

ABSTRACT

HUANG, MINGYAN. Semiparametric Mixed Models for Censored Longitudinal Data. (Under the direction of Dr. Daowen Zhang.)

In longitudinal studies involving laboratory-based outcomes, repeated measurements can be censored due to assay detection limits. For analyzing such data, likelihood based approaches accounting for censoring have been proposed under the linear mixed model (LMM) framework. A key assumption of a LMM is that the response variable is linearly related to all the covariate effects in the model, including the time effect. In many applications, however, the linear parametric form of LMMs appears too restrictive to characterize the complex relationship between a response variable and covariates. More general and robust modeling tools, nonparametric and semiparametric regression models, have become increasingly popular in the last decade. In this dissertation, we propose to use semiparametric mixed models (SPMMs) to analyze censored longitudinal data. The SPMMs extend LMMs and provide more flexible modeling schemes by allowing the time effect and the coefficients of some other covariates to vary nonparametrically over time. We applied two nonparametric smoothing techniques, the regression spline approach with B-splines as basis functions and the smoothing spline approach, for estimation of the smooth functions in our models. For the regression spline approach, we proposed two computational procedures, namely, the direct maximization of likelihood and the expectation-maximization (EM) algorithm, to achieve the maximum likelihood estimates (MLEs) of model parameters. For the smoothing spline method, we introduced an EM version of maximum penalized likelihood estimates (MPLEs) of model parameters and the nonparametric function. We evaluated the performance of the proposed approaches through extensive simulations, and illustrated their applications with data from an AIDS study.

© Copyright 2010 by Mingyan Huang

All Rights Reserved

Semiparametric Mixed Models for Censored Longitudinal Data

by
Mingyan Huang

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2010

APPROVED BY:

Dr. Helen Zhang

Dr. Marie Davidian

Dr. Wenbin Lu

Dr. Daowen Zhang
Chair of Advisory Committee

DEDICATION

To my parents, my husband and my daughter.

BIOGRAPHY

Mingyan Huang was born in Zhuji, Zhejiang, China. She received her B.S. in Biology from Nanjing University in 1992 and her M.S. in Biochemistry from Zhejiang Medical University (currently Zhejiang University, School of Medicine) in 1995. After graduation, she served as a lecturer in the Department of Biochemistry at Zhejiang Medical University for a few years. In January 2001, she was admitted to the graduate program in Statistics at North Carolina State University (NCSU). She earned her M.S. degree in Statistics in May, 2003. Two years later, she decided to continue pursuing her Ph.D degree and was re-enrolled into the statistics program at NCSU. Currently she is a part-time student and working as a biostatistician at the Department of Medical Oncology of Duke University. She is expected to receive her Ph.D degree in 2010.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest appreciation to my advisor, Dr. Daowen Zhang for his constant supports, guidance, encouragement and patience in assisting me with my doctoral research and the preparation of my thesis over the past years. I would also like to thank the rest advisory committee members, Dr. Marie Davidian, Dr. Helen Zhang and Dr. Wenbin Lu for their assistance, service and many valuable suggestions. I am also very grateful to all the faculties in the Department of Statistics at NCSU for their profound impact on my professional growth. My special thanks go to the staff members in the department for all their wonderful help and supports during my study. Finally, I would like to thank my family, especially my husband and my daughter, for their understanding, support, constant love and caring throughout the entire journey.

TABLE OF CONTENTS

List of Tables	vi
List of Figures	viii
Chapter 1 Introduction	1
1.1 Background and Motivation	1
1.2 Statistical Models	5
1.2.1 Semiparametric Mixed Model	5
1.2.2 Time-varying Coefficient Mixed Model	7
1.3 Nonparametric Spline Smoothing in Semiparametric Regression Models	8
1.3.1 Regression Spline Approach	9
1.3.2 Smoothing Spline Approach	11
Chapter 2 Regression Spine Method for Censored Longitudinal Data	15
2.1 Semiparametric Mixed Model	16
2.1.1 Model Specification	16
2.1.2 Estimation Procedures	17
2.1.3 Simulation	22
2.1.4 Application	31
2.2 Time-varying Coefficient Mixed Model	39
2.2.1 Model Specification	39
2.2.2 Estimation Procedure: Hybrid EM/Quasi-Newton Approach	39
2.2.3 Simulation	41
2.2.4 Application	43
2.3 Discussion	55
Chapter 3 Smoothing Spline Method for Censored Longitudinal Data	58
3.1 Maximum Penalized Likelihood Estimates	59
3.1.1 Matrix Notation of SPMM	59
3.1.2 Estimation of Model Parameters	60
3.1.3 Variance for Maximum Penalized Likelihood Estimates	64
3.1.4 Estimation of Smoothing Parameter	65
3.2 Simulation	66
3.3 Application	67
3.4 Discussion	77
References	79
Appendices	83

LIST OF TABLES

Table 2.1	Simulation results from model (2.8). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.	25
Table 2.2	Simulation results from model (2.9). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.	26
Table 2.3	Variance component estimates for ACTG398 viral load data from model (2.10).	33
Table 2.4	Parameter estimates for ACTG398 viral load data from model (2.10).	34
Table 2.5	Simulation results from model (2.12). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.	48
Table 2.6	Simulation results from model (2.13). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.	49
Table 2.7	Variance component estimates for ACTG398 viral load data from model (2.14).	51
Table 2.8	Parameter estimates for ACTG398 viral load data from model (2.14).	51
Table 3.1	Simulation results from model (3.8). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.	70

Table 3.2	Simulation results from model (3.9). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.	71
Table 3.3	Parameter estimates for ACTG398 viral load data from model (3.10).	75

LIST OF FIGURES

Figure 2.1	Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (2.8) with 17% censoring. (a)-(c) for the imputation method; (d)-(f) for the direct optimization approach; (g)-(i) for the EM approach.	27
Figure 2.2	Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (2.8) with 34% censoring. (a)-(c) for the imputation method; (d)-(f) for the direct optimization approach; (g)-(i) for the EM approach.	28
Figure 2.3	Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (2.9) with 17% censoring. (a)-(c) for the imputation method; (d)-(f) for the direct optimization approach; (g)-(i) for the EM approach.	29
Figure 2.4	Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (2.9) with 34% censoring. (a)-(c) for the imputation method; (d)-(f) for the direct optimization approach; (g)-(i) for the EM approach.	30
Figure 2.5	Estimated $\hat{f}(t)$ for different random effect formulations of model (2.10) from three estimation procedures.	34
Figure 2.6	Estimated $\hat{f}(t)$ (solid line) along with its confidence interval (dashed lines) for different random effect formulations of model (2.10) from the EM method.	35
Figure 2.7	Estimated population means of $\log_{10}(\text{RNA copies/ml})$ for the two treatment (DPI versus SPI) groups from the IM, Half IM and EM methods, for patients who had no prior NNRTI treatment ($X_{2i}=0$).	35
Figure 2.8	Estimated population means of $\log_{10}(\text{RNA copies/ml})$ for the two treatment (DPI versus SPI) groups from the Half IM and EM methods, for patients who had no prior NNRTI treatment ($X_{2i}=0$).	36

Figure 2.9	Estimated nonparametric function $\widehat{f}_0(t)$ (solid line) with the true fixed function $f_0(t)$ (dashed line) superimposed (top row panel), empirical biases of $\widehat{f}_0(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f_0(t)$ (bottom row panel) for model (2.12). (a)-(c) the imputation method with 11% censoring; (d)-(f) the imputation method with 22% censoring; (g)-(i) the hybrid EM/Quasi-Newton approach with 11% censoring; (j)-(l) the hybrid EM/Quasi-Newton approach with 22% censoring.	44
Figure 2.10	Estimated nonparametric function $\widehat{f}_1(t)$ (solid line) with the true fixed function $f_1(t)$ (dashed line) superimposed (top row panel), empirical biases of $\widehat{f}_1(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f_1(t)$ (bottom row panel) for model (2.12). (a)-(c) the imputation method with 11% censoring; (d)-(f) the imputation method with 22% censoring; (g)-(i) the hybrid EM/Quasi-Newton approach with 11% censoring; (j)-(l) the hybrid EM/Quasi-Newton approach with 22% censoring.	45
Figure 2.11	Estimated nonparametric function $\widehat{f}_0(t)$ (solid line) with the true fixed function $f_0(t)$ (dashed line) superimposed (top row panel), empirical biases of $\widehat{f}_0(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f_0(t)$ (bottom row panel) for model (2.13). (a)-(c) the imputation method with 11% censoring; (d)-(f) the imputation method with 22% censoring; (g)-(i) the hybrid EM/Quasi-Newton approach with 11% censoring; (j)-(l) the hybrid EM/Quasi-Newton approach with 22% censoring.	46
Figure 2.12	Estimated nonparametric function $\widehat{f}_1(t)$ (solid line) with the true fixed function $f_1(t)$ (dashed line) superimposed (top row panel), empirical biases of $\widehat{f}_1(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f_1(t)$ (bottom row panel) for model (2.13). (a)-(c) the imputation method with 11% censoring; (d)-(f) the imputation method with 22% censoring; (g)-(i) the hybrid EM/Quasi-Newton approach with 11% censoring; (j)-(l) the hybrid EM/Quasi-Newton approach with 22% censoring.	47
Figure 2.13	Estimated $\widehat{\beta}_0(t)$ for different random effect formulations of model (2.14) from three estimation procedures.	51
Figure 2.14	Estimated $\widehat{\beta}_0(t)$ (solid line) along with its confidence interval (dashed lines) for different random effect formulations of model (2.14) from the hybrid EM/Quasi-Newton approach.	52
Figure 2.15	Estimated $\widehat{\beta}_1(t)$ for different random effect formulations of model (2.14) from three estimation procedures.	52
Figure 2.16	Estimated $\widehat{\beta}_1(t)$ from model (2.14) and model (2.10) by the hybrid EM/Quasi-Newton approach.	53

Figure 2.17	Estimated population means of $\log_{10}(\text{RNA copies/ml})$ for the two treatment (DPI versus SPI) groups from the IM, Half IM and hybrid EM/Quasi-Newton methods, for patients who had no prior NNRTI treatment ($X_{2i}=0$).	53
Figure 2.18	Estimated population means of $\log_{10}(\text{RNA copies/ml})$ for the two treatment (DPI versus SPI) groups from the Half IM and hybrid EM/Quasi-Newton methods, for patients who had no prior NNRTI treatment ($X_{2i}=0$).	54
Figure 3.1	Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (3.8) from the proposed approach.(a)-(c) 0.2% censoring; (d)-(f) 17% censoring; (g)-(i) 34% censoring.	69
Figure 3.2	Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (3.8) from the imputation approach.(a)-(c) 0.2% censoring; (d)-(f) 17% censoring; (g)-(i) 34% censoring.	72
Figure 3.3	Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (3.9) from the proposed approach.(a)-(c) 0.2% censoring; (d)-(f) 17% censoring; (g)-(i) 34% censoring.	73
Figure 3.4	Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (3.9) from the imputation approach.(a)-(c) 0.2% censoring; (d)-(f) 17% censoring; (g)-(i) 34% censoring.	74
Figure 3.5	Estimated $\hat{f}(t)$ from the smoothing spline approach (left panel) and the B-spline approach (right panel).	75
Figure 3.6	Estimated $\hat{f}(t)$ (solid line) along with its confidence interval (dashed lines) from the EM methods using two different smoothing approaches.	76
Figure 3.7	Estimated population means of $\log_{10}(\text{RNA copies/ml})$ for the two treatment (DPI versus SPI) groups from two different smoothing approaches, for patients who had no prior NNRTI treatment ($X_{2i}=0$).	76

Chapter 1

Introduction

1.1 Background and Motivation

Longitudinal studies are common in epidemiological and biomedical research. In these studies, subjects are followed over time with repeated measurements of risk factors and health outcomes, and interest often focuses on characterizing the time course of a response variable of interest or demonstrating the dynamic association between a response variable and covariates.

A well-known feature of longitudinal data is the correlation in the multiple measurements within study subjects and this correlation has to be taken into account in statistical modeling to yield valid inference. Parametric regression models, such as linear mixed models (LMMs) (Laird and Ware, 1982; Verbeke and Molenberghs, 2009) have proved to be valuable tools for analyzing continuous longitudinal data. With the incorporation of subject-specific random effects, LMMs can properly model the correlation of longitudinal data. LMMs become increasingly popular with the availability of statistical packages like Splus and SAS for model estimation and inference.

When the response in a longitudinal study is a laboratory-based outcome, censoring may occur due to the upper and/or lower detection limits of an assay. Examples of such censored

longitudinal data arise from a variety of research areas (Barletta et al., 2004; Moulton and Halsey, 1995; Singh and Nocerino, 2002). Censoring could be left, right or interval-censored. Typical examples of left censored longitudinal data are from Human Immunodeficiency Virus (HIV) studies (Barletta et al., 2004), where the detection of viral load (i.e., the virus RNA copy number) in blood compartment is often limited by the sensitivity of a laboratory-performed assay. With the advance of effective antiviral treatments, in some cases the HIV copy number can be extremely low and beyond the detection limit, which leads to left-censoring.

Various statistical approaches have been developed to deal with longitudinal data containing censored measurements within the mixed effects model framework. A simple and *ad hoc* approach is to substitute the censored measurements with the value of the full or half detection limit, which was shown to produce biased estimates (Hughes, 1999; Jacqmin-Gadda et al., 2000). Another method suggested by Paxton et al. (1997) applied an iterative imputation procedure to adjust for the censoring. Although this procedure produces less biased estimates of fixed effects, the variance components estimates are still unreliable as it fails to account for the correlated structure of data and the uncertainty in the imputed data. Recent developments in modeling censored longitudinal data are based on the likelihood of observed data. In particular, two major estimation procedures are available in the LMMs context. One is called Monte Carlo Expectation Maximization (MCEM) algorithm developed by Hughes (1999), where the EM algorithm of Dempster, Rubin and Tsutakawa (1981) was extended to handle left and/or right censored longitudinal data with conditional expectations in the E-step evaluated using Gibbs sampler. Since it involves convergence in the Gibbs sampler for each EM iteration and the convergence in the EM algorithm itself, this MCEM algorithm can be computationally intensive. The second estimation procedure is based on direct optimization of the observed data likelihood, where one first approximates the likelihood of observed data by using certain numerical integration techniques, and then applies optimization subroutines to obtain the maximum likelihood estimates (MLEs) of

model parameters (Jacqmin-Gadda et al., 2000; Lyles et al., 2000). While the direct optimization approach is applicable to LMMs with relatively complex covariance structures and gives better performance than the MCEM algorithm (Jacqmin-Gadda et al., 2000), it requires high-dimensional numerical integrations and its computational complexity greatly increases as the proportion of the censored measurements becomes large, which may lead to a poor approximation of the observed data likelihood and consequently yields biased estimates of model parameters.

Note that the existing approaches developed for censored longitudinal data analysis are mainly based on LMMs. Although LMMs are useful tools for analyzing longitudinal data, an important assumption of LMMs is that the response variable is linearly related to its covariates by a known function. Often times this linear regression function is not straightforward to derive due to the lack of sufficient understanding of scientific problems at hand. In other situations, the linear parametric form of LMMs appears too restrictive to be used to address the complex relationship between a response variable and covariates. To overcome this difficulty, a more general and robust modeling tool is needed, which motivates the development of nonparametric regression models.

In the last decade, nonparametric and semiparametric regression models that provide great flexibility in modeling covariate effects of longitudinal data have been extensively investigated. Instead of using a linear predictor, these models formulate the relationship between the response variable and certain covariates through arbitrary functions and the unknown functions are estimated using nonparametric smoothing techniques. Pure nonparametric regression models have been accused of suffering from the “curse of dimensionality” and are indeed often difficult to implement in practice. Semiparametric regression models hence have gained increasing attention in longitudinal data analysis due to their flexible structure. As implied by the name, semiparametric regression models incorporate both parametric and nonparametric forms of covariate effects, and therefore enjoy the flexibility of nonparametric regression models while retaining nice properties such as easy

implementation and good interpretability of parametric models.

There is a rich literature on the development of semiparametric regression models for longitudinal data analysis. For example, Zeger and Diggle (1994) proposed a semiparametric model where a nonparametric function is used to model the time effect, and a random intercept together with a Gaussian stochastic process is used to account for the within-subject correlation. Extending the method of Zeger and Diggle (1994), Zhang et al. (1998) developed a semiparametric stochastic mixed model that incorporates a general random effects term as well as a stationary or nonstationary process to allow for more flexibility in modeling the within-subject variation. Wang (1998) proposed a smoothing spline model with correlated random errors, which can be applied to longitudinal data. More recently Rice and Wu (2001) introduced a nonparametric model for longitudinal data where both fixed effects and random effects are modeled nonparametrically. It is noteworthy that another special class of semiparametric regression models, time-varying coefficient models, are also becoming increasingly popular (Fan and Zhang, 2000; 2008; Hoover et al., 1998; Zhang, 2004, among others). Time-varying coefficient models are further extensions of the semiparametric regression model which contains only a nonparametric time-effect term. By allowing covariate effects to vary nonparametrically over time, time-varying coefficient models are capable of characterizing the dynamic relationship between a longitudinal response variable and certain covariates of research interest.

All aforementioned semiparametric regression models and their extensions have so far been used in analyzing non-censored longitudinal data. To our knowledge, little effort has been made to model longitudinal data subject to censoring under semiparametric regression model framework. In this dissertation, we propose to use semiparametric mixed models for analyzing censored longitudinal data. We apply both the regression spline and smoothing spline techniques for estimation of the smooth functions in our models. To achieve the MLEs of model parameters, we also propose EM-based estimation procedures which are straightforward to implement. The remainder of this dissertation is organized as follows.

In the following sections of Chapter 1, we first introduce the semiparametric regression models to be used in our research, and then provide a brief review on two commonly used nonparametric smoothing techniques. Detailed exploration on the use of regression splines in semiparametric mixed models for censored longitudinal data analysis is discussed in Chapter 2. Lastly, in Chapter 3 we discuss how to obtain the EM version of penalized likelihood estimates of model parameters by using smoothing splines.

1.2 Statistical Models

1.2.1 Semiparametric Mixed Model

Semiparametric mixed models (SPMMs), in this dissertation, are referred to as a class of models which use an arbitrary smooth function to model the time effect, a parametric linear function to represent covariate effects, and account for the within-subject correlation using random effects. SPMMs are natural extensions of classical LMMs and have been successfully used in longitudinal data analysis (see, e.g., Zeger and Diggle, 1994; Zhang et al., 1998; Diggle, 2002; Li et al., 2010).

The SPMM introduced by Zeger and Diggle (1994), and Zhang et al. (1998) can be generally written as

$$Y_{ij} = f(t_{ij}) + X_{ij}^T \beta + Z_{ij}^T b_i + \epsilon_{ij}, \quad (1.1)$$

where Y_{ij} ($i = 1, \dots, m, j = 1, \dots, n_i$) denotes the j th response of the i th subject observed at time t_{ij} , m is the number of subjects with each subject having n_i observations; β and b_i are the vectors of fixed effects and random effects which are associated with design matrices X_{ij} and Z_{ij} , respectively; and $f(t)$ is an unknown smooth function of time, representing the population baseline mean curve. Assume $b_i \sim N(0, D(\phi))$, where $D(\phi)$ is a positive definite covariance matrix dependent on the vector ϕ of some variance/covariance parameters, $\epsilon_{ij} \sim N(0, \sigma^2)$ are independent measurement errors and are independent of b_i .

This SPMM is particularly useful when the scientific interest is focused on the investigation of the time course of a longitudinal response whereas a linear functional dependence is unavailable or inappropriate. For example, Zeger and Diggle (1994) initially proposed an SPMM for estimating the typical time course of CD4 cell loss in order to closely monitor the disease progression of HIV infection. Zhang et al. (1998) illustrated the usage of SPMMs in a longitudinal hormone study, where the complex progesterone level during a women's menstrual cycle was successfully modeled by a nonparametric function.

Under the SPMM modeling framework, one is interested in accomplishing parameter estimation and model inferences. To facilitate nonparametric estimation, various smoothing techniques have been used to fit these type of SPMMs, including kernel smoothing methods (Diggle, 2002; Lin and Ying, 2001; Lin and Carroll, 2001), smoothing spline methods (Zhang et al., 1998; Li et al., 2010) and regression spline methods (Rice and Wu, 2001). There are one or two so-called smoothing parameters in each of these methods for controlling the model complexity and the trade-off between the bias and variance of estimates. Zeger and Diggle (1994) estimated the nonparametric function of time using a kernel smoother with the smoothing parameter (bandwidth) chosen via cross-validation and the estimates of the variance components were achieved by extrapolating an auto-covariance function. On the basis of the regression spline method, Rice and Wu (2001) discussed that SPMMs can be transformed into a working LMM representation for a given smoothing parameter (knots), therefore model estimation and statistical inference can be obtained by using existing estimation approaches well developed for LMMs. Zhang et al. (1998) estimated the nonparametric baseline function using a smoothing spline by maximizing a penalized likelihood, and obtained the estimates of the smoothing parameter and variance components simultaneously via restricted maximum likelihood (REML).

1.2.2 Time-varying Coefficient Mixed Model

While SPMMs provide more modeling flexibility than LMMs and have proved to be useful in many longitudinal applications, it is evident that model (1.1) is still quite restrictive as it retains the linear parametric assumption on other covariate effects. For problems that involve more than one predictor variable having complicated relationship with the response variable, the SPMMs may not be adequate. Thus another class of semiparametric models, namely time-varying coefficient models, provides supplementary tools for longitudinal data analysis.

The time-varying coefficient model was initially introduced by Hoover et al. (1998) and Wu et al. (1998). Consider a longitudinal dataset consisting of repeated measurements (Y_{ij}, X_{ij}, t_{ij}) , for $i = 1, \dots, m$ and $j = 1, \dots, n_i$, where X_{ij} denote the multivariate design matrix for the i th subject at time point t_{ij} . The time-varying coefficient model proposed by Hoover et al. (1998) and Wu et al. (1998) assumes that the multivariate regression function takes the form

$$Y_{ij} = X_{ij}^T \beta(t_{ij}) + \epsilon_i(t_{ij}), \quad (1.2)$$

where $\beta(t) = \{\beta_0(t), \dots, \beta_p(t)\}^T$, for $p \geq 0$, are the functional coefficients assumed to be smooth nonparametric functions of time and $\epsilon_i(t)$ is a zero-mean stochastic process.

Model (1.2) contains as a special case the SPMMs introduced in the previous section, where only the intercept coefficient is allowed to be time-varying and others are all constants. From statistical modeling point of view, model (1.2) has many appealing features. In particular, it is linear in the regressors and hence possesses the properties, such as simplicity and ease interpretation, as with traditional parametric models. On the other hand, by allowing its coefficients to vary smoothly over time, model (1.2) enjoys certain nonparametric properties, but meanwhile greatly ameliorate the “curse of dimensionality” encountered by most pure nonparametric modeling approaches.

With their flexibility and meaningful interpretability, the time-varying coefficient models have been widely used to explore the dynamic feature which may exist in longitudinal data. One typical example that has been used to illustrate the usage of the time-varying coefficient models is from AIDS clinical studies, where CD4 cell count is considered as a crucial surrogate marker in evaluating antiviral therapies and monitoring the progression of HIV infection. The longitudinal trend of CD4 percentage depletion could be affected by many factors, including cigarette smoking, pre-HIV infection CD4 percentage, age, etc. The impact of these factors on CD4 cell count may not stay constant over time. It has been shown that time-varying coefficient models provided more plausible fitting to the data and were able to reveal the important dynamic patterns of these impacts (Fan and Zhang, 2000; Huang et al., 2002; Wu and Chiang, 2000).

Various nonparametric estimation procedures have been proposed to estimate the functional coefficients in model (1.2) (see, e.g., Brumback and Rice, 1998; Hoover et al., 1998; Fan and Zhang, 2000; Huang et al., 2002). Since the within-subject correlation is an important feature of longitudinal data, efforts have also been made on how to incorporate this correlation structure into estimation procedures (see, e.g., Lin and Ying, 2001; Wang, 2003; Qu and Li, 2006). More recently, inspired by the idea of mixed effects models, some researchers suggested to include a random effects term to efficiently account for the within-subject correlation, which leads to the varying coefficient mixed models. Examples on application of such mixed effects varying coefficient models can be found in Liang et al. (2003), and Zhang (2004), among others.

1.3 Nonparametric Spline Smoothing in Semiparametric Regression Models

Substantial developments of smoothing techniques made nonparametric regression analysis possible. Various smoothing methods are available for different smoothing problem settings.

