Summer 2019

Statistical Analysis of Interval-Censored Data Subject to Additional Complications

Qiang Zheng

Statistical analysis of Interval-Censored Data Subject to Additional Complications

by

Qiang Zheng

Bachelor of Science
Pacific Lutheran University, 2010

Master of Arts
Columbia University, 2012

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Statistics

College of Arts and Sciences

University of South Carolina

2019

Accepted by:

Lianming Wang, Major Professor

John Grego, Committee Member

Dewei Wang, Committee Member

Jiajia Zhang, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

Dedication

This dissertation is dedicated to

My Mother & My Wife

Liyan Xu & Quan Lin

My Father

Xiaoming Zheng

My Friends who Help Me and love Me

Acknowledgments

Firstly and most importantly, I would like to express my deepest and heartfelt thanks to my advisor Dr. Lianming Wang. He is much more than a academic advisor to me. He is a mentor, a good friend and a great person. He always tries his best to help me and support me. I learned so much from him besides my research studies, such as how to face difficulties in my life, how to make choices and how to be peaceful. I receive many suggestions from him and benefit from the suggestions during my Ph.D study.

Secondly, I want to thank my committee Dr. John Grego, Dr. Dewei Wang and Dr. Jiajia Zhang. They provide significantly helpful suggestions to my dissertation.

Thirdly, I want to thank my mother and my wife. My mother is the greatest mother in the world. She always puts me at the first position in all cases. Without her love and support, I cannot finish my bachelor, master and Ph.D. degrees in US. My wife is the most interesting, passionate and beautiful girl I have never met. She makes my life colorful and full of happiness. I also want to thank my father's support.

Lastly, I want to thank my friends, all my friends. I received so much help from my friends through my undergraduate study to my graduate study in my daily life and my study.

Abstract

Survival analysis is an important branch of statistics that studies time to event data (or survival data), in which the response variable is time to a certain event of interest. The most prominent feature of survival data is that the response is not exactly observed due to limits of the study design or nature of the event of interest. Interval-censored data are a common type of survival data and occur frequently in real life studies where subjects are examined at periodical follow ups. The response time is usually not observed, but the status of the event of interest is known at each examination time. In such cases, the response time for each subject is only known to fall within an interval formed by two examination times in which the status of the event has changed. This dissertation proposes new statistical approaches for analyzing real life interval-censored data with additional complications.

Chapter 1 provides an introduction to this dissertation. Firstly, it gives a description of interval-censored data and an explanation of how interval-censored data are obtained with some illustrative examples. Then, a widely used model, the proportional hazards (PH) model, for analyzing interval-censored data is introduced. Thirdly, some literature for fitting the PH model to interval-censored data is reviewed. Fourthly, three additional complications of the analysis of interval-censored data are presented. Lastly, real data sets are given to explain the motivations for studying these complications.

Chapter 2 of this dissertation develops an expectation-maximization (EM) algorithm for analyzing arbitrarily-censored data under the PH model. Arbitrarily-censored data refer to the data sets that include interval-censored observations and

exactly observed failure times. The method developed in Chapter 2 can be considered as an extension of the paper, Wang et al. (2016). The proposed method enjoys all the good properties of Wang's method, such as flexibility, computational efficiency, accuracy, robustness to the choice of initial values, quick convergence and closed-form variance estimation.

Chapter 3 studies current status data, a special case of interval-censored data, with informative censoring. This study was motivated by the tumor studies conducted by the National Toxicology Program (NTP). In such studies, the tumor onset time at a specific important organ of a mice or rat is usually observed but either left- or right-censored at the sacrifice time depending on whether a tumor is found there, resulting in current status data for the tumor onset time. However, the sacrifice time can be correlated to the tumor onset time because some of such animals are killed when they show symptoms of sickness or serious weight loss potential due to the exposure of the substance being tested. This leads to informative censoring problem and ignoring it may cause serious bias and misleading results. In this chapter, a new estimation approach is proposed based on an EM algorithm and has shown excellent performance in the simulation study. The new approach has many good merits such as being robust to initial values, fast to converge, and easy to implement, and providing variance estimates in closed form. The approach is illustrated by applications to two real data sets from NTP studies.

Chapter 4 studies an estimation of system reliability when the status of all components are also known. Both the system and component data are available in such situations, and all these failure times are either left-censored or right-censored at the examination time depending on whether the system and each component has failed. Different strategies are discussed for estimating system reliability: (1) use system data only and (2) use component data. When component data are used, two models are studied under different assumptions on whether component failure times are indepen-

dent or correlated. A new estimation method under the gamma frailty proportional hazards model is proposed to handle the situation when the component failure times are correlated. A detailed comparison is conducted among these different strategies.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Interval-Censored Data

Survival analysis is an important branch of statistics that studies time to event data (or survival data), in which the response variable is time to a certain event of interest. The most prominent feature of survival data is that the response is not exactly observed due to limits of the study design or nature of the event of interest.

Interval-censored data are a common type of survival data and occur frequently in real life studies where subjects are examined at periodical follow ups. The response time, also known as failure time in survival analysis, is usually not observed, but the status of the event of interest is known at each examination time. In such cases, the response time for each subject is only known to fall within an interval formed by two examination times in which the status of the event has changed. For example, in HIV studies, the onset time of HIV for a subject cannot be exactly observed but the status of HIV can be known through laboratory tests, resulting in interval-censored data for the HIV onset time. Interval-censored data consist of left-, interval-, and right-censored observations. A left-censored observation refers to an observation whose failure time is before the first examination time; an interval-censored observation refers to an observation whose failure time is between two examination times; a right-censored observation refers to an observation whose failure time is beyond the last examination time.

1.2 Current Status Data

A well-known special case of interval-censored data is current status data, where each subject is examined only once. The failure time of each subject cannot be observed, but the status of the event of interest is known at the examination time. Therefore, for each subject, one can only know that the failure time occurs before or after the examination time. As a result, current status data only contain left- and right-censored observations. Current status data are commonly seen in epidemiological, medical or toxicology studies. For example, in an experiment of toxicology, lab rats are exposed to a substance with different concentrations for a period of time to study whether the substance was associated with the onset time of tumors in their organs. The onset time of tumors cannot be observed directly, while the status of whether a lab rat has a tumor in its organs can be observed after the rat is sacrificed. In this example, the examination time, or censoring time, is the time when a lab rat is sacrificed. The data obtained in this experiment are current status data.

1.3 Informative Censoring and Non-Informative Censoring

In many interval-censored data or current status data studies, one of the common assumptions is that the failure time is independent of the observational process given covariates, which is known as non-informative censoring. On the contrary, informative censoring refers to those situations when the failure time is correlated with the observational process given covariates. Non-informative censoring is a natural assumption in many interval-censored data studies. For instance, in the HIV study example, it is natural to think that the HIV onset time is independent of the HIV laboratory test time given covariates. However, in some of the cases, informative censoring can be a reasonable assumption. Examples of these cases are given in Section 1.6.

1.4 Models

1.4.1 Notations

The notations used to analyze interval-censored data are as follows. Let T denote the failure time. In interval-censored data, T only can be known to fall into an interval consisting of two examination times, L and R, which are known as the left censoring time and the right censoring time. L is assumed to be strictly less than R in most literature. For the ith subject in a sample with size n, T_i can be observed in three different forms through L_i and R_i: $(0, R_i)$, (L_i, R_i) or $(L_i, +\infty)$, for $i = 1, 2, ..., n$. For an observation in the form $(0, R_i)$, $L_i = 0$ and the observation is left censored. For an observation in the form (L_i, R_i), $0 < L_i < R_i < +\infty$ and the observation is interval censored. For an observation in the form $(L_i, +\infty)$, $R_i = +\infty$ and the observation is right censored.

1.4.2 the Proportional Hazards (PH) model

The PH model is a popular and widely used model to analyze interval-censored data in survival analysis proposed by Cox (1972). The 'Hazards' in the PH model refers to the hazard function denoted as $\lambda(\cdot)$. It is defined as follows,

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leqslant T < t + \Delta t | T \geqslant t)}{\Delta t},$$

where T is the failure time, Δt is a very small time range, $P(\cdot|\cdot)$ is a conditional probability. It can be seen that the hazard function is not a density nor a probability. It can be thought as the likelihood of an observation's failure time falling between t and $t + \Delta t$ given that it has survived up to time t. In this sense, the hazard function is a measure of risk, namely instant death or failure. The greater value the hazard function takes, the greater risk of a subject has. Additionally, It can be shown that $\lambda(t) = \frac{f(t)}{S(t)}$, where $f(\cdot)$ is a probability density function (pdf) and $S(\cdot)$ is a survival function. This relation is going to be used in Chapter 3.

The PH model assumes that the hazard function given a covariate vector $\mathbf{x}$ is proportional to the baseline hazard function, $\lambda_0(\cdot)$. The rate equals to the exponential of a linear combination of a covariate vector $\mathbf{x}$.

$$\lambda(t|\mathbf{x}) = \lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta}),$$

where $\boldsymbol{\beta}$ is a vector of regression parameters. In the PH model, the baseline hazard function $\lambda_0(t)$ and $\boldsymbol{\beta}$ need to be estimated.

There are several studies which analyze interval-censored data under the PH model. In the paper of Wang et al. (2016) and McMahan et al. (2013), they fitted the PH model to the interval-censored data and current status data respectively. The baseline hazard functions were modeled by spline functions. EM algorithms were developed to find the maximum likelihood estimators (MLE) of the regression parameters and the spline functions' parameters jointly. In the paper of Cai et al. (2011), the PH model is fitted to the current status data. The baseline hazard function was also modeled by spline functions. A Bayesian method was developed to find the estimates of the parameters with posterior means. Betensky et al. (2002) fitted the PH models to right-censored and interval-censored data. They developed an EM algorithm to find the estimates of the parameters. Devarajan and Ebrahimi (2011) proposed a generalized version of the PH model by adding a power function to the baseline hazard function. They argued that it can allow the correlation between covariates and the baseline hazard function in the PH model. Tian et al. (2005) added time-varying parameters into the PH model and a kernel-weighted partial likelihood approach was applied. Pan (1999) proposed a generalized gradient projection method by reformulating the iterative convex minorant algorithm as an extension of the PH model. Murphy and Vaart (1997) studied the confidence interval for the real parameter presence in finite parameters cases. Even though the nuisance parameters are not guaranteed to be normal distributed, the likelihood ratio statistics is still asymptotic chi-squared distributed. Moreover, they provided an example to test the significance of the regression parameters in the PH model with current status data. Satten (1996) fitted the PH model to interval-censored data and used a marginal likelihood approach. The method did not need specification of the baseline hazard function. Goggins et al. (1998) fitted the PH model to interval-censored data and developed a Monte Carlo EM (MCEM) algorithm to find the parameter estimates. Huang (1996) studied the PH model's efficiency with interval-censored data. It was shown that with finite number of parameters, the MLE for the PH model is asymptotically efficient but with infinite-dimensional parameters the MLE converges slower than $\sqrt{n}$. The regression analysis of fitting the PH model to interval-censored data was first addressed by Finkelstein (1986) .

1.4.3 The Gamma-frailty Proportional Hazards model

The Gamma-frailty PH model is one type of frailty models to fit correlated survival data. The Gamma-frailty PH model keeps the structure of the PH model and adds an extra gamma frailty term to capture the correlation.

Let T_1 and T_2 be two correlated failure times of interest. Let η be a frailty term distributed as $Gamma(\nu, \nu)$, where $\nu > 0$. Under the Gamma-frailty PH model, given the frailty term η and a covariate vector $\mathbf{x}$, the hazard functions of T_1, T_2 can be expressed as follows,

$$\lambda_1(t|\mathbf{x}, \eta) = \lambda_{01}(t) \exp(\mathbf{x}'\boldsymbol{\beta_1})\eta,$$
$$\lambda_2(t|\mathbf{x}, \eta) = \lambda_{02}(t) \exp(\mathbf{x}'\boldsymbol{\beta_2})\eta,$$

where $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are regression parameters, $\lambda_{01}(t)$ is the baseline hazard function for T_1, $\lambda_{02}(t)$ is the baseline hazard function for T_2.

The Gamma-frailty PH model is often used to study multivariate current status problems. In the paper of Wang et al. (2015), it used the Gamma-frailty PH model to fit bi-variate current status data. An EM algorithm was developed to find the

MLE of the parameters. In Chapter 3, the Gamma-frailty PH model is fitted to the current status data with informative censored observations and non-informative censored observations.

1.5 Additional Complication I

For the interval-censored data, left censoring time is strictly less than right censoring time, i.e., $L < R$. The case that $L = R$ is not considered in interval-censored data. It is because when $L = R$, it indicates that exact failure time is observed, but exact failure time cannot be observed due to periodical examinations in most of the study designs or the nature of the event of interest. However, in some of the situations, besides censored observations, exact failure times are available for a part of the observations in the data. These data are called arbitrarily-censored data. It can be seen that arbitrarily-censored data contain exactly observed failure times, left-, interval- and right-censored observations. Most methods used to analyze interval-censored data can deal with exactly observed failure times by an approximation with an interval. For example, one can use exactly observed failure time as a lower bound and add a very small number to the lower bound to obtain an upper bound. As a result, exactly observed failure times can be turned into interval-censored observations. However, doing that adds uncertainty to the existing data and will cause to overestimate the variance of the regression parameter estimates. The overestimation may cause substantial problem when the proportion of exactly observed failure times is large in a data set. In the following examples, the data include a large proportion of exactly observed failure times.

1.5.1 Diabetes data

This study was conducted in the Steno Memorial Hospital in Denmark from 1933 to 1984. It studied the onset of diabetic nephronpathy. The survival time was the time from onset of diabetes to the onset of diabetic nephronpathy, a major complication of Type I diabetes. The data can be found in **'icenReg'** package and **'glrt'** package in R. The data set contains 731 patients and one covariate, gender (454 males and 277 females). In this data set, there are 595 exactly observed failure times, 1 left-censored observations, 135 interval-censored observations and no right-censored observations. The exactly observed failure times are the greatest proportion in the data.

1.5.2 Childhood morality data

Under 5 mortality rate is a key indicator of the development for the overall child health for a country, . The childhood morality data, children dying between the 1st and 5th birthdays, are from Demographic and Health Surveys in Nigeria in 2003. The study sought for the factors that effect the childhood morality such as the number of breastfeeding months, a mother's educational level, body mass index (BMI) and so on. The death time is the failure time. In the survey, if a mother remembered the exact time of her children's death then the failure time was exact observed. Otherwise a mother provided a time range which her children died from. Therefore, these observations were interval censored. The children's death rate of this data set is 0.059. This data set contains 5890 observations with 11 covariates and 2766 compete cases. The exact failure times are also the largest proportion in this data set.

1.6 Additional Complication II

One of the common assumptions made when one analyzes interval-censored data or current status data is that the observational process is independent of the failure time given covariates, which is known as non-informative censoring. While in some of the cases, it is more reasonable to assume the observational process is correlated with the failure time given covariates, which is known as informative censoring. A real data example is as follows.

1.6.1 The National Toxicology Program (NTP), Study tr-476

The NTP, an inter-agency program whose mission is to evaluate agents of public health concern by developing and applying tools of modern toxicology and molecular biology (from NTP website), performed studies about toxicology and carcinogens of chloroprene and provided a report in September 1998. In the manufacture of neoprene, chloroprene, the 2-chloro analogue of 1, 3-butadiene, a potent, multi-species, multi-organ carcinogen, is only used but with high production and not much information about its carcinogenic potential (from the report of NTP). In a 2-years mice study, groups of 50 male and 50 female mice were exposed to chloroprene at concentrations of control (0 ppm), low dose (12.8ppm), medium dose (32ppm), or high dose (80 ppm) by inhalation, 6 hours per day, 5 days per week, for 2 years. During the experiment, the mice were removed from the study due to accidentally kill, natural death, terminal sacrifice or moribund sacrifice to be observed for whether tumors existed in their organs.

The motivation of the study in Chapter 3 is from the NTP experiments. In the example above, it is natural to consider that given the level of concentration, natural death, accidentally kill and terminal sacrifice were not related to tumors in the organs of mice while moribund sacrifice was. Therefore the correlation between the observational process and the failure time needs to be taken into consideration in this case.

1.7 Additional Complication III

Fitting the PH model to current status data is a well-studied problem. In NTP studies, the PH model can be fitted to the data to analyze whether an experimented substance is toxic to lab mice. In these analysis, it is worth to notice that a lab mouse is essentially a system with many components, i.e., its organs such as a liver or a lung. Therefore, in NTP studies, the PH model is fitted to system data to analyze the system reliability. In addition to system data, the current status data of each component are also available in NTP studies so that one can use component data to analyze the reliability of the system. In Chapter 4, both methods are used to analyze the reliability of a system and the risk of fitting the PH model to system data is discussed.

Chapter 2

Fitting the Proportional Hazards Model to Arbitrarily-Censored Data

2.1 Introduction

Arbitrarily-censored data refer to the data sets that contain exactly observed failure times, left-, interval-, and right-censored observations. In this chapter, an efficient and flexible algorithm is developed for fitting the PH model to arbitrarily-censored data.

Although fitting the PH model to interval-censored data is a well-studied problem, there are only a few studies analyzing arbitrarily-censored data under the PH model. Clifford Anderson-Bergman (2018) fitted the PH model to arbitrarily-censored data and developed an algorithm with two steps for the estimation. One step was to estimate the regression parameters with the conditional Newton Raphson algorithm. The other step was to estimate the baseline survival parameters with the iterative convex minorant (ICM) algorithm. When exactly observed failure times existed, a gradient desccent was used to update the baseline parameters, especially in the case that the proportion of exactly observed failure times was large. The method can be applied through the package '**icenReg**'.

In this chapter, the PH model is fitted to arbitrarily-censored data and formulated in a fashion with finite parameters. An EM algorithm is developed to find the MLE of the parameters. The details of the proposed methodology are provided in Section 2.2 - Section 2.6. The details include modeling the baseline hazard function and the cumulative baseline hazard function with spline functions, a data augmentation including 2-stage Poisson latent variables and multinomial latent variables, the development of the EM algorithm, the asymptotic distribution and variance estimation. Several simulation studies and real data applications are performed in Section 2.7 and Section 2.8 to evaluate the performance of the proposed method.

2.2 Models, Notations and Observed Likelihood

Let T and $\mathbf{x}$ denote the failure time and the covariate vector respectively. Let $F(t|\mathbf{x})$, $S(t|\mathbf{x})$ and $f(t|\mathbf{x})$ be the cumulative distribution function (CDF), the survival function and the probability density function (pdf) for the failure time given the covariate vector $\mathbf{x}$. $\lambda(t|\mathbf{x})$ and $\Lambda(t|\mathbf{x})$ are the hazard function and the cumulative hazard function for the failure time given the covariate vector $\mathbf{x}$. Under the PH model,

$$\lambda(t|\mathbf{x}) = \lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta}),$$
$$\Lambda(t|\mathbf{x}) = \Lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta}),$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)'$ is a vector of regression parameters. The baseline hazard function $\lambda_0(t)$ and the cumulative baseline hazard function $\Lambda_0(t)$ only depend on the failure time T. The relation between $\lambda(t|\mathbf{x})$ and $\Lambda(t|\mathbf{x})$ is that $\Lambda(t|\mathbf{x}) = \int_0^t \lambda(u|\mathbf{x})du$. Under the PH model, the CDF, the survival function and the pdf for the failure time given the covariate vector $\mathbf{x}$ can be written as

$$F(t|\mathbf{x}) = 1 - \exp\left\{-\Lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\right\},$$
$$S(t|\mathbf{x}) = \exp\left\{-\Lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\right\},$$
$$f(t|\mathbf{x}) = \lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\exp\left\{-\Lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\right\}.$$

One assumption is made that given the covariate vector $\mathbf{x}$, the failure time is independent of the observational process. For the ith observation in a sample, L_i and R_i are left censoring time and right censoring time, for $i = 1, 2, ..., n$. Note that for a left-censored (right-censored) observation, $L_i = 0$ ($R_i = \infty$). For a exactly observed failure time, it can be considered as its left censoring time and right censoring time are the same, that is $L_i = R_i$. Let δ_{i0}, δ_{i1}, δ_{i2} and δ_{i3} be the indicators for the ith observation to be exact failure time, left-, interval- or right-censored, for $i = 1, 2, ..., n$. Note that $\delta_{i0} + \delta_{i1} + \delta_{i2} + \delta_{i3} = 1$. With these notations and the assumption, the observed likelihood function can be written as

$$\mathcal{L}_{obs} = \prod_{i=1}^{n} f(R_i|\mathbf{x}_i)^{\delta_{i0}} F(R_i|\mathbf{x}_i)^{\delta_{i1}} \left\{F(R_i|\mathbf{x}_i) - F(L_i|\mathbf{x}_i)\right\}^{\delta_{i2}} \left\{1 - F(L_i|\mathbf{x}_i)\right\}^{\delta_{i3}}.$$

Under the PH model, the observed likelihood function can be further written as

$$\begin{aligned}\mathcal{L}_{obs} = \prod_{i=1}^{n} & [\lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\exp\left\{-\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\right\}]^{\delta_{i0}}[1 - \exp\left\{-\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\right\}]^{\delta_{i1}} \\ & [\exp\left\{-\Lambda_0(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\right\} - \exp\left\{-\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\right\}]^{\delta_{i2}} \\ & [\exp\left\{-\Lambda_0(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\right\}]^{\delta_{i3}}.\end{aligned}$$

In the observed likelihood function, $\lambda_0(\cdot)$, $\Lambda_0(\cdot)$ and the vector of regression parameters $\boldsymbol{\beta}$ need to be estimated.

2.3 The Hazard Function $\lambda_0(\cdot)$ and the Cumulative Hazard Function $\Lambda_0(\cdot)$

Since both $\lambda_0(\cdot)$ and $\Lambda_0(\cdot)$ are of infinite dimensions, they can be difficult to estimate. The baseline hazard function $\lambda_0(\cdot)$ is proposed to be modeled by widely used spline functions, M-splines. M-splines are chosen here because that they are usually used to model a positive function and the baseline hazard function $\lambda_0(\cdot)$ only can take positive values. The approximation of the cumulative baseline hazard function is proposed to use the monotone spline functions, I-spline functions, because the I-spline functions are integration of M-spline functions, which naturally fits the relation between the baseline hazard function and the baseline cumulative hazard function. This method was used in several existing literature such as Cai et al. (2011), McMahan et al. (2013) and Wang et al. (2016).

With I-spline functions and M-spline functions, the baseline hazard function and the baseline cumulative hazard function can be expressed as follows,

$$\Lambda_0(\cdot) = \sum_{l=1}^{k} \gamma_l I_l(\cdot),$$
$$\lambda_0(\cdot) = \sum_{l=1}^{k} \gamma_l M_l(\cdot).$$

In these two functions, $I_l(\cdot)$'s and $M_l(\cdot)$'s are polynomial functions called basis functions. $M_l(\cdot)$'s are the derivatives of $I_l(\cdot)$'s. γ_l's are non-negative coefficients. To build these splines, one needs to specify the degree of the basis functions and choose an increasing sequence of knots within a certain range (Ramsay, 1988). The degree of the basis functions controls the smoothness of the basis functions and the placement of knots determines the flexibility of the basis functions. For example, one can put more knots in the range where more observations fall in to catch the fluctuation of the target function. For the basis functions, the 2nd and 3rd degrees are most commonly used to control the smoothness of the basis functions, which stand for quadratic basis functions and cubic basis functions. For one particular data set, different choices of knots lead to different models. To determine the number of knots for a single data set, the Akaike information criterion (AIC) can be used as a model selection criteria.

With splines functions, the observed likelihood function can be finally written as

$$\begin{aligned}
\mathcal{L}_{obs} = \prod_{i=1}^{n} & \left[\sum_{l=1}^{k} \gamma_l M_l(R_i) \exp\left\{\mathbf{x}_i'\boldsymbol{\beta} - \sum_{l=1}^{k} \gamma_l I_l(R_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})\right\}\right]^{\delta_{i0}} \\
& \left[1 - \exp\left\{-\sum_{l=1}^{k} \gamma_l I_l(R_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})\right\}\right]^{\delta_{i1}} \\
& \left[\exp\left\{-\sum_{l=1}^{k} \gamma_l I_l(R_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})\right\} - \exp\left\{-\sum_{l=1}^{k} \gamma_l I_l(R_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})\right\}\right]^{\delta_{i2}} \\
& \left[\exp\left\{-\sum_{l=1}^{k} \gamma_l I_l(R_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})\right\}\right]^{\delta_{i3}}.
\end{aligned}$$

The parameters in the likelihood function above are coefficients of the spline functions γ_l's and the regression parameter vector $\boldsymbol{\beta}$. One can try to estimate them by maximizing the observed likelihood function directly. However it can be seen that there are summations and subtractions inside products, which makes it difficult to maximize $\mathcal{L}_{obs}$. It turns out that maximizing the observed likelihood function is not workable. Therefore an EM algorithm is going to be developed to solve the maximization problem and find the MLE of the parameters.

2.4 Data Augmentation for the EM Algorithm

The EM algorithm is used to find the MLE of the regression parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)'$ and the parameter vector of the splines $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_k)'$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')$.

The derivation of the EM algorithm starts with a 2-stage data augmentation. This idea was firstly discussed in the paper, Wang et al. (2016). At stage 1, Poisson latent variables Z_i and W_i are introduced with mean $\sum_{l=1}^{k} \gamma_l I_l(t_{i1}) \exp(\mathbf{x}_i'\boldsymbol{\beta})$ and $\sum_{l=1}^{k} \gamma_l \{I_l(t_{i2}) - I_l(t_{i1})\} \exp(\mathbf{x}_i'\boldsymbol{\beta})$ respectively, for $i = 1, 2, ..., n$, where $t_{i1} = R_i 1_{(\delta_{i1}=1)} + L_i 1_{(\delta_{i2}=1)} + L_i 1_{(\delta_{i3}=1)}$, $t_{i2} = R_i 1_{(\delta_{i2}=1)} + L_i 1_{(\delta_{i3}=1)}$.

$$Z_i \sim Poisson \left\{ \sum_{l=1}^{k} \gamma_l I_l(t_{i1}) \exp(\mathbf{x}_i'\boldsymbol{\beta}) \right\},$$
$$W_i \sim Poisson \left[\sum_{l=1}^{k} \gamma_l \{I_l(t_{i2}) - I_l(t_{i1})\} \exp(\mathbf{x}_i'\boldsymbol{\beta}) \right].$$

At stage 2, for each i, latent variables Z_i and W_i are further decomposed as summations of k independent Poisson random variables, $Z_i = \sum_{l=1}^{k} Z_{il}$ and $W_i = \sum_{l=1}^{k} W_{il}$, where the means of Z_{il} and W_{il} are $\gamma_l I_l(t_{i1}) \exp(\mathbf{x}_i'\boldsymbol{\beta})$ and $\gamma_l \{I_l(t_{i2}) - I_l(t_{i1})\} \exp(\mathbf{x}_i'\boldsymbol{\beta})$ respectively, for $l = 1, 2, ..., k$.

$$Z_{il} \sim Poisson \{\gamma_l I_l(t_{i1}) \exp(\mathbf{x}_i'\boldsymbol{\beta})\},$$
$$W_{il} \sim Poisson [\gamma_l \{I_l(t_{i2}) - I_l(t_{i1})\} \exp(\mathbf{x}_i'\boldsymbol{\beta})]$$

with the restriction $\sum_{l=1}^{k} Z_{il} = Z_i, \sum_{l=1}^{k} W_{il} = W_i$.

With latent variables Z_i's and W_i's , the augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$ can be expressed as

$$\mathcal{L}_1(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left[\sum_{l=1}^{k} \{\gamma_l M_l(R_i)\} \right]^{\delta_{i0}} \exp[\delta_{i0} \{\mathbf{x}_i'\boldsymbol{\beta} - \Lambda_0(R_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})\}]$$
$$P_{Z_i}(Z_i)^{(1-\delta_{i0})} P_{W_i}(W_i)^{(\delta_{i2}+\delta_{i3})},$$

where $P_X(\cdot)$ denotes the probability mass function for the random variable X, $Z_i > 0$ if $\delta_{i1} = 1$, $Z_i = 0$ and $W_i > 0$ if $\delta_{i2} = 1$,$Z_i = 0$ and $W_i = 0$ if $\delta_{i3} = 1$. By integrating Z_i's and W_i's out of $\mathcal{L}_1(\boldsymbol{\theta})$, one can obtain the observed likelihood function $\mathcal{L}_{obs}(\boldsymbol{\theta})$. With the stage 2 latent variables Z_{il}'s and W_{il}'s, the augmented likelihood function $\mathcal{L}_2(\boldsymbol{\theta})$ can be expressed as

$$\mathcal{L}_2(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left[\sum_{l=1}^{k} \{\gamma_l M_l(R_i)\} \right]^{\delta_{i0}} \exp[\delta_{i0} \{\mathbf{x}_i'\boldsymbol{\beta} - \Lambda_0(R_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})\}]$$
$$\prod_{l=1}^{k} P_{Z_{il}}(Z_{il})^{(1-\delta_{i0})} P_{W_{il}}(W_{il})^{(\delta_{i2}+\delta_{i3})},$$

where $Z_i > 0$ if $\delta_{i1} = 1$, $Z_i = 0$ and $W_i > 0$ if $\delta_{i2} = 1$,$Z_i = 0$ and $W_i = 0$ if $\delta_{i3} = 1$, $\sum_{l=1}^{k} Z_{il} = Z_i, \sum_{l=1}^{k} W_{il} = W_i$. By integrating Z_{il}'s and W_{il}'s out of $\mathcal{L}_2(\boldsymbol{\theta})$, one can obtain the augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$.

It can be noticed that the summation in the first term of $\mathcal{L}_2(\boldsymbol{\theta})$ makes it difficult to maximize the augmented likelihood function $\mathcal{L}_2(\boldsymbol{\theta})$. Therefore latent random vectors V_i's are introduced to deal with this problem. $V_i = (V_{i1}, V_{i2}, ..., V_{ik})$ has a multinomial distribution, for $i = 1, 2, ..., n$, which can be expressed as

$$V_i = (V_{i1}, V_{i2}, ..., V_{ik}) \sim Multinomial \left\{ 1, (\frac{1}{k}, \frac{1}{k}, ..., \frac{1}{k}) \right\}.$$

Note that $\sum_{l=1}^{k} V_{il} = 1$, for $i = 1, 2, ..., n$. With the latent variable V_{il}'s, the complete likelihood function $\mathcal{L}_{com}(\boldsymbol{\theta})$ can be obtained as follows,

$$\mathcal{L}_{com}(\boldsymbol{\theta}) =$$
$$\prod_{i=1}^{n} \left[\exp[\delta_{i0} \{\mathbf{x}_i'\boldsymbol{\beta} - \Lambda_0(R_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})\}] \left\{ \prod_{l=1}^{k} [\gamma_l M_l(R_i)]^{\delta_{i0} V_{il}} P_{Z_{il}}(Z_{il})^{(1-\delta_{i0})} P_{W_{il}}(W_{il})^{(\delta_{i2}+\delta_{i3})} \right\} \right].$$

where $Z_i > 0$ if $\delta_{i1} = 1$, $Z_i = 0$ and $W_i > 0$ if $\delta_{i2} = 1$,$Z_i = 0$ and $W_i = 0$ if $\delta_{i3} = 1$, $\sum_{l=1}^{k} Z_{il} = Z_i, \sum_{l=1}^{k} W_{il} = W_i$. In the complete likelihood function $\mathcal{L}_{com}(\boldsymbol{\theta})$, the latent variables V_{il}'s, W_{il}'s, Z_{il}'s are treated as missing data. It can be seen that by integrating out V_{ij}'s in $\mathcal{L}_{com}(\boldsymbol{\theta})$, one can get the augmented likelihood function $\mathcal{L}_2(\boldsymbol{\theta})$. Then, by integrating out Z_{il}'s and W_{il}'s in $\mathcal{L}_2(\boldsymbol{\theta})$, one can obtain the augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$. Lastly, by integrating out Z_i's, W_i's in $\mathcal{L}_1(\boldsymbol{\theta})$, one can obtain the observed likelihood function $\mathcal{L}_{obs}(\boldsymbol{\theta})$. Consequently, $\mathcal{L}_{com}(\boldsymbol{\theta})$ is viewed as the complete data likelihood with all V_{il}'s, Z_i's, W_i's, Z_{il}'s and W_{il}'s missing.

2.5 The EM Algorithm

With the latent variables, the EM algorithm can be developed to find the MLE of $\boldsymbol{\theta}$ with two steps, an expectation step (E-step) and a maximization step (M-step).

