Title

Multilevel Time-Varying Joint Models for Longitudinal and Survival Outcomes

Author

Quintanilla Salinas, Isaac

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Multilevel Time-Varying Joint Models for Longitudinal and Survival Outcomes

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Isaac Noe Quintanilla Salinas

September 2022

Dissertation Committee:

Dr. Esra Kürüm, Chairperson
Dr. Weixin Yao
Dr. Analisa Flores

The Dissertation of Isaac Noe Quintanilla Salinas is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I want to thank Dr. Esra Kürüm for her patience and support in my graduate studies at UC Riverside. I appreciate all the hours of mentorship you have provided in both teaching and research. Also, I am grateful for all the hours you spent reading my dissertation and presentation. Without you, none of this would be possible.

I would like to thank my remaining committee members Drs. Weixin Yao, Chioun Lee, Analisa Flores, Daniel Jeske and Padmini Varadarajan, for your guidance through my dissertation process. I am especially grateful to Dr. Analisa Flores for all your support as a teaching supervisor and mentor, and thank you for supporting me with all my ideas.

I would like to thank my fellow graduate students, Dr. Lauren Cappiello, Dr. Debaleena Sain, Dr. Samantha VanSchalkwyk, Dr. Dylan Friel, Dr. Nathan Robertson, Bradley Lubich, Rebecca Kurtz-Garcia, Noé Vidales, Jeff Ellis, and Brian Tran for their friendship and support. I am very grateful to Debaleena for all her support during the research process.

I want to thank everyone who has supported me throughout the years, Alejandra Cabrera, Allan Coello, Dr. Arlene Haffa, Dr. Bryan Juarez, Cynthia Salazar, Douglas Morales, Dr. Esmeralda Iñiguez-Stevens, Dr. Georgia Sadler, Jessica Salazar, Dr. John Bellettiere, Juan Dominguez, Dr. Katherine Vuchkov, Luis Magallanes, Maritza Herrera, Dr. Maggie Santibáñez, Dr. Rado Vuchkov, and Tiffany Ta. I appreciate everything you all have done for me during my time in school.

Lastly I would like to thank my parents, Hugo and Lorena Quintanilla, as well as my siblings, Cindy, Frank, and Emely, for all their support when I was in graduate school. This has been a long journey, and I am glad to have all of you by my side.

This project was supported by a research grant from the National Institute of Diabetes and Digestive and Kidney Diseases (R01 DK092232).

To my parents for all their support.

ABSTRACT OF THE DISSERTATION

Multilevel Time-Varying Joint Models for Longitudinal and Survival Outcomes

by

Isaac Noe Quintanilla Salinas

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, September 2022
Dr. Esra Kürüm, Chairperson

Motivated by the United States Renal Data System (USRDS), we propose a joint modeling framework for longitudinal and survival outcomes that accounts for time-dynamic associations. In this population of patients, two outcomes are of interest, hospitalization, a longitudinal binary outcome, which is a major source of death risk, and mortality, which is higher in this population than in other comparable populations, including Medicare patients with cancer. Therefore, it is of interest to identify the patient-and dialysis facility-level risk factors that jointly affect these outcomes. Furthermore, studies have shown the effect of risk factors changes as a patient undergoes dialysis; therefore, it is necessary to model the associations as a function of time. Additionally, we incorporate multilevel random effects and multilevel covariates, at both the patient and facility levels, to account for the hierarchical data structure. An approximate Expectation-Maximization algorithm is developed for estimation and inference, where the fully exponential Laplace approximation is employed to address the hierarchical structure, and spline models are utilized to incorporate a time-dynamic association. We demonstrate the finite sample performance of our approach via simulation studies. We apply our proposed model to USRDS data to identify significant time-varying associations.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

End-stage renal disease (ESRD) is the final stage of chronic kidney disease where an individual's kidneys fail. This requires individuals to undergo either dialysis treatment or a kidney transplant. In the United States (U.S.), there were over 726,000 individuals with ESRD at the end of 2016 with 70% of them undergoing dialysis treatment (United States Renal Data System, 2018). Compared to other morbid populations, dialysis patients have a higher level of mortality risk. Additionally, dialysis patients are frequently hospitalized due to the nature of dialysis treatment and other comorbidities these patients have, such as diabetes or cardiovascular disease (United States Renal Data System, 2018). Thus, for this population of patients, frequent hospitalizations (collected longitudinally) and mortality are associated outcomes, and it is of interest to understand the relative contribution of risk factors to this association after the initiation of dialysis. Potential risk factors affecting this relationship are both patient- and dialysis facility-level, such as facility staffing (e.g., the ratio of nurse-to-patients). Our goal in this project is to jointly model hospitalization (binary longitudinal outcome) and time-to-death (survival outcome) in the dialysis population while taking into account the multilevel/hierarchical structure of the data (repeated measurements nested within patients and patients further nested within facilities). These models will be applied to the data from the United States Renal Data System (USRDS), a large national database. In terms of covariates, this database includes patient demographics and baseline comorbidities at the subject-level, and facility staffing, such as nurse-to-patient ratio, at the facility-level.

Joint modeling of longitudinal and survival outcomes have been extensively studied in the literature. Researchers have modeled the association using random effect models (De Gruttola and Tu, 1994), proportional hazard models (Tsiatis et al., 1995), shared-parameter models (Wulfsohn and Tsiatis, 1997), and conditional score functions (Tsiatis and Davidian, 2001). De Gruttola and Tu (1994) model the association between the longitudinal and survival outcomes using subject-level random effects, which are assumed to follow a multivariate normal distribution. The authors used the EM (Expectation-Maximization) algorithm to maximize the full joint likelihood function to estimate the parameters. Tsiatis et al. (1995) take a different approach where the longitudinal outcome is modeled with a random effects model, and the association between the longitudinal and survival outcomes are fitted with a proportional hazards model. The parameter in the hazard model are estimated in two steps, an empirical Bayes estimate is used for the longitudinal process, and in the second step, the longitudinal estimates are used as a time-dependent covariate to fit the proportional hazard model. Additionally, the authors argue this method produces better estimators than the maximization of the likelihood of the marginal proportional hazard model. Tsiatis and Davidian (2001) propose an alternative approach for estimating the parameters of the joint model. The authors relax the normality assumption for the random effects by conditioning the survival model with the complete sufficient statistic of the random effects. The parameters are then estimated from the conditional score estimating equations. Wulfsohn and Tsiatis (1997) used shared random effects to model both the longitudinal and survival outcome. This allows the outcome to be jointly modeled. The parameters are then estimated via the EM algorithm. The authors argue this method utilizes the information of both outcomes to jointly obtain more accurate estimates of the parameters. Liu et al. (2008) extend the shared-parameter model for hierarchical data. The authors use random effects to account for the variation of the different levels in the data. The parameters are estimated similarly to Wulfsohn and Tsiatis (1997). Additionally, the shared-parameter model has been extended to handle generalized outcomes (Larsen, 2004; Li et al., 2009; Rizopoulos et al., 2008). The methods described provide a framework to model dependent outcomes; however, the authors do not allow for a dynamic association.

In longitudinal studies, such as the USRDS data, the association between outcomes, longitudinal and time-to-event, or the relationship between a response and its predictors may change over time. The inability of traditional parametric regression models to capture this dynamic structure of the data led Nan et al. (2005), Song and Wang (2008), and Andrinopoulou et al. (2018) to implement varying coefficients in the joint models. Nan et al. (2005) implemented the varying coefficient as a time-dependent covariate in the proportional hazard model. The time-varying coefficient is approximated with a natural cubic B-spline basis. The parameters are then estimated with the maximum partial likelihood estimators. Song and Wang (2008) implemented a time-varying coefficient in the proportional hazards model. The authors approximated the time-varying coefficient with a local linear model. The parameters are then estimated using the corrected score or conditional score local estimating equations. Andrinopoulou et al. (2018) used time-varying coefficients in the shared-parameter model to describe the association of the dependent outcomes. The authors approximated the time-varying coefficients with Bayesian P-Splines (Lang and Brezger, 2004). *While all these methods proposed in the literature allow for a time-varying association between the longitudinal and survival outcomes, they do not implement time-varying coefficients between the risk factors and each outcome; in other words, they cannot accommodate dynamic response-predictor relationships. Additionally, the models do not allow for generalized outcomes and hierarchical data.*

In this dissertation, we develop a flexible multilevel joint modeling approach that accounts for the three-level hierarchy of the USRDS data, that is, longitudinal measurements, hospitalizations measured over time, nested within subjects and subjects further nested within dialysis facilities. In addition, our joint modeling approach accommodates all dynamic associations that may exist in a longitudinal study. *To our knowledge, although time-dynamic effects in joint modeling have been studied in the literature before, there has not been any work that combined these key elements: (1) incorporate time-varying effects of predictors (risk factors) on both the longitudinal and survival outcomes as well as the time-dynamic effect of the longitudinal process on the survival outcome and (2) accommodate complex multilevel data structures while including both subject- and facility-level risk* *factors. Our work will fill this gap in the literature while handling generalized longitudinal outcomes.*

The proposed joint model accounts for all manners of associations, that is, dependence among the repeated measurements within a subject and the correlation between the longitudinal and survival outcomes, by including multilevel random effects. Therefore, given the random effects, we assume that the outcomes are independent, leading to a submodel for each response. The longitudinal submodel is a generalized time-varying linear mixed effects model (Li et al., 2020) and a time-varying proportional hazard model (Cox, 1972) is employed for the survival outcome. The time-varying coefficients within each model allows us to explore the dynamic response-response and response-predictor associations. We demonstrate the estimation of these time-varying coefficients via the P-splines models (Eilers and Marx, 2010) and random-coefficient spline models (Ruppert et al., 2003).

Estimation in our modeling scheme is based on an EM algorithm (Dempster et al., 1977), where we treat the multilevel random effects as missing data. At the expectation step (E-step), the posterior mean and variance of the random effects are estimated, whereas the maximization step (M-step) involves maximizing the joint log-likelihood to obtain the estimated model parameters. One major challenge in the implementation of joint models to our three-level hierarchical data is due to the high-dimensional vector of random effects (of order n_i+1) at the facility level with the facility-level random effects as well as subject-level random effects for n_i patients receiving dialysis at the ith facility. This challenge is compounded especially when the size of the data is large. Our analysis of the USRDS data includes over 292,000 observations on $\sim$34,000 patients in more than 500 facilities, where the number of patients within a facility, denoted by n_i, ranges between 50 to 162. Therefore, we adopt the fully exponential approach proposed by Tierney et al. (1989) to address this computational challenge. It is shown that the fully exponential Laplace approximation is advantageous over the standard Laplace approximation as it leads to lower order approximation errors and reliable estimation results especially when modeling sparse longitudinal outcomes with few repeated measurements within a subject. Although, the fully exponential Laplace approximation have been employed previously (Rizopoulos et al., 2009; Kürüm et al., 2021) in a joint modeling context; these works modeled time-invariant relationships, whereas we demonstrate the use of this approximation in a time-varying joint modeling framework.

The remainder of this dissertation is organized as follows. Chapter 2 includes literature review on joint models and varying-coefficient models. In Chapter 3, we introduce our proposed time-varying joint modeling approach. We describe our estimation procedure based on P-spline and random-coefficient spline models, and discuss a bootstrap approach for inference. In Chapter 4, we present simulation studies designed to demonstrate the finite sample behavior of our estimators. Chapter 5 illustrates our proposed methodology using the USRDS data set. Finally, in Chapter 6, we give our conclusions and outline some future research topics.

Chapter 2

Literature Review

In this chapter, we briefly discuss the statistical concepts that are relevant for our proposed methodology. Section 2.1 provides details on longitudinal data analysis and mixed effects models. Section 2.2 reviews survival data analysis and the Cox proportional hazard model. In Section 2.3, we focus on joint longitudinal-survival models and their corresponding estimation procedures. Lastly, in section 2.4, we present a brief summary of time-varying coefficient models.

2.1 Longitudinal Data Analysis

Longitudinal studies involve repeated measurements of the same subjects over a period of time. These type of studies enable researchers to investigate how the effects of risk factors on an outcome change over time. One challenge in the analysis of longitudinal data arises due to the association among the repeated measurements. In particular, although the subjects are assumed to be independent of each other, due to the dependence among the repeated measurements within a subject, traditional regression models, where all the observations are assumed to be independent, cannot be employed. Therefore, methods that account for this dependence among the repeated measurements have been proposed, such as generalized least squares and mixed effects models.

Linear mixed effects models are used in longitudinal studies to model both the population- and individual-level effects. Let $\boldsymbol{Y}_i = (Y_{i1}, ..., Y_{in_i})^{\mathrm{T}}$, for $i = 1, ..., n$, denote the vector of responses for the ith subject measured at time points $\boldsymbol{t}_i = (t_{i1}, ..., t_{in_i})^{\mathrm{T}}$. The predictor variables at time point t_{ij} are denoted as $\boldsymbol{X}_{ij} = (X_{ij0}, X_{ij1}, ..., X_{ijp})^{\mathrm{T}}$ with $j = 1, \ldots, n_i$, and $\boldsymbol{X}_i$ is a $(p+1) \times n_i$ matrix containing all the predictor values for the ith subject. A linear mixed effects model is composed of both fixed effects and random effects:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)^{\mathrm{T}}$ represents the fixed effects, $\boldsymbol{D}_i = (\boldsymbol{D}_{i1}, ..., \boldsymbol{D}_{in_i})$ with each $\boldsymbol{D}_{ij}$ representing a vector predictors for the random effects, $\boldsymbol{b}_i = (b_{i1}, b_{i2}, ..., b_{iq})^{\mathrm{T}}$ are the subject-level random effects following a normal distribution with $E(\boldsymbol{b}_i) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{b}_i) = \boldsymbol{G}$, and $\boldsymbol{\epsilon}_i$ is the error term, following a normal distribution with $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\epsilon}_i) = \sigma^2\boldsymbol{I}_{n_i}$. Given model (2.1) and the distributional assumptions, $\boldsymbol{Y}_i$ follows a normal distribution such that $\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}, \boldsymbol{V}_i)$, where

$$\boldsymbol{V}_i = \mathrm{Cov}(\boldsymbol{Y}_i) = \boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{G}\boldsymbol{D}_i + \sigma^2\boldsymbol{I}_{n_i}. \tag{2.2}$$

The covariance matrix in (2.2) is parameterized based on the random effects and the time points for the ith subject.

The parameters $\boldsymbol{\beta}$ can be estimated using the weighted least squares (WLS) estimator

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{V}_i^{-1} \boldsymbol{X}_i^{\mathrm{T}} \right)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{V}_i^{-1} \boldsymbol{Y}_i. \tag{2.3}$$

Note that the estimator (2.3) is a function of the covariance matrix $\boldsymbol{V}_i$. If $\boldsymbol{V}_i$ is unknown, ideally, it should be estimated with either a maximum likelihood or restricted maximum likelihood approach. However, it is known that the maximum likelihood approach produces biased estimates for σ^2 when the sample size is small; therefore, it is recommended to use the restricted maximum likelihood function to estimate $\boldsymbol{V}_i$. Let $\boldsymbol{\alpha}$ denote the full parameter vector for the covariance $\boldsymbol{G}$, then the restricted log-likelihood function is written as

$$\begin{aligned} \ell(\boldsymbol{\alpha}, \sigma^2) &= -\sum_{i=1}^{n} \log |\boldsymbol{V}_i| - \frac{1}{2} \left\{ \sum_{i=1}^{n} \left(\boldsymbol{Y}_i - \boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}} \right)^{\mathrm{T}} \boldsymbol{V}_i^{-1} \left(\boldsymbol{Y}_i - \boldsymbol{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}} \right) \right\} \\ &\quad -\frac{1}{2} \log \left| \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{V}_i^{-1} \boldsymbol{X}_i^{\mathrm{T}} \right)^{-1} \right|. \end{aligned} \tag{2.4}$$

An optimization algorithm is used to estimate the values of $\boldsymbol{\alpha}$ and σ^2 that maximize the restricted log-likelihood function (2.4). The estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\sigma}^2$ are then used to obtain $\hat{\boldsymbol{V}}_i$, which can be substituted in (2.3) to estimate $\boldsymbol{\beta}$. Diggle (2002) and Fitzmaurice (2004) provides more details on linear mixed effects models.

In terms of inference, it is shown that the asymptotic distribution for $\hat{\boldsymbol{\beta}}$ is a normal distribution with mean $\boldsymbol{\beta}$ and $Cov(\hat{\boldsymbol{\beta}})$, where the covariance for $\hat{\boldsymbol{\beta}}$ can be estimated as $\widehat{Cov}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{n} \boldsymbol{X}_i \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{X}_i^{\mathrm{T}} \right)^{-1}$. Note that Verbeke and Lesaffre (1997) showed that when the random effects are normally distributed, the maximum likelihood estimates are consistent and asymptotically normal with the inverse Fisher's information matrix as the asymptotic covariance matrix. However, when the random effects are not normally distributed, a sandwich type correction for the Fisher's information matrix is required to obtain an appropriate asymptotic covariance matrix.

Linear mixed effects models can be extended to accommodate response variables that follow any exponential family distribution, known as generalized linear mixed effects models (GLMM). These models involve the following assumptions:

- the conditional distribution of $Y_{ij}|\boldsymbol{b}_i$ follows an exponential family distribution for $i = 1, ..., n$ and $j = 1, ..., n_i$,
- given the random effects ($\boldsymbol{b}_i$), the repeated measurements in $\boldsymbol{Y}_i$ are independent,
- $\boldsymbol{b}_i$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{G}$.

The GLMM framework shows that the expectation of Y_{ij}, given the random effects, is linearly associated with the fixed effects and random effects via a link function $g(\cdot)$:

$$g\{E(Y_{ij}|\boldsymbol{b}_i)\} = \boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{D}_{ij}^{\mathrm{T}}\boldsymbol{b}_i,$$

where $\boldsymbol{X}_{ij}$ and $\boldsymbol{D}_{ij}$ are predictor variables, and $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)^{\mathrm{T}}$ are the regression coefficients for the fixed effects.

The parameters $\boldsymbol{\beta}$ and $\boldsymbol{G}$ can be estimated via a maximum likelihood approach. The random effects are treated as latent variables and integrating out to construct the log-likelihood function leads to

$$\ell(\boldsymbol{\beta}, \boldsymbol{G}; \boldsymbol{Y}) = \log L(\boldsymbol{\beta}, \boldsymbol{G}; \boldsymbol{Y}) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \log \int p(Y_{ij} | \boldsymbol{b}_i; \boldsymbol{\beta}) p_{\boldsymbol{b}_i}(\boldsymbol{b}_i; \boldsymbol{\delta}) d\boldsymbol{b}_i, \quad (2.5)$$

where $L(\boldsymbol{\beta}, \boldsymbol{G}; \boldsymbol{Y})$ is the likelihood function, $p(Y_{ij} | \boldsymbol{b}_i; \boldsymbol{\beta})$ is the conditional density function for Y_{ij}, $p_{\boldsymbol{b}_i}(\boldsymbol{b}_i; \boldsymbol{\delta})$ is the density function of $\boldsymbol{b}_i$, and $\boldsymbol{\delta}$ is a vector of parameters involved in $\boldsymbol{G}$.

The maximum likelihood estimates for $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ are obtained by setting the following observed score functions to zero:

$$\boldsymbol{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\delta} | \boldsymbol{Y}) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \boldsymbol{X}_{ij} \left[Y_{ij} - E\{\mu_{ij}(\boldsymbol{b}_i) | \boldsymbol{Y}_{ij}\} \right] = 0 \quad (2.6)$$

$$\text{and } \boldsymbol{S}_{\boldsymbol{\delta}}(\boldsymbol{\beta}, \boldsymbol{\delta} | \boldsymbol{Y}) = 0.5 \boldsymbol{G}^{-1} \left\{ \sum_{i=1}^{n} E(\boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{T}} | \boldsymbol{Y}_i) \right\} \boldsymbol{G}^{-1} - \frac{n}{2} \boldsymbol{G}^{-1} = 0, \quad (2.7)$$

where $\mu_{ij}(\boldsymbol{b}_i) = g^{-1}(\boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \boldsymbol{D}_{ij}^{\mathrm{T}} \boldsymbol{b}_i)$. The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ can be estimated using the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) on equations (2.6) and (2.7). First, the expectations in the score functions are evaluated using the current estimates of the parameters to target the random effects (E-step); afterwards, the score functions are maximized with respect to the parameters to update the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ (M-step). The EM algorithm iterates between the E- and M-step until a convergence criteria is met. For inferential purposes, the covariance matrices of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$ can be estimated using the observed Fisher information, that is, by plugging the estimated values of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ to the inverse of the negative Hessian matrix of the log-likelihood given in (2.5).

2.2 Survival Analysis

Survival data analysis methods are widely used in engineering and medical fields to identify factors associated with time to an event of interest such as death, occurrence of a disease, or failure of a machine. Traditional regression models are not suitable for this type of data mainly due to censoring. The defining feature of censoring is that the time to an event is not observable for all subjects; in other words, some subjects never developed a disease or experienced death during the study, so their true time-to-event is missing in the data set. Traditional statistical models assume that we have complete information on all subjects, and therefore, when applied to this type of data they would produce biased estimates of the distribution of the event times and possibly incorrect inference.

There are two types of censoring classifications in survival analysis: left/right/interval censoring and informative/noninformative censoring. The first classification of censoring is based on the positioning of the true event time. Right censoring occurs when the event is observed after a pre-specified censoring time, such as the end of the study. Examples of right censoring are lost to follow-up or a subject reaching the end of a study without observing the event. Left censoring occurs when the event of interest happens before a pre-specified censoring time. This is common in adolescent studies where a participant is enrolled after experiencing the event. Interval censoring results when the event occurs between an interval of time points. This is common in longitudinal studies where participants experience the event between two measurements.

The second classification of censoring is related to whether the probability of censoring an individual is independent to the probability of observing the desired event, known as informative and noninformative censoring. Informative censoring occurs when a participant withdraws from the study due to expecting an upcoming time-to-event. For example, a participant may leave a study when their health has deteriorated beyond a certain point due to the disease of interest. Noninformative censoring occurs when a participant leaves the study for non study-related reasons. When informative censoring occurs, there is not enough information to model the censoring mechanism. Therefore, noninformative censoring is essential for survival analysis. For the rest of this section, we only consider situations where noninformative right censoring is present.

In survival analysis, our main interest is to study the distribution of the true event times T^* using the following basic functions: survival function, hazard function, and cumulative hazard function. The survival function is the probability that T^* is larger than a certain time point t:

$$\mathcal{S}(t) = \Pr(T^* > t) = \int_t^\infty p(x)dx,$$

where $p(x)$ denotes the probability density function of T^*. If the event of interest is death, the survival function gives the probability that death occurs after time t. The survival function must be nonincreasing as t increases and $\mathcal{S}(t = 0) = 1$. The hazard (risk) function depicts the conditional probability that the event occurs in the next moment (instant) given that it has not occurred up to time t:

$$h(t) = \lim_{\Delta \to 0} \frac{\Pr(t \leq T^* < t + \Delta | T^* \geq t)}{\Delta}.$$

Furthermore, the hazard function can be expressed in terms of the survival and probability density functions: $\mathscr{h}(t) = \frac{\mathscr{p}(t)}{\mathscr{S}(t)}$. The cumulative hazard function, which is used to describe the accumulated risk up to time t, can be obtained either via the hazard function or the survival function: $\mathscr{H}(t) = \int_0^t \mathscr{h}(x)dx$ and $\mathscr{H}(t) = -\log\{\mathscr{S}(t)\}$, respectively.

As we mentioned above, the main challenge in survival analysis is censoring, that is, we might not observe the true event time for all subjects. Therefore, we introduce the "observed event time" variable T_i, which is defined as the minimum of the potential censoring time C_i and the true event time T_i^* with $i = 1, \ldots, n$. In addition, the event indicator is given as $\delta_i = I(T_i^* < C_i)$ with $I(\cdot)$ as the indicator function. Under this framework, our main objective in survival analysis becomes exploring the distribution of the true event times T_i^* using the available information (T_i, δ_i). Several methods have been developed for this purpose, in particular, to estimate survival and (cumulative) hazard functions.

We discuss the nonparametric method introduced by Kaplan and Meier (1958), known as the Product-Limit estimator, to estimate the survival function. Let $\{t_j, d_j, R_j\}_{j=1}^D$ denote the survival data, where $t_1 < t_2 < \cdots < t_D$ are the ordered distinct observed event times, d_j represents the number of events at time point t_j, and R_j denotes the number of subjects still at risk of experiencing the event at t_j. The Product-Limit estimator is defined as

$$\hat{\mathscr{S}}(t) = \begin{cases} 1, & \text{if } \ t < t_1 \\ \prod_{t_j \leq t}(1 - \frac{d_j}{R_j}), & \text{if } \ t_1 \leq t. \end{cases}$$

Greenwood (1926) provides the estimated variance of $\hat{\mathcal{S}}(t)$ as

$$\widehat{Var}\{\hat{\mathcal{S}}(t)\} = \hat{\mathcal{S}}^2(t) \sum_{t_j \leq t} \frac{d_j}{R_j(R_j - d_j)}.$$

Using the Product-Limit estimator, the cumulative hazard function can be estimated as $\hat{\mathcal{H}}(t) = -\log\{\hat{\mathcal{S}}(t)\}$.

2.2.1 Cox Proportional Hazard Models

Although the Product-Limit estimator provides an unbiased estimate of the survival function, it is limited to exploring the effects of categorical risk factors on the outcome. To overcome this drawback, parametric models, which can be used to model the association between a set of predictors (numeric and categorical) and the survival time, were proposed. However, as the name implies, these models assume a parametric distribution for the true time-to-event T_i^*, which may lead to biased results and incorrect inference, if incorrectly specified. An alternative method, proposed by Cox (1972), is the Cox proportional hazard model, where the estimation is performed via a partial likelihood approach. The partial likelihood procedure allows one to estimate the regression coefficients without making distributional assumptions for the true event times.

Cox (1972) proposed the following model, where it is assumed that the covariates have a multiplicative effect on the hazard of an event,

$$\hbar(t|\boldsymbol{X}, \boldsymbol{\beta}) = \hbar_0(t) \exp\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}\right) \tag{2.8}$$

with $\mathscr{h}_0(t)$ as the baseline hazard function, $\boldsymbol{X} = (X_1, X_2, ..., X_p)^{\mathrm{T}}$ as the vector of predictor variables, and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^{\mathrm{T}}$ denoting the corresponding regression coefficients. Using the hazard function (2.8), the survival function is constructed as

$$\mathcal{S}(t|\boldsymbol{X}, \boldsymbol{\beta}) = \exp\left\{-\int_0^t \mathscr{h}_0(u) \exp(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}) du\right\},$$

where $\boldsymbol{\beta}$ can be estimated using a maximum likelihood approach using the likelihood construction discussed by Kalbfleisch (2002), where noninformative right censoring is assumed. Given the survival data framework $\{T_i^*, \boldsymbol{X}_i, \delta_i\}_{i=1}^n$, the time-to-event T_i^* has a probability density function $p_T(t; \boldsymbol{\beta})$ and survival function $\mathcal{S}_T(t; \boldsymbol{\beta})$, and similarly, the censoring time C_i has a probability density function $p_C(c; \boldsymbol{\theta}_C)$ and survival function $\mathcal{S}_C(c; \boldsymbol{\theta}_C)$, with $\boldsymbol{\theta}_C$ is a vector of parameters for the censoring mechanism. Under noninformative censoring, the time-to-event distribution and the censoring distribution do not share any parameters and are independent; therefore,

$$\begin{aligned}
\Pr\{T_i^* \in (t, t+\Delta), \delta_i = 1; \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\theta}_C\} &= \Pr\{T_i^* \in (t, t+\Delta), C_i > t; \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\theta}_C\} = p_T(t; \boldsymbol{\beta})\Delta\mathcal{S}_C(t; \boldsymbol{\theta}_C) \\
\Pr\{T_i^* \in (t, t+\Delta), \delta_i = 0; \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\theta}_C\} &= \Pr\{T_i^* > t, C_i \in (t, t+\Delta); \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\theta}_C\} = p_C(t; \boldsymbol{\theta}_C)\Delta\mathcal{S}_T(t; \boldsymbol{\beta}).
\end{aligned}$$

The likelihood function is then constructed as

$$\mathscr{L}(\boldsymbol{\beta}, \boldsymbol{\theta}_C) = \prod_{i=1}^n \{p_T(T_i^*; \boldsymbol{\beta})\Delta\mathcal{S}_C(T_i^*; \boldsymbol{\theta}_C)\}^{\delta_i}\{p_C(T_i^*; \boldsymbol{\theta}_C)\Delta\mathcal{S}_T(T_i^*; \boldsymbol{\beta})\}^{1-\delta_i}.$$

Furthermore, we can define the following log-likelihood as the parameters of the censoring distribution are not the main interest,

$$\mathscr{L}(\boldsymbol{\beta}) \propto \prod_{i=1}^{n} \{p_T(T_i^*; \boldsymbol{\beta})\}^{\delta_i} \{\mathcal{S}_T(T_i^*; \boldsymbol{\beta})\}^{1-\delta_i}. \tag{2.9}$$

Replacing $p_T(T_i; \boldsymbol{\beta})$ with $\hbar(T_i^* | \boldsymbol{X}_i, \boldsymbol{\beta}) \mathcal{S}(T_i^* | \boldsymbol{X}_i, \boldsymbol{\beta})$ in (2.9) yields

$$\mathscr{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\{\hbar(T_i^* | \boldsymbol{X}_i, \boldsymbol{\beta})\}^{\delta_i} \{\mathcal{S}(T_i^* | \boldsymbol{X}_i, \boldsymbol{\beta})\} \right], \tag{2.10}$$

where we remove the subscript T in the survival function to align with our previous notation. Given the model form in (2.8), the likelihood function (2.10) requires estimating both the regression coefficients and baseline hazard function. The initial step in this estimation procedure would be to choose a form for the baseline hazard function, which can be specified parametrically or nonparametrically. The former is the aforementioned parametric models approach, which might produce incorrect results, and the latter results in a semiparametric estimation procedure, which can be computationally burdensome. In order to avoid these drawbacks, Cox (1972) proposed following a partial likelihood approach, which avoids estimating the baseline hazard function.

In the partial likelihood approach, $D = \sum_{i=1}^{n} \delta_i$ denotes the number of participants experiencing the event, $\{t_j\}_{j=1}^{D}$ represents the ordered and distinct time-to-event for participants observing the event, $\boldsymbol{X}_{(j)}$ are the predictors for the participant at time point t_j, and $R(t_j)$ is the set of individuals at risk prior to t_j. The partial log-likelihood function is given as

$$\ell(\boldsymbol{\beta}) = \log \mathcal{L}(\boldsymbol{\beta}) = \log \left[\prod_{j=1}^{D} \frac{\exp\{\boldsymbol{X}_{(j)}^{\mathrm{T}} \boldsymbol{\beta}\}}{\sum_{k \in R(t_j)} \exp\{\boldsymbol{X}_k^{\mathrm{T}} \boldsymbol{\beta}\}} \right]. \tag{2.11}$$

In order to find the estimates that maximize the partial log-likelihood, first, the partial derivatives are taken with respect to $\boldsymbol{\beta}$, denoted as $U(\boldsymbol{\beta}) = \frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\beta}}$. Second, the partial maximum likelihood estimates that satisfy $U(\hat{\boldsymbol{\beta}}) = 0$ are obtained using a numerical approach, such as the Newton-Raphson algorithm. For inferential purposes, the covariance matrix of $\hat{\boldsymbol{\beta}}$ is estimated by finding the inverse of the negative Hessian matrix of the partial log-likelihood function (2.11) and evaluating it at $\hat{\boldsymbol{\beta}}$: $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}) = -\left\{ \frac{\partial^2 log(\mathcal{L})}{\partial \boldsymbol{\beta}^{\mathrm{T}} \partial \boldsymbol{\beta}} \right\}^{-1} |_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}}$.

The partial log-likelihood function (2.11) assumes that the time-to-events are distinct, that is, having only one subject experience the event at the specific time point. However, when this assumption is not met, the partial likelihood function does not include information from all participants that experienced the event at the distinct time point, in particular, information from only one subject is included in the calculation of (2.11). Breslow (1974) and Efron (1977) proposed alternative partial likelihood functions for non-distinct survival times, where the numerator and denominator include information from all participants experiencing the event at the distinct time point and the partial likelihood function also incorporates the number of subjects that experience the event at that time point.

As we discussed above, the partial likelihood approach does not make any distributional assumptions while estimating the regression coefficients; however, researchers might be interested in constructing the survival function and this would require estimating the baseline hazard function first. The profile maximum likelihood estimator, proposed by

Breslow (1974), can be used for this purpose. Given the maximum likelihood estimates of $\boldsymbol{\beta}$ from the partial log-likelihood (2.11), the baseline hazard function at time t_j is computed as

$$\hat{h}_0(t_j; \hat{\boldsymbol{\beta}}) = \frac{1}{(t_j - t_{j-1}) \sum_{k \in R(t_j)} \exp\{\boldsymbol{X}_k^{\mathrm{T}} \hat{\boldsymbol{\beta}}\}},$$

where $R(t_j)$ is the set of individuals at risk prior to time t_j. The estimate for the baseline hazard function can be used to estimate the cummulative baseline hazard and survival functions $\hat{\mathcal{H}}_0(t; \hat{\boldsymbol{\beta}}) = \sum_{t_j \geq t} \hat{h}_0(t_j; \hat{\boldsymbol{\beta}})$ and $\hat{\mathcal{S}}(t; \hat{\boldsymbol{\beta}}) = \exp\{-\hat{\mathcal{H}}_0(t; \hat{\boldsymbol{\beta}}) \exp(\boldsymbol{X}^{\mathrm{T}} \hat{\boldsymbol{\beta}})\}$, respectively.

The Cox model defined in (2.8) assumes that all the covariates are time-invariant, that is, collected at the baseline; however, studies may collect data at different time points until a patient experiences an event of interest. In this type of data, it would be of interest to incorporate the repeated measurements, known as time-dependent covariates, to the survival model and explore their effects on the time-to-event. Let $W_i(t)$, for $i = 1, \ldots, n$, denote the time-dependent covariate, which can be classified as either exogenous, the value of $W_i(t)$ is known and uninfluenced by the occurrence of the time-to-event, or endogenous, the value of $W_i(t)$ is unknown and might be influenced by the occurrence of the time-to-event. Let $\mathcal{W}_i(t) = \{W_i(s) : 0 \leq s < t\}$ denote the covariate's history, an exogenous time-dependent covariate satisfies the following condition: $\Pr\{\mathcal{W}_i(t) | \mathcal{W}(s), T_i^* \geq s\} = \Pr\{\mathcal{W}_i(t) | \mathcal{W}(s), T_i^* = s\}$ for $s \leq t$, indicating that $W_i(\cdot)$ is not associated with the occurrence of the time-to-event. On the contrary, an endogenous covariate will not satisfy the condition; therefore, modeling the survival function with an endogenous time-dependent covariate requires careful construction of the likelihood function (further discussed in Section 2.3).

To account for exogenous time-dependent covariates, the Cox model in (2.8) is extended, using a counting process, to incorporate the time-dependent covariate (Andersen and Gill, 1982). Let $\{N_i(t), R_i(t)\}$ be the ith participant's event process, where $N_i(t)$ denotes the number of events for the ith participant at time point t, and $R_i(t)$ represents if the ith participant is at risk at time t ($R_i(t) = 1$) or not ($R_i(t) = 0$). The Cox model (2.8) can be rewritten as

$$h_i(t|\mathcal{W}_i(t), \boldsymbol{X}_i) = h_0(t) \exp\left\{\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \alpha W_i(t)\right\},$$

where α is the regression coefficient for the time-dependent covariate $W_i(t)$. The corresponding partial log-likelihood function with the counting process integral is derived as

$$\mathcal{L}(\boldsymbol{\beta}, \alpha) = \sum_{i=1}^{n} \int_0^{\infty} \left(R_i(t) \exp\{\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \alpha W_i(t)\} - \log\left[\sum_j R_j(t) \exp\{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta} + \alpha W_j(t)\}\right]\right) dN_i(t). \tag{2.12}$$

The regression coefficients $\boldsymbol{\beta}$ and α are estimated by maximizing the partial log-likelihood function (2.12) with respect to the parameters using a numerical approximation algorithm. Inference is similar to the Cox model with baseline covariates, that is, performed via evaluating the inverse Hessian matrix at the estimated values.

2.3 Joint Modeling of Longitudinal and Survival Data

Increasingly, studies involve collecting data on multiple outcomes, usually of different types. That is, within a single study, some outcomes are measured at several time points (longitudinal), whereas other endpoints are measured at a single time point, such as the time until an event of interest occurs (survival). Sections 2.1 and 2.2 described statistical techniques to answer research questions where separate analysis of longitudinal and survival data can be employed. However, in many situations, exploring the association between these outcomes might be of interest, which requires modeling these outcomes jointly. In this section, we discuss joint modeling techniques for longitudinal and survival outcomes. Note that, a naïve approach to study the relationship between these outcomes would be including the longitudinal outcome as a time-varying predictor in a Cox model, similar to the approach described in Section 2.2.1. However, the time-dependent Cox model would only produce accurate estimates and inference when we have exogenous time-varying covariates. As shown by Sweeting and Thompson (2011) and Tsiatis and Davidian (2004), when we have an endogenous time-varying covariate, these models would lead to biased estimates and incorrect inference. In this section, we focus on joint analysis of endogenous time-varying covariates, that is, our longitudinal outcome and a time-to-event outcome.

Several methods have been developed to jointly model longitudinal and survival outcomes, such as the shared-parameter models, where the two outcomes are jointly modeled via a common set of random effects (Wulfsohn and Tsiatis, 1997; Song et al., 2002; Henderson et al., 2000; Hsieh et al., 2006; Rizopoulos et al., 2008); the random effects models, which are similar to the shared-parameter models, but with different set of random effects underlying each outcome (De Gruttola and Tu, 1994); proportional hazard models with a two-stage estimation procedure, where the longitudinal outcome is modeled in the first stage and a time-dependent Cox model with the predicted longitudinal outcomes as a covariate is fit in the second stage (Tsiatis et al., 1995); and conditional score functions, where the survival models are conditioned on the complete sufficient statistics of the random effects (Tsiatis and Davidian, 2001). Among these methods, shared-parameter models are the most frequently utilized method due to their performance in estimation and inference, and their computational feasibility. In this section, we focus on model formulation and estimation under this framework.

