SEMIPARAMETRIC ANALYSIS OF TIME-TO-EVENT DATA AND LONGITUDINAL DATA

A Dissertation presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Ruiwen Zhou

Dr. (Tony) Jianguo Sun, Dissertation Supervisor

May 2021

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

SEMIPARAMETRIC ANALYSIS OF TIME-TO-EVENT DATA AND LONGITUDINAL DATA

presented by Ruiwen Zhou,

a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Jianguo Sun

Dr. Shih-Kang Chao

Dr. Shawn Ni

Dr. Yushu Shi

ACKNOWLEDGMENTS

First and foremost, my most tremendous gratitude goes to my dear advisor Dr. (Tony) Jianguo Sun for his extensive support, patience, guidance, and encouragement during my Ph.D. study. I am so grateful for his invaluable training on critical thinking and information integration ability, which is more important than the knowledge that I have learned in my doctoral studies and will be my lifelong treasure. His encouragement and patience ignite my passion for research and always motivate me to move forward. It was my great privilege and honor to work and study under his guidance.

I extend my gratitude to my doctoral committee members, Drs. Shih-Kang Chao, Xiaoguang Ni, and Yushu Shi for their insightful comments and support. Special thanks to Dr. Huiqiong Li for her great collaboration.

I would also like to thank all the professors, staff, and graduate students in the Department of Statistics. I benefited a lot from our professors' inspiring teaching and encouragement. I also appreciate the generous help from our great staff Judy Dooley and Abbie Van Nice Booher. Besides, I want to thank all my classmates and friends for accompanying me through this adventure, sharing love, laughter, tears, confusion, support, and courage with me for fighting this good fight.

Finally, my greatest gratitude goes to my parents, parents-in-law, and my husband, Dr. Xuefeng Hou. I am grateful for their unconditional love, support, and caring. Special thanks to my husband for holding my hands steadfastly during this journey, for never losing faith in me, and for being my strength in pursuing my dream.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

SEMIPARAMETRIC ANALYSIS OF TIME-TO-EVENT DATA AND LONGITUDINAL DATA

Ruiwen Zhou

Dr. (Tony) Jianguo Sun, Dissertation Supervisor

ABSTRACT

Interval-censored failure time data are commonly observed in demographical, epidemiological, financial, medical, and sociological studies. It is well-known that the proportional hazards model is one of the most used regression models for the analysis of failure time data, and significant literature has been established for fitting it to interval-censored data. Many authors have discussed the problem when complete information on the covariates is available, or the missing is completely at random. Nevertheless, an established method for the situation where the missing is at random does not seem to exist. The first part of this dissertation discusses fitting the proportional hazards model to interval-censored failure time data when there may exist missing on covariates. A sieve maximum likelihood estimation approach is proposed with the use of I-splines to approximate the unknown cumulative baseline hazard function. For the implementation of the method, we develop an EM algorithm based on two-stage data augmentation. Furthermore, we show that the proposed estimators of regression parameters are consistent and asymptotically normal.

Many authors have discussed the joint analysis of longitudinal data and time-to-event data, but most of the existing methods are the hazard-based approach for

the failure time of interest. It is well-known that sometimes the mean residual life (MRL) model, which measures the remaining life expectancy, may be of more interest. To address this issue, the second and third parts of this dissertation consider an MRL-based method for the joint analysis, which gives a meaningful and informative alternative to the hazard-based approach. In the second part, we propose to utilize the proportional mean residual life (PMRL) function with latent random effect to jointly access the observed baseline prognostic factors and continuous longitudinal risk factor on the MRL function. The proposed method extends the conventional proportional mean residual model to accommodate a latent random effect that links the time to event with longitudinal measurement. For the parameter estimation, we propose an extended estimating equation approach. The simulation study shows that the performance of the proposed method is satisfactory. We then apply the proposed method to the ADNI study that reveals insights into critical factors that influence the progression time from MCI status to AD conversion.

To further accommodate binary longitudinal outcome, in the third part, the proportional mean residual model and the generalized linear mixed model are employed to model the failure time of interest and the longitudinal variable, respectively. For estimation, a quasi-likelihood approach is developed with the use of Laplace approximation. A simulation study is conducted, and the proposed method is applied to a set of real data.

Chapter 1

Introduction

1.1 Estimation of the Proportional Hazards Model based on Interval-censored Data with Missing Covariates

It is well-known that the proportional hazards model is one of the mostly used regression models for the analysis of failure time data and a great literature has been established for fitting it to right-censored or interval-censored data, especially the former. By the latter, we mean that the failure time of interest is observed only to belong to an interval instead of being known exactly and it is apparent that the latter includes the former as a special case (Sun, 2006). Among others, the fields that generate interval-censored data include demographical, epidemiological, financial, medical and sociological studies.

As discussed by many authors, missing data can arise due to many circumstances and in general, their analysis highly depends on the censoring mechanism (Little

and Rubin, 2002). For the situation, a naive approach is the so-called complete-case (CC) method, which bases the analysis only on the complete part of the data or throw away the subjects with missing information. It is apparent that this may not only be inefficient but also yield biased estimation when the missing data mechanism depends on the observed data such that covariates may be missing at random (MAR) (Ibrahim and Chen, 1999; Little and Rubin, 2002; Qi and Prentice, 2005). Instead of the CC method, some alternatives could be the multiple imputation procedure and the estimating equation approach. As pointed by many authors, when the missing is MAR, the maximum likelihood approach may be preferred or should be used.

Several maximum likelihood methods have been proposed for regression analysis of right-censored failure time data with missing covariates under the proportional hazards model when the missing is MAR (Chen and Ying, 2002; Chen and Little, 1999; Zhou and Pepe, 1995). However, it does not seem to exist an established approach for interval-censored data with missing covariates except Wen and Lin (2011). In Wen and Lin (2011), they proposed a semiparametric maximum likelihood estimation procedure for regression analysis of current status data, a special case of interval-censored data, with missing covariates under the proportional hazards model.

To fill the research gap, we will consider the estimation of the proportional hazards model when one faces case II interval-censored data with missing covariates and propose a sieve maximum likelihood estimation approach in Chapter 2. The method can be easily implemented and makes use of I-spline functions to approximate the underlying cumulative hazard function.

1.2 Mean Residual Life Model

The mean residual life (MRL) model is a meaningful and informative alternative to hazard-based models. In clinical studies, the researchers and patients may be more interested in knowing how much the treatment can influence the remaining life expectancy rather than the hazards given the patient's current situation. For a nonnegative survival time T with finite expectation, the MRL function at time $t \geq 0$ is defined as

$$m(t) = E(T - t|T > t)$$

and can be interpreted as the remaining life expectancy of a subject given survival up to t. Oakes and Dasu (1990) first proposed the two samples proportional mean residual life model, which provides an alternative to the Cox proportional hazards model. Later on, various studies have been conducted for regression analysis of the MRL function $m(t|Z)$. Maguluri and Zhang (1994) studied the proportional mean residual life model (PMRL) defined by

$$m(t|\mathbf{Z}) = m_0(t) \exp(\boldsymbol{\beta}'\mathbf{Z})$$

without censoring.They utilized the relationship between the PMRL function and proportional hazards function under Hall-Wellner class of distributions to formulate the estimating equations.

Chen and Cheng (2005) later proposed a semiparametric inference procedure for coefficients and baseline mean residual life with censored data. They developed ad-hoc estimation for β and $m_0(t)$. Later, to fulfill the monotonic nondecreasing requirement for the life expectancy, Chen and Cheng (2006) proposed a linear mean residual life

model

$$m(t|Z) = m_0(t) + \beta^{\mathrm{T}} Z$$

and developed estimating equations based on counting processes, extended Buckley-James estimation, and Quasi partial score estimation. To have a more general model, Sun and Zhang (2009) proposed a class of transformed mean residual life models for fitting survival data under right censoring

$$m(t|Z) = g\left\{m_0(t) + \beta' Z\right\},$$

where $g : \Re \mapsto \Re$ is a prespecified link function. They utilized the inverse probability of censoring weighting approach and developed the estimating equations under independent censoring and dependent censoring.

Regression analysis of covariate effects on mean residual life has been well studied. However, the joint modeling of longitudinal outcomes and survival time has never been introduced into the MRL regression framework. Considering joint modeling under the MRL regression framework will provide an important alternative that allows people to look from a different perspective since the MRL model offers another interpretation and has been commonly used in survival analysis.

1.3 Simultaneous Analysis of Longitudinal and Survival Outcomes

Simultaneously observing longitudinal outcomes and survival endpoints on the same subject over time is quite common in the biomedical field or public health field. For

example, in the public health field, we can observe both the life quality and the death time for each subject, along with the subject's health indicators and covariates for disease history. It might be interesting to research the impacts of different patients' health indicators or disease history on life quality and death time simultaneously. For this purpose, joint modeling is necessary since it allows the researchers to take the dependence between the two types of outcomes within the same subject into account.

The joint modeling of survival and longitudinal data has been studied by many authors in the literature. One of the most commonly used models is the shared parameter model, where a random effect is included in the model to link the longitudinal outcome and survival time. Under the shared parameter model, there are mainly two subtypes, selection model and mixture model. Model selection between those two subtypes is subject to the specific research questions. The selection model focuses on estimating the distribution of survival time given the longitudinal data, so research interest lies in covariates' impacts on survival time given other longitudinal indicators. The selection model with categorical longitudinal data was considered by Faucett (1998), Huang et al. (2001), Xu and Zeger (2001), Lin and Mayne (2002), Chen and Ying (2002), Larsen (2004), Chakraborty and Das (2010). However, the mixture model is more interested in estimating the parameters of the longitudinal regression analysis given survival time. As a result, it can be applied to make inferences on longitudinal data with informative dropouts. Pulkstenis et al. (1998), considered the pattern mixture model of binary longitudinal outcomes with informative dropout. Albert and Follmann (2000) proposed to model repeated count data subject to informative dropout. Albert et al. (2002), and Albert and Follmann (2007) studied binary longitudinal data with informative missingness.

Despite the development of joint modeling of longitudinal data and survival time under the hazards function framework. However, the joint modeling of longitudinal outcomes and survival time has never been introduced into the MRL regression framework. As a result, in Chapter 3, we propose a framework for the analysis of time to event data and continuous longitudinal data under proportional mean residual life model and linear mixed model. Later, in Chapter 4, to generalize the method proposed in Chapter 3 to binary longitudinal outcome, we focus on the proportional mean residual model for survival outcome and used the generalized linear mixed model to model longitudinal data.

Chapter 2

A New Approach to Estimation of the Proportional Hazards Model based on Interval-censored Data with Missing Covariates

2.1 Introduction

As described in Section 1.1, interval-censored data with missing covariates is commonly observed in different areas. Many methods have been developed for regression analysis of interval-censored data with missing covariates under missing completely at random mechanism. When the missing is missing at random, comparing to the multiple imputation procedure and the estimating equation approach, the maximum likelihood approach may be preferred with less restriction on the missing-data mechanism.

In this chapter, we will discuss the fitting of the proportional hazards model

to interval-censored failure time data when there may exist missing on covariates under missing at random mechanism, using a sieve maximum likelihood estimation approach. For the estimation, I-spline functions are employed to approximate the unknown baseline cumulative hazard function, and a Poisson-based EM algorithm is developed. The proposed estimator of regression parameters is shown to be consistent and asymptotically normal. The simulation studies indicate that the proposed method seems to work well empirically.

The rest of the Chapter 2 is organized as follows. We will begin in Section 2.2 by introducing the model and assumption that will be used throughout the paper and then presenting the resulting likelihood functions. The proposed sieve maximum likelihood estimation approach will be derived in Section 2.3, and in particular, for the determination of the proposed estimators, an EM algorithm is developed. Section 2.4 establishes the asymptotic properties of the proposed estimators of regression parameters. Some results obtained from a simulation study are presented in Section 2.5 and suggest that the proposed approach works well in practical situations. Section 2.6 provides an application and some discussion and concluding remarks are given in Section 2.7.

2.2 Models, Assumptions and Likelihood

Consider a failure time study that involves n independent subjects and let T_i and $\mathbf{X_i}$ denote the failure time of interest and a p-dimensional vector of covariates associated with subject i. In the following, suppose that for each subject, there exist two monitoring variables or observation times U_i and V_i with $U_i < V_i$ and instead of

observing T_i, one observes only U_i and V_i and the indicator variables $\delta_{1i} = I(T_i < U_i)$, $\delta_{2i} = I(U_i \leq T_i < V_i)$ and $\delta_{3i} = 1 - \delta_{1i} - \delta_{2i}$. That is, we only know if the failure for subject i has occurred before U_i, during the examination interval $[U_i, V_i)$ or after V and observe case II interval-censor (Sun, 2006).

For the covariate effect on T_i, we will assume that given the covariates $\mathbf{X_i}$, the cumulative hazard function of T_i has the form

$$\Lambda_i(t|\mathbf{X_i}) = \Lambda_0(t) \exp\{\beta' \mathbf{X_i}\}\,, \tag{2.1}$$

where $\Lambda_0(t)$ denotes an unspecified baseline cumulative hazard function and β a p-dimensional vector of regression parameters. That is, T_i follows the proportional hazards model. In the following, we will assume that given the covariate $\mathbf{X_i}$, the failure time T_i is independent of observation times U_i and V_i or we have the independent interval censoring.

Under the assumptions above, if there is no missing covariate, the likelihood function would have the form

$$L_c(\beta, \gamma, \Lambda_0) = \prod_{i=1}^{n} f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i}; \beta, \Lambda(t))\, f(\mathbf{X_i}; \gamma)\,,$$

where $f(\mathbf{X_i}; \gamma)$ denotes the density function of the covariate with the unknown parameter γ and

$$f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i}) \propto [1 - \exp\{-\Lambda_0(V_i)\exp(\beta' \mathbf{X_i})\}]^{\delta_{1i}}$$

$$\times [\exp\{-\Lambda_0(U_i)\exp(\beta' \mathbf{X_i})\} - \exp\{-\Lambda_0(V_i)\exp(\beta' \mathbf{X_i})\}]^{\delta_{2i}}$$

$$\times[\exp\{-\Lambda_0(U_i)\exp(\beta'\mathbf{X_i})\}]^{\delta_{3i}}, \quad i = 1, \ldots, n. \tag{2.2}$$

It follows that we would have the log likelihood function

$$l_n(\beta, \gamma, \Lambda_0) = \log[L_c(\theta, \beta, \gamma, \Lambda(t))]$$

$$= \sum_{i=1}^{n} \log[f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i}; \beta, \Lambda(t))] + \sum_{i=1}^{n} \log[f(\mathbf{X_i}; \gamma)]$$

$$= l_1(\beta, \Lambda_0) \; + \; l_2(\gamma)\,. \tag{2.3}$$

It is easy to see that one can maximize $l_1(\beta, \Lambda_0)$ and $l_2(\gamma)$ separately if the goal is to estimate β, γ and Λ_0, or can ignore $l_2(\gamma)$ since γ is usually not of interest. As will be seen below, we have to estimate β, γ and Λ_0 together when there are missing covariates.

Now suppose that some covariates may be missing and the covariate can be written as $\mathbf{X_i}' = (\mathbf{X_i^{obs\prime}}, \mathbf{X_i^{mis\prime}})$, where $\mathbf{X_i^{obs}}$ denotes the components of the covariates that are known or can be observed and $\mathbf{X_i^{mis}}$ the components of the covariates that are missing. Also suppose that we can write the density function of the covariates as

$$f(\mathbf{X_i^{obs}}, \mathbf{X_i^{mis}}; \gamma) \propto \mathbf{f}(\mathbf{X_i^{obs}})\mathbf{f}(\mathbf{X_i^{mis}}|\mathbf{X_i^{obs}}; \gamma)\,.$$

Let $R_i = (R_{i1}, ..., R_{ip})'$ denote the missing indicator with $R_{ij} = 1$ if the jth component of the covariate associated with subject i is observed and 0 otherwise. In the following,

we will assume that the covariate is missing at random, meaning that

$$f(R_i|U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}, \mathbf{X_i^{mis}}, \mathbf{X_i^{obs}}) = f(R_i|U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}, \mathbf{X_i^{obs}})$$

for the conditional density function of R_i. Then the observed likelihood function has the form

$$L_o(\theta) = \prod_{i=1}^{n} \int f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|\mathbf{X_i^{mis}}, \mathbf{X_i^{obs}}; \beta, \Lambda(t)) f(\mathbf{X_i^{obs}}, \mathbf{X_i^{mis}}; \gamma) \mathbf{dX_i^{mis}},$$

where $\theta = (\beta, \gamma, \Lambda_0)$. In the next session, we will discuss estimation of θ by maximizing $L_o(\theta)$.

2.3 Sieve Maximum Likelihood Estimation

In this section, we will discuss estimation of θ by maximizing $L_o(\theta)$ with the focus on making inference about β. For this, it is apparent that it would be difficult directly to maximize it and thus we will develop an EM algorithm. Before presenting the algorithm, we will first discuss the use of the sieve approach and then the data augmentation.

It is well-known that the sieve approach can be used to approximate an unknown function in order to reduce the number of unknown parameters and the computational burden (Ma and Sun (2015); Zhao et al. (2015); Li et al. (2017)). More specifically, for the estimation here, we suggest to first approximate the baseline cumulative hazard

function $\Lambda_0(t)$ by monotone splines such as

$$\Lambda_n(t) = \sum_{l=1}^{s+k_n} \alpha_l I_l(t)$$

(Ramsay, 1988). In the above, $\{I_l(t), l = 1, \ldots, s + k_n\}$ are integrated spline basis functions with the order s and the number of knots k_n, and the α_l's are nonnegative coefficients that ensure monotonicity of $\Lambda_n(t)$. The degree s determines the smoothness of the true baseline cumulative hazard function and is often taken to be 1, 2, or 3, which corresponds to linear, quadratic, or cubic basis functions, respectively. In practice, for the choice of s and k_n, one commonly used method is to try different values of them and compare the obtained results. As an alternative, one could also use the AIC to choose the values of s and k_n that give the smallest AIC and more discussion on this is given below.

Now we discuss the data augmentation and for this, we will assume that all covariates have been observed. Then the log likelihood function $l_1(\beta, \Lambda(t))$ would have the form

$$l_1(\beta, \Lambda_0) = \sum_{i=1}^{n} \log\{[1 - \exp\{-\Lambda_0(V_i)\exp(\beta'\mathbf{X_i})\}]^{\delta_{1i}} \times [\exp\{-\Lambda_0(U_i)\exp(\beta'\mathbf{X_i})\}$$

$$-\exp\{-\Lambda_0(V_i)\exp(\beta'\mathbf{X_i})\}]^{\delta_{2i}} [\exp\{-\Lambda_0(U_i)\exp(\beta'\mathbf{X_i})\}]^{\delta_{3i}}\}. \tag{2.4}$$

By replacing Λ_0 by Λ_n, we have that

$$l_1^*(\beta, \alpha_l) = \sum_{i=1}^{n} \log\left\{[1 - \exp\{-(\sum_{l=1}^{s+k_n} \alpha_l I_l(V_i))\exp(\beta'\mathbf{X_i})\}]^{\delta_{1i}}\right.$$

$$\times[\exp\{-(\sum_{l=1}^{s+k_n}\alpha_l I_l(U_i))\exp(\beta'\mathbf{X_i})\}-\exp\{-(\sum_{l=1}^{s+k_n}\alpha_l I_l(V_i))\exp(\beta'\mathbf{X_i})\}]^{\delta_{2i}}$$

$$\times[\exp\{-(\sum_{l=1}^{s+k_n}\alpha_l I_l(U_i))\exp(\beta'\mathbf{X_i})\}]^{\delta_{3i}}\Bigg\}. \tag{2.5}$$

Note that as pointed out by McMahan and Tebbs (2013), the direct maximization of the function above with the traditional algorithm would suffer numerical instability. Also one may often get local maximizers and have other issues like convergence. In the following, we will further augment the observed data.