2.5.1 E-step

The derivation of E-step starts with taking the logarithm of the complete likelihood function. Then one needs to find the expectation of the logarithm of the complete likelihood function with respect to all the latent variables given the covariate vector $\mathbf{x}$ and the current parameter $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)\prime}, \boldsymbol{\gamma}^{(d)\prime})'$.

$$\begin{aligned}\log \mathcal{L}_{com}(\boldsymbol{\theta}) =& \sum_{i=1}^{n}\sum_{l=1}^{k} \{\delta_{i0}V_{il} + (1-\delta_{i0})Z_{il} + (\delta_{i2}+\delta_{i3})W_{il}\} (\log \gamma_l + \mathbf{x}_i'\boldsymbol{\beta}) \\ &- \sum_{i=1}^{n}\sum_{l=1}^{k} \{\delta_{i0}I_l(R_i) + (\delta_{i1}+\delta_{i2})b_l(R_i) + \delta_{i3}b_l(L_i)\} \gamma_l \exp(\mathbf{x}_i'\boldsymbol{\beta}) + \mathcal{L}(\boldsymbol{\theta}^{(d)}),\end{aligned}$$

where $\mathcal{L}(\boldsymbol{\theta}^{(d)})$ is a function of $\boldsymbol{\theta}^{(d)}$ without $\boldsymbol{\theta}$.

Since the only interested parameter is $\boldsymbol{\theta}$, then $\mathcal{L}(\boldsymbol{\theta}^{(d)})$ can be treated as a constant. Therefore it will not have contributions to the estimate of $\boldsymbol{\theta}$. Hence $\mathcal{L}(\boldsymbol{\theta}^{(d)})$ can be dropped here. Then the $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ function, the expectation of $\log \mathcal{L}_{com}(\boldsymbol{\theta})$ with respect to all the latent variables Z_i's, Z_{il}'s, W_i's, W_{il}'s and V_i's given the covariate vector $\mathbf{x}$ and the current parameter $\boldsymbol{\theta}^{(d)}$, can be expressed as

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) &= E\left\{\log \mathcal{L}_{com}(\boldsymbol{\theta}) | \boldsymbol{x}, \boldsymbol{\theta}^{(d)}\right\} \\
&= \sum_{i=1}^{n} \sum_{l=1}^{k} \{\delta_{i0} E(V_{il}) + (1-\delta_{i0}) E(Z_{il}) + (\delta_{i2}+\delta_{i3}) E(W_{il})\} (\log \gamma_l + \mathbf{x}_i' \boldsymbol{\beta}) \\
&\quad - \sum_{i=1}^{n} \sum_{l=1}^{k} \{\delta_{i0} I_l(R_i) + (\delta_{i1}+\delta_{i2}) b_l(R_i) + \delta_{i3} b_l(L_i)\} \gamma_l \exp(\mathbf{x}_i' \boldsymbol{\beta})
\end{aligned}
$$

To maximize $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$, the expectations of the latent variables' need to be obtained first. Note that all these expectations are posterior expectations given the covariate vector $\mathbf{x}$ and the current parameter $\boldsymbol{\theta}^{(d)}$. The complete likelihood function and the augmented likelihood functions can be used to find the posterior distributions of the latent variables.

In the complete likelihood function, it can be found that the posterior distribution of V_i's are also multinomial distributed given the covariate vector $\mathbf{x}$ and the current parameter $\boldsymbol{\theta}^{(d)}$ according to its kernel,

$$V_i = (V_{i1}, V_{i2}, ..., V_{ik}) \sim Multinomial(1, \tilde{p_i}),$$

where $\tilde{p_i} = (\tilde{p_{i1}}, \tilde{p_{i2}}, ..., \tilde{p_{ik}})$, $\tilde{p_{il}} = \frac{\gamma_l M_l(t_i)}{\sum_{l=1}^{k} \gamma_l M_l(t_i)}$. Therefore the expectation of V_i's are as follows,

$$E(V_{il}) = \frac{\gamma_l M_l(t_i)}{\sum_{j=1}^{k} \gamma_j M_j(R_i)},$$

for $l = 1, 2, ..., k$ and $i = 1, 2, ..., n$.

The posterior distribution of Z_i's, Z_{il}'s, W_i's, W_{il}'s can be found with the complete likelihood function $\mathcal{L}_{com}(\boldsymbol{\theta})$ and the augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$. By observing the augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$, it can be found that both Z_i's and W_i's follow truncated Poisson distributions. Additionally, given Z_i's and W_i's, Z_{il}'s and W_{il}'s follow multinomial distributions. Therefore, the posterior expectations of Z_i's and W_i's, Z_{il}'s and W_{il}'s can be expressed as follows,

$$
\begin{aligned}
E(Z_i) &= \frac{\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\delta_{i1}}{1-\exp\left\{\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\right\}},\\
E(Z_{il}) &= \frac{\gamma_l I_l(R_i)}{\Lambda_0(R_i)}\cdot\frac{\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\delta_{i1}}{1-\exp\left\{\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\right\}},\\
E(W_i) &= \frac{\left\{\Lambda_0(R_i)-\Lambda_0(L_i)\right\}\exp(\mathbf{x}_i'\boldsymbol{\beta})\delta_{i2}}{1-\exp[-\left\{\Lambda_0(R_i)-\Lambda_0(L_i)\right\}\exp(\mathbf{x}_i'\boldsymbol{\beta})]},\\
E(W_{il}) &= \frac{\gamma_l\left\{I_l(R_i)-I_l(R_i)\right\}}{\Lambda_0(R_i)-\Lambda_0(L_i)}\cdot\frac{\left\{\Lambda_0(R_i)-\Lambda_0(L_i)\right\}\exp(\mathbf{x}_i'\boldsymbol{\beta})\delta_{i2}}{1-\exp[-\left\{\Lambda_0(R_i)-\Lambda_0(L_i)\right\}\exp(\mathbf{x}_i'\boldsymbol{\beta})]},
\end{aligned}
$$

for $l = 1, 2, ..., k$ and $i = 1, 2, ..., n$.

2.5.2 M-step

The next step, M-step, is to maximize the $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ function respect to the parameter $\boldsymbol{\theta}$. To maximize $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\theta}$, firstly one can take partial derivatives of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ respect to $\boldsymbol{\theta}$.

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial\gamma_l} &= \sum_{i=1}^{n}\gamma_l^{-1}\left\{\delta_{i0}E(V_{il})+(1-\delta_{i0})E(Z_{il})+(\delta_{i2}+\delta_{i3})E(W_{il})\right\}\\
&\quad-\sum_{i=1}^{n}\left\{(1-\delta_{i3})I_l(R_i)+\delta_{i3}b_l(L_i)\right\}\exp(\mathbf{x}_i'\boldsymbol{\beta}), \qquad (1)
\end{aligned}
$$

for $l = 1, 2, ..., k$.

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial\boldsymbol{\beta}} &= \sum_{i=1}^{n}\sum_{l=1}^{k}\left\{\delta_{i0}E(V_{il})+(1-\delta_{i0})E(Z_{il})+(\delta_{i2}+\delta_{i3})E(W_{il})\right\}\mathbf{x}_i\\
&\quad-\sum_{i=1}^{n}\sum_{l=1}^{k}\left\{(1-\delta_{i3})b_l(R_i)+\delta_{i3}b_l(L_i)\right\}\gamma_l\exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i. \qquad (2)
\end{aligned}
$$

Then set these partial derivative equations to zeros. $\boldsymbol{\theta}$ can be solved with unique solutions. Firstly, γ_l's can be solved in closed forms as functions of $\boldsymbol{\beta}$ using (1),

$$
\gamma_l = \frac{\sum_{i=1}^{n}\left\{\delta_{i0}E(V_{il})+(1-\delta_{i0})E(Z_{il})+(\delta_{i2}+\delta_{i3})E(W_{il})\right\}}{\sum_{i=1}^{n}\left\{(1-\delta_{i3})I_l(R_i)+\delta_{i3}b_l(L_i)\right\}\exp(\mathbf{x}_i'\boldsymbol{\beta})},\ \text{for } l = 1, 2, ..., k.
$$

Then plug the obtained γ_l's to (2) so that (2) only contains the parameter $\boldsymbol{\beta}$.

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{l=1}^{k}\left\{\delta_{i0}E(V_{il})+(1-\delta_{i0})E(Z_{il})+(\delta_{i2}+\delta_{i3})E(W_{il})\right\}\mathbf{x}_i\\
&\quad=\sum_{i=1}^{n}\sum_{l=1}^{k}\left\{(1-\delta_{i3})b_l(R_i)+\delta_{i3}b_l(L_i)\right\}\gamma_l(\boldsymbol{\beta})\exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i.
\end{aligned}
$$

Numerical methods such as Newton-Raphson method can be used to find the solution of $\boldsymbol{\beta}$ from the equation above. Then plug the solution of $\boldsymbol{\beta}$ back into the close form solutions of γ_l's to obtain the solution of γ_l's.

2.5.3 The summary of the EM algorithm

With the results obtained in E-step and M-step, the EM-algorithm can be constructed by the following steps:

Step 1: Assign initial values to $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)\prime}, \boldsymbol{\gamma}^{(d)\prime})'$ and set $d = 0$.

Step 2: Obtain $\boldsymbol{\beta}^{(d+1)}$ by solving the following equation for $\boldsymbol{\beta}$,

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{l=1}^{k}\left\{\delta_{i0}E(V_{il}) + (1-\delta_{i0})E(Z_{il}) + (\delta_{i2}+\delta_{i3})E(W_{il})\right\}\mathbf{x}_i \\ &\quad = \sum_{i=1}^{n}\sum_{l=1}^{k}\left\{(1-\delta_{i3})I_l(R_i) + \delta_{i3}I_l(L_i)\right\}\gamma_l^{(d)}(\boldsymbol{\beta})\exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i,\end{aligned}$$

where

$$\gamma_l^{(d)}(\boldsymbol{\beta}) = \frac{\sum_{i=1}^{n}\left\{\delta_{i0}E(V_{il}) + (1-\delta_{i0})E(Z_{il}) + (\delta_{i2}+\delta_{i3})E(W_{il})\right\}}{\sum_{i=1}^{n}\left\{(1-\delta_{i3})I_l(R_i) + \delta_{i3}b_l(L_i)\right\}\exp(\mathbf{x}_i'\boldsymbol{\beta})}.$$

Step 3: Obtain $\gamma_l^{(d+1)} = \gamma_l^{(d)}(\boldsymbol{\beta}^{(d+1)})$.

Step 4: Repeat step 2-3 until $|\boldsymbol{\theta}^{(d+1)} - \boldsymbol{\theta}^{(d)}|$ is smaller than a tolerance value.

The algorithm is robust to initial values of $\gamma_l^{(d)}$'s because the values of $\gamma_l^{(d)}$'s during the iterations are all positive due to the close form solutions of $\gamma_l^{(d)}$'s so that they cannot be too far away from the truth such as being negative values. The solutions obtained by the developed EM algorithm, denoted as $\hat{\boldsymbol{\theta}}$, is the MLE of $\boldsymbol{\theta}$.

2.6 Asymptotic Properties and Variance Estimation

Under standard regularity conditions, the MLE enjoys the property of the asymptotic normality. That is, as $n \to +\infty$,

$$n^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N\left\{\mathbf{0}, I^{-1}(\boldsymbol{\theta})\right\},$$

where $I(\boldsymbol{\theta})$ is the fisher information matrix. To estimate the variance covariance matrix of $\hat{\boldsymbol{\theta}}$, Louis's method (Louis, 1982) is adopted to obtain $\hat{I}(\hat{\boldsymbol{\theta}})$ which is an estimation of $I(\boldsymbol{\theta})$.

$$\hat{I}(\hat{\boldsymbol{\theta}}) = -\frac{\partial^2 \log \mathcal{L}_{obs}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}},$$

where

$$-\frac{\partial^2 \log \mathcal{L}_{obs}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \text{var}\left\{\frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}.$$

Both $\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ and $\text{var}\left\{\frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}$ have close forms. The technical details can be found in Chapter 2 Supplementary Materials.

2.7 Simulation Study

Simulation studies were conducted to evaluate the performance of the proposed methodology with different proportion of exactly observed failure times, 0%, 5%, 20% and 50%. The simulations were based on the following true distribution of the failure time T.

$$F_T(t|\mathbf{x}) = 1 - \exp\left\{-\Lambda_0(t)\exp(x_1\beta_1 + x_2\beta_2)\right\},$$

where $\Lambda_0(t) = \log(t+1) + t^2$, $x_1 \sim Bernoulli(0.5)$ and $x_2 \sim N(0, 0.5^2)$. The sample size n was chosen to be 200 and all possible combinations of $\beta_1 = \{-1, 1\}$ and $\beta_2 = \{-1, 1\}$ were considered, resulting in four parameter configurations.

The simulation process was as follows. For the ith observation, a uniform random variable with support from 0 to 1, $U_i \sim U(0,1)$, was generated to determine that an observation was exactly observed failure time or censored. If U_i was smaller than the proportion of exactly observed failure times in the data then let the observation be exactly observed failure time, otherwise be censored. For example, in the 5% proportion of exactly observed failure times case, $U_i \sim U(0, 1)$ was generated. If $U_i < 0.05$, then let the observation be exactly observed failure time otherwise be censored. If an observation was censored, then the number of examination times were generated from a $1 + Poisson(3)$ random variable to guarantee that the subject has at least one examination time. The time gap between two examination times was generated from a $Exponetial(3)$ random variable. The observed interval for the subject was determined by the two consecutive examination times whose interval contained T_i. If T_i was less (greater) than the smallest (largest) examination time then the lower (upper) bound of the observed interval was 0 (∞). The inverse CDF method was used to find T_i by solving $F_{T_i}(t_i|\mathbf{x}_i) = v_i$ numerically, where $v_i \sim U(0, 1)$, for $i = 1, 2, ..., n$. If an observation was exactly observed failure time, then the solution of T_i was directly used as the observed value. These distributions and their parameters were chosen in order to obtain similar amount of left-censored, interval-censored and right-censored observations .

The proposed method and a existing method were applied to the simulated data sets for comparison. The existing method by Clifford Anderson-Bergman fitted the PH model to arbitrarily-censored data, which can be applied through the '**icenReg**' package in R. The output is in Table 2.1. For the proposed method, it can be seen that firstly, the regression parameter estimates are very close to the true values of the parameters. Secondly, the sample standard deviations are very close to the average estimated standard errors, which shows that Louis's method performs well in estimating the asymptotic approximation of the variance covariance matrix. Thirdly, the coverage probabilities of 95% Wald confidence intervals cover around 95 percent of the true values indicating that 95% Wald confidence interval can be used to evaluate the estimates obtained by the developed EM algorithm. The existing method also performs well in finding regression parameter estimates, estimated standard errors and 95% coverage probabilities of the true parameters as well. By comparing these two methods, it can be found that although both methods have good performances, the bias of the estimates and the average estimated standard errors of the proposed method are uniformly smaller than the existing method. Moreover, the proposed method is 10-20 times faster the existing method. The main reason of the existing method cost more time was that it used bootstrap method to estimate standard deviations. When sample size became large, it was time consuming. In the simulation study, as the sample sizes approached 6000, the existing method tended to be much slower than the proposed method. Consequently, the propose method is more accurate and more efficient than the compare method.

2.8 Real Data Application

The real data set is from Demographic and Health Surveys in Nigeria in 2003 studying the childhood morality, children dying between the 1st and 5th birthdays. Under 5 mortality rate is a key indicator of the development for the overall child health for a country. The study seeks for the factors that effect the childhood morality such as the number of breastfeeding months, a mother's educational level, body mass index (BMI) and so on. The failure time is a child's death time. If a mother remembered the exact time of her child's death then the failure time was exactly observed. Otherwise, a mother provided a time range which her children died from so the observation was censored. The children's death rate of this data set is 0.059. This data set contains 5890 observations with 11 covariates and 2766 compete cases. Exactly observed failure times take a great proportion in the data set. The output of both methods applied on this data set is in Table 2.2.

In the output, both methods provided similar estimates for all regression param-

Table 2.1: The estimation results on the regression parameters for simulation study based on 500 replications. p denotes the proportion of exactly observed failure times, $\overline{\text{Bias}}$ denotes the empirical bias, SSD denotes the sample standard deviations of 500 point estimates, ESE denotes the average of the 500 estimated standard errors, CP95 denotes the coverage probability with 95% Wald confidence interval.

		Proposed Method				icenReg			
p	Parameters	Bias	SSD	ESE	CP95	Bias	SSD	ESE	CP95
0%	β_1 = 1	0.0240	0.1394	0.1379	0.95	0.0481	0.1455	0.1441	0.95
	β_2 = 1	0.0159	0.1511	0.1450	0.94	0.0409	0.1517	0.1500	0.95
	β_1 = 1	0.0201	0.1438	0.1401	0.94	0.0506	0.1454	0.1466	0.93
	β_2 = -1	0.0032	0.1348	0.1320	0.95	0.0477	0.1504	0.1463	0.94
	β_1 = -1	0.0160	0.1408	0.1413	0.96	0.0331	0.1474	0.1438	0.96
	β_2 = 1	0.0205	0.1512	0.1452	0.93	0.0366	0.1519	0.1557	0.94
	β_1 = -1	0.0200	0.1287	0.1260	0.95	0.0363	0.1475	0.1483	0.94
	β_2 = -1	0.0273	0.1417	0.1395	0.94	0.0366	0.1526	0.1557	0.95
5%	β_1 = 1	0.0224	0.1347	0.1360	0.95	0.0332	0.1424	0.1378	0.96
	β_2 = 1	0.0166	0.1473	0.1406	0.94	0.0292	0.1476	0.1500	0.94
	β_1 = 1	0.0043	0.1310	0.1357	0.96	0.0039	0.1312	0.1295	0.96
	β_2 = -1	0.0152	0.1317	0.1394	0.96	0.0280	0.1464	0.1489	0.96
	β_1 = -1	0.0048	0.1382	0.1388	0.96	0.0331	0.1474	0.1438	0.96
	β_2 = 1	0.0056	0.1477	0.1422	0.94	0.0366	0.1519	0.1557	0.94
	β_1 = -1	0.0076	0.1463	0.1389	0.92	0.0340	0.1416	0.1428	0.94
	β_2 = -1	0.0061	0.1422	0.1422	0.95	0.0039	0.1319	0.1264	0.97
20%	β_1 = 1	0.0102	0.1308	0.1303	0.96	0.0060	0.1290	0.1248	0.94
	β_2 = 1	0.0198	0.1450	0.1340	0.92	0.0086	0.1303	0.1350	0.94
	β_1 = 1	0.0006	0.1378	0.1301	0.94	0.0039	0.1312	0.1295	0.96
	β_2 = -1	0.0216	0.1358	0.1342	0.95	0.0280	0.1464	0.1489	0.96
	β_1 = -1	0.0146	0.1320	0.1330	0.96	0.0331	0.1474	0.1438	0.96
	β_2 = 1	0.0273	0.1459	0.1363	0.93	0.0366	0.1519	0.1557	0.94
	β_1 = -1	0.0117	0.1295	0.1328	0.95	0.0340	0.1416	0.1428	0.94
	β_2 = -1	0.0038	0.1363	0.1354	0.94	0.0039	0.1319	0.1264	0.97
50%	β_1 = 1	0.0064	0.1219	0.1240	0.94	0.0183	0.1233	0.1344	0.93
	β_2 = 1	0.0079	0.1317	0.1250	0.94	0.0138	0.1259	0.1242	0.95
	β_1 = 1	0.0112	0.1195	0.1218	0.94	0.0128	0.1235	0.1197	0.95
	β_2 = -1	0.0070	0.1220	0.1234	0.95	0.0086	0.1258	0.1219	0.95
	β_1 = -1	0.0075	0.1250	0.1229	0.96	0.0099	0.1245	0.1256	0.95
	β_2 = 1	0.0079	0.1246	0.1249	0.95	0.0104	0.1266	0.1259	0.94
	β_1 = -1	0.0192	0.1170	0.1231	0.96	0.0213	0.1243	0.1182	0.97
	β_2 = -1	0.0040	0.1196	0.1245	0.96	0.0063	0.1267	0.1208	0.97

Table 2.2: Data analysis of childhood morality data: the estimated regression parameters (PointEst), the standard error (ESE), test statistics (Z-value) and P-value.

	Proposed Method				icenReg			
Covariate	PointEst	ESE	Z-value	P-value	PointEst	ESE	Z-value	P-value
AgeInterview	0.3379	0.0637	5.3045	0.000	0.2913	0.1082	2.6930	0.007
AgeBirth	-0.3157	0.0669	-4.7190	0.000	-0.2693	0.1149	-2.3450	0.019
BMI	0.0298	0.0155	1.9226	0.055	0.0271	0.0257	1.0520	0.293
BreastfeedMonth	-0.2895	0.0148	-19.561	0.000	-0.2901	0.0265	-10.950	0.000
PrecedingInterval	-0.0045	0.0041	-1.0976	0.272	-0.0043	0.0055	-0.7718	0.440
AntenatalVisits	-0.0243	0.0183	-1.3279	0.184	-0.0242	0.0211	-1.1450	0.252
HospitalDelivery	-0.9274	0.2347	-3.9514	0.000	-0.9180	0.2753	-3.3350	0.001
Male	-0.1169	0.1560	-0.7494	0.454	-0.1174	0.1856	-0.6325	0.527
MotherEducation	-0.1220	0.1958	-0.6231	0.533	-0.1197	0.2388	-0.5014	0.616
Urban	-0.4305	0.2052	-2.0980	0.036	-0.4106	0.2327	-1.7640	0.078
State	-0.0153	0.0071	-2.1549	0.031	-0.0168	0.0099	-1.7020	0.089

eters. Since the proposed method had smaller standard deviation estimates, then it uniformly provided smaller P-value for each covariate than the existing method. Both methods identified that a mother's age during the interview, the age of birth, the number of breastfeeding months, hospital delivery or not, and locations were significant risk factors associated with children's death.

2.9 Discussion

This chapter proposes a method to analyze arbitrarily-censored data under the PH model and an EM algorithm was developed to find the MLE of the parameters. The idea of this study was inspired by the paper, Wang et al. (2016). In Wang's paper, it introduced 2-stage homogeneous Poisson random variables and developed an efficient, flexible EM algorithm to analyze interval-censored data. In this study, besides censored observations, multinomial random variables were introduced in the likelihood function to deal with exactly observed failure times in arbitrarily-censored data. The method proposed in this paper enjoys all the good properties of Wang's method and can be viewed as a more general case of Wang's study.

Chapter 3

Proportional Hazards Model for Current Status Data with Informative Censoring

3.1 Introduction

In the NTP experiments, lab rats were exposed to a substance with different levels of concentrations for a period of time to study the toxicity of the substance. During the experiments, some rats died of natural causes. Some of them were accidentally killed. Some of them were sacrificed because they were moribund during the experiments. The rest of them were sacrificed at the end of the experiments. When a rat died or was sacrificed, researchers examined the rat's organs to observe whether a tumor was found there, resulting in current status data for the tumor onset time. In such data, the failure time is the onset time of a tumor in a rat's organs. The censoring time is the death time of a rat. If a rat died from natural causes, was accidentally killed or was sacrificed at the end of the study, the censoring time can be considered to be independent of the failure time given the concentration level. However, if a rat was sacrificed because it was moribund during the experiment, then it is more reasonable to consider that the censoring time is correlated with the failure time given the concentration level. One of the assumptions made when one analyzes current status data is that the censoring time is independent of the failure time given covariates but in cases such as the NTP studies, the independent assumption fails and may lead the analysis severely away from the truth. In this chapter, a method is going to be developed to take informative censoring into consideration.

There are several existing methods analyzing current status with informative censoring. For example, Chen et al. (2012) took informative censoring into consideration in their model. They used the proportional odds model with a log-normal frailty term to characterize the correlation between the failure time and the informative censoring time. An EM-algorithm was developed to find the MLE of the parameters. The method was computationally intensive and involved approximations, such as using Monte-Carlo simulations to obtain the posterior expectations of the latent variables'. In this chapter, a new efficient and accurate method is developed to analyze the current status data with informative censoring under the Gamma-frailty PH model.

3.2 Models, Notations and Observed Likelihood

Let T denote the failure time. Assume that there are two potential censoring times C and C^*, where C is correlated with the failure time T given the covariate vector $\mathbf{x}$ and C^* is uncorrelated with the failure time T given the covariate vector $\mathbf{x}$. Under the Gamma-frailty PH model, given the frailty term η and the covariate vector $\mathbf{x}$, the hazard functions of T and C are defined as follows,

$$\lambda_T(t|\eta, \mathbf{x}) = \lambda_{0T}(t) \exp(\mathbf{x}'\boldsymbol{\beta_T})\eta,$$
$$\lambda_C(c|\eta, \mathbf{x}) = \lambda_{0C}(c) \exp(\mathbf{x}'\boldsymbol{\beta_C})\eta,$$

where $\lambda_{0T}(t)$ is the baseline hazard function for for the failure time, $\lambda_{0C}(c)$ is the baseline hazard function for for the informative censoring time, $\boldsymbol{\beta_T}$ is a vector of parameters for the failure time, $\boldsymbol{\beta_C}$ is a vector of parameters for the informative censoring time. It can be seen that T and C are correlated due to the existence of the frailty term η which is assumed to follow a gamma distribution with both the shape parameter and the rate parameter being τ. It can be seen that given η, T and C are independent.

Let $\tilde{C}$ be the smaller value between C and C^*, i.e., $\tilde{C}_i = \min(C, C^*)$. Let ξ be an indicator of informative censoring time C before (after) non-informative censoring time C^*, i.e., $\xi = I(C \leq C^*)$. Let δ be the indicator of a left- (right-) censored observation, i.e., $\delta = I(T \leq \tilde{C})$. Under these notations, what can be observed is $\{(\tilde{c}_i, \delta_i, \xi_i, \boldsymbol{x_i}), i = 1, ..., n\}$. These are independent realizations of $(\tilde{C}, \delta, \xi, \mathbf{X})$.

Let $g(\eta|\tau, \tau)$ be the pdf of gamma distribution with both the shape parameter and the rate parameter being τ so that the mean of η always equals to 1 and the value of τ only determines the variability of η. The reason of the mean designed to be 1 is to make the solution of τ identifiable. Let $S_T(\cdot|\eta)$ be the survival function for the failure time given the frailty term η, $S_C(\cdot|\eta)$ be the survival function for the informative censoring time given the frailty term η, $f_C(\cdot|\eta)$ be the pdf for the informative censoring time given the frailty term η. Under these notations, the observed likelihood function $\mathcal{L}_{obs}$ can be written as,

$$\mathcal{L}_{obs} = \prod_{i=1}^{n} \int g(\eta_i|\tau, \tau) \{1 - S_T(\tilde{c}_i|\eta_i)\}^{\delta_i} S_T(\tilde{c}_i|\eta_i)^{1-\delta_i} f_C(\tilde{c}_i|\eta_i)^{\xi_i} S_C(\tilde{c}_i|\eta_i)^{1-\xi_i} d\eta_i.$$

Note that the observed likelihood function does not contain the distribution of C^*. It is because that C^* is a non-informative censoring time so that its distribution does not have any contribution to the parameter estimates. Therefore an assumption is made that the distribution of C^* is known without interested parameters.

Under the Gamma-frailty PH model, $S_T(t|\eta) = \exp\{-\Lambda_{0T}(t) \exp(\boldsymbol{x'\beta}_T)\eta\}$ and $S_C(c|\eta) = \exp\{-\Lambda_{0C}(c) \exp(\boldsymbol{x'\beta}_C)\eta\}$, where $\Lambda_{0T}(t)$ is the baseline cumulative hazard function for the failure time and $\Lambda_{0C}(c)$ is the baseline cumulative hazard function for the informative censoring time. Additionally, the pdf of the gamma frailty term is given by $g(\eta|\tau, \tau) = \frac{\tau^\tau}{\Gamma(\tau)} \eta^{\tau-1} e^{-\tau\eta}$, where $\tau > 0$.

To model the baseline cumulative hazard functions $\Lambda_{0T}(\cdot)$, $\Lambda_{0C}(\cdot)$ and the baseline hazard functions $\lambda_{0T}(\cdot)$, $\lambda_{0C}(\cdot)$ turns out to be a challenging task because they can be infinite dimensions. According to previous work of Cai et al. (2011), McMahan et al. (2013) and Wang et al. (2016), the baseline cumulative hazard functions $\Lambda_{0T}(\cdot)$,

$\Lambda_{0C}(\cdot)$ can be modeled by I-splines as,

$$\Lambda_{0T}(\cdot) = \sum_{j=1}^{k} \gamma_{Tj} I_j(\cdot),$$
$$\Lambda_{0C}(\cdot) = \sum_{j=1}^{k} \gamma_{Cj} I_j(\cdot).$$

The first derivatives of I-splines are M-splines then the baseline hazard functions $\lambda_{0T}(\cdot)$, $\lambda_{0C}(\cdot)$ can be modeled by M-splines as follows,

$$\lambda_{0T}(\cdot) = \sum_{j=1}^{k} \gamma_{Tj} M_j(\cdot),$$
$$\lambda_{0C}(\cdot) = \sum_{j=1}^{k} \gamma_{Cj} M_j(\cdot).$$

The advantage of this method is that both $\Lambda_{0T}(\cdot)$, $\lambda_{0T}(\cdot)$ and $\Lambda_{0C}(\cdot)$, $\lambda_{0C}(\cdot)$ share the same sets of non-negative coefficients γ_{Tj}'s and γ_{Cj}'s. Furthermore, it naturally provides the relation between the baseline cumulative hazard functions and the baseline hazard functions. The technical details of applying spline functions involves specifying the degree of the basis functions and choosing an increasing sequence of knots within a certain range (Ramsay, 1988). The degree of the basis functions controls the smoothness of the basis functions and the placement of knots determines the flexibility of the basis functions. With the spline functions, after integrating out the unobserved frailty term η_i's, the observed likelihood function can be expressed as

$$\begin{aligned}
\mathcal{L}_{obs} = & \prod_{i \in A_1}^{n} \left[1 + \tau^{-1}\Lambda_{0T}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T) + \tau^{-1}\Lambda_{0C}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_C)\right]^{-\tau} \\
& \prod_{i \in A_2}^{n} \Big[\left\{1 + \tau^{-1}\Lambda_{0C}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_C)\right\}^{-\tau} - \\
& \qquad \left\{1 + \tau^{-1}\Lambda_{0T}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T) + \tau^{-1}\Lambda_{0C}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_C)\right\}^{-\tau}\Big] \\
& \prod_{i \in A_3}^{n} \tau\lambda_{0C}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_C)\left\{1 + \tau^{-1}\Lambda_{0T}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T) + \tau^{-1}\Lambda_{0C}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_C)\right\}^{-\tau} \\
& \prod_{i \in A_4}^{n} \tau\lambda_{0C}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_C)\Big[\left\{1 + \tau^{-1}\Lambda_{0C}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_C)\right\}^{-\tau} - \\
& \qquad \left\{1 + \tau^{-1}\Lambda_{0T}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T) + \tau^{-1}\Lambda_{0C}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_C)\right\}^{-\tau}\Big],
\end{aligned}$$

where $A_1 = \{\delta_i = 0, \xi_i = 0\}, A_2 = \{\delta_i = 1, \xi_i = 0\}, A_3 = \{\delta_i = 0, \xi_i = 1\}, A_4 = \{\delta_i = 1, \xi_i = 1\}$, for $i = 1, 2, ..., n$.

The parameters in the observed likelihood function is $\boldsymbol{\theta} = (\boldsymbol{\beta}'_T, \boldsymbol{\beta}'_C, \boldsymbol{\gamma}'_T, \boldsymbol{\gamma}'_C, \tau)$, where $\boldsymbol{\gamma}_T = (\gamma_{T1}, \gamma_{T2}, ..., \gamma_{Tk})', \boldsymbol{\gamma}_C = (\gamma_{C1}, \gamma_{C2}, ..., \gamma_{Ck})'$. In order to find the MLE of the vector of parameters $\boldsymbol{\theta}$ one can try to maximize $\mathcal{L}_{obs}(\boldsymbol{\theta})$ directly but $\mathcal{L}_{obs}(\boldsymbol{\theta})$ is in a very complicated form, which makes the computation difficult. It can be noticed that the difficulty of this maximization problem is caused by the summations and subtractions inside the products. Therefore an EM algorithm is going to be developed to solve this problem and find the MLE of $\boldsymbol{\theta}$.

