Joint analysis of a quantile of longitudinal outcomes and multiple time to events with censoring

by

©Xiaoming Lu

A Thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics and Statistics

Memorial University of Newfoundland

05/2020

St. John's Newfoundland

Abstract

It is very common in health science studies that we observe both longitudinal and survival data, within which different types of data are correlated and need to be analysed together to draw accurate conclusions. In this thesis, we propose a new method to jointly analyse observations of a longitudinal outcome and occurring times for multiple right- and interval-censored events to capture the underlying effects between them. In order to have a more complete view, we apply the quantile regression techniques to measure the effects of covariates on the longitudinal observations and then the effects of longitudinal observations on the occurring times of events at different levels of quantile. Semi-parametric proportional hazards models are proposed for both right- and interval-censored events with a vector of possible time-varying covariates shared with the quantile regression model for the longitudinal outcome. We also assume a variable of random effects in the survival models to measure the dependence between different events. We develop a Monte Carlo Expectation Maximization (MCEM) algorithm for computing non-parametric maximum likelihood estimators of parameters. Our estimators are proved to be consistent and asymptotically normally distributed. Furthermore, our proposed joint model is illustrated through a series of extensive simulation studies and an application to a data set from a French cohort study, *PAQUID*, aiming at studying the cognitive decline, such as the disease of dementia, among the elderly.

Acknowledgments

I would like to express my sincere appreciation to my supervisors, Dr. Zhaozhi Fan and Dr. Deping Ye, for their guidance through the learning and writing process of this thesis. This thesis would not have been done without the extensive support from Dr. Zhaozhi Fan, both professionally and personally, throughout the whole period of my PhD study. I am grateful to having Dr. Deping Ye as my co-supervisor who have taught me a great deal about academic writing, research habits, and pursuing career goals. Furthermore I would like to thank my cousin Yunqi Ji and sister-in-law Bingrui Sun for giving me constant encouragement and persistent help in academic study and daily life. I am also thankful to many faculty and staff members in the Department of Mathematics and Statistics at Memorial University, for introducing me to the statistical world and all the support along the journey. I would like to thank all my loved ones, my parents Yingchen Lv and Fengling Xu, Aunt Fengqiu Xu, and other relatives and friends who have supported me throughout these years of graduate study. I am grateful to all others who have helped.

Table of Contents

Abstract **ii**

Acknowledgments **iii**

Table of Contents **vi**

List of Tables **vii**

List of Figures **x**

1 Introduction **1**

2 Quantile regression and Survival analysis **5**

2.1 Quantile regression 6

2.1.1 Quantiles 6

2.1.2 Quantile regression for independent data 10

2.1.3 Quantile Regression for longitudinal data 13

2.2 Survival analysis 18

2.2.1 Basic quantities in survival analysis 19

2.2.2 Censoring and likelihoods 21

2.2.3 Cox's proportional hazards model for survival data 23

3 Proposed joint model of a quantile of longitudinal outcome and multiple-censored survival times **26**

3.1 Data and Model 27

3.2 The likelihood functions 29

3.2.1 Likelihood attributed to longitudinal observations 29

3.2.2 Likelihood attributed to asymptomatic events 30

3.2.3 Likelihood attributed to symptomatic events 31

3.2.4 Observed joint likelihood 32

3.3 Estimation 33

3.3.1 Monte Carlo E-step 37

3.3.2 M-step 39

3.4 Asymptotics 44

3.4.1 Consistency 46

3.4.2 Asymptotic normality 57

4 Numerical Study **65**

4.1 Simulation 65

4.1.1 Simulation setup 66

4.1.2 Simulation results 67

4.2 An application to dementia: what about the elderly? 81

4.2.1 Data 82

4.2.2 Joint model of IST, dementia and death 84

4.2.3 Joint model of IST, dementia, death and dependency 89

5 Discussion **93**

Bibliography **95**

Appendix **99**

List of Tables

4.1 Simulation Results 150 ald 68

4.2 Simulation Results 300 ald 71

4.3 Simulation Results 150 aldT 76

4.4 Simulation Results 300 aldT 79

4.5 Estimates for the joint model of IST scores, dementia time and death time 88

4.6 Estimates for the joint model of IST scores, dementia time, death time and dependency status 90

A.1 Simulation Results 150 Nor 99

A.2 Simulation Results 300 Nor 102

A.3 Simulation Results 150 NorT 105

A.4 Simulation Results 300 NorT 108

A.5 Estimates of parameter, standard error and confidence interval for the joint model of IST scores, dementia time and death time 123

A.6 Estimates of parameter, standard error and confidence interval for the joint model of IST scores, dementia time, death time and dependency status 127

List of Figures

2.1 Cumulative distribution function for standard normal variable 7

2.2 Cumulative distribution function and quantile function 8

2.3 Quantile Functions for Standard Normal and a Skewed Distribution . 9

4.1 Simulated baseline survival functions for n=150, $\tau = 0.25$ and an ALD error 74

4.2 Box plots of IST for low and high education levels 84

4.3 IST scores against ages for all subjects and a random sample of 10 subjects 85

4.4 Box plots of IST for dementia and no dementia 86

4.5 Kaplan-Meier curves for male and female subjects 87

4.6 Estimated baseline survival curve for dementia and death 91

4.7 Estimated baseline survival curve for dementia, dependency and death 92

A.1 Simulated baseline survival functions for n=150, $\tau = 0.50$ and an ALD error 111

A.2 Simulated baseline survival functions for n=150, $\tau = 0.85$ and an ALD error 111

A.3 Simulated baseline survival functions for n=300, $\tau = 0.25$ and an ALD error 112

A.4 Simulated baseline survival functions for n=300, $\tau = 0.50$ and an ALD error 112
A.5 Simulated baseline survival functions for n=300, $\tau = 0.85$ and an ALD error 113
A.6 Simulated semi-competing baseline survival functions for n=150, $\tau =$ 0.25 and an ALD error 113
A.7 Simulated semi-competing baseline survival functions for n=150, $\tau =$ 0.50 and an ALD error 114
A.8 Simulated semi-competing baseline survival functions for n=150, $\tau =$ 0.85 and an ALD error 114
A.9 Simulated semi-competing baseline survival functions for n=300, $\tau =$ 0.25 and an ALD error 115
A.10 Simulated semi-competing baseline survival functions for n=300, $\tau =$ 0.50 and an ALD error 115
A.11 Simulated semi-competing baseline survival functions for n=300, $\tau =$ 0.85 and an ALD error 116
A.12 Simulated baseline survival functions for n=150, $\tau = 0.25$ and an Normal error 116
A.13 Simulated baseline survival functions for n=150, $\tau = 0.50$ and an Normal error 117
A.14 Simulated baseline survival functions for n=150, $\tau = 0.85$ and an Normal error 117
A.15 Simulated baseline survival functions for n=300, $\tau = 0.25$ and an Normal error 118
A.16 Simulated baseline survival functions for n=300, $\tau = 0.50$ and an Normal error 118

A.17 Simulated baseline survival functions for n=300, $\tau = 0.85$ and an Normal error 119
A.18 Simulated semi-competing baseline survival functions for n=150, $\tau =$ 0.25 and an Normal error 119
A.19 Simulated semi-competing baseline survival functions for n=150, $\tau =$ 0.50 and an Normal error 120
A.20 Simulated semi-competing baseline survival functions for n=150, $\tau =$ 0.85 and an Normal error 120
A.21 Simulated semi-competing baseline survival functions for n=300, $\tau =$ 0.25 and an Normal error 121
A.22 Simulated semi-competing baseline survival functions for n=300, $\tau =$ 0.50 and an Normal error 121
A.23 Simulated semi-competing baseline survival functions for n=300, $\tau =$ 0.85 and an Normal error 122

Chapter 1

Introduction

In clinical or medicine studies, the data collected is often very complex which may include many types of data, such as longitudinal measurements and time-to-event data. It is a common objective to characterize the dependence between different types of data. For example, in a longitudinal study, observations may be lost due to the occurrence of some events that are associated with the response of interest. Without considering the dropout, the model may result in biased estimates for the longitudinal analysis. Also, in survival analysis, a longitudinal covariate may be observed over time and the trend can be used to predict the risk of an event or the risks of several events. For example, a steadily decline in scores of Isaacs Set Test (IST) is predictive of cognitive diseases and death among the elderly.

It is very common, in many application studies, that a subject is under risks of two or more types of events that could be symptomatic or asymptomatic, where the time to the occurrence of a symptomatic event is observed exactly or right-censored if the event does not occur till the end of the study, whereas the time to the development of an asymptomatic event can only be located between two time points (i.e., within an interval of time) by periodically visits or examinations. For example, in a dementia

study for the elderly, psychologists visit subjects every 2 or 3 years and conduct a battery of psychometric tests which helps to diagnose the disease. The exact time to the development of dementia can only be diagnosed between two visits, resulting in interval-censored times, while the death times are reported exactly for those who died or right-censored for those who still alive at the end of study. Moreover, different types of events may be highly correlated with each other and share some common risk factors. Failure to take into account for the correlation between events may lead to a biased effect of a common risk factor on the event of interest. For example, dementia is a chronic disease among the elderly where people are also under high risk of death. Moreover, some risk factors, such as sex and age, have effects on both events.

Another motivation of this thesis is that the mean or the expected value of the longitudinal response may not be the summary of interest. When the distribution of the longitudinal response is highly skewed or contains non-negligible outliers, the conditional median of the longitudinal response is more robust and preferred than the conditional mean. Also, in many clinical and epidemiological studies, researchers are more interested in the tails of a distribution and covariates may have different effects on different quantiles of the longitudinal response distribution. For example, in a study of the spread of sexually transmitted diseases, we may focus on the effect of predictors on people with a great number of sexual partners, since they are the main group of people spreading the disease. Moreover, the effects of a longitudinal outcome on the events of interest may be significant only when the value exceeds or below a threshold. Quantile regression has been extended to longitudinal analysis by Jung (1996) who firstly proposed a quasi-likelihood method in the median regression, Lipsitz et al. (1997) who extended Jung's work to a weighted GEE model, and Lu and Fan (2015) who proposed a weighted quantile regression model applying a general stationary auto-correlation structure for the covariance matrix. Also, different types

of mixed quantile regression models have been proposed to account for within- and/or between-subject correlations using random effects, see, among others, Koenker (2004), Geraci and Bottai (2007), Liu and Bottai (2009), and Geraci and Bottai (2014). The induced smoothing method (Brown and Wang, 2005) has been used to redefine smoothed objective functions in quantile regression by Fu and Wang (2012), Leng and Zhang (2012), and Lu and Fan (2015).

Many methods were proposed in the literature to account for informative dropouts in a longitudinal study by simultaneously modelling the longitudinal outcome and the time to dropout, see Little (1995), Tsiatis and Davidian (2004) and references therein. Farcomeni and Viviani (2015) proposed a joint model for a quantile of the longitudinal outcome and a right-censored time-to-event outcome to account for informative dropout. There are also many methods proposed to jointly analyse correlated right-censored events (Hougaard, 2012), correlated interval-censored events (Zeng et al., 2017), and correlated right- and interval-censored events (Gao et al., 2018). However, no existing literature has proposed a method to jointly modelling a quantile of the longitudinal outcome, right-censored time-to-event outcome, and interval-censored time-to-event outcome. The three types of data are frequently observed together in cohort studies and a joint model of these three types of data would allow us to evaluate the effects of predictors on longitudinal response and both symptomatic and asymptomatic events at different levels of quantile. Furthermore, given the fitted model, we can predict the occurrence of an symptomatic event based on the history of the other events and the location (quantile level) of the longitudinal observations.

In this thesis, we propose a joint model of a quantile of longitudinal outcome, and multiple right-censored and interval-censored time-to-event outcomes.

Our joint model contains two parts, the first part is a proposed quantile regression submodel for the longitudinal outcome which is assumed to follow an asymmetric

Laplace distribution (ALD). The second part consists of proposed semi-parametric proportional hazards models for the events of interest. The survival regression models share a common vector of time-dependent covariates with the longitudinal quantile regression model to measure the effects of longitudinal quantiles on the risks of events. The dependence between events are captured by a shared random effect and estimated by unknown coefficients. By adding a terminal event, our proposed joint models can handle semi-competing risks. We derive non-parametric maximum likelihood estimates by setting up a Monte Carlo expectation maximization algorithm. We show that the derived estimators are consistent and asymptotically normal. Finally, our proposed joint model for a quantile of longitudinal observations and multiple right-censored and interval-censored event times are illustrated through intensive simulation studies and an application to a dementia dataset from a French cohort study, *PAQUID*, aiming at studying cognitive decline among the elderly.

The remainder of this thesis proceeds as follows. In Chapter 2, we give an introduction to quantile regression and its application for longitudinal data along with some basic quantities for survival analysis and the Cox's proportional hazards models. Chapter 3 describes the proposed joint model for three types of data: longitudinal outcome, right-censored time-to-event, and interval-censored time-to-event. A MCEM algorithm is developed for computing non-parametric maximum likelihood estimates. We then show and prove some asymptotic properties of the resulting estimators. In Chapter 4, we illustrates the performance of our proposed joint model by carrying out some extensive simulation studies and applying the method to a dataset from a dementia study. Finally, in Chapter 5, a general discussion summarizes the advantages of our proposed joint model, as well as some possible extensions and perspectives.

Chapter 2

Quantile regression and Survival analysis

In this chapter, we discuss the basics of quantile regression models for a response variable and the Cox's proportional hazards models for time-to-event data, separately. Introduced by Koenker and Bassett Jr (1978), quantile regression is an extension of the traditional mean regression which provides a more complete view of the distribution of the response variable. It has become a very popular approach and applied to a wide range of studies, including biomedicine, epidemiology, ecology, agriculture, econometrics and finance. We will focus on the application of quantile regression in longitudinal studies. Proposed by Cox (1972), the proportional hazards model has been used primarily in biomedicine studies to model the effect of secondary variables on survival time. Unlike a specific life distribution model, the Cox proportional hazards model does not require any specific assumptions of the life distribution in modelling and testing many inferences about survival. We will describe the Cox proportional hazards models for both symptomatic and asymptomatic events (right- and interval-censored survival data).

2.1 Quantile regression

In traditional mean regression, the mean and the standard deviation are two essential measures used to describe a distribution. The mean describes the central location of one distribution, and the standard deviation describes the dispersion. However, focusing on the mean and standard deviation alone will lead us to ignore other important properties which offer more insights into the distribution. Self-thinning of tropical plants (Cade and Guo, 2000) is a very interesting example, where the effects of increasing germination densities of seedlings on the reduction in densities of mature plants were best revealed at the higher plant densities with intense intraspecific competition. Also, in social science, researchers often have data sets with skewed distribution which could not be well characterized by the mean and the standard deviation. To describe the distributional attributes of asymmetric response data sets, quantile regression is developed based on quantiles of the response distribution and measures the effect of covariates on the entire response distribution.

2.1.1 Quantiles

For any real-valued random variable Y, its cumulative distribution function is defined as

$$F(y) = P(Y \leq y),$$

where y is a specified value within the range of Y and P is the probability measure. The cumulative function $F(\cdot)$ is monotonic increasing and has limits 0 and 1 at $-\infty$ and ∞ respectively. The τth quantile of Y, denoted as $Q_\tau(Y)$, is defined as the smallest value of y such that the probability of $Y \leq y$ is τ, where τ is the level of the

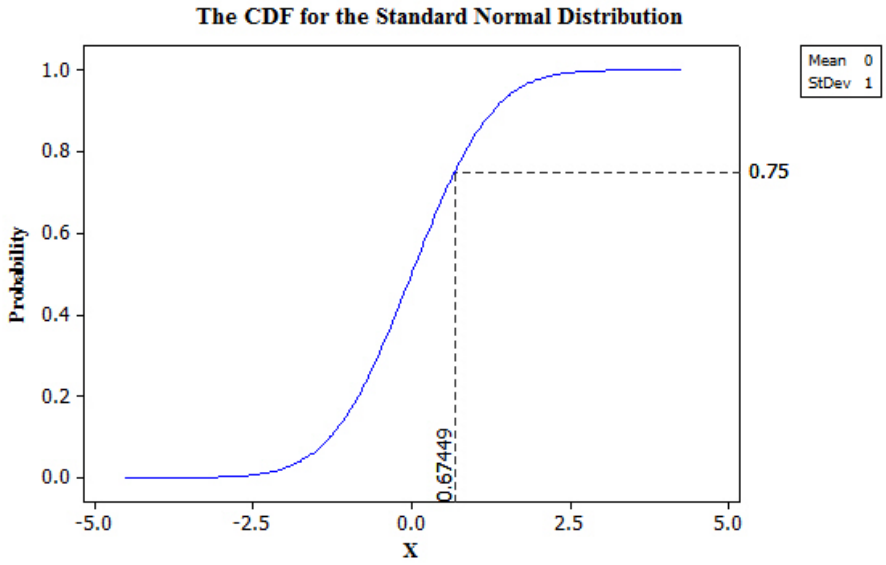

Figure 2.1: The cumulative distribution function for the standard normal variable X and the 75% quantile of X.

quantile which takes values between 0 and 1. That is,

$$Q_\tau(Y) = F^{-1}(\tau) = \inf\{y : F(y) \leq \tau\}.$$

For example, for the standard normal variable X, as shown in Figure 2.1, $F(0.67449) = 0.75$, so $Q_{0.75} = 0.67449$.

We define the quantile function, $Q_{(\cdot)}(Y)$, as a function of τ corresponding to the $F(Y)$. For the same variable Y, the value of Q_τ increases as τ increases indicating that the quantile function is also monotonic increasing. An example of the quantile function is shown in Figure 2.2 along with the corresponding cumulative distribution function for the standard normal distribution. By allowing the quantile level τ to vary between 0 and 1, Q_τ give us the ability to examine a distribution at different locations not just at the center (e.g. the mean for a symmetric distribution and the median for an asymmetric distribution). For example, one may be interested in examining a

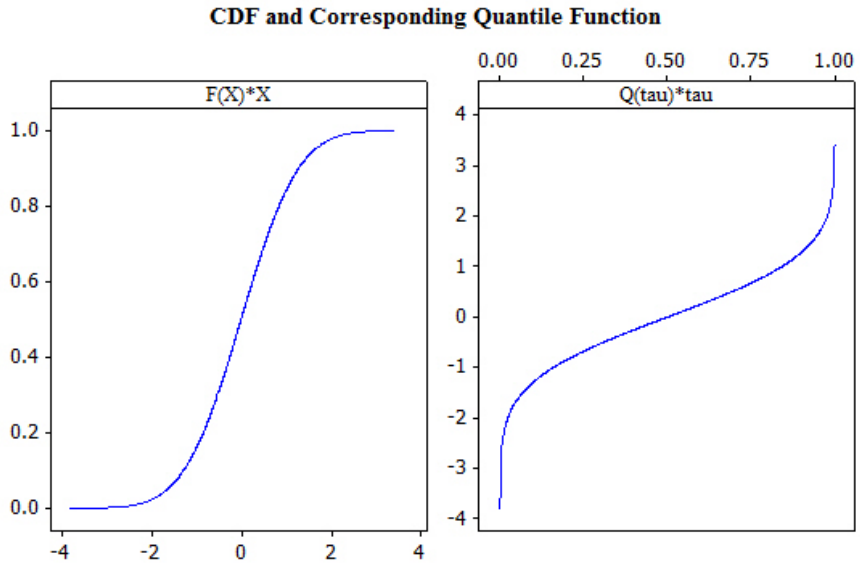

Figure 2.2: The cumulative distribution function and the quantile function for the standard normal variable X.

location at the lower tail (e.g. 0.15th quantile) or upper tail (e.g. 0.90th quantile) of a distribution.

Standard deviation is a commonly used measure to describe the scale or spread of a symmetric distribution. However, when the distribution becomes highly asymmetric or heavy-tailed, a quantile-based scale measure will characterize the scale better. We define

$$QSC_\tau = Q_{1-\tau} - Q_\tau$$

as the scale measure for skewed and heavy-tailed distributions, where τ is selected and less than 0.5. Therefore, we can obtain the spread of any desirable middle $100(1-2\tau)\%$ of a distribution by QSC_τ. For example, the conventional interquartile range (IQR) is actually $QSC_{0.25} = Q_{0.75} - Q_{0.25}$ which measures the spread of the middle 50% of the population.

We then can describe the skewness of a distribution as the level of imbalance

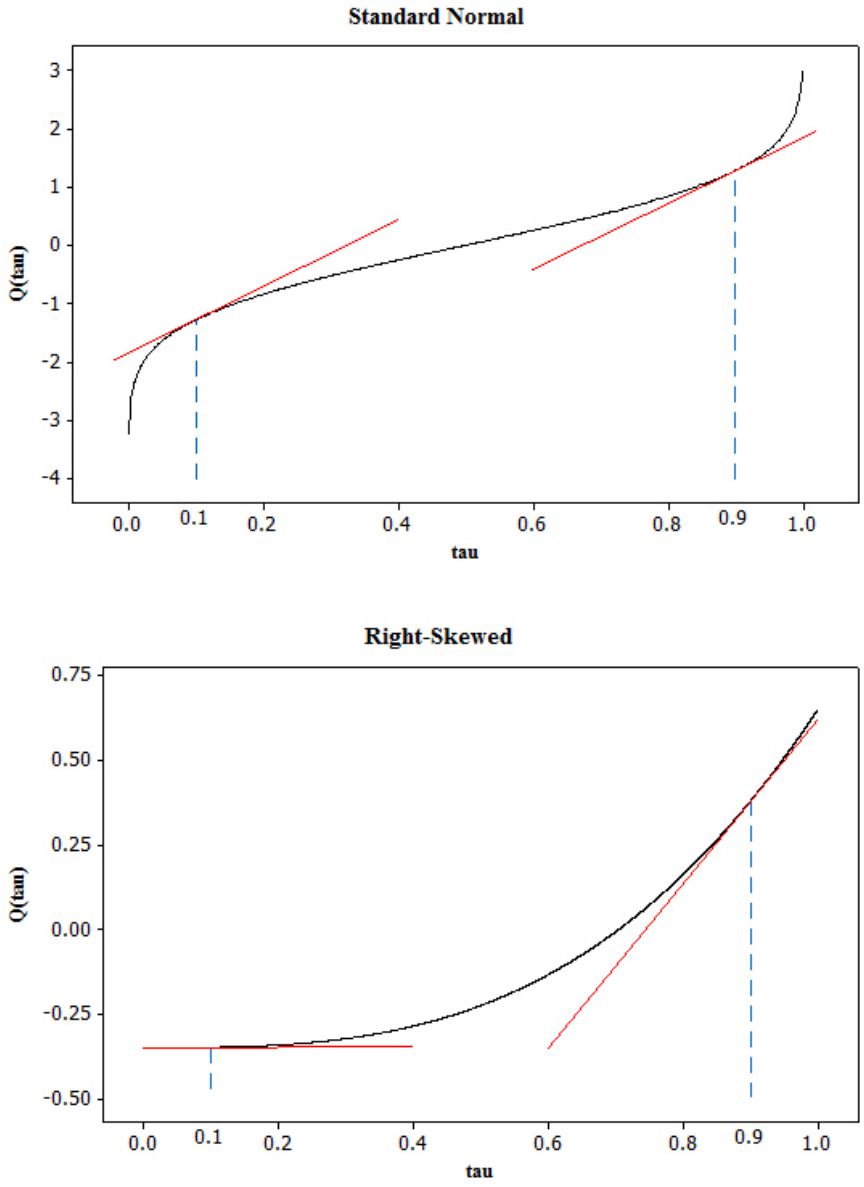

Figure 2.3: Quantile Functions for Standard Normal and a Skewed Distribution

between the scales above and below the median. The upper scale is characterized by $Q_{1-\tau}-Q_{0.5}$ and the lower scale is measured by $Q_{0.5}-Q_{\tau}$. For a symmetric distribution, the upper and lower scales should be the same for any $\tau < 0.5$. If the distribution is skewed, the quantile function will be asymmetric about the median and the difference between upper and lower scales will become large when the distribution becomes less symmetric (a positive difference indicating right skewness and a negative difference indicating left skewness). As shown in Figure 2.3, the slopes of the quantile function at any pair of $(Q_{\tau}, Q_{1-\tau})$ around the median are the same for the standard normal (symmetric) distribution. However, the slope at Q_{τ} is less than the slope at $Q_{1-\tau}$ for a right skewed distribution and $\tau < 0.5$. A measure of quantile-based skewness, QSK_{τ}, is defined as an expression of the ratio between the upper and lower scales. That is,

$$QSK_{\tau} = \frac{Q_{1-\tau} - Q_{0.5}}{Q_{0.5} - Q_{\tau}} - 1$$

for $\tau < 0.5$. Thus, the quantity QSK_{τ} takes the value zero for a symmetric distribution, a negative value for a left-skewed distribution and a positive value for a right-skewed distribution.

2.1.2 Quantile regression for independent data

Similar to the traditional linear regression models which are based on the mean, we try to construct regression models that are based on the quantiles of the response variable. The mean of the distribution of Y can be obtained by minimizing the mean

squared deviation $E\left[(Y-\mu)^2\right]$, where

$$\begin{aligned} E[(Y-\mu)^2] &= E[Y^2] - 2E[Y]\mu + \mu^2 \\ &= (\mu - E[Y])^2 + (E[Y^2] - (E[Y])^2) \\ &= (\mu - E[Y])^2 + Var(Y). \end{aligned}$$

Since $Var(Y)$ is constant, we minimize $E[(Y-\mu)^2]$ by taking $\mu = E[Y]$. For a sample of n realizations of the variable Y, $y_1, \ldots, y_n$, the sample mean can be obtained by seeking the point μ that minimizes the mean squared distance $\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2$. To similarly define quantiles as a solution to a minimization problem, we use the following quantile loss function

$$\rho_\tau(u) = u(\tau - I(u < 0))$$

which gives u a weight of τ if $u \geq 0$ and a weight of $\tau - 1$ if $u < 0$. We seek to minimize the expected loss, $E\left[\rho_\tau(Y-\hat{y})\right] = \int_{-\infty}^{+\infty}\rho_\tau(y-\hat{y})dF(y)$, by differentiating with respect to $\hat{y}$ and setting the partial derivative to zero. That is

$$\begin{aligned} \frac{\partial}{\partial\hat{y}}E\left[\rho_\tau(Y-\hat{y})\right] &= \frac{\partial}{\partial\hat{y}}(\tau-1)\int_{-\infty}^{\hat{y}}(y-\hat{y})dF(y) + \frac{\partial}{\partial\hat{y}}\tau\int_{\hat{y}}^{+\infty}(y-\hat{y})dF(y). \\ &= (1-\tau)\int_{-\infty}^{\hat{y}}dF(y) + \tau\int_{\hat{y}}^{+\infty}dF(y) \\ &= \int_{-\infty}^{\hat{y}}dF(y) - \tau\left\{\int_{-\infty}^{\hat{y}}dF(y) + \int_{\hat{y}}^{+\infty}dF(y)\right\} \\ &= F(\hat{y}) - \tau\int_{-\infty}^{+\infty}dF(y) \\ &= F(\hat{y}) - \tau \\ &\overset{\text{set}}{=} 0. \end{aligned}$$

When the solution is unique, $Q_\tau(y) = \hat{y} = F^{-1}(\tau)$; otherwise, we choose the smallest value from a set of τth quantiles. Thus, the τth sample quantile can be expressed as

the solution to the following minimization problem

$$\min_{\hat{y}\in\mathbb{R}} \sum_{i=1}^{n} \rho_\tau(y_i - \hat{y}). \tag{2.1}$$

If the conditional mean of Y given X is linear and expressed as $E(Y|X) = X^T\beta$, then β can be estimated by solving

$$\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} (y_i - x_i^T\beta)^2,$$

where β denotes a vector of unknown, fixed, parameters summarizing the effects of X on the conditional mean of the response Y. Similarly, since the τth sample quantile solves the problem in (2.1), we are willing to specify the following quantile regression model

$$Q_\tau(Y|X) = X^T\beta_\tau, \tag{2.2}$$

where β_τ denotes the vector of parameters that summarizes the effects of X on the τth conditional quantile of the response Y. Further, the model in (2.2) can also be formulated in the form of a conventional linear model

$$Y = X^T\beta_\tau + \varepsilon, \tag{2.3}$$

where ε denotes a random error term with $Q_\tau(\varepsilon|\beta_\tau, X) = 0$. We can proceed the estimation of β_τ by solving

$$\hat{\beta}_\tau = \arg\min_{\beta_\tau\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T\beta_\tau). \tag{2.4}$$

Optimal solutions to the above problem can be derived using appropriate algorithms, see in Koenker and Bassett Jr (1978), Koenker and D'Orey (1987) and others. There

exits a natural link between the minimization of the quantile loss function and the maximum likelihood theory if the error term in (2.3) is assumed to follow an asymmetric Laplace distribution (ALD), see, among others, Koenker and Machado (1999). If a random variable Y is ALD distributed, then its density is

$$f_Y(Y|\mu, \varrho, \tau) = \frac{\tau(1-\tau)}{\varrho} \exp\left\{-\rho_\tau\left(\frac{Y-\mu}{\varrho}\right)\right\},$$

where $\rho_\tau(u)$ is the quantile loss function defined previously, μ is the location parameter, τ determines the skewness, and $\varrho > 0$ is a scale parameter. For a sample of independent observations, $y_1, \ldots, y_n$, assuming that $Y_i \sim ALD(\mu_i, \varrho, \tau)$ and $\mu_i = X_i^T\beta$, the likelihood function can be derived as

$$L(\beta, \varrho, \tau) = \left[\frac{\tau(1-\tau)}{\varrho}\right]^n \exp\left\{-\sum_{i=1}^{n} \rho_\tau\left(\frac{y_i-\mu_i}{\varrho}\right)\right\},$$

The assumption of ALD errors allows us to recast quantile regression optimization in a (pseudo) maximum likelihood framework. The estimates under the ALD assumption are robust to misspecification of error distributions. Furthermore, such a distributional assumption allows several extensions of the basic framework, including modelling dependent observations.

2.1.3 Quantile Regression for longitudinal data

It is very common in many application studies that data are collected repeatedly on individuals over time. We call this type of data as longitudinal data. For each subject, we may have more than one observations. Under this scenario, the independence assumption between observations may no longer hold. The dependence can be influenced by variabilities coming from three sources: between-subject variability,

within-subject variability, and random error. Failing to take the dependence into consideration in a model setup may lead to severely biased parameter estimates. There are two common approaches to deal with the dependence between longitudinal observations, which essentially fall into the families of marginal and conditional models. The first approach is to specify explicitly an association structure between repeated observations together with the model for the response quantiles. The second approach is to jointly specify the response quantiles and the dependence between longitudinal observations by introducing subject-specific parameters. We discuss the first approach firstly.

Let $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$ be $n_i \geq 1$ repeated measures observed from the ith subject, for $i = 1, \ldots, n$ where n is the number of subjects. Let $x_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$ be the p-dimensional covariate vector associated with the parameter β_τ. A marginal model can be specified as

$$Q_\tau(y_{ij}|x_{ij}) = x_{ij}^T \beta_\tau,$$

or equivalently as

$$y_{ij} = x_{ij}^T \beta_\tau + \varepsilon_{ij},$$

where $Q_\tau(\varepsilon_{ij}|\beta_\tau, x_{ij}) = 0$ and the error terms are independent over different subjects but dependent over repeated measurements on the same subject. When a working independence is assumed between repeated responses from the same individual, we can estimate β_τ by minimizing the following objective function

$$S(\beta_\tau) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \rho_\tau(y_{ij} - x_{ij}^T \beta_\tau)$$

with derived estimating equation as

$$\frac{\partial S(\beta_\tau)}{\partial \beta_\tau} = \sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij}\psi_\tau(y_{ij} - x_{ij}^T\beta_\tau) = \sum_{i=1}^{m} X_i^T\psi_\tau(y_i - X_i\beta_\tau) = 0,$$

where $X_i = [x_{i1}, \ldots, x_{in_i}]^T$ is the $n_i \times p$ matrix of covariates, $y_i = (y_{i1}, \ldots, y_{in_i})^T$ is the $n_i \times 1$ vector of the variable of repeated measures for the ith individual, $\psi_\tau(u) = \rho'_\tau(u) = \tau - I(u < 0)$, and $\psi_\tau(y_i - X_i\beta_\tau) = (\psi_\tau(y_{i1} - x_{i1}^T\beta_\tau), \ldots, \psi_\tau(y_{in_i} - x_{in_i}^T\beta_\tau))^T$ is a $n_i \times 1$ vector. There is an efficient algorithm (Koenker and D'Orey, 1987) to obtain an estimate of β_τ by solving the above equation, which is available in statistical software R (package "quantreg").

Jung (1996) introduced a quasi-likelihood method to take the within-subject correlations into consideration for median regression. Let $f_{ij}(\cdot)$ be an unknown density of ε_{ij}. A quasi-likelihood based estimating equation for β_τ is derived as

$$\sum_{i=1}^{m} X_i^T \Gamma_i V_i^{-1}\psi_\tau(y_i - X_i\beta_\tau) = 0,$$

where $V_i = \mathrm{cov}\left(\psi_\tau(y_i - X_i\beta_\tau)\right)$ and $\Gamma_i = \mathrm{diag}\left[f_{i1}(0), \ldots, f_{in_i}(0)\right]$ is to account for possible overdispersion in the error distribution. However, the estimation of the covariance matrix V_i becomes much complicated when quasi-likelihood method is applied. Whatever correlation matrix that ε_i has, the correlation matrix of $\psi_\tau(\varepsilon_i)$ is no longer the same one, and its correlation structure may be very difficult to specify. To overcome this difficulty, Lu and Fan (2015) proposed a general stationary autocorrelation structure for V_i and estimated the parameters by solving the following equation

$$U(\beta_\tau) = \sum_{i=1}^{m} X_i^T \Gamma_i \Sigma_i^{-1}(\rho)\psi_\tau(y_i - X_i\beta_\tau) = 0, \tag{2.5}$$

where $\Sigma_i(\rho)$ is the covariance matrix of $\psi_\tau(\varepsilon_i)$ that can be expressed as $\Sigma_i(\rho) =$

$A_i^{\frac{1}{2}} C_i(\rho) A_i^{\frac{1}{2}}$ with $A_i = \text{diag}[\sigma_{i11}, \ldots, \sigma_{1n_in_i}]$ being an $n_i \times n_i$ diagonal matrix, $\sigma_{ijj} = \text{var}(\psi_\tau(\varepsilon_{ij}))$ and $C_i(\rho)$ as the correlation matrix of $\psi_\tau(\varepsilon_i)$, ρ being a correlation index parameter. The matrix A_i can be estimated theoretically by $\tilde{\sigma}_{ijj} = \tau(1-\tau)$ or empirically by $\hat{\sigma}_{ijj} = \frac{1}{m}\sum_{i=1}^{m} I\left(y_{ij} < x_{ij}^T \beta_\tau\right)\left(1 - \frac{1}{m}\sum_{i=1}^{m} I\left(y_{ij} < x_{ij}^T \beta_\tau\right)\right)$. The stationary autocorrelation structure of $C_i(\rho)$ is given by

$$C_i(\rho) = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n_i-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n_i-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{n_i-1} & \rho_{n_i-2} & \rho_{n_i-3} & \cdots & 1 \end{pmatrix}$$

where ρ_ℓ is estimated by

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i-\ell} \tilde{y}_{ij}\tilde{y}_{i,j+\ell}/m(n_i-\ell)}{\sum_{i=1}^{m}\sum_{j=1}^{n_i} \tilde{y}_{ij}^2/mn_i}$$

for $\ell = 1, \ldots, n_i - 1$ with $\tilde{y}_{ij} = \left\{\psi_\tau(y_{ij} - x_{ij}^T\beta_\tau)\right\}/\sqrt{\sigma_{ijj}}$. Since the objective function $U(\beta_\tau)$ in (2.5) is non-continuous and can not be differentiated, an induced smoothing method is applied and leads to a smoothed estimating function

$$\tilde{U}(\beta_\tau) = \sum_{i=1}^{m} X_i^T \Gamma_i \Sigma_i^{-1}(\rho)\tilde{\psi}_\tau(y_i - X_i\beta_\tau),$$

where $\tilde{\psi}_\tau(\varepsilon_{ij}) = \tau - 1 + \Phi(\frac{\varepsilon_{ij}}{r_{ij}})$ with Φ being the cumulative distribution function of the standard normal distribution, $r_{ij} = \sqrt{x_{ij}^T \Omega x_{ij}}$, and Ω being an estimate of the covariance matrix of β_τ. We can use $\partial\tilde{U}(\beta_\tau)/\partial\beta_\tau$ as an approximation of $\partial U(\beta_\tau)/\partial\beta_\tau$. It can be derived that

$$\frac{\partial\tilde{U}(\beta_\tau)}{\partial\beta_\tau} = -\sum_{i=1}^{m} X_i^T \Gamma_i \Sigma_i^{-1}(\rho)\tilde{\Lambda}_i X_i,$$

where $\tilde{\Lambda}_i$ is an $n_i \times n_i$ diagonal matrix with the jth diagonal element $\frac{1}{r_{ij}}\phi((y_{ij} - x_{ij}^T\beta_\tau)/r_{ij})$, and ϕ is the density of the standard normal distribution.