Let $N_i(t)$ denote the latent Poisson process with the mean function $\Lambda_n(t)\exp\{\beta'\mathbf{X_i}\}$, $i = 1,\ldots,n$, and define $Z_i = N_i(t_{1i})$ and $W_i = N_i(t_{2i}) - N_i(t_{1i})$ for $\delta_{1i} = 0$, where $t_{1i} = V_i I(\delta_{1i} = 1) + U_i I(\delta_{1i} = 0)$, and $t_{2i} = V_i I(\delta_{2i} = 1) + U_i I(\delta_{3i} = 1)$. Then Z_i and W_i are Poisson random variables with means $\Lambda_n(t_{1i})\exp\{\beta'\mathbf{X_i}\}$ and $\{\Lambda_n(t_{2i}) - \Lambda_n(t_{1i})\}\exp\{\beta'\mathbf{X_i}\}$, respectively, and they are independent given $\delta_{1i} = 0$. Furthermore, note that if T_i is left-censored or interval-censored, we have that

$$P(T_i \le t_{1i}) = P(N_i(t_{1i}) > 0) = P(Z_i > 0) = 1 - \exp\{-\Lambda_n(V_i)\exp(\beta'\mathbf{X_i})\},$$

or

$$P(t_{1i} < T_i \le t_{2i}) = P\{N_i(t_{1i}) = 0, N_i(t_{2i}) > 0\} = P(Z_i = 0, W_i > 0)$$

$$= \exp\{-\Lambda_n(U_i)\exp(\beta'\mathbf{X_i})\} - \exp\{-\Lambda_n(V_i)\exp(\beta'\mathbf{X_i})\},$$

and for right-censored T_i, we have that

$$P(T_i \ge t_{2i}) = P\{N_i(t_{2i}) = 0\} = P(Z_i = 0, W_i = 0) = \exp\{-\Lambda_n(U_i)\exp(\beta'\mathbf{X_i})\}.$$

Thus if the Z_i's and W_i's were observed, the log likelihood function corresponding to $l_1^*(\beta, \alpha_l)$ would have the form

$$l_1^{**}(\beta, \alpha_l) = \sum_{i=1}^{n} \log\{P_{Z_i}(Z_i) P_{W_i}(W_i)^{\delta_{2i}+\delta_{3i}} \{\delta_{1i} I(Z_i > 0)$$

$$+\delta_{2i} I(Z_i = 0, W_i > 0) + \delta_{3i} I(Z_i = 0, W_i = 0)\}\}.$$

In the above, $P_A(.)$ denotes the probability function associated with the random variable A.

In addition, note that one can decompose or write Z_i and W_i as $Z_i = \sum_{l=1}^{k} Z_{il}$ and $W_i = \sum_{l=1}^{k} W_{il}$, the summation of k independent Poisson random variables Z_{il}'s and W_{il}'s with means $\alpha_l I_l(t_{1i}) \exp(\beta' \mathbf{X_i})$ and $\alpha_l \{I_l(t_{2i}) - I_l(t_{1i})\} \exp(\beta' \mathbf{X_i})$, respectively. Then by treating $\{ (Z_i, W_i, Z_{il}, W_{il}, \mathbf{X_i^{mis}}) \}$ to be known, we would have the complete log likelihood function

$$l_1^{***}(\beta, \alpha_l) = \sum_{i=1}^{n} \sum_{l=1}^{k} \log \big\{ P_{Z_{il}}(Z_{il}) P_{W_{il}}(W_{il})^{\delta_{2i}+\delta_{3i}}$$

$$\times \{\delta_{1i} I(Z_i > 0) + \delta_{2i} I(Z_i = 0, W_i > 0) + \delta_{3i} I(Z_i = 0, W_i = 0)\}\}$$

corresponding to corresponding to $l_1^*(\beta, \alpha_l)$. Now we are ready to discuss the two steps of the proposed EM algorithm. Let $\mathbf{O_i} = (U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}, \mathbf{X_i^{obs}}, \mathbf{R_i})$ denote the observed data on subject i and $\theta^{(\mathbf{d})} = (\beta^{(\mathbf{d})'}, \alpha_\mathbf{l}^{(\mathbf{d})'}, \gamma^{(\mathbf{d})'})'$ the estimator of the parameters given after the d iterations. In the E-step of the $(d+1)$th iteration, we

need to determine the expectation $Q(\theta|\theta^{(d)}) = E[l_1^{***}(\beta, \alpha_l) + l_2(\gamma)|\mathbf{O_i}, \theta^{(\mathbf{d})}]$ or

$$Q(\theta|\theta^{(\mathbf{d})}) = \sum_{\mathbf{i=1}}^{\mathbf{n}} \sum_{\mathbf{l=1}}^{\mathbf{k}} [\{\mathbf{E(Z_{il}|O_i}, \theta^{(\mathbf{d})}) + (\delta_{\mathbf{2i}} + \delta_{\mathbf{3i}})\mathbf{E(W_{il}|O_i}, \theta^{(\mathbf{d})})\} \times \{\mathbf{log}(\alpha_{\mathbf{l}}) + \beta_{\mathbf{1}}' \mathbf{X_i^{obs}}\}$$

$$+\{\beta_2' E(Z_{il}\mathbf{X_i^{mis}}|\mathbf{O_i}, \theta^{(d)}) + \beta_2'(\delta_{i2} + \delta_{i3}) E(W_{il}\mathbf{X_i^{mis}}|\mathbf{O_i}, \theta^{(d)})\}$$

$$-\alpha_l \exp(\beta_1' \mathbf{X_i^{obs}}) E(\exp(\beta_2' \mathbf{X_i^{mis}}|\mathbf{O_i}, \theta^{(d)}))\{(\delta_{1i} + \delta_{2i}) I_l(V_i) + \delta_{3i} I_l(U_i)\}]$$

$$+\sum_{i=1}^{n} \int \log\{f(\mathbf{X_i^{obs}}, \mathbf{X_i^{mis}}; \gamma)\} \mathbf{f}(\mathbf{X_i^{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})}) \mathbf{dX_i^{mis}} + \mathbf{l}(\theta^{(\mathbf{d})}),$$

In the above, β_1 and β_2 denote the components of β corresponding to the observed and missing covariates, respectively, and $l(\theta^{(\mathbf{d})})$ is a function of $\theta^{(\mathbf{d})}$ free of θ.

For the determination of the expectation above, we need to calculate

$$E(Z_{il}|\mathbf{O_i}, \theta^{(d)}) = \frac{\alpha_l^{(d)} I_l(V_i) E(Z_i|\mathbf{O_i}, \theta^{(\mathbf{d})})}{\Lambda^{(d)}(V_i)},$$

and

$$E(W_{il}|\mathbf{O_i}, \theta^{(d)}) = \frac{\alpha_l^{(d)}\{I_l(V_i) - I_l(U_i)\} \times E(W_i|\mathbf{O_i}, \theta^{(\mathbf{d})})}{\Lambda^{(d)}(V_i) - \Lambda^{(d)}(U_i)},$$

where $\Lambda^{(d)}(.) = \sum_{l=1}^{k} \alpha_l^{(d)} I_l(.)$. Note that if there were no missing covariates, by following Wang et al. (2016), we would have that

$$E(Z_i|\mathbf{O_i}, \theta^{(\mathbf{d})}) = \frac{\Lambda^{(d)}(V_i) \exp(\beta_1^{(d)'} \mathbf{X_i^{obs}} + \beta_2^{(d)'} \mathbf{X_i^{mis}}) \delta_{1i}}{1 - \exp\{-\Lambda^{(d)}(V_i) \exp(\beta_1^{(d)'} \mathbf{X_i^{obs}} + \beta_2^{(d)'} \mathbf{X_i^{mis}})\}},$$

and

$$E(W_i|\mathbf{O_i}, \theta^{(\mathbf{d})}) = \frac{\{\Lambda^{(d)}(U_i) - \Lambda^{(d)}(V_i)\} \exp(\beta_1^{(d)'} \mathbf{X_i^{obs}} + \beta_2^{(d)'} \mathbf{X_i^{mis}}) \delta_{2i}}{1 - \exp[-\{\Lambda^{(d)}(U_i) - \Lambda^{(d)}(V_i)\} \exp(\beta_1^{(d)'} \mathbf{X_i^{obs}} + \beta_2^{(d)'} \mathbf{X_i^{mis}})]}.$$

When there exist missing categorical covariates, by following Lipsitz and Ibrahim (1998), we have that

$$E(Z_i|\mathbf{O_i},\theta^{(\mathbf{d})}) = \sum_{x_i^{mis}(j)} \frac{\Lambda^{(d)}(V_i)\exp(\beta_1^{(d)'}\mathbf{X_i^{obs}} + \beta_2^{(d)'}x_i^{mis}(j))\delta_{1i}p_{ij}}{1-\exp\{-\Lambda^{(d)}(V_i)\exp(\beta_1^{(d)'}\mathbf{X_i^{obs}} + \beta_2^{(d)'}x_i^{mis}(j))\}},$$

and

$$E(W_i|\mathbf{O_i},\theta^{(\mathbf{d})}) = \sum_{x_i^{mis}(j)} \frac{\{\Lambda^{(d)}(U_i)-\Lambda^{(d)}(V_i)\}\exp(\beta_1^{(d)'}\mathbf{X_i^{obs}} + \beta_2^{(d)'}x_i^{mis}(j))\delta_{2i}p_{ij}}{1-\exp[-\{\Lambda^{(d)}(U_i)-\Lambda^{(d)}(V_i)\}\exp(\beta_1^{(d)'}\mathbf{X_i^{obs}} + \beta_2^{(d)'}x_i^{mis}(j))]}.$$

Here $x_i^{mis}(j)$ denotes the jth possible missing data pattern for subject i and p_{ij} the conditional probability of a given missing data pattern, which can be estimated in the dth iteration of the EM algorithm by

$$p_{ij} = P(\mathbf{X_i^{mis}} = x_i^{mis}(j)|\mathbf{O_i},\theta^{(\mathbf{d})})$$

$$= \frac{f(U_i,V_i,\delta_{1i},\delta_{2i},\delta_{3i}|\mathbf{x_i^{obs}},x_i^{mis}(j))f(\mathbf{x_i^{obs}},x_i^{mis}(j);\gamma^{(d)})}{\sum_{x_i^{mis}(j)} f(U_i,V_i,\delta_{1i},\delta_{2i},\delta_{3i}|\mathbf{x_i^{obs}},x_i^{mis}(j))f(\mathbf{x_i^{obs}},x_i^{mis}(j);\gamma^{(d)})}.$$

For the situation where missing covariates are continuous, the calculation, which will be described at Appendix I, will involve integrations and do not have the closed forms.

In the M-step of the $(d+1)$th iteration, we need to maximize $Q(\theta,\theta^{(d)})$. For this,

one can solve the following score equations

$$\frac{\partial Q}{\partial \beta_1} = \sum_{i=1}^{n}[\{E(Z_i|\mathbf{O_i},\theta^{(\mathbf{d})}) + \delta_{2i}E(W_i|\mathbf{O_i},\theta^{(\mathbf{d})})\} - \{(\delta_{2i}+\delta_{1i})\Lambda(V_i) + \delta_{3i}\Lambda(U_i)\} \tag{2.6}$$

$$\exp(\beta_1'\mathbf{X_i^{obs}})E(\exp(\beta_2'\mathbf{X_i^{mis}}))|\mathbf{O_i},\theta^{(\mathbf{d})})]\mathbf{X_i^{obs}} = 0,$$

$$\frac{\partial Q}{\partial \beta_2} = \sum_{i=1}^{n}[\{E(Z_i\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})}) + \delta_{2i}E(W_i\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})})\} - \{(\delta_{2i}+\delta_{1i})\Lambda(V_i) + \delta_{3i}\Lambda(U_i)\}$$

$$\times\exp(\beta_1'\mathbf{X_i^{obs}})\frac{\partial E(\exp(\beta_2'\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})}))}{\partial\beta_2}] = 0, \tag{2.7}$$

$$\frac{\partial Q}{\partial \alpha_l} = \sum_{i=1}^{n}[\alpha_l^{-1}\{E(Z_{il}|\mathbf{O_i},\theta^{(\mathbf{d})}) + \delta_{2i}E(W_{il}|\mathbf{O_i},\theta^{(\mathbf{d})})\} - \{(\delta_{2i}+\delta_{1i})I_l(V_i) + \delta_{3i}I_l(U_i)\}$$

$$\times\exp(\beta_1'\mathbf{X_i^{obs}})E(\exp(\beta_2'\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})}))] = 0, \tag{2.8}$$

$$\frac{\partial Q}{\partial \gamma} = \sum_{i=1}^{n}\frac{\partial[\int\log\{f(\mathbf{X_i^{obs}},\mathbf{X_i^{mis}};\gamma)\}\mathbf{f}(\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})})\mathbf{dX_i^{mis}}]}{\partial\gamma^{(d)}} = 0\,. \tag{2.9}$$

In the above,

$$\frac{\partial E(\exp(\beta_2'\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})}))}{\partial\beta_2} = \sum_{x_i^{mis}(j)}\exp\left(\beta_2'x_i^{mis}(j)\right)x_i^{mis}(j)p_{ij},$$

$$E(\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})}) = \sum_{x_i^{mis}(j)}x_i^{mis}(j)p_{ij},$$

$$E(Z_i\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})}) = \sum_{x_i^{mis}(j)}\frac{\Lambda^{(d)}(V_i)\exp(\beta_1^{(d)'}\mathbf{X_i^{obs}} + \beta_2^{(d)'}x_i^{mis}(j))x_i^{mis}(j)\delta_{1i}p_{ij}}{1-\exp\{-\Lambda^{(d)}(V_i)\exp(\beta_1^{(d)'}\mathbf{X_i^{obs}} + \beta_2^{(d)'}x_i^{mis}(j))\}},$$

and

$$E(W_i\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})}) = \sum_{x_i^{mis}(j)}\frac{\{\Lambda^{(d)}(U_i)-\Lambda^{(d)}(V_i)\}\exp(\beta_1^{(d)'}\mathbf{X_i^{obs}} + \beta_2^{(d)'}x_i^{mis}(j))x_i^{mis}(j)\delta_{2i}p_{ij}}{1-\exp[-\{\Lambda^{(d)}(U_i)-\Lambda^{(d)}(V_i)\}\exp(\beta_1^{(d)'}\mathbf{X_i^{obs}} + \beta_2^{(d)'}x_i^{mis}(j))]}.$$

The proposed EM algorithm can be summarized as follows.

Step 1. Select the initial estimates $\beta_1^{(0)}, \beta_2^{(0)}, \alpha_l^{(0)}$ and $\gamma^{(0)}$.

Step 2. At the $(d+1)$th iteration, compute the conditional expectations $E(Z_i\mathbf{X_i^{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})})$, $E(W_i\mathbf{X_i^{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})})$, $E(Z_{il}|\mathbf{O_i}, \theta^{(\mathbf{d})})$, $E(W_{il}|\mathbf{O_i}, \theta^{(\mathbf{d})})$, $E(Z_i|\mathbf{O_i}, \theta^{(\mathbf{d})})$, and $E(W_i|\mathbf{O_i}, \theta^{(\mathbf{d})})$.

Step 3. Obtain $\hat{\beta_1}^{(d+1)}$ and $\hat{\beta_2}^{(d+1)}$ by solving the equations (2.6) and (2.7) with

$$\alpha_l^{*(d)}(\beta) = \frac{\sum_{i=1}^{n}\{E(Z_{il}|\mathbf{O_i}, \theta^{(\mathbf{d})}) + \delta_{2i}E(W_{il}|\mathbf{O_i}, \theta^{(\mathbf{d})})\}}{\sum_{i=1}^{n}[\{(\delta_{2i}+\delta_{1i})I_l(V_i) + \delta_{3i}I_l(U_i)\}\exp(\beta_1'\mathbf{X_i^{obs}})E(\exp(\beta_2'\mathbf{X_i^{mis}}|\mathbf{O_i}, \theta^{(\mathbf{d})}))]} \quad (2.10)$$

Step 4. Obtain $\hat{\alpha}_l^{(d+1)}(\beta)$ by solving the equation (2.8) and applying the Quasi-Newton method or the equation (2.10) given $\hat{\beta_1}^{(d+1)}$, $\hat{\beta_2}^{(d+1)}$.

Step 5. Obtain $\hat{\gamma}^{(d+1)}$ by solving the equation (2.9).

Step 6. Repeat Steps 2 - 5 until a pre-specified converge criterion is satisfied.

Let $\hat{\theta}_n = (\hat{\beta}_n, \hat{\gamma}_n, \hat{\Lambda}_n)$ denote the the maximum likelihood estimator of θ given by the EM algorithm above and $\hat{\theta}_n^* = (\hat{\beta}_n, \hat{\Lambda}_n)$. In the next section, we will establish the asymptotic properties of $\hat{\theta}_n^*$.

2.4 Asymptotic Properties

To describe the asymptotic properties of $\hat{\theta}_n^*$, let $\theta_0^* = (\beta_0, \Lambda_0)$ denote the true value of $\theta^* = (\beta, \Lambda_0)$ and define the distance between $\theta^1 = ({\beta_1}^1, {\beta_2}^1, \Lambda^1)$ and $\theta^2 = ({\beta_1}^2, {\beta_2}^2, \Lambda^2)$ as

$$d(\theta^1, \theta^2) = \{\, ||{\beta_1}^1 - {\beta_1}^2||^2 + ||{\beta_2}^1 - {\beta_2}^2||^2 + ||\Lambda^1 - \Lambda^2||_2^2 \,\}^{1/2}.$$

In the above, $||v||$ denotes the Euclidean norm of a vector v and $||\Lambda^1 - \Lambda^2||_2^2 = \int[\{\Lambda^1(u) - \Lambda^2(u)\}^2 + \{\Lambda^1(v) - \Lambda^2(v)\}^2]df(u,v)$, where $f(u,v)$ represents the joint density function of

U and V. Then we have the following consistency and asymptotic normality results.

Theorem 1. Assume that the regularity conditions given in Appendix II hold. Then as $n \to \infty$, we have that $d(\hat{\theta}_n, \theta_0) \to 0$ almost surely and

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \to N(0, \Sigma)$$

in distribution with Σ given in Appendix II.

