Semiparametric inference for proportional mean residual life models with clustered survival data

Huang, Rui

2020

Huang, R. (2020). Semiparametric inference for proportional mean residual life models with clustered survival data. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/143100>

<https://doi.org/10.32657/10356/143100>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 20:57:58 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**SEMIPARAMETRIC INFERENCE FOR
PROPORTIONAL MEAN RESIDUAL LIFE
MODELS WITH CLUSTERED SURVIVAL
DATA**

HUANG RUI

**SCHOOL OF PHYSICAL AND MATHEMATICAL
SCIENCES**

2020

**SEMIPARAMETRIC INFERENCE
FOR PROPORTIONAL MEAN
RESIDUAL LIFE MODELS WITH
CLUSTERED SURVIVAL DATA**

HUANG RUI

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2020

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

16 Jan 2020

.....
Date

.....
HUANG RUI

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

16 Jan 2020

.....
Date

.....
XIANG LIMING

Authorship Attribution Statement

Chapter 2 of this thesis contains materials from the following published paper at a peer-reviewed journal, where I am listed as the first author.

- Huang Rui, Xiang Liming and Ha Il Do. Frailty proportional mean residual life regression for clustered survival data: A hierarchical quasi-likelihood method. *Statistics in Medicine*, 38:4854–4870, 2019. DOI:10.1002/sim.8338.

The contributions of the co-authors are as follows:

Prof. Xiang suggested the study area, proposed the research topic, provided meritorious instruction throughout the study of this topic, and offered incredible help on revising the manuscript draft.

I derived all methodological results and performed all simulation studies under guidance of Prof. Xiang.

Prof. Ha guided the development of hierarchical likelihood estimation for variance component.

I drafted the initial version of the paper, which was edited and finalized by Prof. Xiang and Prof. Ha.

16 Jan 2020

.....
Date

.....
HUANG RUI

Abstract

Mean residual life regression models attract considerable attention in the recent literature of survival data analysis due to their straightforward and easy-to-understand interpretations. All existing works on the regression analysis of mean residual life focus on independent survival data. Clustered data are frequently encountered in biomedical studies, such as multi-center clinical trials where patient information is collected from different hospitals and medical institutions, leading to potential correlations within clusters while independence between clusters. Ignoring such heterogeneity among data would result in invalid inference especially when the association between individuals from the same cluster is strong. This thesis aims to develop novel methodologies for the analysis of clustered survival data based on the mean residual life models.

To address the association among survival times within clusters, a frailty proportional mean residual life regression model is firstly proposed. A hierarchical inference approach is developed by introducing the technique of estimating equations and the idea of hierarchical likelihood into the construction of the hierarchical quasi-likelihood function, resulting in quasi-score estimating equations for both fixed and random effects. To complete the estimation procedure, the frailty component is estimated by maximizing an adjusted hierarchical quasi-likelihood function. This proposed hierarchical inference procedure can overcome the computational intractability caused by the mean residual life and high-dimensional random effects. It also allows any parametric distribution for the frailty term whereas the explicit distribution assumption is usually hard to be verified in

practice.

Further, a more flexible model framework based on the proportional mean residual life regression is proposed for the analysis of clustered survival data. To make inferences about the extended model, penalized quasi-score estimating equations are derived from the penalized conditional quasi-likelihood function, which incorporates a penalty term into the conditional quasi-likelihood function to ensure the identifiability of parameters in the model. The penalized conditional quasi-likelihood method sets the distributional assumption for random effects free and makes the theoretical justification for consistency and asymptotic normality properties of the estimated fixed effects possible. Thus, the extended frailty model allowing non-intercept random effects is expected to have a broader range of applications.

Extensive simulation studies have been conducted to demonstrate the efficiency of both proposed methodologies, leading to results with favorable finite sample performance.

For the purpose of clarifying the utility of the proposed models and inference procedures, an example of breast cancer data collected from multi-center trials is applied to the real data analysis.

Acknowledgement

The four-year PhD study will come to an end. During this period, there have been happiness, sorrows, enjoyments after success, and frustrations after failures. But when I look back on these, I knew all was worthy.

First of all, I would avail myself of this opportunity to give my heartfelt thanks to my supervisor Prof. Xiang Liming, who provided me such a wonderful opportunity to pursue the heights of science. She is rigorous and demanding in my research, every small error won't escape from her supervision. Meanwhile, she is patient and inspiring, whenever I need help, she can always encourage me and give me useful advice. Her meticulous instruction would be an invaluable treasure for my future academic life. Besides, she cares about my after-school life and my career plan, tries her best to prepare me for my job pursuit. In one word, she is a scholarly mentor and beneficial friend.

I would like to express my sincere thanks to the current and previous group members: Wang Shanshan, Peng Mengjiao, Yu Tonghui, Liu Xiaoyu and Zsolt Szab, for their friendship and help. Thank you guys for always being friendly and helpful, which makes the four-year hard work full of happiness and memories.

My heartfelt thanks also go to many friends: Gao Yang, Liu Shu, Yang Qing, Zeng Neng, Yang Zhiguo, Liu Yiming, Yuan Chen, Dharmaliangam Vijai Anand and Ng Tze Yang Justin. You brought lots of fun to my stressful PhD life. I cannot imagine how boring my research will be without you.

In particular, I would like to express my appreciation to the division management of MAS: Prof. Ling San, Prof. Chan Song Heng, Prof. Privault Nicolas, Prof.

Oggier Frandrique Elise, and Prof. Ku Cheng Yeaw. I am also grateful to Prof. Xing Chaoping, Prof. Pan Guangming, Prof. Wang Li-lian, Prof. Wu Guohua, Prof. Peyrin Thomas, Prof. Bei Xiaohui, Prof. Xia Kelin and Prof. Tong Ping for their interesting courses and rigorous scientific attitude. In addition, I am also would like to show my appreciation to SPMS IT team and all administration staff for their support and help.

I wish to acknowledge SPMS and NTU to provide me the research scholarship as my financial support.

Special thanks would go to Meng Zhenyu, who puts a lot of effort during the preparation of my thesis. Your companion gives me the courage to move forward.

Last but not least, my greatest thankfulness would go to my families for supporting me spiritually. Thank you for your love and continued encouragement.

The help from all my beloved reminds me to endeavor for a higher peak. I will not let you down. Thanks again for all of you.

Contents

Abstract	1
Acknowledgement	3
List of Abbreviations and Symbols	7
List of Figures	9
List of Tables	11
1 Introduction	13
1.1 Basic Concepts in Survival Analysis	13
1.2 Nonparametric Estimation	15
1.3 Semiparametric Regression Models	16
1.4 Clustered Data and Frailty Models	19
1.5 Conditional Inference	22
1.6 Outline	25
2 Hierarchical Inference for Frailty Proportional Mean Residual Life Models	27
2.1 Methodology and Estimation Procedures	28
2.1.1 Model formulation	28
2.1.2 Maximum hierarchical quasi-likelihood estimation	31

2.1.3	Fitting algorithm	35
2.2	Simulation Studies	37
2.3	Real Data Analysis	49
3	Conditional Inference for Clustered Survival Data Based on Mean Residual Life Models	63
3.1	Modeling Methodologies	64
3.1.1	Model elaboration	64
3.1.2	Penalized conditional quasi-likelihood estimation	65
3.2	Implementation	67
3.3	Notation and Conditions	68
3.3.1	Derivatives and notation	68
3.3.2	Regularity conditions and lemma	70
3.4	Theoretical Results	72
3.5	Simulation Studies	78
3.6	Application to Breast Cancer Data	85
4	Conclusion and Future Work	101
4.1	Main Contributions	101
4.2	Future Work	103
	Bibliography	107

List of Abbreviations and Symbols

Abbreviations

AHQL	Adjusted Hierarchical Quasi-Likelihood
BIC	Bayesian Information Criterion
GEE	Generalized Estimating Equation
GLMM	Generalized Linear Mixed Model
HL	Hierarchical Likelihood
HQL	Hierarchical Quasi-Likelihood
IPCW	Inverse Probability of Censoring Weighting
KM	Kaplan-Meier
LASSO	Least Absolute Shrinkage and Selection Operator
ML	Maximum Likelihood
MPCQLE	Maximum Penalized Conditional Quasi-Likelihood Estimator
MRL	Mean Residual Life
PCQL	Penalized Conditional Quasi-Likelihood
PH	Proportional Hazards
PMRL	Proportional Mean Residual Life
QIF	Quadratic Inference Function
SCAD	Smoothly Clipped Absolute Deviation

Statistical Symbols

Cov	Covariance
$\det$	Determinant
$\exp$	Exponential
i.i.d.	Independent and Identically Distributed
$\inf$	Infimum (greatest lower bound)
$\lim$	Limit
$\log$	Logarithm of base e
$\max$	Maximum
$\min$	Minimum
$\ \cdot\ _2$	$\mathbf{L}_2$ norm for a vector
$\ \cdot\ _F$	Frobenius norm of a matrix
$\xrightarrow{d}$	Convergence in distribution
$\xrightarrow{p}$	Convergence in probability
$I(\cdot)$	Indicator function