Among them, spline smoothing is one of the most popular smoothing techniques. In this section, we will give a brief introduction on two spine smoothing approaches widely used in nonparametric regression analysis of longitudinal data, namely, regression splines and smoothing splines.

1.3.1 Regression Spline Approach

Regression splines are a special type of spline functions, which usually is represented as a linear combination of a set of basis functions that span a particular linear function space specified by a small set of so-called knots (see below). Consider an interval, say $[\tau_0, \tau_{K+1}]$. Suppose there is a set of distinct interior points $\boldsymbol{\tau} = (\tau_1, \dots, \tau_K)$ (known as knots) with $\tau_0 < \tau_1 < \dots < \tau_K < \tau_{K+1}$, which divide $[\tau_0, \tau_{K+1}]$ into $K + 1$ subintervals ($[\tau_j, \tau_{j+1}], j = 0, \dots, K$). Then a spline of degree $d \geq 0$ on $[\tau_0, \tau_{K+1}]$ for the fixed knot sequence $\boldsymbol{\tau}$, is defined as a function that consists of a polynomial of degree d on each of these subintervals and has possibly discontinuous d th order derivatives at the knot points where two adjacent polynomials meet. The set of all such spline functions form a linear function space, say \mathcal{G} , and it has been shown that any spline function in \mathcal{G} can be uniquely determined by a set of suitable basis functions that span \mathcal{G} and the corresponding coefficients. Therefore, for an arbitrary smooth function $f(t)$ defined on $[\tau_0, \tau_{K+1}]$, one can approximate it using a spline estimate in the form

$$f(t) \approx \sum_{l=1}^L B_l(t)\alpha_l, \quad (1.3)$$

where $\{B_l(\cdot), l = 1, \dots, L\}$ is a set of basis functions with $L = K + d + 1$ and α_l 's are the associated coefficients to be estimated. Regression splines often use a relatively small number of knots (e.g., $K = 3$ to 6). With the number and location of the knots fixed, it is clear that a regression spline approximation (1.3) takes the linear parametric form. As such, nonparametric or semiparametric regression models introduced previously can be

easily converted into fully parametric models by replacing the nonparametric terms with regression spline approximations, and the subsequent model fitting can then proceed by following standard parametric procedures. This model simplification feature makes the regression spline very attractive in practice since it greatly eases the computation.

Different sets of basis functions can be used for regression spline estimation. One of the most commonly used bases is called truncated power basis. Consider a linear function space of splines with degree d and a knot sequence $\boldsymbol{\tau}$. The truncated power basis for this space is

$$1, t, \dots, t^d, (t - \tau_1)_+^d, \dots, (t - \tau_K)_+^d,$$

where $a_+ = a$ if $a \geq 0$ and 0 otherwise. Based on this truncated power basis, a regression spline in this function space can be expressed as

$$g(t) = \sum_{j=0}^d \beta_j t^j + \sum_{j=1}^K \delta_j (t - \tau_j)_+^d,$$

where $\boldsymbol{\alpha} = (\beta_0, \dots, \beta_d, \delta_1, \dots, \delta_K)$ is a vector of unknown parameters to be estimated. The truncated power basis is simple and can be viewed as an intuitive extension of the Taylor polynomial expansion. However, it is well known that these bases suffer from rather poor numerical properties. An important alternative of basis functions is the so-called B-spline basis (Boor, 1978). The definition of B-spline basis functions is given as:

$$B_{j,0}(t) = \begin{cases} 1 & \text{if } \tau_j \leq t < \tau_{j+1}, \\ 0 & \text{otherwise,} \end{cases}$$

$$B_{j,d}(t) = \frac{t - \tau_j}{\tau_{j+d} - \tau_j} B_{j,d-1}(t) + \frac{\tau_{j+d+1} - t}{\tau_{j+d+1} - \tau_{j+1}} B_{j+1,d-1}(t).$$

According to this definition, the B-splines of degree d can be constructed recursively in

terms of lower-degree B-splines. Note that $B_{j,d}(\cdot)$ is non-negative for all $0 \leq j \leq K$ and $d \geq 0$. More importantly, on any span $\tau_j \leq t < \tau_{j+1}$, there are at most $d + 1$ non-zero basis functions. For example, in the cubic B-spline case where $d = 3$, only $B_{j-3,3}(t), B_{j-2,3}(t), B_{j-1,3}(t)$ and $B_{j,3}(t)$ can be non-zero in $[\tau_j, \tau_{j+1})$. This latter local support properties of the B-splines is very appealing since the sparse matrix technology can then be used to speed up computation. The B-spline basis has also been shown to provide more stable numerical solutions than the truncated power basis. Detailed description of B-spline bases is given in Boor (1978).

Besides the decision on which type of basis functions to use, one also needs to choose the number and the location of the knots to apply the regression spline. Proper selection of the knot sequence is critical for good performance of regression spline estimates. Due to computational complexity, it is often impractical to select the number and the location of knots simultaneously. In most statistical literature, one often decides the knot placing method first, either choosing equally spaced knots or using sample quantiles of the data as knots, and then select the number of knots, K , through certain smoothing parameter selection criteria, such as cross-validation, Akaike information Criterion (AIC), and Bayesian information criterion (BIC). More data-adaptive knot allocation strategies, for example, free-knot splines, have also been discussed (Stone et al., 1997; Hansen and Kooperberg, 2002; Stone and Huang, 2002). Compared to the knot sequence, the degree of a regression spline, d , is less crucial, and the most commonly used splines are cubic splines (i.e., $d = 3$).

1.3.2 Smoothing Spline Approach

In nonparametric regression problems, the smoothing spline is a smoothing method which also uses spline functions to estimate regression curves. However, instead of using a limited small number of knots, a smoothing spline uses all distinct data points as its knots while introducing a penalty to control the lack of smoothness.

To motivate the idea, let's consider a simple nonparametric regression problem. Suppose

that we have scalar continuous observations Y_i at design points t_i with $a < t_i < b$ for $i = 1, \dots, n$, and the observations are assumed to satisfy

$$Y_i = f(t_i) + \epsilon_i,$$

where ϵ_i 's are independent random errors with $E(\epsilon_i) = 0$ and $f(t)$ is a smooth curve on the interval $[a, b]$. The smoothing spline estimate of the regression curve $f(t)$ is defined as the minimizer of a penalized least squares score given by,

$$\sum_{i=1}^n \{Y_i - f(t_i)\}^2 + \lambda \int_a^b \{f^{(m)}(t)\}^2 dt. \quad (1.4)$$

The first term in (1.4) is the residual sum of squares which is used to measure goodness-of-fit to the data; the second term is a penalty which is used to restrict the amount of curvature of the fitted function. The smoothing parameter $\lambda \geq 0$ controls the trade-off between the fidelity of the model fitting to the data and the smoothness of the function estimate of $f(t)$. When $\lambda = 0$, the smoothing spline estimate \hat{f}_λ simply interpolates the data points; and as $\lambda \rightarrow \infty$, \hat{f}_λ converges to a linear least squares estimate. Thus it is clear that the basic idea of the smoothing spline method is to find a function that is sufficiently flexible to capture the key features of the data while retaining certain degree of smoothness through the roughness penalty approach. Since $f(t)$ is infinite-dimensional, minimization of (1.4) is usually taken over functions with square integrable m th derivatives, i.e., functions that are members of the Sobolev space of functions over $[a, b]$:

$$W_2^m = \{f(t) : f(t), f'(t), \dots, f^{(m-1)}(t) \text{ are absolutely continuous, } \int_a^b \{f^{(m)}(t)\}^2 dt < \infty\}.$$

Given $f(t) \in W_2^m$, for a fixed λ , the minimizer \hat{f}_λ of (1.4) turns out to be a natural polynomial spline of degree $(2m - 1)$, with knots at each distinct data points t_1, t_2, \dots, t_n . For example, when $m = 2$, the corresponding \hat{f}_λ is called a natural cubic smoothing spline. A “natural” cubic smoothing spline is a cubic spline function which also satisfies $f''(a) = f''(b) = 0$, that is, the second derivatives at the endpoints of the interval are zero. The

cubic smoothing spline is the most popular smoothing spline in statistical applications. One important reason for its popularity is due to its computational advantages. Green and Silverman (1994) showed that the roughness penalty term $\int_a^b f''(t)^2 dt$ associated with the natural cubic spline can be written as

$$\int_a^b f''(t)^2 dt = f^T K f,$$

where $f = \{f(t_1), \dots, f(t_n)\}^T$ is the vector of $f(\cdot)$ evaluated at the data points, and K is a nonnegative definite matrix, often referred to as the roughness penalty matrix. Let $Y = (Y_1, \dots, Y_n)^T$. Minimization of (1.4) for $m = 2$ is equivalent to minimizing the penalized residual

$$(Y - f)^T(Y - f) + \lambda f^T K f,$$

over $f \in R^n$. The solution to this minimization problem is straightforward. One can easily show that the cubic smoothing spline estimate \hat{f}_λ is a linear smoother, in the sense that it is a linear function of observed data Y , given by

$$\hat{f}_\lambda = (\mathbf{I}_n - \lambda K)^{-1} Y,$$

where \mathbf{I}_n denotes the identity matrix of dimension n and $A_\lambda = (\mathbf{I}_n - \lambda K)^{-1}$ is known as the smoothing matrix. Note that $\mathbf{I}_n - \lambda K$ is an $n \times n$ matrix, so its inversion may be computationally demanding. To further simplify computation, Green and Silverman (1994) showed that the penalty matrix K can be decomposed as $K = QR^{-1}Q^T$, where Q and R are two band matrices defined as follows. Set $h_i = t_{i+1} - t_i$ for $i = 1, \dots, n - 1$. Define Q as an $n \times (n - 2)$ lower tridiagonal matrix with entries q_{ij} being specified as

$$q_{j-1,j} = h_{j-1}^{-1}, \quad q_{j,j} = -(h_{j-1}^{-1} + h_j^{-1}), \quad q_{j+1,j} = h_j^{-1},$$

for $i = 1, \dots, n$ and $j = 2, \dots, n-1$. $q_{ij} = 0$ if $|i-j| \geq 2$. The matrix R is an $(n-2) \times (n-2)$ tridiagonal symmetric matrix with its nonzero entries r_{ij} given as

$$r_{j,j} = \frac{1}{3}(h_{j-1} + h_j), j = 2, \dots, n-1,$$

$$r_{j,j+1} = r_{j+1,j} = \frac{1}{6}h_j, j = 2, \dots, n-2,$$

for $|i-j| < 2$. Due to the sparse property of the band matrices Q and R , \hat{f}_λ can be computed indirectly to avoid the intensive matrix inversion; see Green and Silverman (1994) for more details.

For illustration, the residual sum of squares in (1.4) is used to measure the distance between data and estimates. In many statistical applications, more general forms of measures may be used. For example, the weighted residual sum of squares is often used in weighted least squares analysis; in problems formulated on the basis of likelihood, log-likelihood would be the appropriate measure of fidelity to data. Although the formula of (1.4) varies in different statistical problems, its minimizing solutions all fall into the class of cubic smoothing splines as long as the quadratic penalty term $\int_a^b \{f''(t)\}^2 dt$ is used.

Unlike the regression spline, smoothing splines use all distinct data points as knots. Therefore no choice of knot sequence is needed. However, smoothing splines require the smoothing parameter λ to be known in order to accomplish the estimation of nonparametric functions. Proper selection of the smoothing parameter is essential for good performance of spline estimates. The value of λ could be either chosen, or estimated using methods such as cross-validation, generalized cross-validation (GCV, Wahba, 1990), general maximum likelihood (Wahba, 1985), and restricted maximum likelihood (REML, Zhang et al., 1998). We will use GCV for the smoothing parameter estimation in our SPMM model setting for censored longitudinal data.

Chapter 2

Regression Spline Method for Censored Longitudinal Data

In this chapter, we propose to use two nonparametric regression models, the SPMM (Zhang et al., 1998) and the time-varying coefficient mixed model (Liang et al., 2003), to model the censored longitudinal data with the nonparametric functions estimated by using the regression spline method. As described in Section 1.3.1, regression splines can be viewed as piecewise polynomials, and are generally represented as a linear combination of a set of basis functions. Truncated power basis functions and B-spline basis functions are the two most popular ones utilized in various nonparametric regression analysis. Of them, the B-spline basis function is often preferred in practice due to its many good properties. Therefore, in the following discussion, we will focus our attention on the use of the B-spline smoothing technique for nonparametric function estimation.

There are two sections in Chapter 2. The SPMM is the main focus of the first section which includes, from Section 2.1.1 through Section 2.1.4, an introduction of SPMMs, a detailed description of two estimation procedures proposed for statistical inferences, a simulation study designed to assess the performance of the proposed method, and an illustration of the proposed method with a real dataset. The second section discusses time-varying coef-

ficient mixed models, which is organized in the same way as the first section, including model specification in Section 2.2.1, estimation procedures in Section 2.2.2, simulation studies in Section 2.2.3 and an application to a real dataset in Section 2.2.4. At last, we concluded this chapter with some remarks and possible directions for the future research in Section 2.3.

2.1 Semiparametric Mixed Model

2.1.1 Model Specification

Suppose there are m independent subjects in the data. For the i th subject ($i = 1, 2, \dots, m$), let Y_{ij} be the j th underlying true value of the response of interest ($j = 1, 2, \dots, n_i$) taken at time t_{ij} ($t_{ij} \in [0, T]$). Assume that Y_{ij} is from the following SPMM (Zhang et al., 1998),

$$Y_{ij} = f(t_{ij}) + S_{ij}^T \delta + Z_{ij}^T b_i + \epsilon_{ij}, \quad (2.1)$$

where $f(t)$ is an arbitrary smooth function of t , δ is a $p \times 1$ vector of fixed effects, b_i is a $q \times 1$ vector of subject-specific random effects used to model between-subject variation and hence the induced correlation in the data, S_{ij} and Z_{ij} are the corresponding covariates of fixed effects δ and random effects b_i respectively, and ϵ_{ij} 's are independent measurement errors. Further assume that $b_i \sim N\{\mathbf{0}, D(\phi)\}$ with a positive definite covariance matrix $D(\phi)$ dependent on the vector ϕ of some variance/covariance parameters, ϵ_{ij} is distributed as $N(0, \sigma_\epsilon^2)$, and b_i and ϵ_{ij} 's are independent of each other.

We are interested in making inference on the nonparametric function $f(t)$, the fixed effects δ and the variance components $(\phi, \sigma_\epsilon^2)$.

If all the underlying true values Y_{ij} 's had been measured, we could fit model (2.1) using well-established procedures as discussed in Zhang et al. (1998) or Rice and Wu (2001). However, due to assay sensitivity, Y_{ij} may be left-censored at its detection limit, denoted by c_{ij} (we allow the possibility that different values may have different detection limits).

The presence of censoring makes it challenging to fit model (2.1) for a given sample. A naive approach is to impute the left-censored observation by the detection limit or half of the detection limit and fit model (2.1) to the imputed data using existing procedures. However, such practice lacks statistical justification and does not take the imputation error into account, which leads to biased estimation of and invalid inference on model parameters. In the next subsection, we develop a likelihood based estimation and inference procedure for model parameters.

2.1.2 Estimation Procedures

To estimate the nonparametric function $f(t)$, we follow the regression spline method of Rice and Wu (2001) by approximating $f(t)$ using a linear combination of B-spline basis functions of degree d as given in expression (1.3). B-spline basis functions are determined by the degree d of the corresponding B-spline, the number of interior knots, K , in $[0, T]$ and the location of the knots. Here we choose B-splines with $d = 3$, i.e., the cubic B-spline. Given K and d , the knots are often chosen to be equally spaced time points in $[0, T]$. Alternatively, knots can be selected based on sample quantiles of $\{t_{ij}\}$'s, which is usually preferred in situations when data are not collected at equally spaced time points during the follow-up period (Eubank, 1999; Ramsay, 1988).

Substituting expression (1.3) back into model (2.1), and further denoting $\alpha = (\alpha_1, \dots, \alpha_L)^T$, $B(t) = \{B_1(t), \dots, B_L(t)\}^T$, $\beta = (\alpha^T, \delta^T)^T$ and $X_{ij} = \{B(t_{ij})^T | S_{ij}^T\}^T$, the original model (2.1) reduces to the following working linear mixed model for the underlying true data Y_{ij} ,

$$Y_{ij} = X_{ij}^T \beta + Z_{ij}^T b_i + \epsilon_{ij}, \quad (2.2)$$

where β is the new fixed effects. It is evident that the quality of the estimates of the nonparametric function $f(t)$ and other model parameters depends on the adequacy of the approximation (1.3), and a larger K will yield a better approximation for $f(t)$. On the

other hand, as K increases, the number of model parameters in (2.2) also increases, which may cause greater variance and instability in parameter estimates. In order to balance the bias and variance in the parameter estimates, model selection criterion such as AIC or BIC may be employed to select an optimal value of K . This procedure requires maximum likelihood fitting of model (2.2) to the observed data for several different K 's. The fitting is not straightforward when censoring occurs. Here we present two algorithms for maximum likelihood inference.

Direct Optimization Approach

We adopt the notations of Jacqmin-Gadda et al. (2000) for left-censored longitudinal data. For the i th subject, partition the true response vector Y_i into $[Y_i^o, Y_i^c]$, where Y_i^o refers to an $n_i^o \times 1$ vector of completely observed outcomes and Y_i^c represents an $n_i^c \times 1$ vector of censored observations with c_i being the corresponding vector of censoring values ($n_i = n_i^o + n_i^c$). Similarly, let $X_i = [X_i^o, X_i^c]$, $Z_i = [Z_i^o, Z_i^c]$ and V_i^o be the covariance matrix of Y_i^o . Denote by $\theta = (\beta^T, \phi^T, \sigma_\epsilon)$, the model parameters in the working LMM (2.2). Then the likelihood function of θ can be written as

$$\mathcal{L}(\theta; Data) = \prod_{i=1}^m f(Y_i^o | \theta) P(Y_i^c \leq c_i | Y_i^o; \theta), \quad (2.3)$$

where $Y_i^c \leq c_i$ is interpreted as an element-wise inequality.

Under the working LMM (2.2), the conditional distribution of Y_i^c given Y_i^o is normal with a mean vector denoted by $\mu_{i,c}$ and a variance matrix denoted by $V_{i,c}$. The complexity of the calculation of the probability $P(Y_i^c \leq c_i | Y_i^o; \theta)$ depends on n_i^c , the number of censored data points for subject i . When $n_i^c \leq q$, the dimension of the random effects b_i , we may calculate the probability directly (for known values of θ). When $n_i^c > q$, we can calculate the probability as follows to minimize the computational burden.

It can be shown that, under the working LMM (2.2), $P(Y_i^c \leq c_i | Y_i^o; \theta)$ can be written

as

$$P(Y_i^c \leq c_i | Y_i^o; \theta) = E \left\{ \prod_{j=1}^{n_i^c} \Phi \left(\frac{c_{ij} - X_{ij}^{cT} \beta - Z_{ij}^{cT} b_i}{\sigma_\epsilon} \right) \middle| Y_i^o; \theta \right\}, \quad (2.4)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and the expectation is taken with respect to the conditional distribution of b_i given Y_i^o , which is also normal with a mean vector denoted by $\mu_{b_i|Y_i^o}$ and a variance matrix denoted by $V_{b_i|Y_i^o}$.

The expectation (2.4) can be calculated using numerical integration techniques such as Gauss-Hermit approach (Naylor and Smith, 1982) when the dimension q is low, or Monte Carlo methods such as Gibbs sampling or importance sampling when q is large. Here we consider the situation where q is relatively small and choose Gauss-Hermit approach for calculating the expectation (2.4). Denote the function in the expectation (2.4) by $g(b_i)$, and let L_i be the root of $V_{b_i|Y_i^o}$ by cholesky decomposition. Next define the triples (N, e, w) to be the number of quadrature points, vectors of quadrature nodes and quadrature weights respectively. Then the expectation (2.4) can be approximated by

$$E\{g(b_i) | Y_i^o; \theta\} \approx \sum_{v_1=1}^N \cdots \sum_{v_q=1}^N w_{v_1} \cdots w_{v_q} g\{\mu_{b_i|Y_i^o} + L_i e(v)\},$$

where $e(v) = (e_{v_1}, \dots, e_{v_q})^T$, e_k and w_k are the k th elements of e and w respectively.

With this approximation, the log-likelihood function of θ can be approximated by

$$\begin{aligned} \ell(\theta; Data) \approx & \sum_{i=1}^m \left\{ -\frac{1}{2} \log |V_i^o| - \frac{1}{2} (Y_i^o - X_i^o \beta)^T V_i^{o-1} (Y_i^o - X_i^o \beta) \right\} \\ & + \sum_{i=1}^m \log \left[\sum_{v_1=1}^N \cdots \sum_{v_q=1}^N w_{v_1} \cdots w_{v_q} g\{\mu_{b_i|Y_i^o} + L_i e(v)\} \right]. \end{aligned} \quad (2.5)$$

Then the working maximum likelihood estimators (MLEs) of the model parameters θ can be obtained via standard optimization procedures. We use the SAS IML optimizer NLPQN

to maximize (2.5). We select K that minimizes the following BIC criterion

$$BIC = -2\ell(\hat{\theta}; Data) + K\log(n),$$

where $\hat{\theta}$ is the maximizer of (2.5) for a given K , and $n = \sum_{i=1}^m n_i$ is the total number of observations in the sample. Conservatively, one may also use $n = \sum_{i=1}^m n_i^o$, the total number of un-censored observations. After K is selected, the inference of the model parameters can be based on the negative inverse of the Hessian matrix from (2.5).

EM Approach

The preceding direct optimization approach is straightforward. However, the resulting estimates may not be stable due to the complicated form of the approximated log-likelihood function (2.5). An alternative to the direct optimization approach is the expectation maximization (EM) algorithm. EM is an iterative optimization method often used for finding MLEs when there are incomplete or missing data (Dempster et al., 1977). EM alternates between an expectation (E) step and a maximization (M) step. The E-step computes the expected value of the complete data log-likelihood with respect to the unobserved data given the observed data and the current parameter estimates, and the M-step computes the updates of model parameters by maximizing the expected log-likelihood obtained from the E step. These two steps are iterated until convergence is reached. A major advantage of the EM algorithm is that it is stable since usually the M-step produces closed form updates for most parameter estimates and the algorithm will always increase the log-likelihood function.

