-0.22	0.22	0.10	-1.20	0.15	1.47
$\mu = 0.6$	0.20	0.02	0.04	0.35	0.08	0.13	0.14	0.07	0.02
$\gamma_0 = -1$	2.28	0.41	5.38	0.36	0.85	0.86	-0.28	0.35	0.20
$\gamma_1 = 1$	0.01	0.52	0.27	-2.29	4.54	25.84	-0.97	0.23	1.00
$\gamma_2 = 1$	0.05	0.51	0.26	1.83	4.35	22.30	-0.94	0.21	0.92
$\gamma_3 = 1$	-0.99	0.26	1.05	1.78	3.05	12.47	-0.10	0.22	0.06
$\gamma_4 = 1$	0.05	0.67	0.46	-0.11	1.49	2.24	-0.10	0.23	0.06
$\sigma = 0.5$	2.25	0.32	5.15	0.92	1.16	2.21	0.34	0.13	0.13
Time (Sec)	1644.74			299.28			16.48		

method and the approximate methods might yield in greatly biased estimation when variance is too high. In this situation, some “exact” method, such as EM algorithm, might be necessary.

In general, we may conclude that

- TS is risky for unbiased estimation of joint modeling of longitudinal process and time-to-event process, especially when strong association exists in between.
- Approximate simultaneous inference methods are promising to provide satisfactory estimation for both longitudinal processes and time-to-event processes when the measurement errors problem or the model mis-specification problem is tolerable.
- LAP generally performs more stably than LIN. In fact, when more subjects are given, or the between-subject variance falls in an appropriate range, LAP is likely to provide quite reliable estimation.

- LIN can be considered as a valuable alternative of LAP when more subjects are involved, especially considering its fast speed. However, this algorithm requires further study before being widely promoted, since its estimate for some parameters, for example β_5 , might be biased.

Chapter 6

Conclusion and Discussion

The computation associated with the joint modeling of longitudinal process and survival process can be extremely intensive and may lead to convergence problems (Tsiatis and Davidian, 2004; Wu et al., 2008). The situation can be even severe when several nonlinear and semiparametric/nonparametric models are involved in dealing with complex data, such as measurement errors or missing data are present. Previous researches have mainly used EM algorithm or Laplace approximation for making inference of several simultaneously developing processes. Besides the technical details of these two popular methods, we were more impressed by these two different approaches of handling the sharing parameters of the longitudinal studies.

In this thesis, following those two distinct computational frameworks, we proposed an approximate statistical inference method for jointly modeling longitudinal process and time-to-event process based on a NLME model and a parametric AFT model. By linearizing the joint model, we designed a new strategy of updating the random effects that connect two processes, and proposed two approaches for different scenarios of likelihood function. Roughly speaking, one approach is widely applicable and easily implementable when the likelihood function is comparatively complex. The other approach utilizes more information of the estimated random effects when the likelihood function is comparatively simple, and therefore is expected to be more efficient. Both approaches approximate the multidimensional integral in the observed-data joint likelihood by an analytic

expression, which greatly reduce the computational intensity of the complex joint modeling problem.

The proposed inference method, LIN, was applied to an HIV study along with three other methods, TS, MTS and LAP. The inference results obtained from different methods gave significant different conclusions towards the HIV study. Specifically,

- Two simultaneous inference methods suggested that CD4 is significant to the viral load, which was missed by the TS based methods (TS and MTS). Based on the simultaneous inference results, the lower the CD4 is, the higher viral load the patient might have.
- Two simultaneous inference methods also indicated that the viral decay rate is significant to the first decline time of the CD4/CD8 ratio. The lower the viral load is, the later the CD4/CD8 may firstly decline. TS based method failed again in detecting this important feature.

We also compared these inference methods based on simulation results. Generally speaking,

- It is risky to use TS method for unbiased estimation of the joint modeling of longitudinal process and time-to-event process.
- LAP is generally more reliable than the other two methods in providing a valid estimation of the joint models.
- LIN might become a valuable alternative to LAP in terms of estimation results and computation time, especially when more subjects are given.

Linearization methods have achieved numerous successes for solving nonlinear problems in real world applications, according to which we believe that the linearization strategy may also help us in jointly modeling longitudinal data and

time-to-event data. Our initial trial in this area seems promising and we believe that the following research directions might be interesting in the future.