3.3 Data Augmentation for the EM Algorithm

Since the gamma frailty term η_i's cannot be observed, then they are considered as missing data. The augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$ with the latent variables η_i's can be rewritten as

$$\mathcal{L}_1(\boldsymbol{\theta}) = \prod_{i=1}^{n} [1 - \exp\{-\Lambda_{T0}(\tilde{c}_i) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_T)\eta_i\}]^{\delta_i} \exp\{-\Lambda_{T0}(\tilde{c}_i) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_T)\eta_i\}^{1-\delta_i}$$
$$\frac{\tau^\tau}{\Gamma(\tau)} \eta_i^{\tau+\xi_i-1} e^{-\tau\eta_i} \{\lambda_{C0}(\tilde{c}_i)\}^{\xi_i} \exp\{\xi_i \boldsymbol{x}_i' \boldsymbol{\beta}_C - \Lambda_{C0}(\tilde{c}_i) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_C)\eta_i\}.$$

By integrating out η_i's from $\mathcal{L}_1(\boldsymbol{\theta})$, one can obtain the observed likelihood function $\mathcal{L}_{obs}(\boldsymbol{\theta})$. Based on the structure of the augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$, the further data augmentations are provided separately in three parts as the following,

I. The first part is $[1 - \exp\{-\Lambda_{T0}(\tilde{c}_i) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_T)\eta_i\}]^{\delta_i} \exp\{-\Lambda_{T0}(\tilde{c}_i) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_T)\eta_i\}^{1-\delta_i}$. This part owns all and only the parameters related to failure time T_i, for $i = 1, 2, ..., n$. It can be seen that the difficulty of maximization in this part is caused by the term $[1 - \exp\{-\Lambda_{T0}(\tilde{c}_i) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_T)\eta_i\}]^{\delta_i}$. Due to the previous work by Cai et al (2011), 2-stage Poisson random variables are introduced to deal with this problem. At stage

1, Poisson latent variables Z_i's are introduced as the following,

$$Z_i|\eta_i \sim Poisson\left\{\Lambda_{T0}(\tilde{c}_i)\exp\left(\boldsymbol{x_i'}\boldsymbol{\beta}_T\right)\eta_i\right\}, \quad \delta_i = 1_{(Z_i>0)}.$$

With latent variables Z_i's, the augmented likelihood function $\mathcal{L}_2(\boldsymbol{\theta})$ can be expressed as

$$\mathcal{L}_2(\boldsymbol{\theta}) = \prod_{i=1}^{n}\left[\delta_i^{1_{(Z_i>0)}}(1-\delta_i)^{1_{(Z_i=0)}}P_{Z_i}(Z_i)\right]\left\{\lambda_{C0}(\tilde{c}_i)\right\}^{\xi_i} g(\eta_i|\tau+\xi_i,\tau)$$
$$\exp\left\{\xi_i\boldsymbol{x_i'}\boldsymbol{\beta}_C - \Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)\eta_i\right\},$$

where $P_X(\cdot)$ denotes probability mass function for the random variable X, $g(\eta_i|\tau+\xi_i,\tau)$ denotes the pdf of a gamma distribution with the shape parameter $\tau+\xi_i$ and the rate parameter τ. By integrating Z_i's out of $\mathcal{L}_2(\boldsymbol{\theta})$, one can obtain $\mathcal{L}_1(\boldsymbol{\theta})$. At stage 2, for each i, the latent variable Z_i is further decomposed as a summation of k independent Poisson random variables, $Z_i = \sum_{j=1}^{k} Z_{ij}$, where the mean of Z_{ij} is $\gamma_{Tj}I_j(\tilde{c}_i)\exp\left(\boldsymbol{x_i'}\boldsymbol{\beta}_T\right)\eta_i$, for $j = 1, 2, ..., k$.

$$Z_{ij}|\eta_i \sim Poisson\left\{\gamma_{Tj}I_j(\tilde{c}_i)\exp\left(\boldsymbol{x_i'}\boldsymbol{\beta}_T\right)\eta_i\right\}, \quad j = 1, 2, ..., k,$$

with the restriction $\sum_{j=1}^{n} Z_{ij} = Z_i$.

With latent variables Z_{ij}'s, the augmented likelihood associated with stage 2 latent variables is given by,

$$\mathcal{L}_3(\boldsymbol{\theta}) = \prod_{i=1}^{n}\left[\delta_i^{1_{(\sum_{j=1}^{k} Z_{ij}>0)}}(1-\delta_i)^{1_{(\sum_{j=1}^{k} Z_{ij}=0)}}\prod_{j=1}^{k}P_{Z_{ij}}(Z_{ij})\right]\left\{\lambda_{C0}(\tilde{c}_i)\right\}^{\xi_i}$$
$$\exp\left\{\xi_i\boldsymbol{x_i'}\boldsymbol{\beta}_C - \Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)\eta_i\right\} g(\eta_i|\tau+\xi_i,\tau).$$

By integrating Z_{ij}'s out of $\mathcal{L}_3(\boldsymbol{\theta})$, one can obtain $\mathcal{L}_2(\boldsymbol{\theta})$.

II. The second part is $\{\lambda_{C0}(\tilde{c}_i)\}^{\xi_i}$. Note that $\lambda_{C0}(\tilde{c}_i)$ is modeled by M-spline functions as $\sum_{j=1}^{k}\boldsymbol{\gamma}_j M_j(\tilde{c}_i)$. Therefore, this part is presented as

$$\left\{\sum_{j=1}^{k}\boldsymbol{\gamma}_j M_j(\tilde{c}_i)\right\}^{\xi_i}.$$

When $\xi_i = 0$, this term is 1. When $\xi_i = 1$, latent multinomial random vectors V_i's are introduced as

$$V_i = (V_{i1}, V_{i2}, ..., V_{ik}) \sim Multinomial(1, [\frac{1}{k}, \frac{1}{k}, ..., \frac{1}{k}]).$$

Note that $\sum_{j=1}^{k} V_{ij} = 1$ for $i = 1, 2, ..., n$. With the latent variables V_{ij}'s, the augmented likelihood function $\mathcal{L}_4(\boldsymbol{\theta})$ can be expressed as

$$\mathcal{L}_4(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left[\delta_i^{1_{(\sum_{j=1}^{k} Z_{ij} > 0)}} (1-\delta_i)^{1_{(\sum_{j=1}^{k} Z_{ij} = 0)}} \prod_{j=1}^{k} P_{Z_{ij}}(Z_{ij}) \right] \left[\prod_{j=1}^{k} \{\gamma_{Cj} M_j(\tilde{c}_i)\}^{V_{ij}\xi_i} \right]$$
$$\exp\{\xi_i \boldsymbol{x_i'\beta}_C - \Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'\beta}_C)\eta_i\}\, g(\eta_i|\tau + \xi_i, \tau).$$

By integrating V_{ij}'s out of $\mathcal{L}_4(\boldsymbol{\theta})$, one can obtain $\mathcal{L}_3(\boldsymbol{\theta})$.

III. The last part is $\exp\{\xi_i \boldsymbol{x_i'\beta}_C - \Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'\beta}_C)\eta_i\}\, g(\eta_i|\tau + \xi_i, \tau)$. Once one takes logarithm with this part, it will be in linear form.

With the latent variables Z_i's, Z_{ij}'s, V_{ij}'s and η_i's, the augmented likelihood function $\mathcal{L}_4(\boldsymbol{\theta})$ can be written as the complete likelihood function as follows,

$$\mathcal{L}_{com}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \delta_i^{1_{(\sum_{j=1}^{k} Z_{ij} > 0)}} (1-\delta_i)^{1_{(\sum_{j=1}^{k} Z_{ij} = 0)}} \exp\{\xi_i \boldsymbol{x_i'\beta}_C - \Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'\beta}_C)\eta_i\}$$
$$g(\eta_i|\tau + \xi_i, \tau) \prod_{j=1}^{k} P_{Z_{ij}}(Z_{ij}) \{\gamma_{Cj} M_j(\tilde{c}_i)\}^{V_{ij}\xi_i}.$$

By integrating V_{ij}'s out of the complete likelihood function $\mathcal{L}_{com}(\boldsymbol{\theta})$, one can get the augmented likelihood function $\mathcal{L}_3(\boldsymbol{\theta})$. Then, by integrating Z_{ij}'s out of $\mathcal{L}_3(\boldsymbol{\theta})$, one can get the augmented likelihood function $\mathcal{L}_2(\boldsymbol{\theta})$. Further, by integrating Z_i's out of $\mathcal{L}_2(\boldsymbol{\theta})$, one can get the augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$. Finally, by integrating η_i's out of the augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$, the observed likelihood function $\mathcal{L}_{obs}(\boldsymbol{\theta})$ can be obtained. Consequently, to develop the EM-algorithm with the complete likelihood function $\mathcal{L}_{com}(\boldsymbol{\theta})$, all the latent variables Z_{ij}'s, Z_i's, V_{ij}'s and η_i's are viewed as missing data.

3.4 The EM Algorithm

3.4.1 E-step

It follows the derivation of the EM algorithm. The logarithm of the complete likelihood function is as follows,

$$
\begin{aligned}
&\log\,\{\mathcal{L}_{com}(\boldsymbol{\theta})\}\\
=&\sum_{i=1}^{n}\left[-\log\{\Gamma(\tau)\}+\tau\log\{\tau\}+(\tau-1)\log\eta_i-\tau\eta_i-\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_C)\eta_i+\xi_i\boldsymbol{x_i'}\boldsymbol{\beta}_C\right]\\
&+\sum_{i=1}^{n}\sum_{j=1}^{k}\left[-\lambda_{ij}+Z_{ij}\log\lambda_{ij}-\log Z_{ij}!+V_{ij}\xi_i\log\gamma_{Cj}+V_{ij}\xi_i\log M_j(\tilde{c}_i)+V_{ij}\xi_i\log\eta_i\right],
\end{aligned}
$$

where $\lambda_{ij}=\gamma_{Tj}I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\eta_i$. The expectation of the logarithm of the complete likelihood function with respect to the conditional expectations of Z_i's, Z_{ij}'s, V_{ij}'s, η_i's given the observed data and the current parameter $\boldsymbol{\theta}^{(d)}=(\boldsymbol{\beta}_T^{(d)'},\boldsymbol{\beta}_C^{(d)'},\boldsymbol{\gamma}_T^{(d)'},\boldsymbol{\gamma}_C^{(d)'},\tau^{(d)})'$, which yields $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})=E[\log\,\{\mathcal{L}_{com}(\boldsymbol{\theta})\}\,|\boldsymbol{x},\boldsymbol{\theta}^{(d)}]$, are provided as follows,

$$
\begin{aligned}
Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})&=E[\log\,\{\mathcal{L}_{com}(\boldsymbol{\theta})\}\,|\boldsymbol{x},\boldsymbol{\theta}^{(d)}]\\
&=H_1(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})+H_2(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})+H_3(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})+H_4(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})
\end{aligned}
$$

where,

$$
\begin{aligned}
H_1(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})&=-n\log\{\Gamma(\tau)\}+n\tau\log(\tau)+\tau\sum_{i=1}^{n}[E(\log\eta_i)-E(\eta_i)],\\
H_2(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})&=-\sum_{i=1}^{n}\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)E(\eta_i)+\sum_{i=1}^{n}\sum_{j=1}^{k}E(V_{ij})\xi_i\log\gamma_{Cj}+\sum_{i=1}^{n}\xi_i\boldsymbol{x_i'}\boldsymbol{\beta}_C,\\
H_3(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})&=-\sum_{i=1}^{n}\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)E(\eta_i)+\sum_{i=1}^{n}\sum_{j=1}^{k}E(Z_{ij})\log\gamma_{Tj}+\sum_{i=1}^{n}E(Z_i)\boldsymbol{x_i'}\boldsymbol{\beta}_T
\end{aligned}
$$

and

$H_4(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})$ is free of $\boldsymbol{\theta}$.

This completes the E-step of the EM algorithm. The reason $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})$ is written in four parts is because it will be more convenient and easier to maximize $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})$ in M-step.

3.4.2 M-step

Since $H_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ has only one the parameter τ and τ only exists in $H_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$, then it becomes a univariate maximization problem with respect to τ. The maximization problem can be solved by using constrained maximization routines (optim in R).

To find the maximization of $H_2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ and $H_3(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\beta}_C, \boldsymbol{\gamma}_C$ and $\boldsymbol{\beta}_T, \boldsymbol{\gamma}_T$, the following partial derivatives are given,

$$\frac{\partial H_2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_C} = \sum_{i=1}^{n} [-\Lambda_{C0}(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C) E(\eta_i) + \xi_i] \boldsymbol{x_i}, \tag{1}$$

$$\frac{\partial H_2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \gamma_{Cj}} = \sum_{i=1}^{n} [-\gamma_{Cj} I_j(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C) E(\eta_i) + E(V_{ij})\xi_i], \tag{2}$$

$$\frac{\partial H_3(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_T} = \sum_{i=1}^{n} [-\Lambda_{T0}(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T) E(\eta_i) + E(Z_i)] \boldsymbol{x_i}, \tag{3}$$

$$\frac{\partial H_3(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \gamma_{Tj}} = \sum_{i=1}^{n} [-I_j(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T) E(\eta_i) + E(Z_{ij})\gamma_{Tj}^{-1}], \tag{4}$$

for $j = 1, 2, ..., k$.

By setting all these four partial derivatives to zeros, the four vectors of parameters $\boldsymbol{\beta}_C$, $\boldsymbol{\beta}_T$, $\boldsymbol{\gamma}_C$ and $\boldsymbol{\gamma}_T$ can be solved with unique solutions. Firstly, solve (2), (4) for $\boldsymbol{\gamma}_C$ and $\boldsymbol{\gamma}_T$ as functions of $\boldsymbol{\beta}_C$ and $\boldsymbol{\beta}_T$ in closed forms,

$$\gamma_{Cj} = \frac{\sum_{i=1}^{n} E(V_{ij})\xi_i}{\sum_{i=1}^{n} I_j(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C) E(\eta_i)},$$
$$\gamma_{Tj} = \frac{\sum_{i=1}^{n} E(Z_{ij})}{\sum_{i=1}^{n} I_j(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T) E(\eta_i)},$$

for $j = 1, 2, ..., k$. Then plug the solutions of $\boldsymbol{\gamma}_C$ and $\boldsymbol{\gamma}_T$ into (1), (3) to solve for $\boldsymbol{\beta}_C$, $\boldsymbol{\beta}_T$. At last, plug the results of $\boldsymbol{\beta}_C$, $\boldsymbol{\beta}_T$ back into the two equations above to obtain $\boldsymbol{\gamma}_C$, $\boldsymbol{\gamma}_T$.

Note that the expectations of the latent variables' are posterior expectations given the covariate vector $\mathbf{x}$ and current parameter $\boldsymbol{\theta}^{(d)}$. The posterior expectations can be found with their posterior distributions. The complete likelihood $\mathcal{L}_{com}(\boldsymbol{\theta})$ and augmented likelihood functions, $\mathcal{L}_1(\boldsymbol{\theta})$, $\mathcal{L}_2(\boldsymbol{\theta})$, $\mathcal{L}_3(\boldsymbol{\theta})$ can be used to obtain the posterior distributions of the latent variables'. The details are in Section 3.4.3 - Section 3.4.5.

3.4.3 The latent variable $V_i = (V_{i1}, V_{i2}, ..., V_{ik})$

In the complete likelihood function $\mathcal{L}_{com}(\boldsymbol{\theta})$, only the term $\{\gamma_{Cj} M_j(\tilde{c}_i)\}^{V_{ij}\xi_i}$ contains the latent variable V_{ij}'s . Hence, the kernel of the posterior probability density function is $\{\gamma_{Cj} M_j(\tilde{c}_i)\}^{V_{ij}\xi_i}$. That is,

$$f(V_i|\boldsymbol{x}_i) \propto \{\gamma_{Cj} M_j(\tilde{c}_i)\}^{V_{ij}\xi_i}.$$

It is easy to recognize that this is the kernel of a multinomial distribution,

$$V_{i1}, V_{i2}, ..., V_{ik}|\boldsymbol{x}_i \sim$$
$$Multinomial\left[1, \left\{\frac{\gamma_{C1} M_1(\tilde{c}_i)}{\sum_{j=1}^k \gamma_{Cj} M_j(\tilde{c}_i)}, \frac{\gamma_{C2} M_2(\tilde{c}_i)}{\sum_{j=1}^k \gamma_{Cj} M_j(\tilde{c}_i)}, ..., \frac{\gamma_{Ck} M_k(\tilde{c}_i)}{\sum_{j=1}^k \gamma_{Cj} M_j(\tilde{c}_i)}\right\}\right].$$

Therefore the posterior expectation of the latent variable V_{ij} is,

$$E(V_{ij}|\boldsymbol{x}_i) = \frac{\gamma_{Cj} M_j(\tilde{c}_i)}{\sum_{l=1}^k \gamma_{Cl} M_l(\tilde{c}_i)},$$

for $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$.

3.4.4 The latent variable η_i

Note that in the augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$, when $\delta_i = 0$,

$$\mathcal{L}_1(\boldsymbol{\theta}) = \prod_{i=1}^n \frac{\tau^\tau}{\Gamma(\tau)} \eta_i^{\tau-1} e^{-\tau\eta_i} \exp\{-\Lambda_{T0}(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\eta_i\}^{1-\delta_i}$$
$$\{\lambda_{C0}(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)\eta_i\}^{\xi_i} \exp\{-\Lambda_{C0}(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)\eta_i\}.$$

Therefore,

$$f(\eta_i|\boldsymbol{x}_i) \propto \eta_i^{\tau+\xi_i-1} \exp(-b_i\eta_i),$$

where $b_i = \tau + \Lambda_{T0}(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T) + \Lambda_{C0}(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)$. This is the kernel of a gamma distribution. Hence, when $\delta_i = 0$, the posterior distribution of η_i given $\mathbf{x}$ is a gamma distribution as follows,

$$\eta_i|\boldsymbol{x}_i \sim Gamma(\xi_i + \tau, b_i),$$

Therefore, when $\delta_i = 0$,

$$E(\eta_i|\boldsymbol{x}_i) = \frac{\tau + \xi_i}{b_i}, \quad E(\log \eta_i|\boldsymbol{x}_i) = \psi(\tau + \xi_i) - \log b_i,$$

for $i = 1, 2, ..., n$, where $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$.

When $\delta_i = 1$, it is worth to notice that the distribution of η_i given $\mathbf{x}$ consists of summations of two gamma kernel functions.

$$f(\eta_i|\boldsymbol{x}_i) \propto \eta_i^{\tau+\xi_i-1} \exp(-b_i\eta_i) + \eta_i^{\tau+\xi_i-1} \exp(-d_i\eta_i),$$

where $d_i = \tau + \Lambda_{C0}(\tilde{c}_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)$. Therefore,

$$E(\eta_i|\boldsymbol{x}_i) = \frac{(\tau + \xi_i)}{d_i} \frac{1 - (\frac{d_i}{b_i})^{\tau+1+\xi_i}}{1 - (\frac{d_i}{b_i})^{\tau+\xi_i}},$$

$$E(\log \eta_i|\boldsymbol{x}_i) = \psi(\tau + \xi_i) - \frac{b_i^{\tau+\xi_i} \log d_i - d_i^{\tau+\xi_i} \log b_i}{b_i^{\tau+\xi_i} - d_i^{\tau+\xi_i}},$$

for $i = 1, 2, ..., n$. The technical details can be found in Chapter 3 Supplementary Materials. Combining these two cases $\delta_i = 0$ and $\delta_i = 1$ together, the following expectations can be obtained,

$$E(\eta_i|\boldsymbol{x}_i) = (1 - \delta_i)\frac{\tau + \xi_i}{b_i} + \delta_i \frac{(\tau + \xi_i)}{d_i} \frac{1 - (\frac{d_i}{b_i})^{\tau+1+\xi_i}}{1 - (\frac{d_i}{b_i})^{\tau+\xi_i}},$$

$$E(\log \eta_i|\boldsymbol{x}_i) = (1 - \delta_i)\{\psi(\tau + \xi_i) - \log b_i\} + \delta_i \left\{\psi(\tau + \xi_i) - \frac{b_i^{\tau+\xi_i} \log d_i - d_i^{\tau+\xi_i} \log b_i}{b_i^{\tau+\xi_i} - d_i^{\tau+\xi_i}}\right\},$$

for $i = 1, 2, ..., n$.

3.4.5 The latent variables Z_i and Z_{ij}

Since $\sum_{j=1}^{k} Z_{ij} = Z_i$, then Z_{ij} is multinomial distributed given Z_i for $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$. Therefore one can obtain the following relationship by applying the law of iterative rule.

$$E(Z_{ij}|\boldsymbol{x}_i) = E\{E(Z_{ij}|\boldsymbol{x}_i, Z_i)\} = E\left\{\frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} Z_i|\boldsymbol{x}_i\right\} = \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} E(Z_i|\boldsymbol{x}_i).$$

For the latent variable Z_i, it follows a truncated Poisson distribution given η_i with a support of all positive integers when $\delta_i = 1$ and degenerates at 0 when $\delta_i = 0$ for $i = 1, 2, ..., n$. By applying the law of iterative rule again,

$$E(Z_i|\boldsymbol{x}_i) = E\left\{E(Z_i|\boldsymbol{x}_i, \eta_i)\right\} = E\left\{\frac{\eta_i \delta_i \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)}{1-\exp\left[-\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\eta_i\right]}|\boldsymbol{x}_i\right\}.$$

Since the distribution of $\eta_i|\boldsymbol{x}_i$ can be obtained from the previous section, this expectation can be evaluated. It turns out to be,

$$E(Z_i|\boldsymbol{x}_i) = \delta_i \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\frac{\tau+\xi_i}{d_i}\frac{b_i^{\tau+\xi_i}}{b_i^{\tau+\xi_i}-d_i^{\tau+\xi_i}}.$$

Additionally,

$$\begin{aligned} E(Z_{ij}|\boldsymbol{x}_i) &= \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} E(Z_i|\boldsymbol{x}_i) = \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} \delta_i \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\frac{\tau+\xi_i}{d_i}\frac{b_i^{\tau+\xi_i}}{b_i^{\tau+\xi_i}-d_i^{\tau+\xi_i}} \\ &= \delta_i \gamma_{Tj} I(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\frac{\tau+\xi_i}{d_i}\frac{b_i^{\tau+\xi_i}}{b_i^{\tau+\xi_i}-d_i^{\tau+\xi_i}}, \end{aligned}$$

for $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$. The technical details can be found in Chapter 3 Supplementary Materials.

3.4.6 A summary of the EM algorithm

With all the results obtained from E-step and M-step, the EM algorithm can be summarized as follows,

---

Step 1. Set $d = 0$ and initial values of $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}_T^{(d)'}, \boldsymbol{\beta}_C^{(d)'}, \boldsymbol{\gamma}_T^{(d)'}, \boldsymbol{\gamma}_C^{(d)'}, \tau^{(d)})'$.

Step 2. Obtain $\tau^{(d+1)}$ by maximizing

$$-n\log\left\{\Gamma(\tau)\right\} + n\tau\log(\tau) + \tau\sum_{i=1}^{n}[E(\log\eta_i) - E(\eta_i)].$$

Step 3. Obtain $\boldsymbol{\beta}_T^{(d+1)}$ by solving the following equation,

$$\sum_{i=1}^{n}[-\sum_{j=1}^{k}\boldsymbol{\gamma}_{Tj}^{(d)}(\boldsymbol{\beta}_T)I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)E(\eta_i) + E(Z_i)]\boldsymbol{x_i} = 0,$$

where

$$\boldsymbol{\gamma}_{Tj}^{(d)}(\boldsymbol{\beta}_T) = \frac{\sum_{i=1}^{n} E(Z_{ij})}{\sum_{i=1}^{n} I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)E(\eta_i)}.$$

Obtain

$$\gamma_{Tj}^{(d+1)} = \gamma_{Tj}^{(d)}(\boldsymbol{\beta}_T^{(d+1)}).$$

Step 4. Obtain $\boldsymbol{\beta}_C^{(d+1)}$ by solving the following equation,

$$\sum_{i=1}^{n}[-\sum_{j=1}^{k}\boldsymbol{\gamma}_{Cj}^{(d)}(\boldsymbol{\beta}_C)I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)E(\eta_i) + \xi_i]\boldsymbol{x_i} = 0,$$

where

$$\boldsymbol{\gamma}_{Cj}^{(d)}(\boldsymbol{\beta}_C) = \frac{\sum_{i=1}^{n} E(V_{ij})\xi_i}{\sum_{i=1}^{n} I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)E(\eta_i)}.$$

Obtain

$$\gamma_{Cj}^{(d+1)} = \gamma_{Cj}^{(d)}(\boldsymbol{\beta}_C^{(d+1)}).$$

Step 5. Repeat step 2- 4 until $|\boldsymbol{\theta}^{(d+1)} - \boldsymbol{\theta}^{(d)}|$ is smaller than a tolerance value.

The solutions obtained by the developed EM algorithm, denoted as $\hat{\boldsymbol{\theta}}$, is the MLE of $\boldsymbol{\theta}$.

3.5 Asymptotic Properties and Variance Estimation

Under standard regularity conditions, the MLE enjoys the property of the asymptotic normality. That is, as $n \to +\infty$,

$$n^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N\left\{\mathbf{0}, I^{-1}(\boldsymbol{\theta})\right\},$$

where $I(\boldsymbol{\theta})$ is the fisher information matrix. To estimate the variance covariance matrix of $\hat{\boldsymbol{\theta}}$, Louis's method (Louis, 1982) is adopted to obtain $\hat{I}(\hat{\boldsymbol{\theta}})$ which is an estimation of $I(\boldsymbol{\theta})$.

$$\hat{I}(\hat{\boldsymbol{\theta}}) = -\frac{\partial^2 \log \mathcal{L}_{obs}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} | \boldsymbol{\theta} = \hat{\boldsymbol{\theta}},$$

where

$$-\frac{\partial^2 \log \mathcal{L}_{obs}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \text{var}\left\{ \frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\}.$$

Both terms $\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ and $\text{var}\left\{ \frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\}$ have close forms. The technical details of the calculations can be found in Chapter 3 Supplementary Materials.

3.6 Simulation Study

A series of simulations studies were conducted to evaluate the performance of the proposed method. Three scenarios were performed based on the different values of the gamma frailty parameter $\tau = 0.5$, 1, 2. The following models were considered as the true distributions for failure time T and informative censoring time C,

$$F_T(t|x_1, x_2, \eta) = 1 - \exp\left\{-\Lambda_{T0}(t) \exp(x_1 \boldsymbol{\beta}_{T1} + x_2 \boldsymbol{\beta}_{T2})\eta\right\},$$
$$F_C(c|x_1, x_2, \eta) = 1 - \exp\left\{-\Lambda_{C0}(c) \exp(x_1 \boldsymbol{\beta}_{C1} + x_2 \boldsymbol{\beta}_{C2})\eta\right\},$$

where $x_1 \sim Bernoulli(0.5)$, $x_2 \sim N(0, 0.5^2)$, $\eta \sim Gamma(\tau, \tau)$, $\Lambda_{T0}(t) = \log(1+t) + t^2$, $\Lambda_{C0}(c) = \log(1+c)$. The non-informative censoring time C^* followed a truncated exponential distribution with mean 1 and upper bound 10. The regression parameters were specified as $\boldsymbol{\beta}_{T1} = \boldsymbol{\beta}_{C1} \in \{-1, 1\}$ and $\boldsymbol{\beta}_{T2} = \boldsymbol{\beta}_{C2} \in \{-1, 1\}$. To obtain the failure time T and informative censoring time C, the inverse CDF method was applied to solve the following equations numerically, $F_T(t|x_1, x_2, \eta) = u, F_C(c|x_1, x_2, \eta) = v$, where $u \sim U(0, 1)$ and $v \sim U(0, 1)$. For each case of the simulation, 500 data sets were generated with sample sizes $n = 500$.

Table 3.1: Under three different true values of gamma frailty parameter $\tau = 0.5$, 1, 2, the proposed method was applied to estimate the regression parameters of the failure time, the regression parameters of the informative censoring time and the gamma frailty parameter. The summary includes the average 500 estimates bias (Bias) and the sample standard deviation (SSD), the average estimated standard error (ESE) and the empirical 95% Wald confidence interval coverage probabilities (CP95) for all the parameters

	$\tau = 0.5$				$\tau = 1$				$\tau = 2$			
	$\overline{\text{Bias}}$	SSD	ESE	CP95	$\overline{\text{Bias}}$	SSD	ESE	CP95	$\overline{\text{Bias}}$	SSD	ESE	CP95
$\beta_{T1} = 1$	0.033	0.244	0.240	94%	0.021	0.190	0.188	94%	0.011	0.130	0.151	94%
$\beta_{T2} = 1$	0.037	0.259	0.242	93%	0.011	0.189	0.190	96%	0.018	0.165	0.160	95%
$\beta_{C1} = 1$	0.045	0.285	0.277	97%	0.030	0.211	0.222	92%	0.023	0.220	0.211	94%
$\beta_{C2} = 1$	0.063	0.300	0.291	95%	0.035	0.247	0.240	94%	0.030	0.229	0.231	94%
$\hat{\tau}$	0.003	0.108	0.110	93%	0.013	0.328	0.299	91%	0.235	1.021	0.983	86%
$\beta_{T1} = -1$	0.044	0.265	0.248	94%	0.043	0.257	0.256	95%	0.006	0.179	0.178	94%
$\beta_{T2} = 1$	0.037	0.253	0.251	95%	0.039	0.250	0.249	95%	0.015	0.183	0.180	94%
$\beta_{C1} = -1$	0.055	0.314	0.298	95%	0.063	0.320	0.300	94%	0.016	0.231	0.229	93%
$\beta_{C2} = 1$	0.045	0.315	0.299	96%	0.043	0.315	0.298	95%	0.037	0.248	0.239	94%
$\hat{\tau}$	0.010	0.145	0.137	91%	0.013	0.148	0.141	91%	0.420	1.973	1.884	87%
$\beta_{T1} = 1$	0.041	0.223	0.220	95%	0.038	0.197	0.181	94%	0.010	0.167	0.160	96%
$\beta_{T2} = -1$	0.031	0.230	0.232	95%	0.008	0.193	0.192	95%	0.020	0.176	0.173	95%
$\beta_{C1} = 1$	0.045	0.258	0.266	94%	0.045	0.266	0.265	95%	0.022	0.225	0.203	92%
$\beta_{C2} = -1$	0.055	0.276	0.280	96%	0.031	0.240	0.238	95%	0.037	0.212	0.211	94%
$\hat{\tau}$	0.000	0.120	0.104	88%	0.032	0.309	0.281	86%	0.333	2.052	1.521	86%
$\beta_{T1} = -1$	0.037	0.253	0.247	96%	0.001	0.201	0.200	92%	0.011	0.179	0.182	97%
$\beta_{T2} = -1$	0.019	0.261	0.259	96%	0.016	0.200	0.203	96%	0.030	0.189	0.180	94%
$\beta_{C1} = -1$	0.050	0.298	0.293	94%	0.022	0.260	0.254	94%	0.010	0.214	0.218	96%
$\beta_{C2} = -1$	0.037	0.307	0.305	96%	0.025	0.269	0.267	95%	0.041	0.259	0.244	93%
$\hat{\tau}$	0.002	0.121	0.130	89%	0.058	0.441	0.453	91%	0.918	4.831	2.511	86%

For the spline functions, the degrees of the basis splines were set to be 3 and 5 inner knots were placed with equal space in the interval between the minimum observed time and maximum observed time. The choice of number of knots for a single data set can be determined by using criteria such as AIC. In the EM algorithm the initial values of the parameters in the spline functions were set to be 0.5 and the initial values of regression parameters were set to be 0.

The simulation study results are in Table 3.1. In the table, it can be found that firstly, all the regression parameter estimates are very close to the true values of the regression parameters. Secondly, the averaged standard errors of the 500 estimates agree with the sample standard deviation. It indicates that the Louis's

Table 3.2: Results of estimated regression parameters with three different methods, Proposed Method, the PH Model and the marginal GORH model Under three different true values of gamma frailty parameter $\tau = 0.5,\ 1,\ 2:$ The summary includes the average 500 estimates bias (Bias), the sample standard deviation (SSD), the average estimated standard error (ESE) and the empirical 95% Wald confidence interval coverage probabilities (CP95).