$\mathbf{a}^T, \mathbf{A}^T$ Transpose of a vector $\mathbf{a}$ or matrix $\mathbf{A}$

$x^{-1}, \mathbf{A}^{-1}$ Multiplicative inverse of a number x or inverse of a matrix $\mathbf{A}$

Suppose $\{\mathbf{a}_n\}$ and $\{\mathbf{X}_n\}$ are a nonstochastic sequence of real number vectors and a sequence of random vectors for integers n , respectively. Then

$\mathbf{a}_n = \mathcal{O}(1)$ $\{\mathbf{a}_n\}$ is bounded

$\mathbf{a}_n = \mathcal{O}(n)$ $\mathbf{a}_n/n = \mathcal{O}(1)$

$\mathbf{X}_n = \mathcal{O}_p(1)$ $\mathbf{X}_n$ is bounded in probability as $n \rightarrow \infty$

$\mathbf{X}_n = \mathcal{O}_p(n)$ $\mathbf{X}_n/n = \mathcal{O}_p(1)$

$\mathbf{X}_n = o_p(1)$ $\mathbf{X}_n$ converges in probability to 0 as $n \rightarrow \infty$

List of Figures

2.1	Real data analysis for breast cancer data with focus on the overall survival time: Kaplan-Meier survival curves of the censoring time by categorized tumor size and grade.	55
2.2	Real data analysis for breast cancer data with focus on the overall survival time: plots of log cumulative hazard functions (left panel) and log mean residual life functions (right panel) across two different levels of size.	55
2.3	Real data analysis for breast cancer data with focus on the overall survival time: plots of A estimated baseline mean residual life function based on the proposed model and B comparison between the log baseline mean residual life function and log marginal mean residual life function.	56
2.4	Real data analysis for breast cancer data with focus on the overall survival time: plot of 95% confidence intervals for random effects in the proposed model over different centers, where the center number 1 stands for CAL, 2 for NKI, 3 for PNC, 4 for STNO, 5 for TRANSBIG, 6 for UCSF and 7 for UNC.	57

3.1	Application to distant metastasis-free survival: survival curves based on the KM estimates for censoring times over different co- variates.	94
3.2	Application to distant metastasis-free survival: curves of log cu- mulative hazards (left panel) and log mean residual life functions (right panel) over different classes of categorized size.	95
3.3	Application to distant metastasis-free survival: curves of log cu- mulative hazards (left panel) and log mean residual life functions (right panel) over different classes of node.	95
3.4	Application to distant metastasis-free survival: curves of log cu- mulative hazards (left panel) and log mean residual life functions (right panel) over different classes of grade.	96
3.5	Application to distant metastasis-free survival: curves of log mean residual life functions over different classes of node across 15 clusters.	97
3.6	Application to distant metastasis-free survival: curves of log mean residual life functions over different classes of grade across 15 clusters.	98
3.7	Application to distant metastasis-free survival: plots of the pro- posed estimated baseline mean residual life function (left panel) and the estimated log baseline and marginal mean residual life functions (right panel).	99
3.8	Application to distant metastasis-free survival: plot of 95% boot- strap confidence intervals for the predicted random effects across 15 clusters, where 1 denotes for CAL, 2 for DFHCC, 3 for MAINZ, 4 for NCI, 5 for NKI, 6 for PNC, 7 for STK, 8 for SUPER- TAM_HGU133A; 9 for SUPERTAM_HGU133PLUS2, 10 for TRANS- BIG, 11 for UCSF, 12 for UNC, 13 for UNT, 14 for UPP and 15 for VDX.	100

List of Tables

2.1	Simulation results using hierarchical quasi-likelihood method based on frailty and conventional PMRL models: lognormal distributed frailty with true variance parameter $\alpha=0.0, 0.25, 0.5$	41
2.2	Simulation results using hierarchical quasi-likelihood method based on frailty and conventional PMRL models: gamma distributed frailty with true variance parameter $\alpha=0.0, 0.25, 0.5$	43
2.3	Simulation results using hierarchical quasi-likelihood method based on frailty and conventional PMRL models: mixture normal distributed log-frailty with $2/3N(0, 0.25)+1/3N(0, 1)$	45
2.4	Simulation results using hierarchical quasi-likelihood method: obtained by a misspecified gamma frailty PMRL model.	46
2.5	Simulation results using hierarchical quasi-likelihood method: obtained by a misspecified lognormal frailty PMRL model.	47
2.6	Simulation results using hierarchical quasi-likelihood method based on the lognormal frailty PMRL model with the true value of the frailty component $\alpha = 0.25$ and censoring rate to be 50%.	48

2.7	Real data analysis for breast cancer data with focus on the overall survival time: parameter estimates based on the proposed frailty PMRL, conventional PMRL, frailty AFT and frailty Cox PH models.	54
3.1	Simulation results using penalized conditional quasi-likelihood method based on frailty and conventional PMRL models: lognormal frailty with variance component $\alpha=0.0, 0.25, 0.5$.	81
3.2	Simulation results using penalized conditional quasi-likelihood method based on frailty and conventional PMRL models: gamma frailty with variance component $\alpha=0.0, 0.25, 0.5$.	83
3.3	Application to distant metastasis-free survival: covariate information for total 2,668 patients in the valid dataset, the number (percentage) of observations for different covariates over 15 clusters.	91
3.4	Application to distant metastasis-free survival: parameter estimates based on the proposed frailty PMRL, the conventional PMRL and the frailty Cox PH models.	92
3.5	Application to distant metastasis-free survival: adding covariates stepwise.	93

Introduction

In this chapter, we present the background concepts, models and notation used for the analysis of survival data, which are relevant to the main topics of this thesis. The essential concepts and typical notation used in survival analysis are firstly introduced in Section 1.1. Section 1.2 introduces the best known nonparametric estimator, the Kaplan-Mier estimator, for the survival function. The most commonly used semiparametric survival models based on the hazard function and the mean residual life function are then described in Section 1.3. After a review of the frailty hazard-based models for the analysis of clustered data, Section 1.4 highlights the gaps and challenges of incorporating both fixed and random effects into mean residual life models, which motivate our work in Chapter 2. In Section 1.5, we review works about generalized linear mixed models with application of conditional inference, where the distributional assumption of the frailty can be further released and theoretical properties of the estimated fixed effects can be established. The outline of this thesis are summarized in the last section.

1.1 Basic Concepts in Survival Analysis

As an important branch of statistical analysis, survival analysis is widely applied in engineering, epidemiology, biomedical and sociology with the presence

of survival data. The survival time, also known as failure time, is defined as a time duration from a particular starting point to the occurrence of an event of interest such as death, the recurrence of a disease, remission after a treatment and so forth. Let T be the survival time of an individual from a homogenous population. Suppose that the nonnegative random variable T has a probability density function $f(t)$, which is defined to be the instantaneous occurrence rate of the event of interest at any time point t ($t \geq 0$).

The survival function $S(t)$ is primely employed to characterize the distribution of survival time T and defined as the probability that an individual survives to a time point t :

$$S(t) = P(T \geq t) = \int_t^{\infty} f(\omega) d\omega, \quad t \geq 0. \quad (1.1)$$

It is obvious that the survival function $S(t)$ is a monotonically nonincreasing function of time t with $S(0) = 1$ and $S(\infty) = 0$. The survival time T is assumed to be continuous in this thesis. Then, it can be easily obtained from (1.1) that $f(t) = -\frac{dS(t)}{dt}$.

In survival analysis, another fundamental distribution of the survival time T is the hazard function $\lambda(t)$, which specifies the instantaneous failure rate for an individual that has survived until time t and is defined as

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \quad t \geq 0. \quad (1.2)$$

Based on Equations (1.1) and (1.2), we have $S(t) = \exp\{-\int_0^t \lambda(\omega) d\omega\}$ and the cumulative hazard function is defined as $\Lambda(t) = \int_0^t \lambda(\omega) d\omega$.

The mean residual life (MRL) function refers to the expectation of remaining life given that the event of interest has not occurred at a specific time point t . It is usually denoted by $m(t)$ and defined as follows:

$$m(t) = E(T - t | T \geq t) = \frac{\int_t^{\infty} (\omega - t) f(\omega) d\omega}{S(t)}, \quad t \geq 0. \quad (1.3)$$

As one of the essential characteristics of the survival time distribution, the MRL function has gained quite a lot attention from researchers in recent years because

it is easy to be understood and interpreted. It plays as the main role of the work throughout the thesis. The hazard function can be expressed in terms of the MRL function based on Equations (1.2) and (1.3):

$$\lambda(t) = \frac{1}{m(t)} \left[\frac{dm(t)}{dt} + 1 \right]. \quad (1.4)$$

Similarly, according to (1.1) and (1.3), the survival function can be rewritten as follows:

$$S(t) = \frac{m(0)}{m(t)} \exp \left\{ - \int_0^t \frac{d\omega}{m(\omega)} \right\}. \quad (1.5)$$

A particular feature in the analysis of survival data is to handle censoring under which survival times of some individuals are not exactly observed in studies. Traditional statistical methods developed for complete data are not applicable. There are three major types of censoring: right censoring, left censoring and interval censoring. Among them, right censoring is most commonly encountered in biomedical studies. It occurs when an individual does not experience the event of interest or is lost to follow-up within the study duration. For example, due to the shortage of manpower and capital, the examiner has to suspend a breast cancer clinical research before all patients in the study die from recurrence of breast cancer. Or, some breast cancer patients may quit the study in the halfway for some personal reasons.