2.3.1 Shared-parameter Models

Under this approach, to obtain the joint distribution of the longitudinal and survival outcomes, we start by defining the longitudinal and survival submodels. Consider the following data for the ith subject, $\{T_i, \delta_i, \boldsymbol{Y}_i, \boldsymbol{t}_i, \boldsymbol{X}_i\}$, where T_i is the observed time-to-event, δ_i is the event indicator, $\boldsymbol{Y}_i = (Y_{i1}, ..., Y_{in_i})^{\mathrm{T}}$ denotes the repeated measurements with $Y_{ij} = Y_i(t_{ij})$ as the longitudinal outcome at time point t_{ij}, $\boldsymbol{t}_i = (t_{i1}, t_{i2}, ..., t_{in_i})^{\mathrm{T}}$ are the measurement times, and $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip-2})^{\mathrm{T}}$ is a time-invariant vector of predictors with $i = 1, ..., n$. The longitudinal submodel is formulated as a linear mixed effects model

$$\begin{aligned} Y_{ij} &= m_i(t_{ij}) + \epsilon_i(t_{ij}), \\ m_i(t_{ij}) &= \boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{\mathrm{T}}\boldsymbol{b}_i, \\ \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{G}) &\quad \text{and} \quad \epsilon_i(t_{ij}) \sim N(0, \sigma^2), \end{aligned} \tag{2.13}$$

where $\boldsymbol{X}_{ij} = (1, t_{ij}, \boldsymbol{X}_i^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$ are the fixed regression coefficients, and $\boldsymbol{Z}_{ij}$ is a subset of $\boldsymbol{X}_{ij}$ representing the design matrix for the random effects $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{iq})^{\mathrm{T}}$. Additionally, both the random effects and the error term are assumed to be mutually independent.

The survival submodel is constructed using the Cox proportional hazard model,

$$\begin{aligned} h_i\{t|\mathcal{M}_i(t), \boldsymbol{X}_i\} &= \lim_{\Delta \to 0} \frac{\Pr\{t \le T_i < t + \Delta | T_i \ge t, \mathcal{M}_i(t), \boldsymbol{X}_i\}}{\Delta} \\ &= h_0(t) \exp\{\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\gamma} + \alpha m_i(t)\}, \end{aligned} \tag{2.14}$$

where $h_0(t)$ denotes the baseline hazard function, $\boldsymbol{\gamma}$ and α are the regression coefficients, and $\mathcal{M}_i(t) = \{m_i(s) : 0 \le s < t\}$ is the history of the longitudinal outcome until time t. Using equation (2.14), the survival function is expressed as

$$S_i\{t|\boldsymbol{X}_i, \mathcal{M}_i(t)\} = \exp\left[-\int_0^t h_0(s) \exp\{\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\gamma} + \alpha m_i(s)\} ds\right].$$

The joint distribution of the longitudinal and survival outcomes are obtained under the full conditional independence assumption, that is, given the random effects, the outcomes are assumed to be independent. More specifically, the random effects account for the association between the longitudinal and survival outcomes, and they also explain the correlation between the repeated measurements of the longitudinal outcome within a subject. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_T^{\mathrm{T}}, \boldsymbol{\theta}_Y^{\mathrm{T}}, \boldsymbol{\theta}_b^{\mathrm{T}})^{\mathrm{T}}$ denote the full parameter vector where $\boldsymbol{\theta}_T = (\boldsymbol{\gamma}^{\mathrm{T}}, \alpha, \boldsymbol{\theta}_h^T)^{\mathrm{T}}$ is the vector of parameters for the survival submodel with $\boldsymbol{\theta}_h$ as the vector of parameters involved in the baseline hazard function, $\boldsymbol{\theta}_Y = (\boldsymbol{\beta}^{\mathrm{T}}, \sigma^2)^{\mathrm{T}}$ is the vector of parameters in the longitudinal submodel, and $\boldsymbol{\theta}_b$ is the vector of parameters involved in the covariance matrix $\boldsymbol{G}$. The joint density function for the ith subject is formulated as

$$p(T_i, \delta_i, \boldsymbol{Y}_i, \boldsymbol{b}_i; \boldsymbol{\theta}) = p(T_i, \delta_i | \boldsymbol{b}_i; \boldsymbol{\theta}) p(\boldsymbol{Y}_i | \boldsymbol{b}_i; \boldsymbol{\theta}) p(\boldsymbol{b}_i; \boldsymbol{\theta}),$$

where the density for the survival outcome is given as

$$p\{T_i, \delta_i | \boldsymbol{b}_i; \boldsymbol{\theta}, m_i(t)\} = \left[h_0(T_i)\exp\{\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\gamma} + \alpha m_i(T_i)\}\right]^{\delta_i} \exp\left[-\int_0^{T_i} h_0(s)\exp\{\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\gamma} + \alpha m_i(s)\}ds\right],$$

and the joint density for the longitudinal outcome and the random effects is expressed as

$$\begin{aligned} p(\boldsymbol{Y}_i|\boldsymbol{b}_i;\boldsymbol{\theta})p(\boldsymbol{b}_i;\boldsymbol{\theta}) &= \prod_{j=1}^{n_i} p(Y_{ij}|\boldsymbol{b}_i;\boldsymbol{\theta})p(\boldsymbol{b}_i;\boldsymbol{\theta}) \\ &= (2\pi\sigma^2)^{-n_i/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{n_i}\left\{Y_{ij} - \boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{Z}_{ij}^{\mathrm{T}}\boldsymbol{b}_i\right\}^2\right] \\ &\quad \times(2\pi)^{-q/2}|\boldsymbol{G}|^{-1/2}\exp(-\boldsymbol{b}_i\boldsymbol{G}^{-1}\boldsymbol{b}_i/2) \end{aligned}$$

with $\boldsymbol{G}$ as the covariance matrix of the random effects $\boldsymbol{b}_i$.

The parameters are estimated by maximizing the observed (incomplete) log-likelihood function $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\log\int p(T_i, \delta_i, \boldsymbol{Y}_i, \boldsymbol{b}_i; \boldsymbol{\theta})d\boldsymbol{b}_i$ with respect to $\boldsymbol{\theta}$ using optimization methods such as the Newton-Raphson algorithm or the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). Due to the unknown random effects, the EM algorithm is favored over the Newton-Raphson algorithm due to its ability to handle 'missing' data, and the variances have closed-form solutions.

In the EM algorithm, the expectation step (E-step) is used to target the random effects $\{\boldsymbol{b}_i\}_{i=1}^{n}$, and the maximization step (M-step) will update the estimates of $\boldsymbol{\theta}$. The algorithm will repeat the steps until a convergence criteria is met. In the E-step, we compute the expected value of the random effects by taking the expectation of the complete log-likelihood function:

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= \sum_{i=1}^{n} \int \log\{p(T_i, \delta_i, \boldsymbol{Y}_i, \boldsymbol{b}_i; \boldsymbol{\theta})\} p(\boldsymbol{b}_i | T_i, \delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(k)}) d\boldsymbol{b}_i \\
&= \sum_{i=1}^{n} \int \Big[\{\log p(T_i, \delta_i, |\boldsymbol{b}_i; \boldsymbol{\theta}) + \log p(\boldsymbol{Y}_i | \boldsymbol{b}_i; \boldsymbol{\theta}_y) \\
&\quad + \log p(\boldsymbol{b}_i; \boldsymbol{\theta}_b)\} p(\boldsymbol{b}_i | T_i, \delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(k)})\Big] d\boldsymbol{b}_i, \qquad (2.15)
\end{aligned}
$$

where $p(\boldsymbol{b}_i|T_i, \delta_i, \boldsymbol{Y}_i; \theta^{(k)})$ is the conditional density function of $\boldsymbol{b}_i$ and $\boldsymbol{\theta}^{(k)}$ is the current value of $\boldsymbol{\theta}$ at the kth iteration of the EM algorithm. As the integral does not have a closed form, it should be evaluated with numerical integration techniques such as the Gaussian quadrature (Press et al., 1992) or the Laplace approximation methods. (Tierney et al., 1989). Gaussian quadrature techniques are used when the dimensionality of the random effects and the number of repeated measurements are small (Wulfsohn and Tsiatis, 1997). On the contrary, when these quantities are large, Laplace approximation method is recommended (Rizopoulos et al., 2009).

The M-step identifies the estimates that maximize (2.15). The variances for the random effects and error terms can be obtained with closed-form solutions:

$$
\begin{aligned}
\hat{\sigma}^2 &= \left(\sum_{i=1}^{n} n_i\right)^{-1} \sum_{i=1}^{n} \int (\boldsymbol{Y}_i - \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{b}_i)^{\mathrm{T}} (\boldsymbol{Y}_i - \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{b}_i) p(\boldsymbol{b}_i|T_i, \delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(k)}) d\boldsymbol{b}_i \\
&= \left(\sum_{i=1}^{n} n_i\right)^{-1} \sum_{i=1}^{n} \Big[(\boldsymbol{Y}_i - \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta})^{\mathrm{T}} (\boldsymbol{Y}_i - \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta} - 2\boldsymbol{D}_i^{\mathrm{T}}\tilde{\boldsymbol{b}}_i) + \mathrm{tr}\{\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{D}_i \mathrm{Var}(\tilde{\boldsymbol{b}}_i)\} + \tilde{\boldsymbol{b}}_i^{\mathrm{T}}\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{D}_i\tilde{\boldsymbol{b}}_i\Big] \\
\hat{\boldsymbol{G}} &= n^{-1} \sum_{i=1}^{n} \mathrm{Var}(\tilde{\boldsymbol{b}}_i) + \tilde{\boldsymbol{b}}_i^{\mathrm{T}}\tilde{\boldsymbol{b}}_i,
\end{aligned}
$$

where $\tilde{\boldsymbol{b}}_i = \int \boldsymbol{b}_i p(\boldsymbol{b}_i|T_i, \delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(k)}) d\boldsymbol{b}_i$ and $\mathrm{Var}(\tilde{\boldsymbol{b}}_i) = \int (\boldsymbol{b}_i - \tilde{\boldsymbol{b}}_i) p(\boldsymbol{b}_i|T_i, \delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(k)}) d\boldsymbol{b}_i$ are the posterior mean and variance of $\boldsymbol{b}_i$, respectively, which are computed in the E-step. The remaining parameters can be estimated using a Newton-Raphson algorithm:

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \hat{\boldsymbol{\beta}}^{(k)} - \left\{ \partial \mathcal{A}\left(\hat{\boldsymbol{\beta}}^{(k)}\right) / \partial \boldsymbol{\beta} \right\}^{-1} \mathcal{A}\left(\hat{\boldsymbol{\beta}}^{(k)}\right),$$

$$\hat{\boldsymbol{\theta}}_t^{(k+1)} = \hat{\boldsymbol{\theta}}_t^{(k)} - \left\{ \partial \mathcal{A}\left(\hat{\boldsymbol{\theta}}_t^{(k)}\right) / \partial \boldsymbol{\theta}_t \right\}^{-1} \mathcal{A}\left(\hat{\boldsymbol{\theta}}_t^{(k)}\right),$$

where $\mathcal{A}(\cdot)$ is the partial derivative of the log-likelihood function (2.15) with respect to the parameter of interest.

The hazard function in equation (2.14) requires defining a form for the baseline hazard function even though it is possible to leave the baseline hazard function completely unspecified in traditional survival models (see Section 2.2.1). However, in a joint modeling setting, Hsieh et al. (2006) demonstrated that leaving the baseline hazard function unspecified leads to the underestimation of the standard errors of parameter estimators. Therefore, it is advisable to define an explicit form for $\hbar_0(t)$ by either choosing a form corresponding to a parametric distribution (Weibull, log-normal, or Gamma) or parametric but flexible form such as linear splines, B-splines or piecewise constant functions.

For inferential procedures, the standard error of the estimators must be obtained. As we mentioned above, Hsieh et al. (2006) notes that for accurate inference, the standard errors must be obtained using a parametric structure for the baseline hazard function. However, even in that case, under some circumstances, observed Fisher's information might lead to incorrect results. Additionally, Rizopoulos et al. (2008) highlights possible issues due to misspecification of the random effects distribution. Therefore, Hsieh et al. (2006)

and Rizopoulos et al. (2008) recommend following a bootstrap approach to estimate the standard errors of the estimators.

2.4 Varying-Coefficient Models

Parametric regression models are well established in literature to study the relationship between a response and its corresponding set of predictors. Although these models are easy to implement and mostly computationally feasible, they cannot be used to observe dynamic trends in real-world applications. Fan and Zhang (2008) points this limitation using a respiratory study from Hong Kong, where the primary interest is to assess the affects of daily measurements of pollutants, such as Sulfur Dioxide and dust, on the number of daily hospital admissions from 1994 to 1995. In this case, modeling the association between the pollutants and the number of daily hospital admissions as a constant would be inappropriate since daily factors may cause the association to vary. Kürüm et al. (2014) provide details of a study that explores the nonlinear association between the net ecosystem exchange of CO_2 (NEE) and the photosynthetically active radiation (PAR). The researchers demonstrate that the relationship between NEE and PAR depends on temperature; therefore, modeling the regression coefficients as a constant would be improper and the relationship between NEE and PAR should be allowed to change with temperature. These examples show real-life situations where allowing an association to be dynamic is necessary and traditional regression models cannot be employed to model these relationships. With the aim of increasing the flexibility of traditional regression models and reducing the modeling bias, varying-coefficient models were proposed.

Let Y denote the response variable and $\boldsymbol{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$ denote a vector of covariates, respectively, a varying-coefficient model is given as

$$Y = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}(U) + \epsilon, \tag{2.16}$$

where $\boldsymbol{\beta}(U) = \{\beta_1(U), \ldots, \beta_p(U)\}^{\mathrm{T}}$ is referred as the varying-coefficient functions and ϵ is the error term with $E(\epsilon|U) = 0$ and $Var(\epsilon|U) = \sigma^2(U)$. These models were first introduced by Cleveland et al. (1992) and became popular in the statistical literature due to the work by Hastie and Tibshirani (1993). Varying-coefficient models explore the dynamic features of the data by allowing the regression coefficients to change over a covariate U, such as time and temperature. At a fixed $U = u$ value, the model coefficients can be interpreted in a similar way to a linear regression model.

Time-varying coefficient models is a special case of the model presented in (2.16), where the dynamic association is modeled as a function of time. These models are particularly useful in longitudinal studies and they were first proposed by Hoover et al. (1998), where the authors discuss a longitudinal HIV study collected from infants born from HIV-infected mothers in Africa. In this study, the interest was to explore the relationship of the weight of the infant (response) with gender, HIV status, and maternal vitamin A levels (predictors). Studying these associations via a traditional regression model would be inappropriate as it would ignore an infant's development over time and would not allow the response-predictor relationships to change over time. Therefore, to model these dynamic trends in the data and reduce modeling bias, a time-varying coefficient model is needed.

Under the framework provided by Hoover et al. (1998), for the ith subject, $Y_{ij} = Y_i(t)$ and $\boldsymbol{X}_{ij} = \boldsymbol{X}_i(t)$ denote the outcome and a vector of covariates, respectively, at time $t = t_{ij}$ with $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. The time-varying coefficient model is expressed as

$$Y_{ij} = \boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta}(t) + \epsilon_{ij},$$

where $\boldsymbol{\beta}(t) = \{\beta_1(t), \ldots, \beta_p(t)\}$ is the time-varying coefficient function and $\epsilon_{ij} \sim N(0, \sigma^2)$ is the error term at time t_{ij}.

Estimation in varying-coefficient models have been studied extensively. In particular, there are three approaches to estimate the functional form of $\boldsymbol{\beta}(t)$: local polynomial techniques (Fan and Zhang, 1999; Hoover et al., 1998; Wu et al., 1998), polynomial splines (Huang et al., 2002, 2004), and smoothing splines (Hastie and Tibshirani, 1993; Hoover et al., 1998; Chiang et al., 2001). Fan and Zhang (2008) provide an excellent review of the literature on estimation and inference procedures for varying-coefficient models. In this chapter, we focus on spline-based methods for time-varying coefficient models, in particular, penalized B-splines.

Spline models are a global smoothing approach used to model the trend in the relationship between a covariate and an outcome by allowing the slope to change at various values, referred as knots, of the covariate. Spline models achieve this result by utilizing basis functions that alter the slope at the specified knots. A commonly used basis function is the truncated power functions as that is non-zero if a value of X is greater than a knot κ: $(X - \kappa)_+ = (X - \kappa)I((X > \kappa))$, where $I(\cdot)$ is the indicator function. B-spline functions is an alternative basis function that provide more numerical stable results than the truncated power functions, especially due to their performance at the boundaries (Ruppert et al., 2003). Although both methods, spline models with truncated power functions or the B-spline basis functions, can be used to approximate the time-varying coefficients, due to their desirable features, B-splines are more commonly used.

Estimation via the B-splines involve approximating the varying-coefficient function, $\beta_l(t)$, for $l = 1, \ldots, p$, as

$$\beta_l(t) = \boldsymbol{\theta}_l^{\mathrm{T}} \boldsymbol{B}(t),$$

where $\boldsymbol{B}(t) = \{B_1(t), \ldots, B_R(t)\}^{\mathrm{T}}$ are the B-spline basis functions and $\boldsymbol{\theta}_l = (\theta_{l1}, \ldots, \theta_{lR})^{\mathrm{T}}$ are the corresponding R-dimensional spline coefficients. The B-spline basis functions are computed by implementing the recursive algorithm as described by de Boor (1978) and Eilers et al. (2015).

In practice, application of the B-splines approach require specification of the number and the location of the knots. Misspecification of these quantities would possibly lead to biased estimates and inaccurate inference. Additionally, a small number of knots would lead to an undersmooth function, whereas a large number would produce an overfitted model. To overcome this challenge, Eilers and Marx (2010) recommends choosing a large number of knots and implementing a penalty term to control for overfitting the model. This leads to P-spline models where the likelihood function contains a roughness penalty term. Furthermore, in terms of the location of the knots, Eilers and Marx (2010) shows, via extensive simulation studies, that equally-spaced knots perform better than quantile-based knots, which have been shown to hinder the performance of the penalty term. For n subjects, the penalized log-likelihood function is defined as

$$\ell(\boldsymbol{\theta},\sigma^2) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta},\sigma^2) - \sum_{l=1}^{p} \frac{\lambda_l}{2} \boldsymbol{\theta}_l^{\mathrm{T}} \mathscr{D}^{(2)\mathrm{T}} \mathscr{D}^{(2)} \boldsymbol{\theta}_l, \tag{2.17}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)^{\mathrm{T}}$ are the spline coefficients, λ_l is the penalty term for the lth varying-coefficient function, and $\mathscr{D}^{(2)}$ is the matrix of the 2nd-order difference operator, $\Delta^2\theta_r = \Delta\theta_r - \Delta\theta_{r-1} = \theta_r - 2\theta_{r-1} + \theta_{r-2}$ for $r = 3, \ldots, p$, as defined by Eilers and Marx (1996), and

$$\ell_i(\boldsymbol{\theta},\sigma^2) = \sum_{i=1}^{n_i} -\frac{1}{2\sigma^2} \left\{ Y_{ij} - \sum_{l=1}^{p} X_{ijl} \boldsymbol{\theta}_l \boldsymbol{B}(t) \right\}^2 - \frac{1}{2}\log(2\pi\sigma^2),$$

is the log-likelihood contribution of the ith subject. The maximum likelihood estimates $\hat{\boldsymbol{\theta}}$ are found by maximizing (2.17) with respect $\boldsymbol{\theta}$ using a numerical approximation algorithm such as the Newton-Raphson.

An important component of the P-spline estimation method is choosing the optimal values of the penalty terms $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^{\mathrm{T}}$. Eilers and Marx (2010) recommends a leave-one-out cross-validation (LOOCV) approach to obtain the optimal values of $\boldsymbol{\lambda}$. The LOOCV approach identifies the optimal $\boldsymbol{\lambda}$ that minimizes $CV = \sum_{i=1}^{n} \sum_{j=1}^{n} \{Y_{ij} - \hat{Y}_{-(i)j}\}^2 / \sum_{i=1}^{n} n_i$, where $\hat{Y}_{-(i)j}$ is the predicted value of Y_{ij} from a model fitted from a data set without the ith subject. Although this approach is easy to implement and produces accurate results, it may become computationally challenging when the model involves a large number of covariates. An alternative method in this case is the random-coefficient splines model, where the optimal penalty term is obtained as part of the estimation procedure (Brumback et al., 1999; Ruppert et al., 2003; Goldsmith et al., 2011, 2012). Under this approach, $\beta_l(t)$, for $l = 1, \ldots, p$, can be approximated as

$$\beta_l(t) = \boldsymbol{\theta}_l \boldsymbol{B}_F(t) + \boldsymbol{\vartheta}_l \boldsymbol{B}_R(t),$$

where $\boldsymbol{\theta}_l = (\theta_{l1}, \theta_{l2})^{\mathrm{T}}$ and $\boldsymbol{\vartheta}_l \sim N(\mathbf{0}, \sigma^2_{\kappa l} \boldsymbol{I}_{n_\kappa})$ are the fixed and random effects, respectively, $\boldsymbol{B}_F(t) = (1, t)^{\mathrm{T}}$ and $\boldsymbol{B}_R(t) = \{(t - \kappa_1)_+, \ldots, (t - \kappa_{n_\kappa})_+\}^{\mathrm{T}}$ represent the spline basis functions with $\{\kappa_\varkappa\}_{\varkappa=1}^{n_\kappa}$ as the n_κ equally-spaced knots, $(a)_+ = (a)I\{(a) > 0\}$, and $I(\cdot)$ as the indicator function. The log-likelihood function is expressed as

$$\ell(\boldsymbol{\theta}, \boldsymbol{\vartheta}, \sigma^2, \boldsymbol{\sigma}^2_\kappa) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}, \boldsymbol{\vartheta}, \sigma^2) - \sum_{l=1}^{p} \frac{1}{2} \log(2\pi\sigma^2_{\kappa l}) - \frac{1}{2\sigma^2_{\kappa l}} \boldsymbol{\vartheta}_l^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\vartheta}_l, \qquad (2.18)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)^{\mathrm{T}}$, $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_p)^{\mathrm{T}}$, $\boldsymbol{\sigma}^2_\kappa = (\sigma^2_{\kappa 1}, \ldots, \sigma^2_{\kappa p})^{\mathrm{T}}$, and

$$\ell_i(\boldsymbol{\theta}, \boldsymbol{\vartheta}, \sigma^2) = \sum_{j=1}^{n_i} -\frac{1}{2\sigma^2} \left[Y_{ij} - \sum_{l=1}^{p} X_{ijl} \{\boldsymbol{\theta}_l \boldsymbol{B}_F(t) + \boldsymbol{\vartheta}_l \boldsymbol{B}_R(t)\} \right]^2 - \frac{1}{2} \log(2\pi\sigma^2)$$

is the ith subject's contribution to the log-likelihood function Using the methods described in Section 2.1 for estimation under mixed effects models, we obtain $\hat{\boldsymbol{\theta}}$ and predict the random effects $\boldsymbol{\vartheta}$. Additionally, the variance terms $\boldsymbol{\sigma}^2_\kappa$ in (2.18) act as the penalty term for the random coefficients, controlling the smoothness of the time-varying coefficient functions and preventing overfitting. Therefore, utilizing mixed effects models produce smooth functions as part of the estimation procedure without the need to perform a cross validation approach to find the optimal penalty. For inferential procedures, the standard errors for time-varying coefficient functions can be obtained using a negative inverse Hessian matrix, when P-splines models are used, or a bootstrap approach, when random-coefficient spline models are utilized.

Chapter 3

Multilevel Time-Varying Joint Models for Longitudinal and Survival Outcomes

3.1 Model

Consider the following framework, let $i = 1, \ldots, n$ index a cluster (facility); $j = 1, \ldots, n_i$ index the subjects (patients) within the ith cluster; and $k = 1, \ldots, n_{ij}$ index the recorded observations for the jth subject at time t_{ijk}. In our framework, each subject has a recorded longitudinal outcome $Y_{ijk} \equiv Y_{ij}(t_{ijk})$ and an observed event time T_{ij}. Let $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijp})^{\mathrm{T}}$ and $\boldsymbol{Z}_{i(j)} = \{Z_{i(j)1}, \ldots, Z_{i(j)q}\}^{\mathrm{T}}$ denote the subject- and facility-level predictors, respectively. In our motivating problem, the facility-level characteristics are reported every year, and $\boldsymbol{Z}_{i(j)}$ denotes those characteristics recorded in the calendar year prior to the jth patient's dialysis initiation. Thus, the facility-level predictors have both dialysis facility index i and the subject level index j.

The proposed joint modeling framework is composed of a submodel for each outcome. The longitudinal submodel is formulated with a generalized time-varying mixed effects model

$$m_{ij}(t) = E\{Y_{ij}(t)|\boldsymbol{X}_{ij}, \boldsymbol{Z}_{i(j)}, b_{ij}, b_i\} = g^{-1}\left\{\boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_X(t) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}}\boldsymbol{\beta}_Z(t) + b_{ij} + b_i\right\}, \quad (3.1)$$

where $g(\cdot)$ denotes the canonical link function, $\boldsymbol{\beta}_X(t) = \{\beta_{X_1}(t), \ldots, \beta_{X_p}(t)\}^{\mathrm{T}}$ and $\boldsymbol{\beta}_Z(t) = \{\beta_{Z_1}(t), \ldots, \beta_{Z_q}(t)\}^{\mathrm{T}}$ are the subject- and facility-level time-varying coefficients, respectively, b_{ij} and b_i are the subject- and facility-level random effects (RE) such that $b_{ij} \sim N(0, \sigma_S^2)$ and $b_i \sim N(0, \sigma_F^2)$, respectively. The random effects are assumed to be independent of each other. Although we impose distributional assumptions on the random effects, parameter estimation and inference in joint modeling are shown to be robust to misspecification of distribution of the random effects (Song et al., 2002; Hsieh et al., 2006; Rizopoulos et al., 2008). For the USRDS data application, $Y_{ij}(t)$ is a binary longitudinal outcome defined as the indicator of at least one hospitalization in a 3-month follow-up window with midpoint t for the subject j at the facility i. Thus, for the data analysis, $g(\cdot)$ takes the form of the logit link function $g(p) = \log(p/1-p)$.

For the survival submodel, let T_{ij}^* and C_{ij} denote the true event time and potential censoring time, respectively. The observed event time is defined as $T_{ij} = \min(T_{ij}^*, C_{ij})$ and δ_{ij} denotes the event indicator such that $\delta_{ij} = I(T_{ij}^* \leq C_{ij})$ with $I(\cdot)$ as the indicator function. The survival submodel is formulated as a Cox model with time-varying coefficients:

$$\begin{aligned}
h_{ij}\{t|\mathcal{M}_{ij}(t),\boldsymbol{X}_{ij},\boldsymbol{Z}_{i(j)}\} &= \lim_{\varepsilon\to 0}\Pr\{t\le T^*_{ij}<t+\varepsilon|T^*_{ij}\ge t,\mathcal{M}_{ij}(t),\boldsymbol{X}_{ij},\boldsymbol{Z}_{i(j)}\} \\
&= h_0(t)\exp\{\boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_X(t)+\boldsymbol{Z}_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_Z(t)+\alpha(t)m_{ij}(t)\}, \qquad (3.2)
\end{aligned}$$

where $h_0(t)$ denotes the baseline hazard function, $\mathcal{M}_{ij}(t)=\{m_{ij}(s),0\le s<t\}$ represents the history of the longitudinal outcome, $\boldsymbol{\gamma}_X(t)=\{\gamma_{X_1}(t),\ldots,\gamma_{X_p}(t)\}^{\mathrm{T}}$ and $\boldsymbol{\gamma}_Z(t)=\{\gamma_{Z_1}(t),\ldots,\gamma_{Z_q}(t)\}^{\mathrm{T}}$ are the subject- and facility-level time-varying coefficients, respectively, and $\alpha(t)$ represents the time-varying effect of the longitudinal outcome on the risk of an event. The corresponding survival function is constructed from the hazard function (3.2):

$$\begin{aligned}
\mathcal{S}_{ij}\{t|\mathcal{M}_{ij}(t),\boldsymbol{X}_{ij},\boldsymbol{Z}_{i(j)}\} &= \Pr\{T^*_{ij}>t|\mathcal{M}_{ij}(t),\boldsymbol{X}_{ij},\boldsymbol{Z}_{i(j)}\} \\
&= \exp\left[-\int_0^t h_0(s)\exp\{\boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_X(s)+\boldsymbol{Z}_{i(j)}^{\mathrm{T}}\boldsymbol{\gamma}_Z(s)+\alpha(s)m_{ij}(s)\}ds\right].
\end{aligned} \qquad (3.3)$$

In our framework, we assume that random effects account for the association between the longitudinal and survival outcomes. In other words, given the random effects, the longitudinal and survival outcomes are assumed to be independent; therefore, the joint density function is represented as

$$p(T_{ij},\delta_{ij},\boldsymbol{Y}_{ij},b_{ij},b_i;\boldsymbol{\theta})=p(T_{ij},\delta_{ij}|b_{ij},b_i;\boldsymbol{\theta})p(\boldsymbol{Y}_{ij}|b_{ij},b_i;\boldsymbol{\theta})p(b_{ij},b_i;\boldsymbol{\theta}), \qquad (3.4)$$

where $p(\cdot)$ is the probability density function, $\boldsymbol{Y}_{ij}=(Y_{ij1},\ldots,Y_{ijn_{ij}})^{\mathrm{T}}$, $\boldsymbol{\theta}=(\boldsymbol{\theta}_L^{\mathrm{T}},\boldsymbol{\theta}_S^{\mathrm{T}},\sigma_S^2,\sigma_F^2)^{\mathrm{T}}$ with $\boldsymbol{\theta}_L$ and $\boldsymbol{\theta}_S$ denoting the vector of parameters for the longitudinal and survival submodels, respectively. We assume that, in addition to the time-varying effects, the random effects also explain the association between the repeated measurements within a subject; therefore, the density functions for the longitudinal submodel and random effects are constructed as

$$\begin{aligned} p(\boldsymbol{Y}_{ij}|b_{ij}, b_i; \boldsymbol{\theta})p(b_{ij}, b_i; \boldsymbol{\theta}) &= \left[\prod_{k=1}^{n_{ij}} p\{Y_{ijk}|b_{ij}, b_i; \boldsymbol{\theta}\}\right] p(b_{ij}, b_i; \boldsymbol{\theta}) \\ &= \left(\prod_{k=1}^{n_{ij}} \frac{\exp[\{\boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_X(t_{ijk}) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}}\boldsymbol{\beta}_Z(t_{ijk}) + b_{ij} + b_i\}Y_{ijk}]}{1 + \exp\{\boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_X(t_{ijk}) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}}\boldsymbol{\beta}_Z(t_{ijk}) + b_{ij} + b_i\}} \right. \\ &\quad \left. \times (2\pi\sigma_S^2)^{-1/2} \exp\left(\frac{-b_{ij}^2}{2\sigma_S^2}\right)\right) (2\pi\sigma_F^2)^{-1/2} \exp\left(\frac{-b_i^2}{2\sigma_F^2}\right). \end{aligned}$$

The probability density function for the survival submodel is derived using the hazard function in (3.2) and survival function in (3.3):

$$\begin{aligned} p(T_{ij}, \delta_{ij}|b_{ij}, b_i; \boldsymbol{\theta}) &= h_{ij}\{t|\mathcal{M}_{ij}(t), \boldsymbol{X}_{ij}, \boldsymbol{Z}_{i(j)}\}^{\delta_{ij}} S_{ij}\{t|\mathcal{M}_{ij}(t), \boldsymbol{X}_{ij}, \boldsymbol{Z}_{i(j)}\} \\ &= \left[h_0(T_{ij}) \exp\left\{\boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_X(T_{ij}) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}}\boldsymbol{\gamma}_Z(T_{ij}) + \alpha(T_{ij})m_{ij}(T_{ij})\right\}\right]^{\delta_{ij}} \\ &\quad \times \exp\left[-\int_0^{T_{ij}} h_0(s) \exp\{\boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_X(s) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}}\boldsymbol{\gamma}_Z(s) + \alpha(s)m_{ij}(s)\}ds\right]. \end{aligned}$$

3.2 Estimation

In this section, we discuss two approaches to estimate the parameters involved in (3.4): P-spline models (described in Section 3.2.1) and random-coefficient spline models (described in Section 3.2.2). For more information on the computational details of each method, please refer to Appendix A.

3.2.1 P-Spline Model

We propose estimating the Multilevel Time-Varying Joint Model (MTJM) with an approximate Expectation-Maximization (EM) algorithm (Dempster et al., 1977), where the expectation step (E-step) treats the REs as missing, and the maximization step (M-step) estimates the parameters. Due to the high dimensionality of the random effects, specifically integrating the joint log-likelihood function with respect to the REs for each facility (for the USRDS data, the minimum and maximum number of patients (n_i) is 50 and 160, respectively), the fully exponential Laplace approximation is utilized to reduce the computational burden during the E-step. The EM algorithm iterates between the E-step, which computes the expected value of the random effects, and the M-step, which maximizes the parameters from the approximate expected complete likelihood function.

Let $\boldsymbol{u} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)^{\mathrm{T}}$ with $\boldsymbol{u}_i = (b_{i1}, b_{i2}, \cdots, b_{in_i}, b_i)^{\mathrm{T}}$ be the vector of all random effects (subject- and facility-level), and $\boldsymbol{\theta} = (\boldsymbol{\theta}_L^{\mathrm{T}}, \boldsymbol{\theta}_S^{\mathrm{T}}, \sigma_F^2, \sigma_S^2)^{\mathrm{T}}$ are the parameters involved in approximating the time-varying coefficient models and variances, the complete joint log-likelihood function is characterized as $\ell(\boldsymbol{u}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{u}_i, \boldsymbol{\theta})$ where

$$\begin{aligned}
\ell_i(\boldsymbol{u}_i, \boldsymbol{\theta}) &= \sum_{j=1}^{n_i} \log p(T_{ij}, \delta_{ij} | b_{ij}, b_i; \boldsymbol{\theta}) + \log p(\boldsymbol{Y}_{ij} | b_{ij}, b_i; \boldsymbol{\theta}) + \log p(b_{ij}, b_i; \boldsymbol{\theta}) \\
&= \sum_{j=1}^{n_i} \Bigg[\delta_{ij} \{ \log h_0(T_{ij}) + \boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_X(T_{ij}) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}} \boldsymbol{\gamma}_Z(T_{ij}) + \alpha(T_{ij}) m_{ij}(T_{ij}) \} \\
&\quad - \int_0^{T_{ij}} h_0(s) \exp\{ \boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_X(s) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}} \boldsymbol{\gamma}_Z(s) + \alpha(s) m_{ij}(s) \} ds \\
&\quad + \sum_{k=1}^{n_{ij}} \left\{ g(m_{ijk}) Y_{ijk} + \log(q_{ijk}) \right\} - \frac{b_{ij}^2}{2\sigma_S^2} - \frac{1}{2} \log(2\pi\sigma_S^2) \Bigg] \\
&\quad - \frac{b_i^2}{2\sigma_F^2} - \frac{1}{2} \log(2\pi\sigma_F^2), \qquad (3.5)
\end{aligned}$$

is the contribution of the ith facility to the joint log-likelihood, $m_{ijk} = m_{ij}(t_{ijk})$, $m_{ij}(t) = g^{-1}\{\boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_X(t) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}}\boldsymbol{\beta}_Z(t) + b_{ij} + b_i\}$, and $q_{ijk} = 1 - m_{ijk}$. The incomplete likelihood function, $L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int L_i(\boldsymbol{u}_i, \boldsymbol{\theta}) d\boldsymbol{u}_i$ with $L_i(\boldsymbol{u}_i, \boldsymbol{\theta})$ as the likelihood contribution of the ith facility, is used to compute the expected value and variance of the random effects.

The time-varying coefficients $\alpha(t)$, $\boldsymbol{\beta}_X(t)$, $\boldsymbol{\beta}_Z(t)$, $\boldsymbol{\gamma}_X(t)$, and $\boldsymbol{\gamma}_Z(t)$ are estimated via a P-splines approach (Eilers and Marx, 1996). P-splines are used to achieve sufficient smoothing with a high number of equally spaced knots and avoid over fitting with a penalty term on the differences of the adjacent B-spline coefficients. The time-varying functions (VCF) in the longitudinal and survival submodels are approximated as

$$\begin{aligned}
&\boldsymbol{\beta}_{X_\omega}(t) = \boldsymbol{\tau}_{X_\omega}^{\mathrm{T}}\boldsymbol{B}(t), \;\; \boldsymbol{\beta}_{Z_\nu}(t) = \boldsymbol{\tau}_{Z_\nu}^{\mathrm{T}}\boldsymbol{B}(t), \\
&\boldsymbol{\gamma}_{X_\omega}(t) = \boldsymbol{\psi}_{X_\omega}^{\mathrm{T}}\boldsymbol{B}(t), \;\; \boldsymbol{\gamma}_{Z_\nu}(t) = \boldsymbol{\psi}_{Z_\nu}^{\mathrm{T}}\boldsymbol{B}(t), \;\; \alpha(t) = \boldsymbol{\psi}_\alpha \boldsymbol{B}(t),
\end{aligned} \tag{3.6}$$

where $\boldsymbol{B}(t) = \{B_1(t), B_2(t), \ldots, B_R(t)\}^{\mathrm{T}}$ represents the B-spline basis functions, and $\boldsymbol{\tau}_{X_\omega}$, $\boldsymbol{\tau}_{Z_\nu}$, $\boldsymbol{\psi}_{X_\omega}, \boldsymbol{\psi}_{Z_\nu}, \boldsymbol{\psi}_\alpha$ are the R-dimensional spline coefficients with $\omega = 1, \ldots, p$ and $\nu = 1, \ldots, q$. The baseline hazard function, $h_0(t)$, is approximated using the same P-splines approach such that $\log\{h_0(t)\} = \boldsymbol{\psi}_h^{\mathrm{T}}\boldsymbol{B}(t)$, where $\boldsymbol{\psi}_h = (\psi_{h_1}, \ldots, \psi_{h_R})^{\mathrm{T}}$.

Under the P-splines approach, the log-likelihood in (3.5) is rewritten as

$$\ell_i(\boldsymbol{u}_i, \boldsymbol{\theta}_{\mathcal{P}}) = \sum_{j=1}^{n_i} \log p(T_{ij}, \delta_{ij} | b_{ij}, b_i; \boldsymbol{\theta}_{\mathcal{P}}) + \log p(\boldsymbol{Y}_{ij} | b_{ij}, b_i; \boldsymbol{\theta}_{L\mathcal{P}}) + \log p(b_{ij}, b_i; \boldsymbol{\theta}_{\mathcal{P}}), \tag{3.7}$$

where $\boldsymbol{\theta}_{\mathcal{P}} = (\boldsymbol{\theta}_{L\mathcal{P}}, \boldsymbol{\theta}_{S\mathcal{P}}, \sigma_s^2, \sigma_F^2)^{\mathrm{T}}$ denotes the vector of parameters including the P-spline coefficients involved in the longitudinal and survival submodels

$\boldsymbol{\theta}_{L\mathscr{P}} = (\boldsymbol{\tau}_{X_1}^{\mathrm{T}}, \ldots, \boldsymbol{\tau}_{X_p}^{\mathrm{T}}, \boldsymbol{\tau}_{Z_1}^{\mathrm{T}}, \ldots, \boldsymbol{\tau}_{Z_q}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\theta}_{S\mathscr{P}} = \left(\boldsymbol{\psi}_{\alpha}^{\mathrm{T}}, \boldsymbol{\psi}_{\hbar}^{\mathrm{T}}, \boldsymbol{\psi}_{X_1}^{\mathrm{T}}, \ldots, \boldsymbol{\psi}_{X_p}^{\mathrm{T}}, \boldsymbol{\psi}_{Z_1}^{\mathrm{T}}, \ldots, \boldsymbol{\psi}_{Z_q}^{\mathrm{T}}\right)^{\mathrm{T}}$, respectively. Similarly, the incomplete likelihood is redefined as $L(\boldsymbol{\theta}_{\mathscr{P}}) = \sum_{i=1}^{n} \int L_i(\boldsymbol{u}_i, \boldsymbol{\theta}_{\mathscr{P}}) d\boldsymbol{u}_i$.