- Our proposed method is based on a model linearization and a likelihood function approximation. The asymptotic property of the estimator is naturally of interest. That is to say, we need to search for some theoretical justifications for our approach.
- The model that we adopted in our modeling for time-to-event process, Weibull model, is a specific choice for event time analysis. As we introduced in Section 1.1.3, Cox models and general AFT models are also extremely popular in practice. Thus, one might be interested to generalize the linearization strategy to those situations where Cox models or general AFT models are involved in the joint modeling.
- In many practical longitudinal studies, the response variables are recorded as various types of category. Thus, the NLME in our modeling might need to be replaced by generalized mixed effects model (GLMM). Generalizing the linearization strategy to joint modeling with GLMM may be valuable to practical application.
- Without a doubt, the missing data problem is always one of the cores of longitudinal studies. Our approach is based on likelihood, which determines that the missing data problems could be naturally incorporated into our framework. However, as more statistical problems are taken into account, the joint modeling itself is getting more and more complex in terms of both theoretical justification and computation. These could be the focuses of our future research.

Bibliography

- Albert, P. S. and Shih, J. H. (2009). On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics*, pages DOI 10.1111/j.1541-0420.2009.01324.x.
- Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Ser: B*, 61:265 – 285.
- Brown, E. R. and Ibrahim, J. G. (2003). A bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59:221 – 228.
- Brown, E. R., Ibrahim, J. G., and DeGruttola, V. (2005). A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61:64 – 73.
- Carroll, R. J., Ruppert, D., and Stefanski, L. A. (2006). *Measurement Error in Nonlinear Models*. Chapman & Hall/CRC, London, 2nd edition edition.
- Chambers, J. M. (1991). *Statistical Models in S*. Duxbury Press.
- Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurements Data*. Chapman & Hall/CRC, London.
- Davidian, M. and Giltinan, D. M. (2003). Nonlinear models for repeated measurements data: an overview and update. *Journal of Agricultural, Biological and Environmental Statistics*, 8:387 – 419.

- de Boor, C. (1978). *A Practical Guide to Splines*. Springer-Verlag, New York.
- DeGruttola, V. and Tu, X. M. (1994). Modeling progression of cd-4 lymphocyte count and its relationship to survival time. *Biometrics*, 50:1003 – 1014.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Ser: B*, 39:1 – 38.
- Diggle, P., Heagerty, P., Liang, K. Y., and Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press, 2nd edition.
- Ding, A. and Wu, H. (2001). Assessing antiviral potency of anti-hiv therapies in vivo by comparing viral decay rates in viral dynamic models. *Biostatistics*, 2:13 – 29.
- Ding, J. M. and Wang, J. L. (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics*, 64:546 – 556.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC.
- Faucett, C. J. and Thomas, D. C. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: A gibbs sampling approach. *Statistics in Medicine*, 15:1663 – 1685.
- Fletcher, R. (1987). *Practical Methods of Optimization*. John Wiley & Sons, 2nd edition.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal desities. *Journal of the American Statistical Association*, 85:398 – 409.

- Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1, 4:465 – 480.
- Hsieh, F. S., Tseng, Y. K., and Wang, J. L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62:1037 – 1043.
- Huang, X. Z., Stefanski, L. A., and Davidian, M. (2009). Latent-model robustness in joint models for a primary endpoint and a longitudinal process. *Biometrics*, 65:719 – 727.
- Ibrahim, J. G., Chen, M. H., and Lipsitz, S. R. (2001a). Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable. *Biometrika*, 88:551 – 564.
- Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001b). *Bayesian Survival Analysis*. Springer, New York.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38:963 – 974.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley Series in Probability and Statistics. Wiley Interscience, 2nd edition.
- Lee, Y., Nelder, J. A., and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*. Chapman & Hall/CRC.
- Li, L., Hu, B., and Greene, T. (2009). A semiparametric joint model for longitudinal and survival data with application to hemodialysis study. *Biometrics*, 65:737 – 745.
- Li, N., Elashoff, R. M., Li, G., and Saverc, J. (2010). Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the ninds rt-pa stroke trial. *Statistics in Medicine*, 29:546 – 557.

Bibliography

- Lindstrom, M. J. and Bates, D. M. (1988). Newton-raphson and em algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83:1014 – 1022.
- Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46:673 – 687.
- Little, R. J. A. (1992). Regression with missing x's: a review. *Journal of the American Statistical Association*, 87:1227 – 1237.
- Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated measures studies. *Journal of the American Statistical Association*, 90:1112 – 1121.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Values*. Wiley, New York, 2nd edition.
- Liu, B. and Muller, H. G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *Journal of the American Statistical Association*, 104:704 – 717.
- Liu, L. (2009). Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in Medicine*, 28:972 – 986.
- Liu, W. and Wu, L. (2007). Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. *Biometrics*, 63:342 – 350.
- Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society, Series: B*, 44:226 – 233.
- Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, New York, 2nd edition.