To implement the EM algorithm, first we treat the censored response values $\{Y_i^c\}$ and the random effects $\{b_i\}$ as missing data, and consider the situation where D , the variance of the random effects b_i , is unstructured. Then the complete data log-likelihood function is

equal to

$$\begin{aligned}\ell(\theta; Y, b) &= -\frac{n}{2} \log \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} (Y - X\beta - Zb)^T (Y - X\beta - Zb) \\ &\quad - \frac{m}{2} \log |D| - \frac{1}{2} \sum_{i=1}^m b_i^T D^{-1} b_i,\end{aligned}\tag{2.6}$$

where $Y = (Y_1^T, \dots, Y_m^T)^T$ and $b = (b_1^T, \dots, b_m^T)^T$ with (Y, b) referred to as the “complete” data, X and Z are the matrix notations similarly defined. In the E-step, given the observed data and the current values of the model parameter estimates $\theta^{(r)}$, the expected log-likelihood function Q can be derived as follows

$$\begin{aligned}Q(\theta|\theta^{(r)}) &= E\{\ell(\theta; Y, b)|Y^o, Y^c \leq c, \theta^{(r)}\} \\ &= -\frac{n}{2} \log \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} E\{(Y - X\beta - Zb)^T (Y - X\beta - Zb)|Y^o, Y^c \leq c, \theta^{(r)}\} \\ &\quad - \frac{m}{2} \log |D| - \frac{1}{2} \sum_{i=1}^m E(b_i^T D^{-1} b_i | Y_i^o, Y_i^c \leq c_i, \theta^{(r)}),\end{aligned}\tag{2.7}$$

where $c = (c_1^T, \dots, c_m^T)^T$. Denote $\hat{Y} = E(Y|Y^o, Y^c \leq c, \theta^{(r)})$, $\hat{b} = E(b|Y^o, Y^c \leq c, \theta^{(r)})$ and $M_i = E(b_i b_i^T | Y_i^o, Y_i^c \leq c_i, \theta^{(r)})$. Since D is unstructured, maximizing $Q(\theta|\theta^{(r)})$ with respect to β, D and σ_ϵ^2 gives the following closed form updates

$$\begin{aligned}\hat{\beta}^{(r+1)} &= (X^T X)^{-1} X^T (\hat{Y} - Z\hat{b}), \\ \hat{D}^{(r+1)} &= \frac{1}{m} \sum_{i=1}^m M_i, \\ \hat{\sigma}_\epsilon^{2(r+1)} &= \frac{1}{n} \sum_{i=1}^m \left[\sum_{j=1}^{n_i^o} E\{(Y_{ij}^o - X_{ij}^{oT} \beta - Z_{ij}^{oT} b_i)^2 | Y_i^o, Y_i^c \leq c_i, \theta^{(r)}\} \right. \\ &\quad \left. + \sum_{j=1}^{n_i^c} E\{(Y_{ij}^c - X_{ij}^{cT} \beta - Z_{ij}^{cT} b_i)^2 | Y_i^o, Y_i^c \leq c_i, \theta^{(r)}\} \Big|_{\beta=\hat{\beta}^{(r+1)}} \right].\end{aligned}$$

In order to get the above updates for the model parameter estimates, we need to calculate \hat{Y}_i^c , \hat{b}_i , M_i , $E(Y_{ij}^c b_i | Y_i^o, Y_i^c \leq c_i, \theta^{(r)})$ and $E\{(Y_{ij}^c)^2 | Y_i^o, Y_i^c \leq c_i, \theta^{(r)}\}$. Again we apply the

Gauss-Hermit approach to numerically evaluate these expectations. For example, when $q = 1$,

$$\begin{aligned}
\widehat{b}_i &= E(b_i | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}) \\
&= \int b_i f(b_i | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}) db_i \\
&= \frac{\int b_i f(Y_i^o, Y_i^c \leq c_i | b_i; \theta^{(r)}) f(b_i | \theta^{(r)}) db_i}{\int f(Y_i^o, Y_i^c \leq c_i | b_i; \theta^{(r)}) f(b_i | \theta^{(r)}) db_i} \\
&= \frac{\int b_i g_1(b_i) f(b_i | \theta^{(r)}) db_i}{\int g_1(b_i) f(b_i | \theta^{(r)}) db_i},
\end{aligned}$$

where $g_1(b_i)$ has the following expression

$$\begin{aligned}
g_1(b_i) &= f(Y_i^o, Y_i^c \leq c_i | b_i; \theta^{(r)}) = f(Y_i^o | b_i; \theta^{(r)}) P(Y_i^c \leq c_i | b_i; \theta^{(r)}) \\
&= \prod_{j=1}^{n_i^o} \varphi \left(\frac{Y_{ij}^o - X_{ij}^{oT} \beta^{(r)} - Z_{ij}^{oT} b_i}{\sigma_\epsilon^{(r)}} \right) \prod_{j=1}^{n_i^c} \Phi \left(\frac{c_{ij} - X_{ij}^{cT} \beta^{(r)} - Z_{ij}^{cT} b_i}{\sigma_\epsilon^{(r)}} \right),
\end{aligned}$$

and $\varphi(\cdot)$ is the density function of the standard normal distribution. Both expectations in the numerator and denominator of \widehat{b}_i can be calculated using the Gauss-Hermit approach. Similar strategy can be used to calculate other expectations required in the EM iterations. A detailed derivation is given in Appendix B.

The MLEs of θ are obtained by iterating the E-step and M-step until convergence. The variance-covariance matrix of the EM estimates are computed in exactly the same way as in the direct optimization approach. It can be seen that the above EM algorithm will be preferred when the dimension of random effects is small.

2.1.3 Simulation

We conducted a simulation study to evaluate the performance of our procedures. For comparison, we also implemented the *ad hoc* imputation method, in which half of the detection limit is substituted for the censored observations.

We generated data according to the following SPMM with a random intercept only

$$Y_{ij} = f(t_{ij}) + X_{ij}^T \beta + b_i + \epsilon_{ij}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n_i, \quad (2.8)$$

and the following SPMM with a random intercept and a random slope of t

$$Y_{ij} = f(t_{ij}) + X_{ij}^T \beta + b_{0i} + t_{ij} b_{1i} + \epsilon_{ij}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n_i. \quad (2.9)$$

In all cases, a total of 100 equally-spaced time points $\{t_{ij}\}$ was generated in $[0,8]$. Each simulated data set consists of $m = 100$ subjects, with each subject having $n_i = 5$ repeated measurements uniformly located at these time points. For instance, the first subject was measured at the ordered time points $\{1, 21, 41, 61, 81\}$, the second subject was at $\{2, 22, 42, 62, 82\}$, and so on. We took the true function $f(t)$ to be $f(t) = \sin(t)$, generated covariate X_{ij} from $N(2, 1)$ and set the true value of β to be 1.5. The measurement error ϵ was generated from $N(0, 1)$. In model (2.8), the random intercept b_i was generated from $N(0, 0.5^2)$, and in model (2.9), the bivariate random effects $(b_{0i}, b_{1i})^T$ were generated from $N(\mathbf{0}, D)$, with $D_{00} = \sigma_{b_0}^2 = 0.5^2$, $D_{01} = -0.045$, and $D_{11} = \sigma_{b_1}^2 = 0.3^2$ (so $\rho = \text{corr}(b_{0i}, b_{1i}) = -0.3$). Two detection limits were chosen such that the censoring percentages of Y were around 17% and 34%, the same levels reported by Jacqmin-Gadda et al. (2000). In the simulation, we used the cubic B-spline basis functions with equally spaced knots in $[0, 8]$ to estimate $f(t)$. It is time consuming to search for the optimal number of knots (K) that minimize the observed BIC (see Section 2.1.2), and our preliminary simulation indicated that the optimal K ranged from 3 to 5 for the current setting, we hence conservatively set $K = 6$ for our simulation study.

Tables 2.1 and 2.2 summarize the simulation results based on 200 Monte Carlo data sets for the model parameters $(\beta, D, \sigma_\epsilon)$ from model (2.8) and model (2.9), respectively. As expected, the naive imputation method produced biased parameter estimates and incorrect standard error estimates for the parameter estimates, yielding invalid inference. The direct

optimization and EM algorithm produced almost identical results when data were generated from model (2.8). There are some discrepancies, however, in the results when data were generated from the more complicated model (2.9). This is partly due to the fact that the likelihood function of model (2.9) is complicated, and hence the direct optimization in some cases is not stable. In general, both algorithms produced unbiased estimates of the fixed effects and variance components for all cases, and the standard error estimates agreed well with the sampling standard deviation of these parameter estimates, yielding empirical coverage probabilities of the Wald confidence intervals close to the nominal level (95%).

Figure 2.1 to 2.4 display and compare the estimated nonparametric function $\hat{f}(t)$, its empirical bias and point-wise 95% coverage probabilities of the true fixed function $f(t)$ across three methods for the different scenarios considered in this study. Figure 2.1 and 2.2 are generated from model (2.8) with data consisting of 17% and 34% censored observations respectively. Figure 2.3 and 2.4 are generated similarly from model (2.9). It is clear that in all cases, the naive imputation method produced significantly biased (upwards) estimate $\hat{f}(t)$, leading to invalid confidence intervals for $f(t)$ and yielding low coverage probabilities for the confidence intervals. In contrast, both the direct optimization approach and the EM algorithm produced virtually unbiased nonparametric function estimate $\hat{f}(t)$ and correct standard error estimates (results not shown). The coverage probabilities of the point-wise Wald confidence intervals of $f(t)$ are close to the nominal level (95%) with slightly increased fluctuation at the high censoring rate. Overall, there is no noticeable difference in the results on $f(t)$ between the direct optimization procedure and the EM algorithm. In summary, our simulation demonstrates that SPMs, coupled with the direct optimization procedure or the EM algorithm, provide efficient parameter estimates and inference in the analysis of left-censored longitudinal data.

Table 2.1: Simulation results from model (2.8). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.

Censoring%	Parameters	Method	Mean	SD	SE	Bias	CP(95%)
17%	β	Imputation	1.364	0.051	0.046	-0.136	.20
		Direct Opt.	1.505	0.056	0.054	0.005	.95
		EM	1.505	0.056	0.054	0.005	.95
	σ_b	Imputation	0.448	0.061	0.057	-0.052	.80
		Direct Opt.	0.491	0.067	0.069	-0.009	.96
		EM	0.491	0.067	0.069	-0.009	.96
	σ_ϵ	Imputation	0.963	0.037	0.066	-0.037	.98
		Direct Opt.	0.989	0.038	0.039	-0.011	.95
		EM	0.989	0.038	0.039	-0.011	.95
34%	β	Imputation	1.314	0.049	0.049	-0.186	.05
		Direct Opt.	1.506	0.066	0.062	0.006	.94
		EM	1.506	0.066	0.062	0.006	.94
	σ_b	Imputation	0.428	0.068	0.059	-0.072	.71
		Direct Opt.	0.495	0.071	0.075	-0.005	.97
		EM	0.495	0.071	0.075	-0.005	.97
	σ_ϵ	Imputation	1.025	0.037	0.075	0.026	.99
		Direct Opt.	0.986	0.044	0.044	-0.014	.94
		EM	0.986	0.044	0.044	-0.014	.94

Table 2.2: Simulation results from model (2.9). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.

Censoring%	Parameters	Method	Mean	SD	SE	Bias	CP(95%)
17%	β	Imputation	1.326	0.057	0.050	-0.174	.09
		Direct Opt.	1.494	0.057	0.057	-0.006	.96
		EM	1.494	0.057	0.057	-0.006	.96
	σ_{b0}	Imputation	0.469	0.168	0.177	-0.032	.96
		Direct Opt.	0.491	0.159	0.187	-0.009	.97
		EM	0.498	0.150	0.184	-0.002	.98
	ρ	Imputation	-0.280	0.286	0.771	0.020	.92
		Direct Opt.	-0.273	0.280	0.341	0.027	.93
		EM	-0.277	0.260	0.315	0.023	.93
	σ_{b1}	Imputation	0.252	0.028	0.029	-0.048	.61
		Direct Opt.	0.297	0.030	0.033	-0.003	.97
		EM	0.297	0.030	0.033	-0.003	.97
	σ_ϵ	Imputation	0.961	0.038	0.039	-0.039	.82
		Direct Opt.	0.981	0.041	0.045	-0.019	.92
		EM	0.982	0.042	0.045	-0.018	.93
34%	β	Imputation	1.262	0.057	0.053	-0.238	.02
		Direct Opt.	1.499	0.072	0.066	-0.001	.94
		EM	1.499	0.072	0.066	-0.001	.94
	σ_{b0}	Imputation	0.434	0.174	0.207	-0.066	.98
		Direct Opt.	0.505	0.152	0.201	0.005	.97
		EM	0.507	0.149	0.197	0.007	.98
	ρ	Imputation	-0.280	0.382	0.700	0.020	.94
		Direct Opt.	-0.248	0.315	0.423	0.052	.94
		EM	-0.250	0.305	0.401	0.050	.94
	σ_{b1}	Imputation	0.237	0.029	0.030	-0.063	.47
		Direct Opt.	0.294	0.038	0.036	-0.006	.95
		EM	0.294	0.038	0.036	-0.006	.95
	σ_ϵ	Imputation	1.027	0.041	0.042	0.027	.91
		Direct Opt.	0.979	0.051	0.051	-0.021	.93
		EM	0.980	0.051	0.051	-0.021	.92

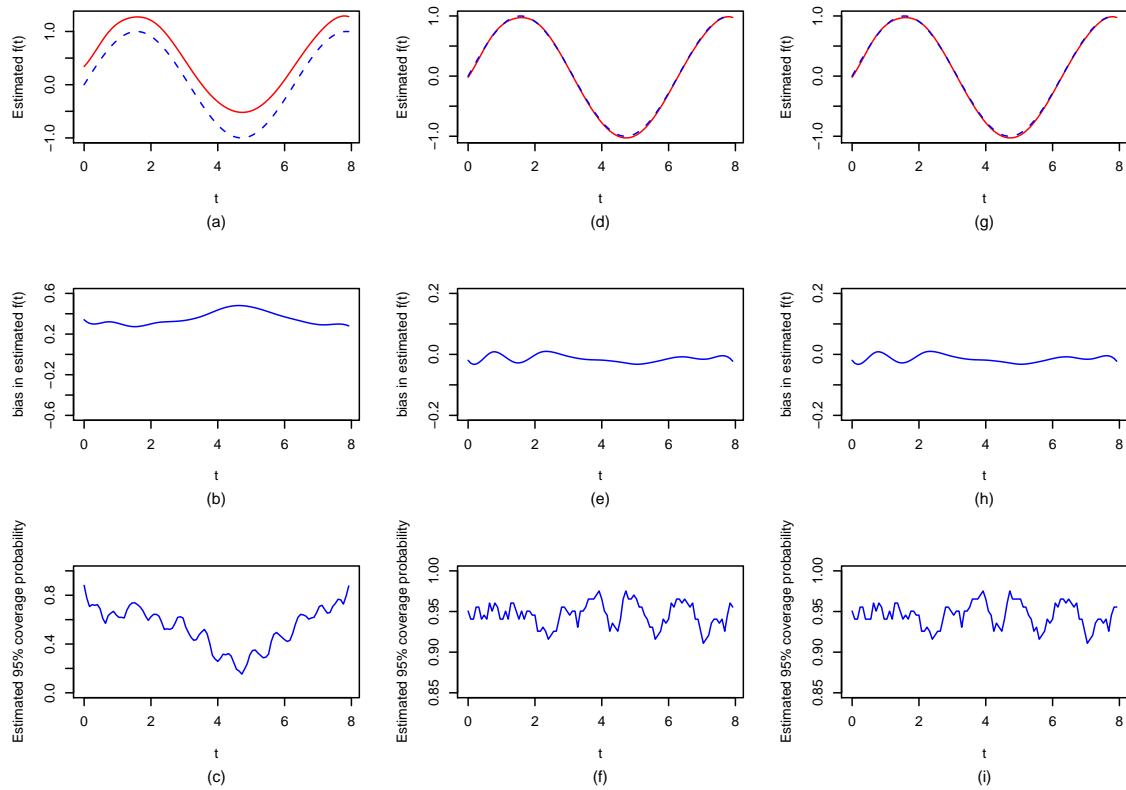


Figure 2.1: Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (2.8) with 17% censoring. (a)-(c) for the imputation method; (d)-(f) for the direct optimization approach; (g)-(i) for the EM approach.

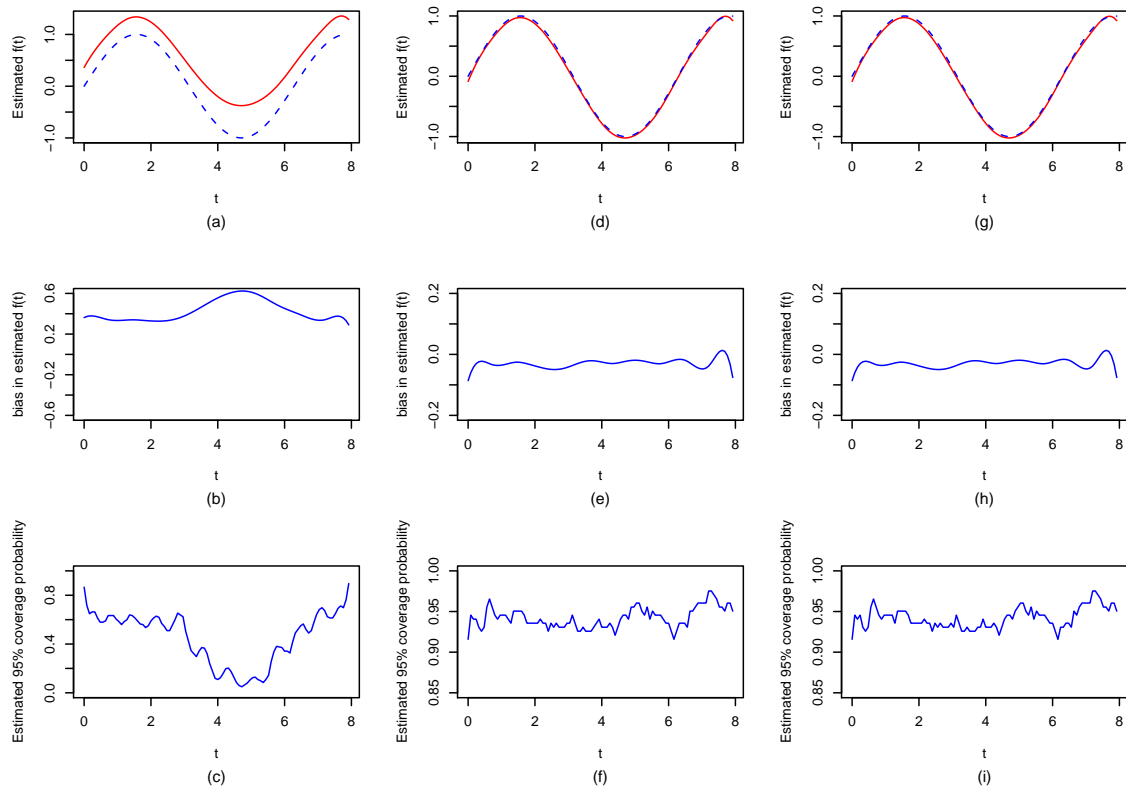


Figure 2.2: Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (2.8) with 34% censoring. (a)-(c) for the imputation method; (d)-(f) for the direct optimization approach; (g)-(i) for the EM approach.

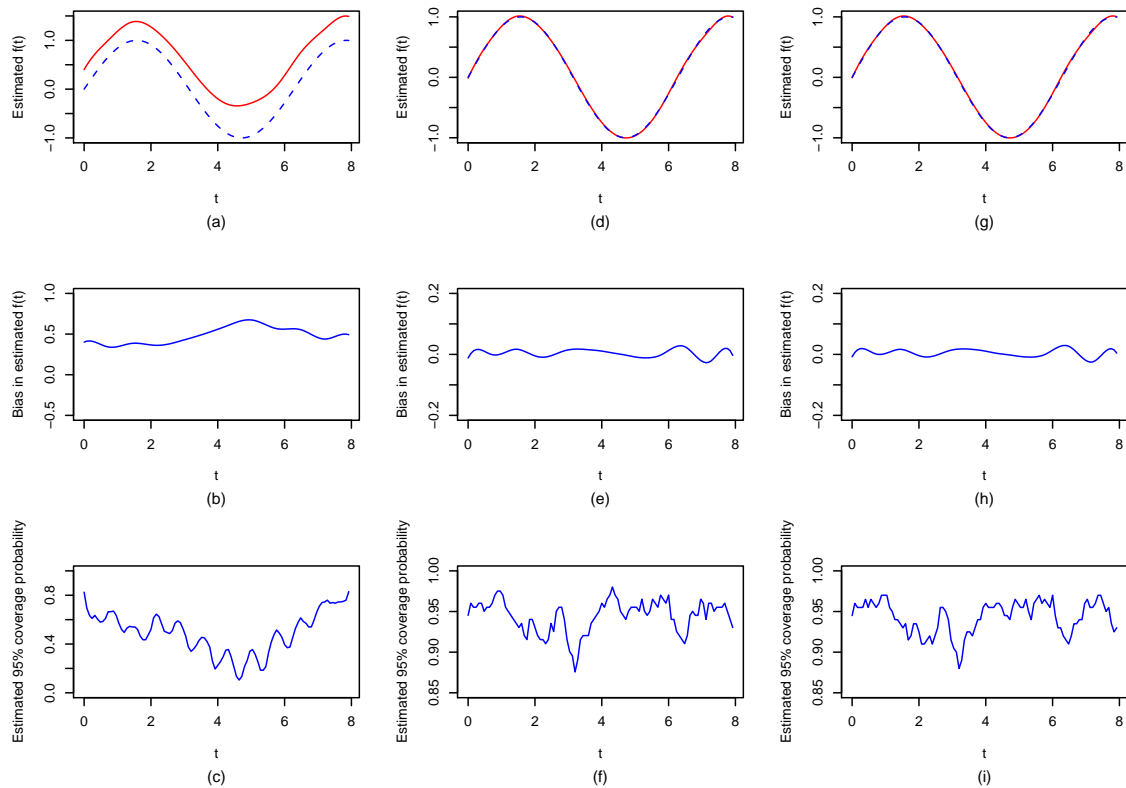


Figure 2.3: Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (2.9) with 17% censoring. (a)-(c) for the imputation method; (d)-(f) for the direct optimization approach; (g)-(i) for the EM approach.

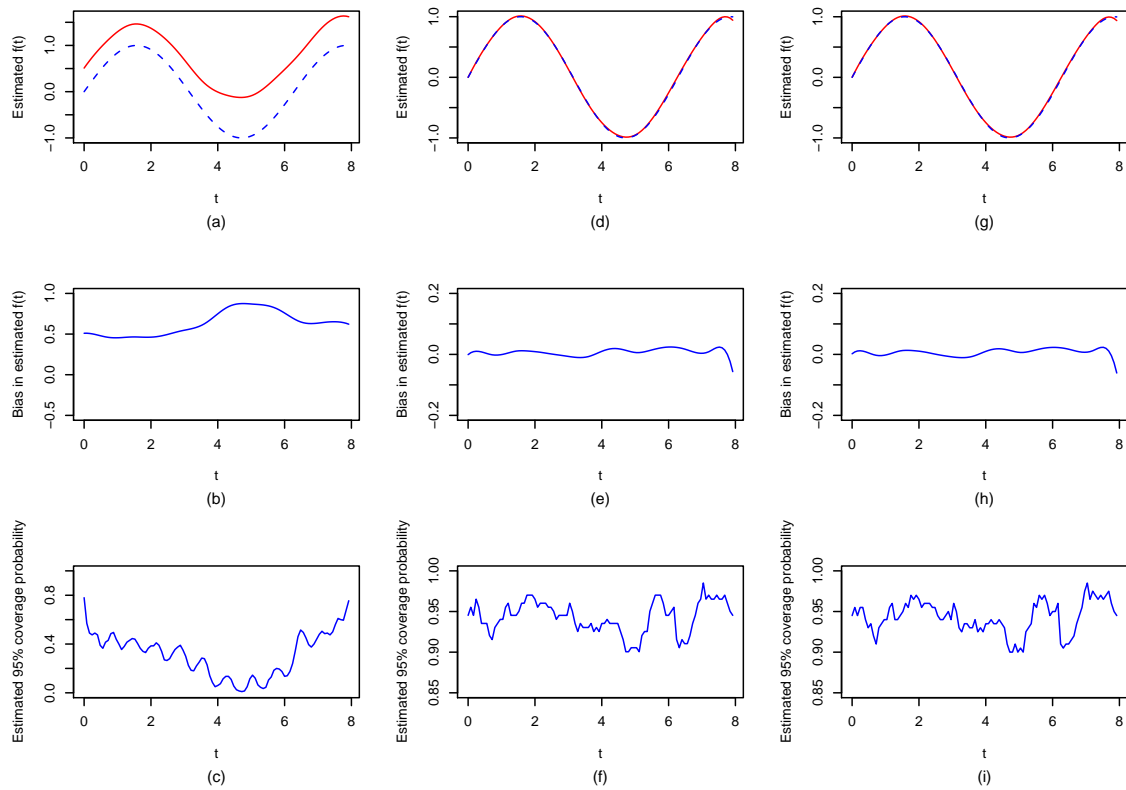


Figure 2.4: Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (2.9) with 34% censoring. (a)-(c) for the imputation method; (d)-(f) for the direct optimization approach; (g)-(i) for the EM approach.

2.1.4 Application

We apply the proposed EM approach to the HIV viral load data from the AIDS Clinical Trials Group (ACTG) 398 study. One of the primary objectives of this study was to determine whether the dual protease inhibitor (DPI) regimen improves anti-viral efficacy of the existing single protease inhibitor (SPI) regimen in patients with anti-retroviral treatment failure. A total of 481 HIV-infected patients with baseline viral load above 1000 copies/ml were enrolled, of whom, 324 patients were assigned to receive the DPI regimen and 157 to receive the SPI regimen. Patients' plasma HIV-1 RNA copies (viral load) were measured at time zero, and approximately at weeks 2, 4, 8, 16 and 24 thereafter. Since each individual patient might not exactly follow the designed schedule for measurements, the actual clinical visit times for individual patients varied from each other, and sometimes missing clinical visits also occurred. A total of 2626 measurements were available in the dataset, with the number of observations from each individual ranging from 1 to 6. Due to the quantification limit of the assay, 22% of the observations were censored at the detection limit of 200 copies/ml. In addition to the treatment assignment, some patients previously also received non-nucleoside reverse transcriptase inhibitors (NNRTIs), which was believed to have an impact on the viral load response to the DPI/SPI regimen. More detailed description about this study can be found in Hammer et al. (2002).