The following steps provide an overview of our estimation procedure to obtain $\hat{\boldsymbol{\theta}}_{\mathscr{P}} = (\hat{\boldsymbol{\theta}}_{L\mathscr{P}}, \hat{\boldsymbol{\theta}}_{S\mathscr{P}}, \hat{\sigma}_s^2, \hat{\sigma}_F^2)^{\mathrm{T}}$:

1. The initial values of $\boldsymbol{\theta}_{L\mathscr{P}}$ and $\boldsymbol{\theta}_{S\mathscr{P}}$ are vectors with all elements set to 0. The initial values of subject- and facility-level random effects variances σ_S^2 and σ_F^2, respectively, are obtained using a generalized multilevel linear mixed effects model.

2. (E-step) A fully exponential Laplace approximation is used to obtain the estimates of the posterior mean and variance for the random effects $\boldsymbol{u}_i$. This leads to the approximated expected likelihood.

3. (M-step) The expected complete log-likelihood function is maximized to obtain the closed-form solutions of the current estimates of σ_S^2 and σ_F^2. A Newton-Raphson algorithm is used to maximize the approximate expected complete log-likelihood function to obtain the current estimates of $(\boldsymbol{\theta}_{L\mathscr{P}}^{\mathrm{T}}, \boldsymbol{\theta}_{S\mathscr{P}}^{\mathrm{T}})^{\mathrm{T}}$.

4. The EM algorithm iterates between the E-step and M-step until the difference between two consecutive log-likelihood values are less than a pre-defined tolerance level ϵ.

E-step and the Fully Exponential Laplace Approximation

In the E-step, the posterior mean and variance of the random effects are computed as

$$\boldsymbol{u}_{i0} = \frac{\int \boldsymbol{u}_i L_i(\boldsymbol{u}_i; \boldsymbol{\theta}_{\mathcal{P}}) d\boldsymbol{u}_i}{\int L_i(\boldsymbol{u}_i; \boldsymbol{\theta}_{\mathcal{P}}) d\boldsymbol{u}_i} \text{ and } \boldsymbol{v}_{i0} = \frac{\int (\boldsymbol{u}_i - \boldsymbol{u}_{i0})(\boldsymbol{u}_i - \boldsymbol{u}_{i0})^{\mathrm{T}} L_i(\boldsymbol{u}_i; \boldsymbol{\theta}_{\mathcal{P}}) d\boldsymbol{u}_i}{\int L_i(\boldsymbol{u}_i; \boldsymbol{\theta}_{\mathcal{P}}) d\boldsymbol{u}_i}, \quad (3.8)$$

respectively. The integrals involved in (3.8) do not have closed-form solutions and are potentially high-dimensional due to the random effects $\boldsymbol{u}_i = (\mathscr{b}_{i1}, \ldots, \mathscr{b}_{in_i}, \mathscr{b}_i)^{\mathrm{T}}$ (of dimensions 51 - 163 in our motivating data). Therefore, for approximating these integrals, we utilize the fully exponential Laplace approximation (Tierney et al., 1989). However, the fully exponential Laplace approximation can only be used for strictly positive function and the integrands in our estimation might not satisfy this condition. To avoid this drawback, we follow the approach discussed in Rizopoulos et al. (2009) and use the cumulant generating function, $\log[E\{\exp(\boldsymbol{c}^{\mathrm{T}}\boldsymbol{u}_i)\}]$ (where $\boldsymbol{c} = (c_1, \ldots, c_{n_i+1})^{\mathrm{T}}$ is a constant vector), which is always positive, to estimate the posterior mean and variance. Under this scenario, the posterior mean and variance are obtained by differentiating and evaluating the cumulant generating function at $\boldsymbol{c} = \boldsymbol{0}$, $\boldsymbol{u}_{i0} = \partial \log[E\{\exp(\boldsymbol{c}^{\mathrm{T}}\boldsymbol{u}_i)\}]/\partial \boldsymbol{c}^{\mathrm{T}}|_{\boldsymbol{c}=\boldsymbol{0}}$ and $\boldsymbol{v}_{i0} = \partial^2 \log[E\{\exp(\boldsymbol{c}^{\mathrm{T}}\boldsymbol{u}_i)\}]/\partial \boldsymbol{c}^{\mathrm{T}}\partial \boldsymbol{c}|_{\boldsymbol{c}=\boldsymbol{0}}$.

The fully exponential Laplace approximation is conducted in two steps: the mode and correction steps. In the mode step, the modes for $\boldsymbol{u}_{i0}$, that is, $\hat{\boldsymbol{u}}_i = \hat{\boldsymbol{u}}_i^{(c)}|_{\boldsymbol{c}=\boldsymbol{0}}$ where $\hat{\boldsymbol{u}}_i^{(c)}|_{\boldsymbol{c}=\boldsymbol{0}} = \operatorname{argmax}_{\boldsymbol{u}_i} \left\{\ell_i(\boldsymbol{u}_i, \boldsymbol{\theta}_{\mathcal{P}}) + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{u}_i\right\}$, are obtained, with $\ell_i(\boldsymbol{u}_i, \boldsymbol{\theta}_{\mathcal{P}})$ is defined in (3.7). This maximization is implemented via a safeguarded Newton-Raphson algorithm where $\hat{\boldsymbol{u}}_i^{it}$ is updated as

$$\hat{\boldsymbol{u}}_i^{it+1} = \hat{\boldsymbol{u}}_i^{it} - s\{\mathscr{H}_{\boldsymbol{u}_i}(\hat{\boldsymbol{u}}_i^{it})\}^{-1}\mathscr{G}_{\boldsymbol{u}_i}(\hat{\boldsymbol{u}}_i^{it}),$$

where $\mathcal{H}_{\boldsymbol{u}_i}(\hat{\boldsymbol{u}}_i^{it}) = \boldsymbol{\Sigma}_i^{(c)}|_{(\boldsymbol{c},\boldsymbol{u}_i)=(\boldsymbol{0},\hat{\boldsymbol{u}}_i^{it})}$ with

$\boldsymbol{\Sigma}_i^{(c)} = -\partial^2 \left\{\ell_i(\boldsymbol{u}_i, \boldsymbol{\theta}_{\mathcal{P}}) + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{u}_i\right\} / \partial \boldsymbol{u}_i^{\mathrm{T}} \partial \boldsymbol{u}_i$, $\mathcal{G}_{\boldsymbol{u}_i}(\hat{\boldsymbol{u}}_i^{it}) = -\partial \ell_i(\boldsymbol{u}_i, \boldsymbol{\theta}_{\mathcal{P}})/\partial \boldsymbol{u}_i^{\mathrm{T}}|_{\boldsymbol{u}_i=\hat{\boldsymbol{u}}_i^{it}}$, and s denotes the step size.

In the correction step, the posterior mean and variance are approximated using the modes from the first step. Differentiating the cumulant-generating function and evaluating at $\boldsymbol{c} = \boldsymbol{0}$, the posterior mean and variance are computed as

$$\boldsymbol{u}_{i0} = \hat{\boldsymbol{u}}_i - \frac{1}{2}\mathrm{tr}(\mathcal{V}) \text{ and } \boldsymbol{v}_{i0} = \boldsymbol{\Sigma}_i^{-1} - \frac{1}{2}\mathrm{tr}\left\{ -\mathcal{V}\mathcal{V}^{\mathrm{T}} + \boldsymbol{\Sigma}_i^{-1} \left.\frac{\partial^2 \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial \boldsymbol{c}}\right|_{(\boldsymbol{c},\boldsymbol{u}_i)=(\boldsymbol{0},\hat{\boldsymbol{u}}_i)} \right\},$$

where $\mathcal{V} = \boldsymbol{\Sigma}_i^{-1}\{\partial \boldsymbol{\Sigma}_i^{(c)}/\partial \boldsymbol{c}^{\mathrm{T}}\}|_{(\boldsymbol{c},\boldsymbol{u}_i)=(\boldsymbol{0},\hat{\boldsymbol{u}}_i)}$, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i^{(c)}|_{\boldsymbol{c}=\boldsymbol{0}}$, with $\hat{\boldsymbol{u}}_i$ and $\boldsymbol{\Sigma}_i^{-1}$ are the modes and the inverse of $\boldsymbol{\Sigma}_i^{it}$ from the last iteration of the Newton-Raphson algorithm, respectively, from the first step. More information on the fully exponential Laplace approximation is provided in Appendix A.1.

After estimating the posterior mean and variance of the random effects, the expectation of the complete joint log-likelihood function is approximated in the E-step. However, the closed-form expression of the expectation, denoted as

$\sum_{i=1}^{n} E\{\ell_i(\boldsymbol{u}_i, \boldsymbol{\theta}_{\mathcal{P}})|\boldsymbol{Y}_i, \boldsymbol{T}_i, \boldsymbol{\delta}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}_{\mathcal{P}}^*\}$ with $\boldsymbol{\theta}_{\mathcal{P}}^*$ representing the current estimates of $\boldsymbol{\theta}_{\mathcal{P}}$, is intractable. Therefore, we employ a second degree Taylor's expansion around the estimated posterior mean of $\boldsymbol{u}_{i0}$ to approximate the expected log-likelihood as follows:

$$\begin{aligned}
\ell^*(\boldsymbol{\theta}^*_{\mathcal{P}}) &= \sum_{i=1}^{n} \ell_i(\boldsymbol{u}^*_{i0}, \boldsymbol{\theta}^*_{\mathcal{P}}) + \ell'_i(\boldsymbol{u}^*_{i0}, \boldsymbol{\theta}^*_{\mathcal{P}}) E(\boldsymbol{u}_i - \boldsymbol{u}^*_{i0}) - \frac{1}{2} E(\boldsymbol{u}_i - \boldsymbol{u}^*_{i0})^{\mathrm{T}} \boldsymbol{\Sigma}^*_i E(\boldsymbol{u}_i - \boldsymbol{u}^*_{i0}) \\
&= \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n_i} \left[\delta_{ij} \{ \log h^*_0(T_{ij}) + \boldsymbol{X}^{\mathrm{T}}_{ij} \boldsymbol{\gamma}^*_X(T_{ij}) + \boldsymbol{Z}^{\mathrm{T}}_{i(j)} \boldsymbol{\gamma}^*_Z(T_{ij}) + \alpha^*(T_{ij}) m^*_{ij}(T_{ij}) \} \right.\right. \\
&\quad - \int_0^{T_{ij}} h^*_0(s) \exp\{ \boldsymbol{X}^{\mathrm{T}}_{ij} \boldsymbol{\gamma}^*_X(s) + \boldsymbol{Z}^{\mathrm{T}}_{i(j)} \boldsymbol{\gamma}^*_Z(s) + \alpha^*(s) m^*_{ij}(s) \} ds \\
&\quad \left. + \sum_{k=1}^{n_{ij}} \left\{ g(m^*_{ijk}) Y_{ijk} + \log(q^*_{ijk}) \right\} - \frac{(b^*_{0ij})^2 + v^*_{b,ij0}}{2\sigma^{2*}_S} - \frac{1}{2} \log(2\pi\sigma^{2*}_S) + \mathcal{R}^*_{ij} \Delta^{(2)*}_{ij} \right] \\
&\quad \left. - \frac{(b^*_{i0})^2 + v^*_{b,i0}}{2\sigma^{2*}_F} - \frac{1}{2} \log(2\pi\sigma^{2*}_F) \right\}, \qquad (3.9)
\end{aligned}$$

where $m^*_{ij}(t) = g^{-1}\left\{ \boldsymbol{X}^{\mathrm{T}}_{ij} \boldsymbol{\beta}^*_X(t) + \boldsymbol{Z}^{\mathrm{T}}_{i(j)} \boldsymbol{\beta}^*_Z(t) + b^*_{ij0} + b^*_{i0} \right\}$, $m_{ijk} = m^*_{ij}(t_{ijk})$,, $q_{ijk} = 1 - m_{ijk}$, $E(\boldsymbol{u}_i - \boldsymbol{u}^*_{i0}) = 0$, $\Delta^{(2)*}_{ij}$ is $\Delta^{(2)}_{ij}$ (defined in the Appendix A.1), and $\boldsymbol{\Sigma}^*_i = -\partial^2 \ell_i(\boldsymbol{u}_i, \boldsymbol{\theta}^*_{\mathcal{P}}) / \partial \boldsymbol{u}^{\mathrm{T}}_i \partial \boldsymbol{u}_i |_{(\boldsymbol{u}_i) = (\boldsymbol{u}^*_{i0})}$. Moreover, $\boldsymbol{u}^*_{i0}$ denotes the estimated posterior mean of $\boldsymbol{u}_{i0}$ obtained in the E-step, $\mathcal{R}^*_{ij} = \frac{v^*_{b,ij0} + 2\eta^*_{ij0} + v^*_{b,i0}}{2}$ with $v^*_{b,ij0}$, $v^*_{b,i0}$, and η^*_{ij0} representing the posterior variance for the subject- and facility-level random effects, and the posterior covariance between the subject- and facility-level random effects, respectively.

M-step

The variance for the subject- and facility-level random effects are estimated by setting the score functions of the incomplete log-likelihood $\sum_{i=1}^{n} \log \int L_i(\boldsymbol{u}_i; \boldsymbol{\theta}_{\mathcal{P}}) d\boldsymbol{u}_i$ to zero. The score functions of the incomplete log-likelihood function with respect to σ^2_S and σ^2_F are given as

$$V(\sigma_S^2) = \sum_{i=1}^{n} \frac{\partial}{\partial \sigma_S^2} \log\left\{\int L_i(\boldsymbol{u}_i, \boldsymbol{\theta}_{\mathcal{P}}) d\boldsymbol{u}_i\right\} = \sum_{i=1}^{n} \int \sum_{j=1}^{n_i} \left(\frac{b_{ij}^2}{2\sigma_S^2} - \frac{1}{\sigma_S^2}\right) \mathcal{L}(\boldsymbol{u}_i) d\boldsymbol{u}_i = \sum_{i=1}^{n} V_i(\sigma_S^2)$$

and

$$V(\sigma_F^2) = \sum_{i=1}^{n} \frac{\partial}{\partial \sigma_F^2} \log\left\{\int L_i(\boldsymbol{u}_i, \boldsymbol{\theta}_{\mathcal{P}}) d\boldsymbol{u}_i\right\} = \sum_{i=1}^{n} \int \left(\frac{b_i^2}{2\sigma_F^2} - \frac{1}{\sigma_F^2}\right) \mathcal{L}(\boldsymbol{u}_i) d\boldsymbol{u}_i = \sum_{i=1}^{n} V_i(\sigma_F^2),$$

where $\mathcal{L}(\boldsymbol{u}_i) = L_i(\boldsymbol{u}_i; \boldsymbol{\theta}_{\mathcal{P}}) / \int L_i(\boldsymbol{u}_i; \boldsymbol{\theta}_{\mathcal{P}}) d\boldsymbol{u}_i$ is the posterior density function of $\boldsymbol{u}_i$. Setting $V(\sigma_S^2)$ and $V(\sigma_F^2)$ to zero, the estimates of σ_S^2 and σ_F^2 for the current iteration are given as

$$\sigma_S^{2*} = \left(\sum_{i=1}^{n} n_i\right)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \{(b_{ij0}^*)^2 + \nu_{b,ij0}^*\} \text{ and } \sigma_F^{2*} = n^{-1} \sum_{i=1}^{n} \{(b_{i0}^*)^2 + \nu_{b,i0}^*\}.$$

The parameters $\boldsymbol{\theta}_{\mathcal{P}}^{\backslash\sigma} = (\boldsymbol{\theta}_{L\mathcal{P}}^{\mathrm{T}}, \boldsymbol{\theta}_{S\mathcal{P}}^{\mathrm{T}})^{\mathrm{T}}$ are estimated by maximizing the penalized log-likelihood

$$\ell_\lambda^*(\boldsymbol{\theta}_{\mathcal{P}}^{\backslash\sigma *}) = \ell^*(\boldsymbol{\theta}_{\mathcal{P}}^{\backslash\sigma *}) - \frac{1}{2}\boldsymbol{\lambda}\mathrm{diag}\left(\Xi^* \mathcal{D}^{(2)\mathrm{T}} \mathcal{D}^{(2)} \Xi^{*\mathrm{T}}\right), \tag{3.10}$$

where $\ell^*(\cdot)$ is the log-likelihood function specified in (3.9), $\boldsymbol{\lambda}$ is a vector of penalty terms for each time-varying coefficient function, $\mathcal{D}^{(2)}$ is the second order difference matrix (defined in Section 2.4), and Ξ^* is a matrix with each row corresponding to the coefficients for the P-spline functions defined in (3.6) and the baseline hazard function (the penalized log-likelihood is further defined in the Appendix A.1.2). Note that the parameters in $\boldsymbol{\theta}_{\mathcal{P}}^{\backslash\sigma}$ do not have closed-form solutions; therefore, we employ the Newton-Raphson algorithm and update $\hat{\boldsymbol{\theta}}_{\mathscr{P}}^{\setminus\sigma}$ through

$$\hat{\boldsymbol{\theta}}_{\mathscr{P}}^{\setminus\sigma(it+1)} = \hat{\boldsymbol{\theta}}_{\mathscr{P}}^{\setminus\sigma(it)} - \boldsymbol{\mathscr{H}}_{\boldsymbol{\theta}_{\mathscr{P}}^{\setminus\sigma}}^{(it)} \boldsymbol{\mathscr{G}}_{\boldsymbol{\theta}_{\mathscr{P}}^{\setminus\sigma}}^{(it)},$$

where 'it' is the current iteration, $\boldsymbol{\mathscr{G}}_{\boldsymbol{\theta}_{\mathscr{P}}^{\setminus\sigma}}^{(it)}$ and $\boldsymbol{\mathscr{H}}_{\boldsymbol{\theta}_{\mathscr{P}}^{\setminus\sigma}}^{(it)}$ are the gradient and Hessian of the penalized log-likelihood function (3.10) with respect to $\boldsymbol{\theta}_{\mathscr{P}}^{\setminus\sigma}$, respectively, evaluated at the current estimates $\hat{\boldsymbol{\theta}}_{\mathscr{P}}^{\setminus\sigma(it)}$.

3.2.2 Random-Coefficient Spline Model

Although the P-splines approach described in the previous section is very flexible and performs well in most cases, as the number of covariates get larger, determining the optimal penalties for each time-varying coefficient function becomes computationally infeasible. To overcome this challenge, we propose to utilize mixed effects models that estimate the penalty as part of the estimation procedure. This leads to our Multilevel Time-Varying Joint Model with random-coefficient splines (MTJMRE). Under the random-coefficient spline approach, the time-varying functions in the longitudinal and survival submodels ((3.1), (3.2)) are approximated as

$$\boldsymbol{\beta}_{X_\omega}(t) = \boldsymbol{\phi}_{X_\omega}^{\mathrm{T}} \boldsymbol{B}_F(t) + \boldsymbol{\varphi}_{X_\omega}^{\mathrm{T}} \boldsymbol{B}_R(t), \;\; \boldsymbol{\beta}_{Z_\nu}(t) = \boldsymbol{\phi}_{Z_\nu}^{\mathrm{T}} \boldsymbol{B}_F(t) + \boldsymbol{\varphi}_{Z_\omega}^{\mathrm{T}} \boldsymbol{B}_R(t),$$

$$\boldsymbol{\gamma}_{X_\omega}(t) = \boldsymbol{\rho}_{X\omega}^{\mathrm{T}} \boldsymbol{B}_F(t) + \boldsymbol{\varrho}_{X\omega}^{\mathrm{T}} \boldsymbol{B}_R(t), \;\; \boldsymbol{\gamma}_{Z_\nu}(t) = \boldsymbol{\rho}_{Z\omega}^{\mathrm{T}} \boldsymbol{B}_F(t) + \boldsymbol{\varrho}_{Z\omega}^{\mathrm{T}} \boldsymbol{B}_R(t),$$

$$\alpha(t) = \boldsymbol{\rho}_{\alpha}^{\mathrm{T}} \boldsymbol{B}_F(t) + \boldsymbol{\varrho}_{\alpha}^{\mathrm{T}} \boldsymbol{B}_R(t),$$

where $\boldsymbol{B}_F(t) = \{1, t\}^{\mathrm{T}}$ and $\boldsymbol{B}_R(t) = \{(t-\kappa_1)_+, \ldots, (t-\kappa_{n_\kappa})_+\}^{\mathrm{T}}$ with $(a)_+ = (a)I\{(a) > 0\}$ represents the spline basis functions, $\boldsymbol{\phi}_{X_\omega}, \boldsymbol{\phi}_{Z_\nu}, \boldsymbol{\rho}_{X_\omega}, \boldsymbol{\rho}_{Z_\nu}, \boldsymbol{\rho}_\alpha$ are the 2-dimensional spline coefficients, and $\boldsymbol{\varphi}_{X_\omega} \sim N(\mathbf{0}, \sigma^2_{\varphi_{x_\omega}} \boldsymbol{I}_{n_\kappa})$, $\boldsymbol{\varphi}_{Z_\nu} \sim N(\mathbf{0}, \sigma^2_{\varphi_{z_\nu}} \boldsymbol{I}_{n_\kappa})$, $\boldsymbol{\varrho}_{X_\omega} \sim N(\mathbf{0}, \sigma^2_{\varrho_{x_\omega}} \boldsymbol{I}_{n_\kappa})$, $\boldsymbol{\varrho}_{Z_\nu} \sim N(\mathbf{0}, \sigma^2_{\varrho_{z_\omega}} \boldsymbol{I}_{n_\kappa})$, $\boldsymbol{\varrho}_{\alpha} \sim N(\mathbf{0}, \sigma^2_{\alpha} \boldsymbol{I}_{n_\kappa})$ are the n_κ-dimensional spline random effect (S-RE) coefficients with $\omega = 1, \ldots, p$ and $\nu = 1, \ldots, q$, and $\boldsymbol{I}_{n_\kappa}$ is the $n_\kappa \times n_\kappa$ identity matrix. The baseline hazard function, $\hbar_0(t)$, is approximated using the same random-coefficient spline approach such that $\log\{\hbar_0(t)\} = \boldsymbol{\rho}_{\hbar}^{\mathrm{T}} \boldsymbol{B}_F(t) + \boldsymbol{\varrho}_{\hbar}^{\mathrm{T}} \boldsymbol{B}_R(t)$, where $\boldsymbol{\rho}_{\hbar}$ is the 2-dimensional coefficients and $\boldsymbol{\varrho}_{\hbar} \sim N(\mathbf{0}, \sigma^2_{\hbar} \boldsymbol{I}_{n_\kappa})$ is the n_κ-dimensional random effects.

Let $\boldsymbol{\theta}_{\mathcal{R}} = (\boldsymbol{\theta}_{L\mathcal{R}}^{\mathrm{T}}, \boldsymbol{\theta}_{S\mathcal{R}}^{\mathrm{T}}, \boldsymbol{\theta}_{\sigma}^{\mathrm{T}}, \sigma_F^2, \sigma_S^2)^{\mathrm{T}}$ be the full parameter vector with $\boldsymbol{\theta}_{L\mathcal{R}} = (\boldsymbol{\phi}_{X_1}^{\mathrm{T}}, \ldots, \boldsymbol{\phi}_{X_p}^{\mathrm{T}}, \boldsymbol{\phi}_{Z_1}^{\mathrm{T}}, \ldots, \boldsymbol{\phi}_{Z_p}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\theta}_{S\mathcal{R}} = (\boldsymbol{\rho}_{\alpha}^{\mathrm{T}}, \boldsymbol{\rho}_{\hbar}^{\mathrm{T}}, \boldsymbol{\rho}_{X_1}^{\mathrm{T}}, \ldots, \boldsymbol{\rho}_{X_p}^{\mathrm{T}}, \boldsymbol{\rho}_{Z_1}^{\mathrm{T}}, \ldots, \boldsymbol{\rho}_{Z_q}^{\mathrm{T}})^{\mathrm{T}}$ as the spline RE coefficients in the longitudinal and survival submodels, respectively, and $\boldsymbol{\theta}_{\sigma} = \{\sigma^2_{\alpha}, \sigma^2_{\hbar}, (\boldsymbol{\sigma}^2_x)^{\mathrm{T}}, (\boldsymbol{\sigma}^2_z)^{\mathrm{T}}\}^{\mathrm{T}}$ as the vector of all their corresponding variances, where $\boldsymbol{\sigma}^2_x = (\sigma^2_{\varphi_{x_1}}, \ldots, \sigma^2_{\varphi_{x_p}}, \sigma^2_{\varrho_{x_1}}, \ldots, \sigma^2_{\varrho_{x_p}})^{\mathrm{T}}$ and $\boldsymbol{\sigma}^2_z = (\sigma^2_{\varphi_{z_1}}, \ldots, \sigma^2_{\varphi_{z_q}}, \sigma^2_{\varrho_{z_1}}, \ldots, \sigma^2_{\varrho_{z_q}})^{\mathrm{T}}$ represent all the patient- and facility-level spline RE coefficient variances, respectively. In addition, similar to Section 3.2.1, we denote $\boldsymbol{u} = (\boldsymbol{u}_1^{\mathrm{T}}, \ldots \boldsymbol{u}_n^{\mathrm{T}})^{\mathrm{T}}$, with $\boldsymbol{u}_i = (b_{i1}, \ldots, b_{in_i}, b_i)^{\mathrm{T}}$ denoting the REs for the ith facility, as the vector of all subject and facility-level random effects and $\boldsymbol{\vartheta} = (\boldsymbol{\varphi}_X^{\mathrm{T}}, \boldsymbol{\varphi}_Z^{\mathrm{T}}, \boldsymbol{\varrho}_{\alpha}^{\mathrm{T}}, \boldsymbol{\varrho}_{\hbar}^{\mathrm{T}}, \boldsymbol{\varrho}_X^{\mathrm{T}}, \boldsymbol{\varrho}_Z^{\mathrm{T}})^{\mathrm{T}}$ as the vector of all spline RE coefficients with $\boldsymbol{\varphi}_X = (\boldsymbol{\varphi}_{X_1}^T, \ldots, \boldsymbol{\varphi}_{X_p}^T)^{\mathrm{T}}$, $\boldsymbol{\varphi}_Z = (\boldsymbol{\varphi}_{Z_1}^T, \ldots, \boldsymbol{\varphi}_{Z_q}^T)^{\mathrm{T}}$, $\boldsymbol{\varrho}_X = (\boldsymbol{\varrho}_{X_1}^T, \ldots, \boldsymbol{\varrho}_{X_p}^T)^{\mathrm{T}}$, $\boldsymbol{\varrho}_Z = (\boldsymbol{\varrho}_{Z_1}^T, \ldots, \boldsymbol{\varrho}_{Z_q}^T)^{\mathrm{T}}$. Under the random-coefficient spline approximation for each varying-coefficient function and baseline hazard function, the joint density function in (3.4) is redefined as

$$
\begin{aligned}
p(T_{ij}, \delta_{ij}, \boldsymbol{Y}_{ij}, b_i, b_{ij}, \boldsymbol{\vartheta}; \boldsymbol{\theta}_{\mathcal{R}}) \;=\;& p(T_{ij}, \delta_{ij} | b_{ij}, b_i; \boldsymbol{\vartheta}, \boldsymbol{\theta}_{\mathcal{R}}) p(\boldsymbol{Y}_{ij} | b_{ij}, b_i; \boldsymbol{\vartheta}, \boldsymbol{\theta}_{\mathcal{R}}) p(b_{ij}, b_i; \boldsymbol{\theta}_{\mathcal{R}}) \times \\
& p(\boldsymbol{\varrho}_{\alpha}; \sigma^2_{\alpha}) p(\boldsymbol{\varrho}_{\hbar}; \sigma^2_{\hbar}) \prod_{\omega=1}^{p} p(\boldsymbol{\varphi}_{X\omega}; \sigma^2_{\varphi_{x_\omega}}) p(\boldsymbol{\varrho}_{X_\omega}; \sigma^2_{\varrho_{x_\omega}}) \\
& \prod_{\nu=1}^{q} p(\boldsymbol{\varphi}_{Z_\nu}; \sigma^2_{\varphi_{z_\nu}}) p(\boldsymbol{\varrho}_{Z_\nu}; \sigma^2_{\varrho_{z_\nu}}).
\end{aligned}
$$

We propose to estimate the parameters $\boldsymbol{\theta}_{\mathcal{R}} = (\boldsymbol{\theta}_{L\mathcal{R}}^{\mathrm{T}}, \boldsymbol{\theta}_{S\mathcal{R}}^{\mathrm{T}}, \boldsymbol{\theta}_{\sigma}^{\mathrm{T}}, \sigma_F^2, \sigma_S^2)^{\mathrm{T}}$ via an EM algorithm:

1. The initial values for $\boldsymbol{\theta}_{L\mathcal{R}}$ and $\boldsymbol{\theta}_{S\mathcal{R}}$ are vectors with all the elements set to 0. The initial values for the σ_S^2 and σ_F^2 are obtained from a generalized linear mixed effects model. The variances associated with the S-RE are obtained from generalized additive models.

2. (E-step) A fully exponential Laplace approximation is used to obtain the posterior mean and variance for the REs $\boldsymbol{u}$ and $\boldsymbol{\vartheta}$. This leads to the approximate expected log-likelihood function.

3. (M-step) The parameters in $\boldsymbol{\theta}_{\mathcal{R}}$ are estimated. The S-RE variances $\boldsymbol{\theta}_{\sigma}$, σ_S^2 and σ_F^2 are estimated with closed-form solutions. A Newton-Raphson algorithm is used to maximize the approximate expected log-likelihood function to estimate $\boldsymbol{\theta}_{L\mathcal{R}}$ and $\boldsymbol{\theta}_{S\mathcal{R}}$.

4. The algorithm iterates between the E-step and M-step until the relative difference between two consecutive log-likelihood values are less than a pre-defined value.

E-Step and the Fully Exponential Laplace Approximation

In the E-step, we first compute the posterior mean and variance of the subject- and facility-level REs $\boldsymbol{u} = (\boldsymbol{u}_1^{\mathrm{T}}, \dots \boldsymbol{u}_n^{\mathrm{T}})^{\mathrm{T}}$, which are defined as

$$\boldsymbol{u}_{i0} = \frac{\int \boldsymbol{u}_i L_i(\boldsymbol{u}_i; \boldsymbol{\vartheta}, \boldsymbol{\theta}_{\mathcal{R}}) d\boldsymbol{u}_i}{\int L_i(\boldsymbol{u}_i; \boldsymbol{\vartheta}, \boldsymbol{\theta}_{\mathcal{R}}) d\boldsymbol{u}_i} \text{ and } \boldsymbol{v}_{i0} = \frac{\int (\boldsymbol{u}_i - \boldsymbol{u}_{i0})(\boldsymbol{u}_i - \boldsymbol{u}_{i0})^{\mathrm{T}} L_i(\boldsymbol{u}_i; \boldsymbol{\vartheta}, \boldsymbol{\theta}_{\mathcal{R}}) d\boldsymbol{u}_i}{\int L_i(\boldsymbol{u}_i; \boldsymbol{\vartheta}, \boldsymbol{\theta}_{\mathcal{R}}) d\boldsymbol{u}_i},$$

respectively, where

$$
\begin{aligned}
L_i(\boldsymbol{u}_i;\boldsymbol{\vartheta},\boldsymbol{\theta}_{\mathcal{R}}) \;=\; & \prod_{j=1}^{n_i} p(T_{ij},\delta_{ij}|b_{ij},b_i;\boldsymbol{\vartheta},\boldsymbol{\theta}_{\mathcal{R}})p(\boldsymbol{Y}_{ij}|b_{ij},b_i;\boldsymbol{\vartheta},\boldsymbol{\theta}_{\mathcal{R}})p(b_{ij},b_i;\boldsymbol{\theta}_{\mathcal{R}}) \\
& \times p(\boldsymbol{\varrho}_{\alpha};\sigma^2_{\alpha})p(\boldsymbol{\varrho}_{\hbar};\sigma^2_{\hbar})\prod_{\omega=1}^{p} p(\boldsymbol{\varphi}_{X\omega};\sigma^2_{\varphi_{x_\omega}})p(\boldsymbol{\varrho}_{X_\omega};\sigma^2_{\varrho_{x_\omega}}) \\
& \times \prod_{\nu=1}^{q} p(\boldsymbol{\varphi}_{Z_\nu};\sigma^2_{\varphi_{z_\nu}})p(\boldsymbol{\varrho}_{Z_\nu};\sigma^2_{\varrho_{z_\nu}}) \qquad (3.11)
\end{aligned}
$$

is the likelihood function for the ith facility. Second, we calculate the posterior mean and variance of the S-RE and baseline hazard function RE, $\boldsymbol{\vartheta} = (\boldsymbol{\varrho}_{\alpha}^{\mathrm{T}},\boldsymbol{\varrho}_{\hbar}^{\mathrm{T}},\boldsymbol{\varrho}_{X}^{\mathrm{T}},\boldsymbol{\varrho}_{Z}^{\mathrm{T}},\boldsymbol{\varphi}_{X}^{\mathrm{T}},\boldsymbol{\varphi}_{Z}^{\mathrm{T}})^{\mathrm{T}}$, using the following formulas

$$
\boldsymbol{\vartheta}_0 = \frac{\int \boldsymbol{\vartheta} L(\boldsymbol{\vartheta};\boldsymbol{u},\boldsymbol{\theta}_{\mathcal{R}})d\boldsymbol{\vartheta}}{\int L(\boldsymbol{\vartheta};\boldsymbol{u},\boldsymbol{\theta}_{\mathcal{R}})d\boldsymbol{\vartheta}} \quad \text{and} \quad \mathcal{V}_{\boldsymbol{\vartheta}0} = \frac{\int (\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0)(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0)^{\mathrm{T}} L(\boldsymbol{\vartheta};\boldsymbol{u},\boldsymbol{\theta}_{\mathcal{R}})d\boldsymbol{\vartheta}}{\int L(\boldsymbol{\vartheta};\boldsymbol{u},\boldsymbol{\theta}_{\mathcal{R}})d\boldsymbol{\vartheta}},
$$

where

$$
\begin{aligned}
L(\boldsymbol{\vartheta};\boldsymbol{u},\boldsymbol{\theta}_{\mathcal{R}}) \;=\; & \prod_{i=1}^{n}\prod_{j=1}^{n_i} p(T_{ij},\delta_{ij}|b_{ij},b_i;\boldsymbol{\vartheta},\boldsymbol{\theta}_{\mathcal{R}})p(\boldsymbol{Y}_{ij}|b_{ij},b_i;\boldsymbol{\vartheta},\boldsymbol{\theta}_{\mathcal{R}})p(b_{ij},b_i;\boldsymbol{\theta}_{\mathcal{R}}) \\
& \times p(\boldsymbol{\varrho}_{\alpha};\sigma^2_{\alpha})p(\boldsymbol{\varrho}_{\hbar};\sigma^2_{\hbar})\prod_{\omega=1}^{p} p(\boldsymbol{\varphi}_{X\omega};\sigma^2_{\varphi_{x_\omega}})p(\boldsymbol{\varrho}_{X_\omega};\sigma^2_{\varrho_{x_\omega}}) \\
& \times \prod_{\nu=1}^{q} p(\boldsymbol{\varphi}_{Z_\nu};\sigma^2_{\varphi_{z_\nu}})p(\boldsymbol{\varrho}_{Z_\nu};\sigma^2_{\varrho_{z_\nu}}) \qquad (3.12)
\end{aligned}
$$

is the likelihood function. Lastly, similar to Section 3.2.1, since the expected log-likelihood function does not have a closed form, it will be approximated around the posterior means of $(\boldsymbol{u}^{\mathrm{T}},\boldsymbol{\vartheta}^{\mathrm{T}})^{\mathrm{T}}$ with a second-order Taylor's expansion as explained in Section 3.2.1.