An alternative approach to account for the dependence between repeated observations is to include a measure of the unobserved heterogeneity in the quantile regression models. This heterogeneity comes either from unobserved covariates or from a different effect of measured covariates on the response due to genetic, environmental, social and/or economic factors. We can define a conditional quantile regression model as

$$Q_\tau(y_{ij}|b_i, x_{ij}) = b_i + x_{ij}^T\beta_\tau,$$

or equivalently as

$$y_{ij} = b_i + x_{ij}^T\beta_\tau + \varepsilon_{ij},$$

where b_i denote subject-specific parameters that could be distribution free or independent and identically distributed random variables. We can also assume a q-dimensional vector of subject-specific random parameters, $b_i = (b_{i1}, \ldots, b_{iq})$. Therefore, a linear quantile mixed model is defined by

$$y_{ij} = x_{ij}^T\beta_\tau + z_{ij}^T b_i + \varepsilon_{ij},$$

where z_{ij} denotes a subset of x_{ij}. As mentioned by Geraci and Bottai (2014), the random structure above allows to account for between-subject heterogeneity associated with given explanatory variables and does not require orthogonality between the observed and the unobserved covariates. The estimation of parameters can proceed through a maximum likelihood method. Let $f_b(\cdot; \Sigma_\tau)$ be the density of $\boldsymbol{b}_i$ with a

covariance matrix Σ_τ. The likelihood function is defined by

$$L(\beta_\tau, \varrho, \Sigma_\tau, \tau) = \prod_{i=1}^{n} \int_{\boldsymbol{b}_i} \prod_{j=1}^{n_i} f_Y\left(y_{ij}|\boldsymbol{b}_i, \beta_\tau, \varrho, \tau\right) f_b\left(\boldsymbol{b}_i; \Sigma_\tau\right) d\boldsymbol{b}_i.$$

The integral in the expression above does not have a closed form solution and numerical integration methods are required. A Monte Carlo Expectation and Maximization (MCEM) method has been derived and Gaussian quadrature methods are suggested to reduce the computational burden, see among others, Liu and Bottai (2009), Geraci and Bottai (2014).

2.2 Survival analysis

Time-to-event data arises in many applied studies, such as medicine, biology, health science, epidemiology, engineering, economics, and demography. The time that takes for a well-defined event to occur is termed as survival time. Survival analysis examines and models the survival data which contains the response of time and explanatory or predictor variables. Observations of event time are censored if an event is known to occur only in a certain period of time, or in other words, for some subjects the event has not occurred at the end of study. Possible types of censoring are right censoring, where all that is known is that the subject has not experienced the event at a given time, left censoring, where event has occurred prior to the start of the study, or interval censoring, where event can only be known to occur between two time points. Within many well-known methods in survival analysis for estimating the distribution of survival times, some focus on estimating unconditional survival distributions, however the most interesting method is to examine the relationship between survival times and one or more predictors. One most widely used method of survival analysis is the Cox proportional hazards regression model introduced by Cox

(1972).

2.2.1 Basic quantities in survival analysis

Let T be the time until some well-defined event. This event may be death, the development of some disease, equipment breakdown, and so forth. We will assume that T is a non-negative continuous random variable from a homogeneous population with probability density function $f_T(t)$ and cumulative distribution function $F_T(t) = P(T < t)$. The density $f_T(t)$ gives the unconditional probability of the event's occurring at time t and the distribution function $F_T(t)$ is the probability that the event has occurred by time t.

It will often be convenient to work with the survival function which is the complement of the cumulative distribution function $F_T(t) = P(T < t)$, giving the probability of a subject experiencing the event after time t. It is defined as

$$S(t) = P(T \geq t) = 1 - F_T(t) = \int_t^\infty f_T(s)ds$$

which is continuous and strictly decreasing. Thus, we have

$$f_T(t) = \frac{-dS(t)}{dt}. \tag{2.6}$$

Another basic quantity in survival analysis is the hazard function, also known as the conditional failure rate in reliability, the age-specific failure rate in epidemiology, the inverse of the Mill's ratio in economics, or simply referred as the hazard rate. The hazard function is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P\left\{t \leq T < t + \Delta t | T \geq t\right\}}{\Delta t}.$$

One can see that $\lambda(t)\Delta t$ may be viewed as the conditional probability that the event will occur in the next instant given that it has not occurred before the current time t. As Δt goes down to zero, we obtain an instantaneous rate of occurrence. The above expression of hazard function can be further written as

$$\begin{aligned}\lambda(t) &= \lim_{\Delta t \to 0} \frac{P\{t \leq T < t + \Delta t, T \geq t\}}{\Delta t} \frac{1}{P\{T \geq t\}} \\ &= \lim_{\Delta t \to 0} \frac{P\{t \leq T < t + \Delta t\}}{\Delta t} \frac{1}{P\{T \geq t\}} \\ &= \frac{f_T(t)}{S(t)}.\end{aligned}$$

Along with (2.6), we have

$$\lambda(t) = \frac{-d \log[S(t)]}{dt}. \tag{2.7}$$

By introducing the boundary of survival function at time zero to be one ($S(0) = 1$), the integration of both sides of (2.7) gives a formula for the survival function in terms of the hazard:

$$S(t) = \exp\left\{-\int_0^t \lambda(s)ds\right\}. \tag{2.8}$$

The integral in the equation (2.8) is the cumulative hazard function (or cumulative risk function), denoted by

$$\Lambda(t) = \int_0^t \lambda(s)ds.$$

The other important quantity in survival analysis, especially in life length studies, is the mean residual life function. It is defined as $mrl(t) = E(T - t|T > t)$, measuring the expected remaining lifetime for a subject at time t. It then follows that

$$mrl(t) = \frac{\int_t^\infty (s-t) f_T(s)ds}{S(t)} = \frac{\int_t^\infty S(s)ds}{S(t)}.$$

Furthermore, the mean or expected lifetime by definition is

$$E(T) = mrl(0) = \frac{\int_0^\infty S(t)dt}{S(0)} = \int_0^\infty S(t)dt,$$

followed by the variance of T:

$$Var(T) = 2\int_0^\infty tS(t)dt - \left[\int_0^\infty tS(t)dt\right]^2.$$

2.2.2 Censoring and likelihoods

A distinguishing feature often present in survival analysis, is known as censoring, which occurs when some event times can only be observed within certain intervals. The first category of censoring we consider is the right censoring. There are three types of right censoring, the first is the *Type I* censoring where the exact event time is observed only if it occurs prior to some pre-specified time, the second is the *Type II* censoring in which the study continues until a certain number of subjects has experienced the event, and the *Type III* is a random censoring where some subjects may experience some competing event causing them to be removed from the study and whose times for the event of interest be right censored. Let C_r be the potential censoring time and T is the time variable. The right censored data can be conveniently represented by pairs of (D, Δ), where $D = \min(T, C_r)$ and $\Delta = I(T \leq C_r)$. The event indicator Δ (= 0 or 1) indicates whether the exact event time is observed ($\Delta = 1$, $T \leq C_r$) or the event is censored ($\Delta = 0$, $T > C_r$).

The other type of censoring is the left censoring in which the subjects has already experienced the event of interest sometime before time C_l (denoting left censoring time). Similar to the right censoring, the left censored data can be represented by pairs of (D, Υ), where $D = \max(T, C_l)$ and $\Upsilon = I(T \geq C_l)$. The exact time is

observed if $T \geq C_l$ with the event indicator $\Upsilon = 1$, while the event is left censored if $T \leq C_l$ with $\Upsilon = 0$.

A more general type of censoring in survival data is the interval censoring where the event of interest is only known to occur within an interval of time. This type of censoring often occurs in clinical or longitudinal studies where subjects have periodic follow-up and their event time can only be known to fall in an interval, say $(L, R]$ (L represents the left time point and R represents the right time point for the censoring interval).

When constructing likelihood functions for survival models, we should very carefully consider the censoring mechanisms. If an observation reflecting the exact event time, it provides information on the probability that the event is occurring at this time (approximately equal to $f_T(T)$). Right censoring observations provide information of the survival function $S(C_r)$ that the event time is larger than the right censoring time. For left-censored observations, the contribution to the likelihood is the probability that the event has already occurred before C_l which is equal to the cumulative distribution function $F_T(C_l) = 1 - S(C_l)$. Finally, interval-censored observations provide information on the probability that the event occurred within the interval $(L, R]$. Now, suppose a study involves n subjects with T denoting the variable of time for an event of interest. Let $\mathcal{T}$ be the set of exact event times, $\mathcal{R}$ be the set of right-censored observations, $\mathcal{L}$ be the set of left-censored observations, and $\mathcal{I}$ be the set of interval-censored observations. The likelihood function can be constructed by combining all information on the survival time as

$$L = \prod_{i \in \mathcal{T}} f_T(t_i) \prod_{i \in \mathcal{R}} S(C_{ri}) \prod_{i \in \mathcal{L}} [1 - S(C_{li})] \prod_{i \in \mathcal{I}} [S(L_i) - S(R_i)] \,,$$

or equivalently,

$$L = \prod_{i \in \mathcal{T}} \lambda(t_i) S(t_i) \prod_{i \in \mathcal{R}} S(C_{ri}) \prod_{i \in \mathcal{L}} [1 - S(C_{li})] \prod_{i \in \mathcal{I}} [S(L_i) - S(R_i)] ,$$

based on the relationship between the density, hazard and survival functions.

2.2.3 Cox's proportional hazards model for survival data

Survival analysis typically models the effects of covariates or explanatory variables on the survival time. Many survival models focus directly on the hazard function. The most common method is to specify a linear-like model for the logarithm of the hazard function. For example, a parametric model based on the exponential distribution may be written as

$$\log \lambda(t|X) = \alpha + X^T \beta,$$

or, equivalently,

$$\lambda(t|X) = \exp\left\{\alpha + X^T \beta\right\},$$

as a linear model for the log-hazard function or as a multiplicative model for the hazard function, where α is a consant and X is a vector of covariates whose effect on survival time is measured by β. The constant α represents baseline log-hazard when all covariates are zero, $\log \lambda(t|X = \mathbf{0}) = \alpha$, or $\lambda(t|X = \mathbf{0}) = e^{\alpha}$.

Cox (1972) introduced a family of survival models that leave the baseline hazard function unspecified:

$$\log \lambda(t|X) = \alpha(t) + X^T \beta,$$

or, equivalently,

$$\lambda(t|X) = \lambda_0(t) \exp\left\{X^T \beta\right\}. \tag{2.9}$$

This model is semi-parametric because the baseline hazard does not have any parametric assumptions while the covariates enter the model linearly. The Cox model is often called a proportional hazards model, if we consider two subjects with different covariate values X and X^*, the hazard ratio for these two subjects is

$$\frac{\lambda(t|X)}{\lambda(t|X^*)} = \frac{\lambda_0(t)\exp\left\{X^T\beta\right\}}{\lambda_0(t)\exp\left\{X^{*T}\beta\right\}} = \exp\left\{(X - X^*)^T\beta\right\}$$

which is independent of time t. By integrating the hazard function in (2.9), we obtain the cumulative hazard

$$\Lambda(t|X) = \Lambda_0(t)\exp\left\{X^T\beta\right\}$$

which is also proportional, where $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$. Then the survival function follows as

$$S(t|X) = \exp\left\{-\Lambda_0(t)\exp\left\{X^T\beta\right\}\right\}.$$

Typically in many survival studies, subjects are monitored during the study, and other covariate variables are recorded whose values may change over time. The Cox proportional hazards model is also possible to include these time-dependent covariates. Let $X(t)$ denote a set of covariates or risk factors at time t which may effect the survival distribution of T. The proportional hazards model can be generalized to

$$\lambda(t|X(t)) = \lambda_0(t)\exp\left\{X(t)^T\beta\right\}.$$

A general form of the likelihood function for a Cox's proportional hazards model

with possible time-dependent covariates thus can be constructed as

$$
\begin{aligned}
L(\beta) = & \prod_{i \in \mathcal{T}} \lambda_0(t_i) \exp\left\{X(t_i)^T \beta\right\} \exp\left\{-\int_0^{t_i} \lambda_0(s) \exp\left(X(s)^T \beta\right) ds\right\} \\
& \prod_{i \in \mathcal{R}} \exp\left\{-\int_0^{C_{ri}} \lambda_0(s) \exp\left(X(s)^T \beta\right) ds\right\} \\
& \prod_{i \in \mathcal{L}} \left[1 - \exp\left\{-\int_0^{C_{li}} \lambda_0(s) \exp\left(X(s)^T \beta\right) ds\right\}\right] \\
& \prod_{i \in \mathcal{I}} \left[\exp\left\{-\int_0^{L_i} \lambda_0(s) \exp\left(X(s)^T \beta\right) ds\right\} - \exp\left\{-\int_0^{R_i} \lambda_0(s) \exp\left(X(s)^T \beta\right) ds\right\}\right].
\end{aligned}
$$

Estimates of β then can be obtained through non-parametric maximum likelihood method.

Chapter 3

Proposed joint model of a quantile of longitudinal outcome and multiple-censored survival times

In this chapter, we introduce a joint model for multiple types of right- and interval-censored event times and quantiles of a longitudinal response. The joint distribution of these types of data are related with potentially time-dependent covariates and latent variables through a linear quantile regression model and proportional hazards models. The longitudinal and survival processes share some common predictors and a shared random effect is assumed between all right- and interval-censored events, which results in a hybrid of the shared-parameter model and the joint model. We develop a Monte Carlo expectation maximization (MCEM) algorithm to obtain the maximum likelihood estimates of parameters. We show that the estimators are consistent, efficient and asymptotically normal.

3.1 Data and Model

Let $T_1, \ldots, T_{K_1}$ denote the occurring times for a number of K_1 types of asymptomatic events, and $T_{K_1+1}, \ldots, T_K$ are observed failure times for a number of K_2 symptomatic events, where $K_2 = K - K_1$. The longitudinal response Y_{ij} is repeatedly observed at visits $j = 1, \ldots, n_i$ during the follow-up for the ith individual, $i = 1, \ldots, n$. We assume that the longitudinal outcome is associated with event times $T_1, \ldots, T_K$, but is independent of censoring times.

We let X_{ij} be a vector of covariates used to model only longitudinal response and W_{ik} be a vector of covariates used to predict only the time of the kth event. The longitudinal and survival processes shared a common vector of covariates, $H_i(t)$, which is possibly dependent on time t. Further, different types of survival events are supposed to be dependent and the dependencies are captured by random effect b_i. Conditional on covariates X_{ij}, W_{ik}, $H_i(t)$ and random effect b_i, our model consists of two types of equations, one is a linear equation for the longitudinal response and the other one is the hazard function of T_k, $k = 1, \ldots, K$:

$$\begin{cases} Y_{ij} = \eta^T X_{ij} + \delta^T H_i(t) + \epsilon_{ij} = \tilde{Q}_{\tau ij} + \epsilon_{ij} \\ \lambda_k(t_i; \tilde{Q}_{\tau t_i}, W_{ik}, b_i) = e^{\beta_k^T W_{ik} + \alpha_k \delta^T H_i(t_i) + \zeta_k b_i} \lambda_{k0}(t_i). \end{cases} \tag{3.1}$$

where $\boldsymbol{\eta}$, $\boldsymbol{\delta}$, $\boldsymbol{\beta}_k$, α_k, ζ_k are unknown regression parameters associated with fixed and random effects, $\tilde{Q}_{\tau ij}$ denotes any specified τth quantile of the longitudinal outcome with the τth quantile of the distribution of the error ϵ_{ij} set to be 0, and $\lambda_{k0}(\cdot)$ is the baseline hazard function for the kth event. Furthermore, the model is based on the assumption that the change of the quantile of the longitudinal outcome has effects on the development of survival events, that is, the risk of survival events are conditional on the history of longitudinal process up to the current time which is denoted as $\tilde{Q}_{\tau t_i}$.

The random effect b_i is a zero-centred normal variable with variance σ^2.

Unknown parameters $\boldsymbol{\eta}$ and $\boldsymbol{\beta}_k$ can be estimated to measure the effects of covariates, X_{ij}, used only in longitudinal model and the effects of covariates, W_{ik}, used only in the survival processes. The effects of the time-dependent predictor $H_i(t)$ on the longitudinal observations can be estimated as $\boldsymbol{\delta}$. The contribution of the longitudinal process to the risk of the kth event is explained through $\boldsymbol{\delta}^T H_i(t)$ and measured by a scalar parameter α_k. The effects of the covariate $H_i(t)$ itself on the log-hazard ratios then can be measured by $\alpha_k \boldsymbol{\delta}$. By adding a latent normal random variable b_i, we tend to capture some underlying effects for the development of both asymptomatic and symptomatic events, and thus makes our survival models as mixed effects proportional hazard models.

Remark 1. The random effect b_i characterizes some common unobserved or omitted covariates that also affects the risks of both asymptomatic and symptomatic events. For example, in a cognitive ageing study, b_i represents the underlying health conditions for the development of dementia and death, such as social environment, depression, physical activity, and/or genetic factor. The effect of b_i on the log-hazard ratios of the kth event is measured by a scalar parameter ζ_k.

In the proposed joint model, we usually assume that K_1 and K_2 are greater than or equal to one, that is, we have one or more than one events of each type. However, the model can be reduced to one that contains only symptomatic events or one that contains only asymptomatic events by setting $K_1 = 0$ or $K_2 = 0$. Moreover, depending on the purpose of our study, the joint model can be modified to weigh one type of process more than the other one. For example, if our interest is in modelling the quantiles of the longitudinal response with dropout, we may use the survival models to measure the informative dropout, and thus reduce the bias in the estimation of parameters of the longitudinal process. In this scenario, we simplify the survival

models by removing some covariates and/or even the random effects and propose a more sophisticated (e.g. mixed effects) regression model for the longitudinal data. If we aim to get accurate prediction of event times but with a mismeasured longitudinal covariate, the longitudinal process in the joint model can be used to deal with measurement errors. In this scenario, we weigh survival models more than the longitudinal process, and a quantile regression is considered in the longitudinal process to manage possibly skewed measurement errors. In this work, we focus on predicting failure times for survival events and use longitudinal observations to reduce the bias and deal with measurement errors. For example, in a cognitive ageing study, we jointly analyse repeated assessment of a psychometric test with survival times for dementia and death. More discussions on potential extensions and modifications of our proposed joint model of longitudinal quantiles and survival times are addressed in Chapter 5.

3.2 The likelihood functions

We first describe the sub-models for each type of data separately, and then link them in the observed likelihood of the full joint model.

3.2.1 Likelihood attributed to longitudinal observations

As discussed in Chapter 2, the estimation of quantile regression parameters can proceed by solving the minimization problem (2.4), and thus lead to minimizing the quantile loss function. Similar properties to this minimization problem can be found in asymmetric least square estimations (Newey and Powell, 1987). Specifically, minimizing the quantile loss function is equivalent to maximizing the likelihood of an asymmetric Laplace distribution (ALD) (Koenker and Machado, 1999). Based on

this, we assume that the distribution of ϵ_{ij} is an ALD, and then $\tilde{Q}_{\tau ij}$ represents the conditional τth quantile of Y_{ij} for a specified and fixed $0 < \tau < 1$. Conditional on X_{ij} and $H_i(t)$, the distribution of Y_{ij} has a density function

$$f_Y(Y_{ij}|X_{ij}, H_i(t)) = \frac{\tau(1-\tau)}{\varrho} \exp\left\{-\rho\left(\frac{Y_{ij} - \eta^T X_{ij} - \delta^T H_i(t)}{\varrho}\right)\right\}, \tag{3.2}$$

where $\rho(s) = s(\tau - I(s < 0))$ is the quantile loss function, τ determines the skewness, and $\varrho > 0$ is a scale parameter. The assumption of an ALD will result in a pseudo-likelihood function for the longitudinal process when the error is not ALD distributed (e.g. normal, as illustrated in chapter 4 through simulation studies).

3.2.2 Likelihood attributed to asymptomatic events

Suppose that the monitoring times for detecting asymptomatic events are arbitrary for each subject and independent of the event time T_k ($k = 1, \ldots, K_1$). For the ith subject, we let $(L_{ik}, R_{ik}]$ be an interval with the lower bound L_{ik} being the largest monitoring time point below T_{ik} and the upper bound R_{ik} being the smallest monitoring time point above the event time T_{ik}. If L_{ik} is the last monitoring time during the follow-up study, we let $R_{ik} = \infty$ indicating that the asymptomatic event does not occur at least during the period of study. Conditioning on W_{ik}, $H_i(t)$ and b_i, the likelihood for the kth asymptomatic event can be expressed as the difference between the values of the cumulative distribution function $F_k(\cdot)$ at R_{ik} and L_{ik}, or equivalently be written as

the difference between the values of the survival function $S_k(\cdot)$ at L_{ik} and R_{ik}:

$$\begin{aligned} & F_k(R_{ik}; W_{ik}, H_i(t), b_i) - F_k(L_{ik}; W_{ik}, H_i(t), b_i) \\ & \quad = S_k(L_{ik}; W_{ik}, H_i(t), b_i) - S_k(R_{ik}; W_{ik}, H_i(t), b_i) \\ & \quad = \exp\left\{ - \int_0^{L_{ik}} e^{\beta_k^T W_{ik} + \alpha_k \delta^T H_i(t) + \zeta_k b_i} d\Lambda_k(t) \right\} \\ & \quad\quad - \exp\left\{ - \int_0^{R_{ik}} e^{\beta_k^T W_{ik} + \alpha_k \delta^T H_i(t) + \zeta_k b_i} d\Lambda_k(t) \right\}, \end{aligned} \tag{3.3}$$

where $\Lambda_k(t) = \int_0^t \lambda_{k0}(s) ds$.

3.2.3 Likelihood attributed to symptomatic events

For $k = K_1+1, \ldots, K$, let C_k be the censoring time of the kth symptomatic event such that we observe the event time as $D_k = \min(T_k, C_k)$. Further, let $\Delta_k = I(T_k \leqslant C_k)$ denote the symptomatic event indicator, where $I(\cdot)$ is the indicator function. Thus, we have that the survival function for the kth symptomatic event and the ith subject is

$$S_k(D_{ik}; W_{ik}, H_i(t), b_i) = \exp\left\{ - \int_0^{D_{ik}} e^{\beta_k^T W_{ik} + \alpha_k \delta^T H_i(t) + \zeta_k b_i} d\Lambda_k(t) \right\},$$

where $\Lambda_k(t) = \int_0^t \lambda_{k0}(s) ds$. Thus, the distribution of the symptomatic event time can be written as

$$f_k(D_{ik}, \Delta_{ik}; b_i) = \lambda_k(D_{ik}; W_{ik}, H_i(t), b_i)^{\Delta_{ik}} S_k(D_{ik}; W_{ik}, H_i(t), b_i), \tag{3.4}$$

where $\lambda_k(D_{ik}; W_{ik}, H_i(t), b_i)$ is given by the hazard function in (3.1).

3.2.4 Observed joint likelihood

We denote the data collected from a random sample of n subjects as $\{O_i : i = 1, \ldots, n\}$, where O_i is the union of $\{Y_{ij}, X_{ij}, H_i(\cdot) : j = 1, \ldots, n_i\}$, $\{L_{ik}, R_{ik}, W_{ik} : k = 1, \ldots, K_1\}$ and $\{D_{ik}, \Delta_{ik}, W_{ik}, k = K_1 + 1, \ldots, K\}$. We assume that the monitoring times for the asymptomatic events and censoring times for symptomatic events are independent of event time $T_{ik}(k = 1, \ldots, K)$ and random effect b_i. Further, we leave the baseline hazard function $\lambda_{k0}(\cdot)$ unspecified leading to a semi-parametric joint model and non-parametric maximum likelihood estimation (NPMLE) as in section 3.3. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_K^T)^T$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^T$ and $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_K)^T$. Concerning parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\zeta}, \varrho, \sigma^2)$ and $\mathcal{A} \equiv (\Lambda_1, \ldots, \Lambda_K)$, the joint likelihood of the longitudinal and survival processes is obtained as

$$L_n(\boldsymbol{\theta}, \mathcal{A}) = \prod_{i=1}^{n} \int_{b_i \in \mathbb{R}} \prod_{j=1}^{n_i} f_Y(Y_{ij}) \prod_{k=1}^{K_1} [S_k(L_{ik}) - S_k(R_{ik})] \prod_{k=K_1+1}^{K} f_k(D_{ik}) f_b(b_i) db_i, \quad (3.5)$$

where $f_Y(\cdot)$, $S_k(\cdot)$ and $f_k(\cdot)$ are the conditional functions defined in (3.2), (3.3) and (3.4), and $f_b(\cdot)$ is the density function of the normal distribution with mean 0 and variance σ^2.

When a study involves one terminal event (e.g., death), the study is terminated if that event occurs, such that no information is collected after the occurrence of the terminal event. Without loss of generality, let the Kth event is terminal in our joint model setting. Then T_K is bigger than any monitoring times for the asymptomatic events, and all the other symptomatic event times $T_k(k = K_1 + 1, \ldots, K - 1)$ are censored at $\min(C_k, T_K)$. The joint likelihood for this semi-competing risks setting is the same as the one in (3.5), due to the fact that non-terminal events are mutually independent, and $T_k(k = 1, \ldots, K - 1)$ are independent of monitoring times and C_k.

3.3 Estimation

We use the maximum likelihood estimation method for estimating parameters. To evaluate the integrations over b_i involved in the joint likelihood in (3.5), quadrature methods are often adopted. However, quadrature methods have limitations. One is that, a quadrature method works well for one type of the distribution of random effects but is not promised to be appropriate for other types of distribution. For example, we use Gauss-Hermite quadrature when the distribution of random effects is Gaussian and use Gauss-Laguerre quadrature when the support of random effects is $(0, \infty)$. The other drawback of the quadrature methods is that it becomes too slow or less accurate when the dimensionality of random effects is large. This may happen when we need multiple-dimensional random effects to accommodate multiple types of survival events in a joint model or to characterize unobserved properties at different times when used in the longitudinal model.

To give our proposed joint model the possibility and flexibility to be modified or extended under different purposes and settings of study (as discussed in Chapter 5), we propose a Monte Carlo expectation maximization (MCEM) algorithm for computation. The MCEM method can be applied to fit the model with any assumptions on the random effects and any forms of regression functions (linear, non-linear). The MCEM algorithm consists of two steps. One step is to take a random sample of random effects and estimate the conditional expectation of the complete data log-likelihood using current estimates of parameters (E-step), and the other step is to maximize the obtained expectation (M-step). To start, we set the initial values of $\boldsymbol{\eta}$, $\boldsymbol{\delta}$, $\boldsymbol{\beta}_k$, α_k as the estimates obtained by fitting separate longitudinal and survival models with only fixed effects. When we fit asymptomatic event times with traditional Cox models, practically, we can use the mean of L_{ik} and R_{ik} as an approximate event time of type k ($k = 1, \ldots, K_1$) for the ith subject. Let $\zeta_k = 1$ on the first run. We keep

alternating the E- and M-steps until convergence. The standard errors are estimated through a bootstrap procedure. In particular, we draw a simple random sample of size n by sampling subjects with replacement rather than observations. Inferences then can be made based on the standard errors and Wald statistics.

For fitting our proposed joint model with an MCEM algorithm, we first re-express the cumulative hazard functions in the joint likelihood function $L(\boldsymbol{\theta}, \mathcal{A})$ in (3.5) as summations of history hazards up to the desired times. Using the information of all subjects, we sort all the distinct interval boundaries of $(L_{ik}, R_{ik}]$ $(i = 1, \ldots, n; R_{ik} < \infty)$ for the kth $(k = 1, \ldots, K_1)$ asymptomatic event from the smallest to the largest as $t_{k1} < t_{k2} < \cdots < t_{km_k}$. For $k = K_1{+}1, \ldots, K$, all uncensored symptomatic event times D_{ik} (corresponding $\Delta_{ik} = 1$) are sorted in the same way as $t_{k1} < t_{k2} < \cdots < t_{km_k}$. Further, we estimate Λ_k $(k = 1, \ldots, K)$ with a step function, Λ_k, which jumps only at $t_{k1} < t_{k2} < \cdots < t_{km_k}$ with respective jump sizes of $\lambda_{k1}, \ldots, \lambda_{k,m_k}$. The objective function need to be maximized thus becomes the following

$$
\begin{aligned}
L_n(\boldsymbol{\theta}, \mathcal{A}) = & \prod_{i=1}^{n} \int_{b_i \in \mathbb{R}} \prod_{j=1}^{n_i} \left[\frac{\tau(1-\tau)}{\varrho} \exp\left\{ -\rho\left(\frac{Y_{ij} - \eta^T X_{ij} - \delta^T H_i(t)}{\varrho} \right) \right\} \right] \\
& \times \prod_{k=1}^{K_1} \left[\exp\left\{ - \sum_{t_{kl} \le L_{ik}} e^{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k b_i} \lambda_{kl} \right\} \right. \\
& \qquad \left. - I(R_{ik} < \infty) \exp\left\{ - \sum_{t_{kl} \le R_{ik}} e^{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k b_i} \lambda_{kl} \right\} \right] \\
& \times \prod_{k=K_1+1}^{K} \left[\left\{ e^{\beta_k^T W_{ik} + \alpha_k \delta^T H_i(D_{ik}) + \zeta_k b_i} \Lambda_k(D_{ik}) \right\}^{\Delta_{ik}} \right. \\
& \qquad \left. \times \exp\left\{ - \sum_{t_{kl} \le D_{ik}} e^{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k b_i} \lambda_{kl} \right\} \right] \\
& \times f_b(b_i) db_i.
\end{aligned}
\tag{3.6}
$$

where W_{ikl} is the vector of covariates at time t_{kl} for the ith individual for $l = 1, \ldots, m_k$, $k = 1, \ldots, K$; W_{ik} is the vector of covariates at time D_{ik} and $\Lambda_k(D_{ik})$ is the jump size

of Λ_k at time D_{ik} for the ith individual for $k = K_1 + 1, \ldots, K$.

Direct maximization of (3.6) is difficult, specifically, when we take the logarithms of the likelihoods for asymptomatic event times which involve subtractions of two exponential functions. To make the maximization feasible, we apply a Poisson process to derive a likelihood equivalent to the objective function. For $k = 1, \ldots, K_1$, we denote $N_{ik}(t)$ as a non-homogeneous Poisson process with intensity function the same as the hazard function of T_{ik}. For the ith subject, if $R_{ik} = \infty$, no event of type k occur before L_{ik}, and the likelihood is

$$\exp\left\{-\sum_{t_{kl}\le L_{ik}} e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k b_i}\lambda_{kl}\right\}.$$

If $R_{ik} < \infty$, no event of type k occur before L_{ik} but occur in interval $(L_{ik}, R_{ik}]$, and the likelihood is

$$\begin{aligned}&\exp\left\{-\sum_{t_{kl}\le L_{ik}} e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k b_i}\lambda_{kl}\right\}\\ &\times\left(1-\exp\left\{-\sum_{L_{ik}<t_{kl}\le R_{ik}} e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k b_i}\lambda_{kl}\right\}\right).\end{aligned}$$

We can write the above two likelihoods in a general form as

$$\begin{aligned}&\exp\left\{-\sum_{t_{kl}\le L_{ik}} e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k b_i}\lambda_{kl}\right\}\\ &-I(R_{ik}<\infty)\exp\left\{-\sum_{t_{kl}\le R_{ik}} e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k b_i}\lambda_{kl}\right\}\end{aligned}\tag{3.7}$$

which is the same as the asymptomatic part in the objective function (3.6). Based on this fact, we let P_{ikl} ($l = 1, \ldots, m_k$, $t_{kl} \le R^*_{ik}$, $R^*_{ik} = I(R_{ik} = \infty)L_{ik} + I(R_{ik} < \infty)R_{ik}$) be independent Poisson random variables with means $\lambda_{kl}\exp\{\beta_k^T W_{ikl} + \alpha_k\delta^T H_i(t_{kl}) + \zeta_k b_i\}$. Therefore, we propose a MCEM algorithm treating P_{ikl} as latent variables, and

we work on $f_P(P_{ikl})$ defined as the likelihood function of $\{P_{ikl}, l = 1, \ldots, m_k, t_{kl} \leq R^*_{ik}\}$ with the following expression

$$\prod_{l=1, t_{kl} \leq R^*_{ik}}^{m_k} \left\{ \frac{1}{P_{ikl}!} \left(\lambda_{kl} e^{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k b_i} \right)^{P_{ikl}} \exp\left(-\lambda_{kl} e^{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k b_i} \right) \right\}.$$

The likelihood for $A_{ik} = \sum_{t_{kl} < L_{ik}} P_{ikl} = 0$ and $B_{ik} = I(R_{ik} < \infty) \sum_{L_{ik} < t_{kl} \leq R_{ik}} P_{ikl} > 0$ given b_i is equal to (3.7).