		Proposed Method				PH Model				Marginal GORH			
β_T		$\overline{\text{Bias}}$	SSD	ESE	CP95	$\overline{\text{Bias}}$	SSD	ESE	CP95	$\overline{\text{Bias}}$	SSD	ESE	CP95
$\tau = 0.5$													
1	1	0.033	0.244	0.240	94%	0.519	0.172	0.179	16%	0.284	0.256	0.250	78%
		0.037	0.259	0.242	93%	0.499	0.170	0.178	20%	0.257	0.266	0.260	83%
-1	1	0.044	0.265	0.248	94%	0.464	0.197	0.201	35%	0.261	0.272	0.264	83%
		0.037	0.253	0.251	95%	0.471	0.194	0.200	34%	0.270	0.274	0.268	83%
1	-1	0.041	0.223	0.220	95%	0.515	0.176	0.173	17%	0.278	0.252	0.244	80%
		0.031	0.230	0.232	95%	0.516	0.173	0.182	16%	0.279	0.261	0.256	81%
-1	-1	0.037	0.253	0.247	96%	0.484	0.197	0.190	31%	0.283	0.269	0.263	81%
		0.025	0.257	0.252	95%	0.478	0.195	0.204	31%	0.273	0.274	0.277	80%
$\tau = 1$													
1	1	0.021	0.190	0.188	94%	0.438	0.159	0.160	23%	0.276	0.208	0.205	73%
		0.011	0.189	0.190	96%	0.428	0.164	0.170	27%	0.262	0.217	0.214	78%
-1	1	0.043	0.257	0.256	95%	0.381	0.183	0.177	44%	0.213	0.232	0.218	85%
		0.039	0.250	0.249	95%	0.381	0.186	0.194	45%	0.211	0.235	0.243	83%
1	-1	0.038	0.197	0.181	94%	0.431	0.159	0.172	26%	0.266	0.208	0.220	72%
		0.008	0.193	0.192	95%	0.437	0.166	0.170	24%	0.278	0.217	0.211	77%
-1	-1	0.001	0.201	0.200	92%	0.388	0.187	0.193	45%	0.223	0.233	0.236	81%
		0.016	0.200	0.203	96%	0.381	0.185	0.196	48%	0.218	0.234	0.243	81%
$\tau = 2$													
1	1	0.011	0.130	0.151	94%	0.320	0.156	0.165	45%	0.196	0.188	0.191	80%
		0.018	0.165	0.160	95%	0.309	0.165	0.160	52%	0.184	0.196	0.190	83%
-1	1	0.006	0.179	0.178	94%	0.268	0.182	0.190	64%	0.152	0.210	0.216	88%
		0.015	0.183	0.180	94%	0.251	0.184	0.191	71%	0.132	0.213	0.218	91%
1	-1	0.010	0.167	0.160	96%	0.334	0.157	0.164	41%	0.207	0.187	0.196	78%
		0.020	0.176	0.173	95%	0.316	0.165	0.172	49%	0.190	0.195	0.203	81%
-1	-1	0.011	0.179	0.182	97%	0.259	0.179	0.178	69%	0.146	0.209	0.201	89%
		0.030	0.189	0.180	94%	0.242	0.184	0.202	72%	0.123	0.213	0.230	88%

method performs well in estimating the standard error with a finite sample size, $n = 500$. Thirdly, the empirical 95% Wald confidence intervals for all the regression parameters cover 93% - 97% of the true values, which suggests that Wald confidence intervals can be used as an inference method to evaluate the performance of the developed EM algorithm. Lastly, it can be seen that the estimates of the parameter τ are close to the true values. As τ gets larger, the bias of the estimates becomes higher. The 95% Wald confidence intervals cover around 90% of the true values. This is caused by the lack of information for the parameter τ in the data but it does not effect the performance of the estimates of the regression parameters which are the most interested parameters in the study.

For the purpose of comparison, two other commonly used models in literature for the analysis of current status data, the PH model and the generalized odds rate hazards (GORH) model, were applied to the same data sets. The PH model is one of the most widely used model to analyze current status data so that it can be considered as a benchmark. It assumes the censoring time is independent of the failure time given covariates. A method developed by McMahan et al. (2013) fitted the PH model to current status data. The method can be implemented via the R package by McMahan and Wang. The results are shown in the Table 3.2 called the PH model. All the regression parameter estimates are far from the true values and the 95% Wald confidence intervals have low coverage probabilities of the true values. The coverage probabilities increase as τ gets larger because the correlation between the informative censoring time and the failure time becomes weaker when τ goes up. Hence, the ignorance of the correlation between the censoring time and the failure time can lead to large errors of the parameter estimates.

The other model is the GORH model which is the marginal model of the Gamma-frailty PH model. The GORH model is an appropriate model here because the true data were generated from the Gamma-frailty PH model. From the output of the marginal GORH column in Table 3.2, it can be seen that the bias of regression parameters is around 20% and the coverage probabilities are near 90%. Although it performs much better than the PH model, the results are still not satisfactory. This is because that even though the correct model is applied, it only utilizes the information from the marginal distribution, the distribution of the failure time T, but lack of the information from the distribution of the informative censoring time C.

Table 3.3: Regression parameter estimates with the PH model: The data were generated from the Gamma-frailty PH model with the parameter $\tau = 10$, 100. The summary includes the average 500 estimates bias (Bias) and the sample standard deviation (SSD), the average estimated standard error (ESE) and the empirical 95% Wald confidence interval coverage probabilities (CP95) for all the parameters.

		$\hat{\beta}_T$			
τ	β_T	$\overline{\text{Bias}}$	SSD	ESE	CP95
10	-1	0.0921	0.1631	0.1700	89%
	1	0.0995	0.1645	0.1730	88%
	1	0.1155	0.1651	0.1639	90%
	1	0.1178	0.1647	0.1654	88%
100	-1	0.0268	0.1699	0.1618	96%
	1	0.0166	0.1708	0.1752	95%
	1	0.0241	0.1731	0.1760	95%
	1	0.0205	0.1727	0.1768	93%

From Table 3.2, in the output of the PH model column, it can be seen that as the frailty parameter gets larger, the bias of the estimates gets smaller. This is because the distribution of the frailty random variable has the expectation 1 and the variance τ^{-1}. As τ increases, the variance goes down and the correlation between the failure time and the informative censoring time becomes weaker. When τ approaches infinity the distribution degenerates at 1, leading to no correlation between the failure time and the informative censoring time. Besides, the Gamma-frailty PH model has a explicit criteria to quantify the statistical association called Kendall's τ, $\tau = (1+\tau)^{-1}$. Consequentially, for a large value of τ, the PH model should be very close to the Gamma-frailty PH model. In Table 3.3, other two simulations were made with larger values of τ, $\tau = 10$ and $\tau = 100$, to demonstrate this situation. The true regression parameters were chosen to be $\{-1, 1\}$. The true model was the Gamma-frailty PH model but the PH model was fitted to the data. The output is in Table 3.3. When $\tau = 10$, comparing to the output of the PH model in Table 3.2, the bias of the estimates decreases from around 50% to around 10% and the empirical 95% Wald confidence interval coverage probabilities increase from around 10% to around 90%. Furthermore, as $\tau = 100$, the PH model performs as well as the proposed method.

3.7 Real Data Application

All three methods were applied to two real data sets from the NTP, $\boldsymbol{tr-486}$ and $\boldsymbol{tr-467}$.

3.7.1 $\boldsymbol{tr-467}$

In the manufacture of neoprene, chloroprene, the 2-chloro analogue of 1,3-butadiene, a potent, multi-species, multi-organ carcinogen, is only used but with high production and not much information about its carcinogenic potential (from the report of NTP). The NTP, an inter-agency program whose mission is to evaluate agents of public health concern by developing and applying tools of modern toxicology and molecular biology (from NTP website), performed studies about toxicology and carcinogens of chloroprene and provided a report in September 1998. In the 2-years mice study, groups of 50 male and 50 female mice were exposed to chloroprene at concentrations of control (0 ppm), low dose (12.8ppm), medium dose (32ppm), or high dose (80 ppm) by inhalation, 6 hours per day, 5 days per week, for 2 years. The mice were removed from the study because of accidentally kill, natural death, terminal sacrifice or moribund sacrifice after certain amount of days to be observed for whether Alveolar/Bronchiolar Adenoma was in their organs. it is natural to consider that given the level of concentration, natural death, accidentally kill and terminal sacrifice were not related to the onset time of Alveolar/Bronchiolar Adenoma in mice's organs while moribund sacrifice was. Hence accidentally kill, natural death and terminal sacrifice are considered as non-informative censoring while moribund sacrifice is considered as as informative censoring. In this analysis, we focused on whether chloroprene was associated with the onset time of Alveolar/Bronchiolar Adenoma in mice's lungs.

Additionally, the association between chloroprene and the informative censoring time was another interesting aspect to study. The concentration levels of chloroprene were treated as factors because it was not tested continuously.

From the output in Table 3.4, there are several things worth to be noticed. Firstly, all three concentration levels of chloroprene had significant effects on the onset time of Alveolar/Bronchiolar Adenoma in mice's lungs. Secondly, for the same concentration level of chloroprene, gender did not have a significant effect on the onset time of Alveolar/Bronchiolar Adenoma in mice's lungs but it did on the mice's moribundity. It indicates that if both male and female mice had Alveolar/Bronchiolar Adenoma in their lungs, female mice had higher probability to be moribund than male mice. Thirdly, all of the methods provided very close estimates. This is because the correlation between the failure time and the informative censoring time was weak due to large value of τ. This real data application shows that the proposed method works as good as the other two methods when the correlation between failure time and the informative censoring time is weak.

Table 3.4: $\boldsymbol{tr-467}$ data analysis: The summary includes the estimated regression parameters (Est), the estimated standard error (SE) and P-value. The estimated gamma frailty parameter τ is 13.856 with estimated standard error 24.429.

	Proposed Method						PH Model			Marginal GORH		
	$\hat{\beta}_T$			$\hat{\beta}_C$			$\hat{\beta}_T$			$\hat{\beta}_T$		
	Est	SE	P-value	Est	SE	P-value	Est	SE	P-value	Est	SE	P-value
Low	1.551	0.243	<0.01	0.664	0.165	<0.01	1.509	0.182	<0.01	1.844	0.260	<0.01
Medium	2.177	0.179	<0.01	1.718	0.115	<0.01	2.071	0.117	<0.01	2.647	0.265	<0.01
High	2.270	0.186	<0.01	1.889	0.103	<0.01	2.168	0.133	<0.01	2.775	0.260	<0.01
Gender	-0.192	0.183	0.148	-0.827	0.151	<0.01	-0.125	0.175	0.237	0.033	0.270	0.451

3.7.2 $\boldsymbol{tr-486}$

Isoprene was evaluated for toxicity in this study because its structure is similar to 1,3 -butadiene and a large amount of production with potential exposure to human in this study. This analysis focused whether Isoprene was associated with the onset time of Leukemia Mononuclear in livers of female rats after a two-year period study with four different exposure levels, control (0 ppm), low dose (220 ppm), medium dose (700 ppm) , high dose (7000 ppm).

Table 3.5: $\boldsymbol{tr-486}$ data analysis: The summary includes the estimated regression parameters (Est), the estimated standard error (SE) and P-value. The estimated τ, gamma frailty parameter, is 0.572 with standard error 0.382.

	Proposed Method						PH Model			Marginal GORH		
	$\hat{\boldsymbol{\beta}}_T$			$\hat{\boldsymbol{\beta}}_C$			$\hat{\boldsymbol{\beta}}_T$			$\hat{\boldsymbol{\beta}}_T$		
	Est	SE	P-value	Est	SE	P-value	Est	SE	P-value	Est	SE	P-value
Low	-0.022	0.575	0.515	-0.242	0.331	<0.01	0.070	0.382	0.427	0.123	0.159	0.439
Medium	0.744	0.466	0.055	0.164	0.284	<0.01	0.529	0.337	0.058	0.770	0.137	<0.01
High	0.462	0.301	0.063	0.604	0.253	<0.01	0.337	0.201	0.017	0.496	0.205	0.016

In Table 3.5, the results of all methods indicate that the low dose (220 ppm) exposure level of Isoprene did not have significant effects on the onset time of Leukemia Mononuclear in female rats' livers. But when the exposure level increased to medium dose (700 ppm) or high dose (7000 ppm), the effect became significant. Although all of the methods suggested the same trend, the proposed method and the GORH model provided higher estimates of the regression parameters than the PH model. Therefore, if the correlation between the informative censoring time and the failure time was ignored (in the PH model), then the estimated effect might be lower than the truth.

3.7.3 Data application summary

In these two data applications, the proposed method provided similar estimates of regression parameters as the other two methods in $\boldsymbol{tr-467}$ study but gave different ones in $\boldsymbol{tr-486}$ study. That is because the correlation between the failure time and the informative censoring time was strong in $\boldsymbol{tr-486}$ study but weak in $\boldsymbol{tr-467}$

study, which can be seen from the gamma frailty parameter estimates of these two data sets with the proposed method. Therefore, when the failure time and the informative censoring time are independent or the correlation is weak given covariates, the proposed method can provide almost the same parameter estimates as the other methods. When the failure time and the informative censoring time are not independent given covariates, it can capture the correlation and make better estimations. Moreover, the proposed method also provides the estimates of informative censored parameters, which shows the relation between the covariates and the informative censoring time.

3.8 Discussion

In previous literature, for the analysis of current status data, either all the censored observations are considered to be non-informative or the developed methods involve approximations. This chapter develops a new method to analyze the current status data with informative censoring under the Gamma-frailty PH model. The proposed method is efficient, accurate and easy to apply.

Chapter 4

Statistical Analysis of System Reliability for Current Status Data with the PH Model

4.1 Introduction

In NTP studies, the data obtained are current status data. The analysis for current status data in Chapter 3 focused on whether an experimented substance was harmful to one specific organ of a lab mouse, such as a liver or a lung. However, the mission of the NTP is to evaluate whether a substance is harmful to a lab mouse, not just to one specific organ. One can consider that as long as a substance is harmful to one organ of a lab mouse, then it is harmful to the mouse. In this situation, a lab mouse can be considered as a system, where its organs are components of the system. As long as one of the components fails, the system fails. The analysis needs to focus on the reliability of the system.

To perform statistical analysis of system reliability, one way is to estimate the survival function of a system. The survival function of a system can be estimated by using system data or component data. One can fit the PH model to system data to estimate the system survival function directly or fit the PH model to component data to estimate the system survival function under certain assumptions. The advantage of the first strategy is that it needs less data, less assumptions and fewer estimations. The advantage of the second strategy is that the analysis uses more information which may increase the accuracy of estimations, and it can provide more information about how each component effects the reliability of a system. In a system, some of the components may be more important to the system than the others. For example, in NTP studies, a tumor in the heart of a lab mouse is much more lethal than a tumor in the liver of a lab mouse. Therefore, to analyze the reliability of a system through each component can help us understand the reliability of a system more deeply. In this chapter, several methods are developed to estimate the survival function of a system with system data and component data. The methods are compared with each other and the best strategy for analyzing system reliability is discussed.

The structure of this chapter is as follows. In Section 4.2, the notations are introduced. In Section 4.3, the PH model is fitted to system data and a method is developed by McMahan et al. (2013) is used to estimate the survival function of the system. In Section 4.4, the PH model is fitted to the data of each component. Under the assumption that all components of a system are independent from each other, the survival function of a system is estimated. In Section 4.5, all components of a system are assumed to be correlated with each other. The Gamma-frailty PH model is fitted to the data of all components. A method is developed to estimate the survival function of the system. In Section 4.6, several simulations are made to evaluate the performance of the three methods. In Section 4.7, these methods are applied to a real data set form the NTP. Since most data sets from the NTP are in the same structure, then these methods can be widely applied to them.

4.2 Notations and Assumptions

Let T and C denote the failure time and the censoring time of a system. Let δ be an indicator of left (right) censored observation, i.e., $\delta = I(T < C)$. Assume that the system consists of k components. For the jth component of the system, let T_j, C_j and δ_j denote the failure time, the censoring time and the indicator of left (right) censored observation, i.e., $\delta_j = I(T_j < C_j)$, for $j = 1, 2, ..., k$. Let $S(t|\mathbf{x})$ be the survival function of a system given the covariate $\mathbf{x}$ and $S_j(t|\mathbf{x})$ be the survival function of each component given the covariate $\mathbf{x}$. Let $\Lambda(t|\mathbf{x})$ be the cumulative hazard function of a system given the covariate $\mathbf{x}$ and $\Lambda_j(t|\mathbf{x})$ be the cumulative hazard function of each component given the covariate $\mathbf{x}$. Let $\Lambda_0(t|\mathbf{x})$ be the baseline cumulative hazard function of a system given the covariate $\mathbf{x}$ and $\Lambda_{0j}(t|\mathbf{x})$ be the baseline cumulative hazard function of each component given the covariate $\mathbf{x}$.

We consider the following situations. Firstly, all the components are censored at the same time so that $C_1 = C_2 = ... = C_k = C$. It is because in NTP studies, once a mouse was censored then all its organs were censored at the same time. More generally, if a system fails down, then one will check all its components for the problems. Secondly, as long as one component fails, the system fails. That is, $T = \min\{T_1, T_2, ..., T_k\}$. It is because in NTP studies, it is reasonable to think that as long as a substance is harmful to one organ of a mouse, the substance is harmful to the mouse.

4.3 The PH Model with System Data

Under the notations in Section 4.2, what can be observed for system data in a sample with n observations is $\{(c_i, \delta_i, \boldsymbol{x_i}), i = 1, ..., n\}$. These are independent realizations of $\{(C, \delta, \mathbf{X})\}$.

4.3.1 The observed likelihood function, the augmented likelihood function and the complete likelihood function

Under the PH model, the survival function of the system failure time T can be written as

$$S(t|\mathbf{x}) = \exp\{-\Lambda_0(t)\exp(\boldsymbol{x'\beta})\}.$$

With the observed data, the observed likelihood function can be written as

$$\begin{aligned}\mathcal{L}_{obs} &= \prod_{i=1}^{n} \{1 - S(c_i)\}^{\delta_i} S(c_i)^{1-\delta_i} \\ &= \prod_{i=1}^{n} [1 - \exp\{-\Lambda_0(c_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta})\}]^{\delta_i} \exp\{-\Lambda_0(c_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta})\}^{1-\delta_i}.\end{aligned}$$

The cumulative baseline function $\Lambda_0(\cdot)$ is modelled with I-splines as

$$\Lambda_0(\cdot) = \sum_{l=1}^{m} \gamma_l I_l(\cdot),$$

where γ_l's are non-negative coefficients and $I_l(\cdot)$'s are basis functions. With the splines, the observed likelihood function can be further written as

$$\mathcal{L}_{obs} = \prod_{i=1}^{n} \left[1 - \exp\left\{-\sum_{l=1}^{m} \gamma_l I_l(c_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta})\right\}\right]^{\delta_i} \exp\left\{-\sum_{l=1}^{m} \gamma_l I_l(c_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta})\right\}^{1-\delta_i}.$$

The parameters in the observed likelihood function are coefficients of the spline functions γ_l's and the regression parameter vector $\boldsymbol{\beta}$. Let $\boldsymbol{\theta} = (\gamma_1, \gamma_2, ..., \gamma_m, \boldsymbol{\beta})$. An EM algorithm, first developed by McMahan et al. (2013), is used find the MLE of the parameters.

4.3.2 Data Augmentation and the EM Algorithm

The data argumentation starts with 2-stage latent variables. At stage 1, Poisson latent variables Z_i's are introduced as the following,

$$Z_i \sim Poisson\left\{\sum_{l=1}^{m} \gamma_l I_l(c_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta})\right\}, \quad \delta_i = 1_{(Z_i>0)}.$$

With latent variables Z_i's, the augmented likelihood function $\mathcal{L}_2(\boldsymbol{\theta})$ can be expressed as

$$\mathcal{L}_1(\boldsymbol{\theta}) = \prod_{i=1}^{n} \delta_i^{1_{(Z_i>0)}}(1-\delta_i)^{1_{(Z_i=0)}} P_{Z_i}(Z_i),$$

where $P_X(\cdot)$ denotes probability mass function for the random variable X. By integrating Z_i's out of $\mathcal{L}_1(\boldsymbol{\theta})$, one can obtain $\mathcal{L}_{obs}(\boldsymbol{\theta})$. At stage 2, for each i, the latent variable Z_i is further decomposed as a summation of m independent Poisson random variables, $Z_i = \sum_{l=1}^{m} Z_{il}$, where the mean of Z_{il} is $\gamma_l I_l(c_i)\exp(\boldsymbol{x_i'\beta})$, for $l = 1, 2, ..., m$.

$$Z_{il} \sim Poisson\left\{\gamma_l I_l(c_i)\exp(\boldsymbol{x_i'\beta})\right\}, \quad l = 1, 2, ..., m,$$

with the restriction $\sum_{l=1}^{m} Z_{il} = Z_i$.

With latent variables Z_{il}'s, the complete likelihood associated with stage 2 latent variables is given by,

$$\mathcal{L}_{com}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \delta_i^{1\left(\sum_{l=1}^{m} Z_{il}>0\right)} (1-\delta_i)^{1\left(\sum_{l=1}^{m} Z_{il}=0\right)} \prod_{l=1}^{m} P_{Z_{il}}(Z_{il}).$$

In the complete likelihood function $\mathcal{L}_{com}(\boldsymbol{\theta})$, the latent variables Z_i's, Z_{il}'s are treated as missing data. It can be seen that by integrating out Z_{ij}'s in $\mathcal{L}_{com}(\boldsymbol{\theta})$, one can get the augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$. Then, by integrating out Z_i's in $\mathcal{L}_1(\boldsymbol{\theta})$, one can obtain the observed likelihood function $\mathcal{L}_{obs}(\boldsymbol{\theta})$. Consequently, $\mathcal{L}_{com}(\boldsymbol{\theta})$ is viewed as the complete data likelihood with Z_i's and Z_{il}'s missing.

4.3.3 The EM Algorithm

The derivation of the EM algorithm starts with E-step. In E-step, one needs to take the expectation of $\mathcal{L}_{com}(\boldsymbol{\theta})$ with respect to all the latent variables Z_i's, Z_{il}'s given the observed data and the current parameter $\boldsymbol{\theta}^{(d)} = (\gamma_1^{(d)}, \gamma_2^{(d)}, ..., \gamma_m^{(d)}, \boldsymbol{\beta}^{(d)})$, which yields the Q function.

$$\log\left[\mathcal{L}_{com}(\boldsymbol{\theta})\right] = \sum_{i=1}^{n}\sum_{l=1}^{m} \log \frac{\exp(-\lambda_{il})\lambda_{il}^{Z_{il}}}{Z_{il}!} = \sum_{i=1}^{n}\sum_{l=1}^{m}\left[(-\lambda_{il}) + Z_{il}\log\lambda_{il} - \log Z_{il}!\right],$$

where $\lambda_{il} = \gamma_l I_l(c_i)\exp(\boldsymbol{x_i'\beta})$, for $i = 1, 2, ...n$, $l = 1, 2, ..., m$.

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = E[\log\left\{\mathcal{L}_{com}(\boldsymbol{\theta})\right\}|\boldsymbol{x}, \boldsymbol{\theta}^{(d)}] = \sum_{i=1}^{n}\sum_{l=1}^{q}\left[-\lambda_{il} + E(Z_{il})\log\lambda_{il} - E(\log Z_{il}!)\right]$$

$$= \sum_{i=1}^{n}\sum_{l=1}^{m}\left[-\gamma_l I_l(c_i)\exp(\boldsymbol{x_i'\beta}) + E(Z_{il})\left\{\log\gamma_l + \boldsymbol{x_i'\beta}\right\} + \log\left\{I_l(c_i)\right\} - E(\log Z_i!)\right],$$

where $E(Z_i) = \frac{\delta_i \sum_{l=1}^{m}\gamma_l^{(d)} I_l(c_i)\exp(\boldsymbol{x_i'\beta}^{(d)})}{1-\exp\left\{-\sum_{l=1}^{m}\gamma_l^{(d)} I_l(c_i)\exp(\boldsymbol{x_i'\beta}^{(d)})\right\}}$, $E(Z_{il}) = \frac{\gamma_l^{(d)} I_l(c_i)E(Z_i)}{\sum_{l'=1}^{m}\gamma_{l'}^{(d)} I_{l'}(c_i)}$. Note that all expectations of the latent variables' are conditional expectations given the observed data and current parameters. They can be obtain with the augmented function and complete likelihood function.

In the next step, M-step, one needs solve for the parameter $\boldsymbol{\theta}$ by maximizing $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$. Firstly, take partial derivatives of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\theta}$.

$$\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \sum_{l=1}^{m} \left[-\gamma_l I_l(c_i) \exp(\boldsymbol{x_i'\beta}) + E(Z_{il})\right] \boldsymbol{x_i},$$
$$\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \gamma_l} = \sum_{i=1}^{n} \left[-I_l(c_i) \exp(\boldsymbol{x_i'\beta}) + E(Z_{il})\gamma_l^{-1}\right],$$

for $l = 1, 2, ..., m$. Then, set these partial derivatives to zeros. For the 2nd equation, γ_l's can be solved as a function of $\boldsymbol{\beta}$ in close forms as follows,

$$\gamma_l = \frac{\sum_{i=1}^{n} E(Z_{il})}{\sum_{i=1}^{n} I_l(t_i) \exp(\boldsymbol{x_i'\beta})}, \text{ for } l = 1, 2, ..., m.$$

Thirdly, plug the solution of γ_l's back into the 1st equation so that the 1st equation only has the parameter $\boldsymbol{\beta}$. Numerical methods can be used to solve for $\boldsymbol{\beta}$. Lastly, by plugging the solution of $\boldsymbol{\beta}$ into the equation above, γ_l's can be obtained.

With the results from E-step and M-step, the EM algorithm can be summarized as follows,

---

Step 1. Set $d = 0$ and initial values of $\boldsymbol{\theta}^{(d)}$.

Step 2. Obtain $\boldsymbol{\beta}^{(d+1)}$ by solving the following equations,

$$\sum_{i=1}^{n} \left[-\sum_{l=1}^{m} \gamma_l^{(d)}(\boldsymbol{\beta}) I_l(c_i) \exp(\boldsymbol{x_i'\beta}_j) + E(Z_{ij})\right] \boldsymbol{x_i} = 0,$$

where

$$\gamma_l^{(d)} = \frac{\sum_{i=1}^{n} E(Z_{ijl})}{\sum_{i=1}^{n} I_l(c_i) \exp(\boldsymbol{x_i'\beta})}.$$

Step 3. Obtain

$$\boldsymbol{\gamma}_l^{(d+1)} = \boldsymbol{\gamma}_l^{(d)}(\boldsymbol{\beta}^{(d+1)}).$$

Step 4. Repeat steps 2- 3 until convergence.

The solutions obtained by the developed EM algorithm, denoted as $\hat{\boldsymbol{\theta}}$, is the MLE of $\boldsymbol{\theta}$. Therefore, the estimated survival function of the system $\hat{S}_{sys}(t|\mathbf{x})$ is as follows,

$$\hat{S}_{sys}(t|\mathbf{x}) = \exp\left\{-\sum_{l=1}^{m} \hat{\gamma}_l I_l(t) \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}})\right\}.$$

4.4 The PH Models with Component Data

Under the notations in Section 4.2, what can be observed for component data in a sample with n observations is $\{(c_i, \delta_{ij}, \boldsymbol{x_i}), j = 1, 2, ...k, i = 1, ..., n\}$. These are independent realizations of $\{(C, \delta_j, \mathbf{X}), j = 1, 2, ...k\}$. With an assumption that all components of a system are independent from each other, the survival function of the system failure time T can be written as

$$S(t|\mathbf{x}) = \prod_{j=1}^{k} S_j(t|\mathbf{x}).$$

Under the PH model, the survival function of each component time T_j can be expressed as

$$S_j(t|\mathbf{x}) = \exp\{-\Lambda_{0j}(t) \exp(\boldsymbol{x'}\boldsymbol{\beta}_j)\}, \text{ for } j = 1, 2, ..., k.$$

The survival function of the system can be expressed as

$$S(t|\mathbf{x}) = \exp\left\{-\sum_{j=1}^{k} \Lambda_{0j}(t) \exp(\boldsymbol{x'}\boldsymbol{\beta}_j)\right\},$$

where $\boldsymbol{\beta}_j$'s are the regression parameters for each component. The baseline cumulative hazard function $\Lambda_{0j}(\cdot)$ is modelled by I-splines for $j = 1, 2, ..., k$ as follows,

$$\Lambda_{0j}(\cdot) = \sum_{l=1}^{m} \gamma_{jl} I_{jl}(\cdot),$$

where γ_{jl}'s are non-negative parameters and $I_{jl}(\cdot)$'s are basis functions. Then the observed likelihood function can be written as

$$\begin{aligned}\mathcal{L}_{obs} &= \prod_{i=1}^{n}\prod_{j=1}^{k}\{1-S_j(c_i|x)\}^{\delta_{ij}}\, S_j(c_i|x)^{1-\delta_{ij}}\\ &= \prod_{j=1}^{k}\mathcal{L}_{obs}^{j},\end{aligned}$$

where $\mathcal{L}_{obs}^{j} = \prod_{i=1}^{n}\{1-S_j(c_i|x)\}^{\delta_{ij}}\, S_j(c_i|x)^{1-\delta_{ij}}$, for $j = 1, 2, ..., k$. Since all components are independent from each other, then maximizing $\mathcal{L}_{obs}$ is the same as maximizing all $\mathcal{L}_{obs}^{j}$'s separately. It can be seen that each $\mathcal{L}_{obs}^{j}$ is the same as the observed likelihood function of the PH model with system data. Therefore same method, the EM algorithm, can be used to find the MLE of the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', ..., \boldsymbol{\gamma}_k')$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', ..., \boldsymbol{\beta}_k')$, $\boldsymbol{\gamma}_j = (\boldsymbol{\gamma}_{j1}', \boldsymbol{\gamma}_{j2}', ..., \boldsymbol{\gamma}_{jm}')$ for $j = 1, 2, ..., k$. The current parameter in the EM algorithm is denoted as $\boldsymbol{\theta}^{(d)}$, where $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)\prime}, \boldsymbol{\gamma}_1^{(d)\prime}, \boldsymbol{\gamma}_2^{(d)\prime}, ..., \boldsymbol{\gamma}_k^{(d)\prime})$, $\boldsymbol{\beta}^{(d)} = (\boldsymbol{\beta}_1^{(d)\prime}, \boldsymbol{\beta}_2^{(d)\prime}, ..., \boldsymbol{\beta}_k^{(d)\prime})$, $\boldsymbol{\gamma}_j^{(d)} = (\boldsymbol{\gamma}_{j1}^{(d)\prime}, \boldsymbol{\gamma}_{j2}^{(d)\prime}, ..., \boldsymbol{\gamma}_{jm}^{(d)\prime})$ for $j = 1, 2, ..., k$.

4.4.1 A summary of the EM algorithm

The EM algorithm can be summarized as follows,

Step 1. Set $d = 0$ and initial values of $\boldsymbol{\theta}^{(d)}$.

Step 2. Obtain $\boldsymbol{\beta}_j^{(d+1)}$ by solving the following equations,

$$\sum_{i=1}^{n}\left[-\sum_{l=1}^{p}\gamma_{jl}^{(d)}(\boldsymbol{\beta}_j)I_{jl}(t_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_j) + E(Z_{ij})\right]\boldsymbol{x_i} = 0, \text{ for } j = 1, 2, ..., k,$$

where

$$\gamma_{jl}^{(d)} = \frac{\sum_{i=1}^{n}E(Z_{ijl})}{\sum_{i=1}^{n}I_{jl}(t_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_j)}.$$

Step 3. Obtain

$$\boldsymbol{\gamma}_{jl}^{(d+1)} = \boldsymbol{\gamma}_{jl}^{(d)}(\boldsymbol{\beta}_j^{(d+1)}), \text{ for } j = 1, 2, ..., k, \quad l = 1, 2, ..., m.$$

Step 4. Repeat steps 2- 3 until convergence.