This thesis focuses on right-censored survival data. Let C be the potential censoring time for an individual. Subject to right censoring, the observation of survival time is given by $Y = \min(T, C)$ with a censoring indicator $\delta = I(T \leq C)$, which equals to 1 when the exact failure time T is observed and 0 otherwise.

1.2 Nonparametric Estimation

Suppose that there are N observations of survival times in the study and D distinct failure times $0 < t_{(1)} < t_{(2)} < \dots < t_{(D)}$ ($D \leq N$) among them. We define the number of events at time $t_{(k)}$ as $d_{(k)}$, that is, the number of individuals

who experience the event of interest at $t_{(k)}$, and the number of individuals who are at risk at time $t_{(k)}$ as $n_{(k)}$, which means there are $n_{(k)}$ individuals are alive and uncensored just ahead of the time point $t_{(k)}$.

Under the assumption that the event of interest and the censoring mechanism are independent, the Kaplan-Meier (KM) estimator (Kaplan and Meier, 1958), which is the best known nonparametric estimator of the survival function $S(t)$, is defined by:

$$\hat{S}(t) = \prod_{k:t_{(k)} \leq t} \left(1 - \frac{d_{(k)}}{n_{(k)}}\right), \quad (1.6)$$

The KM estimator is consistent for large samples under right censoring and thus has been widely employed to estimate the survival function.

Nonparametric estimators usually characterize main information about survival times of the studied population in survival analysis. For some other nonparametric estimations of the basic survival distributions, we refer interested readers to Aalen (1978), Klein (1991), Nelson (1972) and Yang et al. (1978). Although few assumptions, if not none, are required for survival data in nonparametric estimation procedures, nonparametric estimations are likely to cause biases in comparing distributions of survival times for different groups due to limited information.

1.3 Semiparametric Regression Models

In applications of survival analysis, researchers have more information about study individuals in hand that might be relevant to the survival times, including individuals' clinical factors (such as various physical and clinical measurements) and prognostic factors (such as the progress of treatment and the stages of disease) recorded in medical trials. In order to investigate the heterogeneity in a population, it is an important way to employ such factors as covariates in regression models. In survival analysis, it is often interesting to explore effects of these covariates on the survival distributions of an individual. In the past

decades, semiparametric models have been proposed in many ways and some special methodologies are adopted to accommodate the censoring structure of survival data.

Among commonly used semiparametric models in survival analysis, the Cox proportional hazards (PH) model proposed by Cox (1972) is well known and more appealing since it provides clear and direct explanation for relative risks. With covariates, the Cox PH model assumes the conditional hazard function in the form of

$$\lambda(t|\mathbf{X}) = \lambda_0(t) \exp(\mathbf{X}^T \boldsymbol{\psi}), \quad (1.7)$$

where the unspecified function $\lambda_0(t)$ is called the baseline hazard function, $\mathbf{X}$ is the vector of covariates and $\boldsymbol{\psi}$ is the vector of unknown regression parameters. The Cox PH model (1.7) specifies that the covariates $\mathbf{X}$ have multiplicative effects on the hazard function of the survival time.

When the proportional hazards assumption in the Cox PH model is violated, some alternative models have been proposed to capture the nonproportionality, such as the proportional odds model (Bennett, 1983; Murphy et al., 1997), the accelerated failure time model (Buckley and James, 1979; Lin and Chen, 2012) and the additive hazards model (Cox and Oakes, 1984; Lin et al., 1995). Though reference and studies on the hazard-based models are ubiquitous in traditional survival analysis field, the hazard function is conceptually difficult to be understood by nonstatisticians. For example, patients and doctors concern more about how long a patient could survive after commencing treatments rather than the relative risk of death or recurrence.

In recent years, more and more attentions have been paid to models based on the mean residual life function on account of its understandability and interpretability and the related studies have achieved great progress. By concentrating on the persistency of an individual, the mean residual life model provides a simple idea to interpret the relationship between the expectation of the remaining lifetime and prognostic factors. As an alternative to the Cox PH model, the

proportional mean residual life (PMRL) model is naturally developed from the perspective of residual lifetime instead of instantaneous failure rate. Oakes and Dasu (1990) firstly proposed a PMRL model based on the mean residual life for a two-sample case without considering any covariate. To associate the PMRL model with related covariates, Maguluri and Zhang (1994) extended Oakes and Dasu (1990)'s model to a regression model in the form of

$$m(t|\mathbf{X}) = m_0(t) \exp(\mathbf{X}^T \boldsymbol{\beta}), \quad (1.8)$$

where $m_0(t)$ is the baseline mean residual life function, $\mathbf{X}$ is the covariate vector and $\boldsymbol{\beta}$ is the vector of unknown regression coefficients.

To predict the expectancy life of an individual based on (1.8), the regression parameters $\boldsymbol{\beta}$ need to be estimated first. Abundant work has been done in the literature with focus on estimation procedures, either the conditional likelihood (Cox, 1972) or the partial likelihood (Cox, 1975), for the inference about parameters in the Cox PH model. However, these commonly used estimation procedures can not be applied to estimate the regression parameters in the PMRL model. The reason is when constructing the likelihood function, the derivative in (1.4) and the integral in (1.5) would make the computation complicated especially when the baseline mean residual life function $m_0(t)$ is unknown.

To avoid such a tricky problem in computation, estimation procedures directly based on estimating equations have been developed so that the construction of the likelihood can be skipped. Several authors have made significant contributions in the related area. For the analysis of the PMRL regression model, Chen and Cheng (2005) constructed generalized partial scores via counting processes to estimate the model parameters involving the regression coefficients and the unknown baseline MRL function, the consistency and asymptotic normality properties of the estimated fixed effects were further investigated. Chen et al. (2005) proposed weighted estimating equations to estimate regression coefficients in the PMRL model. They incorporated the inverse probability of censoring weighting (IPCW) into the difference between the observation and the corresponding expectation of

the MRL, the asymptotic property of the proposed estimator was studied based on martingale theory. Sun and Zhang (2009) developed a class of estimating equations for the analysis of a transformed MRL regression models by making use of the IPCW technique, and studied efficiency and asymptotic properties of the corresponding regression coefficient estimators. Please refer to Andersen and Gill (1982) and Fleming and Harrington (2011) for more information about the theory of counting processes, and Robins and Rotnitzky (1992) and Robins et al. (1994) for the IPCW technique.

1.4 Clustered Data and Frailty Models

All the semiparametric regression models introduced in Section 1.3 are generally applied to model homogenous survival data. However, data collection methods commonly used in biomedical studies, such as multi-center clinical trials, produce correlated survival data due to some common unobservable factors. An example of breast cancer data is studied by Haibe-Kains et al. (2012), where clinical prognostic factors for 5,715 patients from 36 breast cancer centers in different geographic locations are recorded. When we aim to explore the relationship between the clinical prognostic covariates and survival times in the above databases, it is naturally expected that there exist some potential correlations among patients from the same center because of the similar environment around and medical treatment conditions shared within centers. In survival analysis, we call such dependent survival data as clustered survival data, where the cluster refers to the center or institute. The conventional models, such as the Cox PH model and the PMRL model, are incapable of addressing the potential correlations caused by the clustered structure of survival data from the motivating example. The inference procedures based on these models would yield misleading results, especially when the correlation within clusters is sufficiently strong.

Extensive works have been done in last decades by extending conventional

regression models based on the relative risks to the frailty regression models, in which the frailty (or random effects) is employed to handle the potential dependence within clusters and describe how the cluster-specific association influences the hazard rate. Among these models, the frailty Cox PH model has been constructed as follows based on the classical Cox PH model (1.7) by imposing an unobservable random-effects variable u , which is used to characterize the clustered structure:

$$\lambda(t|\mathbf{X}) = \lambda_0(t) \exp(\mathbf{X}^T \boldsymbol{\psi})u. \quad (1.9)$$

From model (1.9), we can see that random effects act multiplicatively on the relative risks. By taking the frailty into consideration, the frailty Cox PH model is able to handle potential dependence among data caused by the clustered structure. This advantage makes it very popular in the analysis of clustered survival data.