Based on the above, we obtain the following complete-data log-likelihood

$$\begin{aligned}
l_c(\boldsymbol{\theta}, \mathcal{A}) =& \sum_{i=1}^{n} \Big[\sum_{j=1}^{n_i} \log f_Y(Y_{ij}) + \sum_{k=1}^{K_1} \log f_P(P_{ikl}) + \sum_{k=K_1+1}^{K} \log f_k(D_{ik}) + \log f_b(b_i) \Big] \\
=& -\log \varrho \sum_{i=1}^{n} n_i - \sum_{i=1}^{n} \sum_{j=1}^{n_i} \rho \left(\frac{Y_{ij} - \eta^T X_{ij} - \delta^T H_i(t)}{\varrho} \right) \\
&+ \sum_{i=1}^{n} \sum_{k=1}^{K_1} \sum_{l=1}^{m_k} I(t_{kl} \leq R^*_{ik}) \Big[P_{ikl} (\log \lambda_{kl} + \beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k b_i) \\
&\quad - \lambda_{kl} \exp\{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k b_i\} \Big] \\
&+ \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \Big[\Delta_{ik} \{\log \Lambda_k(D_{ik}) + \beta_k^T W_{ik} + \alpha_k \delta^T H_i(D_{ik}) + \zeta_k b_i\} \\
&\quad - \sum_{t_{kl} \leq D_{ik}} \lambda_{kl} \exp\{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k b_i\} \Big] \\
&+ \sum_{i=1}^{n} \log f_b(b_i). \qquad (3.8)
\end{aligned}$$

3.3.1 Monte Carlo E-step

The conditional expected value of the complete-data log-likelihood function in (3.8) is

$$
\begin{aligned}
& E[l_c(\boldsymbol{\theta}, \mathcal{A})] \\
=& \sum_{i=1}^{n} E\Big[\sum_{j=1}^{n_i} \log f_Y(Y_{ij}) + \sum_{k=1}^{K_1} \sum_{l=1}^{m_k} \log f_P(P_{ikl}) + \sum_{k=K_1+1}^{K} \log f_k(D_{ik}) + \log f_b(b_i) \Big] \\
=& \sum_{i=1}^{n} \int_{b_i \in \mathbb{R}} \Big[\sum_{j=1}^{n_i} \log f_Y(Y_{ij}) + \sum_{k=1}^{K_1} \sum_{l=1}^{m_k} \log f_P(P_{ikl}) + \sum_{k=K_1+1}^{K} \log f_k(D_{ik}) + \log f_b(b_i) \Big] \\
& \times f_b(b_i \mid \mathcal{O}_i, \boldsymbol{\theta}^r, \mathcal{A}^r) db_i,
\end{aligned}
\tag{3.9}
$$

where the expectation is obtained given the observed data $\mathcal{O}_i (i = 1, \ldots, n)$, with $\mathcal{O}_i = \{Y_{ij}, X_{ij}, H_i(\cdot) : j = 1, \ldots, n_i\} \bigcup \{A_{ik} = 0, B_{ik} > 0, W_{ik} : k = 1, \ldots, K_1\} \bigcup \{D_{ik}, \Delta_{ik}, W_{ik} : k = K_1 + 1, \ldots, K\}$, and $(\boldsymbol{\theta}^r, \mathcal{A}^r)$ denote the values of the parameters at the current rth iteration of the algorithm. The posterior distribution of the random effect b_i given $\mathcal{O}_i$ is proportional to the complete joint likelihood for the ith subject, that is

$$
f_b(b_i \mid \mathcal{O}_i, \boldsymbol{\theta}^r, \mathcal{A}^r) \propto \prod_{j=1}^{n_i} f_Y(Y_{ij}) \prod_{k=1}^{K_1} [S_k(L_{ik}) - S_k(R_{ik})] \prod_{k=K_1+1}^{K} f_k(D_{ik}) f_b(b_i) \Big|_r,
$$

where $|_r$ means that the computations are based on estimates of parameters at the rth iteration.

We propose a Monte Carlo simulation based on a Uniform distribution to approximate the integrals over the random effects in (3.9). At each iteration, we draw a simple random sample $(c_{i1}, \ldots, c_{im})$ from a Uniform distribution with a possible support $(-a, a)$ and assign each sample point a weight, where the weights are calculated

by the following

$$w_{if} = \frac{f_b(c_{if} \mid \mathcal{O}_i, \boldsymbol{\theta}^r, \mathcal{A}^r)}{\sum_{f=1}^{m} f_b(c_{if} \mid \mathcal{O}_i, \boldsymbol{\theta}^r, \mathcal{A}^r)}$$

for $f = 1, \ldots, m$. By doing this, we try to use $\{(c_{i1}, w_{i1}), \ldots, (c_{im}, w_{im})\}$ to mimic the entire conditional posterior distribution of b_i given the observed data and thus to approximate the expectation of $l_c(\boldsymbol{\theta}, \mathcal{A})$ at the rth iteration. The approximation is

$$\begin{aligned}
E[l_c(\boldsymbol{\theta}, \mathcal{A})] \approx & -\log \varrho \sum_{i=1}^{n} n_i - \sum_{i=1}^{n} \sum_{j=1}^{n_i} \rho\left(\frac{Y_{ij} - \eta^T X_{ij} - \delta^T H_i(t)}{\varrho}\right) \\
& + \sum_{i=1}^{n} \sum_{k=1}^{K_1} \sum_{l=1}^{m_k} \sum_{f=1}^{m} w_{if} I(t_{kl} \le R_{ik}^*) \Big[\hat{E}[P_{ikl}](\log \lambda_{kl} + \beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) \\
& \quad + \zeta_k c_{if}) - \lambda_{kl} \exp\{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k c_{if}\} \Big] \\
& + \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \sum_{f=1}^{m} w_{if} \Big[\Delta_{ik} \{\log \Lambda_k(D_{ik}) + \beta_k^T W_{ik} + \alpha_k \delta^T H_i(D_{ik}) + \zeta_k c_{if}\} \\
& \quad - \sum_{t_{kl} \le D_{ik}} \lambda_{kl} \exp\{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k c_{if}\} \Big] \\
& + \sum_{i=1}^{n} \sum_{f=1}^{m} w_{if} \log f_b(c_{if}),
\end{aligned} \tag{3.10}$$

where the conditional expectation of P_{ikl} given $\mathcal{O}_i$ and b_i is

$$\hat{E}[P_{ikl}] = I(L_{ik} < t_{kl} \le R_{ik} < \infty) \frac{\lambda_{kl} \exp\{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k b_i\}}{1 - \exp\{-\sum_{L_{ik} < t_{kl'} \le R_{ik}} \lambda_{kl'} e^{\beta_k^T W_{ikl'} + \alpha_k \delta^T H_i(t_{kl'}) + \zeta_k b_i}\}}.$$

Remark 2. The support range $(-a, a)$ of the Uniform distribution used in the Monte Carlo simulations can be determined by fitting separate traditional Cox models for all events with only fixed effects. For $k = 1, \ldots, K_1$, we assume $I(R_{ik} < \infty) = \hat{\Lambda}_{ik} e^{e_{ik}}$ where $\hat{\Lambda}_{ik}$ denotes the estimated cumulative hazard of the kth event for the ith subject using traditional Cox model and $e^{e_{ik}}$ denotes the exponential contribution of some underlying random effects. We know that the martingale residual is defined

as $r_{mik} = I(R_{ik} < \infty) - \hat{\Lambda}_{ik}$, thus we have

$$\begin{aligned} e_{ik} &= log(I(R_{ik} < \infty)/\hat{\Lambda}_{ik}) \\ &= log\left(\frac{I(R_{ik} < \infty)}{I(R_{ik} < \infty) - r_{mik}}\right) \end{aligned}$$

Similarly, for $k = K_1 + 1, \ldots, K$, we have

$$e_{ik} = log\left(\frac{\Delta_{ik}}{\Delta_{ik} - r_{mik}}\right)$$

Since we use $e^{e_{ik}}$ to approximate $e^{\zeta_k b_i}$ and set the initial value of ζ_k to 1, we choose a positive number of a such that $(-a, a)$ covers a majority of $\{e_{ik} : i = 1, \ldots, n, k = 1, \ldots, K\}$ (e.g., a is three times of the standard deviation of e_{ik}'s). Also, this can be used to set the initial value of σ^2 as the variance of e_{ik}'s at the first iteration.

3.3.2 M-step

In the M-step, we maximize the approximated conditional expectation of the complete-data log-likelihood. By zeroizing the differentiations of $E[l_c(\boldsymbol{\theta}, \mathcal{A})]$ in (3.10) with respect to each type of parameters, we obtain the estimating equations for corresponding parameters. Differentiating (3.10) with respect to ϱ and λ_{kl}, we update ϱ and λ_{kl}, dependent on the other parameters, with explicit expressions:

$$\varrho = \frac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \rho(Y_{ij} - \eta^T X_{ij} - \delta^T H_i(t)), \tag{3.11}$$

$$\lambda_{kl} = \frac{\sum_{i=1}^{n} \sum_{f=1}^{m} w_{if} I(t_{kl} \leq R_{ik}^*) \hat{E}[P_{ikl}]}{\sum_{i=1}^{n} \sum_{f=1}^{m} w_{if} I(t_{kl} \leq R_{ik}^*) e^{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k c_{if}}} \tag{3.12}$$

for $k = 1, \ldots, K_1$, $l = 1, \ldots, m_k$, and

$$\lambda_{kl} = \frac{\sum_{i=1}^{n} \Delta_{ik} I(D_{ik} = t_{kl})}{\sum_{i=1}^{n} \sum_{f=1}^{m} I(D_{ik} \geq t_{kl}) w_{if} e^{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k c_{if}}} \tag{3.13}$$

for $k = K_1+1, \ldots, K$, $l = 1, \ldots, m_k$. We then update the other parameters through a one-step Newton-Raphson algorithm by using the updated estimates in (3.11), (3.12) and (3.13). Specifically, the estimating equation for $\boldsymbol{\eta}$ is

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} \psi_\tau(Y_{ij} - \eta^T X_{ij} - \delta^T H_i(t)) X_{ij} = 0$$

where $\psi_\tau(s) = \rho'(s) = \tau - I(s < 0)$ is the derivative of the quantile loss function. To make Newton-Raphson method useable, Lu and Fan (2015) derived an equivalent smoothed estimating equation

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} X_{ij}(\tau - 1 + \Phi(\epsilon_{ij}/r_{ij}^X)) = 0,$$

where $\Phi(\cdot)$ is the cumulative probability function of the standard normal, $r_{ij}^X = \sqrt{X_{ij}^T \Omega_\eta X_{ij}}$ and Ω_η is an estimate of the covariance matrix of parameter $\boldsymbol{\eta}$. The corresponding second order derivative of (3.10) with respect to $\boldsymbol{\eta}$ then follows as

$$-\sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{\phi(\epsilon_{ij}/r_{ij}^X)}{r_{ij}^X} X_{ij} X_{ij}^T$$

where $\phi(\cdot)$ is the standard normal density function. The objective function for $\boldsymbol{\delta}$ is

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n_i}\psi_\tau(Y_{ij}-\eta^T X_{ij}-\delta^T H_i(t))H_i(t)\\
&+\sum_{i=1}^{n}\sum_{k=1}^{K_1}\sum_{l=1}^{m_k}\sum_{f=1}^{m} w_{if}I(t_{kl}\le R^*_{ik})\{\hat{E}[P_{ikl}]-\lambda_{kl}e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}}\}\alpha_k H_i(t_{kl})\\
&-\sum_{i=1}^{n}\sum_{k=K_1+1}^{K}\sum_{f=1}^{m} w_{if}\left(\sum_{t_{kl}\le D_{ik}}\lambda_{kl}e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}}\alpha_k H_i(t_{kl})\right)\\
&+\sum_{i=1}^{n}\sum_{k=K_1+1}^{K}\Delta_{ik}\alpha_k H_i(t_{kl}),
\end{aligned}$$

followed by the second order derivative of (3.10) with respect to $\boldsymbol{\delta}$ as

$$\begin{aligned}
&-\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{\phi(\epsilon_{ij}/r_{ij}^H)}{r_{ij}^H}H_i(t)H_i^T(t)\\
&-\sum_{i=1}^{n}\sum_{k=1}^{K_1}\sum_{l=1}^{m_k}\sum_{f=1}^{m} w_{if}I(t_{kl}\le R^*_{ik})\lambda_{kl}e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}}\alpha_k^2 H_i(t_{kl})H_i^T(t_{kl})\\
&-\sum_{i=1}^{n}\sum_{k=K_1+1}^{K}\sum_{f=1}^{m} w_{if}\left(\sum_{t_{kl}\le D_{ik}}\lambda_{kl}e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}}\alpha_k^2 H_i(t_{kl})H_i^T(t_{kl})\right).
\end{aligned}$$

where $r_{ij}^H=\sqrt{H_i^T(t)\Omega_\delta H_i(t)}$ and Ω_δ is an estimate of the covariance matrix of parameter $\boldsymbol{\delta}$. For $k=1,\ldots,K_1$, the objective function for $\boldsymbol{\beta}_k$ is

$$\sum_{i=1}^{n}\sum_{l=1}^{m_k}\sum_{f=1}^{m} w_{if}I(t_{kl}\le R^*_{ik})\{\hat{E}[P_{ikl}]-\lambda_{kl}e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}}\}W_{ikl}$$

followed by the second order derivative of (3.10) with respect to $\boldsymbol{\beta}_k$ as

$$-\sum_{i=1}^{n}\sum_{l=1}^{m_k}\sum_{f=1}^{m} w_{if}I(t_{kl}\le R^*_{ik})\lambda_{kl}e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}}W_{ikl}W_{ikl}^T.$$

For $k = K_1 + 1, \ldots, K$, the objective function for $\boldsymbol{\beta}_k$ is

$$-\sum_{i=1}^{n}\sum_{f=1}^{m} w_{if}\left(\sum_{t_{kl}\le D_{ik}} \lambda_{kl} e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}} W_{ikl}\right) + \sum_{i=1}^{n}\Delta_{ik} W_{ikl}$$

followed by the second order derivative of (3.10) with respect to $\boldsymbol{\beta}_k$ as

$$-\sum_{i=1}^{n}\sum_{f=1}^{m} w_{if}\left(\sum_{t_{kl}\le D_{ik}} \lambda_{kl} e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}} W_{ikl} W_{ikl}^T\right).$$

For $k = 1, \ldots, K_1$, the objective function for α_k is

$$\sum_{i=1}^{n}\sum_{l=1}^{m_k}\sum_{f=1}^{m} w_{if} I(t_{kl} \le R_{ik}^*)\{\hat{E}[P_{ikl}] - \lambda_{kl} e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}}\}\delta^T H_i(t_{kl})$$

followed by the second order derivative of (3.10) with respect to α_k as

$$-\sum_{i=1}^{n}\sum_{l=1}^{m_k}\sum_{f=1}^{m} w_{if} I(t_{kl} \le R_{ik}^*)\lambda_{kl} e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}}\delta^T H_i(t_{kl}) H_i^T(t_{kl})\delta.$$

For $k = K_1 + 1, \ldots, K$, the objective function for α_k is

$$-\sum_{i=1}^{n}\sum_{f=1}^{m} w_{if}\left(\sum_{t_{kl}\le D_{ik}} \lambda_{kl} e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}}\delta^T H_i(t_{kl})\right) + \sum_{i=1}^{n}\Delta_{ik}\delta^T H_i(t_{kl})$$

followed by the second order derivative of (3.10) with respect to α_k as

$$-\sum_{i=1}^{n}\sum_{f=1}^{m} w_{if}\left(\sum_{t_{kl}\le D_{ik}} \lambda_{kl} e^{\beta_k^T W_{ikl}+\alpha_k\delta^T H_i(t_{kl})+\zeta_k c_{if}}\delta^T H_i(t_{kl}) H_i^T(t_{kl})\delta\right)$$

For $k = 1, \ldots, K_1$, the objective function for ζ_k is

$$\sum_{i=1}^{n}\sum_{l=1}^{m_k}\sum_{f=1}^{m} w_{if} I(t_{kl} \leq R_{ik}^*)\{\hat{E}[P_{ikl}] - \lambda_{kl} e^{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k c_{if}}\} c_{if}$$

followed by the second order derivative of (3.10) with respect to ζ_k as

$$-\sum_{i=1}^{n}\sum_{l=1}^{m_k}\sum_{f=1}^{m} w_{if} I(t_{kl} \leq R_{ik}^*) \lambda_{kl} e^{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k c_{if}} c_{if}^2.$$

For $k = K_1 + 1, \ldots, K$, the objective function for ζ_k is

$$\begin{aligned} &-\sum_{i=1}^{n}\sum_{f=1}^{m} w_{if} \left(\sum_{t_{kl} \leq D_{ik}} \lambda_{kl} e^{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k c_{if}} c_{if} \right) \\ &+\sum_{i=1}^{n} \Delta_{ik} c_{if} \end{aligned}$$

followed by the second order derivative of (3.10) with respect to ζ_k as

$$-\sum_{i=1}^{n}\sum_{f=1}^{m} w_{if} \left(\sum_{t_{kl} \leq D_{ik}} \lambda_{kl} e^{\beta_k^T W_{ikl} + \alpha_k \delta^T H_i(t_{kl}) + \zeta_k c_{if}} c_{if}^2 \right).$$

Finally, the variance of the random effect b_i can be updated by the weighted empirical variance of $\{c_{i1}, \ldots, c_{im}, i = 1, \ldots, n\}$, that is

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{f=1}^{m} w_{if} c_{if}^2$$

We iterate between the Monte Carlo E-step and the M-step until convergence. The final non-parametric maximum likelihood estimators for $\boldsymbol{\theta}$ and $\mathcal{A}$ are denoted as $\hat{\boldsymbol{\theta}}$ and $\hat{\mathcal{A}}$.

3.4 Asymptotics

In this section, we establish and prove the consistency and asymptotic normality of the non-parametric maximum likelihood estimators, $\hat{\boldsymbol{\theta}}$ and $\hat{\mathcal{A}}$, for our proposed joint model of quantiles of longitudinal observations and multiple right- and interval-censored survival times. For $k = 1, \ldots, K_1$, let $\mathfrak{I}_k$ be the support for monitoring times with the least upper bound $\mathfrak{t}_k$. For $k = K_1 + 1, \ldots, K$, let $\mathfrak{I}_k = [0, \mathfrak{t}_k]$ where $\mathfrak{t}_k$ is the study duration time. We assume throughout that the following regularity conditions are satisfied:

C1. For any subject i, the number of longitudinal measurements n_i is bounded and $sup_i\|X_i\| < \infty$ with the dimension of X_{ij} fixed as p_1, where $\|\cdot\|$ is the Euclidean norm. The cumulative distribution function $F_{ij}(z) = P(Y_{ij} - \eta^T X_{ij} - \delta^T H_i(t) \leq z | X_{ij}, H_i(t))$ is absolutely continuous with continuous density f_{ij} and its first derivative being uniformly bounded away from 0 and ∞ at zero, for $j = 1, \ldots, n_i$.

C2. With probability 1, $H_i(\cdot)$ and W_{ik} have bounded variation in $\mathfrak{I}_k$ with fixed dimensions p_2 and p_3. If there exists a deterministic function $a_1(t)$ and constant vectors a_2, a_3 such that $a_1(t) + a_2^T W_{ik} + a_3^T H_i(t) = 0$ with probability 1, then $a_1(t) = 0$, $a_2 = 0$, and $a_3 = 0$ for any t in $\mathfrak{I}_k$ and $k = 1, \ldots, K$.

C3. The true value of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}_0 = (\boldsymbol{\eta}_0, \boldsymbol{\delta}_0, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{\zeta}_0, \varrho_0, \sigma_0^2)$, are interior points of known compact sets of the space $\Theta = \mathfrak{E} \times \mathfrak{D} \times \mathfrak{B} \times \mathfrak{A} \times \mathfrak{Z} \times \mathfrak{R} \times \mathfrak{S}$, where $\mathfrak{E} \subset \mathbb{R}^{p_1}$, $\mathfrak{D} \subset \mathbb{R}^{p_2}$, $\mathfrak{B} \subset \mathbb{R}^{p_3}$, $\mathfrak{A} \subset \mathbb{R}^4$, $\mathfrak{Z} \subset \mathbb{R}^4$, $\mathfrak{R} \subset (0, \infty)$, and $\mathfrak{S} \subset (0, \infty)$.

C4. For any $X_i = (X_{i1}^T, \ldots, X_{in_i}^T)$ and $H_i = (H_i^T(t_1), \ldots, H_i^T(t_{n_i}))$, the following conditions are satisfied:

(a) For any positive definite matrix Σ_i, $\frac{1}{m}\sum_{i=1}^m X_i^T \Sigma_i \Gamma_i X_i$ converges to a positive definite matrix; where Γ_i is an $n_i \times n_i$ diagonal matrix with the jth

diagonal element $f_{ij}(0)$.

(b) For any positive definite matrix Σ_i, $\frac{1}{m}\sum_{i=1}^{m} H_i^T \Sigma_i \Gamma_i H_i$ converges to a positive definite matrix.

C5. For $k = 1, \ldots, K$, the true value of $\Lambda_k(\cdot)$, denoted by $\Lambda_{k0}(\cdot)$, is strictly increasing and continuously differentiable with positive derivatives in $\mathfrak{I}$, and $\Lambda_{k0}(0) = 0$.

C6. For $k = 1, \ldots, K_1$, the number of possible monitoring times M_{ik} is bigger than zero with a finite mean. For two monitoring times next to each other, say $I_{ik,m}$ and $I_{ik,m+1}$, $P\{\min_{0 \le m < m_k}(I_{ik,m+1} - I_{ik,m}) \ge \varphi | M_{ik}, W_{ik}, H_i\} = 1$ for some positive constant φ. In addition, there exists a probability measure μ_k in $\mathfrak{I}_k$ such that the conditional bivariate density of $(I_{ik,m}, I_{ik,m+1})$ is dominated by $\mu_k \times \mu_k$ and its Radon-Nikodym derivative, $\tilde{f}_{km}(u, v | M_{ik}, W_{ik}, H_i)$, can be expanded to a positive and twice continuously differentiable function with respect to u and v when $v - u \ge \varphi$.

C7. For $k = K_1 + 1, \ldots, K$, $P(C_k \ge \mathfrak{t}_k | W_{ik}, H_i) = P(C_k = \mathfrak{t}_k | W_{ik}, H_i) \ge \vartheta$ for some positive constant ϑ.

Remark 3. Condition C3 is standard assumptions for parameters in the joint regression models of longitudinal and time-to-event data. Conditions C1 and C4 are required for the consistency and convergence of estimators of the parameters using quantile regression models for longitudinal data. Condition C6 ensures that the distance between any two adjacent monitoring times is at least φ, resulting in no exact observations for asymptomatic events. Condition C7 allows a positive probability for the kth symptomatic event to be observed during the study and that some individuals are still at risk of the kth event at the end $\mathfrak{t}_k$ of the study.

For $k = 1, \ldots, K$, we denote $\boldsymbol{W}_{ik} = E_{1k} \times (W_{i1}^T, \ldots, W_{iK}^T)^T$, $\boldsymbol{H}_{ik} = E_{2k} \times (\delta^T H_i(t), \ldots, \delta^T H_i(t))^T$, and $\boldsymbol{b}_{ik} = E_{2k} \times (b_i, \ldots, b_i)^T$, where E_{1k} is a diagonal matrix

with diagonal elements $e_k \otimes \mathbf{1}_{p_3}$, E_{2k} is a diagonal matrix with diagonal elements e_k, e_k is the kth canonical vector in $\mathbb{R}^K$, $\otimes$ represents the Kronecker product, and $\mathbf{1}_{p_3}$ denotes a p_3 dimensional vector with all ones. Therefore, the proposed hazard function for the kth event time thus be

$$\lambda_k(t; \tilde{Q}_{\tau t}, \boldsymbol{W}_{ik}, \boldsymbol{b}_{ik}) = e^{\boldsymbol{\beta}^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}^T \boldsymbol{b}_{ik}} \lambda_{k0}(t).$$

3.4.1 Consistency

Let $\|\cdot\|_{\infty(\mathfrak{I}_k)}$ denote the supremum norm on $\mathfrak{I}_k$, $\|\cdot\|$ be the Euclidean norm, $\mathbb{P}_n$ represent the empirical measure for n independent individuals, $\mathbb{P}$ denote the true probability measure, and $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - \mathbb{P})$ be the empirical process. We first state and prove the almost sure (a.s.) consistency of $(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})$. Let $(\boldsymbol{\theta}_0, \mathcal{A}_0)$ be the true values of the parameters. We have the following theorem.

Theorem 3.4.1. *Under regularity conditions $C1-C7$, as $n \to \infty$, the non-parametric maximum likelihood estimator $(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})$ is consistent. That is, $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \to_{a.s.} 0$, and $\|\hat{\Lambda}_k - \Lambda_{k0}\|_{\infty(\mathfrak{I}_k)} \to_{a.s.} 0$ for $k = 1, \ldots, K$.*

The consistency in Theorem 3.4.1 can be proved by verifying the following steps: First, we show that the non-parametric maximum likelihood estimators of $\boldsymbol{\theta}$ and $\mathcal{A}$ exist. Second, we prove that, $\limsup_n \hat{\Lambda}_k(\mathfrak{t}_k - \xi) < \infty$ $(k = 1, \ldots, K_1)$ with probability 1 for any $\xi > 0$ and $\limsup_n \hat{\Lambda}_k(\mathfrak{t}_k) < \infty$ $(k = K_1 + 1, \ldots, K)$ with probability 1. Third, based on the second step, we select a subsequence of $\hat{\Lambda}_k$ such that $\hat{\Lambda}_k$ converges to some right-continuous monotone function Λ_{k*} with probability 1. We can choose a sub-subsequence and further assume that $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_*$. By showing that $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$ and $\mathcal{A}_* = \mathcal{A}_0$, we conclude that $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_0$ and $\hat{\mathcal{A}}$ converges to $\mathcal{A}_0$.

Proof of Theorem 3.4.1. We first prove the existence of the NPMLEs, $\hat{\boldsymbol{\theta}}$ and $\hat{\mathcal{A}}$. For

a specified value of quantile $0 < \tau < 1$ and any $(\boldsymbol{\theta}, \mathcal{A})$ in Θ, the ith term of the likelihood for longitudinal response is

$$\Psi_i(\mathcal{O}_i; \boldsymbol{\eta}, \boldsymbol{\delta}, \varrho) = \prod_{j=1}^{n_i} \frac{\tau(1-\tau)}{\varrho} \exp\left\{-\rho\left(\frac{Y_{ij} - \eta^T X_{ij} - \delta^T H_i(t)}{\varrho}\right)\right\} \tag{3.14}$$

Since $\rho(s) = s(\tau - I(s < 0)) \geq 0$ and ϱ is a scale parameter greater than zero, the likelihood in (3.14) is less than or equal to $\Upsilon_i = \{\tau(1-\tau)/\varrho\}^{n_i} < \infty$. According to conditions C1 and C3, there exists a constant $\widehat{M}$ such that $\Upsilon_i \leq \widehat{M}$ for any $i \subset (1, \ldots, n)$. Let $\widetilde{M} = \sum_{k=1}^{K} \sup_{t \in \mathfrak{I}_k} \sup_{\boldsymbol{W}_k, \boldsymbol{H}_k, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\alpha}, \boldsymbol{\zeta}} \{|\boldsymbol{\beta}^T \boldsymbol{W}_k| + |\boldsymbol{\alpha}^T \boldsymbol{H}_k| + |\zeta_k|\}$ which is finite under Conditions C2 and C3. For the joint likelihood $L_n(\boldsymbol{\theta}, \mathcal{A})$, the ith integrand is bounded by

$$\begin{aligned}
&\widehat{M} \prod_{k=1}^{K_1} \left[\exp\left\{ - \int_0^{L_{ik}} e^{\boldsymbol{\beta}^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}^T \boldsymbol{b}_{ik}} d\Lambda_k(s) \right\} \right. \\
&\quad \left. - \exp\left\{ - \int_0^{R_{ik}} e^{\boldsymbol{\beta}^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}^T \boldsymbol{b}_{ik}} d\Lambda_k(s) \right\} \right] \\
&\times \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}^T \boldsymbol{b}_{ik}} \Lambda_k(D_{ik}) \right\}^{\Delta_{ik}} \right. \\
&\quad \left. \times \exp\left\{ - \int_0^{D_{ik}} e^{\boldsymbol{\beta}^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}^T \boldsymbol{b}_{ik}} d\Lambda_k(s) \right\} \right] \\
&\times f_b(b_i; \sigma^2).
\end{aligned} \tag{3.15}$$

Since, for $k = 1, \ldots, K_1$,

$$\exp\left\{ - \int_0^{L_{ik}} e^{\boldsymbol{\beta}^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}^T \boldsymbol{b}_{ik}} d\Lambda_k(s) \right\} - \exp\left\{ - \int_0^{R_{ik}} e^{\boldsymbol{\beta}^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}^T \boldsymbol{b}_{ik}} d\Lambda_k(s) \right\}$$

$$\leq 1,$$

the ith integrand of the joint likelihood in (3.15) is further bounded by

$$\widehat{M}\prod_{K_1+1}^{K}\left[\left(\Lambda_k\{D_{ik}\}e^{\widetilde{M}|b_i|}\right)^{\Delta_{ik}}\left\{1+\int_0^{D_{ik}}e^{\boldsymbol{\beta}^T\boldsymbol{W}_{ik}+\boldsymbol{\alpha}^T\boldsymbol{H}_{ik}+\boldsymbol{\zeta}^T\boldsymbol{b}_{ik}}d\Lambda_k(s)\right\}^{-\Delta_{ik}}\right]f_b(b_i;\sigma^2).$$

Therefore, the joint likelihood attains the maximum only if the values of $\hat{\Lambda}_k$ ($k = K_1+1,\ldots,K$) were finite, and there exists a non-parametric maximum likelihood estimator $(\hat{\boldsymbol{\theta}},\hat{\mathcal{A}})$ by allowing $\hat{\Lambda}_k(\mathfrak{t}_k)=\infty$ ($k=1,\ldots,K_1$).

The next step is to show that the above conditions for the existence of $(\hat{\boldsymbol{\theta}},\hat{\mathcal{A}})$ are satisfied. That is, with probability one, $\hat{\Lambda}_k(\mathfrak{t}_k)$ is bounded as $n\to\infty$ for $k = K_1+1,\ldots,K$, and $\hat{\Lambda}_k(\mathfrak{t}_k-\xi)$ is bounded as $n\to\infty$ for any $\xi>0$ when $k = 1,\ldots,K_1$. Otherwise, as proved latter, there will be a contradiction to the fact that $L_n(\hat{\boldsymbol{\theta}},\hat{\mathcal{A}})-L_n(\boldsymbol{\theta},\mathcal{A})\geq 0$ for any $(\boldsymbol{\theta},\mathcal{A})$ in the parameter space Θ.

For $k=1,\ldots,K_1$, we let $\tilde{\Lambda}_k$ be a step function with $\tilde{\Lambda}_k(t_{kl})=\Lambda_{k0}(t_{kl})$ such that it converges uniformly to Λ_{k0} for $l=1,\ldots,m_k$. For $k=K_1+1,\ldots,K$, we define a function $\tilde{\Lambda}_k$ as the solution of the following equation which is constructed by differentiating the joint likelihood $L_n(\boldsymbol{\theta},\mathcal{A})$ with respect to $\Lambda_k(D_{ik})$ and setting it to be equal to zero. That is, $\tilde{\Lambda}_k$ ($k=K_1+1,\ldots,K$) satisfies

$$\begin{aligned}\frac{\Delta_{ik}}{\tilde{\Lambda}_k(D_{ik})}&=\sum_{j=1}^{n}\frac{\int_b \boldsymbol{g}_1\left(b,\mathcal{O}_j;\boldsymbol{\theta}_0,\mathcal{A}_0\right)\boldsymbol{g}_{2k}\left(D_{ik},b,\mathcal{O}_j;\boldsymbol{\theta}_0,\mathcal{A}_0\right)f_b(b;\sigma_0^2)db\Psi(\mathcal{O}_j;\boldsymbol{\theta}_0)}{\int_b \boldsymbol{g}_1\left(b,\mathcal{O}_j;\boldsymbol{\theta}_0,\mathcal{A}_0\right)f_b(b;\sigma_0^2)db\Psi(\mathcal{O}_j;\boldsymbol{\theta}_0)}\\&=\sum_{j=1}^{n}\frac{\int_b \boldsymbol{g}_1\left(b,\mathcal{O}_j;\boldsymbol{\theta}_0,\mathcal{A}_0\right)\boldsymbol{g}_{2k}\left(D_{ik},b,\mathcal{O}_j;\boldsymbol{\theta}_0,\mathcal{A}_0\right)f_b(b;\sigma_0^2)db}{\int_b \boldsymbol{g}_1\left(b,\mathcal{O}_j;\boldsymbol{\theta}_0,\mathcal{A}_0\right)f_b(b;\sigma_0^2)db},\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{g}_1\left(b,\mathcal{O};\boldsymbol{\theta},\mathcal{A}\right)=&\prod_{k=1}^{K_1}\Big[\exp\Big\{-\int_0^{L_k}e^{\boldsymbol{\beta}^T\boldsymbol{W}_k+\boldsymbol{\alpha}^T\boldsymbol{H}_k+\boldsymbol{\zeta}^T\boldsymbol{b}_k}d\Lambda_k(s)\Big\}\\&-\exp\Big\{-\int_0^{R_k}e^{\boldsymbol{\beta}^T\boldsymbol{W}_k+\boldsymbol{\alpha}^T\boldsymbol{H}_k+\boldsymbol{\zeta}^T\boldsymbol{b}_k}d\Lambda_k(s)\Big\}\Big]\\\times&\prod_{k=K_1+1}^{K}\Big[\Big\{e^{\boldsymbol{\beta}^T\boldsymbol{W}_k+\boldsymbol{\alpha}^T\boldsymbol{H}_k+\boldsymbol{\zeta}^T\boldsymbol{b}_k}\Lambda_k(D_k)\Big\}^{\Delta_k}\exp\Big\{-\int_0^{D_k}e^{\boldsymbol{\beta}^T\boldsymbol{W}_k+\boldsymbol{\alpha}^T\boldsymbol{H}_k+\boldsymbol{\zeta}^T\boldsymbol{b}_k}d\Lambda_k(s)\Big\}\Big],\end{aligned}$$

and

$$\boldsymbol{g}_{2k}\left(t,b,\mathcal{O};\boldsymbol{\theta},\mathcal{A}\right)=I(D_k\geq t)e^{\boldsymbol{\beta}^T\boldsymbol{W}_k+\boldsymbol{\alpha}^T\boldsymbol{H}_k+\boldsymbol{\zeta}^T\boldsymbol{b}_k}.$$

By Lemma 1 of Gao et al. (2018), the following classes of functions

$$\mathcal{C}_1=\left\{\int_b\boldsymbol{g}_1\left(b,\mathcal{O};\boldsymbol{\theta},\mathcal{A}\right)f_b(b;\sigma^2)db:\boldsymbol{\theta}\in\Theta,\mathcal{A}\in\mathcal{D}_1\right\}\tag{3.16}$$

and

$$\mathcal{C}_2=\left\{\int_b\boldsymbol{g}_1\left(b,\mathcal{O};\boldsymbol{\theta},\mathcal{A}\right)\boldsymbol{g}_{2k}\left(t,b,\mathcal{O};\boldsymbol{\theta},\mathcal{A}\right)f_b(b;\sigma^2)db:\boldsymbol{\theta}\in\Theta,t\in\mathfrak{I}_k,\mathcal{A}\in\mathcal{D}_1\right\}\tag{3.17}$$

for $k=K_1+1,\ldots,K$ are Glivenko-Cantelli, where $\mathcal{D}_1=\mathcal{D}_{1,\infty}\times,\ldots,\mathcal{D}_{K_1,\infty}\times\mathcal{D}_{K_1+1,M}\times,\ldots,\mathcal{D}_{K,M}$, $\mathcal{D}_{k,c}=\{\Lambda:\Lambda\text{ is increasing with }\Lambda(0)=0\text{ and }\Lambda(\mathfrak{t}_k)\leq c\}$ and M is a finite constant. Therefore, $\tilde{\Lambda}_k$, as an expression formed by Glivenko-Cantelli functions, converges to the true value Λ_{k0} uniformly as $n\to\infty$ for $k=K_1+1,\ldots,K$. We write $\tilde{\mathcal{A}}=(\tilde{\Lambda}_1,\ldots,\tilde{\Lambda}_K)$.

