UC Riverside
UC Riverside Electronic Theses and Dissertations

Title
Subgroup Analysis of Survival Data With Interval Censoring and Time Varying Covariates

Permalink
https://escholarship.org/uc/item/02j0g9z5

Author
Brannan, Michael Tyler

Publication Date
2021

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Subgroup Analysis of Survival Data With Interval Censoring and Time Varying Covariates

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

Doctor of Philosophy

in

Applied Statistics

by

Michael Tyler Brannan

March 2021

Dissertation Committee:

Dr. Yehua Li, Chairperson
Dr. Xinping Cui
Dr. Weixin Yao

Copyright by
Michael Tyler Brannan
2021

The Dissertation of Michael Tyler Brannan is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am grateful to my adviser, Dr. Yehua Li, whose guidance was instrumental in completing my degree. I am also grateful to my High School, Undergraduate, and Graduate Teachers, Professors, and Mentors who answered my questions and allowed me to achieve a deep understanding of the material. I am especially grateful to Austin Chi and Dr. Nathan Robertson, without both, I would not have been able to achieve what has been accomplished here. Their countless hours of assistance on both theoretical as well as computational aspects of this model allowed it to develop as quickly as it has. Austin Chi was instrumental in achieving the structure, layout, and speed needed when coding the package. His wealth of knowledge in computer science was invaluable. Dr. Nathan Robertson was instrumental as a sounding board for theoretical ideas along with expanding my knowledge in Monte Carlo techniques. I am also grateful to Matthew McNab for editing this document.

To my parents for all their support.

ABSTRACT OF THE DISSERTATION

Subgroup Analysis of Survival Data With Interval Censoring and Time Varying Covariates

by

Michael Tyler Brannan

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, March 2021
Dr. Yehua Li, Chairperson

We developed a model that performs unsupervised clustering of survival times in a joint survival-longitudinal framework with known longitudinal trajectories and all forms of censoring and truncation. The model allows for data that is observed, left censored, right censored, interval censored, along with all forms of truncation. From simulation studies, the model correctly identifies the parameter estimates in any level of censoring quickly. When clustering is present, we use variations of AIC and BIC for identifying the correct number of clusters. From simulation studies, we find that BIC correctly identifies the right number of clusters within multiple levels of censoring greater than 90% of the time along with correctly estimating the parameter estimates. All the analysis is performed in the R package currently being developed, which performs the analysis relatively quickly. We applied the model to the Study of Women's Health Across the Nation (SWAN) dataset. We used this data set for detecting Alzheimer's disease and to decipher what covariates are linked to an increased risk for developing Alzheimer's disease.

Contents

List of Figures

List of Tables

Chapter 1

Introduction, Background, Literature Review, and Model Framework

1.1 Introduction

My research addresses the topic of unsupervised clustering of survival times in a joint survival-longitudinal modeling environment where the longitudinal trajectories are known. From the other published works we have reviewed, no one has attempted this before. Others in the field have developed joint survival-longitudinal models including clustering, but none performed clustering in an unsupervised setting. Moreover, this model allows observed data, left censored, right censored, interval censored, and all forms of truncation. Most models only allow observed and right censored data with a few also including interval censored. However, we have not found any that include all possible data forms. Since our model contains a non-parametric baseline function, we decided to use a method that allows estimation of the parameters without needing to include additional parameter constraints. Therefore, we followed a framework by Cai and Betensky (2003) instead of adding constraints or reparameterizing the baseline parameters as in Rizopoulos et al. (2009).

This research harnesses ideas from survival analysis, longitudinal analysis, and mixture modeling. We modified and fused techniques from all three areas to more closely model the real world. The longitudinal part of my model uses B-spline basis functions to model both the mean and individual trajectories. The longitudinal framework is modeled after the form in Rice and Wu (2001). This allows the B-spline basis function knots to differ between the fixed and random effects. The added flexibility for the individual knots is different from the technique used in other joint survival-longitudinal models, like one found in Brown et al. (2005).

For the clustering aspect of the model, we used the Expectation-Maximization (EM) algorithm by Dempster et al. (1977) along with an extension of Akaike Information Criterion (AIC) by Akaike (1973) and Bayesian Information Criterion (BIC) by Schwarz (1978) for determining the number of clusters. This has been applied in a mixture regression setting with normal densities as described by Naik et al. (2007). However, we have not seen this done in a joint survival-longitudinal setting.

I am also building an R package to allow others access to this model. I have coded the model using the R package Rcpp by Eddelbuettel and François (2011), Eddelbuettel (2013), and Eddelbuettel and Balamuta (2017). Rcpp allows the integration of R by R Core

Team (2020) and c++. I also used Rcpp Armadillo by Eddelbuettel and Sanderson (2014). This allowed me to write the package in c++ to gain speed while still being able to use objects specific to R by R Core Team (2020).

From the parts of the model that are included in the package, the package runs quickly. Given reasonable starting values, it can find the estimates for 36 parameters in under 10 minutes for 200 individuals and under 20 minutes for 500 individuals.

To demonstrate that the survival portion of the model estimates the parameters correctly, we simulated data sets with two longitudinal covariates and two time independent covariates. We had 200 replicates of data sets with 200 individuals and 500 individuals. From these findings, we were able to show that the model has very little bias and is not greatly affected by increasing censoring, specifically interval censoring.

Additionally, we demonstrated that the unsupervised clustering aspect of the survival model with known longitudinal covariates correctly identifies the number of clusters along with the correct parameter estimates for each cluster. We ran 200 replicates with both 300 and 900 individuals. For the ones that converged, we found that it estimated the correct number of clusters, which was 2, and correctly estimated the 32 baseline parameters and 4 cluster specific parameters.

This survival model with known longitudinal covariates has potential to promote immense cross-fertilization of ideas or application for other disciplines. We applied it to the Study of Women's Health Across the Nation (SWAN) dataset. We used this data set for detecting Alzheimer's disease and to decipher what covariates are linked to an increase risk in developing Alzheimer's disease. Covariates given include blood pressure, medical conditions, smoking status, relationships, BMI, and numerous vitamin levels.

1.2 Structure

This dissertation is structured into 6 chapters. The first chapter is a background of methods used, a literature review of techniques for survival, longitudinal, clustering, and joint modeling, and the framework of our model. The second chapter is forming a joint survival-longitudinal model assuming the longitudinal part is fully known and there is no clustering. The third chapter is performing unsupervised clustering of the joint survival-longitudinal model assuming the longitudinal part is fully known. The fourth chapter is a look into the R Package I developed. The fifth chapter is using the model to analyze the SWAN dataset. The sixth chapter is current and future work.

1.3 Background

1.3.1 Numerical Integration

There are multiple numerical integration techniques available. We described the five techniques as stated in Section 7.7 of Stewart (2008), which are left endpoint approximation, right endpoint approximation, Midpoint Rule, Trapezoidal Rule, and Simpson's Rule. Given that $f(x)$ is the function value at point x in the interval $[a = x_0, b = x_n]$, $\Delta x = \frac{b-a}{n}$, $x_i = a + i\Delta x$, $\bar{x}_i = \frac{x_{i-1}+x_i}{2}$, and n is the number of segments $[a, b]$ is divided into, the formulas for the numerical integration techniques are:

Technique	Formula
Left endpoint approximation	$\int_a^b f(x)\,dx = \sum_{i=1}^n f(x_{i-1})\,\Delta x$
Right endpoint approximation	$\int_a^b f(x)\,dx = \sum_{i=1}^n f(x_i)\,\Delta x$
Midpoint Rule	$\int_a^b f(x)\,dx = \sum_{i=1}^n f(\bar{x}_i)\,\Delta x$
Trapezoidal Rule	$\int_a^b f(x)\,dx = \sum_{i=1}^n \left(f(x_{i-1}) + f(x_i)\right)\frac{\Delta x}{2}$
Simpson Rule	$\int_a^b f(x)\,dx = \sum_{i=1}^n \left(f(x_{i-1}) + 4f(\bar{x}_i) + f(x_i)\right)\frac{\Delta x}{3}$.

Figures 1.1-1.5 show how each numerical integration technique performs at estimating the integral of $f(x) = e^x$.

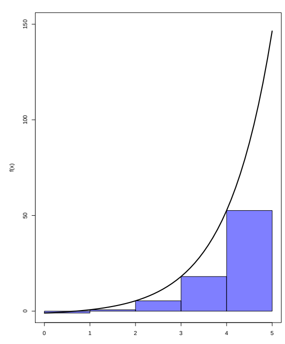

Figure 1.1: Left Endpoint Approximation

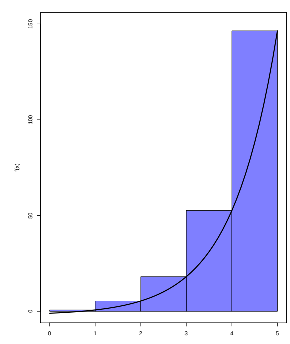

Figure 1.2: Right Endpoint Approximation

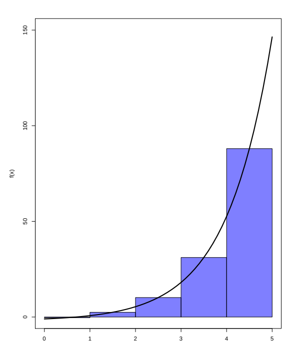

Figure 1.3: Midpoint Rule

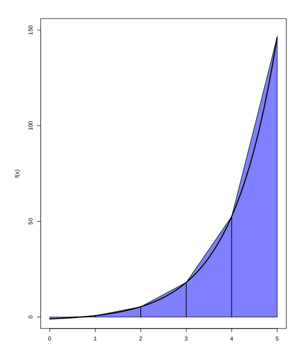

Figure 1.4: Trapezoidal Rule

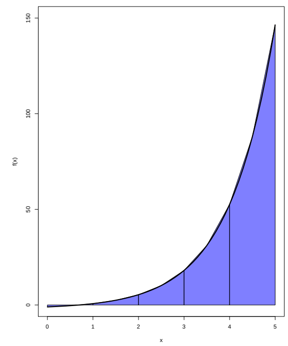

Figure 1.5: Simpson's Rule

Since the Midpoint Rule always gives an equal or better numerical approximation to the left and right endpoint approximations, we will not be discussing them from now on. Therefore, the error rates for Midpoint Rule, Trapezoidal Rule, and Simpson's Rule are:

Technique	Error Rate
Midpoint Rule	$\|E_M\| = \frac{\|f''(x)\|(b-a)^3}{24n^2}$
Trapezoidal Rule	$\|E_T\| = \frac{\|f''(x)\|(b-a)^3}{12n^2}$
Simpson's Rule	$\|E_S\| = \frac{\left\|f^{(4)}(x)\right\|(b-a)^5}{180n^4}$,

where $f''(x)$ is the second derivative of f and $f^{(4)}$ is the fourth derivative of f. Additionally, n must always be even when using any formulas involving Simpson's Rule.

From looking at the error rates, we can see that Simpson's Rule's error rate is much smaller than the Midpoint Rule's error rate and the Trapezoidal Rule's error rate. Since Simpson's Rule only required calculating one additional point per interval more than the Trapezoidal Rule, we felt that the extra computation time was worth the decrease in error rate. Even though Midpoint Rule only required calculating one point per interval, the error rate was much larger compared to Simpson's Rule. There are other techniques with even smaller error rates, but they require calculating even more function values, like Boole's Rule and Hardy's Rule that require five and seven function values per segment, respectively. We felt that they would take too much computational time for a diminishing decrease in error rate. Additionally, we allow the user to increase the number of segments per interval if higher precision is desired. As can be see from the figures of e^x, we can see that Simpson's Rule does a better job estimating the integral of e^x than does the Midpoint Rule or Trapezoidal Rule, since it more closely follows the function's value. Since our function is a form of e^x, we conclude that Simpson's Rule would also give a close approximation for our function.

1.3.2 Newton Optimization

We begin by explaining Newton's method for a function $f(x)$ as given in Section 4.8 of Stewart (2008). Newton's method uses the idea that given a specific point, x_1, on the function $f(x)$, the tangent line's x-intercept at x_1 is close to the root of $f(x)$. Therefore, we can find the x-intercept of the tangent line easily. The slope of the tangent line at x_1 is $f'(x_1)$. Since we know that the y-value of the x-intercept of the tangent line is 0, we find x_2 using the the point slope equation and get the following equation for x_2:

$$\begin{aligned} y - f(x_1) &= f'(x_1)(x_2 - x_1) \\ 0 - f(x_1) &= f'(x_1)(x_2 - x_1) \\ (x_2 - x_1) &= -\frac{f(x_1)}{f'(x_1)} \\ x_2 &= x_1 - \frac{f(x_1)}{f'(x_1)}. \end{aligned}$$

Having found x_2, we now use it as our starting value for the next round. The general formula is:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

We continue the process until the difference between x_n and x_{n+1} is less than some prespecified threshold. The application of the method can be seen in Figure 1.6 for estimating the root for the function $e^x - 2$.

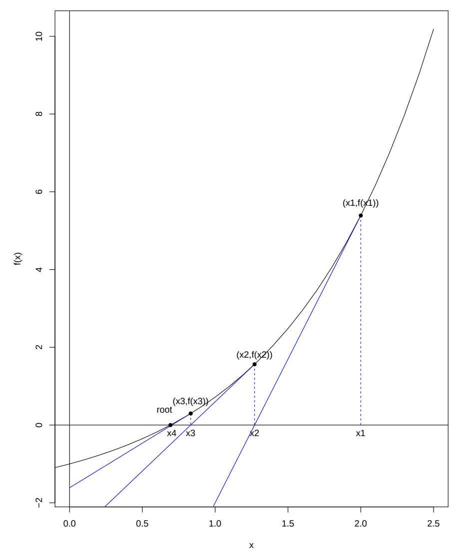

Figure 1.6: Plot of Newton's method for $e^x - 2$.

For Newton optimization, we use this same formula, but instead of finding the roots of $f(x)$, we find the roots of $f'(x)$. We make this change since we are interested in finding the maximum or minimum of $f(x)$. These points are the location where $f'(x) = 0$, which are the roots of $f'(x)$. Therefore, the formula in the univarate case becomes:

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)},$$

where f'' is the second derivative of f. In the multivariate case, the general formula becomes:

$$X_{n+1} = X_n - \left[f''\left(X_n\right)\right]^{-1} f'\left(X_n\right),$$

where X is a p-dimensional vector, $f'\left(X\right)$ is the p-dimensional gradient vector or vector of first derivatives of $f\left(X\right)$, and $f''\left(X\right)$ is the $p \times p$ Hessian matrix or matrix of second derivatives of $f\left(X\right)$.

1.3.3 B-Splines

In our model, we make use of basis splines, referred to as B-splines. They are the basis functions for the spline space and are used due to the fact that they are numerically stable and uniquely defined by their knot sequence. Additionally, the sum of the B-splines is always 1 and they are piece-wise polynomials of order n and degree $p = n - 1$. The locations where the piece-wise polynomials meet are called knots. Given the knots are all distinct, the B-spline meets the criteria of being continuous up to degree at least $n-2$. We find that B-splines of degree 0 have the following form:

$$B_{i,0}\left(x\right) = \begin{cases} 1 & \text{if } k_i \leq x < k_{i+1}. \\ 0 & \text{otherwise,} \end{cases}$$

where k_i is the i^{th} knot where $i = 0, \ldots, K$. This means that if the value of x is in the knot segment the value of the B-spline is 1, otherwise it is 0. Using this result, the Cox de-Boor recursion formula from De Boor (1978) for representing B-splines of degree p is:

$$B_{i,p} = \frac{x - k_i}{k_{i+p} - k_i} B_{i,p-1}(x) + \frac{k_{i+p+1} - x}{k_{i+p+1} - k_{i+1}} B_{i+1,p-1}(x).$$

Since a single B-spline extends over $p + 2$ knots, as can be seen by $B_{i,0}$ extending over the two knots k_i and k_{i+1}, internal knots have to be extended by p in order to calculate the first and last knot. This is usually done by repeating the first and last knots each p times more, sometimes called boundary knots.

1.3.4 Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation (MLE) is a method of estimating the parameters of a probability distribution by maximizing the likelihood function. This is done by finding the parameter values that make the observed data most probable. The process usually starts with forming the likelihood and then either taking the derivative of the likelihood or more commonly the loglikelihood. When the first derivative exists, we then set the derivative with respect to the parameter equal to 0 and then solve for the parameter of interest in terms of the remaining parameters and the observed data. The point found, along with boundary points when applicable, are checked to see if the point indeed maximizes the likelihood or loglikelihood function. This is the method used to estimate the parameters in our model and full derivations are given in the Appendices. A more detailed background with examples can be found in Section 7.2.2 of Casella and Berger (2002).

1.3.5 Survival Analysis

The following overview is a brief introduction of what is described in further detail in Chapters 1, 2, 3, and 8 of Klein and Moeschberger (2011). Survival analysis was designed to analyze problems involving time to event data. Many times, the data can have observations that are censored or truncated. Censored data occurs when the event happens outside of the given period of time. For example, if a study was conducted and individuals were watched for a year to see if a certain event occurred, right censored observations would mean the event did not occur during the year, left censored observations would mean the event occurred before the year began, and interval censored observations would mean the event occurred during the year, but the observers do not know preciously when. Right censoring means the event did not occur before some prespecified time, left censoring means the event occurred before they are observed in the study, and interval censored means the event is known to occur between two specified times. Truncation means that only individuals within a certain observation window are observed. The difference between truncation and censoring is that with truncation, any individual who does not have an event in the observation window, gives no information to the observer. Censoring, on the other hand, does give partial information to the observer, since individuals may have already had the event occur before the observation windows or still have not had it occur by the time the observation window ends. Therefore, truncated data causes conditional estimation.

Left truncation means that we only observe the event if it is greater than the truncation time. An example of this occurs with microscopic particles. The scientist can only see particles that are large enough to be discerned by the microscope. Anything smaller is not observed by the scientist. For humans, this could be exposure to some disease, diagnosis of the disease, or entry into a retirement home. For instance, if I wanted to conduct a study looking at the age when a member of the retirement home died, only individuals who entered the retirement home are included. Individuals who died before being able to enter the retirement home are not included, since the truncation point was age entering the retirement. Right truncation occurs when the event occurs before the truncation time. This occurs with star viewing, since stars too far away to be seen are not included. In humans, AIDS studies commonly use right truncation. Studies that look at AIDS caused by transfusions take the time from the transfusion until a certain time, and include individuals who have had AIDS occur by that date. Therefore, this causes right truncation as individuals who had not had AIDS occur at this point are not included in the study.

Before we show how censoring and trunction are written mathematically, we will briefly review some of the basics of survival analysis. First we define the survival function. The survival function is defined as the probability that an individual survives beyond a certain value, $Pr\left(X > x\right)$. When X is a continuous random variable, the survival function is a continuous, strictly decreasing function. It is the complement of the cumulative distribution function, $F\left(X\right)$, meaning $S\left(X\right) = 1 - F\left(X\right)$. Now,

$$S\left(X\right) = Pr\left(X > x\right) = \int_x^\infty f\left(t\right) dt \text{ and}$$
$$f\left(x\right) = -\frac{dS\left(X\right)}{dx},$$

where $f(x)$ is the probability density function of X. The hazard function is the probability that an individual at time x will experience the event in the next instant. It is defined as:

$$h(x) = \lim_{\Delta x \to 0} \frac{Pr(x \leq X < x + \Delta x | X \geq x)}{\Delta x}.$$

Given X is a continuous random variable,

$$h(x) = \frac{f(x)}{S(X)} = -\frac{d \log(S(x))}{dx}.$$

The cumulative hazard function, $H(X)$ is defined as:

$$H(X) = \int_0^x h(t)\, dt = -\log(S(x)).$$

This means the following relationship exists:

$$S(x) = e^{-H(X)} = e^{-\int_0^x h(t)dt}.$$

Lastly, since we assume a Cox proportional hazards model in our model framework, we explain the basics of the Cox proportional hazards model. The model, developed by Cox in Cox (1972), is a way of predicting the survival time by a set of explanatory variables. The basic model, as discussed in Klein and Moeschberger (2011) is:

$$h(t|Z) = h_0(t)\, c\left(\beta^T Z\right),$$

where $h(t|Z)$ is the hazard rate at time t for an individual with covariates Z, $h_0(t)$ is the baseline hazard rate, β is the vector of parameters, and $c\left(\beta^T Z\right)$ is a known function. This type of model is a semi-parametric model, since a parametric model is assumed only for the covariate effect, $c\left(\beta^T Z\right)$, whereas the baseline hazard rate is treated non-parametrically. What is commonly used is $c\left(\beta^T Z\right) = e^{\beta^T Z}$.

With this background, we have the following list of how each observed, censored, and truncated observation is written:

observed	$f(x)$
right-censored	$S(C_r)$
left-censored	$F(C_l)$
interval-censored	$S(T_l) - S(T_r)$
left-truncation	$\frac{f(x)}{S(L)}$
right-truncation	$\frac{f(x)}{S(R)}$
interval-truncation	$\frac{f(x)}{S(L)-S(R)}$.

Now C_l and C_r are the left and right censoring times, respectively, and L and R are the right and left truncation times, respectively. For ease of writing, we used $f(x)$ for the numerator of all the truncation equations. However, right, left, and interval censored data can be truncated as well.

1.3.6 Mixed Models applied to Longitudinal Data

In this section, we give a brief overview of mixed models as they relate to longitudinal data analysis. As stated in Chapter 4 of Ruppert et al. (2003), mixed models are regression models that incorporate random effects along with fixed effects. The general form, using standard regression notation, is:

$$Y = X\beta + Zu + \epsilon,$$

where Y is the vector of response values for the individuals, X is the matrix of fixed effects, β is vector of fixed effect parameters, Z is the matrix of random effects, u is the vector of random effect parameters, and ϵ is the random error component. The assumptions on u and ϵ are:

$$E\begin{bmatrix} u \\ \epsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } \mathrm{Cov}\begin{bmatrix} u \\ \epsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix},$$

where G is the covariance-variance matrix for u and $R = \sigma_\epsilon^2 I$. Since longitudinal data is data with repeated measures and can have elements that are the same for all members, fixed effects, and those that are individual specific, random effects, the mixed model fits this type of data well. The random effects are considered to have come from a distribution with a shared covariance-variance matrix, G. An example would be blood pressure, since it is usually measured multiple times over some defined period. Blood pressure differs per individual, but there is an assumption that all people come from a common distribution with a shared covariance-variance matrix. Therefore, the fixed effects could be each week the measurement was taken or estimating the overall mean of the population. The random effect could be as simple as a random intercept for each person or the more complicated estimation of each person's blood-pressure trajectory from the overall mean trajectory.

In our model, looking at one longitudinal covariate, j, for a single individual, i, this would be measured m_{ij} times. There would be a component that is fixed, which is the mean trajectory function, and a random component, which is the individual trajectory. The overview of how this is written in our model is:

$$\boldsymbol{Y}_{ij} = \boldsymbol{X}_{ij} + \boldsymbol{\varepsilon}_{ij},$$

where each is a vector of length m_{ij} and each row of $\boldsymbol{X}_{ij}$ is $X_{ij} = \boldsymbol{\beta}_j^T \overline{\boldsymbol{B}}_j + \boldsymbol{\xi}_{ij}^T \boldsymbol{B}_j$, where $\boldsymbol{\beta}_j^T \overline{\boldsymbol{B}}_j$ is the fixed effect part where $\boldsymbol{\beta}_j$ is the j^{th} mean trajectory coefficients shared by all individuals and $\boldsymbol{\xi}_{ij}^T \boldsymbol{B}_j$ is the random effect part where $\boldsymbol{\xi}_{ij}$ is the i^{th} individual's j^{th} individual trajetory coefficients. Moreover, $\overline{\boldsymbol{B}}_j$ and $\boldsymbol{B}_j$ are the vectors of B-spline basis for longitudinal trajectory j for the mean and random effects, respectively. Additionally,the random effects share a common covariance-variance matrix, $\Sigma_{\boldsymbol{\xi}_j}$, $\boldsymbol{\varepsilon}_{ij} \sim N_{m_{ij}}\left(0, \sigma_{Y_j}^2 I\right)$, and $\boldsymbol{\xi}_{ij}$ and $\boldsymbol{\varepsilon}_{ij}$ are independent. Full details are given in the Section 1.5.1.

1.3.7 Mixture Models

We briefly describe mixture models as stated in Naik et al. (2007). A mixture model is a probability distribution for representing sub-populations in the overall population. We are interested in finite mixture regression models, which means that the number of regression models, C, in the mixture is not infinite. Using standard regression notation, the mixture regression model is:

$$f\left(y; \boldsymbol{x}, \phi\right) = \sum_{c=1}^{C} \omega_c f_c\left(y; \boldsymbol{x}, \boldsymbol{\beta}_c, \sigma_c\right),$$

where $0 \leq \omega_c \leq 1$, $\sum_{c=1}^{C} \omega_c = 1$, and $f_c\left(y; \boldsymbol{x}, \boldsymbol{\beta}_c, \sigma_c\right)$ is a normal density with mean $\boldsymbol{x}^T\boldsymbol{\beta}_c$ and variance σ_c. Now $\boldsymbol{x}$ is a $p \times 1$ vector of explanatory variables, which are considered fixed, $\boldsymbol{\beta}_c$ is a $p \times 1$ parameter vector, and $\phi_c = \{\omega_c, \boldsymbol{\beta}_c, \sigma_c\}$ for $c = 1, \ldots, C$. The idea is that each of the sub-populations' densities is weighted by the amount of the population it makes up, given by ω_c, and each of the C densities have their own parameter estimates ϕ_c. Since we do not know for sure which individuals belong to each sub-population, each individual's density is a mixture of all C densities with corresponding weights $\boldsymbol{\omega}$.

1.3.8 K-Means Clustering

We briefly describe K-Means Clustering as stated in Section 10.3.1 of James et al. (2013). The idea is to partition the data into K distinct clusters where the observations within each cluster is similar by some measure. The algorithm starts by randomly assigning each observation to one of the K clusters. Assuming that each of the observations is of dimension p, we then calculate each cluster's centroid. The centroid of a cluster is the vector of p-dimensional means of the observations in that cluster. We then assign, using Euclidean distance, each observation to the centroid it is closest to. We repeat the process until there is no observations that change clusters. Figure 1.7 shows three iterations of the K-means algorithm, which is calculating the centroids and then reassigning the observations to the new centroids.

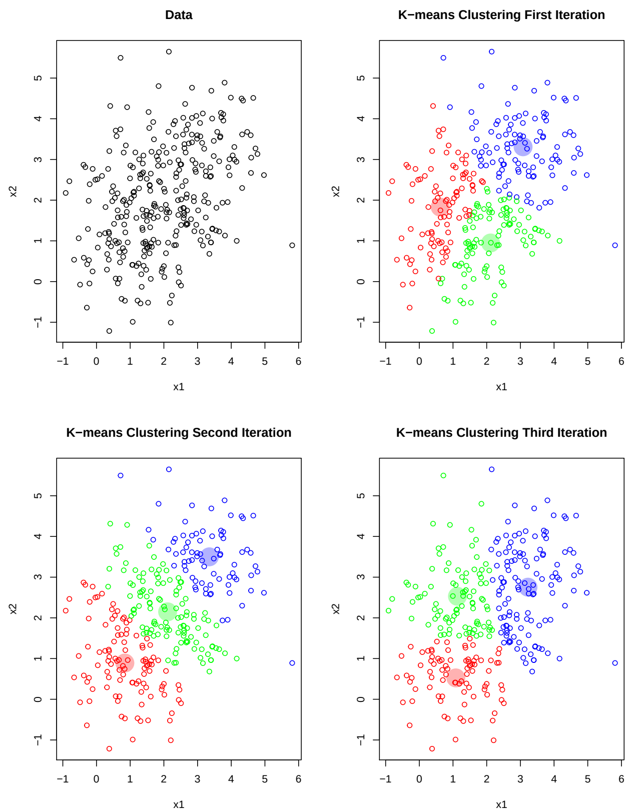

Figure 1.7: First three iterations of the K-means algorithm

1.3.9 Expectation-Maximization (EM) Algorithm

Since we used a mixture model as part of our model framework, we had latent variables for which group an individual was assigned to that needed to be estimated along with the parameters. Therefore, we use the Expectation-Maximization (EM) Algorithm to find the maximum of the loglikelihood function as stated in Dempster et al. (1977). In this context, the idea behind the EM algorithm is that in the E-step, we take the expected value of the loglikelihood with the current parameter values, which means finding the expected values of the latent variables, since the latent variables state the probabilities per cluster for each individual. Once we have those values for each individual, we then maximize the expected loglikelihood with respect to the parameters of interest. Since the EM algorithm is an iterative method, we continue this process until the difference between the previous and the current iterations is below a preset threshold. We give the EM example presented in Naik et al. (2007) in Section 1.4.5 along with the our EM algorithm in Chapter 3 with the full derivations in Appendix G.

1.3.10 Joint Model

A joint model combines the probability distributions of multiple distributions into one, hence joining them. In our case, we are combining a mixed model that represents the longitudinal data, a survival model given by a Cox proportional hazards model, as well as a mixture model to account for the subgroups. The reason why we use a joint model over just estimating a longitudinal model and then estimating a survival model is that parameters shared across models can be better estimated since information from both models is used. For instance, the parameter coefficeints of the B-spline that estimate a longitudinal element, such as blood pressure, are used in both the longitudinal submodel, to estimate the observed measurements, as well as in the survival submodel, to estimate the measurement at the time of the event of interest, per each identified cluster.

1.4 Literature Review

Standardized notation was used throughout this paper. The same notation used in our model is what is used as closely as possible for the models used in the papers mentioned below. Terms are defined each time and a list of them are given in Appendix A.

1.4.1 Cai and Betensky (2003)

In Cai and Betensky (2003), they had the following survival model:

$$T_i \sim \lambda_i\left(t|\boldsymbol{Z}_i\right) = \lambda_0\left(t\right)e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}},$$

where $\lambda_i\left(t|\boldsymbol{Z}_i\right)$ is the hazard function for the i^{th} individual, $\lambda_0\left(t\right)$ is the baseline hazard function, $\boldsymbol{\zeta}$ is the coefficients for the baseline covariates, $\boldsymbol{Z}_i = \begin{pmatrix} Z_{i1} \\ Z_{i2} \\ \vdots \\ Z_{iq} \end{pmatrix}$ is the baseline covariates for the i^{th} individual where there are q baseline covariates, and Z_{ij} is the j^{th} baseline covariate for the i^{th} individual where $j = 1, \ldots, q$. For the baseline hazard function $\lambda_0\left(t\right)$, they used the following form:

$$\lambda_0\left(t\right) = e^{\alpha_0+\alpha_1 t+\sum_{k=1}^{K} b_k(t-\kappa_k)_+} \text{ or}$$
$$\log\lambda_0\left(t\right) = \eta_0\left(t\right) = \alpha_0 + \alpha_1 t + \sum_{k=1}^{K} b_k\left(t-\kappa_k\right)_+,$$

where $\left(t-\kappa_k\right)_+ \equiv \max\left(0, \left(t-\kappa_k\right)\right)$. For the spline model, they treated the b_k's, the coefficient parameters for each of the k knots where $k = 1, \ldots, K$, as random effects distributed as follows:

$$\boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_K \end{bmatrix} \sim N_K(\boldsymbol{0}, \sigma_b^2\boldsymbol{I}_K).$$

Their model could handle survival times that are observed, right censored, or interval censored. Furthermore, their baseline hazard form avoids needing to add constraints or reparameterize the baseline coefficient parameters. However, they did not have time dependent or longitudinal covariates.

1.4.2 Brown et al. (2005)

In Brown et al. (2005), they used the following joint longitudinal and survival model in a Bayesian setting. They let the longitudinal model have the following form:

$$Y_{ijl} = X_{ij}\left(t_{ijl}\right) + \varepsilon_{ijl}, \text{ where } i = 1, \ldots, n,\ j = 1, \ldots, p, \text{ and } l = 1, \ldots, m_{ij}.$$

Now $X_{ij}\left(t_{ijl}\right)$ is the j^{th} trajectory function for the i^{th} individual and t_{ijl} is the l^{th} observed time point for the i^{th} individual's j^{th} trajectory. Additionally, n is the number of individuals, p is the number of longitudinal trajectories, and m_{ij} is the number of observed time points for the i^{th} individual's j^{th} trajectory.

Furthermore, $\boldsymbol{X}_{il}\left(t\right) = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{pmatrix}\left(t_{il}\right)$ and $X_{ijl}\left(t\right) = \sum_{k'=1}^{J} \xi_{ijk'} B_{k'}\left(t_{ijl}\right)$, where J is the number of basis knots for the individual trajectory functions, $\xi_{ijk'}$ is the k' individual basis knot's coefficient for the i^{th} individual's j^{th} individual trajectory function, and $B_{k'}\left(t\right)$ is the k' basis knot for the individual trajectory function. They let $\boldsymbol{\xi}_{ik'} \sim N\left(\boldsymbol{\beta}_{0k'} + \boldsymbol{Z}_i^T \boldsymbol{\zeta}, \boldsymbol{V}_{0k'}\right)$, where $\boldsymbol{\beta}_{0k'}$ and $\boldsymbol{V}_{0k'}$ are the p dimensional k'^{th} basis spline's mean and covariance-variance matrix, respectively, and both of which have priors. Also, $\boldsymbol{\zeta}$

is the coefficients for the baseline covariates, $\boldsymbol{Z}_i = \begin{pmatrix} Z_{i1} \\ Z_{i2} \\ \vdots \\ Z_{iq} \end{pmatrix}$ is the baseline covariates for the i^{th} individual where there are q baseline covariates, and Z_{ij} is the j^{th} baseline covariate for the i^{th} individual where $j = 1, \ldots, q$. Now $\boldsymbol{Y}_{il} \sim N_p\left(\boldsymbol{X}_{il}\left(t\right), \boldsymbol{\Sigma}\right)$ where $\boldsymbol{\Sigma}$ has a prior distribution.