To make inferences about both fixed and random effects in the frailty Cox PH model, estimation procedures based on the marginal likelihood have been frequently used when a gamma distribution is assumed to the frailty, see for example Klein (1992), Nielsen et al. (1992) and Parner et al. (1998). However, an intractable computation may be involved in the marginal likelihood function because the integrals over random effects would get even more complicated as the dimension of random effects goes higher. The hierarchical likelihood approach, which is firstly introduced by Lee and Nelder (1996) for the analysis of generalized linear mixed models, deals with the above mentioned issue by avoiding the computational complexity of integrals over random effects. The hierarchical likelihood method allows the frailty to take any distribution by loosening the distribution constraint on gamma or lognormal. In addition, it can be implemented via a simple algorithm such as the Newton-Raphson method. When censoring is presented in clustered data, a great deal of papers have been published in the past few years, where the hierarchical likelihood approach is applied to make inferences about frailty models.

Suppose that there are m clusters with q_i individuals in the i -th cluster for the clustered data subject to right censoring. The observation of the survival time for the j -th individual in the i -th cluster can be denoted by (Y_{ij}, δ_{ij}) , where $Y_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = I(T_{ij} \leq C_{ij})$, the indicator function δ_{ij} equals to 1 if the event of interest occurs and 0 otherwise. Combining the covariate information, we denote the observed data for all individuals as $\{(Y_{ij}, \delta_{ij}, \mathbf{X}_{ij}) : i = 1, \dots, m, j = 1, \dots, q_i\}$. Given the above observations, Ha et al. (2001) proposed the following hierarchical likelihood function based on the frailty Cox PH model (1.9):

$$h(\boldsymbol{\beta}, \Lambda_0, \alpha) = \sum_{ij} [\delta_{ij}(\log \lambda_0(Y_{ij}) + \eta_{ij}) - \Lambda_0(Y_{ij}) \exp(\eta_{ij})] + \sum_i l_{2i}(\alpha; v_i), \quad (1.10)$$

where $\eta_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\psi} + v_i$, $\Lambda_0(t) = \int_0^t \lambda_0(\omega) d\omega$ is the cumulative baseline hazard function, and $l_{2i}(\cdot)$ is the logarithm of the probability density function of the log-frailty $v_i = \log u_i$. They further estimated fixed effects $\boldsymbol{\psi}$ and random effects $\mathbf{v}$ based on the hierarchical likelihood function (1.10). The variance parameter of random effects was estimated by maximizing an adjusted profile hierarchical likelihood function, which was constructed by incorporating the first-order Laplace approximation in (1.10) and substituting $\boldsymbol{\psi}$, $\mathbf{v} = (v_1, \dots, v_m)^T$ and $\Lambda_0(\cdot)$ with the corresponding estimators. Ha and Lee (2003) showed that the frailty Cox PH model (1.9) can be fitted by using a Poisson generalized linear model. Ha et al. (2011) generalized the frailty Cox PH model from one component to two components for investigating the heterogeneity of survival data from multi-center clinical trials, and developed the hierarchical likelihood function for inferences about the frailty model. For an overview of the hierarchical likelihood method, we refer interested readers to the book by Ha et al. (2017) and the references therein.

In spite of the substantial literature about the frailty Cox PH models for the analysis of clustered survival data, to the best of our knowledge, few papers have considered a proportional mean residual life (PMRL) model with frailty to address the clustered structure. Therefore, a frailty model based on the PMRL is proposed

in Chapter 2 for the analysis of clustered survival data, which is able to evaluate impacts of the fixed covariates and the cluster factors on the remaining lifetimes of individuals from different clusters. By making use of the idea of hierarchical likelihood, a hierarchical quasi-likelihood function is proposed in Chapter 2. An inference procedure based on estimating equations is simultaneously developed for the estimation of both fixed and random effects.

1.5 Conditional Inference

In the analysis of clustered survival data, the hierarchical approach provides an appealing advantage by avoiding the computational complexity problem which arises when the dimension of random effects goes higher. Although the hierarchical likelihood method assumes the frailty distribution does not necessarily have to be gamma or lognormal, which is frequently assumed in the related literature, it still requires an explicit parametric distribution for the frailty term. Such a parametric distribution assumption is often hard to be verified in practice because little information about the frailty distribution is available in observations. The theoretical performance of the estimators from the hierarchical likelihood function is still open and very challenging to be justified rigorously even for clustered data without censoring. It is thus important to further investigate an alternative method to relax the parametric distribution assumption for the frailty term and meanwhile be able to produce estimators with well-established properties. To this end, conditional inference is consequently considered in this thesis.

Conditional inference is a useful method to handle clustered data, originally proposed by Verbeke et al. (2001) for linear mixed models, and Jiang (1999) for generalized linear mixed models (GLMMs). It basically treats random effects as nuisance parameters and performs an inference based on a conditional likelihood function. Let's consider a generalized linear mixed model for clustered data

y_i ($i = 1, \dots, N$). Suppose that $\alpha = (\alpha_1, \dots, \alpha_m)^T$ is the vector of unobservable random effects with mean 0, x_i and z_i are known vectors with p -dimension and m -dimension, respectively, and y_i ($i = 1, \dots, N$) are conditional independent responses given α . Let $b_i(\cdot)$ be a differentiable function. The conditional expectation of y_i is assumed to follow a GLMM in the form of

$$E(y_i|\alpha) = b'_i(\eta_i),$$

where $\eta_i = x_i^T \beta + z_i^T \alpha$ and β is the vector of unknown fixed effects. It is noticed that fixed effects β and random effects α can not to be uniquely determined by η_i since there may exist another parametric pair $(\tilde{\beta}, \tilde{\alpha})$ such that $\eta_i = x_i^T \tilde{\beta} + z_i^T \tilde{\alpha} = x_i^T \beta + z_i^T \alpha$. To remedy such identifiability problem, a class of linear constraints $P_A \alpha = 0$, which utilizes information provided by the zero-mean assumption of random effects α , is introduced in the inference. So that fixed effects β and random effects α can be identified in a restricted parameter space $\{(\beta, \alpha) : P_A \alpha = 0\}$. The projection matrix P_A is known with $A \in \mathfrak{B}(\mathcal{N}(P_{X^\perp} Z))$, where $P_{X^\perp} = I_N - X(X^T X)^{-1} X^T$, I_N is an $N \times N$ identity matrix, X and Z are matrices consisting of x_i and z_i , respectively. The construction of a projection matrix is briefly described as follows. Let M and V be any matrix and vector space, respectively. Define $\mathfrak{B}(V) = \{B : B \text{ is a matrix whose columns constitute a base for } V\}$ and $\mathcal{N}(M) = \{v : Mv = 0\}$ is the null space of M .

In constrained least squares, $\lambda |P_A \alpha|^2 / 2$ with a tuning parameter λ is usually chosen as the penalty term according to the Lagrangian multiplier method. By analogy with the weighted least squares method, Jiang (1999) constructed the following penalized conditional quasi-log-likelihood:

$$l_p(\beta, \alpha) = \sum_i \omega_i (y_i \eta_i - b_i(\eta_i)) - \frac{\lambda}{2} |P_A \alpha|^2, \quad (1.11)$$

where ω_i ($i = 1, \dots, N$) are weights. Then the penalized generalized weighted least square estimators $\hat{\beta}$ and $\hat{\alpha}$ are obtained by maximizing (1.11) with respect to β and α .

Jiang (1999) established the asymptotic property for $\hat{\beta}$ under some conditions and pointed out that the choice of λ involved in the penalty would not affect the consistency of the estimator. Wang et al. (2012) extended Jiang (1999)'s idea of conditional inference to the quadratic inference function (Qu et al., 2000) for longitudinal data and developed penalized quadratic inference functions (QIFs) to estimate fixed effects and random effects iteratively. They provided a closed-form asymptotic variance matrix of the estimated fixed effects when the predicted random effects were consistent. Based on the penalized QIFs approach, the assumption of an explicit distribution for random effects was completely relaxed and both the serial correlation and heterogeneous variation arose in longitudinal data were addressed. Other applications of conditional inference can be found in recent literature, including Cho et al. (2017) and Xing et al. (2018) for improving the efficiency of estimators, Yu et al. (2018) for establishing asymptotic properties of estimators under misspecified GLMMs and Liu and Xiang (2019) for the inference of GLMMs with the presence of missing covariates and without distributional assumption of random effects.

To the best of our knowledge, there is no application of conditional inference in the survival data setting. Based on the penalized conditional quasi-log-likelihood methods developed by Jiang (1999), Chapter 3 aims to propose a more flexible methodology for the analysis of clustered survival data. A penalty term associated with random effects is incorporated into a conditional quasi-likelihood, which is derived from the quasi score of fixed effects, to account for within-cluster association, while allowing distribution-free random effects in the model. Given some regularity conditions, the consistency and the asymptotic normality of the estimated fixed effects are further investigated with the advantage of good properties of counting processes and proven techniques used for GLMMs.

1.6 Outline

We end this chapter by highlighting main contents of this thesis.