The proof of the results above is sketched in Appendix II. For inference about β, it is apparent that one needs to estimate Σ and one common approach would be to employ the Louis's Formula. However, it can be seen below that this would be computationally intensive for the situation considered here and thus instead by following Wen and Lin (2011) and others, we propose to employ the nonparametric bootstrap method (Efron, 1981; Su and Wang, 2016). Specifically, let Q be an integer and for each $1 \leq q \leq Q$, draw a new data set, denoted by $O^{(q)}$, of the sample size n with replacement from the original observed data $\{ O_i; i = 1, \ldots, n \}$. Let $\hat{\beta}_n^q$ denote the estimator of β defined above based on the bootstrap samples $O^{(q)}$, $q = 1, ..., Q$. respectively. Then one can estimate the covariance matrix of $\hat{\beta}_n$ by using the sample covariance matrix of the $\hat{\beta}_n^{(q)}$'s and the numerical results below suggest that it seems to work well.

2.5 A Simulation Study

In this section, we present some results obtained from a simulation study conducted to evaluate the finite sample performance of the sieve maximum likelihood estimation procedure proposed in the previous sections. In the study, it was assumed that there exist two covariates X^{obs} and X^{miss} that followed the Bernoulli distribution with the success probabilities

0.6 and

$$\frac{\exp(1 - X^{obs})}{(1 + \exp(1 - X^{obs}))},$$

respectively. Given the covariates, the failure times of interest T_i's were generated based on model (1) with $\Lambda_0(t) = t^3$ or t.

For the missing mechanism, we considered the following two situations

$$P(R_i = 1|O_i) = \frac{\exp\{U + V + X_i^{obs}\}}{\{1 + \exp\{U + V + X_i^{obs}\}\}},$$

and

$$P(R_i = 1|O_i) = \frac{\exp\{0.22U + 0.22V + 0.22X_i^{obs}\}}{\{1 + \exp\{0.22U + 0.22V + 0.22X_i^{obs}\}\}},$$

which correspond to the missing rates of 30% and 40%, respectively. For the generation of the observation times or censoring intervals, it was assumed that the U_i's and V_i's follow the uniform distribution over the region $\{(u, v) : 0 \leq u \leq 0.28, u + 0.8 \leq v \leq 1.2\}$. The results given below are based on the sample size $n = 200$ with 1000 replications.

Table 2.1 gives the obtained results on estimation of the regression parameters β_1 and β_2 with their true values being $\{0.2, 0.5\}$ and $\{0.5, 1\}$, respectively, and the 30% missing rate. Here for the I-spline approximation to the cumulative baseline hazards function, we took $s = 3$, the degree or order of the spline basis functions, and $k_n = 5$, the number of knots is 5, and chose the knots equally spaced between the smallest and largest observation times by following Wang et al. (2016). In the table, we calculated the estimated bias given by the average of the estimates minus the true value (Bias), the sample standard error (SE), the average of the estimated standard error, and the 95% empirical coverage probability (CP). For comparison, we also applied the naive or complete data approach, denoted by CC in the table, that deleted the subjects with missing covariates and the full data approach, denoted by Full in the table, that assumed no missing covariates.

One can see from Table 2.1 that the proposed method and the Full approach gave similar performance and both seem to do better than the CC method. In particular, the proposed estimator seems to be unbiased and the bootstrapping variance estimation performed well for these situations. Also the results on the coverage probabilities indicate that the normal approximation to the distribution of the proposed estimator appears to be reasonable. To further see this, we investigated the quantile plots of the standardized estimator against the standard normal distribution and present in Figure 2.1 the plots corresponding to the situations considered in Table 2.1, which again suggest the normal approximation is appropriate. Table 2.2 displays the estimation results obtained as above except for 40% missing rate, and the results in Table 2.3 were also obtained as above except that $\Lambda_0(t) = t$. It is apparent that they gave similar conclusions as with Table 2.1 and again suggest that one should not apply the CC approach when there are missing covariates. We also considered some other set-ups and obtained similar results.

2.6 An Application

Now we apply the sieve maximum likelihood estimation procedure proposed in the previous sections to a set of data arising from Alzhehelmer's Disease Neuroimaging Initiative, discussed by Li et al. (2020) among others. The original study is a longitudinal study and among others, one variable of interest is the Alzheimers disease (AD) conversion. Due to the nature of the study, only interval-censored data are available on the occurrence time of the AD conversion, and the participants in the study are classified into three groups based on their cognitive conditions, cognitively normal, mild cognitive impairment and Alzheimer's disease. By following Li et al. (2020) and others, we will focus on the patients in the mild cognitive impairment group to determine the baseline prognostic factors or covariates for the AD conversion.

Li et al. (2020) considered five baseline covariates. They are the Rey Auditory Verbal Learning Test (RAVLT), the Middle temporal gyrus (MidTemp) from Neuroimaging, the participants's Alzheimer's Disease Assessment Scale 13 items (ADAS13), the functional assessment questionnaire score (FAQ), and the participant's baseline age (Age). Among the 396 participants in the mild cognitive impairment group, around 20% of them missed the information on the MidTemp. Also there are 3 subjects with missing ADAS13 and 3 subjects with missing FAQ, and in the analysis below, we will exclude these six subjects for simplicity.

Table 2.4 presents the analysis results given by the proposed sieve maximum likelihood estimation procedure, including the estimated covariate effect (Estimate), the estimated standard error (SE) and the p-value for testing the covariate effect being zero. For comparison, we also include in the table the results given by Li et al. (2020) based on the 316 subjects with complete information on the MidTemp. One can see that the proposed method suggests that except Age, all other four covariates, RAVLT, MidTemp, ADAS13 and FAQ, had significant effects on the AD conversion. In contrast, the approach that ignored the missing information indicates that MidTemp may only have some mild effect and ADAS13 had no effect on predicting the AD conversion. In addition, as expected, the proposed method gave more efficient estimates than Li et al. (2020) for all covariates.

2.7 Discussion and Concluding Remarks

In this chapter, we discussed the inference about the proportional hazards model when one faces interval-censored failure time data with missing covariates and for the problem, a sieve maximum likelihood estimation procedure was proposed. In the method, I-spline functions were employed to approximate the unknown baseline cumulative hazard function and a Poisson-based EM algorithm was developed. The proposed estimator of regression param-

eters was shown to be consistent and asymptotically normal, and the numerical studies indicated that the proposed method seems to work well for practical situations and should be used when covariates are missing at random.

Table 2.1: Estimation of regression parameters β_1 and β_2 with $\Lambda_0(t) = t^3$ and 30% missing covariates.

True values		Method	$\hat{\beta}_1$				$\hat{\beta}_2$			
β_1	β_2		Bias	SE	ESE	CP	Bias	SE	ESE	CP
0.5	0.5	Proposed	−0.021	0.199	0.196	95.2	−0.018	0.225	0.227	94.7
		CC	0.064	0.254	0.261	93.7	0.041	0.248	0.240	94.6
		Full	−0.026	0.196	0.205	94.2	0.000	0.197	0.200	95.8
0.2	0.5	Proposed	−0.024	0.198	0.198	94.7	−0.010	0.226	0.232	94.4
		CC	0.041	0.246	0.246	94.2	0.038	0.242	0.248	94.6
		Full	−0.017	0.196	0.195	95.4	0.014	0.193	0.196	95.4
0.5	1	Proposed	−0.093	0.211	0.212	92.8	−0.052	0.229	0.228	95.2
		CC	0.091	0.306	0.306	93.4	0.125	0.307	0.313	92.2
		Full	−0.032	0.212	0.212	94.8	−0.009	0.211	0.212	95.2

Table 2.2: Estimation of regression parameters β_1 and β_2 with $\Lambda_0(t) = t^3$ and 40% missing covariates.

True values		Method	$\hat{\beta}_1$				$\hat{\beta}_2$			
β_1	β_2		Bias	SE	ESE	CP	Bias	SE	ESE	CP
0.5	0.5	Proposed	−0.031	0.200	0.198	94.9	−0.021	0.245	0.243	94.6
		CC	0.036	0.279	0.285	94.8	0.062	0.275	0.285	94.5
		Full	−0.023	0.196	0.196	95.4	−0.024	0.197	0.198	94.6
0.2	0.5	Proposed	−0.035	0.199	0.199	94.5	−0.010	0.246	0.251	95.1
		CC	0.014	0.269	0.265	94.6	0.026	0.267	0.274	93.4
		Full	−0.017	0.195	0.196	95.4	−0.014	0.196	0.195	95.4
0.5	1	Proposed	−0.117	0.213	0.212	92.8	−0.070	0.247	0.250	93.0
		CC	0.074	0.349	0.339	94.5	0.124	0.398	0.350	92.9
		Full	−0.055	0.212	0.210	94.1	−0.037	0.211	0.204	95.4

Table 2.3: Estimation of regression parameters β_1 and β_2 with $\Lambda_0(t) = t$ and 30% missing covariates.

True values		Method	$\hat{\beta}_1$				$\hat{\beta}_2$			
β_1	β_2		Bias	SE	ESE	CP	Bias	SE	ESE	CP
0.5	0.5	Proposed	0.021	0.203	0.213	94.7	0.049	0.243	0.247	94.3
		CC	0.101	0.250	0.248	93.3	0.082	0.245	0.237	93.0
		Full	0.016	0.194	0.196	94.4	0.005	0.194	0.189	95.3
0.2	0.5	Proposed	−0.015	0.196	0.196	95.4	0.004	0.235	0.230	94.8
		CC	0.056	0.242	0.238	94.5	0.064	0.239	0.227	93.3
		Full	0.007	0.191	0.189	95.2	−0.000	0.191	0.188	95.1
0.5	1	Proposed	0.013	0.250	0.247	95.2	0.148	0.290	0.278	92.5
		CC	0.150	0.297	0.308	92.0	0.198	0.303	0.312	89.9
		Full	0.071	0.223	0.221	94.0	0.108	0.222	0.227	92.6

Table 2.4: Analysis results of Alzhehelmer's Disease data

Covariate	Method	Estimate	SE	p-value
RAVLT	Li et al. (2020)	−0.679	0.324	0.018
	Proposed	−0.305	0.096	0.001
Midtemp	Li et al. (2020))	−0.434	0.290	0.072
	Proposed	−0.291	0.075	0.000
ADAS13	Li et al. (2020)	0.380	0.690	0.291
	Proposed	0.410	0.094	0.000
FAQ	Li et al. (2020)	0.426	0.244	0.040
	Proposed	0.410	0.071	0.000
Age	Li et al. (2020)	−0.364	0.274	0.092
	Proposed	−0.087	0.083	0.147

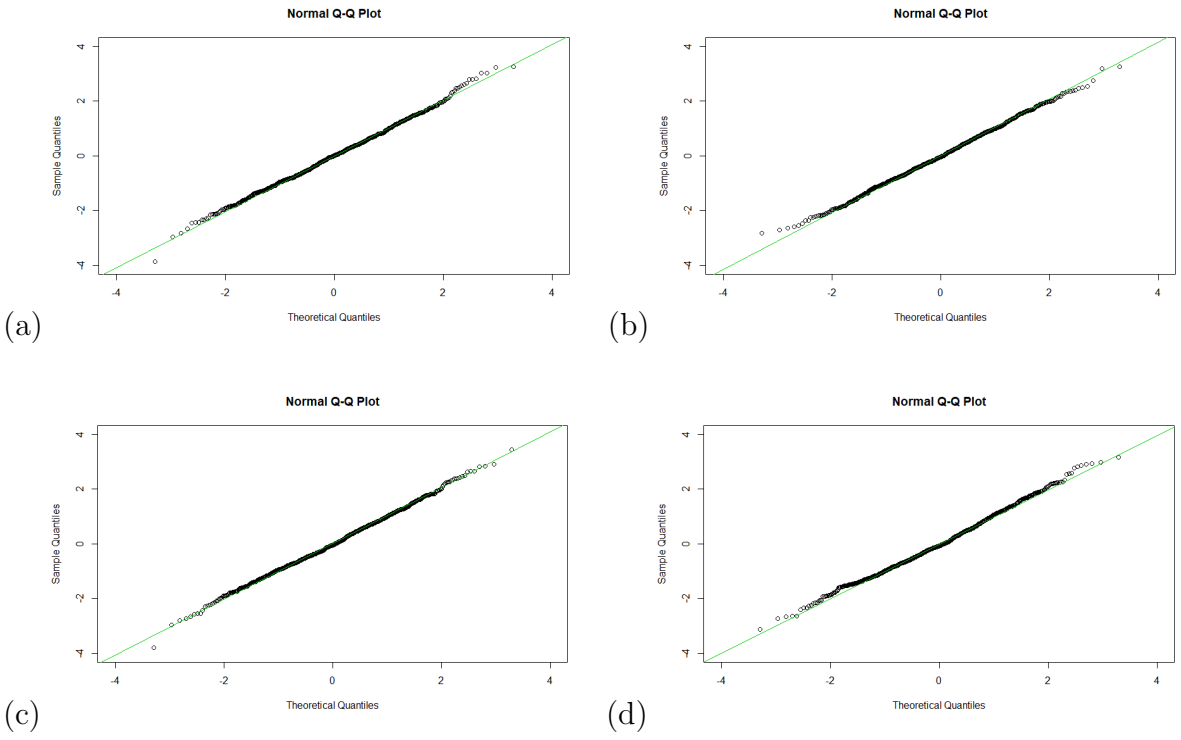

Figure 2.1: Quantile plots of the standardized estimates of (a) β_1 with $\beta_1 = \beta_2 = 0.5$, (b) β_2 with $\beta_1 = \beta_2 = 0.5$, (c) β_1 with $\beta_1 = 0.5$ and $\beta_2 = 1$, and (d) β_2 with $\beta_1 = 0.5$ and $\beta_2 = 1$.

Chapter 3

Joint Modeling of Continuous Longitudinal and Survival Outcomes under the Proportional Mean Residual Life Model

3.1 Introduction

In survival analysis literature, the additive hazards and proportional hazards models are commonly used models to search for prognostic factor effects on the hazard rate of the event of interest. However, when the mean residual life (MRL) is of interest rather than the hazard, the mean residual life model is necessary to show the correlations between risk factors and the mean residual life function. The mean residual function at time $t \geq 0$ is defined as $m(t) = E(T - t \mid T > t)$, where T denotes the survival time of interest. Since Oakes and Dasu first proposed MRL function, various works have been done. Maguluri and Zhang (1994), Chen and Cheng (2005), considered the proportional MRL model for

survival outcome without and with censoring. Later, Chen and Cheng (2006), Chen (2007) developed an additive RML model to further loosen restrictions in the proportional MRL (PMRL) model.

MRL function (MRLF) is widely applied in many fields, such as survival analysis and reliability research. The methodology developed in this project is motivated by the phase-1 of ADNI study. Within the 800 study subjects, about 200 subjects are cognitive normal at baseline, 400 patients are with mild cognitive impairment (MCI) status at baseline, and 200 AD conversion patients at baseline. The participants were checked at 0, 6, 12, 18, 24, and 36 month. At each visit, their cognitive status, neuropsychological assessments, and clinical measurements were recorded. In the literature, many pieces of research have been done on the MCI population. MCI is a risky transitional state between normal cognition and AD. About 32% of MCI patients will transit to Alzheimer's dementia within five years. At present, there is no cure existing for AD. Medications and treatments can only temporarily improve the symptoms (Lin et al., 2020). Therefore, it is important for the researchers to know the prognostic factors and evaluate the treatments that can prolong the progression time from MCI to AD. To this end, the mean residual time function for AD onset time is more informative than the hazards of AD development when we are interested in prolonging the progression time from MCI to AD conversion.

Joint models of longitudinal and survival outcomes were discussed by many authors. (Wulfsohn and Tsiatis, 1997; Liu et al., 2004; Elashoff, Li, et al., 2016; Shen et al., 2019). Under the shared parameter model, the assumed submodels for survival outcome and longitudinal outcomes are linked by using a common latent structure. Compared to only include baseline longitudinal information as covariates in the survival model, joint modeling with time-dependent longitudinal outcome including the dependency of the two types of outcomes will reduce the bias for parameter estimation and improve statistical inference efficiency. However, there seems not exist work that focuses on joint model longitudinal

outcomes and survival outcome under MRL model. The development of methodology in this chapter will provide an informative alternative for the joint analysis of two types of outcomes.

As a result, in this chapter, we provide a unified framework that simultaneous regression analysis of time-to-event data and longitudinal data under proportional mean residual life model and linear mixed model. For the estimation process, we propose a two-step procedure, which use the Restricted Maximum Likelihood to estimate the parameters and latent random variable of the longitudinal model in the first step and estimate the parameters in the proportional MRL (PMRL) model in the second step with extended estimation equation. In the second step, similar to He et al. (2017), the baseline mean residual life function and the parameters are all involving in the exponential function of latent variables, which adding additional computational difficulty and requiring technically demanding estimating equations.

The rest of this chapter is organized as followed. Section 3.2 presents the models for longitudinal and survival outcomes. The two-step estimation procedure for unknown parameters and baseline MRL function is developed in Section 3.3. Section 3.4 shows the simulation studies to exam the empirical performance of the proposed method with finite sample sizes. An application to the ADNI study for the MCI patients is presented in Section 3.5 We conclude the chapter in Section 3.6 with some discussions about the present work.

3.2 Model Formulation and Notation

Let $\boldsymbol{Y}(t)$ denote the longitudinal process at time t and suppose that $\boldsymbol{Y}(t)$ follow the linear mixed model. We use T to denote the survival time, and the survival time is subject to right-censoring. Suppose that a random sample of subjects with size n is followed during time $[0, \tau]$, where τ is the maximum follow-up time. Subject-specific random effects are denoted

by $\boldsymbol{b}_i, i = 1, ..., n$, and we assume that the $\boldsymbol{b}_i's$ are mutually independent and identically distributed from $MVN(0, \boldsymbol{\Sigma}_b)$.

Under the linear mixed model assumption, given the random effect $\boldsymbol{b}_i$, the longitudinal outcome $\boldsymbol{Y}_i(t)$ at time t for subject i is from the following

$$\mathbf{Y_i(t)} = \mathbf{X_i^T}\boldsymbol{\beta} + \tilde{\mathbf{X}}_\mathbf{i}^\mathbf{T}\mathbf{b_i} + \epsilon_\mathbf{i}. \tag{3.1}$$

In the above, we denote $\boldsymbol{X}_i(t)$ and $\widetilde{\boldsymbol{X}}_i(t)$ as the vector of observed covariates for subject i at time t, $\boldsymbol{X}_i(t)$ represent the covariates that only affect longitudinal outcome, and $\widetilde{\boldsymbol{X}}_i(t)$ represent the covariates that affect both longitudinal and survival outcomes. The covariates $\boldsymbol{X}_i(t)$ and $\widetilde{\boldsymbol{X}}_i(t)$ can be completely different or share some components. The parameter $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_{p_n})$ is a $p_n + 1$ dimensional vector of unknown parameters and ϵ_i denotes the measurement error process and from $MVN(0, \Psi_\epsilon)$.