The same dataset has been analyzed using a semiparametric time-varying coefficients model by Sun and Wu (2005), where the effects of time and NNRTI were modeled nonparametrically and the effect of the treatment was assumed to follow a given parametric form. In their study, however, the censoring feature of the data was ignored and they only restricted their analysis to those responses that were within the detectable range. We shall see that by discarding censored observations, such an analysis can significantly bias the statistical inferences, including the estimation of the treatment effect. We attempt to re-analyze this dataset and compare the anti-retroviral effects between the DPI and SPI regimen while taking into account the censoring process.

Our preliminary investigation, along with the analysis presented in Sun and Wu (2005), indicates that the NNRTI effect remains nearly constant over the entire treatment course. Thus we propose to fit the data with the following SPMM,

$$Y_{ij} = f(t_{ij}) + \beta_1 X_{1i} t_{ij} I(t_{ij} > 0) + \beta_2 X_{2i} + Z_{ij} b_i + \epsilon_{ij}, \quad (2.10)$$

where Y_{ij} is $\log_{10}(\text{RNA copies/ml})$, a transformation commonly used for HIV data analysis, $X_{1i} = 1$ if patients are from the DPI group, and $X_{1i} = 0$ if patients from SPI group, X_{2i} is an indicator variable for NNRTI (1, if patients had prior NNRTI treatments; 0, if not), and $t_{ij} = \log_{10}(\text{day of actual visit} + 40) - \log_{10}(40)$ is the transformed time (day) after initiation of the treatment. This transformation makes the time points more evenly distributed and thus helps stabilize computation (Sun and Wu, 2005). The indicator function $I(t_{ij} > 0)$ ensures that there is no treatment effect before the time 0 because the baseline HIV measurements for some patients were collected before the scheduled time zero ($t < 0$ in those cases). Note that compared with the model used by Sun and Wu (2005), this model has an extra random effects term $Z_{ij} b_i$, which allows for a more appropriate account for the between-subject variation of longitudinal data. For this dataset, we consider cases where the dimension of b_i is less than 3 for computational convenience. To examine the impact of different formulations of random effects on the model parameter estimates, we have considered three scenarios: $Z_{ij} = (1)$, $Z_{ij} = (1 - X_{1i}, X_{1i})$ and $Z_{ij} = (1, t_{ij})$. The first scenario is the simplest case containing a random intercept only. The second one assumes that the random intercepts in the two treatment arms have different variances, and the last formulation is the most commonly used one in many practical data analysis, which has a random intercept as well as a random slope of time.

We set the location of knots for the B-spline basis functions to be the sample quantiles of t_{ij} 's (the transformed time), and let the optimal number of knots be determined by the BIC model selection criterion. For comparison and illustration purpose, we also conducted

the analysis using the imputation method, where the censored observations were imputed by either the detection limit (referred to as IM hereafter) or the half of the detection limit (referred to as Half IM hereafter).

Table 2.3: Variance component estimates for ACTG398 viral load data from model (2.10).

Random Effects & BIC	Model Parameters	IM		Half IM		EM	
		Estimates	SE	Estimates	SE	Estimates	SE
$Z_{ij} = (1)$	σ_b	0.846	0.031	1.032	0.039	0.979	0.037
	σ_e	0.707	0.011	0.940	0.014	0.856	0.015
BIC		6690.4		8116.6		6992.6	
$Z_{ij} = (1 - X_{1i}, X_{1i})$	σ_{b0}	0.843	0.054	1.005	0.066	0.948	0.063
	σ_{b1}	0.847	0.038	1.045	0.048	0.993	0.046
	σ_e	0.707	0.011	0.940	0.014	0.856	0.015
BIC		6692.4		8118.3		6994.3	
$Z_{ij} = (1, t_{ij})$	σ_{b0}	0.676	0.030	0.762	0.036	0.693	0.034
	σ_{b1}	1.518	0.075	2.229	0.100	2.500	0.126
	ρ	0.118	0.071	0.076	0.069	0.097	0.069
	σ_e	0.579	0.010	0.726	0.013	0.662	0.013
BIC		6255.6		7477.0		6322.1	

We first select the number of knots (K) using the observed BIC described in Section 2.1.2. For all the scenarios considered here, the SPMMs with knots $K = 5$ are optimal regardless of the random effect formulations. Table 2.3 presents the variance components estimates of model (2.10) under three different random formulations from three analysis procedures. The model with the random term $Z_{ij} = (1, t_{ij})$ seems to be most favorable among the three formulations which is indicated by its minimum BIC score. Note that the estimates of σ_{b0} and σ_{b1} in the second formulation $Z_{ij} = (1 - X_{1i}, X_{1i})$ are nearly identical, suggesting that the random intercepts in the two treatment arms have the same variance. Thus, the first and second formulations are indeed equivalent in this case. From Table 2.3, one can also observe that the imputation method (IM) always resulted in smaller overall variance than the EM approach, while the half imputation method (Half IM) in general

Table 2.4: Parameter estimates for ACTG398 viral load data from model (2.10).

Random Effects	Model Parameters	IM		Half IM		EM	
		Estimates	SE	Estimates	SE	Estimates	SE
$Z_{ij} = (1)$	β_1	-0.481	0.109	-0.754	0.143	-0.680	0.136
	β_2	0.610	0.083	0.831	0.102	0.763	0.097
$Z_{ij} = (1 - X_{1i}, X_{1i})$	β_1	-0.481	0.109	-0.752	0.142	-0.681	0.136
	β_2	0.609	0.083	0.830	0.102	0.761	0.097
$Z_{ij} = (1, t_{ij})$	β_1	-0.416	0.179	-0.674	0.251	-0.726	0.281
	β_2	0.459	0.073	0.553	0.086	0.454	0.080

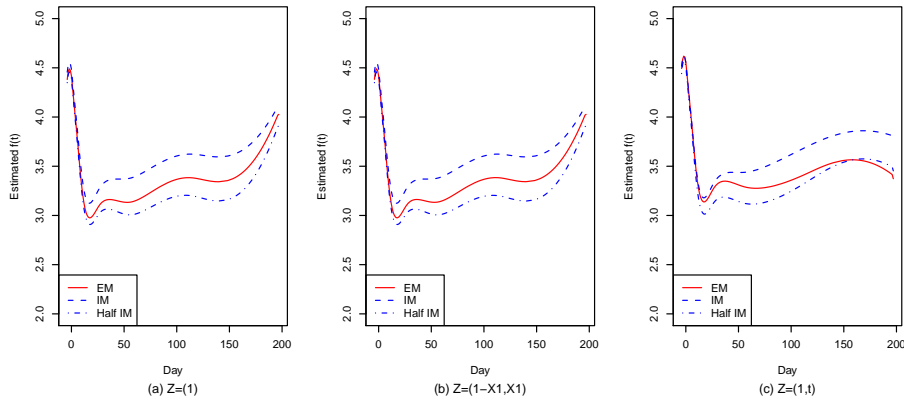


Figure 2.5: Estimated $\hat{f}(t)$ for different random effect formulations of model (2.10) from three estimation procedures.

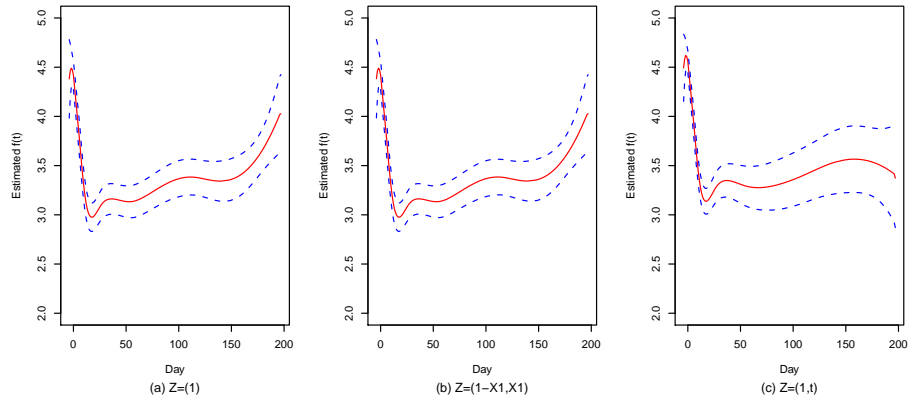


Figure 2.6: Estimated $\hat{f}(t)$ (solid line) along with its confidence interval (dashed lines) for different random effect formulations of model (2.10) from the EM method.

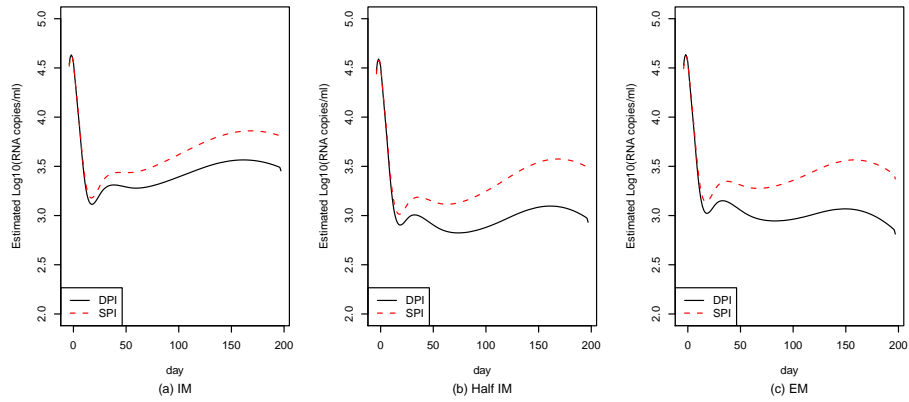


Figure 2.7: Estimated population means of $\log_{10}(\text{RNA copies/ml})$ for the two treatment (DPI versus SPI) groups from the IM, Half IM and EM methods, for patients who had no prior NNRTI treatment ($X_{2i}=0$).

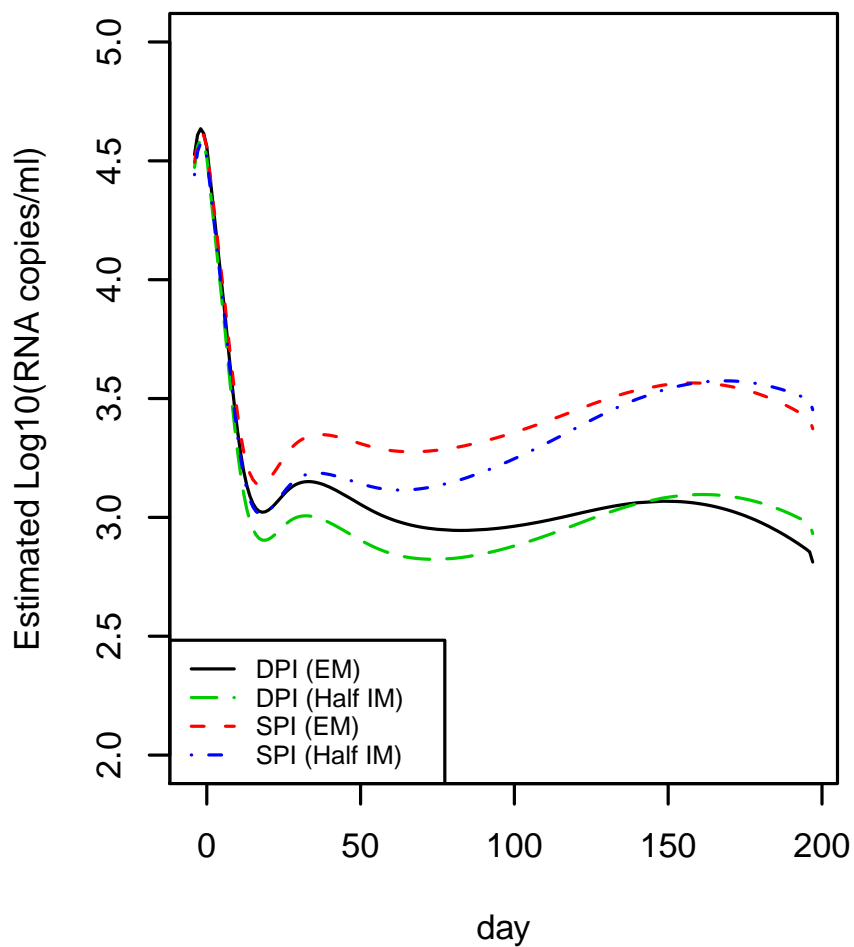


Figure 2.8: Estimated population means of $\log_{10}(\text{RNA copies/ml})$ for the two treatment (DPI versus SPI) groups from the Half IM and EM methods, for patients who had no prior NNRTI treatment ($X_{2i}=0$).

caused inflated estimate of variance.

The estimates of β 's in model (2.10) are listed in Table 2.4. Again the formulations $Z_{ij} = (1)$ and $Z_{ij} = (1 - X_{1i}, X_{1i})$ give the nearly identical estimates of β 's. These estimates, however, observably differ from the estimates of the formulation $Z_{ij} = (1, t_{ij})$. This implies that comparing with the models of $Z_{ij} = (1)$ and $Z_{ij} = (1 - X_{1i}, X_{1i})$, the model with the random term $Z_{ij} = (1, t_{ij})$ tends to account for the data variation properly, thus leading to more accurate estimates of the treatment effects. The estimates of β 's indicate a positive relation between $\log_{10}(\text{RNA copies/ml})$ and NNRTI and a negative relation between $\log_{10}(\text{RNA copies/ml})$ and the treatment regimen. Table 2.4 also shows that the estimates of the IM method are different from that of the EM method, especially the estimate of β_1 , the treatment effect. Note that since the IM method did not correct for the missing information attributable to censoring, its standard errors of the estimates are consistently biased downward. The half IM method, to some extent, improved the estimate of β_1 , but its estimate of β_2 deviates significantly from the corresponding EM estimates.

Figure 2.5 displays and compares the estimates of $f(t)$ from two imputation methods and our EM method under three different random effects settings. As expected, the formulations $Z_{ij} = (1)$ and $Z_{ij} = (1 - X_{1i}, X_{1i})$ give visually indistinguishable results. Even though the overall trend of $\hat{f}(t)$ among the three random formulations are quite similar, the estimated curve from the formulation $Z_{ij} = (1, t_{ij})$ shaped differently in the middle and tail sections. It is evident from Figure 2.5 that compared with the EM approach, the IM method consistently overestimates $f(t)$ for the entire treatment period, whereas the half IM method appears to over-correct the missing information and thus causes underestimated $f(t)$. We also present the point-wise 95% confidence intervals associated with the estimated $f(t)$ from the EM method (Figure 2.6). The $\hat{f}(t)$ curve reflects the average trajectory of $\log_{10}(\text{RNA copies/ml})$ over time in the population who had no prior NNRTI experience and only received the SPI regimen, which also represents the baseline level of $\log_{10}(\text{RNA copies/ml})$ with respect to the treatment effect of DPI. Patients in the SPI group experienced sharp drop in blood HIV

viral load during the first 2 weeks. After that the viral load level maintains a steady rise, with gradually increased variation indicated by its point-wise confidence interval until the end of follow-ups.

To further demonstrate the importance of properly accounting for censoring in assessing the treatment effect, we display the estimated population curves of $\log_{10}(\text{RNA copies/ml})$ for the two treatments (DPI versus SPI) groups based on the optimal model, i.e., $Z_{ij} = (1, t_{ij})$ (see Figure 2.7 and Figure 2.8). To be concise, we only present the results from patients receiving no prior NNRTI treatment. Results from the EM approach show that the two regimens are equally effective in bringing down the HIV virus levels right after the initiation of the treatment (Figure 2.7, right panel). After 2 weeks, the viral load gradually rises in the SPI treated patients, whereas it remains stable or continues to decrease in patients receiving DPI regimen, indicating that overall the DPI is superior over the SPI in suppressing virus replication. In comparing with the EM approach, the IM method leads to higher estimated population means of viral load in both treatment groups, and also greatly attenuates the magnitude of the treatment difference between the DPI and SPI regimens. Although the half IM method improves the estimate and its results seem quite pleasing in Figure 2.7, it is clear that a significant downward trend remains in the estimates for both treatment arms (Figure 2.8).

Model (2.10) provides a reasonable fit to these HIV data and well reflects the trajectory of viral load changes over the entire treatment course. However, it is noticed that model (2.10) assumes that the treatment effect follows a linear function of time. A question raised naturally is whether such a parametric assumption is too restrictive in practice. In the next section, we tried to investigate a time-varying coefficient mixed model, where the treatment effect is modeled by an unspecified arbitrary function of time.

2.2 Time-varying Coefficient Mixed Model

2.2.1 Model Specification

Similarly, consider a longitudinal study of m independent subjects. Let Y_{ij} denote the j th underlying true observation of the response from the i th subject ($i = 1, 2, \dots, m$; $j = 1, 2, \dots, n_i$), and X_i , a scalar, be the covariate of particular interest, which may have potentially complex dynamic relationship with the response Y_{ij} . To explore such a dynamic feature of the X_i effect, we consider the following time-varying coefficient mixed model:

$$Y_{ij} = f_0(t_{ij}) + X_i f_1(t_{ij}) + S_{ij}^T \delta + Z_{ij}^T b_i + \epsilon_{ij}, \quad (2.11)$$

where $f_0(t)$ and $f_1(t)$ are two arbitrary but smooth functions of t , representing a nonparametric time effect and a time-varying fixed effect of the covariate X_i respectively. Similar to the notations previously defined in model (2.1), S_{ij} is a $p \times 1$ covariate vector of other fixed effects with the corresponding coefficients denoted by δ , Z_{ij} is a known $q \times 1$ vector associated with the individual random effects $b_i \sim N\{\mathbf{0}, D(\phi)\}$, where $D(\phi)$ is a positive definite $q \times q$ matrix specified by a vector of unknown parameters ϕ , and ϵ_{ij} 's are random measurement errors assumed to have a distribution $N(0, \sigma_\epsilon^2)$ and to be independent of b_i .

Model (2.11) is closely related to the varying-coefficient models introduced by Hoover et al. (1998) and Wu et al. (1998). It contains as a special case the SPMMs introduced in the previous section, where only the intercept is allowed to be time-varying and other coefficients are all constants.

2.2.2 Estimation Procedure: Hybrid EM/Quasi-Newton Approach

To fit model (2.11), we again apply the regression spline method and approximate the time-varying coefficient curves through a linear combination of B-spline basis functions as shown in equation (1.3). To accommodate the fact that each individual coefficient curves may require different amounts of smoothing, K , the number of knots, as well as the location

of the knots are allowed to be selected differently in order to provide adequate smoothing for each coefficient curves. Suppose the location of knots is selected based on sample quantiles of t_{ij} 's, and K_0 and K_1 are the numbers of knots chosen for $f_0(t)$ and $f_1(t)$ respectively, through the BIC model selection criterion. Similar as in Section 2.1.2, let $\{B_{sl}(t)\}$ ($l = 1, \dots, L_s; s = 0, 1$) be L_s B-spline basis functions evaluated at t , and α_{sl} 's be the associated coefficients. Then one can approximate model (2.11) by

$$\begin{aligned} Y_{ij} &= \sum_{l=1}^{L_0} B_{0l}(t_{ij})\alpha_{0l} + X_i * \sum_{l=1}^{L_1} B_{1l}(t_{ij})\alpha_{1l} + S_{ij}^T\delta + Z_{ij}^T b_i + \epsilon_{ij} \\ &= B_0(t_{ij})^T \alpha_0 + X_i * B_1(t_{ij})^T \alpha_1 + S_{ij}^T\delta + Z_{ij}^T b_i + \epsilon_{ij}. \end{aligned}$$

Again, if we redefine $X_{ij} = \{B_0(t_{ij})^T | X_i * B_1(t_{ij})^T | S_{ij}^T\}^T$ as the design matrix of fixed effects, then in the same way as in Section 2.1.2, model (2.11) also reduces to the working linear mixed model as given in equation (2.2). Therefore, the estimation procedure developed for model (2.1) can be directly applied to model (2.11), except that the dimension of β is now getting larger.

While the two estimation approaches, the direct optimization and the EM approach, showed great performance in parameter estimation and statistical inferences, both of them have their own advantages and disadvantages. The direct optimization by the Quasi-Newton optimizer is well acknowledged by its quadratic convergence rate. One major disadvantage of the direct optimization method, however, is that it is not guaranteed to converge when the log-likelihood is not concave. The EM algorithm has been highly appreciated due to its many appealing features, such as guaranteed increase in the likelihood and good performance for ill-conditioned problems. However, the significant drawback of the EM algorithm is its slow convergence near the MLEs and lack of information matrix at the convergence point. Motivated by the work of Redner and Walker (1984) and Aitkin and Aitkin (1996), here we propose a hybrid EM/Quasi-Newton algorithm for model inference. The hybrid EM/Quasi-Newton algorithm is essentially a combination of the EM and direct optimization

approaches, which first takes advantage of the monotonic convergence property of the EM algorithm to exploit the global optimal region and then applies the direct optimization with the use of the Quasi-Newton optimizer to quickly locate the optimal parameter values.

2.2.3 Simulation

We now conduct a simulation study to assess the performance of the hybrid EM/Quasi-Newton approach in the context of time-varying coefficient mixed models. Consider two models:

$$Y_{ij} = f_0(t_{ij}) + X_i f_1(t_{ij}) + b_i + \epsilon_{ij}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n_i, \quad (2.12)$$

and

$$Y_{ij} = f_0(t_{ij}) + X_i f_1(t_{ij}) + b_{0i} + t_{ij} b_{1i} + \epsilon_{ij}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n_i. \quad (2.13)$$

Both models allow the intercept as well as the fixed effect coefficient to vary over time, with model (2.12) including a random intercept only and model (2.13) including a random intercept and a random time slope. The covariate X_i is a binary treatment indicator, taking the value 1 or 0 with probability 0.5. The true functions are taken to be $f_0(t) = \cos(t)$ and $f_1(t) = \sin(t)$. Following the simulation in Section 2.1.3, the measurement error ϵ_{ij} is generated from $N(0, 1)$, the random intercept b_i is assumed to follow $N(0, 0.5^2)$, and the bivariate random effects $(b_{0i}, b_{1i})^T$ are jointly normally distributed with mean vector $\mathbf{0}$ and variance matrix D ($D_{00} = \sigma_{b_0}^2 = 0.5^2$, $D_{01} = -0.045$, and $D_{11} = \sigma_{b_1}^2 = 0.3^2$, thus $\rho = \text{corr}(b_{0i}, b_{1i}) = -0.3$). We also set the designed time points to be a set of 100 equally spaced points $\{t_1, \dots, t_{100}\}$ in $[0, 8]$. For each simulation, 200 replicates of samples were generated according to the above models. Each sample consists of $m = 100$ subjects, with each subject having $n_i = 5$ repeated measurements evenly distributed at these designed time points (see Section 2.1.3). To examine the impact of censoring proportions on

statistical inferences, we consider three different quantification limits such that the censoring percentages of Y are 0%, 11% or 22% (The dataset ACTG398 in Section 2.1.4 consists of 22% censored observations).

In the B-spline smoothing step, different values of K_s ($s = 0, 1$) and different sets of basis functions could be used for the approximation of $f_0(t)$ and $f_1(t)$. For simplicity, following the suggestion of Rice and Wu (2001), we use the same basis functions for $f_0(t)$ and $f_1(t)$, and set $K_0 = K_1 = K$. The number of knots K is determined by the model selection criterion BIC as given previously. Our simulation shows that $K = 3$ is optimal for all settings considered here. Additionally, Aitkin and Aitkin (1996), for their hybrid EM/Gauss-Newton algorithm, suggested to start with five EM iterations to approach the MLEs, and then switch to Gauss-Newton algorithm. We adopt their idea and use the estimates from the fifth EM iteration as the initial values to start the subsequent direct optimization approach.

Table 2.5 and 2.6 report the estimates of the variance components for model (2.12) and model (2.13) respectively. For comparison, we also implemented the imputation method, with values of censored Y_{ij} 's imputed to be half of the detection limit. Except for 0% censoring rate (the imputation and the hybrid EM/Quasi-Newton approaches are indeed equivalent in this case), the imputation method always resulted in significant biases in model parameter estimates, and hence incorrect Wald confidence intervals with erroneously low coverage probabilities. It is also evident that the estimates by the imputation method became more biased as the proportion of the censored observations increases. In contrast, the hybrid EM/Quasi-Newton approach consistently produced virtually unbiased estimates of the variance components regardless of models or censoring rates, and the corresponding empirical coverage probabilities of the Wald confidence intervals were close to the nominal level (95%) for all cases.