Let us start with the first step where the posterior mean and variance of the random effects are targeted. We compute these values with a fully exponential Laplace approximation as described in Section 3.2.1. Using the likelihood function in (3.11), the log-likelihood function is expressed as

$$
\begin{aligned}
\ell_i(\boldsymbol{u}_i; \boldsymbol{\vartheta}, \boldsymbol{\theta}_{\mathscr{R}}) \;=\; & \sum_{j=1}^{n_i} \Bigg[\delta_{ij}\{\log \mathscr{h}_0(T_{ij}) + \boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_X(T_{ij}) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}} \boldsymbol{\gamma}_Z(T_{ij}) + \alpha(T_{ij}) \mathscr{m}_{ij}(T_{ij})\} \\
& - \int_0^{T_{ij}} \mathscr{h}_0(s) \exp\{\boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_X(s) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}} \boldsymbol{\gamma}_Z(s) + \alpha(s) \mathscr{m}_{ij}(s)\} ds \\
& + \sum_{k=1}^{n_{ij}} \left\{ g(\mathscr{m}_{ijk}) Y_{ijk} + \log(\mathscr{q}_{ijk}) \right\} - \frac{\mathscr{b}_{ij}^2}{2\sigma_S^2} - \frac{1}{2}\log(2\pi\sigma_S^2) \Bigg] - \frac{\mathscr{b}_i^2}{2\sigma_F^2} - \frac{1}{2}\log(2\pi\sigma_F^2) \\
& - \frac{1}{2} \left(\frac{1}{\sigma_\alpha^2} \boldsymbol{\varrho}_\alpha^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varrho}_\alpha + \log\left|2\pi\sigma_\alpha^2 \boldsymbol{I}_{n_\kappa}\right| + \frac{1}{\sigma_{\mathscr{h}}^2} \boldsymbol{\varrho}_{\mathscr{h}}^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varrho}_{\mathscr{h}} + \log\left|2\pi\sigma_{\mathscr{h}}^2 \boldsymbol{I}_{n_\kappa}\right| \right) \\
& - \frac{1}{2} \sum_{\omega=1}^{p} \left(\frac{1}{\sigma_{\varphi_{x_\omega}}^2} \boldsymbol{\varphi}_{X_\omega}^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varphi}_{X_\omega} + \log\left|2\pi\sigma_{\varphi_{x_\omega}}^2 \boldsymbol{I}_{n_\kappa}\right| + \frac{1}{\sigma_{\varrho_{x_\omega}}^2} \boldsymbol{\varrho}_{X_\omega}^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varrho}_{X_\omega} + \log\left|2\pi\sigma_{\varrho_{x_\omega}}^2 \boldsymbol{I}_{n_\kappa}\right| \right) \\
& - \frac{1}{2} \sum_{\nu=1}^{p} \left(\frac{1}{\sigma_{\varphi_{z_\nu}}^2} \boldsymbol{\varphi}_{Z_\nu}^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varphi}_{Z_\nu} + \log\left|2\pi\sigma_{\varphi_{z_\nu}}^2 \boldsymbol{I}_{n_\kappa}\right| + \frac{1}{\sigma_{\varrho_{z_\nu}}^2} \boldsymbol{\varrho}_{Z_\nu}^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varrho}_{Z_\nu} + \log\left|2\pi\sigma_{\varrho_{z_\nu}}^2 \boldsymbol{I}_{n_\kappa}\right| \right).
\end{aligned}
$$

Similar to the random effects $\boldsymbol{u}$, the S-RE are also computed using a fully exponential Laplace approximation. In the mode step, the modes for $\boldsymbol{\vartheta}$ are computed by finding the values $\hat{\boldsymbol{\vartheta}} = \hat{\boldsymbol{\vartheta}}^{(\boldsymbol{c})}|_{\boldsymbol{c}=\boldsymbol{0}}$ where $\hat{\boldsymbol{\vartheta}}^{(\boldsymbol{c})}|_{\boldsymbol{c}=\boldsymbol{0}} = \operatorname{argmax}_{\boldsymbol{\vartheta}}\{\ell(\boldsymbol{\vartheta}; \boldsymbol{u}_0^*, \boldsymbol{\theta}_{\mathscr{R}}) + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\vartheta}\}$, with $\boldsymbol{c} = \{c_1, \ldots, c_{(2p+2q+2)n_\kappa}\}^{\mathrm{T}}$ is a constant vector, and

$$
\begin{aligned}
\ell(\boldsymbol{\vartheta}; \boldsymbol{u}_0^*, \boldsymbol{\theta}_{\mathscr{R}}) \;=\; & \sum_{i=1}^{n} \Bigg(\sum_{j=1}^{n_i} \Bigg[\delta_{ij}\{\log \mathscr{h}_0(T_{ij}) + \boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_X(T_{ij}) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}} \boldsymbol{\gamma}_Z(T_{ij}) + \alpha(T_{ij}) \mathscr{m}_{ij}^*(T_{ij})\} \\
& - \int_0^{T_{ij}} \mathscr{h}_0(s) \exp\{\boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_X(s) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}} \boldsymbol{\gamma}_Z(s) + \alpha(s) \mathscr{m}_{ij}^*(s)\} ds \\
& + \sum_{k=1}^{n_{ij}} \left\{ g(\mathscr{m}_{ijk}^*) Y_{ijk} + \log(\mathscr{q}_{ijk}^*) \right\} - \frac{(\mathscr{b}_{ij}^*)^*}{2\sigma_S^2} - \frac{1}{2}\log(2\pi\sigma_S^2) \Bigg] - \frac{(\mathscr{b}_i^*)^2}{2\sigma_F^2} - \frac{1}{2}\log(2\pi\sigma_F^2) \Bigg) \\
& - \frac{1}{2} \left(\frac{1}{\sigma_\alpha^2} \boldsymbol{\varrho}_\alpha^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varrho}_\alpha + \log\left|2\pi\sigma_\alpha^2 \boldsymbol{I}_{n_\kappa}\right| + \frac{1}{\sigma_{\mathscr{h}}^2} \boldsymbol{\varrho}_{\mathscr{h}}^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varrho}_{\mathscr{h}} + \log\left|2\pi\sigma_{\mathscr{h}}^2 \boldsymbol{I}_{n_\kappa}\right| \right) \\
& - \frac{1}{2} \sum_{\omega=1}^{p} \left(\frac{1}{\sigma_{\varphi_{x_\omega}}^2} \boldsymbol{\varphi}_{X_\omega}^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varphi}_{X_\omega} + \log\left|2\pi\sigma_{\varphi_{x_\omega}}^2 \boldsymbol{I}_{n_\kappa}\right| + \frac{1}{\sigma_{\varrho_{x_\omega}}^2} \boldsymbol{\varrho}_{X_\omega}^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varrho}_{X_\omega} + \log\left|2\pi\sigma_{\varrho_{x_\omega}}^2 \boldsymbol{I}_{n_\kappa}\right| \right) \\
& - \frac{1}{2} \sum_{\nu=1}^{p} \left(\frac{1}{\sigma_{\varphi_{z_\nu}}^2} \boldsymbol{\varphi}_{Z_\nu}^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varphi}_{Z_\nu} + \log\left|2\pi\sigma_{\varphi_{z_\nu}}^2 \boldsymbol{I}_{n_\kappa}\right| + \frac{1}{\sigma_{\varrho_{z_\nu}}^2} \boldsymbol{\varrho}_{Z_\nu}^{\mathrm{T}} \boldsymbol{I}_{n_\kappa} \boldsymbol{\varrho}_{Z_\nu} + \log\left|2\pi\sigma_{\varrho_{z_\nu}}^2 \boldsymbol{I}_{n_\kappa}\right| \right),
\end{aligned}
$$

$m^*_{ij}(t) = g^{-1}\left\{\boldsymbol{X}^{\mathrm{T}}_{ij}\boldsymbol{\beta}_X(t) + \boldsymbol{Z}^{\mathrm{T}}_{i(j)}\boldsymbol{\beta}_Z(t) + b^*_{ij0} + b^*_{i0}\right\}$, $m^*_{ijk} = m^*_{ij}(t_{ijk})$, $q^*_{ijk} = 1 - m^*_{ijk}$, is the log-likelihood function. The values $\hat{\boldsymbol{\vartheta}}$ can be computed from a safeguarded Newton-Raphson algorithm with $\hat{\boldsymbol{\vartheta}}^{it}$ updated as

$$\hat{\boldsymbol{\vartheta}}^{it+1} = \hat{\boldsymbol{\vartheta}}^{it} - s\{\mathcal{H}_{\boldsymbol{\vartheta}}(\hat{\boldsymbol{\vartheta}}^{it})\}^{-1}\mathcal{G}_{\boldsymbol{\vartheta}}(\hat{\boldsymbol{\vartheta}}^{it}),$$

where 'it' denotes the current iteration, $\mathcal{H}_{\boldsymbol{\vartheta}}(\hat{\boldsymbol{\vartheta}}^{it}) = \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}} = \boldsymbol{\Sigma}^{(c)}_{\boldsymbol{\vartheta}}|_{(\boldsymbol{c},\boldsymbol{\vartheta})=(\boldsymbol{0},\hat{\boldsymbol{\vartheta}}^{it})}$ with $\boldsymbol{\Sigma}^{(c)}_{\boldsymbol{\vartheta}} = -\partial^2\left\{\ell(\boldsymbol{\vartheta};\boldsymbol{u}^*_0,\boldsymbol{\theta}_{\mathcal{R}}) + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\vartheta}\right\}/\partial\boldsymbol{\vartheta}^{\mathrm{T}}\partial\boldsymbol{\vartheta}$, and $\mathcal{G}_{\boldsymbol{\vartheta}}(\hat{\boldsymbol{\vartheta}}^{it}) = -\partial\ell(\boldsymbol{\vartheta};\boldsymbol{u}^*_0,\boldsymbol{\theta}_{\mathcal{R}})/\partial\boldsymbol{\vartheta}^{\mathrm{T}}|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}^{it}}$, and s denotes the step size.

In the correction step, the posterior mean and variance for the S-RE are computed from the modes. Differentiating the cumulant-generating function, $\log E\{\exp(\boldsymbol{c}^T\boldsymbol{\vartheta})\}$, and evaluating at $\boldsymbol{c} = \boldsymbol{0}$ leads to the following the posterior mean and variance

$$\boldsymbol{\vartheta}_0 = \hat{\boldsymbol{\vartheta}} - \frac{1}{2}\mathrm{tr}(\mathcal{V}_{\boldsymbol{\vartheta}}) \text{ and } \mathcal{V}_{\vartheta 0} = \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\vartheta}} - \frac{1}{2}\mathrm{tr}\left\{-\mathcal{V}_{\boldsymbol{\vartheta}}\mathcal{V}^{\mathrm{T}}_{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\vartheta}}\frac{\partial^2\boldsymbol{\Sigma}^{(c)}_{\boldsymbol{\vartheta}}}{\partial\boldsymbol{c}^{\mathrm{T}}\partial\boldsymbol{c}}\bigg|_{(\boldsymbol{c},\boldsymbol{\vartheta})=(\boldsymbol{0},\hat{\boldsymbol{\vartheta}})}\right\}, \tag{3.13}$$

where $\mathcal{V}_{\boldsymbol{\vartheta}} = \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\vartheta}}\{\partial\boldsymbol{\Sigma}^{(c)}_{\boldsymbol{\vartheta}}/\partial\boldsymbol{c}^{\mathrm{T}}\}|_{(\boldsymbol{c},\boldsymbol{\vartheta})=(\boldsymbol{0},\hat{\boldsymbol{\vartheta}})}$, $\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}} = \boldsymbol{\Sigma}^{(c)}_{\boldsymbol{\vartheta}}|_{\boldsymbol{c}=\boldsymbol{0}}$, with $\hat{\boldsymbol{\vartheta}}$ and $\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\vartheta}}$ as the modes and the inverse of $\boldsymbol{\Sigma}^{it}_{\boldsymbol{\vartheta}}$, respectively, obtained from the last iteration of the Newton-Raphson algorithm from the mode step. More information on the fully exponential Laplace approximation is provided in the Appendix A.2.

The expectation of the complete joint log-likelihood function in the E-step is approximated using the estimated posterior mean and variance of $\boldsymbol{\theta}_{RE} = (\boldsymbol{u}^{\mathrm{T}}, \boldsymbol{\vartheta}^{\mathrm{T}})^{\mathrm{T}}$. However, the closed-expression of the expectation of the log-likelihood function is intractable. Therefore, we employ a second degree Taylor's expansion around the estimated posterior mean of $\boldsymbol{\theta}_{RE}$, denoted as $\boldsymbol{\theta}^*_{0RE}$,

$$
\begin{aligned}
\ell(\boldsymbol{\theta}_{RE}, \boldsymbol{\theta}_{\mathcal{R}}) &\approx \ell(\boldsymbol{\theta}^*_{0RE}, \boldsymbol{\theta}^*_{\mathcal{R}}) + \ell'(\boldsymbol{\theta}^*_{0RE}, \boldsymbol{\theta}^*_{\mathcal{R}}) E(\boldsymbol{\theta}_{RE} - \boldsymbol{\theta}^*_{0RE}) - \frac{1}{2} E(\boldsymbol{\theta}_{RE} - \boldsymbol{\theta}^*_{0RE})^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{\theta}_{RE}} E(\boldsymbol{\theta}_{RE} - \boldsymbol{\theta}^*_{0RE}) \\
\ell^*(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*_{\mathcal{R}}) &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n_i} \left[\delta_{ij} \{ \log h^*_0(T_{ij}) + \boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\gamma}^*_X(T_{ij}) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}} \boldsymbol{\gamma}^*_Z(T_{ij}) + \alpha^*(T_{ij}) m^*_{ij}(T_{ij}) \} \right.\right. \\
&\quad - \int_0^{T_{ij}} h^*_0(s) \exp\{ \boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\gamma}^*_X(s) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}} \boldsymbol{\gamma}^*_Z(s) + \alpha^*(s) m^*_{ij}(s) \} ds \\
&\quad \left. + \sum_{k=1}^{n_{ij}} \left\{ g(m^*_{ijk}) Y_{ijk} + \log(q^*_{ijk}) \right\} - \frac{(b^*_{0ij})^2 + v^*_{b,ij0}}{2\sigma^{2*}_S} - \frac{1}{2}\log(2\pi\sigma^{2*}_S) + \mathcal{R}^*_{ij} \Delta^{(2)*}_{ij} \right] \\
&\quad \left. - \frac{(b^*_i)^2 + v^*_{b,i0}}{2\sigma^{2*}_F} - \frac{1}{2}\log(2\pi\sigma^{2*}_F) \right) + p_\ell(\boldsymbol{\varrho}^*_\alpha; \sigma^{2*}_\alpha) + p_\ell(\boldsymbol{\varrho}^*_h; \sigma^{2*}_h) \\
&\quad + \sum_{\omega=1}^{p} \left\{ p_\ell(\boldsymbol{\varphi}^*_{X_\omega}; \sigma^{2*}_{\varphi_{x_\omega}}) + p_\ell(\boldsymbol{\varrho}^*_{X_\omega}; \sigma^{2*}_{\varrho_{x_\omega}}) \right\} + \sum_{\nu=1}^{q} \left\{ p_\ell(\boldsymbol{\varphi}^*_{Z_\nu}; \sigma^{2*}_{\varphi_{z_\nu}}) + p_\ell(\boldsymbol{\varrho}^*_{Z_\nu}; \sigma^{2*}_{\varrho_{z_\nu}}) \right\} \\
&\quad + \sum_{\varkappa=1}^{n_\kappa} \left[\frac{\varsigma^*_{h\varkappa}}{2} \Lambda^*(\varrho_{h\varkappa}) + \frac{\varsigma^*_{\alpha\varkappa}}{2} \Lambda^*(\varrho_{h\varkappa}) + \sum_{\omega=1}^{p} \left\{ \frac{\varsigma^*_{\varphi_{x_\omega}\varkappa}}{2} \Lambda^*(\varphi_{X_\omega \varkappa}) + \frac{\varsigma^*_{\varrho_{x_\omega}\varkappa}}{2} \Lambda^*(\varrho_{X_\omega \varkappa}) \right\} \right. \\
&\quad \left. + \sum_{\nu=1}^{q} \left\{ \frac{\varsigma^*_{\varphi_{z_\nu}\varkappa}}{2} \Lambda^*(\varphi_{Z_\nu \varkappa}) + \frac{\varsigma^*_{\varrho_{z_\nu}\varkappa}}{2} \Lambda^*(\varrho_{Z_\nu \varkappa}) \right\} \right], \qquad (3.14)
\end{aligned}
$$

where $m^*_{ij}(t) = g^{-1}\left\{ \boldsymbol{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta}^*_X(t) + \boldsymbol{Z}_{i(j)}^{\mathrm{T}} \boldsymbol{\beta}^*_Z(t) + b^*_{ij0} + b^*_{i0} \right\}$, $m^*_{ijk} = m^*_{ij}(t_{ijk})$, $q^*_{ijk} = 1 - m^*_{ijk}$, $E(\boldsymbol{\theta}_{RE} - \boldsymbol{\theta}^*_{0RE}) = 0$, $\Delta^{(2)*}_{ij}$ is $\Delta^{(2)}_{ij}$ evaluated at $\boldsymbol{\theta}^*_{RE}$ and $\boldsymbol{\theta}^*_{\mathcal{R}}$ (defined in the Appendix A.1), $\boldsymbol{\Sigma}^*_{\boldsymbol{\theta}_{RE}} = -\partial^2\ell/\partial\boldsymbol{\theta}^{\mathrm{T}}_{RE}\partial\boldsymbol{\theta}_{RE}|_{(\boldsymbol{\theta}_{RE})=(\boldsymbol{\theta}^*_{RE0})}$, and $p_\ell(\cdot;\cdot)$ is the log of the multivariate normal density function. Additionally, $\Lambda^*(\cdot)$ indicates the second derivative of the log-likelihood function with respect to the S-RE's random effects (defined in Appendix A.2), evaluated at current estimates of $\boldsymbol{\theta}_{RE}$ and $\boldsymbol{\theta}_{\mathcal{R}}$, and $\varsigma^*_{\alpha\varkappa}$, $\varsigma^*_{h\varkappa}$, $\varsigma^*_{\varphi_{x_\omega}\varkappa}$, $\varsigma^*_{\varphi_{z_\nu}\varkappa}$, $\varsigma^*_{\varrho_{x_\omega}\varkappa}$, and $\varsigma^*_{\varrho_{z_\nu}\varkappa}$ indicate the posterior variance of each S-RE. Moreover, $\boldsymbol{u}^*_{i0}$ denotes the estimated posterior mean of $\boldsymbol{u}_{i0}$ obtained in the E-step, $\mathcal{R}^*_{ij} = \frac{v^*_{b,ij0} + 2\eta^*_{ij0} + v^*_{b,i0}}{2}$ with $v^*_{b,ij0}$, $v^*_{b,i0}$, and η^*_{ij0} represents the posterior variance for the subject- and facility-level random effects, and posterior covariance between the subject- and facility-level random effects, respectively.

M-Step

The subject- and facility-level variances are computed as described in Section 3.2.1. The S-RE variances are obtained by setting the score functions of the incomplete log-likelihood function, provided in Appendix A.2.3, to zero. The closed-form solutions for σ_{α}^2, $\sigma_{\hbar}^2$, $\sigma_{\varphi_{x_\omega}}^2$, $\sigma_{\varphi_{z_\nu}}^2$, $\sigma_{\varrho_{x_\omega}}^2$ and $\sigma_{\varrho_{z_\nu}}^2$ are expressed as

$$\sigma_{\alpha}^{2*} = n_\kappa^{-1} \sum_{\varkappa=1}^{n_\kappa} \left\{ (\varrho_{\alpha\varkappa}^*)^2 + \varsigma_{\alpha\varkappa}^* \right\}, \quad \sigma_{\hbar}^{2*} = n_\kappa^{-1} \sum_{\varkappa=1}^{n_\kappa} \left\{ (\varrho_{\hbar\varkappa}^*)^2 + \varsigma_{\hbar\varkappa}^* \right\},$$
$$\sigma_{\varphi_{x_\omega}}^{2*} = n_\kappa^{-1} \sum_{\varkappa=1}^{n_\kappa} \left\{ (\varphi_{X_\omega\varkappa}^*)^2 + \varsigma_{\varphi_{x_\omega}\varkappa}^* \right\}, \quad \sigma_{\varrho_{x_\omega}}^{2*} = n_\kappa^{-1} \sum_{\varkappa=1}^{n_\kappa} \left\{ (\varrho_{X_\omega\varkappa}^*)^2 + \varsigma_{\varrho_{x_\omega}\varkappa}^* \right\},$$
$$\sigma_{\varphi_{z_\nu}}^{2*} = n_\kappa^{-1} \sum_{\varkappa=1}^{n_\kappa} \left\{ (\varphi_{Z_\nu\varkappa}^*)^2 + \varsigma_{\varphi_{z_\nu}\varkappa}^* \right\}, \quad \sigma_{\varrho_{z_\nu}}^{2*} = n_\kappa^{-1} \sum_{\varkappa=1}^{n_\kappa} \left\{ (\varrho_{Z_\nu\varkappa}^*)^2 + \varsigma_{\varrho_{z_\nu}\varkappa}^* \right\},$$

respectively. The parameters $\boldsymbol{\theta}_{\mathcal{R}}^{\backslash\sigma} = (\boldsymbol{\theta}_{L\mathcal{R}}^{\mathrm{T}}, \boldsymbol{\theta}_{S\mathcal{R}}^{\mathrm{T}})^{\mathrm{T}}$ are estimated by maximizing the approximated expected log-likelihood function defined in (3.14). Since the parameters in $\boldsymbol{\theta}_{\mathcal{R}}^{\backslash\sigma}$ do not have closed-form solutions, the Newton-Raphson algorithm is employed with the update as

$$\boldsymbol{\theta}_{\mathcal{R}}^{\backslash\sigma(it+1)} = \boldsymbol{\theta}_{\mathcal{R}}^{\backslash\sigma(it)} - \mathcal{H}_{\boldsymbol{\theta}_{\mathcal{R}}^{\backslash\sigma}}^{(it)} \mathcal{G}_{\boldsymbol{\theta}_{\mathcal{R}}^{\backslash\sigma}}^{(it)},$$

where '*it*' is the current iteration, $\mathcal{G}_{\boldsymbol{\theta}_{\mathcal{R}}^{\backslash\sigma}}^{(it)}$ and $\mathcal{H}_{\boldsymbol{\theta}_{\mathcal{R}}^{\backslash\sigma}}^{(it)}$ are the gradient and Hessian of the log-likelihood function (3.14) with respect to $\boldsymbol{\theta}_{\mathcal{R}}^{\backslash\sigma}$, respectively, evaluated at the current estimates $\boldsymbol{\theta}_{\mathcal{R}}^{\backslash\sigma(it)}$.

3.3 Inference

For inference under both the P-spline and random-coefficient spline approaches, we suggest using bootstrap-based standard errors as empirical results show that likelihood-based standard errors are biased in estimation of the true standard errors (Hsieh et al., 2006; Kürüm et al., 2021). Let $\hat{f}(t)$ denote a single estimated time-varying coefficient function from the longitudinal submodel ($\hat{\beta}_*(t)$) or survival ($\alpha(t)$, $\hbar(t)$, $\gamma_*(t)$) submodel, the following setup is used to obtain the bootstrap-based standard errors of $\hat{f}(t)$ under both our estimation procedures:

1. Draw a bootstrap sample by sampling n facilities from the data with replacement.

2. Apply the EM algorithm in Section 3.2.1 to the bootstrap sample and obtain parameter estimates, $\hat{\boldsymbol{\theta}}_{\mathscr{P}}$, using MTJM (P-splines approach). Similarly, perform the EM algorithm presented in Section 3.2.2 to the bootstrap sample and obtain parameter estimates, $\hat{\boldsymbol{\theta}}_{\mathscr{R}}$, and predicted posterior means of $\boldsymbol{\theta}^*_{0RE}$ using MTJMRE (random-coefficient spline technique).

3. Repeat Steps 1 and 2 for B bootstrap samples.

4. Compute $\hat{f}_b(t)$, the estimated function using the fitted model from sample b, $b = 1, \ldots, B$, and obtain the standard error of $\hat{f}(t)$:

$$SE\{\hat{f}(t)\} = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\{\hat{f}_b(t) - \bar{f}(t)\}^2},$$

with $\bar{f}(t) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}_b(t)$.

Chapter 4

Simulation Study

This chapter provides the results of the simulation studies conducted for the MTJM (P-splines Models; 4.1) and MTJMRE (Random-Coefficient Spline Models; 4.2). Each simulation study utilized the `Armadillo` library (Sanderson and Curtin, 2016, 2018) via the `RcppArmadillo` package (Eddelbuettel and Sanderson, 2014) and analyzed in R (version 3.6.3) (R Core Team, 2022).

4.1 P-Splines Models

We generated 50 Monte Carlo data sets to assess the efficacy of MTJM, as described in 3.2.1, in estimating the varying-coefficient functions as well as the bootstrap-based standard errors. Each data set contains $n = 200$ facilities, with each facility containing $n_i = 20$ patients, each patient having 20 observations, measured at equally-spaced time-points in the interval [0,1], before censoring from the survival submodel.

Each subject had two covariates, the subject-level X_{ij1} and facility-level $Z_{i(j)1}$, both generated from a normal distribution with a mean 0 and variance 1. The subject- and facility-level random effects (REs) were simulated independently from a normal distribution with mean 0 and variances 1.4 and 0.3, respectively.

The true time-to-event T^*_{ij} was generated using the inverse probability integral transformation and Weibull baseline hazard function, $\hbar_0(t) = 1.1t^{0.1}$, as described in Bender et al. (2005). The censoring time was $C_{ij} = \min(C^*_{ij}, 1)$, where C^*_{ij} was generated from an exponential distribution with a mean of 2.5. The observed time-to-event and event indicator were obtained as $T_{ij} = \min(T^*_{ij}, C_{ij})$ and $\delta_{ij} = I(T^*_{ij} < C_{ij})$, respectively. The overall censoring rate was approximately 71%. The binary longitudinal outcome at time t, $Y_{ij}(t)$, was simulated using an underlying normal latent variable Y^*_{ij}, where $Y_{ij}(t) = I(Y^*_{ij} > 0)$, and the longitudinal submodel served as the mean of the latent variable. The overall hospitalization rate was approximately 27%. The time-varying coefficient functions for the longitudinal submodel were $\beta_0(t) = cos(2\pi t) - 1.5$, $\beta_{X_1}(t) = \sin(\pi t)$, and $\beta_{Z_1}(t) = \sin(2\pi t)$, and the time-varying coefficient functions for the survival submodel were $\alpha(t) = \sin(2\pi t)$, $\gamma_{X_1}(t) = -\cos(2\pi t)/2$, and $\gamma_{Z_1}(t) = -\sin(2\pi t)/2$.

The time-varying coefficient and baseline hazard functions were both modeled with B-spline functions using cubic truncated power functions and 5 interior knots, and estimated using the procedure in Section 3.2.1. The bootstrap-based confidence intervals were obtained using the approach described in Section 3.3 based on $B = 48$ samples.

Results

Table 4.1 provides the bias, true standard deviation (SD), bootstrap-based standard error (Boot_{SE}), and standard deviation of the bootstrap-based standard errors ($\text{Boot}_{\text{SD}_{\text{SE}}}$) for subject- and facility-level RE variances. The estimated RE variances perform well with small bias; however, the bootstrap-based standard errors deviate from the true standard deviation, but within margin of error. Tables 4.2 and 4.3 provide the bias, true standard deviation (SD), bootstrap-based standard error (Boot_{SE}), and standard deviation of the bootstrap-based standard errors ($\text{Boot}_{\text{SD}_{\text{SE}}}$) for each time-varying coefficient functions in longitudinal and survival submodels, respectively, measured at time points 0.25, 0.5, and 0.75. The tables show a relatively small bias for all functions; furthermore, the bootstrap-based standard errors perform well in estimating the theoretical standard deviation.

Figures 4.1 and 4.2 provide the estimated time-varying functions (dashed line) and true functions (solid line) for the longitudinal and survival submodels, respectively. Both figures show the estimated function performs well in aligning with the true functions; furthermore, the bootstrap-based confidence intervals (dotted) capture the true function. We observe a boundary effect on the function $\alpha(t)$ as the confidence intervals begin to widen.

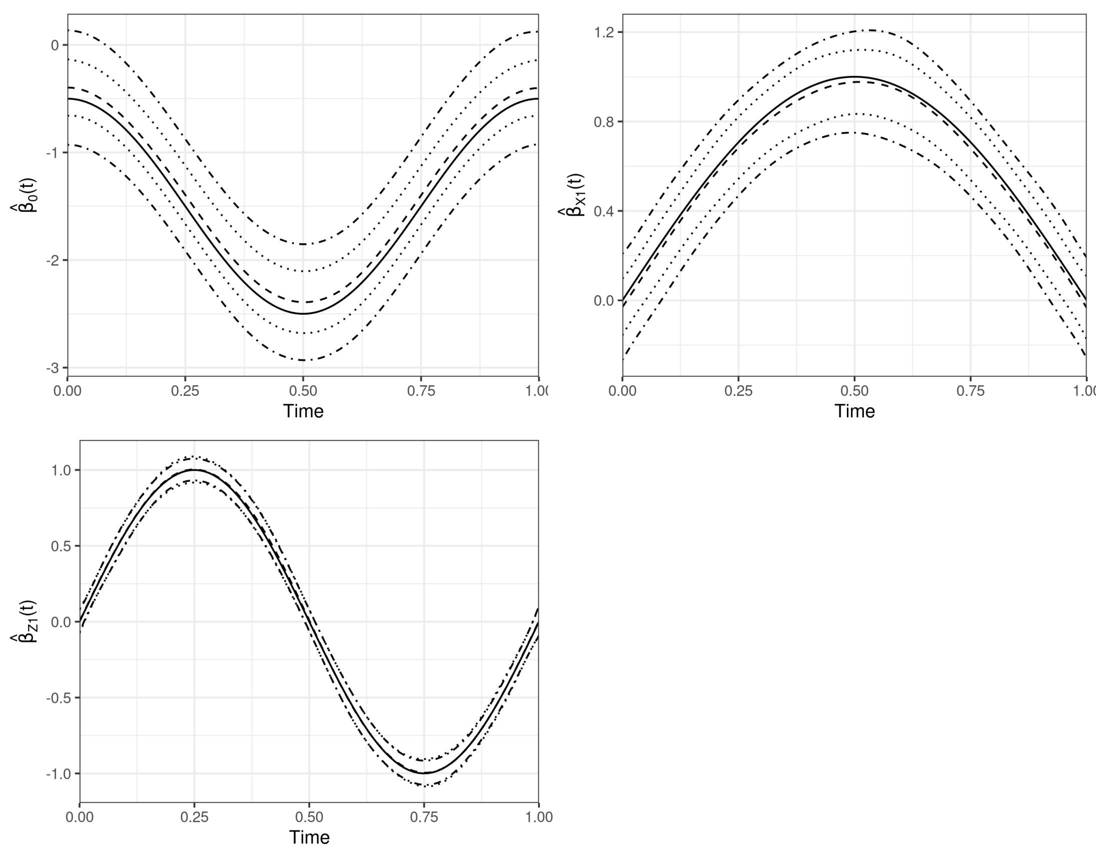

Figure 4.1: Estimated time-varying functions (dashed line) in the longitudinal submodel, based on (3.1), for $n = 200$ facilities overlaying the true functions (solid line) along with 95% bootstrap-based (dotted) and mean theoretical (dashed-dotted) confidence intervals.

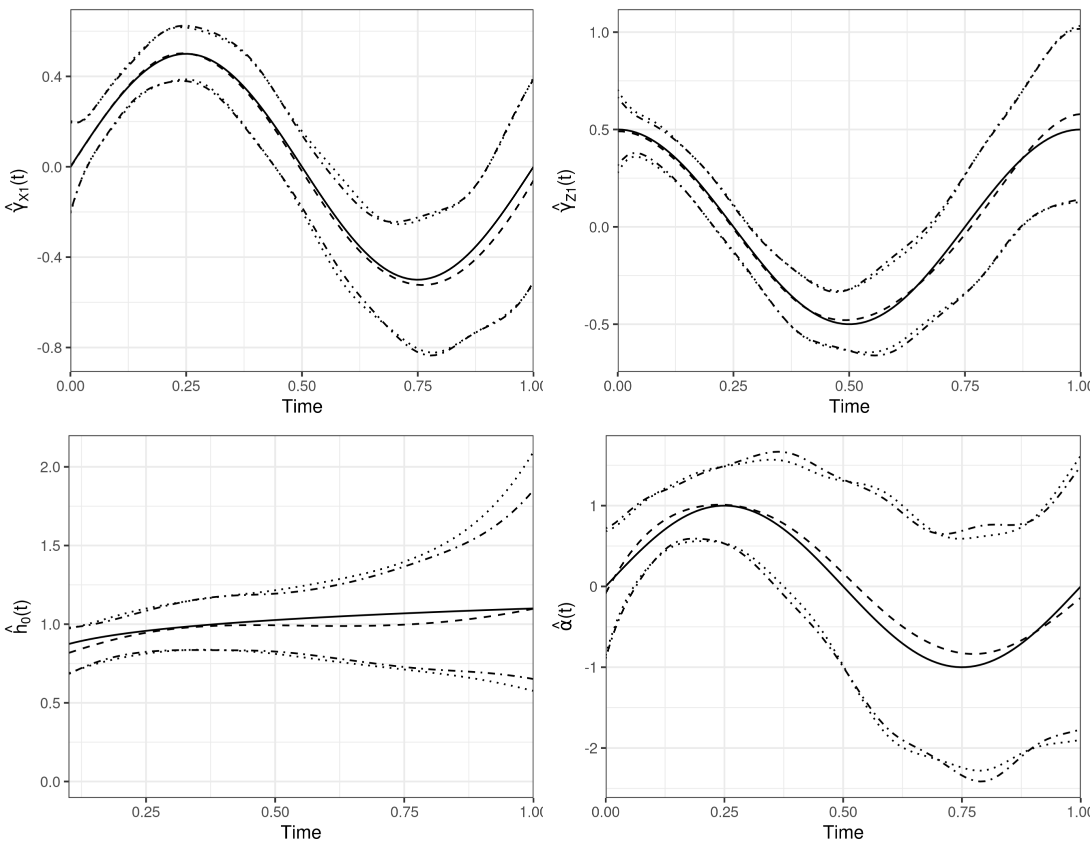

Figure 4.2: Estimated time-varying functions (dashed line) in the survival submodel, based on (3.2), for $n = 200$ facilities overlaying the true functions (solid line) along with 95% bootstrap-based (dotted) and mean theoretical (dashed-dotted) confidence intervals.

Parameter	True value	Bias	SD	Boot_{SE}	$\text{Boot}_{\text{SD}_{\text{SE}}}$
σ_S^2	1.4	0.02	0.06	0.15	0.11
σ_F^2	0.3	0.11	0.31	0.16	0.04

Table 4.1: The bias, true standard deviation, bootstrap-based standard error, and standard deviation of the bootstrap-based standard error are provided for the subject- (σ_S^2) and facility-level (σ_F^2) variances.

Function	Time	True value	Bias	SD	Boot_{SE}	$\text{Boot}_{\text{SD}_{\text{SE}}}$
	0.25	-1.50	0.11	0.27	0.14	0.04
$\beta_0(t)$	0.50	-2.50	0.11	0.27	0.15	0.05
	0.75	-1.50	0.10	0.28	0.14	0.04
	0.25	0.71	-0.02	0.11	0.07	0.02
$\beta_{X_1}(t)$	0.50	1.00	-0.02	0.12	0.07	0.02
	0.75	0.71	-0.03	0.11	0.07	0.02
	0.25	1.00	0.00	0.04	0.04	0.01
β_{Z_1}	0.50	0.00	0.01	0.04	0.04	0.00
	0.75	-1.00	0.00	0.04	0.05	0.01

Table 4.2: The bias, theoretical standard deviation, bootstrap-based standard error, and the standard deviation of the bootstrap-based standard error for the longitudinal submodel.

Function	Time	True value	Bias	SD	Boot_{SE}	$\text{Boot}_{\text{SD}_{\text{SE}}}$
	0.25	1.00	0.01	0.25	0.24	0.03
$\alpha(t)$	0.50	0.00	0.16	0.59	0.58	0.11
	0.75	-1.00	0.17	0.77	0.72	0.13
	0.25	0.96	-0.01	0.07	0.08	0.01
$\hbar_0(t)$	0.50	1.03	-0.03	0.09	0.10	0.01
	0.75	1.07	-0.07	0.16	0.17	0.02
	0.25	0.50	0.00	0.06	0.06	0.01
$\gamma_{X_1}(t)$	0.50	0.00	-0.02	0.08	0.09	0.02
	0.75	-0.50	-0.02	0.15	0.14	0.02
	0.25	0.00	-0.01	0.06	0.06	0.01
$\gamma_{Z_1}(t)$	0.50	-0.50	0.02	0.08	0.08	0.01
	0.75	0.00	-0.04	0.16	0.15	0.02

Table 4.3: The bias, theoretical standard deviation, bootstrap-based standard error, and the standard deviation of the bootstrap-based standard error for the survival submodel.

4.2 Random-Coefficient Spline Models

We conducted a simulation study to assess the efficacy of the proposed MTJMRE model in estimating the time-varying coefficient functions as well as the bootstrap-based standard errors. Performance of the model was assessed at $n = 250$ and $n = 500$ facilities, with each facility containing $n_i = 50$ patients. The maximum number of repeated measurements for each patient was 25 observations, measured at equally-spaced time-points in the interval [0,1], before censoring from the survival submodel. Reports for each case were based on 50 Monte Carlo data sets.

The subject-level covariates, $\boldsymbol{X}_{ij} = (X_{ij1}, X_{ij2})^{\mathrm{T}}$, were simulated from a normal distribution with means 0 and 1.5 and variances 1 and 0.5, respectively. The facility-level covariates, $\boldsymbol{Z}_{i(j)} = (Z_{i(j)1}, Z_{i(j)2})^{\mathrm{T}}$, were simulated from a normal distribution with means -0.3 and 1.5 and variances 1 and 0.5, respectively. The subject- and facility-level random effects (REs) were simulated independently from a normal distribution with mean 0 and variances 1.3 and 0.2, respectively.

For each subject, the true time-to-event T^*_{ij} was generated using the inverse probability integral transformation with the Weibull baseline hazard function as $\hbar_0(t) = 1.5t^{0.5}$ (Bender et al., 2005). The censoring time was $C_{ij} = \min(C^*_{ij}, 1)$, where C^*_{ij} was generated from an exponential distribution with a mean of 1.5. The observed time-to-event and event indicator were obtained as $T_{ij} = \min(T^*_{ij}, C_{ij})$ and $\delta_{ij} = I(T^*_{ij} < C_{ij})$, respectively. The binary longitudinal outcome at time t, $Y_{ij}(t)$, was simulated using an underlying normal latent variable Y^*_{ij}, where $Y_{ij}(t) = I(Y^*_{ij} > 0)$, and the mean of the latent variable was determined using the longitudinal submodel. The overall censoring and hospitalization rate were approximately 58% and 31%, respectively.

The subject-specific and facility-specific time-varying coefficient functions for the longitudinal submodel were $\boldsymbol{\beta}_X(t) = \{\beta_0(t), \beta_{X_1}(t), \beta_{X_2}(t)\}^{\mathrm{T}}$ with $\beta_0(t) = \cos(3\pi t/2) - 0.5$, $\beta_{X_1}(t) = \sin(2\pi t - 1/8)$, and $\beta_{X_2}(t) = -sin(2\pi t - 1/8)$; and $\boldsymbol{\beta}_Z(t) = \{\beta_{Z_1}(t), \beta_{Z_2}(t)\}^{\mathrm{T}}$ with $\beta_{Z_1}(t) = cos(\pi t - 1/2)$, and $\beta_{Z_2}(t) = -cos(\pi t - 1/2)$, respectively. The time-varying coefficient functions for the survival submodel were defined as $\alpha(t) = \sin(2\pi t)$, $\boldsymbol{\gamma}_X(t) = \{\gamma_{X_1}(t), \gamma_{X_2}(t)\}^{\mathrm{T}}$ specifying the subject-level functions with $\gamma_{X_1}(t) = -\cos(2\pi t)$ and $\gamma_{X_2}(t) = \cos(2\pi t)$, and $\boldsymbol{\gamma}_Z(t) = \{\gamma_{Z_1}(t), \gamma_{Z_2}(t)\}^{\mathrm{T}}$ were the facility-level functions with $\gamma_{Z_1}(t) = -\sin(3\pi t/4)$ and $\gamma_{Z_2}(t) = \sin(3\pi t/4)$.

The time-varying coefficient and baseline hazard functions were both modeled with the random-coefficient spline models and estimated using the procedure in Section 3.2.2. In terms of inference, the bootstrap-based confidence intervals were obtained using the approach described in Section 3.3 based on $B = 25$ samples.