The survival part of the model is defined as follows:

$$T_i \sim \lambda_i\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) = \lambda_0\left(t\right) e^{\boldsymbol{X}_i^T(t)\boldsymbol{\gamma} + \boldsymbol{Z}_i^T\boldsymbol{\zeta}},$$

where $\lambda_i\left(t|\boldsymbol{X}_{il}, \boldsymbol{Z}_i\right)$ is the hazard function for the i^{th} individual, $\lambda_0\left(t\right)$ is the baseline hazard function, $\boldsymbol{X}_i\left(t\right) = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{pmatrix}(t)$ is the vector of trajectory functions at survival time t, $\boldsymbol{\gamma}$ is the coefficients for the longitudinal trajectory functions, $\boldsymbol{\zeta}$ is the coefficients for the baseline covariates, $\boldsymbol{Z}_i = \begin{pmatrix} Z_{i1} \\ Z_{i2} \\ \vdots \\ Z_{iq} \end{pmatrix}$ is the baseline covariates for the i^{th} individual where there are q baseline covariates, and Z_{ij} is the j^{th} baseline covariate for the i^{th} individual where $j = 1, \ldots, q$.

To form the joint likelihood function for individual i, they used the product of the longitudinal and survival likelihoods, thus $f(\boldsymbol{Y}_i, T_i) = f(T_i|\boldsymbol{X}_i, \boldsymbol{Z}_i)\, f(\boldsymbol{Y}_i)$, where $f(\boldsymbol{Y}_i)$ is the longitudinal likelihood function and $f(T_i|\boldsymbol{X}_i, \boldsymbol{Z}_i)$ is the survival likelihood function.

Their model assumes the baseline hazard is piecewise constant, allows only observed and right censored data, and the same basis functions are used for the mean trajectory and the individual trajectories. Furthermore, they applied the trapezoidal rule to approximate their integral of the cumulative hazard function. Also, the model works when given specific prior distributions, such as $\boldsymbol{\Sigma}^{-1} \sim$ Wishart and $\boldsymbol{\beta}_{0j} \sim N_p(A_0, A_1)$ where A_0 and A_1 are user chosen to keep generality while maintaining proper and conjugate priors. Lastly, the authors stated that their coded version of the the model does not always converge.

1.4.3 Rice and Wu (2001)

In Brown et al. (2005), they reference Rice and Wu (2001) as a frequentist's example of a longitudinal model, which had the following form:

$$Y_{ijl} = X_{ij}(t_{ijl}) + \varepsilon_{ijl}, \text{ where } i = 1, \ldots, n, \ j = 1, \ldots, p, \text{ and } l = 1, \ldots, m_{ij}.$$

Now $X_{ij}(t_{ijl})$ is the j^{th} trajectory function for the i^{th} individual and t_{ijl} is the l^{th} observed time point for the i^{th} individual's j^{th} trajectory. Additionally, n is the number of individuals, p is the number of longitudinal trajectories, and m_{ij} is the number of observed time points for the i^{th} individual's j^{th} trajectory.

Furthermore, $\boldsymbol{X}_i(t) = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{pmatrix}(t)$ and $X_{ij}(t) = \psi_j(t) + \sum_{k'=1}^{J} \xi_{ijk'} B_{k'}(t)$ where $\psi_j(t)$ is the j^{th} mean trajectory function and is given by $\psi_j(t) = \sum_{k'=1}^{\overline{J}} \beta_{jk'} \overline{B}_{k'}(t)$. Lastly, $\overline{J}$ and J are the number of basis knots for the mean and individual trajectory functions, respectively, $\beta_{jk'}$ is the k' mean basis knot's coefficient for the j^{th} mean trajectory function, $\overline{B}_{k'}(t)$ is k' basis knot for the mean trajectory function, $\xi_{ijk'}$ is the k' individual basis knot's coefficient for the i^{th} individual's j^{th} individual trajectory function, and $B_{k'}(t)$ is the k' basis knot for the individual trajectory function.

This model allows for the flexibility in the number of basis knots between the mean and individual longitudinal trajectory functions. We adopted this structure in our longitudinal model.

1.4.4 Rizopoulos et al. (2009)

In Rizopoulos et al. (2009), they performed a joint survival-longitudinal model. The longitudinal model has the following form:

$$Y_{il} = X_i(t_{il}) + \varepsilon_{il}, \text{ where } i = 1, \ldots, n \text{ and } l = 1, \ldots, m_i.$$

Now $X_i(t_{il})$ is the trajectory function for the i^{th} individual and t_{il} is the l^{th} observed time point for the i^{th} individual's trajectory. Additionally, n is the number of individuals and m_i is the number of observed time points for the i^{th} individual's trajectory. Furthermore, $X_i(t) = \boldsymbol{x}_i(t)\boldsymbol{\beta} + \boldsymbol{z}_i(t)\boldsymbol{\xi}_i$ where $\boldsymbol{x}_i(t)$ and $\boldsymbol{z}_i(t)$ are the rows for the fixed and random effects for an individual, respectively. They state spline functions can be used for $\boldsymbol{x}_i(t)$ and $\boldsymbol{z}_i(t)$. Additionally, $\boldsymbol{\beta}$ and $\boldsymbol{\xi}_i$ are the coefficients for the fixed and random effects, respectively, where $i = 1, \ldots, n$.

For the survival part of the model, the survival times, T_i, have the following form:

$$T_i \sim \log \Lambda_i(t|X_i, \boldsymbol{Z}_i) = \log \Lambda_0(t) + X_i(t)\gamma + \boldsymbol{Z}_i^T \boldsymbol{\zeta},$$

where $\Lambda_i(t|X_i, \boldsymbol{Z}_i)$ is the cumulative hazard function for the i^{th} individual, $\Lambda_0(t)$ is the baseline cumulative hazard function, γ is the coefficient for the longitudinal trajectory function, $\boldsymbol{\zeta}$ is the coefficients for the baseline covariates, $\boldsymbol{Z}_i = \begin{pmatrix} Z_{i1} \\ Z_{i2} \\ \vdots \\ Z_{iq} \end{pmatrix}$ is the baseline covariates for the i^{th} individual where there are q baseline covariates, and Z_{ij} is the j^{th} baseline covariate for the i^{th} individual where $j = 1, \ldots, q$. The log baseline cumulative hazard function, $\log \Lambda_0(t)$, is modeled by a B-spline as follows:

$$\log \Lambda_0(t) = \varpi_0 + \sum_{k=1}^{m} \varpi_k B_k \{\log(t), q\},$$

where $\boldsymbol{\varpi}^T = (\varpi_0, \varpi_1, \ldots, \varpi_m)$ are the B-spline coefficients, $B\{\log(t), q\}$ is the B-spline basis knots and q is the degree of the B-spline. To maintain monotonicity of $\log \Lambda_0(t)$, the following reparameterization of ϖ is done:

$$\varpi_1 = \varpi_1^* \text{ and } \varpi_k = \varpi_{k-1} + \exp(\varpi_k^*),$$

where ϖ_k^* is the unconstrained parameter values and ϖ_k is the constrained parameter.

The joint model is $p(\boldsymbol{Y}_i, T_i) = \int p(T_i|X_i, \boldsymbol{Z}_i, \boldsymbol{\xi}_i) \, p(\boldsymbol{Y}_i|\boldsymbol{\xi}_i) \, p(\boldsymbol{\xi}_i) \, d\boldsymbol{\xi}_i$ where $p(T_i|X_i, \boldsymbol{Z}_i, \boldsymbol{\xi}_i)$ is the survival probability density function, $p(\boldsymbol{Y}_i|\boldsymbol{\xi}_i)$ is the longitudinal probability density function, and $p(\boldsymbol{\xi}_i)$ is the random effects probability density function, which they assumed was a normal density function.

This model does perform joint survival-longitudinal modeling; however, it only allows for one longitudinal covariate. Moreover, it requires constraining the baseline B-spline coefficients, $\boldsymbol{\varpi}$. Cai and Betensky (2003) on the other hand avoids having to due this by how they designed the cumulative hazard function.

1.4.5 Naik et al. (2007)

For determining how to cluster the data, we decided to follow the technique similar to Naik et al. (2007), even though they used a normal regression setting for their density framework. Since they used a normal regression setting, we use standard regression notation for terms not shared in our model. For their model, they assumed a finite-mixture regression model having the following density function:

$$f(y; \boldsymbol{x}, \phi) = \sum_{c=1}^{C} \omega_c f_c(y; \boldsymbol{x}, \boldsymbol{\beta}_c, \sigma_c),$$

where $0 \leq \omega_c \leq 1$, $\sum_{c=1}^{C} \omega_c = 1$, and $f_c(y; \boldsymbol{x}, \boldsymbol{\beta}_c, \sigma_c)$ is a normal density with mean $\boldsymbol{x}^T\boldsymbol{\beta}_c$ and variance σ_c. Now $\boldsymbol{x}$ is a $p \times 1$ vector of explanatory variables, which are considered fixed, $\boldsymbol{\beta}_c$ is a $p \times 1$ parameter vector, and $\phi = \{\omega_c, \boldsymbol{\beta}_c, \sigma_c \text{ for } c = 1, \ldots, C\}$. Letting $\boldsymbol{V}$ be an $n \times C$ indicator matrix with ic^{th} element, v_{ic}, equaling:

$$v_{ic} = \begin{cases} 1 & \text{if } i \text{ is in cluster } c. \\ 0 & \text{if } i \text{ is not in cluster } c. \end{cases}$$

Therefore, for a given set of data $(y_i, \boldsymbol{x}_i, v_i), i = 1, \ldots, n$, the complete data loglikelihood is:

$$\ell(\phi; \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{X}) = \sum_{c=1}^{C} \sum_{i=1}^{n} v_{ic} \left\{ \log(\omega_c) + \log f_c(y_i; \boldsymbol{x}_i, \boldsymbol{\beta}_c, \sigma_c) \right\},$$

where $\boldsymbol{Y} = (y_1, \ldots, y_n)^T$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ is an $n \times p$ matrix of explanatory vectors $\boldsymbol{x}_i$, and $\boldsymbol{V}$ is the $n \times C$ indicator matrix defined above.

They applied the expectation-maximization (EM) algorithm to estimate the model parameters. They let $\phi^{(m)} = \left\{ \omega_c^{(m)}, \boldsymbol{\beta}_c^{(m)}, \sigma_c^{(m)} \text{ for } c = 1, \ldots, C \right\}$ for the m^{th} iteration of the EM algorithm and defined $Q\left(\phi; \phi^{(m)}\right) = E\left[\ell | \boldsymbol{Y}, \boldsymbol{X}, \phi^{(m)}\right]$. In the E-step, $Q\left(\phi; \phi^{(m)}\right)$ is obtained by replacing v_{ic} with the expected value $\tau_{ic} = E\left[v_{ic} | y_i\right]$. τ_{ic} is given by

$$\tau_{ic}^{(m)} = \frac{\omega_c^{(m)} f_c\left(y_i; \boldsymbol{x}_i, \boldsymbol{\beta}_c^{(m)}, \sigma_c^{(m)}\right)}{\sum_{c=1}^{C} \omega_c^{(m)} f_c\left(y_i; \boldsymbol{x}_i, \boldsymbol{\beta}_c^{(m)}, \sigma_c^{(m)}\right)}.$$

In the M-step, they maximized $Q\left(\phi;\phi^{(m)}\right)$ with respect to $(\omega_c, \boldsymbol{\beta}_c, \sigma_c)$, giving closed-form estimates for the $(m+1)$ iteration of:

$$\begin{aligned}
\omega_c^{(m+1)} &= \sum_{i=1}^{n} \frac{\tau_{ic}^{(m)}}{n}, \\
\boldsymbol{\beta}_c^{(m+1)} &= \left(\tilde{\boldsymbol{X}}_c^{(m)T} \tilde{\boldsymbol{X}}_c^{(m)}\right)^{-1} \tilde{\boldsymbol{X}}_c^{(m)T} \tilde{\boldsymbol{Y}}_c^{(m)}, \text{and} \\
\sigma_c^{2(m+1)} &= \frac{\tilde{\boldsymbol{Y}}_c^{(m)T}\left(\boldsymbol{I} - \tilde{\boldsymbol{H}}_c^{(m)}\right)\tilde{\boldsymbol{Y}}_c^{(m)}}{\operatorname{tr}\left(\boldsymbol{W}_c^{(m)}\right)},
\end{aligned}$$

for $c = 1, \ldots, C$. Now $\boldsymbol{W}_c^{(m)} = \operatorname{diag}\left(\boldsymbol{\tau}_c^{(m)}\right)$, where $\boldsymbol{\tau}_c$ is the vector of probabilities that each individual belongs to cluster c, $\tilde{\boldsymbol{X}}_c^{(m)} = \boldsymbol{W}_c^{(m)1/2}\boldsymbol{X}, \tilde{\boldsymbol{Y}}_c^{(m)} = \boldsymbol{W}_c^{(m)1/2}\boldsymbol{Y}$, and $\tilde{\boldsymbol{H}}_c^{(m)} = \tilde{\boldsymbol{X}}_c^{(m)}\left(\tilde{\boldsymbol{X}}_c^{(m)T}\tilde{\boldsymbol{X}}_c^{(m)}\right)^{-1}\tilde{\boldsymbol{X}}_c^{(m)T}$.

The EM algorithm stops once the value of $\log\left\{f\left(\boldsymbol{Y};\boldsymbol{X},\phi^{(m+1)}\right)/f\left(\boldsymbol{Y};\boldsymbol{X},\phi^{(m)}\right)\right\}$ decreases below a pre-specified threshold. In order to initialize the EM algorithm, $\tau_{ic}^{(0)}$ for each individual is initialized by partitioning $\boldsymbol{X}$ into C clusters and they performed this initialization using the K-means clustering method. They do state that partitioning can be done either randomly or using the K-means clustering method.

They developed the Mixture Regression Criterion (MRC) and were able to write it in the following form:

$$MRC = \sum_{c=1}^{C} \hat{n}_c \log\left(\hat{\sigma}_c^2\right) + \sum_{c=1}^{C} \frac{\hat{n}_c\left(\hat{n}_c + p_c\right)}{\hat{n}_c - p_c - 2} - 2\sum_{c=1}^{C} \hat{n}_c \log\left(\hat{\omega}_c\right),$$

where $p_c = \operatorname{tr}\left(\hat{\boldsymbol{H}}_c\right)$ and $\hat{n}_c = \operatorname{tr}\left(\hat{\boldsymbol{W}}_c\right)$.

The authors state that the first term measures the lack of fit, the second term is a penalty for overfitting the model by adding more variables, and the third term is a clustering penalty function to mitigate for overclustering. When comparing models with different amounts of clusters and variables, the model with the lowest MRC is to be chosen.

The authors state that in order to determine how many clusters and variables to keep, they implemented a two-stage procedure. The first stage keeps all variables and determines the number of clusters. Once the number of clusters is decided, then the number of variables is decided for each cluster. The authors stated that simulation results showed the two-stage procedure worked well when compared to an exhaustive search.

The authors compared MRC to $AIC = -2\log f\left(\boldsymbol{Y};\boldsymbol{X},\hat{\phi}\right) + 2p$ and $BIC = -2\log f\left(\boldsymbol{Y};\boldsymbol{X},\hat{\phi}\right) + p\log(n)$. However, since these are the expressions for one cluster, to account for having C clusters and p variables per cluster, they replaced p with $d = (C-1) + C(p+1)$ when comparing to the MRC.

1.5 The Model

1.5.1 Longitudinal Submodel

The longitudinal portion of the model is defined as:

$$Y_{ijl} = X_{ij}\left(t_{ijl}\right) + \varepsilon_{ijl}, \text{ where } i = 1, \ldots, n,\ j = 1, \ldots, p, \text{ and } l = 1, \ldots, m_{ij}.$$

Now $X_{ij}\left(t_{ijl}\right)$ is the j^{th} trajectory function for the i^{th} individual and t_{ijl} is the l^{th} observed time point for the i^{th} individual's j^{th} trajectory. Additionally, n is the number of individuals, p is the number of longitudinal trajectories, and m_{ij} is the number of observed time points for the i^{th} individual's j^{th} trajectory.

Furthermore, $\boldsymbol{X}_i\left(t\right) = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{pmatrix}\left(t\right)$ and $X_{ij}\left(t\right) = \psi_j\left(t\right) + \sum_{k'=1}^{J} \xi_{ijk'} B_{k'}\left(t\right)$ where $\psi_j\left(t\right)$ is the j^{th} mean trajectory function and is given by $\psi_j\left(t\right) = \sum_{k'=1}^{\overline{J}} \beta_{jk'} \overline{B}_{k'}\left(t\right)$. Lastly, $\overline{J}$ and J are the number of basis knots for the mean and individual trajectory functions, respectively, $\beta_{jk'}$ is the k' mean basis knot's coefficient for the j^{th} mean trajectory function, $\overline{B}_{k'}\left(t\right)$ is k' basis knot for the mean trajectory function, $\xi_{ijk'}$ is the k' individual basis knot's coefficient for the i^{th} individual's j^{th} individual trajectory function, and $B_{k'}\left(t\right)$ is the k' basis knot for the individual trajectory function. We assume that each j^{th} set of individual trajectory coefficients have a shared covariance-variance matrix, $\Sigma_{\boldsymbol{\xi}_j}$. Additionally, $\varepsilon_{ijl} \sim N\left(0, \sigma^2_{Y_j}\right)$ and $\boldsymbol{\xi}_{ij}$ and $\boldsymbol{\varepsilon}_{ij}$ are independent.

1.5.2 Survival Submodel

For the survival portion of the model, the survival times, T_i, have the following model:

$$T_i \sim \lambda_i\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) = \lambda_0\left(t\right) e^{\boldsymbol{X}_i^T(t)\boldsymbol{\gamma}+\boldsymbol{Z}_i^T\boldsymbol{\zeta}},$$

where $\lambda_i\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ is the hazard function for the i^{th} individual, $\lambda_0\left(t\right)$ is the baseline hazard function, $\boldsymbol{\gamma}$ is the coefficients for the longitudinal trajectory functions, $\boldsymbol{X}_i$ is the p-dimensional vector of longitudinal trajectories for the i^{th} individual, $\boldsymbol{\zeta}$ is the coefficients for the baseline covariates, $\boldsymbol{Z}_i = \begin{pmatrix} Z_{i1} \\ Z_{i2} \\ \vdots \\ Z_{iq} \end{pmatrix}$ is the baseline covariates for the i^{th} individual where there are q baseline covariates, and Z_{ij} is the j^{th} baseline covariate for the i^{th} individual where $j = 1, \ldots, q$. For the baseline hazard function $\lambda_0\left(t\right)$, we used the following form:

$$\lambda_0\left(t\right) = e^{\alpha_0+\alpha_1 t+\sum_{k=1}^{K} b_k (t-\kappa_k)_+} \text{ or}$$
$$\log \lambda_0\left(t\right) = \eta_0\left(t\right) = \alpha_0 + \alpha_1 t + \sum_{k=1}^{K} b_k\left(t-\kappa_k\right)_+ ,$$

where $\left(t-\kappa_k\right)_+ \equiv \max\left(0, \left(t-\kappa_k\right)\right)$. In our spline model, we are going to treat the b_k's, the coefficient parameters for each of the k knots where $k = 1, \ldots, K$, as random effects distributed as follows:

$$\boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_K \end{bmatrix} \sim N_K(\mathbf{0}, \sigma_b^2 \boldsymbol{I}_K).$$

Our survival times, T_i, can be observed, left censored, right censored, or interval censored. Additionally, they can be left and right truncated. We denoted observed and right censored times by T_i^r and left censored times by T_i^l. For truncation, we denote left truncated times by L_i and right truncated times by R_i.

1.5.3 Clustering Submodel

When there are clusters, the survival times, T_i, have the following model:

$$T_i \sim \sum_{c=1}^{C} v_{ic}\lambda_{ic}\left(t|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right) = \sum_{c=1}^{C} v_{ic}\lambda_0\left(t\right) e^{\boldsymbol{X}_{ic}^T(t)\boldsymbol{\gamma}_c + \boldsymbol{Z}_i^T \boldsymbol{\zeta}_c},$$

where $\lambda_{ic}\left(t|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)$ is the hazard function for the i^{th} individual for cluster c, $\lambda_0\left(t\right)$ is the baseline hazard function, $\boldsymbol{X}_{ic}$ is the p-dimensional vector of longitudinal trajectories for the i^{th} individual for cluster c, $\boldsymbol{\gamma}_c$ is the coefficients for the longitudinal trajectory functions for cluster c, $\boldsymbol{\zeta}_c$ is the coefficients for the baseline covariates for cluster c where $c = 1, \ldots, C$ and C is the number of clusters.

The longitudinal part of the model, if it has clustering, has the following form:

$$Y_{ijl} = \sum_{c=1}^{C} v_{ic}\left(X_{ijc}\left(t_{ijl}\right) + \varepsilon_{ijlc}\right), \text{ where } i = 1, \ldots, n,$$
$$j = 1, \ldots, p,\ l = 1, \ldots, m_{ij}, \text{ and } c = 1, \ldots, C.$$

Now $X_{ijc}\left(t_{ijl}\right)$ is the j^{th} trajectory function for cluster c for the i^{th} individual and t_{ijl} is the l^{th} observed time point for the i^{th} individual's j^{th} trajectory. Additionally, n is the number of individuals, p is the number of longitudinal trajectories, m_{ij} is the number of observed time points for the i^{th} individual's j^{th} trajectory.

Now,

$$v_{ic} = \begin{cases} 1 & \text{if } i \in c. \\ 0 & \text{if } i \notin c. \end{cases}$$

Furthermore, $\boldsymbol{X}_{ic}(t) = \begin{pmatrix} X_{i1c} \\ X_{i2c} \\ \vdots \\ X_{ipc} \end{pmatrix}(t)$ and $X_{ijc}(t) = \psi_{jc}(t) + \sum_{k'=1}^{J} \xi_{ijck'} B_{k'}(t)$ where $\psi_{jc}(t)$ is the j^{th} mean trajectory function for cluster c and is given by $\psi_{jc}(t) = \sum_{k'=1}^{\overline{J}} \beta_{jck'} \overline{B}_{k'}(t)$. Lastly, $\overline{J}$ and J are the number of basis knots for the mean and individual trajectory functions, respectively, $\beta_{jck'}$ is the k' mean basis knot's coefficient for the j^{th} mean trajectory function for cluster c, $\overline{B}_{k'}(t)$ is k' basis knot for the mean trajectory function, $\xi_{ijck'}$ is the k' individual basis knot's coefficient for the i^{th} individual's j^{th} individual trajectory function for cluster c, and $B_{k'}(t)$ is the k' basis knot for the individual trajectory function.

1.5.4 Joint Model

The complete data likelihood for the joint model is:

$$
\begin{aligned}
&L\left(\Theta;\boldsymbol{T}^l,\boldsymbol{T}^r,\boldsymbol{L},\boldsymbol{R},\boldsymbol{\delta}_O,\boldsymbol{\delta}_R,\boldsymbol{\delta}_L,\boldsymbol{\delta}_I,\boldsymbol{Z},\boldsymbol{Y},\boldsymbol{V},\boldsymbol{\xi}\right)=\\
&\prod_{i=1}^{n}\prod_{c=1}^{C}\left\{\omega_c\left(\frac{\lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)e^{-\Lambda_c(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{O_i}}\left(\frac{e^{-\Lambda_c(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{R_i}}\right.\\
&\qquad\times\left(\frac{1-e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{L_i}}\left(\frac{e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}-e^{-\Lambda_c(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{I_i}}\\
&\qquad\left.\times\prod_{j=1}^{p}\left(\sigma^2_{Y_jc}\right)^{-\frac{m_{ij}}{2}}e^{\left(-\|\boldsymbol{Y}_{ij}-\boldsymbol{X}_{ijc}\|^2/2\sigma^2_{Y_jc}\right)}\det\left(\Sigma_{\boldsymbol{\xi}_jc}\right)^{-\frac{1}{2}}e^{-\frac{\boldsymbol{\xi}_{ij}^T\Sigma^{-1}_{\boldsymbol{\xi}_jc}\boldsymbol{\xi}_{ij}}{2}}\right\}^{v_{ic}},
\end{aligned}
$$

where $\Theta=\left(\alpha_0,\alpha_1,\boldsymbol{b},\omega_c,\boldsymbol{\gamma}_c,\boldsymbol{\zeta}_c,\sigma_b^2,\boldsymbol{\sigma}_Y^2,\boldsymbol{\beta}_{jc},\Sigma_{\boldsymbol{\xi}_jc}\right)$, $c=1,\dots,C$, and $j=1,\dots,p$. Now T_i^l and T_i^r are the left and right time to event values, respectively, L_i and R_i are the left and right truncation times, respectively,

$$\delta_{O_i}=\begin{cases}1 \text{ if } T_i \text{ is observed.}\\ 0 \text{ otherwise.}\end{cases},\qquad \delta_{R_i}=\begin{cases}1 \text{ if } T_i \text{ is right censored.}\\ 0 \text{ otherwise.}\end{cases},$$

$$\delta_{L_i}=\begin{cases}1 \text{ if } T_i \text{ is left censored.}\\ 0 \text{ otherwise.}\end{cases},\qquad \delta_{I_i}=\begin{cases}1 \text{ if } T_i \text{ is interval censored.}\\ 0 \text{ otherwise.}\end{cases},$$

and $\Lambda_c\left(t|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)=\int_0^t\lambda_c\left(t|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)$ is the cumulative hazard function. Additionally, $\lambda_c\left(t|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)=\lambda_0\left(t\right)e^{\boldsymbol{X}_{ic}^T\boldsymbol{\gamma}_c+\boldsymbol{Z}_i^T\boldsymbol{\zeta}_c}$, where $\lambda_0\left(t\right)$ is represented by a linear spline model. The model is:

$$\lambda_0\left(t\right)=e^{\alpha_0+\alpha_1 t+\sum_{k=1}^{K}b_k(t-\kappa_k)_+},$$

where K is the number of basis knots and length of vector $\boldsymbol{b}$, κ_k is the k^{th} basis knot, $\boldsymbol{b}$ is the vector of b_k's, where b_k is the coefficient for the k^{th} basis knot, α_0 is the intercept term, and α_1 is the coefficient for $\kappa_0=0$. Lastly, $\boldsymbol{\gamma}_c$ is the coefficients for the time-dependent covariates, $\boldsymbol{X}$, per cluster c, $\boldsymbol{\zeta}_c$ is the coefficients for the baseline covariates, $\boldsymbol{Z}$, per cluster c, and $\boldsymbol{X}_{ic}$ is the p-dimensional vector of longitudinal trajectory functions for individual i for cluster c, $c = 1, \ldots, C$. Note: $(t - \kappa_k)_+ \equiv \max(0, (t - \kappa_k))$ and when the time value is observed, $T_i^l = T_i^r = T_i$ and thus T_i^r is used.

Additionally, since we now have clustering in this model, we have that $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_C)$ is the vector of cluster proportions, where C is the number of clusters, ω_c is the cluster proportion for cluster c, $0 \leq \omega_c \leq 1$, and $\sum_{c=1}^{C} \omega_c = 1$. Now $\boldsymbol{V}$ is the $n \times C$ matrix of indicator values were:

$$v_{ic} = \begin{cases} 1 & \text{if } i \in c. \\ 0 & \text{if } i \notin c. \end{cases}$$

Chapter 2

Survival Analysis with Time Varying Covariates Under Various Censoring & Truncation Schemes

2.1 Forming The Loglikelihood

The likelihood function is given by:

$$L\left(\boldsymbol{\theta};\sigma_b^2\right)=\prod_{i=1}^{n}\left(\frac{\lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)e^{-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}{e^{-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}-e^{-\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}}\right)^{\delta_{O_i}}\left(\frac{e^{-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}{e^{-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}-e^{-\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}}\right)^{\delta_{R_i}}$$
$$\times\left(\frac{1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}{e^{-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}-e^{-\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}}\right)^{\delta_{L_i}}\left(\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-e^{-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}{e^{-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}-e^{-\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}}\right)^{\delta_{I_i}},$$

where $\boldsymbol{\theta}=\left(\boldsymbol{\alpha}^T,\boldsymbol{b}^T,\boldsymbol{\gamma}^T,\boldsymbol{\zeta}^T\right)^T$ with $\boldsymbol{\alpha}=(\alpha_0,\alpha_1)^T$, T_i^l and T_i^r are the left and right time to event values, respectively, L_i and R_i are the left and right truncation times, respectively,

$$\delta_{O_i} = \begin{cases} 1 \text{ if } T_i \text{ is observed.} \\ \\ 0 \text{ otherwise.} \end{cases}, \qquad \delta_{R_i} = \begin{cases} 1 \text{ if } T_i \text{ is right censored.} \\ \\ 0 \text{ otherwise.} \end{cases},$$

$$\delta_{L_i} = \begin{cases} 1 \text{ if } T_i \text{ is left censored.} \\ \\ 0 \text{ otherwise.} \end{cases}, \qquad \delta_{I_i} = \begin{cases} 1 \text{ if } T_i \text{ is interval censored.} \\ \\ 0 \text{ otherwise.} \end{cases},$$

and $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) = \int_0^t \lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ is the cumulative hazard function. Additionally, $\lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) = \lambda_0\left(t\right) e^{\boldsymbol{X}_i^T \boldsymbol{\gamma} + \boldsymbol{Z}_i^T \boldsymbol{\zeta}}$, where $\lambda_0\left(t\right)$ is represented by a linear spline model. The model is:

$$\lambda_0\left(t\right) = e^{\alpha_0 + \alpha_1 t + \sum_{k=1}^{K} b_k (t - \kappa_k)_+},$$

where K is the number of basis knots and length of vector $\boldsymbol{b}$, κ_k is the k^{th} basis knot, $\boldsymbol{b}$ is the vector of b_k's, where b_k is the coefficient for the k^{th} basis knot, α_0 is the intercept term, and α_1 is the coefficient for $\kappa_0 = 0$. Lastly, $\boldsymbol{\gamma}$ is the coefficients for the time-dependent covariates, $\boldsymbol{X}$, and $\boldsymbol{\zeta}$ is the coefficients for the baseline covariates, $\boldsymbol{Z}$. Note: $(t - \kappa_k)_+ \equiv \max\left(0, (t - \kappa_k)\right)$ and when the time value is observed, $T_i^l = T_i^r = T_i$ and thus T_i^r is used.

Now the loglikelihood is:

$$
\begin{aligned}
\ell_0 =& \sum_{i=1}^{n} \Big[\log \left(\lambda_0 \left(T_i^r \right) e^{\boldsymbol{X}_i^T \left(T_i^r \right) \boldsymbol{\gamma} + \boldsymbol{Z}_i^T \boldsymbol{\zeta}} \right) \delta_{O_i} + \log \left(e^{-\Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} \right) \delta_{O_i} + \\
& \log \left(e^{-\Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} \right) \delta_{R_i} + \log \left(1 - e^{-\Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} \right) \delta_{L_i} + \\
& \log \left(e^{-\Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} - e^{-\Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} \right) \delta_{I_i} - \\
& \log \left(e^{-\Lambda \left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} - e^{-\Lambda \left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} \right) \left(\delta_{O_i} + \delta_{R_i} + \delta_{L_i} + \delta_{I_i} \right) \Big] \\
=& \sum_{i=1}^{n} \Big[\log \left(\lambda_0 \left(T_i^r \right) \right) \delta_{O_i} + \left(\boldsymbol{X}_i^T \left(T_i^r \right) \boldsymbol{\gamma} + \boldsymbol{Z}_i^T \boldsymbol{\zeta} \right) \delta_{O_i} - \Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) \delta_{O_i} - \\
& \Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) \delta_{R_i} + \log \left(1 - e^{-\Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} \right) \delta_{L_i} - \\
& \Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) \delta_{I_i} + \log \left(1 - e^{\Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) - \Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} \right) \delta_{I_i} + \\
& \Lambda \left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) - \log \left(1 - e^{\Lambda \left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) - \Lambda \left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} \right) \Big] \\
=& \sum_{i=1}^{n} \Big[\eta_0 \left(T_i^r \right) \delta_{O_i} + \left(\boldsymbol{X}_i^T \left(T_i^r \right) \boldsymbol{\gamma} + \boldsymbol{Z}_i^T \boldsymbol{\zeta} \right) \delta_{O_i} - \Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) \left(\delta_{O_i} + \delta_{R_i} \right) + \\
& \log \left(1 - e^{-\Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} \right) \delta_{L_i} - \Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) \delta_{I_i} + \\
& \log \left(1 - e^{\Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) - \Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} \right) \delta_{I_i} + \\
& \Lambda \left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) - \log \left(1 - e^{\Lambda \left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) - \Lambda \left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i \right)} \right) \Big],
\end{aligned}
$$

where $\delta_{O_i} + \delta_{R_i} + \delta_{L_i} + \delta_{I_i} = 1$ for $i = 1, \ldots, n$. Since we want to estimate $\boldsymbol{\theta} = \left(\boldsymbol{\alpha}^T, \boldsymbol{b}^T, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T \right)^T$, we will use the penalized loglikelihood, which is given by:

$$
\ell_p \left(\boldsymbol{\theta}; \sigma_b^2 \right) = \ell_0 \left(\boldsymbol{\theta}; \sigma_b^2 \right) - \frac{\boldsymbol{b}^T \boldsymbol{b}}{2 \sigma_b^2},
$$

where this means that the variance, σ_b^2, controls the amount of smoothing. Under the given mixed model, the loglikelihood becomes:

$$\ell\left(\left(\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T\right)^T; \sigma_b^2\right) = -\frac{K}{2}\log\sigma_b^2 + \int \ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right) d\boldsymbol{b}.$$

Since this is a K-dimensional intractable integral, we adapt a penalized quasilikelihood (PQL) approach to approximate $\ell\left(\left(\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T\right)^T; \sigma_b^2\right)$ giving:

$$\ell\left(\widehat{\boldsymbol{\theta}}; \sigma_b^2\right) \simeq -\frac{K}{2}\log\sigma_b^2 + \ell_p\left(\widehat{\boldsymbol{\theta}}; \sigma_b^2\right).$$

The derivation and a brief explanation of PQL is given in Appendix B. In order to estimate $\ell\left(\widehat{\boldsymbol{\theta}}; \sigma_b^2\right)$, we must find $\boldsymbol{Q}$ and $\mathbb{Q}$, where $\boldsymbol{Q}$ is the $(2+K+p+q)\times 1$ vector of first-order partial derivatives of $\ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right)$ and $\mathbb{Q}$ is the $(2+K+p+q)\times(2+K+p+q)$ matrix of the second-order partial derivatives of $\ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right)$. The first step to finding $\boldsymbol{Q}$ and $\mathbb{Q}$ is finding the first and second partial derivatives of $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$.