Given the random effect $\boldsymbol{b_i}$ and the observed covariates, the conditional mean residual life time function for the survival time T_i for subject i at time t is assumed to follow the proportional mean residual life model

$$m(t|\boldsymbol{b}_i, \mathbf{Z}(t)) = m_0(t) \exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i(t) + \tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\psi} \circ \boldsymbol{b}_i)\right). \tag{3.2}$$

Here we denote $\boldsymbol{Z}_i(t)$ and $\widetilde{\boldsymbol{Z}}_i(t)$ as the vector of observed covariates for subject i at time t, $\boldsymbol{Z}_i(t)$ represent the covariates that only affect survival outcome, and $\widetilde{\boldsymbol{Z}}_i(t)$ represent the covariates that affect both longitudinal and survival outcomes. The covariates $\boldsymbol{Z}_i(t)$ and $\widetilde{\boldsymbol{Z}}_i(t)$ can be completely different or share some components. Additionally, $m_0(t)$ denotes the unknown baseline mean residual life function, $\boldsymbol{\gamma}$ is the vector of parameter for the fixed effect, $\boldsymbol{\psi}$ is the vector of parameter for the random effect, and $\boldsymbol{\psi} \circ \boldsymbol{b}_i$ denotes the component-wise product. Under model (3.1) and model (3.2), we assume that the longitudinal outcome $\boldsymbol{Y}_i(t)$ and survival time T_i are independent given the random effect $\mathbf{b}_i$ and the observed

covariates. The parameter $\boldsymbol{\psi}$ indicate the dependence between longitudinal outcome and survival time linked by the latent random variable $\mathbf{b}_i$. If $\boldsymbol{\psi}$ is significantly not equal to 0, it means the longitudinal outcome and survival time are correlated by the latent random variable. If not, the dependence between the longitudinal outcome and survival time is not caused by random latent variable.

3.3 Two Step Estimation Procedure

Let $\mathbf{Y} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_n), \boldsymbol{b} = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n)$, and $\boldsymbol{\theta}$ be a vector containing unknown components of $\{\boldsymbol{\Psi}_\epsilon, \boldsymbol{\Sigma}_b\}$. The full likelihood for longitudinal observation is

$$Q(\boldsymbol{\theta}, \boldsymbol{\beta}) = \int l(\boldsymbol{\theta}, \boldsymbol{\beta} \mid \boldsymbol{b}) p(\boldsymbol{b})\, d\boldsymbol{b}.$$

Here, $l(\boldsymbol{\theta} \mid \boldsymbol{b})$ is the complete-data log-likelihood function with the form of

$$\begin{aligned} l(\boldsymbol{\theta}, \boldsymbol{\beta} \mid \boldsymbol{b}) = -\frac{1}{2}\{&n(q+n_i)\log(2\pi) + n\log|\boldsymbol{\Psi}_\varepsilon| + n\log|\boldsymbol{\Sigma}_{\mathbf{b}}| \\ &+ \sum_{i=1}^{n}\left(\mathbf{Y}_i - \mathbf{x_i^T}\beta - \tilde{\mathbf{x}}_\mathbf{i}^\mathbf{T}\mathbf{b_i}\right)^T \boldsymbol{\Psi}_\varepsilon^{-1}\left(\mathbf{Y}_i - \mathbf{x_i^T}\beta - \tilde{\mathbf{x}}_\mathbf{i}^\mathbf{T}\mathbf{b_i}\right) + \sum_{i=1}^{n} b_i^T {\boldsymbol{\Sigma}_\mathbf{b}}^{-1} b_i\}, \end{aligned}$$

where $p(\boldsymbol{b}_i \mid \mathbf{Y}, \boldsymbol{\theta}) \overset{D}{=} N\left(\boldsymbol{\mu}_{b_i}, \boldsymbol{\Sigma}_*\right)$ with $\boldsymbol{\mu}_{b_i} = \boldsymbol{\Sigma}_b \tilde{\mathbf{X}}^T \boldsymbol{\Psi}_\varepsilon^{-1}(\mathbf{Y}_i - \mathbf{x_i^T}\beta)$ and $\boldsymbol{\Sigma}_* = \left({\boldsymbol{\Sigma}_\mathbf{b}}^{-1} + \tilde{\mathbf{X}}^T\boldsymbol{\Psi}_\varepsilon^{-1}\tilde{\mathbf{X}}\right)^{-1}$. The estimated parameter $\hat{\boldsymbol{\theta}}$ can be obtained using restricted maximum likelihood estimation, and fixed effect parameter $\hat{\boldsymbol{\beta}}$ can be estimated by $\hat{\boldsymbol{\beta}} = \left\{\boldsymbol{X}^T\boldsymbol{\Psi}_\varepsilon^{-1}\boldsymbol{X}\right\}^{-1}\boldsymbol{X}^T\boldsymbol{\Psi}_\varepsilon^{-1}\boldsymbol{Y}$. For the inference of the proportional mean residual life model, following He et al. (2017), we propose the borrow-strength estimation method given the latent random variable $\boldsymbol{b}$.

Let C_i denote the censoring time for subject i, and $\Delta_i = I(T_i \leq C_i)$ denote the censoring indicator. The observed time is $V_i(t) = min\{T_i, C_i\}$. Let $R_i(t) = I(V_I > t)$ denote the

at-risk process. We assume that T_i and C_i are independent given covariates and latent random effect $\boldsymbol{b}$. For the estimation of baseline mean residual life function $m_0(t)$, along the same line of the method in the work of Huang et al. (2019), we formulate the estimating equations by IPCW technique with a class of zero-mean stochastic process

$$M_i(t) = \frac{\Delta_i I\left(V_i > t\right)}{G\left(V_i\right)}\left[\left(V_i - t\right) - m_0(t)\exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i(t) + \tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\psi}\circ\boldsymbol{b}_i)\right)\right]. \tag{3.3}$$

In the above, $G(\cdot)$ denotes the distribution function for censoring time C. The censoring survival function $G(V_i)$ can be estimated based on $\{(V_i, 1-\Delta_i) : i = 1, \ldots, n, j = 1,\}$ by using Kaplan-Meier (K-M) estimator $\hat{G}(t) = \prod_{k:C_k \le t}\left(1 - \frac{d_k^c}{n_k^c}\right)$,
where $d_k^c = \sum_i \left(1-\Delta_i\right) I\left(V_i = C_k\right), n_k^c = \sum_i I\left(V_i \ge C_k\right)$ and $0 < C_1 < C_2 < \ldots < C_D$ are distinct censoring times.

Plugging the K-M estimators to equation (3.3), by assuming that we know the parameters, covariates, and latent random effect, we can obtain an estimating equation for $m_0(t)$ as

$$\frac{\Delta_i I\left(V_i > t\right)}{\hat{G}\left(V_i\right)}\left[\left(V_i - t\right) - m_0(t)\exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i(t) + \tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\psi}\circ\boldsymbol{b}_i)\right)\right] = 0. \tag{3.4}$$

Here, $0 \le t \le \tau$ and $0 \le \tau = \inf\{t : P(T \ge t) = 0\} < \infty$. Then we can have a closed form for $m_0(t)$ below

$$\hat{m}_0(t;\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{b}) = \frac{\sum_i \frac{\Delta_i I(V_i > t)}{\hat{G}(V_i)}\left(V_i - t\right)}{\sum_i \frac{\Delta_i I(V_i > t)}{\hat{G}(V_i)}\exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i(t) + \tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\psi}\circ\boldsymbol{b}_i)\right)}, \quad 0 \le t \le \tau.$$

However, the random effect $\boldsymbol{b}_i's$ are latent variable and can not be observed. As the result, we can not apply estimation equation (3.4) directly. To address this issue, following He et al. (2017), we apply the estimator of $\boldsymbol{b}_i$ based on longitudinal model (3.1) with known

$\boldsymbol{\theta}$, which takes the following form,

$$\hat{\boldsymbol{b}}_i(\boldsymbol{\theta}) = \boldsymbol{\Gamma}(\boldsymbol{\theta})(Y - X\boldsymbol{\beta}),$$

where $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \left(\tilde{\mathbf{X}}^T \boldsymbol{\Psi}_\varepsilon^{-1} \tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^T \boldsymbol{\Psi}_\varepsilon^{-1}$ is a $q \times q$ matrix of $\boldsymbol{\theta}$, and q is the dimension of random effect $\boldsymbol{b}_i$. With the estimated random effect $\hat{\boldsymbol{b}}_i(\boldsymbol{\theta})$, we can estimate $m_0(t)$ by

$$\hat{m}_0(t; \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{b}) = \frac{\sum_i \frac{\Delta_i I(V_i > t)}{\hat{G}(V_i)} (V_i - t)}{\sum_i \frac{\Delta_i I(V_i > t)}{\hat{G}(V_i)} \exp\left(\boldsymbol{\gamma}' \mathbf{Z}_i(t) + \tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\psi} \circ \hat{\boldsymbol{b}}_i(\theta))\right)}, \quad 0 \le t \le \tau.$$

Given the estimated baseline MRL function $\hat{m}_0(t)$ and latent random effect variable $\boldsymbol{b}$, according to Sun and Zhang (2009), we can formulate the estimating equation from stochastic process (3.3) for the parameter $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$

$$\boldsymbol{D}(\boldsymbol{\gamma}, \boldsymbol{\psi} \mid \boldsymbol{b}) = \sum_i \int_0^\tau \frac{\Delta_i I\left(V_i > t\right) Z_i}{\hat{G}\left(V_i\right)} \left[(V_i - t) - \hat{m}_0(t) \exp\left(\boldsymbol{\gamma}' \mathbf{Z}_i + \tilde{\boldsymbol{Z}}_i(\boldsymbol{\psi} \circ \boldsymbol{b}_i)\right)\right] \mathrm{d}\mathcal{H}(t). \tag{3.5}$$

In the above, $\mathcal{H}(t)$ is an increasing weight function and can be specified in different forms. We set $\mathcal{H}(t)$ to be the observed counting process $N(t) = \sum_{ij} I\left(Y_{ij} \le t\right) \delta_{ij}$ due to its good performance as shown in the simulation study by Sun and Zhang (2009). Note that $E(\boldsymbol{D}(\boldsymbol{\gamma}, \boldsymbol{\psi} \mid \boldsymbol{b})) = 0$. As a result, we can estimate $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$ by solving estimating equation $\boldsymbol{D}(\boldsymbol{\gamma}, \boldsymbol{\psi} \mid \boldsymbol{b}) = \mathbf{0}$.

However, even though $E(\hat{\boldsymbol{b}} \mid \boldsymbol{b}) = \boldsymbol{b}$, simply replacing $\boldsymbol{b}$ by $\hat{\boldsymbol{b}}$ in $\boldsymbol{D}(\boldsymbol{\gamma}, \boldsymbol{\psi} \mid \boldsymbol{b})$ would lead to biased estimator since

$$E\left[\exp\left\{\boldsymbol{\psi}^T \hat{\boldsymbol{b}}_i(\boldsymbol{\theta})\right\} \mid \boldsymbol{b}_i\right] = \exp\left\{\boldsymbol{\psi}^T \boldsymbol{b}_i + \frac{1}{2} \boldsymbol{\psi}^T \mathbf{D}(\boldsymbol{\theta}) \boldsymbol{\psi}\right\} \tag{3.6}$$

and

$$E\left[\exp\left\{\psi^T\hat{b}_i(\theta)\right\}\hat{b}_i(\theta)\mid b_i\right]=\exp\left\{\psi^T b_i+\frac{1}{2}\psi^T\mathbf{D}(\theta)\psi\right\}\{b_i+\mathbf{D}(\theta)\psi\} \tag{3.7}$$

where $D(\theta)=\left(\tilde{\mathbf{X}}^T\boldsymbol{\Psi}_\varepsilon^{-1}\tilde{\mathbf{X}}\right)^{-1}$.

To eliminate the bias, using equations (3.6) and (3.7), we can develop the following extended estimating equations for the parameters in the proportional mean residual model with

$$\begin{aligned}\boldsymbol{D}(\boldsymbol{\gamma}\mid\boldsymbol{b})=&\sum_i\int_0^\tau\frac{\Delta_i I\left(V_i>t\right)\bar{Z}_i}{\hat{G}\left(V_i\right)}[(V_i-t)-\hat{m}_0(t)\exp(\boldsymbol{\gamma}'\mathbf{Z}_i+\tilde{\boldsymbol{Z}}_i(\boldsymbol{\psi}\boldsymbol{b}_i)-\\&\frac{\psi\tilde{Z}_i^T D(\theta)\psi\tilde{Z}_i}{2})]\mathrm{d}\mathcal{H}(t),\end{aligned} \tag{3.8}$$

$$\begin{aligned}&\boldsymbol{D}(\boldsymbol{\psi}\mid\boldsymbol{b})\\&=\sum_i\int_0^\tau\frac{\Delta_i I\left(V_i>t\right)b_i^*\tilde{Z}_i^*}{\hat{G}\left(V_i\right)}[(V_i-t)-\hat{m}_0(t)\exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i+\tilde{\boldsymbol{Z}}_i(\boldsymbol{\psi}\circ\boldsymbol{b}_i)\right.\\&-\frac{\psi\tilde{Z}_i^T D(\theta)\psi\tilde{Z}_i}{2}]\mathrm{d}\mathcal{H}(t)+\sum_i\int_0^\tau\frac{\Delta_i I\left(V_i>t\right)b_i\bar{Z}_i^*}{\hat{G}\left(V_i\right)}[\hat{m}_0(t)\\&\exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i+\tilde{\boldsymbol{Z}}_i(\boldsymbol{\psi}\circ\boldsymbol{b}_i)-\frac{\psi\tilde{Z}_i^T D(\theta)\psi\tilde{Z}_i}{2}\right)\psi D(\theta)\tilde{Z}_i]\mathrm{d}\mathcal{H}(t).\end{aligned} \tag{3.9}$$

With $\boldsymbol{\theta}$, we can estimate the parameters for proportional MRL by solving $\boldsymbol{D}(\boldsymbol{\gamma}\mid\boldsymbol{b})=\mathbf{0}$ and $\boldsymbol{D}(\boldsymbol{\psi}\mid\boldsymbol{b})=\mathbf{0}$ with Newton-Raphson algorithm. Then we can update $\hat{m}_0(t,\boldsymbol{\gamma},\boldsymbol{\psi})$ with closed form equation (3.10), for $0\le t\le\tau$.

$$\hat{m}_0(t;\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{b}) = \frac{\sum_i \frac{\Delta_i I(V_i>t)}{\hat{G}(V_i)}\,(V_i - t)}{\sum_i \frac{\Delta_i I(V_i>t)}{\widehat{G}(V_i)} \exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i + \tilde{\boldsymbol{Z}}_i(\boldsymbol{\psi}\circ\boldsymbol{b}_i) - \frac{\boldsymbol{\psi}\tilde{\boldsymbol{Z}}_i^T D(\theta)\boldsymbol{\psi}\tilde{\boldsymbol{Z}}_i}{2}\right)}, \quad 0 \le t \le \tau. \quad (3.10)$$

For the variance estimation, we use bootstrapping method to sample with replacement from the observed data. Then we can estimate the variance by the empirical variance estimator from the bootstrap sample estimation.

3.4 A Simulation Study

In this section, we conduct a simulation study to show the empirical performance of the proposed method. In this simulation study, we assume that the longitudinal outcome Y_{ij} is continuous variable generated from

$$Y_{ij}(t) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3ij} + b_i + \epsilon_i, \quad (3.11)$$

for $i = 1, ..., n$, and $j = 1, ..., n_i$. The survival outcome were generated from proportional MRL model as follow,

$$m\left(t \mid b_i\right) = m_0(t) \exp\left\{\gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \psi b_i * Z_{3i}\right\},$$

where $b_i \sim N\left(0, \sigma_b^2\right), \epsilon_i \sim N\left(0, \sigma^2\right)$, $X_{1i} \equiv Z_{1i}$ was generated from a Bernoulli(0.5), and $X_{2i} \equiv Z_{2i}$ was simulated from the Unif$(0, 1)$. They were included in both hazard and longitudinal models. The covariate Z_{i3} was generated from the standard normal distribution, and it was included only in the mean residual model. We denote X_{3ij} as the time at measurement generated uniformly from 0 to maximum observation time $\tau = 2.4$ with the number of observation times $n_i = 10$, and it was included only in the longitudinal model.

Table 3.1 and Table 3.2 present the results on estimation of the regression parameters β_0, β_1, β_2, β_3, γ_1, γ_2, ϕ, and σ_b^2 with baseline mean residual function $m_0(t) = -0.5t + 0.5$, $m_0(t) = -0.5t + 1$, and 10%, 20% censoring rate, respectively. The sample size in the Table 3.1 and Table 3.2 is $n = 500$. In the Table 3.3, we increased the sample size to $n = 1200$.

In Tables $3.1 - 3.3$, we calculated the estimated bias given the average of the estimates over 500 replication minus the true value (Bias), the sample standard deviation (SE), the average of estimated standard error (ESE), and the 95% empirical coverage probability (CP). From Tables $3.1 - 3.3$, we can see that the bias of the proposed estimation is small, and the bootstrapping variance estimation performs well in different setups. The coverage probabilities that the normal approximation to the parameter distribution seems to be reasonable. Furthermore, as shown in Table 3.3, when the sample size increased from 500 to 1200, the parameter estimation bias is closer to zero with a smaller standard error.

3.5 Real Data Example

In this section, we will apply the proposed joint modeling approach to analyze the data from the first phase of the Alzheimer's Disease Neuroimaging Initiative study (http://adni.loni.usc.edu/) described in section 3.1. We are interested in identifying the important risk factors that have prognostic effects on the progression time-to-AD in the MCI population.

For the analysis below, we include 351 subjects with at least one follow-up and with complete information on 22 covariates and risk factors. Since the exact date of conversion from MCI to AD was unknown, following Li et al. (2018), we define the survival endpoint as the date of the first follow-up with an AD diagnosis. Subjects that did not convert were censored at the date of their last interview. For the application of the proposed method, T_i denotes the AD conversion time for subject i, $m(t) = E(T - t \mid T > t)$ denotes the

expectancy of progression time to AD diagnosis given the subject with MCI as initial status progresses to AD after time t, and $Y_i(t)$ is the the Alzheimer's Disease Assessment Scale Score 13 (ADAS13).

ADAS13 represents a composite score of 13 items and ranges from 0 to 85 measuring the cognitive functions. Higher ADAS13 score indicates poor cognitive function. According to Li et al. (2017), ADAS13 is the strongest longitudinal predictor of MCI-to-AD conversion diagnosis. Figure 3.1 displays the smooth curves of ADAS13 over time for three subgroups in the sample: red curve represents for participants with less than 3 years follow-up time; green curve represents for participants with more than 3 years and less than 6 years follow-up time; blue curve represents participants with more than 6 years follow-up time. Shaded regions are 95% pointwise confidence intervals. Three curves all show that the ADAS13 score increases with the follow-up time in three subgroups. However, ADAS13 score tends to be higher with a shorter follow-up time, which means the patients with higher ADAS13 score and more severe cognitive impairment have a shorter progression time to AD conversion. This figure indicates a strong correlation between ADAS13 and progression time to AD conversion. This correlation between longitudinal outcome and survival outcome indicates dependence. As a result, separate modeling longitudinal outcome and the terminal event will ignore the dependence and lead to a biased inference.