The solutions obtained by the developed EM algorithm, denoted as $\hat{\boldsymbol{\theta}}$, is the MLE of $\boldsymbol{\theta}$. Therefore, the estimated survival function of the system $\hat{S}_{ind}(t|\mathbf{x})$ is as follows,

$$\hat{S}_{ind}(t|\mathbf{x}) = \exp\left\{-\sum_{j=1}^{k}\sum_{l=1}^{m}\hat{\gamma}_{jl}I_{jl}(t)\exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}}_j)\right\}.$$

4.5 The Gamma-frailty PH Model with Component Data

Under the notations Section 4.2, what can be observed for component data in a sample with n observations is $\{(c_i, \delta_{ij}, \boldsymbol{x_i}), j = 1, 2, ...k, i = 1, ..., n\}$. These are independent realizations of $\{(C, \delta_j, \mathbf{X}), j = 1, 2, ...k\}$. Under the Gamma-frailty PH model, the survival function of the failure time for the j the component T_j given the frailty term η and the covariate $\mathbf{x}$ can be written as

$$S_j(t|\boldsymbol{x}, \eta) = \exp\left\{-\eta\Lambda_{0j}(t)\exp(\boldsymbol{x}'\boldsymbol{\beta}_j)\right\}, \text{ for } j = 1, 2, ..., k,$$

where $\eta \sim Ga(\tau, \tau)$, $\tau > 0$. The cumulative baseline function $\Lambda_0(\cdot)$ is modelled with I-splines as follows,

$$\Lambda_{0j}(\cdot) = \sum_{l=1}^{m}\gamma_{jl}I_{jl}(\cdot),$$

where γ_{jl}'s are non-negative coefficients and $I_{jl}(\cdot)$'s are basis functions. The survival function of the system failure time T can be written as

$$\begin{aligned}
S(t|\mathbf{x}) &= P(T > t) \\
&= P(T_1 > t, T_2 > t, ..., T_k > t|\mathbf{x}) \\
&= \int P(T_1 > t, T_2 > t, ..., T_k > t|\mathbf{x}, \eta)g(\eta|\tau, \tau)d\eta \\
&= \int P(T_1 > t|\mathbf{x}, \eta)P(T_2 > t|\mathbf{x}, \eta)...P(T_k > t|\mathbf{x}, \eta)g(\eta|\tau, \tau)d\eta \\
&= \int g(\eta|\tau, \tau)\prod_{j=1}^{k}S_j(t|\mathbf{x}, \eta)d\eta.
\end{aligned}$$

By integrating η out, the survival function of the system can be obtained as follows,

$$S(t|\mathbf{x}) = \int \frac{\tau^\tau}{\Gamma(\tau)} \eta^{\tau-1} e^{-\tau\eta} \prod_{j=1}^{k} \exp\left\{-\eta \Lambda_{0j}(t) \exp(\boldsymbol{x'}\boldsymbol{\beta}_j)\right\} d\eta$$

$$= \left\{ \tau^{-1} \sum_{j=1}^{k} \sum_{l=1}^{m} \gamma_{jl} I_{jl}(t) \exp(\boldsymbol{x'}\boldsymbol{\beta}_j) + 1 \right\}^{-\tau}.$$

The parameters are $\tau, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_k, \gamma_{11}, \gamma_{12}, ..., \gamma_{km}$.

4.5.1 The Conditional Likelihood and the Observed Likelihood

For the ith observation, the conditional likelihood given η_i is as follows,

$$\mathcal{L}_i(\boldsymbol{\gamma}, \boldsymbol{\beta}, \tau|\eta_i) = \prod_{j \in L_i^C} S_j(c_i|\mathbf{x}_i, \eta_i) \prod_{j \in L_i} \left\{1 - S_j(c_i|\mathbf{x}_i, \eta_i)\right\}$$

$$= \exp(-\eta_i \sum_{j \in L_i^C} H_{ij}) \prod_{j \in L_i} \left\{1 - S_j(c_i|\mathbf{x}_i, \eta_i)\right\},$$

where $H_{ij} = \Lambda_{0j}(c_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)$, $L_i = \{j \in \{1, 2, ..., k\} : \delta_{ij} = 1\}$, L_i^C is the complement of L_i and the complete set is $\{1, 2, ..., k\}$.

To write the conditional likelihood in a general form, let $\mathcal{B}(L_i)$ be the set containing all subsets of L_i and A_{ip} be the pth element in $\mathcal{B}(L_i)$, for $p = 1, 2, ..., 2^{d_i}$, where d_i is the number of elements in L_i, i.e, $\mathcal{B}(L_i) = \{A_{i1}, A_{i2}, ..., A_{i2^{d_i}}\}$. Then, the conditional likelihood function can be written as

$$\mathcal{L}_i(\boldsymbol{\gamma}, \boldsymbol{\beta}, \tau|\eta_i) = \sum_{p=1}^{2^{d_i}} (-1)^{|L_i| - |A_{ip}|} \exp(-\eta_i \sum_{j \in A_{ip}^C} H_{ij}),$$

where A_{ip}^C 's are the complement sets of A_{ip}'s with the complete set being $\{1, 2, ..., k\}$. If $A_{ip}^C = \emptyset$, define $\sum_{j \in A_{ip}^C} H_{ij} = 0$.

For example, assume that there is a system with 5 components. The ith observation is $\{c_i, \mathbf{x}_i, \delta_{i1} = 1, \delta_{i2} = 1, \delta_{i3} = 1, \delta_{i4} = 0, \delta_{i5} = 0\}$. For this observation, $L_i = \{1, 2, 3\}$, $L_i^C = \{4, 5\}$, $d_i = 3$ and $2^{d_i} = 8$. The set, $\mathcal{B}(L_i)$, containing all subsets of L_i can be presented as follows,

$$\mathcal{B}(L_i) = \mathcal{B}(\{1, 2, 3\}) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

The elements in $\mathcal{B}(L_i)$ are as follows,

$$A_{i1} = \emptyset, A_{i2} = \{1\}, A_{i3} = \{2\}, A_{i4} = \{3\}, A_{i5} = \{1,2\},$$
$$A_{i6} = \{1,3\}, A_{i7} = \{2,3\}, A_{i8} = \{1,2,3\}.$$

Since the complete set is $\{1,2,3,4,5\}$, then the complement sets of A_{ip}'s are as follows,

$$A_{i1}^C = \{1,2,3,4,5\}, A_{i2}^C = \{2,3,4,5\}, A_{i3}^C = \{1,3,4,5\}, A_{i4}^C = \{1,2,4,5\},$$
$$A_{i5}^C = \{3,4,5\}, A_{i6}^C = \{2,4,5\}, A_{i7}^C = \{1,4,5\}, A_{i8}^C = \{4,5\}.$$

Additionally,

$$|L_i| - |A_{i1}| = 3, |L_i| - |A_{i2}| = 2, |L_i| - |A_{i3}| = 2, |L_i| - |A_{i4}| = 2,$$
$$|L_i| - |A_{i5}| = 1, |L_i| - |A_{i6}| = 1, |L_i| - |A_{i7}| = 1, |L_i| - |A_{i8}| = 0.$$

Then, the conditionally likelihood for this observation can be written as

$$\begin{aligned}
\mathcal{L}_i(\boldsymbol{\gamma}, \boldsymbol{\beta}, \tau|\eta_i) &= \sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} \exp(-\eta_i \sum_{j \in A_{ip}^C} H_{ij}) \\
&= (-1)^3 \exp(-\eta_i \sum_{j\in\{1,2,3,4,5\}} H_{ij}) + (-1)^2 \exp(-\eta_i \sum_{j\in\{2,3,4,5\}} H_{ij}) + \\
&\quad (-1)^2 \exp(-\eta_i \sum_{j\in\{1,3,4,5\}} H_{ij}) + (-1)^2 \exp(-\eta_i \sum_{j\in\{1,2,4,5\}} H_{ij}) + \\
&\quad (-1)^1 \exp(-\eta_i \sum_{j\in\{3,4,5\}} H_{ij}) + (-1)^1 \exp(-\eta_i \sum_{j\in\{2,4,5\}} H_{ij}) + \\
&\quad (-1)^1 \exp(-\eta_i \sum_{j\in\{1,4,5\}} H_{ij}) + (-1)^0 \exp(-\eta_i \sum_{j\in\{4,5\}} H_{ij}) \\
&= -\exp(-\eta_i \sum_{j\in\{1,2,3,4,5\}} H_{ij}) + \exp(-\eta_i \sum_{j\in\{2,3,4,5\}} H_{ij}) + \\
&\quad \exp(-\eta_i \sum_{j\in\{1,3,4,5\}} H_{ij}) + \exp(-\eta_i \sum_{j\in\{1,2,4,5\}} H_{ij}) - \\
&\quad \exp(-\eta_i \sum_{j\in\{3,4,5\}} H_{ij}) - \exp(-\eta_i \sum_{j\in\{2,4,5\}} H_{ij}) - \\
&\quad \exp(-\eta_i \sum_{j\in\{1,4,5\}} H_{ij}) + \exp(-\eta_i \sum_{j\in\{4,5\}} H_{ij}),
\end{aligned}$$

where $H_{ij} = \Lambda_{0j}(c_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)$. The distribution of η_i is $g(\eta_i|\tau) = \frac{\tau^\tau}{\Gamma(\tau)}\eta_i^{\tau-1}e^{-\tau\eta_i}$. Therefore, the observed likelihood for the ith observation can be written as

$$\begin{aligned}
\mathcal{L}_i(\boldsymbol{\gamma},\boldsymbol{\beta},\tau) &= \int \frac{\tau^\tau}{\Gamma(\tau)}\eta_i^{\tau-1}e^{-\tau\eta_i}\sum_{p=1}^{2^{d_i}}(-1)^{|L_i|-|A_{ip}|}\exp(-\eta_i\sum_{j\in A_{ip}^C}H_{ij})d\eta_i \\
&= \sum_{p=1}^{2^{d_i}}(-1)^{|L_i|-|A_{ip}|}\int \frac{\tau^\tau}{\Gamma(\tau)}\eta_i^{\tau-1}\exp\left\{-(\tau+\sum_{j\in A_{ip}^C}H_{ij})\eta_i\right\}d\eta_i \\
&= \sum_{p=1}^{2^{d_i}}\frac{(-1)^{|L_i|-|A_{ip}|}\tau^\tau}{(\tau+\sum_{j\in A_{ip}^C}H_{ij})^\tau} \\
&= \sum_{p=1}^{2^{d_i}}(-1)^{|L_i|-|A_{ip}|}(1+\tau^{-1}\sum_{j\in A_{ip}^C}H_{ij})^{-\tau}.
\end{aligned}$$

Hence the observed likelihood function for all observations can be written as

$$\begin{aligned}
\mathcal{L}_{obs}(\boldsymbol{\gamma},\boldsymbol{\beta},\tau) &= \prod_{i=1}^{n}\mathcal{L}_i(\boldsymbol{\gamma},\boldsymbol{\beta},\tau) \\
&= \prod_{i=1}^{n}\left\{\sum_{p=1}^{2^{d_i}}(-1)^{|L_i|-|A_{ip}|}(1+\tau^{-1}\sum_{j\in A_{ip}^C}H_{ij})^{-\tau}\right\}.
\end{aligned}$$

Since the observed likelihood function is in a complex form, an EM algorithm is developed to find the MLE of the parameter $\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\tau,\boldsymbol{\beta}',\boldsymbol{\gamma}')$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,...,\boldsymbol{\beta}_k)'$, $\boldsymbol{\gamma} = (\gamma_{11},\gamma_{12},...,\gamma_{km})'$. The derivation of the algorithm is based on the following data augmentation.

4.5.2 Data Augmentation for the EM Algorithm

Since the gamma frailty term η_i's can not be observed, then they are considered as missing data. The augmented likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$ with the latent variables η_i's can be rewritten as

$$\begin{aligned}
\mathcal{L}_1(\boldsymbol{\theta}) &= \prod_{i=1}^{n}g(\eta_i|\tau,\tau)\prod_{j=1}^{k}\{1-S_j(c_i|x)\}^{\delta_{ij}}\,S_j(c_i|x)^{1-\delta_{ij}}d\eta_i \\
&= \prod_{i=1}^{n}g(\eta_i|\tau,\tau)\prod_{j=1}^{k}\left[1-\exp\{-\eta_i\Lambda_{0j}(c_i)\exp(\boldsymbol{x}'\boldsymbol{\beta}_j)\}\right]^{\delta_{ij}} \\
&\qquad \left[\exp\{-\eta_i\Lambda_{0j}(c_i)\exp(\boldsymbol{x}'\boldsymbol{\beta}_j)\}\right]^{1-\delta_{ij}},
\end{aligned}$$

where $g(\eta_i|\tau,\tau)$ is the pdf of the gamma distribution with both shape and rate parameters being τ. By integrating out η_i's from $\mathcal{L}_1(\boldsymbol{\theta})$, one can obtain the observed likelihood function $\mathcal{L}_{obs}(\boldsymbol{\theta})$. To maximize $\mathcal{L}_1(\boldsymbol{\theta})$, due to the previous works by Cai et al. (2011) and McMahan et al. (2013), 2-stage Poisson random variables can be introduced to facilitate the computation. At stage 1, Poisson latent variables Z_{ij}'s are introduced as the following,

$$Z_{ij} \sim Poisson\left\{\Lambda_{0j}(c_i)\exp(\boldsymbol{x}'\boldsymbol{\beta}_j)\eta_i\right\}, \quad \delta_i = 1_{(Z_{ij}>0)}.$$

With the latent variables Z_{ij}'s, the augmented likelihood functions can be written as

$$\mathcal{L}_2(\boldsymbol{\theta}) = \prod_{i=1}^{n} g(\eta_i|\tau,\tau)\prod_{j=1}^{k}\delta_i^{1(Z_{ij}>0)}(1-\delta_i)^{1(Z_{ij}=0)}P_{Z_{ij}}(Z_{ij}),$$

where $P_X(\cdot)$ denotes probability mass function for the random variable X. By integrating Z_{ij}'s out of $\mathcal{L}_2(\boldsymbol{\theta})$, one can obtain $\mathcal{L}_1(\boldsymbol{\theta})$. At stage 2, for each latent variable Z_{ij}, it can be further decomposed as a summation of m independent Poisson random variables, $Z_{ij} = \sum_{l=1}^{m} Z_{ijl}$, where the mean of Z_{ijl} is $\gamma_{jl}I_{jl}(c_i)\exp(\boldsymbol{x}'\boldsymbol{\beta}_j)\eta_i$, for $l = 1, 2, ..., m$.

$$Z_{ijl} \sim Poisson\left\{\gamma_{jl}I_{jl}(c_i)\exp(\boldsymbol{x}'\boldsymbol{\beta}_j)\eta_i\right\}, \quad l = 1, 2, ..., m,$$

with restriction $Z_{ij} = \sum_{l=1}^{m} Z_{ijl}$. With the latent variables Z_{ijl}'s, the complete likelihood function can be written as

$$\mathcal{L}_{com}(\boldsymbol{\theta}) = \prod_{i=1}^{n} g(\eta_i|\tau,\tau)\prod_{j=1}^{k}\delta_i^{1(Z_{ij}>0)}(1-\delta_i)^{1(Z_{ij}=0)}1_{(\sum_{l=1}^{m} Z_{ijl}=Z_{ij})}\prod_{l=1}^{m} P_{Z_{ijl}}(Z_{ijl}).$$

By integrating Z_{ijl}'s out of $\mathcal{L}_{com}(\boldsymbol{\theta})$, one can obtain $\mathcal{L}_2(\boldsymbol{\theta})$. Then, integrating Z_{ij}'s out of $\mathcal{L}_2(\boldsymbol{\theta})$, one can obtain $\mathcal{L}_1(\boldsymbol{\theta})$. Then, integrating η_i's out of $\mathcal{L}_1(\boldsymbol{\theta})$, one can obtain the observed likelihood function $\mathcal{L}_{obs}(\boldsymbol{\theta})$. Consequently, $\mathcal{L}_{com}(\boldsymbol{\theta})$ is viewed as the complete likelihood function and η_i's, Z_{ij}'s, Z_{ijl}'s are missing data.

4.5.3 The EM Algorithm

It follows the E-step of the EM algorithm. The E-step is to find the $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ function, the expectation of the logarithm of the complete likelihood function with respect to all the latent variables given the covariate $\boldsymbol{X}$ and the current parameters $\boldsymbol{\theta}^{(d)}$.

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = E[\log\{\mathcal{L}_{com}(\boldsymbol{\theta})\} | \boldsymbol{x}, \boldsymbol{\theta}^{(d)}] = H_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) + H_2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) + H_3(\boldsymbol{\theta}^{(d)}),$$

where

$$H_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = \sum_{i=1}^{n} [-\log\{\Gamma(\tau)\} + \tau \log\{\tau\} + (\tau - 1)E(\log \eta_i) - \tau E(\eta_i)]$$

$$H_2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = \sum_{i=1}^{n} \sum_{j=1}^{k} \sum_{l=1}^{m} [-E(\eta_i) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_j) \gamma_{jl} I_{jl}(c_i) + E(Z_{ijl}) \boldsymbol{x}_i' \boldsymbol{\beta}_j + E(Z_{ijl}) \log(\gamma_{jl})] .$$

and $H_3(\boldsymbol{\theta}^{(d)})$ is free of $\boldsymbol{\theta}$.

In the next step, M-step, one needs solve for the parameter $\boldsymbol{\theta}$ by maximizing $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$. Since $H_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ has only one the parameter τ and τ only exists in $H_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$, it becomes a univariate maximization problem with respect to τ. The maximization problem can be solved by using constrained maximization routines (optim in R).

To maximize $H_2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\beta}_j$'s, $\boldsymbol{\gamma}_{jl}$'s , the following partial derivatives are given,

$$\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{n} [-E(\eta_i) \Lambda_{0j}(c_i) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_j) + E(Z_{ij})] \boldsymbol{x}_i', \quad (1)$$

$$\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \gamma_{jl}} = \sum_{i=1}^{n} \left[-E(\eta_i) I_{jl}(c_i) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_j) + E(Z_{ijl}) \gamma_{jl}^{-1}\right] . \quad (2)$$

Set these partial derivatives to zeros and solve (2) for γ_{jl},

$$\gamma_{jl} = \frac{\sum_{i=1}^{n} E(Z_{ijl})}{\sum_{i=1}^{n} E(\eta_i) I_{jl}(c_i) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_j)}, \quad (3)$$

for $j = 1, 2, ..., k$ and $l = 1, 2, ..., m$. Then plug the solution of γ_{jl}'s into (1) and solve the following equation numerically for $\boldsymbol{\beta}_j$'s,

$$\sum_{i=1}^{n}\left[-\sum_{l=1}^{p}\gamma_{jl}(\boldsymbol{\beta}_j)E(\eta_i)I_{jl}(c_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_j)+E(Z_{ij})\right]\boldsymbol{x_i'}=0.$$

Finally, plug the solution of $\boldsymbol{\beta}_j$'s to (3) to calculate γ_{jl}'s. The solution of $\boldsymbol{\theta}$ can be obtained.

Note that all the expectations of the latent variables are conditional expectation given data and the current parameter $\boldsymbol{\theta}^{(d)}$. Using the augmented likelihood function and the complete likelihood function the expectations can be obtain as the following,

$$E(Z_{ij}|\boldsymbol{x})=\delta_{ij}H_{ij}\frac{\sum_{q=1}^{2^{d_i-1}}(-1)^{|L_i|-|B_{iq}|-1}(1+\tau^{-1}\sum_{r\in B_{iq}^C}H_{ir})^{-\tau-1}}{\sum_{p=1}^{2^{d_i}}(-1)^{|L_i|-|A_{ip}|}(1+\tau^{-1}\sum_{r\in A_{ip}^C}H_{ir})^{-\tau}},$$

A_{ip} is the pth element of the set containing all subsets of L_i, A_{ip}^C is the complement of A_{ip} with the complete set being $\{1, 2, .., k\}$, for $p = 1, 2, ..., 2^{d_i}$ and d_i is the number of elements in L_i. B_{iq} is the qth element of the set containing all subsets of $L_i \setminus \{j\}$, B_{iq}^C is the complement of B_{iq} with the complete set being $\{1, 2, .., k\} \setminus \{j\}$, for $q = 1, 2, ..., 2^{d_i-1}$.

$$E(Z_{ijl}|\boldsymbol{x})=\frac{\gamma_{jl}I_{jl}(c_i)}{\Lambda_{0j}(c_i)}E(Z_{ij}|\boldsymbol{x}),$$

for $j = 1, 2, ..., k$ and $l = 1, 2, ..., m$.

$$E(\eta_i|\boldsymbol{x})=\frac{\sum_{p=1}^{2^{d_i}}(-1)^{|L_i|-|A_{ip}|}(1+\tau^{-1}\sum_{j\in A_{ip}^C}H_{ij})^{-\tau-1}}{\sum_{p=1}^{2^{d_i}}(-1)^{|L_i|-|A_{ip}|}(1+\tau^{-1}\sum_{j\in A_{ip}^C}H_{ij})^{-\tau}},$$

$$E(\log\eta_i|\boldsymbol{x})=\frac{\sum_{p=1}^{2^{d_i}}(-1)^{|L_i|-|A_{ip}|}(1+\tau^{-1}\sum_{j\in A_{ip}^C}H_{ij})^{-\tau}\left[\psi(\tau)-\log(\tau+\sum_{j\in A_{ip}^C}H_{ij})\right]}{\sum_{p=1}^{2^{d_i}}(-1)^{|L_i|-|A_{ip}|}(1+\tau^{-1}\sum_{j\in A_{ip}^C}H_{ij})^{-\tau}},$$

where $\psi(\tau) = \frac{\Gamma'(\tau)}{\Gamma(\tau)}$.

4.5.4 A summary of the EM algorithm

With all conditional expectations of the latent variables and the results from E-step and M-step, the EM algorithm can be summarized as follows,

---

Step 1. Set $d = 0$ and initial values of $\boldsymbol{\theta}^{(d)}$.

Step 2. Obtain $\tau^{(d+1)}$ by maximizing,

$$-n \log \{\Gamma(\tau)\} + n\tau \log(\tau) + \tau \sum_{i=1}^{n} [E(\log \eta_i) - E(\eta_i)].$$

Step 3. Obtain $\boldsymbol{\beta}_j^{(d+1)}$ by solving the following equations,

$$\sum_{i=1}^{n} \left[-\sum_{l=1}^{p} \gamma_{jl}^{(d)}(\boldsymbol{\beta}_j) I_{jl}(c_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_j) + E(Z_{ij}) \right] \boldsymbol{x_i} = 0, \text{ for } j = 1, 2, ..., k,$$

where

$$\gamma_{jl}^{(d)}(\boldsymbol{\beta}_j) = \frac{\sum_{i=1}^{n} E(Z_{ijl})}{\sum_{i=1}^{n} I_{jl}(c_i) \exp(\boldsymbol{x_i'}\boldsymbol{\beta}_j)}.$$

Step 4. Obtain

$$\boldsymbol{\gamma}_{jl}^{(d+1)} = \boldsymbol{\gamma}_{jl}^{(d)}(\boldsymbol{\beta}_j^{(d+1)}), \text{ for } j = 1, 2, ..., k, \quad l = 1, 2, ..., m.$$

Step 5. Repeat step 2- 4 until convergence.

The solutions obtained by the developed EM algorithm, denoted as $\hat{\boldsymbol{\theta}}$, is the MLE of $\boldsymbol{\theta}$. Therefore, the estimated survival function of the system $\hat{S}_{dep}(t|\mathbf{x})$ is as follows,

$$\hat{S}_{dep}(t|\mathbf{x}) = \left\{ \hat{\tau}^{-1} \sum_{j=1}^{k} \sum_{l=1}^{m} \hat{\gamma}_{jl} I_{jl}(t) \exp(\boldsymbol{x'}\hat{\boldsymbol{\beta}}_j) + 1 \right\}^{-\hat{\tau}}.$$

This section essentially develops a method for fitting the Gamma-frailty PH model to multivariate current status data. In previous literature, there was a method developed by Wang et al. (2015) fitting the Gamma-frailty PH model to bivariate current status data, which could only analyze bivariate current status data. Both methods used the EM algorithm to find the MLE of the parameters in the model. For the EM algorithm method developed by Wang et al. (2015), when the dimension of data is higher than two, the conditional expectations in the EM algorithm becomes difficult to calculate due to its complex form. The method developed in this section finds a way to write the conditional expectations in a general form so that one can still use the EM algorithm to find the MLE of the parameters in the model in multidimensional cases.

4.6 Simulation Study

A series of simulation studies were conducted to evaluate the three performance of the methods under two scenarios. In Scenario I, the failure times of all components in a system were considered to be independent from each other. In Scenario II, all components of a system were considered to be correlated with each other. For both scenarios, all three methods were applied to simulated data to estimate the survival function of a system.

4.6.1 Scenario I

The simulation is based on the following true distributions of the component failure times T_j's in a system and T_j's are independent from each other,

$$F_{T_j}(t|\mathbf{x}) = 1 - \exp\left\{-\Lambda_{0j}(t)\exp(x_1\beta_{j1} + x_2\beta_{j2})\right\}, \text{ for } j = 1, 2, 3, 4,$$

where $\mathbf{x} = (x_1, x_2)'$, $\Lambda_{01}(t) = \log(t+1) + t^{1/2}$, $\Lambda_{02}(t) = \log(t+1) + t$, $\Lambda_{03}(t) = \log(t+1) + t^{3/2}$, $\Lambda_{04}(t) = \log(t+1) + t^{5/2}$, $x_1 \sim Bernoulli(0.5)$ and $x_2 \sim N(0, 0.5^2)$. The sample size n was chosen to be 200 with 500 replications and $\boldsymbol{\beta}_{11} = \boldsymbol{\beta}_{12} = 0.5$, $\boldsymbol{\beta}_{21} = \boldsymbol{\beta}_{22} = 0.7$, $\boldsymbol{\beta}_{31} = \boldsymbol{\beta}_{32} = 1$, $\boldsymbol{\beta}_{41} = \boldsymbol{\beta}_{42} = 1$ were considered. The inverse CDF method was used to compute T_j's by solving $F_{T_j}(t|\mathbf{x}_i) = u_j$ numerically, where $u_j \sim U(0,1)$, for $j = 1,2,3,4$. The censoring time C followed a truncated exponential distribution with mean 0.1 and upper bound 2. The true distribution of the system failure time T is $\min\{T_j : j = 1, .., 4\}$.

For the plots of the estimated survival functions in Figure 4.1 and Figure 4.2, the survival function of the system estimated by the independent PH model with component data and the Gamma-frailty PH model with component data were very close to each other so that they were plotted as one dotted line marked as 'Component models'. The survival function of the system estimated by the PH model with system data was marked as 'System model'. It can be seen that the survival functions estimated by the PH model with system data were farther from the truth than other two methods and had the largest MSE in all cases. Especially, when the covariates took negative values. Therefore, the simulation study shows that in some cases, directly fitting the PH model to system data may lead to biased estimation of the system reliability function.

4.6.2 Scenario II

The simulation is based on the following true distributions of the component failure times T_j's in a system,

$$F_{T_j}(t|\mathbf{x}, \eta) = 1 - \exp\left\{-\Lambda_{0j}(t)\exp(x_1\beta_{j1} + x_2\beta_{j2})\eta\right\}, \text{for } j = 1,2,3,4,$$

where $\mathbf{x} = (x_1, x_2)'$, $\Lambda_{01}(t) = \log(t+1) + t^{1/2}$, $\Lambda_{02}(t) = \log(t+1) + t$, $\Lambda_{03}(t) = \log(t+1) + t^{3/2}$, $\Lambda_{04}(t) = \log(t+1) + t^{5/2}$, $x_1 \sim Bernoulli(0.5)$ and $x_2 \sim N(0, 0.5^2)$. The sample size n was chosen to be 250 with 500 replications and $\boldsymbol{\beta}_{j1} = 0.7, \boldsymbol{\beta}_{j2} = 1$

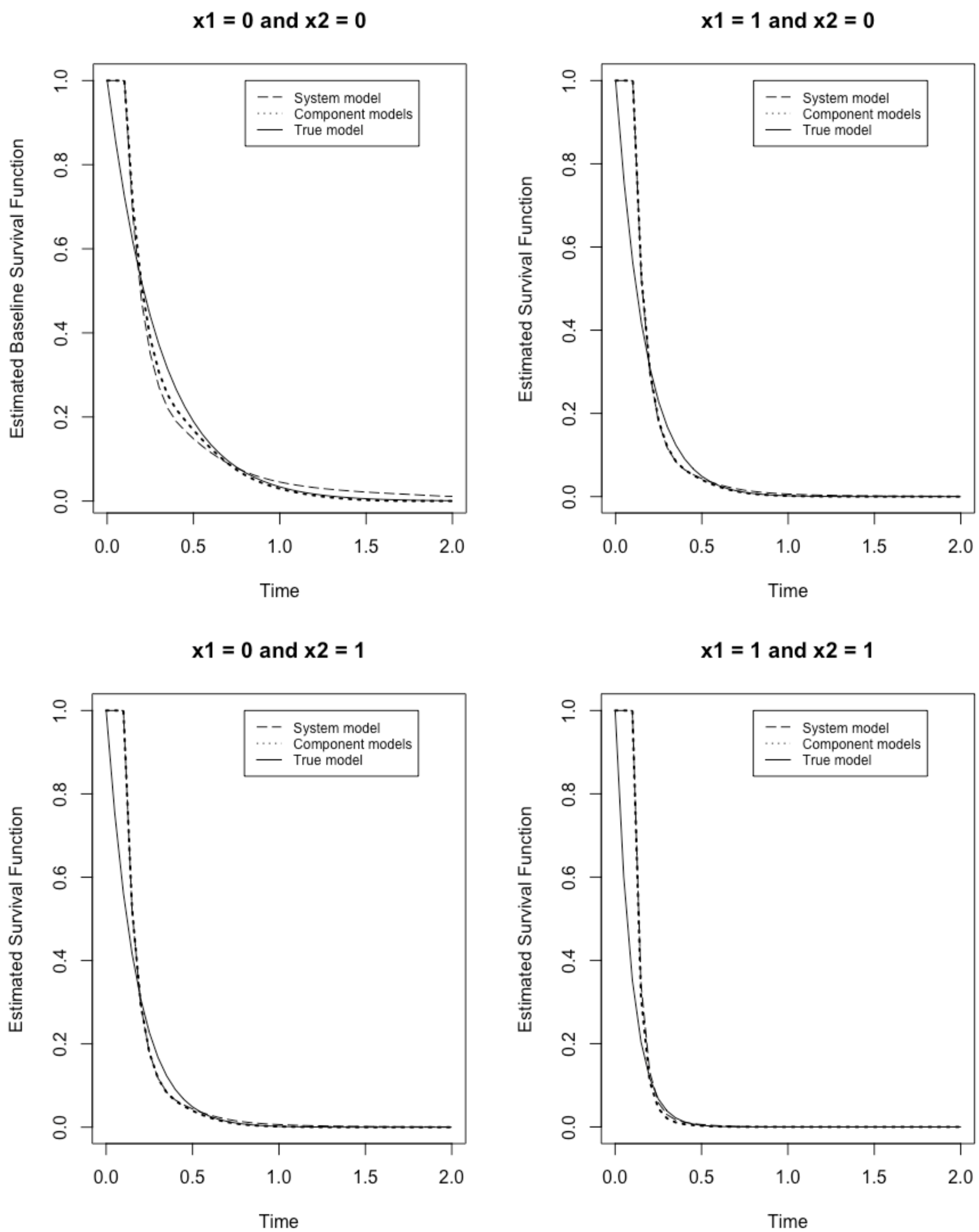

Figure 4.1: The estimated survival functions for a system using the PH model with system data (System model), the independent PH model with component data (Component models) and the Gamma-frailty PH model with component data (Component models) with the covariate $(x_1, x_2) \in \{(0,0),(1,0),(0,1),(1,1)\}$ in Scenario I.