2.2 $\Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i)$ & 1^{st} Derivatives of $\Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i)$

We first decided to divide up the space by knot segments to integrate over $\Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i)$. Therefore,

$$\begin{aligned}\Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i) &= \sum_{k=0}^{k_t^*-1} \int_{\kappa_k}^{\kappa_{k+1}} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}+\boldsymbol{Z}_i^T\boldsymbol{\zeta}} du \\ &\quad + \int_{\kappa_{k_t^*}}^{t} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}+\boldsymbol{Z}_i^T\boldsymbol{\zeta}} du \\ &= e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} \sum_{k=0}^{k_t^*-1} \int_{\kappa_k}^{\kappa_{k+1}} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}} du \\ &\quad + e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} \int_{\kappa_{k_t^*}}^{t} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}} du,\end{aligned}$$

where $\kappa_0 = 0$ and $k_t^* = \max(k : \kappa_k < t, 1 \le k \le K)$.

The derivatives of Λ with respect to α_0 and b_j are given here. All the derivatives are found in Section C.1 of the Appendix.

$$\begin{aligned}\frac{\partial}{\partial\alpha_0}\Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i) &= \Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i) \\ \frac{\partial}{\partial b_j}\Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i) &= e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} I(t > \kappa_j) \sum_{k=0}^{k_t^*-1} \int_{\kappa_k}^{\kappa_{k+1}} (u-\kappa_j)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}} du \\ &\quad + e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} I(t > \kappa_j) \int_{\kappa_{k_t^*}}^{t} (u-\kappa_j)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}} du,\end{aligned}$$

where b_j is the j^{th} element of the coefficient vector for the knots where $j = 1, \ldots, K$. Since the integrals do not have analytic solutions when the degree of $\boldsymbol{X}_i^T(t)$ is greater than one, we use Simpson's Rule to integrate the derivatives.

2.3 1^{st} Derivatives of The Loglikelihood

Using the above results, we find the first derivatives of $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)=\ell_0-\frac{\boldsymbol{b}^T\boldsymbol{b}}{2\sigma_b^2}$, where $\ell_0=\sum_{i=1}^{n}\left[\eta_0\left(T_i^r\right)\delta_{O_i}+\left(\boldsymbol{X}_i^T\left(T_i^r\right)\boldsymbol{\gamma}+\boldsymbol{Z}_i^T\boldsymbol{\zeta}\right)\delta_{O_i}-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\right.$ $\log\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\delta_{L_i}-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}+$ $\log\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\delta_{I_i}+\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-$ $\left.\log\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\right]$. The first derivative with respect to α_0 is given here and all the derivatives are found in Section C.2 of the Appendix.

$$
\begin{aligned}
\frac{\partial}{\partial\alpha_0}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial}{\partial\alpha_0}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right)\\
&= \sum_{i=1}^{n}\left[\delta_{O_i}-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)\right.\\
&\quad+\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\delta_{L_i}-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\delta_{I_i}\\
&\quad+\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\\
&\quad\left.+\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\right]\\
&= \sum_{i=1}^{n}\left[\delta_{O_i}-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\right.\\
&\quad-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}-\frac{\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{I_i}\\
&\quad\left.+\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\right].
\end{aligned}
$$

The final equation for the derivative is better computationally and comes from multiplying by a form of 1 on certain terms. For example, for the left censor term we multiplied by $\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}$.

2.4 2^{nd} Derivatives of $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$

The second derivatives of $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ with respect to α_1 are given here. All the derivatives can be found in Section C.3 of the Appendix.

$$
\begin{aligned}
\frac{\partial^2}{\partial \alpha_1^2}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) &= e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} u^2 e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du \\
&+ e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t} u^2 e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du, \\
\frac{\partial^2}{\partial \alpha_1 \partial b_j}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) &= e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} u\left(u-\kappa_j\right)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du \\
&+ e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t} u\left(u-\kappa_j\right)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du, \\
\frac{\partial^2}{\partial \alpha_1 \partial \gamma_j}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) &= e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} uX_{ij}\left(u\right) e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du \\
&+ e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t} uX_{ij}\left(u\right) e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du, \\
\frac{\partial^2}{\partial \alpha_1 \partial \zeta_j}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) &= Z_{ij}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} u e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du \\
&+ Z_{ij}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t} u e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du.
\end{aligned}
$$

Now, b_j is the j^{th} element of the coefficient vector for the baseline knots where $j = 1, \ldots, K$, γ_j is the j^{th} element of the coefficient vector for the trajectory functions where $j = 1, \ldots, p$, $X_{ij}\left(t\right)$ is the j^{th} trajectory function for the i^{th} individual at time t where $j = 1, \ldots, p$, ζ_j is the j^{th} element of the covariate coefficient vector where $j = 1, \ldots, q$, and Z_{ij} is the j^{th} element of the covariate vector for the i^{th} individual where $j = 1, \ldots, q$. Since again the integrals do not have analytic solutions when the degree of $\boldsymbol{X}_i^T(t)$ is greater than one, we use Simpson's Rule to integrate the derivatives.

2.5 2^{nd} Derivatives of the Loglikelihood

Using the above results, we find the second derivatives for $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) = \ell_0 - \frac{\boldsymbol{b}^T\boldsymbol{b}}{2\sigma_b^2}$, where $\ell_0 = \sum_{i=1}^n \left[\eta_0\left(T_i^r\right)\delta_{O_i} + \left(\boldsymbol{X}_i^T\left(T_i^r\right)\boldsymbol{\gamma} + \boldsymbol{Z}_i^T\boldsymbol{\zeta}\right)\delta_{O_i} - \Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) + \log\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\delta_{L_i} - \Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} + \log\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\delta_{I_i} + \Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) - \log\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\right]$. One of the the second derivatives with respect to α_0 is given here. All the second derivatives of the loglikelihood are found in

Section C.4 of the Appendix.

$$
\begin{aligned}
\frac{\partial^2}{\partial \alpha_0^2}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial \alpha_0^2}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^n \Bigg[-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) \\
&\quad + \frac{\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} - \frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i} \\
&\quad - \frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} - \frac{\left(\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{I_i} \\
&\quad + \frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad + \frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\Bigg] \\
&= \sum_{i=1}^n \Bigg[-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) \\
&\quad + \frac{\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} - \frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad - \frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} - \frac{\left(\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{I_i} \\
&\quad + \frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad + \frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\Bigg].
\end{aligned}
$$

The final equation is better computationally and comes from multiplying by a form of 1 on certain terms. For example, for the second left censor term we multiplied by $\left(\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\right)^2$.

2.6 Estimating the Smoothing Parameter

Since σ_b^2 controls the amount of smoothing, we can automatically select the smoothness by using the marginal loglikelihood. Cai and Betensky (2003) stated that although a restricted maximum likelihood (REML) is sought for estimating the variance component, it is not well defined for a non-Gaussian mixed model. However, Harville (1974) showed that the REML for Gaussian models is the same as the marginal likelihood when the regression parameters are integrated with a flat prior. Therefore, we need to maximize the marginal loglikelihood, given by:

$$\ell_{marg}\left(\sigma_b^2\right) = -\frac{K}{2}\log\left(\sigma_b^2\right) + \log\int \exp\left[\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)\right] d\boldsymbol{\theta}.$$

We apply Laplace's method to approximate $\ell_{marg}\left(\sigma_b^2\right)$, leading to:

$$\ell_{marg}\left(\sigma_b^2\right) \simeq -\frac{K}{2}\log\left(\sigma_b^2\right) + \ell_p\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\} - \frac{1}{2}\log\left|-\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right|,$$

where $\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right)$ is the solution to $\boldsymbol{Q}\left(\boldsymbol{\theta};\sigma_b^2\right) = \mathbf{0}$. Additionally, $\boldsymbol{Q}$ is the $(2+K+p+q)\times 1$ vector of first-order partial derivatives of $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)$ and $\mathbb{Q}$ is the $(2+K+p+q)\times(2+K+p+q)$ matrix of the second-order partial derivatives of $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)$. By maximizing the approximated marginal loglikelihood we obtain an estimate for σ_b^2. We used Newton optimization to find the σ_b^2 that maximizes the marginal loglikelihood. The short and detailed derivations of the marginal loglikelihood can be found in Appendix D. The derivations for its first and second derivatives of the marginal loglikelihood can be found in Appendix E.

Since an exhaustive method for trying every set of $\boldsymbol{\theta}$ and every σ_b^2 is not feasible, we decided to perform an iterative estimating between $\boldsymbol{\theta}$ and σ_b^2. The steps are:

1. Start with initial estimates of $\boldsymbol{\theta}$ and σ_b^2.

2. Conduct one iteration of Newton optimization for $\boldsymbol{\theta}$ with fixed σ_b^2.

3. With updated $\boldsymbol{\theta}$ estimate, find the σ_b^2 that maximizes $\ell_{marg}\left(\sigma_b^2\right)$.

4. Repeat steps 2 and 3 until the difference between the previous σ_b^2 and current σ_b^2 is less than a specific threshold.

5. If the initial estimates for $\boldsymbol{\theta}$ are not good starting values, it can cause σ_b^2 to tend towards 0. To stabilize the estimation, the number of iterations for $\boldsymbol{\theta}$ are increased from 1 to 10. From simulation experience, this change does not effect the results for σ_b^2.

2.7 Simulation Setting

2.7.1 Data Simulation Structure

The following structure is used for simulating data:

n # of individuals.

q # of baseline covariates.

p # of longitudinal trajectory functions.

$f(t)$ The baseline hazard function.

K The number of knots, in our case we used the min $\left(\left\lfloor \frac{n}{4} \right\rfloor, 30\right)$ as in Cai and Betensky (2003).

κ_k's The baseline knots where $k = 1, \ldots, K$. Chosen based on data and usually spaced on unique quantiles of the survival times T_i^l's, T_i^r's, and $\frac{T_i^l + T_i^r}{2}$'s.

α_0 The intercept of the estimated baseline hazard function.

α_1 The κ_0 coefficient of the estimated baseline hazard function. Note: $\kappa_0 = 0$.

b_k's Distributed $N_K(\mathbf{0}, \sigma_b^2 \boldsymbol{I}_K)$ where $k = 1, \ldots, K$. They are the coefficients of the $(t - \kappa_k)_+$'s, $k = 1, \ldots, K$. Used in estimating the baseline hazard function.

$\boldsymbol{\zeta}$ The coefficients for the baseline covariates. Chosen by the user.

$\boldsymbol{\gamma}$ The coefficients for the longitudinal trajectory functions. Chosen by the user.

$\boldsymbol{Z}$ The baseline covariate matrix that is $n \times q$.

$\boldsymbol{X}_i(t)$ The set of p longitudinal trajectory functions for the i^{th} individual. Reminder that $X_{ij}(t) = \psi_j(t) + \sum_{k'=1}^{J} \xi_{ijk'} B_{k'}(t)$ where $j = 1, \ldots, p$.

Additionally, the j^{th} mean trajectory function is given by

$\psi_j(t) = \sum_{k'=1}^{\overline{J}} \beta_{jk'} \overline{B}_{k'}(t)$.

$\boldsymbol{\beta}_j$'s The mean longitudinal basis knots' coefficients.

$\boldsymbol{\xi}_{ij}$'s Distributed $N_J\left(\boldsymbol{\mu}_{\boldsymbol{\xi}_j}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_j}\right)$ where J is the number of individual longitudinal basis knots and $\boldsymbol{\mu}_{\boldsymbol{\xi}_j}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_j}$ are mean and covariance-variance matrix for the $\boldsymbol{\xi}_{ij}$'s. For the simulation, since the trajectory functions are all known, we decided to let the individual longitudinal coefficients all have the same mean and covariance-variance matrix.

The covariance-variance matrix we decided on is:

$$\Sigma_{\boldsymbol{\xi}_j} = \sigma^2_{\xi_j} \begin{pmatrix} 1 & \rho_{\xi_j} & \rho_{\xi_j} & \cdots & \rho_{\xi_j} \\ \rho_{\xi_j} & 1 & \rho_{\xi_j} & \cdots & \rho_{\xi_j} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{\xi_j} & \rho_{\xi_j} & \rho_{\xi_j} & \cdots & 1 \end{pmatrix}$$

for each j, where $\sigma^2_{\xi_j}$ is the variance and ρ_{ξ_j} is the correlation.

2.7.2 Simulating Actual Times

To simulate the survival times, T_i, $i = 1, \ldots, n$, we used the fact that $S(t|\boldsymbol{X}_i, \boldsymbol{Z}_i) = e^{-\Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i)}$, where $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) = \int_0^t \lambda\left(t\right)$. This method is similar to the one used by Brown, Ibrahim, and DeGruttola (2005) to simulate their joint survival-longitudinal data. The method is:

1. Simulate a random survival probability, $s \sim U(0,1)$.

2. Use R's *Uniroot* function to solve $\log\left(s\right) + \Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ for where the function is 0 along with using R's *Integrate* function to integrate $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ from 0 to the value of t such that the above equation is true.

2.7.3 Simulation Study Setup

We conducted a simulation study where we performed 200 replicates for both 200 and 500 individuals. We simulated each replicate in R with a different seed from $i = 0, \ldots, 199$ and chose the following values for the inputs and parameters:

n 200 or 500.

q 2.

p 2.

$f\left(t\right)$ $\frac{(t-20)^4}{10^8} + 0.05$.

K 30, since $\left\lfloor \frac{n}{4} \right\rfloor > 30$.

κ_k's	Changed with each of the 200 replicates for each sample size n. We used the unique quantiles of the survival times T_i^l's, T_i^r's, and $\frac{T_i^l+T_i^r}{2}$'s.
α_0	Estimated in each of the replicates.
α_1	Estimated in each of the replicates.
b_k's	Estimated in each of the replicates and 30 used since $K = 30$.
$\boldsymbol{\zeta}$	The coefficients are 1.1 and 0.9.
$\boldsymbol{\gamma}$	The coefficients are 1 and 1.2.
$\boldsymbol{Z}$	The matrix that is $n \times 2$ of the time independent covariates. Each $Z_{ij} \sim Beta(2,2) * 4 - 2$ for $i = 1, \dots, n$ and $j = 1, 2$. Unlike a normal distribution, this beta only allows for values between -2 and 2 and is still mound-shaped symmetric.
$\boldsymbol{X}_i\left(t\right)$	Is $\psi_j\left(t\right) + \sum_{k'=1}^{J} \xi_{ijk'} B_{k'}\left(t\right)$ where $\overline{J} = J = 8$ since we have 6 distinct knots at 0, 20, 40, 60, 80, and 100 and are using a cubic B-spline. Additionally, the j^{th} mean trajectory function is given by $\psi_j\left(t\right) = \sum_{k'=1}^{\overline{J}} \beta_{jk'} \overline{B}_{k'}\left(t\right)$.

$$\boldsymbol{\beta} \quad \begin{bmatrix} -1.41 & -0.80 & -0.25 & -0.49 & 0.30 & 0.49 & 0.76 & 1.42 \\ \\ 0.42 & 1.17 & 0.11 & -0.29 & -0.38 & -0.32 & -0.69 & -1.21 \end{bmatrix}.$$

$$\boldsymbol{\xi}_{ij}\text{'s} \quad \text{Distributed } N_J\left(\boldsymbol{\mu}_{\boldsymbol{\xi}_j} = \mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_j} = 3\boldsymbol{I}\right).$$

This gave us survival times between $0 - 100$ years old with baseline hazard decreasing until 20 and then increasing from then on. Given in Figures 2.1-2.2 are plots of the survival times for $n = 200$ and $n = 500$ for one replicate.

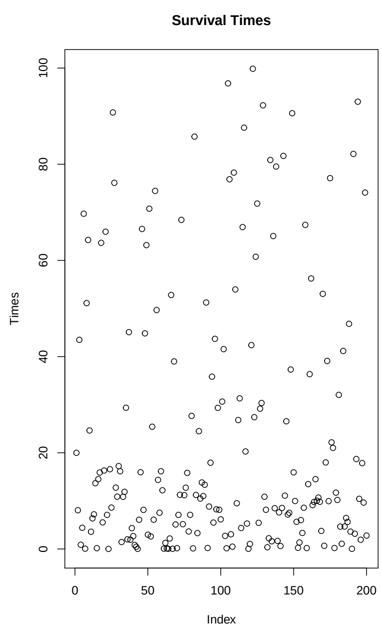

Figure 2.1: Plot of 200 survival times

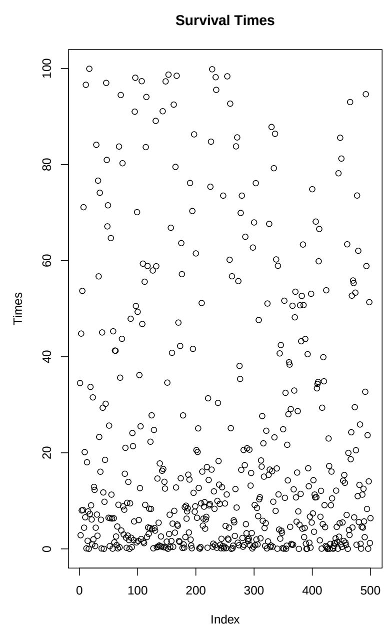

Figure 2.2: Plot of 500 survival times

2.7.4 Censoring Time's Design and Values

We performed five different levels of censoring. They were 0% censored, 50% censored, 75% censored, 90% censored, and 100% censored. We let T_i^l and T_i^r be the left and right time values, respectively, for the i^{th} individual. We let C_l and C_r be the left and right censoring times, respectively, for the data sets. For interval censored data, we set the T_i^l and T_i^r to be the floor and ceiling, respectively, of the observed time value. For instance, if $T_i = 2.4$, we let $T_i^l = 2$ and $T_i^r = 3$. To control the amount of censoring, we used a uniform random variable, U, and set the threshold accordingly. We only allowed observed values that were above C_l and below C_r. Therefore, we used U with the observed data and any remaining times that were not right censored and left censored became interval censored. For example, if we wanted 75% of the data censored, then we set the threshold for $U < .25$ and only chose times that were between C_l and C_r. Therefore, once the 25% randomly selected times were found, the remaining times that were not greater than C_r and less than C_l were interval censored to achieve 75% censoring. The design for the 75% is:

$T_i^l = T_i^r = T_i$ if $(U < .25 \;\&\; C_l < T_i \;\&\; C_r > T_i)$

$T_i^l = T_i^r = C_r$ if $(C_r < T_i)$

$T_i^l = T_i^r = C_l$ if $(C_l > T_i)$

$T_i^l = \lfloor T_i \rfloor$, $T_i^r = \lceil T_i \rceil$ otherwise.

In our simulation study, we set $C_l = 1$ and $C_r = 75$. This seemed to be realistic of the human race.

2.7.5 Performing the Simulation Study

Given that we are assuming that the longitudinal trajectory functions are known, we use the true trajectory function, $X_{ij}(t) = \psi_j(t) + \sum_{k'=1}^{J} \xi_{ijk'} B_{k'}(t)$. When conducting the actual simulation study, we chose starting values that we believed would be reasonable for α_0, α_1, b_k's, ζ's, and γ's and performed Newton optimization. However, even if poor starting values were chosen, the method finds the values that maximize the likelihood function. We used the method described above to find both the best estimate of σ_b^2 as well as the best estimates for $\boldsymbol{\theta}$. The simulations with 200 individuals took under 10 minutes per replicate and the simulations with 500 individuals took under 20 minutes per replicate. This speed is probably due to the fact that all of the coding is done using Rcpp and all functions are written in c++.

2.8 Results

Parameter Simulation Results

p_{cens}		$n = 200$				$n = 500$			
		RBias	ECP	MESE	ESE	RBias	ECP	MESE	ESE
0	$\gamma_1(1.0)$	0.00	0.97	0.07	0.07	-0.01	0.94	0.05	0.04
	$\gamma_2(1.2)$	0.00	0.97	0.08	0.08	-0.02	0.95	0.05	0.05
	$\zeta_1(1.1)$	-0.01	0.92	0.10	0.10	-0.03	0.91	0.06	0.06
	$\zeta_2(0.9)$	-0.03	0.92	0.09	0.10	-0.03	0.93	0.06	0.06
50	$\gamma_1(1.0)$	0.01	0.97	0.08	0.08	0.00	0.97	0.05	0.05
	$\gamma_2(1.2)$	0.01	0.94	0.09	0.10	0.00	0.96	0.06	0.06
	$\zeta_1(1.1)$	-0.01	0.95	0.11	0.11	-0.02	0.94	0.07	0.07
	$\zeta_2(0.9)$	-0.02	0.93	0.10	0.11	-0.02	0.95	0.06	0.06
75	$\gamma_1(1.0)$	0.01	0.97	0.08	0.08	0.00	0.97	0.05	0.05
	$\gamma_2(1.2)$	0.01	0.95	0.09	0.10	0.00	0.96	0.06	0.06
	$\zeta_1(1.1)$	-0.01	0.95	0.11	0.11	-0.02	0.94	0.07	0.07
	$\zeta_2(0.9)$	-0.02	0.94	0.10	0.11	-0.02	0.95	0.06	0.06
90	$\gamma_1(1.0)$	0.01	0.97	0.08	0.08	0.00	0.98	0.05	0.05
	$\gamma_2(1.2)$	0.01	0.94	0.09	0.10	0.00	0.96	0.06	0.06
	$\zeta_1(1.1)$	-0.01	0.96	0.11	0.11	-0.02	0.94	0.07	0.07
	$\zeta_2(0.9)$	-0.02	0.93	0.10	0.11	-0.02	0.95	0.06	0.06
100	$\gamma_1(1.0)$	0.01	0.97	0.08	0.08	0.00	0.98	0.05	0.05
	$\gamma_2(1.2)$	0.01	0.94	0.09	0.10	0.00	0.95	0.06	0.06
	$\zeta_1(1.1)$	-0.01	0.96	0.11	0.11	-0.02	0.94	0.07	0.07
	$\zeta_2(0.9)$	-0.02	0.93	0.10	0.11	-0.02	0.95	0.06	0.06

Table 2.1: Parameter Results for Survival Analysis with Time Varying Covariates Under Various Censoring & Truncation Schemes

We see in Table 2.1 that for all the different levels of censoring, p_{cens}, the results show there is minimal relative bias (RBias), the empirical 95% coverage probabilities (ECPs) are around 95%, and the mean estimated standard errors (MESEs) and empirical standard errors (ESEs) have values that are close to each other. The mean estimated standard errors and empirical standard errors decrease as the sample size increases. The formulas for RBias, ECP, MESE, and ESE along with the derivation for the standard errors is given in F of the Appendix.

σ_b^2 Simulation Results

	$n = 200$		$n = 500$	
p_{cens}	Average	ESE	Average	ESE
0	0.04	0.02	0.04	0.02
50	0.06	0.02	0.05	0.01
75	0.06	0.02	0.05	0.02
90	0.06	0.02	0.05	0.01
100	0.06	0.02	0.05	0.02

Table 2.2: σ_b^2 Results for Survival Analysis with Time Varying Covariates Under Various Censoring & Truncation Schemes

We can see in Table 2.2 that the value of σ_b^2 increases with censoring, meaning less smoothing is needed. However, in both sample sizes and all 5 levels of censoring the values of σ_b^2 are very close to 0. This means a large level of smoothing was required for the $\boldsymbol{b}$'s.

Chapter 3

Survival Analysis with Time Varying Covariates Under Various Censoring & Truncation Schemes With Clustering

The likelihood, assuming that the longitudinal trajectories are completely known, is:

$$
\begin{aligned}
&L\left(\Theta; \boldsymbol{T}^l, \boldsymbol{T}^r, \boldsymbol{L}, \boldsymbol{R}, \boldsymbol{\delta}_O, \boldsymbol{\delta}_R, \boldsymbol{\delta}_L, \boldsymbol{\delta}_I, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{V}\right) = \\
&\prod_{i=1}^{n}\prod_{c=1}^{C}\left\{\omega_c \left(\frac{\lambda_c\left(T_i^r|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right) e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i)}}\right)^{\delta_{O_i}} \left(\frac{e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i)}}\right)^{\delta_{R_i}}\right. \\
&\qquad \left. \times \left(\frac{1 - e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i)}}\right)^{\delta_{L_i}} \left(\frac{e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)} - e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i)}}\right)^{\delta_{I_i}}\right\}^{v_{ic}},
\end{aligned}
$$

where $\Theta = (\alpha_0, \alpha_1, \boldsymbol{b}, \omega_c, \boldsymbol{\gamma}_c, \boldsymbol{\zeta}_c)$, $c = 1, \ldots, C$, and $j = 1, \ldots, p$. Now T_i^l and T_i^r are the left and right time to event values, respectively, L_i and R_i are the left and right truncation times, respectively,

$$\delta_{O_i} = \begin{cases} 1 \text{ if } T_i \text{ is observed.} \\ \\ 0 \text{ otherwise.} \end{cases}, \qquad \delta_{R_i} = \begin{cases} 1 \text{ if } T_i \text{ is right censored.} \\ \\ 0 \text{ otherwise.} \end{cases},$$

$$\delta_{L_i} = \begin{cases} 1 \text{ if } T_i \text{ is left censored.} \\ \\ 0 \text{ otherwise.} \end{cases}, \qquad \delta_{I_i} = \begin{cases} 1 \text{ if } T_i \text{ is interval censored.} \\ \\ 0 \text{ otherwise.} \end{cases},$$

and $\Lambda_c \left(t|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right) = \int_0^t \lambda_c \left(t|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)$ is the cumulative hazard function. Additionally, $\lambda_c \left(t|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right) = \lambda_0 \left(t\right) e^{\boldsymbol{X}_{ic}^T \boldsymbol{\gamma}_c + \boldsymbol{Z}_i^T \boldsymbol{\zeta}_c}$, where $\lambda_0 \left(t\right)$ is represented by a linear spline model. The model is:

$$\lambda_0 \left(t\right) = e^{\alpha_0 + \alpha_1 t + \sum_{k=1}^{K} b_k (t - \kappa_k)_+},$$

where K is the number of basis knots and length of vector $\boldsymbol{b}$, κ_k is the k^{th} basis knot, $\boldsymbol{b}$ is the vector of b_k's, where b_k is the coefficient for the k^{th} basis knot, α_0 is the intercept term, and α_1 is the coefficient for $\kappa_0 = 0$. Lastly, $\boldsymbol{\gamma}_c$ is the coefficients for the time-dependent covariates, $\boldsymbol{X}$, per cluster c, $\boldsymbol{\zeta}_c$ is the coefficients for the baseline covariates, $\boldsymbol{Z}$, per cluster c, and $\boldsymbol{X}_{ic}$ is the p-dimensional vector of longitudinal trajectory functions for individual i for cluster c, $c = 1, \ldots, C$. Note: $\left(t - \kappa_k\right)_+ \equiv \max\left(0, \left(t - \kappa_k\right)\right)$ and when the time value is observed, $T_i^l = T_i^r = T_i$ and thus T_i^r is used.

Since we now have clustering in this model, we have that $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_C)$ is the vector of cluster proportions, where C is the number of clusters, ω_c is the cluster proportion for cluster c, $0 \leq \omega_c \leq 1$, and $\sum_{c=1}^{C} \omega_c = 1$. Now $\boldsymbol{V}$ is the $n \times C$ matrix of indicator values where:

$$v_{ic} = \begin{cases} 1 & \text{if } i \in c. \\ \\ 0 & \text{if } i \notin c. \end{cases}$$

The loglikelihood is:

$$\begin{aligned} \ell\left(\Theta; \boldsymbol{T}^l, \boldsymbol{T}^r, \boldsymbol{L}, \boldsymbol{R}, \boldsymbol{\delta}_O, \boldsymbol{\delta}_R, \boldsymbol{\delta}_L, \boldsymbol{\delta}_I, \boldsymbol{Z}, \boldsymbol{V}\right) = \sum_{i=1}^{n}\sum_{c=1}^{C} \Bigg[v_{ic}\log\left(\omega_c\right) \\ + v_{ic}\log\Bigg\{ \left(\frac{\lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{O_i}} \left(\frac{e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{R_i}} \\ \times \left(\frac{1 - e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{L_i}} \left(\frac{e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)} - e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{I_i}} \Bigg\}\Bigg]. \end{aligned}$$

3.1 EM Layout

In order to estimate the mixture survival models, we decided on using a version of the expectation-maximization (EM) algorithm by Dempster et al. (1977) as discussed in Naik et al. (2007). We denote $\ell\left(\Theta; \boldsymbol{T}^l, \boldsymbol{T}^r, \boldsymbol{L}, \boldsymbol{R}, \boldsymbol{\delta}_O, \boldsymbol{\delta}_R, \boldsymbol{\delta}_L, \boldsymbol{\delta}_I, \boldsymbol{Z}, \boldsymbol{V}\right)$ as ℓ and we let

$$\begin{aligned} S_c\left(\Theta\right) = \prod_{i=1}^{n} S_{ic}\left(\Theta\right) = \prod_{i=1}^{n} S_c\left(\Theta; T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right) \\ = \prod_{i=1}^{n} \left(\frac{\lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)e^{-\Lambda_c(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{O_i}} \left(\frac{e^{-\Lambda_c(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{R_i}} \\ \times \left(\frac{1 - e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{L_i}} \left(\frac{e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)} - e^{-\Lambda_c(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)} - e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{I_i}} \end{aligned}$$

for ease of reading. We let $\Theta^{(m)} = \left(\alpha_0^{(m)}, \alpha_1^{(m)}, \boldsymbol{b}^{(m)}, \boldsymbol{\gamma}_c^{(m)}, \boldsymbol{\zeta}_c^{(m)}, \omega_c^{(m)}\right)$ where $c = 1, \ldots, C$ be the estimates of the parameters of interest at the m^{th} iteration. We let $Q\left(\Theta; \Theta^{(m)}\right) = E\left(\ell|\Theta^{(m)}, \boldsymbol{T}^l, \boldsymbol{T}^r, \boldsymbol{L}, \boldsymbol{R}, \boldsymbol{\delta}_O, \boldsymbol{\delta}_R, \boldsymbol{\delta}_L, \boldsymbol{\delta}_I, \boldsymbol{Z}\right)$. Thus for the E-step, we obtain $Q\left(\Theta; \Theta^{(m)}\right)$ by substituting v_{ic} with $\tau_{ic} = E\left[v_{ic}|T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{X}_i, \boldsymbol{Z}_i\right]$. We find that τ_{ic} for each iteration m is given by:

$$\widehat{\tau}_{ic}^{(m)} = \frac{\omega_c^{(m)} S_{ic}\left(\Theta^{(m)}\right)}{\sum_{c'=1}^{C} \omega_{c'}^{(m)} S_{ic'}\left(\Theta^{(m)}\right)}.$$

In the M-step, unlike Naik et al. (2007) where each parameter had a closed-form solution, the only closed-form solution we have is:

$$\widehat{\omega}_c^{(m)} = \frac{\sum_{i=1}^{n} \tau_{ic}^{(m)}}{\sum_{i=1}^{n} \sum_{c'=1}^{C} \tau_{ic'}^{(m)}} = \frac{\sum_{i=1}^{n} \tau_{ic}^{(m)}}{n}.$$

The derivations of these terms is given in Appendix G. To estimate the other parameters of Θ, we again use Newton optimization as we did in Chapter 2 to estimate $\boldsymbol{Q}$ and $\mathbb{Q}$. Now though, we estimate $\boldsymbol{Q}$ and $\mathbb{Q}$ for each of the C clusters. The EM algorithm stops once the value of $\log \frac{\sum_{c=1}^{C} \omega_c S_c\left(\Theta^{(m+1)}\right)}{\sum_{c'=1}^{C} \omega_{c'} S_{c'}\left(\Theta^{(m)}\right)}$ decreases below a pre-specified threshold. In order to initialize the EM algorithm, $\tau_{ic}^{(0)}$ for each individual is initialized by partitioning $(\boldsymbol{X}, \boldsymbol{Z})$ into C clusters using K-means clustering method, where $\boldsymbol{X}$ is the matrix of longitudinal covariate values for all individuals at their survival times and $\boldsymbol{Z}$ is the baseline covariates for all individuals.