Results

Figures 4.3 and 4.4 display the estimated time-varying functions (dashed) and true functions (solid) for the longitudinal and survival submodels based on $n = 250$ facilities, respectively; additionally, figures 4.5 and 4.6 display the estimated time-varying functions (dashed) and true functions (solid) for the longitudinal and survival submodels based on $n = 500$ facilities, respectively. For $n = 250$, there is a noticeable bias in estimating the time-varying functions, and the width of the bootstrap-based confidence intervals increase at the boundaries for the longitudinal submodel; however, as expected, the shape, bias and confidence intervals improve when the number of facilities increases to $n = 500$. In terms of the survival submodel, overall, our method performs well, where the estimates target the true functions, and the bootstrap-based confidence intervals cover the true function. For both scenarios ($n = 250$ and $n = 500$), we observe slight boundary effects in our estimation such that the estimated functions deviate from the truth at these points. In addition, for both longitudinal and survival submodels, bootstrap-based confidence intervals are close to the theoretical confidence intervals (dotted lines). Overall, the results show that our method leads to lower bias and narrower confidence intervals as the number of facilities increases.

Table 4.4 provides the estimated bias, theoretical standard deviation, bootstrap-based standard error, and standard deviation of the bootstrap-based standard error for the subject- and facility-level variances. The table shows that both variances are estimated well regardless of the number facilities; however, as the number of facilities increases, the bootstrap-based standard errors perform better in estimating the true standard deviation. Tables 4.5 and 4.6 provide the estimated bias, theoretical standard deviation, bootstrap-based standard error, and standard deviation of the bootstrap-based standard error for the time-varying functions in the longitudinal and survival submodel, respectively, at time points 0.25, 0.5, and 0.75. Overall, the tables indicate that the bias decreases as the number of facilities increases to $n = 500$. Furthermore, the bootstrap-based standard errors target the true standard deviation, that is, the SD is captured within ± 2 the standard deviation of the bootstrap-based standard errors for all functions and time points.

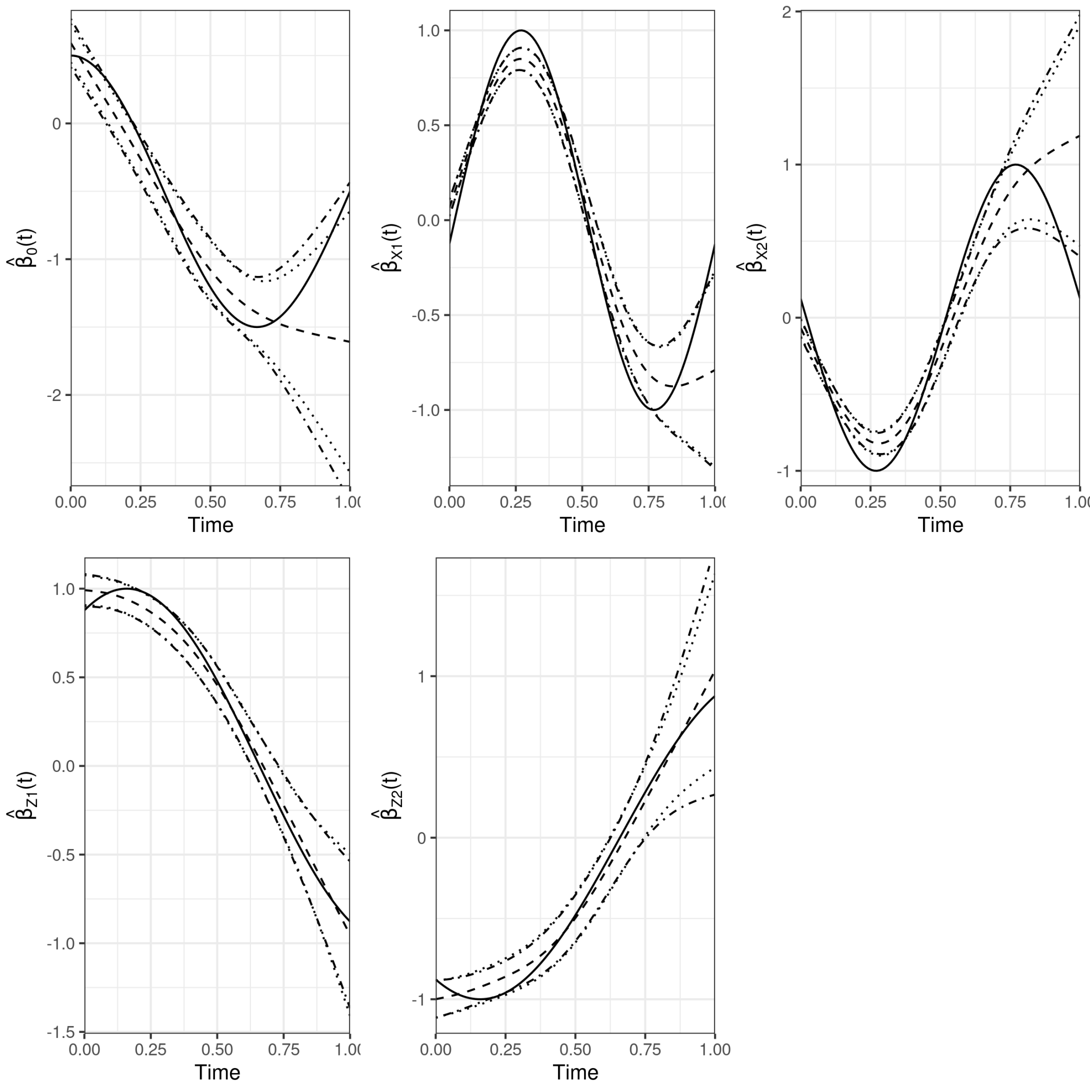

Figure 4.3: Estimated time-varying functions (dashed line) in the longitudinal submodel, based on (3.1), simulation runs for $n = 250$ facilities overlaying the true functions (solid line) along with 95% bootstrap-based (dotted) and mean theoretical (dashed-dotted) confidence intervals.

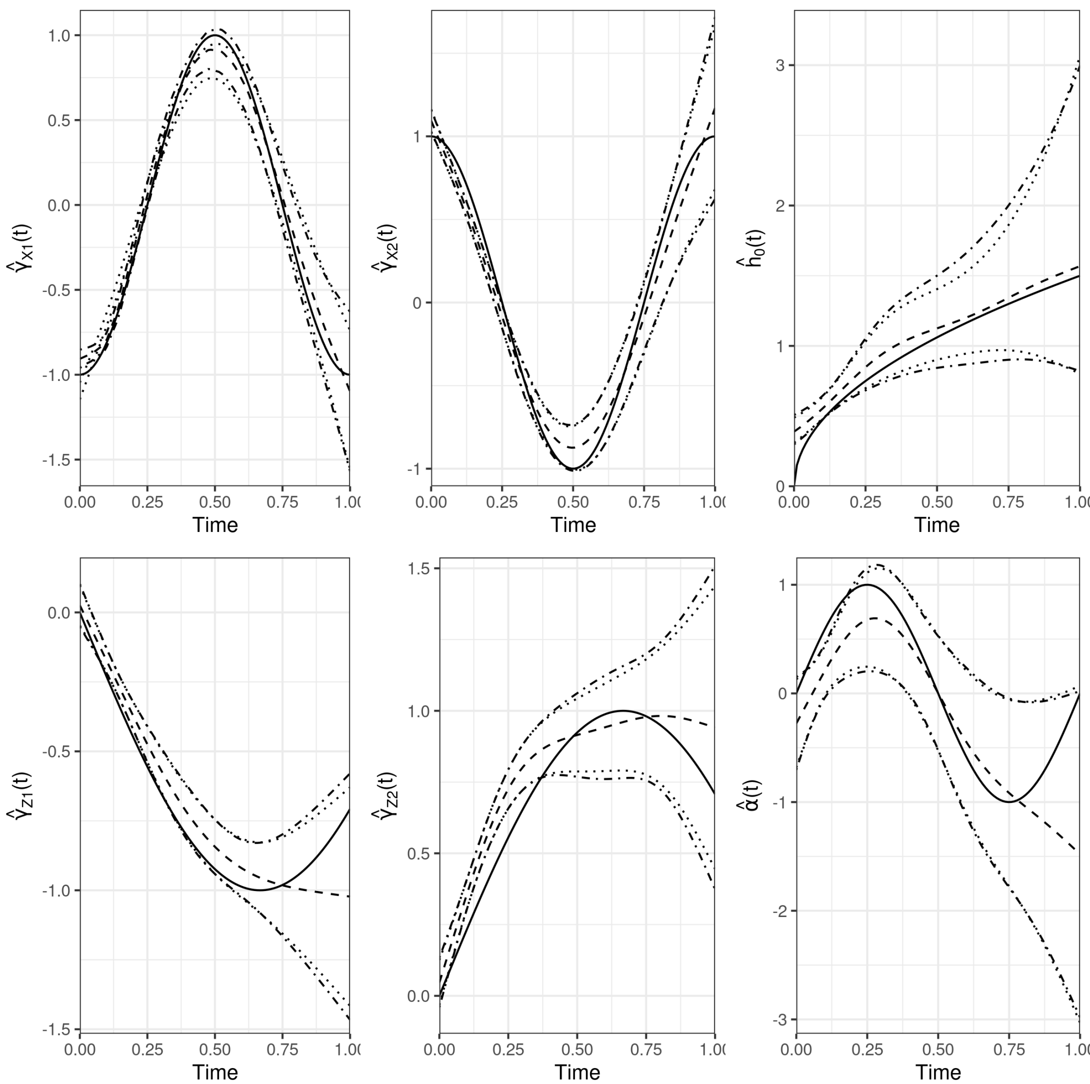

Figure 4.4: Estimated time-varying functions (dashed line) of the survival submodel, based on (3.2), simulation runs for $n = 250$ facilities overlaying the true functions (solid line) along with 95% bootstrap-based (dotted) and mean theoretical (dashed-dotted) confidence intervals.

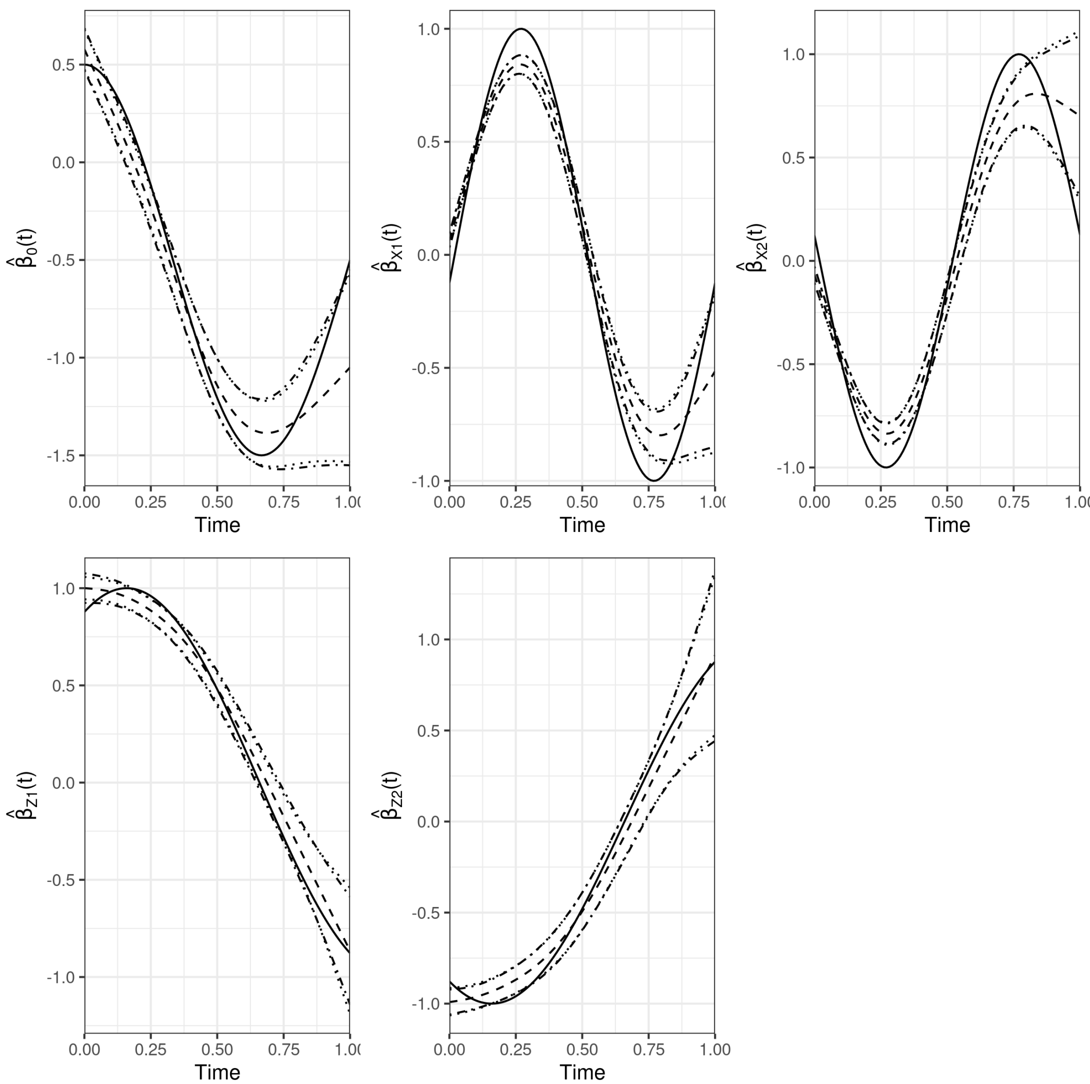

Figure 4.5: Estimated time-varying functions (dashed line) in the longitudinal submodel, based on (3.1), simulation runs for $n = 500$ facilities overlaying the true functions (solid line) along with 95% bootstrap-based (dotted) and mean theoretical (dashed-dotted) confidence intervals.

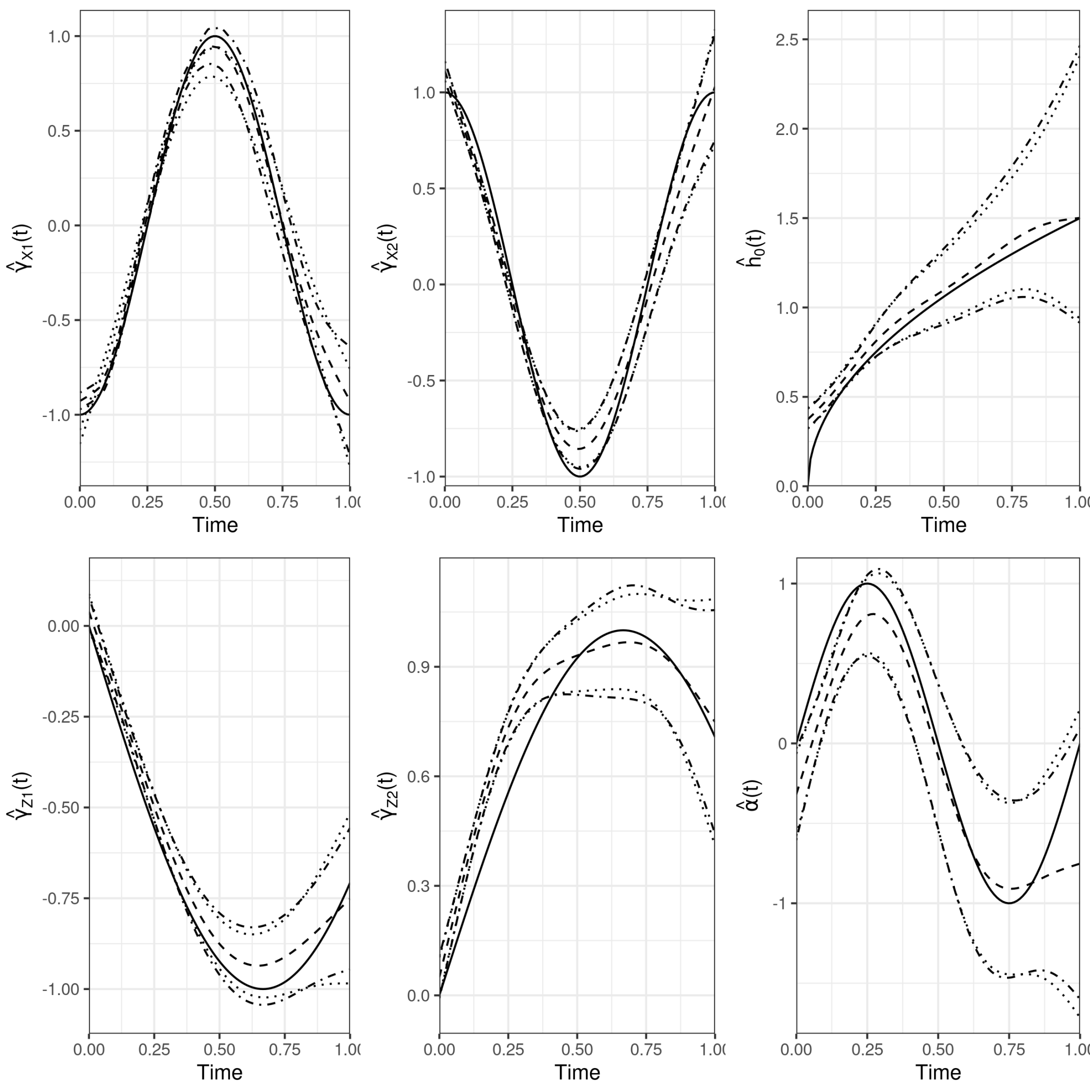

Figure 4.6: Estimated time-varying functions (dashed line) in the survival submodel, based on (3.2), simulation runs for $n = 500$ facilities overlaying the true functions (solid line) along with 95% bootstrap-based (dotted) and mean theoretical (dashed-dotted) confidence intervals.

Parameter	True value	Number of facilities	Bias	SD	Boot_{SE}	$\text{Boot}_{\text{SD}_{\text{SE}}}$
σ_S^2	1.3	250	0.03	0.07	0.06	0.01
		500	0.03	0.05	0.04	0.005
σ_F^2	0.2	250	0.02	0.03	0.03	0.04
		500	0.02	0.02	0.02	0.003

Table 4.4: The true and estimated values, as well as the true and bootsrap-based percentiles, are provided for the subject- (σ_S^2) and facility-level (σ_F^2) variances.

Function	Time	True value	Number of facilities	Bias	SD	Boot_{SE}	$\text{Boot}_{\text{SD}_{\text{SE}}}$
$\beta_0(t)$	0.25	-0.12	250	-0.12	0.09	0.08	0.01
			500	0.06	0.06	0.05	0.01
	0.5	-1.21	250	0.11	0.13	0.12	0.02
			500	-0.03	0.07	0.08	0.01
	0.75	-1.42	250	-0.05	0.21	0.18	0.04
			500	-0.06	0.11	0.10	0.01
$\beta_{X_1}(t)$	0.25	0.99	250	-0.06	0.03	0.03	0.00
			500	0.08	0.02	0.03	0.00
	0.5	0.12	250	0.03	0.06	0.06	0.01
			500	-0.02	0.04	0.04	0.01
	0.75	-0.99	250	0.11	0.10	0.10	0.02
			500	-0.13	0.06	0.07	0.01
$\beta_{X_1}(t)$	0.25	-0.99	250	0.10	0.03	0.04	0.01
			500	-0.06	0.03	0.03	0.01
	0.5	-0.12	250	-0.10	0.06	0.07	0.01
			500	0.05	0.05	0.05	0.01
	0.75	0.99	250	-0.11	0.14	0.12	0.02
			500	0.17	0.07	0.07	0.01
$\beta_{Z_1}(t)$	0.25	0.96	250	-0.10	0.04	0.04	0.01
			500	0.09	0.03	0.03	0.01
	0.5	0.48	250	-0.02	0.06	0.06	0.02
			500	-0.01	0.05	0.04	0.01
	0.75	-0.28	250	0.06	0.10	0.09	0.02
			500	-0.10	0.08	0.06	0.01
$\beta_{Z_2}(t)$	0.25	-0.96	250	0.13	0.05	0.06	0.01
			500	-0.10	0.04	0.04	0.01
	0.5	-0.48	250	-0.03	0.08	0.09	0.02
			500	0.03	0.06	0.06	0.01
	0.75	0.28	250	-0.04	0.13	0.12	0.03
			500	0.09	0.08	0.08	0.01

Table 4.5: The bias, theoretical standard deviation, bootstrap-based standard error, and the standard deviation of the bootstrap-based standard error for the longitudinal submodel

Function	Time	True value	Number of facilities	Bias	SD	Boot_{SE}	$\text{Boot}_{\text{SD}_{\text{SE}}}$
$\alpha(t)$	0.25	1	250	-0.17	0.30	0.27	0.06
			500	0.02	0.15	0.14	0.02
	0.5	0	250	-0.05	0.29	0.29	0.06
			500	0.13	0.25	0.26	0.04
	0.75	-1	250	0.07	0.45	0.45	0.08
			500	-0.04	0.31	0.29	0.04
$\hbar_0(t)$	0.25	0.75	250	0.14	0.12	0.11	0.02
			500	-0.09	0.05	0.07	0.01
	0.5	1.06	250	0.06	0.17	0.12	0.02
			500	-0.02	0.11	0.09	0.01
	0.75	1.30	250	0.08	0.28	0.17	0.03
			500	-0.10	0.20	0.12	0.02
$\gamma_{X_1}(t)$	0.25	0	250	0.06	0.04	0.04	0.01
			500	-0.05	0.03	0.03	0.00
	0.5	1	250	-0.09	0.06	0.06	0.01
			500	0.06	0.05	0.05	0.01
	0.75	0	250	0.05	0.10	0.11	0.02
			500	-0.02	0.08	0.07	0.01
$\gamma_{X_1}(t)$	0.25	0	250	-0.07	0.04	0.04	0.01
			500	0.07	0.03	0.03	0.00
	0.5	-1	250	0.06	0.08	0.07	0.01
			500	-0.08	0.06	0.06	0.01
	0.75	0	250	-0.05	0.11	0.12	0.02
			500	0.03	0.08	0.09	0.01
$\gamma_{Z_1}(t)$	0.25	-0.56	250	0.08	0.03	0.03	0.01
			500	-0.05	0.02	0.02	0.00
	0.5	-0.93	250	0.08	0.05	0.05	0.01
			500	-0.04	0.05	0.04	0.01
	0.75	-0.98	250	0.00	0.09	0.09	0.02
			500	-0.07	0.06	0.05	0.01
$\gamma_{Z_2}(t)$	0.25	0.56	250	0.21	0.04	0.04	0.01
			500	-0.22	0.02	0.03	0.00
	0.5	0.93	250	-0.03	0.08	0.07	0.01
			500	0.03	0.06	0.06	0.01
	0.75	0.98	250	0.01	0.12	0.11	0.02
			500	0.00	0.09	0.08	0.02

Table 4.6: The bias, theoretical standard deviation, bootstrap-based standard error, and the standard deviation of the bootstrap-based standard error for the survival submodel

Chapter 5

Application to USRDS Data: Joint Modeling of Hospitalization and Survival Outcomes

5.0.1 USRDS Study Cohort and Patient- and Facility-level Risk Factors

We applied the MTJMRE to the United States Renal Data System (USRDS) data, a national database collecting information on nearly all U.S. patients with end-stage renal disease on dialysis. The study collected information on patients who were at least 18 years old and initiated dialysis between January 1, 2006 and December 31, 2008. These patients were followed for 5 years, with the last follow-up date as December 31, 2013, or until a patient switched facilities. The inclusion criteria were: (1) patients who survived the first 90 days, did not recover any kidney function, and did not have a kidney transplant, and (2) patients were covered by Medicare as their primary payer on Day 91.

Per the recommendation of the USRDS researcher's guide "90-day rule", the first day of study follow-up started on Day 91 to allow for the completion of the Medicare eligibility application and establish a stable dialysis treatment modality (United States Renal Data System, 2018). The final study cohort included 292,672 observations, 34,030 patients, and 520 facilities, where each facility approximately contained 50 to 162 patients (median is 61, first [Q1] and third [Q3] quartile are 54-71, respectively).

The patient mean age was 65 years old with a standard deviation of 15, and 45% of patients were recorded as female. The common baseline comorbidites are reported as follows: chronic obstructive pulmonary disease (COPD; 18.7%), septicemia (10.2%), other infectious disease (23.1%), cardiorespiratory failure (12%), coagulopathy (7.9%), and psychiatric conditions (11.2%). Among 520 facilities, the median follow-up time was 24.3 months (Q1-Q3: 21.1-27.4), and the mean number of hospitalizations was 1.8 person-years with a standard deviation 2.2. The median unadjusted marginal survival time was calculated as 46.5 months. The mean ratios for nurses to patients and patient care technician (PCT) to patients is 7.6% and 9.4% with standard deviations of 3.2 and 2.9, respectively.

The proposed MTJMRE was fitted to study the time-varying effects of the patient- and facility-level covariates for the longitudinal and survival submodels. The longitudinal submodel included age (centered), sex, baseline comorbidities (COPD, septicemia, other infectious diseases, cardiorespiratory failure, coagulopathy, and psychiatric conditions), nurse-to-patient ratio, and PCT-to-patient ratio. The survival submodel includes age (centered), sex, baseline comorbidities, nurse-to-patient ratio, PCT-to-patient ratio, and hospitalization risk score (longitudinal outcome) as a covariate. The time-varying coefficients were estimated using a random-coefficient spline models with 20 equally-spaced knots. The baseline hazard function was fitted using a Weibull model. All analysis was conducted in R (version 3.6.3) (R Core Team, 2022) and utilized the `Armadillo` library (Sanderson and Curtin, 2016, 2018) via the `RcppArmadillo` package (Eddelbuettel and Sanderson, 2014).

5.0.2 Analysis

The estimated time-varying effects of the patient-level risk factors (solid line) on longitudinal hospitalizations are displayed in figure 5.1, along with the 95% confidence intervals (dashed lines), and a reference line at 1 (dotted line). The results demonstrate that older age at transition to dialysis is associated with a higher odds of hospitalization starting at about 12 months (i.e., after the fragile first year transition to dialysis time period; figure 5.1 (b)). The effect of the psychiatric conditions and other infectious diseases/pneumonia is relatively stable over time and they are significantly associated with higher odds of hospitalization approximately until the end of the third and fourth year of dialysis, respectively. The effect of COPD (figure 5.1 (d)) becomes significant after the first year of dialysis and similar to other chronic and acute conditions, its effect can be considered to stay stable during the first five years of transition to dialysis. The estimated subject- and facility-level variances were $\sigma_S^2 = 1.15$ and $\sigma_F^2 = 0.82$ with a 95% confidence interval of (0.83, 1.46) and (0.73, 0.92), respectively.

The estimated time-varying effects (solid line) of the patient-level covariates on the survival submodel is shown in figure 5.2, with the 95% confidence intervals (dashed lines), and a reference line at 1 (dotted line). As expected, older age at dialysis transition is associated with an increased hazard of death (figure 5.2 (a)). We observe that between first and fourth years of dialysis, females have a lower risk of death compared to males (figure 5.2 (a)). The chronic obstructive pulmonary disease is significantly associated with risk of death during the second and third years of dialysis and these conditions both increase the hazard of death. We observe a similar result in the effect of cardiorespiratory failure and psychiatric conditions (figure 5.2 (f,h)). The estimated time-varying effect of hospitalization risk score on the hazard of death is presented in figure 5.2 (i). According to the bootstrap-based confidence intervals, hospitalization risk score have a significant effect on survival between months 38 and 48 with the highest point estimate of HR(t)$\approx$ 5.0 at about 47 months post-dialysis. The estimated baseline hazard parameter was $\lambda = 1.004$ with a 95% confidence interval of (1, 1.04).

Figure 5.3 displays the time-varying effects (solid line) along with the bootstrap-based confidence intervals (dashed lines) and the reference line (dotted line). The estimated facility-level time-varying effects indicate that both nurse-to-patient ratio and PCT-to-patient ratio are associated with significantly lower risk of death. In addition, we observe that the hazard ratios for both factors remain constant throughout the study period, HR(t) $\approx$ 0.8 and HR(t) $\approx$ 0.75, respectively, for $0 < t \leq 60$. In terms of hospitalization, between months 25 and 40 after transition to dialysis, both nurse-to-patient and PCT-to-patient ratios are associated with a slightly lower the risk of hospitalization.

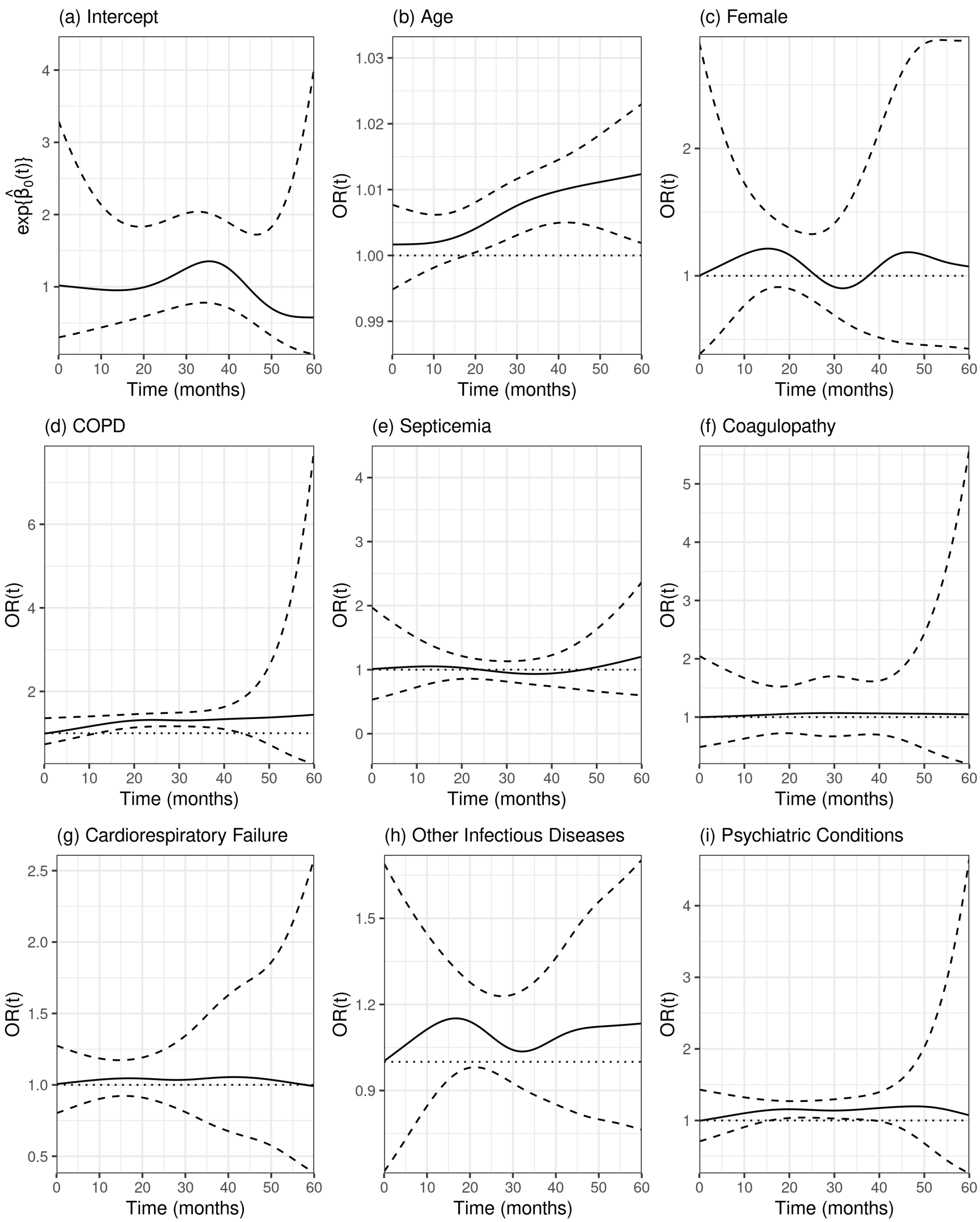

Figure 5.1: Estimated patient-level effects on hospitalization, time-varying odds ratio $OR(t) = \exp\{\hat{\beta}_{X_\omega}(t)\}$, (solid) along with their 95% bootstrap-based confidence intervals (dashed).

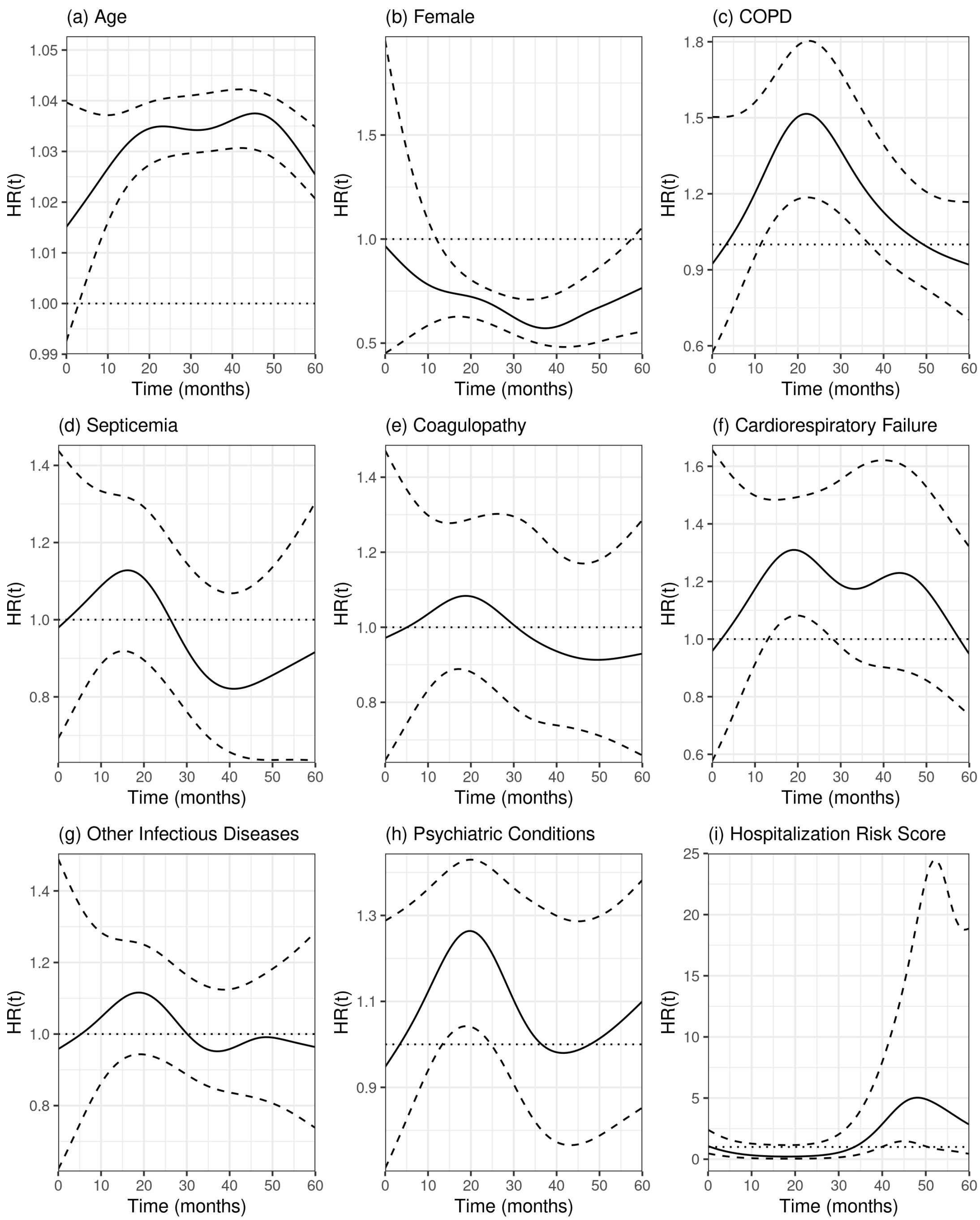

Figure 5.2: Estimated patient-level effects on survival, time-varying hazard ratios $HR(t) = \exp\{\hat{\gamma}_{X_\omega}(t)\}$ and $HR(t) = \exp\{\hat{\alpha}(t)\}$, (solid) along with their 95% bootstrap-based confidence intervals (dashed).

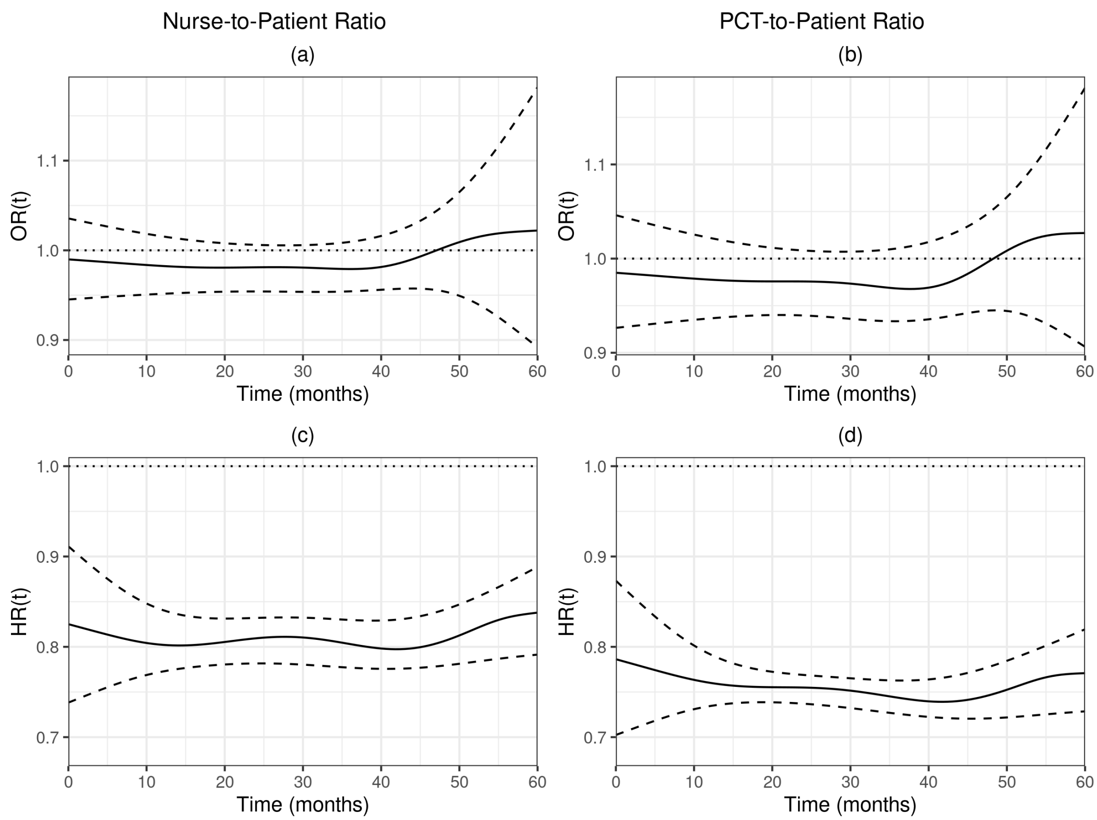

Figure 5.3: Estimated facility-level effects on (a,b) hospitalization, time-varying odds ratio $OR(t) = \exp\{\hat{\beta}_{Z_\nu}(t)\}$, and (c,d) on survival, time-varying hazard ratio $HR(t) = \exp\{\hat{\gamma}_{Z_\nu}(t)\}$, (solid) along with their 95% bootstrap-based confidence intervals (dashed).