To identify the covariates and risk factors that have effects on the progression time to AD conversion. Following Yi and Sun (2020), the longitudinal and survival models are

assumed to be

$$
\begin{aligned}
y_i\left(t_{ij}\right) =& \alpha_0+\alpha_1 t_{ij}+\alpha_2 \text{ Age }_i+\alpha_3 \text{ APOE } \epsilon 4_i+\alpha_4 \text{ CDRSB } \\
&+\alpha_5 \text{ DIGIRSCOR }+\alpha_6 \text{ TRABSCOR }+\alpha_7 \text{ FAQ}+\alpha_8 \text{ Ventricles } \\
&+\alpha_9 \text{ Hippocampus}+\alpha_{10} \text{ WholeBrain }+\alpha_{11} \text{ Entorhinal }+\alpha_{12} \text{ Fusiform } \\
&+\alpha_{13} \text{ MidTemp}+\alpha_{14} \text{ ICV}+\alpha_{15} \text{ RAVLT.i}+\alpha_{16} \text{ RAVLT.l}+\alpha_{17} \text{ RAVLT.f } \\
&+\alpha_{18} \text{ RAVLT.perc.f}+b_i+\varepsilon_i\left(t_{ij}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
m_i(t) =& m_0(t) \exp(\beta_1 t_i+\beta_2 \text{ Age }_i+\beta_3 \text{ APOE } \epsilon 4_i+\beta_4 \text{ CDRSB }+\beta_5 \text{ DIGIRSCOR } \\
&+\beta_6 \text{ TRABSCOR }+\beta_7 \text{ FAQ}+\beta_8 \text{ Ventricles }+\beta_9 \text{ Hippocampus } \\
&+\beta_{10} \text{ WholeBrain }+\beta_{11} \text{ Entorhinal }+\beta_{12} \text{ Fusiform } \\
&+\beta_{13} \text{ MidTemp }+\beta_{14} ICV+\beta_{15} \text{ RAVLT.i }+\beta_{16} \text{ RAVLT.l } \\
&+\beta_{17} \text{ RAVLT.f }+\beta_{18} \text{ RAVLT.perc.f }+\beta_{19} \text{ PTMARRY } \\
&+\beta_{20} \text{ PTEDUCA }_i+\beta_{21} \text{ MMSE }+\beta_{22} \text{ ADASQ4}+\gamma b_i).
\end{aligned}
$$

The results in Table 3.4 show that ADAS13 was clearly correlated with observation time, age, APOE $\epsilon 4_i$, DIGIRSCOR (participant's digit symbol substitution test score), TRABSCOR (Trails B score), Entorhinal (MRI entorhinal volume), Midtemp, RAVLT.i (Rey auditory verbal learning test score of immediate recall). With time progression, the presence of at least one apolipoprotein E allele (APOE $\mathcal{E}4$), higher TRABSCOR, the patients is more likely to have higher ADAS 13 score with poorer cognitive function. Besides, elder group, people with higher DIGIRSCOR, larger Entorhinaln volume, larger Midtemp (middle temporal gyrus), and higher RAVLT.i score will have smaller ADAS13 score with better cognitive function. Furthermore, the larger observation time, higher RAVLT.f score, and higher education level will increase the progression time from MCI to AD conversion. The

presence of at least one apolipoprotein E allele (APOE $\mathcal{E}4$), larger MRI entorhinal volume, and higher score of RAVLT.perc.f (higher forgetting percent) will decrease the progression time to AD conversion.

The conclusions from the above analysis are similar to Yi and Sun (2020) with DIGIRSCOR and RAVLT.i score. Both show that a high DIGIRSCOR and a small RAVLT.i score indicate a large ADAS13 score. However, we find significant effects of observation time, age, APOE $\epsilon 4_i$, TRABSCOR (Trails B score), Entorhinal (MRI entorhinal volume), and Midtemp on the ADAS13 score, which were shown to be insignificant in Yi and Sun (2020). For the survival outcome, similar to Yi and Sun (2020), the analysis result with the proposed method also indicates that higher score of RAVLT.perc.f (higher forgetting percent) will decrease the progression time to AD conversion, therefore have higher hazard of developing AD conversion as stated in Yi and Sun (2020). On the contrary, with the proposed method, we find that the baseline time, the presence of at least one apolipoprotein E allele, entorhinal volume, and RAVLT.f have significant effects on the expectancy progression time-to-AD, but not on the hazards of AD conversion.

3.6 Discussion

This chapter discussed the estimation for joint analysis of longitudinal data and right-censored failure time data under a proportional MRL model. As mentioned above, the MRL function is an important alternative method to the hazard model. When MRL function or life expectancy is of interest, the proposed approach can increase estimation accuracy and efficiency by joint modeling longitudinal outcome and survival outcome. Under the shared parameter framework, a common latent random effect is applied to link the longitudinal model and time to event model and take the dependency between two types of outcomes comes from sampling on the same subject into account. We proposed a two-step estimation

procedure with an extended estimation equation to perform parameter estimation, and the numerical studies indicated that it works well in practice.

Table 3.1: Estimation of regression parameters with $m_0(t) = -0.5t + 0.5$ and 10% censoring rate and $n = 500$ and 500 replications.

Par.	True	MLE			
		Bias	SE	ESE	CP
β_0	-0.5	-0.001	0.077	0.072	93.2
β_1	0.5	0.003	0.064	0.066	96.4
β_2	-0.5	0.006	0.113	0.113	94.6
β_3	-0.2	-0.000	0.009	0.009	94.8
γ_1	-0.1	-0.018	0.048	0.049	95.8
γ_2	-0.1	-0.054	0.049	0.055	97.2
ψ	-0.1	0.098	0.045	0.043	94.0
σ_b^2	0.5	0.001	0.034	0.034	96.0
σ^2	0.25	0.000	0.033	0.035	95.4

Table 3.2: Estimation of regression parameters with $m_0(t) = -0.5t + 1$ and 20% censoring rate and $n = 500$ and 500 replications.

Par.	True	MLE			
		Bias	SE	ESE	CP
β_0	-0.5	-0.001	0.077	0.072	93.2
β_1	0.5	0.003	0.064	0.066	96.4
β_2	-0.5	0.006	0.113	0.113	94.6
β_3	-0.2	-0.000	0.009	0.009	94.8
γ_1	-0.1	-0.018	0.057	0.055	94.2
γ_2	-0.1	-0.052	0.058	0.052	91.8
ψ	-0.1	0.098	0.053	0.047	90.6
σ_b^2	0.5	0.001	0.034	0.034	96.0
σ^2	0.25	0.000	0.033	0.035	95.4

Table 3.3: Estimation of regression parameters with $m_0(t) = -0.5t + 1$ and 20% censoring rate and $n = 1200$ and 500 replications.

Par.	True	MLE			
		Bias	SE	ESE	CP
β_0	-0.5	-0.001	0.049	0.047	94.2
β_1	0.5	-0.003	0.041	0.044	96.6
β_2	-0.5	0.000	0.070	0.073	96.6
β_3	-0.2	0.000	0.006	0.006	95.6
γ_1	-0.1	-0.003	0.039	0.035	92.0
γ_2	-0.1	-0.005	0.057	0.050	92.2
ψ	-0.1	0.097	0.031	0.031	94.0
σ_b^2	0.5	-0.001	0.021	0.022	95.8
σ^2	0.25	0.001	0.021	0.022	96.4

Table 3.4: Analysis results of Alzhehelmer's Disease data

Longitudinal model	ADAS13		
	Estimate	SE	p-value
Intercept	0.227	0.037	0.000
t_{ij}	0.285	0.037	0.000
Age	−0.133	0.053	0.001
APOE$\mathcal{E}$4	0.129	0.039	0.000
CDRSB	0.062	0.043	0.143
DIGIRSCOR	−0.098	0.046	0.003
TRABSCOR	0.118	0.047	0.011
FAQ	0.056	0.046	0.224
Ventricles	0.020	0.063	0.749
Hippocampus	−0.116	0.065	0.073
WholeBrain	0.032	0.099	0.743
Entorhinal	−0.086	0.044	0.048
Fusiform	−0.076	0.049	0.121
Midtemp	−0.125	0.058	0.033
ICV	0.027	0.087	0.755
RAVLT.i	−0.240	0.054	0.000
RAVLT.l	−0.019	0.048	0.686
RAVLT.f	−0.031	0.076	0.683
RAVLT.perc.f	0.084	0.095	0.377
σ^2	0.083	0.021	0.000
σ_b^2	0.287	0.037	0.000
Proportional Mean residual Model			
	Estimate	SE	p-value
t_i	0.504	0.035	0.000
Age	0.001	0.031	0.484
APOE$\mathcal{E}$4	−0.008	0.028	0.000
CDRSB	0.022	0.032	0.238
DIGIRSCOR	0.010	0.028	0.357
TRABSCOR	0.028	0.029	0.172
FAQ	−0.019	0.027	0.246
Ventricles	−0.032	0.042	0.221
Hippocampus	0.415	0.040	0.154
WholeBrain	0.014	0.055	0.400
Entorhinal	−0.051	0.028	0.033
Fusiform	−0.032	0.034	0.175
Midtemp	0.010	0.033	0.383
ICV	−0.037	0.059	0.264
RAVLT.i	−0.022	0.042	0.299
RAVLT.l	−0.057	0.045	0.102
RAVLT.f	0.186	0.049	0.000
RAVLT.perc.f	−0.182	0.072	0.000
PTEDUCA	0.061	0.004	0.000
MMSE	−0.002	0.028	0.460
ADASQ4	0.022	0.033	0.252

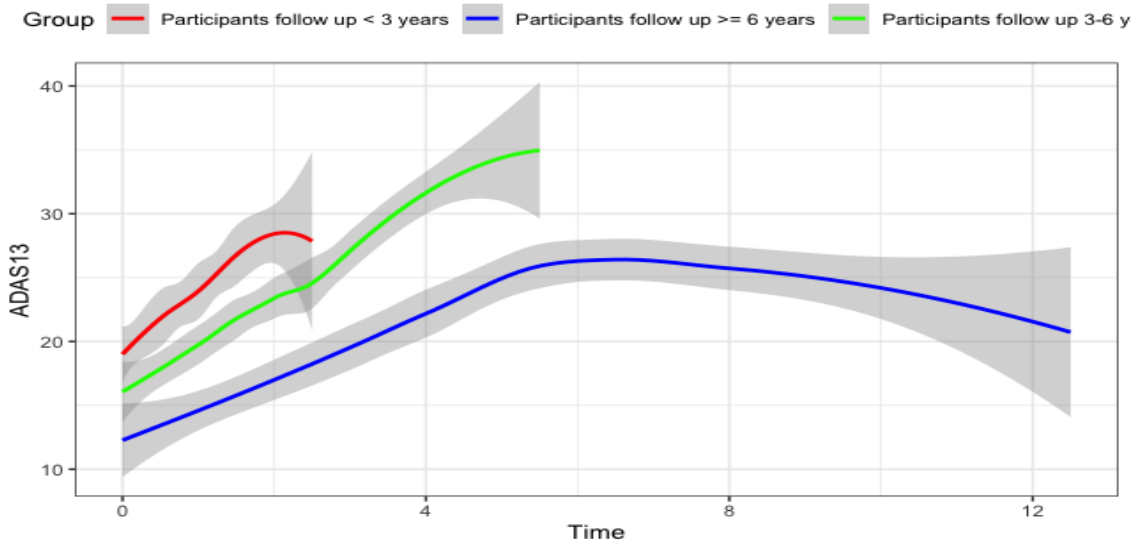

Figure 3.1: Smooth curve of ADAS-Cog 13 over time for MCI patient. Shaded regions are 95% pointwise confidence intervals.

Chapter 4

Joint Modeling of Binary Longitudinal and Survival Outcome under the Proportional Mean Residual Life Model

4.1 Introduction

As introduced in Section 1.3, joint analysis of longitudinal data and survival outcome is necessary to allow one to take the dependence of two types of outcomes within the same subject into account. Utilizing the mean residual life model for survival outcome in the joint modeling offers informative insights on the remaining life expectancy given the current status. In Chapter 3, we have discussed the simultaneous modeling of continuous longitudinal outcome and right-censored survival outcome under the proportional mean residual life model.

In this chapter, we will focus on the proportional mean residual model for survival

outcome and use the generalized linear mixed model to incorporate both categorical and continuous longitudinal data, although we mainly focus on the binary longitudinal data. Under the shared parameter framework, we assume that the random effects are from a multivariate normal distribution. The maximum likelihood approach using the Expectation-Maximization (EM) algorithm will provide estimators that are consistent and asymptotically follow normal distribution (Zeng and Cai, 2005). However, the EM algorithm requires integration that can be computationally intensive with high sample size and a larger number of observations on longitudinal outcomes. As a result, in the estimation procedure, we utilize the relationship between MRL function and hazard function to formulate the likelihood function and propose a quasi-likelihood approach to develop an efficient estimation procedure for parameter inference.

The rest of this chapter is organized as follows. We will begin in Section 4.2 by introducing the model and assumption that will be used throughout the paper. The likelihood for simultaneous modeling survival outcome and longitudinal outcomes will be derived in Section 4.3, and in particular, an iterative estimation procedure with the proposed quasi-likelihood estimation approach is developed. In Section 4.4, we will show some simulation results obtained from a simulation study. Section 4.5 will provide a real data application. Some discussion and concluding remarks will be given in Section 4.6.

4.2 Model Formulation and Notation

We denote $\boldsymbol{Y(t)}$ as the longitudinal process at time t. Suppose that $\boldsymbol{Y(t)}$ is from the exponential family to include both continuous and discrete variables. We use T to denote the survival time, and the survival time is subject to right-censoring. Suppose that a random sample of subjects with size n are followed during time $[0, \tau]$, where τ is the maximum follow-up time. Subject-specific random effects are denoted by $\boldsymbol{b}_i, i = 1, ..., n$, and we assume that

$\boldsymbol{b}_i's$ are mutually independent and identically distributed from $MVN(0, \boldsymbol{\Sigma}_b)$, and the vector length of $\boldsymbol{b}_i$ is d_b.

Under the exponential family distribution assumption, given the random effect $\boldsymbol{b}_i$, the longitudinal outcome $Y_i(t)$ at time t for subject i has the following density

$$\exp\left\{\frac{y_i\eta_i\left(t|\boldsymbol{b}_i\right) - B\left(\eta_i\left(t|\boldsymbol{b}_i\right)\right)}{A\left(D_i(t;\psi)\right)} + C\left(y_i, D_i(t;\psi)\right)\right\} \tag{4.1}$$

with

$$\eta_i\left(t|\boldsymbol{b}_i\right) = g\left(\mu_i\left(t|\boldsymbol{b}_i\right)\right) = \boldsymbol{X}_i(t)\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_i(t)\boldsymbol{b}_i.$$

In the above, $\mu_i\left(t|\boldsymbol{b}_i\right) = \mathrm{E}\left(Y_i(t)|\boldsymbol{b}_i\right) = B'\left(\eta_i\left(t|\boldsymbol{b}_i\right)\right)$, $v_i\left(t|\boldsymbol{b}_i\right) = \mathrm{Var}\left(Y_i(t)|\boldsymbol{b}_i\right) = B''\left(\eta_i\left(t|\boldsymbol{b}_i\right)\right)A\left(D_i(t;\psi)\right) = v\left(\mu_i\left(t|\boldsymbol{b}_i\right)\right)A\left(D_i(t;\psi)\right)$, and the functions $A(\cdot), B(\cdot), C(\cdot)$, and $D(\cdot)$ are known and ψ is the dispersion parameter. We use $g(\cdot)$ and $v(\cdot)$ to denote the known link and variance functions, respectively. The covariates $\boldsymbol{X}_i(t)$ and $\widetilde{\boldsymbol{X}}_i(t)$ are the observed covariates for subject i, $\boldsymbol{X}_i(t)$ represent the covariates that only affect longitudinal outcome, and $\widetilde{\boldsymbol{X}}_i(t)$ represent the covariates that affect both longitudinal and survival outcomes. The covariates $\boldsymbol{X}_i(t)$ and $\widetilde{\boldsymbol{X}}_i(t)$ can be completely different or share some components. The parameter $\boldsymbol{\beta}$ is the parameter for fixed effect, and $\boldsymbol{b}_i$ represents the within-subject random effect.

Given the random effect $\boldsymbol{b_i}$ and the observed covariates, the conditional mean residual life time function for the survival time T_i for subject i at time t is assumed to follow the proportional mean residual life model

$$m(t|\boldsymbol{b}_i, \mathbf{Z}(t)) = m_0(t)\exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i(t) + \tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\phi}\circ\boldsymbol{b}_i)\right). \tag{4.2}$$

The covariates $\boldsymbol{Z}_i(t)$ and $\widetilde{\boldsymbol{Z}}_i(t)$ denote the vector of observed covariates for subject i at time t, $\boldsymbol{Z}_i(t)$ represent the covariates that only affect survival outcome, and $\widetilde{\boldsymbol{Z}}_i(t)$ represent the

covariates that affect both longitudinal and survival outcomes. The covariates $\boldsymbol{Z}_i(t)$ and $\widetilde{\boldsymbol{Z}}_i(t)$ can be completely different or share some components. The function $m_0(t)$ denotes the baseline mean residual life function, $\boldsymbol{\gamma}$ is the vector of parameter for the fixed effect, $\boldsymbol{\phi}$ is the vector of parameter for the random effect, and $\boldsymbol{\phi} \circ \boldsymbol{b}_i$ denotes the component-wise product.

Under model (4.1) and model (4.2), we assume that the longitudinal outcome $Y_i(t)$ and survival time T_i are independent given the random effect $\mathbf{b}_i$ and the observed covariates. The parameter $\boldsymbol{\phi}$ represents the dependence level between longitudinal outcome and survival time linked by the latent random variable $\mathbf{b}_i$. If $\boldsymbol{\phi}$ is significantly not equal to 0, it means that the longitudinal outcome and survival time are correlated by the latent random variable. If not, the dependence between the longitudinal outcome and survival time is not caused by the random latent variable.

4.3 Quasi Likelihood Estimation

Suppose that for each subject i, we have n_i repeated measurements for the longitudinal outcome. We assume that the observation time is non-informative and the number of repeated measurements is bounded. For the missingness of $Y_i(t)$, we assume it to be non-informative as well. The observed longitudinal data for n subjects is $\left(n_i, Y_{ij}, \boldsymbol{X}_{ij}, \widetilde{\boldsymbol{X}}_{ij}\right), i = 1, \ldots, n, j = 1, \ldots, n_i$. The observed survival data for n subject is $\left(V_i, \Delta_i, \left\{\left(\boldsymbol{Z}_i(t), \widetilde{\boldsymbol{Z}}_i(t)\right) : t \leq V_i\right\}\right), i = 1, \ldots, n$, where $V_i = \min\left(T_i, C_i\right)$, $\Delta_i = I\left(T_i \leq C_i\right)$. We denote C_i as the censoring time for subject i and we assume that the censoring time is independent of T_i and $Y_i(t)$ given the random effect and observed covariates.

The goal of the joint modeling is to estimate and make inference on the parameters $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^T, \boldsymbol{\phi}^T, \text{Vech}\left(\boldsymbol{\Sigma}_b\right)^T \boldsymbol{\psi}^T, \boldsymbol{\gamma}^T\right)^T$. For the mean residual life model, we also need to estimate the baseline MRL function $m_0(t)$. The Vech$(\cdot)$ operator creates a column vector

containing the unknown parameters in the matrix.