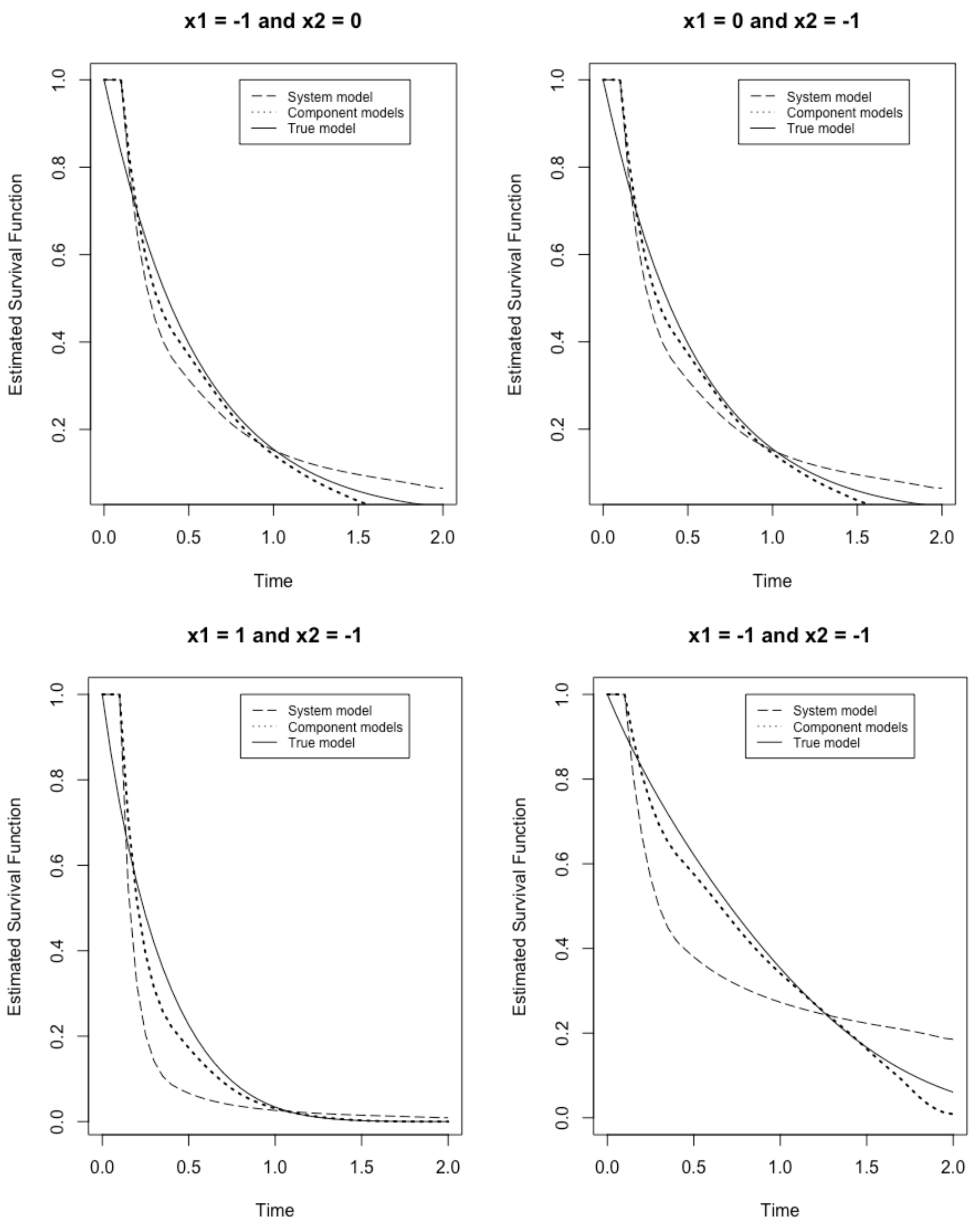

Figure 4.2: The estimated survival functions for a system using the PH model with system data (System model), the independent PH model with component data (Component models) and the Gamma-frailty PH model with component data (Component models) with the covariate $(x_1, x_2) \in \{(-1, 0), (0, -1), (1, -1), (-1, -1)\}$ in Scenario I.

were considered, for $j = 1, 2, 3, 4$. The parameter of the frailty term τ was set to be 1. The inverse CDF method was used to compute T_j's by solving $F_{T_j}(t|\mathbf{x}_i, \eta) = u_j$ numerically, where $u_j \sim U(0, 1)$, for $j = 1, 2, 3, 4$. The censoring time C followed a truncated exponential distribution with mean 0.1 and upper bound 3. The true distribution of the system failure time T is $\min\{T_j : j = 1, 2, 3, 4\}$.

According to the plots of the estimated survival functions in Figure 4.3 and Figure 4.4, it can be seen that the independent PH model with component data had the worst performance, thought it used all component data. Therefore, the independent assumption had a great effect on the performance of this method. When the assumption was violated, the method was worse than only using system data. Hence, it is important to verify the independent assumption before this method is applied. The PH model with system data performed better than the independent PH model but it was still not satisfactory in most of the cases. The system survival function estimated by the Gamma-frailty model with component data performed the best and it was very close to the true model. In this scenario, the simulation study shows that the independent assumption had a great effect on the performance of estimating the system survival function using the independent PH model with component, but it did not effect the performance of the Gamma-frailty PH model with component data.

Estimating the system survival function using the Gamma-frailty PH model with component data had the best performance in both Scenario I and II, which was the best choice among the three.

4.7 Real Data Application

The real data set is $\boldsymbol{tr-467}$ from the NTP, the same data set used in Chapter 3 real data application. The analysis of $\boldsymbol{tr-467}$ in Chapter 3 focused on whether there was an association between chloroprene and the onset time of Alveolar/Bronchiolar Adenoma in a mouse's lung. In this Chapter, a lab mouse's liver, pituitary gland

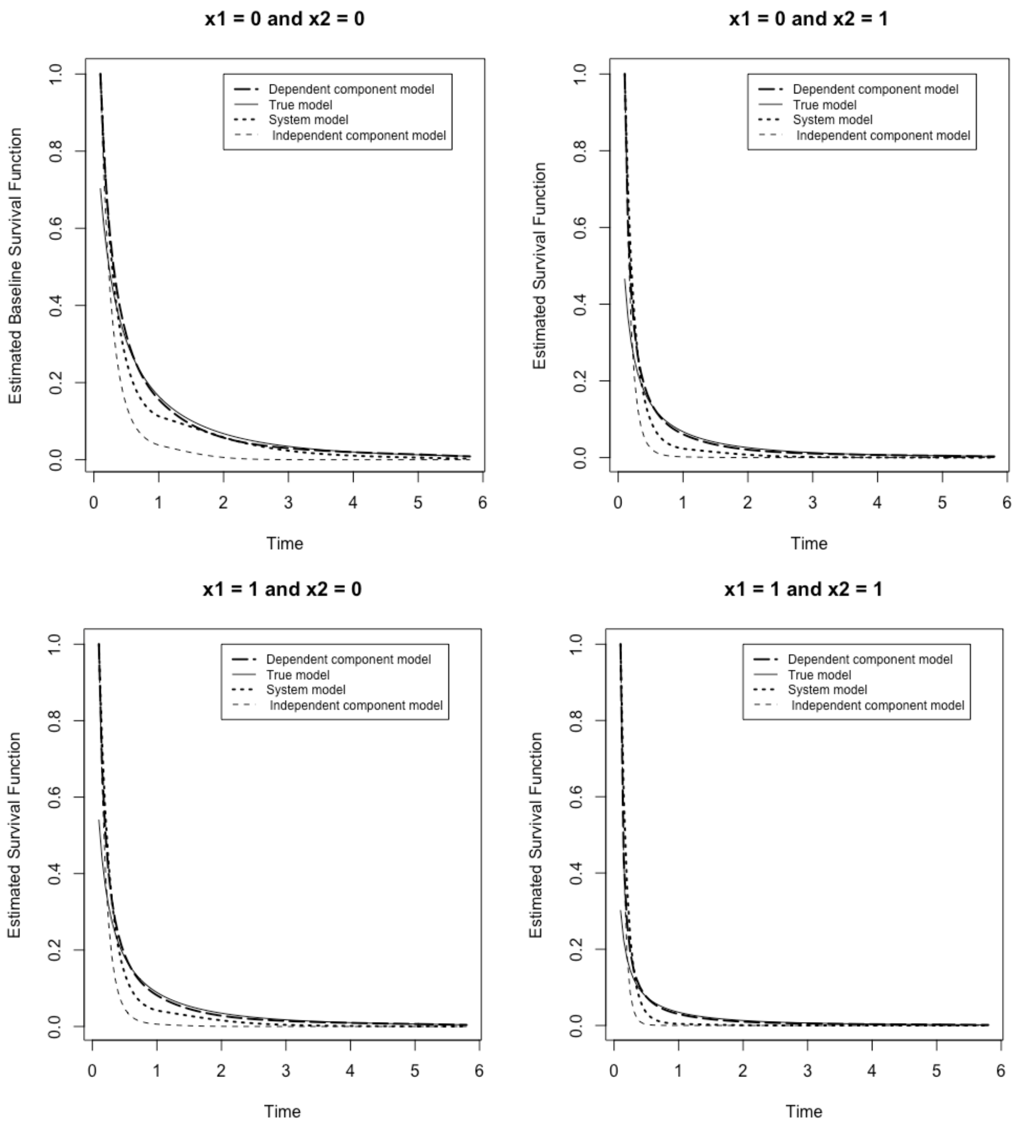

Figure 4.3: The estimated survival functions for a system using the PH model with system data (System model), the independent PH model with component data (Independent model) and the Gamma-frailty PH model with component data (Dependent model) with the covariate $(x_1, x_2) \in \{(-1, 0), (0, -1), (1, -1), (-1, -1)\}$ in Scenario II.

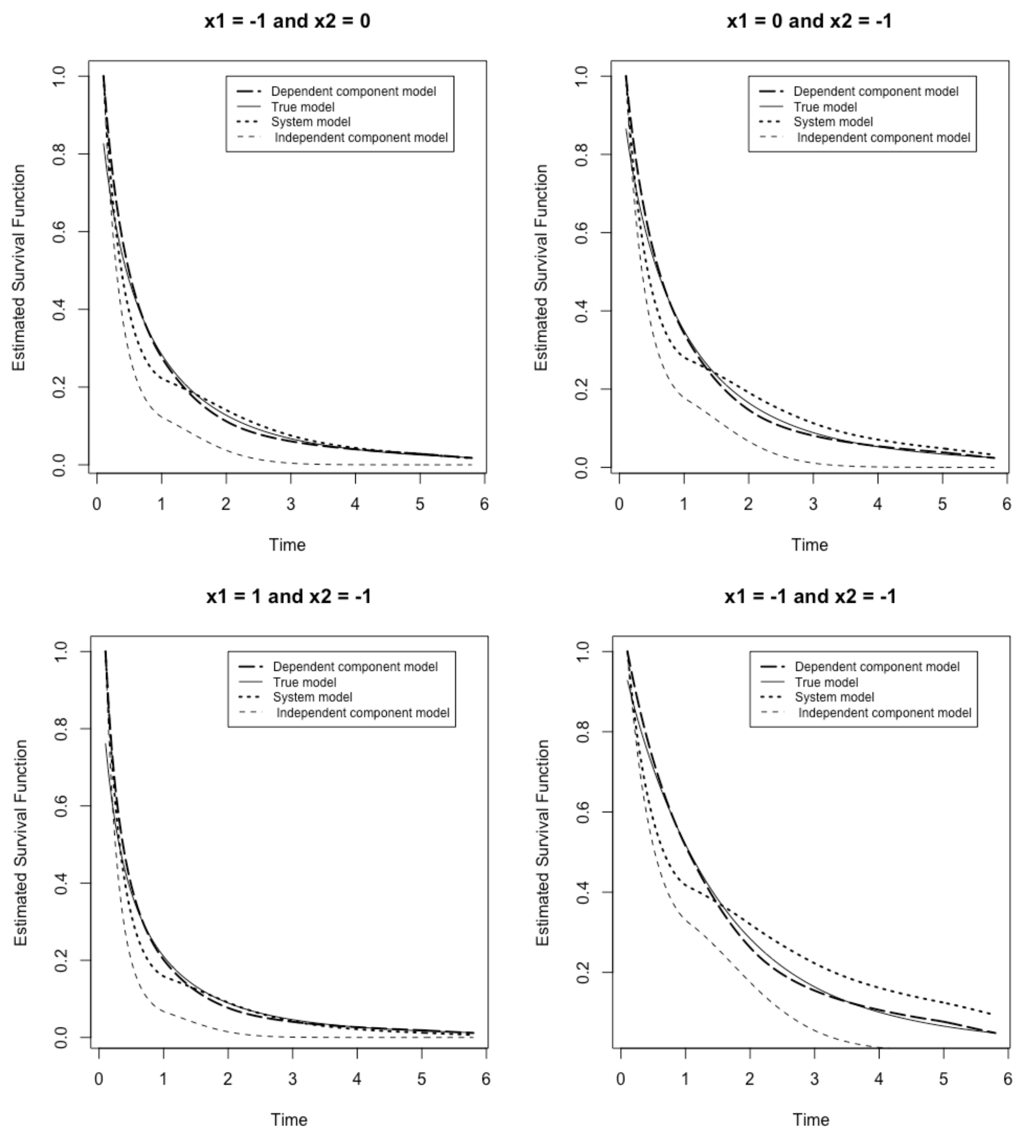

Figure 4.4: The estimated survival functions for a system using the PH model with system data (System model), the independent PH model with component data (Independent model) and the Gamma-frailty PH model with component data (Dependent model) with the covariate $(x_1, x_2) \in \{(-1, 0), (0, -1), (1, -1), (-1, -1)\}$ in Scenario II.

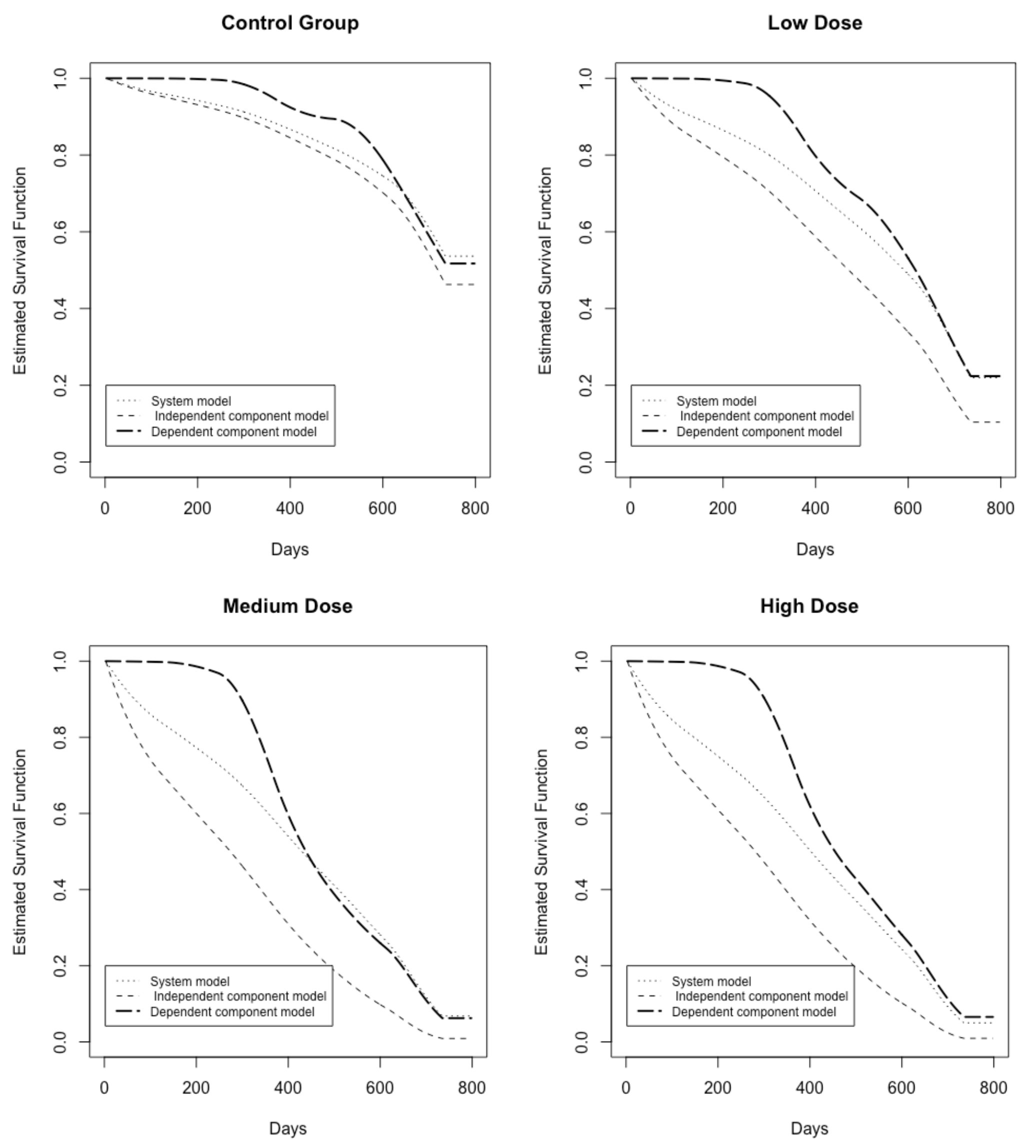

Figure 4.5: Statistical analysis of system reliability of $\boldsymbol{tr-467}$: estimated survival functions using the PH model with system data (System model), the independent PH model with component data (Independent model) and the Gamma-frailty PH model with component data (Dependent model) in four dose levels, control group, low dose, medium dose, high dose.

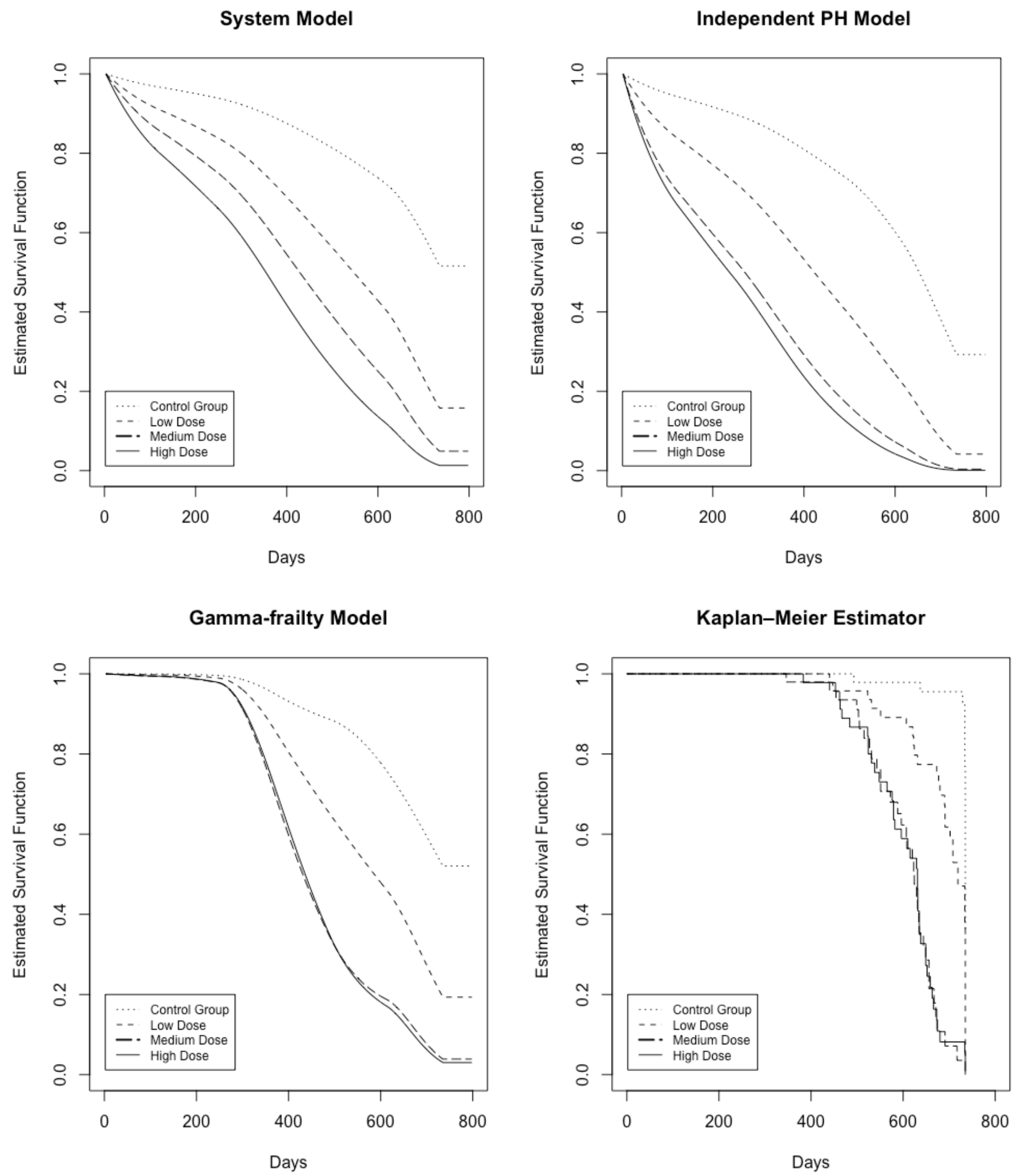

Figure 4.6: Statistical analysis of system reliability of $\boldsymbol{tr-467}$: estimated survival functions using the PH model with system data (System model), the independent PH model with component data (Independent model), the Gamma-frailty PH model with component data (Dependent model) and Kaplan-Meier estimator with system data (Kaplan-Meier Estimator) in four dose levels, control group, low dose, medium dose, high dose.

Table 4.1: Mean square error (MSE) of the estimated system survival functions with the three methods in Scenario I and II.

Method	(x_1, x_2)	Mean square error Scenario I	Mean square error Scenario II
System model	(0, 0)	0.0066	0.0019
	(1, 0)	0.0010	0.0036
	(1, 1)	0.0131	0.0061
	(-1, -1)	0.0196	0.0044
Independent component model	(0, 0)	0.0052	0.0057
	(1, 0)	0.0062	0.0045
	(1, 1)	0.0127	0.0052
	(-1, -1)	0.0013	0.0132
Dependent component model	(0, 0)	0.0052	0.0015
	(1, 0)	0.0062	0.0026
	(1, 1)	0.0127	0.0047
	(-1, -1)	0.0012	0.0003

```
 [1] 735+ 734+ 681+ 735+ 635+ 735  493  735+ 735  477+ 735+ 729
[13] 735+ 734+ 735  735  734+ 735+ 734+ 735  735+ 734+ 734  663+
[25] 735+ 610+ 705+ 735+ 734  527+ 734  706+ 677+ 735  541+ 734
[37] 735  486+ 735  735+  35+ 735  735  735  734  637  734+ 734
[49] 735  734+
```

(a) Control Group

```
 [1] 607  735  523  624  734+ 702  735  691+ 734+ 680  440  735
[13] 673  677  551+ 548+ 735  285+ 691  735  447  735  690+ 621
[25] 631  735  531  551  674+ 318+ 719  691  734  708  734  735
[37] 735  735+ 735  690  719  482+ 636+ 540+ 698+ 708  623  734
[49] 631+ 321+
```

(b) Low Dose

```
 [1] 631  440  537+ 539+ 216+ 503  608  649  635  463+ 543  499
[13] 674  735  691  634  656  527  531  588  487+ 582+ 515  505
[25] 604+ 607  623  551  472+ 425+ 524+ 346  572  671  389+ 596
[37] 644  656  620  394+ 629+ 629  717  624  378+ 551  523+ 454
[49] 667  652+
```

(c) Medium Dose

```
 [1] 467  579  607+ 576  453  734  652  659  523  565  252+ 624+
[13] 610  582  531+ 463  524  631+ 735  735  443+ 524+ 674  638
[25] 324+ 531  632  615  383  596  673  631  635  336+ 484  632
[37] 629    3+ 663  548  462  649  680  666  538  579  635  649
[49] 524  631
```

(d) High Dose

Figure 4.7: System data of $\boldsymbol{tr-467}$.

and lung were considered as a system. The analysis focuses on whether there was an association between chloroprene and the minimum onset time of Alveolar/Bronchiolar Adenoma in a mouse's lung, Adenoma in a mouse's pituitary gland, Hepatocellular Carcinoma in a mouse's liver and Hepatocellular Adenoma in a mouse's liver.

Figure 4.5 provides the estimated system survival functions in four dose levels. It can be seen that the system survival functions estimated by the three methods were different for all four dose level groups. Especially, between 0 - 300 hundred days, the system survival functions estimated by the Gamma-frailty model with component data were almost 1, which were much higher than the ones estimated by the other two methods. From the real system data shown in Figure 4.7, it can be seen that for all the observations in four groups, no tumors were observed within 300 days (no left-censored observations within 300 days). The smallest left-censored observation was 493 days for control group, 440 days for low dose group, 346 days for medium dose group and 383 days for high dose group. Moreover, it is more reasonable to consider that it will take a period of time to develop a tumor in a mouse. Therefore, the system survival function estimated by the Gamma-frailty model with component data is closest to the truth than the other two methods.

The analysis in this study sought for whether the system survival functions were different among the four groups. In Figure 4.6, besides the three methods, the Kaplan-Meier estimator was also applied to estimate the survival functions of the system using system data for the four groups. The advantage of Kaplan-Meier estimator is that it is a non-parametric estimator so that it is not limited to model structures. The Kaplan-Meier estimator showed that there was no difference of the survival functions between medium and high dose group, which agreed with the Gamma-frailty method, but the PH model using system data suggested that there existed a difference. All models suggested that there existed a difference of the survival functions among control, low dose and medium dose group.

4.8 Discussion

From the angle of real application, firstly, three methods are developed to analyze system reliability for current status data. All these methods can be widely applied to NTP data. Secondly, in some cases, analyze system reliability for current status data using system data may lead to biased estimation of the system reliability function. In these cases, using component data to analyze system reliability can be a better strategy. Thirdly, the method developed using component data to analyze system reliability can also be used as an optional method to confirm the analysis result. Lastly, this project considers a simple system. More complex systems are going to be considered in future work.

From the angle of methodology, a new method is developed for fitting the Gamma-frailty PH model to multivariate current status data. In previous literature, Wang et al. (2015) developed a similar method for analyzing bi-variate current status data. The method can not handle complex likelihood functions in higher dimensional situations because they involved too many terms. This study develops an EM-algorithm and finds a way to write the the conditional expectations in a general form and make the computation feasible. The method is not restricted to the situations discussed in this project. It can be used to analyze general multivariate current status data.

Bibliography

Betensky, Rebecca A, Jane C Lindsey, Louise M Ryan, and MP Wand (2002). "A local likelihood proportional hazards model for interval censored data". In: *Statistics in Medicine* 21.2, pp. 263–275.

Cai, Bo, Xiaoyan Lin, and Lianming Wang (2011). "Bayesian proportional hazards model for current status data with monotone splines". In: *Computational Statistics & Data Analysis* 55.9, pp. 2644–2651.

Chen, Chyong-Mei, Tai-Fang C Lu, Man-Hua Chen, and Chao-Min Hsu (2012). "Semiparametric transformation models for current status data with informative censoring". In: *Biometrical Journal* 54.5, pp. 641–656.

Devarajan, Karthik and Nader Ebrahimi (2011). "A semi-parametric generalization of the Cox proportional hazards regression model: Inference and applications". In: *Computational statistics & data analysis* 55.1, pp. 667–676.

Finkelstein, Dianne M (1986). "A proportional hazards model for interval-censored failure time data." In: *Biometrics* 42.4, pp. 845–854.

Goggins, William B, Dianne M Finkelstein, David A Schoenfeld, and Alan M Zaslavsky (1998). "A Markov chain Monte Carlo EM algorithm for analyzing interval-censored data under the Cox proportional hazards model". In: *Biometrics*, pp. 1498–1507.

Huang, Jian et al. (1996). "Efficient estimation for the proportional hazards model with interval censoring". In: *The Annals of Statistics* 24.2, pp. 540–568.

McMahan, Christopher S, Lianming Wang, and Joshua M Tebbs (2013). "Regression analysis for current status data using the EM algorithm". In: *Statistics in medicine* 32.25, pp. 4452–4466.

Murphy, Susan A, Aad W van der Vaart, et al. (1997). "Semiparametric likelihood ratio inference". In: *The Annals of Statistics* 25.4, pp. 1471–1509.

Pan, Wei (1999). “Extending the iterative convex minorant algorithm to the Cox model for interval-censored data”. In: *Journal of Computational and Graphical Statistics* 8.1, pp. 109–120.

Satten, Glen A (1996). “Rank-based inference in the proportional hazards model for interval censored data”. In: *Biometrika* 83.2, pp. 355–370.

Tian, Lu, David Zucker, and LJ Wei (2005). “On the Cox model with time-varying regression coefficients”. In: *Journal of the American statistical Association* 100.469, pp. 172–183.

Wang, Lianming, Christopher S McMahan, Michael G Hudgens, and Zaina P Qureshi (2016). “A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data”. In: *Biometrics* 72.1, pp. 222–231.

Wang, Naichen, Lianming Wang, and Christopher S McMahan (2015). “Regression analysis of bivariate current status data under the Gamma-frailty proportional hazards model using the EM algorithm”. In: *Computational Statistics & Data Analysis* 83, pp. 140–150.

Appendix A

Chapter 2 Supplementary Materials

A.1 Variance Estimation with Louis's Method

Louis (1982) gives a formula as follows,

$$\hat{I}(\hat{\boldsymbol{\theta}}) = -\frac{\partial^2 \log \mathcal{L}_{obs}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} | \boldsymbol{\theta} = \hat{\boldsymbol{\theta}},$$

where

$$-\frac{\partial^2 \log \mathcal{L}_{obs}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - Var\left\{\frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}.$$

To evaluate the first term $\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$, the $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ function is as follows,

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = & \sum_{i=1}^{n} \sum_{l=1}^{k} \left\{\delta_{i0} E(V_{il}) + (1-\delta_{i0}) E(Z_{il}) + (\delta_{i2}+\delta_{i3}) E(W_{il})\right\} \left(\log \gamma_l + \mathbf{x}_i' \boldsymbol{\beta}\right) \\ & - \sum_{i=1}^{n} \sum_{l=1}^{k} \left\{\delta_{i0} I_l(R_i) + (\delta_{i1}+\delta_{i2}) I_l(R_i) + \delta_{i3} I_l(L_i)\right\} \gamma_l \exp(\mathbf{x}_i' \boldsymbol{\beta}). \end{aligned}$$

The first derivative of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\theta}$ is as follows,

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}} = & \sum_{i=1}^{n} \sum_{l=1}^{k} \left\{\delta_{i0} E(V_{il}) + (1-\delta_{i0}) E(Z_{il}) + (\delta_{i2}+\delta_{i3}) E(W_{il})\right\} \mathbf{x}_i \\ & - \sum_{i=1}^{n} \left\{(1-\delta_{i3}) \Lambda_0(R_i) + \delta_{i3} \Lambda_0(L_i)\right\} \exp(\mathbf{x}_i' \boldsymbol{\beta}) \mathbf{x}_i, \end{aligned}$$

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \gamma_l} = & \sum_{i=1}^{n} \gamma_l^{-1} \left\{\delta_{i0} E(V_{il}) + (1-\delta_{i0}) E(Z_{il}) + (\delta_{i2}+\delta_{i3}) E(W_{il})\right\} \\ & - \sum_{i=1}^{n} \left\{(1-\delta_{i3}) I_l(R_i) + \delta_{i3} I_l(L_i)\right\} \exp(\mathbf{x}_i' \boldsymbol{\beta}). \end{aligned}$$

The second derivative of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\theta}$ is as follows,

$$\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\sum_{i=1}^{n} \left\{(1-\delta_{i3}) \Lambda_0(R_i) + \delta_{i3} \Lambda_0(L_i)\right\} \exp(\mathbf{x}_i' \boldsymbol{\beta}) \mathbf{x}_i \mathbf{x}_i',$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta} \partial \gamma_l} = -\sum_{i=1}^{n} \{(1-\delta_{i3}) I_l(R_i) + \delta_{i3} I_l(L_i)\} \exp(\mathbf{x}_i' \boldsymbol{\beta}) \mathbf{x}_i,$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \gamma_l \partial \gamma_{l'}} = -\sum_{i=1}^{n} \gamma_l^{-2} \{\delta_{i0} E(V_{il}) + (1-\delta_{i0}) E(Z_{il}) + (\delta_{i2}+\delta_{i3}) E(W_{il})\} I(l = l').$$

To evaluate the second term $Var\left\{\frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}$, the $\log \mathcal{L}_{com}(\boldsymbol{\theta})$ is as follows,

$$\begin{aligned}\log \mathcal{L}_{com}(\boldsymbol{\theta}) =& \sum_{i=1}^{n}\sum_{l=1}^{k} \{V_{il}\delta_{i0} + (1-\delta_{i0}) Z_{il} + (\delta_{i2}+\delta_{i3}) W_{il}\} \log \gamma_l \\ &+ \sum_{i=1}^{n} \{\delta_{i0} + (1-\delta_{i0}) Z_i + (\delta_{i2}+\delta_{i3}) W_i\} \mathbf{x}_i' \boldsymbol{\beta} \\ &- \sum_{i=1}^{n}\sum_{l=1}^{k} \{\delta_{i0} I_l(R_i) + (\delta_{i1}+\delta_{i2}) b_l(Ri) + \delta_{i3} b_l(L_i)\} \gamma_l \exp(\mathbf{x}_i' \boldsymbol{\beta}).\end{aligned}$$

The first derivative of $\log \mathcal{L}_{com}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is,

$$\begin{aligned}\frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} =& \sum_{i=1}^{n} \{\delta_{i0} + (1-\delta_{i0}) Z_i + (\delta_{i2}+\delta_{i3}) W_i)\} \mathbf{x}_i \\ &- \sum_{i=1}^{n}\sum_{l=1}^{k} \{\delta_{i0} I_l(R_i) + (\delta_{i1}+\delta_{i2}) I_l(R_i) + \delta_{i3} I_l(L_i)\} \gamma_l \exp(\mathbf{x}_i' \boldsymbol{\beta}) \mathbf{x}_i,\end{aligned}$$

$$\begin{aligned}\frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \gamma_l} =& \sum_{i=1}^{n} \{\delta_{i0} V_{il} + (1-\delta_{i0}) Z_{il} + (\delta_{i2}+\delta_{i3}) W_{il}\} \gamma_l^{-1} \\ &- \sum_{i=1}^{n} \{\delta_{i0} I_l(R_i) + (\delta_{i1}+\delta_{i2}) b_l(Ri) + \delta_{i3} b_l(L_i)\} \exp(\mathbf{x}_i' \boldsymbol{\beta}).\end{aligned}$$