Chapter 6

Conclusion

Motivated by the data from United States Data Renal System (USRDS), we introduced a multilevel time-varying joint longitudinal-survival model to describe the dynamic associations between each outcome (longitudinal and survival) and their corresponding multilevel predictors (subject- and facility-level risk factors). We adopted a varying-coefficient modeling scheme for each outcome to explore these time-varying relationships. In terms of estimation, we proposed two approaches, based on P-splines and random-coefficient splines. Under each method, estimation and inference were performed via a proposed approximate Expectation-Maximization algorithm, where at the E-step, the random effects were targeted, while at the M-step, model parameters were estimated via a Newton-Raphson algorithm. We investigated the finite sample capabilities of both estimation procedures through extensive simulation studies, in which we recommended estimating the standard errors using a bootstrap approach. Note that, although both procedures performed well and were demonstrated to be capable of estimating time-dynamic associations in a joint modeling setting, we suggest using the P-splines method when the number of covariates in both submodels are small.

In addition to the work presented in this dissertation, future work on Multilevel Time-varying Joint Models are as follows:

1. Extend the models to incorporate multiple longitudinal outcomes as well as competing risk type of survival outcomes.

2. Incorporate different functional forms of $m_{ij}(t)$, such as longitudinal trajectory $(dm_{ij}(t)/dt)$ or cumulative effect $(\int m_{ij}(t)/dt)$, into the survival submodel.

Bibliography

Andersen, P. K. and Gill, R. D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, 10(4):1100–1120.

Andrinopoulou, E.-R., Eilers, P. H. C., Takkenberg, J. J. M., and Rizopoulos, D. (2018). Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using P-splines. *Biometrics*, 74(2):685–693.

Bender, R., Augustin, T., and Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, 24(11):1713–1723.

Breslow, N. (1974). Covariance Analysis of Censored Survival Data. *Biometrics*, 30(1):89–99.

Brumback, B. A., Ruppert, D., and Wand, M. P. (1999). Comment. *Journal of the American Statistical Association*, 94(447):794–797.

Chiang, C.-T., Rice, J. A., and Wu, C. O. (2001). Smoothing Spline Estimation for Varying Coefficient Models with Repeatedly Measured Dependent Variables. *Journal of the American Statistical Association*, 96(454):605–619.

Cleveland, W. S., Grosse, E., and Shyu, W. M. (1992). Local Regression Models. In *Statistical Models in S*. Routledge. Num Pages: 68.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

de Boor, C. (1978). *A Practical Guide to Splines*. Springer New York, NY.

De Gruttola, V. and Tu, X. M. (1994). Modelling Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time. *Biometrics*, 50(4):1003.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Diggle, P. (2002). *Analysis of longitudinal data.* Number 25 in Oxford statistical science series. Oxford University Press, Oxford ;, 2nd ed. / peter j. diggle ... [et al.]. edition.

Eddelbuettel, D. and Sanderson, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063.

Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association*, 72(359):557–565.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2).

Eilers, P. H. C. and Marx, B. D. (2010). Splines, knots, and penalties. *WIREs Computational Statistics*, 2(6):637–653.

Eilers, P. H. C., Marx, B. D., and Durban, M. (2015). Twenty years of P-splines. *SORT-Statistics and Operations Research Transactions*, 39(3):149–186.

Fan, J. and Zhang, W. (1999). Statistical Estimation in Varying Coefficient Models. *The Annals of Statistics*, 27(5):1491–1518.

Fan, J. and Zhang, W. (2008). Statistical Methods with Varying Coefficient Models. *Statistics and its interface*, 1(1):179–195.

Fitzmaurice, G. M. (2004). *Applied longitudinal analysis.* Wiley series in probability and statistics. Wiley-Interscience, Hoboken, N.J.

Givens, G. H. and Hoeting, J. (2012). *Computational Statistics.* Wiley Series in Computational Statistics. Wiley, Chicester, 2nd ed edition.

Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2011). Penalized Functional Regression. *Journal of computational and graphical statistics : a joint publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America*, 20(4):830–851.

Goldsmith, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2012). Longitudinal Penalized Functional Regression for Cognitive Outcomes on Neuronal Tract Measurements. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 61(3):453–469.

Greenwood, M. (1926). *A Report on the Natural Duration of Cancer.* London HMSO. Publisher: London: H.M.S.O.

Hastie, T. and Tibshirani, R. (1993). Varying-Coefficient Models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779.

Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480.

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85(4):809–822.

Hsieh, F., Tseng, Y.-K., and Wang, J.-L. (2006). Joint Modeling of Survival and Longitudinal Data: Likelihood Approach Revisited. *Biometrics*, 62(4):1037–1043.

Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-Coefficient Models and Basis Function Approximations for the Analysis of Repeated Measurements. *Biometrika*, 89(1):111–128.

Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial Spline Estimation and Inference for Varying Coefficient Models with Longitudinal Data. *Statistica Sinica*, 14(3):763–788.

Kalbfleisch, J. D. (2002). *The statistical analysis of failure time data.* Wiley series in probability and statistics. JWiley, Hoboken, N.J, 2nd ed. edition.

Kaplan, E. L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282):457–481.

Kürüm, E., Li, R., Wang, Y., and Şentürk, D. (2014). Nonlinear Varying-Coefficient Models with Applications to a Photosynthesis Study. *Journal of Agricultural, Biological, and Environmental Statistics*, 19(1):57–81.

Kürüm, E., Nguyen, D. V., Li, Y., Rhee, C. M., Kalantar-Zadeh, K., and Şentürk, D. (2021). Multilevel joint modeling of hospitalization and survival in patients on dialysis. *Stat*, 10(1):e356.

Lang, S. and Brezger, A. (2004). Bayesian P-Splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212.

Larsen, K. (2004). Joint Analysis of Time-to-Event and Multiple Binary Indicators of Latent Classes. *Biometrics*, 60(1):85–92.

Li, N., Elashoff, R. M., Li, G., and Saver, J. (2009). Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Statistics in Medicine*, pages n/a–n/a.

Li, Y., Nguyen, D. V., Kürüm, E., Rhee, C. M., Chen, Y., Kalantar-Zadeh, K., and Şentürk, D. (2020). A multilevel mixed effects varying coefficient model with multilevel predictors and random effects for modeling hospitalization risk in patients on dialysis. *Biometrics*, 76(3):924– 938.

Liu, L., Ma, J. Z., and O'Quigley, J. (2008). Joint analysis of multi-level repeated measures data and survival: an application to the end stage renal disease (ESRD) data. *Statistics in Medicine*, 27(27):5679–5691.

Nan, B., Lin, X., Lisabeth, L. D., and Harlow, S. D. (2005). A Varying-Coefficient Cox Model for the Effect of Age at a Marker Event on Age at Menopause. *Biometrics*, 61(2):576–583. tex.ids: nanVaryingCoefficientCoxModel2005a publisher: [Wiley, International Biometric Society].

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1992). *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*. Cambridge University Press, 2 edition. Published: Hardcover.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rizopoulos, D., Verbeke, G., and Lesaffre, E. (2009). Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):637–654.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika*, 95(1):63–74.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Sanderson, C. and Curtin, R. (2016). Armadillo: a template-based C++ library for linear algebra. *The Journal of Open Source Software*, 1(2):26.

Sanderson, C. and Curtin, R. (2018). A User-Friendly Hybrid Sparse Matrix Class in C++. In Davenport, J. H., Kauers, M., Labahn, G., and Urban, J., editors, *Mathematical Software – ICMS 2018*, volume 10931, pages 422–430. Springer International Publishing, Cham. Series Title: Lecture Notes in Computer Science.

Song, X., Davidian, M., and Tsiatis, A. A. (2002). A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data. *Biometrics*, 58(4):742–753.

Song, X. and Wang, C. Y. (2008). Semiparametric Approaches for Joint Modeling of Longitudinal and Survival Data with Time-Varying Coefficients. *Biometrics*, 64(2):557–566.

Sweeting, M. J. and Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal. Biometrische Zeitschrift*, 53(5):750–763.

Tierney, L., Kass, R. E., and Kadane, J. B. (1989). Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions. *Journal of the American Statistical Association*, 84(407):710–716.

Tsiatis, A. A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88(2):447–458. Publisher: Oxford Academic.

Tsiatis, A. A. and Davidian, M. (2004). Joint Modeling of Longitudinal and Time-to-Event Data: An Overview. *Statistica Sinica*, 14(3):809–834.

Tsiatis, A. A., DeGruttola, V., and Wulfsohn, M. S. (1995). Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS. *Journal of the American Statistical Association*, 90(429):27–37.

United States Renal Data System (2018). USRDS Annual Report: End Stage Renal Disease in the United States. Technical report, National Institute of Health, National Institutle of Diabetes and Digestive and Kidney Diseases.

Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis*, 23(4):541–556.

Wu, C. O., Chiang, C.-T., and Hoover, D. R. (1998). Asymptotic Confidence Regions for Kernel Smoothing of a Varying-Coefficient Model with Longitudinal Data. *Journal of the American Statistical Association*, 93(444):1388–1402.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–339.

Appendix A

EM Algorithm Details

A.1 P-Spline Models

A.1.1 E-Step Functions

The score function $\mathcal{G}_{\boldsymbol{u}_i}$ is written as

$$\mathcal{G}_{\boldsymbol{u}_i} = -\left(\Delta_{i1}^{(1)} - \frac{b_{i1}}{\sigma_S^2}, \Delta_{i2}^{(1)} - \frac{b_{i1}}{\sigma_S^2}, \cdots, \Delta_{in_i}^{(1)} - \frac{b_{in_i}}{\sigma_S^2}, \sum_{j=1}^{n_i}\left\{\Delta_{ij}^{(1)}\right\} - \frac{b_i}{\sigma_F^2}\right)^{\mathrm{T}},$$

where

$$\Delta_{ij}^{(1)} = \delta_{ij}\alpha(T_{ij})m_{ij}(T_{ij})q_{ij}(T_{ij}) - \int_0^{T_{ij}} h_0(s)\alpha(s)m_{ij}(s)q_{ij}(s)\mathcal{E}_{ij}ds + \sum_{k=1}^{n_{ij}}(Y_{ijk} - m_{ijk})$$

with $q_{ij}(T_{ij}) = 1 - m_{ij}(T_{ij})$ and $\mathcal{E}_{ij} = exp\{\boldsymbol{Z}_{i(j)}^{\mathrm{T}}\gamma_Z(t) + \boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_X(t) + \alpha(t)m_{ij}(t)\}$.

The Hessian matrix $\boldsymbol{\mathcal{H}}_{\boldsymbol{u}_i}$ is written as

$$\boldsymbol{\mathcal{H}}_{\boldsymbol{u}_i} = \boldsymbol{\Sigma}_i = -\left\{ \begin{array}{ccccc} \frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_{i1}^2} & 0 & \cdots & 0 & \frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_{i1}\partial \mathscr{b}_i} \\ 0 & \frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_{i2}^2} & \cdots & 0 & \frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_{i2}\partial \mathscr{b}_i} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_{in_i}^2} & \frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_{in_i}\partial \mathscr{b}_i} \\ \frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_{i1}\partial \mathscr{b}_i} & \frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_{i2}\partial \mathscr{b}_i} & \cdots & \frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_{in_i}\partial \mathscr{b}_i} & \frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_i^2} \end{array} \right\}, \tag{A.1}$$

where $\frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_{ij}^2} = \Delta_{ij}^{(2)} - \frac{1}{\sigma_S^2}$, $\frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_i^2} = \sum_{j=1}^{n_i} \left\{ \Delta_{ij}^{(2)} \right\} - \frac{1}{\sigma_F^2}$, and $\frac{\partial^2 \ell_i(\boldsymbol{u}_i,\boldsymbol{\theta})}{\partial \mathscr{b}_{ij}\partial \mathscr{b}_i} = \Delta_{ij}^{(2)}$ with

$$\begin{aligned} \Delta_{ij}^{(2)} \quad = \quad & \delta_{ij}\alpha(T_{ij})\mathscr{m}_{ij}(T_{ij})\mathscr{q}_{ij}(T_{ij})\left\{\mathscr{q}_{ij}(T_{ij}) - \mathscr{m}_{ij}(T_{ij})\right\} \\ & - \int_0^{T_{ij}} \mathscr{h}_0(s)\alpha(s)\mathscr{m}_{ij}(s)\mathscr{q}_{ij}(s)\mathcal{E}_{ij}\{\alpha(s)\mathscr{m}_{ij}(s)\mathscr{q}_{ij}(s) + \mathscr{q}_{ij}(s) - \mathscr{m}_{ij}(s)\}ds \\ & - \sum_{k=1}^{n_{ij}} \mathscr{m}_{ijk}\mathscr{q}_{ijk}. \end{aligned} \tag{A.2}$$

To obtain the values involved in $\mathcal{V}$, we must first obtain $\frac{\partial \boldsymbol{\Sigma}_i}{\partial \boldsymbol{u}_i^{\mathrm{T}}}$, $\frac{\partial^2 \boldsymbol{\Sigma}_i}{\partial \boldsymbol{u}_i^{\mathrm{T}} \partial \boldsymbol{u}_i}$, $\frac{\partial \hat{\boldsymbol{u}}_i^{(c)}}{\partial \boldsymbol{c}^T}$, $\frac{\partial^2 \hat{\boldsymbol{u}}_i^{(c)}}{\partial \boldsymbol{c}^T \partial \boldsymbol{c}}$, and $\left(\frac{\partial \hat{\boldsymbol{u}}_i^{(c)}}{\partial \boldsymbol{c}^T}\right)\left(\frac{\partial \hat{\boldsymbol{u}}_i^{(c)}}{\partial \boldsymbol{c}}\right)$. The first derivative of $\boldsymbol{\Sigma}_i$ with respect to $\boldsymbol{u}_i^{\mathrm{T}}$ is a $(n_i+1)^2 \times (n_i+1)$ matrix from the concatenation of $(n_i+1)\times(n_i+1)$ matrices representing the $\frac{\partial \boldsymbol{\Sigma}_i}{\partial \mathscr{b}_{ij}}$ and $\frac{\partial \boldsymbol{\Sigma}_i}{\partial \mathscr{b}_i}$. Let $\Psi_{ij}^{(1)}$ represent the matrix involved in $\frac{\partial \boldsymbol{\Sigma}_i}{\partial \mathscr{b}_{ij}}$.

The only nonzero elements involved in the $\Psi_{ij}^{(1)}$ are the $(j,j)^{\text{th}}$, $(j,n_i+1)^{\text{th}}$, $(n_i+1,j)^{\text{th}}$, and $(n_i+1,n_i+1)^{\text{th}}$ which are formulated as

$$\begin{aligned}
\Psi_{ij}^{(1)} &= -\delta_{ij}\alpha(T_{ij})m_{ij}(T_{ij})q_{ij}(T_{ij})\left[\left\{m_{ij}(T_{ij})-q_{ij}(T_{ij})\right\}^2-2m_{ij}(T_{ij})q_{ij}(T_{ij})\right] \\
&\quad +\int_0^{T_{ij}} h_0(s)\alpha(s)m_{ij}(s)q_{ij}(s)\mathcal{E}_{ij}\times \\
&\quad \times\left[\left\{\alpha(s)m_{ij}(s)q_{ij}(s)+q_{ij}(s)-m_{ij}(s)\right\}^2\right. \\
&\quad \left.+m_{ij}(s)q_{ij}(s)\left\{\alpha(s)q_{ij}(s)-\alpha(s)m_{ij}(s)-2\right\}\right]ds \\
&\quad +\sum_{k=1}^{n_{ij}}\left\{m_{ijk}q_{ijk}^2-m_{ijk}^2q_{ijk}\right\}.
\end{aligned}$$

The second derivative of $\boldsymbol{\Sigma}_i$ with respect to $\boldsymbol{u}_i^{\mathrm{T}}$ and $\boldsymbol{u}_i$ is a $(n_i+1)^2\times(n_i+1)^2$ matrix compiled of block matrices of size $(n_i+1)\times(n_i+1)$:

$$\frac{\partial^2\boldsymbol{\Sigma}_i}{\partial\boldsymbol{u}_i^T\partial\boldsymbol{u}_i}=\begin{pmatrix}
\Psi_{i1}^{(2)} & \mathbf{0} & \cdots & \mathbf{0} & \Psi_{i1}^{(2)} \\
\mathbf{0} & \Psi_{i2}^{(2)} & \cdots & \mathbf{0} & \Psi_{i2}^{(2)} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \Psi_{in_i}^{(2)} & \Psi_{in_i}^{(2)} \\
\Psi_{i1}^{(2)} & \Psi_{i2}^{(2)} & \cdots & \Psi_{in_i}^{(2)} & \sum_{j=1}^{n_i}\Psi_{ij}^{(2)}
\end{pmatrix},$$

where $\Psi_{ij}^{(2)}$ is an $(n_i+1)\times(n_i+1)$ matrix with the only non-zero elements being in the $(j,j)^{\text{th}}$, $(j,n_i+1)^{\text{th}}$, $(n_i+1,j)^{\text{th}}$, and $(n_i+1,n_i+1)^{\text{th}}$ position formulated as

$$\begin{aligned}\Psi_{ij}^{(2)} &= -\delta_{ij}\alpha(T_{ij})m_{ij}(T_{ij})q_{ij}(T_{ij}) \\ &\quad \times\left\{q_{ij}^3(T_{ij}) - 11m_{ij}(T_{ij})q_{ij}^2(T_{ij}) + 11m_{ij}^2(T_{ij})q_{ij}(T_{ij}) - m_{ij}^3(T_{ij})\right\} \\ &\quad + \int_0^{T_{ij}} h_0(s)\alpha(s)m_{ij}(s)q_{ij}(s)\mathcal{E}_{ij}\Big[\left\{\alpha(s)m_{ij}(s)q_{ij}(s) + q_{ij}(s) - m_{ij}(s)\right\}^3 \\ &\quad + m_{ij}(s)q_{ij}(s)\left\{\alpha(s)q_{ij}(s) - \alpha(s)m_{ij}(s) - 2\right\}\left\{3\alpha(s)m_{ij}(s)q_{ij}(s) + 4q_{ij}(s) - 4m_{ij}(s)\right\} \\ &\quad - 2\alpha(s)m_{ij}^2(s)q_{ij}^2(s)\Big]ds + \sum_{k=1}^{n_{ij}}\left(m_{ijk}q_{ijk}^3 - 4m_{ijk}^2q_{ijk}^2 + m_{ijk}^3q_{ijk}\right).\end{aligned}$$

The first derivative of $\hat{\boldsymbol{u}}_i^{(c)}$ with respect to $\boldsymbol{c}$ is a $(n_i+1)\times(n_i+1)$ matrix denoted as

$$\frac{\partial\hat{\boldsymbol{u}}_i^{(c)}}{\partial\boldsymbol{c}^{\mathrm{T}}} = \begin{pmatrix} \frac{\partial\hat{b}_{i1}^{(c)}}{\partial c_1} & \cdots & \frac{\partial\hat{b}_{i1}^{(c)}}{\partial c_{n_i+1}} \\ \vdots & \vdots & \vdots \\ \frac{\partial\hat{b}_{in_i}^{(c)}}{\partial c_1} & \cdots & \frac{\partial\hat{b}_{in_i}^{(c)}}{\partial c_{n_i+1}} \\ \frac{\partial\hat{b}_{i}^{(c)}}{\partial c_1} & \cdots & \frac{\partial\hat{b}_{i}^{(c)}}{\partial c_{n_i+1}} \end{pmatrix},$$

and the second derivative of ${\hat{\boldsymbol{u}}_i}^{(c)}$ respect to $\boldsymbol{c}$ is a $(n_i+1)^2\times(n_i+1)$ matrix denoted as

$$\frac{\partial^2\hat{\boldsymbol{u}}_i^{(c)}}{\partial\boldsymbol{c}^{\mathrm{T}}\partial\boldsymbol{c}} = \begin{pmatrix} \frac{\partial^2\hat{\boldsymbol{u}}_i^{(c)}}{\partial\boldsymbol{c}^{\mathrm{T}}\partial c_1} \\ \frac{\partial^2\hat{\boldsymbol{u}}_i^{(c)}}{\partial\boldsymbol{c}^{\mathrm{T}}\partial c_2} \\ \vdots \\ \frac{\partial^2\hat{\boldsymbol{u}}_i^{(c)}}{\partial\boldsymbol{c}^{\mathrm{T}}\partial c_{n_i+1}} \end{pmatrix}.$$

Rizopoulos et al. (2009) provided the first and second partial derivatives for $\hat{\boldsymbol{u}}_i$ with respect to $\boldsymbol{c}$ as $\frac{\partial \hat{\boldsymbol{u}}_i^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}}|_{\boldsymbol{c}=\boldsymbol{0}} = \boldsymbol{\Sigma}_i^{-1}$ and $\frac{\partial^2 \hat{\boldsymbol{u}}_i^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial \boldsymbol{c}}|_{\boldsymbol{c}=\boldsymbol{0}} = \boldsymbol{\Sigma}_i^{-1}(-\partial \boldsymbol{\Sigma}_i / \partial \boldsymbol{u}_i^{\mathrm{T}}) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\Sigma}_i^{-1}$, respectively, where $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}^{(c)}|_{\boldsymbol{c}=\boldsymbol{0}}$.

To obtain the partial derivatives $\frac{\partial \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}}$ and $\frac{\partial^2 \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial \boldsymbol{c}}$, Rizopoulos et al. (2009) expanded the derivatives using the chain rule:

$$
\begin{aligned}
\left.\frac{\partial \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{c}^T}\right|_{(\boldsymbol{c},\boldsymbol{u}_i)=(0,\hat{\boldsymbol{u}}_i)} &= \left.\frac{\partial \boldsymbol{\Sigma}_i^{(c)}}{\partial \hat{\boldsymbol{u}}_i^{(c)}} \frac{\partial \hat{\boldsymbol{u}}_i{}^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}}\right|_{(\boldsymbol{c},\boldsymbol{u}_i)=(0,\hat{\boldsymbol{u}}_i)} = \left.\frac{\partial \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{u}_i^{\mathrm{T}}} \frac{\partial \hat{\boldsymbol{u}}_i{}^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}}\right|_{\boldsymbol{c}=0}, \\
\left.\frac{\partial^2 \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial \boldsymbol{c}}\right|_{(\boldsymbol{c},\boldsymbol{u}_i)=(0,\hat{\boldsymbol{u}}_i)} &= \left.\frac{\partial^2 \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{u}_i^{\mathrm{T}} \partial \boldsymbol{u}_i} \left(\frac{\partial \hat{\boldsymbol{u}}_i{}^{(c)}}{\partial \boldsymbol{c}} \frac{\partial \hat{\boldsymbol{u}}_i{}^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}} \right)\right|_{\boldsymbol{c}=0} + \left.\frac{\partial \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{u}_i^{\mathrm{T}}} \frac{\partial^2 \hat{\boldsymbol{u}}_i^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial \boldsymbol{c}}\right|_{\boldsymbol{c}=0}.
\end{aligned}
$$

The first derivative of $\boldsymbol{\Sigma}_i^{(c)}$ with respect to $\boldsymbol{c}$ is a $(n_i+1)^2 \times (n_i+1)$ matrix composed of concatenated $(n_i+1) \times (n_i+1)$ matrices obtained from $\frac{\partial \boldsymbol{\Sigma}_i^{(c)}}{\partial c_j}$ for $j \in \{1, ..., n_i+1\}$, where the matrix $\frac{\partial \boldsymbol{\Sigma}_i^{(c)}}{\partial c_j} = \frac{\boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{u}_i} \frac{\partial \boldsymbol{u}_i^{(c)}}{\partial c_j} = \sum_{j\prime=1}^{n_i} \frac{\partial \boldsymbol{\Sigma}_i^{(c)}}{\partial b_{ij\prime}} \frac{\partial b_{ij\prime}^{(c)}}{\partial c_j} + \frac{\partial \boldsymbol{\Sigma}_i^{(c)}}{\partial b_i} \frac{\partial b_i^{(c)}}{\partial c_j}$. The second derivative of $\boldsymbol{\Sigma}_i^{(c)}$ with respect to $\boldsymbol{c}$ is a $(n_i+1)^2 \times (n_i+1)^2$ matrix composed of $(n_i+1)^2$ matrices representing $\frac{\partial^2 \boldsymbol{\Sigma}_i^{(c)}}{\partial c_j \partial c_{j\prime}}$ of size $(n_i+1) \times (n_i+1)$ for $j, j\prime \in \{1, ..., n_i+1\}$, where $\frac{\partial^2 \boldsymbol{\Sigma}_i^{(c)}}{\partial c_j \partial c_{j\prime}}$ is formulated as

$$
\left.\frac{\partial^2 \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{u}_i^{\mathrm{T}} \partial \boldsymbol{u}_i} \left(\frac{\partial \hat{\boldsymbol{u}}_i{}^{(c)}}{\partial c_j} \frac{\partial \hat{\boldsymbol{u}}_i{}^{(c)}}{\partial c_{j\prime}} \right)\right|_{\boldsymbol{c}=\boldsymbol{0}} + \left.\frac{\partial \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{u}_i^{\mathrm{T}}} \frac{\partial^2 \hat{\boldsymbol{u}}_i^{(c)}}{\partial c_j \partial c_{j\prime}}\right|_{\boldsymbol{c}=\boldsymbol{0}}. \tag{A.3}
$$

The components in (A.3) are obtained as

$$
\begin{aligned}
\left.\frac{\partial^2 \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{u}_i^{\mathrm{T}} \partial \boldsymbol{u}_i} \left(\frac{\partial \hat{\boldsymbol{u}}_i{}^{(c)}}{\partial c_j} \frac{\partial \hat{\boldsymbol{u}}_i{}^{(c)}}{\partial c_{j\prime}} \right)\right|_{\boldsymbol{c}=\boldsymbol{0}} = &\left[\sum_{j^*=1}^{n_i} \left\{ \frac{\partial^2 \boldsymbol{\Sigma}^{(c)}}{\partial b_{ij^*}^2} \frac{\partial \hat{b}_{ij^*}^{(c)}}{\partial c_j} \frac{\partial \hat{b}_{ij^*}^{(c)}}{\partial c_{j'}} + \frac{\partial^2 \boldsymbol{\Sigma}^{(c)}}{\partial b_{ij^*} \partial b_i} \frac{\partial \hat{b}_{ij^*}^{(c)}}{\partial c_j} \frac{\partial \hat{b}_i^{(c)}}{\partial c_{j'}} \right.\right. \\
&\left.\left.\left. + \frac{\partial^2 \boldsymbol{\Sigma}^{(c)}}{\partial b_{ij^*} \partial b_i} \frac{\partial \hat{b}_i^{(c)}}{\partial c_j} \frac{\partial \hat{b}_{ij^*}^{(c)}}{\partial c_{j'}} \right\} + \frac{\partial^2 \boldsymbol{\Sigma}^{(c)}}{\partial b_i^2} \frac{\partial \hat{b}_i^{(c)}}{\partial c_j} \frac{\partial \hat{b}_i^{(c)}}{\partial c_{j'}} \right] \right|_{\boldsymbol{c}=\boldsymbol{0}}
\end{aligned}
$$

and

$$\left.\frac{\partial \boldsymbol{\Sigma}_i^{(c)}}{\partial \boldsymbol{u}_i^{\mathrm{T}}} \frac{\partial^2 \hat{\boldsymbol{u}}_i^{(c)}}{\partial c_j \partial c_{j\prime}}\right|_{\boldsymbol{c}=\mathbf{0}} = \left.\left\{ \sum_{j^*=1}^{n_i} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \mathscr{b}_{ij^*}} \frac{\partial^2 \hat{\mathscr{b}}_{ij^*}^{(c)}}{\partial c_j \partial c_{j\prime}} + \frac{\partial \boldsymbol{\Sigma}_i}{\partial \mathscr{b}_i} \frac{\partial^2 \hat{\mathscr{b}}_i^{(c)}}{\partial c_j \partial c_{j\prime}} \right\}\right|_{\boldsymbol{c}=\mathbf{0}}.$$

A.1.2 M-Step

The penalized likelihood function (3.10) contains a penalty term $\boldsymbol{\lambda} = (\lambda_\alpha, \lambda_{\mathscr{h}}, \boldsymbol{\lambda}_{\tau_x}^{\mathrm{T}}, \boldsymbol{\lambda}_{\tau_z}^{\mathrm{T}}, \boldsymbol{\lambda}_{\psi_x}^{\mathrm{T}}, \boldsymbol{\lambda}_{\psi_z})^{\mathrm{T}}$ that controls the roughness of each time-varying coefficient function with $\boldsymbol{\lambda}_{\tau_x} = (\lambda_{\tau_{x_1}}, \ldots, \lambda_{\tau_{x_p}})^{\mathrm{T}}$, $\boldsymbol{\lambda}_{\tau_z} = (\lambda_{\tau_{z_1}}, \ldots, \lambda_{\tau_{z_q}})^{\mathrm{T}}$, $\boldsymbol{\lambda}_{\psi_x} = (\lambda_{\psi_{x_1}}, \ldots, \lambda_{\psi_{x_p}})^{\mathrm{T}}$, and $\boldsymbol{\lambda}_{\psi_z} = (\lambda_{\psi_{z_1}}, \ldots, \lambda_{\psi_{z_q}})^{\mathrm{T}}$; and the matrix Ξ^* contains the time-varying coefficients formulated as

$$\Xi^* = \begin{pmatrix} \psi_{\alpha 1} & \cdots & \psi_{\alpha R} \\ \psi_{\hbar 1} & \cdots & \psi_{\hbar R} \\ \tau_{X_1 1} & \cdots & \tau_{X_1 R} \\ \vdots & \ddots & \vdots \\ \tau_{X_p 1} & \cdots & \tau_{X_p R} \\ \tau_{Z_1 1} & \cdots & \tau_{Z_1 R} \\ \vdots & \ddots & \vdots \\ \tau_{Z_q 1} & \cdots & \tau_{Z_q R} \\ \psi_{X_1 1} & \cdots & \psi_{X_1 R} \\ \vdots & \ddots & \vdots \\ \psi_{X_p 1} & \cdots & \psi_{X_p R} \\ \psi_{Z_1 1} & \cdots & \psi_{Z_1 R} \\ \vdots & \ddots & \vdots \\ \psi_{Z_q 1} & \cdots & \psi_{Z_q R} \end{pmatrix}.$$

The maximum likelihood estimates for the parameters in $\boldsymbol{\theta}_{\mathscr{P}}$ are obtained by maximizing the penalized likelihood function (3.10). However, computing the Hessian matrix can be burdensome for the M-Step. Therefore, a quasi-Newton method is used with a BFGS update to approximate the Hessian matrix. Approximating the update for the Hessian matrix can be found in Givens and Hoeting (2012). Below are the first derivatives for the M-Step.

$$
\begin{aligned}
\frac{\partial \ell_\lambda^*(\boldsymbol{\theta}_{\mathscr{P}})}{\partial \boldsymbol{\psi}_\alpha} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\boldsymbol{B}(T_{ij})\delta_{ij}m_{ij}^*(T_{ij}) - \int_0^{T_{ij}} \boldsymbol{B}(s)h_0^*(s)\mathcal{E}_{ij}^* m_{ij}^*(s)ds + \mathcal{R}_{ij}^*\frac{\partial \Delta_{ij}^{(2)*}}{\partial \boldsymbol{\theta}_{\alpha\xi}}\right\} \\
&\quad -\frac{\lambda_\alpha}{2}\mathcal{D}^{(2)\mathrm{T}}\mathcal{D}^{(2)}\boldsymbol{\psi}_\alpha \\
\frac{\partial \ell_\lambda^*(\boldsymbol{\theta}_{\mathscr{P}})}{\partial \boldsymbol{\psi}_h} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\delta_{ij}\boldsymbol{B}(T_{ij}) - \int_0^{T_{ij}} \boldsymbol{B}(s)h_0^*(s)\mathcal{E}_{ij}^* ds + \mathcal{R}_{ij}^*\frac{\partial \Delta_{ij}^{(2)*}}{\partial \boldsymbol{\psi}_h}\right\} \\
&\quad -\frac{\lambda_h}{2}\mathcal{D}^{(2)\mathrm{T}}\mathcal{D}^{(2)}\boldsymbol{\psi}_h \\
\frac{\partial \ell_\lambda^*(\boldsymbol{\theta}_{\mathscr{P}})}{\partial \boldsymbol{\tau}_{X_\omega}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big\{X_{ij\omega}\otimes\boldsymbol{B}(T_{ij})\delta_{ij}\alpha^*(T_{ij})m_{ij}^*(T_{ij})q_{ij}^*(T_{ij}) \\
&\quad -\int_0^{T_{ij}} X_{ij\omega}\otimes\boldsymbol{B}(s)h_0^*(s)\mathcal{E}_{ij}^*\alpha^*(s)m_{ij}^*(s)q_{ij}^*(s)ds \\
&\quad +\sum_{k=1}^{n_{ij}} X_{ij\omega}\otimes\boldsymbol{B}(t_{ijk})\left(Y_{ijk}-m_{ijk}^*\right) + \mathcal{R}_{ij}^*\frac{\partial \Delta_{ij}^{(2)*}}{\partial \boldsymbol{\tau}_{X_\omega}}\Bigg\} \\
&\quad -\frac{\lambda_{\tau_{x_\omega}}}{2}\mathcal{D}^{(2)\mathrm{T}}\mathcal{D}^{(2)}\boldsymbol{\tau}_{X_\omega} \\
\frac{\partial \ell_\lambda^*(\boldsymbol{\theta}_{\mathscr{P}})}{\partial \boldsymbol{\tau}_{Z_\nu}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big\{Z_{i(j)\nu}\otimes\boldsymbol{B}(T_{ij})\delta_{ij}\alpha^*(T_{ij})m_{ij}^*(T_{ij})q_{ij}^*(T_{ij}) \\
&\quad -\int_0^{T_{ij}} Z_{i(j)\nu}\otimes\boldsymbol{B}(s)h_0^*(s)\mathcal{E}_{ij}^*\alpha^*(s)m_{ij}^*(s)q_{ij}^*(s)ds \\
&\quad +\sum_{k=1}^{n_{ij}} Z_{i(j)\nu}\otimes\boldsymbol{B}(t_{ijk})\left(Y_{ijk}-m_{ijk}^*\right) + \mathcal{R}_{ij}^*\frac{\partial \Delta_{ij}^{(2)*}}{\partial \boldsymbol{\tau}_{Z_\nu}}\Bigg\} \\
&\quad -\frac{\lambda_{\tau_{z_\nu}}}{2}\mathcal{D}^{(2)\mathrm{T}}\mathcal{D}^{(2)}\boldsymbol{\tau}_{Z_\nu} \\
\frac{\partial \ell_\lambda^*(\boldsymbol{\theta}_{\mathscr{P}})}{\partial \boldsymbol{\psi}_{X_\omega}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\delta_{ij}X_{ij\omega}^{\mathrm{T}}\otimes\boldsymbol{B}(T_{ij}) - \int_0^{T_{ij}} X_{ij\omega}\otimes\boldsymbol{B}(s)h_0^*(s)\mathcal{E}_{ij}^* ds + \mathcal{R}_{ij}^*\frac{\partial \Delta_{ij}^{(2)*}}{\partial \boldsymbol{\psi}_{X_\omega}}\right\} \\
&\quad -\frac{\lambda_{\psi_{x_\omega}}}{2}\mathcal{D}^{(2)\mathrm{T}}\mathcal{D}^{(2)}\boldsymbol{\psi}_{X_\omega} \\
\frac{\partial \ell_\lambda^*(\boldsymbol{\theta}_{\mathscr{P}})}{\partial \boldsymbol{\psi}_{Z_\nu}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\delta_{ij}Z_{i(j)\nu}^{\mathrm{T}}\otimes\boldsymbol{B}(T_{ij}) - \int_0^{T_{ij}} Z_{i(j)\nu}\otimes\boldsymbol{B}(s)h_0^*(s)\mathcal{E}_{ij}^* ds + \mathcal{R}_{ij}^*\frac{\partial \Delta_{ij}^{(2)*}}{\partial \boldsymbol{\psi}_{Z_\nu}}\right\} \\
&\quad -\frac{\lambda_{\psi_{z_\nu}}}{2}\mathcal{D}^{(2)\mathrm{T}}\mathcal{D}^{(2)}\boldsymbol{\psi}_{Z_\nu}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \Delta_{ij}^{(2)*}}{\partial \boldsymbol{\psi}_{\alpha}} &= \boldsymbol{B}(T_{ij})\delta_{ij}\mathpzc{m}_{ij}^{*}(T_{ij})\mathpzc{q}_{ij}^{*}(T_{ij})\{\mathpzc{q}_{ij}^{*}(T_{ij}) - \mathpzc{m}_{ij}^{*}(T_{ij})\} - \int_{0}^{T_{ij}} \boldsymbol{B}(s)\mathpzc{h}_{0}^{*}(s)\mathcal{E}_{ij}^{*}\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s) \\
&\quad \times \left\{\alpha^{*2}(s)\mathpzc{m}_{ij}^{2*}(s)\mathpzc{q}_{ij}^{*}(s) + 3\alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s) - \alpha^{*}(s)\mathpzc{m}_{ij}^{2*} + \mathpzc{q}_{ij}^{*}(s) - \mathpzc{m}_{ij}^{*}(s)\right\} ds \\
\frac{\partial \Delta_{ij}^{(2)*}}{\partial \boldsymbol{\psi}_{\mathpzc{h}}} &= -\int_{0}^{T_{ij}} \boldsymbol{B}(s)\mathpzc{h}_{0}^{*}(s)\alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s)\mathcal{E}_{ij}^{*}\{\alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s) + \mathpzc{q}_{ij}^{*}(s) - \mathpzc{m}_{ij}^{*}(s)\}ds \\
\frac{\partial \Delta_{ij}^{(2)*}}{\partial \boldsymbol{\tau}_{X_{\omega}}} &= X_{ij\omega} \otimes \boldsymbol{B}(T_{ij})\delta_{ij}\alpha^{*}(T_{ij})\mathpzc{m}_{ij}^{*}(T_{ij})\mathpzc{q}_{ij}^{*}(T_{ij}) \\
&\quad \times \left[\left\{\mathpzc{q}_{ij}^{*}(T_{ij}) - \mathpzc{m}_{ij}^{*}(T_{ij})\right\}^{2} - 2\mathpzc{m}_{ij}^{*}(T_{ij})\mathpzc{q}_{ij}^{*}(T_{ij})\right] \\
&\quad - \int_{0}^{T_{ij}} X_{ij} \otimes \boldsymbol{B}(s)\mathpzc{h}_{0}^{*}(s)\mathcal{E}_{ij}^{*}\alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s) \times \\
&\quad \times \left[\left\{\alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s) + \mathpzc{q}_{ij}^{*}(s) - \mathpzc{m}_{ij}^{*}(s)\right\}^{2}\right. \\
&\quad \left. + \mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s)\left\{\alpha^{*}(s)\mathpzc{q}_{ij}^{*}(s) - \alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s) - 2\right\}\right] ds \\
&\quad - \sum_{k=1}^{n_{ij}} X_{ij\omega} \otimes \boldsymbol{B}(t_{ijk})\left\{\mathpzc{m}_{ijk}\mathpzc{q}_{ijk}^{*2} - \mathpzc{m}_{ijk}^{*2}\mathpzc{q}_{ijk}^{*}\right\} \\
\frac{\partial \Delta_{ij}^{(2)*}}{\partial \boldsymbol{\tau}_{Z_{\nu}}} &= Z_{i(j)\nu} \otimes \boldsymbol{B}(T_{ij})\delta_{ij}\alpha^{*}(T_{ij})\mathpzc{m}_{ij}^{*}(T_{ij})\mathpzc{q}_{ij}^{*}(T_{ij}) \\
&\quad \times \left[\left\{\mathpzc{q}_{ij}^{*}(T_{ij}) - \mathpzc{m}_{ij}^{*}(T_{ij})\right\}^{2} - 2\mathpzc{m}_{ij}^{*}(T_{ij})\mathpzc{q}_{ij}^{*}(T_{ij})\right] \\
&\quad - \int_{0}^{T_{ij}} Z_{i(j)} \otimes \boldsymbol{B}(s)\mathpzc{h}_{0}^{*}(s)\mathcal{E}_{ij}^{*}\alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s) \times \\
&\quad \times \left[\left\{\alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s) + \mathpzc{q}_{ij}^{*}(s) - \mathpzc{m}_{ij}^{*}(s)\right\}^{2}\right. \\
&\quad \left. + \mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s)\left\{\alpha^{*}(s)\mathpzc{q}_{ij}^{*}(s) - \alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s) - 2\right\}\right] ds \\
&\quad - \sum_{k=1}^{n_{ij}} Z_{i(j)\nu} \otimes \boldsymbol{B}(t_{ijk})\left\{\mathpzc{m}_{ijk}\mathpzc{q}_{ijk}^{*2} - \mathpzc{m}_{ijk}^{*2}\mathpzc{q}_{ijk}^{*}\right\} \\
\frac{\partial \Delta_{ij}^{(2)*}}{\partial \psi_{X_{\omega}}} &= -\int_{0}^{T_{ij}} X_{ij\omega}^{\mathrm{T}} \otimes \boldsymbol{B}(s)\mathpzc{h}_{0}^{*}(s)\mathcal{E}_{ij}^{*}\alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s) \\
&\quad \times \left\{\alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s) + \mathpzc{q}_{ij}^{*}(s) - \mathpzc{m}_{ij}^{*}(s)\right\} ds \\
\frac{\partial \Delta_{ij}^{(2)*}}{\partial \psi_{Z_{\nu}}} &= -\int_{0}^{T_{ij}} Z_{i(j)\nu}^{\mathrm{T}} \otimes \boldsymbol{B}(s)\mathpzc{h}_{0}^{*}(s)\mathcal{E}_{ij}^{*}\alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s) \\
&\quad \times \left\{\alpha^{*}(s)\mathpzc{m}_{ij}^{*}(s)\mathpzc{q}_{ij}^{*}(s) + \mathpzc{q}_{ij}^{*}(s) - \mathpzc{m}_{ij}^{*}(s)\right\} ds,
\end{aligned}
$$