For n subjects, we have longitudinal outcomes $\boldsymbol{Y} = \left(\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_n^T\right)^T$, where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$, survival time observations, $\boldsymbol{V} = (V_1, \ldots, V_n)^T$, and the subject-specific random effects $\boldsymbol{b} = \left(\boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_n^T\right)^T$. Utilizing the relationship between MRL function and hazard function and the relationship between MRL function and survival function from Oakes and Dasu (1990), we have

$$\lambda\left(t|\boldsymbol{Z}_i(t), \widetilde{\boldsymbol{Z}}_i(t), \boldsymbol{b}_i\right) = \frac{1}{m_0(t)}\left(\frac{\mathrm{d}m_0(t)}{\mathrm{d}t} + \exp\left(-(\boldsymbol{\gamma}'\mathbf{Z}_i(t) + \tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\phi} \circ \boldsymbol{b}_i))\right)\right) \tag{4.3}$$

and

$$S\left(t|\boldsymbol{Z}_i(t), \widetilde{\boldsymbol{Z}}_i(t), \boldsymbol{b}_i\right) = \frac{m_0(0)}{m_0(t)} \exp\left\{-\int_0^t \frac{\exp\left(-(\boldsymbol{\gamma}'\mathbf{Z}_i(\omega) + \tilde{\boldsymbol{Z}}_i(\omega)(\boldsymbol{\phi} \circ \boldsymbol{b}_i))\right)\mathrm{d}\omega}{m_0(\omega)}\right\} \tag{4.4}$$

The full likelihood function of observed data $(\boldsymbol{Y}, \boldsymbol{V})$ has the form

$$L_f(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V}) = \int_{\boldsymbol{b}} L_c(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b}) d\boldsymbol{b},$$

where the complete likelihood function $L_c(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b})$ for longitudinal data, survival data and unobserved random effect can be expressed as

$$L_c(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{b}) = \prod_{i=1}^{n} \left[f\left(\boldsymbol{Y}_i, V_i|\boldsymbol{b}_i\right) f\left(\boldsymbol{b}_i\right)\right] = \prod_{i=1}^{n} f\left(\boldsymbol{Y}_i|\boldsymbol{b}_i\right) f\left(V_i|\boldsymbol{b}_i\right) f\left(\boldsymbol{b}_i\right)$$

$$= \prod_{i=1}^{n} f\left(\boldsymbol{Y}_i|\boldsymbol{b}_i\right)\left(\left[\lambda\left(V_i|\boldsymbol{Z}_i(t), \widetilde{\boldsymbol{Z}}_i(t), b_i\right)^{\Delta_i} S(V_i|\boldsymbol{Z}_i(t), \widetilde{\boldsymbol{Z}}_i(t), b_i)\right]\right) f\left(\boldsymbol{b}_i\right)$$

$$= \prod_{i=1}^{n} \exp\left\{ \sum_{j=1}^{n_i} \left[\frac{Y_{ij}\left(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \tilde{\boldsymbol{X}}_{ij}(t)\boldsymbol{b}_i \right) - B_{ij}\left(\boldsymbol{\beta}; \boldsymbol{b}_i\right)}{A\left(D_i\left(t_j;\phi\right)\right)} + C\left(Y_{ij}; D_i\left(t_j;\phi\right)\right) \right] \right\}$$

$$\times \frac{m_0(0)}{m_0(V_i)^{1+\Delta_i}} \left(\frac{\mathrm{d}m_0(t)}{\mathrm{d}t}|_{t=V_i} + \exp\left(-(\boldsymbol{\gamma}'\mathbf{Z}_i(V_i) + \tilde{\boldsymbol{Z}}_i(V_i)(\boldsymbol{\phi} \circ \boldsymbol{b}_i)) \right) \right)^{\Delta_i}$$

$$\times \exp\left\{ - \int_0^{V_i} \frac{\exp\left(-(\boldsymbol{\gamma}'\mathbf{Z}_i(\omega) + \tilde{\boldsymbol{Z}}_i(\omega)(\boldsymbol{\phi} \circ \boldsymbol{b}_i)) \right) \mathrm{d}\omega}{m_0(\omega)} \right\}$$

$$\times (2\pi)^{-d_b/2} \left|\boldsymbol{\Sigma}_b\right|^{-1/2} \exp\left\{ -\frac{1}{2}\boldsymbol{b}_i^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{b}_i \right\}.$$

With the double integration, directly maximizing the full likelihood function can be challenging. When we assume that the baseline MRL function $m_0(t)$ is known, the estimate of regression parameters $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ can be obtained through an EM algorithm by considering random effect $\boldsymbol{b}$ as missing data. However, the integration in the E-step will cause intensive computation burden with a large sample size n, longitudinal data observation times n_i, and the number of parameters. With the iterations, the integration will potentially slow down the EM algorithm convergence. To estimate $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\phi}$ with less computational burden, following Choi and Zeng (2017), we propose to utilize first-order Laplace approximation to approximate the full likelihood in order to avoid integration over the random effect.

To derive the approximated likelihood function, we rewrite the full likelihood as

$$L_f(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V}) = (2\pi)^{-nd_b/2} \left|\boldsymbol{\Sigma}_b\right|^{-n/2} \int_{\boldsymbol{b}} \exp\left\{ \sum_{i=1}^{n} \left[l_{i|b_i}\left(\boldsymbol{\theta}, m_0(t)\right) - \frac{1}{2}\boldsymbol{b}_i^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{b}_i \right] \right\} d\boldsymbol{b},$$

where the logarithm of the conditional joint density given an unobserved random effect $\boldsymbol{b_i}$

is

$$l_{i|b_i}\left(\boldsymbol{\theta}, m_0(t)\right) = \sum_{j=1}^{n_i}\left[\frac{Y_{ij}\left(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_{ij}(t)\boldsymbol{b}_i\right) - B_{ij}\left(\boldsymbol{\beta}; \boldsymbol{b}_i\right)}{A\left(D_i\left(t_j;\phi\right)\right)} + C\left(Y_{ij}; D_i\left(t_j;\phi\right)\right)\right] +$$

$$\log\{m_0(0)\} - (1+\Delta_i)\log\{m_0(V_i)\} + \Delta_i \log\{\frac{\mathrm{d}m_0(t)}{\mathrm{d}t}|_{t=V_i} + \exp(-(\boldsymbol{\gamma}'\mathbf{Z}_i + \tilde{\boldsymbol{Z}}_i(\boldsymbol{\psi}\circ\boldsymbol{b}_i)))\}$$

$$-\int_0^{V_i}\frac{\exp\left(-(\boldsymbol{\gamma}'\mathbf{Z}_i + \tilde{\boldsymbol{Z}}_i(\boldsymbol{\psi}\circ\boldsymbol{b}_i))\right)\mathrm{d}\omega}{m_0(\omega)}.$$

Fowllowing Choi and Zeng (2017), we define

$$-\kappa(b) = \sum_{i=1}^{n}\left[l_{i|b_i}\left(\boldsymbol{\theta}, \Lambda_s\right) - \frac{1}{2}\boldsymbol{b}_i^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_i\right] = \sum_{i=1}^{n}\left[-\boldsymbol{\kappa}_i\left(\boldsymbol{b}_i\right)\right]$$

and apply Laplace's approximation as following,

$$-\kappa_i\left(\boldsymbol{b}_i\right) \approx -\boldsymbol{\kappa}_i\left(\tilde{\boldsymbol{b}}_i\right) - \frac{1}{2}\left(\boldsymbol{b} - \widetilde{\boldsymbol{b}}_i\right)^T\boldsymbol{\kappa}_i''\left(\tilde{\boldsymbol{b}}_i\right)\left(\boldsymbol{b}_i - \tilde{\boldsymbol{b}}_i\right)$$

where $\boldsymbol{\kappa}'$ and $\boldsymbol{\kappa}''$ denote the d_b vector and $d_b \times d_b$ dimensional matrix of first and second-order partial derivatives of $\boldsymbol{\kappa}$ with respect to $\boldsymbol{b}$ and $\tilde{\boldsymbol{b}}$.

Then the the full likelihood $L_f(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V})$ can be approximated by

$$L_p(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V})$$

$$= \prod_{i=1}^{n}\left[(2\pi)^{-d_b/2}\left|\boldsymbol{\Sigma}_b\right|^{-1/2}\int_{\boldsymbol{b}}\exp\left\{-\boldsymbol{\kappa}_i\left(\tilde{\boldsymbol{b}}_i\right) - \frac{1}{2}\left(\boldsymbol{b}_i - \tilde{\boldsymbol{b}}_i\right)^T\boldsymbol{\kappa}_i''\left(\tilde{\boldsymbol{b}}_i\right)\left(\boldsymbol{b}_i - \tilde{\boldsymbol{b}}_i\right)\right\}d\boldsymbol{b}\right]$$

$$= \left|\boldsymbol{\Sigma}_b\right|^{-n/2}\exp\left\{\sum_{i=1}^{n}\left[-\boldsymbol{\kappa}_i\left(\tilde{\boldsymbol{b}}_i\right) - \frac{1}{2}\log\left|\boldsymbol{\kappa}_i''\left(\tilde{\boldsymbol{b}}_i\right)\right|\right]\right\}$$

$$= \exp\left\{\sum_{i=1}^{n}\left[-\frac{1}{2}\log\left|\boldsymbol{I}_{d_b} - \boldsymbol{\Sigma}_b \tilde{l}''_{i|b_i}\left(\boldsymbol{\theta}, m_0(t)\right)\right| + \tilde{l}_{i|b_i}\left(\boldsymbol{\theta}, m_0(t)\right) - \frac{1}{2}\tilde{\boldsymbol{b}}_i^T \boldsymbol{\Sigma}_b^{-1}\tilde{\boldsymbol{b}}_i\right]\right\}.$$

With known baseline mean residual function $m_0(t)$, we could maximize the approximated likelihood function $L_p(\cdot)$ derived above and apply the iterative procedure to get parameter estimation. However, $m_0(t)$ is usually unknown in most situation. It is still intractable to maximize the approximated likelihood with an unknown baseline mean residual function. Therefore, in the following, we propose a quasi-likelihood approach that is easy to compute.

To formulate the quasi-likelihood, we consider formulating the quasi-score using stochastic processes. Following Sun and Zhang (2009), a class of zero-mean stochastic processes can be established with the IPCW technique,

$$M_i(t) = \frac{\Delta_i I\left(V_i > t\right)}{G\left(V_i\right)}\left[\left(V_i - t\right) - m_0(t)\exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i(t) + \tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\phi}\circ\boldsymbol{b}_i)\right)\right], \tag{4.5}$$

where $G(\cdot)$ denotes the distribution function for censoring time C. We assume that the censoring time C is independent of failure time T given observed covariates and random effects. The censoring survival function $G(V_i)$ can be estimated based on $\{(V_i, 1-\Delta_i) : i = 1, \ldots, n, j = 1,\}$ by using Kaplan-Meier (K-M) estimator $\hat{G}(t) = \prod_{k:C_k \le t}\left(1 - \frac{d_k^c}{n_k^c}\right)$, where $d_k^c = \sum_i \left(1-\Delta_i\right) I\left(V_i = C_k\right), n_k^c = \sum_i I\left(V_i \ge C_k\right)$ and $0 < C_1 < C_2 < \ldots < C_D$ are distinct censoring times. After plugging in the K-M estimators to equation (4.5), we obtain an estimating equation

$$\frac{\Delta_i I\left(V_i > t\right)}{\hat{G}\left(V_i\right)}\left[\left(V_i - t\right) - m_0(t)\exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i(t) + \tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\phi}\circ\boldsymbol{b}_i)\right)\right] = 0, \tag{4.6}$$

where $0 \le t \le \tau$, where $0 \le \tau = \inf\{t : P(T \ge t) = 0\} < \infty$.

Given $\boldsymbol{\gamma}, \boldsymbol{\phi}$, and $\boldsymbol{b}$, (4.6) becomes an estimating equation for $m_0(t)$. We can show that with this estimating equation (4.6), we can derive the closed form solution for baseline mean

residual life function as

$$\hat{m}_0(t;\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{b}) = \frac{\sum_i \frac{\Delta_i I(V_i>t)}{\hat{G}(V_i)}\,(V_i - t)}{\sum_i \frac{\Delta_i I(V_i>t)}{\hat{G}(V_i)} \exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i(t) + \tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\phi}\circ\boldsymbol{b}_i)\right)}, \quad 0 \le t \le \tau.$$

Following Sun and Zhang (2009), we can construct the quasi-score for $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ given random effect $\boldsymbol{b}$ by plugging in $\hat{m}_0(t;\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{b})$ and $\hat{G}(t)$ into stochastic processes (4.5). The quasi score can be written as

$$\boldsymbol{D}(\boldsymbol{\gamma},\boldsymbol{\phi} \mid \boldsymbol{b}) = \sum_i \int_0^\tau \frac{\Delta_i I\left(V_i > t\right) Z_i}{\hat{G}\left(V_i\right)} \left[(V_i - t) - \hat{m}_0(t) \exp\left(\boldsymbol{\gamma}'\mathbf{Z}_i + \tilde{\boldsymbol{Z}}_i(\boldsymbol{\phi}\circ\boldsymbol{b}_i)\right)\right] \mathrm{d}\mathcal{H}(t). \tag{4.7}$$

Here, $\mathcal{H}(t)$ is an increasing weight function and can be specified in different forms. We set $\mathcal{H}(t)$ to be the observed counting process $N(t) = \sum_{ij} I\left(Y_{ij} \le t\right)\delta_{ij}$ due to its good performance as shown in the simulation study by Sun and Zhang (2009).

With quasi-score in equation (4.7), we derive the following quasi conditional log-likelihood function $l_q(\cdot)$ given below to approximate the true conditional log-likelihood function $l_{i|b_i}(\cdot)$ in $L_p(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V})$. The quasi conditional log-likelihood $l_q(\boldsymbol{\theta} \mid \boldsymbol{b})$ is constructed with the quasi score function in (4.7), and it is defined as

$$l_{i|\boldsymbol{b}_i}(\boldsymbol{\theta}, m_0(t)) \approx l_q(\boldsymbol{\theta} \mid \boldsymbol{b}) = \sum_{i=1}^{n} \ell_{iq} = \sum_{i=1}^{n} (\ell_{i1} + \ell_{i2q}).$$

Here,

$$\ell_{i1} = \sum_{j=1}^{n_i} \left[\frac{Y_{ij}\left(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \widetilde{\boldsymbol{X}}_{ij}(t)\boldsymbol{b}_i\right) - B_{ij}\left(\boldsymbol{\beta};\boldsymbol{b}_i\right)}{A\left(D_i\left(t_j;\psi\right)\right)} + C\left(Y_{ij}; D_i\left(t_j;\psi\right)\right)\right]$$

and

$$\ell_{i2q} = \sum_i \int_0^\tau \frac{\Delta_i I\left(V_i > t\right)}{G\left(V_i\right)} \left(V_i - t\right) \left[\eta_i - \log \sum_{ij} \frac{\Delta_i I\left(V_i > t\right) \exp\left(\eta_i\right)}{G\left(V_i\right)}\right] \mathrm{d}N(t)$$

is the conditional quasi log-likelihood function for $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ given $\boldsymbol{b}$ corresponding to $\boldsymbol{D}(\boldsymbol{\gamma}, \boldsymbol{\phi} \mid \boldsymbol{b})$. A noteworthy feature of the proposed quasi likelihood is that it does not depend on unknown $m_0(t)$, leading to an efficient quasi likelihood inference approach. With the derived quasi conditional log-likelihood $\ell_q(\boldsymbol{\theta}|b)$ and first order Laplace approximation, we can get quasi-likelihood $L_{pq}(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V})$ defined below as a further approximation for $L_f(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V})$. We define $L_{pq}(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V})$ as

$$L_f(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V}) \approx L_{pq}(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V})$$

$$= \prod_{i=1}^{n} \left[(2\pi)^{-d_b/2} \left|\boldsymbol{\Sigma}_b\right|^{-1/2} \int_{\boldsymbol{b}} \exp\left\{ -\boldsymbol{\kappa}_i\left(\tilde{\boldsymbol{b}}_i\right) - \frac{1}{2}\left(\boldsymbol{b}_i - \tilde{\boldsymbol{b}}_i\right)^T \boldsymbol{\kappa}_i''\left(\tilde{\boldsymbol{b}}_i\right)\left(\boldsymbol{b}_i - \tilde{\boldsymbol{b}}_i\right) \right\} d\boldsymbol{b} \right]$$

$$= \left|\boldsymbol{\Sigma}_b\right|^{-n/2} \exp\left\{ \sum_{i=1}^{n} \left[-\boldsymbol{\kappa}_i\left(\tilde{\boldsymbol{b}}_i\right) - \frac{1}{2}\log\left|\boldsymbol{\kappa}_i''\left(\tilde{\boldsymbol{b}}_i\right)\right| \right] \right\}$$

$$= \exp\left\{ \sum_{i=1}^{n} \left[-\frac{1}{2}\log\left| \boldsymbol{I}_{d_b} - \boldsymbol{\Sigma}_b \tilde{l}_{iq}''\left(\boldsymbol{\theta}, m_0(t)\right)\right| + \tilde{l}_{iq}\left(\boldsymbol{\theta}, m_0(t)\right) - \frac{1}{2}\tilde{\boldsymbol{b}}_i^T \boldsymbol{\Sigma}_b^{-1} \tilde{\boldsymbol{b}}_i \right] \right\}.$$

We utilize one-step Newton-Raphson method to obtain estimation for $\boldsymbol{\theta}$ and $\boldsymbol{b}$. The procedure is a three-step iteration until the pre-specified parameter convergence is achieved.

- Step 1: At the k-th interation, conduct one-step Newton-Raphson iteration to obtain the solution $\widetilde{\boldsymbol{b}}$ of $\boldsymbol{\kappa}'(\boldsymbol{b}) = 0$. The $(k+1)$ -th estimate is $\tilde{\boldsymbol{b}}^{(k+1)} = \widetilde{\boldsymbol{b}}^{(k)} - \left[\boldsymbol{\kappa}''\left(\tilde{\boldsymbol{b}}^{(k)}\right)\right]^{-1}\left[\boldsymbol{\kappa}'\left(\tilde{\boldsymbol{b}}^{(k)}\right)\right]^T$, where $\tilde{\boldsymbol{b}}^{(k)} = \tilde{\boldsymbol{b}}^{(k)}\left(\widehat{\boldsymbol{\theta}}^{(k-1)}\right)$,
  $\boldsymbol{\kappa}'(\boldsymbol{b}) = \left(\boldsymbol{\kappa}_1'\left(\boldsymbol{b}_1\right)^T, \ldots, \boldsymbol{\kappa}_n'\left(\boldsymbol{b}_n\right)^T\right)^T$ and $\boldsymbol{\kappa}''(\boldsymbol{b}) = \left(\boldsymbol{\kappa}_1''\left(\boldsymbol{b}_1\right)^T, \ldots, \boldsymbol{\kappa}_n''\left(\boldsymbol{b}_n\right)^T\right)^T$. $\boldsymbol{\kappa}_n'$ and $\boldsymbol{\kappa}_n''$ can be calculated by following equations

$$\boldsymbol{\kappa}_i'\left(\widetilde{\boldsymbol{b}}_i\right) = -\widetilde{l}_{iq}'\left(\boldsymbol{\theta}\right) + \boldsymbol{\Sigma}_b^{-1}\widetilde{\boldsymbol{b}}_i,$$

$$\boldsymbol{\kappa}_i''\left(\widetilde{\boldsymbol{b}}_i\right) = -\widetilde{l}_{ip}''\left(\boldsymbol{\theta}\right) + \boldsymbol{\Sigma}_b^{-1},$$

where $\widetilde{l}_{ip}'\left(\boldsymbol{\theta}\right)$ and $\widetilde{l}_{ip}''\left(\boldsymbol{\theta}\right)$ are the first and second derivatives of conditional quasi log

likelihood function $l_{iq}(\boldsymbol{\theta})$ with respect to $\boldsymbol{b}_i$ evaluated at $\widetilde{\boldsymbol{b}}_i$.