Let $l_n(\boldsymbol{\theta},\mathcal{A})=\log L_n(\boldsymbol{\theta},\mathcal{A})$ be the joint log likelihood. By definition and the fact

that $e^{-|x|}(1+y) \leq 1 + e^x y \leq e^{|x|}(1+y)$, we have

$$
\begin{aligned}
0 \leq & n^{-1} l_n(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) - n^{-1} l_n(\boldsymbol{\theta}_0, \tilde{\mathcal{A}}) \\
\leq & \mathcal{M}_1 + n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \log \left\{ n \hat{\Lambda}_k(D_{ik}) \right\} \\
& + n^{-1} \sum_{i=1}^{n} \left[\log \int_{b_i} \prod_{k=K_1+1}^{K} \left\{ \frac{e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{W}_{ik} + \hat{\boldsymbol{\alpha}}^T \boldsymbol{H}_{ik} + \hat{\boldsymbol{\zeta}}^T \boldsymbol{b}_{ik}}}{1 + \int_0^{D_{ik}} e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{W}_{ik} + \hat{\boldsymbol{\alpha}}^T \boldsymbol{H}_{ik} + \hat{\boldsymbol{\zeta}}^T \boldsymbol{b}_{ik}} d\hat{\Lambda}_k(s)} \right\}^{\Delta_{ik}} f_b(b_i; \hat{\sigma}^2) db_i \right] \\
\leq & \mathcal{M}_1 + n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \log \left\{ n \hat{\Lambda}_k(D_{ik}) \right\} \\
& + n^{-1} \sum_{i=1}^{n} \left[\log \int_{b_i} \prod_{k=K_1+1}^{K} \left\{ \frac{e^{\widetilde{M} \|\boldsymbol{b}_{ik}\|}}{e^{-\widetilde{M} \|\boldsymbol{b}_{ik}\|} \left\{ 1 + \hat{\Lambda}_k(D_{ik}) \right\}} \right\}^{\Delta_{ik}} f_b(b_i; \hat{\sigma}^2) db_i \right] \\
\leq & \mathcal{M}_2 + n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \log \left\{ n \hat{\Lambda}_k(D_{ik}) \right\} - n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \left[\Delta_{ik} \log \left\{ 1 + \hat{\Lambda}_k(D_{ik}) \right\} \right],
\end{aligned}
$$

where $\mathcal{M}_1$ and $\mathcal{M}_2$ are constants. Applying the partitioning method of Murphy (1994), we choose a sequence of time $u_{k0} = \mathfrak{t}_k > u_{k1} > \cdots > u_{k,Q_k} = 0$ and have the following inequality

$$
\begin{aligned}
& n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \log \left\{ n \hat{\Lambda}_k(D_{ik}) \right\} - n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \left[\Delta_{ik} \log \left\{ 1 + \hat{\Lambda}_k(D_{ik}) \right\} \right] \\
\leq & \sum_{k=K_1+1}^{K} \sum_{q=0}^{Q_k - 1} n^{-1} \sum_{i=1}^{n} I\left(D_{ik} \in [u_{k,q+1}, u_{kq})\right) \log \left\{ n \hat{\Lambda}_k(D_{ik}) \right\} \\
& - \sum_{k=K_1+1}^{K} n^{-1} \sum_{i=1}^{n} I\left(D_{ik} = \mathfrak{t}_k\right) \Delta_{ik} \log \left\{ 1 + \hat{\Lambda}_k(\mathfrak{t}_k) \right\} \\
& - \sum_{k=K_1+1}^{K} \sum_{q=0}^{Q_k - 1} n^{-1} \sum_{i=1}^{n} \Delta_{ik} I\left(D_{ik} \in [u_{k,q+1}, u_{kq})\right) \log \left\{ 1 + \hat{\Lambda}_k(u_{k,q+1}) \right\}
\end{aligned}
$$

which is further bounded by

$$\begin{aligned}
&\mathcal{M}_3 - (2n)^{-1} \sum_{k=K_1+1}^{K} \sum_{i=1}^{n} I\left(D_{ik} = \mathfrak{t}_k\right) \Delta_{ik} \log\left\{1 + \hat{\Lambda}_k(\mathfrak{t}_k)\right\} \\
&- \sum_{K_1+1}^{K} \left\{ (2n)^{-1} \sum_{i=1}^{n} I\left(D_{ik} = \mathfrak{t}_k\right) \Delta_{ik} - n^{-1} \sum_{i=1}^{n} \Delta_{ik} I\left(D_{ik} \in [u_{k1}, u_{k0})\right) \right\} \log\left\{1 + \hat{\Lambda}_k(\mathfrak{t}_k)\right\} \\
&- \sum_{K_1+1}^{K} \sum_{q=1}^{Q_k-1} \left\{ n^{-1} \sum_{i=1}^{n} I\left(D_{ik} \in [u_{kq}, u_{k,q-1})\right) \Delta_{ik} - n^{-1} \sum_{i=1}^{n} \Delta_{ik} I\left(D_{ik} \in [u_{k,q+1}, u_{kq})\right) \right\} \\
&\times \log\left\{1 + \hat{\Lambda}_k(u_{kq})\right\},
\end{aligned}$$

where $\mathcal{M}_3$ is a constant. Using a selected sequence of $\{u_{kq}; q = 0, \ldots, Q_k\}$, we can make all the coefficients of $\log\left\{1 + \hat{\Lambda}_k(u_{kq})\right\}$ terms negative such that $n^{-1} l_n(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) - n^{-1} l_n(\boldsymbol{\theta}_0, \tilde{\mathcal{A}}) \leq \mathcal{M}_2 + \mathcal{M}_3$. However, if $\hat{\Lambda}_k(\mathfrak{t}_k)$ diverges to ∞, then $\log\left\{1 + \hat{\Lambda}_k(\mathfrak{t}_k)\right\}$ diverges to ∞ and $n^{-1} l_n(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) - n^{-1} l_n(\boldsymbol{\theta}_0, \tilde{\mathcal{A}})$ diverges to $-\infty$, which is a contradiction. Hence, we conclude that $\limsup_n \hat{\Lambda}_k(\mathfrak{t}_k) < \infty$ for $k = K_1 + 1, \ldots, K$.

Now, we let $\tilde{\mathcal{A}}^* = (\tilde{\Lambda}_1, \ldots, \tilde{\Lambda}_{K_1}, \hat{\Lambda}_{K_1+1}, \ldots, \hat{\Lambda}_K)$ and $d_{ikm} = I(I_{ik,m} < T_{ik} \leq I_{ik,m+1})$ for $i = 1, \ldots, n$, $k = 1, \ldots, K_1$, $m = 0, \ldots, M_{ik}$, where $I_{ik,M_{ik}+1} = \infty$. Then, we have the following inequalities

$$\begin{aligned}
0 \leq & n^{-1} l_n(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) - n^{-1} l_n(\boldsymbol{\theta}_0, \tilde{\mathcal{A}}^*) \\
\leq & \mathcal{M}_4 + n^{-1} \sum_{i=1}^{n} \left(\log \int_{b_i} \prod_{k=1}^{K_1} \left[\exp\left\{ -e^{\widetilde{M}\|\boldsymbol{b}_{ik}\|} \hat{\Lambda}_k(I_{ik,M_{ik}}) \right\} \right]^{d_{ik,M_{ik}}} f_b(b_i; \hat{\sigma}^2) db_i \right) \\
\leq & \mathcal{M}_4 + n^{-1} \sum_{i=1}^{n} \left(\log \int_{\|b_i\| \leq 1} \prod_{k=1}^{K_1} \left[\exp\left\{ -e^{\widetilde{M}\|\boldsymbol{b}_{ik}\|} \hat{\Lambda}_k(I_{ik,M_{ik}}) \right\} \right]^{d_{ik,M_{ik}}} f_b(b_i; \hat{\sigma}^2) db_i \right) \\
& + n^{-1} \sum_{i=1}^{n} \left(\log \int_{\|b_i\| \geq 1} f_b(b_i; \hat{\sigma}^2) db_i \right) \\
\leq & \mathcal{M}_5 - \mathcal{M}_6 n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_1} d_{ik,M_{ik}} e^{\widetilde{M}} \hat{\Lambda}_k(I_{ik,M_{ik}}),
\end{aligned}$$

where $\mathcal{M}_4$, $\mathcal{M}_5$ and $\mathcal{M}_6$ are positive constants. If $\limsup_n \hat{\Lambda}_k(\mathfrak{t}_k - \xi) = \infty$, then

$e^{\widetilde{M}}\hat{\Lambda}_k(\mathfrak{t}_k - \xi) = \infty$ indicating a contradiction. Therefore, $\limsup_n \hat{\Lambda}_k(\mathfrak{t}_k - \xi) < \infty$ with probability one for any $\xi > 0$ and $k = 1, \ldots, K_1$. From Helly's selection lemma, along a selected sequence of ξ decreasing to 0, $\hat{\Lambda}_k \to \Lambda_{k*}$ point wise in $\mathfrak{I}_k$ and $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_*$. We denote $\mathcal{A}_* = (\Lambda_{1*}, \ldots, \Lambda_{K*})$.

Proof of the theorem will be completed if we can show that $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$ and $\mathcal{A}_* = \mathcal{A}_0$. For $k = K_1 + 1, \ldots, K$, since $\tilde{\Lambda}_k$ is defined by imitating $\hat{\Lambda}_k$, $\hat{\Lambda}_k(t)$ is absolutely continuous with respect to $\tilde{\Lambda}_k$, and

$$\hat{\Lambda}_k(t) = \int_0^t \frac{\mathbb{P}_n \nu_k(s, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)}{|\mathbb{P}_n \nu_k(s, \mathcal{O}; \hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})|} d\tilde{\Lambda}_k(s), \tag{3.18}$$

where

$$\nu_k(t, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) = \frac{\int_b \boldsymbol{g}_1(b, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A})\, \boldsymbol{g}_{2k}(D_k, b, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A})\, f_b(b; \sigma^2) db}{\int_b \boldsymbol{g}_1(b, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A})\, f_b(b; \sigma^2) db},$$

Following from the Glivenko-Cantelli property of the function classes $\mathcal{C}_1$ and $\mathcal{C}_2$, we have that

$$\sup_{t \in \mathfrak{I}_k} |\mathbb{P}_n \nu_k(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0) - \mathbb{P}\nu_k(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)| \to_{a.s.} 0$$

and

$$\sup_{t \in \mathfrak{I}_k} |\mathbb{P}_n \nu_k(t, \mathcal{O}; \hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) - \mathbb{P}\nu_k(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*)| \to_{a.s.} 0.$$

Further, for any $\xi > 0$,

$$\limsup_n \hat{\Lambda}_k(\mathfrak{t}_k) \geq \int_0^{\mathfrak{t}_k} \frac{\mathbb{P}\nu_k(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)}{\xi + |\mathbb{P}\nu_k(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*)|} d\Lambda_{k0}(t).$$

By letting $\xi \to 0$, it follows from the Monotone Convergence Theorem that

$$\int_0^{\mathfrak{t}_k} \frac{\mathbb{P}\nu_k(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)}{|\mathbb{P}\nu_k(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*)|} d\Lambda_{k0}(t) < \infty.$$

Thus, we conclude that $\min_{t \in \mathfrak{I}_k} |\mathbb{P}\nu_k(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*)| > 0$. We further take the limits

on both sides of (3.18) and have that $\Lambda_{k*}(t)$ is absolutely continuous with respect to $\Lambda_{k0}(t)$ meaning that $\Lambda_{k*}(t)$ is differentiable with respect to t. Moreover, $d\hat{\Lambda}_k(t)/d\tilde{\Lambda}_k(t)$ converges to $d\Lambda_{k*}(t)/d\Lambda_{k0}(t)$.

We define a function

$$m(\boldsymbol{\theta}, \mathcal{A}) = \log \left\{ \frac{L_i(\boldsymbol{\theta}, \mathcal{A}) + L_i(\boldsymbol{\theta}_0, \tilde{\mathcal{A}})}{2} \right\},$$

where $L_i(\boldsymbol{\theta}, \mathcal{A})$ is the likelihood function for the ith subject and $i = 1, \ldots, n$. The class of functions $m(\boldsymbol{\theta}, \mathcal{A})$ is Glivenko-Cantelli. By the concavity of the log function,

$$\mathbb{P}_n m(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) \geq \frac{1}{2} \left\{ \mathbb{P}_n \log L_i(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) + \mathbb{P}_n \log L_i(\boldsymbol{\theta}_0, \tilde{\mathcal{A}}) \right\} \geq \mathbb{P}_n l_{ni}(\boldsymbol{\theta}_0, \tilde{\mathcal{A}}) = \mathbb{P}_n m(\boldsymbol{\theta}_0, \tilde{\mathcal{A}}),$$

Thus, it follows that

$$\begin{aligned}
0 \leq & \mathbb{P}_n m(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) - \mathbb{P}_n m(\boldsymbol{\theta}_0, \tilde{\mathcal{A}}) \\
= & \mathbb{P} \left\{ m(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) - m(\boldsymbol{\theta}_0, \tilde{\mathcal{A}}) \right\} + o_P(1) \\
= & \mathbb{P} \log \left[\frac{L_i(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) + L_i(\boldsymbol{\theta}_0, \tilde{\mathcal{A}})}{2 L_i(\boldsymbol{\theta}_0, \tilde{\mathcal{A}})} \right] + o_P(1) \\
= & \mathbb{P} \log \left[\frac{1}{2} + \frac{\prod_{K_1+1}^{K} \hat{\Lambda}_k(D_{ik})^{\Delta_{ik}} \int_{b_i} \boldsymbol{g}_1 \left(b_i, \mathcal{O}; \hat{\boldsymbol{\theta}}, \hat{\mathcal{A}} \right) f_b(b_i; \hat{\sigma}^2) db_i \Psi_i(\mathcal{O}; \hat{\boldsymbol{\theta}})}{2 \prod_{K_1+1}^{K} \tilde{\Lambda}_k(D_{ik})^{\Delta_{ik}} \int_{b_i} \boldsymbol{g}_1 \left(b_i, \mathcal{O}; \boldsymbol{\theta}_0, \tilde{\mathcal{A}} \right) f_b(b_i; \sigma_0^2) db_i \Psi_i(\mathcal{O}; \boldsymbol{\theta}_0)} \right] + o_P(1) \\
\to & \mathbb{P} \left\{ \log \left[\frac{1}{2} + \frac{\prod_{K_1+1}^{K} \Lambda_{k*}(D_{ik})^{\Delta_{ik}} \int_{b_i} \boldsymbol{g}_1 \left(b_i, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_* \right) f_b(b_i; \sigma_*^2) db_i \Psi_i(\mathcal{O}; \boldsymbol{\theta}_*)}{2 \prod_{K_1+1}^{K} \Lambda_{k0}(D_{ik})^{\Delta_{ik}} \int_{b_i} \boldsymbol{g}_1 \left(b_i, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0 \right) f_b(b_i; \sigma_0^2) db_i \Psi_i(\mathcal{O}; \boldsymbol{\theta}_0)} \right] \right\},
\end{aligned}$$

indicating a positive value of the negative Kullback-Leibler information. Thus, we

have

$$
\begin{aligned}
&\Psi_i(\mathcal{O}_i;\boldsymbol{\eta}_*,\boldsymbol{\delta}_*,\varrho_*)\int_{b_i}\prod_{k=1}^{K_1}\left(\sum_{m=0}^{M_{ik}} d_{ikm}\left[\exp\left\{-\int_0^{I_{ikm}} e^{\boldsymbol{\beta}_*^T\boldsymbol{W}_{ik}+\boldsymbol{\alpha}_*^T\boldsymbol{H}_{ik}+\boldsymbol{\zeta}_*^T\boldsymbol{b}_{ik}}d\Lambda_{k*}(s)\right\}\right.\right.\\
&\qquad\left.\left.-\exp\left\{-\int_0^{I_{ik,m+1}} e^{\boldsymbol{\beta}_*^T\boldsymbol{W}_{ik}+\boldsymbol{\alpha}_*^T\boldsymbol{H}_{ik}+\boldsymbol{\zeta}_*^T\boldsymbol{b}_{ik}}d\Lambda_{k*}(s)\right\}\right]\right)\\
&\quad\times\prod_{k=K_1+1}^{K}\left[\left\{e^{\boldsymbol{\beta}_*^T\boldsymbol{W}_{ik}+\boldsymbol{\alpha}_*^T\boldsymbol{H}_{ik}+\boldsymbol{\zeta}_*^T\boldsymbol{b}_{ik}}\Lambda_{k*}(D_{ik})\right\}^{\Delta_{ik}}\right.\\
&\qquad\left.\times\exp\left\{-\int_0^{D_{ik}} e^{\boldsymbol{\beta}_*^T\boldsymbol{W}_{ik}+\boldsymbol{\alpha}_*^T\boldsymbol{H}_{ik}+\boldsymbol{\zeta}_*^T\boldsymbol{b}_{ik}}d\Lambda_{k*}(s)\right\}\right]f_b(b_i;\sigma_*^2)db_i\\
=\;&\Psi_i(\mathcal{O}_i;\boldsymbol{\eta}_0,\boldsymbol{\delta}_0,\varrho_0)\int_{b_i}\prod_{k=1}^{K_1}\left(\sum_{m=0}^{M_{ik}} d_{ikm}\left[\exp\left\{-\int_0^{I_{ikm}} e^{\boldsymbol{\beta}_0^T\boldsymbol{W}_{ik}+\boldsymbol{\alpha}_0^T\boldsymbol{H}_{ik}+\boldsymbol{\zeta}_0^T\boldsymbol{b}_{ik}}d\Lambda_{k0}(s)\right\}\right.\right.\\
&\qquad\left.\left.-\exp\left\{-\int_0^{I_{ik,m+1}} e^{\boldsymbol{\beta}_0^T\boldsymbol{W}_{ik}+\boldsymbol{\alpha}_0^T\boldsymbol{H}_{ik}+\boldsymbol{\zeta}_0^T\boldsymbol{b}_{ik}}d\Lambda_{k0}(s)\right\}\right]\right)\\
&\quad\times\prod_{k=K_1+1}^{K}\left[\left\{e^{\boldsymbol{\beta}_0^T\boldsymbol{W}_{ik}+\boldsymbol{\alpha}_0^T\boldsymbol{H}_{ik}+\boldsymbol{\zeta}_0^T\boldsymbol{b}_{ik}}\Lambda_{k0}(D_{ik})\right\}^{\Delta_{ik}}\right.\\
&\qquad\left.\times\exp\left\{-\int_0^{D_{ik}} e^{\boldsymbol{\beta}_0^T\boldsymbol{W}_{ik}+\boldsymbol{\alpha}_0^T\boldsymbol{H}_{ik}+\boldsymbol{\zeta}_0^T\boldsymbol{b}_{ik}}d\Lambda_{k0}(s)\right\}\right]f_b(b_i;\sigma_0^2)db_i
\end{aligned}
$$

with probability one, where $d_{ikm} = I(I_{ikm} < T_{ik} \leq I_{ik,m+1})$ for $k = 1,\ldots,K_1$ and $m = 0,\ldots,M_{ik}$ with $I_{ik,M_{ik}+1} = \infty$. For any k from 1 to K_1 and m from 0 to M_{ik}, we let $d_{ikm'}$ equals 0 for $m' < m$ and equals 1 for $m' \geq m$. By taking the sum of $M_{ik}+1$

terms of likelihoods for $k = 1, \ldots, K_1$, we obtain

$$\begin{aligned}
&\Psi_i(\mathcal{O}_i; \boldsymbol{\eta}_*, \boldsymbol{\delta}_*, \varrho_*) \int_{b_i} \prod_{k=1}^{K_1} \exp\left\{ - \int_0^{I_{ikm}} e^{\boldsymbol{\beta}_*^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_*^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_*^T \boldsymbol{b}_{ik}} d\Lambda_{k*}(s) \right\} \\
&\quad \times \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_*^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_*^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_*^T \boldsymbol{b}_{ik}} \Lambda_{k*}(D_{ik}) \right\}^{\Delta_{ik}} \right. \\
&\quad \left. \times \exp\left\{ - \int_0^{D_{ik}} e^{\boldsymbol{\beta}_*^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_*^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_*^T \boldsymbol{b}_{ik}} d\Lambda_{k*}(s) \right\} \right] f_b(b_i; \sigma_*^2) db_i \\
= \quad & \Psi_i(\mathcal{O}_i; \boldsymbol{\eta}_0, \boldsymbol{\delta}_0, \varrho_0) \int_{b_i} \prod_{k=1}^{K_1} \exp\left\{ - \int_0^{I_{ikm}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} d\Lambda_{k0}(s) \right\} \\
&\quad \times \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} \Lambda_{k0}(D_{ik}) \right\}^{\Delta_{ik}} \right. \\
&\quad \left. \times \exp\left\{ - \int_0^{D_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} d\Lambda_{k0}(s) \right\} \right] f_b(b_i; \sigma_0^2) db_i.
\end{aligned}$$

Because m is randomly chosen, we can use any t_{ik} in $\mathfrak{I}_k$ to replace I_{ikm}. We then let $\Delta_{ik} = 1$ for all $k \in \{K_1 + 1, \ldots, K\}$. By integrating D_{ik} from 0 to t_{ik}, we have

$$\begin{aligned}
&\Psi_i(\mathcal{O}_i; \boldsymbol{\eta}_*, \boldsymbol{\delta}_*, \varrho_*) \int_{b_i} \exp\left\{ - \sum_{k=1}^{K_1} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_*^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_*^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_*^T \boldsymbol{b}_{ik}} d\Lambda_{k*}(s) \right. \\
&\quad \left. - \sum_{k=K_1+1}^{K} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_*^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_*^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_*^T \boldsymbol{b}_{ik}} d\Lambda_{k*}(s) \right\} f_b(b_i; \sigma_*^2) db_i \\
= \quad & \Psi_i(\mathcal{O}_i; \boldsymbol{\eta}_0, \boldsymbol{\delta}_0, \varrho_0) \int_{b_i} \exp\left\{ - \sum_{k=1}^{K_1} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} d\Lambda_{k0}(s) \right. \\
&\quad \left. - \sum_{k=K_1+1}^{K} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} d\Lambda_{k0}(s) \right\} f_b(b_i; \sigma_0^2) db_i,
\end{aligned}$$

which can be further written as

$$\begin{aligned}
&\Psi_i(\mathcal{O}_i; \boldsymbol{\eta}_*, \boldsymbol{\delta}_*, \varrho_*) \int_{b_i} \exp\left\{ - \sum_{k=1}^{K} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_*^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_*^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_*^T \boldsymbol{b}_{ik}} d\Lambda_{k*}(s) \right\} f_b(b_i; \sigma_*^2) db_i \\
= \quad & \Psi_i(\mathcal{O}_i; \boldsymbol{\eta}_0, \boldsymbol{\delta}_0, \varrho_0) \int_{b_i} \exp\left\{ - \sum_{k=1}^{K} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} d\Lambda_{k0}(s) \right\} f_b(b_i; \sigma_0^2) db_i.
\end{aligned} \tag{3.19}$$

By letting $t_{ik} = 0$ for all $k = 1, \ldots, K$ in (3.19), we obtain the following equation

$$\Psi_i(\mathcal{O}_i; \boldsymbol{\eta}_*, \boldsymbol{\delta}_*, \varrho_*) = \Psi_i(\mathcal{O}_i; \boldsymbol{\eta}_0, \boldsymbol{\delta}_0, \varrho_0). \tag{3.20}$$

Using the proof arguments of Theorem 3.1 in Lu and Fan (2015), we conclude that $\boldsymbol{\eta}_* = \boldsymbol{\eta}_0$, $\boldsymbol{\delta}_* = \boldsymbol{\delta}_0$, and $\varrho_* = \varrho_0$. For any $k \in \{1, \ldots, K\}$, we let $t_{ik'} = 0$ for all $k' \neq k$ in (3.19). Along with the result in (3.20), we have the following equation

$$\begin{aligned} & \int_{b_i} \exp\left\{ - e^{\boldsymbol{\zeta}_*^T \boldsymbol{b}_{ik}} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_*^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_*^T \boldsymbol{H}_{ik}} d\Lambda_{k*}(s) \right\} f_b(b_i; \sigma_*^2) db_i \\ = & \int_{b_i} \exp\left\{ - e^{\boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik}} d\Lambda_{k0}(s) \right\} f_b(b_i; \sigma_0^2) db_i. \end{aligned}$$

Following the results of Theorem 1 in Elbers and Ridder (1982), we can conclude that $\sigma_*^2 = \sigma_0^2$, $\boldsymbol{\zeta}_* = \boldsymbol{\zeta}_0$, and

$$\int_0^{t_{ik}} e^{\boldsymbol{\beta}_*^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_*^T \boldsymbol{H}_{ik}} d\Lambda_{k*}(s) = \int_0^{t_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik}} d\Lambda_{k0}(s). \tag{3.21}$$

By differentiating both sides of the above equation with respect to t_{ik} and then taking the logarithm, we have the following equation

$$\boldsymbol{\beta}_*^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_*^T \boldsymbol{H}_{ik} + \lambda_{k*}(t_{ik}) = \boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \lambda_{k0}(t_{ik}). \tag{3.22}$$

By condition C2, we conclude that $\boldsymbol{\beta}_* = \boldsymbol{\beta}_0$, $\boldsymbol{\alpha}_* = \boldsymbol{\alpha}_0$, and $\lambda_{k*}(t_{ik}) = \lambda_{k0}(t_{ik})$. Further, by letting $\boldsymbol{W}_{ik} = \boldsymbol{H}_{ik} = \boldsymbol{0}$, we obtain that $\Lambda_{k*}(t_{ik}) = \Lambda_{k0}(t_{ik})$ for $t_{ik} \in \mathfrak{I}_k$ and $k = 1, \ldots, K$. The above implies that $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$ and $\mathcal{A}_* = \mathcal{A}_0$. We now have proved that $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \to 0$, $|\hat{\Lambda}_k(t) - \Lambda_{k0}(t)| \to 0$, and thus $\hat{\mathcal{A}}$ converges uniformly to $\mathcal{A}_0$ on $\prod_k \mathfrak{I}_k$. □

3.4.2 Asymptotic normality

The asymptotic normality of the non-parametric maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is stated in the following theorem.

Theorem 3.4.2. *Under regularity conditions $C1 - C7$, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges weakly to a multivariate zero-mean normal vector with a covariance matrix that attains the semi-parametric efficiency bound.*

Once the consistency of $(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})$ is proved, the normality of $\hat{\boldsymbol{\theta}}$ can be verified through empirical processes, the Taylor expansion of the score functions for $\hat{\boldsymbol{\theta}}$ and $\hat{\mathcal{A}}$ around the true parameters $\boldsymbol{\theta}_0$ and $\mathcal{A}_0$, and the Donsker properties of the score functions.

Proof of Theorem 3.4.2. For convenience, we process with individual (ith) terms of the joint log likelihood $l_n(\boldsymbol{\theta}, \mathcal{A})$ and define the following terms

$$\mathcal{H}_{ij\tau}(t, \mathcal{O}; \boldsymbol{\theta}) = \tau - I\left\{(Y_{ij} - \eta^T X_{ij} - \delta^T H_i(t)) < 0\right\},$$

$$\mathcal{H}_{i1k}(t, u, v, b_i, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) = \frac{\boldsymbol{g}_1\left(b_i, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}\right) \boldsymbol{q}_1(t, u, v, b_i, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) f_b(b_i; \sigma^2)}{\int_{b_i'} \boldsymbol{g}_1\left(b_i', \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}\right) f_b(b_i'; \sigma^2) db_i'}$$

for $k = 1, \ldots, K_1$, and

$$\mathcal{H}_{i2k}(t, b_i, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) = \frac{\boldsymbol{g}_1\left(b_i, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}\right) \boldsymbol{q}_2(t, D_{ik}, b_i, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) f_b(b_i; \sigma^2)}{\int_{b_i'} \boldsymbol{g}_1\left(b_i', \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}\right) f_b(b_i'; \sigma^2) db_i'}$$

for $k = K_1 + 1, \ldots, K$, where

$$
\begin{aligned}
&\boldsymbol{q}_1(t, u, v, b, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) \\
=&e^{\boldsymbol{\beta}^T \boldsymbol{W}_k + \boldsymbol{\alpha}^T \boldsymbol{H}_k + \boldsymbol{\zeta}^T \boldsymbol{b}_k} \\
&\times \left[\frac{I(v \geq t) \exp\left\{ - \int_0^v e^{\boldsymbol{\beta}^T \boldsymbol{W}_k + \boldsymbol{\alpha}^T \boldsymbol{H}_k + \boldsymbol{\zeta}^T \boldsymbol{b}_k} d\lambda_k(s) \right\}}{\exp\left\{ - \int_0^u e^{\boldsymbol{\beta}^T \boldsymbol{W}_k + \boldsymbol{\alpha}^T \boldsymbol{H}_k + \boldsymbol{\zeta}^T \boldsymbol{b}_k} d\lambda_k(s) \right\} - \exp\left\{ - \int_0^v e^{\boldsymbol{\beta}^T \boldsymbol{W}_k + \boldsymbol{\alpha}^T \boldsymbol{H}_k + \boldsymbol{\zeta}^T \boldsymbol{b}_k} d\lambda_k(s) \right\}} \right. \\
&\quad \left. - \frac{I(u \geq t) \exp\left\{ - \int_0^u e^{\boldsymbol{\beta}^T \boldsymbol{W}_k + \boldsymbol{\alpha}^T \boldsymbol{H}_k + \boldsymbol{\zeta}^T \boldsymbol{b}_k} d\lambda_k(s) \right\}}{\exp\left\{ - \int_0^u e^{\boldsymbol{\beta}^T \boldsymbol{W}_k + \boldsymbol{\alpha}^T \boldsymbol{H}_k + \boldsymbol{\zeta}^T \boldsymbol{b}_k} d\lambda_k(s) \right\} - \exp\left\{ - \int_0^v e^{\boldsymbol{\beta}^T \boldsymbol{W}_k + \boldsymbol{\alpha}^T \boldsymbol{H}_k + \boldsymbol{\zeta}^T \boldsymbol{b}_k} d\lambda_k(s) \right\}} \right],
\end{aligned}
$$

and

$$
\boldsymbol{q}_2(t, u, b, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) = -I(u \geq t) e^{\boldsymbol{\beta}^T \boldsymbol{W}_k + \boldsymbol{\alpha}^T \boldsymbol{H}_k + \boldsymbol{\zeta}^T \boldsymbol{b}_k}.
$$

By differentiating the log-likelihood function with respect to $\boldsymbol{\theta}$, we derive that the individual score function for parameter $\boldsymbol{\theta}$ is

$$
\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \mathcal{A}) = \left(\boldsymbol{l}_{\boldsymbol{\eta}}(\boldsymbol{\theta}, \mathcal{A})^T, \boldsymbol{l}_{\boldsymbol{\delta}}(\boldsymbol{\theta}, \mathcal{A})^T, \boldsymbol{l}_{\boldsymbol{\beta}}(\boldsymbol{\theta}, \mathcal{A})^T, \boldsymbol{l}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}, \mathcal{A})^T, \boldsymbol{l}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}, \mathcal{A})^T, l_{\varrho}(\boldsymbol{\theta}, \mathcal{A}), l_{\sigma^2}(\boldsymbol{\theta}, \mathcal{A}) \right)^T,
$$

where

$$
\boldsymbol{l}_{\boldsymbol{\eta}}(\boldsymbol{\theta}, \mathcal{A}) = \sum_{j=1}^{n_i} \mathcal{H}_{ij\tau}(t, \mathcal{O}; \boldsymbol{\theta}) X_{ij},
$$

$$
\begin{aligned}
\boldsymbol{l}_{\boldsymbol{\delta}}(\boldsymbol{\theta}, \mathcal{A}) =& \sum_{j=1}^{n_i} \mathcal{H}_{ij\tau}(t, \mathcal{O}; \boldsymbol{\theta}) H_i(t) \\
&+ \sum_{k=1}^{K_1} \sum_{m=0}^{M_{ik}} d_{ikm} \int_0^{\mathfrak{t}_k} \int_{b_i} \mathcal{H}_{i1k}(t, I_{ikm}, I_{ik,m+1}, b_i, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) db_i \alpha_k H_i(t) d\Lambda_k(s) \\
&+ \sum_{k=K_1+1}^{K} \left\{ \Delta_{ik} \alpha_k H_i(D_{ik}) + \int_0^{\mathfrak{t}_k} \int_{b_i} \mathcal{H}_{i2k}(t, b_i, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) db_i \alpha_k H_i(t) d\Lambda_k(s) \right\},
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{l}_{\boldsymbol{\beta}}(\boldsymbol{\theta}, \mathcal{A}) =& \sum_{k=1}^{K_1} \sum_{m=0}^{M_{ik}} d_{ikm} \int_0^{\mathfrak{t}_k} \int_{b_i} \mathcal{H}_{i1k}(t, I_{ikm}, I_{ik,m+1}, b_i, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) db_i \boldsymbol{W}_{ik} d\Lambda_k(s) \\
&+ \sum_{k=K_1+1}^{K} \left\{ \Delta_{ik} \boldsymbol{W}_{ik} + \int_0^{\mathfrak{t}_k} \int_{b_i} \mathcal{H}_{i2k}(t, b_i, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) db_i \boldsymbol{W}_{ik} d\Lambda_k(s) \right\},
\end{aligned}
$$

$$\boldsymbol{l}_{\boldsymbol{\alpha}}(\boldsymbol{\theta},\mathcal{A}) = \sum_{k=1}^{K_1}\sum_{m=0}^{M_{ik}} d_{ikm}\int_0^{\mathfrak{t}_k}\int_{b_i}\mathcal{H}_{i1k}(t,I_{ikm},I_{ik,m+1},b_i,\mathcal{O};\boldsymbol{\theta},\mathcal{A})db_i\boldsymbol{H}_{ik}d\Lambda_k(s)$$
$$+\sum_{k=K_1+1}^{K}\left\{\Delta_{ik}\boldsymbol{H}_{ik}+\int_0^{\mathfrak{t}_k}\int_{b_i}\mathcal{H}_{i2k}(t,b_i,\mathcal{O};\boldsymbol{\theta},\mathcal{A})db_i\boldsymbol{H}_{ik}d\Lambda_k(s)\right\},$$

$$\boldsymbol{l}_{\boldsymbol{\zeta}}(\boldsymbol{\theta},\mathcal{A}) = \sum_{k=1}^{K_1}\sum_{m=0}^{M_{ik}} d_{ikm}\int_0^{\mathfrak{t}_k}\int_{b_i}\mathcal{H}_{i1k}(t,I_{ikm},I_{ik,m+1},b_i,\mathcal{O};\boldsymbol{\theta},\mathcal{A})db_i\boldsymbol{b}_{ik}d\Lambda_k(s)$$
$$+\sum_{k=K_1+1}^{K}\left\{\Delta_{ik}\boldsymbol{b}_{ik}+\int_0^{\mathfrak{t}_k}\int_{b_i}\mathcal{H}_{i2k}(t,b_i,\mathcal{O};\boldsymbol{\theta},\mathcal{A})db_i\boldsymbol{b}_{ik}d\Lambda_k(s)\right\},$$

$$l_{\varrho}(\boldsymbol{\theta},\mathcal{A}) = \sum_{j=1}^{n_i}\left\{\varrho-\rho\left(Y_{ij}-\eta^T X_{ij}-\delta^T H_i(t)\right)\right\},$$

$$l_{\sigma^2}(\boldsymbol{\theta},\mathcal{A}) = \frac{\int_{b_i}\boldsymbol{g}_1\left(b_i,\mathcal{O};\boldsymbol{\theta},\mathcal{A}\right)f'_{b,\sigma^2}(b_i;\sigma^2)db_i}{\int_{b_i}\boldsymbol{g}_1\left(b_i,\mathcal{O};\boldsymbol{\theta},\mathcal{A}\right)f_b(b_i;\sigma^2)db_i},$$

and $f'_{b,\sigma^2}(b_i;\sigma^2)$ is the derivative of $f_b(b_i;\sigma^2)$ with respect to σ^2. Let $l_{\mathcal{A}}(\boldsymbol{\theta},\mathcal{A})(\boldsymbol{h})$ be the score operator for $\mathcal{A}$ along the submodel $d\mathcal{A}_{\epsilon,\boldsymbol{h}} = ((1+\epsilon h_1)d\Lambda_1,\ldots,(1+\epsilon h_K)d\Lambda_K)^T$ for $\boldsymbol{h} = (h_1,\ldots,h_K)$ where $h_k \in L_2(\mu_k)$ for $k = 1,\ldots,K_1$ and h_k is in the set of functions on $\mathfrak{I}_k$ with total variation bounded by one for $k = K_1+1,\ldots,K$. We have

$$l_{\mathcal{A}}(\boldsymbol{\theta},\mathcal{A})(\boldsymbol{h}) = \sum_{k=1}^{K_1}\sum_{m=0}^{M_{ik}} d_{ikm}\int_0^{\mathfrak{t}_k}\int_{b_i}\mathcal{H}_{i1k}(t,I_{ikm},I_{ik,m+1},b_i,\mathcal{O};\boldsymbol{\theta},\mathcal{A})db_i h_k(t)d\Lambda_k(s)$$
$$+\sum_{k=K_1+1}^{K}\left\{\Delta_{ik}h_k(D_{ik})+\int_0^{\mathfrak{t}_k}\int_{b_i}\mathcal{H}_{i2k}(t,b_i,\mathcal{O};\boldsymbol{\theta},\mathcal{A})db_i h_k(t)d\Lambda_k(s)\right\}.$$