Therefore the covariance can be obtained as follows,

$$\begin{aligned}&\operatorname{cov}\left\{\frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}, \frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}\right\} \\ &= \sum_{i=1}^{n} \{(1-\delta_{i0}) \operatorname{var}(Z_i) + (\delta_{i2}+\delta_{i3}) \operatorname{var}(W_i)\} \mathbf{x}_i \mathbf{x}_i', \\ &\operatorname{cov}\left\{\frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}, \frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \gamma_l}\right\} \\ &= \sum_{i=1}^{n} \gamma_l^{-1} \{(1-\delta_{i0}) \operatorname{cov}(Z_i, Z_{il}) + (\delta_{i2}+\delta_{i3}) \operatorname{cov}(W_i, W_{il})\} \mathbf{x}_i, \\ &\operatorname{cov}\left\{\frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \gamma_l}, \frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \gamma_{l'}}\right\} \\ &= \sum_{i=1}^{n} (\gamma_l, \gamma_{l'})^{-1} \{\delta_{i0} \operatorname{cov}(V_{il}, V_{il'}) + (1-\delta_{i0}) \operatorname{cov}(Z_{il}, Z_{il'}) + (\delta_{i2}+\delta_{i3}) \operatorname{cov}(W_{il}, W_{il'})\} \mathbf{x}_i.\end{aligned}$$

Appendix B

Chapter 3 Supplementary Materials

B.1 The Conditional Distribution of V_i

In the complete likelihood function, only the last term contains the latent variable V_i. Hence the conditional probability density function, $f(V_i|\boldsymbol{x_i})$ is proportional to it which can be expressed as follows,

$$f(V_i|\boldsymbol{x}_i) \propto \prod_{j=1}^{k} \{\gamma_{Cj} M_j(\tilde{c}_i)\}^{V_{ij}\xi_i} .$$

It is easy to recegnize that this is the kernel of multinomial distribution, that is $V_i|\boldsymbol{x}_i \sim Multinomial[1, (\frac{\gamma_{C1}M_1(\tilde{c}_i)}{\sum_{j=1}^{k}\gamma_{Cj}M_j(\tilde{c}_i)}, \frac{\gamma_{C2}M_2(\tilde{c}_i)}{\sum_{j=1}^{k}\gamma_{Cj}M_j(\tilde{c}_i)}, ..., \frac{\gamma_{Ck}M_k(\tilde{c}_i)}{\sum_{j=1}^{k}\gamma_{Cj}M_j(\tilde{c}_i)})]$.
Therefore the conditional expectation, variance and covariance can be found as follows,

$$E(V_{ij}|\boldsymbol{x}_i) = \frac{\gamma_{Cj}M_j(\tilde{c}_i)}{\sum_{j=1}^{k}\gamma_{Cj}M_j(\tilde{c}_i)},$$

$$\text{var}(V_{ij}|\boldsymbol{x}_i) = \frac{\gamma_{Cj}M_j(\tilde{c}_i)}{\sum_{j=1}^{k}\gamma_{Cj}M_j(\tilde{c}_i)} \left\{1 - \frac{\gamma_{Cj}M_j(\tilde{c}_i)}{\sum_{j=1}^{k}\gamma_{Cj}M_j(\tilde{c}_i)}\right\},$$

$$\text{cov}(V_{ij}, V_{ij'}|\boldsymbol{x}_i) = -\frac{\gamma_{Cj}M_j(\tilde{c}_i)}{\sum_{j=1}^{k}\gamma_{Cj}M_j(\tilde{c}_i)} \frac{\gamma_{Cj'}M_{j'}(\tilde{c}_i)}{\sum_{j=1}^{k}\gamma_{Cj'}M_{j'}(\tilde{c}_i)} + \frac{\gamma_{Cj}M_j(\tilde{c}_i)}{\sum_{j=1}^{k}\gamma_{Cj}M_j(\tilde{c}_i)} I(j = j').$$

B.2 The Conditional Distribution of η_i

To find the posterior distribution of η_i, the following augmented Likelihood function is used,

$$\mathcal{L}_1(\boldsymbol{\theta}) = \prod_{i=1}^{n}[1-\exp\{-\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x'}\boldsymbol{\beta}_T)\eta_i\}]^{\delta_i}\exp\{-\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x'}\boldsymbol{\beta}_T)\eta_i\}^{1-\delta_i}$$
$$\frac{\tau^{\tau}}{\Gamma(\tau)}\eta_i^{\tau+\xi_i-1}e^{-\tau\eta_i}\{\lambda_{C0}(\tilde{c}_i)\}^{\xi_i}\exp\{\xi_i\boldsymbol{x'}\boldsymbol{\beta}_C - \Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x'}\boldsymbol{\beta}_C)\eta_i\}.$$

Hence

$$f(\boldsymbol{\eta}|\boldsymbol{x}) \propto \prod_{i=1}^{n}[1-\exp\{-\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\eta_i\}]^{\delta_i}\exp\{-\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\eta_i\}^{1-\delta_i}$$
$$\eta_i^{\xi_i+\tau-1}\exp\{-[\tau+\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)]\eta_i\}.$$

When $\delta_i = 0$,

$$f(\boldsymbol{\eta}|\boldsymbol{x}) \propto \prod_{i=1}^{n}\eta_i^{\xi_i+\tau-1}\exp\{-[\tau+\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)+\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)]\eta_i\},$$
$$\eta_i|\boldsymbol{x}_i \sim Gamma(\xi_i+\tau, \tau+\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)+\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C))$$
$$\sim Gamma(\xi_i+\tau, b_i),$$

where $b_i = \tau+\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)+\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)$.

When $\delta_i = 1$,

$$f(\boldsymbol{\eta}|\boldsymbol{x}) \propto \prod_{i=1}^{n}\eta_i^{\xi_i+\tau-1}[1-\exp\{-\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\eta_i\}]$$
$$\exp\{-[\tau+\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)]\eta_i\}$$
$$\propto \prod_{i=1}^{n}(\eta_i^{\xi_i+\tau-1}\exp\{-[\tau+\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)]\eta_i\}$$
$$-\eta_i^{\xi_i+\tau-1}\exp\{-[\tau+\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)+\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)]\eta_i\}),$$
$$f(\eta_i|\boldsymbol{x}_i) \propto \eta_i^{\xi_i+\tau-1}\exp(-d_i\eta_i) - \eta_i^{\xi_i+\tau-1}\exp(-b_i\eta_i),$$

where $d_i = \tau+\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)$.

The following proofs are needed to find the posterior expectations and covariance related to $\log \eta_i$.

Assume X follows a gamma distribution with parameters $(a,\ 1)$. With the fact $\Gamma^{(n)}(x) = \int (\log t)^n t^{x-1} e^{-t} dt$, the following results can be obtained.

$$
\begin{aligned}
&X \sim Gamma(a, 1),\\
&E(\log X) = \int (\log X) \frac{1}{\Gamma(a)} t^{a-1} e^{-X} dX = \frac{\Gamma'(a)}{\Gamma(a)} = \psi(a),\\
&E(X \log X) = \int (\log X) \frac{1}{\Gamma(a)} X^{a+1-1} e^{-X} dX = \frac{\Gamma'(a+1)}{\Gamma(a)} = a\psi(a) + 1,\\
&E(\log^2 X) = \int (\log X)^2 \frac{1}{\Gamma(a)} X^{a-1} e^{-X} dX = \frac{\Gamma^{(2)}(x)}{\Gamma(x)} = \psi_1(a) + \psi^2(a).
\end{aligned}
$$

Assume Y follows a gamma distribution with parameters $(a,\ b)$ then,

$$
\begin{aligned}
&Y \sim Gamma(a, b),\\
&Y = X \cdot b^{-1},\\
&E[\log(Y)] = E[\log(Xb^{-1})] = E(\log X) + E(\log b^{-1}) = \psi(a) - \log b,\\
&E[\log^2(Y)] = E[\log^2(Xb^{-1})] = E[(\log X - \log b)^2]\\
&= E(\log^2 X) - 2\log(b) E(\log X) + \log^2 b\\
&= \psi_1(a) + \psi^2(a) - 2\log(b)\psi(a) + \log^2 b\\
&= \psi_1(a) + [\psi(a) - \log(b)]^2,\\
&E[Y \log(Y)] = E[b^{-1} X \log(b^{-1} X)] = b^{-1} E(X \log X) - b^{-1} \log b E(X)\\
&= b^{-1}[a\psi(a) + 1] - b^{-1}(\log b) a = b^{-1} a[\psi(a+1) - \log b].
\end{aligned}
$$

These results are going to be used to to find the posterior expectations and covariance related to $\log \eta_i$.

With the results obtained above, when $\delta_i = 0$,

$\eta_i|\boldsymbol{x}_i \sim Gamma(\xi_i + \tau, b_i)$, therefore

$$\begin{aligned}
E(\eta_i|\boldsymbol{x}_i) &= \frac{\tau + \xi_i}{b_i}, \\
E(\log \eta_i|\boldsymbol{x}_i) &= \psi(\tau + \xi_i) - \log b_i, \\
\mathrm{var}(\eta_i|\boldsymbol{x}_i) &= \frac{\tau + \xi_i}{b_i^2}, \\
\mathrm{var}(\log \eta_i|\boldsymbol{x}_i) &= E(\log^2 \eta_i|\boldsymbol{x}_i) - [E(\log \eta_i|\boldsymbol{x}_i)]^2 \\
&= \psi_1(\tau + \xi_i) + [\psi(\tau + \xi_i) - \log b_i]^2 - [\psi(\tau + \xi_i) - \log b_i]^2 \\
&= \psi_1(\tau + \xi_i), \\
\mathrm{cov}(\log \eta_i, \eta_i|\boldsymbol{x}_i) &= E(\eta_i \log \eta_i|\boldsymbol{x}_i) - E(\log \eta_i|\boldsymbol{x}_i)E(\eta_i|\boldsymbol{x}_i) \\
&= b_i^{-1}[(\tau + \xi_i)\psi(\tau + \xi_i) + 1] - b_i^{-1}(\log b_i)(\tau + \xi_i) \\
&\quad - (\tau + \xi_i)b_i^{-1}[\psi(\tau + \xi_i) - \log b_i] \\
&= b_i^{-1}.
\end{aligned}$$

When $\delta_i = 1$,

$$\begin{aligned}
f(\eta_i|\boldsymbol{x}_i) \propto &\quad \eta_i^{\xi_i+\tau-1} \exp(-d_i\eta_i) - \eta_i^{\xi_i+\tau-1} \exp(-b_i\eta_i) \\
\propto &\frac{\Gamma(\tau + \xi_i)}{d_i^{\tau+\xi_i}} \frac{d_i^{\tau+\xi_i}}{\Gamma(\tau + \xi_i)} \eta_i^{\xi_i+\tau-1} \exp(-d_i\eta_i) - \frac{\Gamma(\tau + \xi_i)}{b_i^{\tau+\xi_i}} \frac{b_i^{\tau+\xi_i}}{\Gamma(\tau + \xi_i)} \eta_i^{\xi_i+\tau-1} \exp(-b_i\eta_i) \\
\propto &\frac{\Gamma(\tau + \xi_i)}{d_i^{\tau+\xi_i}} g(\eta_i|\tau + \xi_i, d_i) - \frac{\Gamma(\tau + \xi_i)}{b_i^{\tau+\xi_i}} g(\eta_i|\tau + \xi_i, b_i).
\end{aligned}$$

Therefore when $\delta_i = 1$,

$$\int f(\eta_i|\boldsymbol{x}_i)d\eta_i \propto \frac{\Gamma(\tau+\xi_i)}{d_i^{\tau+\xi_i}} - \frac{\Gamma(\tau+\xi_i)}{b_i^{\tau+\xi_i}}. \tag{B.1}$$

$$\begin{aligned}
&\int \eta_i f(\eta_i|\boldsymbol{x}_i)d\eta_i \\
&\propto \frac{\Gamma(\tau+\xi_i)}{d_i^{\tau+\xi_i}}\frac{\tau+\xi_i}{d_i} - \frac{\Gamma(\tau+\xi_i)}{b_i^{\tau+\xi_i}}\frac{\tau+\xi_i}{b_i} \\
&\propto \frac{\Gamma(\tau+\xi_i+1)}{d_i^{\tau+\xi_i+1}} - \frac{\Gamma(\tau+\xi_i+1)}{b_i^{\tau+\xi_i+1}}.
\end{aligned} \tag{B.2}$$

$$\begin{aligned}
&\int \log\eta_i f(\eta_i|\boldsymbol{x}_i)d\eta_i \\
&\propto \frac{\Gamma(\tau+\xi_i)}{d_i^{\tau+\xi_i}}[\psi(\tau+\xi_i) - \log d_i] - \frac{\Gamma(\tau+\xi_i)}{b_i^{\tau+\xi_i}}[\psi(\tau+\xi_i) - \log b_i].
\end{aligned} \tag{B.3}$$

$$\begin{aligned}
&\int \eta_i^2 f(\eta_i|\boldsymbol{x}_i)d\eta_i \\
&\propto \int \frac{\Gamma(\tau+\xi_i+2)}{d_i^{\tau+\xi_i+2}}\frac{d_i^{\tau+\xi_i+2}}{\Gamma(\tau+\xi_i+2)}\eta_i^{\xi_i+\tau+2-1}\exp(-d_i\eta_i) \\
&\quad - \frac{\Gamma(\tau+\xi_i+2)}{b_i^{\tau+\xi_i+2}}\frac{b_i^{\tau+\xi_i+2}}{\Gamma(\tau+\xi_i+2)}\eta_i^{\xi_i+\tau+2-1}\exp(-b_i\eta_i)d\eta_i \\
&\propto \frac{\Gamma(\tau+\xi_i+2)}{d_i^{\tau+\xi_i+2}} - \frac{\Gamma(\tau+\xi_i+2)}{b_i^{\tau+\xi_i+2}}.
\end{aligned} \tag{B.4}$$

$$\begin{aligned}
&\int \log^2\eta_i f(\eta_i|\boldsymbol{x}_i)d\eta_i \\
&\propto \frac{\Gamma(\tau+\xi_i)}{d_i^{\tau+\xi_i}}\int \log\eta_i g(\eta_i|\tau+\xi+1, d_i)d\eta_i - \frac{\Gamma(\tau+\xi_i)}{b_i^{\tau+\xi_i}}\int \log^2\eta_i g(\eta_i|\tau+\xi_i, b_i)d\eta_i \\
&\propto \frac{\Gamma(\tau+\xi_i)}{d_i^{\tau+\xi_i}}\left\{\psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log d_i]^2\right\} \\
&\quad - \frac{\Gamma(\tau+\xi_i)}{b_i^{\tau+\xi_i}}\left\{\psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log b_i]^2\right\}.
\end{aligned} \tag{B.5}$$

$$\begin{aligned}
&\int \eta_i\log\eta_i f(\eta_i|\boldsymbol{x}_i)d\eta_i \\
&\propto \frac{\Gamma(\tau+\xi_i+1)}{d_i^{\tau+\xi_i+1}}\int \log\eta_i g(\eta_i|\tau+\xi_i+1, d_i)d\eta_i \\
&\qquad - \frac{\Gamma(\tau+\xi_i+1)}{b_i^{\tau+\xi_i+1}}\int \log\eta_i g(\eta_i|\tau+\xi_i+1, b_i)d\eta_i \\
&\propto \frac{\Gamma(\tau+\xi_i+1)}{d_i^{\tau+\xi_i+1}}[\psi(\tau+\xi_i+1) - \log d_i] - \frac{\Gamma(\tau+\xi_i+1)}{b_i^{\tau+\xi_i+1}}[\psi(\tau+\xi_i+1) - \log b_i].
\end{aligned} \tag{B.6}$$

Hence when $\delta_i = 1$,

$$\begin{aligned} E(\eta_i|\boldsymbol{x}_i) &= \frac{\int \eta_i f(\eta_i|\boldsymbol{x}_i) d\eta_i}{\int f(\eta_i|\boldsymbol{x}_i) d\eta_i} = \frac{(\mathbf{B}.2)}{(\mathbf{B}.1)} \\ &= (\tau+\xi_i) \frac{\frac{1}{d_i^{\tau+\xi_i+1}} - \frac{1}{b_i^{\tau+\xi_i+1}}}{\frac{1}{d_i^{\tau+\xi_i}} - \frac{1}{b_i^{\tau+\xi_i}}} \\ &= \frac{(\tau+\xi_i)}{d_i} \frac{1-(\frac{d_i}{b_i})^{\tau+\xi_i+1}}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}}, \end{aligned}$$

$$\begin{aligned} E(\eta_i^2|\boldsymbol{x}_i) &= \frac{\int \eta_i^2 f(\eta_i|\boldsymbol{x}_i) d\eta_i}{\int f(\eta_i|\boldsymbol{x}_i) d\eta_i} = \frac{(\mathbf{B}.4)}{(\mathbf{B}.1)} \\ &= \frac{\frac{\Gamma(\tau+\xi_i+2)}{d_i^{\tau+\xi_i+2}} - \frac{\Gamma(\tau+\xi_i+2)}{b_i^{\tau+\xi_i+2}}}{\frac{\Gamma(\tau+\xi_i)}{d_i^{\tau+\xi_i}} - \frac{\Gamma(\tau+\xi_i)}{b_i^{\tau+\xi_i}}} \\ &= (\tau+\xi_i)(\tau+\xi_i+1) \frac{\frac{1}{d_i^{\tau+\xi_i+2}} - \frac{1}{b_i^{\tau+\xi_i+2}}}{\frac{1}{d_i^{\tau+\xi_i}} - \frac{1}{b_i^{\tau+\xi_i}}} \\ &= \frac{(\tau+\xi_i)(\tau+\xi_i+1)}{d_i^2} \frac{1-(\frac{d_i}{b_i})^{\tau+\xi_i+2}}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}}, \end{aligned}$$

$$\begin{aligned} E(\log\eta_i|\boldsymbol{x}_i) &= \frac{\int \log\eta_i f(\eta_i|\boldsymbol{x}_i) d\eta_i}{\int f(\eta_i|\boldsymbol{x}_i) d\eta_i} = \frac{(\mathbf{B}.3)}{(\mathbf{B}.1)} \\ &= \left\{ \frac{\psi(\tau+\xi_i) - \log d_i}{d_i^{\tau+\xi_i}} - \frac{\psi(\tau+\xi_i) - \log b_i}{b_i^{\tau+\xi_i}} \right\} \Big/ \left\{ \frac{1}{d_i^{\tau+\xi_i}} - \frac{1}{b_i^{\tau+\xi_i}} \right\} \\ &= \frac{b_i^{\tau+\xi_i}(\psi(\tau+\xi_i) - \log d_i) - d_i^{\tau+\xi_i}(\psi(\tau+\xi_i) - \log b_i)}{b_i^{\tau+\xi_i} - d_i^{\tau+\xi_i}} \\ &= \psi(\tau+\xi_i) - \frac{b_i^{\tau+\xi_i}\log d_i - d_i^{\tau+\xi_i}\log b_i}{b_i^{\tau+\xi_i} - d_i^{\tau+\xi_i}} \\ &= \psi(\tau+\xi_i) - \frac{\log d_i - (\frac{d_i}{b_i})^{\tau+\xi_i}\log b_i}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}}, \end{aligned}$$

$$E(\log \eta_i^2|\boldsymbol{x}_i) = \frac{\int \log^2 \eta_i f(\eta_i|\boldsymbol{x}_i)d\eta_i}{\int f(\eta_i|\boldsymbol{x}_i)d\eta_i} = \frac{(\mathbf{B}.5)}{(\mathbf{B}.1)}$$

$$= \left\{ \frac{\Gamma(\tau+\xi_i)}{d_i^{\tau+\xi_i}} \left\{ \psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log d_i]^2 \right\} \right.$$

$$\left. - \frac{\Gamma(\tau+\xi_i)}{b_i^{\tau+\xi_i}} \left\{ \psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log b_i]^2 \right\} \right\} / \left\{ \frac{\Gamma(\tau+\xi_i)}{d_i^{\tau+\xi_i}} - \frac{\Gamma(\tau+\xi_i)}{b_i^{\tau+\xi_i}} \right\}$$

$$= (b_i^{\tau+\xi_i} \left\{ \psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log d_i]^2 \right\}$$

$$- d_i^{\tau+\xi_i} \left\{ \psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log b_i]^2 \right\})(b_i^{\tau+\xi_i} - d_i^{\tau+\xi_i})^{-1}$$

$$= \frac{\{\psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log d_i]^2\} - (\frac{d_i}{b_i})^{\tau+\xi_i} \{\psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log b_i]^2\}}{1 - (\frac{d_i}{b_i})^{\tau+\xi_i}},$$

$$E(\eta_i \log \eta_i|\boldsymbol{x}_i) = \frac{\int \eta_i \log \eta_i f(\eta_i|\boldsymbol{x}_i)d\eta_i}{\int f(\eta_i|\boldsymbol{x}_i)d\eta_i} = \frac{(\mathbf{B}.6)}{(\mathbf{B}.1)}$$

$$= \frac{\frac{\Gamma(\tau+\xi_i+1)}{d_i^{\tau+\xi_i+1}}[\psi(\tau+\xi_i+1) - \log d_i] - \frac{\Gamma(\tau+\xi_i+1)}{b_i^{\tau+\xi_i+1}}[\psi(\tau+\xi_i+1) - \log b_i]}{\frac{\Gamma(\tau+\xi_i)}{d_i^{\tau+\xi_i}} - \frac{\Gamma(\tau+\xi_i)}{b_i^{\tau+\xi_i}}}$$

$$= \frac{\frac{(\tau+\xi_i)}{d_i^{\tau+\xi_i+1}}[\psi(\tau+\xi_i+1) - \log d_i] - \frac{(\tau+\xi_i)}{b_i^{\tau+\xi_i+1}}[\psi(\tau+\xi_i+1) - \log b_i]}{\frac{1}{d_i^{\tau+\xi_i}} - \frac{1}{b_i^{\tau+\xi_i}}}$$

$$= \frac{(\tau+\xi_i)}{d_i} \frac{[\psi(\tau+\xi_i+1) - \log d_i] - (\frac{d_i}{b_i})^{\tau+\xi_i+1}[\psi(\tau+\xi_i+1) - \log b_i]}{1 - (\frac{d_i}{b_i})^{\tau+\xi_i}},$$

$$\text{var}(\eta_i|\boldsymbol{x}_i) = E(\eta_i^2|\boldsymbol{x}_i) - [E(\eta_i|\boldsymbol{x}_i)]^2$$

$$= \frac{(\tau+\xi_i)(\tau+\xi_i+1)}{d_i^2} \frac{1 - (\frac{d_i}{b_i})^{\tau+\xi_i+2}}{1 - (\frac{d_i}{b_i})^{\tau+\xi_i}} - [\frac{(\tau+\xi_i)}{d_i} \frac{1 - (\frac{d_i}{b_i})^{\tau+\xi_i+1}}{1 - (\frac{d_i}{b_i})^{\tau+\xi_i}}]^2,$$

$$\text{var}(\log \eta_i|\boldsymbol{x}_i) = E(\log^2 \eta_i|\boldsymbol{x}_i) - [E(\log \eta_i|\boldsymbol{x}_i)]^2$$

$$= \frac{\{\psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log d_i]^2\} - (\frac{d_i}{b_i})^{\tau+\xi_i} \{\psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log b_i]^2\}}{1 - (\frac{d_i}{b_i})^{\tau+\xi_i}}$$

$$- [\psi(\tau+\xi_i) - \frac{\log d_i - (\frac{d_i}{b_i})^{\tau+\xi_i} \log b_i}{1 - (\frac{d_i}{b_i})^{\tau+\xi_i}}]^2,$$

$$\begin{aligned}
\operatorname{cov}(\eta_i, \log \eta_i | \boldsymbol{x}_i) &= E(\eta_i \log \eta_i | \boldsymbol{x}_i) - E(\eta_i|\boldsymbol{x}_i) E(\log \eta_i | \boldsymbol{x}_i) \\
&= \frac{(\tau+\xi_i)}{d_i} \frac{[\psi(\tau+\xi_i+1) - \log d_i] - (\frac{d_i}{b_i})^{\tau+\xi_i+1}[\psi(\tau+\xi_i+1) - \log b_i]}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}} \\
&- [\frac{(\tau+\xi_i)}{d_i} \frac{1-(\frac{d_i}{b_i})^{\tau+\xi_i+1}}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}}][\psi(\tau+\xi_i) - \frac{\log d_i - (\frac{d_i}{b_i})^{\tau+\xi_i} \log b_i}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}}].
\end{aligned}$$

Combining the cases $\delta_i = 0$ and $\delta_i = 1$ together, the following results can be obtained,

$$E(\eta_i|\boldsymbol{x}_i) = (1-\delta_i)\frac{\tau+\xi_i}{b_i} + \delta_i \frac{(\tau+\xi_i)}{d_i} \frac{1-(\frac{d_i}{b_i})^{\tau+1+\xi_i}}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}},$$

$$\begin{aligned}
E(\log \eta_i|\boldsymbol{x}_i) = (1-\delta_i)\left\{\psi(\tau+\xi_i) - \log b_i\right\} + \\
\delta_i \left\{ \psi(\tau+\xi_i) - \frac{b_i^{\tau+\xi_i}\log d_i - d_i^{\tau+\xi_i}\log b_i}{b_i^{\tau+\xi_i} - d_i^{\tau+\xi_i}} \right\},
\end{aligned}$$

$$\begin{aligned}
\operatorname{var}(\eta_i|\boldsymbol{x}_i) = (1-\delta_i)\frac{\tau+\xi_i}{b_i^2} + \\
\delta_i \left\{ \frac{(\tau+\xi_i)(\tau+\xi_i+1)}{d_i^2} \frac{1-(\frac{d_i}{b_i})^{\tau+\xi_i+2}}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}} - [\frac{(\tau+\xi_i)}{d_i} \frac{1-(\frac{d_i}{b_i})^{\tau+\xi_i+1}}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}}]^2 \right\},
\end{aligned}$$

$$\begin{aligned}
&\operatorname{var}(\log \eta_i|\boldsymbol{x}_i) = (1-\delta_i)\psi_1(\tau+\xi_i) + \\
&\delta_i \frac{\{\psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log d_i]^2\} - (\frac{d_i}{b_i})^{\tau+\xi_i}\{\psi_1(\tau+\xi_i) + [\psi(\tau+\xi_i) - \log b_i]^2\}}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}} \\
&- \delta_i[\psi(\tau+\xi_i) - \frac{\log d_i - (\frac{d_i}{b_i})^{\tau+\xi_i}\log b_i}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}}]^2,
\end{aligned}$$

$$\begin{aligned}
&\operatorname{cov}(\eta_i, \log \eta_i|\boldsymbol{x}_i) = (1-\delta_i) b_i^{-1} + \\
&\delta_i \left\{ \frac{(\tau+\xi_i)}{d_i} \frac{[\psi(\tau+\xi_i+1) - \log d_i] - (\frac{d_i}{b_i})^{\tau+\xi_i+1}[\psi(\tau+\xi_i+1) - \log b_i]}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}} \right. \\
&\left. - [\frac{(\tau+\xi_i)}{d_i} \frac{1-(\frac{d_i}{b_i})^{\tau+\xi_i+1}}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}}][\psi(\tau+\xi_i) - \frac{\log d_i - (\frac{d_i}{b_i})^{\tau+\xi_i}\log b_i}{1-(\frac{d_i}{b_i})^{\tau+\xi_i}}] \right\},
\end{aligned}$$

where

$$b_i = \tau + \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T) + \Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C),$$

$$d_i = \tau + \Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C).$$

B.3 The Conditional Distributions of Z_i and Z_{ij}

Since $\sum_{j=1}^{k} Z_{ij} = Z_i$ then given Z_i, Z_{ij} is multinomial distributed for $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$ then

$$E(Z_{ij}|\boldsymbol{x}_i) = E\left\{E(Z_{ij}|\boldsymbol{x}_i, Z_i)\right\} = E(\frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} Z_i|\boldsymbol{x}_i) = \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} E(Z_i|\boldsymbol{x}_i).$$

For Z_i, given η_i it follows a truncated Poisson distribution with a support of all positive integers when $\delta_i = 1$ and degenerates at 0 when $\delta_i = 0$ for $i = 1, 2, ..., n$. By applying the law of iterative rule again and let $\lambda_i = \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)$,

$$\begin{aligned}
E(Z_i|\boldsymbol{x}_i) &= E\left\{E(Z_i|\boldsymbol{x}_i, \eta_i)\right\} = E\left\{\frac{\eta_i \delta_i \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)}{1 - \exp\left[-\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\eta_i\right]}|\boldsymbol{x}_i\right\}, \\
E(Z_i|\boldsymbol{x}_i) &= E\left\{E(Z_i|\boldsymbol{x}_i, \eta_i)\right\} = E\left\{\frac{\eta_i \delta_i \lambda_i}{1 - \exp\left(-\lambda_i \eta_i\right)}|\boldsymbol{x}_i\right\}, \\
E(Z_i^2|\boldsymbol{x}_i) &= E\left\{E(Z_i^2|\boldsymbol{x}_i, \eta_i)\right\} = E\left\{\frac{\delta_i(\eta_i\lambda_i + \eta_i^2\lambda_i^2)}{1 - \exp\left(-\lambda_i \eta_i\right)}|\boldsymbol{x}_i\right\}.
\end{aligned}$$

Then

$$\begin{aligned}
\text{cov}(Z_{ij}, Z_{ij'}|\boldsymbol{x}_i) &= E(Z_{ij}Z_{ij'}|\boldsymbol{x}_i) - E(Z_{ij}|\boldsymbol{x}_i)E(Z_{ij'}|\boldsymbol{x}_i) \\
&= E\left\{E(Z_{ij}Z_{ij'}|\boldsymbol{x}_i, z_i)\right\} - E\left\{E(Z_{ij}|\boldsymbol{x}_i, z_i)\right\} E\left\{E(Z_{ij'}|\boldsymbol{x}_i, z_i)\right\} \\
&= E\left\{\text{cov}(Z_{ij}, Z_{ij'}|\boldsymbol{x}_i, z_i) + E(Z_{ij}|\boldsymbol{x}_i, z_i)E(Z_{ij'}|\boldsymbol{x}_i, z_i)\right\} \\
&\quad - E\left\{E(Z_{ij}|\boldsymbol{x}_i, z_i)\right\} E\left\{E(Z_{ij'}|\boldsymbol{x}_i, z_i)\right\} \\
&= E[-\frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} \frac{\gamma_{Tj'} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} Z_i + \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} \frac{\gamma_{Tj'} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} Z_i^2] \\
&\quad - \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} \frac{\gamma_{Tj'} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} [E(Z_i|\boldsymbol{x}_i)]^2 \\
&= -\frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} \frac{\gamma_{Tj'} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} E(Z_i|\boldsymbol{x}_i) + \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} \frac{\gamma_{Tj'} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} E(Z_i^2|\boldsymbol{x}_i) \\
&\quad - \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} \frac{\gamma_{Tj'} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} [E(Z_i|\boldsymbol{x}_i)]^2 \\
&= \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} \frac{\gamma_{Tj'} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} [-E(Z_i|\boldsymbol{x}_i) + E(Z_i^2|\boldsymbol{x}_i) - E(Z_i|\boldsymbol{x}_i)^2] \\
&= \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} \frac{\gamma_{Tj'} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} [-E(Z_i|\boldsymbol{x}_i) + \text{var}(Z_i|\boldsymbol{x}_i)],
\end{aligned}$$

$$\mathrm{var}(Z_{ij}|\boldsymbol{x}_i) = \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} E(Z_i|\boldsymbol{x}_i) + [\frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}]^2[-E(Z_i|\boldsymbol{x}_i) + \mathrm{var}(Z_i|\boldsymbol{x}_i)],$$

$$\begin{aligned}
\mathrm{cov}(Z_{ij}, Z_i|\boldsymbol{x}_i) &= E(Z_{ij}Z_i|\boldsymbol{x}_i) - E(Z_{ij}|\boldsymbol{x}_i)E(Z_i|\boldsymbol{x}_i) \\
&= E\left\{z_i E(Z_{ij}|\boldsymbol{x}, z_i)\right\} - \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} E(Z_i|\boldsymbol{x}_i)E(Z_i|\boldsymbol{x}_i) \\
&= E\left\{z_i \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} z_i\right\} - \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}[E(Z_i|\boldsymbol{x}_i)]^2 \\
&= \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} \mathrm{var}(Z_i|\boldsymbol{x}_i).
\end{aligned}$$

If $\delta_i = 0$, then $E(Z_i|\boldsymbol{x}_i)$ is zero.