Section 2.1 first introduces the frailty regression models based on the mean residual life function to analyze clustered survival data, and then constructs quasi scores based on a hierarchical quasi-likelihood function for inferences about both fixed and random effects in the proposed frailty model. To evaluate the proposed method's performance, extensive simulation studies have been carried out in Section 2.2. Section 2.3 applies the proposed frailty model and the hierarchical inference procedure on the breast cancer data from the motivating example. Details of used notation and derivations are provided in the Appendix.

In Section 3.1, we extend the frailty proportional mean residual life regression model to a more flexible model allowing the frailty free of distribution and develop a penalized conditional quasi-likelihood estimation procedure for the estimation of both fixed and random effects. Section 3.2 summarizes implementation steps of an iterative algorithm for the conditional inference method. With related notation and technical conditions given in Section 3.3, Section 3.4 establishes large sample properties for the proposed estimator of fixed effects and provides the detail theoretical justification. For illustration implementation, Section 3.5 conducts simulation studies under two generally used frailty distributions to investigate the performance of the proposed conditional inference approach. In order to exemplify the usefulness of the proposed model and the conditional inference procedure, we apply them to analyze the breast cancer data in Section 3.6.

The thesis ends with some conclusion remarks in the last chapter, including main contributions of Chapters 2 and 3 and some potential future topics based on mean residual life models.

Hierarchical Inference for Frailty Proportional Mean Residual Life Models[†]

As mentioned in Section 1.4, the hierarchical likelihood approach has a great practical use for inferences about generalized linear mixed models (GLMMs) with clustered non-normal data (Lee and Nelder, 1996; Lee et al., 2017) due to its capability of avoiding complex computations for numerical integrals caused by random effects, especially when the dimension of random effects is high, and making inferences about unobserved random effects. Its application in inferences about the frailty Cox proportional hazards (PH) model with survival data has been studied by Ha et al. (2001). More recent developments along this line can be found in the monograph by Ha et al. (2017). Note that the idea of hierarchical likelihood has not been applied on mean residual life models in the existing works, which focus on independent survival times incapable of applying to model clustered data as addressed in the motivation example.

[†]This chapter is based on the paper Huang et al. (2019): “Rui, Huang and Liming, Xiang and Il Do, Ha. Frailty proportional mean residual life regression for clustered survival data: a hierarchical quasi-likelihood method. *Statistics in Medicine*, 38:4854–4870, 2019.”

In this chapter, we are primarily interested in developing a frailty proportional mean residual life (PMRL) model by taking random effects into account to accommodate potential correlations within clusters and thus to predict the remaining survival time of an individual from a clustered cohort.

The frailty PMRL regression model presents unique challenge in the estimation procedure. This is because the marginal likelihood function is formed with integrals over random effects. Moreover, the relationship between the mean residual life function and the survival function would yield another layer of integral and the relationship between the mean residual life function and the hazard function would lead to a derivative in constructing the likelihood function, leading to intractable computation problems. The existing numerical methods for handling integrals caused by unobservable random effects in GLMMs or the frailty Cox PH models fail to overcome such computational challenge.

We propose a novel hierarchical quasi-likelihood method for the inference about the frailty PMRL model. A set of quasi-score estimating equations is in turn developed for parameters of both fixed and random effects in the proposed frailty model, with advantages that they can be solved iteratively using the standard Newton-Raphson algorithm and increase the model flexibility by relaxing the most commonly used restriction of gamma distribution on the frailty. We also construct an adjusted hierarchical quasi-likelihood function to estimate the variance component, which is regarded as a measure of the impact of random effects.

2.1 Methodology and Estimation Procedures

2.1.1 Model formulation

Let T be the study survival time of a random individual with finite expectation and $\mathbf{X}$ be the corresponding covariate vector. We are interested in the mean

residual life (MRL) function at a specific time point t , defined by

$$m(t) = E(T - t | T \geq t), \quad t \geq 0.$$

In practice, T is often censored by an independent random censoring time variable C with an unknown survival distribution $G(\cdot)$. Suppose that there are m clusters in the study. For the j -th individual ($j = 1, \dots, q_i$) in the i -th cluster ($i = 1, 2, \dots, m$), let T_{ij} and C_{ij} represent the corresponding survival time and potential censoring time, respectively, and $\mathbf{X}_{ij}$ be a p -dimensional vector of covariates. Defined $Y_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = I(T_{ij} \leq C_{ij})$, where $I(\cdot)$ is indicator function. Thus, observed data for all individuals are given by $\{(Y_{ij}, \delta_{ij}, \mathbf{X}_{ij}) : i = 1, \dots, m, j = 1, \dots, q_i\}$, thus we have $n = \sum_{i=1}^m q_i$ samples in total.

Let U_i be the unobservable frailty random variable for the i -th cluster. Conditional on the covariates $\mathbf{X}_{ij}$ and frailty $U_i = u_i$, we assume that the MRL function of T_{ij} can be modelled by a proportional MRL (PMRL) regression model, that is, the conditional PMRL function of T_{ij} has the following form:

$$m(t | \mathbf{X}_{ij}, u_i) = m_0(t) \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}) u_i, \quad (2.1)$$

where $m_0(t)$ is an unknown baseline MRL function and $\boldsymbol{\beta}$ is a p -dimensional vector of regression parameters corresponding to fixed effects. We further denote $v_i = \log u_i$ as log-frailty (random effects) associated with cluster i . The expression in (2.1) can be rewritten as a frailty PMRL regression model

$$m(t | \mathbf{X}_{ij}, \mathbf{v}) = m_0(t) \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_{ij}^T \mathbf{v}), \quad (2.2)$$

where $\mathbf{Z}_{ij}$ is an m -dimensional vector of group indicator covariates and $\mathbf{v} = (v_1, \dots, v_m)^T$ is a vector of random effects.

Furthermore, let $\eta_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_{ij}^T \mathbf{v}$. Going through the relationship between the MRL function and the hazard function studied by Oakes and Dasu (1990), from model (2.2) the conditional hazard function $\lambda(t | \mathbf{X}_{ij}, v_i)$ and the conditional

survival function $S(t|\mathbf{X}_{ij}, v_i)$ of T_{ij} can be given as

$$\begin{aligned}\lambda(t|\mathbf{X}_{ij}, v_i) &= \frac{1}{m_0(t)} \left[\frac{dm_0(t)}{dt} + \exp(-\eta_{ij}) \right], \\ S(t|\mathbf{X}_{ij}, v_i) &= \frac{m_0(0)}{m_0(t)} \exp \left\{ - \int_0^t \frac{\exp(-\eta_{ij}) d\omega}{m_0(\omega)} \right\}.\end{aligned}$$

We assume that random effects $V_i = \log U_i$ ($i = 1, 2, \dots, m$) are independent and identically distributed random variables with a density function $f_v(\cdot; \alpha)$, where α is an unknown frailty parameter. The log-gamma and normal distributions (Ripatti and Palmgren, 2000) are two common distributions for V_i though other distributions (e.g. mixture normal) are possible, where $E(U_i) = 1$ and $Var(U_i) = \alpha$ for the gamma frailty, and $V_i = \log U_i \sim N(0, \alpha)$ for the lognormal frailty. Based on the observed data $\{(Y_{ij}, \delta_{ij}, \mathbf{X}_{ij}) : i = 1, \dots, m, j = 1, \dots, q_i\}$, we obtain the marginal likelihood function of $\boldsymbol{\beta}$ and α by integrating out unobserved random effects

$$\begin{aligned}L(\boldsymbol{\beta}, \alpha) &= \prod_i \int \prod_j \lambda^{\delta_{ij}}(Y_{ij}|\mathbf{X}_{ij}, v_i) S(Y_{ij}|\mathbf{X}_{ij}, v_i) f_v(v_i; \alpha) dv_i \\ &= \prod_i \int \prod_j \frac{m_0(0)}{m_0(Y_{ij})^{1+\delta_{ij}}} \left(\frac{dm_0(t)}{dt} \Big|_{t=Y_{ij}} + \exp(-\eta_{ij}) \right)^{\delta_{ij}} \\ &\quad \cdot \exp \left\{ - \int_0^{Y_{ij}} \frac{\exp(-\eta_{ij}) d\omega}{m_0(\omega)} \right\} f_v(v_i; \alpha) dv_i.\end{aligned}$$

In general, it is very hard to find estimators of parameters $\boldsymbol{\beta}$ and α by directly maximizing the likelihood above due to the complicated two-layer integral involved.