3.2 1^{st} and 2^{nd} Derivatives of Loglikelihood

We again used the penalized loglikelihood. Therefore, the penalized loglikelihood for a single cluster, c, is now written as $\ell_{p,c}\left(\Theta; \sigma_b^2\right) = \ell_{0,c} - \frac{\boldsymbol{b}^T \boldsymbol{b}}{2\sigma_b^2}$, where $\ell_{0,c} = \sum_{i=1}^{n} \Big\{ \tau_{ic} \log(\omega_c) + \tau_{ic} \Big[\eta_0(T_i^r)\, \delta_{O_i} + \left(\boldsymbol{X}_{ic}^T(T_i^r)\, \boldsymbol{\gamma}_c + \boldsymbol{Z}_i^T \boldsymbol{\zeta}_c\right) \delta_{O_i} - \Lambda_c(T_i^r | \boldsymbol{X}_{ic}, \boldsymbol{Z}_i)(\delta_{O_i} + \delta_{R_i}) + \log\left(1 - e^{-\Lambda_c\left(T_i^l | \boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}\right) \delta_{L_i} - \Lambda_c\left(T_i^l | \boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right) \delta_{I_i} + \log\left(1 - e^{\Lambda_c\left(T_i^l | \boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right) - \Lambda_c\left(T_i^r | \boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}\right) \delta_{I_i} + \Lambda_c(L_i | \boldsymbol{X}_{ic}, \boldsymbol{Z}_i) - \log\left(1 - e^{\Lambda_c(L_i | \boldsymbol{X}_{ic}, \boldsymbol{Z}_i) - \Lambda_c(R_i | \boldsymbol{X}_{ic}, \boldsymbol{Z}_i)}\right) \Big] \Big\}$. The first derivative with respect to α_0 is now multiplied by τ_{ic} for clustering as given below. The rest, as mentioned in Chapter 2 when there was no clustering, are found in Section C.2 of the Appendix. The only update to them would be multiplying by τ_{ic} and using the parameters estimates for cluster c as given here:

$$\begin{aligned}
\frac{\partial}{\partial \alpha_0}\ell_{p,c}\left(\Theta;\sigma_b^2\right) &= \frac{\partial}{\partial \alpha_0}\ell_{0,c}\left(\Theta;\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\tau_{ic}\Bigg[\delta_{O_i} - \frac{\partial}{\partial \alpha_0}\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) + \frac{\frac{\partial}{\partial \alpha_0}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}{e^{\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad - \frac{\partial}{\partial \alpha_0}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)\delta_{I_i} - \frac{\left(\frac{\partial}{\partial \alpha_0}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right) - \frac{\partial}{\partial \alpha_0}\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)\right)}{e^{\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}-1}\delta_{I_i} \\
&\quad + \frac{\partial}{\partial \alpha_0}\Lambda_c\left(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial}{\partial \alpha_0}\Lambda_c\left(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right) - \frac{\partial}{\partial \alpha_0}\Lambda_c\left(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)\right)}{e^{\Lambda_c\left(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)-\Lambda_c\left(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}-1}\Bigg].
\end{aligned}$$

One of the second derivatives with respect to α_0 now multiplied by τ_{ic} for clustering is given below. The rest, as mentioned in Chapter 2 when there was no clustering, are found in Section C.4 of the Appendix. The only update to them would be multiplying by τ_{ic} and using the parameters estimates for cluster c as given here:

$$\begin{aligned}
\frac{\partial^2}{\partial \alpha_0^2}\ell_{p,c}\left(\Theta;\sigma_b^2\right) &= \frac{\partial^2}{\partial \alpha_0^2}\ell_{0,c}\left(\Theta;\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\tau_{ic}\Bigg[-\frac{\partial^2}{\partial \alpha_0^2}\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) \\
&\quad + \frac{\frac{\partial^2}{\partial \alpha_0^2}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}{e^{\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}-1}\delta_{L_i} - \frac{e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad - \frac{\partial^2}{\partial \alpha_0^2}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)\delta_{I_i} - \frac{\left(\frac{\partial^2}{\partial \alpha_0^2}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial \alpha_0^2}\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)\right)}{e^{\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}-1}\delta_{I_i} \\
&\quad - \frac{e^{\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right) - \frac{\partial}{\partial \alpha_0}\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}\right)^2}\delta_{I_i} \\
&\quad + \frac{\partial^2}{\partial \alpha_0^2}\Lambda_c\left(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial \alpha_0^2}\Lambda_c\left(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial \alpha_0^2}\Lambda_c\left(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)\right)}{e^{\Lambda_c\left(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)-\Lambda_c\left(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}-1} \\
&\quad + \frac{e^{\Lambda_c\left(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)-\Lambda_c\left(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda_c\left(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right) - \frac{\partial}{\partial \alpha_0}\Lambda_c\left(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{\Lambda_c\left(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)-\Lambda_c\left(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}\right)^2}\Bigg].
\end{aligned}$$

3.3 Unsupervised Clustering Methodology

Since we wanted to develop an unsupervised clustering methodology for our survival model that would both determine the correct number of clusters along with correct parameter estimates, we had to decide a threshold to stop conducting more clusters. We decided to use the model selection criterions of AIC Akaike (1973), $AIC = -2\log\left(\sum_{c=1}^{C}\omega_c S_c\left(\Theta\right)\right) + 2p$, and BIC Schwarz (1978), $BIC = -2\log\left(\sum_{c=1}^{C}\omega_c S_c\left(\Theta\right)\right) + p\log\left(n\right)$. However, since these are the expressions for one cluster, we used a version of the modified form mentioned in Naik et al. (2007) to account for having C clusters and p variables per cluster by replacing p with $d = (C-1) + Cp$. Since we know that BIC's penalty is harsher than AIC's penalty, we thought that once the two models with additional clusters have AIC values that are greater the current model, this would be the stopping criterion.

The methodology given current cluster number C is:

- If $AIC_C < AIC_{C+1}$ and $AIC_C < AIC_{C+2}$, then C is the optimal number of clusters.
- If not, keep increasing C until a pre-specified maximum number of clusters is reached.

We chose to check that $AIC_C < AIC_{C+1}$ and $AIC_C < AIC_{C+2}$ to decrease the chance of stopping earlier by having the AIC for the model with $C+1$ clusters have a slighly higher AIC than C clusters due to chance when really more clusters is the better fitting model.

Since we have a baseline hazard function, $\lambda_0\left(t\right)$, that is assumed the same for all groups since it is for the population, then each cluster shares the coefficient estimates of the baseline spline model. Since we still have to estimate the smoothing parameter σ_b^2 along with the fact that we perform the clustering model chronologically, we decided to estimate

σ_b^2 when $n = 1$ and use that estimate for all the other clusters. From simulation results, the method seems to work well.

3.4 Simulation Setting

Note: Since the longitudinal part is assumed known, we only focused on clustering in the survival model and assumed no clustering in the longitudinal part.

3.4.1 Data Simulation Structure

The following structure is used for simulating data:

n # of individuals.

q # of baseline covariates.

p # of longitudinal trajectory functions.

$f(t)$ The baseline hazard function.

K The number of knots, in our case we used the min $\left(\left\lfloor \frac{n}{4} \right\rfloor, 30\right)$ as in Cai and Betensky (2003).

κ_k's The baseline knots where $k = 1, \ldots, K$. Chosen based on data and usually spaced on unique quantiles of the survival times T_i^l's, T_i^r's, and $\frac{T_i^l + T_i^r}{2}$'s.

α_0 The intercept of the estimated baseline hazard function.

α_1	The κ_0 coefficient of the estimated baseline hazard function. Note: $\kappa_0 = 0$.
b_k's	Distributed $N_K(\mathbf{0}, \sigma_b^2 \boldsymbol{I}_K)$ where $k = 1, \ldots, K$. They are the coefficients of the $(t - \kappa_k)_+$'s, $k = 1, \ldots, K$. Used in estimating the baseline hazard function.
C	The number of clusters.
$\boldsymbol{\zeta}$	The $q \times C$ matrix of coefficients for the baseline covariates. Chosen by the user.
$\boldsymbol{\gamma}$	The $p \times C$ matrix of coefficients for the longitudinal trajectory functions. Chosen by the user.
$\boldsymbol{\omega}$	The vector of length C of cluster weights.
$\boldsymbol{Z}$	The baseline covariate matrix that is $n \times q$.
$\boldsymbol{X}_i(t)$	The set of p longitudinal trajectory functions for the i^{th} individual. Reminder that $X_{ij}(t) = \psi_j(t) + \sum_{k'=1}^{J} \xi_{ijk'} B_{k'}(t)$ where $j = 1, \ldots, p$. Additionally, the j^{th} mean trajectory function is given by $\psi_j(t) = \sum_{k'=1}^{\overline{J}} \beta_{jk'} \overline{B}_{k'}(t)$.
$\boldsymbol{\beta}_j$'s	The mean longitudinal basis knots' coefficients.

$\boldsymbol{\xi}_{ij}$'s Distributed $N_J\left(\boldsymbol{\mu}_{\boldsymbol{\xi}_j}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_j}\right)$ where J is the number of individual longitudinal basis knots and $\boldsymbol{\mu}_{\boldsymbol{\xi}_j}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_j}$ are mean and covariance-variance matrix for the $\boldsymbol{\xi}_{ij}$'s. For the simulation, since the trajectory functions are all known, we decided to let the individual longitudinal coefficients all have the same mean and covariance-variance matrix.

The covariance-variance matrix we decided on is:

$$\boldsymbol{\Sigma}_{\boldsymbol{\xi}_j} = \sigma^2_{\xi_j} \begin{pmatrix} 1 & \rho_{\xi_j} & \rho_{\xi_j} & \cdots & \rho_{\xi_j} \\ \rho_{\xi_j} & 1 & \rho_{\xi_j} & \cdots & \rho_{\xi_j} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{\xi_j} & \rho_{\xi_j} & \rho_{\xi_j} & \cdots & 1 \end{pmatrix}$$

for each j, where $\sigma^2_{\xi_j}$ is the variance and ρ_{ξ_j} is the correlation.

3.4.2 Simulating Actual Times

To simulate the survival times, T_i, $i = 1, \ldots, n$, we used the fact that $S(t|\boldsymbol{X}_i, \boldsymbol{Z}_i) = e^{-\Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i)}$, where $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) = \int_0^t \lambda\left(t\right)$. This method is similar to the one used by Brown, Ibrahim, and DeGruttola (2005) to simulate their joint survival-longitudinal data.

The method is:

1. Simulate a random survival probability, $s \sim U(0, 1)$.

2. Use R's *Uniroot* function to solve $\log(s) + \Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i)$ for where the function is 0 along with using R's *Integrate* function to integrate $\Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i)$ from 0 to the value of t such that the above equation is true.

To account for multiple clusters, we applied the following method to each cluster and then took the proportional amount of each cluster that we were using to achieve the desired dataset for each replicate. The method looks like:

1. Generate n samples from each cluster c, $c = 1, \ldots, C$.

2. Using R's $rmultinom\,(1, n, prob = \boldsymbol{\omega})$ function to select the appropriate proportion per cluster.

3.4.3 Simulation Study Setup

We conducted a simulation study where we performed 200 replicates for both 300 and 900 individuals. We simulated each replicate in R with a different seed from $i = 0, \ldots, 199$ and chose the following values for the inputs and parameters:

n 300 or 900.

q 2.

p 2.

$f(t)$ $\frac{(t-20)^4}{10^8} + 0.05$.

K 30, since $\left\lfloor \frac{n}{4} \right\rfloor > 30$.

κ_k's Changed with each of the 200 replicates for each sample size n. We used the unique quantiles of the survival times T_i^l's, T_i^r's, and $\frac{T_i^l+T_i^r}{2}$'s.

α_0 Estimated in each of the replicates.

α_1 Estimated in each of the replicates.

b_k's Estimated in each of the replicates and 30 used since $K = 30$.

C 2.

$\boldsymbol{\zeta}$ $\begin{bmatrix} 1.1 & -1.1 \\ 0.9 & -0.9 \end{bmatrix}$.

$\boldsymbol{\omega}$ $(0.6, 0.4)$.

$\boldsymbol{\gamma}$ $\begin{bmatrix} 1.0 & -1.0 \\ 1.2 & -1.2 \end{bmatrix}$.

$\boldsymbol{Z}$ The matrix that is $n \times 2$ of the time independent covariates.

Each $Z_{ij} \sim Beta(2,2) * 4 - 2$ for $i = 1, \ldots, n$ and $j = 1, 2$.

Unlike a normal distribution, this beta only allows for values between -2 and 2 and is still mound-shaped symmetric.

$\boldsymbol{X}_i(t)$ Is $\psi_j(t) + \sum_{k'=1}^{J} \xi_{ijk'} B_{k'}(t)$ where $\overline{J} = J = 8$ since we have 6 distinct knots at 0, 20, 40, 60, 80, and 100 and are using a cubic B-spline.

Additionally, the j^{th} mean trajectory function is given by

$\psi_j(t) = \sum_{k'=1}^{J} \beta_{jk'} \overline{B}_{k'}(t)$.

$$\boldsymbol{\beta} \quad \begin{bmatrix} -1.41 & -0.80 & -0.25 & -0.49 & 0.30 & 0.49 & 0.76 & 1.42 \\ \\ 0.42 & 1.17 & 0.11 & -0.29 & -0.38 & -0.32 & -0.69 & -1.21 \end{bmatrix}.$$

$\boldsymbol{\xi}_{ij}$'s Distributed $N_J\left(\boldsymbol{\mu}_{\boldsymbol{\xi}_j} = \mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_j} = 3\boldsymbol{I}\right)$.

This gave us survival times between $0 - 100$ years old with baseline hazard decreasing until 20 and then increasing from then on. Given in Figures 3.1-3.4 are the survival times for $n = 300$ and $n = 900$ along with color coding for clusters for one replicate.

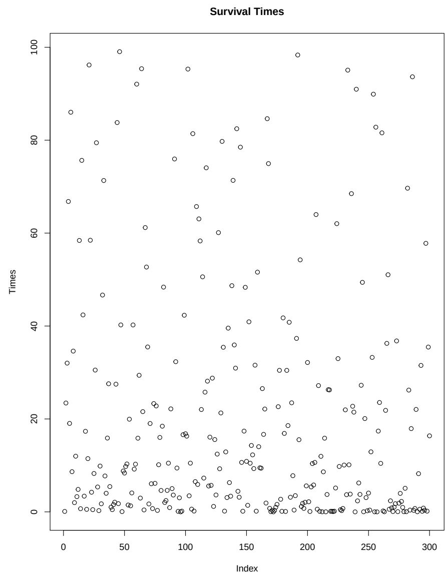

Figure 3.1: Plot of 300 survival times

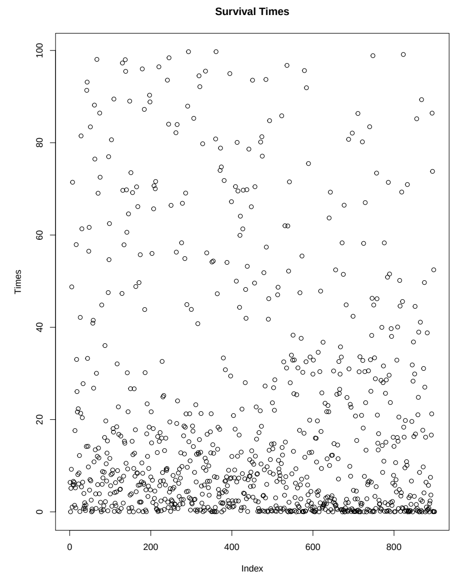

Figure 3.2: Plot of 900 survival times

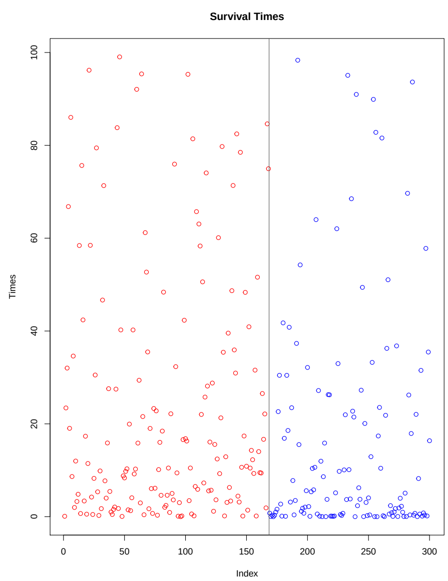

Figure 3.3: Plot of 300 survival times with cluster color coding

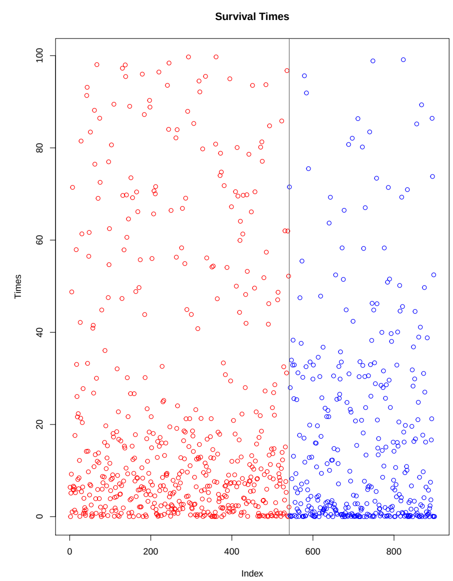

Figure 3.4: Plot of 900 survival times with cluster color coding

3.5 Results for 2 Clusters

Cluster Assignment for $n = 300$

		Cluster Number					
Criterion	p_{cens}	1	**2**	3	4	5	6
AIC	0	1 (0.52%)	146 (76.44%)	36 (18.85%)	6 (3.14%)	1 (0.52%)	1 (0.52%)
	50	0	159 (83.68%)	27 (14.21%)	4 (2.11%)	0	0
	75	0	154 (81.05%)	30 (15.79%)	6 (3.16%)	0	0
	90	0	158 (83.16%)	24 (12.63%)	8 (4.21%)	0	0
	100	0	156 (82.11%)	25 (13.16%)	9 (4.74%)	0	0
BIC	0	1 (0.52%)	178 (93.19%)	10 (5.24%)	2 (1.05%)	0	0
	50	0	189 (99.47%)	1 (0.53%)	0	0	0
	75	0	189 (99.47%)	1 (0.53%)	0	0	0
	90	0	188 (98.95%)	2 (1.05%)	0	0	0
	100	0	187 (98.42%)	3 (1.58%)	0	0	0

Table 3.1: Table of which cluster was chosen best with $n = 300$ by criterion. 2 clusters was the correct cluster number.

Cluster Assignment for $n = 900$

Criterion	p_{cens}	Cluster Number 1	**2**	3	4	5	6
AIC	0	0	157 (78.89%)	34 (17.09%)	5 (2.51%)	3 (1.51%)	0
	50	0	151 (76.26%)	40 (20.20%)	7 (3.54%)	0	0
	75	0	148 (74.75%)	44 (22.22%)	6 (3.03%)	0	0
	90	0	153 (77.27%)	40 (20.20%)	5 (2.53%)	0	0
	100	0	147 (74.24%)	46 (23.23%)	5 (2.53%)	0	0
BIC	0	0	190 (95.48%)	9 (4.52%)	0	0	0
	50	0	196 (98.99%)	2 (1.01%)	0	0	0
	75	0	197 (99.49%)	1 (0.51%)	0	0	0
	90	0	198 (100%)	0	0	0	0
	100	0	198 (100%)	0	0	0	0

Table 3.2: Table of which cluster was chosen best with $n = 900$ by criterion. 2 clusters was the correct cluster number.

Specification Rate

	$n = 300$		$n = 900$	
p_{cens}	Ave	ESE	Ave	ESE
0	0.91	0.05	0.92	0.01
50	0.92	0.02	0.92	0.01
75	0.92	0.01	0.92	0.01
90	0.92	0.02	0.92	0.01
100	0.91	0.04	0.92	0.01

Table 3.3: Table of average specification rate (AVE) along the empirical standard error (ESE) for $n = 300$ and $n = 900$.

Cluster Proportions

	$n = 300$				$n = 900$			
$\boldsymbol{\omega}$	1 (.6)		2 (.4)		1 (.6)		2 (.4)	
p_{cens}	RBias	ESE	RBias	ESE	RBias	ESE	RBias	ESE
0	-0.01	0.03	0.01	0.03	0.00	0.02	0.00	0.02
50	-0.01	0.03	0.01	0.03	0.00	0.02	0.00	0.02
75	-0.01	0.03	0.01	0.03	0.00	0.02	0.00	0.02
90	-0.01	0.03	0.01	0.03	0.00	0.02	0.00	0.02
100	-0.01	0.04	0.01	0.04	0.00	0.02	0.00	0.02

Table 3.4: Table of relative bias (RBias) along the empirical standard error (ESE) for $\boldsymbol{\omega}$ for $n = 300$ and $n = 900$.

Simulation Results - 100% Observed

		$n = 300$				$n = 900$			
Cluster		RBias	ECP	MESE	ESE	RBias	ECP	MESE	ESE
1	γ_1 (1.0)	-0.04	0.86	0.07	0.13	-0.03	0.79	0.04	0.06
	γ_2 (1.2)	-0.02	0.81	0.07	0.15	-0.02	0.73	0.04	0.06
	ζ_1 (1.1)	-0.03	0.85	0.10	0.18	-0.04	0.80	0.05	0.07
	ζ_2 (0.9)	-0.01	0.90	0.09	0.15	-0.04	0.84	0.05	0.06
2	γ_1 (-1.0)	-0.01	0.85	0.07	0.12	-0.01	0.77	0.04	0.05
	γ_2 (-1.2)	-0.02	0.87	0.08	0.18	-0.02	0.85	0.05	0.06
	ζ_1 (-1.1)	-0.02	0.87	0.12	0.19	-0.03	0.89	0.07	0.07
	ζ_2 (-0.9)	-0.03	0.93	0.11	0.17	-0.03	0.90	0.06	0.07

Table 3.5: Table of relative bias (RBias), emperical coverage probabilities (ECP), mean estimated standard errors (MESE), and the empirical standard error (ESE) for the parameter estimates per cluster with 100% observed data for $n = 300$ and $n = 900$.

Simulation Results - 50% Censored

Cluster		$n=300$				$n=900$			
		RBias	ECP	MESE	ESE	RBias	ECP	MESE	ESE
1	γ_1 (1.0)	0.00	0.88	0.09	0.11	-0.01	0.88	0.05	0.06
	γ_2 (1.2)	0.02	0.83	0.08	0.11	0.00	0.79	0.04	0.07
	ζ_1 (1.1)	0.00	0.86	0.11	0.14	-0.02	0.83	0.06	0.08
	ζ_2 (0.9)	0.01	0.90	0.10	0.13	-0.02	0.83	0.06	0.08
2	γ_1 (-1.0)	0.03	0.84	0.09	0.11	0.00	0.83	0.05	0.07
	γ_2 (-1.2)	0.02	0.86	0.11	0.15	-0.01	0.88	0.06	0.08
	ζ_1 (-1.1)	0.01	0.91	0.14	0.16	-0.03	0.90	0.08	0.09
	ζ_2 (-0.9)	-0.01	0.90	0.13	0.16	-0.02	0.90	0.07	0.09

Table 3.6: Table of relative bias (RBias), emperical coverage probabilities (ECP), mean estimated standard errors (MESE), and the empirical standard error (ESE) for the parameter estimates per cluster with 50% censored data for $n = 300$ and $n = 900$.

Simulation Results - 75% Censored

Cluster		$n=300$				$n=900$			
		RBias	ECP	MESE	ESE	RBias	ECP	MESE	ESE
1	γ_1 (1.0)	0.00	0.88	0.09	0.10	0.00	0.87	0.05	0.06
	γ_2 (1.2)	0.01	0.83	0.08	0.11	0.00	0.79	0.04	0.07
	ζ_1 (1.1)	-0.01	0.87	0.11	0.14	-0.02	0.82	0.06	0.08
	ζ_2 (0.9)	0.01	0.92	0.10	0.12	-0.02	0.85	0.06	0.08
2	γ_1 (-1.0)	0.03	0.86	0.09	0.11	0.00	0.81	0.05	0.07
	γ_2 (-1.2)	0.02	0.86	0.11	0.15	-0.01	0.88	0.06	0.08
	ζ_1 (-1.1)	0.01	0.92	0.14	0.17	-0.03	0.89	0.08	0.09
	ζ_2 (-0.9)	-0.01	0.91	0.13	0.16	-0.02	0.91	0.07	0.09

Table 3.7: Table of relative bias (RBias), emperical coverage probabilities (ECP), mean estimated standard errors (MESE), and the empirical standard error (ESE) for the parameter estimates per cluster with 75% censored data for $n = 300$ and $n = 900$.

Simulation Results - 90% Censored

Cluster		$n = 300$ RBias	ECP	MESE	ESE	$n = 900$ RBias	ECP	MESE	ESE
1	γ_1 (1.0)	0.00	0.87	0.09	0.11	0.00	0.87	0.05	0.06
	γ_2 (1.2)	0.02	0.82	0.08	0.11	0.00	0.79	0.04	0.07
	ζ_1 (1.1)	0.00	0.86	0.11	0.14	-0.02	0.83	0.06	0.08
	ζ_2 (0.9)	0.01	0.92	0.10	0.12	-0.02	0.84	0.06	0.08
2	γ_1 (-1.0)	0.03	0.86	0.09	0.11	0.00	0.82	0.05	0.07
	γ_2 (-1.2)	0.02	0.86	0.11	0.15	-0.01	0.87	0.06	0.08
	ζ_1 (-1.1)	0.01	0.91	0.14	0.16	-0.03	0.90	0.08	0.09
	ζ_2 (-0.9)	-0.01	0.90	0.13	0.16	-0.02	0.91	0.07	0.09

Table 3.8: Table of relative bias (RBias), emperical coverage probabilities (ECP), mean estimated standard errors (MESE), and the empirical standard error (ESE) for the parameter estimates per cluster with 90% censored data for $n = 300$ and $n = 900$.

Simulation Results - 100% Censored

Cluster		$n = 300$ RBias	ECP	MESE	ESE	$n = 900$ RBias	ECP	MESE	ESE
1	γ_1 (1.0)	0.00	0.87	0.09	0.11	0.00	0.87	0.05	0.06
	γ_2 (1.2)	0.01	0.79	0.08	0.14	0.00	0.81	0.04	0.06
	ζ_1 (1.1)	-0.01	0.87	0.11	0.15	-0.02	0.83	0.06	0.08
	ζ_2 (0.9)	0.01	0.91	0.10	0.13	-0.02	0.83	0.06	0.08
2	γ_1 (-1.0)	0.02	0.85	0.09	0.13	0.00	0.82	0.05	0.07
	γ_2 (-1.2)	0.02	0.86	0.11	0.19	-0.01	0.88	0.06	0.08
	ζ_1 (-1.1)	0.01	0.89	0.14	0.19	-0.03	0.90	0.08	0.09
	ζ_2 (-0.9)	-0.02	0.89	0.13	0.17	-0.02	0.90	0.07	0.09

Table 3.9: Table of relative bias (RBias), emperical coverage probabilities (ECP), mean estimated standard errors (MESE), and the empirical standard error (ESE) for the parameter estimates per cluster with 100% censored data for $n = 300$ and $n = 900$.

We ran 200 replicates and the results in Tables 3.1-3.9 are for the replicates that converged, which was about 190 when $n = 300$ and about 198 when $n = 900$. We can see from Table 3.1 and Table 3.2 that BIC outperforms AIC at identifying the correct number of clusters and correctly identifies the right cluster greater than 93% of the time. Additionally, the specification rate, which is the number of individuals correctly specified to their cluster divided by the total number of individuals, is greater than 90%. We see that the cluster proportions, $\boldsymbol{\omega}$, have very little relative bias (RBias) and small empirical standard errors (MESEs) in both sample sizes.

In Tables 3.5 - 3.9 for all the different levels of censoring, p_{cens}, the results show there is minimal relative bias (RBias), the empirical 95% coverage probabilities (ECPs) are lower than 95%, and the mean estimated standard errors (MESEs) are smaller than the empirical standard errors (ESEs). The mean estimated standard errors and empirical standard errors decrease as the sample size increases. We believe that the MESEs are underestimated due to only having access to the observed data, as stated in Louis (1982), and therefore causing the ECPs to be lower than 95%. The formulas for RBias, ECP, MESE, ESE, and specification rate along with the derivation for the standard errors are given in Appendix F.

Chapter 4

R Package

4.1 Overview

All of the analysis performed in this dissertation was done using the package I am currently writing using RStudio by RStudio Team (2020). I have developed the package using Rcpp by Eddelbuettel and François (2011), Eddelbuettel (2013), and Eddelbuettel and Balamuta (2017), along with Rcpp Armadillo Eddelbuettel and Sanderson (2014). This has allowed me to use elements from R by R Core Team (2020) while gaining the speed of c++. Due to the nature of this model and since I wrote the package using Rcpp, I had to write most of the functions needed. For example, I wrote a B-spline function, Simpson's Rule function, and Newton optimization function. For functions I wrote that were similar to functions existing in other well-know packages, I tested their results against R packages and confirmed they were working correctly.

The speed of c++ is needed since there are multiple iterations when performing Newton optimization and multiple iterations in the EM algorithm step, which needs to perform the Newton optimization each time. Therefore, using only R based functions would be too slow due to the shear number of iterations.

4.2 Functions

My package has numerous functions that I had to write; however, there are currently only a few functions that I believe the user would want to use. I want to highlight two of them.

The first one is the the function MSCEM2, which stands for Mixture Survival Criterion Expectation-Maximization 2. The idea for the name comes from Naik et al. (2007), since they called their criterion the Mixture Regression Criterion (MRC), and the fact I used the expectation-maximization (EM) algorithm by Dempster et al. (1977). The following is the function along with each input:

```
MSCEM2(NumericVector assigns, NumericMatrix z, NumericVector Tl,
       NumericVector Tr, NumericVector delta_O, NumericVector delta_R,
       NumericVector delta_L, NumericVector delta_I, NumericVector L,
       NumericVector R, NumericMatrix knots_n_p_q,
       NumericVector knots_p, NumericVector beta, NumericMatrix xi,
       NumericMatrix p, NumericMatrix d, NumericVector knots_q,
       NumericMatrix start_vals, double sigma_2_b = 0, int n_simp = 2,
       int maxit_n = 50, int maxit_s = 100, int maxit_em = 50,
       double rel_error_em = .1, double rel_error_n = .1,
       double rel_error_s = .02, double stepsize = 1, int Case = 2,
       int CaseX = 1).
```

The inputs are described below:

assigns The initial assignments of the individuals to each group.

z The $n \times q$ matrix of baseline covariates.

Tl The vector of length n of left censored or the left interval censored time points per individual, input vector of 0's if not used.

Tr The vector of length n of observed, right censored, or the right interval censored time points per individual, input vector of 0's if not used.

delta_O Vector of length n of which individuals have observed times, 1 is yes and 0 is no.

delta_R Vector of length n of which individuals have right censored times. 1 is yes and 0 is no.

delta_L Vector of length n of which individuals have left censored times. 1 is yes and 0 is no.

delta_I Vector of length n of which individuals have interval censored times. 1 is yes and 0 is no.

L Vector of length n of left truncation times, 0 is no left truncation.

R Vector of length n of right truncation times, 0 is no right truncation.

knots_n_p_q The matrix of how many knots each of the longitudinal mean and individual trajectories should have. The first column is for each mean trajectory and the second column is for each individual trajectory.

knots_p — The vector of all the knots for the mean longitudinal trajectories.

beta — The vector of all the coefficient values for the mean longitudinal trajectories.

xi — The matrix of all the coefficient values for the individual trajectories. There are n rows, where each row is the coefficients for all the longitudinal trajectories.

p — The matrix stating the degree of the B-splines for the longitudinal trajectories. The first column is for each mean trajectory and the second column is for each individual trajectory.

d — The matrix stating which derivative of the p^{th} degree B-splines for the longitudinal trajectories to use. The first column is for each mean trajectory and the second column is for each individual trajectory.

knots_q	The vector of all the knots for the individual longitudinal trajectories.
start_vals	User supplied set of starting values for the baseline hazard as well as for the coefficents of interest. If no good initial estimates, then input a matrix of 0's whose rows equal the number of coefficients in the model and whose columns are the number of clusters sought.
sigma_2_b	User specified value of the smoothing parameter σ_b^2. If 1 cluster, σ_b^2 will be estimated using the method stated in Chapter 2. If more than 1 cluster is specified, then the σ_b^2 given will be used. If 0 is given, then σ_b^2=0.2. σ_b^2=0.2 is used since the value was stable in the simulations and lead to reasonable answers even when σ_b^2 was actually much smaller than that value.
n_simp	The number of segments to use for Simpson's rule. Default is 2.
maxit_n	The maximum number of Newton optimization iterations when finding the coefficient values. Default is 50.

maxit_s	The maximum number of Newton optimization iterations when estimating σ_b^2 in the 1 cluster scenario. Default is 100.
maxit_em	The maximum number of EM iterations. The default is 50.
rel_error_em	The threshold for when the EM algorithm has converged. It is the value such that $\log\left(\Theta^{(m+1)}\right) - \log\left(\Theta^{(m)}\right) <$ rel_error_em. Default is 0.1.
rel_error_n	The threshold for when the Newton optimization algorithm for the coefficents has converged. It is the value that every element in the vector $\boldsymbol{Q}$ is less than if Case = 2. Default is 0.1.
rel_error_s	The threshold for when the Newton optimization algorithm for σ_b^2 has converged. It is the value such that the first-derivative of the marginal loglikelihood is less than it if Case = 2. Default is 0.02.
stepsize	The stepsize of the Newton optimization algorithm for both the coefficents and σ_b^2. Default is 1.
Case	Case = 1 means that the Newton optimization algorithm uses the rel_error value to check relative error value between the previous and

current values of the coefficients or σ_b^2, depending on what is being calculated. Case = 2 means that it checks that the 1^{st} derivative is less than the rel_error values.