We use the Taylor's expansions of the following two equations at the true value of parameters, $(\boldsymbol{\theta}_0,\mathcal{A}_0)$,

$$\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}},\hat{\mathcal{A}})\right\} = -\sqrt{n}\mathbb{P}\left\{l_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}},\hat{\mathcal{A}})-l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0,\mathcal{A}_0)\right\},$$

and

$$\mathbb{G}_n\left\{l_{\mathcal{A}}(\hat{\boldsymbol{\theta}},\hat{\mathcal{A}})(\boldsymbol{h})\right\} = -\sqrt{n}\mathbb{P}\left\{l_{\mathcal{A}}(\hat{\boldsymbol{\theta}},\hat{\mathcal{A}})(\boldsymbol{h})-l_{\mathcal{A}}(\boldsymbol{\theta}_0,\mathcal{A}_0)(\boldsymbol{h})\right\}.$$

Based on the properties in Lemma 5 of Gao et al. (2018), the second term of the expansions are bounded by

$$
\begin{aligned}
&O_P(1)\sqrt{n}E\Bigg[\sum_{k=1}^{K_1}\sum_{m=0}^{M_{ik}}\Big\{\hat{\Lambda}_k(I_{ikm})-\Lambda_{k0}(I_{ikm})\Big\}^2+\sum_{K_1+1}^{K}\Big\{\hat{\Lambda}_k(D_{ik})-\Lambda_{k0}(D_{ik})\Big\}^2\\
&+\|\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}_0\|^2+\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_0\|^2+\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2+\|\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|^2+\|\hat{\boldsymbol{\zeta}}-\boldsymbol{\zeta}_0\|^2\\
&+\|\hat{\varrho}-\varrho_0\|^2+\|\hat{\sigma}^2-\sigma_0\|^2\Bigg]\\
=&\sqrt{n}\Big[O_P(n^{-2/3})+O_P(1)\|\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}_0\|^2+O_P(1)\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_0\|^2+O_P(1)\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2\\
&+O_P(1)\|\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|^2+O_P(1)\|\hat{\boldsymbol{\zeta}}-\boldsymbol{\zeta}_0\|^2+O_P(1)\|\hat{\varrho}-\varrho_0\|^2+O_P(1)\|\hat{\sigma}^2-\sigma_0\|^2\Big]\\
=&O_P\Big(n^{-1/6}+\sqrt{n}\|\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}_0\|^2+\sqrt{n}\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_0\|^2+\sqrt{n}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2+\sqrt{n}\|\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|^2\\
&+\sqrt{n}\|\hat{\boldsymbol{\zeta}}-\boldsymbol{\zeta}_0\|^2+\sqrt{n}\|\hat{\varrho}-\varrho_0\|^2+\sqrt{n}\|\hat{\sigma}^2-\sigma_0\|^2\Big).
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}},\hat{\mathcal{A}})\right\}=&-\sqrt{n}\mathbb{P}\left\{l_{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+\boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}(\boldsymbol{h})(\hat{\mathcal{A}}-\mathcal{A}_0)\right\}\\
&+O_P\Big(n^{-1/6}+\sqrt{n}\|\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}_0\|^2+\sqrt{n}\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_0\|^2+\sqrt{n}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2\\
&+\sqrt{n}\|\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|^2+\sqrt{n}\|\hat{\boldsymbol{\zeta}}-\boldsymbol{\zeta}_0\|^2+\sqrt{n}\|\hat{\varrho}-\varrho_0\|^2+\sqrt{n}\|\hat{\sigma}^2-\sigma_0\|^2\Big),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{G}_n\left\{l_{\mathcal{A}}(\hat{\boldsymbol{\theta}},\hat{\mathcal{A}})(\boldsymbol{h})\right\}=&-\sqrt{n}\mathbb{P}\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h})\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h})\left(\hat{\mathcal{A}}-\mathcal{A}_0\right)\right\}\\
&+O_P\Big(n^{-1/6}+\sqrt{n}\|\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}_0\|^2+\sqrt{n}\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_0\|^2+\sqrt{n}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2\\
&+\sqrt{n}\|\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|^2+\sqrt{n}\|\hat{\boldsymbol{\zeta}}-\boldsymbol{\zeta}_0\|^2+\sqrt{n}\|\hat{\varrho}-\varrho_0\|^2+\sqrt{n}\|\hat{\sigma}^2-\sigma_0\|^2\Big),
\end{aligned}
$$

where $l_{\boldsymbol{\theta}\boldsymbol{\theta}}$, $\boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}(\boldsymbol{h})$,$l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h})$, and $l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h})$ are second order derivatives of $\boldsymbol{l}(\boldsymbol{\theta},\mathcal{A})$ and evaluated at $(\boldsymbol{\theta}_0,\mathcal{A}_0)$.

Using the arguments in the proof of Theorem 2 of Gao et al. (2018), we can show

that there exists a set of functions $\boldsymbol{h}^* = (\boldsymbol{h}_1^*, \ldots, \boldsymbol{h}_K^*)$ such that $l_{\mathcal{A}}^* l_{\mathcal{A}}(\boldsymbol{h}^*) = l_{\mathcal{A}}^* l_{\boldsymbol{\theta}}$ with $l_{\mathcal{A}}^*$ being the adjoint operator of $l_{\mathcal{A}}$. Further, $\boldsymbol{h}^*$ can be expanded to be a continuously differentiable function with bounded total variation. Thus, we have

$$
\begin{aligned}
&\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})\right\} - \mathbb{G}_n\left\{l_{\mathcal{A}}(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})(\boldsymbol{h}^*)\right\} \\
=&-\sqrt{n}\mathbb{P}\left\{\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + \boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}\left(\hat{\mathcal{A}} - \mathcal{A}_0\right)\right\} \\
&+\sqrt{n}\mathbb{P}\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h}^*)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h}^*)\left(\hat{\mathcal{A}} - \mathcal{A}_0\right)\right\} \\
&+O_P\Big(n^{-1/6} + \sqrt{n}\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\|^2 + \sqrt{n}\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0\|^2 + \sqrt{n}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 \\
&\quad + \sqrt{n}\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|^2 + \sqrt{n}\|\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\|^2 + \sqrt{n}\|\hat{\varrho} - \varrho_0\|^2 + \sqrt{n}\|\hat{\sigma}^2 - \sigma_0\|^2\Big) \\
=&\sqrt{n}\mathbb{P}\left[\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right) - l_{\mathcal{A}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right)\left(\boldsymbol{h}^*\right)\right\}^{\otimes 2}\right]\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \\
&+O_P\Big(n^{-1/6} + \sqrt{n}\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\|^2 + \sqrt{n}\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0\|^2 + \sqrt{n}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 \\
&\quad + \sqrt{n}\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|^2 + \sqrt{n}\|\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\|^2 + \sqrt{n}\|\hat{\varrho} - \varrho_0\|^2 + \sqrt{n}\|\hat{\sigma}^2 - \sigma_0\|^2\Big).
\end{aligned}
$$

Let $\mathcal{D}_2 = \mathcal{D}_{1,M}\times, \ldots, \mathcal{D}_{K,M}$, where $\mathcal{D}_{k,c} = \{\Lambda : \Lambda \text{ is increasing with } \Lambda(0) = 0 \text{ and } \Lambda(\mathfrak{t}_k) \leq c\}$ and M is a finite constant. By replacing $\mathcal{A} \in \mathcal{D}_1$ with $\mathcal{A} \in \mathcal{D}_2$ in (3.16) and (3.17), we can show that the function classes, $\mathcal{C}_1$ and $\mathcal{C}_2$, are Donsker (Van der Vaart and Wellner, 1996). Therefore, $\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right) - l_{\mathcal{A}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right)\left(\boldsymbol{h}^*\right)$ belongs to a Donsker class and

$$
\sup_{\boldsymbol{h}} \mathbb{P}\left\{l_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) - l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathcal{A}_0) + l_{\mathcal{A}}(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})(\boldsymbol{h}) - l_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{A}_0)(\boldsymbol{h})\right\}^2 \to 0.
$$

As a result of the Donsker property, $\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})\right\} - \mathbb{G}_n\left\{l_{\mathcal{A}}(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})(\boldsymbol{h}^*)\right\}$ converges to a multivariate zero-mean random variable.

Now, we need to show that the matrix

$$
\mathbb{P}\left[\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right) - l_{\mathcal{A}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right)\left(\boldsymbol{h}^*\right)\right\}^{\otimes 2}\right]
$$

is invertible which is equivalent to proving that the matrix is not singular. If the matrix is singular, there exists a non-zero vector $\boldsymbol{v} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5, v_6, v_7)^T$ in the real space with dimension corresponding to $\boldsymbol{\theta}$ such that $\boldsymbol{v}^T E[\{\boldsymbol{l}_{\boldsymbol{\theta}} - l_{\mathcal{A}}(\boldsymbol{h}^*)\}^{\otimes 2}]\boldsymbol{v} = 0$. That is, the score function along the submodel $\{\boldsymbol{\theta}_0 + \epsilon\boldsymbol{v}, \mathcal{A}_\epsilon(-\boldsymbol{v}^T\boldsymbol{h}^*)\}$ is zero with probability one, which leads to the following equation

$$
\begin{aligned}
&\sum_{j=1}^{n_i} \mathcal{H}_{ij\tau}(t, \mathcal{O}; \boldsymbol{\theta}_0) \left\{\boldsymbol{v}_1^T X_{ij} + \boldsymbol{v}_2^T H_i(t)\right\} + v_6 n_i \\
&+ \int_{b_i} \Bigg(\sum_{k=1}^{K_1} \sum_{m=0}^{M_{ik}} d_{ikm} \int_0^{\mathfrak{t}_k} \boldsymbol{q}_1(t, I_{ikm}, I_{ik,m+1}, b, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)\{\boldsymbol{v}_3^T \boldsymbol{W}_{ik} + \boldsymbol{v}_4^T \boldsymbol{H}_{ik} + \boldsymbol{v}_5^T \boldsymbol{b}_{ik} \\
&\quad - \boldsymbol{v}^T \boldsymbol{h}_k^*(s)\} d\Lambda_{k0}(s) + \sum_{k=K_1+1}^{K} \Delta_{ik} \left\{\boldsymbol{v}_3^T \boldsymbol{W}_{ik} + \boldsymbol{v}_4^T \boldsymbol{H}_{ik} + \boldsymbol{v}_5^T \boldsymbol{b}_{ik} - \boldsymbol{v}^T \boldsymbol{h}_k^*(D_{ik})\right\} \\
&\quad + \sum_{k=K_1+1}^{K} \int_0^{\mathfrak{t}_k} \boldsymbol{q}_2(t, D_{ik}, b, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)\{\boldsymbol{v}_3^T \boldsymbol{W}_{ik} + \boldsymbol{v}_4^T \boldsymbol{H}_{ik} + \boldsymbol{v}_5^T \boldsymbol{b}_{ik} - \boldsymbol{v}^T \boldsymbol{h}_k^*(s)\} d\Lambda_{k0}(s) \\
&\quad + v_7 \frac{f'_{b,\sigma^2}(b_i; \sigma^2)}{f_b(b_i; \sigma^2)} \Bigg) \boldsymbol{g}_1\left(b_i, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right) f_b(b_i; \sigma_0^2) db_i = 0
\end{aligned}
$$

with probability one. For any k from 1 to K_1 and m from 0 to M_{ik}, we let $d_{ikm'}$ equals 0 for $m' < m$ and equals 1 for $m' \geq m$. For $k = K_1 + 1, \ldots, K$, we set $\Delta_{ik} = 0$ and let $D_{ik} = t_{ik}$, where $t_{ik} \in \mathfrak{I}_k$. By taking the sum of all $M_{ik} + 1$ terms of d_{ikm} for $k = 1, \ldots, K_1$, we obtain

$$
\begin{aligned}
&\sum_{j=1}^{n_i} \mathcal{H}_{ij\tau}(t, \mathcal{O}; \boldsymbol{\theta}_0) \left\{\boldsymbol{v}_1^T X_{ij} + \boldsymbol{v}_2^T H_i(t)\right\} + v_6 n_i \\
&+ \int_{b_i} \Bigg(\sum_{k=1}^{K_1} \int_0^{I_{ikm}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} \{\boldsymbol{v}_3^T \boldsymbol{W}_{ik} + \boldsymbol{v}_4^T \boldsymbol{H}_{ik} + \boldsymbol{v}_5^T \boldsymbol{b}_{ik} - \boldsymbol{v}^T \boldsymbol{h}_k^*(s)\} d\Lambda_{k0}(s) \\
&\quad + \sum_{k=K_1+1}^{K} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} \{\boldsymbol{v}_3^T \boldsymbol{W}_{ik} + \boldsymbol{v}_4^T \boldsymbol{H}_{ik} + \boldsymbol{v}_5^T \boldsymbol{b}_{ik} - \boldsymbol{v}^T \boldsymbol{h}_k^*(s)\} d\Lambda_{k0}(s) \\
&\quad + v_7 \frac{f'_{b,\sigma^2}(b_i; \sigma^2)}{f_b(b_i; \sigma^2)} \Bigg) \prod_{k=1}^{K_1} \exp\left\{ - \int_0^{I_{ikm}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} d\Lambda_{k0}(s) \right\} \\
&\quad \times \prod_{k=K_1+1}^{K} \exp\left\{ - \int_0^{t_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} d\Lambda_{k0}(s) \right\} f_b(b_i; \sigma_0^2) db_i = 0
\end{aligned}
$$

Because m is randomly chosen, we can use any t_{ik} in $\mathfrak{I}_k$ to replace I_{ikm}. That is,

$$\sum_{j=1}^{n_i} \mathcal{H}_{ij\tau}(t, \mathcal{O}; \boldsymbol{\theta}_0) \left\{ \boldsymbol{v}_1^T X_{ij} + \boldsymbol{v}_2^T H_i(t) \right\} + v_6 n_i$$
$$+ \int_{b_i} \left(\sum_{k=1}^{K} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} \{ \boldsymbol{v}_3^T \boldsymbol{W}_{ik} + \boldsymbol{v}_4^T \boldsymbol{H}_{ik} + \boldsymbol{v}_5^T \boldsymbol{b}_{ik} - \boldsymbol{v}^T \boldsymbol{h}_k^*(s) \} d\Lambda_{k0}(s) \right.$$
$$\left. + v_7 \frac{f'_{b,\sigma^2}(b_i; \sigma^2)}{f_b(b_i; \sigma^2)} \right) \exp \left\{ - \sum_{k=1}^{K} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} d\Lambda_{k0}(s) \right\} f_b(b_i; \sigma_0^2) db_i = 0.$$

Because t_{ik} is arbitrary, we can set $t_{ik} = 0$ for all $k = 1, \ldots, K$ in the above equation to obtain

$$\sum_{j=1}^{n_i} \mathcal{H}_{ij\tau}(t, \mathcal{O}; \boldsymbol{\theta}_0) \left\{ \boldsymbol{v}_1^T X_{ij} + \boldsymbol{v}_2^T H_i(t) \right\} + v_6 n_i = 0.$$

Because $\mathcal{H}_{ij\tau} \neq 0$, $n_i > 0$ and along with conditions C2 and C4, we conclude that $\boldsymbol{v}_1 = \mathbf{0}$, $\boldsymbol{v}_2 = \mathbf{0}$ and $v_6 = 0$. Applying the inverse Laplace transform, we have

$$\sum_{k=1}^{K} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} \{ \boldsymbol{v}_3^T \boldsymbol{W}_{ik} + \boldsymbol{v}_4^T \boldsymbol{H}_{ik} + \boldsymbol{v}_5^T \boldsymbol{b}_{ik} - \boldsymbol{v}^T \boldsymbol{h}_k^*(s) \} d\Lambda_{k0}(s)$$
$$+ v_7 \frac{f'_{b,\sigma^2}(b_i; \sigma^2)}{f_b(b_i; \sigma^2)} = 0$$

for any b_i. It follows that $\boldsymbol{v}_5 = 0$, $v_7 = 0$ and

$$\sum_{k=1}^{K} \int_0^{t_{ik}} e^{\boldsymbol{\beta}_0^T \boldsymbol{W}_{ik} + \boldsymbol{\alpha}_0^T \boldsymbol{H}_{ik} + \boldsymbol{\zeta}_0^T \boldsymbol{b}_{ik}} \{ \boldsymbol{v}_3^T \boldsymbol{W}_{ik} + \boldsymbol{v}_4^T \boldsymbol{H}_{ik} - \boldsymbol{v}^T \boldsymbol{h}_k^*(s) \} d\Lambda_{k0}(s) = 0.$$

By differentiating both sides of the above equation with respect to t_{ik}, we obtain $\{ \boldsymbol{v}_3^T \boldsymbol{W}_{ik} + \boldsymbol{v}_4^T \boldsymbol{H}_{ik} - \boldsymbol{v}^T \boldsymbol{h}_k^*(s) \} = 0$. By condition C2, we claim that $\boldsymbol{v}_3 = 0$ and $\boldsymbol{v}_4 = 0$. Therefore, the matrix $\mathbb{P} \left[\{ \boldsymbol{l}_{\boldsymbol{\theta}} (\boldsymbol{\theta}_0, \mathcal{A}_0) - l_{\mathcal{A}} (\boldsymbol{\theta}_0, \mathcal{A}_0) (\boldsymbol{h}^*) \}^{\otimes 2} \right]$ is invertible.

Using all the above results, it follows that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_P(1/\sqrt{n})$, and

$$\begin{aligned}
&\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \\
&= \left(\mathbb{P}\left[\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right) - l_{\mathcal{A}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right)\left(\boldsymbol{h}^*\right)\right\}^{\otimes 2}\right]\right)^{-1} \mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}}) - l_{\mathcal{A}}(\hat{\boldsymbol{\theta}}, \hat{\mathcal{A}})(\boldsymbol{h}^*)\right\} + o_P(1).
\end{aligned}$$

Since the estimator $\hat{\boldsymbol{\theta}}$ has an efficient influence function, $\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$ converges weakly to a multivariate zero-mean normal random variable with its covariance matrix reaching the semi-parametric efficiency bound. □

Chapter 4

Numerical Study

In this chapter, we illustrate our method through extensive simulation studies and an application to a dementia dataset from a French cohort study: *PAQUID*. In the simulation studies, we assess the performance of our proposed joint model under different assumptions of the distribution of the longitudinal response, for different sample sizes, and under the situation that there is a terminal event. In the real data example, we fit our model to the data considering Isaacs Set Test scores as longitudinal measurements, dementia and dependency status change as interval-censored asymptomatic events, and death as a right-censored and terminal event.

4.1 Simulation

In this section, we setup and conduct some simulation studies and report the results to investigate the performance of our proposed joint method in analysing longitudinal and survival data. We show that a joint analysis of longitudinal and multiple-censored time-to-event data is better than that based on separate models. In Section 4.1.1, we list the process and values of parameters for different setups of the simulation studies. One longitudinal response variable, two asymptomatic events and two symptomatic

events are considered with different sample sizes and error distributions at quantile levels of $\tau = 0.25$, 0.50 and 0.85. The simulation results for different settings are then reported and interpreted in Section 4.1.2.

4.1.1 Simulation setup

We set the sample size n to 150 or 300 and simulate 1000 replicates for different quantiles of longitudinal observations at $\tau = 0.25, 0.5$ and 0.85. We consider two asymptomatic events, two symptomatic events and a maximum of six longitudinal observations for each subject. For $i = 1, \ldots, n$ and $k = 1, 2, 3, 4$, we set $X_i = (1, x_i)^T$, $H_i(t) = h_i * t$ and $W_{ik} = w_{ik}$, where x_i , h_i and w_{ik} are generated from independent standard normal distributions. We assume that the random variable b_i follows a zero-centred normal distribution with a variance $\sigma^2 = 1$. We fix $\boldsymbol{\eta} = (\eta_1, \eta_2)^T = (1, 1)^T$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^T = (1, 1, 0.5, 0.5)^T$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^T = (0.5, 0.5, 1, 1)^T$, and $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \zeta_3, \zeta_4)^T = (0.25, 0.25, 0.5, 0.25)^T$. Further, we let $\Lambda_k(t) = \log(1 + t/k)$ for $k = 1, 2, 3, 4$. All symptomatic- and asymptomatic-event times, T_{ik}, can be obtained by solving the following equations

$$\begin{aligned} s_{ik} &= S_k(T_{ik}; W_{ik}, H_i(t), b_i) \\ &= \exp\left\{ - \int_0^{T_{ik}} e^{\beta_k^T W_{ik} + \alpha_k \delta^T H_i(t) + \zeta_k b_i} d\Lambda_k(t) \right\} \end{aligned}$$

using numerical methods, where s_{ik} is randomly generated from the uniform distribution on $[0, 1]$ interval, that is $s_{ik} \sim Unif[0, 1]$. We then let both symptomatic events be censored by $C_i \sim Unif(M(T_4), 15)$, where $M(T_4)$ is the median of $\{T_{14}, \ldots, T_{n4}\}$ and 15 is set to be the maximum following time. The censoring rates for two symptomatic rates are then around 32% and 36% respectively. For $k = 1, 2$, the potential monitoring times for asymptomatic events are generated through $I_{ik}^d = I_{ik}^{d-1} + 0.1 + Unif(0, 0.5)$

and stopped at C_i for $d \geqslant 1$ and $I^0 = 0$. Thus $(L_{ik}, R_{ik}]$ is the smallest interval constructed by the series of I_{ik}^d and ∞, which contains the true time T_{ik}. For each subject, the longitudinal responses are potentially observed at $t = 0$, and the 5th, 10th, 20th, 40th, 80th quantiles of the series of $\{D_{i4} = \min(T_{i4}, C_i); i = 1, \ldots, n\}$. In order to illustrate the performance of our model under violation of assumptions, we generate longitudinal observations under two distributional settings using the first equation in (3.1), $Y_{ij} = \eta^T X_{ij} + \delta^T H_i(t) + \epsilon_{ij}$. In one setting, we assume an ALD distributed error, ϵ_{ij}, centred at 0 with skewness τ and scale parameter $\varrho = 1$. In another setting, we assume a normal distributed error, $\epsilon_{ij} = \varepsilon_{ij} - q_\tau$, where $\varepsilon_{ij} \sim N(0, 1)$ and q_τ is the τth quantile of the standard normal distribution. We fit the generated data using our proposed joint model and compare the results with those of separate models.

In MCEM iterations, we set the initial values of $\boldsymbol{\eta}$, $\boldsymbol{\delta}$, $\boldsymbol{\beta}_k$, α_k as estimates obtained from separate models with fixed effects, and let $\zeta_k = 1$, $\sigma_b^2 = 1$ and $\lambda_{kl} = 1/m_k$ to start the first run. We use 100 random points generated from $Unif(-3, 3)$ to approximate integrations over random effect b_i within iterations. The variance of parameters are estimated with 100 bootstrap samples.

4.1.2 Simulation results

For the estimates under each setting of the above simulation studies, we report the bias, the bootstrapped standard error (SD), the empirical standard deviation (SE) over 1000 replicates, and the coverage probability (CP) of the true parameter by the 95% confidence interval constructed using the bootstrapped standard error. Table 4.1 summarizes the simulation results for a sample of size $n = 150$ and values of longitudinal response generated from an Asymmetric Laplace Distribution. It can be seen that, for every quantile level (τ= 0.25, 0.50, and 0.85), the biases of parameter estimators are small for both joint and separate models of longitudinal and survival data. Our

Table 4.1: Biases (Bias), empirical standard deviations (SD) over 1000 simulations, averaged Bootstrap standard errors (SE) and coverage probabilities (CP) of the true value based on the 95% CI's for the estimates of parameters using our proposed joint model and separate models with fixed effects are reported. This table lists the results of sample size $n = 150$ and at three quantile levels $\tau = 0.25, 0.50$ and 0.85, where the longitudinal error follows an ALD.

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
150	0.25	η_1	-0.008	0.096	0.097	0.953	-0.003	0.081	0.069	0.891
		η_2	-0.005	0.096	0.093	0.947	-0.002	0.079	0.069	0.895
		δ	0.010	0.074	0.073	0.937	-0.001	0.026	0.022	0.861
		β_1	0.048	0.147	0.152	0.943	-0.074	0.113	0.115	0.893
		β_2	0.051	0.136	0.140	0.933	-0.036	0.125	0.121	0.928
		β_3	-0.001	0.130	0.131	0.965	-0.046	0.112	0.111	0.918
		β_4	0.014	0.133	0.141	0.946	-0.007	0.116	0.115	0.953
		α_1	0.000	0.121	0.125	0.951	-0.022	0.107	0.081	0.846
		α_2	-0.059	0.124	0.120	0.931	-0.071	0.112	0.067	0.654
		α_3	0.001	0.133	0.133	0.950	-0.065	0.110	0.105	0.860
		α_4	0.010	0.139	0.143	0.962	-0.008	0.109	0.105	0.933
		ζ_1	0.015	0.167	0.166	0.947				
		ζ_2	0.044	0.196	0.201	0.960				
		ζ_3	-0.088	0.178	0.176	0.929				
		ζ_4	-0.030	0.168	0.170	0.943				
		ϱ	-0.003	0.038	0.037	0.952				
		σ^2	-0.003	0.052	0.055	0.961				

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
150	0.50	η_1	-0.008	0.082	0.081	0.950	0.001	0.070	0.058	0.880
		η_2	0.008	0.079	0.079	0.951	-0.002	0.071	0.060	0.883
		δ	0.006	0.068	0.069	0.944	0.000	0.021	0.018	0.872
		β_1	0.049	0.134	0.135	0.935	-0.074	0.115	0.115	0.879
		β_2	0.040	0.140	0.137	0.927	-0.040	0.122	0.121	0.933
		β_3	0.001	0.135	0.133	0.946	-0.040	0.116	0.111	0.923
		β_4	-0.003	0.121	0.123	0.954	-0.001	0.120	0.116	0.952
		α_1	-0.008	0.108	0.105	0.935	-0.023	0.101	0.080	0.863
		α_2	-0.063	0.131	0.134	0.927	-0.069	0.114	0.066	0.663
		α_3	-0.024	0.136	0.143	0.968	-0.064	0.111	0.105	0.852
		α_4	0.023	0.143	0.140	0.943	-0.003	0.111	0.106	0.939
		ζ_1	-0.003	0.172	0.168	0.944				
		ζ_2	0.037	0.192	0.190	0.943				
		ζ_3	-0.066	0.164	0.164	0.928				
		ζ_4	-0.013	0.189	0.186	0.961				
		ϱ	-0.001	0.035	0.035	0.963				
		σ^2	-0.004	0.051	0.052	0.960				
150	0.85	η_1	-0.009	0.106	0.104	0.933	-0.005	0.099	0.080	0.862
		η_2	-0.003	0.105	0.101	0.927	0.002	0.099	0.085	0.876
		δ	0.013	0.073	0.071	0.908	-0.002	0.031	0.027	0.855
		β_1	0.059	0.149	0.150	0.926	-0.075	0.121	0.115	0.871
		β_2	0.037	0.138	0.136	0.953	-0.036	0.130	0.122	0.927

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		β_3	-0.007	0.127	0.125	0.945	-0.047	0.113	0.112	0.926
		β_4	0.010	0.129	0.132	0.950	-0.009	0.118	0.116	0.946
		α_1	0.000	0.113	0.114	0.953	-0.023	0.104	0.081	0.868
		α_2	-0.067	0.127	0.126	0.915	-0.071	0.117	0.066	0.640
		α_3	-0.007	0.138	0.136	0.959	-0.064	0.118	0.106	0.857
		α_4	0.024	0.133	0.137	0.952	-0.007	0.112	0.105	0.931
		ζ_1	0.006	0.162	0.163	0.954				
		ζ_2	0.050	0.208	0.215	0.950				
		ζ_3	-0.099	0.191	0.200	0.936				
		ζ_4	-0.010	0.179	0.176	0.954				
		ϱ	-0.002	0.038	0.038	0.941				
		σ^2	0.001	0.051	0.054	0.964				

proposed joint model performed well in estimating the effects of longitudinal response, acted as a covariate in survival sub-model, on survival times for both symptomatic and asymptomatic events. Our joint model has significant smaller biases in estimating α_1 and α_3 than separate models. Further, the bootstrapped variance estimators of our joint model are more accurate than those in separate models. The separate models tend to underestimate the variances of many parameters. The coverage probabilities for all estimates are around 0.95 for the joint model, while the separate models give many coverage probabilities less than 0.90. The difference between our proposed joint model and separate models become more clear when the sample size is large. Table 4.2 shows the simulation results for a sample size $n = 300$. The biases become smaller

Table 4.2: Biases (Bias), empirical standard deviations (SD) over 1000 simulations, averaged Bootstrap standard errors (SE) and coverage probabilities (CP) of the true value based on the 95% CI's for the estimates of parameters using our proposed joint model and separate models with fixed effects are reported. This table lists the results of sample size $n = 300$ and at three quantile levels $\tau = 0.25, 0.50$ and 0.85, where the longitudinal error follows an ALD.

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
300	0.25	η_1	0.007	0.065	0.068	0.959	0.001	0.056	0.060	0.963
		η_2	0.003	0.062	0.062	0.931	0.000	0.060	0.059	0.948
		δ	0.007	0.053	0.049	0.931	0.000	0.017	0.015	0.923
		β_1	0.021	0.089	0.094	0.941	-0.080	0.081	0.080	0.800
		β_2	0.019	0.095	0.100	0.959	-0.058	0.085	0.084	0.890
		β_3	-0.011	0.087	0.092	0.962	-0.052	0.080	0.077	0.871
		β_4	0.004	0.096	0.100	0.946	-0.009	0.080	0.080	0.948
		α_1	-0.022	0.087	0.086	0.921	-0.042	0.081	0.053	0.773
		α_2	-0.101	0.102	0.103	0.948	-0.112	0.093	0.042	0.397
		α_3	-0.014	0.107	0.099	0.940	-0.083	0.083	0.071	0.722
		α_4	0.005	0.093	0.091	0.953	-0.019	0.076	0.071	0.922
		ζ_1	-0.003	0.108	0.108	0.959				
		ζ_2	0.033	0.133	0.138	0.940				
		ζ_3	-0.067	0.113	0.121	0.928				
		ζ_4	-0.015	0.135	0.134	0.940				
		ϱ	-0.001	0.023	0.023	0.950				
		σ^2	-0.020	0.035	0.037	0.943				

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
300	0.50	η_1	0.000	0.053	0.049	0.913	-0.001	0.048	0.052	0.960
		η_2	-0.008	0.054	0.057	0.954	0.003	0.049	0.051	0.956
		δ	0.001	0.047	0.043	0.921	0.001	0.015	0.014	0.934
		β_1	0.020	0.094	0.083	0.928	-0.081	0.084	0.080	0.781
		β_2	0.028	0.106	0.105	0.925	-0.053	0.083	0.085	0.908
		β_3	-0.005	0.095	0.093	0.959	-0.054	0.083	0.077	0.865
		β_4	0.009	0.083	0.088	0.954	-0.012	0.083	0.080	0.941
		α_1	-0.021	0.088	0.080	0.939	-0.043	0.080	0.053	0.771
		α_2	-0.087	0.106	0.094	0.935	-0.107	0.092	0.043	0.409
		α_3	-0.025	0.098	0.102	0.966	-0.085	0.082	0.071	0.737
		α_4	0.009	0.089	0.083	0.933	-0.018	0.082	0.072	0.921
		ζ_1	-0.002	0.127	0.129	0.957				
		ζ_2	0.055	0.152	0.147	0.912				
		ζ_3	-0.074	0.117	0.127	0.934				
		ζ_4	-0.004	0.120	0.121	0.950				
		ϱ	0.001	0.024	0.028	0.966				
		σ^2	-0.015	0.036	0.037	0.948				
300	0.85	η_1	-0.007	0.092	0.092	0.957	-0.002	0.069	0.073	0.956
		η_2	-0.010	0.078	0.078	0.955	-0.001	0.069	0.071	0.957
		δ	0.006	0.058	0.057	0.954	0.000	0.022	0.018	0.872
		β_1	0.043	0.094	0.094	0.938	-0.076	0.082	0.080	0.822
		β_2	0.042	0.093	0.091	0.926	-0.052	0.086	0.084	0.893

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		β_3	-0.002	0.084	0.087	0.951	-0.052	0.078	0.077	0.886
		β_4	0.009	0.082	0.082	0.935	-0.009	0.080	0.080	0.940
		α_1	-0.018	0.093	0.090	0.924	-0.043	0.083	0.053	0.760
		α_2	-0.090	0.104	0.100	0.929	-0.103	0.082	0.043	0.431
		α_3	-0.020	0.101	0.097	0.935	-0.079	0.081	0.071	0.724
		α_4	0.012	0.095	0.095	0.960	-0.018	0.080	0.072	0.901
		ζ_1	-0.005	0.118	0.115	0.927				
		ζ_2	0.039	0.145	0.144	0.926				
		ζ_3	-0.082	0.119	0.117	0.929				
		ζ_4	-0.010	0.142	0.140	0.953				
		ϱ	-0.002	0.025	0.025	0.945				
		σ^2	0.018	0.035	0.035	0.936				

for most of parameter estimators when sample size increases. For $k = 1, 2, 3, 4$, our proposed model gives smaller biases than separate models for almost all estimators of β_k, α_k and ζ_k. The coverage probabilities under separate model are small for many parameters and can be as low as 0.40 for α_2. The above indicates that our joint model is better in capturing the way in which the longitudinal measurements and survival times are connected and how the longitudinal process affects the risks of different events.