- Step 2: The $(k+1)$-estimate of $\boldsymbol{\theta}$ can be obtained by one step Newton-Rapshon method by

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} - \left[S_q'\left(\widehat{\boldsymbol{\theta}}^{(k)T}\right)\right]^{-1}\left[S_q\left(\widehat{\boldsymbol{\theta}}^{(k)T}\right)\right]^{T},$$

  where $S_q\left(\widehat{\boldsymbol{\theta}}^{(k)T}\right)$ and $S_q'\left(\widehat{\boldsymbol{\theta}}^{(k)T}\right)$ are the first and second derivatives of $L_{pq}(\boldsymbol{\theta}, m_0(t); \boldsymbol{Y}, \boldsymbol{V})$ with respect to $\boldsymbol{\theta}$.

- Step 3: With $\theta^{(k+1)}$ and $b^{(k+1)}$, we can update $m_0(t)$ by

$$\hat{m}_0(t; \boldsymbol{\gamma}^{(k+1)}, \boldsymbol{\phi}^{(k+1)}, \boldsymbol{b}^{(k+1)}) = \frac{\sum_i \frac{\Delta_i I(V_i > t)}{\hat{G}(V_i)}(V_i - t)}{\sum_i \frac{\Delta_i I(V_i > t)}{\widehat{G}(V_i)} \exp\left(\boldsymbol{\gamma}'^{(k+1)}\mathbf{Z}_i(t) + \tilde{\boldsymbol{Z}}_i(t)(\boldsymbol{\phi}^{(k+1)} \circ \boldsymbol{b}_i^{(k+1)})\right)}.$$

For variance estimation of the parameters $\hat{\boldsymbol{\theta}}$, we propose to use the nonparametric bootstrap method. Specifically, we sample with replacement from the observed data set for Q times, where Q is an integer to denote the number of bootstrap samples. For each bootstrap sample, we can get an estimation of $\boldsymbol{\theta}^{(j)}, j = 1, ..., Q$ with the proposed quasi-likelihood estimation procedure. Then, the estimate for the variance of $\hat{\boldsymbol{\theta}}$ can be estimated by the empirical variance of the bootstrap sample $\boldsymbol{\theta}^{(j)}, j = 1, ..., Q$.

4.4 A Simulation Study

In this section, we present some results obtained from a simulation study that is conducted to evaluate the finite sample performance of the proposed quasi-likelihood estimation pro-

cedure. In the simulation study, we assume that Y_{ij} is a binary outcome following

$$P\left(Y_{ij} = y_{ij} \mid b_i\right) = \exp\left\{y_{ij}\xi_{ij} - \log\left(1 + \exp\left\{\xi_{ij}\right\}\right)\right\}, \quad y_{ij} = 0, 1$$

with $\xi_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + b_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3ij} + b_i$ for $j = 1, \ldots, n_i$ and

$$m\left(t \mid b_i\right) = m_0(t) \exp\left\{\psi b_i * Z_{3i} + \gamma_1 Z_{1i} + \gamma_2 Z_{2i}\right\}.$$

In the above, $b_i \sim N\left(0, \sigma_b^2\right), X_{1i} \equiv Z_{1i}$ are generated from a Bernoulli(0.5), and $X_{2i} \equiv Z_{2i}$ are simulated from the Unif$(0, 1)$. They are included in both hazard and longitudinal models. The covariate Z_{i3} is generated from the standard normal distribution, and it is included only in the mean residual model. The covariate X_{3ij} denotes the time at measurement, and it is included only in the longitudinal model. Following Choi and Zeng (2017), we generate the measurement time points as every unit of time over the follow-up ranging 0 through 2.4 and simulate four different units which are 0.3, 0.1, 0.05 and 0.03 producing the numbers of longitudinal observations (ni) per subject to be 4, 8,15 and 25, respectively.

Tables $4.1 - 4.3$ show the results on estimation of the regression parameters β_0, β_1, β_2, β_3, γ_1, γ_2, ϕ, and σ_b^2 with the baseline mean residual function $m_0(t) = -0.5t + 1$ and 20% censoring rate. In Tables $4.1 - 4.3$, we calculate the estimated bias given by the average of the estimates over 500 replication minus the true value (Bias), the sample standard deviation (SE), the average of estimated standard error (ESE), and the 95% empirical coverage probability (CP). In Table 4.1, we have assumed that the number of longitudinal observation time for all the subjects is fixed at 20, which may not be true. As the result, in Table 4.2 and Table 4.3, the observation times for each subject were generated uniformly from 0 to 2.4 with a random number of observation times generated uniformly from $15 - 20$ and $10 - 20$. From Tables $4.1 - 4.3$, we can see that the bias of the proposed quasi likelihood is small and the bootstrapping variance estimation performs well for this setup. The coverage

probabilities indicate the normal approximation to the parameter distribution seems to be reasonable.

4.5 Real Data Example

To illustrate the methodology, we applied the proposed method to the data from the first phase of Alzhehelmer's Disease Neuroimaging Initiative (http://adni.loni.usc.edu/).

The participants were followed up for 3.2 years before conversion to AD or censoring and checked periodically for their cognitive status. Among the 804 participants, 402 of them had mild cognitive impairment (MCI) as initial cognitive condition. MCI represents the intermediate stage in the AD progression, and the subjects with MCI have became an increasingly common target population for evaluating prognostic factors for AD conversion. In this real data study, we also target at the MCI population. By the end of the first phase study, 219 of the participants with MCI as initial cognitive condition had AD conversion, 183 participants stayed at MCI and had right censored time for AD conversion.

The goal of this study is to identify the predictors that will affect both the MCI subjects expected progression time to AD conversion and important disease progression biomarkers Functional Assessment Questionnaire (FAQ) and Mini-Mental State Examination (MMSE). Current literature mainly focuses on using biomarkers at baseline to predict the hazards of AD conversion for the subjects in the MCI population. However, the time expectancy for MCI patients staying in MCI status before transferring to AD conversion may be more of interest. Identification of the prognostic factors for the time expectancy is important for developing treatment that can prolong the time patients staying in MCI status before AD conversion. Besides, comparing to method that only utilize biomarkers information at baseline, incorporation of longitudinal data into the modeling will increase the accuracy and efficiency of parameter estimation.

As the result, in this real data analysis, we modeled the longitudinal measurements Functional Assessment Questionnaire (FAQ) and Mini-Mental State Examination (MMSE), respectively, with time-to-AD conversion under the proportional mean residual model while adjusting for age at baseline, the presence of apolipoprotein E (APOE) $\mathcal{E}4$ allele, gender and education level. The time of measurement was also included in the longitudinal model. The longitudinal measurements can be correlated with time-to-AD conversion by a random effect, where the random effect is assumed to be from a standard normal distribution.

In the analysis, the longitudinal measurements Functional Assessment and Mini-Mental State Examination are transferred into binary outcomes. The original FAQ score is based on some self-evaluating questions, such as, "keeping track of current events" and "preparing a balanced meal". The response score ranges from 0 ("Normal" or "Never did [the activity] but could do now") to 3 ("Dependent"). According to Teng et al. (2010), the Functional Activities Questionnaire (FAQ), a standardized assessment of instrumental ADLs (activities of daily living), delineates the clinical distinction between MCI and Alzheimer's with an optimal cut-point of 5. Thus, we dichotomized this FAQ score into 0 ("Low functional dependence", FAQ < 5) and 1 ("High functional dependence", FAQ ≥ 5). We then used this binary outcome in our analysis. The advantage of dichotomizing the measurement is to reduce possible measurement error. Similarly, following Yu et al. (2020), a cutoff score of Mini-Mental State Examination (MMSE) ranging $24 - 27$ can be used to detect cognitive dysfunction with correct rate around 90%. Following their conclusions, we dichotomized the MMSE score with cutoff score of 25. The MMSE score is defined as 0 ("high scores reflect less cognitive impairment" MMSE > 25) or 1 ("lower scores reflect severer cognitive impairment" MMSE ≤ 25).

In the Table 4.4, we show the estimates, the estimated standard errors and p-value of the maximum quasi-likelihood estimation (MQLE). For longitudinal outcome FAQ, among all the covariates, the result suggests that the presence of apolipoprotain E(AOPE) $\mathcal{E}4$

allele and time has significant effects on higher FAQ score or high functional dependence. Under joint modeling of FAQ score and survival outcome, the FAQ score is negatively correlated with the expectancy of MCI participants staying in MCI status, which means higher functional dependence will shorten the expected time for MCI patient progression to AD conversion. For joint modeling of MMSE and survival outcome, the intercept, baseline age, the presence of apolipoprotain E(AOPE) $\mathcal{E}4$ allele, female, and time of measurement have significant effect on lower scores or severer cognitive impairment. In addition, the lower the MMSE score, higher baseline age, the presence of apolipoprotain E(AOPE) $\mathcal{E}4$ allele and less number of education years have significant negative influence on the expectancy time of participants staying in MCI status.

4.6 Conclusion

In this chapter, we have developed a joint regression model for binary longitudinal and survival outcomes, where the generalized linear mixed model and proportional mean residual model were utilized to model the longitudinal outcomes and time to event data. In the method, an iterative estimation procedure with proposed quasi-likelihood approach was developed. The numerical studies showed that the proposed method works well for practical situations.

Table 4.1: Estimation of regression parameters with $m_0(t) = -0.5t + 1$ and 20% censoring rate and $n = 800$ and 500 replications, n_i=20.

Par.	True	PQMLE			
		Bias	SE	ESE	CP
β_0	-0.25	-0.084	0.087	0.090	94.6
β_1	0.5	0.058	0.071	0.081	97.3
β_2	-0.5	-0.053	0.119	0.137	96.4
β_3	-0.2	-0.004	0.022	0.025	97.5
γ_1	-0.1	0.005	0.053	0.056	96.4
γ_2	-0.1	0.001	0.100	0.100	94.6
ϕ	-0.1	-0.071	0.035	0.031	92.4
σ_b^2	0.5	-0.010	0.028	0.025	91.7

Table 4.2: Estimation of regression parameters with $m_0(t) = -0.5t + 1$ and 20% censoring rate and $n = 800$ and 500 replications, n_i is randomly generated from 15 to 20.

Par.	True	PQMLE			
		Bias	SE	ESE	CP
β_0	-0.25	-0.083	0.087	0.093	95.8
β_1	0.5	0.061	0.069	0.081	97.8
β_2	-0.5	-0.055	0.130	0.132	94.8
β_3	-0.2	-0.005	0.026	0.026	95.8
γ_1	-0.1	0.003	0.056	0.062	97.8
γ_2	-0.1	0.001	0.103	0.111	96.4
ϕ	-0.1	-0.071	0.035	0.034	95.4
σ_b^2	0.5	-0.011	0.031	0.031	94.8

Table 4.3: Estimation of regression parameters with $m_0(t) = -0.5t + 1$ and 20% censoring rate and $n = 800$ and 500 replications, n_i is randomly generated from 10 to 20.

Par.	True	PQMLE			
		Bias	SE	ESE	CP
β_0	-0.25	-0.083	0.091	0.097	95.2
β_1	0.5	0.069	0.079	0.077	93.4
β_2	-0.5	-0.065	0.125	0.133	95.4
β_3	-0.2	-0.004	0.029	0.026	92.4
γ_1	-0.1	0.007	0.058	0.055	95.2
γ_2	-0.1	0.013	0.097	0.102	95.8
ϕ	-0.1	-0.067	0.034	0.038	98.2
σ_b^2	0.5	-0.012	0.031	0.025	90.0

Table 4.4: Analysis results of Alzhehelmer's Disease data

Longitudinal model	FAQ			MMSE		
	Estimate	SE	p-value	Estimate	SE	p-value
Intercept	1.096	0.771	0.155	−1.447	0.358	0.000
Age	0.009	0.494	0.985	0.556	0.264	0.035
Education	0.155	0.447	0.728	−0.340	0.256	0.182
APOE$\mathcal{E}$4	1.240	0.434	0.004	1.358	0.317	0.000
Gender	0.970	0.891	0.276	1.469	0.637	0.021
Time at measurement	1.127	0.383	0.003	1.237	0.260	0.000
Variance of random effect	4.475	0.815	0.000	3.700	0.429	0.000
Proportional Mean residual Model						
	Estimate	SE	p-value	Estimate	SE	p-value
Random effect coefficient	−0.010	0.001	0.000	−0.003	0.001	0.000
Age	−0.116	0.075	0.120	−0.158	0.073	0.030
Education	−0.122	0.115	0.289	0.358	0.115	0.002
APOE$\mathcal{E}$4	−0.112	0.080	0.159	−0.4565	0.080	0.000

Chapter 5

Future Research

5.1 Regression Analysis of Interval-Censored Data with Missing Covariates

In Chapter 2, we have focused on the proportional hazards model (2.1). It is apparent that the same type of missing data could happen to other commonly used regression models such as the additive hazards model or the linear transformation model. It would be useful to develop some estimation procedures similar to that proposed above for these latter models. In addition, we have assumed that covariates are time-independent and it is clear that sometimes there may exist time-dependent covariates. Thus it would be helpful to generalize the proposed approach to allow time-dependent covariates. Also in Chapter 2, it has been assumed that we observe case II interval-censored data and as pointed out by some authors, in practice, one may face a more general type of interval-censored data, case K interval-censored data (Sun, 2006; Wang et al., 2016). It is apparent that the method given above cannot be directly applied to this latter situation and in other words, some generalizations

of it are needed.

5.2 Joint Analysis of Longitudinal and Survival Outcome under Mean Residual Life Model

One limitation of the proportional mean residual life model is its monotonically non-decreasing assumption on $m(t|z) + t = m_0(t)exp(\beta^T Z) + t$, which may not always satisfy. As pointed out by Chen and Cheng (2006), although a monotonically non-decreasing $m_0(t)$ may mathematically fulfill the monotonically non-decreasing assumption on $m(t|z) + t$, it may not be applicable for some processes, for example, human aging. Thus, one future research direction is to extend the current proposed approach to additive residual life model or transformed mean residual life model to loose the non-decreasing restriction.

In Chapter 3 and Chapter 4, we have assumed that the failure time of interest was righted censored. However, in practice, one may observe current status data, or case II interval censored data. It would be useful to develop some generalized method to allow other types of censoring mechanism.

In addition, we have assumed that the latent variable and random error are from a multivariate normal distribution, which can be relaxed by some modification of the estimation procedure. A similar two-step estimation procedure could be proposed with an EM algorithm as the first step with no restriction on the distribution of random effects.

Appendix A

Appendix

A.1 Appendix I: E-step of the EM Algorithm for Continuous Covariates in Chapter 2

In the E-step of the EM algorithm developed in Section 3, we need to calculate the expectations $E(Z_i|\mathbf{O_i}, \theta^{(\mathbf{d})})$ and $E(W_i|\mathbf{O_i}, \theta^{(\mathbf{d})})$. As described there, when missing covariates are categorical, they are some summations and can be expressed in the closed form. However, for continuous covariates, this will not be the case and instead we have to deal with the integrals that do not have a closed form. More specifically, we have that

$$E(Z_i|\mathbf{O_i}, \theta^{(\mathbf{d})}) = \int_{\mathbf{X_{miss}}} \frac{\mathbf{\Lambda^{(d)}}(\mathbf{V_i})\exp(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i^{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{X_i^{miss}})\delta_{\mathbf{1i}}}{\mathbf{1} - \exp\{-\mathbf{\Lambda^{(d)}}(\mathbf{V_i})\exp(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i^{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{X_i^{miss}})\}}$$

$$\times p(\mathbf{X_i^{miss}}|\mathbf{O_i}, \theta^{(\mathbf{d})})\mathbf{dX_i^{miss}},$$

and

$$E(W_i|\mathbf{O_i},\theta^{(\mathbf{d})}) = \int_{\mathbf{X_i^{miss}}} \frac{\{\mathbf{\Lambda^{(d)}}(\mathbf{U_i}) - \mathbf{\Lambda^{(d)}}(\mathbf{V_i})\}\exp(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i^{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{X_i^{miss}})\delta_{\mathbf{2i}}}{\mathbf{1} - \exp[-\{\mathbf{\Lambda^{(d)}}(\mathbf{U_i}) - \mathbf{\Lambda^{(d)}}(\mathbf{V_i})\}\exp(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i^{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{X_i^{miss}})]}$$

$$\times p(\mathbf{X_i^{miss}}|\mathbf{O_i},\theta^{(\mathbf{d})})\mathbf{dX_i^{miss}}$$

by using the notation defined before.

To calculate the integrals above, by following Herring and Ibrahim (2001), one can employ the Monte-Carlo estimation approach, which draws the sample from

$$p_{ij} = P(\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})}) = \frac{\mathbf{f}(\mathbf{U_i},\mathbf{V_i},\delta_{\mathbf{1i}},\delta_{\mathbf{2i}},\delta_{\mathbf{3i}}|\mathbf{X_i^{obs}},\mathbf{X_i^{mis}})\mathbf{f}(\mathbf{X_i^{obs}},\mathbf{X_i^{mis}};\gamma^{(\mathbf{d})})}{\int_{\mathbf{X_i^{mis}}}\mathbf{f}(\mathbf{U_i},\mathbf{V_i},\delta_{\mathbf{1i}},\delta_{\mathbf{2i}},\delta_{\mathbf{3i}}|\mathbf{X_i^{obs}},\mathbf{X_i^{mis}})\mathbf{f}(\mathbf{X_i^{obs}},\mathbf{X_i^{mis}};\gamma^{(\mathbf{d})})}$$

$$\propto f(U_i,V_i,\delta_{1i},\delta_{2i},\delta_{3i}|\mathbf{X_i^{obs}},\mathbf{X_i^{mis}})f(\mathbf{X_i^{obs}},\mathbf{X_i^{mis}};\gamma^{(\mathbf{d})})\,.$$

Note that $f(U_i,V_i,\delta_{1i},\delta_{2i},\delta_{3i}|\mathbf{X_i^{obs}},\mathbf{X_i^{mis}})$ is log-concave (Gilks and Wild, 1992) and if $f(\mathbf{X_i^{obs}},\mathbf{X_i^{mis}};\gamma^{(\mathbf{d})})$ belongs to the exponential family, the logrithm of $P(\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})})$ is concave. It follows that one can use the Gibbs sampler (Gilks and Wild, 1992) and adaptive rejection algorithm (Gilks and Wild, 1992) to sample from $P(\mathbf{X_i^{mis}}|\mathbf{O_i},\theta^{(\mathbf{d})})$.