CaseX: CaseX = 1 means that the clusters have the same baseline. Case = 2, which means that baselines are allowed to differ between clusters is still experimental.

With these inputs, the function returns the following outputs:

omega new: The final values of $\boldsymbol{\omega}$.

omega old: The previous round values of $\boldsymbol{\omega}$.

tau new: The final values for $\boldsymbol{\tau}$.

tau old: The previous round values for $\boldsymbol{\tau}$.

total old: The previous round of loglikelihood value.

total new: The current round of loglikelihood value.

count: How many EM rounds it took to converge.

like_old: The previous round expected loglikelihood value.

	Had as check to make sure likelihood increased.
like_new	The current round expected loglikelihood value.
	Had as check to make sure likelihood increased.
AICC	The AIC value with the penalty using d
	instead of p to account for clusters. Hence,
	the extra C to correct for clusters.
BIC	The BIC values with the penalty using d
	instead of p to account for clusters.
X_n	The final values of the coefficients for all C clusters.
X_n_tuned_old	The previous round values of
	the coefficients for all C clusters.
X_n_tuned	Same as X_n.
Var	The variance for the coefficient parameters
	for the C clusters. First column has the baseline
	coefficients' variances for all clusters.
sigma_old	The previous EM round's σ_b^2 value.

Only changes in 1 cluster setting.

sigma_new The current EM round's σ_b^2 value.

Only changes in 1 cluster setting.

This function runs relatively quickly. With $n = 300$ and 1 cluster, the code is done in under 10 minutes. With multiple clusters, the function usually finishes in under 30 minutes.

The other function of interest to users is MSCC, which stands for Mixture Survival Criterion Clustering. This is the function that uses MSCEM2 for each number of clusters, C. The function stops once the AIC is greater for both $C + 1$ and $C + 2$ clusters than the current C clusters or the maxclust value is reached, whichever occurs first. The function starts each cluster number C with dividing the data into C initial assignments using K-means clustering. The version we are currently using is from the *stats* package by R Core Team (2020). The function with its inputs is given here:

```
MSCC(NumericMatrix z, NumericVector Tl, NumericVector Tr,
     NumericVector delta_O, NumericVector delta_R,
     NumericVector delta_L, NumericVector delta_I,
     NumericVector L, NumericVector R,
     NumericMatrix knots_n_p_q, NumericVector knots_p,
     NumericVector beta, NumericMatrix xi, NumericMatrix p,
     NumericMatrix d, NumericVector knots_q, int n_simp = 2,
     int maxclust = 6, int maxit_n = 50, int maxit_s = 100,
     int maxit_em = 50, double rel_error_em = .1,
     double rel_error_n = .1, double rel_error_s = .02,
     double stepsize = 1, int Case = 2, int CaseX = 1,
     int CaseK = 1, int maxrestarts_km=10, int maxit_km = 10).
```

The description of the inputs that differ from MSCEM2 is given below:

maxclust The maximum number of clusters allowed if the AIC threshold is not met. Default is 6.

CaseK CaseK = 1 means that only the covariates are included when performing K-means. CaseK = 2 means that the survival times are also included. Default is CaseK = 1.

maxrestarts_km The maximum number of restarts with different seeds for K-means function. Default is 10, where as 1 the default set by R Core Team (2020).

maxit_km The maximum number of iterations allowed. Default is 10, which is the same default set by R Core Team (2020).

With these inputs, the function returns the following outputs:

$1, \ldots, C+2$ Each value gives all of the information given in MSCEM2 up to the $(C+2)^{th}$ cluster values greater than the C^{th} cluster with lowest AIC value.

AICC The list of all the AIC values for all clusters

	calculated.
BIC	The list of all the BIC values for all clusters calculated.
best number of clusters aicc	The cluster whose AIC value is the lowest.
best number of clusters bic	The cluster whose BIC value is the lowest.

This function with $n = 300$ takes multiple hours to run, due to the nature of running MSCEM2, which takes about 30 minutes, multiple times. This runtime is due to the fact that MSCC runs a minimum of 3 clusters, if the AIC is lowest of 1 cluster. Otherwise, it runs even more clusters.

4.3 Comparison to Other Packages

There is a package in R called *lcmm*, which stands for latent class mixed models, by Proust-Lima et al. (2019) based on Proust-Lima et al. (2017). The package is designed to estimate latent class mixed models (LCMM), joint latent class mixed models (JLCM), and mixed models with univariate and multivariate outcomes. They estimated all parameters using a maximum likelihood framework. The package also contains multiple post fit functions.

The package includes 4 estimation functions:

hlme	Estimation of latent class linear mixed models.
lcmm	Estimation of univariate latent process (and latent class) mixed models.
multlcmm or mlcmm	Estimation of multivariate latent process (and latent class) mixed models.
Jointlcmm or jlcmm	Estimation of joint latent class models for longitudinal and time to event data.

Of these 4 functions, the first three are very similar and deal with univariate and multivariate mixed models with different link functions. However, the Jointlcmm function is the one closest to my function MSCC. The Jointlcmm allows the user, given a specific number of clusters, to specify a survival model with longitudinal and baseline covariates. The function will return the values for the coefficients of the baseline hazard, the coefficient estimates for the longitudinal and baseline covariates, the coefficient estimates for the fixed effects of the longitudinal model, and the variance-covariance matrix of the random effects. The Jointlcmm function supports survival data that is observed, right censored, or left truncated.

In my MSCC function, I allow the user to specify a survival model with longitudinal and baseline covariates. Unlike Jointlcmm, the user currently has to specify the longitudinal fixed and random effects since my package does not yet support joint modeling. My function then outputs the coefficent estimates for the longitudinal and baseline covariates. Unlike Jointlcmm, the user is allowed to provide data that is all forms of censoring and truncation. Also, the user does not have to prespecify the number of latent classes; instead, the model determines which number of clusters is optimal based on AIC or BIC depending on which criteria the user wants to use.

There is another package in R called *JM*, which stand for joint model, by Rizopoulos (2010) based on the work in Rizopoulos et al. (2009). The function in there most similar to MSCC is called jointModel. Given a survival model with normal longitudinal responses and baseline covariates, the function returns the values for the coefficients of the baseline hazard, the coefficient estimates for the longitudinal and baseline covariates, the coefficient estimates for the fixed effects of the longitudinal model, and the variance-covariance matrix of the random effects. The jointModel function supports survival data that is observed and right censored.

The difference between jointModel and MSCC is that, since my package currently does not support joint modeling, the user has to specify the longitudinal fixed and random effects. However, the user is allowed to provide data that is all forms of censoring and truncation. Additionally, jointModel does not perform clustering of any form where as MSCC does unsupervised clustering.

Chapter 5

SWAN Dataset

5.1 Data Description

We were interested in applying our model to the Study of Women's Health Across the Nation (SWAN) dataset. Sutton-Tyrrell et al. (2010) states that the goal of the study "is to help scientists, health care providers, and women learn how mid-life experiences affect health and quality of life during aging." The study focused on women who were ages 40-55, living in the following geographic areas with the research centers given in parentheses:

Ypsilanti and Inkster, MI (University of Michigan)

Boston, MA (Massachusetts General Hospital)

Chicago, IL (Rush Presbyterian-St. Luke's Medical Center)

Alameda and Contra Costa County, CA (UC Davis and Kaiser Permanente)

Los Angeles, CA (UC Los Angeles)

Hackensack, NJ (Hackensack University Medical Center)

Pittsburgh, PA (University of Pittsburgh).

Each woman had to speak either English, Japanese, Cantonese, and/or Spanish and have the cognitive ability to consent to the study. They also had to belong to one of the five targeted racial/ethnic groups, which included African American, Asian American (Chinese and Japanese), Hispanic/Latino American, and White/Non-Hispanic American. They started with 202,985 individuals. After screening, 34,985 were considered eligible for the study. Of the 34,985 eligible, 16,065 completed the survey. Only 3,302 enrolled in the longitudinal survey.

The part of the study we are looking at ranges from 1996-2008 where each member was suppose to be tested 10 times. Even though 3,302 enrolled, only 2,245 completed their tenth visit. The study is still ongoing and is now conducting visits 15 and 16 for those in the original cohort. Furthermore, they have started new cohorts with each funding round.

The study contains three different data sets. One contains the cognitive scores and time-variant variables, another time-invariant variables, and the last dataset has nutrition variables. Complete detail of all variables is provided in Sutton-Tyrrell et al. (2010). The cognitive and time-variant variables include the following:

COGDAY	The day the cognitive function tests were taken. They were only taken on visits 4 and 6-10 if they were even taken.
IMED	Immediate recall of story. There were 12 ideas and the sum of the number of ideas they can recall after 5 seconds is the IMED score.
DELAY	Delayed recall of story. Same as above except that they asked a few moments later instead of seconds.
DIGIT	Number of correct responses to 12 questions which asks for subjects to say certain numbers backwards.

The next 3 terms require a little background. They all refer to the Symbol Digit Modalities Test (SDMT), which is a test for detecting cognitive impairment and takes about 5 minutes to administer. It is meant to catch brain damage and/or cognitive functional changes over time. The test consists of giving an individual a reference key where numbers have certain geometric shapes associated with them. The individual then has 90 seconds to pair a set of given numbers with the geometric shapes and can give responses either in written or oral form. There are 110 such questions per exam. Table 5.1 is the beginning of what a SDMT test would look like. An example test form is given in Langdon et al. (2011).

Example SDMT Test

KEY								
$\heartsuit$	$\int$	α	∞	$\triangle$	★	■	$\clubsuit$	$\diamondsuit$
1	2	3	4	5	6	7	8	9

$\int$	★	$\heartsuit$	$\clubsuit$	$\diamondsuit$	α	■	$\int$	$\clubsuit$

Table 5.1: Example Symbol Digit Modalities Test (SDMT)

However, this strictly gives the raw score. The test itself does not come with a specified cutoff for mental impairment. What is considered a low score is dependent on the condition of the individual and the study. For instance, the opening sentence of Greenslade and Pinot de Moira (2016) states "The SDMT manual does not provide Standard Scores." Therefore, the UK decided on the following formula as the criteria, which is widely used by educational statisticians. However, Greenslade and Pinot de Moira (2016) states that other fields should check the validity of the formula for their area of work. The method is:

1. Take the SDMT and substract the mean given for each age group.

2. Divide by the standard deviation given for each age group.

3. Multiply by 15.

4. Add 100.

5. Round up or down to get standard score.

Greenslade and Pinot de Moira (2016) states that the mean and standard deviation for each age in the UK is available and included the mean scores per age group. The focus of their document was on supporting special needs in schools. The SDMT test is also used for checking cognitive impairment in Multiple Sclerosis (MS) patients. In Sonder et al. (2014), the article is comparing multiple mental cognitive tests to determine which works best at assessing MS patients.

With this background on SDMT, here are the remaining cognitive variables:

SDMTCOR	The number correct $(0-110)$ for the Symbol Digit Modalities Test.
SDMTATM	The number attempted $(0-110)$ for the Symbol Digit Modalities Test.
SDMT	$(\#\ correct\ (0-110)\,/\#\ attempted\ (0-110)) * 12$.
COGSCORE	IMED+DELAY+DIGIT+SDMT.
GLBSCORE	Average of z scores of IMED, DELAY, DIGIT, and SDMT.

The time-variant variables included were BMI, both diastolic and systolic blood pressure, pulse, three different hormone measures, and four different cardiovascular measures, like total cholesterol and glucose. These variables could be taken every visit but may not have. The study also contained 12 time-invariant variables including age, occupation, race, current religious preference, and smoking status. Lastly, the study contained 35 nutritional variables such as alpha carotene, Animal Zinc, and B1 which were taken on visits 0, 5, 9 if they were taken.

Histograms of SDMTCOR, SDMT, COGSCORE, COGSCORE Percentile, and GLBSCORE are in given in Figures 5.1-5.2.

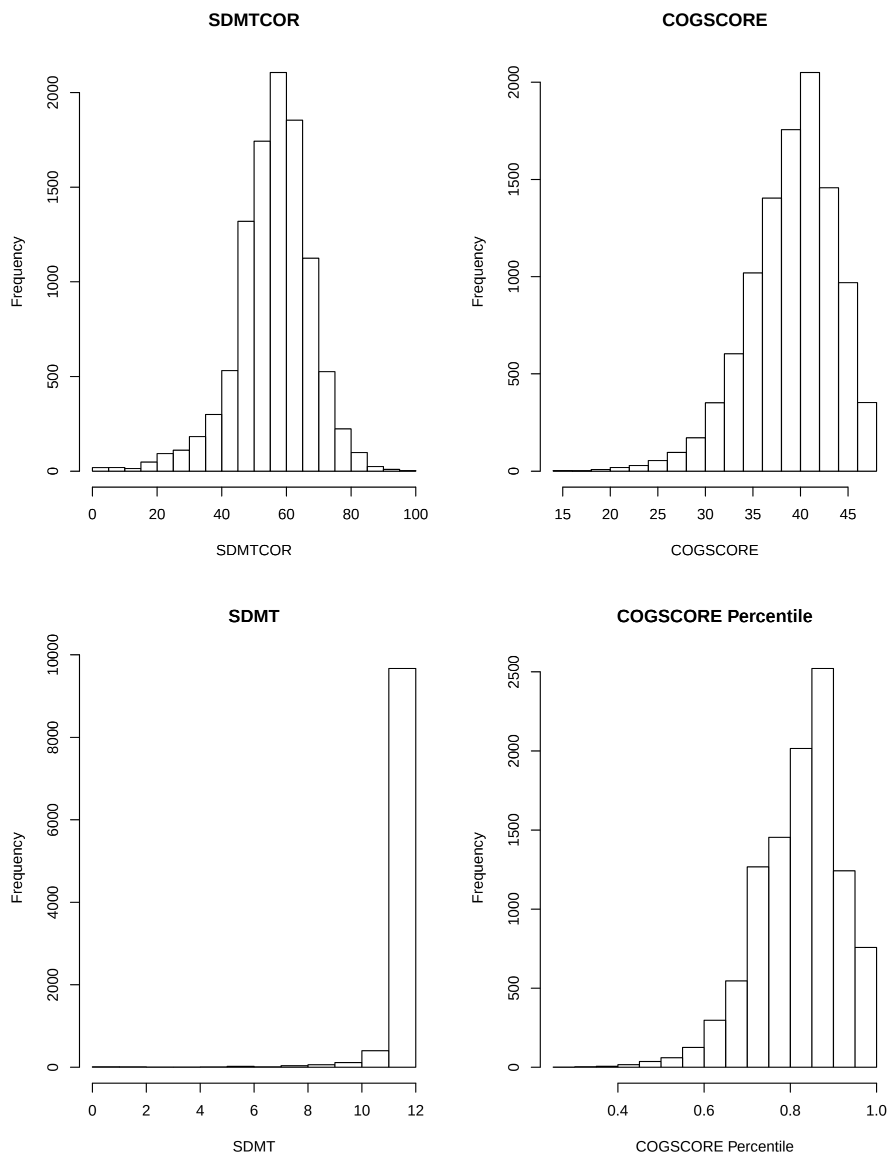

Figure 5.1: SDMTCOR, SDMT, COGSCORE, and COGSCORE divided by 48

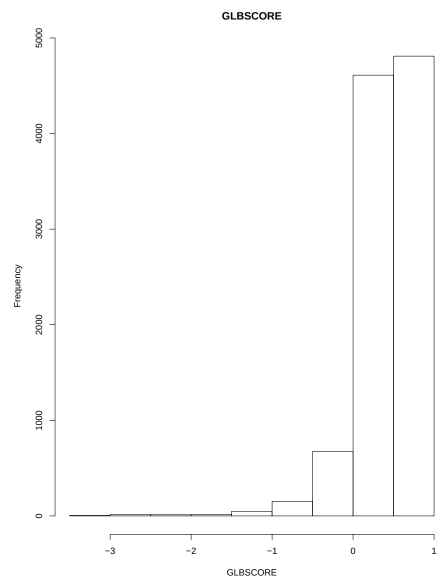

Figure 5.2: GLBSCORE variable for all individuals for all tests

5.2 Data Setup and Reasoning

From the SWAN dataset, we removed any time points with missing values across all variables of interest. Additionally, for the longitudinal variables in our analysis, we only used individuals with at least 4 longitudinal measures. This left 2,643 individuals in our analysis where 753 were African-American, 234 were Chinese, 258 were Japanese, 1,245 were White Non-Hispanic, and 153 were Hispanic. Since there is no set threshold for cognitive impairment when using SDMT, we used the SDMT values given by the variable SDMT in the SWAN dataset as the score received from the test. Additionally, we did not find any literature about what are acceptable values for people not already suffering some sort of disease or impaired mental state as in the cohort from the SWAN dataset. Therefore, we decided to use values less than 9 out of 12 as our cutoff, since that would be the equivalent of a 75% on an exam. In order to structure the data in a survival analysis setting, we decided that the first time an individual's SDMT score was below 9, that would be considered a failure. Using this criteria, we ended up with 2,510 right censored individuals, 52 left censored individuals, and 81 interval censored individuals.

The longitudinal variables we used were diastolic blood pressure (DIABP) and systolic blood pressure (SYSBP). In an article on the Johns Hopkins Medicine (2020) website, it states that having too high of systolic blood pressure can hurt the vessels in the brain increasing the chance of Alzheimer's disease. Both McLeod (2019) in an article on *The Conversation* and Johns Hopkins Medicine (2020) in an article on their website state that having too low of diastolic blood pressure can increase the chance of dementia due to the fact that low diastolic blood pressure means blood is not being pushed through the body strong enough and therefore the brain is not receiving enough oxygenated blood.

The baseline covariates we used were the ones labeled RACE and HOUSEHL. On the Alzheimer's Association (2020) website, it states that older Latinos and African-Americans are one-and-a-half and twice, respectively, as likely to get Alzheimer's diesease as older whites, and therefore we included race. HOUSEHL was a variable stating whether or not the woman lived alone. In Desai et al. (2020), it states that living alone increases the risk of dementia by approximately 30% and therefore we included the variable HOUSEHL.

5.3 Data Analysis

The variables we included in our analysis were:

ID The identification number given the woman when entering the study.

COGYEAR	This was the variable COGDAY, which was the day since entering the study the cognitive function tests were taken, divided by 365.
VISITS	Which visit number the variable was recorded on.
SDMT	$(\#\text{correct on SDMT}\,(0-110)\,/\#\text{attempted on SDMT}\,(0-110)) * 12$.
DIABP	Diastolic blood pressure measurements.
SYSBP	Systolic blood pressure measurements.
RACE	Coded 1-5 for the 5 races included in the study where 1 was African-American, 2 was Chinese, 3 was Japanese, 4 was White Non-Hispanic, and 5 was Hispanic. We coded it as an indicator variable where 0 was Latino or African-American and 1 was Chinese, Japanese, and White Non-Hispanic.
HOUSEHL	Coded where 1 means the person lives alone and 2 means she lives with others. We coded it as an indicator variable where 0 means she lives alone and 1 means she lives with others.

Once we had decided on the variables and their coding scheme, we normalized the values of the variables DIABP and SYSBP. Since these are longitudinal covariates, we decided to treat them as B-splines. We had 3 knots for both the mean trajectory as well as the individual trajectories for each of the splines at times 0, 6, and 12 since the maximum time value was 11.48. We decided to let the mean trajectory be a cubic spline and therefore it has 5 coefficients. We decided that the individual trajectories should be linear splines. We did not believe higher order splines seem reasonable, since we are requiring a minimum number of longitudinal measures per individual of 4. Therefore each individual's trajectory function has 3 coefficients. We then used the R package *lme4* to estimate the mean and longitudinal coefficients. Using these estimates, along with the survival times and baseline covariates, we found the optimal number of clusters and coefficient values.

5.4 Results

Given in Table 5.12 is the AIC and BIC values per cluster.

Cluster	AIC Value	BIC Value
1	1353.88	1377.40
2	1352.12	1405.04
3	1357.03	1439.34
4	1368.95	1480.66

Table 5.12: Cluster Number, AIC Value, and BIC Value

Since the lowest AIC value of 1352.12 occurs with 2 clusters, the method checks that both cluster's 3 and 4 have AIC values greater than 2 before is quits making more clusters. In Table 5.13, the values for $\boldsymbol{\omega}$ and σ_b^2 for the best clusters are given.

AIC			BIC		
Best # of Clusters	$\boldsymbol{\omega}$	σ_b^2	Best # of Clusters	$\boldsymbol{\omega}$	σ_b^2
2	0.56	0.06	1	1.0	0.06
	0.44	0.06			

Table 5.13: Best cluster $\boldsymbol{\omega}$ and σ_b^2 values by AIC and BIC

Since $n = 2,643$, the penalty on BIC is quite large. Therefore, we decided to use the AIC criterion for the optimal number of clusters, which was 2. The coefficient values with their standard errors for the 2 clusters can be seen in Table 5.14.

Variable	Cluster 1	Cluster 2
DIABP	-0.95 (0.26)	0.36 (0.22)
SYSBP	0.69 (0.22)	-0.20 (0.20)
RACE	0.82 (0.37)	-5.43 (1.12)
HOUSEHL	-0.04 (0.42)	2.52 (0.14)

Table 5.14: Coefficient values and standard errors for the 2 clusters

Cluster 2 contains 87 individuals; however, 4 individuals have probabilities of being in the cluster between 0.50 and 0.51 and are not very representative of the rest of the group. 3 of the 4 are right censored and 1 is left censored. The remaining 83 individuals are either African American or Hispanic as they have 0 for RACE. They also all have 1 for the variable HOUSEHL meaning all the individuals live with at least one other person. The survival times are all either left or interval censored, with 35 being left censored and 48 being interval censored. In cluster 1, 16 are left censored and 33 are interval censored. Of the 81 individuals who are left or interval censored, 80 are either Chinese, Japanese, or White Non-Hispanic as their indicator values are 1 for the variable RACE. Most of the women live with at least one other person as the value for HOUSEHL is 1, but some do not. Right censored individuals had probabilities of being in cluster 1 of 0.67 with median value of 0.55. For cluster 2, all three individuals where below 0.51.

Chapter 6

Future Work

6.1 Short-Term Future Work

We are currently looking at accurately estimating the variance as we believe it is currently under estimated. We have looked at Louis (1982), but the covariance-variance matrices for smaller sample sizes have not always been positive-definite. We are now currently looking at implementing the technique in Meng and Rubin (1991). Additionally, we are working on expanding this model to work in a joint modeling setting. We decided to use an MCEM framework for our joint survival-longitudinal model with clustering.

For the package, I am currently working on making the joint model functional. I have been able to get the MCE-step running, and part of the M-step working. However, I am still working getting the rest of the M-step running. Our goal is to apply this joint model to the Study of Women's Health Across the Nation (SWAN) dataset described above. The theory is given below.

6.2 Current MCEM Work

The complete data likelihood for the joint model is:

$$
\begin{aligned}
&L\left(\Theta;\boldsymbol{T}^l,\boldsymbol{T}^r,\boldsymbol{L},\boldsymbol{R},\boldsymbol{\delta}_O,\boldsymbol{\delta}_R,\boldsymbol{\delta}_L,\boldsymbol{\delta}_I,\boldsymbol{Z},\boldsymbol{Y},\boldsymbol{V},\boldsymbol{\xi}\right)= \\
&\prod_{i=1}^{n}\prod_{c=1}^{C}\left\{\omega_c\left(\frac{\lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)e^{-\Lambda_c(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{O_i}}\left(\frac{e^{-\Lambda_c(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{R_i}}\right. \\
&\quad\times\left(\frac{1-e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{L_i}}\left(\frac{e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}-e^{-\Lambda_c(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{I_i}} \\
&\quad\left.\times\prod_{j=1}^{p}\left(\sigma^2_{Y_jc}\right)^{-\frac{m_{ij}}{2}}e^{\left(-\|\boldsymbol{Y}_{ij}-\boldsymbol{X}_{ijc}\|^2/2\sigma^2_{Y_jc}\right)}\det\left(\Sigma_{\boldsymbol{\xi}_jc}\right)^{-\frac{1}{2}}e^{-\frac{\boldsymbol{\xi}_{ij}^T\Sigma^{-1}_{\boldsymbol{\xi}_jc}\boldsymbol{\xi}_{ij}}{2}}\right\}^{v_{ic}},
\end{aligned}
$$

where $\Theta=\left(\alpha_0,\alpha_1,\boldsymbol{b},\omega_c,\boldsymbol{\gamma}_c,\boldsymbol{\zeta}_c,\sigma^2_b,\boldsymbol{\sigma}^2_Y,\boldsymbol{\beta}_{jc},\Sigma_{\boldsymbol{\xi}_jc}\right)$, $c=1,\ldots,C$, and $j=1,\ldots,p$.

In the joint model, along with the terms to estimate in the survival model with clustering, we have to estimate the $C\times p$ matrix of error variances, $\boldsymbol{\sigma}^2_Y$, each of the p covariance-variance matrices for the C clusters, $\Sigma_{\boldsymbol{\xi}_jc}$, and each individual's $\boldsymbol{X}_{ic}$ and $\boldsymbol{X}_{ijc}$. Now in order to estimate $\boldsymbol{X}_{ic}$ and $\boldsymbol{X}_{ijc}$, we have to estimate the mean trajectory coefficients for the longitudinal covariates per cluster, $\boldsymbol{\beta}_{jc}$, along with each individual's individual trajectory coefficients, $\boldsymbol{\xi}_{ij}$, where $j=1,\ldots,p$ and $c=1,\ldots,C$. Since the $\boldsymbol{\xi}_i$ individual trajectory coefficients are random effects, then they are latent variables like the cluster membership, $\boldsymbol{\tau}_i$, for $i=1,\ldots,n$. We chose to use the Monte Carlo Expectation-Maximization (MCEM) algorithm to estimate the parameters as in Huang et al. (2014).

6.2.1 Overview of MCEM Methodology

The overview of the MCEM methodology is as follows:

1. Given that the complete data likelihood has the above form, we take take the loglikelihood. Therefore, $\ell\left(\Theta; \boldsymbol{T}^l, \boldsymbol{T}^r, \boldsymbol{L}, \boldsymbol{R}, \boldsymbol{\delta}_O, \boldsymbol{\delta}_R, \boldsymbol{\delta}_L, \boldsymbol{\delta}_I, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{\xi}\right) =$ $\log\left\{L\left(\Theta; \boldsymbol{T}^l, \boldsymbol{T}^r, \boldsymbol{L}, \boldsymbol{R}, \boldsymbol{\delta}_O, \boldsymbol{\delta}_R, \boldsymbol{\delta}_L, \boldsymbol{\delta}_I, \boldsymbol{Z}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{\xi}\right)\right\}$. From now on, we will write the loglikelihood as $\ell\left(\Theta; \boldsymbol{V}, \boldsymbol{\xi}\right)$ or $\ell\left(\Theta; \boldsymbol{\tau}, \boldsymbol{\xi}\right)$, depending on if the cluster assignment is known or not, for ease of reading.

2. Now, $\mathcal{Q}\left(\Theta|\Theta_{prev}\right) = \sum_{i=1}^{n} E\left[\ell_i\left(\Theta; T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i, \boldsymbol{Y}_i, \boldsymbol{\tau}_i, \boldsymbol{\xi}_i\right) \mid T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i, \boldsymbol{Y}_i, \Theta_{prev}\right]$. Therefore, $\mathcal{Q}$ is the expected loglikelihood and from now on we write $\mathcal{Q}\left(\Theta|\Theta_{prev}\right) = \sum_{i=1}^{n} E\left[\ell_i\left(\Theta; \boldsymbol{\tau}_i, \boldsymbol{\xi}_i\right)|\Theta_{prev}\right]$, for ease of reading.

3. We approximate $\mathcal{Q}\left(\Theta|\Theta_{prev}\right)$ by its Monte Carlo estimate,

$$\widehat{\mathcal{Q}}\left(\Theta|\Theta_{prev}\right) = \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\ell_i\left(\Theta; \boldsymbol{\tau}_i^{(r)}, \boldsymbol{\xi}_i^{(r)}\right),$$

   where R is the number of Monte Carlo replicates and both $\boldsymbol{\tau}_i^{(r)}$ and $\boldsymbol{\xi}_i^{(r)}$ are samples from the conditional distribution $\left[\boldsymbol{\tau}_i^{(r)}, \boldsymbol{\xi}_i^{(r)}|T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i, \boldsymbol{Y}_i, \Theta_{prev}\right]$.

4. Since this distribution does not have a closed form, we use a Gibbs sampler incorporated with a Metropolis-Hastings step.

5. In the M-step, we need to maximize $\widehat{\mathcal{Q}}\left(\Theta|\Theta_{prev}\right)$, where $\Theta = \left(\alpha_0, \alpha_1, \boldsymbol{b}, \omega_c, \boldsymbol{\gamma}_c, \boldsymbol{\zeta}_c, \sigma_b^2, \boldsymbol{\sigma}_Y^2, \boldsymbol{\beta}_{jc}, \Sigma_{\boldsymbol{\xi}_j c}\right)$, $c = 1, \ldots, C$, and $j = 1, \ldots, p$.

6. Now, $\widehat{\mathcal{Q}}\left(\Theta|\Theta_{prev}\right)$ can be factored into the following way:

$$\begin{aligned}\widehat{\mathcal{Q}}\left(\Theta|\Theta_{prev}\right) =& \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(r)}\log\left(\omega_c\right)\\ &+\frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(r)}\log\left[f_c\left(T_i^l,T_i^r,L_i,R_i,\delta_{O_i},\delta_{R_i},\delta_{L_i},\delta_{I_i},\boldsymbol{Z}_i|\boldsymbol{\xi}_i^{(r)},\Theta_{prev}\right)\right]\\ &+\frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(r)}\sum_{j=1}^{p}\log\left[f_c\left(\boldsymbol{Y}_{ij}|\boldsymbol{\xi}_{ij}^{(r)},\Theta_{prev}\right)\right]\\ &+\frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(r)}\sum_{j=1}^{p}\log\left[f_c\left(\boldsymbol{\xi}_{ij}^{(r)}|\Theta_{prev}\right)\right],\end{aligned}$$

where the three density functions per cluster c are defined as:

(a) The survival submodel is denoted by:

$$\begin{aligned}&f_c\left(T_i^l,T_i^r,L_i,R_i,\delta_{O_i},\delta_{R_i},\delta_{L_i},\delta_{I_i},\boldsymbol{Z}_i|\boldsymbol{\xi}_i,\Theta_{prev}\right)=\\ &\left(\frac{\lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{O_i}}\left(\frac{e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{R_i}}\\ &\times\left(\frac{1-e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{L_i}}\left(\frac{e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}-e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)}}{e^{-\Lambda_c(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}-e^{-\Lambda_c(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i)}}\right)^{\delta_{I_i}}.\end{aligned}$$

(b) The longitudinal submodel is denoted by

$$f_c\left(\boldsymbol{Y}_{ij}|\boldsymbol{\xi}_{ij},\Theta_{prev}\right)=\left(\sigma^2_{Y_jc}\right)^{-\frac{n_{ij}}{2}}e^{\left(-\|\boldsymbol{Y}_{ij}-\boldsymbol{X}_{ijc}\|^2/2\sigma^2_{Y_jc}\right)},\text{ for } j=1,\ldots,p.$$

(c) The random effects density is given by

$$f_c\left(\boldsymbol{\xi}_{ij}|\Theta_{prev}\right)=\det\left(\Sigma_{\boldsymbol{\xi}_jc}\right)^{-\frac{1}{2}}e^{-\frac{\boldsymbol{\xi}_{ij}^T\Sigma^{-1}_{\boldsymbol{\xi}_jc}\boldsymbol{\xi}_{ij}}{2}},\text{ for } j=1,\ldots,p.$$

However, the survival and longitudinal densities share $\boldsymbol{\beta}'_{jc}s$, and therefore we cannot maximize them separately. In order to estimate the parameters from Θ, we decided to use Newton optimization, which involves finding $\boldsymbol{Q}$ and $\mathbb{Q}$ of the loglikelihood as before, but they are now augmented in order to also estimate $\boldsymbol{\beta}'_{jc}s$. In addition, we must also estimate $\boldsymbol{\sigma}^2_Y$ and $\Sigma_{\boldsymbol{\xi}_jc}$, $j=1,\ldots,p$ and $c=1,\ldots,C$.