Figures 2.9 to 2.12 present the estimated nonparametric functions $\hat{f}_0(t)$ and $\hat{f}_1(t)$, their empirical biases and point-wise 95% coverage probabilities of the true fixed functions $f_0(t)$

and $f_1(t)$ from two estimation procedures at two censoring levels, 11% and 22%. At 0% censoring rate, the results from the imputation method and the hybrid EM/Quasi Newton approach were identical (results not shown). As the censoring rate increases, the imputation method produced significantly biased estimates, especially when the proportion of the censored observations increases to 22%. The biased estimates and incorrect standard error estimates resulted in unreliable confidence intervals and hence low coverage probabilities for both $f_0(t)$ and $f_1(t)$. In contrast, the hybrid EM/Quasi-Newton approach is quite robust to increased censoring levels. It continuously performs well at all censoring levels, yielding unbiased estimates $\widehat{f}_0(t)$ and $\widehat{f}_1(t)$, and the coverage probabilities of the point-wise Wald confidence intervals of both $f_0(t)$ and $f_1(t)$ are much closer to the nominal level (95%).

In conclusion, our simulation demonstrates that the hybrid EM/Quasi Newton approach provides a practically feasible way to analyze left-censored longitudinal data with time-varying coefficient mixed models. It is also noted that the hybrid EM/Quasi Newton approach doubles the speed of the direct optimization, and saves significant amount of computation time compared to the EM approach.

2.2.4 Application

To illustrate the approach proposed in this subsection, we again consider the HIV viral load data from ACTG398 study, which was described previously in Section 2.1.4. Besides the main objective of this study, i.e., comparing the anti-retroviral effects of two treatments (DPI and SPI) to obtain evidence that the DPI regimen may be superior to the SPI regimen, our additional focus in this analysis is to characterize the dynamic feature of difference in antiviral activity between these two treatment regimens. We propose to fit the data with a time-varying coefficient mixed model given as

$$Y_{ij} = \beta_0(t_{ij}) + \beta_1(t_{ij})X_{1i} + \beta_2 X_{2i} + Z_{ij}b_i + \epsilon_{ij}. \quad (2.14)$$

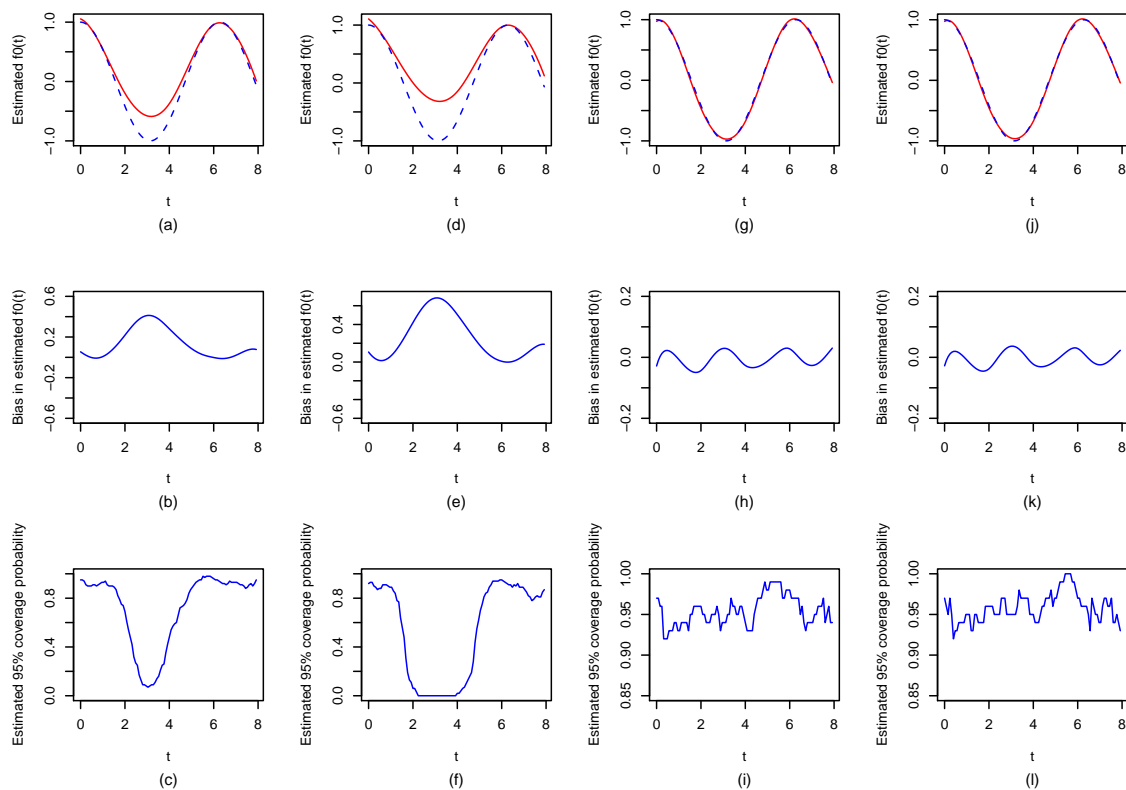


Figure 2.9: Estimated nonparametric function $\hat{f}_0(t)$ (solid line) with the true fixed function $f_0(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}_0(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f_0(t)$ (bottom row panel) for model (2.12). (a)-(c) the imputation method with 11% censoring; (d)-(f) the imputation method with 22% censoring; (g)-(i) the hybrid EM/Quasi-Newton approach with 11% censoring; (j)-(l) the hybrid EM/Quasi-Newton approach with 22% censoring.

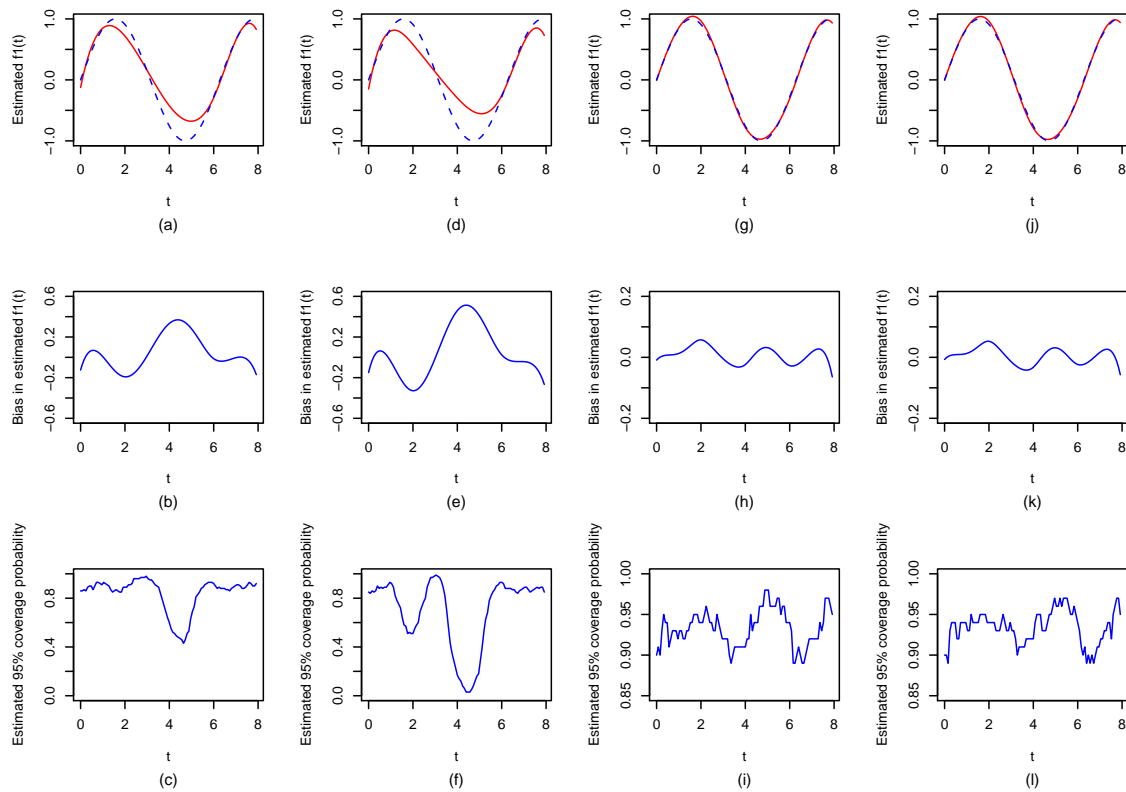


Figure 2.10: Estimated nonparametric function $\hat{f}_1(t)$ (solid line) with the true fixed function $f_1(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}_1(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f_1(t)$ (bottom row panel) for model (2.12). (a)-(c) the imputation method with 11% censoring; (d)-(f) the imputation method with 22% censoring; (g)-(i) the hybrid EM/Quasi-Newton approach with 11% censoring; (j)-(l) the hybrid EM/Quasi-Newton approach with 22% censoring.

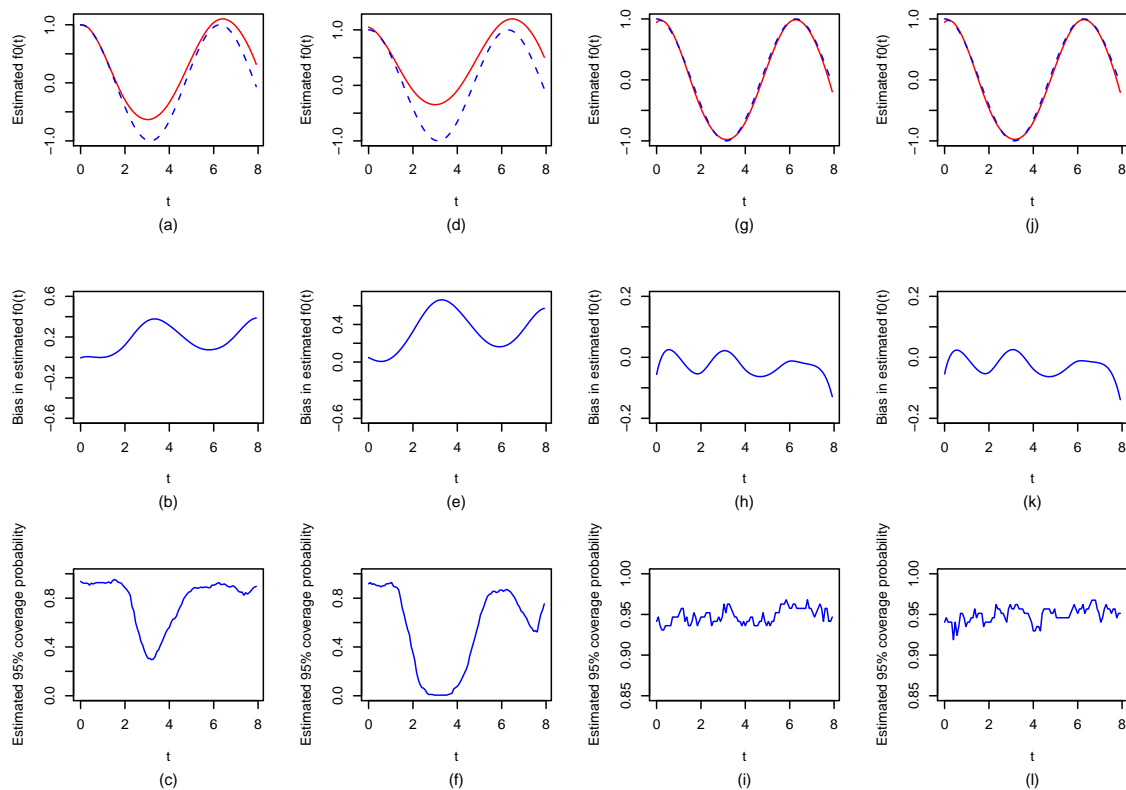


Figure 2.11: Estimated nonparametric function $\hat{f}_0(t)$ (solid line) with the true fixed function $f_0(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}_0(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f_0(t)$ (bottom row panel) for model (2.13). (a)-(c) the imputation method with 11% censoring; (d)-(f) the imputation method with 22% censoring; (g)-(i) the hybrid EM/Quasi-Newton approach with 11% censoring; (j)-(l) the hybrid EM/Quasi-Newton approach with 22% censoring.

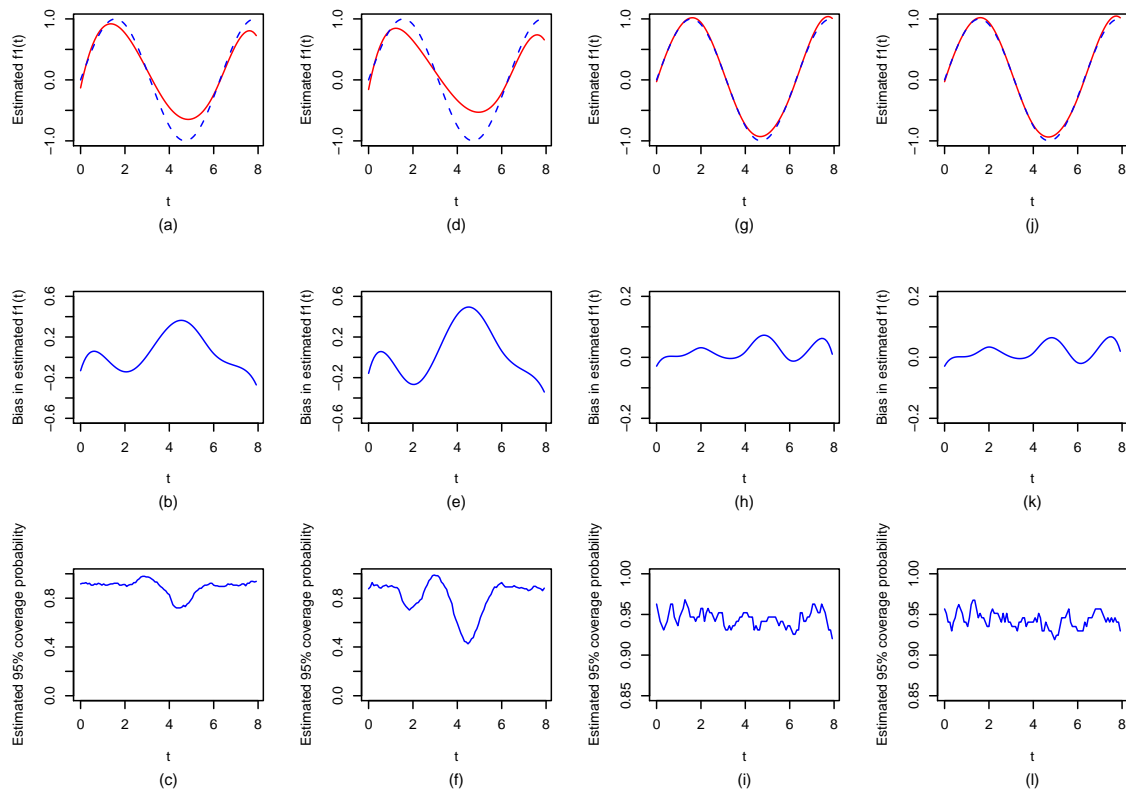


Figure 2.12: Estimated nonparametric function $\hat{f}_1(t)$ (solid line) with the true fixed function $f_1(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}_1(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f_1(t)$ (bottom row panel) for model (2.13). (a)-(c) the imputation method with 11% censoring; (d)-(f) the imputation method with 22% censoring; (g)-(i) the hybrid EM/Quasi-Newton approach with 11% censoring; (j)-(l) the hybrid EM/Quasi-Newton approach with 22% censoring.

Table 2.5: Simulation results from model (2.12). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.

Censoring%	Parameters	Method	Mean	SD	SE	Bias	CP(95%)
0%	σ_b	Hybrid EM	0.477	0.066	0.065	-0.023	0.94
	σ_ϵ	Hybrid EM	0.985	0.034	0.035	-0.015	0.92
11%	σ_b	Imputation	0.369	0.064	0.058	-0.131	0.36
		Hybrid EM	0.477	0.067	0.067	-0.023	0.95
	σ_ϵ	Imputation	0.863	0.032	0.031	-0.137	0.00
		Hybrid EM	0.986	0.035	0.037	-0.014	0.95
22%	σ_b	Imputation	0.321	0.058	0.053	-0.179	0.10
		Hybrid EM	0.475	0.073	0.069	-0.025	0.95
	σ_ϵ	Imputation	0.789	0.029	0.028	-0.211	0.00
		Hybrid EM	0.984	0.037	0.040	-0.016	0.94

Same as in model (2.10) (Section 2.1.4), Y_{ij} represents $\log_{10}(\text{RNA copies/ml})$, X_{1i} is a treatment indicator (1, if the patient is from the DPI group; 0, if the patient is in the SPI group), $X_{2i} = 1$ if a patient was NNRTI-experienced and 0 otherwise, and t_{ij} 's take the same logarithm transformation as previously described. We also considered three different formulations of random effects: $Z_{ij} = (1)$, $Z_{ij} = (1 - X_{1i}, X_{1i})$ and $Z_{ij} = (1, t_{ij})$. Model (2.14) represents a more general case compared to model (2.10) since it allows a nonparametric time-varying coefficient for the treatment effect. We anticipate that the relationship between viral load and treatment effect can be more appropriately characterized with this time-varying coefficient mixed model, and that the resulting estimated $\hat{\beta}_1(t)$ could help us visually determine whether a parametric linear function of time (as in model (2.10)) is adequate to address the treatment effect for this dataset.

Again we use the regression spline method with B-spline basis for nonparametric regression, and the observed BIC for number of knots selection. By setting $K_0 = K_1 = K$ as in the simulation study, the BIC selected models with knots $K = 5$ as the optimal one in all

Table 2.6: Simulation results from model (2.13). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.

Censoring%	Parameters	Method	Mean	SD	SE	Bias	CP(95%)
0%	σ_{b0}	Hybrid EM	0.468	0.158	0.186	-0.032	0.96
	ρ	Hybrid EM	-0.251	0.270	0.313	0.049	0.97
	σ_{b1}	Hybrid EM	0.295	0.028	0.031	-0.005	0.96
	σ_ϵ	Hybrid EM	0.984	0.040	0.040	-0.016	0.93
11%	σ_{b0}	Imputation	0.487	0.145	0.149	-0.013	0.93
		Hybrid EM	0.461	0.177	0.196	-0.039	0.94
	ρ	Imputation	-0.470	0.205	0.209	-0.170	0.67
		Hybrid EM	-0.244	0.310	0.412	0.056	0.95
	σ_{b1}	Imputation	0.255	0.025	0.027	-0.045	0.62
		Hybrid EM	0.295	0.029	0.032	-0.005	0.97
	σ_ϵ	Imputation	0.883	0.035	0.036	-0.117	0.10
		Hybrid EM	0.985	0.043	0.043	-0.015	0.94
22%	σ_{b0}	Imputation	0.482	0.135	0.128	-0.018	0.96
		Hybrid EM	0.466	0.180	0.201	-0.034	0.91
	ρ	Imputation	-0.583	0.172	0.165	-0.283	0.45
		Hybrid EM	-0.275	0.296	0.407	0.025	0.97
	σ_{b1}	Imputation	0.242	0.024	0.025	-0.058	0.34
		Hybrid EM	0.296	0.030	0.033	-0.004	0.97
	σ_ϵ	Imputation	0.805	0.034	0.033	-0.195	0.00
		Hybrid EM	0.984	0.045	0.047	-0.016	0.95

the scenarios considered. The imputation methods of IM and Half IM were also carried out here.

Table 2.7 and 2.8 summarize the estimates of the variance components and β_2 in model (2.14), respectively. These results are almost identical to the estimates presented in Table 2.3 and 2.4 (Section 2.1.4). Likewise, the estimated $\widehat{\beta}_0(t)$ of model (2.14) (Figure 2.13) is also visually indistinguishable to the estimated $\widehat{f}(t)$ of model (2.10) (Figure 2.5). Together, these results imply that in this case, modeling the treatment effect via a time-varying coefficient $\beta_1(t)$ has little impact on other model parameter estimates. To avoid redundant description of the similar results, next we will only focus on our primary interest, i.e., characterize the dynamic feature of difference in antiviral activity between two treatment regimens. It is evident from Figure 2.15 that the treatment effect (treatment difference between DPI and SPI) does not follow a straight line over the entire treatment period as assumed in model (2.10). The time-varying coefficient $\beta_1(t)$ curve jags up and down during the first 6 weeks, showing a great amount of fluctuation. During the first 10 days of treatment, one observed significantly improved antiviral activity of DPI versus SPI, which was followed by a sudden decrease in efficacy from week 2 to 3. After week 3, the DPI regimen appeared to regain its antiviral efficacy from then on. The time-varying coefficient $\beta_1(t)$ curve tended to sustain a steadily declining trend starting from week 6 until the end of the study. The declining trend after week 6 can be well approximated by a straight line (Figure 2.16), hence it is not unexpected that model (2.10) provided a reasonably good fit to this dataset. On the other hand, if the short-term treatment effect is of additional interest, model (2.14) would be a better choice.

It is worth mentioning that in all cases, the two imputation methods again yielded biased estimates compared to the hybrid EM/Quasi-Newton approach. Estimated population means of viral load for the two treatment regimens exhibited similar patterns as that of model (2.10) (see Figure 2.17 and 2.18).

Table 2.7: Variance component estimates for ACTG398 viral load data from model (2.14).

Random Effects & BIC	Model Parameters	IM		Half IM		Hybrid EM	
		Estimates	SE	Estimates	SE	Estimates	SE
$Z_{ij} = (1)$	σ_b	0.846	0.031	1.033	0.039	0.979	0.037
	σ_e	0.706	0.011	0.939	0.014	0.854	0.015
BIC		6696.5		8122.3		6998.4	
$Z_{ij} = (1 - X_{1i}, X_{1i})$	σ_{b0}	0.844	0.054	1.007	0.066	0.952	0.063
	σ_{b1}	0.847	0.038	1.045	0.048	0.993	0.046
	σ_e	0.706	0.011	0.938	0.014	0.854	0.015
BIC		6696.5		8122.1		6998.2	
$Z_{ij} = (1, t_{ij})$	σ_{b0}	0.676	0.030	0.762	0.036	0.693	0.034
	σ_{b1}	1.515	0.075	2.226	0.100	2.495	0.126
	ρ	0.120	0.071	0.078	0.069	0.100	0.069
	σ_e	0.578	0.010	0.725	0.013	0.661	0.013
BIC		6259.2		7480.2		6326.1	

Table 2.8: Parameter estimates for ACTG398 viral load data from model (2.14).

Random Effects	Model Parameters	IM		Half IM		Hybrid EM	
		Estimates	SE	Estimates	SE	Estimates	SE
$Z_{ij} = (1)$	β_2	0.613	0.083	0.836	0.103	0.768	0.097
$Z_{ij} = (1 - X_{1i}, X_{1i})$	β_2	0.613	0.083	0.835	0.102	0.766	0.097
$Z_{ij} = (1, t_{ij})$	β_2	0.463	0.073	0.556	0.086	0.458	0.080

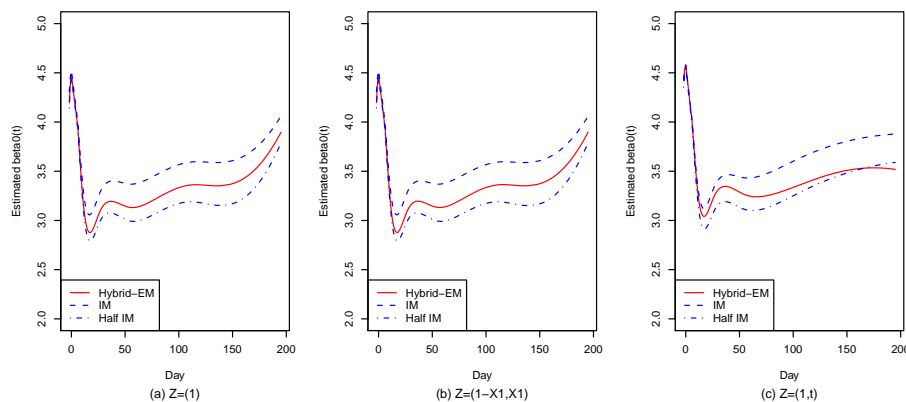


Figure 2.13: Estimated $\hat{\beta}_0(t)$ for different random effect formulations of model (2.14) from three estimation procedures.