Unlike the classical likelihood for fixed parameters $\boldsymbol{\beta}$ only, the hierarchical likelihood function is constructed for both $\boldsymbol{\beta}$ and unobserved random effects $\mathbf{v}$ at the same time and provides a statistically efficient procedure for various random-effects models (Lee and Nelder, 1996). For usual PH-based frailty models, the hierarchical likelihood function is defined by the logarithm of the joint density of $(Y, \delta, \mathbf{X})$ and $\mathbf{v}$ (Ha et al., 2001), and thus the hierarchical likelihood inference naturally avoids the integral caused by random effects. It yields an approximate

maximum likelihood (ML) estimator for β , an approximate restricted ML estimator for α and asymptotic normality for the estimators of realized random effects as the cluster size q_i grows large (Ha et al., 2017; Lee et al., 2017). Along the same lines of the hierarchical likelihood approach, we can define a hierarchical likelihood function of β , $\mathbf{v}$ and α based on model (2.2) in the form of

$$h(\beta, \mathbf{v}, \alpha) = h_1(\beta|\mathbf{v}) + h_2(\alpha; \mathbf{v}), \quad (2.3)$$

where $h_1(\beta|\mathbf{v})$ is the conditional log-likelihood function of β given random effects $\mathbf{v}$, that is

$$h_1(\beta|\mathbf{v}) = \sum_{ij} \left[\delta_{ij} \log \left(\frac{dm_0(t)}{dt} \Big|_{t=Y_{ij}} + \exp(-\eta_{ij}) \right) - \exp(-\eta_{ij}) \int_0^{Y_{ij}} \frac{d\omega}{m_0(\omega)} \right], \quad (2.4)$$

and $h_2(\alpha; \mathbf{v}) = \sum_i \log f_v(v_i; \alpha)$ is the log-likelihood function of α based on the distribution of random effects $\mathbf{v}$.

To some extent, $h(\beta, \mathbf{v}, \alpha)$ relaxes the computational complexity by avoiding the integral over random effect v_i in the likelihood. When the baseline MRL function $m_0(t)$ is known, the estimates of parameters β , α and random effects $\mathbf{v}$ can be obtained through maximizing (2.3). However, in most situations $m_0(t)$ is unknown. Like the usual PMRL regression models, it is still intractable to maximize $h_1(\beta|\mathbf{v})$ in (2.4). The development of an alternative estimation method, which allows us to achieve easy-to-compute estimators of β and α , is therefore essential. To this end, we propose a novel method of estimating equations based on hierarchical quasi-likelihood in the following section.

2.1.2 Maximum hierarchical quasi-likelihood estimation

To overcome the intractable computation problem, we consider the estimating equation method relying on stochastic processes first. Recall that $G(\cdot)$ is the survival function of the censoring time C and assume that C is independent of T , V and $\mathbf{X}$. Along the similar line of the method in Sun and Zhang (2009), by

model (2.2) we define a class of zero-mean stochastic processes using the inverse probability of censoring weighting (IPCW) technique

$$M_{ij}(t) = \frac{\delta_{ij}I(Y_{ij} > t)}{G(Y_{ij})} [(Y_{ij} - t) - m_0(t) \exp(\eta_{ij})] \quad (2.5)$$

for $i=1, \dots, m$, $j=1, \dots, q_i$. Suppose that there are D distinct ordered censoring times: $0 < C_{(1)} < C_{(2)} < \dots < C_{(D)}$. Based on observations $\{(Y_{ij}, 1 - \delta_{ij}) : i=1, \dots, m, j=1, \dots, q_i\}$, the Kaplan-Meier estimator of the censoring survival function $G(t)$ can be obtained by $\hat{G}(t) = \prod_{k: C_{(k)} \leq t} (1 - \frac{d_k^c}{n_k^c})$, where $d_k^c = \sum_{ij} (1 - \delta_{ij})I(Y_{ij} = C_{(k)})$, $n_k^c = \sum_{ij} I(Y_{ij} \geq C_{(k)})$. With $G(Y_{ij})$ replaced by $\hat{G}(Y_{ij})$ in (2.5), we obtain an estimating equation

$$\sum_{ij} \frac{\delta_{ij}I(Y_{ij} > t)}{\hat{G}(Y_{ij})} [(Y_{ij} - t) - m_0(t) \exp(\eta_{ij})] = 0 \quad (2.6)$$

for $0 \leq t \leq \tau$, where $0 \leq \tau = \inf \{t : P(T \geq t) = 0\} < \infty$. Given values of β and $\mathbf{v}$, (2.6) becomes an equation in terms of $m_0(t)$. One can show that this equation has a closed form solution

$$\hat{m}_0(t; \beta, \mathbf{v}) = \frac{\sum_{ij} \frac{\delta_{ij}I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t)}{\sum_{ij} \frac{\delta_{ij}I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \exp(\eta_{ij})}, \quad 0 \leq t \leq \tau. \quad (2.7)$$

Plugging estimates $\hat{m}_0(t; \beta, \mathbf{v})$ and $\hat{G}(Y_{ij})$ into the stochastic processes in (2.5) and following the semiparametric inference in Sun and Zhang (2009), we can construct a quasi score for β given $\mathbf{v}$:

$$\mathbf{D}(\beta|\mathbf{v}) = \sum_{ij} \int_0^\tau \frac{\delta_{ij}I(Y_{ij} > t) \mathbf{X}_{ij}}{\hat{G}(Y_{ij})} [(Y_{ij} - t) - \hat{m}_0(t; \beta, \mathbf{v}) \exp(\eta_{ij})] d\mathcal{H}(t), \quad (2.8)$$

where $\mathcal{H}(t)$ is an increasing weight function and can be specified in different forms. We set $\mathcal{H}(t)$ to be the observed counting process $N(t) = \sum_{ij} I(Y_{ij} \leq t) \delta_{ij}$ due to its good performance as shown in the simulation study by Sun and Zhang (2009). In the following context, we thus substitute $\mathcal{H}(t)$ with $N(t)$ directly. Note that $E(\mathbf{D}(\beta|\mathbf{v})) = \mathbf{0}$. As a result, we propose to estimate β by solving estimating

equations $\mathbf{D}(\boldsymbol{\beta}|\mathbf{v}) = \mathbf{0}$ for given $\mathbf{v}$, similar to the generalized estimating equation (GEE) method defined by Liang and Zeger (1986).

To complete the estimation procedure, we need to estimate random effects $\mathbf{v}$ and parameter α . We start by acting as if there exists a quasi-likelihood function of $\boldsymbol{\beta}$ conditional on $\mathbf{v}$ and associated with the quasi score $\mathbf{D}(\boldsymbol{\beta}|\mathbf{v})$. Inspired by the idea of the hierarchical likelihood in (2.3), we propose a hierarchical quasi-likelihood (HQL) function for $\boldsymbol{\beta}$, $\mathbf{v}$ and α , that is

$$h_q(\boldsymbol{\beta}, \mathbf{v}, \alpha) = h_{q1}(\boldsymbol{\beta}|\mathbf{v}) + h_2(\alpha; \mathbf{v}), \quad (2.9)$$

where

$$h_{q1}(\boldsymbol{\beta}|\mathbf{v}) = \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t) \left[\eta_{ij} - \log \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t) \exp(\eta_{ij})}{\hat{G}(Y_{ij})} \right] dN(t)$$

is the conditional quasi-likelihood function for $\boldsymbol{\beta}$ given $\mathbf{v}$ corresponding to $\mathbf{D}(\boldsymbol{\beta}|\mathbf{v})$, and $h_2(\alpha; \mathbf{v})$ is the log-likelihood function of α based on $\mathbf{v}$. For example,

$$h_2(\alpha; \mathbf{v}) = -\frac{m \log \alpha}{2} - \frac{\sum_i v_i^2}{2\alpha}$$

for lognormal frailty and

$$h_2(\alpha; \mathbf{v}) = \frac{\sum_i (v_i - u_i)}{\alpha} - m \log \Gamma\left(\frac{1}{\alpha}\right) - \frac{m \log \alpha}{\alpha}$$

for gamma frailty. A noteworthy feature of the proposed HQL in (2.9) is that it does not depend on unknown $m_0(t)$, leading to an efficient hierarchical quasi-likelihood inference approach.

Given values of $\boldsymbol{\beta}$ and α , taking derivative of (2.9) with respect to $\mathbf{v}$ yields the quasi score of $\mathbf{v}$

$$\begin{aligned} \mathbf{D}(\mathbf{v}|\boldsymbol{\beta}, \alpha) &= \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{Z}_{ij}}{\hat{G}(Y_{ij})} [(Y_{ij} - t) - \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \exp(\eta_{ij})] dN(t) \\ &\quad + \frac{\partial h_2(\alpha; \mathbf{v})}{\partial \mathbf{v}}. \end{aligned} \quad (2.10)$$

Derivation details for (2.10) can be found in the Appendix. Then, we can obtain the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{v}}$ by alternately solving estimating equations $\mathbf{D}(\boldsymbol{\beta}|\mathbf{v}) = \mathbf{0}$

and $\mathbf{D}(\mathbf{v}|\boldsymbol{\beta}, \alpha) = \mathbf{0}$. This estimation procedure is easy to implement using a standard Newton-Raphson algorithm.