- Pinheiro, J. C. and Bates, D. M. (2002). *Mixed-Effects Models in S and SPLUS*. Springer, New York.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York, 2nd edition.
- Ratcliffe, S. J., Guo, W. S., and Have, T. R. T. (2004). Joint modeling of longitudinal and survival data via a common frailty. *Biometrics*, 60:892 – 899.
- Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed-effects models for unequally sampled noisy curves. *Biometrics*, 57:253 – 259.
- Rizopoulos, D., Verbeke, G., and Lesaffre, E. (2009). Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society, Series: B*, 71:637 – 654.
- Rizopoulos, D., Verbeke, G., Lesaffre, E., and Vanrenterghem, Y. (2008). A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics*, 64:611C619.
- Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2010). Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics*, 66:20 – 729.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC.
- Shah, A., Laird, N., and Schoenfeld, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association*, 92:775 – 779.
- Song, X., Davidian, M., and Tsiatis, A. A. (2002a). An estimator for the pro-

- portional hazards model with multiple longitudinal covariates measured with error. *Biostatistics*, 3, 4:511 – 528.
- Song, X., Davidian, M., and Tsiatis, A. A. (2002b). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58:742 – 753.
- Song, X. and Huang, Y. J. (2006). A corrected pseudo-score approach for additive hazards model with longitudinal covariates measured with error. *Lifetime Data Analysis*, 12:97 – 110.
- Song, X. and Wang, C. Y. (2008). Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. *Biometrics*, 64:557 – 566.
- Tseng, Y. K., Hsieh, F. S., and Wang, J. L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, 92:587 – 603.
- Tsiatis, A. A. and Davidian, M. (2004). An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica*, 14:793 – 818.
- Verbeke, G. and Molenberghs, G. (2001). *Linear Mixed Models for Longitudinal Data*. Springer, New York.
- Vonesh, E. F., Greene, T., and Schluchter, M. D. (2006). Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine*, 25:143 – 163.
- Vonesh, E. F., Wang, H., Nie, L., and Majumdar, D. (2002). Conditional second order generalized estimating equations for generalized linear and nonlinear mixed-effects models. *Journal of the American Statistical Association*, 97:271 – 283.

- Wang, Y. and Taylor, J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96:895 – 905.
- Wei, G. C. and Tanner, M. A. (1990). A monte-carlo implementation of the em algorithm and the poor man’s data augmentation algorithm. *Journal of the American Statistical Association*, 85:699 – 704.
- Wu, H. and Ding, A. (1999). Population hiv-1 dynamics in vivo: applicable models and inferential tools for virological data from aids clinical trials. *Biometrics*, 55:410 – 418.
- Wu, H. and Zhang, J. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches*. Wiley-Interscience.
- Wu, L. (2002). 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*, 97:955 – 964.
- Wu, L. (2009). *Mixed Effects Models for Complex Data*. Chapman & Hall/CRC.
- Wu, L., Hu, X. J., and Wu, H. (2008). Joint inference for nonlinear mixed-effects models and time-to-event at the presence of missing data. *Biostatistics*, 9:308 – 320.
- Wu, L., Liu, W., and Hu, X. J. (2010). Joint inference on hiv viral dynamics and immune suppression in presence of measurement errors. *Biometrics*, 66:327 – 335.
- Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44:175 – 188.

- Wulfshon, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53:300 – 309.
- Xu, J. and Zeger, S. L. (2001a). The evaluation of multiple surrogate endpoints. *Biometrics*, 57:81 – 87.
- Xu, J. and Zeger, S. L. (2001b). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics*, 50:375 – 387.
- Ye, W., Lin, X., and Taylor, J. M. G. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data: A two stage regression calibration approach. *Biometrics*.
- Yu, M. G., Law, N. J., Taylor, J. M. G., and Sandler, H. M. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, 14:835 – 862.
- Yu, M. G., Taylor, J. M. G., and Sandler, H. M. (2008). Individual prediction in prostate cancer studies using a joint longitudinal survivalcure model. *Journal of the American Statistical Association*, 103:178 – 187.
- Zeng, D. L. and Cai, J. W. (2005a). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *Annals of Statistics*, 33:2132 – 2163.
- Zeng, D. L. and Cai, J. W. (2005b). Simultaneous modelling of survival and longitudinal data with an application to repeated quality of life measures. *Lifetime Data Analysis*, 11:151 – 174.
- Zhang, H. P., Ye, Y. Q., Diggle, P., and Shi, J. (2008). Joint modeling of time series measures and recurrent events and analysis of the effects of air quality on respiratory symptoms. *Journal of the American Statistical Association*, 103:48 – 60.

Bibliography

Zhang, S., Muller, P., and Do, K. A. (2010). A bayesian semiparametric survival model with longitudinal markers. *Biometrics*, 66:719 – 727.