A.2 Random-Coefficient: Models

A.2.1 Derivatives to obtain the posterior modes

The derivatives with the respect subject- and facility-level random effects required for Section 3.2.2 are computed by replacing the VCMs from the derivatives in Section A.1.1 with random-coefficient spline models. The score functions $\mathcal{G}_{\vartheta}$ is defined as

$$\mathcal{G}_{\vartheta} = -\left(\mathcal{G}^{\mathrm{T}}_{\varphi_\alpha}, \mathcal{G}^{\mathrm{T}}_{\varphi_\hbar}, \mathcal{G}^{\mathrm{T}}_{\varphi_{X_1}}, \ldots, \mathcal{G}^{\mathrm{T}}_{\varphi_{X_p}}, \mathcal{G}^{\mathrm{T}}_{\varphi_{Z_1}}, \ldots, \mathcal{G}^{\mathrm{T}}_{\varphi_{Z_q}}, \mathcal{G}^{\mathrm{T}}_{\varrho_{X_1}}, \ldots, \mathcal{G}^{\mathrm{T}}_{\varrho_{X_p}}, \mathcal{G}^{\mathrm{T}}_{\varrho_{Z_1}}, \ldots, \mathcal{G}^{\mathrm{T}}_{\varrho_{Z_q}}\right)^{\mathrm{T}},$$

where $\mathcal{G}_{\varphi_\alpha} = \left\{\frac{\partial\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\boldsymbol{\varphi}_{\alpha 1}}, \ldots, \frac{\partial\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi_{\alpha n_\kappa}}\right\}^{\mathrm{T}}$, $\mathcal{G}_{\varphi_\hbar} = \left\{\frac{\partial\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi_{\hbar 1}}, \ldots, \frac{\partial\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi_{\hbar n_\kappa}}\right\}^{\mathrm{T}}$, $\mathcal{G}_{\varphi_{\beta_\eta}} = \left\{\frac{\partial\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi_{\beta_\eta 1}}, \ldots, \frac{\partial\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi_{\beta_\eta n_\kappa}}\right\}^{\mathrm{T}}$, and $\mathcal{G}_{\varphi_{\gamma_\eta}} = \left\{\frac{\partial\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi_{\gamma_\eta 1}}, \ldots, \frac{\partial\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi_{\gamma_\eta n_\kappa}}\right\}^{\mathrm{T}}$, with

$$\frac{\partial\ell(\boldsymbol{\theta}_{RE};\boldsymbol{\theta})}{\partial\varrho_{\alpha\varkappa}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\delta_{ij}m_{ij}(T_{ij})(T_{ij}-\kappa_\varkappa)_+ - \int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}m_{ij}(s)(s-\kappa_\varkappa)_+ ds\right\} - \frac{\varrho_{\alpha\varkappa}}{\sigma^2_\alpha},$$

$$\frac{\partial\ell(\boldsymbol{\theta}_{RE};\boldsymbol{\theta})}{\partial\varrho_{\hbar\varkappa}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\delta_{ij}(T_{ij}-\kappa_\varkappa)_+ - \int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}(s-\kappa_\varkappa)_+ ds\right\} - \frac{\varrho_{\hbar\varkappa}}{\sigma^2_\hbar},$$

$$\begin{aligned}\frac{\partial\ell(\boldsymbol{\theta}_{RE};\boldsymbol{\theta})}{\partial\varphi_{X_\omega\varkappa}} = & \sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{ij\omega}\Bigg(\Bigg\{\delta_{ij}\alpha(T_{ij})m_{ij}(T_{ij})q_{ij}(T_{ij})(T_{ij}-\kappa_\varkappa)_+ \\ & - \int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}\alpha(s)m_{ij}(s)q_{ij}(s)(s-\kappa_\varkappa)_+ ds\Bigg\} \\ & + \left[\sum_{k=1}^{n_{ij}}\{Y_{ijk}(t_{ijk}-\kappa_\varkappa)_+ - m_{ijk}(t_{ijk}-\kappa_\varkappa)_+\}\right]\Bigg) \\ & - \frac{\varphi_{X_\omega\varkappa}}{\sigma^2_{\varphi_{x_\omega}}},\end{aligned}$$

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta}_{RE};\boldsymbol{\theta})}{\partial \varphi_{Z_\nu \varkappa}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} Z_{i(j)\nu}\Bigg(\Bigg\{\delta_{ij}\alpha(T_{ij})m_{ij}(T_{ij})q_{ij}(T_{ij})(T_{ij}-\kappa_\varkappa)_+ \\ &\quad - \int_0^{T_{ij}} h_0(s)\mathcal{E}_{ij}\alpha(s)m_{ij}(s)q_{ij}(s)(s-\kappa_\varkappa)_+ ds\Bigg\} \\ &\quad + \left[\sum_{k=1}^{n_{ij}}\{Y_{ijk}(t_{ijk}-\kappa_\varkappa)_+ - m_{ijk}(t_{ijk}-\kappa_\varkappa)_+\}\right]\Bigg) \\ &\quad - \frac{\varphi_{Z_\nu \varkappa}}{\sigma^2_{\varphi_{z_\nu}}},\end{aligned}$$

$$\frac{\partial \ell(\theta_{RE};\boldsymbol{\theta})}{\partial \varrho_{X_\omega \varkappa}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{ij\omega}\left\{\delta_{ij}(T_{ij}-\kappa_\varkappa)_+ - \int_0^{T_{ij}} h_0(s)\mathcal{E}_{ij}(s-\kappa_\varkappa)_+ ds\right\} - \frac{\varrho_{X_\Omega \varkappa}}{\sigma^2_{x_\nu}}.$$

and

$$\frac{\partial \ell(\theta_{RE};\boldsymbol{\theta})}{\partial \varrho_{Z_\omega \varkappa}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} Z_{i(j)\omega}\left\{\delta_{ij}(T_{ij}-\kappa_\varkappa)_+ - \int_0^{T_{ij}} h_0(s)\mathcal{E}_{ij}(s-\kappa_\varkappa)_+ ds\right\} - \frac{\varrho_{Z_\Omega \varkappa}}{\sigma^2_{z_\nu}}.$$

Due to the random effects in S-RE being independent of each other, the Hessian matrix $\mathcal{H}_\vartheta$ can be written as

$$\mathcal{H}_\vartheta = \boldsymbol{\Sigma}_\vartheta = -\left\{\begin{array}{cccccc} \mathcal{H}_{\varrho_\alpha} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{H}_{\varrho_h} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{H}_{\varphi_X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{H}_{\varphi_Z} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{H}_{\varrho_X} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathcal{H}_{\varrho_Z} \end{array}\right\},$$

where $\boldsymbol{\mathscr{H}}_{\varrho_\alpha} = \text{diag}\left\{\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varrho_{\alpha 1}^2}, \ldots, \frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varrho_{\alpha n_\kappa}^2}\right\}$, $\boldsymbol{\mathscr{H}}_{\varrho_\hbar} = \text{diag}\left\{\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varrho_{\hbar 1}^2}, \ldots, \frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varrho_{\hbar n_\kappa}^2}\right\}$,

$$\boldsymbol{\mathscr{H}}_{\varphi_X} = \begin{Bmatrix} \boldsymbol{\mathscr{H}}_{\varphi_{X_1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathscr{H}}_{\varphi_{X_2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathscr{H}}_{\varphi_{X_p}} \end{Bmatrix},$$

$$\boldsymbol{\mathscr{H}}_{\varphi_Z} = \begin{Bmatrix} \boldsymbol{\mathscr{H}}_{\varphi_{Z_1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathscr{H}}_{\varphi_{Z_2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathscr{H}}_{\varphi_{Z_q}} \end{Bmatrix},$$

$$\boldsymbol{\mathscr{H}}_{\varrho_X} = \begin{Bmatrix} \boldsymbol{\mathscr{H}}_{\varrho_{X_1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathscr{H}}_{\varrho_{X_2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathscr{H}}_{\varrho_{X_q}} \end{Bmatrix},$$

and

$$\boldsymbol{\mathscr{H}}_{\varrho_Z} = \begin{Bmatrix} \boldsymbol{\mathscr{H}}_{\varrho_{Z_1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathscr{H}}_{\varrho_{Z_2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathscr{H}}_{\varrho_{Z_q}} \end{Bmatrix},$$

with $\boldsymbol{\mathcal{H}}_{\varphi_{X_\omega}} = \text{diag}\left\{\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi^2_{X_\omega 1}},\ldots,\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi^2_{X_\omega n_\kappa}}\right\}$, $\boldsymbol{\mathcal{H}}_{\varphi_{Z_\nu}} = \text{diag}\left\{\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi^2_{Z_\nu 1}},\ldots,\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi^2_{Z_\nu n_\kappa}}\right\}$, $\boldsymbol{\mathcal{H}}_{\varrho_{X_\omega}} = \text{diag}\left\{\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varrho^2_{X_\omega 1}},\ldots,\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varrho^2_{X_\omega n_\kappa}}\right\}$ and $\boldsymbol{\mathcal{H}}_{\varrho_{X_\omega}} = \text{diag}\left\{\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varrho^2_{Z_\nu 1}},\ldots,\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varrho^2_{Z_\nu n_\kappa}}\right\}$.
The derivatives in the Hessian matrix $\boldsymbol{\mathcal{H}}_{\boldsymbol{\theta}_\varphi}$ are defined as

$$\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varrho^2_{\alpha\varkappa}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{-\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}m^2_{ij}(s)(s-\kappa_\varkappa)^2_+ ds\right\} - \frac{1}{\sigma^2_\alpha},$$

$$\frac{\partial^2\ell(\boldsymbol{\theta}_{RE};\boldsymbol{\theta})}{\partial\varrho^2_{\hbar\varkappa}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{-\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}(s-\kappa_\vartheta)^2_+ ds\right\} - \frac{1}{\sigma^2_\hbar},$$

$$\begin{aligned}\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi^2_{X_\omega\varkappa}} = & \sum_{i=1}^{n}\sum_{j=1}^{n_i}X^2_{ij\omega}\Bigg(\Bigg[\delta_{ij}\alpha(T_{ij})m_{ij}(T_{ij})q_{ij}(T_{ij})\{q_{ij}(T_{ij})-m_{ij}(T_{ij})\}(T_{ij}-\kappa_\varkappa)^2_+ \\ & -\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}\alpha(s)m_{ij}(s)q_{ij}(s)\{\alpha(s)m_{ij}(s)q_{ij}(s)+q_{ij}(s)-m_{ij}(s)\}(s-\kappa_\varkappa)^2_+ ds\Bigg] \\ & + \left[\sum_{k=1}^{n_{ij}}\{-m_{ijk}q_{ijk}(t_{ijk}-\kappa_\varkappa)_+\}\right]\Bigg) - \frac{1}{\sigma^2_{\varphi_{x_\omega}}},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varphi^2_{Z_\nu\varkappa}} = & \sum_{i=1}^{n}\sum_{j=1}^{n_i}Z^2_{i(j)\nu}\Bigg(\Bigg[\delta_{ij}\alpha(T_{ij})m_{ij}(T_{ij})q_{ij}(T_{ij})\{q_{ij}(T_{ij})-m_{ij}(T_{ij})\}(T_{ij}-\kappa_\varkappa)^2_+ \\ & -\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}\alpha(s)m_{ij}(s)q_{ij}(s)\{\alpha(s)m_{ij}(s)q_{ij}(s)+q_{ij}(s)-m_{ij}(s)\}(s-\kappa_\varkappa)^2_+ ds\Bigg] \\ & + \left[\sum_{k=1}^{n_{ij}}\{-m_{ijk}q_{ijk}(t_{ijk}-\kappa_\varkappa)_+\}\right]\Bigg) - \frac{1}{\sigma^2_{\varphi_{z_\nu}}},\end{aligned}$$

$$\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varrho^2_{X_\omega\varkappa}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}X^2_{ij\omega}\left\{-\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}(s-\kappa_\varkappa)^2_+ ds\right\} - \frac{1}{\sigma^2_{\varrho_{x_\omega}}},$$

and

$$\frac{\partial^2\ell(\boldsymbol{\theta}_{RE},\boldsymbol{\theta})}{\partial\varrho^2_{Z_\nu\varkappa}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}Z^2_{i(j)\nu}\left\{-\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}(s-\kappa_\varkappa)^2_+ ds\right\} - \frac{1}{\sigma^2_{\varrho_{z_\nu}}}.$$

A.2.2 Derivatives for Correction Terms

The correction terms in 3.13 require the following derivatives: $\frac{\partial \boldsymbol{\Sigma}_\varphi}{\partial \boldsymbol{\vartheta}^{\mathrm{T}}}$, $\frac{\partial^2 \boldsymbol{\Sigma}}{\partial \boldsymbol{\vartheta}^{\mathrm{T}} \partial \boldsymbol{\vartheta}}$, $\frac{\partial \hat{\boldsymbol{\vartheta}}^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}}$, and $\frac{\partial^2 \hat{\boldsymbol{\vartheta}}^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial \boldsymbol{c}}$. The first derivative of $\boldsymbol{\Sigma}_\vartheta$ with respect to $\boldsymbol{\vartheta}^{\mathrm{T}}$ is a $\{(p+q+2)n_\kappa\}^2 \times \{(p+q+2)n_\kappa\}$ matrix from the concatenation of $n_\kappa \times n_\kappa$ matrices representing the the derivatice of $\boldsymbol{\Sigma}_\varphi$ with respect to an individual S-RE. The second derivative of $\boldsymbol{\Sigma}_\vartheta$ with respect to $\boldsymbol{\vartheta}^{\mathrm{T}}$ and $\boldsymbol{\vartheta}$ is a $\{(p+q+2)n_\kappa\}^2 \times \{(p+q+2)n_\kappa\}^2$ matrix composed of the concatenation of $n_\kappa \times n_\kappa$ matrices representing the second derivative of $\boldsymbol{\Sigma}_\vartheta$ with respect to an individual S-RE. The remaining sections will derivatives for the corrections terms with respect the VCM's $\alpha(t)$, $\hbar_0(t)$, $\boldsymbol{\beta}_X(t)$, $\boldsymbol{\beta}_Z(t)$, $\boldsymbol{\gamma}_X(t)$, and $\boldsymbol{\gamma}_Z(t)$.

Correction Terms for $\alpha(t)$

The first derivative of $\hat{\boldsymbol{\varrho}}_\alpha^{(c)}$ with respect to $\boldsymbol{c}$ is a $n \times n$ matrix denoted as

$$\frac{\partial \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}} = \begin{pmatrix} \frac{\partial \hat{\varrho}_{\alpha 1}^{(c)}}{\partial c_1} & \cdots & \frac{\partial \hat{\varrho}_{\alpha 1}^{(c)}}{\partial c_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial \hat{\varrho}_{\alpha n_\kappa}^{(c)}}{\partial c_1} & \cdots & \frac{\partial \hat{\varrho}_{\alpha n_\kappa}^{(c)}}{\partial c_n} \end{pmatrix},$$

and the second derivative of $\hat{\varrho}_\alpha^{(c)}$ with respect to $\boldsymbol{c}$ is a $n_\kappa^2 \times n_\kappa$ matrix denoted as

$$\frac{\partial^2 \hat{\varrho}_\alpha^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial \boldsymbol{c}} = \begin{pmatrix} \frac{\partial^2 \hat{\varrho}_\alpha^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial c_1} \\ \frac{\partial^2 \hat{\varrho}_\alpha^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial c_2} \\ \vdots \\ \frac{\partial^2 \hat{\varrho}_\alpha^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial c_n} \end{pmatrix}.$$

The first and second partial derivatives for $\hat{\varrho}_\alpha$ with respect to $\boldsymbol{c}$ as $\frac{\partial \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}}|_{\boldsymbol{c}=\mathbf{0}} = \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{-1}$ and $\frac{\partial^2 \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial \boldsymbol{c}}|_{\boldsymbol{c}=\mathbf{0}} = \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{-1}(-\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}/\partial \boldsymbol{\varrho}_\alpha^{\mathrm{T}})\boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{-1}\boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{-1}$, respectively, where $\boldsymbol{\mathcal{H}}_{\varrho_\alpha} = \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}|_{\boldsymbol{c}=\mathbf{0}}$.

The partial derivatives $\frac{\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}}$ and $\frac{\partial^2 \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial \boldsymbol{c}}$ are obtained using the chain rule:

$$
\begin{aligned}
\left.\frac{\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}}\right|_{(\boldsymbol{c},\boldsymbol{\varrho}_\alpha)=(0,\hat{\boldsymbol{\varrho}}_\alpha)} &= \left.\frac{\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \hat{\boldsymbol{\varrho}}_\alpha^{(c)}} \frac{\partial \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}}\right|_{(\boldsymbol{c},\boldsymbol{\varrho}_\alpha)=(\mathbf{0},\hat{\boldsymbol{\varrho}}_\alpha)} = \left.\frac{\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{\varrho}_\alpha^{T}} \frac{\partial \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}}\right|_{\boldsymbol{c}=\mathbf{0}}, \\
\left.\frac{\partial^2 \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial \boldsymbol{c}}\right|_{(\boldsymbol{c},\boldsymbol{\varrho}_\alpha)=(\mathbf{0},\hat{\boldsymbol{\varrho}}_\alpha)} &= \frac{\partial^2 \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{\varrho}_\alpha^{\mathrm{T}} \partial \boldsymbol{\varrho}_\alpha} \left.\left(\frac{\partial \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial \boldsymbol{c}} \frac{\partial \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}}}\right)\right|_{\boldsymbol{c}=0} + \left.\frac{\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{\varrho}_\alpha^{\mathrm{T}}} \frac{\partial^2 \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial \boldsymbol{c}^{\mathrm{T}} \partial \boldsymbol{c}}\right|_{\boldsymbol{c}=0}.
\end{aligned}
$$

The first derivative of $\boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}$ with respect to $\boldsymbol{c}$ is a $n_\kappa^2 \times n_\kappa$ matrix composed of concatenated $n_\kappa \times n_\kappa$ matrices obtained from $\frac{\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial c_\varkappa}$ for $\varkappa \in \{1, ..., n_\kappa\}$, where the matrix $\frac{\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial c_\varkappa} = \frac{\boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{\varrho}_\alpha} \frac{\partial \boldsymbol{\varrho}_\alpha^{(c)}}{\partial c_\varkappa} = \sum_{\varkappa'=1}^{n_\kappa} \frac{\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \varrho_{\alpha\varkappa'}} \frac{\partial \varrho_{\alpha\varkappa'}^{(c)}}{\partial c_\varkappa}$. The second derivative of $\boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}$ with respect to $\boldsymbol{c}$ is a $n_\kappa^2 \times n_\kappa^2$ matrix composed of n_κ^2 matrices representing $\frac{\partial^2 \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial c_\varkappa \partial c_{\varkappa'}}$ of size $n_\kappa \times n_\kappa$ for $\varkappa, \varkappa' \in \{1, ..., n_\kappa\}$, where $\frac{\partial^2 \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial c_\varkappa \partial c_{\varkappa'}}$ is formulated as

$$
\frac{\partial^2 \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{\varrho}_\alpha^{\mathrm{T}} \partial \boldsymbol{\varrho}_\alpha} \left.\left\{\frac{\partial \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial c_\varkappa} \frac{\partial \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial c_{\varkappa'}}\right\}\right|_{\boldsymbol{c}=\mathbf{0}} + \left.\frac{\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{\varrho}_\alpha^{\mathrm{T}}} \frac{\partial^2 \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial c_\varkappa \partial c_{\varkappa'}}\right|_{\boldsymbol{c}=\mathbf{0}}. \tag{A.4}
$$

The components in (A.4) are obtained as

$$
\frac{\partial^2 \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{\varrho}_\alpha^{\mathrm{T}} \partial \boldsymbol{\varrho}_\alpha} \left.\left\{\frac{\partial \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial c_\varkappa} \frac{\partial \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial c_{\varkappa'}}\right\}\right|_{\boldsymbol{c}=\mathbf{0}} = \sum_{\varkappa^*=1}^{n_\kappa} \left\{\frac{\partial^2 \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \varrho_{\alpha\varkappa^*}^2} \frac{\partial \hat{\varrho}_{\alpha\varkappa^*}^{(c)}}{\partial c_\varkappa} \frac{\partial \hat{\varrho}_{\alpha\varkappa^*}^{(c)}}{\partial c_{\varkappa'}}\right\}
$$

and

$$
\left.\frac{\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}^{(c)}}{\partial \boldsymbol{\varrho}_\alpha^{\mathrm{T}}} \frac{\partial^2 \hat{\boldsymbol{\varrho}}_\alpha^{(c)}}{\partial c_\varkappa \partial c_{\varkappa'}}\right|_{\boldsymbol{c}=\mathbf{0}} = \sum_{\varkappa^*=1}^{n_\kappa} \left.\left\{\frac{\partial \boldsymbol{\mathcal{H}}_{\varrho_\alpha}}{\partial \varrho_{\alpha\varkappa^*}} \frac{\partial^2 \hat{\varrho}_{\alpha\varkappa^*}^{(c)}}{\partial c_\varkappa \partial c_{\varkappa'}}\right\}\right|_{\boldsymbol{c}=\mathbf{0}}.
$$

The formulation for $\frac{\partial \mathcal{H}_{\varrho_\alpha}}{\partial \varrho_{\alpha\varkappa}}$ is $n_\kappa \times n_\kappa$ matrix with the $(\varkappa, \varkappa)^{\text{th}}$ element being the only nonzero element expressed as

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{-\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}m_{ij}^3(s)(s-\kappa_\varkappa)_+^3 ds\right\}.$$

The formulation of $\frac{\partial^2 \mathcal{H}^{(c)}_{\varrho_\alpha}}{\partial \boldsymbol{\varrho}_\alpha^{\text{T}} \partial \boldsymbol{\varrho}_\alpha}$ is a $n_\kappa^2 \times n_\kappa^2$ matrix composed of n_κ^2 block matrices. Each block matrix is a $n_\kappa \times n_\kappa$ matrix with the $(\varkappa, \varkappa)^{\text{th}}$ element being the only nonzero element expressed as

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{-\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}m_{ij}^4(s)(s-\kappa_\varkappa)_+^4 ds\right\}.$$

Correction Terms for $\hbar_0(t)$

The correction terms for $\hbar_0(t)$ can be obtained the same way as the correction terms for $\alpha(t)$ as referred in A.2.2. The formulation for $\frac{\partial \mathcal{H}_{\varrho_\hbar}}{\partial \varrho_{\hbar\varkappa}}$ is $n_\kappa \times n_\kappa$ matrix with the $(\varkappa, \varkappa)^{\text{th}}$ element being the only nonzero element expressed as

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{-\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}(s-\kappa_\varkappa)_+^{3p} ds\right\}$$

The formulation of $\frac{\partial^2 \mathcal{H}^{(c)}_{\varrho_\hbar}}{\partial \boldsymbol{\varrho}_\hbar^{\text{T}} \partial \boldsymbol{\varrho}_\hbar}$ is a $n_\kappa^2 \times n_\kappa^2$ matrix composed of n_κ^2 block matrices. Each block matrix is a $n_\kappa \times n_\kappa$ matrix with the $(\varkappa, \varkappa)^{\text{th}}$ element being the only nonzero element expressed as

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{-\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}(s-\kappa_\varkappa)_+^{4p}ds\right\}$$

Correction Terms for $\beta_{X_\omega}(t)$

The correction terms for $\beta_{X_\omega}(t)$ can be obtained the same way as the correction terms for $\alpha(t)$ as referred in A.2.2. The formulation for $\frac{\partial \mathcal{H}_{\varphi_{X_\omega}}}{\partial \varphi_{X_\omega\varkappa}}$ is $n_\kappa \times n_\kappa$ matrix with the $(\varkappa,\varkappa)^{\text{th}}$ element being the only nonzero element expressed as

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{ij\omega}^3\Bigg(\delta_{ij}\alpha(T_{ij})m_{ij}(T_{ij})q_{ij}(T_{ij})\left[\{q_{ij}(T_{ij})-m_{ij}(T_{ij})\}^2-2m_{ij}(T_{ij})q_{ij}(T_{ij})\right](T_{ij}-\kappa_\varkappa)_+^3\\&-\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}\alpha(s)m_{ij}(s)q_{ij}(s)(s-\kappa_\varkappa)_+^3\times\\&\times\left[\left\{\alpha(s)m_{ij}(s)q_{ij}(s)+q_{ij}(s)-m_{ij}(s)\right\}^2+m_{ij}(s)q_{ij}(s)\left\{\alpha(s)q_{ij}(s)-\alpha(s)m_{ij}(s)-2\right\}\right]ds\\&+\left[\sum_{k=1}^{n_{ij}}\{-m_{ijk}q_{ijk}\left(q_{ijk}-m_{ijk}\right)(t_{ijk}-\kappa_\varkappa)_+^3\}\right]\Bigg)\end{aligned}$$

The formulation of $\frac{\partial^2 \mathcal{H}^{(c)}_{\varphi_{X_\omega}}}{\partial \boldsymbol{\varphi}_{X_\omega}^{\mathrm{T}}\partial \boldsymbol{\varphi}_{X_\omega}}$ is a $n_\kappa^2 \times n_\kappa^2$ matrix composed of n_κ^2 block matrices. Each block matrix is a $n_\kappa \times n_\kappa$ matrix with the $(\varkappa,\varkappa)^{\text{th}}$ element being the only nonzero element expressed as

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{ij\omega}^4\Bigg(\delta_{ij}\alpha(T_{ij})m_{ij}(T_{ij})q_{ij}(T_{ij})\\&\times\left[q_{ij}^3(T_{ij})-11m_{ij}(T_{ij})q_{ij}^2(T_{ij})-11m_{ij}^2(T_{ij})q_{ij}(T_{ij})-m_{ij}^3(T_{ij})\right](T_{ij}-\kappa_\varkappa)_+^4\\&-\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}\alpha(s)m_{ij}(s)q_{ij}(s)(s-\kappa_\varkappa)_+^4\left[\left\{\alpha(s)m_{ij}(s)q_{ij}(s)+q_{ij}(s)-m_{ij}(s)\right\}^3\right.\\&+m_{ij}(s)q_{ij}(s)\left\{\alpha(s)q_{ij}(s)-\alpha(s)m_{ij}(s)-2\right\}\left\{3\alpha(s)m_{ij}(s)q_{ij}(s)+4q_{ij}(s)-4m_{ij}(s)\right\}\\&\left.-2\alpha(s)m_{ij}^2(s)q_{ij}^2(s)\right]ds+\sum_{k=1}^{n_{ij}}\left[-m_{ijk}q_{ijk}\left\{\left(q_{ijk}-m_{ijk}\right)^2-2m_{ijk}q_{ijk}\right\}(t_{ijk}-\kappa_\varkappa)_+^4\right]\Bigg)\end{aligned}$$

Correction Terms for $\beta_{Z_\nu}(t)$

The correction terms for $\beta_{Z_\nu}(t)$ can be obtained the same way as the correction terms for $\alpha(t)$ as referred in A.2.2. The formulation for $\frac{\partial \mathcal{H}_{\varphi_{Z_\nu}}}{\partial \varphi_{Z_\nu \varkappa}}$ is $n_\kappa \times n_\kappa$ matrix with the $(\varkappa, \varkappa)^{\text{th}}$ element being the only nonzero element expressed as

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} Z_{i(j)\nu}^3 \Bigg(\delta_{ij}\alpha(T_{ij}) m_{ij}(T_{ij}) q_{ij}(T_{ij}) \left[\{q_{ij}(T_{ij}) - m_{ij}(T_{ij})\}^2 - 2m_{ij}(T_{ij})q_{ij}(T_{ij})\right](T_{ij}-\kappa_\varkappa)_+^3$$
$$- \int_0^{T_{ij}} h_0(s)\mathcal{E}_{ij}\alpha(s)m_{ij}(s)q_{ij}(s)(s-\kappa_\varkappa)_+^3 \times$$
$$\times \left[\left\{\alpha(s)m_{ij}(s)q_{ij}(s) + q_{ij}(s) - m_{ij}(s)\right\}^2 + m_{ij}(s)q_{ij}(s)\left\{\alpha(s)q_{ij}(s) - \alpha(s)m_{ij}(s) - 2\right\}\right] ds$$
$$+ \left[\sum_{k=1}^{n_{ij}} \left\{-m_{ijk}q_{ijk}\left(q_{ijk} - m_{ijk}\right)(t_{ijk}-\kappa_\varkappa)_+^3\right\}\right]\Bigg)$$

The formulation of $\frac{\partial^2 \mathcal{H}^{(c)}_{\varphi_{Z_\nu}}}{\partial \boldsymbol{\varphi}_{Z_\nu}^{\mathrm{T}} \partial \boldsymbol{\varphi}_{Z_\nu}}$ is a $n_\kappa^2 \times n_\kappa^2$ matrix composed of n_κ^2 block matrices. Each block matrix is a $n_\kappa \times n_\kappa$ matrix with the $(\varkappa, \varkappa)^{\text{th}}$ element being the only nonzero element expressed as

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} Z_{i(j)\nu}^4 \Bigg(\delta_{ij}\alpha(T_{ij}) m_{ij}(T_{ij}) q_{ij}(T_{ij})$$
$$\times \left[q_{ij}^3(T_{ij}) - 11m_{ij}(T_{ij})q_{ij}^2(T_{ij}) - 11m_{ij}^2(T_{ij})q_{ij}(T_{ij}) - m_{ij}^3(T_{ij})\right](T_{ij}-\kappa_\varkappa)_+^4$$
$$- \int_0^{T_{ij}} h_0(s)\mathcal{E}_{ij}\alpha(s)m_{ij}(s)q_{ij}(s)(s-\kappa_\varkappa)_+^4 \left[\left\{\alpha(s)m_{ij}(s)q_{ij}(s) + q_{ij}(s) - m_{ij}(s)\right\}^3\right.$$
$$+m_{ij}(s)q_{ij}(s)\left\{\alpha(s)q_{ij}(s) - \alpha(s)m_{ij}(s) - 2\right\}\left\{3\alpha(s)m_{ij}(s)q_{ij}(s) + 4q_{ij}(s) - 4m_{ij}(s)\right\}$$
$$\left.-2\alpha(s)m_{ij}^2(s)q_{ij}^2(s)\right] ds + \sum_{k=1}^{n_{ij}}\left[-m_{ijk}q_{ijk}\left\{\left(q_{ijk} - m_{ijk}\right)^2 - 2m_{ijk}q_{ijk}\right\}(t_{ijk}-\kappa_\varkappa)_+^4\right]\Bigg)$$

Correction Terms for $\gamma_{X_\omega}(t)$

The correction terms for $\gamma_{X_\omega}(t)$ can be obtained the same way as the correction terms for $\alpha(t)$ as referred in A.2.2. The formulation for $\frac{\partial \mathcal{H}_{\varrho_{X_\omega}}}{\partial \varrho_{X_\omega \varkappa}}$ is $n_\kappa \times n_\kappa$ matrix with the $(\varkappa, \varkappa)^{\text{th}}$ element being the only nonzero element expressed as

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{ij\omega}^3 \Big\{ -\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}(s-\kappa_\varkappa)_+^3 ds \Big\}$$

The formulation of $\frac{\partial^2 \mathcal{H}^{(c)}_{\varrho_{X_\omega}}}{\partial \boldsymbol{\varrho}_{X_\omega}^{\mathrm{T}} \partial \boldsymbol{\varrho}_{X_\omega}}$ is a $n_\kappa^2 \times n_\kappa^2$ matrix composed of n_κ^2 block matrices. Each block matrix is a $n_\kappa \times n_\kappa$ matrix with the $(\varkappa, \varkappa)^{\text{th}}$ element being the only nonzero element expressed as

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{ij\omega}^4 \Big\{ -\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}(s-\kappa_\varkappa)_+^4 ds \Big\}$$

Correction Terms for $\gamma_{Z_\nu}(t)$

The correction terms for $\gamma_{Z_\nu}(t)$ can be obtained the same way as the correction terms for $\alpha(t)$ as referred in A.2.2. The formulation for $\frac{\partial \mathcal{H}_{\varrho_{Z_\nu}}}{\partial \varrho_{Z_\nu \varkappa}}$ is $n_\kappa \times n_\kappa$ matrix with the $(\varkappa, \varkappa)^{\text{th}}$ element being the only nonzero element expressed as

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} Z_{i(j)\nu}^3 \Big\{ -\int_0^{T_{ij}} \hbar_0(s)\mathcal{E}_{ij}(s-\kappa_\varkappa)_+^3 ds \Big\}$$

The formulation of $\frac{\partial^2 \mathscr{H}^{(c)}_{\varrho_{Z_\nu}}}{\partial \boldsymbol{\varrho}^{\mathrm{T}}_{Z_\nu} \partial \boldsymbol{\varrho}_{Z_\nu}}$ is a $n_\kappa^2 \times n_\kappa^2$ matrix composed of n_κ^2 block matrices. Each block matrix is a $n_\kappa \times n_\kappa$ matrix with the $(\varkappa, \varkappa)^{\mathrm{th}}$ element being the only nonzero element expressed as

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} Z^4_{i(j)\nu} \Big\{ - \int_0^{T_{ij}} \hbar_0(s) \mathcal{E}_{ij} (s - \kappa_\varkappa)^4_+ ds \Big\}$$

A.2.3 M-Step

The score functions for the incomplete likelihood function with respect to σ^2_α, $\sigma^2_\hbar$, $\sigma^2_{\varphi_{x_\omega}}$, $\sigma^2_{\varphi_{z_\nu}}$, $\sigma^2_{\varrho_{x_\omega}}$ and $\sigma^2_{\varrho_{z_\nu}}$ are provided as

$$\begin{aligned}
V(\sigma^2_\alpha) &= \frac{\partial}{\partial \sigma^2_\alpha} \log \Big\{ \int L(\boldsymbol{\varrho}_\alpha; \boldsymbol{u}, \boldsymbol{\theta}_{\mathcal{R}}) d\boldsymbol{\varrho}_\alpha \Big\} = \int \sum_{\varkappa=1}^{n_\kappa} \left(\frac{\varrho^2_{\alpha\vartheta}}{2\sigma^4_\alpha} - \frac{1}{\sigma^2_\alpha} \right) \mathcal{L}_\vartheta (\boldsymbol{\varrho}_{\alpha\varkappa}) d\boldsymbol{\varrho}_{\alpha\varkappa} = V_\vartheta(\sigma^2_\alpha), \\
V(\sigma^2_\hbar) &= \frac{\partial}{\partial \sigma^2_\hbar} \log \Big\{ \int L(\boldsymbol{\varrho}_\hbar; \boldsymbol{u}, \boldsymbol{\theta}_{\mathcal{R}}) d\boldsymbol{\varrho}_\hbar \Big\} = \int \sum_{\varkappa=1}^{n_\kappa} \left(\frac{\varrho^2_{\hbar\vartheta}}{2\sigma^4_\hbar} - \frac{1}{\sigma^2_\hbar} \right) \mathcal{L}_\vartheta (\boldsymbol{\varrho}_{\hbar\varkappa}) d\boldsymbol{\varrho}_{\hbar\varkappa} = V_\vartheta(\sigma^2_\hbar), \\
V(\sigma^2_{\varphi_{x_\omega}}) &= \frac{\partial}{\partial \sigma^2_{\varphi_{x_\omega}}} \log \Big\{ \int L(\boldsymbol{\varphi}_{X_\omega}; \boldsymbol{u}, \boldsymbol{\theta}_{\mathcal{R}}) d\boldsymbol{\varphi}_{X_\omega} \Big\} = \int \sum_{\varkappa=1}^{n_\kappa} \left(\frac{\varphi^2_{X_\omega\varkappa}}{2\sigma^4_{\varphi_{x_\omega}}} - \frac{1}{\sigma^2_{\varphi_{x_\omega}}} \right) \mathcal{L}_\vartheta (\boldsymbol{\varphi}_{X_\omega}) d\boldsymbol{\varphi}_{X_\omega} \\
&= V_\vartheta(\sigma^2_{\varphi_{x_\omega}}), \\
V(\sigma^2_{\varrho_{x_\omega}}) &= \frac{\partial}{\partial \sigma^2_{\varrho_{x_\omega}}} \log \Big\{ \int L(\boldsymbol{\varrho}_{X_\omega}; \boldsymbol{u}, \boldsymbol{\theta}_{\mathcal{R}}) d\boldsymbol{\varrho}_{X_\omega} \Big\} = \int \sum_{\varkappa=1}^{n_\kappa} \left(\frac{\varrho^2_{X_\omega\varkappa}}{2\sigma^4_{\varrho_{x_\omega}}} - \frac{1}{\sigma^2_{\varrho_{x_\omega}}} \right) \mathcal{L}_\vartheta (\boldsymbol{\varrho}_{X_\omega}) d\boldsymbol{\varrho}_{X_\omega} \\
&= V_\vartheta(\sigma^2_{\varrho_{x_\omega}}), \\
V(\sigma^2_{\varphi_{z_\nu}}) &= \frac{\partial}{\partial \sigma^2_{\varphi_{z_\nu}}} \log \Big\{ \int L(\boldsymbol{\varphi}_{Z_\nu}; \boldsymbol{u}, \boldsymbol{\theta}_{\mathcal{R}}) d\boldsymbol{\varphi}_{Z_\nu} \Big\} = \int \sum_{\varkappa=1}^{n_\kappa} \left(\frac{\varphi^2_{Z_\nu\varkappa}}{2\sigma^4_{\varphi_{z_\nu}}} - \frac{1}{\sigma^2_{\varphi_{z_\nu}}} \right) \mathcal{L}_\vartheta (\boldsymbol{\varphi}_{Z_\nu}) d\boldsymbol{\varphi}_{Z_\nu} \\
&= V_\vartheta(\sigma^2_{\varphi_{z_\nu}}), \text{ and} \\
V(\sigma^2_{\varrho_{z_\nu}}) &= \frac{\partial}{\partial \sigma^2_{\varrho_{z_\nu}}} \log \Big\{ \int L(\boldsymbol{\varrho}_{Z_\nu}; \boldsymbol{u}, \boldsymbol{\theta}_{\mathcal{R}}) d\boldsymbol{\varrho}_{Z_\nu} \Big\} = \int \sum_{\varkappa=1}^{n_\kappa} \left(\frac{\varrho^2_{Z_\nu\varkappa}}{2\sigma^4_{\varrho_{z_\nu}}} - \frac{1}{\sigma^2_{\varrho_{z_\nu}}} \right) \mathcal{L}_\vartheta (\boldsymbol{\varrho}_{Z_\nu}) d\boldsymbol{\varrho}_{Z_\nu} \\
&= V_\vartheta(\sigma^2_{\varrho_{z_\nu}}),
\end{aligned}$$

respectively, where $\mathcal{L}_\vartheta(\boldsymbol{\vartheta}) = L(\boldsymbol{\vartheta}; \boldsymbol{u}, \boldsymbol{\theta}_{\mathcal{R}}) / \int L(\boldsymbol{\vartheta}; \boldsymbol{u}, \boldsymbol{\theta}_{\mathcal{R}}) d\boldsymbol{\vartheta}$ is the posterior density function of $\boldsymbol{\vartheta}$.