For different simulation replicates, we set some fixed time points uniformly located between 0 and 6 (e.g. $0.0, 0.2, 0.4, \ldots, 5.8, 6.0$) and estimate the survival probabilities at each fixed time point by $\exp(-\sum_{tkl<t} \lambda_{kl})$ for $k = 1, 2, 3, 4$. The survival probabil-

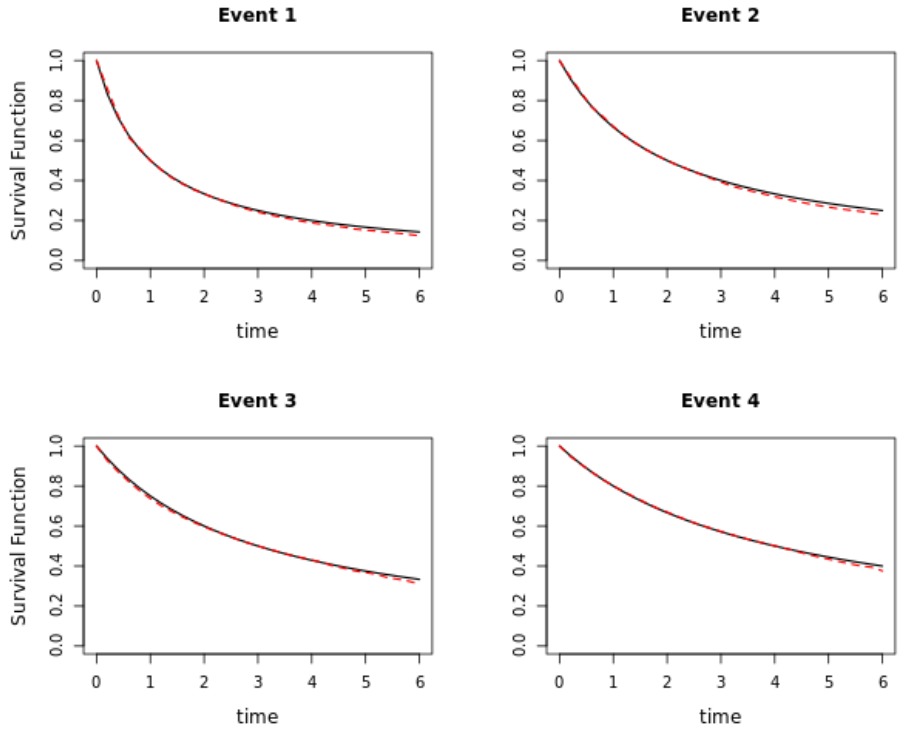

Figure 4.1: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.25$ and ALD distributed longitudinal response variable.

ities at each fixed time point are then averaged over 1000 simulations to approximate the true baseline survival function. Figure 4.1 along Figures A.1, A.2, A.3, A.4 and A.5 in the Appendix show the estimations of the baseline survival functions for the settings of $n = 150$ or 300, $\tau = 0.25, 0.50, 0.85$ and ALD distributed longitudinal response variable. For all levels of quantile, it can be seen that the estimated baseline survival curves are very close to the true curves indicating the estimators are accurate and virtually unbiased, especially at early follow-up time and when sample size is big where a large number of subjects exist and more information about the event risks can be obtained from the data.

We considered a semi-competing risks setup by involving a terminal event. With-

out loss of generality, we set the last event (k=4) to be terminal which can occur after any of the other three events have occurred but its occurrence censors all the other events and terminates the follow-up study. The semi-competing event was also censored by C, such that the censoring rate for itself is still around 36% while the censoring rate for the other symptomatic event changed to be around 58%. Simulation results are shown in Tables 4.3 and 4.4 for sample sizes $n = 150$ and 300 and quantiles $\tau = 0.25, 0.50$ and 0.85. The results of separate models improved due to the increase of the correlation between events and thus relatively reduces the dependency between longitudinal measurements and the risks of events. However, our proposed joint model still performs better in estimating variances of parameters, especially for small samples. The estimation of baseline survival functions for semi-competing risks joint models are shown in Figures A.6, A.7, A.8, A.9, A.10 and A.11 for different sample sizes and levels of quantile. Similar conclusions are made to the case of no competing event. Only difference is that the estimators for the first three events are more biased at later follow-up time points. This is due to that, at later follow-up times, events are more likely censored by either C or the terminal event, resulting less information can be obtained from the data and thus lead to biased estimates, especially when the sample size is small.

In order to illustrate the justifiability of assuming an asymmetric Laplace distributed error for the longitudinal outcome, we run more simulation studies with normally assumed error distributions when generating longitudinal responses, Y_{ij}. The simulation results are summarized in Tables A.1 , A.2, Figures A.12, A.13, A.14, A.15, A.16, A.17 for no competing risks joint models and Tables A.3, A.4, Figures A.18, A.19, A.20, A.21, A.22, A.23 for semi-competing risks joint models. It can be seen that the results are approximately equal to those reported in previous tables and figures for the cases of assuming an ALD distributed error for the longitudinal outcome. The conclusions

are similar and indicating the robustness of the ALD assumption with respect to model misspecification when the true error distribution for the longitudinal outcome is Normal.

Table 4.3: Biases (Bias), empirical standard deviations (SD) over 1000 simulations, averaged Bootstrap standard errors (SE) and coverage probabilities (CP) of the true value based on the 95% CI's for the estimates of parameters using our proposed joint model and separate models with fixed effects are reported. This table lists the results of sample size $n = 150$ and at three quantile levels $\tau = 0.25, 0.50$ and 0.85, where the longitudinal error follows an ALD distribution and the last symptomatic event is terminal.

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
150	0.25	η_1	0.024	0.108	0.106	0.950	0.003	0.079	0.068	0.891
		η_2	0.004	0.099	0.094	0.945	-0.003	0.084	0.069	0.880
		δ	0.018	0.101	0.105	0.932	0.001	0.026	0.022	0.863
		β_1	0.076	0.152	0.139	0.933	-0.055	0.141	0.135	0.898
		β_2	0.059	0.161	0.175	0.924	-0.017	0.151	0.151	0.942
		β_3	0.015	0.164	0.149	0.936	-0.046	0.136	0.140	0.948
		β_4	0.010	0.128	0.145	0.966	-0.008	0.116	0.115	0.952
		α_1	0.041	0.190	0.195	0.927	-0.017	0.146	0.148	0.947
		α_2	0.041	0.170	0.166	0.955	-0.014	0.145	0.141	0.941
		α_3	0.003	0.202	0.189	0.936	-0.049	0.177	0.159	0.894
		α_4	-0.004	0.137	0.134	0.951	-0.008	0.113	0.105	0.927
		ζ_1	0.039	0.203	0.209	0.962				
		ζ_2	0.045	0.217	0.207	0.934				
		ζ_3	-0.063	0.209	0.219	0.945				
		ζ_4	-0.041	0.189	0.202	0.960				

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		ϱ	0.004	0.039	0.036	0.952				
		σ^2	-0.005	0.046	0.043	0.950				
150	0.50	η_1	-0.007	0.096	0.089	0.931	0.000	0.072	0.058	0.884
		η_2	-0.002	0.094	0.088	0.936	-0.002	0.073	0.060	0.874
		δ	-0.006	0.096	0.095	0.942	-0.001	0.022	0.019	0.874
		β_1	0.063	0.146	0.158	0.941	-0.058	0.138	0.135	0.917
		β_2	0.072	0.176	0.178	0.958	-0.022	0.154	0.151	0.942
		β_3	-0.004	0.155	0.158	0.954	-0.034	0.141	0.139	0.935
		β_4	0.008	0.128	0.124	0.960	-0.005	0.116	0.115	0.962
		α_1	0.036	0.175	0.184	0.947	-0.007	0.155	0.150	0.934
		α_2	0.044	0.172	0.165	0.953	-0.011	0.141	0.142	0.953
		α_3	0.025	0.181	0.169	0.942	-0.048	0.164	0.160	0.912
		α_4	0.038	0.144	0.138	0.926	-0.005	0.116	0.106	0.927
		ζ_1	0.039	0.182	0.180	0.951				
		ζ_2	0.049	0.253	0.242	0.947				
		ζ_3	-0.014	0.207	0.215	0.971				
		ζ_4	-0.027	0.187	0.179	0.945				
		ϱ	0.002	0.040	0.038	0.946				
		σ^2	-0.002	0.049	0.051	0.955				
150	0.85	η_1	-0.008	0.112	0.109	0.955	-0.006	0.098	0.080	0.876
		η_2	0.001	0.127	0.131	0.956	-0.001	0.103	0.086	0.873

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		δ	-0.003	0.102	0.104	0.953	-0.001	0.033	0.027	0.844
		β_1	0.081	0.159	0.163	0.926	-0.060	0.136	0.135	0.917
		β_2	0.053	0.168	0.168	0.935	-0.018	0.161	0.150	0.935
		β_3	0.019	0.159	0.150	0.939	-0.035	0.137	0.140	0.949
		β_4	0.009	0.128	0.131	0.948	-0.002	0.119	0.116	0.943
		α_1	0.038	0.186	0.177	0.919	-0.009	0.152	0.148	0.940
		α_2	0.065	0.152	0.152	0.941	-0.002	0.145	0.142	0.948
		α_3	0.021	0.217	0.218	0.955	-0.049	0.172	0.159	0.892
		α_4	0.030	0.147	0.144	0.950	-0.008	0.110	0.105	0.935
		ζ_1	0.003	0.212	0.214	0.954				
		ζ_2	0.033	0.238	0.247	0.966				
		ζ_3	-0.049	0.219	0.212	0.937				
		ζ_4	-0.052	0.193	0.195	0.951				
		ϱ	-0.001	0.039	0.039	0.946				
		σ^2	-0.005	0.045	0.044	0.952				

Table 4.4: Biases (Bias), empirical standard deviations (SD) over 1000 simulations, averaged Bootstrap standard errors (SE) and coverage probabilities (CP) of the true value based on the 95% CI's for the estimates of parameters using our proposed joint model and separate models with fixed effects are reported. This table lists the results of sample size $n = 300$ and at three quantile levels $\tau = 0.25, 0.50$ and 0.85, where the longitudinal error follows an ALD distribution and the last symptomatic event is terminal.

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
300	0.25	η_1	0.002	0.066	0.066	0.954	0.004	0.054	0.060	0.969
		η_2	-0.010	0.064	0.065	0.951	0.000	0.056	0.059	0.954
		δ	0.013	0.067	0.061	0.944	0.000	0.017	0.015	0.909
		β_1	0.035	0.115	0.116	0.954	-0.066	0.092	0.093	0.873
		β_2	0.036	0.119	0.114	0.950	-0.034	0.108	0.104	0.924
		β_3	0.002	0.101	0.101	0.956	-0.044	0.102	0.097	0.909
		β_4	-0.008	0.090	0.083	0.949	-0.011	0.080	0.080	0.948
		α_1	0.012	0.111	0.099	0.918	-0.029	0.105	0.100	0.920
		α_2	0.010	0.109	0.105	0.930	-0.024	0.097	0.097	0.935
		α_3	-0.002	0.126	0.127	0.971	-0.062	0.112	0.108	0.885
		α_4	-0.007	0.097	0.089	0.936	-0.020	0.077	0.071	0.926
		ζ_1	-0.005	0.145	0.152	0.959				
		ζ_2	0.033	0.159	0.159	0.950				
		ζ_3	-0.042	0.151	0.163	0.953				
		ζ_4	-0.028	0.135	0.129	0.947				
		ϱ	0.003	0.028	0.030	0.971				

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		σ^2	-0.021	0.030	0.033	0.938				
300	0.50	η_1	-0.002	0.059	0.059	0.949	0.001	0.050	0.052	0.963
		η_2	-0.003	0.057	0.061	0.951	-0.001	0.048	0.051	0.968
		δ	-0.002	0.068	0.074	0.949	0.000	0.015	0.014	0.934
		β_1	0.015	0.105	0.098	0.928	-0.066	0.093	0.094	0.879
		β_2	0.019	0.114	0.112	0.936	-0.031	0.105	0.104	0.929
		β_3	0.002	0.108	0.108	0.949	-0.044	0.103	0.097	0.918
		β_4	0.011	0.083	0.081	0.962	-0.014	0.080	0.080	0.947
		α_1	0.021	0.138	0.136	0.962	-0.020	0.100	0.101	0.948
		α_2	0.012	0.090	0.089	0.955	-0.011	0.097	0.097	0.949
		α_3	0.014	0.141	0.147	0.943	-0.071	0.118	0.108	0.852
		α_4	0.007	0.102	0.109	0.954	-0.022	0.075	0.072	0.919
		ζ_1	-0.012	0.139	0.126	0.936				
		ζ_2	-0.004	0.150	0.155	0.933				
		ζ_3	-0.054	0.142	0.149	0.937				
		ζ_4	-0.032	0.126	0.127	0.956				
		ϱ	0.003	0.027	0.025	0.936				
		σ^2	-0.026	0.033	0.034	0.931				
300	0.85	η_1	0.004	0.103	0.077	0.922	-0.004	0.067	0.073	0.957
		η_2	0.008	0.106	0.082	0.966	-0.004	0.068	0.071	0.955
		δ	0.003	0.077	0.074	0.928	-0.001	0.021	0.018	0.884
		β_1	0.046	0.103	0.096	0.935	-0.069	0.093	0.093	0.862

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		β_2	0.031	0.112	0.114	0.943	-0.039	0.107	0.104	0.921
		β_3	-0.012	0.113	0.128	0.971	-0.049	0.100	0.097	0.909
		β_4	0.005	0.085	0.084	0.936	-0.013	0.082	0.080	0.946
		α_1	0.004	0.114	0.116	0.948	-0.018	0.102	0.101	0.935
		α_2	0.011	0.104	0.109	0.963	-0.017	0.099	0.097	0.944
		α_3	0.001	0.135	0.132	0.960	-0.066	0.111	0.108	0.876
		α_4	0.018	0.112	0.107	0.954	-0.017	0.075	0.072	0.925
		ζ_1	0.023	0.137	0.132	0.927				
		ζ_2	0.041	0.160	0.152	0.925				
		ζ_3	-0.042	0.141	0.152	0.963				
		ζ_4	-0.035	0.137	0.150	0.950				
		ϱ	0.001	0.029	0.028	0.955				
		σ^2	-0.018	0.030	0.032	0.943				

4.2 An application to dementia: what about the elderly?

Dementia is a common disease in the elderly. As the world's population is ageing, dementia has been recognized as a public health priority by the World Health Organization. Dementia is characterized by a significant cognitive decline which is progressive and steeper than the normal ageing-caused cognitive decline. Some degenerative forms of dementia can be developed among the elderly, such as Alzheimer's disease,

and Parkinson's dementia. The health, social and economic costs are increasing with the disease of dementia, as it heightens the risk of dependency and disability among the older people. Therefore, it is essential to identify risk factors to better understand the natural trajectory of dementia and early detect subjects with high risk of developing dementia.

We apply our proposed joint model to a French prospective cohort: PAQUID ("Personnes Agées QUID?" that is, "what about the elderly?", Letenneur et al. (1994)), which aims at studying cognitive ageing from repeated psychometric tests. The cohort involves 3777 subjects, aged 65 and older at entry, from the Gironde and Dordogne regions of south west France. During more than 25 years follow-up, subjects were visited by a trained psychologist every 2 or 3 years at home and administered a questionnaire that included health information, a battery of cognitive tests, and scales of disability. Dementia was assessed based on a two-stage screening procedure according to the Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised (DSM III R) criteria (American Psychiatric Association, 1987). Age at death was recorded from being informed by families or retrieving death registries.

4.2.1 Data

The dataset we use is a random subsample from the PAQUID cohort study, called "*paquid*". The *paquid* dataset is provided by an R package named as "**lcmm**" (Proust-Lima et al. (2019)) which consists of 500 subjects. Over a maximum period of 20 years, three common cognitive tests were repeatedly measured and scores were recorded. The three cognitive tests are the Mini Mental State Examination ($MMSE$) assessing global cognition, the Isaacs Set Test (IST) evaluating verbal fluency, and the Benton Visual Retention Test ($BVRT$) for visual perception and memory. Measures of physical dependency ($HIER$, 0 =no dependency; 1 =mild dependency; 2 =moderate

dependency; 3 =severe dependency) and depressive symptomatology ($CESD$) were collected at follow-up visits along with age at each visit (age), age at dementia diagnosis or last visit ($agedem$) and indicator of positive diagnosis of dementia (dem). Time-independent socio-demographic information is also provided: indicator of educational level (CEP, 1 =graduated from primary school; 0 =otherwise), sex ($male$, 1 for men; 0 for women), and age at entry (age_init). Since dementia was developed between two periodic visits, we let (L, R) denote the smallest time interval constructed by (age_init, age, Inf) that covers the time to dementia onset.

We then add two new variables, $death$ and $death\_age$, based on subjects' $HIER$ scores and exponential distributions. We assume that the nature life time that all subjects can survive after the last visit follows an exponential distribution with rate 0.5 and all subjects have a maximum age of 105. At the last visit, subjects with no or mild dependency ($HIER = 0$ or 1) can survive another length of time following an exponential distribution with rate 1 and subjects with moderate or severe dependency ($HIER = 2$ or 3) can live for another length of life time following an exponential distribution with rate 1.5. If a subject's simulated life time is less than his or her simulated natural life time, death occurred ($death = 1$) and age at death is observed ($death\_age =$ simulated life time). Otherwise, the event of death is censored at the last visit with indicator $death = 0$. In this way, we achieve a total death rate around 71%, a death rate around 68% among subjects developed dementia and a death rate around 72% among subjects without dementia diagnosed. For a better interpretation of the effect of age, we let $age = age - 65$ to centre the intercept around 65. Then, we divide the new values of age by 10 to reduce numerical problems in computation. Too large ages in the model reduce the attribution of random effects. We do the same procedure to get new values of $agedem$, age_init, (L, R).

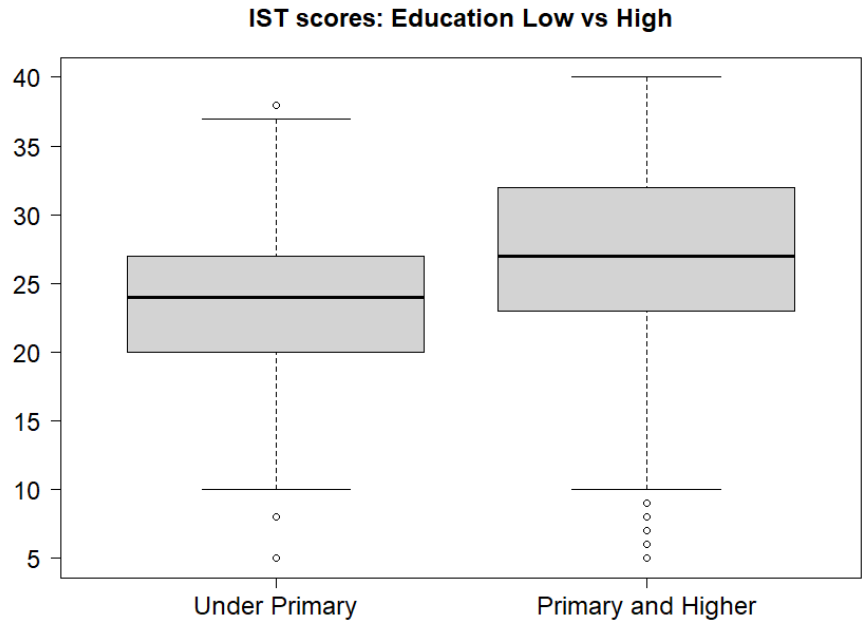

Figure 4.2: Box plots of IST scores for subjects not graduated from primary school and subjects graduated from primary school in the *paquid* dataset.

4.2.2 Joint model of IST, dementia and death

For longitudinal analysis, we work with the repeated measures of the Isaacs Set Test. Partitioner needs to produce up to 10 words within 15 seconds for four different semantic categories. Scores of IST ranges from 0 to 40. In Figure 4.2, comparing the two box plots of IST scores for subjects with and without primary school diploma respectively, we find that the two distributions are different. More specifically, subjects who received higher education tended to perform better than subjects with a low level of education in the Isaacs Set Test. Further, as subjects age, they were more likely to have a worse IST performance. This can be seen from the plots in Figure 4.3. In Figure 4.3, the left scatter plot shows a decreasing trend in IST scores among all subjects as they get older. In the plot on the right side, IST scores for a randomly selected sample of 10 subjects are connected separately for each subject using lines

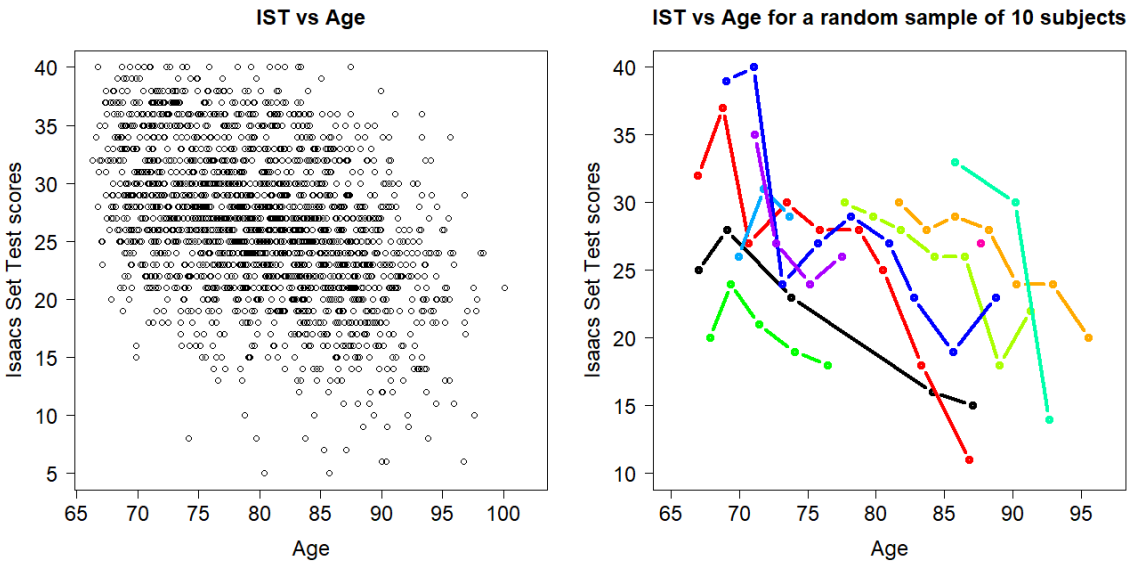

Figure 4.3: Plot of IST scores against ages at the time of visit. The left scatter plot is for all subjects in the *paquid* dataset. The right plot is based on a random subsample of 10 subjects. The IST scores in the right plot are connected with separate lines for each of those 10 subjects.

with different colours. Similarly, for those 10 subjects, as they getting older, they tend to perform worse in a Issacs Set Test. Therefore, we use education level (CEP) and subject's age (age) as two predictors of IST scores. The quantile regression model for longitudinal scores of Isaacs Set Test thus be proposed as

$$IST = \eta_1 + \eta_2 * CEP + \delta * age + \epsilon, \tag{4.1}$$

where the error ϵ is ALD distributed with a specified skewed parameter τ and a scale parameter ϱ.

When setting up survival models, time-to-dementia should not be analysed separately to the longitudinal IST scores as they are highly correlated. The IST score, as a cognitive marker, is a predictor of the risk of dementia. Subjects with lower IST scores are at higher risk of dementia. This can be seen in Figure 4.4 as the group of

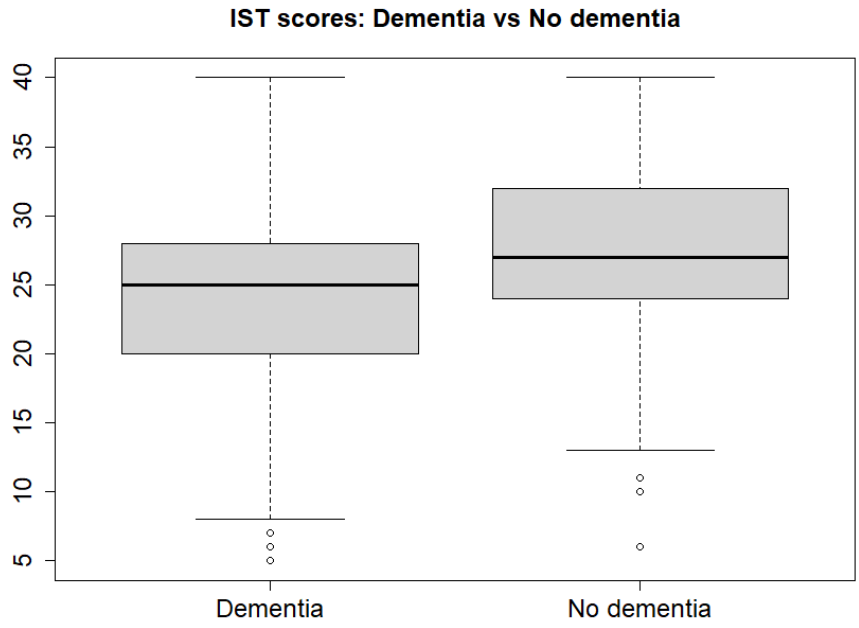

Figure 4.4: Box plots of IST scores for subjects diagnosed with dementia and subjects without dementia diagnosis in the *paquid* dataset.

subjects diagnosed with dementia has lower IST scores than those without dementia diagnosis. Our summary interest may focus on the lower tail of the distribution of IST scores and its effects on the risk of dementia.

As older people are more likely to develop the disease of dementia, the risk of death is non-negligible among the population at risk of dementia. Those two events are correlated and also share some common risk factors, such as sex and age. Figure 4.5 shows the different survival curves for the event of death for male and female subjects respectively and females are more likely to live for a longer time. Moreover, some longitudinal measurements of cognitive markers may be missing in the follow-up due to death. By taking death into consideration, as a semi-competing event, we could reduce the bias in estimating the effect of covariates on the IST scores or the risk of dementia. Further, the joint analysis distinguishes the effects of predictors on the

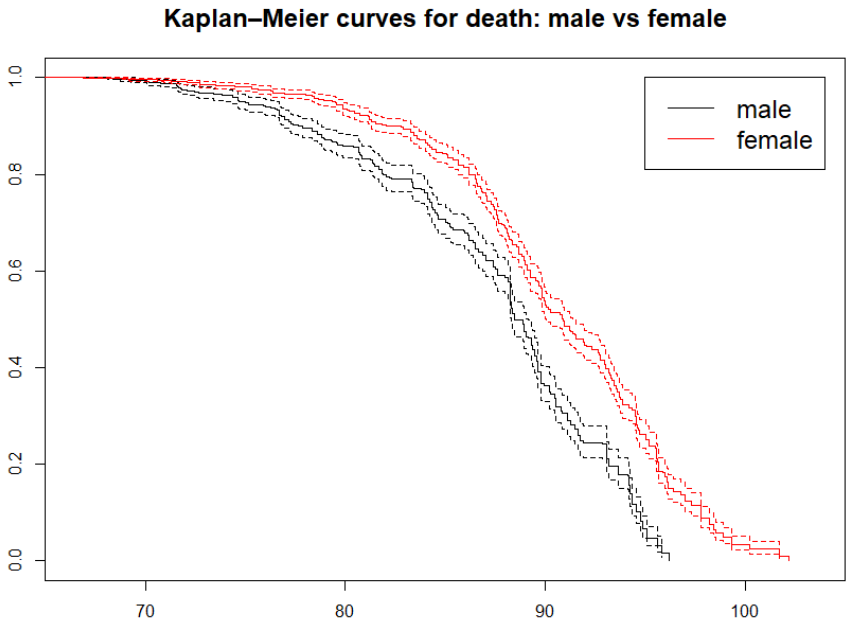

Figure 4.5: Kaplan-Meier death curves for male and female subjects in the simulated *paquid* dataset.

decline of cognitive marker, the risk of dementia and the risk of death respectively. We apply Cox proportional hazard models for the risk of dementia and the risk of death as the followings

$$\lambda_{dem}(t) = \exp\{\beta_1 * male + \alpha_1 * \delta * age + \zeta_1 * b\} * \lambda_{dem0}(t), \tag{4.2}$$

and

$$\lambda_{death}(t) = \exp\{\beta_2 * male + \alpha_2 * \delta * age + \zeta_2 * b\} * \lambda_{death0}(t), \tag{4.3}$$

where the random effect variable b captures possible underlying health conditions that affects both risks of dementia and death.

At different quantile levels, $\tau = 0.10,\ 0.15,\ 0.25,\ 0.35,\ 0.50,\ 0.75$ and 0.85, of the distribution of the IST scores, we report the estimates of parameters and their

Table 4.5: Estimates of regression and dispersion parameters in the joint model of Isaacs Set Test scores, dementia time and death time, at different levels of τ.

	τ						
	0.10	0.15	0.25	0.35	0.50	0.75	0.85
${\eta_1}^*$	23.6644	24.5815	26.3962	27.8512	29.9704	33.1576	34.2338
${\eta_2}^*$	2.3521	2.5598	2.7014	2.6155	2.7481	3.9115	4.7353
δ^*	-4.0997	-3.8345	-3.6812	-3.4971	-3.6152	-3.7238	-3.4048
β_1	-0.3512	-0.3519	-0.3552	-0.3435	-0.3553	-0.3329	-0.3700
${\beta_2}^*$	0.5184	0.5177	0.5107	0.5117	0.5026	0.5254	0.4905
${\alpha_1}^*$	-0.1410	-0.1522	-0.1362	-0.1499	-0.1310	-0.1501	-0.1415
${\alpha_2}^*$	-0.1117	-0.1200	-0.1085	-0.1188	-0.1032	-0.1193	-0.1128
${\zeta_1}^*$	1.0829	1.0598	1.0843	1.1143	1.0896	1.0969	1.0892
${\zeta_2}^*$	0.8489	0.8536	0.8663	0.8714	0.8815	0.8735	0.8833
ϱ^*	1.0011	1.3264	1.7901	2.0688	2.2201	1.7529	1.2501
${\sigma^2}^*$	1.5611	1.5975	1.4169	1.4286	1.3870	1.3962	1.3570

*Estimates in the row are significant at the 5% level for all quantiles τ.

bootstrapped standard error, lower and upper limits of the 95% confidence intervals in Table 4.5 and Table A.5 respectively. As expected, estimates of parameters are different at different quantiles. It can be seen that the estimates of the effect of CEP are significantly positive indicating subjects who received higher education get higher Isaacs Set Test scores than those did not finish their primary school. As subjects getting older, they perform worse in Isaacs Set Tests and have higher risks of dementia and death. Almost all effects of risk factors are significant except for

the effect of *male* on the dementia. Negative estimates of β_1 suggest that males are less likely to develop dementia than females but in an insignificant manner. Males have higher risk of death than females. The estimates of the variance of the random effects, σ^2, are significantly larger than zero, indicating some common underlying health conditions that strongly connect with both dementia and death. Further, the estimates of ζ_1 and ζ_2 are significantly larger than zero, suggesting a strong positive dependence between death and dementia.

4.2.3 Joint model of IST, dementia, death and dependency

Since the status of dependency, death and the disease of dementia are highly correlated, we take the dependency into our consideration in analysing the *paquid* dataset. The status of dependency were recorded at each visit, however the status changes might happen between two visits, making the change of the dependency status as an asymptomatic event with interval-censored times. We define another event as the first status change of dependency from non-severe ($HIER = 0, 1, 2$) to severe ($HIER = 3$) which can only happen between two visits. Without loss of generality, we set the initial ages as the very first monitoring time and the status of dependency at entry for all subjects were non-severe. The interval time points are obtained by the same way as those for dementia. The Cox proportional hazard sub-model for the risk of severe dependency thus be written as

$$\lambda_{den}(t) = \exp\{\beta_3 * male + \alpha_3 * \delta * age + \zeta_3 * b\} * \lambda_{den0}(t). \tag{4.4}$$

Estimation results for the joint analysis of longitudinal IST, dementia, dependency and death are reported in Tables 4.6 and A.6. At all specified levels of quantile, males are at lower risk of severe dependency than women but the effect is insignificant.

Table 4.6: Estimates of regression and dispersion parameters in the joint model of Isaacs Set Test scores, dementia time, death time and dependency status, at different levels of τ.

	τ						
	0.10	0.15	0.25	0.35	0.50	0.75	0.85
$\eta_1{}^*$	23.6521	24.5827	26.3963	27.8508	29.9732	33.1842	34.2020
$\eta_2{}^*$	2.3596	2.5601	2.7019	2.6158	2.7485	3.9276	4.7310
δ^*	-4.0947	-3.8353	-3.6815	-3.4968	-3.6172	-3.7552	-3.3843
$\beta_1{}^*$	-0.5231	-0.5462	-0.5466	-0.5115	-0.5088	-0.5526	-0.5305
$\beta_2{}^*$	0.4634	0.4605	0.4369	0.4795	0.4657	0.4608	0.4399
β_3	-0.2252	-0.2378	-0.2459	-0.2011	-0.2073	-0.2394	-0.2465
$\alpha_1{}^*$	-0.1065	-0.1195	-0.1107	-0.1286	-0.1165	-0.1219	-0.1162
$\alpha_2{}^*$	-0.0847	-0.0937	-0.0858	-0.1019	-0.0931	-0.0977	-0.0905
$\alpha_3{}^*$	-0.0965	-0.1089	-0.0990	-0.1171	-0.1048	-0.1107	-0.1042
$\zeta_1{}^*$	1.7782	1.8045	1.7321	1.7760	1.7571	1.8048	1.7096
$\zeta_2{}^*$	0.8259	0.8418	0.8340	0.8181	0.8154	0.7992	0.8419
$\zeta_3{}^*$	1.7542	1.7624	1.7176	1.7425	1.7396	1.7575	1.7052
ϱ^*	1.0011	1.3264	1.7901	2.0688	2.2201	1.7529	1.2500
σ^{2*}	1.5873	1.5600	1.5426	1.6164	1.4909	1.5759	1.4770

*Estimates in the row are significant at the 5% level for all quantiles τ.

Ageing increases the risk of severe dependency significantly. Moreover, after adding the dependency as an asymptomatic event in the joint model, the effects of *male* on the risk of dementia become significant suggesting females have significantly higher risks to develop dementia than males.