6.2.2 Detailed Description of MCEM Methodology

E-Step Methodology

The E-Step Methodology for $r = 1, \ldots, R$ per individual, i, $i = 1, \ldots, n$ is:

1. The marginal distribution of $\boldsymbol{\tau}_i$ is $\boldsymbol{\tau}_i \sim Multinomial\left[1, (\omega_{1,prev}, \ldots, \omega_{C,prev})\right].$

2. For $\boldsymbol{\tau}_i^{(r+1)}$, we have that $[\boldsymbol{\tau}_i|\Theta_{prev}] \sim Multinomial\left[1, (\widetilde{\omega}_{i1}, \ldots, \widetilde{\omega}_{iC})\right]$ where

$$\widetilde{\omega}_{ic} = \frac{\omega_{c,prev} A_c \left(\boldsymbol{\xi}_1^{(r)}, \ldots, \boldsymbol{\xi}_p^{(r)}, \Theta_{prev}\right)}{\sum_{c'=1}^{C} \omega_{c',prev} A_{c'} \left(\boldsymbol{\xi}_1^{(r)}, \ldots, \boldsymbol{\xi}_p^{(r)}, \Theta_{prev}\right)}.$$

Now, $A_c\left(\boldsymbol{\xi}_1^{(r)}, \ldots, \boldsymbol{\xi}_p^{(r)}, \Theta_{prev}\right)$ equals:

$$A_c\left(\boldsymbol{\xi}_1^{(r)}, \ldots, \boldsymbol{\xi}_p^{(r)}, \Theta_{prev}\right) = f_c\left(T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i|\boldsymbol{\xi}_i, \Theta_{prev}\right) \times \prod_{j=1}^{p} f_c\left(\boldsymbol{Y}_{ij}|\boldsymbol{\xi}_{ij}, \Theta_{prev}\right) f_c\left(\boldsymbol{\xi}_{ij}|\Theta_{prev}\right).$$

3. Since there can be p different $\boldsymbol{\xi}_{ij}$'s that we have to estimate, we update each set separately, holding the other ones fixed. We use the updated values from the $\boldsymbol{\xi}_{ij}$'s already updated when estimating the current one. For a specific $\boldsymbol{\xi}_{ij}$, we have that

$$f\left(\boldsymbol{\xi}_{ij}|T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i, \boldsymbol{Y}_i, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) \propto$$
$$f\left(T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i|\boldsymbol{\xi}_i, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)$$
$$f\left(\boldsymbol{Y}_i|\boldsymbol{\xi}_i, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) f\left(\boldsymbol{\xi}_{ij}|\boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right).$$

Since this function does not have a closed form, we use a Metropolis-Hastings independence sampler. We sample $\boldsymbol{\xi}_{ij}^*$ from $f\left(\boldsymbol{\xi}_{ij}|\boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)$, which is multivariate normal, $N_{n_{ij}}\left(\mathbf{0}, \Sigma_{\boldsymbol{\xi}_{ij}c,prev}\right)$. Since $\boldsymbol{\xi}_i$ stands for all p longitudinal individual trajectory coefficients for an individual, $\boldsymbol{\xi}_i^*$ and $\boldsymbol{\xi}_i^{(r)}$ means that only the current $\boldsymbol{\xi}_{ij}$ that is being updated is changing while all the other $\boldsymbol{\xi}_{ij}'s$ are fixed. Therefore, the acceptance ratio becomes:

$$\frac{f\left(\boldsymbol{\xi}_{ij}^*|T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i, \boldsymbol{Y}_i, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) f\left(\boldsymbol{\xi}_{ij}^{(r)}|\boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)}{f\left(\boldsymbol{\xi}_{ij}^{(r)}|T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i, \boldsymbol{Y}_i, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) f\left(\boldsymbol{\xi}_{ij}^*|\boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)} =$$

$$\frac{f\left(T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i|\boldsymbol{\xi}_i^*, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) f\left(\boldsymbol{Y}_i|\boldsymbol{\xi}_i^*, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)}{f\left(T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i|\boldsymbol{\xi}_i^{(r)}, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) f\left(\boldsymbol{Y}_i|\boldsymbol{\xi}_i^{(r)}, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)}$$

$$\times \frac{f\left(\boldsymbol{\xi}_{ij}^*|\boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) f\left(\boldsymbol{\xi}_{ij}^{(r)}|\boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)}{f\left(\boldsymbol{\xi}_{ij}^{(r)}|\boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) f\left(\boldsymbol{\xi}_{ij}^*|\boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)} =$$

$$\frac{f\left(T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i|\boldsymbol{\xi}_i^*, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) f\left(\boldsymbol{Y}_i|\boldsymbol{\xi}_i^*, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)}{f\left(T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i|\boldsymbol{\xi}_i^{(r)}, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) f\left(\boldsymbol{Y}_i|\boldsymbol{\xi}_i^{(r)}, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)}.$$

The three density functions are defined as follows:

(a) The survival submodel is denoted by:

$$f\left(T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i|\boldsymbol{\xi}_i, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) =$$
$$\prod_{c=1}^{C}\left[f_c\left(T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i|\boldsymbol{\xi}_i, \Theta_{prev}\right)\right]^{\tau_{ic}^{(r+1)}}.$$

(b) The longitudinal submodel is denoted by

$$f\left(\boldsymbol{Y}_i|\boldsymbol{\xi}_i, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) = \prod_{c=1}^{C}\left[\prod_{j=1}^{p} f_c\left(\boldsymbol{Y}_{ij}|\boldsymbol{\xi}_{ij}, \Theta_{prev}\right)\right]^{\tau_{ic}^{(r+1)}}.$$

(c) The random effects density is given by

$$f\left(\boldsymbol{\xi}_{ij}|\boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) = \prod_{c=1}^{C}\left[f_c\left(\boldsymbol{\xi}_{ij}|\Theta_{prev}\right)\right]^{\tau_{ic}^{(r+1)}} \text{ for } j = 1, \ldots, p.$$

We set $\boldsymbol{\xi}_{ij}^{(r+1)} = \boldsymbol{\xi}_{ij}^{*}$ with probability

$$\alpha_{prob} = \min\left\{1, \frac{f\left(T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i | \boldsymbol{\xi}_i^{*}, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) f\left(\boldsymbol{Y}_i | \boldsymbol{\xi}_i^{*}, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)}{f\left(T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{Z}_i | \boldsymbol{\xi}_i^{(r)}, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right) f\left(\boldsymbol{Y}_i | \boldsymbol{\xi}_i^{(r)}, \boldsymbol{\tau}_i^{(r+1)}, \Theta_{prev}\right)}\right\}$$

and $\boldsymbol{\xi}_{ij}^{(r+1)} = \boldsymbol{\xi}_{ij}^{(r)}$ otherwise. We write it as:

$$\boldsymbol{\xi}_{ij}^{(r+1)} = \begin{cases} \boldsymbol{\xi}_{ij}^{*} & \text{with probability } \alpha_{prob}. \\ \boldsymbol{\xi}_{ij}^{(r)} & \text{otherwise.} \end{cases}$$

M-Step Methodology

The M-Step Methodology for $r = 1, \ldots, R$ for maximizing $\widehat{\mathcal{Q}}\left(\Theta|\Theta_{prev}\right)$, where we need to find the estimates of $\Theta = \left(\alpha_0, \alpha_1, \boldsymbol{b}, \omega_c, \boldsymbol{\gamma}_c, \boldsymbol{\zeta}_c, \sigma_b^2, \boldsymbol{\sigma}_Y^2, \boldsymbol{\beta}_{jc}, \Sigma_{\boldsymbol{\xi}_j c}\right)$, $c = 1, \ldots, C$, and $j = 1, \ldots, p$ is:

1. We find that the estimate for ω_c is:

$$\widehat{\omega}_c = \sum_{r=1}^{R} \sum_{i=1}^{n} \frac{\tau_{ic}^{(r)}}{nR} \text{ for } c = 1, \ldots, C.$$

2. We find that the estimate for $\sigma_{Y_j c}^2$ is:

$$\widehat{\sigma}_{Y_j c}^2 = \frac{\sum_{r=1}^{R} \sum_{i=1}^{n} \tau_{ic}^{(r)} \|\boldsymbol{Y}_{ij} - \boldsymbol{X}_{ijc}^{(r)}\|^2}{\sum_{r=1}^{R} \sum_{i=1}^{n} \tau_{ic}^{(r)} m_{ij}}.$$

3. We find that the estimate for $\Sigma_{\boldsymbol{\xi}_j c}$ is:

$$\widehat{\Sigma}_{\boldsymbol{\xi}_j c} = \frac{\sum_{r=1}^{R} \sum_{i=1}^{n} \tau_{ic}^{(r)} \left(\boldsymbol{\xi}_{ij}^{(r)} \boldsymbol{\xi}_{ij}^{T(r)}\right)}{\sum_{r=1}^{R} \sum_{i=1}^{n} \tau_{ic}^{(r)}}.$$

4. Derivations for the estimates of $\sigma^2_{Y_jc}$ and $\Sigma_{\boldsymbol{\xi}_jc}$ are found in Appendix I. The remaining terms of Θ are found using Newton optimization and thus we must estimate $\boldsymbol{Q}$ and $\mathbb{Q}$ per cluster c, $c = 1, \ldots, C$. We will only being looking at a single cluster to demonstrate estimating both $\boldsymbol{Q}$ and $\mathbb{Q}$. Since we have shown how to estimate most of the parameters of Θ before, we only demonstrate here how to find one of first and second derivatives of $\beta_{jk'}$ to account for the joint model. All of the derivatives of $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ and $\widehat{\mathcal{Q}}_p\left(\Theta|\Theta_{prev}\right)$ with respect to $\beta_{jk'}$ are found in Appendix H. The rest of the terms are estimated as found in Chapter 3 except for now they take into account the R replicates.

5. The 1^{st} derivative of $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ with respect to $\beta_{jk'}$, $k' = 1, \ldots, \overline{J}$ is:

$$\begin{aligned}\frac{\partial}{\partial \beta_{jk'}}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) =& e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}}\gamma_j\overline{B}_{jk'}\left(u\right)e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du \\ &+ e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t}\gamma_j\overline{B}_{jk'}\left(u\right)e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du.\end{aligned}$$

6. The 2^{nd} derivative of $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ with respect to $\beta_{jk'}$, $k' = 1, \ldots, \overline{J}$ is:

$$\begin{aligned}\frac{\partial}{\partial \beta_{jk'}\partial \beta_{j'k''}}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) =& e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}}\gamma_j\overline{B}_{jk'}\left(u\right)\gamma_{j'}\overline{B}_{j'k''}\left(u\right)e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du \\ &+ e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t}\gamma_j\overline{B}_{jk'}\left(u\right)\gamma_{j'}\overline{B}_{j'k''}\left(u\right)e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du.\end{aligned}$$

7. The first derivative of $\widehat{\mathcal{Q}}_p\left(\Theta|\Theta_{prev}\right)$,with respect to $\beta_{jk'c}$ is:

$$\begin{aligned}\frac{\partial}{\partial\beta_{jk'c}}\widehat{\mathcal{Q}}_p\left(\Theta|\Theta_{prev}\right)=&\frac{1}{R}\sum_{i=1}^{n}\sum_{r=1}^{R}\tau_{ic}^{(r)}\frac{\overline{\boldsymbol{B}}_{jk'}^{T}\left(\boldsymbol{Y}_{ij}-\boldsymbol{X}_{ijc}^{(r)}\right)}{\sigma_{Y_jc}^2}+\frac{1}{R}\sum_{r=1}^{R}\frac{\partial}{\partial\beta_{jk'c}}\ell_{p,c}^{(r)}\left(\boldsymbol{\theta};\sigma_b^2\right)\\=&\sum_{i=1}^{n}\frac{\overline{\boldsymbol{B}}_{jk'}^{T}\left(\boldsymbol{Y}_{ij}-\overline{\boldsymbol{X}}_{ijc}\right)}{\sigma_{Y_jc}^2}+\frac{1}{R}\sum_{r=1}^{R}\frac{\partial}{\partial\beta_{jk'c}}\ell_{p,c}^{(r)}\left(\boldsymbol{\theta};\sigma_b^2\right),\end{aligned}$$

where $\mathcal{Q}_p$ is $\mathcal{Q}$ now taking into account the penalized survival likelihood and $\ell_{p,c}^{(r)}\left(\boldsymbol{\theta};\sigma_b^2\right)$ is the penalized survival likelihood as stated in Chapter 3 now taking into account the R replicates of $\tau_{ic}^{(r)}$ and $\boldsymbol{\xi}_{ij}^{(r)}$. Now $X_{ijc}^{(r)}(t)=\psi_{jc}(t)+\sum_{k'=1}^{J}\xi_{ijk'}^{(r)}B_{k'}(t)$, where $\psi_{jc}(t)$ is the j^{th} mean trajectory function for cluster c and is given by $\psi_{jc}(t)=\sum_{k'=1}^{\overline{J}}\beta_{jk'c}\overline{B}_{k'}(t)$. Thus, only $\boldsymbol{\xi}_{ij}$ depends on r and in the last line we get $\overline{X}_{ijc}(t)=\sum_{k'=1}^{\overline{J}}\beta_{jk'c}\overline{B}_{k'}(t)+\sum_{k'=1}^{J}\overline{\xi}_{ijk'}B_{k'}(t)$, where $\overline{\xi}_{ijk'}=\sum_{r=1}^{R}\xi_{ijk'}^{(r)}$. Note: R in these formulas is the replicates per individual i that occur in cluster c. Thus R per individual i per cluster c, which we denote by R_{ic}, in practice is calculated by $R_{ic}=\sum_{c=1}^{C}\tau_{ic}^{(r)}$. The first derivative of $\ell_{p,c}\left(\boldsymbol{\theta};\sigma_b^2\right)$ with respect to $\beta_{jk'c}$ is:

$$\begin{aligned}&\frac{\partial}{\partial\beta_{jk'c}}\ell_{p,c}\left(\boldsymbol{\theta};\sigma_b^2\right)=\frac{\partial}{\partial\beta_{jk'}}\ell_{0,c}\left(\boldsymbol{\theta};\sigma_b^2\right)\\&=\sum_{i=1}^{n}\tau_{ic}^{(r)}\Bigg[\left(\gamma_j\overline{B}_{jk'}\left(T_i^r\right)\right)\delta_{O_i}-\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}{e^{\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\&\quad-\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\delta_{I_i}-\frac{\left(\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\right)}{e^{\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}-1}\delta_{I_i}\\&\quad+\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(L_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(L_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(R_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\right)}{e^{\Lambda_c\left(R_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\Lambda_c\left(L_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}-1}\Bigg].\end{aligned}$$

8. The second derivative of $\widehat{\mathcal{Q}}_p\left(\Theta|\Theta_{prev}\right)$ with respect to $\beta_{jk'c}$ and $\beta_{j'k''c}$ is:

$$\begin{aligned}\frac{\partial}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\widehat{\mathcal{Q}}_p\left(\Theta|\Theta_{prev}\right)=&\frac{1}{R}\sum_{i=1}^{n}\sum_{r=1}^{R}-\tau_{ic}^{(r)}\frac{\overline{\boldsymbol{B}}_{jk'}^{T}\overline{\boldsymbol{B}}_{j'k''}}{\sigma_{Y_j}^2}+\frac{1}{R}\sum_{r=1}^{R}\frac{\partial}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\ell_{p,c}\left(\boldsymbol{\theta};\sigma_b^2\right)\\=&\sum_{i=1}^{n}-\frac{\overline{\boldsymbol{B}}_{jk'}^{T}\overline{\boldsymbol{B}}_{j'k''}}{\sigma_{Y_j}^2}+\frac{1}{R}\sum_{r=1}^{R}\frac{\partial}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\ell_{p,c}\left(\boldsymbol{\theta};\sigma_b^2\right),\end{aligned}$$

where the second derivative of $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)$ with respect to $\beta_{jk'c}$ and $\beta_{j'k''c}$ is:

$$
\begin{aligned}
&\frac{\partial^2}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\ell_{p,c}\left(\boldsymbol{\theta};\sigma_b^2\right) = \frac{\partial^2}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\ell_{0,c}\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) + \frac{\frac{\partial^2}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}{e^{\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad -\frac{e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\frac{\partial}{\partial\beta_{j'k''c}}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad -\frac{\left(\frac{\partial^2}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\right)}{e^{\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\beta_{j'k''c}}\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{j'k''c}}\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\Lambda_c\left(L_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right) \\
&\quad +\frac{\left(\frac{\partial^2}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\Lambda_c\left(L_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\beta_{jk'c}\partial\beta_{j'k''c}}\Lambda_c\left(R_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\right)}{e^{\Lambda_c\left(R_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\Lambda_c\left(L_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda_c\left(L_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\Lambda_c\left(R_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(L_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'c}}\Lambda_c\left(R_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda_c\left(L_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\Lambda_c\left(R_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\beta_{j'k''c}}\Lambda_c\left(L_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{j'k''c}}\Lambda_c\left(R_i|\boldsymbol{X}_{ic}^{(r)},\boldsymbol{Z}_i\right)\right)\Bigg].
\end{aligned}
$$

9. Thus, $\boldsymbol{Q}_c = \frac{1}{R}\sum_{i=1}^{n}\sum_{r=1}^{R}\boldsymbol{Q}_{ic}^{(r)}$ and $\mathbb{Q}_c = \frac{1}{R}\sum_{i=1}^{n}\sum_{r=1}^{R}\mathbb{Q}_{ic}^{(r)}$.

10. We continue the process until $\max_j\left(\frac{\Theta_{curr,j}-\Theta_{prev,j}}{\Theta_{prev,j}+\delta_1}\right) < \delta_2$, where j is indexing for all terms in Θ. As stated in Huang et al. (2014), they used $\delta_1 = 0.001$ and $\delta_2 = 0.005$ as recommended in Booth and Hobert (1999).

In Figures 6.1-6.3, we provide trace plots for one individual's Gibbs Sampler for $\boldsymbol{\tau}_i$ and the Metropolis-Hastings independence sampler within a Gibbs Sampler of the first element of two $\boldsymbol{\xi}_{ij}$'s of dimension three where there are two clusters. We plotted every 100^{th} from a run of 100,000 iterations.

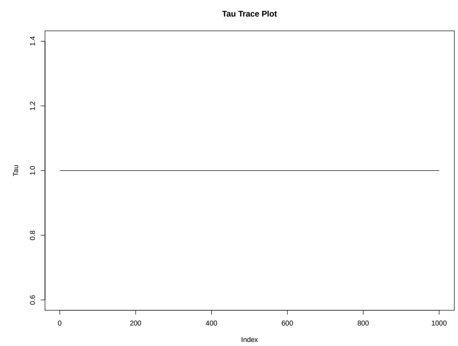

Figure 6.1: Plot of every 100^{th} iteration of τ_{ic}.

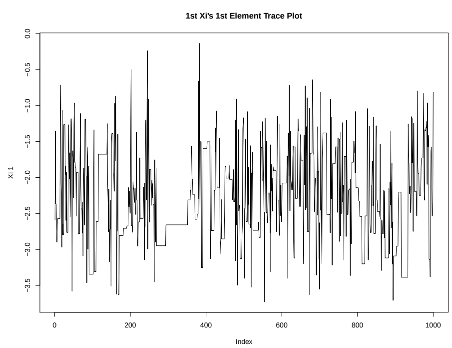

Figure 6.2: Plot of every 100^{th} iteration of the 1^{st} element of 1^{st} $\boldsymbol{\xi}_{ij}$.

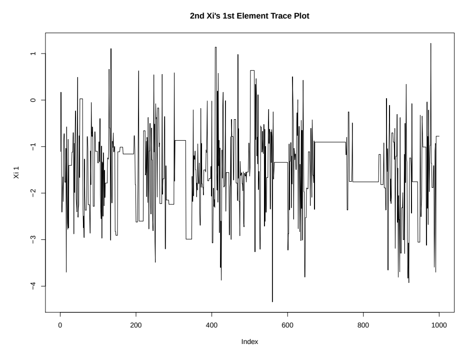

Figure 6.3: Plot of every 100^{th} iteration of the 1^{st} element of 2^{nd} $\boldsymbol{\xi}_{ij}$.

We can see that $\boldsymbol{\tau}_i$ is not mixing between the two clusters, but both of the $\boldsymbol{\xi}_{ij}$'s are mixing well.

6.3 Long-Term Future Work

Long term, once the package performs the joint survival-longitudinal model, I would like to add flexibility to the package to allow time-dependent inputs other than just longitudinal variables represented as B-splines. I would also like to add the ability to plot outputs of the data. For instance, I would like to have a plot of the fitted time values vs. the actual time values as is common in regression settings. Additionally, I would like to add the ability to perform clustering strictly in the longitudinal submodel as I have in the survival submodel. Since time is a factor, we may optimize the package by multi-threading it. This would probably be conducted using RcppParallel by Allaire et al. (2020) and further usage of RcppArmadillo.

In terms of theory, we are looking at adding that cluster membership could be affected by the individual's covariates. It would probably be modeled by logistic regression. However, estimating these parameters along with all the other parameters we are estimating may not be easily accomplished.

Bibliography

Akaike, H. (1973). Information theory and an extension of the maximumlikelihood principle. *The 2nd International Symposium on InformationTheory*, pages 267–281. eds. B. N. Petrov and F. Csaki, Budapest: Akademia Kiado.

Allaire, J., Francois, R., Ushey, K., Vandenbrouck, G., Geelnard, M., and Intel (2020). *RcppParallel: Parallel Programming Tools for 'Rcpp'*. R package version 5.0.1.

Alzheimer's Association (2020). Causes and risk factors for alzheimer's disease. Electronic. Accessed: December 30, 2020.

Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):265–285.

Brown, E. R., Ibrahim, J. G., and DeGruttola, V. (2005). A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73.

Cai, T. and Betensky, R. A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics*, 59(3):570–579.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury/Thompson Learning, 2nd edition.

Cox, D. (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*. Applied mathematical sciences (Springer-Verlag New York Inc.) ; v. 27. Springer-Verlag, New York.

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):1–22.

Desai, R., John, A., Stott, J., and Charlesworth, G. (2020). Living alone and risk of dementia: A systematic review and meta-analysis. *Ageing Research Reviews*, 62:101122.

Eddelbuettel, D. (2013). *Seamless R and C++ Integration with Rcpp*. Springer, New York. ISBN 978-1-4614-6867-7.

Eddelbuettel, D. and Balamuta, J. J. (2017). Extending extitR with extitC++: A Brief Introduction to extitRcpp. *PeerJ Preprints*, 5:e3188v1.

Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

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

Greenslade, S. and Pinot de Moira, A. (2016). *How to calculate standard scores for the Symbol Digit Modalities Test*. communicate-ed.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, 61(2):383–385.

Huang, H., Li, Y., and Guan, Y. (2014). Joint modeling and clustering paired generalized longitudinal trajectories with application to cocaine abuse treatment data. *Journal of the American Statistical Association*, 109(508):1412–1424.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Johns Hopkins Medicine (2020). Blood pressure and alzheimer's risk: What's the connection? Electronic. Accessed:December 30, 2020.

Klein, J. P. and Moeschberger, M. L. (2011). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, 2nd edition.

Langdon, D., Amato, M., Boringa, J., Brochet, B., Foley, F., Fredrikson, S., Hämäläinen, P., Hartung, h.-p., Krupp, L., Penner, I., Reder, A., and Benedict, R. (2011). Recommendations for a brief international cognitive assessment for multiple sclerosis (bicams). *Multiple sclerosis (Houndmills, Basingstoke, England)*, 18:891–8.

Laplace, P. S. (1986). Memoir on the probability of the causes of events. *Statistical Science*, 1(3):364–378.

Louis, T. A. (1982). Finding the observed information matrix when using theEMAlgorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):226–233.

McLeod, K. (2019). Low blood pressure could be a culprit in dementia, studies suggest. Electronic. Accessed:December 30, 2020.

Meng, X.-L. and Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association*, 86(416):899–909.

Naik, P. A., Shi, P., and Tsai, C.-L. (2007). Extending the akaike information criterion to mixture regression models. *Journal of the American Statistical Association*, 102(477):244–254.

Proust-Lima, C., Philipps, V., Diakite, A., and Liquet, B. (2019). *lcmm: Extended Mixed Models Using Latent Classes and Latent Processes*. R package version: 1.8.1.

Proust-Lima, C., Philipps, V., and Liquet, B. (2017). Estimation of extended mixed models using latent classes and latent processes: The R package lcmm. *Journal of Statistical Software*, 78(2):1–56.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57(1):253–259.

Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33.

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.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Number 12. Cambridge University Press.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Sonder, J. M., Burggraaff, J., Knol, D. L., Polman, C. H., and Uitdehaag, B. M. (2014). Comparing long-term results of pasat and sdmt scores in relation to neuropsychological testing in multiple sclerosis. *Multiple Sclerosis Journal*, 20(4):481–488.

Stewart, J. (2008). *Calculus Early Transcendentals Sixth Edition*. Thomson Brooks/Cole, 6 edition.

Sutton-Tyrrell, K., Selzer, F., Sowers, MaryFran, R. (Mary Frances Roy), Neer, R., Powell, L., Gold, E. B., Greendale, G., Weiss, G., Matthews, K. A., and McKinlay, S. (2010). Study of women's health across the nation (swan): Baseline dataset, [united states], 1996-1997.

Appendix A

Terms

The following is the list of terms used throughout the dissertation organized by submodel, starting with survival, then longitudinal, and ending with cluster specific terms. The list is broken into terms used for the group and at the individual level.

A.1 Survival Model

This is the list of the terms used at the group level:

$\boldsymbol{T}^l$ Vector of left survival time points.

$\boldsymbol{T}^r$ Vector of right survival time points.

$\boldsymbol{L}$ Vector of left truncation time points.

$\boldsymbol{R}$ Vector of right truncation time points.

$\boldsymbol{\delta}_O$ Vector of indicators for which time points are observed.

$\boldsymbol{\delta}_R$ Vector of indicators for which time points are right censored.

$\boldsymbol{\delta}_L$ Vector of indicators for which time points are left censored.

$\boldsymbol{\delta}_I$ Vector of indicators for which time points are interval censored.

α_0 Intercept term for baseline hazard function.

α_1 Coefficient for κ_0.

$\boldsymbol{\alpha}$ Vector of α_0 and α_1.

$\boldsymbol{b}$ Vector of baseline hazard spline coefficients.

$\boldsymbol{\gamma}$ Matrix or vector of coefficients for time-dependent covariates.

$\boldsymbol{X}$ Matrix of longitudinal covariates.

$\boldsymbol{\zeta}$ Matrix or vector of coefficient for baseline covariates.

$\boldsymbol{Z}$ Matrix of baseline covariates.

$\boldsymbol{\theta}$ Vector of parameters for survival model.

σ_b^2 Variance of baseline hazard spline coefficients.

$\lambda_0(t)$ Baseline hazard function at generic time point, t.

$\lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i)$ Hazard function at generic time point, t.

$\Lambda(t|\boldsymbol{X}_i, \boldsymbol{Z}_i)$ Cumulative hazard function generic time point, t.

ℓ_0 The survival loglikelihood without penalty term.

ℓ_p The survival penalized loglikelihood.

This is the list of the terms used at the individual level:

T_i^l The left survival time point for individual i.

T_i^r The right survival time point for individual i.

L_i The left truncation time point for individual i.

R_i The right truncation time point for individual i.

δ_{O_i} The indicator for individual i for whether or not the time point is observed.

δ_{R_i} The indicator for individual i for whether or not the time point is right censored.

δ_{L_i} The indicator for individual i for whether or not the time point is left censored.

δ_{I_i} The indicator for individual i for whether or not the time point is interval censored.

$\boldsymbol{Z}_i$ Vector of baseline covariates for individual i.

$\boldsymbol{X}_i$ Vector of longitudinal covariates for individual i for all longitudinal trajectories given a specific time.

$\lambda_0 \left(T_i^r\right)$ Baseline hazard function at time point T_i^r.

$\lambda \left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ Hazard function at time point T_i^l.

$\lambda \left(T_i^r|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ Hazard function at time point T_i^r.

$\Lambda \left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ Cumulative hazard function at time point T_i^l.

$\Lambda \left(T_i^r|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ Cumulative hazard function at time point T_i^r.

$\Lambda \left(L_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ Cumulative hazard function at time point L_i.

$\Lambda \left(R_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ Cumulative hazard function at time point R_i.

A.2 Longitudinal Model

This is the list of terms used at the group level not already specified in the survival model above:

$\boldsymbol{Y}$ Matrix of longitudinal measures.

$\boldsymbol{\sigma}_Y^2$ Matrix of error variances for longitudinal trajectories.

$\sigma^2_{Y_j}$ Error variance for longitudinal trajectory j.

$\boldsymbol{\beta}_j$ The $\overline{J}$ dimensional vector of mean trajectory coefficients for the j^{th} longitudinal covariate.

$\beta_{jk'}$ The k' element of the $\overline{J}$ dimensional vector of mean trajectory coefficients for the j^{th} longitudinal covariate.

$\boldsymbol{\xi}$ Matrix of all individual longitudinal trajectories.

$\Sigma_{\boldsymbol{\xi}_j}$ Covariance-variance matrix for individual longitudinal covariate j.

This is the list of terms used at the individual level level not already specified in the survival model above:

$\boldsymbol{Y}_i$ Matrix or vector of observed longitudinal values for individual i.

$\boldsymbol{Y}_{ij}$ Vector of the longitudinal measurements for individual i and longitudinal trajectory j.

$\boldsymbol{X}_{ij}$ Vector of longitudinal covariates for individual i for longitudinal trajectory j.

$\boldsymbol{\xi}_i$ Matrix of individual longitudinal trajectories for individual i.

$\boldsymbol{\xi}_{ij}$ Vector of individual longitudinal trajectories for individual i for longitudinal trajectory j.

$\overline{\boldsymbol{B}}_{jk'}$ The m_{ij} dimensional vector of k' basis knot values for the j^{th} mean trajectory for individual i at times t_{ijl}, $l = 1, \ldots, m_{ij}$.

$\overline{B}_{jk'}$ The k' element of the basis knots for the j^{th} mean trajectory for individual i at time, t_{ijl}.

A.3 Clustering Model

This is the list of terms used at the group level in the clustering model:

$\boldsymbol{V}$ Matrix of indicators that is $n \times C$, where each row is for an individual telling which cluster, c, the individual belongs to.

$\boldsymbol{\tau}$ Matrix of probabilities that is $n \times C$, where each row is the individual's probabilities of being in each cluster.

$\boldsymbol{\tau}_c$ Vector of length n of probabilities for cluster c.

$\boldsymbol{\omega}$ Vector of weights for each cluster c.

ω_c Weight for cluster c.

$\boldsymbol{\gamma}_c$ Vector of coefficients for time-dependent covariates for cluster c.

$\boldsymbol{\zeta}_c$ Vector of coefficients for baseline covariates for cluster c.

$\ell_{0,c}$ The survival loglikelihood without penalty term for cluster c.

$\ell_{p,c}$ The survival penalized loglikelihood for cluster c.

$\boldsymbol{\sigma}^2_{Yc}$ Vector of variances for longitudinal covariates for cluster c.

$\Sigma_{\boldsymbol{\xi}_j c}$ Covariance-variance matrix for individual longitudinal covariate j for cluster c.

$\boldsymbol{\beta}_{jc}$ The $\overline{J}$ dimensional vector of mean trajectory coefficients for the j^{th} longitudinal covariate for cluster c.

$\beta_{jk'c}$ The k' element of the $\overline{J}$ dimensional vector of mean trajectory coefficients for the j^{th} longitudinal covariate for cluster c.

Θ Vector of all parameters.

This is the list of terms used at the individual level in the clustering model:

v_{ic} Indicator stating if individual i belong to cluster c. Used in the complete data likelihood, and the value is either 0 or 1.

$\boldsymbol{\tau}_i$ Vector of length C of an individual's probabilities of being in each cluster.

τ_{ic} Probability that individual i belongs to cluster c.

$\boldsymbol{X}_{ic}$ Vector of longitudinal covariates for individual i for all longitudinal trajectories given a specific time for cluster c.

$\boldsymbol{X}_{ijc}$ Vector of longitudinal covariates for individual i for longitudinal trajectory j for cluster c.

$\lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)$ Hazard function at time point T_i^l for cluster c.

$\lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)$ Hazard function at time point T_i^r for cluster c.

$\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)$ Cumulative hazard function at time point T_i^l for cluster c.

$\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)$ Cumulative hazard function at time point T_i^r for cluster c.

$\Lambda_c\left(L_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)$ Cumulative hazard function at time point L_i for cluster c.

$\Lambda_c\left(R_i|\boldsymbol{X}_{ic},\boldsymbol{Z}_i\right)$ Cumulative hazard function at time point R_i for cluster c.

$\sigma^2_{Y_jc}$ Variance for longitudinal covariate j for cluster c.

$\boldsymbol{\xi}_{ijc}$ Vector of individual longitudinal trajectories for individual i for longitudinal trajectory j for cluster c.

Appendix B

Derivation of Loglikelihood

We start with the fact that:

$$\ell\left(\left(\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T\right)^T; \sigma_b^2\right) = -\frac{K}{2}\log\sigma_b^2 + \int \ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right) d\boldsymbol{b},$$

where $\ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right) = \ell_0\left(\boldsymbol{\theta}; \sigma_b^2\right) - \frac{\boldsymbol{b}^T\boldsymbol{b}}{2\sigma_b^2}$ and a brief background of what penalized quasilikelihood (PQL) is. Penalized quasilikelihood (PQL), as stated in Section 10.8.2 of Ruppert et al. (2003) with full derivation of PQL given in Section 10.10.4 of Ruppert et al. (2003), is used in mixed models framework. PQL estimates of the parameters are achieved by treating the random effects as fixed parameters and penalizing the likelihood according to the distribution of the random effects. Therefore, in our framework, since the $\boldsymbol{b}$'s are assumed to be distributed $\boldsymbol{b} \sim N_K\left(\mathbf{0}, \sigma_b^2\boldsymbol{I}_K\right)$, the PQL approximation for the penalized loglikelihood is:

$$\int \ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right) d\boldsymbol{b} \simeq \ell_p\left(\widehat{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T\right); \sigma_b^2\right),$$

where $\widehat{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T\right) = \left(\boldsymbol{\alpha}^T, \widehat{\boldsymbol{b}^T}, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T\right)^T$ and $\widehat{\boldsymbol{b}} = \underset{\boldsymbol{b}}{\operatorname{argmax}}\, \ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right)$. Thus, finding the maximum likelihood estimates of $\ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right)$ with respect to $\boldsymbol{\theta}$ leads to maximizing $\ell\left(\left(\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T\right)^T; \sigma_b^2\right)$ with respect to $\left(\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T\right)^T$. Therefore, we can now see that under our given mixed model:

$$\ell\left(\left(\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T\right)^T; \sigma_b^2\right) = -\frac{K}{2}\log\sigma_b^2 + \int \ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right) d\boldsymbol{b}$$

can be approximated using the PQL approach giving:

$$\ell\left(\widehat{\boldsymbol{\theta}}; \sigma_b^2\right) \simeq -\frac{K}{2}\log\sigma_b^2 + \ell_p\left(\widehat{\boldsymbol{\theta}}; \sigma_b^2\right).$$

Appendix C

1^{st} and 2^{nd} Derivatives of $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ and the Loglikelihood

C.1 $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ & 1^{st} Derivatives of $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$

We first decided to divide up the space by knot segments to integrate over $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$. Therefore,

$$\begin{aligned}
\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) &= \sum_{k=0}^{k_t^*-1} \int_{\kappa_k}^{\kappa_{k+1}} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}+\boldsymbol{Z}_i^T\boldsymbol{\zeta}} du \\
&\quad + \int_{\kappa_{k_t^*}}^{t} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}+\boldsymbol{Z}_i^T\boldsymbol{\zeta}} du \\
&= e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} \sum_{k=0}^{k_t^*-1} \int_{\kappa_k}^{\kappa_{k+1}} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}} du \\
&\quad + e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} \int_{\kappa_{k_t^*}}^{t} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^{K} b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}} du,
\end{aligned}$$

where $\kappa_0 = 0$ and $k_t^* = \max\left(k : \kappa_k < t, 1 \leq k \leq K\right)$.