For the estimate of the frailty parameter α , Ha et al. (2001) addressed the bias problem of the maximum hierarchical likelihood estimator in the frailty models. They suggested a correction method by replacing the profile hierarchical likelihood with its adjusted version. In a similar manner, given estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{v}}$, we define an adjusted hierarchical quasi-likelihood (AHQL) function for α with the following form:

$$\begin{aligned} h_q^*(\alpha) &= h_q^*(\alpha; \hat{\boldsymbol{\beta}}, \hat{\mathbf{v}}) = [h_q(\boldsymbol{\beta}, \mathbf{v}, \alpha) + \frac{1}{2} \log\{\det(2\pi \mathbf{H}^{-1})\}]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{v}=\hat{\mathbf{v}}} \\ &= \frac{p+m}{2} \log(2\pi) + [h_q(\boldsymbol{\beta}, \mathbf{v}, \alpha) - \frac{1}{2} \log\{\det(\mathbf{H})\}]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{v}=\hat{\mathbf{v}}}, \end{aligned} \quad (2.11)$$

where $\mathbf{H} = \begin{bmatrix} \mathbf{H}_{bb} & \mathbf{H}_{bv} \\ \mathbf{H}_{vb} & \mathbf{H}_{vv} \end{bmatrix} = \begin{bmatrix} -\frac{\partial \mathbf{D}(\boldsymbol{\beta}|\mathbf{v})}{\partial \boldsymbol{\beta}^T} & -\frac{\partial \mathbf{D}(\boldsymbol{\beta}|\mathbf{v})}{\partial \mathbf{v}^T} \\ -\frac{\partial \mathbf{D}(\mathbf{v}|\boldsymbol{\beta}, \alpha)}{\partial \boldsymbol{\beta}^T} & -\frac{\partial \mathbf{D}(\mathbf{v}|\boldsymbol{\beta}, \alpha)}{\partial \mathbf{v}^T} \end{bmatrix}$ is a $(p+m) \times (p+m)$ matrix, and

$$\det(\mathbf{H}) = \det(\mathbf{H}_{vv}) \det(\mathbf{H}_{bb} - \mathbf{H}_{bv} \mathbf{H}_{vv}^{-1} \mathbf{H}_{vb}).$$

In fact, $\mathbf{H}$ is the negative Hessian matrix of the hierarchical quasi-likelihood function h_q in (2.9). The definition of $\mathbf{H}$ guarantees its existence and positive definiteness so that $\mathbf{H}^{-1}$ exists. The entries of $\mathbf{H}$ and the detailed justification of its positive definiteness are given in the Appendix of this chapter.

The estimator $\hat{\alpha}$ is then obtained by maximizing the AHQL function $h_q^*(\alpha)$ with respect to α . The resulting estimate of $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{v}^T)^T$, denoted by $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\mathbf{v}}^T)^T$, is named as the maximum hierarchical quasi-likelihood estimator, while $\hat{\alpha}$ is called the maximum adjusted hierarchical quasi-likelihood estimator.

Define $\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{H}_{bb}^{inv} & \mathbf{H}_{bv}^{inv} \\ \mathbf{H}_{vb}^{inv} & \mathbf{H}_{vv}^{inv} \end{bmatrix}$, where

$$\mathbf{H}_{bb}^{inv} = (\mathbf{H}_{bb} - \mathbf{H}_{bv} \mathbf{H}_{vv}^{-1} \mathbf{H}_{vb})^{-1},$$

$$\mathbf{H}_{bv}^{inv} = -\mathbf{H}_{bb}^{inv} \mathbf{H}_{bv} \mathbf{H}_{vv}^{-1},$$

$$\mathbf{H}_{vb}^{inv} = (\mathbf{H}_{bv}^{inv})^T,$$

$$\mathbf{H}_{vv}^{inv} = \mathbf{H}_{vv}^{-1} + \mathbf{H}_{vv}^{-1} \mathbf{H}_{vb} \mathbf{H}_{bb}^{inv} \mathbf{H}_{bv} \mathbf{H}_{vv}^{-1}.$$

The covariance matrix of $\hat{\boldsymbol{\theta}}$ can be estimated following the sandwich method of Sun and Zhang (2009)

$$\begin{aligned} \widehat{Cov}(\hat{\boldsymbol{\theta}}) &= \mathbf{H}^{-1} \boldsymbol{\Sigma} \mathbf{H}^{-1} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \alpha=\hat{\alpha}}, \\ &= \begin{pmatrix} \mathbf{H}_{bb}^{inv} & \mathbf{H}_{bv}^{inv} \\ \mathbf{H}_{vb}^{inv} & \mathbf{H}_{vv}^{inv} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{zx} & \boldsymbol{\Sigma}_{zz} \end{pmatrix} \begin{pmatrix} \mathbf{H}_{bb}^{inv} & \mathbf{H}_{bv}^{inv} \\ \mathbf{H}_{vb}^{inv} & \mathbf{H}_{vv}^{inv} \end{pmatrix} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \alpha=\hat{\alpha}}. \end{aligned}$$

As a result, the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$ has the following form:

$$\begin{aligned} \widehat{Cov}(\hat{\boldsymbol{\beta}}) &= (\mathbf{H}_{bb}^{inv} \boldsymbol{\Sigma}_{xx} \mathbf{H}_{bb}^{inv} + \mathbf{H}_{bv}^{inv} \boldsymbol{\Sigma}_{zx} \mathbf{H}_{bb}^{inv} + \mathbf{H}_{bb}^{inv} \boldsymbol{\Sigma}_{xz} \mathbf{H}_{vb}^{inv} \\ &\quad + \mathbf{H}_{bv}^{inv} \boldsymbol{\Sigma}_{zz} \mathbf{H}_{vb}^{inv}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \alpha=\hat{\alpha}} \end{aligned} \quad (2.12)$$

where $\boldsymbol{\Sigma}_{xx} = \sum_{ij} \hat{\boldsymbol{\xi}}_{ij}^{x \otimes 2}$, $\boldsymbol{\Sigma}_{zz} = \sum_{ij} \hat{\boldsymbol{\xi}}_{ij}^{z \otimes 2}$, $\boldsymbol{\Sigma}_{zx} = \sum_{ij} \hat{\boldsymbol{\xi}}_{ij}^z (\hat{\boldsymbol{\xi}}_{ij}^x)^T$, $\boldsymbol{\Sigma}_{xz} = (\boldsymbol{\Sigma}_{zx})^T$, the vectors $\hat{\boldsymbol{\xi}}_{ij}^x$ and $\hat{\boldsymbol{\xi}}_{ij}^z$ ($i=1, \dots, m$, $j=1, \dots, q_i$) are defined in the beginning of the Appendix and $\boldsymbol{\xi}^{\otimes 2} = \boldsymbol{\xi} \boldsymbol{\xi}^T$ stands for an outer product of a vector $\boldsymbol{\xi}$.

Note that estimation of $Cov(\hat{\boldsymbol{\beta}})$ given in Equation (2.12) actually allows covariates to be dependent on random effects when $\mathbf{H}_{vb}^{inv} \neq \mathbf{0}$. This situation could occur in practice when cluster level characteristics such as different environments and life styles over centers from different countries are associated with both cluster level MRL and the disease progression (a possible covariate). Ignoring such association could yield biased estimators of random effects in mixed effects models, as recognized by Neuhaus and McCulloch (2011). On the other hand, when covariate effects are uncorrelated with random effects, the last three terms in the right hand side of Equation (2.12) contribute little, and thus can be ignored in the estimation of $Cov(\hat{\boldsymbol{\beta}})$. In this case, a sandwich form of the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$ can be obtained by $\widehat{Cov}(\hat{\boldsymbol{\beta}}) = \mathbf{H}_{bb}^{inv} \boldsymbol{\Sigma}_{xx} \mathbf{H}_{bb}^{inv} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \alpha=\hat{\alpha}}$.

2.1.3 Fitting algorithm

We employ the Newton-Raphson algorithm to solve the quasi-score estimating equations proposed in Section 2.1.2. The implementation steps are summarized

as follows:

Step 0: Set initial values $\hat{\beta}^{(0)}$, $\hat{\mathbf{v}}^{(0)}$, $\hat{\alpha}^{(0)}$.

Step 1: Denote the k -th step estimates as $\hat{\beta}^{(k)}$, $\hat{\mathbf{v}}^{(k)}$ and $\hat{\alpha}^{(k)}$. We estimate the baseline MRL function at distinct failure times using $\hat{m}_0(\cdot; \hat{\beta}^{(k)}, \hat{\mathbf{v}}^{(k)})$ given by (2.7), and update quasi scores as $\mathbf{D}(\hat{\beta}^{(k)}|\hat{\mathbf{v}}^{(k)})$ and $\mathbf{D}(\hat{\mathbf{v}}^{(k)}|\hat{\beta}^{(k)}, \hat{\alpha}^{(k)})$.