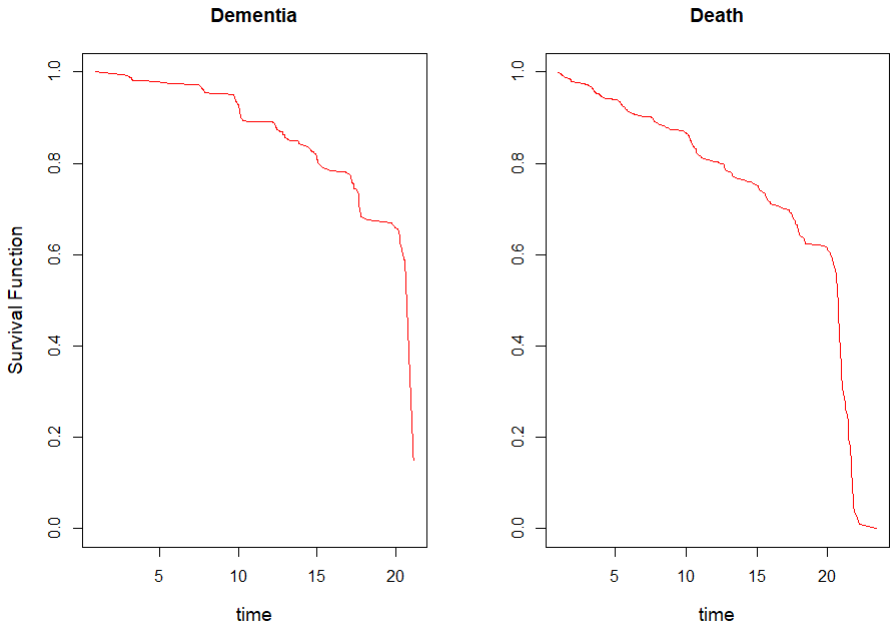

Figure 4.6: Estimated baseline survival curve for dementia and death under the joint model for the simulated *paquid* dataset.

Figures 4.6 and 4.7 show the estimated baseline survival curves of dementia, dependency and death for above two joint models. It can be seen that there is a big drop in survival probabilities for all events at follow-up time around 20. The risk of severe dependency seems to be lower than the risk of dementia at any certain time point.

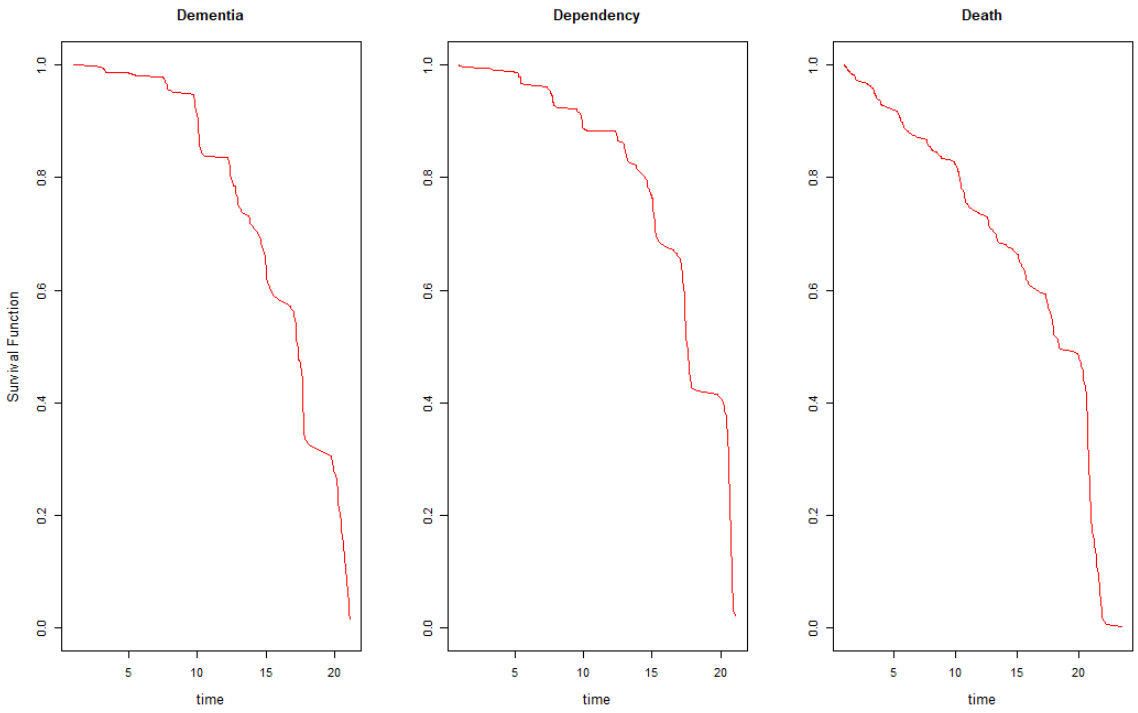

Figure 4.7: Estimated baseline survival curve for dementia, dependency and death under the joint model for the simulated *paquid* dataset.

Chapter 5

Discussion

In this thesis, we proposed a new joint model for a quantile of longitudinal observations and survival times of both asymptomatic and symptomatic events. We measured the effects of covariates on the longitudinal response at a specified τth quantile by assuming an ALD error in the longitudinal regression model. The survival and longitudinal processes are connected through a common vector of time-varying covariates to measure the effects of the longitudinal trend on the occurring times of events at different quantiles of the longitudinal response distribution. We characterized the dependence between and within asymptomatic and symptomatic events through a random effect variable b and used different coefficients to capture the effects of underlying unobserved conditions on the time of occurrence for different events. There is no such proposals that have been previously made in the literature.

We proved that the non-parametric maximum likelihood estimators derived from our proposed joint model satisfies the theories of consistency and normality. Our proposed MCEM method performed well in computing the estimates of parameters and baseline hazards for both the simulation studies and real data analysis. Our joint model outperformed separate models for small samples where information is

lacking and inference is difficult to draw, especially at latter times of a follow-up study. By applying our joint model of quantiles of longitudinal measurements of a psychometric test, development times of dementia, death times, and status change times of dependency, we concluded more insights of the cognitive ageing among older people.

There are many possible expansions of our proposed joint model. For example, in the quantile regression submodel for longitudinal data, we can add a vector of random effects, $\boldsymbol{u}_i$, to capture the heterogeneity between subjects and between repeated observations, resulting in a linear mixed quantile model. Furthermore, the random effects in the longitudinal model also affects the survival events with the effects measured by an unknown coefficient, say χ. We then have the form of proposed joint model as the following

$$\begin{cases} Y_{ij} = \eta^T X_{ij} + \delta^T H_i(t) + \boldsymbol{u}_i^T Z_{ij} + \epsilon_{ij} = \tilde{Q}_{\tau ij} + \epsilon_{ij} \\ \lambda_k(t_i; \tilde{Q}_{\tau t_i}, W_{ik}, b_i) = e^{\beta_k^T W_{ik} + \alpha_k \delta^T H_i(t_i) + \chi_k \boldsymbol{u}_i^T Z_{ij} + \zeta_k b_i} \lambda_{k0}(t_i). \end{cases}$$

where Z_{ij} is a vector of covariates associated with random effects $\boldsymbol{u}_i$. In addition, if the longitudinal response is count or binary, a generalized linear quantile model would be more appropriate for the longitudinal outcome, such as Poisson or logistic models.

The survival part of our proposed joint model can be reduced to contain only interval-censored events or only right-censored (competing) events to accommodate problems in real world survival analysis. We can include more than one terminal events in the joint model with the study ended for the occurrence of any one of them. Also, the joint model can be expanded by adding other types of events, such as left-censored events, truncated events, events that can be reoccurred during the follow-up, and so on. Another interesting extension of our proposed model is to include multi-

stage or multi-level events with the development of stages happened in an interval of time. For example, for the stages 0, I, II, III, and IV of a cancer, a latter stage of the cancer censors the previous stage and the change time of stage is observed between two examinations. By adding a terminal event, the other events including the multi-stage event will be censored by the occurrence of the terminal event no matter what stage the subject is at (e.g. death will censor all stages of a cancer). Our proposed model can also be modified to deal with recurring events.

With the fitted joint model and all its expansion forms, we can do the dynamic predictions for future events of interest by updating the event history. Compared to other joint methods in the literature, our model has the flexibility to choose a level of quantile for the longitudinal response and use the corresponding model for prediction. For example, if the latest longitudinal measurement falls in the first quarter of the distribution, we can set $\tau = 0.25$ to fit the joint model and then perform the dynamic prediction and/or detecting individuals with high risk scores. Moreover, the quantile level τ can be changed with time and newly updated history including newly measured longitudinal outcome.

Bibliography

Brown, B.M., Wang, Y.G., 2005. Standard errors and covariance matrices for smoothed rank estimators. Biometrika 92, 149–158.

Cox, D., 1972. Regression models and life-tables. Journal of the Royal Statistical Society. Series B 34, 187–220.

Elbers, C., Ridder, G., 1982. True and spurious duration dependence: the identifiability of the proportional hazard model. The Review of Economic Studies 49, 403–409.

Farcomeni, A., Viviani, S., 2015. Longitudinal quantile regression in the presence of informative dropout through longitudinal-survival joint modeling. Statistics in Medicine 34, 1199–1213.

Fu, L., Wang, Y.G., 2012. Quantile regression for longitudinal data with a working correlation model. Computational Statistics and Data Analysis 56, 2526–2538.

Gao, F., Zeng, D., Couper, D., Lin, D., 2018. Semiparametric regression analysis of multiple right- and interval-censored events. Journal of the American Statistical Association , online.

Geraci, M., Bottai, M., 2007. Quantile regression for longitudinal data using the asymmetric laplace distribution. Biostatistics 8, 140–154.

Geraci, M., Bottai, M., 2014. Linear quantile mixed models. Statistics and Computing 24, 461–479.

Hougaard, P., 2012. Analysis of multivariate survival data. New York: Springer .

Jung, S.H., 1996. Quasi-likelihood for median regression models. Journal of the American Statistical Association 91, 251–257.

Koenker, R., 2004. Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74–89.

Koenker, R., Bassett Jr, G., 1978. Regression quantiles. Econometrica 46, 33–55.

Koenker, R., D'Orey, V., 1987. Computing regression quantiles. Applied Statistics 36, 383–393.

Koenker, R., Machado, J., 1999. Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association 94, 1296–1309.

Leng, C., Zhang, W., 2012. Smoothing combined estimating equations in quantile regression for longitudinal data. Statistics and Computing 17 October 2012, 1–14.

Letenneur, L., Commenges, D., Dartigues, J.F., Barberger-Gateau, P., 1994. Incidence of dementia and alzheimer's disease in elderly community risidents of south-western france. International Journal of Epidemiology 23, 1256–1261.

Lipsitz, S.R., Fitzmaurice, G.M., Molenberghs, G., Zhao, L.P., 1997. Quantile regression methods for longitudinal data with drop-outs: application to cd4 cell counts of patients infected with the human immunodeficiency virus. Journal of the Royal Statistical Society: Series C 46, 463–476.

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

Liu, Y., Bottai, M., 2009. Mixed-effects models for conditional quantiles with longitudinal data. International Journal of Biostatistics 5, Article 28.

Lu, X., Fan, Z., 2015. Weighted quantile regression for longitudinal data. Computational Statistics 30, 569–592.

Murphy, S., 1994. Consistency in a proportional hazards model incorporating a random effect. The Annals of Statistics 22, 712–731.

Newey, W., Powell, J.L., 1987. Asymmetric least squares estimation and testing. Econometrica 55, 819–47.

Proust-Lima, C., Philipps, V., Diakite, A., Liquet, B., 2019. Extended mixed models using latent classes and latent processes. R Package lcmm, v1.8.1.

Tsiatis, A.A., Davidian, M., 2004. An overview of joint modeling of longitudinal and time-to-event data. Statistica Sinica 14, 739–818.

Van der Vaart, A., Wellner, J., 1996. Weak Convergence and Empirical Processes. volume New York. Springer-Verlag.

Zeng, D., Gao, F., Lin, D., 2017. Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data. Biometrika 104, 505–525.

Appendix

Table A.1: Biases (Bias), empirical standard deviations (SD) over 1000 simulations, averaged Bootstrap standard errors (SE) and coverage probabilities (CP) of the true value based on the 95% CI's for the estimates of parameters using our proposed joint model and separate models with fixed effects are reported. This table lists the results of sample size $n = 150$ and at three quantile levels $\tau = 0.25, 0.50$ and 0.85, where the longitudinal error follows a standard normal distribution.

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
150	0.25	η_1	0.005	0.056	0.057	0.962	0.000	0.046	0.038	0.874
		η_2	-0.004	0.051	0.051	0.943	0.000	0.047	0.038	0.862
		δ	0.007	0.048	0.047	0.935	0.000	0.013	0.011	0.841
		β_1	0.035	0.124	0.131	0.956	-0.073	0.116	0.115	0.891
		β_2	0.044	0.136	0.134	0.934	-0.044	0.124	0.122	0.929
		β_3	0.002	0.135	0.131	0.952	-0.050	0.113	0.112	0.912
		β_4	0.011	0.118	0.116	0.961	-0.006	0.115	0.116	0.953
		α_1	-0.003	0.107	0.105	0.930	-0.023	0.099	0.080	0.877
		α_2	-0.058	0.119	0.121	0.926	-0.069	0.117	0.066	0.665
		α_3	-0.012	0.147	0.144	0.954	-0.063	0.115	0.105	0.857
		α_4	0.011	0.120	0.119	0.932	-0.011	0.112	0.105	0.931

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		ζ_1	-0.013	0.158	0.163	0.934				
		ζ_2	0.013	0.179	0.182	0.955				
		ζ_3	-0.099	0.156	0.162	0.934				
		ζ_4	-0.023	0.169	0.171	0.951				
		ϱ								
		σ^2	-0.014	0.054	0.056	0.975				
150	0.50	η_1	0.000	0.043	0.040	0.941	0.001	0.043	0.035	0.870
		η_2	-0.011	0.048	0.044	0.924	0.000	0.043	0.035	0.885
		δ	0.003	0.042	0.042	0.932	0.000	0.012	0.010	0.846
		β_1	0.058	0.145	0.148	0.915	-0.069	0.115	0.115	0.896
		β_2	0.058	0.145	0.140	0.924	-0.043	0.126	0.122	0.919
		β_3	-0.026	0.112	0.116	0.949	-0.047	0.118	0.111	0.901
		β_4	0.027	0.127	0.130	0.941	-0.014	0.117	0.115	0.942
		α_1	0.006	0.102	0.107	0.958	-0.023	0.107	0.081	0.850
		α_2	-0.062	0.129	0.127	0.927	-0.066	0.110	0.067	0.692
		α_3	-0.009	0.134	0.135	0.958	-0.065	0.112	0.105	0.870
		α_4	0.035	0.124	0.131	0.966	-0.009	0.110	0.105	0.939
		ζ_1	0.012	0.143	0.144	0.924				
		ζ_2	0.024	0.199	0.198	0.932				
		ζ_3	-0.131	0.160	0.171	0.937				
		ζ_4	0.017	0.183	0.181	0.949				

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		ϱ								
		σ^2	-0.009	0.049	0.048	0.941				
150	0.85	η_1	0.009	0.060	0.060	0.964	-0.002	0.050	0.040	0.868
		η_2	-0.001	0.064	0.067	0.973	-0.001	0.053	0.042	0.853
		δ	0.006	0.048	0.047	0.939	0.000	0.015	0.013	0.833
		β_1	0.042	0.126	0.127	0.935	-0.069	0.118	0.115	0.885
		β_2	0.056	0.129	0.128	0.937	-0.043	0.123	0.122	0.935
		β_3	-0.012	0.125	0.122	0.951	-0.049	0.117	0.111	0.902
		β_4	-0.001	0.137	0.131	0.944	-0.008	0.117	0.115	0.947
		α_1	-0.002	0.107	0.103	0.923	-0.023	0.102	0.080	0.857
		α_2	-0.056	0.126	0.127	0.928	-0.073	0.115	0.066	0.649
		α_3	-0.003	0.141	0.141	0.953	-0.065	0.119	0.105	0.857
		α_4	0.011	0.108	0.111	0.951	-0.009	0.112	0.105	0.930
		ζ_1	-0.001	0.153	0.154	0.945				
		ζ_2	0.031	0.175	0.175	0.953				
		ζ_3	-0.080	0.178	0.179	0.921				
		ζ_4	-0.003	0.183	0.179	0.957				
		ϱ								
		σ^2	-0.008	0.051	0.051	0.958				

Table A.2: Biases (Bias), empirical standard deviations (SD) over 1000 simulations, averaged Bootstrap standard errors (SE) and coverage probabilities (CP) of the true value based on the 95% CI's for the estimates of parameters using our proposed joint model and separate models with fixed effects are reported. This table lists the results of sample size $n = 300$ and at three quantile levels $\tau = 0.25, 0.50$ and 0.85, where the longitudinal error follows a standard normal distribution.

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
300	0.25	η_1	0.003	0.039	0.040	0.972	0.000	0.031	0.033	0.959
		η_2	-0.003	0.036	0.035	0.937	0.000	0.032	0.032	0.944
		δ	0.006	0.034	0.034	0.935	0.000	0.009	0.009	0.933
		β_1	0.043	0.094	0.091	0.908	-0.079	0.083	0.080	0.810
		β_2	0.020	0.105	0.106	0.941	-0.057	0.087	0.084	0.875
		β_3	-0.010	0.083	0.081	0.946	-0.052	0.080	0.077	0.882
		β_4	0.004	0.089	0.090	0.961	-0.012	0.081	0.080	0.949
		α_1	-0.025	0.093	0.093	0.930	-0.041	0.083	0.053	0.778
		α_2	-0.096	0.102	0.100	0.935	-0.108	0.097	0.043	0.417
		α_3	-0.022	0.085	0.086	0.935	-0.080	0.080	0.071	0.743
		α_4	0.001	0.090	0.089	0.969	-0.017	0.079	0.072	0.917
		ζ_1	0.003	0.114	0.117	0.959				
		ζ_2	0.030	0.128	0.127	0.947				
		ζ_3	-0.077	0.108	0.115	0.929				
		ζ_4	-0.009	0.132	0.131	0.951				
		ϱ								
		σ^2	-0.020	0.039	0.039	0.948				

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
300	0.50	η_1	0.004	0.036	0.036	0.953	-0.001	0.029	0.030	0.952
		η_2	0.003	0.033	0.033	0.943	-0.001	0.030	0.030	0.947
		δ	0.003	0.029	0.030	0.936	0.000	0.009	0.008	0.924
		β_1	0.028	0.088	0.089	0.937	-0.082	0.081	0.080	0.809
		β_2	0.010	0.104	0.107	0.964	-0.050	0.088	0.085	0.903
		β_3	-0.028	0.086	0.082	0.919	-0.053	0.081	0.077	0.882
		β_4	0.001	0.089	0.088	0.968	-0.012	0.082	0.080	0.934
		α_1	-0.027	0.092	0.089	0.924	-0.045	0.081	0.053	0.765
		α_2	-0.098	0.093	0.094	0.933	-0.108	0.095	0.043	0.416
		α_3	-0.016	0.090	0.088	0.950	-0.081	0.074	0.071	0.764
		α_4	0.006	0.092	0.090	0.947	-0.019	0.079	0.072	0.932
		ζ_1	-0.006	0.094	0.098	0.956				
		ζ_2	0.034	0.117	0.117	0.931				
		ζ_3	-0.080	0.104	0.105	0.916				
		ζ_4	-0.011	0.117	0.118	0.943				
		ϱ								
		σ^2	-0.025	0.035	0.034	0.912				
300	0.85	η_1	0.000	0.038	0.037	0.945	0.001	0.037	0.037	0.948
		η_2	0.002	0.037	0.037	0.941	0.000	0.037	0.036	0.937
		δ	0.001	0.031	0.030	0.943	0.000	0.010	0.009	0.898
		β_1	0.027	0.094	0.096	0.946	-0.081	0.083	0.080	0.797
		β_2	0.016	0.101	0.104	0.971	-0.049	0.090	0.084	0.887

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		β_3	-0.017	0.083	0.081	0.962	-0.053	0.078	0.077	0.873
		β_4	0.000	0.081	0.080	0.961	-0.009	0.082	0.080	0.938
		α_1	-0.026	0.098	0.097	0.923	-0.042	0.082	0.053	0.761
		α_2	-0.084	0.101	0.102	0.937	-0.101	0.094	0.043	0.467
		α_3	-0.018	0.078	0.080	0.953	-0.076	0.078	0.072	0.751
		α_4	0.002	0.087	0.088	0.950	-0.015	0.078	0.072	0.926
		ζ_1	-0.006	0.101	0.102	0.945				
		ζ_2	0.026	0.129	0.130	0.943				
		ζ_3	-0.079	0.109	0.110	0.922				
		ζ_4	-0.008	0.132	0.131	0.938				
		ϱ								
		σ^2	-0.023	0.035	0.035	0.933				

Table A.3: Biases (Bias), empirical standard deviations (SD) over 1000 simulations, averaged Bootstrap standard errors (SE) and coverage probabilities (CP) of the true value based on the 95% CI's for the estimates of parameters using our proposed joint model and separate models with fixed effects are reported. This table lists the results of sample size $n = 150$ and at three quantile levels $\tau = 0.25, 0.50$ and 0.85, where the longitudinal error follows a standard normal distribution and the last symptomatic event is terminal.

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
150	0.25	η_1	0.006	0.057	0.055	0.943	-0.001	0.045	0.038	0.869
		η_2	0.001	0.059	0.060	0.965	-0.003	0.046	0.037	0.855
		δ	0.006	0.057	0.056	0.934	-0.001	0.014	0.011	0.823
		β_1	0.077	0.165	0.163	0.926	-0.055	0.139	0.134	0.912
		β_2	0.070	0.171	0.187	0.947	-0.016	0.152	0.151	0.952
		β_3	0.012	0.166	0.169	0.952	-0.041	0.143	0.139	0.929
		β_4	0.025	0.132	0.132	0.943	-0.011	0.122	0.115	0.932
		α_1	0.056	0.158	0.155	0.930	-0.017	0.149	0.147	0.942
		α_2	0.036	0.154	0.156	0.957	-0.010	0.143	0.141	0.954
		α_3	0.011	0.191	0.181	0.955	-0.047	0.166	0.160	0.918
		α_4	0.006	0.115	0.113	0.950	-0.005	0.108	0.105	0.941
		ζ_1	0.037	0.194	0.208	0.961				
		ζ_2	0.033	0.213	0.217	0.970				
		ζ_3	-0.045	0.212	0.193	0.936				
		ζ_4	-0.028	0.193	0.204	0.948				
		ϱ								

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		σ^2	-0.006	0.049	0.051	0.965				
150	0.50	η_1	0.004	0.058	0.055	0.945	-0.001	0.041	0.035	0.886
		η_2	0.004	0.051	0.047	0.946	0.000	0.043	0.035	0.859
		δ	0.008	0.062	0.062	0.944	0.000	0.012	0.010	0.844
		β_1	0.065	0.168	0.182	0.940	-0.049	0.139	0.135	0.914
		β_2	0.049	0.175	0.182	0.953	-0.029	0.147	0.149	0.948
		β_3	0.001	0.160	0.168	0.961	-0.032	0.144	0.139	0.931
		β_4	0.011	0.118	0.119	0.954	-0.002	0.117	0.115	0.944
		α_1	0.027	0.167	0.163	0.948	-0.022	0.146	0.147	0.946
		α_2	0.039	0.160	0.156	0.951	-0.018	0.138	0.140	0.942
		α_3	-0.001	0.179	0.170	0.943	-0.056	0.170	0.159	0.895
		α_4	0.018	0.127	0.121	0.933	-0.007	0.107	0.105	0.945
		ζ_1	0.020	0.193	0.198	0.945				
		ζ_2	0.077	0.224	0.222	0.940				
		ζ_3	-0.040	0.201	0.206	0.956				
		ζ_4	-0.024	0.174	0.174	0.954				
		ϱ								
		σ^2	-0.001	0.045	0.048	0.965				
150	0.85	η_1	-0.005	0.067	0.065	0.955	-0.001	0.053	0.040	0.840
		η_2	-0.005	0.057	0.057	0.954	-0.001	0.051	0.042	0.877
		δ	-0.002	0.066	0.064	0.920	-0.003	0.015	0.012	0.818
		β_1	0.049	0.152	0.175	0.953	-0.059	0.137	0.134	0.902

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		β_2	0.043	0.173	0.182	0.954	-0.021	0.155	0.150	0.943
		β_3	0.008	0.156	0.147	0.935	-0.039	0.145	0.140	0.930
		β_4	0.008	0.113	0.108	0.925	-0.002	0.125	0.116	0.928
		α_1	0.045	0.174	0.152	0.927	-0.020	0.153	0.147	0.931
		α_2	0.046	0.163	0.192	0.968	-0.012	0.139	0.141	0.941
		α_3	0.018	0.173	0.172	0.945	-0.047	0.168	0.159	0.916
		α_4	0.024	0.132	0.128	0.945	-0.008	0.111	0.105	0.932
		ζ_1	0.039	0.190	0.197	0.952				
		ζ_2	0.040	0.240	0.236	0.961				
		ζ_3	-0.035	0.189	0.190	0.960				
		ζ_4	-0.048	0.184	0.209	0.974				
		ϱ								
		σ^2	-0.007	0.044	0.044	0.965				

Table A.4: Biases (Bias), empirical standard deviations (SD) over 1000 simulations, averaged Bootstrap standard errors (SE) and coverage probabilities (CP) of the true value based on the 95% CI's for the estimates of parameters using our proposed joint model and separate models with fixed effects are reported. This table lists the results of sample size $n = 300$ and at three quantile levels $\tau = 0.25, 0.50$ and 0.85, where the longitudinal error follows a standard normal distribution and the last symptomatic event is terminal.

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
300	0.25	η_1	-0.004	0.041	0.041	0.950	-0.004	0.032	0.033	0.951
		η_2	-0.003	0.037	0.035	0.936	0.000	0.033	0.032	0.940
		δ	0.003	0.038	0.039	0.964	0.000	0.009	0.009	0.917
		β_1	0.018	0.108	0.114	0.963	-0.061	0.095	0.093	0.882
		β_2	0.010	0.122	0.120	0.961	-0.035	0.104	0.104	0.936
		β_3	-0.004	0.109	0.109	0.955	-0.046	0.100	0.097	0.905
		β_4	0.003	0.087	0.090	0.958	-0.012	0.078	0.080	0.952
		α_1	0.010	0.103	0.101	0.948	-0.025	0.103	0.101	0.940
		α_2	0.012	0.102	0.105	0.946	-0.020	0.098	0.097	0.941
		α_3	-0.005	0.124	0.115	0.935	-0.064	0.110	0.109	0.903
		α_4	-0.008	0.093	0.097	0.966	-0.015	0.078	0.072	0.929
		ζ_1	-0.011	0.136	0.136	0.945				
		ζ_2	0.028	0.149	0.149	0.944				
		ζ_3	-0.039	0.133	0.131	0.935				
		ζ_4	-0.027	0.119	0.119	0.932				
		ϱ								

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		σ^2	-0.023	0.032	0.033	0.920				
300	0.50	η_1	-0.002	0.038	0.038	0.940	0.001	0.030	0.030	0.949
		η_2	0.003	0.035	0.035	0.953	-0.001	0.030	0.030	0.948
		δ	0.002	0.035	0.034	0.934	0.000	0.009	0.008	0.921
		β_1	0.028	0.110	0.110	0.932	-0.061	0.095	0.093	0.882
		β_2	0.038	0.121	0.120	0.935	-0.037	0.106	0.104	0.928
		β_3	-0.012	0.098	0.102	0.955	-0.043	0.100	0.097	0.914
		β_4	-0.009	0.077	0.078	0.951	-0.009	0.082	0.080	0.932
		α_1	0.006	0.109	0.102	0.953	-0.017	0.102	0.101	0.934
		α_2	0.021	0.105	0.111	0.964	-0.025	0.099	0.097	0.940
		α_3	0.001	0.122	0.110	0.938	-0.066	0.111	0.108	0.877
		α_4	-0.006	0.091	0.087	0.958	-0.015	0.077	0.072	0.933
		ζ_1	-0.003	0.137	0.137	0.946				
		ζ_2	-0.034	0.139	0.123	0.921				
		ζ_3	-0.048	0.142	0.134	0.927				
		ζ_4	-0.030	0.128	0.129	0.944				
		ϱ								
		σ^2	-0.021	0.032	0.032	0.907				
300	0.85	η_1	0.001	0.049	0.046	0.950	-0.004	0.037	0.037	0.945
		η_2	0.004	0.042	0.043	0.960	0.002	0.036	0.036	0.951
		δ	-0.006	0.042	0.042	0.944	0.001	0.010	0.009	0.897
		β_1	0.031	0.113	0.123	0.969	-0.066	0.098	0.093	0.868

Continued on next page

			Joint Model				Separate Model			
n	τ		Bias	SD	SE	CP	Bias	SD	SE	CP
		β_2	0.041	0.124	0.131	0.933	-0.039	0.103	0.104	0.928
		β_3	-0.001	0.101	0.100	0.955	-0.043	0.104	0.097	0.904
		β_4	-0.010	0.081	0.089	0.960	-0.010	0.080	0.080	0.948
		α_1	0.015	0.110	0.100	0.935	-0.027	0.100	0.101	0.935
		α_2	0.022	0.104	0.095	0.927	-0.019	0.096	0.097	0.929
		α_3	0.016	0.126	0.128	0.921	-0.066	0.108	0.108	0.887
		α_4	0.017	0.094	0.100	0.954	-0.012	0.077	0.072	0.926
		ζ_1	0.016	0.138	0.137	0.950				
		ζ_2	0.039	0.154	0.157	0.935				
		ζ_3	-0.021	0.139	0.135	0.938				
		ζ_4	-0.028	0.127	0.129	0.945				
		ϱ								
		σ^2	-0.018	0.029	0.027	0.902				

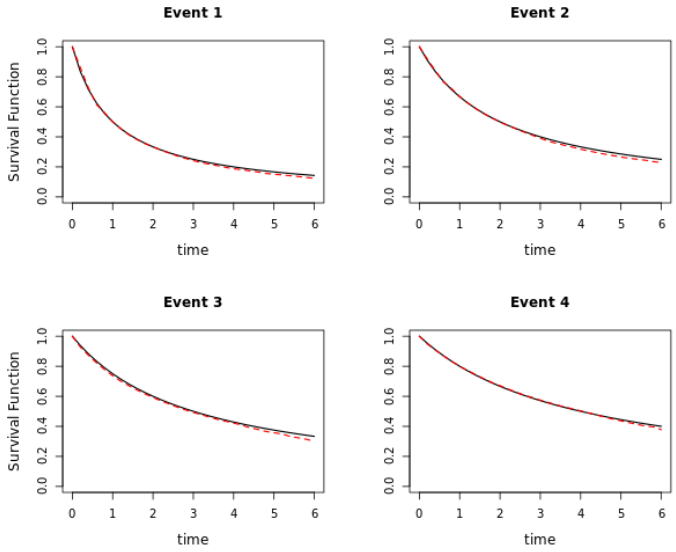

Figure A.1: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.50$ and ALD distributed longitudinal response variable.

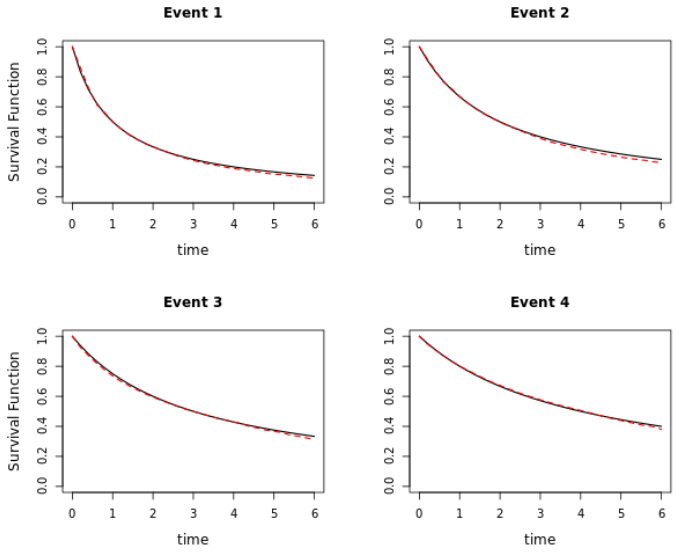

Figure A.2: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.85$ and ALD distributed longitudinal response variable.

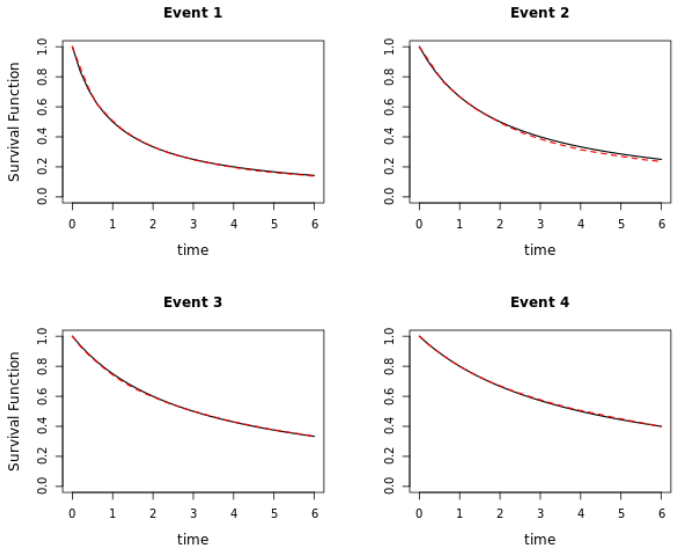

Figure A.3: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.25$ and ALD distributed longitudinal response variable.

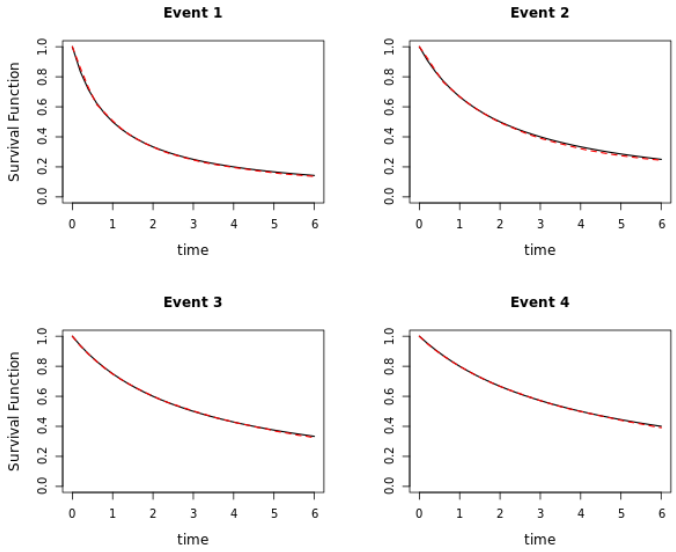

Figure A.4: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.50$ and ALD distributed longitudinal response variable.

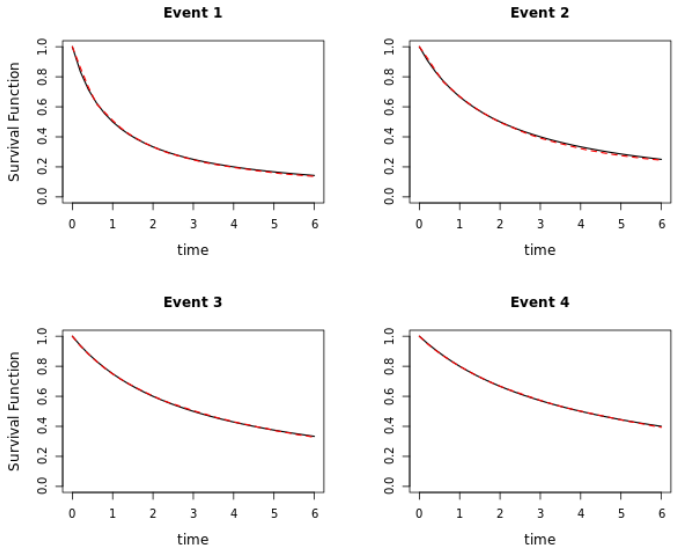

Figure A.5: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.85$ and ALD distributed longitudinal response variable.