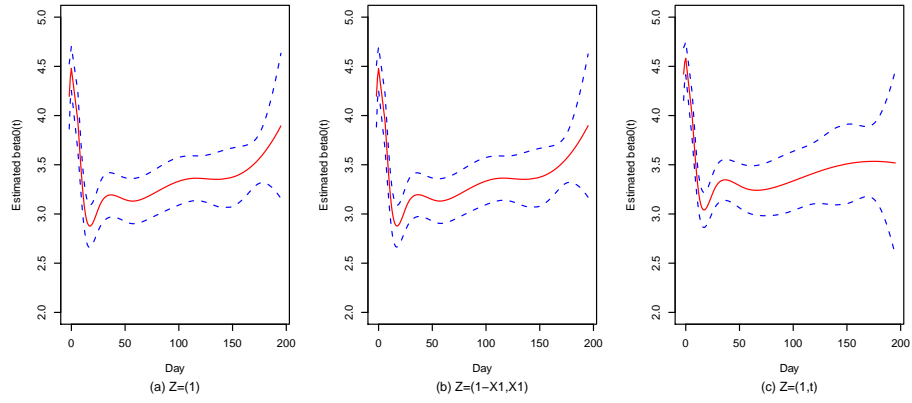


Figure 2.14: Estimated $\hat{\beta}_0(t)$ (solid line) along with its confidence interval (dashed lines) for different random effect formulations of model (2.14) from the hybrid EM/Quasi-Newton approach.

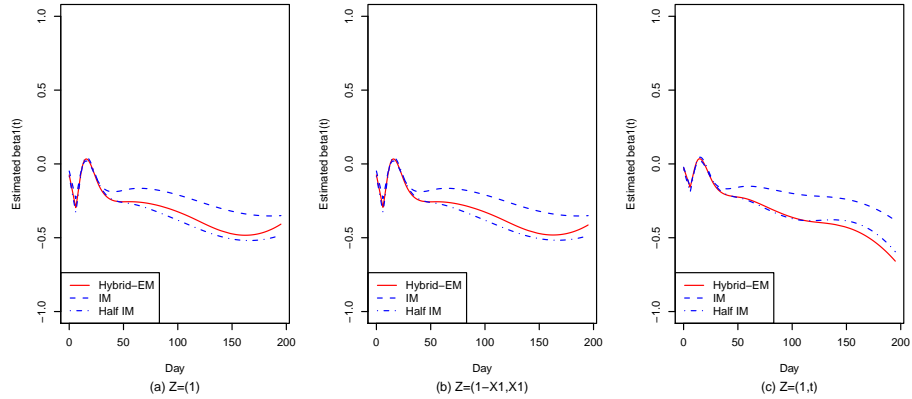


Figure 2.15: Estimated $\hat{\beta}_1(t)$ for different random effect formulations of model (2.14) from three estimation procedures.

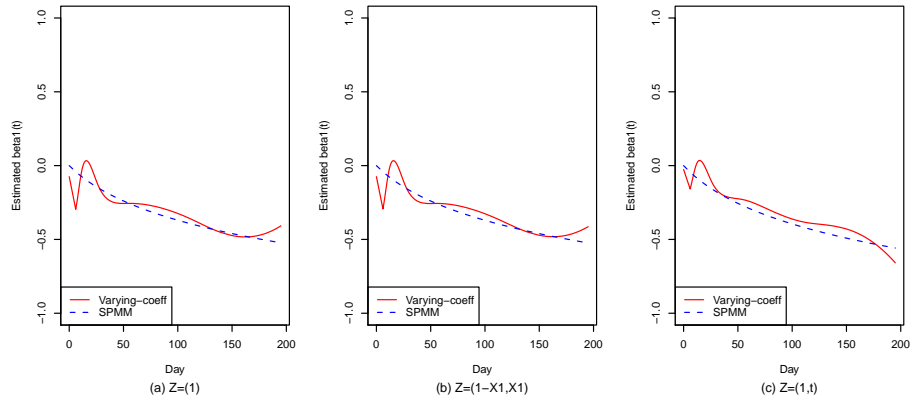


Figure 2.16: Estimated $\hat{\beta}_1(t)$ from model (2.14) and model (2.10) by the hybrid EM/Quasi-Newton approach.

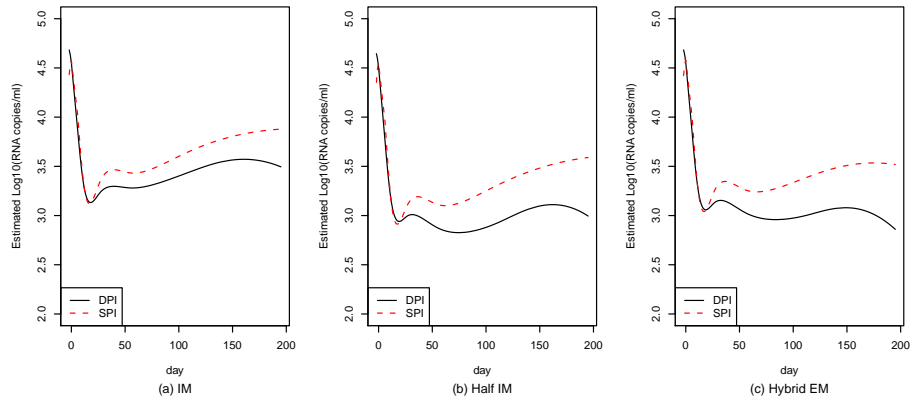


Figure 2.17: Estimated population means of $\log_{10}(\text{RNA copies/ml})$ for the two treatment (DPI versus SPI) groups from the IM, Half IM and hybrid EM/Quasi-Newton methods, for patients who had no prior NNRTI treatment ($X_{2i}=0$).

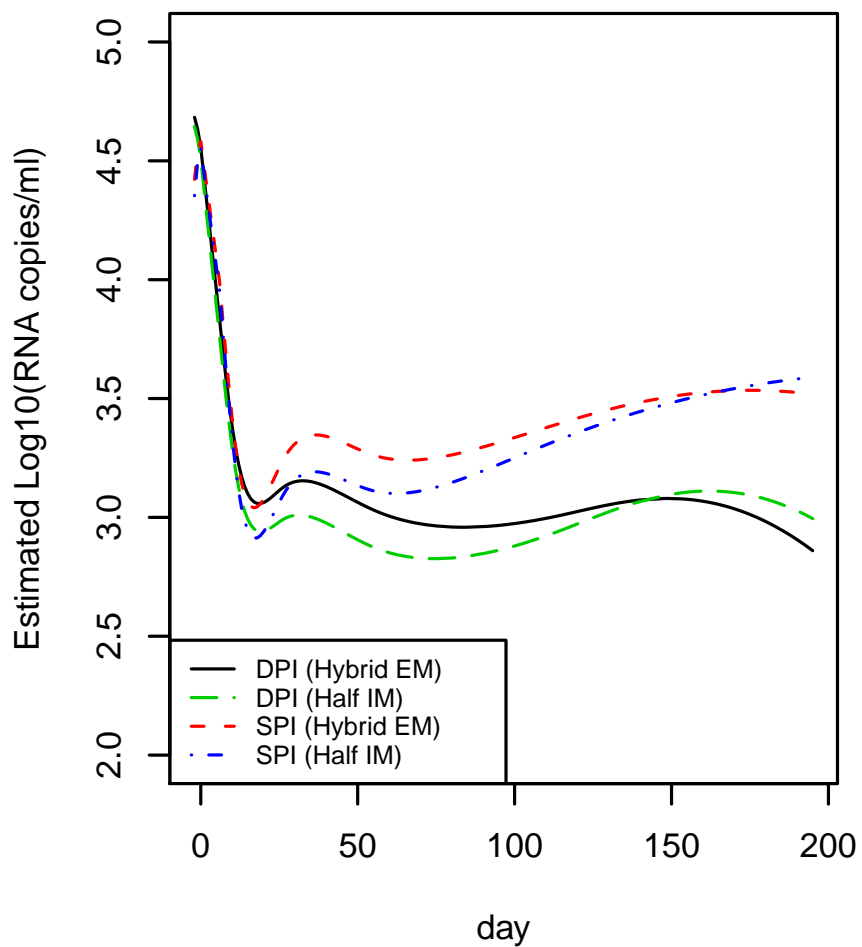


Figure 2.18: Estimated population means of $\log_{10}(\text{RNA copies/ml})$ for the two treatment (DPI versus SPI) groups from the Half IM and hybrid EM/Quasi-Newton methods, for patients who had no prior NNRTI treatment ($X_{2i}=0$).

2.3 Discussion

In this chapter, we first proposed to use a semiparametric mixed model (SPMM) to analyze censored longitudinal data. In comparing with LMMs, SPMMs offer more flexibility in characterizing the feature of the relationship between the response variable and the time effect and other covariate effects. In order to overcome the challenge posed by infinite dimensions of nonparametric functions, we used the regression spline method and approximated the nonparametric function in an SPMM by a linear combination of B-spline basis functions. Compared to alternative smoothing techniques, including the kernel smoothing and spline smoothing, one major advantage of employing the B-spline smoother is that the original SPMM can be transformed into a working LMM representation for a given number of knots while the random effects as well as the covariance matrix of the response remain unchanged. Obtaining the working MLEs based on the LMM representation is fairly straightforward. In particular, with the availability of standard optimization routines, the direct optimization approach is easy to implement if the random effect structure is not too complicated. Otherwise, the EM algorithm can be adapted for more stable parameter estimation.

Although SPMMs provide a certain amount of flexibility in characterizing the relationship between response variables and the time effect, they assume parametric fixed effects for all other covariates. This parametric assumption may be too restrictive in some applications, thus, in the second section, we proposed an alternative extension, varying-coefficient mixed models, which also use nonparametric smooth functions to model the effects of other covariates of interest, allowing one to assess their potentially complex influences on the response variable in more depth. To facilitate the computation, we proposed a hybrid EM/Quasi-Newton approach by combining the direct optimization approach with the EM algorithm, and our simulation demonstrated its successful performance.

Since our estimation procedures were developed based on the working LMM representation, one may be interested to know whether the proposed procedures are comparable or even superior to the existing estimation approaches well developed under the LMM frame-

work, such as the MCEM (Hughes, 1999) and the direct optimization approach (Jacqmin-Gadda et al., 2000). By courtesy of Hughes and Jacqmin-Gadda, we obtained both the R code used to implement the MCEM (Hughes, 1999) and the Fortran code used to implement the direct optimization approach (Jacqmin-Gadda et al., 2000). However, our attempt to compare the hybrid EM/Quasi-Newton approach with these two approaches turned out to be a failure. It was noticed that the Fortran code of Jacqmin-Gadda et al. (2000) is only applicable to models consisting of no more than 5 fixed time-dependent covariates. With the B-spline transformation, however, the resulting working LMMs generally have a relatively large number of fixed time-dependent covariates. With regard to the MCEM, our preliminary experiments showed that the convergence rate of the MCEM is extremely slow for our simulation settings. The major difficulty we experienced with the MCEM is its poor convergence property, for instance, of 100 simulation runs from model (2.9), over 85% did not converge with the MCEM approach (results not shown).

While we focused our discussion on the left censored longitudinal data, the proposed procedures can easily be modified to accommodate right and/or interval censoring (due to upper and/or lower qualification limits). In addition, in this study we only considered the cases where the dimension of the random effects is relatively small, and employed the Gauss-Hermit approach to numerically approximate the expectations involved in the estimation procedure. When the dimension of integrals increases, the Gauss-Hermit approach suffers from the curse of dimensionality, and alternative numerical integration rules need to be explored, such as Monte Carlo importance sampling, Gibbs sampling or Spherical-Radial integration (Monahan and Genz, 1997). It should be mentioned that another class of semiparametric models becomes popular for modeling long-term HIV viral load data, semiparametric nonlinear mixed effects models (Huang et al., 2008; Wu and Zhang, 2002). By introducing an extra random effects term into the nonparametric component, this type of models can appropriately account for the high between-subject variation and help the development of individualized treatment strategies for AIDS patients. Dealing with the cen-

soring issue under this class of semiparametric models is even more challenging and could be of future research interest.

Chapter 3

Smoothing Spline Method for Censored Longitudinal Data

While the B-spline smoothing approach is attractive due to its easy implementation, concerns may be raised on the methods chosen for locating the knots. For example, equally-spaced knots are only appropriate for ideal situations where measurements of subjects are quite evenly distributed over time. The knots locating method based on sample quantiles can be used in general cases and is shown to improve smoothing results, but it is still rather arbitrary and not fully driven by data. An alternative to the B-spline smoothing method is the smoothing spline approach. Smoothing splines arise naturally as a solution to optimization problems in a roughness penalty setting. A smoothing spline uses all the distinct data points as knots and meanwhile introduces a penalty for controlling the smoothness of the estimates, and hence it is considered to be a more data-adaptive smoothing method. In this chapter, we consider the application of the smoothing spline method in nonparametric regression analysis of censored longitudinal data.

The remainder of this chapter is organized as follows. In Section 3.1 we describe the penalized likelihood approach and its EM extension as well as the GCV algorithm for estimating smoothing parameters. Section 3.2 reports the results from simulation studies.

We illustrate our method using a real data example in Section 3.3, and conclude the chapter with a discussion in Section 3.4.

3.1 Maximum Penalized Likelihood Estimates

3.1.1 Matrix Notation of SPMM

We use the same setup as in Section 2.1 of Chapter 2. Suppose there is a random sample of m subjects. For the i th subject, let Y_{ij} be the j th observation of the response variable, X_{ij} be a $p \times 1$ vector of fixed covariates and Z_{ij} be a $q \times 1$ vector of random covariates, measured at time point t_{ij} ($i = 1, \dots, m; j = 1, \dots, n_i$ and $t_{ij} \in [0, T]$). We consider the same semiparametric mixed model (SPMM) as in Section 2.1.1,

$$Y_{ij} = f(t_{ij}) + X_{ij}^T \beta + Z_{ij}^T b_i + \epsilon_{ij}. \quad (3.1)$$

Note that the notations of model (3.1) are slightly different from that of model (2.1). Instead of using $S_{ij}^T \delta$, we take $X_{ij}^T \beta$ to be the fixed-effects term in concordant with conventional notations used for mixed effects models. If we define $t^0 = (t_1^0, \dots, t_r^0)$ as an r -dimensional vector of ordered distinct values of the time points t_{ij} , and N_i as the $n_i \times r$ incidence matrix for the i th subject such that the (j, k) th element of N_i is 1 if $t_{ij} = t_k^0$ ($j = 1, \dots, n_i; k = 1, \dots, r$), and 0 otherwise. Then, model (3.1) can be written in a vector notation as

$$Y_i = N_i f + X_i \beta + Z_i b_i + \epsilon_i,$$

where $f = \{f(t_1^0), \dots, f(t_r^0)\}^T$ and $Y_i = (Y_{i1}, \dots, Y_{in_i})^T$. Similar notations apply to X_i , Z_i and ϵ_i as well. If we further denote $Y = (Y_1^T, \dots, Y_m^T)^T$, and N , X , ϵ similarly, and $Z =$

$\text{diag}(Z_1, \dots, Z_m)$, model (3.1) can be represented in a matrix format as

$$Y = Nf + X\beta + Zb + \epsilon, \quad (3.2)$$

where $b = (b_1^T, \dots, b_m^T)^T \sim N\{\mathbf{0}, G(\phi)\}$ with $G(\phi) = \text{diag}(D, \dots, D)$, $\epsilon = (\epsilon_1^T, \dots, \epsilon_m^T)^T$ with $\epsilon \sim N\{\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_n\}$ and $n = \sum_{i=1}^m n_i$ being the total number of observations.

3.1.2 Estimation of Model Parameters

For noncensored longitudinal data, Zhang et al. (1998) adopted a penalized likelihood approach based on smoothing splines, and showed that for given variance components $\theta = (\phi, \sigma_\epsilon^2)$, the maximum penalized likelihood estimates (MPLEs) of β and f in model (3.2) can be obtained as the maximizer of the penalized log-likelihood function, given by

$$\ell(f, \beta, \theta; Y) - \frac{\lambda}{2} \int_0^T \{f''(t)\}^2 dt = \ell(f, \beta, \theta; Y) - \frac{\lambda}{2} f^T K f, \quad (3.3)$$

where $\lambda > 0$ is the smoothing parameter controlling the trade-off between the fidelity to the data, measured by $\ell(f, \beta, \theta; Y)$, the log-likelihood contributed by data, and the roughness of the estimated $f(t)$, measured by $\int_0^T \{f''(t)\}^2 dt$, and K is the nonnegative definite smoothing matrix defined in equation (2.3) of Green and Silverman (1994). It is well known that the resulting estimator of $f(t)$ is a natural cubic smoothing spline which can be uniquely determined by its values evaluated at the distinct knots, i.e., the ordered distinct values of the time points t^0 in our setting (Green and Silverman, 1994).

Censoring poses significant computational challenges in parameter estimation and statistical inference. In this section, we extend the approach of Zhang et al. (1998) and propose an EM algorithm to fit model (3.1) for censored longitudinal data. If we view the observed data as incomplete data and assume the complete data consisting of Y and b , then for a given λ , the MPLEs of β and f as well as the variance components $\theta = (\phi, \sigma_\epsilon^2)$ can be readily updated by maximizing the complete data penalized log-likelihood function in each

EM iteration. Upon convergence, this EM algorithm will lead to an EM version of MPLEs of the model parameters and the nonparametric function.

Under model (3.2), the penalized log-likelihood function of the complete data (Y, b) has the form

$$\ell_p = \ell(f, \beta, \theta; Y, b) - \frac{\lambda}{2} f^T K f, \quad (3.4)$$

where $\ell(f, \beta, \theta; Y, b)$, the complete data log-likelihood function from (Y, b) is

$$\begin{aligned} \ell(f, \beta, \theta; Y, b) &= -\frac{n}{2} \log(2\pi\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} (Y - Nf - X\beta - Zb)^T \\ &\quad \times (Y - Nf - X\beta - Zb) - \frac{m}{2} \log |D| - \frac{1}{2} \sum_{i=1}^m b_i^T D^{-1} b_i. \end{aligned}$$

For a given smoothing parameter λ , the EM estimates of the model parameters can be achieved by iterating between the following E-step and M-step until convergence. In E-step, taking expectation of (3.4) with respect to censored observations and b yields a corresponding penalized Q function which we denoted as $Q_p(f, \beta, \theta | f^{(r)}, \beta^{(r)}, \theta^{(r)})$,

$$\begin{aligned} Q_p(f, \beta, \theta | f^{(r)}, \beta^{(r)}, \theta^{(r)}) &= -\frac{n}{2} \log(2\pi\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} E\{(Y - Nf - X\beta - Zb)^T \\ &\quad \times (Y - Nf - X\beta - Zb)\} - \frac{m}{2} \log |D| \\ &\quad - \frac{1}{2} \sum_{i=1}^m E\{b_i^T D^{-1} b_i\} - \frac{\lambda}{2} f^T K f. \end{aligned} \quad (3.5)$$

In M-step, differentiation of this Q_p function with respect to f and β leads to

$$\begin{aligned} \frac{\partial Q_p}{\partial \beta} &= -\frac{1}{\sigma_\epsilon^2} \{X^T X\beta + X^T Nf - X^T E(Y) + X^T ZE(b)\} \\ \frac{\partial Q_p}{\partial f} &= -\frac{1}{\sigma_\epsilon^2} \{N^T X\beta + N^T Nf - N^T E(Y) + N^T ZE(b)\} - \frac{\lambda}{2} (2K)f. \end{aligned}$$

Hence the MPLEs of β and f solve

$$\begin{bmatrix} X^T X & X^T N \\ N^T X & N^T N + \sigma_\epsilon^2 \lambda K \end{bmatrix} \begin{bmatrix} \beta \\ f \end{bmatrix} = \begin{bmatrix} X^T E(Y) - X^T Z E(b) \\ N^T E(Y) - N^T Z E(b) \end{bmatrix},$$

where $E(Y) = E(Y|Y^o, Y^c \leq c; f^{(r)}, \beta^{(r)}, \theta^{(r)}, \lambda)$ and $E(b) = E(b|Y^o, Y^c \leq c; f^{(r)}, \beta^{(r)}, \theta^{(r)}, \lambda)$.

To estimate the variance components $\theta = (\phi, \sigma_\epsilon^2)$, Zhang et al. (1998) wrote model (3.2) as an LMM representation. In this LMM representation, the smoothing parameter λ acts as an additional variance component, which makes it feasible to estimate λ jointly with other variance components θ through the restricted maximum likelihood (REML) approach.

For censored data, theoretically, an EM version of REML estimates of θ and λ can be derived in the similar fashion as the EM version of the MPLES of β and f . However, we have found that the LMM representation of expression (12) in Zhang et al. (1998) has a high-dimensional random effects term, which increases the computational cost significantly for censored data. To overcome this difficulty, we propose to estimate θ and λ in two separate steps. The similar maximum penalized likelihood approach as that for β and f will be used to estimate θ . For λ , we introduce an approach called generalized cross-validation (GCV) in Section 3.1.4. It can be shown that maximizing the Q_p function (3.5) with respect to σ_ϵ^2 and D results in the following closed form updates,

$$\begin{aligned} \hat{\sigma}_\epsilon^2 &= \frac{1}{n} E\{(Y - Nf - X\beta - Zb)^T (Y - Nf - X\beta - Zb)\} \\ \hat{D} &= \frac{1}{m} \sum_{i=1}^m E(b_i b_i^T | Y_i^o, Y_i^c \leq c_i, f^{(r)}, \beta^{(r)}, \theta^{(r)}, \lambda), \end{aligned}$$

where the expectations are taken with respect to Y^c and b for given observed data.

To get the updates for the model parameter estimates, we need to calculate several expectations required in each EM iteration. Adopting the same notations as in Chapter 2, the

expectations need to be evaluated include $E(Y)$, $E(b)$, $E(b_i b_i^T | Y_i^o, Y_i^c \leq c_i, f^{(r)}, \beta^{(r)}, \theta^{(r)}, \lambda)$, $E\{(Y_{ij}^c)^2 | Y_i^o, Y_i^c \leq c_i, f^{(r)}, \beta^{(r)}, \theta^{(r)}, \lambda\}$ and $E(Y_{ij}^c b_i | Y_i^o, Y_i^c \leq c_i, f^{(r)}, \beta^{(r)}, \theta^{(r)}, \lambda)$. When the dimension of the random effects, q , is less than 3, Gauss-Hermit approach could be used for calculating these expectations. For example,

$$\begin{aligned}
E(b_i) &= E(b_i | Y_i^o, Y_i^c \leq c_i; f^{(r)}, \beta^{(r)}, \theta^{(r)}) \\
&= \int b_i f(b_i | Y_i^o, Y_i^c \leq c_i; f^{(r)}, \beta^{(r)}, \theta^{(r)}) db_i \\
&= \frac{\int b_i f(Y_i^o, Y_i^c \leq c_i | b_i; f^{(r)}, \beta^{(r)}, \theta^{(r)}) f(b_i | \theta^{(r)}) db_i}{\int f(Y_i^o, Y_i^c \leq c_i | b_i; f^{(r)}, \beta^{(r)}, \theta^{(r)}) f(b_i | \theta^{(r)}) db_i} \\
&= \frac{\int b_i g_1(b_i) f(b_i | \theta^{(r)}) db_i}{\int g_1(b_i) f(b_i | \theta^{(r)}) db_i},
\end{aligned}$$

where $g_1(b_i)$ has the following expression

$$\begin{aligned}
g_1(b_i) &= f(Y_i^o, Y_i^c \leq c_i | b_i; f^{(r)}, \beta^{(r)}, \theta^{(r)}) \\
&= f(Y_i^o | b_i; f^{(r)}, \beta^{(r)}, \theta^{(r)}) P(Y_i^c \leq c_i | b_i; f^{(r)}, \beta^{(r)}, \theta^{(r)}) \\
&= \varphi \left(\frac{Y_i^o - N_i^o f^{(r)} - X_i^o \beta^{(r)} - Z_i^o b_i}{\sigma_\epsilon^{(r)}} \right) \prod_{j=1}^{n_i^c} \Phi \left(\frac{c_{ij} - N_{ij}^c f^{(r)} - X_{ij}^c \beta^{(r)} - Z_{ij}^c b_i}{\sigma_\epsilon^{(r)}} \right),
\end{aligned}$$

where N_{ij}^c is referred to as the j th row of N_i^c . As before, expectations in the numerator and denominator of $E(b_i)$ can then be calculated using the Gauss-Hermit approach. Same strategies can be applied to calculate other expectations as well.