The first derivatives are:

$$\frac{\partial}{\partial \alpha_0}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)=\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)$$

$$\frac{\partial}{\partial \alpha_1}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)=e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} u e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du + e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t} u e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du,$$

$$\frac{\partial}{\partial b_j}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)=e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}I\left(t>\kappa_j\right)\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} \left(u-\kappa_j\right)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du + e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}I\left(t>\kappa_j\right)\int_{\kappa_{k_t^*}}^{t} \left(u-\kappa_j\right)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du,$$

$$\frac{\partial}{\partial \gamma_j}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)=e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} X_{ij} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du + e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t} X_{ij} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du,$$

$$\frac{\partial}{\partial \zeta_j}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)=Z_{ij}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du + Z_{ij}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t} e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l (u-\kappa_l)_+ + \boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du,$$

where b_j is the j^{th} element of the coefficient vector for the baseline knots where $j=1,\ldots,K$, γ_j is the j^{th} element of the coefficient vector for the trajectory functions where $j=1,\ldots,p$, $X_{ij}\left(t\right)$ is the j^{th} trajectory function for the i^{th} individual at time t where $j=1,\ldots,p$, ζ_j is the j^{th} element of the covariate coefficient vector where $j=1,\ldots,q$, and Z_{ij} is the j^{th} element of the covariate vector for the i^{th} individual where $j=1,\ldots,q$. Since the integrals do not have analytic solutions when the degree of $\boldsymbol{X}_i^T\left(t\right)$ is greater than one, we use Simpson's Rule to integrate the derivatives.

C.2 1^{st} Derivatives of the Loglikelihood

Using the results from Section C.1, we find that for $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)=\ell_0-\frac{\boldsymbol{b}^T\boldsymbol{b}}{2\sigma_b^2}$, where $\ell_0=\sum_{i=1}^{n}\left[\eta_0\left(T_i^r\right)\delta_{O_i}+\left(\boldsymbol{X}_i^T\left(T_i^r\right)\boldsymbol{\gamma}+\boldsymbol{Z}_i^T\boldsymbol{\zeta}\right)\delta_{O_i}-\right.$ $\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\log\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\delta_{L_i}-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}+$ $\log\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\delta_{I_i}+\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-$ $\left.\log\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\right]$, the first derivatives are:

$$\begin{aligned}
\frac{\partial}{\partial\alpha_0}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)&=\frac{\partial}{\partial\alpha_0}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right)\\
&=\sum_{i=1}^{n}\Bigg[\delta_{O_i}-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)\\
&\qquad+\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\delta_{L_i}-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\qquad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\delta_{I_i}\\
&\qquad+\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\\
&\qquad+\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\Bigg]\\
&=\sum_{i=1}^{n}\Bigg[\delta_{O_i}-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\qquad-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}-\frac{\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{I_i}\\
&\qquad+\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\Bigg]
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial}{\partial \alpha_1} \ell_p \left(\boldsymbol{\theta}; \sigma_b^2\right) &= \frac{\partial}{\partial \alpha_1} \ell_0 \left(\boldsymbol{\theta}; \sigma_b^2\right) \\
&= \sum_{i=1}^{n} \Bigg[t\delta_{O_i} - \frac{\partial}{\partial \alpha_1} \Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \left(\delta_{O_i} + \delta_{R_i}\right) \\
&\qquad + \frac{e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \frac{\partial}{\partial \alpha_1} \Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{1 - e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}} \delta_{L_i} - \frac{\partial}{\partial \alpha_1} \Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} \\
&\qquad - \frac{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial \alpha_1} \Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \alpha_1} \Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{1 - e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}} \delta_{I_i} \\
&\qquad + \frac{\partial}{\partial \alpha_1} \Lambda \left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \\
&\qquad + \frac{e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial \alpha_1} \Lambda \left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \alpha_1} \Lambda \left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{1 - e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}} \Bigg] \\
&= \sum_{i=1}^{n} \Bigg[t\delta_{O_i} - \frac{\partial}{\partial \alpha_1} \Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \left(\delta_{O_i} + \delta_{R_i}\right) + \frac{\frac{\partial}{\partial \alpha_1} \Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{L_i} \\
&\qquad - \frac{\partial}{\partial \alpha_1} \Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} - \frac{\left(\frac{\partial}{\partial \alpha_1} \Lambda \left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \alpha_1} \Lambda \left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{I_i} \\
&\qquad + \frac{\partial}{\partial \alpha_1} \Lambda \left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial}{\partial \alpha_1} \Lambda \left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \alpha_1} \Lambda \left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \Bigg]
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial}{\partial b_j}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial}{\partial b_j}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) - \frac{b_j}{\sigma_b^2} \\
&= \sum_{i=1}^{n}\Bigg[(t-\kappa_j)_+\,\delta_{O_i} - \frac{\partial}{\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) \\
&\quad + \frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\delta_{L_i} - \frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\delta_{I_i} \\
&\quad + \frac{\partial}{\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) \\
&\quad + \frac{e^{\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}\left(\frac{\partial}{\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{1-e^{\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}}\Bigg] - \frac{b_j}{\sigma_b^2} \\
&= \sum_{i=1}^{n}\Bigg[(t-\kappa_j)_+\,\delta_{O_i} - \frac{\partial}{\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) + \frac{\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad - \frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} - \frac{\left(\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{I_i} \\
&\quad + \frac{\partial}{\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial}{\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}-1}\Bigg] - \frac{b_j}{\sigma_b^2}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial}{\partial \gamma_j} \ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right) &= \frac{\partial}{\partial \gamma_j} \ell_0\left(\boldsymbol{\theta}; \sigma_b^2\right) \\
&= \sum_{i=1}^{n} \left[X_{ij}\left(T_i^r\right) \delta_{O_i} - \frac{\partial}{\partial \gamma_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left(\delta_{O_i} + \delta_{R_i}\right) \right. \\
&\quad + \frac{e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \frac{\partial}{\partial \gamma_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{1 - e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}} \delta_{L_i} - \frac{\partial}{\partial \gamma_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \gamma_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \gamma_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{1 - e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}} \delta_{I_i} \\
&\quad + \frac{\partial}{\partial \gamma_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \\
&\quad \left. + \frac{e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \gamma_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \gamma_j} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{1 - e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}} \right] \\
&= \sum_{i=1}^{n} \left[X_{ij}\left(T_i^r\right) \delta_{O_i} - \frac{\partial}{\partial \gamma_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left(\delta_{O_i} + \delta_{R_i}\right) + \frac{\frac{\partial}{\partial \gamma_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{L_i} \right. \\
&\quad - \frac{\partial}{\partial \gamma_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} - \frac{\left(\frac{\partial}{\partial \gamma_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \gamma_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{I_i} \\
&\quad \left. + \frac{\partial}{\partial \gamma_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial}{\partial \gamma_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \gamma_j} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \zeta_j}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial}{\partial \zeta_j}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\left[Z_{ij}\delta_{O_i} - \frac{\partial}{\partial \zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)\right. \\
&\qquad + \frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial \zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\delta_{L_i} - \frac{\partial}{\partial \zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\qquad - \frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\delta_{I_i} \\
&\qquad + \frac{\partial}{\partial \zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) \\
&\qquad \left. + \frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\right] \\
&= \sum_{i=1}^{n}\left[Z_{ij}\delta_{O_i} - \frac{\partial}{\partial \zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) + \frac{\frac{\partial}{\partial \zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\right. \\
&\qquad - \frac{\partial}{\partial \zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} - \frac{\left(\frac{\partial}{\partial \zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{I_i} \\
&\qquad \left. + \frac{\partial}{\partial \zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial}{\partial \zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\right],
\end{aligned}
$$

where b_j is the j^{th} element of the coefficient vector for the baseline knots where $j = 1, \ldots, K$, γ_j is the j^{th} element of the coefficient vector for the trajectory functions where $j = 1, \ldots, p$, $X_{ij}(t)$ is the j^{th} trajectory function for the i^{th} individual at time t where $j = 1, \ldots, p$, ζ_j is the j^{th} element of the covariate coefficient vector where $j = 1, \ldots, q$, and Z_{ij} is the j^{th} element of the covariate vector for the i^{th} individual where $j = 1, \ldots, q$. The final equation for each derivative is better computationally and comes from multiplying by a form of 1 on certain terms. For example, for the left censor term we multiplied by $\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}$.

C.3 2^{nd} Derivatives of $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$

The second derivatives are:

$$\begin{aligned}
\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) &= \frac{\partial}{\partial \alpha_0}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) = \Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\\
\frac{\partial^2}{\partial \alpha_0 \partial \alpha_1}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) &= \frac{\partial}{\partial \alpha_1}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\\
\frac{\partial^2}{\partial \alpha_0 \partial b_j}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) &= \frac{\partial}{\partial b_j}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\\
\frac{\partial^2}{\partial \alpha_0 \partial \gamma_j}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) &= \frac{\partial}{\partial \gamma_j}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\\
\frac{\partial^2}{\partial \alpha_0 \partial \zeta_j}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) &= \frac{\partial}{\partial \zeta_j}\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)
\end{aligned}$$

$$\frac{\partial^2}{\partial \alpha_1^2}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) = e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} \sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} u^2 e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$+ e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t} u^2 e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du,$$

$$\frac{\partial^2}{\partial \alpha_1 \partial b_j}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) = e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} \sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} u\left(u-\kappa_j\right)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$+ e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t} u\left(u-\kappa_j\right)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du,$$

$$\frac{\partial^2}{\partial \alpha_1 \partial \gamma_j}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) = e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} \sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} uX_{ij}\left(u\right) e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$+ e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t} uX_{ij}\left(u\right) e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du,$$

$$\frac{\partial^2}{\partial \alpha_1 \partial \zeta_j}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) = Z_{ij}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} \sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} u e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$+ Z_{ij}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t} u e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$\frac{\partial^2}{\partial b_j \partial b_{j'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) = e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} I\left(t>\kappa_j\right) \sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} \left(u-\kappa_j\right)_+ \left(u-\kappa_{j'}\right)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$+ e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} I\left(t>\kappa_j\right)\int_{\kappa_{k_t^*}}^{t} \left(u-\kappa_j\right)_+ \left(u-\kappa_{j'}\right)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du,$$

$$\frac{\partial^2}{\partial b_j \partial \gamma_{j'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) = e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} I\left(t>\kappa_j\right) \sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} \left(u-\kappa_j\right)_+ X_{ij'}\left(u\right) e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$+ e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} I\left(t>\kappa_j\right)\int_{\kappa_{k_t^*}}^{t} \left(u-\kappa_j\right)_+ X_{ij'}\left(u\right) e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du,$$

$$\frac{\partial^2}{\partial b_j \partial \zeta_{j'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) = Z_{ij'}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} I\left(t>\kappa_j\right) \sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}} \left(u-\kappa_j\right)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$+ Z_{ij'}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}} I\left(t>\kappa_j\right)\int_{\kappa_{k_t^*}}^{t} \left(u-\kappa_j\right)_+ e^{\alpha_0+\alpha_1 u+\sum_{l=1}^K b_l(u-\kappa_l)_+ +\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)=e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}}X_{ij}\left(u\right)X_{ij'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^{K}b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$+e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t}X_{ij}\left(u\right)X_{ij'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^{K}b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du,$$

$$\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)=Z_{ij'}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}}X_{ij}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^{K}b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$+Z_{ij'}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t}X_{ij}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^{K}b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)=Z_{ij}Z_{ij'}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}}e^{\alpha_0+\alpha_1u+\sum_{l=1}^{K}b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du$$

$$+Z_{ij}Z_{ij'}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t}e^{\alpha_0+\alpha_1u+\sum_{l=1}^{K}b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du.$$

Now, b_j is the j^{th} element of the coefficient vector for the baseline knots where $j=1,\ldots,K$, γ_j is the j^{th} element of the coefficient vector for the trajectory functions where $j=1,\ldots,p$, $X_{ij}\left(t\right)$ is the j^{th} trajectory function for the i^{th} individual at time t where $j=1,\ldots,p$, ζ_j is the j^{th} element of the covariate coefficient vector where $j=1,\ldots,q$, and Z_{ij} is the j^{th} element of the covariate vector for the i^{th} individual where $j=1,\ldots,q$. Since again the integrals do not have analytic solutions when the degree of $\boldsymbol{X}_i^T\left(t\right)$ is greater than one, we use Simpson's Rule to integrate the derivatives.

C.4 2^{nd} Derivatives of the Loglikelihood

Using the results from Sections C.1 and C.3, we find that for $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)=\ell_0-\frac{\boldsymbol{b}^T\boldsymbol{b}}{2\sigma_b^2}$, where $\ell_0=\sum_{i=1}^n\Big[\eta_0\left(T_i^r\right)\delta_{O_i}+\left(\boldsymbol{X}_i^T\left(T_i^r\right)\boldsymbol{\gamma}+\boldsymbol{Z}_i^T\boldsymbol{\zeta}\right)\delta_{O_i}-$ $\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\log\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\delta_{L_i}-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}+$ $\log\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\delta_{I_i}+\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-$ $\log\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\Big]$, the second derivatives are as follows:

$$
\begin{aligned}
\frac{\partial^2}{\partial \alpha_0^2}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial \alpha_0^2}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) \\
&\quad +\frac{\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}-\frac{\left(\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\Bigg] \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) \\
&\quad +\frac{\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}-\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}-\frac{\left(\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial \alpha_0^2}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial \alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad -\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg] \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad -\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad -\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha_0\partial b_j}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\alpha_0\partial b_j}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right)\\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]\\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i} \\
&\quad-\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad-\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad\times\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad+\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad+\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad\times\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg] \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad-\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad-\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad\times\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad+\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad+\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad\times\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad -\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg] \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad -\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad -\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha_1^2}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\alpha_1^2}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\left[-\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)\right. \\
&\quad +\frac{\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}-\frac{\left(\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}-1} \\
&\quad \left.+\frac{e^{\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(e^{\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}-1\right)^2}\right] \\
&= \sum_{i=1}^{n}\left[-\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)\right. \\
&\quad +\frac{\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}-\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}-\frac{\left(\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1^2}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}-1} \\
&\quad \left.+\frac{e^{\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)^2}{\left(1-e^{\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}\right)^2}\right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha_1\partial b_j}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\alpha_1\partial b_j}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right)\\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]\\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad -\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg] \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad -\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad -\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right)\\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]\\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial b_j \partial b_{j'}} \ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right) &= \frac{\partial^2}{\partial b_j \partial b_{j'}} \ell_0\left(\boldsymbol{\theta}; \sigma_b^2\right) - \frac{1}{\sigma_b^2} I\left(j = j'\right) \\
&= \sum_{i=1}^{n} \Bigg[-\frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left(\delta_{O_i} + \delta_{R_i}\right) + \frac{\frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{L_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \frac{\partial}{\partial b_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2} \delta_{L_i} \\
&\quad - \frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} \\
&\quad - \frac{\left(\frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - 1}} \delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial b_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \delta_{I_i} \\
&\quad + \frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \\
&\quad + \frac{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial b_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)\Bigg] - \frac{1}{\sigma_b^2} I\left(j = j'\right) \\
&= \sum_{i=1}^{n} \Bigg[-\frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left(\delta_{O_i} + \delta_{R_i}\right) + \frac{\frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{L_i} \\
&\quad - \frac{e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \frac{\partial}{\partial b_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{\left(1 - e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2} \delta_{L_i} \\
&\quad - \frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} \\
&\quad - \frac{\left(\frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - 1}} \delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(1 - e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial b_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \delta_{I_i} \\
&\quad + \frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial b_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \\
&\quad + \frac{e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(1 - e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial b_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)\Bigg] - \frac{1}{\sigma_b^2} I\left(j = j'\right)
\end{aligned}
$$

$$\frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \ell_p \left(\boldsymbol{\theta}; \sigma_b^2\right) = \frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \ell_0 \left(\boldsymbol{\theta}; \sigma_b^2\right)$$

$$= \sum_{i=1}^n \Bigg[-\frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left(\delta_{O_i} + \delta_{R_i}\right) + \frac{\frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{L_i}$$

$$- \frac{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \frac{\partial}{\partial \gamma_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2} \delta_{L_i}$$

$$- \frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i}$$

$$- \frac{\left(\frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - 1}} \delta_{I_i}$$

$$- \frac{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2}$$

$$\times \left(\frac{\partial}{\partial \gamma_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \gamma_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \delta_{I_i}$$

$$+ \frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1}$$

$$+ \frac{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2}$$

$$\times \left(\frac{\partial}{\partial \gamma_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \gamma_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \Bigg]$$

$$= \sum_{i=1}^n \Bigg[-\frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left(\delta_{O_i} + \delta_{R_i}\right) + \frac{\frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{L_i}$$

$$- \frac{e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \frac{\partial}{\partial \gamma_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{\left(1 - e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2} \delta_{L_i}$$

$$- \frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i}$$

$$- \frac{\left(\frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - 1}} \delta_{I_i}$$

$$- \frac{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(1 - e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2}$$

$$\times \left(\frac{\partial}{\partial \gamma_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \gamma_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \delta_{I_i}$$

$$+ \frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial \gamma_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1}$$

$$+ \frac{e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(1 - e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2}$$

$$\times \left(\frac{\partial}{\partial \gamma_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \gamma_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \Bigg]$$

$$
\begin{aligned}
\frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right) &= \frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \ell_0\left(\boldsymbol{\theta}; \sigma_b^2\right) \\
&= \sum_{i=1}^n \Bigg[-\frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left(\delta_{O_i} + \delta_{R_i}\right) + \frac{\frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{L_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \frac{\partial}{\partial \zeta_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2} \delta_{L_i} \\
&\quad - \frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} \\
&\quad - \frac{\left(\frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - 1}} \delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial \zeta_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \zeta_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \delta_{I_i} \\
&\quad + \frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \\
&\quad + \frac{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial \zeta_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \zeta_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \Bigg] \\
&= \sum_{i=1}^n \Bigg[-\frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left(\delta_{O_i} + \delta_{R_i}\right) + \frac{\frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{L_i} \\
&\quad - \frac{e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \frac{\partial}{\partial \zeta_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{\left(1 - e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2} \delta_{L_i} \\
&\quad - \frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} \\
&\quad - \frac{\left(\frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - 1}} \delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(1 - e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial \zeta_{j'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \zeta_{j'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \delta_{I_i} \\
&\quad + \frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial b_j \partial \zeta_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \\
&\quad + \frac{e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial b_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial b_j} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(1 - e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial \zeta_{j'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \zeta_{j'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right)\\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\gamma_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\gamma_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\gamma_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]\\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\gamma_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\gamma_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\gamma_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\gamma_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad -\frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg] \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad -\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad -\frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\zeta_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right)\\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]\\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\zeta_j\partial\zeta_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\zeta_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\zeta_{j'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg].
\end{aligned}
$$

The final equation for each derivative is better computationally and comes from multiplying by a form of 1 on certain terms. For example, for the second left censor term we multiplied by $\left(\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\right)^2$.

Appendix D

Derivation of Marginal Loglikelihood

D.1 Background and Short Derivation of Marginal Loglikelihood

In order to estimate the smoothing parameter, we use the restricted maximum likelihood (REML) since Harville (1974) showed that the REML for Gaussian models is the same as the marginal likelihood when the regression parameters are integrated with a flat prior. Thus, we find that the marginal loglikelihood is:

$$\ell_{marg}\left(\sigma_b^2\right) = -\frac{K}{2}\log\left(\sigma_b^2\right) + \log\int_{\mathbb{R}^{2+K+p+q}} \exp\left[\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)\right] d\boldsymbol{\theta}.$$

Since the likelihood of the function is given by:

$$
\begin{aligned}
L\left(\boldsymbol{\theta};\sigma_b^2\right) &= \exp\left\{\ell\left(\boldsymbol{\theta};\sigma_b^2\right)\right\} \\
&= \left(\frac{1}{\sigma_b^2}\right)^{\frac{K}{2}} \exp\left\{\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)\right\},
\end{aligned}
$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\alpha}^T, \boldsymbol{b}^T, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T\right)^T$ and $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^T$, plugging in we see that:

$$
\begin{aligned}
\ell_{marg}\left(\sigma_b^2\right) &= \log \int_{\mathbb{R}^{2+K+p+q}} \left(\frac{1}{\sigma_b^2}\right)^{\frac{K}{2}} \exp\left\{\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)\right\} d\boldsymbol{\alpha} d\boldsymbol{b} d\boldsymbol{\gamma} d\boldsymbol{\zeta} \\
&= -\frac{K}{2}\log\left(\sigma_b^2\right) + \log \int_{\mathbb{R}^{2+K+p+q}} \exp\left\{\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)\right\} d\boldsymbol{\theta}.
\end{aligned}
$$

We apply Laplace's method as stated in Laplace (1986), which states that:

$$
\int_a^b e^{Mf(x)}dx \approx \sqrt{\frac{2\pi}{M|f''(x)|}}e^{Mf(x_0)},
$$

where M is a constant, f'' is the 2^{nd} derivative of f, and x_0 is the maximum of f. Therefore, the approximation of $\ell_{marg}\left(\sigma_b^2\right)$ is:

$$
\ell_{marg}\left(\sigma_b^2\right) \simeq -\frac{K}{2}\log\left(\sigma_b^2\right) + \ell_p\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\} - \frac{1}{2}\log\left|-\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right|,
$$

where $\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right)$ is the solution to $\boldsymbol{Q}\left(\boldsymbol{\theta};\sigma_b^2\right) = \mathbf{0}$. Additionally, $\boldsymbol{Q}$ is the $(2+K+p+q) \times 1$ vector of first-order partial derivatives of $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)$ and $\mathbb{Q}$ is the $(2+K+p+q) \times (2+K+p+q)$ matrix of the second-order partial derivatives of $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)$.

D.2 Detailed Derivation of Marginal Loglikelihood

We start with the fact that $L\left(\boldsymbol{\theta};\sigma_b^2\right)=\left(\frac{1}{\sigma_b^2}\right)^{\frac{K}{2}}\exp\left\{\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)\right\}$, where $\boldsymbol{\theta}=\left(\boldsymbol{\alpha}^T,\boldsymbol{b}^T,\boldsymbol{\gamma}^T,\boldsymbol{\zeta}^T\right)^T$ with $\boldsymbol{\alpha}=(\alpha_0,\alpha_1)^T$. We also know that the Taylor series expansion around $\widehat{\boldsymbol{\theta}}$ for $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)\simeq\ell_p\left(\widehat{\boldsymbol{\theta}};\sigma_b^2\right)+\boldsymbol{Q}\left(\widehat{\boldsymbol{\theta}};\sigma_b^2\right)(\theta-\widehat{\theta})+\frac{1}{2}(\theta-\widehat{\theta})^T\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}(\theta-\widehat{\theta})$. Since $\widehat{\theta}$ is the solution to $\boldsymbol{Q}\left(\boldsymbol{\theta};\sigma_b^2\right)=\mathbf{0}$, the equation becomes $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)\simeq\ell_p\left(\widehat{\boldsymbol{\theta}};\sigma_b^2\right)+\frac{1}{2}(\theta-\widehat{\theta})^T\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}(\theta-\widehat{\theta})$. Now using a similar argument as the derivation of PQL, we find:

$$
\begin{aligned}
L_{marg}\left(\sigma_b^2\right)&=\int_{\mathbb{R}^{2+K+p+q}}\left(\frac{1}{\sigma_b^2}\right)^{\frac{K}{2}}\exp\left\{\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)\right\}d\boldsymbol{\theta}\\
&\simeq(2\pi)^{\frac{2+K+p+q}{2}}|\Sigma|^{\frac{1}{2}}\left(\frac{1}{\sigma_b^2}\right)^{\frac{K}{2}}\exp\left\{\ell_p\left(\widehat{\boldsymbol{\theta}};\sigma_b^2\right)\right\}\\
&\quad\cdot\int_{\mathbb{R}^{2+K+p+q}}(2\pi)^{-\frac{2+K+p+q}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left\{\frac{1}{2}(\theta-\widehat{\theta})^T\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}(\theta-\widehat{\theta})\right\}d\boldsymbol{\theta}\\
&=(2\pi)^{\frac{2+K+p+q}{2}}|\Sigma|^{\frac{1}{2}}\left(\frac{1}{\sigma_b^2}\right)^{\frac{K}{2}}\exp\left\{\ell_p\left(\widehat{\boldsymbol{\theta}};\sigma_b^2\right)\right\}\\
&=(2\pi)^{\frac{2+K+p+q}{2}}\left|-\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right|^{-\frac{1}{2}}\left(\frac{1}{\sigma_b^2}\right)^{\frac{K}{2}}\exp\left\{\ell_p\left(\widehat{\boldsymbol{\theta}};\sigma_b^2\right)\right\}.
\end{aligned}
$$

The second step uses the Taylor series approximation. The third step uses that fact the integral of a $N\left(\widehat{\theta},\Sigma\right)$ over the whole space is 1 and that the Hessian matrix $\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}=-\Sigma^{-1}$ or $\Sigma=-\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}^{-1}$. Taking the log of both sides and removing the constant we get:

$$
\ell_{marg}\left(\sigma_b^2\right)\simeq-\frac{K}{2}\log\left(\sigma_b^2\right)+\ell_p\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}-\frac{1}{2}\log\left|-\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right|.
$$

Appendix E

Derivations of 1^{st} and 2^{nd} Derivatives of Marginal Loglikelihood

E.1 Derivation of 1^{st} Derivative of Marginal Loglikelihood

We know that

$$\ell_{marg}\left(\sigma_b^2\right) \simeq -\frac{K}{2}\log\left(\sigma_b^2\right) + \ell_p\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\} - \frac{1}{2}\log\left|-\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right|.$$

Additionally, $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) = \ell_0 - \frac{\boldsymbol{b}^T\boldsymbol{b}}{2\sigma_b^2}$ and ℓ_0 does not have any dependence on σ_b^2. Lastly, Jacobi's formula states that $\frac{d}{dx}\left(\det A\left(x\right)\right) = \det A\left(x\right) \cdot tr\left(A^{-1}\left(x\right)\frac{d}{dx}A\left(x\right)\right)$. Therefore, the first derivative is found as follows:

$$\begin{aligned}\frac{d}{d\sigma_b^2}\ell_{marg}\left(\sigma_b^2\right) &\simeq -\frac{K}{2\sigma_b^2}+\frac{1}{2\left(\sigma_b^2\right)^2}\boldsymbol{b}^T\boldsymbol{b}-\frac{1}{2\left|-\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right|}\cdot\left|-\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right| \\ &\quad\cdot tr\left(-\mathbb{Q}^{-1}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\frac{d}{d\sigma_b^2}\left(-\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right)\right) \\ &= -\frac{K}{2\sigma_b^2}+\frac{1}{2\left(\sigma_b^2\right)^2}\boldsymbol{b}^T\boldsymbol{b}-\frac{1}{2}tr\left(\mathbb{Q}^{-1}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\frac{d}{d\sigma_b^2}\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right).\end{aligned}$$

E.2 Derivation of 2^{nd} Derivative of Marginal Loglikelihood

Before we find the second derivative, we first state that the derivative of the trace of a matrix is the same as the trace of the derivative of the matrix, thus $\frac{d}{dx}tr\left(A\right)=tr\left(\frac{d}{dx}A\right)$. Secondly, we derive that $\frac{d}{dx}A^{-1}=-A^{-1}\left(\frac{d}{dx}A\right)A^{-1}$, where A is a square matrix whose inverse exists.

Derivation:

$$\begin{aligned}I &= AA^{-1} \\ \frac{d}{dx}I &= \frac{d}{dx}\left(AA^{-1}\right) \\ 0 &= \left(\frac{d}{dx}A\right)A^{-1}+A\frac{d}{dx}\left(A^{-1}\right) \\ A\frac{d}{dx}\left(A^{-1}\right) &= -\left(\frac{d}{dx}A\right)A^{-1} \\ \frac{d}{dx}\left(A^{-1}\right) &= -A^{-1}\left(\frac{d}{dx}A\right)A^{-1}.\end{aligned}$$

Now, the second derivative of the marginal loglikelihood, using the first derivative is:

$$\begin{aligned}
\frac{d^2}{d\left(\sigma_b^2\right)^2}\ell_{marg}\left(\sigma_b^2\right) \simeq & \frac{K}{2\left(\sigma_b^2\right)^2} - \frac{1}{\left(\sigma_b^2\right)^3}\boldsymbol{b}^T\boldsymbol{b} \\
& - \frac{1}{2}tr\left(\frac{d}{d\sigma_b^2}\left(\mathbb{Q}^{-1}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\frac{d}{d\sigma_b^2}\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right)\right) \\
= & \frac{K}{2\left(\sigma_b^2\right)^2} - \frac{1}{\left(\sigma_b^2\right)^3}\boldsymbol{b}^T\boldsymbol{b} \\
& + \frac{1}{2}tr\left(\mathbb{Q}^{-1}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\frac{d}{d\sigma_b^2}\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right. \\
& \left.\times\mathbb{Q}^{-1}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\frac{d}{d\sigma_b^2}\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right) \\
& - \frac{1}{2}tr\left(\mathbb{Q}^{-1}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\frac{d^2}{d\left(\sigma_b^2\right)^2}\mathbb{Q}\left\{\widehat{\boldsymbol{\theta}}\left(\sigma_b^2\right);\sigma_b^2\right\}\right).
\end{aligned}$$

Appendix F

Formulas and Derivations for Results

F.1 Formulas for Results

We use θ here as a general parameter we are estimating.

RBias Equal to $\frac{\overline{\theta}-\theta}{\theta}$, where $\overline{\theta} = \frac{\sum_{i=1}^{r} \widehat{\theta}_i}{r}$, θ is the true parameter value, and r is the number of replicates.

ECP Equal to $\frac{\sum_{i=1}^{r} I(U_i > \theta > L_i)}{r}$, where $U_i = \widehat{\theta}_i + z_{\alpha/2}\widehat{\sigma}_{\widehat{\theta}_i}$ and $L_i = \widehat{\theta}_i - z_{\alpha/2}\widehat{\sigma}_{\widehat{\theta}_i}$. Now $\widehat{\sigma}_{\widehat{\theta}_i}$ is the estimated standard error from each of the r runs and $z_{\alpha/2}$ is the confidence level, in our case $z_{\alpha/2} = 1.96$.

MESE Denoted by $\overline{\sigma}_{\widehat{\theta}}$ and equal to $\overline{\sigma}_{\widehat{\theta}} = \frac{\sum_{i=1}^{r} \widehat{\sigma}_{\widehat{\theta}_i}}{r}$.

ESE Denoted by $\widehat{\sigma}_{\widehat{\theta}}$ and equal to $\widehat{\sigma}_{\widehat{\theta}} = \sqrt{\frac{\sum_{i=1}^{r}\left(\widehat{\theta}_i - \overline{\theta}\right)^2}{r-1}}$.

Specification Rate Equal to $\frac{\sum_{i=1}^{n}\sum_{c=1}^{C} I(v_{ic}=1 \cap \tau_{ic}=\max(\boldsymbol{\tau}_i))}{n}$. The numerator states that in order to count towards the specification rate, an individual must have come from cluster c and be identified as belonging to cluster c.