More specifically for the determination of $E(Z_i|\mathbf{O_i},\theta^{(\mathbf{d})})$, for each subject with missing covariate $\mathbf{X_i^{miss}}$, we first apply the Gibbs sampler and adaptive reject algorithm to draw the sample $s_{i,1},...,s_{i,n_i}$ of size n_i from $p(\mathbf{X_i^{miss}}|\mathbf{O_i},\theta^{(\mathbf{d})})$. Then the conditional expectation can be approximated by

$$E(Z_i|\mathbf{O_i},\theta^{(\mathbf{d})}) = \frac{\mathbf{1}}{\mathbf{n_i}}\sum_{\mathbf{k=1}}^{\mathbf{n_i}} \frac{\mathbf{\Lambda^{(d)}}(\mathbf{V_i})\exp(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i^{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{s_{i,k}})\delta_{\mathbf{1i}}}{\mathbf{1} - \exp\{-\mathbf{\Lambda^{(d)}}(\mathbf{V_i})\exp(\beta_{\mathbf{1}}^{(\mathbf{d})'}\mathbf{X_i^{obs}} + \beta_{\mathbf{2}}^{(\mathbf{d})'}\mathbf{s_{i,k}})\}}\,.$$

In comparison to the categorical covariate situation, the above operation can be regarded as replacing each x_i^{miss} by n_i sampled values with equal weight. It is apparent that $E(W_i|\mathbf{O_i},\theta^{(\mathbf{d})})$ can be calculated similarly.

A.2 Appendix II: Proofs of the Asymptotic Properties in Chapter 2

In this Appendix, we will sketch the proof for the consistency and asymptotic normality of the proposed estimators given in Theorem 1 by employing the empirical process theory and nonparametric techniques. Define $Pf = \int f(x)dP(x)$ and $P_n f = n^{-1}\sum_{i=1}^{n} f(X_i)$ for a function f, a probability function P and a sample $X_1, \ldots, X_n$. For the proof, we need the following regularity conditions.

(A1) Assume that $\Lambda(\tau_1) < \infty$, $\Lambda(\tau_2) < \infty$, and there exists a positive constant a such that $P(V - U > a) > 0$. Also the union of the supports of U and V is contained in the interval $[r_1, r_2]$ with $0 < r_1 < r_2 < +\infty$.

(A2) The function Λ_0 is continuously differentiable on $[r_1, r_2]$, and satisfies $M^{-1} < \Lambda_0(r_1) < \Lambda_0(r_2) < M$ for some positive constant M.

(A3) The set of covariates (X, Z) has bounded support.

(A4) The conditional distribution $f(\mathbf{X_i^{mis}}|\mathbf{X_i^{obs}};\gamma)$ is identifiable and has continuous seconder-order derivatives with respect to γ, and $-E_0[\partial^2/\partial\gamma^2)\log f(\mathbf{X_i^{mis}}|\mathbf{X_i^{obs}};\boldsymbol{\gamma_0})]$ is positive definite.

(A5) For any $(\theta, \boldsymbol{\Lambda})$ near $(\theta_{\mathbf{0}}, \boldsymbol{\Lambda_0})$, $P_0(\log L(\theta, \boldsymbol{\Lambda}) - \log L(\theta_{\mathbf{0}}, \boldsymbol{\Lambda_0}) \leq -K(||\theta - \theta_{\mathbf{0}}||^{\mathbf{2}} + ||\boldsymbol{\Lambda} - \boldsymbol{\Lambda_0}||^{\mathbf{2}})$ for a fixed constant $K > 0$.

First we will prove the consistency and for this, we will verify the conditions of Theorem 5.7 of Vaart (1998). Let $BV_\omega[r_1, r_2]$ denote the functions whose total variation in $[r_1, r_2]$ are bounded by a given constant. Then the class of functions

$$F_\omega = \left\{ \int_0^{U_k} \exp\{\beta^T X_i\} d\Lambda(s) : \Lambda \in BV_\omega[r_1, r_2] \right\}$$

is a convex hull of functions $\{I(U_k \geq s)\exp\{\beta^T X_i\}$ and thus it is a Donsker class. Further-

more,

$$\exp\left(-\int_0^{U_k}\exp\{\beta^T X_i\}d\Lambda(s)\right)-\exp\left(-\int_0^{U_{k+1}}\exp\{\beta^T X_i\}d\Lambda(s)\right)$$

is bounded away from zero. Therefore, $l(\theta,\hat{\alpha}|\mathbf{O})=\log\mathbf{L}(\theta,\hat{\alpha}|\mathbf{O})$ belongs to some Donsker class due to the preservation property of the Donsker class under Lipschitz-continuous transformations. Then we can conclude that

$\sup_{\theta\in\Theta_n}|P_n l(\theta,\hat{\alpha}|\mathbf{O})-P_n l(\theta_0,\hat{\alpha}|\mathbf{O})|$ converges in probability to 0 as $n\to 0$.

Now we verify that another condition of Theorem 5.7 of Vaart (1998) also holds. That is, for any $\varepsilon>0$, we have

$$\sup_{d(\theta,\theta_0)>\varepsilon} Pl(\theta,\hat{\alpha}|\mathbf{O})<Pl(\theta_0,\hat{\alpha}|\mathbf{O})\,.$$

Note that this condition is satisfied if we can prove the model is identifiable. According to condition (A4) and similar arguments to the proof of Theorem 2.1 of Chang and Wu (2007), we can show the identifiability of the model parameters. Now, by Theorem 5.7 of Vaart (1998), we have $d(\hat{\theta}_n,\theta_0)=o_p(1)$, which completes the proof of consistency.

Before proving the asymptotic normality, we will need to establish the convergence rate. For this, we will first define the covering number of the class $\mathcal{L}=\{l(\theta,\hat{\alpha}|\mathbf{O}):\theta\in\Theta\}$ and establish a needed lemma.

Lemma 1. Assume that Conditions (A1), (A3)-(A4) hold. Then the covering number of the class $\mathcal{L}=\{l(\theta,\hat{\alpha}|\mathbf{O}):\theta\in\Theta\}$ satisfies

$$N(\epsilon,\mathcal{L},L_2(P))=O(\epsilon^{-1}).$$

Proof of Lemma 1: The proof is similar to that of Zeng et al. (2016) and Hu et al. (2017) and thus omitted.

To establish the convergence rate, for any $\eta>0$, define the class $\mathcal{F}_\eta=\{l(\theta_{n0},\hat{\alpha}|\mathbf{O})-$

$l(\theta, \hat{\alpha}|\mathbf{O}) : \theta \in \Theta, d(\theta, \theta_{n0}) \leq \eta\}$ with $\theta_{n0} = (\beta_0, \Lambda_{n0})$. Following the calculation of Shen and Wong (1994), P.597, we can establish that $\log N_{[]}(\epsilon, \mathcal{F}_\eta, \| \ . \ \|_2) \leq CN\log(\eta/\epsilon)$ with $N = m + 1$, where $N_{[]}(\epsilon, \mathcal{F}_\eta, d)$ denotes the bracketing number (see the Definition 2.1.6 in Vaart and Wellner, 1996) with respect to the metric or semi-metric d of a function class $\mathcal{F}$. Moreover, some algebraic calculations lead to $\| \ l(\theta_{n0}, \hat{\alpha}|\mathbf{O}) - l(\theta, \hat{\alpha}|\mathbf{O}) \ \|_2^2 \leq C\eta^2$ for any $l(\theta_{n0}, \hat{\alpha}|\mathbf{O}) - l(\theta, \hat{\alpha}|\mathbf{O}) \in \mathcal{F}_\eta$. Therefore, by Lemma 3.4.2 of Vaart and Wellner (1996), we obtain

$$E_p \| \ n^{1/2}(P_n - P) \ \|_{\mathcal{F}_\eta} \leq CJ_\eta(\epsilon, \mathcal{F}_\eta, \| \ . \ \|_2)\{1 + \frac{J_\eta(\epsilon, \mathcal{F}_\eta, \| \ . \ \|_2)}{\eta^2 n^{1/2}}\}, \qquad (S)$$

where $J_\eta(\eta, \mathcal{F}_\eta, \| \ . \ \|_2) = \int_0^\eta \{log N_{[]}(\epsilon, \mathcal{F}_\eta, \| \ . \ \|_2)\}^{1/2} d\epsilon$. The right-hand side of (S) yields $\phi_n(\eta) = C\eta^{1/2}(1 + \frac{\eta^{1/2}}{\eta^2 n^{1/2}} M_1)$, where M_1 is a positive constant. Then $\phi_n(\eta)/\eta$ is a decreasing function, and $n^{2/3}\phi_n(-1/3) = O(n^{1/2})$. According the theorem 3.4.1 of Vaart and Wellner (1996), we can conclude that $d(\hat{\theta}, \theta_0) = O_p(n^{-1/3})$.

Now we prove the asymptotic normality of $\hat{\beta}_n$. Following the proof of Theorem 2 in Zeng et al. (2016), one can obtain that

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = (E[\{l_\beta - l_\Lambda(s^*)\}\{l_\beta - l_\Lambda(s^*)\}^T)^{-1} G_n\{l_\beta - l_\Lambda(s^*)\} + o_p(1),$$

where l_β is the score function for β, $l_\Lambda(s^*)$ is the score function along this submodel $d\Lambda_{\epsilon,s^*} = (1 + \epsilon s^*)d\Lambda$. This implies that the influence function for $\hat{\beta}_n$ is exactly the efficient influence function, so that $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges to a zero-mean normal random vector whose covariance matrix attains the semiparametric efficiency bound.

Bibliography

Albert, PS., and Follmann, DA. Random effects and latent processes approaches for analyzing binary longitudinal data with missingness: a comparison of approaches using opiate clinical trial data. *Statistical Methods in Medical Research* 2007; **16(5)**: 417-439.

Albert, PS., and Follmann, DA. Modeling repeated count data subject to informative dropout. *Biometrics* 2000; **56(3)**:667-677.

Albert, PS., Follmann, DA., Wang, SA., and Suh, EB. A latent autoregressive model for longitudinal binary data subject to informative missingness. *Biometrics* 2002; **58(3)**: 631-642.

Choi, J., Cai, J., Zeng, D., et al. Joint analysis of survival time and longitudinal categorical outcomes[J]. *Statistics in Biosciences* 2015; **7(1)**: 19-47.

Chakraborty, A., and Das, K. Inferences for joint modelling of repeated ordinal scores and time to event data[J]. *Computational and mathematical methods in medicine* 2010; **11(3)**: 281-295.

Choi, J., Cai, J., Zeng, D.. Penalized likelihood approach for simultaneous analysis of survival time and binary longitudinal outcome[J]. *Sankhya B* 2017; **79(2)**: 190-216.

Chen, YQ. Additive expectancy regression. *Journal of the American Statistical Association* 2007; **102(477)**:153-166.

Chen, YQ., and Cheng, S. Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika* 2005; **92(1)**:19-29.

Chen, YQ., and Cheng, S. Linear life expectancy regression with censored data. *Biometrika* 2006; **93(2)**: 303-313.

Chen, MH., Ibrahim, JG., Lipsitz SR. Bayesian methods for missing covariates in cure rate models[J]. *Lifetime Data Analysis* 2002; **8(2)**: 117-146.

Chen, K., Jin, Z., and Ying, Z. Semiparametric analysis of transformation models with censored data. *Biometrika* 2002; **89**: 659-668.

Chen, HY., and Little, RJA. Proportional hazards regression with missing covariates. *Journal of the American Statistical Association* 1999; **94(447)**: 896-908.

Chang, IS., Wen, CC., and Wu, YJ. A profile likelihood theory for thecorrelated gamma-frailty model with current status family data. *Statistica Sinica* 2007; **17**: 1023-1046.

Ding, J., and Wang, JL. Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics* 2008; **64(2)**, 546-556.

Efron, B. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika* 1981; **68(3)**: 589-599.

Elashoff, RM., Li, G., and Li, N. An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in Medicine* 2007; **26(14)**: 2813-2835.

Faucett, CL., Schenker, N., and Elashoff, RM. Analysis of censored survival data with intermittently observed time-dependent binary covariates[J]. *Journal of the American Statistical Association* 1998; **93(442)**: 427-437.

Gilks, WR., and Wild, P. Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 1992; **41**: 337-348.

He, H., Cai, J., and Song, X., and Sun, L. Analysis of proportional mean residual life model with latent variables. *Statistics in medicine* 2017; **36(5)**: 813-826.

Herring, AH., and Ibrahim JG. Likelihood-based methods for missing covariates in the Cox proportional hazards model. *Journal of the American Statistics Association* 2001; **96**: 292-302.

Hogan, JW., and Laird, NM. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in medicine* 1997; **16(3)**: 239-257.

Huang, X., Li, G., Elashoff, RM., and Pan, J. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime data analysis* 2011; **17(1)**: 80-100.

Huang, W., Zeger, SL., Anthony, JC., and Garrett, E. Latent variable model for joint analysis of multiple repeated measures and bivariate event times. *Journal of the American Statistical Association* 2001; **96(455)**: 906-914.

Hu, T., Zhou, Q., and Sun, J. Regression analysis of bivariate current status data under the proportional hazards model. *The Canadian Journal of Statistics* 2017; **45**: 410-424.

Ibrahim, JG., Lipsitz, SR., and Chen, MH. Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 1999; **61(1)**: 173-190.

Larsen, K. Joint analysis of time-to-event and multiple binary indicators of latent classes[J]. *Biometrics* 2004; **60(1)**: 85-92.

Li, S., Hu, T., Wang, P., et al. Regression analysis of current status data in the presence of dependent censoring with applications to tumorigenicity experiments. *Computational Statistics and Data Analysis* 2017; **110**: 75-86.

Lipsitz, SR., Ibrahim, JG., and Zhao, LP. A weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *Journal of the American Statistics Association* 1999; **94**: 1147-1160.

Li, K., and Luo, S. Functional joint model for longitudinal and time-to-event data: an application to Alzheimer's disease. *Statistics in medicine* 2017; **36(22)**: 3560-3572.

Lin, H., McCulloch, CE., and Mayne, ST. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Statistics in Medicine* 2002; **21(16)**: 2369-2382.

Little, RJA., and Rubin, AB. *Statistical Analysis with Missing data*. New York, NY: Johan Wiley & Sons, 2002.

Li, K., O'Brien, R., Lutz, M., et al. A prognostic model of Alzheimer's disease relying on multiple longitudinal measures and time-to-event data[J]. *Alzheimer's Dementia* 2018; **14(5)**: 644-651.

Li, S., Wu, Q., and Sun, J. Penalized estimation of semiparametric transformation models with interval-censored data and application to Alzheimer's disease. *Statistical Methods in Medical Research* 2019; **29**: 0962280219884720.

Ma, L., Hu, T., and Sun, JG. Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika* 2015; **102**: 731-738.

McMahan, CS., Wang, L., and Tebbs, JM. Regression analysis for current status data using the EM algorithm. *Statistics in medicine* 2013; **32**: 4452-4466.

Maguluri, G., and Zhang, CH. Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society: Series B (Methodological)* 1994; **56(3)**: 477-489.

Oakes, D., and Dasu, T. A note on residual life[J]. *Biometrika* 1990; **77(2)**: 409-410.

O'Bryant, SE., Humphreys, JD., Smith, GE., et al. Detecting dementia with the mini-mental state examination in highly educated individuals[J]. *Archives of neurology* 2008; **65(7)**: 963-967.

Parner, E. Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics* 1998; **26(1)**: 183-214.

Pan, D., He, H., Song, X., and Sun, L. Regression analysis of additive hazards model with latent variables. *Journal of the American Statistical Association* 2015; **110(511)**: 1148-1159.

Pulkstenis, EP., Ten Have, TR., and Landis, JR. Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication[J]. *Journal of the American Statistical Association* 1998; **93(442)**: 438-450.

Qi, L., Wang, CY., and Prentice, RL. Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistics Association* 2005; **100**: 1250-1263.

Ramsay, JO. Monotone regression splines in action. *Statistical science* 1988; **3(4)**: 425-441.

Roy, J., and Lin, X. Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* 2000; **56(4)**: 1047-1054.

Roy, J., and Lin, X. Analysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates: changes in methadone treatment practices. *Journal of the American Statistical Association* 2002; **97(457)**: 40-52.

Sun, JG. *The statistical Analysis of Interval-censored Failure Time Data*. New York, NY: Springer, 2006.

Sun, L., Song, X., and Zhang, Z. Mean residual life models with time-dependent coefficients under right censoring. *Biometrika* 2012; **99(1)**: 185-197.

Shen, X. and Wong, WH. *Convergence rate of sieve estimates. Annals of Statistics* 1994; **22**: 580-615.

Su, YR. and Wang, JL. Semiparametric efficient estimation for shared-frailty models with doubly-censored clustered data. *Annals of statistic* 2016; **44(3)**: 1298-1331.

Sun, L., and Zhang, Z. A class of transformed mean residual life models with censored survival data. *Journal of the American Statistical Association* 2009; **104(486)**: 803-815.

Teng, E., Becker, BW., Woo, E., et al. Utility of the Functional Activities Questionnaire for distinguishing mild cognitive impairment from very mild Alzheimer's disease[J]. *Alzheimer disease and associated disorders* 2010; **24(4)**: 348.

Van Der Vaart, AW. *Asymptotic Statistic*. Cambridge University Press, 1998.

Van Der Vaart, AW., and Wellner, JA. *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York, NY: Springer-Verlag, 1996.

Wen, CC., and Lin, CT. Analysis of current status data with missing covariates. *Biometrics* 2011; **67**: 760-769.

Wang, LM., McMahan, CS.and Hudgens, MG., et al. A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics* 2016; **72**: 222-231.

Xu, J., and Zeger, SL. Joint analysis of longitudinal data comprising repeated measures and times to events[J]. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2001; **50(3)**: 375-387.

Yu, RL., Lee, WJ., Li, JY., et al. Evaluating mild cognitive dysfunction in patients with Parkinson's disease in clinical practice in Taiwan[J]. *Scientific reports*, 2020; **10(1)**: 1-9.

Yi, F., Tang, N., and Sun, J. Simultaneous variable selection and estimation for joint models of longitudinal and failure time data with interval censoring. *Biometrics* 2020;.

Zhou, H., and Pepe, MS. Auxiliary covariate data in failure time regression. *Biometrik* 1995; **82(1)**: 139-149.

Zhao S., Hu T., Ma L., et al.Regression analysis of informative current status data with the additive hazards model. *Lifetime data analysis* 2015; **21**: 241-258.

Zeng, D., Mao, L., and Lin, DY. Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika* 2016; **103**: 253-271.

VITA

Ruiwen Zhou received her Bachelor of Science in 2014 from Southwest Jiaotong University, Chengdu, China. She received her Master of Art in 2016 in the Department of Statistics from the University of Missouri-Columbia. Then she joined the Ph.D. program in the Department of Statistics at the University of Missouri-Columbia in August 2016. She will start a postdoctoral position at Department of Statistics in Duke University in June 2021.