To obtain starting values for the EM algorithm, one may apply the existing software, such as SAS macro SPMM (Zhang, 1998), to the imputed data. That is, the censored observations were substituted by the detection limit and the resulting imputed data were then treated as being noncensored.

3.1.3 Variance for Maximum Penalized Likelihood Estimates

In regular penalized likelihood problems, the observed information matrix obtained by treating the penalized likelihood as a usual likelihood, has been used to derive variance estimates for MPLEs (Wahba, 1983; Silverman, 1985; Segal et al., 1994). Following the similar idea, to develop statistical inferences for the above EM version MPLEs, we consider the penalized log-likelihood function of observed data given by

$$\ell_{op} = \ell_o(f, \beta, \theta; Data) - \frac{\lambda}{2} f^T K f. \quad (3.6)$$

Note that due to the missing information caused by the censored observations, rather than the complete data log-likelihood, the observed data log-likelihood should be used here, which is of the form

$$\begin{aligned} \ell_o(f, \beta, \theta; Data) &= \sum_{i=1}^m \{ \log f(Y_i^o | f, \beta, \theta) + \log P(Y_i^c \leq c_i | Y_i^o; f, \beta, \theta) \} \\ &= \sum_{i=1}^m \left\{ -\frac{1}{2} \log |V_i^o| - \frac{1}{2} (Y_i^o - N_i^o f - X_i^o \beta)^T V_i^{o-1} (Y_i^o - N_i^o f - X_i^o \beta) \right\} \\ &\quad + \sum_{i=1}^m \log \left\{ E \left[\prod_{j=1}^{n_i^c} \Phi \left(\frac{c_{ij} - N_{ij}^c f - X_{ij}^{cT} \beta - Z_{ij}^{cT} b_i}{\sigma_\epsilon} \right) \middle| Y_i^o; f, \beta, \theta \right] \right\}. \end{aligned}$$

We estimate the variance of MPLEs using the information matrix of (3.6) with f, β, θ being the estimates from the last EM iteration. Here SAS NLPFDD subroutine is implemented for computing the second-order derivatives of the function (3.6). Again $P(Y_i^c \leq c_i | Y_i^o; f, \beta, \theta)$, a conditional expectation of b_i given Y_i^o , can be numerically evaluated by using Gauss-Hermit approach (Chapter 2). The performance of this variance estimation approach will be further demonstrated by the simulation study in Section 3.2.

3.1.4 Estimation of Smoothing Parameter

The proper selection of smoothing parameters is essential for good performance of spline estimates. As mentioned previously, Zhang et al. (1998) treated the smoothing parameter as an additional variance component and estimated it with other variance components simultaneously using REML. Although extension of their approach to censored data through EM algorithm is theoretically possible, practical implementation is difficult due to the high dimensionality of the random effects. To avoid the computational difficulty involved in the E-step of EM algorithm, we propose to estimate the smoothing parameter separately via GCV. The GCV method is a widely used data-driven procedure for estimating smoothing parameters. For noncensored data, the GCV estimate of smoothing parameters is the minimizer of the objective function $V(\lambda)$ (Gu, 2002; Wahba, 1990), given by

$$V(\lambda) = \frac{n^{-1}Y^T\{\mathbf{I} - A(\lambda)\}^2Y}{\{n^{-1}\text{trace}(\mathbf{I} - A(\lambda))\}^2}, \quad (3.7)$$

where n is the total number of observations in data and $A(\lambda)$ is known as the smoothing matrix generally defined by $\hat{Y} = A(\lambda)Y$ with the fitted values $\hat{Y} = N\hat{f} + X\hat{\beta} + Z\hat{b}$. The explicit expression of $A(\lambda)$ is derived in Appendix C.

Direct application of (3.7) to censored data produces biased estimates (Huang and Harrington, 2002). One method proposed to circumvent this problem is to transform the data in an unbiased way to account for the censoring. In our case, we first replace censored observations by their expected values $E(Y)$, i.e., $E(Y^o) = Y^o$ and $E(Y^c)$ estimated as previously stated in the E-step, and then apply the GCV to the transformed data as if they were completely observed. Standard optimization methods such as the Quasi-Newton procedure can be used to solve (3.7) for λ . In this study we use SAS IML optimization subroutine NLPQN.

Now putting all pieces together, we summarize the entire estimation procedure as follows.

1. Get initial values of f , β and θ ;

2. Iterate between step (a) and step (b) until convergence:
 - (a) Select the smoothing parameter λ through the GCV method;
 - (b) Given the current estimate of λ , update f , β and θ by maximizing the expected penalized log-likelihood function conditional on the observed data.

3.2 Simulation

To assess the performance of the aforementioned estimation procedure, we carried out a simulation study with the same parameter settings as described in Section 2.1.3. Again, data are generated according to the two SPMMs given by

$$Y_{ij} = f(t_{ij}) + X_{ij}\beta + b_i + \epsilon_{ij}, \quad (3.8)$$

$$Y_{ij} = f(t_{ij}) + X_{ij}\beta + b_{0i} + t_{ij}b_{1i} + \epsilon_{ij}. \quad (3.9)$$

Each simulated dataset consists of 100 subjects with each subject having 5 repeated measurements at t_{ij} ($i = 1, \dots, m = 100; j = 1, \dots, n_i = 5$). As before, we set the true $f(t) = \sin(t)$, generated X_{ij} from $N(2, 1)$ and chose the true value of β to be 1.5. The measurement error ϵ_{ij} was drawn from $N(0, 1)$ and b_i from $N(0, 0.5^2)$ for model (3.8). In model (3.9), the bivariate random effects $(b_{0i}, b_{1i})^T$ were generated from $N(\mathbf{0}, D)$ with $D_{00} = \sigma_{b_0}^2 = 0.5^2$, $D_{01} = -0.045$, and $D_{11} = \sigma_{b_1}^2 = 0.3^2$ (so $\rho = \text{corr}(b_{0i}, b_{1i}) = -0.3$). Three detection limits were chosen so that the censoring percentages of outcomes are around 0.2%, 17% and 34% accordingly.

Estimation results of the model parameters from (3.8) and (3.9) are summarized in Table 3.1 and Table 3.2, respectively. In parallel, the results from the imputation method (half detection limit) are also presented for comparison purpose. Note that we denote the

smoothing parameter to be $\tau = 1/\lambda$, which is consistent with the notation used by Zhang et al. (1998). In the case of 0.2% censoring (an ideal scenario considered to be equivalent to non-censoring situations), the EM and imputation methods showed nearly identical good performance as anticipated. With the increased number of left-censored observations, the imputation method again leads to biased estimates and unreliable confidence intervals; especially the fixed regression coefficient β is significantly underestimated. In contrast, the estimates from the EM algorithm are almost all unbiased, and the associated coverage probability rates are close to the nominal level. Furthermore, compared to the EM algorithm, the imputation method consistently underestimated the smoothing parameter τ . This may be mainly due to the arbitrary use of the half detection limit as the imputation values, as it appears that most of censored observations in the simulated dataset have true values smaller than the half detection limit. Thus the imputation process over-valued those censored observations and caused the curvature of $\hat{f}(t)$ underestimated (see Figure 3.2 and 3.4). Recall that we proposed to estimate variance of MPLEs via the information matrix of the observed data penalized log-likelihood function. Despite some observed deviations between the estimated standard error and the empirical standard deviation (e.g., for parameter ρ), overall this variance estimation approach works well in our simulation setting.

Figure 3.1 to Figure 3.4 display and compare the estimated nonparametric function $\hat{f}(t)$, its empirical bias and point-wise 95% coverage probabilities of the true fixed function $f(t)$ at different censoring levels. These results are very similar to the ones shown previously in Section 2.1.3. Again the EM approach yields smaller overall biases than the imputation method in all censoring cases, and it continues to perform well with data containing as high as 34% censored observations.

3.3 Application

We illustrate the proposed smoothing spline EM method by applying it to the ACTG398 HIV viral load data described in Chapter 2. To compare the different smoothing methods

in estimating the model parameters and the nonparametric function $f(t)$, we fit the data with the same SPMM as stated previously (see Section 2.1.4), in which the time effect is modeled via a nonparametric function while the effects of other covariates are assumed to follow conventional parametric forms. We anticipate that the results from the two different smoothing approaches will be highly concordant.

We restate the model to be used in this example

$$Y_{ij} = f(t_{ij}) + \beta_1 X_{1i} t_{ij} I(t_{ij} > 0) + \beta_2 X_{2i} + Z_{ij} b_i + \epsilon_{ij}. \quad (3.10)$$

Again Y_{ij} is $\log_{10}(\text{RNA copies/ml})$, $X_{1i} = 1$ if patients are from the DPI group, and $X_{1i} = 0$ otherwise. X_{2i} is an indicator variable for NNRTI (1, if patients had prior NNRTI treatments; 0, if not), and $t_{ij} = \log_{10}(\text{day of actual visit} + 40) - \log_{10}(40)$ is the transformed time (day) after initiation of the treatment. To further stabilize computation, $t_{ij} \times 10$, rather than t_{ij} , was used to fit the above model. We set the random effects to be $Z_{ij} = (1, t_{ij})$ since it has been proven to be the optimal choice according to the previous analysis in Chapter 2.

Presented in Table 3.3 are the estimates of the regression coefficients and the variance components as well as their corresponding standard error estimates from the imputation (the detection limit) and the EM methods. The results of the B-spline smoothing from Section 2.1.4 are also listed for comparison purpose. As expected, the results of the smoothing spline EM method show strong agreement with those from the B-spline EM method. Again we see that the imputation method yields significant biases in both the parameter estimates and the corresponding standard error estimates. The estimated $\hat{f}(t)$ along with its 95% point-wise confidence interval, displayed in Figure 3.5 and 3.6, shows the same patterns as those observed in the B-spline smoothing analysis and is again significantly biased upward by the imputation method. Note that $\hat{f}(t)$ based on the smoothing spline approach exhibits more detailed curve features than the B-spline approach, which may be explained by the

fact that the smoothing spline uses much more knots than the B-spline smoother and hence is able to increase the fit of the spline function to the data.

Similarly, we also present in Figure 3.7 the estimated population means of $\log_{10}(\text{RNA copies/ml})$ for the two treatment groups. Although the overall profiles of these fitted curves are slightly different from those of the B-spline EM method, the results and conclusions drawn from the two analyses are virtually the same.

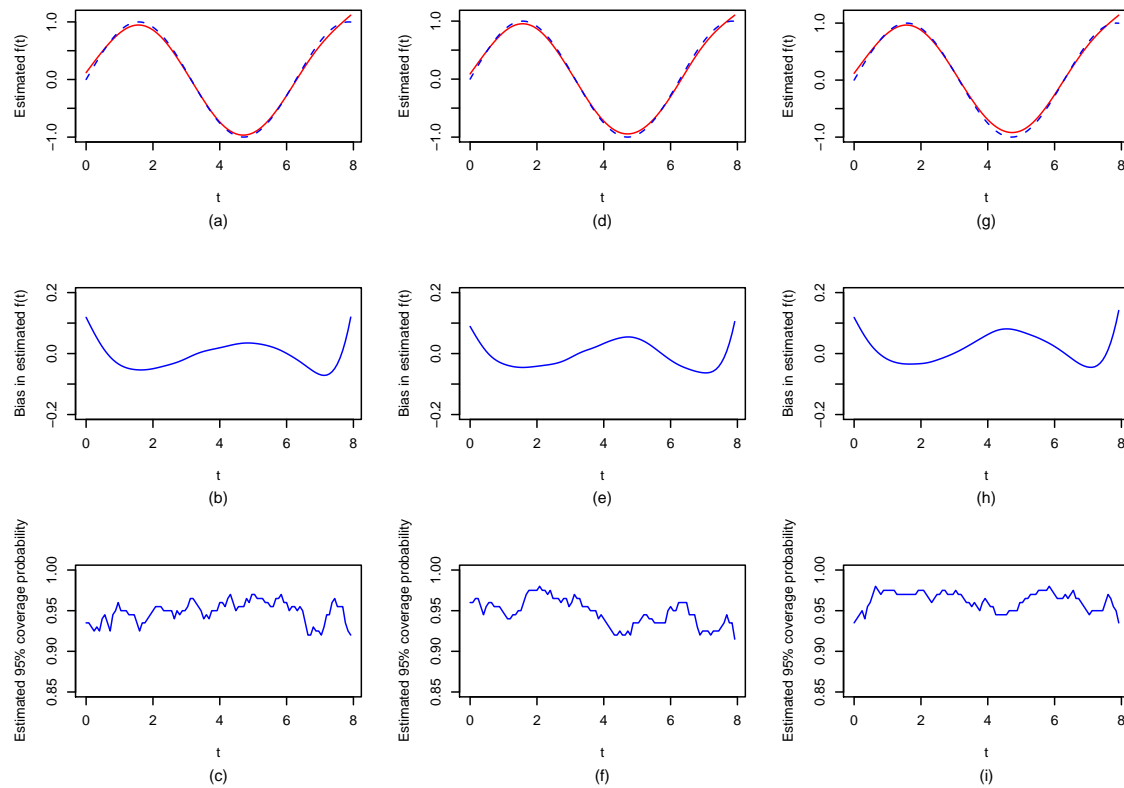


Figure 3.1: Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (3.8) from the proposed approach. (a)-(c) 0.2% censoring; (d)-(f) 17% censoring; (g)-(i) 34% censoring.

Table 3.1: Simulation results from model (3.8). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.

Censoring%	Parameters	Method	Mean	SD	SE	Bias	CP(95%)	
0.2%	β	Imputation	1.502	0.046	0.047	0.002	.97	
		EM	1.500	0.047	0.048	0.000	.97	
	σ_b	Imputation	0.488	0.061	0.065	-0.012	.97	
		EM	0.498	0.079	0.067	-0.002	.95	
	σ_ϵ	Imputation	0.987	0.037	0.035	-0.013	.92	
		EM	1.002	0.040	0.035	0.002	.92	
	τ	Imputation	0.637	0.245				
		EM	0.611	0.225				
	17%	β	Imputation	1.367	0.047	0.046	-0.133	.18
			EM	1.500	0.058	0.054	-0.000	.95
σ_b		Imputation	0.440	0.060	0.064	-0.060	.86	
		EM	0.492	0.079	0.069	-0.008	.94	
σ_ϵ		Imputation	0.967	0.036	0.034	-0.034	.80	
		EM	1.000	0.044	0.037	-0.000	.90	
τ		Imputation	0.482	0.286				
		EM	0.614	0.362				
34%		β	Imputation	1.307	0.044	0.049	-0.193	.01
			EM	1.506	0.064	0.063	0.006	.96
	σ_b	Imputation	0.427	0.068	0.070	-0.073	.86	
		EM	0.500	0.081	0.071	-0.000	.96	
	σ_ϵ	Imputation	1.035	0.037	0.037	0.035	.87	
		EM	0.995	0.046	0.041	-0.005	.93	
	τ	Imputation	0.473	0.314				
		EM	0.623	0.534				

Table 3.2: Simulation results from model (3.9). Mean is the Monte Carlo mean of 200 estimates, SD is their Monte Carlo standard deviation, SE is the average of the 200 estimated standard errors, Bias is the difference between the mean estimates and true values and CP(95%) is the Monte Carlo coverage probability of the true parameter values based on 95% Wald confidence intervals.

Censoring%	Parameters	Method	Mean	SD	SE	Bias	CP(95%)	
0.2%	β	Imputation	1.491	0.051	0.053	-0.009	.94	
		EM	1.503	0.051	0.052	0.003	.95	
	σ_{b0}	Imputation	0.480	0.145	0.212	-0.020	.96	
		EM	0.487	0.159	0.195	-0.013	.94	
	ρ	Imputation	-0.270	0.299	0.687	0.030	.95	
		EM	-0.272	0.332	0.578	0.028	.93	
	σ_{b1}	Imputation	0.296	0.030	0.031	-0.004	.93	
		EM	0.299	0.033	0.031	-0.001	.93	
	σ_{ϵ}	Imputation	0.989	0.040	0.041	-0.011	.92	
		EM	0.988	0.041	0.041	-0.012	.91	
	τ	Imputation	0.497	0.612				
		EM	0.510	0.651				
	17%	β	Imputation	1.331	0.054	0.050	-0.169	.13
			EM	1.500	0.062	0.057	0.000	.94
σ_{b0}		Imputation	0.470	0.157	0.183	-0.030	.95	
		EM	0.505	0.157	0.200	0.005	.95	
ρ		Imputation	-0.275	0.323	0.560	0.025	.89	
		EM	-0.286	0.327	0.601	0.014	.93	
σ_{b1}		Imputation	0.258	0.028	0.029	-0.042	.71	
		EM	0.299	0.034	0.033	-0.001	.96	
σ_{ϵ}		Imputation	0.974	0.039	0.040	-0.026	.91	
		EM	0.988	0.047	0.045	-0.012	.92	
τ		Imputation	0.427	0.578				
		EM	0.541	0.666				
34%		β	Imputation	1.266	0.052	0.053	-0.234	.01
			EM	1.493	0.064	0.065	-0.007	.93
	σ_{b0}	Imputation	0.459	0.176	0.236	-0.041	.93	
		EM	0.516	0.177	0.238	0.016	.91	
	ρ	Imputation	-0.310	0.355	0.876	-0.010	.90	
		EM	-0.312	0.295	0.965	-0.012	.90	
	σ_{b1}	Imputation	0.243	0.031	0.035	-0.057	.53	
		EM	0.302	0.033	0.037	0.002	.96	
	σ_{ϵ}	Imputation	1.035	0.048	0.052	0.035	.92	
		EM	0.977	0.049	0.051	-0.024	.94	
	τ	Imputation	0.397	0.569				
		EM	0.630	0.893				

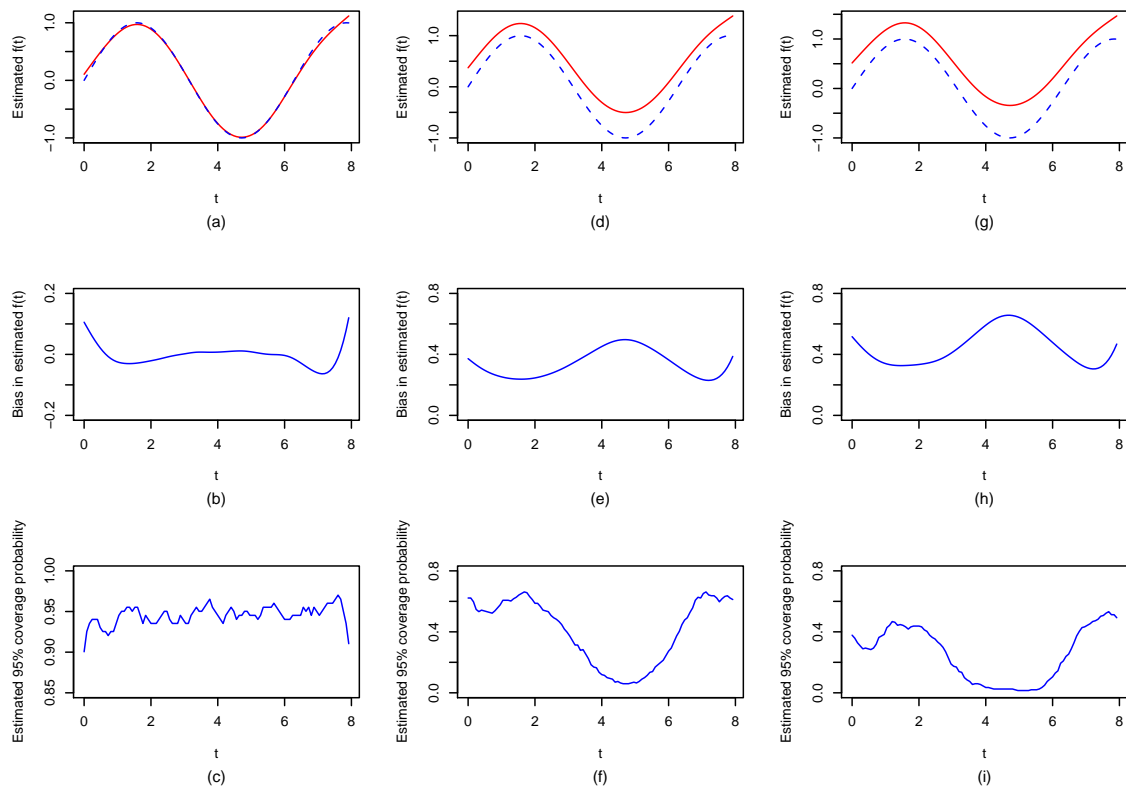


Figure 3.2: Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $\hat{f}(t)$ (bottom row panel) for model (3.8) from the imputation approach. (a)-(c) 0.2% censoring; (d)-(f) 17% censoring; (g)-(i) 34% censoring.

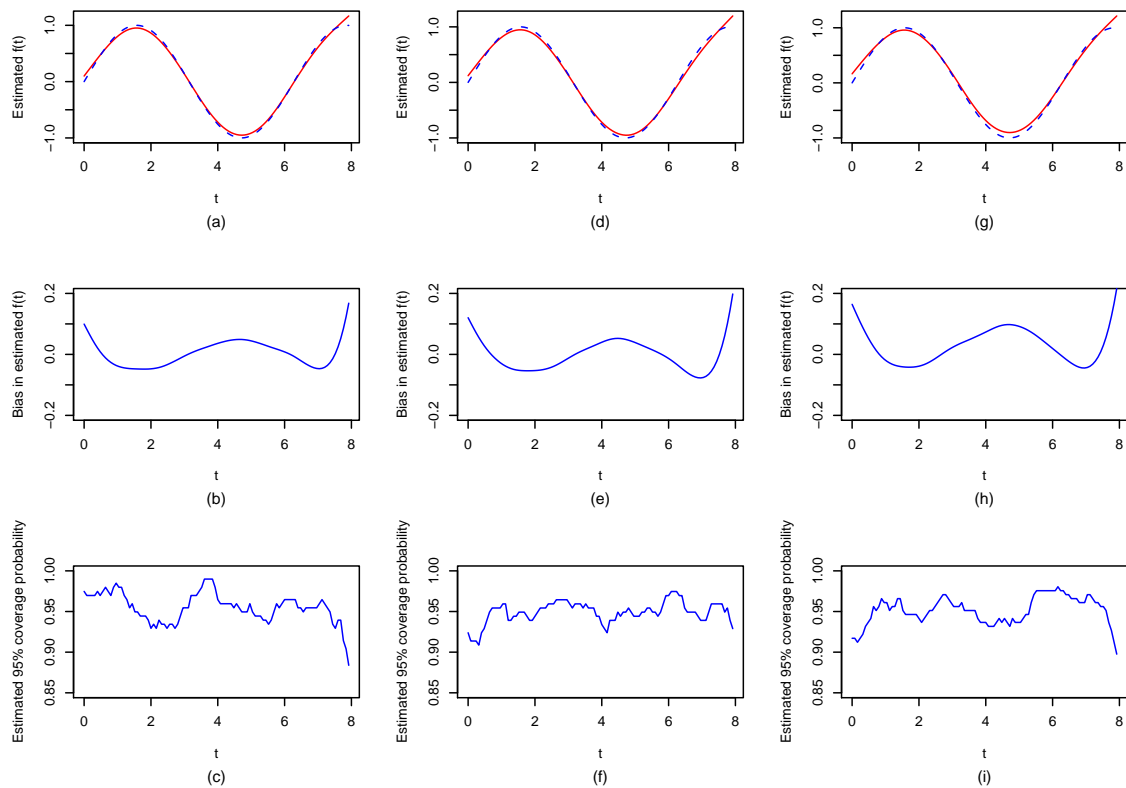


Figure 3.3: Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $f(t)$ (bottom row panel) for model (3.9) from the proposed approach. (a)-(c) 0.2% censoring; (d)-(f) 17% censoring; (g)-(i) 34% censoring.