If $\delta_i = 1$, then

$$\begin{aligned}
&E(Z_i|\boldsymbol{x}_i) = E\left\{\frac{\eta_i\delta_i\lambda_i}{1-\exp(-\lambda_i\eta_i)}|\boldsymbol{x}_i\right\} \\
&= \frac{\int \frac{\eta_i\delta_i\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T)}{1-\exp[-\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T)\eta_i]} L_1(\eta_i|\boldsymbol{x}_i)d\eta_i}{\int L_1(\eta_i|\boldsymbol{x}_i)d\eta_i} \\
&= \frac{\int \frac{\eta_i\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T)}{1-\exp[-\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T)\eta_i]} g(\eta_i|\tau)(1-S_T(\tilde{c}_i|\eta_i))\lambda_C(\tilde{c}_i|\eta_i)^{\xi_i} S_C(\tilde{c}_i|\eta_i)d\eta_i}{\int g(\eta_i|\tau)(1-S_T(\tilde{c}_i|\eta_i))\lambda_C(\tilde{c}_i|\eta_i)^{\xi_i} S_C(\tilde{c}_i|\eta_i)d\eta_i} \\
&= \frac{\int \eta_i\eta_i^{\tau-1}e^{-\eta_i\tau}\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T)\left\{\lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_C)\eta_i\right\}^{\xi_i} S_C(\tilde{c}_i|\eta_i)d\eta_i}{\int \eta_i^{\tau-1}e^{-\eta_i\tau}(1-S_T(\tilde{c}_i|\eta_i))\left\{\lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_C)\eta_i\right\}^{\xi_i} S_C(\tilde{c}_i|\eta_i)d\eta_i} \\
&= \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T)\frac{\int \eta_i^{\tau+1+\xi_i-1}e^{-\eta_i\tau}S_C(\tilde{c}_i|\eta_i)d\eta_i}{\int \eta_i^{\tau+\xi_i-1}e^{-\eta_i\tau}(1-S_T(\tilde{c}_i|\eta_i))S_C(\tilde{c}_i|\eta_i)d\eta_i} \\
&= \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T)\frac{\int \eta_i^{\tau+1+\xi_i-1}e^{-d_i\eta_i}d\eta_i}{\int \eta_i^{\tau+\xi_i-1}e^{-d_i\eta_i}d\eta_i - \int \eta_i^{\tau+\xi_i-1}e^{-b_i\eta_i}d\eta_i} \\
&= \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T)\frac{\Gamma(\tau+1+\xi_i)/d_i^{\tau+1+\xi_i}}{\Gamma(\tau+\xi_i)/d_i^{\tau+\xi_i} - \Gamma(\tau+\xi_i)/b_i^{\tau+\xi_i}} \\
&= \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T)(\tau+\xi_i)\frac{1/d_i^{\tau+1+\xi_i}}{1/d_i^{\tau+\xi_i} - 1/b_i^{\tau+\xi_i}} \\
&= \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T)\frac{\tau+\xi_i}{d_i}\frac{1/d_i^{\tau+\xi_i}}{1/d_i^{\tau+\xi_i} - 1/b_i^{\tau+\xi_i}} \\
&= \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_T)\frac{\tau+\xi_i}{d_i}\frac{b_i^{\tau+\xi_i}}{b_i^{\tau+\xi_i} - d_i^{\tau+\xi_i}},
\end{aligned}$$

$$
\begin{aligned}
E(Z_i^2|\boldsymbol{x}_i) &= E\left\{\frac{\delta_i(\eta_i\lambda_i+\eta_i^2\lambda_i^2)}{1-\exp\left(-\lambda_i\eta_i\right)}|\boldsymbol{x}_i\right\} \\
&= E\left\{\frac{\delta_i\eta_i\lambda_i}{1-\exp\left(-\lambda_i\eta_i\right)}|\boldsymbol{x}_i\right\} + E\left\{\frac{\delta_i\eta_i^2\lambda_i^2}{1-\exp\left(-\lambda_i\eta_i\right)}|\boldsymbol{x}_i\right\},
\end{aligned}
$$

$$
\begin{aligned}
E\left\{\frac{\delta_i\eta_i^2\lambda_i^2}{1-\exp\left(-\lambda_i\eta_i\right)}|\boldsymbol{x}_i\right\} &= \lambda_i^2\frac{\int \eta_i^{\tau+2+\xi_i-1}e^{-\eta_i\tau}S_C(\tilde{c}_i|\eta_i)d\eta_i}{\int \eta_i^{\tau+\xi_i-1}e^{-\eta_i\tau}(1-S_T(\tilde{c}_i|\eta_i))S_C(\tilde{c}_i|\eta_i)d\eta_i} \\
&= \lambda_i^2\frac{\int \eta_i^{\tau+2+\xi_i-1}e^{-d_i\eta_i}d\eta_i}{\int \eta_i^{\tau+\xi_i-1}e^{-d_i\eta_i}d\eta_i - \int \eta_i^{\tau+\xi_i-1}e^{-b_i\eta_i}d\eta_i} \\
&= \lambda_i^2\frac{\Gamma(\tau+2+\xi_i)/d_i^{\tau+2+\xi_i}}{\Gamma(\tau+\xi_i)/d_i^{\tau+\xi_i}-\Gamma(\tau+\xi_i)/b_i^{\tau+\xi_i}} \\
&= \lambda_i^2(\tau+\xi_i+1)(\tau+\xi_i)\frac{1/d_i^{\tau+2+\xi_i}}{1/d_i^{\tau+\xi_i}-1/b_i^{\tau+\xi_i}} \\
&= \lambda_i^2\frac{(\tau+\xi_i+1)(\tau+\xi_i)}{d_i^2}\frac{1/d_i^{\tau+\xi_i}}{1/d_i^{\tau+\xi_i}-1/b_i^{\tau+\xi_i}} \\
&= \lambda_i^2\frac{(\tau+\xi_i+1)(\tau+\xi_i)}{d_i^2}\frac{b_i^{\tau+\xi_i}}{b_i^{\tau+\xi_i}-d_i^{\tau+\xi_i}}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
E(Z_i|\boldsymbol{x}_i) &= \delta_i\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\frac{\tau+\xi_i}{d_i}\frac{b_i^{\tau+\xi_i}}{b_i^{\tau+\xi_i}-d_i^{\tau+\xi_i}}, \\
E(Z_i^2|\boldsymbol{x}_i) &= \delta_i[\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)]^2\frac{(\tau+\xi_i+1)(\tau+\xi_i)}{d_i^2}\frac{b_i^{\tau+\xi_i}}{b_i^{\tau+\xi_i}-d_i^{\tau+\xi_i}} + E(Z_i|\boldsymbol{x}_i), \\
\mathrm{var}(Z_i|\boldsymbol{x}_i) &= E(Z_i^2|\boldsymbol{x}_i) - [E(Z_i|\boldsymbol{x}_i)]^2, \\
\mathrm{cov}(Z_{ij}, Z_i|\boldsymbol{x}_i) &= \frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}\mathrm{var}(Z_i|\boldsymbol{x}_i), \\
\mathrm{cov}(Z_{ij}, Z_{ij'}|\boldsymbol{x}_i) &= \frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}\frac{\gamma_{Tj'}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}[-E(Z_i|\boldsymbol{x}_i)+\mathrm{var}(Z_i|\boldsymbol{x}_i)], \\
\mathrm{var}(Z_{ij}|\boldsymbol{x}_i) &= \frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}E(Z_i|\boldsymbol{x}_i) + [\frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}]^2[-E(Z_i|\boldsymbol{x}_i)+\mathrm{var}(Z_i|\boldsymbol{x}_i)],
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{cov}(Z_{ij}, Z_{ij'}|\boldsymbol{x}_i) = \\
&\quad \frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}\frac{\gamma_{Tj'}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}[-E(Z_i|\boldsymbol{x}_i)+\mathrm{var}(Z_i|\boldsymbol{x}_i)] + \frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}E(Z_i|\boldsymbol{x}_i)I(j=j'),
\end{aligned}
$$

where $b_i = \tau + \Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T) + \Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C), d_i = \tau + \Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)$. The expectation, covariance between Z_i and η_i, Z_{ij} and η_i are as follows,

$$
\begin{aligned}
\operatorname{cov}(Z_i, \eta_i|\boldsymbol{x}_i) &= E(Z_i\eta_i|\boldsymbol{x}_i) - E(Z_i|\boldsymbol{x}_i)E(\eta_i|\boldsymbol{x}_i), \\
E(Z_i\eta_i|\boldsymbol{x}_i) &= E[E(Z_i\eta_i|\boldsymbol{x}_i, \eta_i)] = E[\eta_i E(Z_i|\boldsymbol{x}_i, \eta_i)] \\
&= E[\eta_i \frac{\eta_i\delta_i\lambda_i}{1-\exp(-\lambda_i\eta_i)}|\boldsymbol{x}_i] = \delta_i\lambda_i \frac{(\tau+\xi_i+1)(\tau+\xi_i)}{d_i^2}\frac{b_i^{\tau+\xi_i}}{b_i^{\tau+\xi_i} - d_i^{\tau+\xi_i}},
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{cov}(Z_{ij}, \eta_i|\boldsymbol{x}_i) &= E(Z_{ij}\eta_i|\boldsymbol{x}_i) - E(Z_{ij}|\boldsymbol{x}_i)E(\eta_i|\boldsymbol{x}_i), \\
E(Z_{ij}\eta_i|\boldsymbol{x}_i) &= E[\eta_i E(Z_{ij}|\boldsymbol{x}_i, \eta_i)] = E\left\{\eta_i E[E(Z_{ij}|\boldsymbol{x}_i, \eta_i, Z_i)]\right\} \\
&= E\left\{\eta_i E[\frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}Z_i|\boldsymbol{x}_i, \eta_i]\right\} = \frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}E\left\{\eta_i E[Z_i|\boldsymbol{x}_i, \eta_i]\right\} \\
&= \frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}E(Z_{ij}\eta_i|\boldsymbol{x}_i), \\
\operatorname{cov}(Z_{ij}, \eta_i|\boldsymbol{x}_i) &= \frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}E(Z_{ij}\eta_i|\boldsymbol{x}_i) - \frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}E(Z_i|\boldsymbol{x}_i)E(\eta_i|\boldsymbol{x}_i) \\
&= \frac{\gamma_{Tj}I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)}\operatorname{cov}(Z_i, \eta_i|\boldsymbol{x}_i),
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{cov}(Z_i, \log\eta_i|\boldsymbol{x}_i) &= E(Z_i\log\eta_i|\boldsymbol{x}_i) - E(Z_i|\boldsymbol{x}_i)E(\log\eta_i|\boldsymbol{x}_i), \\
E(Z_i\log\eta_i|\boldsymbol{x}_i) &= E[E(Z_i\log\eta_i|\boldsymbol{x}_i, \eta_i)] = E[\log\eta_i E(Z_i|\boldsymbol{x}_i, \eta_i)] \\
&= E[\log\eta_i \frac{\eta_i\delta_i\lambda_i}{1-\exp(-\lambda_i\eta_i)}|\boldsymbol{x}_i] \\
&= \delta_i\lambda_i \frac{\int \log\eta_i \eta_i^{\tau+1+\xi_i-1}e^{-d_i\eta_i}d\eta_i}{\int \eta_i^{\tau+\xi_i-1}e^{-d_i\eta_i}d\eta_i - \int \eta_i^{\tau+\xi_i-1}e^{-b_i\eta_i}d\eta_i} \\
&= \delta_i\lambda_i \frac{(\tau+\xi_i)[\psi(\tau+\xi+1) - \log d_i]}{d_i}\frac{b_i^{\tau+\xi_i}}{b_i^{\tau+\xi_i} - d_i^{\tau+\xi_i}},
\end{aligned}
$$

$$\operatorname{cov}(Z_{ij}, \log \eta_i | \boldsymbol{x}_i) = E(Z_{ij} \log \eta_i | \boldsymbol{x}_i) - E(Z_{ij} | \boldsymbol{x}_i) E(\log \eta_i | \boldsymbol{x}_i),$$

$$\begin{aligned}
E(Z_{ij} \log \eta_i | \boldsymbol{x}_i) &= E[\log \eta_i E(Z_{ij} | \boldsymbol{x}_i, \eta_i)] = E\left\{\log \eta_i E[E(Z_{ij} | \boldsymbol{x}_i, \eta_i, Z_i)]\right\} \\
&= E\left\{\log \eta_i E[\frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} Z_i | \boldsymbol{x}_i, \eta_i]\right\} = \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} E\left\{\log \eta_i E[Z_i | \boldsymbol{x}_i, \eta_i]\right\} \\
&= \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} E(Z_{ij} \log \eta_i | \boldsymbol{x}_i), \\
\operatorname{cov}(Z_{ij}, \log \eta_i | \boldsymbol{x}_i) &= \frac{\gamma_{Tj} I(\tilde{c}_i)}{\Lambda_{T0}(\tilde{c}_i)} \operatorname{cov}(Z_i, \log \eta_i | \boldsymbol{x}_i).
\end{aligned}$$

B.4 Variance Estimation with Louis's Method

Louis (1982) gives a formula as follows,

$$\hat{I}(\hat{\boldsymbol{\theta}}) = -\frac{\partial^2 \log \mathcal{L}_{obs}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} | \boldsymbol{\theta} = \hat{\boldsymbol{\theta}},$$

where

$$-\frac{\partial^2 \log \mathcal{L}_{obs}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \operatorname{var}\left\{\frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}.$$

To evaluate the first term $\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$, the $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ function is as follows,

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = H_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) + H_2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) + H_3(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) + H_4(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}).$$

The term $\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ can be obtained as follows,

$$\begin{aligned}
H_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) &= -n \log\left\{\Gamma(\tau)\right\} + n\tau \log(\tau) + \tau \sum_{i=1}^{n} [E(\log \eta_i) - E(\eta_i)], \\
\frac{\partial H_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \tau} &= -n\psi(\tau) + n \log(\tau) + n + \sum_{i=1}^{n} [E(\log \eta_i) - E(\eta_i)], \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \tau \partial \tau} &= \frac{\partial^2 H_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \tau \partial \tau} = -n\psi_1(\tau) + n\tau^{-1}.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_C} &= \frac{\partial H_2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_C} = \sum_{i=1}^{n} [-\Lambda_{C0}(\tilde{c}_i) \exp(\boldsymbol{x}_{\boldsymbol{i}}' \boldsymbol{\beta}_C) E(\eta_i) + \xi_i] \boldsymbol{x}_{\boldsymbol{i}}, \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_C \partial \boldsymbol{\beta}_C'} &= \frac{\partial^2 H_2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_C \partial \boldsymbol{\beta}_C'} = \sum_{i=1}^{n} -\Lambda_{C0}(\tilde{c}_i) \exp(\boldsymbol{x}_{\boldsymbol{i}}' \boldsymbol{\beta}_C) E(\eta_i) \boldsymbol{x}_{\boldsymbol{i}} \boldsymbol{x}_{\boldsymbol{i}i}', \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_C \partial \gamma_{Cj}} &= \frac{\partial^2 H_2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_C \partial \gamma_{Cj}} = \sum_{i=1}^{n} -I_j(\tilde{c}_i) \exp(\boldsymbol{x}_{\boldsymbol{i}}' \boldsymbol{\beta}_C) E(\eta_i) \boldsymbol{x}_{\boldsymbol{i}}.
\end{aligned}$$

$$\frac{\partial Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \gamma_{Cj}} = \frac{\partial H_2(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \gamma_{Cj}} = \sum_{i=1}^{n}[-I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)E(\eta_i) + E(V_{ij})\xi_i\gamma_{Cj}^{-1}],$$

$$\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \gamma_{Cj}\partial \gamma_{Cj'}} = \frac{\partial^2 H_2(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \gamma_{Cj}\partial \gamma_{Cj'}} = \sum_{i=1}^{n} -\gamma_{Cj}^{-2}\xi_i E(V_{ij})I(j=j').$$

$$\frac{\partial Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_T} = \frac{\partial H_3(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_T} = \sum_{i=1}^{n}[-\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)E(\eta_i) + E(Z_i)]\boldsymbol{x_i},$$

$$\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_T\partial \boldsymbol{\beta}_T'} = \frac{\partial^2 H_3(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_T\partial \boldsymbol{\beta}_T'} = \sum_{i=1}^{n} -\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)E(\eta_i)\boldsymbol{x_i}\boldsymbol{x_i'}{}_i,$$

$$\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_T\partial \gamma_{Tj}} = \frac{\partial^2 H_3(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}_T\partial \gamma_{Tj}} = \sum_{i=1}^{n} -I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)E(\eta_i)\boldsymbol{x_i}.$$

$$\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \gamma_{Tj}} = \frac{\partial H_3(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \gamma_{Tj}} = \sum_{i=1}^{n}[-I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)E(\eta_i) + E(Z_{ij})\gamma_{Tj}^{-1}],$$

$$\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \gamma_{Tj}\partial \gamma_{Tj'}} = \frac{\partial^2 H_3(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})}{\partial \gamma_{Tj}\partial \gamma_{Tj'}} = \sum_{i=1}^{n} -\gamma_{Tj}^{-2}E(Z_{ij})I(j=j').$$

The term var $\left\{\frac{\partial \log \mathcal{L}_{com}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}$ can be obtained as follows,

$$\frac{\partial \log(\mathcal{L}_{com})}{\partial \tau} = -n\psi(\tau) + n + n\log(\tau) + \sum_{i=1}^{n}[\log \eta_i - \eta_i],$$

$$\begin{aligned}\frac{\partial \log(\mathcal{L}_{com})}{\partial \boldsymbol{\beta}_C} &= \sum_{i=1}^{n}[-\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)\eta_i\boldsymbol{x_i} + \xi_i\boldsymbol{x_i}] \\ &= \sum_{i=1}^{n}[-\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)\eta_i + \xi_i]\boldsymbol{x_i},\end{aligned}$$

$$\frac{\partial \log(\mathcal{L}_{com})}{\partial \gamma_{Cj}} = \sum_{i=1}^{n}[-I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)\eta_i + V_{ij}\xi_i\gamma_{Cj}^{-1}],$$

$$\begin{aligned}\frac{\partial \log(\mathcal{L}_{com})}{\partial \boldsymbol{\beta}_T} &= \sum_{i=1}^{n}\sum_{j=1}^{k}[-\gamma_{Tj}I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\eta_i + Z_{ij}]\boldsymbol{x_i} \\ &= \sum_{i=1}^{n}[-\Lambda_{T0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\eta_i + Z_i]\boldsymbol{x_i},\end{aligned}$$

$$\frac{\partial \log(\mathcal{L}_{com})}{\partial \gamma_{Tj}} = \sum_{i=1}^{n}[-I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\eta_i + Z_{ij}\gamma_{Tj}^{-1}].$$

$$\begin{aligned}
&\text{var}\left\{\frac{\partial \log(\mathcal{L}_{com})}{\partial \tau}\right\} = \text{var}\left\{\sum_{i=1}^{n}(\log \eta_i - \eta_i)\right\} = \sum_{i=1}^{n}\text{var}(\log \eta_i - \eta_i) \\
&= \sum_{i=1}^{n}\left\{\text{var}(\log \eta_i) + \text{var}(\eta_i) - 2\text{cov}(\log \eta_i, \eta_i)\right\}, \\
&\text{cov}\left\{\frac{\partial \log(\mathcal{L}_{com})}{\partial \tau}, \frac{\partial \log(\mathcal{L}_{com})}{\partial \boldsymbol{\beta}_C}\right\} = \sum_{i=1}^{n} -\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)\boldsymbol{x_i}[\text{cov}(\log \eta_i, \eta_i) - \text{var}(\eta_i)], \\
&\text{cov}\left\{\frac{\partial \log(\mathcal{L}_{com})}{\partial \tau}, \frac{\partial \log(\mathcal{L}_{com})}{\partial \boldsymbol{\beta}_T}\right\} = \sum_{i=1}^{n} -\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)\boldsymbol{x_i}[\text{cov}(\log \eta_i, \eta_i) - \text{var}(\eta_i)] \\
&\qquad + \sum_{i=1}^{n}\boldsymbol{x_i}[\text{cov}(\log \eta_i, Z_i) - \text{cov}(\eta_i, Z_i)], \\
&\text{cov}\left\{\frac{\partial \log(\mathcal{L}_{com})}{\partial \tau}, \frac{\partial \log(\mathcal{L}_{com})}{\partial \gamma_{Cj}}\right\} = \sum_{i=1}^{n} -I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)[\text{cov}(\log \eta_i, \eta_i) - \text{var}(\eta_i)], \\
&\text{cov}\left\{\frac{\partial \log(\mathcal{L}_{com})}{\partial \tau}, \frac{\partial \log(\mathcal{L}_{com})}{\partial \gamma_{Tj}}\right\} = \sum_{i=1}^{n} -I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_T)[\text{cov}(\log \eta_i, \eta_i) - \text{var}(\eta_i)] \\
&\qquad + \sum_{i=1}^{n}\gamma_{Tj}^{-1}[\text{cov}(\log \eta_i, Z_{ij}) - \text{cov}(\eta_i, Z_{ij})].
\end{aligned}$$

$$\begin{aligned}
\text{var}\left\{\frac{\partial \log(\mathcal{L}_{com})}{\partial \boldsymbol{\beta}_C}\right\} &= \sum_{i=1}^{n}[\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)\boldsymbol{x_i}]^2\text{var}(\eta_i), \\
\text{cov}\left\{\frac{\partial \log(\mathcal{L}_{com})}{\partial \boldsymbol{\beta}_C}, \frac{\partial \log(\mathcal{L}_{com})}{\partial \gamma_{Cj}}\right\} &= \sum_{i=1}^{n}\Lambda_{C0}(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)\boldsymbol{x_i}I_j(\tilde{c}_i)\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)\text{var}(\eta_i), \\
\text{cov}\left\{\frac{\partial \log(\mathcal{L}_{com})}{\partial \gamma_{Cj}}, \frac{\partial \log(\mathcal{L}_{com})}{\partial \gamma_{Cj'}}\right\} &= \\
&\sum_{i=1}^{n}\Big\{(\gamma_{Cj}\gamma_{Cj'})^{-1}\xi_i\text{cov}(V_{ij}, V_{ij'}) + I_j(\tilde{c}_i)I_{j'}(\tilde{c}_i)[\exp(\boldsymbol{x_i'}\boldsymbol{\beta}_C)]^2\text{var}(\eta_i).
\end{aligned}$$

Appendix C

Chapter 4 Supplementary Materials

Since $\sum_{l=1}^{p} Z_{ijl} = Z_{ij}$, then the conditional distribution of Z_{ijl} is multinomial distributed given Z_{ij}. Therefore one can obtain the following relationship by applying the law of iterative rule.

$$E(Z_{ijl}|\boldsymbol{x}) = E\left\{E(Z_{ijl}|\boldsymbol{x}, Z_{ij})\right\} = E\left\{\frac{\boldsymbol{\gamma}_{jl} I_{jl}(c_i)}{\Lambda_{0j}(c_i)} Z_{ij}|\boldsymbol{x}\right\} = \frac{\boldsymbol{\gamma}_{jl} I_{jl}(c_i)}{\Lambda_{0j}(c_i)} E(Z_{ij}|\boldsymbol{x}).$$

For the latent variable Z_{ij}, given η_i it follows a truncated Poisson distribution with when $\delta_i = 1$ it takes all positive integers, when $\delta_i = 0$ it is 0. By applying the law of iterative rule again,

$$E(Z_{ij}|\boldsymbol{x}) = E\left\{E(Z_{ij}|\boldsymbol{x}, \eta_i)\right\} = E\left\{\frac{\eta_i \delta_{ij} \Lambda_{0j}(c_i) \exp(\boldsymbol{x'}\boldsymbol{\beta}_j)}{1 - \exp\left[-\Lambda_{0j}(c_i) \exp(\boldsymbol{x'}\boldsymbol{\beta}_j)\eta_i\right]}|\boldsymbol{x}\right\}.$$

Let $H_{ij} = \Lambda_{0j}(c_i) \exp(\boldsymbol{x'}\boldsymbol{\beta}_j)$. When $\delta_{ij} = 1$,

$$E\left\{\frac{\eta_i H_{ij}}{1 - \exp\left(-H_{ij}\eta_i\right)}|\boldsymbol{x}\right\} = \frac{\int \frac{\eta_i H_{ij}}{1-\exp\left(-H_{ij}\eta_i\right)} \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i}{\int \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i},$$

where

$$\mathcal{L}_1(\boldsymbol{\theta}) = g(\eta_i|\tau, \tau) \prod_{j=1}^{k} \left\{1 - S_j(c_i|x)\right\}^{\delta_{ij}} S_j(c_i|x)^{1-\delta_{ij}}.$$

The numerator is,

$$
\begin{aligned}
&\int \frac{\eta_i H_{ij}}{1-\exp(-H_{ij}\eta_i)} \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i \\
=&\int \frac{\eta_i H_{ij}}{1-\exp(-H_{ij}\eta_i)} g(\eta_i|\tau,\tau) \prod_{r=1}^{k} \{1 - S_r(c_i|x)\}^{\delta_{ir}} S_r(c_i|x)^{1-\delta_{ir}} d\eta_i \\
=&\int \frac{g(\eta_i|\tau+1,\tau) H_{ij}}{1-\exp(-H_{ij}\eta_i)} \prod_{r=1}^{k} \{1 - \exp(-H_{ir}\eta_i)\}^{\delta_{ir}} \exp(-H_{ir}\eta_i)^{1-\delta_{ir}} d\eta_i \\
=&H_{ij} \int g(\eta_i|\tau+1,\tau) \prod_{\substack{r=1 \\ r\neq j}}^{k} \{1 - \exp(-H_{ir}\eta_i)\}^{\delta_{ir}} \exp(-H_{ir}\eta_i)^{1-\delta_{ir}} d\eta_i \\
=&H_{ij} \int g(\eta_i|\tau+1,\tau) \sum_{q=1}^{2^{d_i-1}} (-1)^{|L_i|-|B_{iq}|-1} \exp(-\eta_i \sum_{j\in B_{iq}^C} H_{ij}) d\eta_i \\
=&H_{ij} \sum_{q=1}^{2^{d_i-1}} (-1)^{|L_i|-|B_{iq}|-1} (1+\tau^{-1} \sum_{j\in B_{iq}^C} H_{ij})^{-\tau-1},
\end{aligned}
$$

where B_{iq} is the qth element of the set containing all subsets of $L_i \setminus \{j\}$, B_{iq}^C is the complement of B_{iq} with the complete set being $\{1,2,..,k\} \setminus \{j\}$, for $q = 1,2,...,2^{d_i-1}$.

The denominator is,

$$
\begin{aligned}
\int \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i &= \int \frac{\tau^\tau}{\Gamma(\tau)} \eta_i^{\tau-1} e^{-\tau\eta_i} \sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} \exp(-\eta_i \sum_{j\in A_{ip}^C} H_{ij}) d\eta_i \\
&= \sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} \int \frac{\tau^\tau}{\Gamma(\tau)} \eta_i^{\tau-1} \exp\left\{-(\tau + \sum_{j\in A_{ip}^C} H_{ij})\eta_i\right\} d\eta_i \\
&= \sum_{p=1}^{2^{d_i}} \frac{(-1)^{|L_i|-|A_{ip}|}\tau^\tau}{(\tau + \sum_{j\in A_{ip}^C} H_{ij})^\tau} \\
&= \sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} (1+\tau^{-1} \sum_{j\in A_{ip}^C} H_{ij})^{-\tau},
\end{aligned}
$$

where A_{ip} is the pth element of the set containing all subsets of L_i, A_{ip}^C is the complement of A_{ip}, for $p = 1,2,...,2^{d_i}$ and the complete set is $\{1,2,..,k\}$.

Therefore,

$$
\begin{aligned}
E(Z_{ij}|\boldsymbol{x}) &= \delta_{ij} \frac{\int \frac{\eta_i H_{ij}}{1-\exp(-H_{ij}\eta_i)} \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i}{\int \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i} \\
&= \delta_{ij} H_{ij} \frac{\sum_{p=1}^{2^{d_i-1}} (-1)^{|L_i|-|B_{iq}|-1} (1+\tau^{-1} \sum_{r\in B_{iq}^C} H_{ir})^{-\tau-1}}{\sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} (1+\tau^{-1} \sum_{r\in A_{ip}^C} H_{ir})^{-\tau}}.
\end{aligned}
$$

The conditional expectation of η_i and $\log \eta_i$ given data, $E(\eta_i|\boldsymbol{x})$ and $E(\log \eta_i|\boldsymbol{x})$, can be expressed as follows,

$$E(\eta_i|\boldsymbol{x}) = \frac{\int \eta_i \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i}{\int \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i}, \quad E(\log \eta_i|\boldsymbol{x}) = \frac{\int \log \eta_i \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i}{\int \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i}.$$

$$\begin{aligned}
\int \eta_i \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i &= \int \eta_i g(\eta_i|\tau,\tau) \prod_{j=1}^{k} \{1 - S_j(c_i|\mathbf{x}_i, \eta_i)\}^{\delta_{ij}} S_j(c_i|\mathbf{x}_i, \eta_i)^{1-\delta_{ij}} d\eta_i \\
&= \int \frac{\tau^\tau}{\Gamma(\tau)} \eta_i^\tau e^{-\tau\eta_i} \sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} \exp(-\eta_i \sum_{j \in A_{ip}^C} H_{ij}) d\eta_i \\
&= \frac{\tau^\tau}{\Gamma(\tau)} \sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} \int \eta_i^\tau \exp\left\{-(\tau + \sum_{j \in A_{ip}^C} H_{ij})\eta_i\right\} d\eta_i \\
&= \tau^{\tau+1} \sum_{p=1}^{2^{d_i}} \int \frac{(-1)^{|L_i|-|A_{ip}|}}{\Gamma(\tau+1)} \eta_i^\tau \exp\left\{-(\tau + \sum_{j \in A_{ip}^C} H_{ij})\eta_i\right\} d\eta_i \\
&= \sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} (1 + \tau^{-1} \sum_{j \in A_{ip}^C} H_{ij})^{-\tau-1}.
\end{aligned}$$

$$\begin{aligned}
&\int \log \eta_i \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i \\
&= \int \log \eta_i g(\eta_i|\tau,\tau) \prod_{j=1}^{k} \{1 - S_j(c_i|\mathbf{x}_i, \eta_i)\}^{\delta_{ij}} S_j(c_i|\mathbf{x}_i, \eta_i)^{1-\delta_{ij}} d\eta_i \\
&= \int \frac{\tau^\tau}{\Gamma(\tau)} (\log \eta_i) \eta_i^{\tau-1} e^{-\tau\eta_i} \sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} \exp(-\eta_i \sum_{j \in A_{ip}^C} H_{ij}) d\eta_i \\
&= \sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} \int \frac{\tau^\tau}{\Gamma(\tau)} (\log \eta_i) \eta_i^{\tau-1} \exp\left\{-(\tau + \sum_{j \in A_{ip}^C} H_{ij})\eta_i\right\} d\eta_i \\
&= \tau^\tau \sum_{p=1}^{2^{d_i}} \int \log \eta_i \frac{1}{\Gamma(\tau)} \eta_i^{\tau-1} \exp\left\{-(\tau + \sum_{j \in A_{ip}^C} H_{ij})\eta_i\right\} d\eta_i \\
&= \sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} (1 + \tau^{-1} \sum_{j \in A_{ip}^C} H_{ij})^{-\tau} \left[\psi(\tau) - \log\left\{(\tau + \sum_{j \in A_{ip}^C} H_{ij})\eta_i\right\}\right].
\end{aligned}$$

Therefore,

$$E(\eta_i|\boldsymbol{x}) = \frac{\int \eta_i \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i}{\int \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i} = \frac{\sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} (1 + \tau^{-1} \sum_{j \in A_{ip}^C} H_{ij})^{-\tau-1}}{\sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} (1 + \tau^{-1} \sum_{j \in A_{ip}^C} H_{ij})^{-\tau}},$$

$$
\begin{aligned}
E(\log \eta_i|\boldsymbol{x}) &= \frac{\int \eta_i \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i}{\int \mathcal{L}_1(\boldsymbol{\theta}) d\eta_i} \\
&= \frac{\sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} (1 + \tau^{-1} \sum_{j \in A_{ip}^C} H_{ij})^{-\tau} \left[\psi(\tau) - \log(\tau + \sum_{j \in A_{ip}^C} H_{ij})\right]}{\sum_{p=1}^{2^{d_i}} (-1)^{|L_i|-|A_{ip}|} (1 + \tau^{-1} \sum_{j \in A_{ip}^C} H_{ij})^{-\tau}}.
\end{aligned}
$$