The maximum likelihood estimates for the parameters in $\boldsymbol{\theta}_{\mathscr{R}}$ are obtained by maximizing the log-likelihood function (3.14). Similarly to the P-splines approach, computing the Hessian matrix can be burdensome for the M-Step. Therefore, a quasi-Newton method is used with a BFGS update to approximate the Hessian matrix. Approximating the update for the Hessian matrix can be found in Givens and Hoeting (2012). Below are the first derivatives for the M-Step.

Derivatives for $\alpha(t)$

$$\frac{\partial \ell^*(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \rho_{\alpha l}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{ T_{ij}^{1-l} m_{ij}^*(T_{ij}) - \int_0^{T_{ij}} h_0^*(s)\mathcal{E}_{ij}^* s^{1-l} m_{ij}^*(s)ds + \mathcal{R}_{ij}^* \frac{\partial \Delta_{ij}^{(2)*}}{\partial \rho_{\alpha l}} \right\}$$

$$+ \sum_{\varkappa=1}^{n_\kappa}\left[\frac{\varsigma_{\alpha\varkappa}^*}{2} \frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{\alpha\varkappa}}}{\partial \rho_{\alpha l}} + \frac{\varsigma_{h\varkappa}^*}{2} \frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{h\varkappa}}}{\partial \rho_{\alpha l}} \right.$$

$$+ \sum_{\omega=1}^{p}\left\{ \frac{\varsigma_{\varphi_{x_\omega}\varkappa}^*}{2} \frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varphi^2_{X_\omega\varkappa}}}{\partial \rho_{\alpha l}} + \frac{\varsigma_{\varrho_{x_\omega}\varkappa}^*}{2} \frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{X_\omega\vartheta}}}{\partial \rho_{\alpha l}} \right\}$$

$$\left. + \sum_{\nu=1}^{q}\left\{ \frac{\varsigma_{\varphi_{z_\nu}\varkappa}^*}{2} \frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varphi^2_{z_\nu\varkappa}}}{\partial \rho_{\alpha l}} + \frac{\varsigma_{\varrho_{z_\nu}\varkappa}^*}{2} \frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{z_\nu\vartheta}}}{\partial \rho_{\alpha l}} \right\}\right]$$

$$\frac{\partial \Delta_{ij}^{(2)*}}{\partial \rho_{\alpha l}} = \delta_{ij} T_{ij}^{1-l} m_{ij}^*(T_{ij}) q_{ij}^*(T_{ij}) \left\{ q_{ij}^*(T_{ij}) - m_{ij}^*(T_{ij}) \right\}$$

$$- \int_0^{T_{ij}} s^{1-l} h_0^*(s) m_{ij}^*(s) q_{ij}^*(s) \mathcal{E}_{ij}$$

$$\times \left\{ \alpha^{2*}(s) m_{ij}^{2*}(s) q_{ij}^*(s) + 3\alpha^*(s) m_{ij}^*(s) q_{ij}^*(s) - \alpha^*(s) m_{ij}^{2*}(s) + q_{ij}^*(s) - m_{ij}^*(s) \right\} ds$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{\alpha\varkappa}}}{\partial \rho_{\alpha l}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} - \int_0^{T_{ij}} h_0^*(s) \mathcal{E}_{ij}^* m_{ij}^{3*}(s)(s-\kappa_\varkappa)_+^{2p} s^{1-l} ds$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{h\vartheta}}}{\partial \rho_{\alpha l}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} - \int_0^{T_{ij}} h_0^*(s) \mathcal{E}_{ij}^* (s-\kappa_\varkappa)_+^{2} s^{1-l} ds$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varphi^2_{X_\omega\vartheta}}}{\partial \rho_{\alpha l}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{ij\omega}^2 \Big[\delta_{ij} m_{ij}^*(T_{ij}) q_{ij}^*(T_{ij}) \{ q_{ij}^*(T_{ij}) - m_{ij}^*(T_{ij}) \} (T_{ij}-\kappa_\varkappa)_+^2 T_{ij}^{1-l}$$

$$- \int_0^{T_{ij}} h_0^*(s) \mathcal{E}_{ij}^* m_{ij}^*(s) q_{ij}^*(s) \left\{ \alpha^{2*}(s) m_{ij}^{2*}(s) q_{ij}^*(s) \right.$$

$$\left. + 3\alpha^*(s) m_{ij}^*(s) q_{ij}^*(s) - \alpha^*(s) m_{ij}^{2*}(s) + q_{ij}^*(s) - m_{ij}^*(s) \right\} (s-\kappa_\varkappa)_+^2 s^{1-l} ds \Big]$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varphi^2_{Z_\nu\vartheta}}}{\partial \rho_{\alpha l}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} Z_{i(j)\nu}^2 \Big[\delta_{ij} m_{ij}^*(T_{ij}) q_{ij}^*(T_{ij}) \{ q_{ij}^*(T_{ij}) - m_{ij}^*(T_{ij}) \} (T_{ij}-\kappa_\varkappa)_+^2 T_{ij}^{1-l}$$

$$- \int_0^{T_{ij}} h_0^*(s) \mathcal{E}_{ij}^* m_{ij}^*(s) q_{ij}^*(s) \left\{ \alpha^{2*}(s) m_{ij}^{2*}(s) q_{ij}^*(s) \right.$$

$$\left. + 3\alpha^*(s) m_{ij}^*(s) q_{ij}^*(s) - \alpha^*(s) m_{ij}^{2*}(s) + q_{ij}^*(s) - m_{ij}^*(s) \right\} (s-\kappa_\varkappa)_+^2 s^{1-l} ds \Big]$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{X_\omega\vartheta}}}{\partial \rho_{\alpha l}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} - X_{ij\omega}^2 \int_0^{T_{ij}} h_0^*(s) \mathcal{E}_{ij}^* (s-\kappa_\varkappa)_+^{2} s^{1-l} ds$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{Z_\nu\vartheta}}}{\partial \rho_{\alpha l}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} - Z_{i(j)\nu}^2 \int_0^{T_{ij}} h_0^*(s) \mathcal{E}_{ij}^* (s-\kappa_\varkappa)_+^{2} s^{1-l} ds$$

Derivatives for $\hbar_0(t)$

$$
\begin{aligned}
\frac{\partial \ell^*(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \rho_{\hbar l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\delta_{ij}T_{ij}^{1-l} - \int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}s^{1-l}ds + \mathcal{R}^*_{ij}\frac{\partial \Delta_{ij}^{(2)*}}{\partial \rho_{\hbar l}}\right\} \\
&\quad + \sum_{\varkappa=1}^{n_\kappa}\left[\frac{\varsigma^*_{\alpha\varkappa}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{\alpha\varkappa}}}{\partial \rho_{\hbar l}} + \frac{\varsigma^*_{\hbar\varkappa}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{\hbar\vartheta}}}{\partial \rho_{\hbar l}}\right. \\
&\quad + \sum_{\omega=1}^{p}\left\{\frac{\varsigma^*_{\varphi_{x_\omega}\varkappa}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varphi^2_{X_\omega\varkappa}}}{\partial \rho_{\hbar l}} + \frac{\varsigma^*_{\varrho_{x_\omega}\varkappa}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{X_\omega\varkappa}}}{\partial \rho_{\hbar l}}\right\} \\
&\quad \left. + \sum_{\nu=1}^{q}\left\{\frac{\varsigma^*_{\varphi_{z_\nu}\varkappa}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varphi^2_{Z_\nu\varkappa}}}{\partial \rho_{\hbar l}} + \frac{\varsigma^*_{\varrho_{z_\nu}\varkappa}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{Z_\nu\varkappa}}}{\partial \rho_{\hbar l}}\right\}\right] \\
\frac{\partial \Delta_{ij}^{(2)*}}{\partial \rho_{\alpha l}} &= -\int_0^{T_{ij}} s^{1-\varphi}\hbar_0^*(s)\alpha^*(s)\mathscr{m}^*_{ij}(s)\mathscr{q}^*_{ij}(s)\mathcal{E}^*_{ij}\left\{\alpha^*(s)\mathscr{m}^*_{ij}(s)\mathscr{q}^*_{ij}(s) + \mathscr{q}^*_{ij}(s) - \mathscr{m}^*_{ij}(s)\right\}ds \\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{\alpha\varkappa}}}{\partial \rho_{\hbar l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}\mathscr{m}^{2*}_{ij}(s)s^{1-l}(s-\kappa_\varkappa)^2_+ds \\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{\hbar\varkappa}}}{\partial \rho_{\hbar l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}s^{1-l}(s-\kappa_\varkappa)^2_+ds \\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varphi^2_{X_\omega\varkappa}}}{\partial \rho_{\hbar l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -X^2_{ij\omega}\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}\alpha^*(s)\mathscr{m}^*_{ij}(s)\mathscr{q}^*_{ij}(s) \\
&\quad \times\{\alpha^*(s)\mathscr{m}^*_{ij}(s)\mathscr{q}^*_{ij}(s) + \mathscr{q}^*_{ij}(s) - \mathscr{m}^*_{ij}(s)\}s^{1-l}(s-\kappa_\varkappa)^2_+ds \\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varphi^2_{Z_\nu\varkappa}}}{\partial \rho_{\hbar l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -Z^2_{i(j)\nu}\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}\alpha^*(s)\mathscr{m}^*_{ij}(s)\mathscr{q}^*_{ij}(s) \\
&\quad \times\{\alpha^*(s)\mathscr{m}^*_{ij}(s)\mathscr{q}^*_{ij}(s) + \mathscr{q}^*_{ij}(s) - \mathscr{m}^*_{ij}(s)\}s^{1-l}(s-\kappa_\varkappa)^2_+ds \\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{X_\omega\varkappa}}}{\partial \rho_{\hbar l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -X^2_{ij\varkappa}\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}s^{1-l}(s-\kappa_\varkappa)^2_+ds \\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{Z_\nu\varkappa}}}{\partial \rho_{\hbar l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -Z^2_{i(j)\varkappa}\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}s^{1-l}(s-\kappa_\varkappa)^2_+ds.
\end{aligned}
$$

Derivatives for $\beta_{X_{\omega'}}(t)$

$$\frac{\partial \ell^*(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \phi_{X_{\omega'}l}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\Bigg[X_{ij\omega'}\Bigg\{\delta_{ij}\alpha^*(T_{ij})m^*_{ij}(T_{ij})q^*_{ij}(T_{ij})T_{ij}^{1-l} - \int_0^{T_{ij}} h^*_0(s)\mathcal{E}^*_{ij}m^*_{ij}(s)q^*_{ij}(s)s^{1-l}ds$$
$$+\sum_{k=1}^{n_{ij}}\left(Y_{ijk}-m^*_{ijk}\right)t_{ijk}^{1-l}\Bigg\} + \mathcal{R}^*_{ij}\frac{\partial \Delta_{ij}^{(2)*}}{\partial \phi_{X_{\omega'}l}}\Bigg]$$
$$+\sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma^*_{\alpha_\varkappa}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{\alpha\varkappa}}}{\partial \phi_{X_{\omega'}l}} + \frac{\varsigma^*_{h\varkappa}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{h\varkappa}}}{\partial \phi_{X_{\omega'}l}}\right\}$$
$$+\sum_{\omega=1}^{p}\sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma^*_{\varphi_{x_\omega\varkappa}}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varphi^2_{X_\omega\varkappa}}}{\partial \phi_{X_{\omega'}l}} + \frac{\varsigma^*_{\varrho_{x_\omega\varkappa}}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{X_\omega\varkappa}}}{\partial \phi_{X_{\omega'}l}}\right\}$$
$$+\sum_{\nu=1}^{p}\sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma^*_{\varphi_{z_\nu\varkappa}}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varphi^2_{Z_\nu\varkappa}}}{\partial \phi_{X_{\omega'}l}} + \frac{\varsigma^*_{\varrho_{z_\nu\varkappa}}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{Z_\nu\varkappa}}}{\partial \phi_{X_{\omega'}l}}\right\}$$

$$\frac{\partial \Delta_{ij}^{(2)}}{\partial \phi_{X_{\omega'}l}} = X_{ij\omega'}\Bigg(\delta_{ij}\alpha^*(T_{ij})m^*_{ij}(T_{ij})q^*_{ij}(T_{ij})\left[\left\{m^*_{ij}(T_{ij})-q^*_{ij}(T_{ij})\right\}^2 - 2m^*_{ij}(T_{ij})q^*_{ij}(T_{ij})\right]T_{ij}^{1-l}$$
$$-\int_0^{T_{ij}} h^*_0(s)\alpha^*(s)m^*_{ij}(s)q^*_{ij}(s)\mathcal{E}^*_{ij}s^{1-l}\times$$
$$\times\left[\left\{\alpha^*(s)m^*_{ij}(s)q^*_{ij}(s)+q_{ij}(s)-m^*_{ij}(s)\right\}^2\right.$$
$$\left.+m_{ij}(s)q^*_{ij}(s)\left\{\alpha^*(s)q^*_{ij}(s)-\alpha^*(s)m^*_{ij}(s)-2\right\}\right]ds$$
$$-\sum_{k=1}^{n_{ij}}\left\{m^*_{ijk}(q^*_{ijk})^2-(m^*_{ijk})^2q^*_{ijk}\right\}t_{ijk}^{1-l}\Bigg)$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{\alpha\vartheta}}}{\partial \phi_{X_{\omega'}l}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} -X_{ij\omega'}\int_0^{T_{ij}} h^*_0(s)\mathcal{E}^*_{ij}m^*_{ij}(s)\left[\{m^*_{ij}(s)\}^2+2m^*_{ij}(s)q^*_{ij}(s)\right](s-\kappa_\varkappa)_+^{2p}s^{1-l}ds,$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varrho^2_{h\varkappa}}}{\partial \phi_{X_{\omega'}l}} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} -X_{ij\omega'}\int_0^{T_{ij}} h^*_0(s)\mathcal{E}^*_{ij}(s-\kappa_\varkappa)_+^{2p}s^{1-l}ds,$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial \varphi^2_{X_\omega\varkappa}}}{\partial \phi_{X_{\omega'}l}} = X_{ij\omega'}X^2_{ij\omega}\Bigg(\delta_{ij}\alpha^*(T_{ij})m^*_{ij}(T_{ij})q^*_{ij}(T_{ij})$$
$$\times\left[\left\{m^*_{ij}(T_{ij})-q^*_{ij}(T_{ij})\right\}^2-2m^*_{ij}(T_{ij})q^*_{ij}(T_{ij})\right]T_{ij}^{1-l}(T_{ij}-\kappa_\varkappa)_+^{2p}$$
$$-\int_0^{T_{ij}} h^*_0(s)\alpha^*(s)m^*_{ij}(s)q^*_{ij}(s)\mathcal{E}^*_{ij}s^{1-l}s^{1-l}(s-\kappa_\varkappa)_+^{2p}\times$$
$$\times\left[\left\{\alpha^*(s)m^*_{ij}(s)q^*_{ij}(s)+q^*_{ij}(s)-m^*_{ij}(s)\right\}^2\right.$$
$$\left.+m^*_{ij}(s)q^*_{ij}(s)\left\{\alpha^*(s)q_{ij}(s)-\alpha^*(s)m_{ij}(s)-2\right\}\right]ds$$
$$-\sum_{k=1}^{n_{ij}}\left\{m_{ijk}(q^*_{ijk})^2-(m^*_{ijk})^2q_{ijk}\right\}t_{ijk}^{1-l}(s-\kappa_\varkappa)_+^{2p}\Bigg).$$

$$
\begin{aligned}
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varphi^2_{Z_\nu \varkappa}}}{\partial \phi_{X_{\omega'} l}} &= X_{ij\omega'} Z^2_{i(j)\nu} \Bigg(\delta_{ij} \alpha^*(T_{ij}) m^*_{ij}(T_{ij}) q^*_{ij}(T_{ij}) \\
&\times \left[\left\{ m^*_{ij}(T_{ij}) - q^*_{ij}(T_{ij}) \right\}^2 - 2 m^*_{ij}(T_{ij}) q^*_{ij}(T_{ij}) \right] T^{1-l}_{ij} (T_{ij} - \kappa_\varkappa)^{2p}_+ \\
&- \int_0^{T_{ij}} h^*_0(s) \alpha^*(s) m^*_{ij}(s) q^*_{ij}(s) \mathcal{E}^*_{ij} s^{1-l} s^{1-l} (s - \kappa_\varkappa)^{2p}_+ \times \\
&\times \left[\left\{ \alpha^*(s) m^*_{ij}(s) q^*_{ij}(s) + q^*_{ij}(s) - m^*_{ij}(s) \right\}^2 \right. \\
&\left. + \, m^*_{ij}(s) q^*_{ij}(s) \left\{ \alpha^*(s) q_{ij}(s) - \alpha^*(s) m_{ij}(s) - 2 \right\} \right] ds \\
&- \sum_{k=1}^{n_{ij}} \left\{ m_{ijk} (q^*_{ijk})^2 - (m^*_{ijk})^2 q_{ijk} \right\} t^{1-l}_{ijk} (s - \kappa_\varkappa)^{2p}_+ \Bigg) \\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varrho^2_{X_\omega \varkappa}}}{\partial \phi_{X_{\omega'} l}} &= \sum_{i=1}^{n} \sum_{j=1}^{n_i} -X_{ij\omega'} X^2_{ij\omega} \int_0^{T_{ij}} h^*_0(s) \mathcal{E}^*_{ij} (s - \kappa_\varkappa)^{2p}_+ s^{1-l} ds, \\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varrho^2_{Z_\nu \varkappa}}}{\partial \phi_{X_{\omega'} l}} &= \sum_{i=1}^{n} \sum_{j=1}^{n_i} -X_{ij\omega'} Z^2_{i(j)\nu} \int_0^{T_{ij}} h^*_0(s) \mathcal{E}^*_{ij} (s - \kappa_\varkappa)^{2p}_+ s^{1-l} ds,
\end{aligned}
$$

Derivatives for $\beta_{Z_{\nu'}}(t)$

$$
\begin{aligned}
\frac{\partial \ell^*(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \phi_{Z_{\nu'} l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\Bigg[Z_{i(j)\nu'}\Bigg\{\delta_{ij}\alpha^*(T_{ij}) m^*_{ij}(T_{ij}) q^*_{ij}(T_{ij}) T_{ij}^{1-l} - \int_0^{T_{ij}} h^*_0(s)\mathcal{E}^*_{ij} m^*_{ij}(s) q^*_{ij}(s) s^{1-l} ds \\
&+ \sum_{k=1}^{n_{ij}} \left(Y_{ijk} - m^*_{ijk}\right) t_{ijk}^{1-l}\Bigg\} + \mathcal{R}^*_{ij}\frac{\partial \Delta^{(2)*}_{ij}}{\partial \phi_{Z_{\nu'} l}}\Bigg] \\
&+ \sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma^*_{\alpha_\varkappa}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varrho^2_{\alpha\vartheta}}}{\partial \phi_{Z_{\nu'} l}} + \frac{\varsigma^*_{h_\varkappa}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varrho^2_{h\varkappa}}}{\partial \phi_{Z_{\nu'} l}}\right\} \\
&+ \sum_{\omega=1}^{p}\sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma^*_{\varphi_{x_\omega \varkappa}}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varphi^2_{X_\omega \varkappa}}}{\partial \phi_{Z_{\nu'} l}} + \frac{\varsigma^*_{\varrho_{x_\omega \varkappa}}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varrho^2_{X_\omega \varkappa}}}{\partial \phi_{Z_{\nu'} l}}\right\} \\
&+ \sum_{\nu=1}^{p}\sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma^*_{\varphi_{z_\nu \varkappa}}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varphi^2_{Z_\nu \varkappa}}}{\partial \phi_{Z_{\nu'} l}} + \frac{\varsigma^*_{\varrho_{z_\nu \varkappa}}}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varrho^2_{Z_\nu \varkappa}}}{\partial \phi_{Z_{\nu'} l}}\right\} \\
\frac{\partial \Delta^{(2)}_{ij}}{\partial \phi_{Z_{\nu'} l}} &= Z_{i(j)\nu'}\Bigg(\delta_{ij}\alpha^*(T_{ij}) m^*_{ij}(T_{ij}) q^*_{ij}(T_{ij})\left[\left\{m^*_{ij}(T_{ij}) - q^*_{ij}(T_{ij})\right\}^2 - 2m^*_{ij}(T_{ij}) q^*_{ij}(T_{ij})\right] T_{ij}^{1-l} \\
&- \int_0^{T_{ij}} h^*_0(s)\alpha^*(s) m^*_{ij}(s) q^*_{ij}(s)\mathcal{E}^*_{ij} s^{1-l} \times \\
&\times \Big[\left\{\alpha^*(s) m^*_{ij}(s) q^*_{ij}(s) + q_{ij}(s) - m^*_{ij}(s)\right\}^2 \\
&+ m_{ij}(s) q^*_{ij}(s)\left\{\alpha^*(s) q^*_{ij}(s) - \alpha^*(s) m^*_{ij}(s) - 2\right\}\Big] ds \\
&- \sum_{k=1}^{n_{ij}}\left\{m^*_{ijk}(q^*_{ijk})^2 - (m^*_{ijk})^2 q^*_{ijk}\right\} t_{ijk}^{1-l}\Bigg) \\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varrho^2_{\alpha\vartheta}}}{\partial \phi_{Z_{\nu'} l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} - Z_{i(j)\nu'}\int_0^{T_{ij}} h^*_0(s)\mathcal{E}^*_{ij} m^*_{ij}(s)\left[\{m^*_{ij}(s)\}^2 + 2m^*_{ij}(s) q^*_{ij}(s)\right](s-\kappa_\varkappa)_+^{2p} s^{1-l} ds, \\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varrho^2_{h\varkappa}}}{\partial \phi_{Z_{\nu'} l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} - Z_{i(j)\nu'}\int_0^{T_{ij}} h^*_0(s)\mathcal{E}^*_{ij}(s-\kappa_\varkappa)_+^{2p} s^{1-l} ds, \\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varphi^2_{X_\omega \varkappa}}}{\partial \phi_{Z_{\nu'} l}} &= Z_{i(j)\nu'} X^2_{ij\omega}\Bigg(\delta_{ij}\alpha^*(T_{ij}) m^*_{ij}(T_{ij}) q^*_{ij}(T_{ij}) \\
&\times\left[\left\{m^*_{ij}(T_{ij}) - q^*_{ij}(T_{ij})\right\}^2 - 2m^*_{ij}(T_{ij}) q^*_{ij}(T_{ij})\right] T_{ij}^{1-l}(T_{ij}-\kappa_\varkappa)_+^{2p} \\
&- \int_0^{T_{ij}} h^*_0(s)\alpha^*(s) m^*_{ij}(s) q^*_{ij}(s)\mathcal{E}^*_{ij} s^{1-l} s^{1-l}(s-\kappa_\varkappa)_+^{2p} \times \\
&\times \Big[\left\{\alpha^*(s) m^*_{ij}(s) q^*_{ij}(s) + q^*_{ij}(s) - m^*_{ij}(s)\right\}^2 \\
&+ m^*_{ij}(s) q^*_{ij}(s)\left\{\alpha^*(s) q_{ij}(s) - \alpha^*(s) m_{ij}(s) - 2\right\}\Big] ds \\
&- \sum_{k=1}^{n_{ij}}\left\{m_{ijk}(q^*_{ijk})^2 - (m^*_{ijk})^2 q_{ijk}\right\} t_{ijk}^{1-l}(s-\kappa_\varkappa)_+^{2p}\Bigg)
\end{aligned}
$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varphi^2_{Z_\nu \varkappa}}}{\partial \phi_{Z_{\nu'} l}} = Z_{i(j)\nu'} Z^2_{i(j)\nu} \Bigg(\delta_{ij} \alpha^*(T_{ij}) m^*_{ij}(T_{ij}) q^*_{ij}(T_{ij})$$

$$\times \left[\left\{ m^*_{ij}(T_{ij}) - q^*_{ij}(T_{ij}) \right\}^2 - 2 m^*_{ij}(T_{ij}) q^*_{ij}(T_{ij}) \right] T^{1-l}_{ij} (T_{ij} - \kappa_\varkappa)^{2p}_+$$

$$- \int_0^{T_{ij}} h^*_0(s) \alpha^*(s) m^*_{ij}(s) q^*_{ij}(s) \mathcal{E}^*_{ij} s^{1-l} s^{1-l} (s - \kappa_\varkappa)^{2p}_+ \times$$

$$\times \left[\left\{ \alpha^*(s) m^*_{ij}(s) q^*_{ij}(s) + q^*_{ij}(s) - m^*_{ij}(s) \right\}^2 \right.$$

$$\left. + m^*_{ij}(s) q^*_{ij}(s) \left\{ \alpha^*(s) q_{ij}(s) - \alpha^*(s) m_{ij}(s) - 2 \right\} \right] ds$$

$$- \sum_{k=1}^{n_{ij}} \left\{ m_{ijk} (q^*_{ijk})^2 - (m^*_{ijk})^2 q_{ijk} \right\} t^{1-l}_{ijk} (s - \kappa_\varkappa)^{2p}_+ \Bigg)$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varrho^2_{X_\omega \varkappa}}}{\partial \phi_{Z_{\nu'} l}} = \sum_{i=1}^{n} \sum_{j=1}^{n_i} - Z_{i(j)\nu'} X^2_{ij\omega} \int_0^{T_{ij}} h^*_0(s) \mathcal{E}^*_{ij} (s - \kappa_\varkappa)^{2p}_+ s^{1-l} ds,$$

$$\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \varrho^2_{Z_\nu \varkappa}}}{\partial \phi_{Z_{\nu'} l}} = \sum_{i=1}^{n} \sum_{j=1}^{n_i} - Z_{i(j)\nu'} Z^2_{i(j)\nu} \int_0^{T_{ij}} h^*_0(s) \mathcal{E}^*_{ij} (s - \kappa_\varkappa)^{2p}_+ s^{1-l} ds,$$

Derivatives for $\gamma_{X_{\omega'}l}(t)$

$$
\begin{aligned}
\frac{\partial \ell^*(\boldsymbol{\theta}^*_{RE}, \boldsymbol{\theta}^*)}{\partial \rho_{X_{\omega'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[X_{ij\omega'}\left\{\delta_{ij}T_{ij}^{1-l} - \int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}s^{1-l}ds + \mathcal{R}^*_{ij}\frac{\partial \Delta^{(2)*}_{ij}}{\partial \rho_{X_{\omega'}l}}\right.\right] \\
&+ \sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma^*_{\alpha_\varkappa}}{2}\frac{\partial\frac{\partial^2\ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial\varrho^2_{\alpha\varkappa}}}{\partial\rho_{X_{\omega'}l}} + \frac{\varsigma^*_{\hbar_\varkappa}}{2}\frac{\partial\frac{\partial^2\ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial\varrho^2_{\hbar\varkappa}}}{\partial\rho_{X_{\omega'}l}}\right\} \\
&+ \sum_{\omega=1}^{p}\sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma^*_{\varphi_{x_\omega\varkappa}}}{2}\frac{\partial\frac{\partial^2\ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial\varphi^2_{X_\omega\varkappa}}}{\partial\rho_{X_{\omega'}l}} + \frac{\varsigma^*_{\varrho_{x_\omega\varkappa}}}{2}\frac{\partial\frac{\partial^2\ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial\varrho^2_{X_\omega\varkappa}}}{\partial\rho_{X_{\omega'}l}}\right\} \\
&+ \sum_{\nu=1}^{p}\sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma^*_{\varphi_{z_\nu\varkappa}}}{2}\frac{\partial\frac{\partial^2\ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial\varphi^2_{Z_\nu\varkappa}}}{\partial\rho_{X_{\omega'}l}} + \frac{\varsigma^*_{\varrho_{z_\nu\varkappa}}}{2}\frac{\partial\frac{\partial^2\ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial\varrho^2_{Z_\nu\varkappa}}}{\partial\rho_{X_{\omega'}l}}\right\} \\
\frac{\partial \Delta^{(2)*}_{ij}}{\partial \rho_{X_{\omega'}l}} &= -X_{ij\omega'}\int_0^{T_{ij}} s^{1-l}\hbar_0^*(s)\alpha^*(s)m^*_{ij}(s)q^*_{ij}(s)\mathcal{E}^*_{ij}\left\{\alpha^*(s)m^*_{ij}(s)q^*_{ij}(s) + q^*_{ij}(s) - m^*_{ij}(s)\right\}ds \\
\frac{\partial\frac{\partial^2\ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial\varrho^2_{\alpha\varkappa}}}{\partial\rho_{X_{\omega'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -X_{ij\omega'}\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}\{m^*_{ij}(s)\}^2 s^{1-l}(s-\kappa_\varkappa)_+^{2p}ds, \\
\frac{\partial\frac{\partial^2\ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial\varphi^2_{X_\omega\varkappa}}}{\partial\rho_{X_{\omega'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -X_{ij\omega'}X^2_{ij\omega}\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}\alpha^*(s)m^*_{ij}(s)q^*_{ij}(s) \\
&\times\{\alpha^*(s)m^*_{ij}(s)q^*_{ij}(s) + q^*_{ij}(s) - m^*_{ij}(s)\}s^{1-l}(s-\kappa_\varkappa)_+^{2p}ds \\
\frac{\partial\frac{\partial^2\ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial\varphi^2_{Z_\nu\varkappa}}}{\partial\rho_{X_{\omega'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -X_{ij\omega'}Z^2_{i(j)\nu}\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}\alpha^*(s)m^*_{ij}(s)q^*_{ij}(s) \\
&\times\{\alpha^*(s)m^*_{ij}(s)q^*_{ij}(s) + q^*_{ij}(s) - m^*_{ij}(s)\}s^{1-l}(s-\kappa_\varkappa)_+^{2p}ds \\
\frac{\partial\frac{\partial^2\ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial\varrho^2_{X_\omega\varkappa}}}{\partial\rho_{X_{\omega'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -X_{ij\omega'}X^2_{ij\omega}\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}s^{1-l}(s-\kappa_\varkappa)_+^{2p}ds, \\
\frac{\partial\frac{\partial^2\ell(\boldsymbol{\theta}^*_{RE},\boldsymbol{\theta}^*)}{\partial\varrho^2_{Z_\nu\varkappa}}}{\partial\rho_{X_{\omega'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -X_{ij\omega'}Z^2_{i(j)\nu}\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}^*_{ij}s^{1-l}(s-\kappa_\varkappa)_+^{2p}ds,
\end{aligned}
$$

Derivatives for $\gamma_{Z_{\nu'}l}(t)$

$$
\begin{aligned}
\frac{\partial \ell^*(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \rho_{Z_{\nu'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[X_{ij\omega'}\left\{\delta_{ij}T_{ij}^{1-l} - \int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}_{ij}^* s^{1-l} ds + \mathcal{R}_{ij}^*\frac{\partial \Delta_{ij}^{(2)*}}{\partial \rho_{Z_{\nu'}l}}\right.\right]\\
&+ \sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma_{\alpha_\varkappa}^*}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \varrho_{\alpha\varkappa}^2}}{\partial \rho_{Z_{\nu'}l}} + \frac{\varsigma_{\hbar_\varkappa}^*}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \varrho_{\hbar\varkappa}^2}}{\partial \rho_{Z_{\nu'}l}}\right\}\\
&+ \sum_{\omega=1}^{p}\sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma_{\varphi_{x_\omega\varkappa}}^*}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \varphi_{X_\omega\varkappa}^2}}{\partial \rho_{Z_{\nu'}l}} + \frac{\varsigma_{\varrho_{x_\omega\varkappa}}^*}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \varrho_{X_\omega\varkappa}^2}}{\partial \rho_{Z_{\nu'}l}}\right\}\\
&+ \sum_{\nu=1}^{p}\sum_{\varkappa=1}^{n_\kappa}\left\{\frac{\varsigma_{\varphi_{z_\nu\varkappa}}^*}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \varphi_{Z_\nu\varkappa}^2}}{\partial \rho_{Z_{\nu'}l}} + \frac{\varsigma_{\varrho_{z_\nu\varkappa}}^*}{2}\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \varrho_{Z_\nu\varkappa}^2}}{\partial \rho_{Z_{\nu'}l}}\right\}\\
\frac{\partial \Delta_{ij}^{(2)*}}{\partial \rho_{Z_{\nu'}l}} &= -Z_{i(j)\nu'}\int_0^{T_{ij}} s^{1-l}\hbar_0^*(s)\alpha^*(s)m_{ij}^*(s)q_{ij}^*(s)\mathcal{E}_{ij}^*\left\{\alpha^*(s)m_{ij}^*(s)q_{ij}^*(s) + q_{ij}^*(s) - m_{ij}^*(s)\right\}ds\\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \varrho_{\alpha\varkappa}^2}}{\partial \rho_{X_{\omega'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -X_{ij\omega'}\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}_{ij}^*\{m_{ij}^*(s)\}^2 s^{1-l}(s-\kappa_\varkappa)_+^{2p}ds,\\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \varphi_{X_\omega\varkappa}^2}}{\partial \rho_{Z_{\nu'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -Z_{i(j)\nu'}X_{ij\omega}^2\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}_{ij}^*\alpha^*(s)m_{ij}^*(s)q_{ij}^*(s)\\
&\times\{\alpha^*(s)m_{ij}^*(s)q_{ij}^*(s) + q_{ij}^*(s) - m_{ij}^*(s)\}s^{1-l}(s-\kappa_\varkappa)_+^{2p}ds\\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \varphi_{Z_\nu\varkappa}^2}}{\partial \rho_{Z_{\nu'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -Z_{i(j)\nu'}Z_{i(j)\nu}^2\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}_{ij}^*\alpha^*(s)m_{ij}^*(s)q_{ij}^*(s)\\
&\times\{\alpha^*(s)m_{ij}^*(s)q_{ij}^*(s) + q_{ij}^*(s) - m_{ij}^*(s)\}s^{1-l}(s-\kappa_\varkappa)_+^{2p}ds\\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \varrho_{X_\omega\varkappa}^2}}{\partial \rho_{Z_{\nu'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -Z_{i(j)\nu'}X_{ij\omega}^2\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}_{ij}^* s^{1-l}(s-\kappa_\varkappa)_+^{2p}ds,\\
\frac{\partial \frac{\partial^2 \ell(\boldsymbol{\theta}_{RE}^*, \boldsymbol{\theta}^*)}{\partial \varrho_{Z_\nu\varkappa}^2}}{\partial \rho_{Z_{\nu'}l}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} -Z_{i(j)\nu'}Z_{i(j)\nu}^2\int_0^{T_{ij}} \hbar_0^*(s)\mathcal{E}_{ij}^* s^{1-l}(s-\kappa_\varkappa)_+^{2p}ds,
\end{aligned}
$$