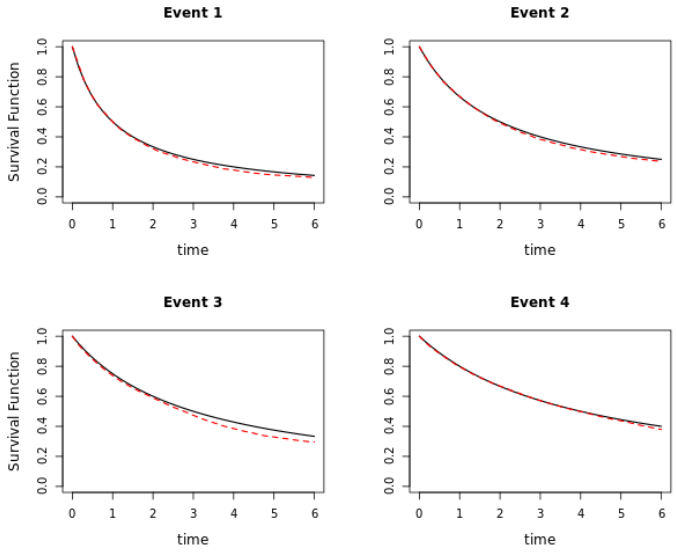

Figure A.6: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.25$ and ALD distributed longitudinal response variable.

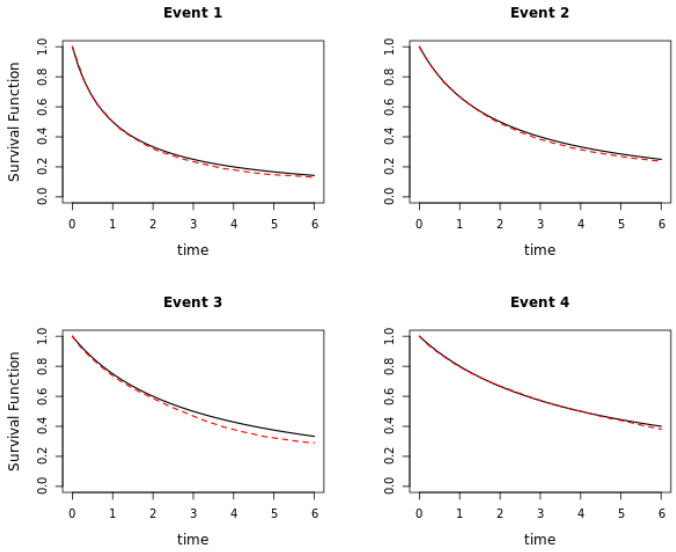

Figure A.7: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.50$ and ALD distributed longitudinal response variable.

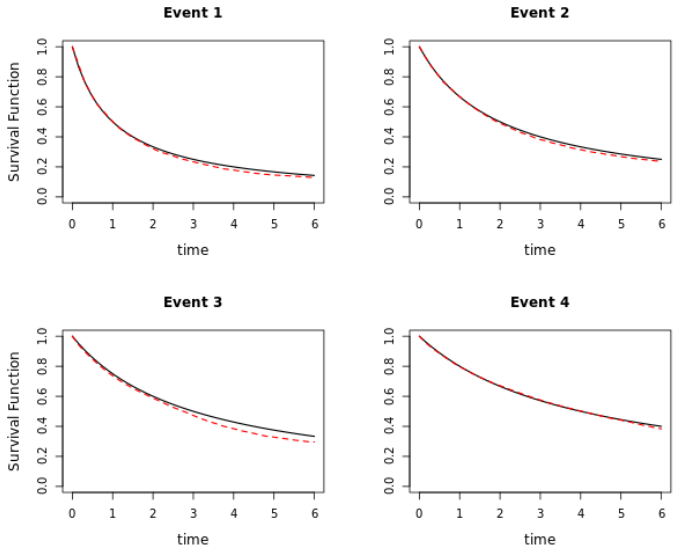

Figure A.8: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.85$ and ALD distributed longitudinal response variable.

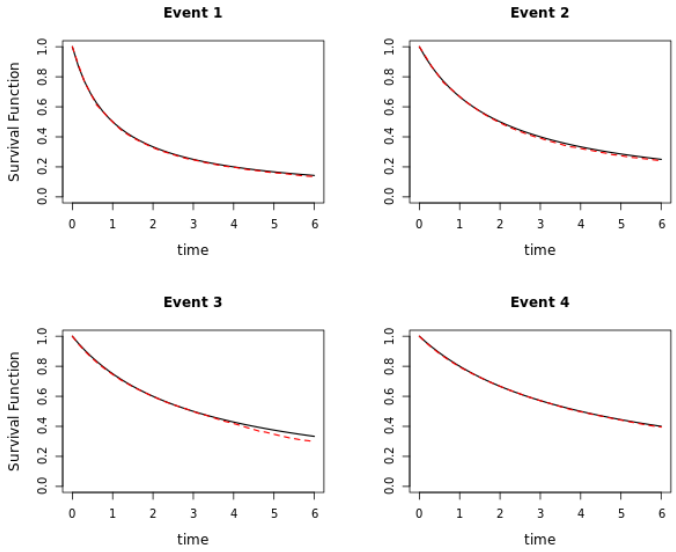

Figure A.9: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.25$ and ALD distributed longitudinal response variable.

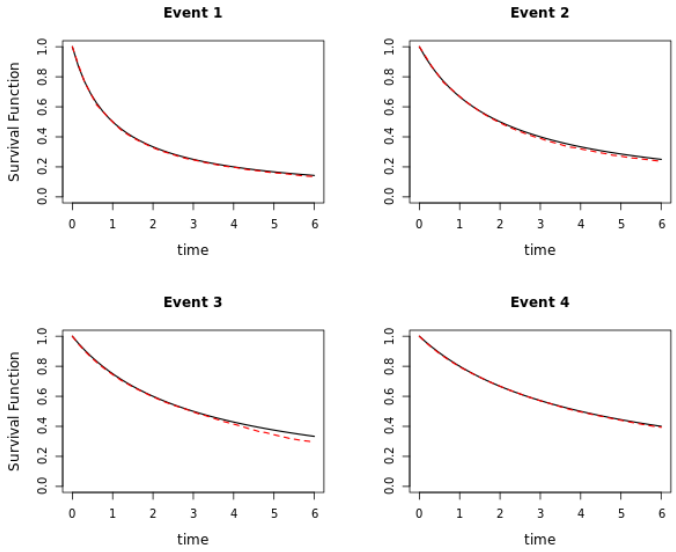

Figure A.10: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.50$ and ALD distributed longitudinal response variable.

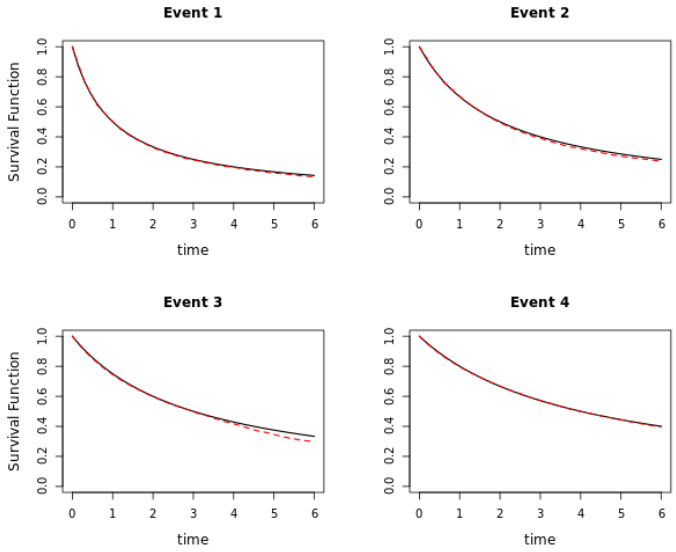

Figure A.11: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.85$ and ALD distributed longitudinal response variable.

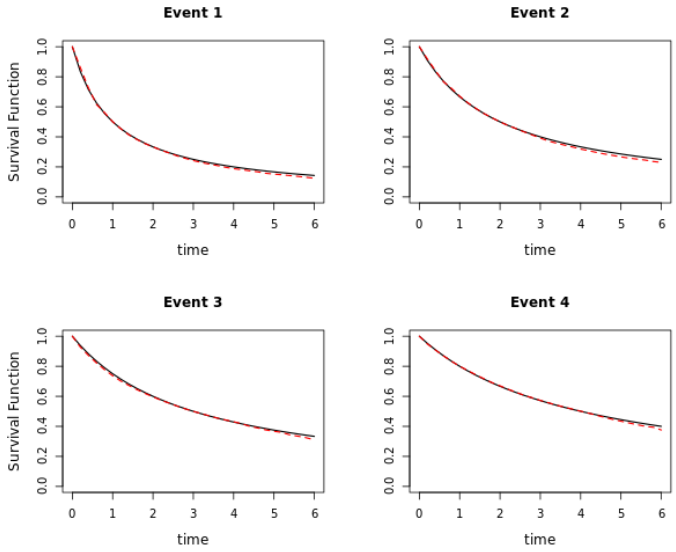

Figure A.12: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.25$ and Normal distributed longitudinal response variable.

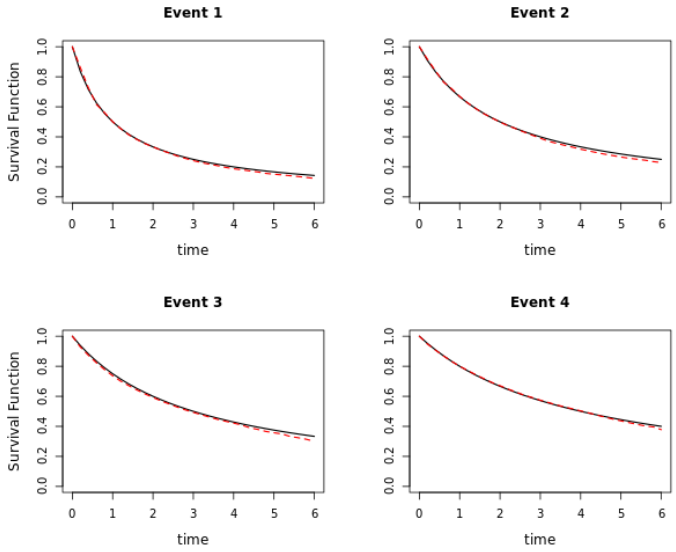

Figure A.13: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.50$ and Normal distributed longitudinal response variable.

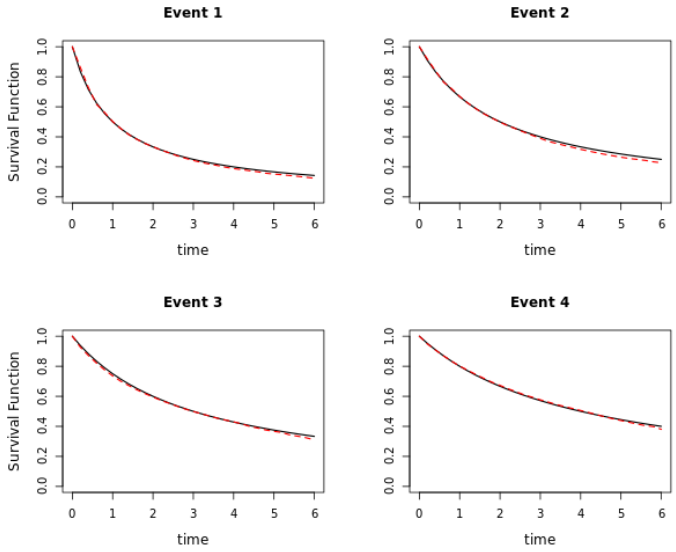

Figure A.14: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.85$ and Normal distributed longitudinal response variable.

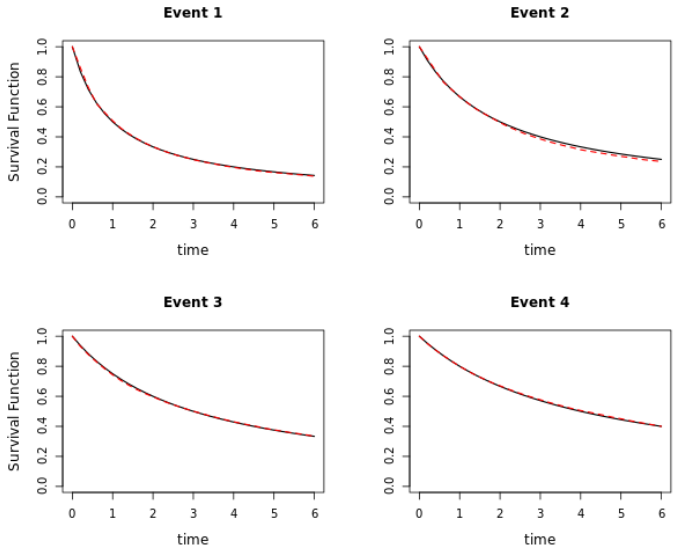

Figure A.15: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.25$ and Normal distributed longitudinal response variable.

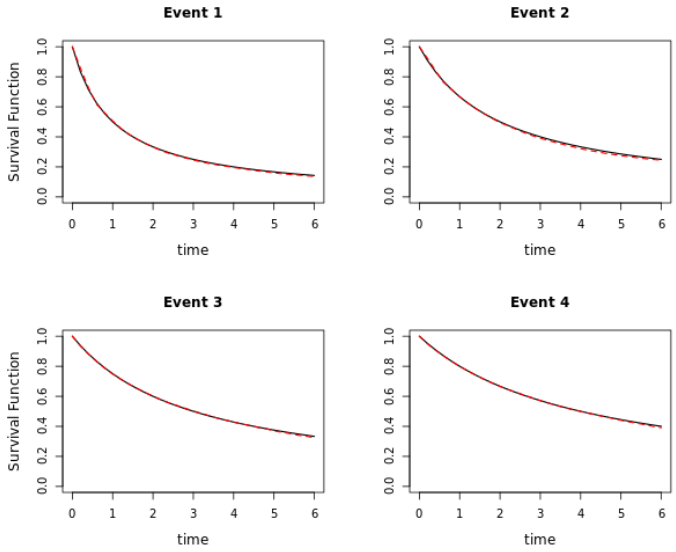

Figure A.16: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.50$ and Normal distributed longitudinal response variable.

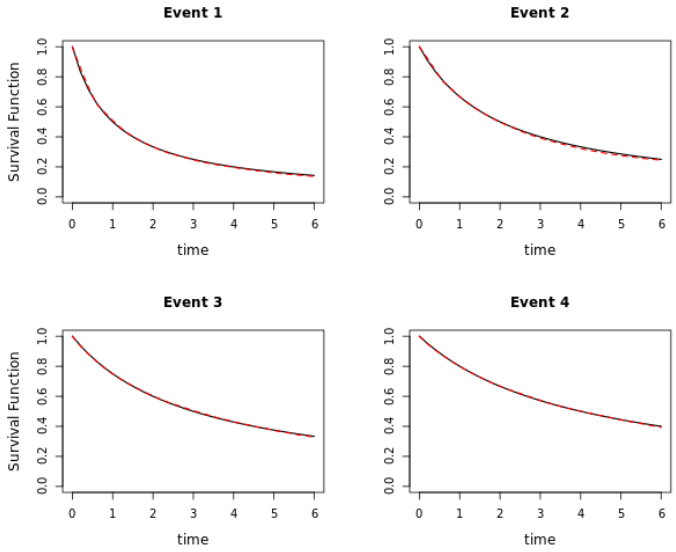

Figure A.17: The estimation of baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.85$ and Normal distributed longitudinal response variable.

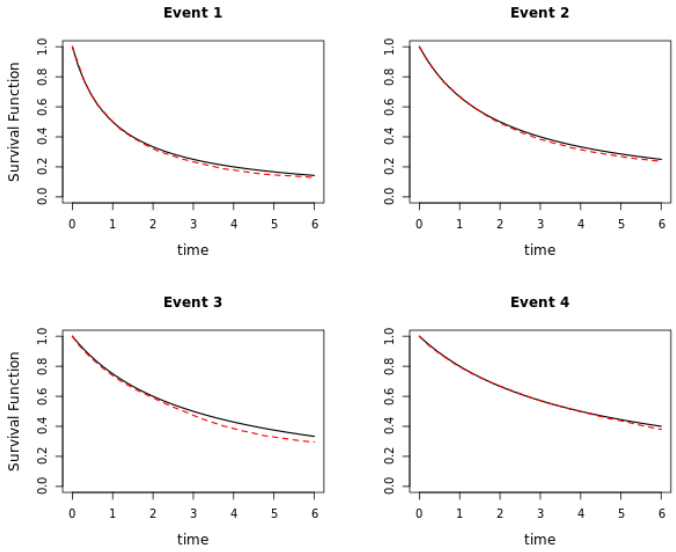

Figure A.18: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.25$ and Normal distributed longitudinal response variable.

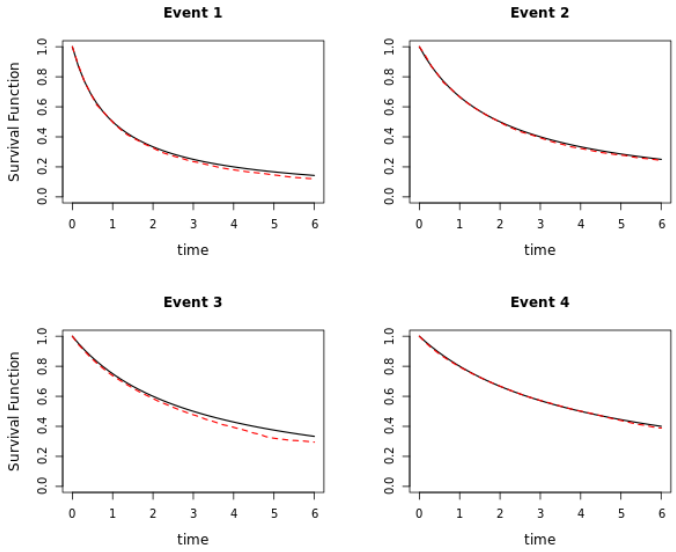

Figure A.19: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.50$ and Normal distributed longitudinal response variable.

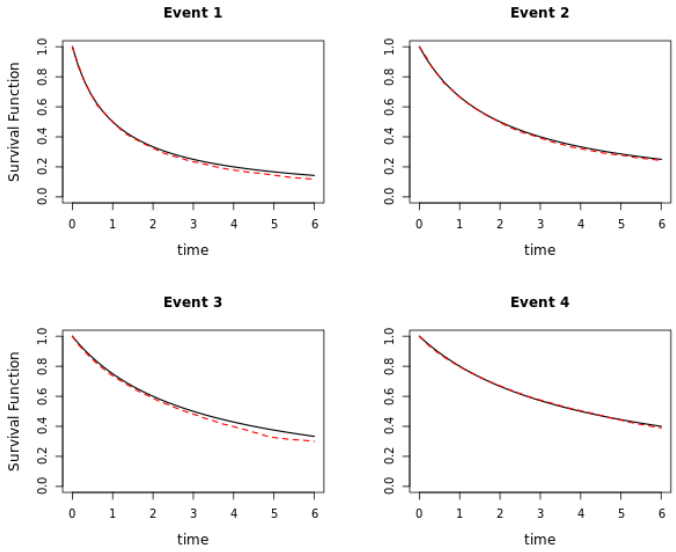

Figure A.20: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 150$, $\tau = 0.85$ and Normal distributed longitudinal response variable.

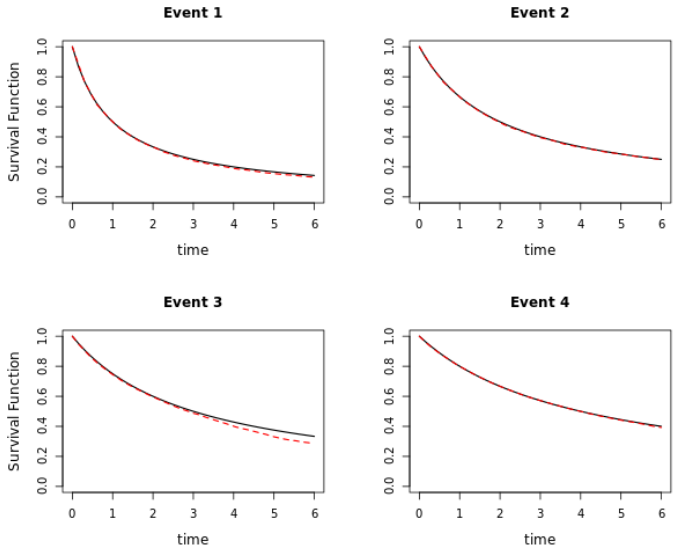

Figure A.21: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.25$ and Normal distributed longitudinal response variable.

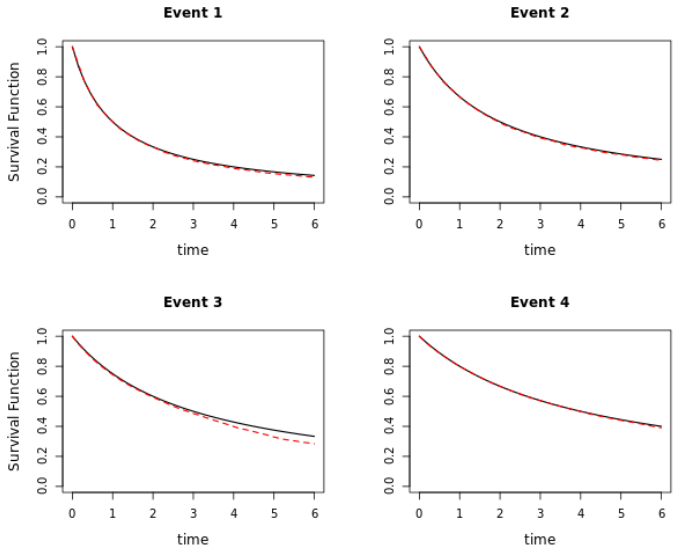

Figure A.22: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.50$ and Normal distributed longitudinal response variable.

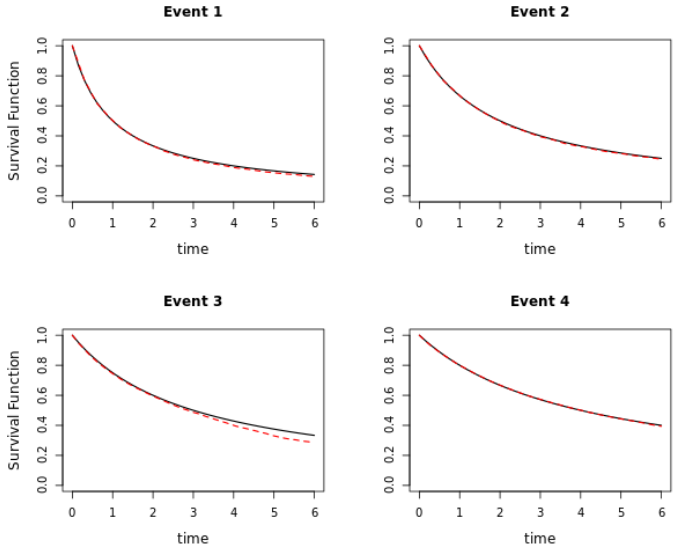

Figure A.23: The estimation of semi-competing baseline survival functions average over 1000 simulations for a sample size of $n = 300$, $\tau = 0.85$ and Normal distributed longitudinal response variable.

Table A.5: Estimates of regression and dispersion parameters with their bootstrapped standard errors (SD) and 95% confidence intervals (LL: lower limit; UL: upper limit) for the joint model of Isaacs Set Test scores, dementia time and death time, at different levels of τ.

		τ						
		0.10	0.15	0.25	0.35	0.50	0.75	0.85
${\eta_1}^*$		23.6644	24.5815	26.3962	27.8512	29.9704	33.1576	34.2338
	SD	0.3717	0.3757	0.3565	0.3026	0.2763	0.2401	0.3486
	LL	22.9359	23.8452	25.6975	27.2581	29.4287	32.6870	33.5506
	UL	24.3929	25.3179	27.0949	28.4443	30.5120	33.6281	34.9171
${\eta_2}^*$		2.3521	2.5598	2.7014	2.6155	2.7481	3.9115	4.7353
	SD	0.4275	0.4473	0.4071	0.3502	0.3305	0.3335	0.4000
	LL	1.5141	1.6831	1.9035	1.9292	2.1002	3.2579	3.9513
	UL	3.1900	3.4364	3.4994	3.3018	3.3959	4.5651	5.5192
δ^*		-4.0997	-3.8345	-3.6812	-3.4971	-3.6152	-3.7238	-3.4048
	SD	0.0833	0.0919	0.1031	0.1104	0.1150	0.1078	0.0986
	LL	-4.2630	-4.0146	-3.8832	-3.7136	-3.8405	-3.9351	-3.5979
	UL	-3.9364	-3.6544	-3.4792	-3.2806	-3.3898	-3.5125	-3.2116

Continued on next page

		τ						
		0.10	0.15	0.25	0.35	0.50	0.75	0.85
β_1		-0.3512	-0.3519	-0.3552	-0.3435	-0.3553	-0.3329	-0.3700
	SD	0.2614	0.2647	0.2778	0.2676	0.2801	0.2723	0.2755
	LL	-0.8635	-0.8707	-0.8997	-0.8680	-0.9044	-0.8666	-0.9099
	UL	0.1611	0.1669	0.1893	0.1811	0.1937	0.2009	0.1698
$\beta_2{}^*$		0.5184	0.5177	0.5107	0.5117	0.5026	0.5254	0.4905
	SD	0.1175	0.1195	0.1178	0.1171	0.1181	0.1179	0.1182
	LL	0.2881	0.2836	0.2798	0.2822	0.2712	0.2942	0.2588
	UL	0.7487	0.7519	0.7415	0.7412	0.7341	0.7566	0.7222
$\alpha_1{}^*$		-0.1410	-0.1522	-0.1362	-0.1499	-0.1310	-0.1501	-0.1415
	SD	0.0149	0.0160	0.0178	0.0180	0.0183	0.0170	0.0194
	LL	-0.1702	-0.1836	-0.1712	-0.1852	-0.1670	-0.1833	-0.1795
	UL	-0.1117	-0.1209	-0.1012	-0.1147	-0.0951	-0.1168	-0.1036
$\alpha_2{}^*$		-0.1117	-0.1200	-0.1085	-0.1188	-0.1032	-0.1193	-0.1128
	SD	0.0085	0.0091	0.0095	0.0102	0.0098	0.0094	0.0105
	LL	-0.1283	-0.1378	-0.1272	-0.1388	-0.1224	-0.1377	-0.1334

Continued on next page

		τ						
		0.10	0.15	0.25	0.35	0.50	0.75	0.85
	UL	-0.0950	-0.1022	-0.0899	-0.0987	-0.0840	-0.1009	-0.0922
$\zeta_1{}^*$		1.0829	1.0598	1.0843	1.1143	1.0896	1.0969	1.0892
	SD	0.1128	0.1159	0.1250	0.1180	0.1262	0.1242	0.1279
	LL	0.8617	0.8327	0.8394	0.8830	0.8422	0.8534	0.8385
	UL	1.3040	1.2870	1.3293	1.3456	1.3369	1.3403	1.3398
$\zeta_2{}^*$		0.8489	0.8536	0.8663	0.8714	0.8815	0.8735	0.8833
	SD	0.0880	0.0863	0.0937	0.0931	0.0943	0.0939	0.0964
	LL	0.6764	0.6845	0.6826	0.6889	0.6966	0.6894	0.6943
	UL	1.0214	1.0227	1.0499	1.0539	1.0663	1.0575	1.0723
ϱ^*		1.0011	1.3264	1.7901	2.0688	2.2201	1.7529	1.2501
	SD	0.0232	0.0310	0.0403	0.0458	0.0501	0.0423	0.0318
	LL	0.9557	1.2656	1.7111	1.9790	2.1219	1.6700	1.1877
	UL	1.0465	1.3872	1.8692	2.1586	2.3183	1.8358	1.3124
σ^{2*}		1.5611	1.5975	1.4169	1.4286	1.3870	1.3962	1.3570

Continued on next page

	τ						
	0.10	0.15	0.25	0.35	0.50	0.75	0.85
SD	0.7713	0.7892	0.6745	0.6829	0.6435	0.6703	0.6358
LL	0.0494	0.0501	0.0949	0.0902	0.1258	0.0824	0.1108
UL	3.0727	3.2000	2.7390	2.7670	2.6483	2.7100	2.6033

*Estimates in the row are significant at the 5% level for all quantiles τ.

Table A.6: Estimates of regression and dispersion parameters with their bootstrapped standard errors (SD) and 95% confidence intervals (LL: lower limit; UL: upper limit) for the joint model of Isaacs Set Test scores, dementia time, death time and dependency status, at different levels of τ.

		τ						
		0.10	0.15	0.25	0.35	0.50	0.75	0.85
${\eta_1}^*$		23.6521	24.5827	26.3963	27.8508	29.9732	33.1842	34.2020
	SD	0.3712	0.3756	0.3565	0.3027	0.2766	0.2374	0.3522
	LL	22.9244	23.8465	25.6976	27.2575	29.4311	32.7189	33.5117
	UL	24.3797	25.3189	27.0950	28.4440	30.5153	33.6495	34.8923
${\eta_2}^*$		2.3596	2.5601	2.7019	2.6158	2.7485	3.9276	4.7310
	SD	0.4274	0.4472	0.4071	0.3502	0.3308	0.3297	0.4040
	LL	1.5219	1.6836	1.9040	1.9294	2.1002	3.2814	3.9392
	UL	3.1974	3.4366	3.4998	3.3023	3.3968	4.5739	5.5229
δ^*		-4.0947	-3.8353	-3.6815	-3.4968	-3.6172	-3.7552	-3.3843
	SD	0.0838	0.0914	0.1030	0.1098	0.1156	0.1072	0.0995
	LL	-4.2591	-4.0144	-3.8834	-3.7120	-3.8438	-3.9653	-3.5793
	UL	-3.9304	-3.6562	-3.4795	-3.2815	-3.3907	-3.5451	-3.1892

Continued on next page

		τ						
		0.10	0.15	0.25	0.35	0.50	0.75	0.85
$\beta_1{}^*$		-0.5231	-0.5462	-0.5466	-0.5115	-0.5088	-0.5526	-0.5305
	SD	0.1979	0.1919	0.1996	0.1969	0.2042	0.1892	0.2082
	LL	-0.9110	-0.9224	-0.9378	-0.8974	-0.9091	-0.9234	-0.9386
	UL	-0.1352	-0.1700	-0.1553	-0.1255	-0.1085	-0.1817	-0.1224
$\beta_2{}^*$		0.4634	0.4605	0.4369	0.4795	0.4657	0.4608	0.4399
	SD	0.1237	0.1236	0.1224	0.1240	0.1220	0.1233	0.1241
	LL	0.2208	0.2181	0.1970	0.2365	0.2266	0.2191	0.1966
	UL	0.7059	0.7028	0.6767	0.7226	0.7047	0.7024	0.6832
β_3		-0.2252	-0.2378	-0.2459	-0.2011	-0.2073	-0.2394	-0.2465
	SD	0.1953	0.1869	0.2034	0.1865	0.2011	0.1808	0.2117
	LL	-0.6079	-0.6042	-0.6446	-0.5668	-0.6013	-0.5937	-0.6614
	UL	0.1575	0.1286	0.1529	0.1645	0.1868	0.1149	0.1685
$\alpha_1{}^*$		-0.1065	-0.1195	-0.1107	-0.1286	-0.1165	-0.1219	-0.1162
	SD	0.0095	0.0101	0.0116	0.0114	0.0113	0.0103	0.0135
	LL	-0.1252	-0.1394	-0.1335	-0.1509	-0.1387	-0.1421	-0.1427

Continued on next page

		τ						
		0.10	0.15	0.25	0.35	0.50	0.75	0.85
	UL	-0.0878	-0.0996	-0.0880	-0.1062	-0.0943	-0.1017	-0.0896
$\alpha_2{}^*$		-0.0847	-0.0937	-0.0858	-0.1019	-0.0931	-0.0977	-0.0905
	SD	0.0068	0.0074	0.0079	0.0081	0.0079	0.0075	0.0087
	LL	-0.0980	-0.1081	-0.1013	-0.1177	-0.1086	-0.1125	-0.1076
	UL	-0.0714	-0.0793	-0.0703	-0.0861	-0.0776	-0.0829	-0.0733
$\alpha_3{}^*$		-0.0965	-0.1089	-0.0990	-0.1171	-0.1048	-0.1107	-0.1042
	SD	0.0092	0.0094	0.0096	0.0097	0.0105	0.0083	0.0127
	LL	-0.1146	-0.1273	-0.1178	-0.1361	-0.1253	-0.1270	-0.1291
	UL	-0.0785	-0.0905	-0.0802	-0.0981	-0.0842	-0.0944	-0.0793
$\zeta_1{}^*$		1.7782	1.8045	1.7321	1.7760	1.7571	1.8048	1.7096
	SD	0.0512	0.0473	0.0506	0.0480	0.0521	0.0478	0.0592
	LL	1.6778	1.7119	1.6330	1.6820	1.6550	1.7111	1.5936
	UL	1.8786	1.8972	1.8312	1.8699	1.8592	1.8986	1.8257
$\zeta_2{}^*$		0.8259	0.8418	0.8340	0.8181	0.8154	0.7992	0.8419
	SD	0.0722	0.0706	0.0737	0.0712	0.0748	0.0710	0.0760

Continued on next page

		τ						
		0.10	0.15	0.25	0.35	0.50	0.75	0.85
	LL	0.6843	0.7033	0.6895	0.6786	0.6688	0.6601	0.6929
	UL	0.9675	0.9802	0.9785	0.9577	0.9620	0.9382	0.9908
$\zeta_3{}^*$		1.7542	1.7624	1.7176	1.7425	1.7396	1.7575	1.7052
	SD	0.0648	0.0616	0.0662	0.0595	0.0625	0.0619	0.0691
	LL	1.6271	1.6416	1.5877	1.6259	1.6171	1.6361	1.5698
	UL	1.8813	1.8831	1.8474	1.8592	1.8622	1.8788	1.8407
ϱ^*		1.0011	1.3264	1.7901	2.0688	2.2201	1.7529	1.2500
	SD	0.0234	0.0312	0.0407	0.0462	0.0502	0.0424	0.0319
	LL	0.9551	1.2652	1.7104	1.9782	2.1217	1.6699	1.1875
	UL	1.0470	1.3876	1.8698	2.1593	2.3185	1.8359	1.3126
σ^{2*}		1.5873	1.5600	1.5426	1.6164	1.4909	1.5759	1.4770
	SD	0.6894	0.6675	0.6687	0.7050	0.6388	0.6725	0.6415
	LL	0.2360	0.2516	0.2320	0.2345	0.2389	0.2578	0.2196
	UL	2.9385	2.8684	2.8531	2.9983	2.7429	2.8940	2.7345

*Estimates in the row are significant at the 5% level for all quantiles τ.