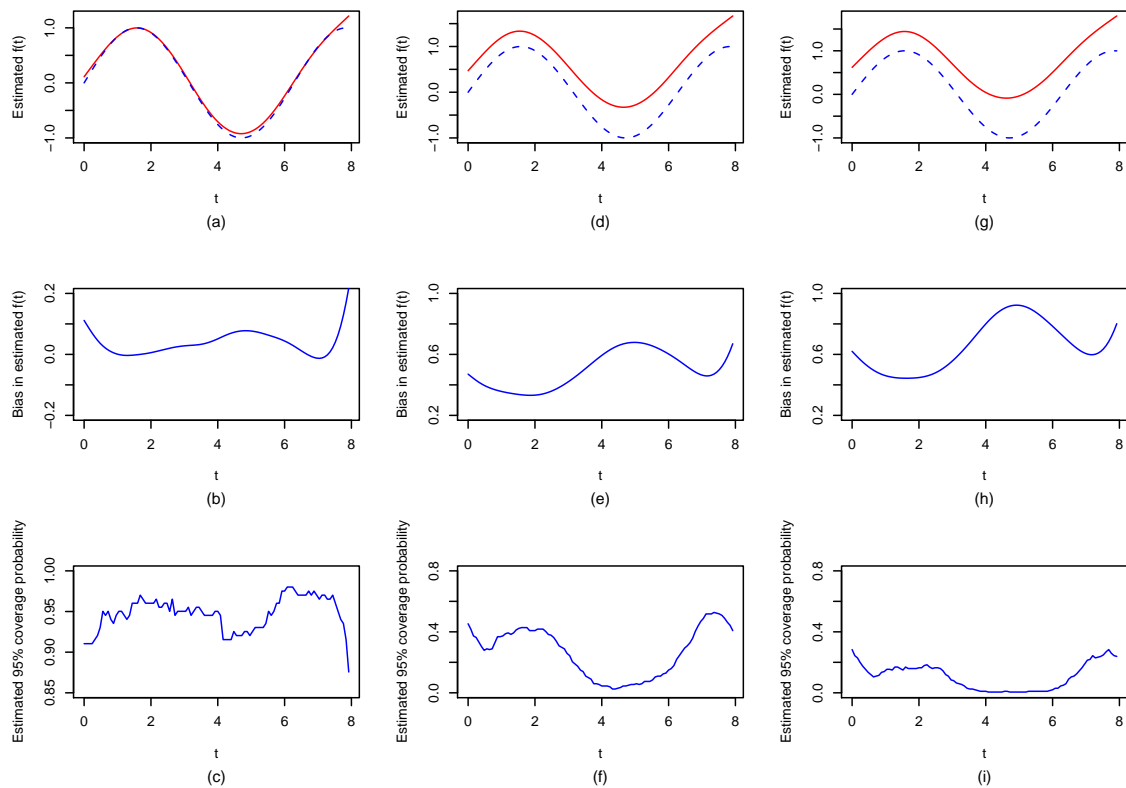


Figure 3.4: Estimated nonparametric time function $\hat{f}(t)$ (solid line) with the true fixed function $f(t)$ (dashed line) superimposed (top row panel), empirical biases of $\hat{f}(t)$ (middle row panel) and pointwise 95% coverage probabilities of $\hat{f}(t)$ (bottom row panel) for model (3.9) from the imputation approach. (a)-(c) 0.2% censoring; (d)-(f) 17% censoring; (g)-(i) 34% censoring.

Table 3.3: Parameter estimates for ACTG398 viral load data from model (3.10).

Model Parameters	Method	Smoothing Spline		B-Spline	
		Estimates	SE	Estimates	SE
β_1	Imputation	-0.039	0.018	-0.042	0.018
	EM	-0.071	0.028	-0.072	0.028
β_2	Imputation	0.455	0.073	0.459	0.073
	EM	0.469	0.081	0.454	0.080
σ_{b0}	Imputation	0.678	0.030	0.677	0.030
	EM	0.697	0.034	0.695	0.034
σ_{b1}	Imputation	0.151	0.008	0.152	0.007
	EM	0.249	0.013	0.250	0.013
ρ	Imputation	0.112	0.071	0.117	0.071
	EM	0.088	0.070	0.096	0.070
σ_ϵ	Imputation	0.577	0.010	0.579	0.010
	EM	0.661	0.013	0.662	0.013
τ	Imputation	0.7505			
	EM	0.9402			

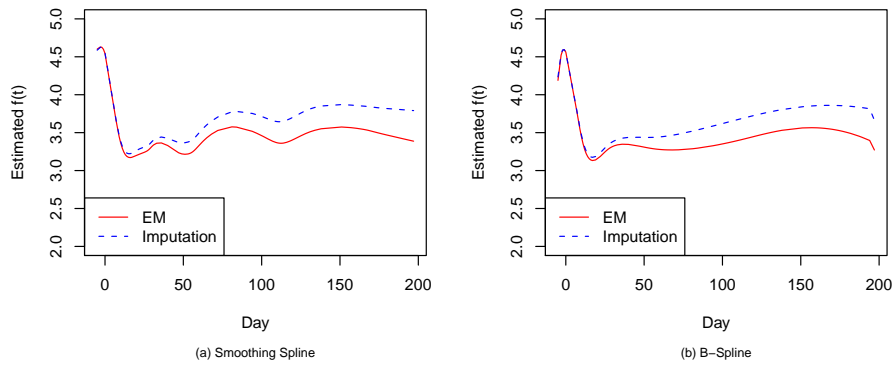


Figure 3.5: Estimated $\hat{f}(t)$ from the smoothing spline approach (left panel) and the B-spline approach (right panel).

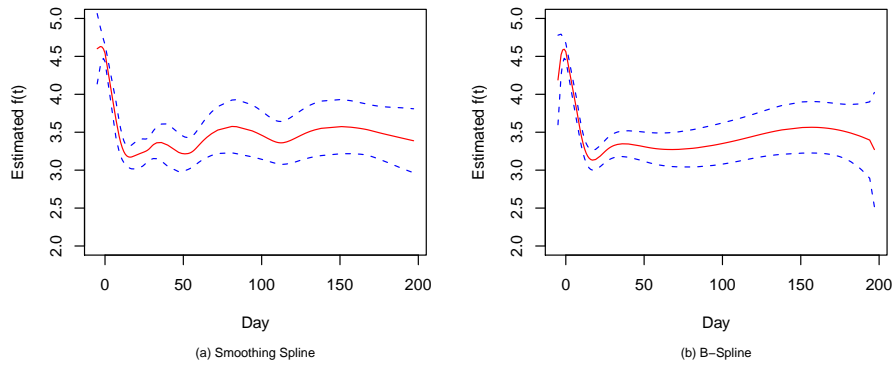


Figure 3.6: Estimated $\hat{f}(t)$ (solid line) along with its confidence interval (dashed lines) from the EM methods using two different smoothing approaches.

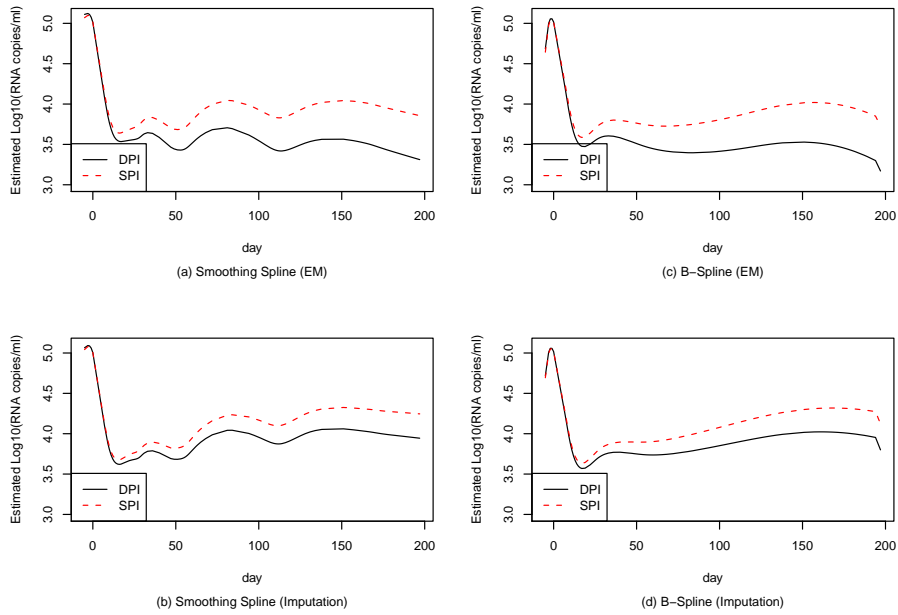


Figure 3.7: Estimated population means of $\log_{10}(\text{RNA copies/ml})$ for the two treatment (DPI versus SPI) groups from two different smoothing approaches, for patients who had no prior NNRTI treatment ($X_{2i}=0$).

3.4 Discussion

In this chapter, we have demonstrated the usage of smoothing splines in analyzing censored longitudinal data within the semiparametric mixed model framework (3.1). By extending the maximum penalized likelihood method well developed for noncensored data (Zhang et al., 1998), we proposed an EM algorithm for computing the MPLEs of model parameters and the nonparametric function $f(t)$. Simulation studies indicate that the proposed method performed well. A similar EM algorithm has been used as a means for ill-posed problems in the context of penalized likelihood formulations by Green (1990), where one can find rigorous theoretical proofs of such an EM algorithm and its convergence properties. To estimate the smoothing parameter, we proposed an approximated generalized cross-validation (GCV) approach based on the commonly used GCV for noncensored data. Such approximated GCV greatly facilitated the entire estimation procedure, and its performance has been well demonstrated by the fitted $f(t)$ curves in our simulation study.

As a natural solution to the roughness penalty problem, smoothing splines and their statistical properties have been intensively studied in literature. Hastie and Tibshirani (1990) also showed that for a small to mediate degree of smoothness, a smoothing spline has better properties than a regression spline with respect to locality of influence. Moreover, for automatic selection of the amount of smoothing, smoothing splines have the additional advantage over regression splines in that the smoothing parameter λ is continuous as opposed to K (the number of knots, must be an integer in regression spline analysis), making the fine-tuning of smoothness feasible. Therefore, the smoothing spline method in this sense is a more data-adaptive smoothing approach. However, it is evident that compared to regression splines presented in Chapter 2, fitting smoothing splines is more difficult and generally requires expensive computation.

We have used the information matrix of the observed data penalized log-likelihood to assess the variability of the EM version MPLEs. The performance of this variance estimation approach was confirmed by the simulation study. Some deviations were observed

between the estimated standard error and the empirical standard deviation for the variance components estimates $(\widehat{\sigma}_{b_0}, \widehat{\rho})$ under the relatively complicated model (3.9). This may be caused partly by the inaccuracy of numerical approximations in second-order derivative computation, given that the observed data penalized log-likelihood function of model (3.9) is complicated. Segal et al. (1994) introduced an alternative way to compute variance for MPLEs obtained through the EM algorithm. Since their approach is tailored specifically for EM version MPLEs, it may provide better performance than ours, and is worthy for further investigation.

REFERENCES

- Aitkin, M. and Aitkin, I. (1996). A hybrid EM/Gauss-Newton algorithm for maximum likelihood in mixture distributions. *Statistics and Computing*, 6(2):127-130.
- Barletta, J. M., Edelman, D. C., and Constantine, N. T. (2004). Lowering the detection limits of HIV-1 viral load using Real-Time Immuno-PCR for HIV-1 p24 antigen. *American Journal of Clinical Pathology*, 122(1):20-27.
- Boor, C. (1978). A practical guide to splines. *Applied Mathematical Sciences*, 27.
- Brumback, B. A. and Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, 93(443):961-976.
- Dempster, A. P., Laird, N. M., Rubin, D. B., et al. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):138.
- Dempster, A. P., Rubin, D. B., and Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association*, 76(374):341-353.
- Diggle, P. (2002). *Analysis of longitudinal data*. Oxford University Press.
- Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC.
- Fan, J. and Zhang, J. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 62(2):303-322.
- Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its interface*, 1(1):179-195.
- Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. CRC Press.
- Gu, C. (2002). *Smoothing spline ANOVA models*. Springer Verlag.
- Hammer, S. M., Vaida, F., Bennett, K. K., Holohan, M. K., Sheiner, L., Eron, J. J., Wheat, L. J., Mitsuyasu, R. T., Gulick, R. M., Valentine, F. T., et al. (2002). Dual vs single protease inhibitor therapy following antiretroviral treatment failure: a randomized trial. *Jama*, 288(2):169.
- Hansen, M. H. and Kooperberg, C. (2002). Spline adaptation in extended linear models. *Statistical Science*, 17(1):2-20.
- Harville, D. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *The Annals of Statistics*, 4(2):384-395.

- Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. CRC Press.
- Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L. (1998). Nonparametric smoothing estimates of Time-Varying coefficient models with longitudinal data. *Biometrika*, 85(4):809–822.
- Huang, J. and Harrington, D. (2002). Penalized partial likelihood regression for Right-Censored data with bootstrap selection of the penalty parameter. *Biometrics*, 58(4):781–791.
- Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-Coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128.
- Huang, Y., Liang, H., and Wu, H. (2008). Identifying significant covariates for anti-HIV treatment response: mechanism-based differential equation models and empirical semi-parametric regression models. *Statistics in medicine*, 27(23):4722–4739.
- Hughes, J. P. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics*, 55(2):625–629.
- Jacqmin-Gadda, H., Thiebaut, R., Chene, G., and Commenges, D. (2000). Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics*, 1(4):355–368.
- Laird, N. M. and Ware, J. H. (1982). Random-Effects models for longitudinal data. *Biometrics*, 38(4):963–974.
- Li, Y., Lin, X., and Mller, P. (2010). Bayesian inference in semiparametric mixed models for longitudinal data. *Biometrics*, 66(1):70–78.
- Liang, H., Wu, H., and Carroll, R. J. (2003). The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error. *Biostatistics*, 4(2):297.
- Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 96(453):103–113.
- Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, 96(455):1045–1056.
- Lyles, R. H., Lyles, C. M., and Taylor, D. J. (2000). Random regression models for human immunodeficiency virus ribonucleic acid data subject to left censoring and informative drop-outs. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(4):485–497.
- Monahan, J. and Genz, A. (1997). Spherical-radial integration rules for bayesian computation. *Journal of the American Statistical Association*, 92(438):664674.

- Moulton, L. H. and Halsey, N. A. (1995). A mixture model with detection limits for regression analyses of antibody response to vaccine. *Biometrics*, 51(4):1570–1578.
- Naylor, J. C. and Smith, A. F. M. (1982). Applications of a method for the efficient computation of posterior distributions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 31(3):214–225.
- Paxton, W. B., Coombs, R. W., McElrath, M. J., Keefer, M. C., Hughes, J., Sinangil, F., Chernoff, D., Demeter, L., Williams, B., and Corey, L. (1997). Longitudinal analysis of quantitative virologic measures in human immunodeficiency Virus-Infected subjects with 400 CD4 lymphocytes: Implications for applying measurements to individual patients. *The Journal of Infectious Diseases*, 175(2):247–254.
- Qu, A. and Li, R. (2006). Quadratic inference functions for Varying-Coefficient models with longitudinal data. *Biometrics*, 62(2):379–391.
- Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science*, 3(4):425–441.
- Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM review*, 26(2):195–239.
- Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57(1):253–259.
- Segal, M. R., Bacchetti, P., and Jewell, N. P. (1994). Variances for maximum penalized likelihood estimates obtained via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(2):345–352.
- Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(1):152.
- Singh, A. and Nocerino, J. (2002). Robust estimation of mean and variance using environmental data sets with below detection limit observations. *Chemometrics and Intelligent Laboratory Systems*, 60(1-2):69–86.
- Stone, C. J., Hansen, M. H., Kooperberg, C., and Truong, Y. K. (1997). Polynomial splines and their tensor products in extended linear modeling. *The Annals of Statistics*, 25(4):1371–1425.
- Stone, C. J. and Huang, J. Z. (2002). Free knot splines in concave extended linear modeling. *Journal of Statistical Planning and Inference*, 108(1-2):219–253.
- Sun, Y. and Wu, H. (2005). Semiparametric time-varying coefficients regression model for longitudinal data. *Scandinavian Journal of Statistics*, 32(1):214–7.
- Verbeke, G. and Molenberghs, G. (2009). *Linear Mixed Models for Longitudinal Data*. Springer.

- Wahba, G. (1983). Bayesian” confidence intervals” for the cross-validated smoothing spline. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(1):133-150.
- Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics*, 13(4):1378-1402.
- Wahba, G. (1990). *Spline models for observational data*. SIAM.
- Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*, 90(1):43-52.
- Wang, Y. (1998). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, 93(441):341-348.
- Wu, C. O., Chiang, C., and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a Varying-Coefficient model with longitudinal data. *Journal of the American Statistical Association*, 93(444):1388-1402.
- Wu, C. O. and Chiang, C. T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statistica Sinica*, 10(2):433-456.
- Wu, H. and Zhang, J. T. (2002). The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models. *Statistics in medicine*, 21(23):3655-3675.
- Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, 50(3):689-699.
- Zhang, D. (2004). Generalized linear mixed models with varying coefficients for longitudinal data. *Biometrics*, 60(1):815.
- Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, 93(442):710-719.

APPENDICES

Appendix A: Derivation of expression (2.4)

$$\begin{aligned}
P(Y_i^c \leq c_i | Y_i^o; \theta) &= E[I(Y_i^c \leq c_i) | Y_i^o; \theta] \\
&= E\{E[I(Y_i^c \leq c_i) | Y_i^o, b_i] | Y_i^o; \theta\} \\
&= E\{E[I(Y_i^c \leq c_i) | b_i] | Y_i^o; \theta\} \\
&= E\{P(Y_i^c \leq c_i | b_i) | Y_i^o; \theta\} \\
&= E\left\{\prod_{j=1}^{n_i^c} P(Y_{ij}^c \leq c_{ij} | b_i) | Y_i^o; \theta\right\} \\
&= E\left\{\prod_{j=1}^{n_i^c} \Phi\left(\frac{c_{ij} - X_{ij}^c T \beta - Z_{ij}^c T b_i}{\sigma_\epsilon}\right) \middle| Y_i^o; \theta\right\}.
\end{aligned}$$

Appendix B: Derivation of required expectations for the EM approach

$$\begin{aligned}
E(b_i b_i^T | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}) &= \int b_i b_i^T f(b_i | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}) db_i \\
&= \frac{\int b_i b_i^T f(Y_i^o, Y_i^c \leq c_i | b_i; \theta^{(r)}) f(b_i | \theta^{(r)}) db_i}{\int f(Y_i^o, Y_i^c \leq c_i | b_i; \theta^{(r)}) f(b_i | \theta^{(r)}) db_i} \\
&= \frac{\int b_i b_i^T g_1(b_i) f(b_i | \theta^{(r)}) db_i}{\int g_1(b_i) f(b_i | \theta^{(r)}) db_i}.
\end{aligned}$$

Similarly

$$\begin{aligned}
E(Y_{ij}^c | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}) &= E[E(Y_{ij}^c | Y_i^o, Y_i^c \leq c_i, b_i; \theta^{(r)}) | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}] \\
&= E[E(Y_{ij}^c | Y_{ij}^c \leq c_{ij}, b_i; \theta^{(r)}) | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}] \\
&= E(g_2(b_i) | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}) \\
&= \frac{\int g_2(b_i) g_1(b_i) f(b_i | \theta^{(r)}) db_i}{\int g_1(b_i) f(b_i | \theta^{(r)}) db_i},
\end{aligned}$$

$$\begin{aligned}
E(Y_{ij}^{c2}|Y_i^o, Y_i^c \leq c_i; \theta^{(r)}) &= E[E(Y_{ij}^{c2}|Y_i^o, Y_i^c \leq c_i, b_i; \theta^{(r)})|Y_i^o, Y_i^c \leq c_i; \theta^{(r)}] \\
&= E[E(Y_{ij}^{c2}|Y_{ij}^c \leq c_{ij}, b_i; \theta^{(r)})|Y_i^o, Y_i^c \leq c_i; \theta^{(r)}] \\
&= E(g_3(b_i)|Y_i^o, Y_i^c \leq c_i; \theta^{(r)}) \\
&= \frac{\int g_3(b_i)g_1(b_i)f(b_i|\theta^{(r)})db_i}{\int g_1(b_i)f(b_i|\theta^{(r)})db_i},
\end{aligned}$$

$$\begin{aligned}
E(b_i Y_{ij}^c | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}) &= E[E(b_i Y_{ij}^c | Y_i^o, Y_i^c \leq c_i, b_i; \theta^{(r)}) | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}] \\
&= E[b_i E(Y_{ij}^c | Y_{ij}^c \leq c_{ij}, b_i; \theta^{(r)}) | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}] \\
&= E(b_i g_2(b_i) | Y_i^o, Y_i^c \leq c_i; \theta^{(r)}) \\
&= \frac{\int b_i g_2(b_i) g_1(b_i) f(b_i | \theta^{(r)}) db_i}{\int g_1(b_i) f(b_i | \theta^{(r)}) db_i}.
\end{aligned}$$

Note that $g_2(b_i) = E(Y_{ij}^c | Y_{ij}^c \leq c_{ij}, b_i; \theta^{(r)})$ and $g_3(b_i) = E(Y_{ij}^{c2} | Y_{ij}^c \leq c_{ij}, b_i; \theta^{(r)})$. These two expectations are calculated based on the truncated normal distribution with $Y_{ij}^c | b_i \sim N(X_{ij}^{cT} \beta + Z_{ij}^{cT} b_i, \sigma_\epsilon^2)$.

Appendix C: Derivation of $A(\lambda)$ in expression (3.7)

Consider model (3.2). If the censored observations Y^c are treated as missing data, the complete data penalized log-likelihood function can be written as

$$\begin{aligned}
\ell &= -\frac{1}{2} \log |V| - \frac{1}{2} (Y - Nf - X\beta)^T \\
&\quad \times V^{-1} (Y - Nf - X\beta) - \frac{\lambda}{2} f^T K f,
\end{aligned} \tag{11}$$

where $V = \text{diag}(V_1, \dots, V_m)$ with $V_i = Z_i D(\phi) Z_i^T + \sigma_\epsilon^2 \mathbf{I}_{n_i}$ is the covariance of the complete data Y .

For a fixed smoothing parameter λ and variance components $\theta = (\phi, \sigma_\epsilon^2)$, the EM updates

of (β, f) can be easily computed. In the E-step, taking expectation of (11) with respect to censored observations will yield a Q function which can be defined as $Q(\beta, f|\beta^{(r)}, f^{(r)})$. In the M-step, differentiation of this Q function with respect to β and f and setting the derivatives to 0, one has

$$\begin{bmatrix} X^T V^{-1} X & X^T V^{-1} N \\ N^T V^{-1} X & N^T V^{-1} N + \lambda K \end{bmatrix} \begin{bmatrix} \beta \\ f \end{bmatrix} = \begin{bmatrix} X^T V^{-1} E(Y) \\ N^T V^{-1} E(Y) \end{bmatrix}, \quad (12)$$

where $E(Y) = E(Y|Y^o, Y^c \leq c; \beta^{(r)}, f^{(r)}, \theta)$. To simplify the notations, one may denote the coefficient matrix of equation (12) by H , and define $X_* = [X, N]$ and $\beta_* = (\beta^T, f^T)^T$. If X_* is of full rank, the linear system (12) has a unique solution which is given as $\hat{\beta}_* = H^{-1} X_*^T V^{-1} E(Y)$.

In the case of non-censoring, the subject-specific random effects b in model (3.2) could be estimated by the well-known empirical Bayesian estimator (Harville, 1976), i.e.,

$$\hat{b} = E(b|Y, \beta_*, \theta) = G(\phi) Z^T V^{-1} (Y - X_* \beta_*),$$

where $G(\phi) = \text{diag}(D, \dots, D)$. When censoring occurs, appropriate statistical justification is needed in estimating \hat{b} . Note that the expectation of $E(b|Y, \beta_*, \theta)$ given observed data is of the form

$$E\{E(b|Y, \beta_*, \theta)|Y^o, Y^c \leq c, \beta_*, \theta\} = G(\phi) Z^T V^{-1} \{E(Y|Y^o, Y^c \leq c, \beta_*, \theta) - X_* \beta_*\}.$$

Therefore instead of Y , $E(Y)$ should be plugged into the empirical Bayesian formula to obtain valid estimate of b . Furthermore, one can replace β_* by its estimate $\hat{\beta}_* = H^{-1} X_*^T V^{-1} E(Y)$ and write \hat{b} in a linear function of $E(Y)$ as $\hat{b} = G(\phi) Z^T V^{-1} \{\mathbf{I}_n - X_* H^{-1} X_*^T V^{-1}\} E(Y)$.

Under model (3.2), the fitted values are $\hat{Y} = X_* \hat{\beta}_* + Z \hat{b}$, and thus by the definition

$\widehat{Y} = A(\lambda)Y$ and with Y replaced by $E(Y)$ for censored cases, one will get

$$A(\lambda) = (X_*, Z) \times \begin{pmatrix} H^{-1}X_*^T V^{-1} \\ G(\phi)Z^T V^{-1}(\mathbf{I}_n - X_* H^{-1} X_*^T V^{-1}) \end{pmatrix}.$$