F.2 Derivation of Standard Errors

In order to calculate empirical coverage probabilities (ECP) or mean estimated standard errors (MESE), we need to calculate the covariance-variance matrix for $\boldsymbol{\theta} = \left(\boldsymbol{\alpha}^T, \boldsymbol{b}^T, \boldsymbol{\gamma}^T, \boldsymbol{\zeta}^T\right)^T$ with $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^T$. The covariance-variance matrix of $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ is:

$$\operatorname{cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \\ \widehat{\boldsymbol{b}} - \boldsymbol{b} \\ \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \\ \widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta} \end{bmatrix} | \sigma_b^2 \right) = \operatorname{cov}\left(\begin{bmatrix} \widehat{\boldsymbol{\alpha}} \\ \widehat{\boldsymbol{b}} - \boldsymbol{b} \\ \widehat{\boldsymbol{\gamma}} \\ \widehat{\boldsymbol{\zeta}} \end{bmatrix} | \sigma_b^2 \right),$$

since only $\boldsymbol{b}$ depends on σ_b^2. Before finding the estimate for the covariance-variance matrix of $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$, we first present the Taylor series approximation for the first derivatives of the loglikelihood, or score function, denoted by $\boldsymbol{Q}(\boldsymbol{\theta})$ and centered at $\boldsymbol{\theta}$. It is:

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} l\left(\widehat{\boldsymbol{\theta}}\right) &= \boldsymbol{Q}\left(\widehat{\boldsymbol{\theta}}\right) = \mathbf{0} \\
\boldsymbol{Q}\left(\widehat{\boldsymbol{\theta}}\right) &\approx \boldsymbol{Q}(\boldsymbol{\theta}) + \mathbb{Q}(\boldsymbol{\theta})\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \\
\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) &= -\mathbb{Q}^{-1}(\boldsymbol{\theta})\, \boldsymbol{Q}(\boldsymbol{\theta}),
\end{aligned}
$$

where $\mathbb{Q}(\boldsymbol{\theta})$ is the second derivatives of the loglikelihood or Hessian matrix. The other piece we need to remember is that $\text{Var}(AX) = AXA^T$.

With this background, we find that the approximate covariance-variance matrix of $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ given σ_b^2 is:

$$
\begin{aligned}
\text{cov}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} | \sigma_b^2\right) &= \text{cov}\left(-\mathbb{Q}^{-1}(\boldsymbol{\theta})\, \boldsymbol{Q}(\boldsymbol{\theta})\, | \sigma_b^2\right) \\
&= \left\{-\mathbb{Q}\left(\boldsymbol{\theta}; \sigma_b^2\right)\right\}^{-1} \text{Var}\left\{\boldsymbol{Q}\left(\boldsymbol{\theta}; \sigma_b^2\right)\right\} \left\{-\mathbb{Q}\left(\boldsymbol{\theta}; \sigma_b^2\right)\right\}^{-1} \\
&= -\mathbb{Q}\left(\boldsymbol{\theta}; \sigma_b^2\right)^{-1},
\end{aligned}
$$

since for likelihood functions the $\text{Var}\left\{\boldsymbol{Q}\left(\widehat{\boldsymbol{\theta}}; \sigma_b^2\right)\right\} = -\mathbb{Q}\left(\widehat{\boldsymbol{\theta}}; \sigma_b^2\right)$. Now using the above results, the estimated covariance-variance matrix of $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ given σ_b^2 is:

$$
\widehat{\text{cov}}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} | \sigma_b^2\right) = -\mathbb{Q}\left(\widehat{\boldsymbol{\theta}}; \sigma_b^2\right)^{-1}.
$$

Appendix G

Derivations of $\boldsymbol{\tau}_i$ and $\boldsymbol{\omega}$ for EM Algorithm

G.1 Derivation of $\boldsymbol{\tau}_i$

Given that v_{ic} is unknown, we have to estimate it by its expected value, therefore, $\tau_{ic} = E\left[v_{ic}|T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{X}_i, \boldsymbol{Z}_i\right]$. Therefore,

$$\begin{aligned}\tau_{ic} &= E\left[v_{ic}|T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{X}_i, \boldsymbol{Z}_i\right] \\ &= 1 \cdot \Pr\left(i \in c\right) + 0 \cdot \Pr\left(i \notin c\right) \\ &= \Pr\left(i \in c\right).\end{aligned}$$

Now, the probability that the individual, i, belongs to cluster c is equal to the density value of cluster c of the individual weighted by the probability of the cluster times a normalizing constant. In our case, this is the sum of the probability that the individual belongs in each of the C clusters. This can be written as:

$$\widehat{\tau}_{ic} = \frac{\omega_c S_{ic}\left(\Theta\right)}{\sum_{c'=1}^{C} \omega_{c'} S_{ic'}\left(\Theta\right)},$$

where

$$\begin{aligned} S_c\left(\Theta\right) &= \prod_{i=1}^{n} S_{ic}\left(\Theta\right) = \prod_{i=1}^{n} S_c\left(\Theta; T_i^l, T_i^r, L_i, R_i, \delta_{O_i}, \delta_{R_i}, \delta_{L_i}, \delta_{I_i}, \boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right) \\ &= \prod_{i=1}^{n} \left(\frac{\lambda_c\left(T_i^r | \boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right) e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}}{e^{-\Lambda_c\left(L_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)} - e^{-\Lambda_c\left(R_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}} \right)^{\delta_{O_i}} \left(\frac{e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}}{e^{-\Lambda_c\left(L_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)} - e^{-\Lambda_c\left(R_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}} \right)^{\delta_{R_i}} \\ &\quad \times \left(\frac{1 - e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}}{e^{-\Lambda_c\left(L_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)} - e^{-\Lambda_c\left(R_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}} \right)^{\delta_{L_i}} \left(\frac{e^{-\Lambda_c\left(T_i^l|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)} - e^{-\Lambda_c\left(T_i^r|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}}{e^{-\Lambda_c\left(L_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)} - e^{-\Lambda_c\left(R_i|\boldsymbol{X}_{ic}, \boldsymbol{Z}_i\right)}} \right)^{\delta_{I_i}}. \end{aligned}$$

We let $\Theta = \left(\alpha_0, \alpha_1, \boldsymbol{b}, \boldsymbol{\gamma}_c, \boldsymbol{\zeta}_c, \omega_c\right)$ where $c = 1, \ldots, C$ be the estimates of the parameters of interest.

G.2 Derivation of $\boldsymbol{\omega}$

Solving for ω_c, we start with the fact that we have the constraint that $\sum_{c=1}^{C} \omega_c = 1$ and the expected value of the likelihood is $E\left(\ell\left(\Theta; \boldsymbol{T}^l, \boldsymbol{T}^r, \boldsymbol{L}, \boldsymbol{R}, \boldsymbol{\delta}_O, \boldsymbol{\delta}_R, \boldsymbol{\delta}_L, \boldsymbol{\delta}_I, \boldsymbol{Z}, \boldsymbol{V}\right)\right) = \sum_{i=1}^{n} \sum_{c=1}^{C} \left[\tau_{ic} \log\left(\omega_c\right) + \tau_{ic} \log\left(S_{ic}\left(\Theta\right)\right)\right]$. We now use a Lagrange multiplier, which is where we maximize a function subject to a contraint. In its general form, it is written as $L\left(x, \lambda\right) = f\left(x\right) + \lambda\left(K - g\left(x\right)\right)$, where $g\left(x\right) = K$ and K is a constant. More background can be found in section 14.8 of Stewart (2008). With this background, we now solve the

Lagrange multiplier with the constraint $\sum_{c=1}^{C} \omega_c = 1$, leading to:

$$\ell\left(\boldsymbol{\omega}, \lambda\right) = \sum_{i=1}^{n} \sum_{c=1}^{C} \left[\tau_{ic} \log\left(\omega_c\right) + \tau_{ic} \log\left(S_{ic}\left(\Theta\right)\right)\right] + \lambda \left(1 - \sum_{c=1}^{C} \omega_c\right).$$

Taking the derivative of $\ell\left(\boldsymbol{\omega}, \lambda\right)$ with respect to ω_c and setting it equal to 0, we get:

$$\begin{aligned}
\frac{\partial}{\partial \omega_c} \ell\left(\boldsymbol{\omega}, \lambda\right) &= \sum_{i=1}^{n} \frac{\tau_{ic}}{\omega_c} - \lambda \\
\lambda &= \sum_{i=1}^{n} \frac{\tau_{ic}}{\omega_c} \\
\omega_c &= \sum_{i=1}^{n} \frac{\tau_{ic}}{\lambda}.
\end{aligned}$$

Now, we need to find λ. Therefore, we use the fact that $\sum_{c=1}^{C} \tau_{ic} = 1$ and $\sum_{c=1}^{C} \omega_c = 1$, and find that:

$$\begin{aligned}
\sum_{c=1}^{C} \omega_c &= \sum_{i=1}^{n} \sum_{c=1}^{C} \frac{\tau_{ic}}{\lambda} \\
1 &= \frac{1}{\lambda} \sum_{i=1}^{n} \sum_{c=1}^{C} \tau_{ic} \\
\lambda &= \sum_{i=1}^{n} \sum_{c=1}^{C} \tau_{ic} \\
\lambda &= \sum_{i=1}^{n} 1 \\
\lambda &= n.
\end{aligned}$$

Therefore,

$$\widehat{\omega}_c = \frac{\sum_{i=1}^{n} \tau_{ic}}{\sum_{i=1}^{n} \sum_{c'=1}^{C} \tau_{ic'}} = \frac{\sum_{i=1}^{n} \tau_{ic}}{n}.$$

Appendix H

Derivations of $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ and Loglikelihood Involving $\beta_{jk'}$

Since every term from before that does not involve $\beta_{jk'}$ is solved for in the general setting in Section C of the Appendix, we do not present them here. Therefore, we will only be looking at terms involving $\beta_{jk'}$ when giving the first and second derivatives of $\Lambda\left(t|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)$ and the loglikelihood in this section. They will be presented as given in Section C of Appendix, without clustering or Monte Carlo replicates.

H.1 1^{st} and 2^{nd} Derivatives of $\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)$

The 1^{st} derivative $\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)$ with respect to $\beta_{jk'}$, $k'=1,\ldots,\overline{J}$ is:

$$\begin{aligned}\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)&=e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}}\gamma_j\overline{B}_{jk'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du\\&+e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t}\gamma_j\overline{B}_{jk'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du.\end{aligned}$$

The 2^{nd} derivatives of $\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)$ with respect to $\beta_{jk'}$, $k'=1,\ldots,\overline{J}$ is:

$$\begin{aligned}\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)&=\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\\
\frac{\partial}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)&=e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}}u\gamma_j\overline{B}_{jk'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du\\
&+e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t}u\gamma_j\overline{B}_{jk'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du\\
\frac{\partial}{\partial b_j\partial\beta_{j'k'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)&=e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}}\left(u-\kappa_j\right)_+\gamma_{j'}\overline{B}_{j'k'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du\\
&+e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t}\left(u-\kappa_j\right)_+\gamma_{j'}\overline{B}_{j'k'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du\\
\frac{\partial}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)&=e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}}\left(\psi_j\left(u\right)\gamma_{j'}+I\left(j=j'\right)\right)\overline{B}_{j'k'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du\\
&+e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t}\left(\psi_j\left(u\right)\gamma_{j'}+I\left(j=j'\right)\right)\overline{B}_{j'k'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du\\
\frac{\partial}{\partial\zeta_j\partial\beta_{j'k'}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)&=Z_{ij}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}}\gamma_{j'}\overline{B}_{j'k'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du\\
&+Z_{ij}e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t}\gamma_{j'}\overline{B}_{j'k'}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du\\
\frac{\partial}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(t|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)&=e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\sum_{k=0}^{k_t^*-1}\int_{\kappa_k}^{\kappa_{k+1}}\gamma_j\overline{B}_{jk'}\left(u\right)\gamma_{j'}\overline{B}_{j'k''}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du\\
&+e^{\boldsymbol{Z}_i^T\boldsymbol{\zeta}}\int_{\kappa_{k_t^*}}^{t}\gamma_j\overline{B}_{jk'}\left(u\right)\gamma_{j'}\overline{B}_{j'k''}\left(u\right)e^{\alpha_0+\alpha_1u+\sum_{l=1}^K b_l(u-\kappa_l)_++\boldsymbol{X}_i^T(u)\boldsymbol{\gamma}}du.\end{aligned}$$

H.2 1^{st} Derivative of Loglikelihood

The first derivative of the complete data loglikelihood with respect to $\beta_{jk'}$ is:

$$\frac{\partial}{\partial \beta_{jk'}} \log \left(L\left(\Theta\right)\right) = \sum_{i=1}^{n} \frac{\overline{\boldsymbol{B}}_{jk'}^{T} \left(\boldsymbol{Y}_{ij} - \boldsymbol{X}_{ij}\right)}{\sigma_{Y_j}^2} + \frac{\partial}{\partial \beta_{jk'}} \ell_p \left(\boldsymbol{\theta}; \sigma_b^2\right).$$

Now each element of $\boldsymbol{X}_{ij}$ equals $X_{ij}\left(t\right) = \psi_j\left(t\right) + \sum_{k'=1}^{J} \xi_{ijk'} B_{k'}\left(t\right)$, where $\psi_j\left(t\right)$ is the j^{th} mean trajectory function and is given by $\psi_j\left(t\right) = \sum_{k'=1}^{J} \beta_{jk'} \overline{B}_{k'}\left(t\right)$. Remembering that $\ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right) = \ell_0 - \frac{\boldsymbol{b}^T \boldsymbol{b}}{2\sigma_b^2}$, where $\ell_0 = \sum_{i=1}^{n} \Big[\eta_0\left(T_i^r\right) \delta_{O_i} + \left(\boldsymbol{X}_i^T \left(T_i^r\right) \boldsymbol{\gamma} + \boldsymbol{Z}_i^T \boldsymbol{\zeta}\right) \delta_{O_i} -$ $\Lambda\left(T_i^r|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left(\delta_{O_i} + \delta_{R_i}\right) + \log\left(1 - e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right) \delta_{L_i} - \Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} +$ $\log\left(1 - e^{\Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right) \delta_{I_i} + \Lambda\left(L_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \log\left(1 - e^{\Lambda\left(L_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)\Big]$, the first derivative of $\ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right)$ with respect to $\beta_{jk'}$ is:

$$\begin{aligned}
\frac{\partial}{\partial \beta_{jk'}} \ell_p\left(\boldsymbol{\theta}; \sigma_b^2\right) &= \frac{\partial}{\partial \beta_{jk'}} \ell_0\left(\boldsymbol{\theta}; \sigma_b^2\right) \\
&= \sum_{i=1}^{n} \Bigg[\left(\gamma_j \overline{B}_{jk'}\left(T_i^r\right)\right) \delta_{O_i} - \frac{\partial}{\partial \beta_{jk'}} \Lambda\left(T_i^r|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left(\delta_{O_i} + \delta_{R_i}\right) \\
&\quad + \frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \frac{\partial}{\partial \beta_{jk'}} \Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{1 - e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)}} \delta_{L_i} - \frac{\partial}{\partial \beta_{jk'}} \Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial \beta_{jk'}} \Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \beta_{jk'}} \Lambda\left(T_i^r|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{1 - e^{\Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)}} \delta_{I_i} \\
&\quad + \frac{\partial}{\partial \beta_{jk'}} \Lambda\left(L_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) \\
&\quad + \frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial \beta_{jk'}} \Lambda\left(L_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \beta_{jk'}} \Lambda\left(R_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{1 - e^{\Lambda\left(L_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)}}\Bigg] \\
&= \sum_{i=1}^{n} \Bigg[\left(\overline{B}_{jk'}\left(T_i^r\right) \gamma_j\right) \delta_{O_i} - \frac{\partial}{\partial \beta_{jk'}} \Lambda\left(T_i^r|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left(\delta_{O_i} + \delta_{R_i}\right) + \frac{\frac{\partial}{\partial \beta_{jk'}} \Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{L_i} \\
&\quad - \frac{\partial}{\partial \beta_{jk'}} \Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} - \frac{\left(\frac{\partial}{\partial \beta_{jk'}} \Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \beta_{jk'}} \Lambda\left(T_i^r|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{I_i} \\
&\quad + \frac{\partial}{\partial \beta_{jk'}} \Lambda\left(L_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial}{\partial \beta_{jk'}} \Lambda\left(L_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \beta_{jk'}} \Lambda\left(R_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1}\Bigg].
\end{aligned}$$

The final equation is better computationally and comes from multiplying by a form of 1 on certain terms. For example, for the left censor term we multiplied by $\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}$.

H.3 2^{nd} Derivative of Loglikelihood

Since every second derivative of $\beta_{jk'}$, except of $\beta_{jk'}$ with itself, only involves the penalized loglikelihood of the survival model, $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)$, we are only going to be at that term. Remembering that $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) = \ell_0 - \frac{\boldsymbol{b}^T\boldsymbol{b}}{2\sigma_b^2}$, where $\ell_0 = \sum_{i=1}^n [\eta_0\left(T_i^r\right)\delta_{O_i} +$ $\left(\boldsymbol{X}_i^T\left(T_i^r\right)\boldsymbol{\gamma} + \boldsymbol{Z}_i^T\boldsymbol{\zeta}\right)\delta_{O_i} - \Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) + \log\left(1 - e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\delta_{L_i} -$ $\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} + \log\left(1 - e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)\delta_{I_i} + \Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) -$

$\log\left(1-e^{\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}\right)\Big]$, the second derivatives are:

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i} \\
&\quad-\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad-\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad\times\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad+\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}-1} \\
&\quad+\frac{e^{\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}-1\right)^2} \\
&\quad\times\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg] \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad-\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad-\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad\times\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad+\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_0\partial\beta_{jk'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}-1} \\
&\quad+\frac{e^{\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}\left(\frac{\partial}{\partial\alpha_0}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_0}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)-\Lambda(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)}\right)^2} \\
&\quad\times\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right)\\
&= \sum_{i=1}^n\Bigg[-\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]\\
&= \sum_{i=1}^n\Bigg[-\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\alpha_1\partial\beta_{jk'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\alpha_1}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\alpha_1}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}} \ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial b_j \partial \beta_{j'k'}} \ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial \beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad -\frac{\left(\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial \beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial \beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \beta_{j'k'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg] \\
&= \sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i} \\
&\quad -\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial \beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad -\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad -\frac{\left(\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad -\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial b_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial \beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad +\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial b_j \partial \beta_{j'k'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad +\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial b_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial b_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial \beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial \beta_{j'k'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right) &= \frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right) \\
&= \sum_{i=1}^{n}\left[\overline{B}_{j'k'}\left(T_i^r\right)I\left(j=j'\right)\delta_{O_i} - \frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) + \frac{\frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\right. \\
&\quad - \frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i} \\
&\quad - \frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad - \frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad + \frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad + \frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2} \\
&\quad \left.\times\left(\frac{\partial}{\partial\beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{j'k'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\right] \\
&= \sum_{i=1}^{n}\left[\overline{B}_{j'k'}\left(T_i^r\right)I\left(j=j'\right)\delta_{O_i} - \frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right) + \frac{\frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\right. \\
&\quad - \frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i} \\
&\quad - \frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i} \\
&\quad - \frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times\left(\frac{\partial}{\partial\beta_{j'k'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{j'k'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i} \\
&\quad + \frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\gamma_j\partial\beta_{j'k'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1} \\
&\quad + \frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\gamma_j}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\gamma_j}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \left.\times\left(\frac{\partial}{\partial\beta_{j'k'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{j'k'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \ell_p \left(\boldsymbol{\theta}; \sigma_b^2\right) &= \frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \ell_0 \left(\boldsymbol{\theta}; \sigma_b^2\right) \\
&= \sum_{i=1}^{n} \Bigg[-\frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \left(\delta_{O_i} + \delta_{R_i}\right) + \frac{\frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{L_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \frac{\partial}{\partial \zeta_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \frac{\partial}{\partial \beta_{j'k'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2} \delta_{L_i} \\
&\quad - \frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} \\
&\quad - \frac{\left(\frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - 1}} \delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial \zeta_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \zeta_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial \beta_{j'k'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \beta_{j'k'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \delta_{I_i} \\
&\quad + \frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \\
&\quad + \frac{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial \zeta_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \zeta_j} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial \beta_{j'k'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \beta_{j'k'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \Bigg] \\
&= \sum_{i=1}^{n} \Bigg[-\frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \left(\delta_{O_i} + \delta_{R_i}\right) + \frac{\frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \delta_{L_i} \\
&\quad - \frac{e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \frac{\partial}{\partial \zeta_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \frac{\partial}{\partial \beta_{j'k'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}{\left(1 - e^{-\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2} \delta_{L_i} \\
&\quad - \frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) \delta_{I_i} \\
&\quad - \frac{\left(\frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - 1}} \delta_{I_i} \\
&\quad - \frac{e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial \zeta_j} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \zeta_j} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(1 - e^{\Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial \beta_{j'k'}} \Lambda\left(T_i^l | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \beta_{j'k'}} \Lambda\left(T_i^r | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \delta_{I_i} \\
&\quad + \frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) + \frac{\left(\frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial^2}{\partial \zeta_j \partial \beta_{j'k'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} - 1} \\
&\quad + \frac{e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)} \left(\frac{\partial}{\partial \zeta_j} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \zeta_j} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right)}{\left(1 - e^{\Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)}\right)^2} \\
&\quad \times \left(\frac{\partial}{\partial \beta_{j'k'}} \Lambda\left(L_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right) - \frac{\partial}{\partial \beta_{j'k'}} \Lambda\left(R_i | \boldsymbol{X}_i, \boldsymbol{Z}_i\right)\right) \Bigg].
\end{aligned}
$$

The second derivative of the complete data loglikelihood with respect to $\beta_{jk'}$ and $\beta_{j'k''}$ is:

$$\frac{\partial}{\partial\beta_{jk'}\partial\beta_{j'k''}}\log\left(L\left(\Theta\right)\right) = \sum_{i=1}^{n} -\frac{\overline{\boldsymbol{B}}_{jk'}^{T}\overline{\boldsymbol{B}}_{j'k''}}{\sigma_{Y_j}^{2}} + \frac{\partial}{\partial\beta_{jk'}\partial\beta_{j'k''}}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right),$$

where the second derivative of $\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)$ with respect to $\beta_{jk'}$ and $\beta_{j'k''}$ is:

$$
\begin{aligned}
&\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\ell_p\left(\boldsymbol{\theta};\sigma_b^2\right)=\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\ell_0\left(\boldsymbol{\theta};\sigma_b^2\right)\\
&=\sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\beta_{j'k''}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\beta_{j'k''}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{j'k''}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\beta_{j'k''}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{j'k''}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg]\\
&=\sum_{i=1}^{n}\Bigg[-\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\delta_{O_i}+\delta_{R_i}\right)+\frac{\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\delta_{L_i}\\
&\quad-\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\frac{\partial}{\partial\beta_{j'k''}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}{\left(1-e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\delta_{L_i}\\
&\quad-\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\delta_{I_i}\\
&\quad-\frac{\left(\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-1}}\delta_{I_i}\\
&\quad-\frac{e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\beta_{j'k''}}\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{j'k''}}\Lambda\left(T_i^r|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\delta_{I_i}\\
&\quad+\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)+\frac{\left(\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial^2}{\partial\beta_{jk'}\partial\beta_{j'k''}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{e^{\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}-1}\\
&\quad+\frac{e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\left(\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{jk'}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)}{\left(1-e^{\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}\right)^2}\\
&\quad\times\left(\frac{\partial}{\partial\beta_{j'k''}}\Lambda\left(L_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\frac{\partial}{\partial\beta_{j'k''}}\Lambda\left(R_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right)\Bigg].
\end{aligned}
$$

The final equation for each derivative is better computationally and comes from multiplying by a form of 1 on certain terms. For example, for the second left censor term we multiplied by $\left(\frac{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}{e^{-\Lambda\left(T_i^l|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)}}\right)^2$.

Appendix I

Derivation of $\sigma^2_{Y_jc}$ and $\Sigma_{\boldsymbol{\xi}_jc}$

We need to estimate all $\sigma^2_{Y_jc}$'s that are part of $\boldsymbol{\sigma}^2_Y$ along with each $\Sigma_{\boldsymbol{\xi}_jc}$, where $c = 1, \ldots, C$ and $j = 1, \ldots, p$. Below we work through finding the maximum likelihood estimate for one $\sigma^2_{Y_jc}$ and one $\Sigma_{\boldsymbol{\xi}_jc}$ below.

I.1 Derivation of $\sigma^2_{Y_jc}$

We know that $\boldsymbol{Y}_{ij}$ is the i^{th} individual's j^{th} longitudinal vector of measures of length m_{ij}. Also, we know that $\boldsymbol{Y}_{ij} \sim N_{m_{ij}}\left(\boldsymbol{X}_{ijc}, \sigma^2_{Y_jc}\right)$. Now, for each cluster c, each $X_{ij}(t) = \psi_j(t) + \sum_{k'=1}^{J} \xi_{ijk'} B_{k'}(t)$ where $\psi_j(t)$ is the j^{th} mean trajectory function and is given by $\psi_j(t) = \sum_{k'=1}^{\overline{J}} \beta_{jk'} \overline{B}_{k'}(t)$. Therefore, the part of the loglikelihood that depends on $\sigma^2_{Y_j c}$ is:

$$\begin{aligned}
\ell\left(\sigma^2_{Y_j c}\right) &= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\ell_i\left(\sigma^2_{Y_j c};\boldsymbol{Y}_{ij},\boldsymbol{\beta}_j,\boldsymbol{\xi}^{(r)}_{ij},\boldsymbol{\tau}^{(r)}_i\right)\\
&= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\log\left[\left(\sigma^2_{Y_j c}\right)^{\frac{m_{ij}}{2}} e^{-\frac{\|\boldsymbol{Y}_{ij}-\boldsymbol{X}^{(r)}_{ijc}\|^2}{2\sigma^2_{Y_j c}}}\right]^{\tau^{(r)}_{ic}}\\
&= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\left[\frac{m_{ij}\tau^{(r)}_{ic}}{2}\log\left(\sigma^2_{Y_j c}\right)-\frac{\|\boldsymbol{Y}_{ij}-\boldsymbol{X}^{(r)}_{ijc}\|^2\tau^{(r)}_{ic}}{2\sigma^2_{Y_j c}}\right].
\end{aligned}$$

Taking the derivative with respect to $\sigma^2_{Y_j c}$, setting it equal to 0, and solving, we get:

$$\begin{aligned}
\frac{d\ell\left(\sigma^2_{Y_j c}\right)}{d\sigma^2_{Y_j c}} &= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}-\frac{m_{ij}\tau^{(r)}_{ic}}{2\sigma^2_{Y_j c}}+\frac{\|\boldsymbol{Y}_{ij}-\boldsymbol{X}^{(r)}_{ijc}\|^2\tau^{(r)}_{ic}}{2\left(\sigma^2_{Y_j c}\right)^2}\\
\frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\frac{m_{ij}\tau^{(r)}_{ic}}{2\sigma^2_{Y_j c}} &= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\frac{\|\boldsymbol{Y}_{ij}-\boldsymbol{X}^{(r)}_{ijc}\|^2\tau^{(r)}_{ic}}{2\left(\sigma^2_{Y_j c}\right)^2}\\
\sigma^2_{Y_j c}\sum_{r=1}^{R}\sum_{i=1}^{n}m_{ij}\tau^{(r)}_{ic} &= \sum_{r=1}^{R}\sum_{i=1}^{n}\|\boldsymbol{Y}_{ij}-\boldsymbol{X}^{(r)}_{ijc}\|^2\tau^{(r)}_{ic}\\
\widehat{\sigma}^2_{Y_j c} &= \frac{\sum_{r=1}^{R}\sum_{i=1}^{n}\|\boldsymbol{Y}_{ij}-\boldsymbol{X}^{(r)}_{ijc}\|^2\tau^{(r)}_{ic}}{\sum_{r=1}^{R}\sum_{i=1}^{n}m_{ij}\tau^{(r)}_{ic}}.
\end{aligned}$$

I.2 Derivation of $\Sigma_{\boldsymbol{\xi}_j c}$

Now we work through finding the maximum likelihood estimate of $\Sigma_{\boldsymbol{\xi}_j c}$ below. The part of the loglikelihood that depends on $\Sigma_{\boldsymbol{\xi}_j c}$ is:

$$\begin{aligned}
\ell\left(\Sigma_{\boldsymbol{\xi}_j c}\right) &= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\ell_i\left(\Sigma_{\boldsymbol{\xi}_j c}; \boldsymbol{\xi}_{ij}^{(r)}, \boldsymbol{\tau}_i^{(r)}\right) \\
&= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\log\left[\det\left(\Sigma_{\boldsymbol{\xi}_j c}\right)^{-\frac{1}{2}} e^{-\frac{\boldsymbol{\xi}_{ij}^{T(r)}\Sigma_{\boldsymbol{\xi}_j c}^{-1}\boldsymbol{\xi}_{ij}^{(r)}}{2}}\right]^{\tau_{ic}^{(r)}} \\
&= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\left[\frac{\tau_{ic}^{(r)}}{2}\log\left(\det\left(\Sigma_{\boldsymbol{\xi}_j c}^{-1}\right)\right) - \frac{\tau_{ic}^{(r)}}{2}\left(\boldsymbol{\xi}_{ij}^{T(r)}\Sigma_{\boldsymbol{\xi}_j c}^{-1}\boldsymbol{\xi}_{ij}^{(r)}\right)\right] \\
&= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\frac{\tau_{ic}^{(r)}}{2}\left[\log\left(\det\left(\Sigma_{\boldsymbol{\xi}_j c}^{-1}\right)\right) - \operatorname{tr}\left(\boldsymbol{\xi}_{ij}^{T(r)}\Sigma_{\boldsymbol{\xi}_j c}^{-1}\boldsymbol{\xi}_{ij}^{(r)}\right)\right] \\
&= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\frac{\tau_{ic}^{(r)}}{2}\left[\log\left(\det\left(\Sigma_{\boldsymbol{\xi}_j c}^{-1}\right)\right) - \operatorname{tr}\left(\Sigma_{\boldsymbol{\xi}_j c}^{-1}\boldsymbol{\xi}_{ij}^{(r)}\boldsymbol{\xi}_{ij}^{T(r)}\right)\right].
\end{aligned}$$

We know that the trace of a square matrix is equal to the sum of the diagonals. Therefore, the 2^{nd} to last line above is due to the fact that $\left(\boldsymbol{\xi}_{ij}^{T(r)}\Sigma_{\boldsymbol{\xi}_j}^{-1}\boldsymbol{\xi}_{ij}^{(r)}\right)$ is a 1×1 matrix and therefore it is equal to its trace. The trace also does not change under cyclic permutations, meaning $\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB)$, which we used in the last line. To find the derivative with respect to $\Sigma_{\boldsymbol{\xi}_j}^{-1}$, we use Jacobi's formula, which states that $\frac{d}{dx}(\det A(x)) = \det A(x) \cdot tr\left(A^{-1}(x)\frac{d}{dx}A(x)\right)$. We also use the fact that the derivative of the trace is equal to the trace of the derivative, $\frac{d}{dA}\operatorname{tr}(A) = \operatorname{tr}\left(\frac{d}{dA}A\right)$. Taking the derivative with respect to

$\Sigma_{\boldsymbol{\xi}_j}^{-1}$, setting it equal to 0, and solving, we get:

$$\begin{aligned}
\frac{d\ell\left(\Sigma_{\boldsymbol{\xi}_j c}^{-1}\right)}{d\Sigma_{\boldsymbol{\xi}_j c}^{-1}} &= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\frac{\tau_{ic}^{(r)}}{2}\left[\frac{1}{\det\left(\Sigma_{\boldsymbol{\xi}_j c}^{-1}\right)}\det\left(\Sigma_{\boldsymbol{\xi}_j c}^{-1}\right)\operatorname{tr}\left(\Sigma_{\boldsymbol{\xi}_j c}\right) - \operatorname{tr}\left(\boldsymbol{\xi}_{ij}^{(r)}\boldsymbol{\xi}_{ij}^{T(r)}\right)\right] \\
&= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\frac{\tau_{ic}^{(r)}}{2}\left[\operatorname{tr}\left(\Sigma_{\boldsymbol{\xi}_j c}\right) - \operatorname{tr}\left(\boldsymbol{\xi}_{ij}^{(r)}\boldsymbol{\xi}_{ij}^{T(r)}\right)\right] \\
&= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\frac{\tau_{ic}^{(r)}}{2}\operatorname{tr}\left(\Sigma_{\boldsymbol{\xi}_j c} - \boldsymbol{\xi}_{ij}^{(r)}\boldsymbol{\xi}_{ij}^{T(r)}\right) \\
&= \operatorname{tr}\left[\frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\frac{\tau_{ic}^{(r)}}{2}\left(\Sigma_{\boldsymbol{\xi}_j c} - \boldsymbol{\xi}_{ij}^{(r)}\boldsymbol{\xi}_{ij}^{T(r)}\right)\right].
\end{aligned}$$

This derivative is only 0 if the trace is 0. Thus, solving for $\Sigma_{\boldsymbol{\xi}_j c}$ within the trace gives:

$$\begin{aligned}
\frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\frac{\tau_{ic}^{(r)}}{2}\Sigma_{\boldsymbol{\xi}_j c} &= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\frac{\tau_{ic}^{(r)}}{2}\left(\boldsymbol{\xi}_{ij}^{(r)}\boldsymbol{\xi}_{ij}^{T(r)}\right) \\
\sum_{r=1}^{R}\sum_{i=1}^{n}\tau_{ic}^{(r)}\Sigma_{\boldsymbol{\xi}_j c} &= \sum_{r=1}^{R}\sum_{i=1}^{n}\tau_{ic}^{(r)}\left(\boldsymbol{\xi}_{ij}^{(r)}\boldsymbol{\xi}_{ij}^{T(r)}\right) \\
\widehat{\Sigma}_{\boldsymbol{\xi}_j c} &= \frac{\sum_{r=1}^{R}\sum_{i=1}^{n}\tau_{ic}^{(r)}\left(\boldsymbol{\xi}_{ij}^{(r)}\boldsymbol{\xi}_{ij}^{T(r)}\right)}{\sum_{r=1}^{R}\sum_{i=1}^{n}\tau_{ic}^{(r)}}.
\end{aligned}$$