Step 2: Update $\hat{\beta}^{(k+1)}$ and $\hat{\mathbf{v}}^{(k+1)}$ by

$$\begin{bmatrix} \hat{\beta}^{(k+1)} \\ \hat{\mathbf{v}}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \hat{\beta}^{(k)} \\ \hat{\mathbf{v}}^{(k)} \end{bmatrix} + \mathbf{H}^{-1} \begin{bmatrix} \mathbf{D}(\beta|\mathbf{v}) \\ \mathbf{D}(\mathbf{v}|\beta, \alpha) \end{bmatrix}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(k)}, \alpha=\hat{\alpha}^{(k)}},$$

equivalently

$$\begin{cases} \hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + [\mathbf{H}_{bb}^{inv} \mathbf{D}(\beta|\mathbf{v}) + \mathbf{H}_{bv}^{inv} \mathbf{D}(\mathbf{v}|\beta, \alpha)]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(k)}, \alpha=\hat{\alpha}^{(k)}}, \\ \hat{\mathbf{v}}^{(k+1)} = \hat{\mathbf{v}}^{(k)} + [\mathbf{H}_{vb}^{inv} \mathbf{D}(\beta|\mathbf{v}) + \mathbf{H}_{vv}^{inv} \mathbf{D}(\mathbf{v}|\beta, \alpha)]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(k)}, \alpha=\hat{\alpha}^{(k)}}. \end{cases}$$

Step 3: With $\hat{\beta}^{(k+1)}$ and $\hat{\mathbf{v}}^{(k+1)}$, update $\hat{\alpha}^{(k+1)}$ by maximizing the AHQL function $h_q^*(\alpha; \hat{\beta}^{(k+1)}, \hat{\mathbf{v}}^{(k+1)})$ in (2.11).

Step 4: Repeat Steps 1-3 until the convergence is achieved.

We take $\hat{\beta}^{(0)} = \mathbf{0}$, $\hat{\mathbf{v}}^{(0)} = \mathbf{0}$ and $\hat{\alpha}^{(0)} = 0.1$ as initial values in the numerical studies. Alternatively, the estimate of β obtained from the conventional PMRL model is a better choice for $\hat{\beta}^{(0)}$, which could lead to a faster convergence of the algorithm. Based on our experience in simulations, several choices of the initial values had been tried and all led to the same estimation result although the algorithm would require less iteration steps when the initial values were chosen closer to the true values of parameters. The convergence can be monitored by adopting a stopping rule that $\max \{|\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)}|, |\hat{\alpha}^{(k+1)} - \hat{\alpha}^{(k)}|\} < \epsilon$, where ϵ is a threshold value and is taken as 10^{-4} in the simulation studies.

2.2 Simulation Studies

In this section, we conduct simulation studies to assess the performance of the proposed method and compare with results from the conventional PMRL model, which is incapable of accommodating dependence within clusters. Two common frailty distributions, the lognormal and gamma distributions, are considered. In addition, we also consider a mixture normal distribution for log-frailty (random effects). The purpose of taking such a nonstandard distribution is to show that the proposed method can handle any distributions for the frailty or random effects. Misspecification of the frailty distribution is also taken into consideration in the simulation to evaluate the robustness of the estimated fixed effects.

First, we generate data of failure time T from the working PMRL model with lognormal frailty

$$m(t|X, v) = m_0(t) \exp(X\beta + v), \quad (2.13)$$

where the true value of the regression coefficient is taken to be $\beta = 2.0$, X is generated from a Bernoulli distribution with probability of success = 0.5 and v is generated from a normal distribution with mean 0 and variance α . We consider different levels for the variance component, that is, $\alpha = 0.0, 0.25, 0.5$, representing a none, medium or strong association within clusters, respectively. The baseline MRL function is chosen from the Hall-Wellner family as is in the work of Sun and Zhang (2009) in the form of $m_0(t) = a_1 t + a_2$ with $a_1 = -0.05$ and $a_2 = 0.1$, which yields a rescaled beta distribution $B(1, 19)$ with support $[0, 2]$. The censoring time C is generated from an exponential distribution with mean λ_c , which is determined to yield the desired censoring rates of 10%, 20% and 30% approximately. The observed time Y is obtained by taking the minimum of T and C . We set the number of clusters to be $m = 5, 10, 20$ with each cluster including the same number of individuals $q = 50$.

We simulate data with 1,000 replications for each combination of m and the censoring rate, and analyze each simulated data set using the proposed frailty

PMRL model and the conventional PMRL model (ignoring clustered data structure). Estimation results are summarized in terms of the mean of bias in percentage (%Bias), mean of relative bias in percentage (%Rbias), sample standard deviation (SD), mean of estimated standard error (SE), root of mean square error (RMSE) and empirical coverage probability of a nominal 95% confidence interval in percentage (CP%), where the standard error of an estimated regression coefficient is computed by the square root of the sandwich variance $\mathbf{H}_{bb}^{inv} \boldsymbol{\Sigma}_{xx} \mathbf{H}_{bb}^{inv}$. Results obtained with three different levels of frailty ($\alpha = 0.0, 0.25, 0.5$) are reported in Table 2.1.

When there is no cluster-specific association among data, that is, $\alpha = 0$, it can be seen that the proposed estimator of β , the parameter of interest, performs competitively compared with its counterpart in the conventional PMRL model in terms of RMSE. Although the bias of $\hat{\beta}$ is sometimes doubled that of the corresponding estimator in the conventional model when $m = 5$, it is still of very small magnitude. These findings illustrate the validity of our method in the case with zero variance component. We observe that the estimator of α may be slightly biased in this case, but the RMSE of $\hat{\alpha}$ decreases as the number of clusters m increases. When $\alpha = 0.25$, data have relatively low dependence within clusters, simulation results indicate that the proposed model outperforms the conventional model in estimating β with considerably less biases, lower SEs (or SDs) and thus lower RMSEs in all simulation cases. When the number of clusters is fixed, the proposed estimators for α and β tend to perform better as the censoring rate decreases. For a given censoring rate, as the number of clusters increases, relative biases, SDs and RMSEs of all parameter estimates decrease in the proposed model, while estimates from the conventional model do not exhibit such a trend due to ignoring within-cluster association. When $\alpha = 0.5$, that is, the impact of random effects increases, the relative biases and SDs of estimated α and β increase as expected. With either increasing sample size or decreasing censoring rate, the %Rbias, SD and RMSE tend to be reduced. Although the

bias and CP show some discrepancy for data with higher dependence and the censoring rate up to 30%, the RMSE always clearly shows a decreasing trend as m increases regardless the values of α and censoring rate.

It is also found from Table 2.1 that the SDs and SEs of estimated β are close to each other and their difference becomes smaller as the sample size gets larger. This finding indicates that the proposed SE works well. The CPs based on the proposed method are almost all close to the nominal level 95%. However, the conventional method produces biased estimates with greater SDs, leading to poor CPs, especially when frailty parameter becomes greater, say $\alpha=0.5$.

Furthermore, we consider another scenario with gamma frailty. The frailty term $\exp(v)$ is generated from the $\text{gamma}(\frac{1}{\alpha}, \alpha)$ distribution with mean 1 and variance α . We present the simulation results obtained from the PMRL model with gamma frailty and the conventional PMRL model in Table 2.2 for three different choices of α . We see that the same trends and performances in terms of Bias (Rbias), SD and CP hold for the case with gamma frailty, as reflected in Table 2.1.

In addition, we examine the performance of the proposed method when the frailty distribution is nonstandard. In particular, we assume a mixture of two Gaussian distributions $\frac{2}{3}N(0, \frac{1}{2}\alpha) + \frac{1}{3}N(0, 2\alpha)$ for the log-frailty, where the variance of the log-frailty still equals α , and its true value of α is set to be 0.5. We generate failure times from the working model (2.13) again, using the same parameter design as before except that log-frailty v is generated from a mixture of $N(0, 0.25)$ and $N(0, 1)$ distributions with mixing probabilities $2/3$ and $1/3$, respectively. Both the frailty and conventional PMRL models are applied to each simulated sample. Results summarized in Table 2.3 indicate that the proposed method can still produce efficient estimates of β in this nonstandard frailty case though relative biases of estimates of α sometimes become quite big, especially for small samples. It is worth noting that the conventional method performs poorly in this case, by providing apparently biased estimates and the resultant 95% confidence

intervals with serious under-coverage of the true value of β regardless of the sample size and censoring rate.

At last, we examine the performance of the proposed estimators under misspecified distributions of random effects. Table 2.4 reports estimation results of the gamma frailty PMRL model when the true distribution of log-frailty is a normal distribution $N(0, 0.25)$ or a mixture of $N(0, 0.25)$ and $N(0, 1)$ with mixing probabilities $2/3$ and $1/3$. In both two misspecification cases, although the estimator of α performs poorly as expected, the proposed estimator of β works surprisingly well with small biases and SEs similar to those in Tables 2.1-2.3. We further provide results in Table 2.5 obtained by a misspecified lognormal frailty PMRL model when the true distribution for random effects is a mixture of normal distributions. Similar to those shown in Table 2.4, a robust estimation trend is also found for the estimates of β .

For the illustration purpose, we conduct simulations, where the lognormal frailty with variance $\alpha = 0.25$ and the censoring rate of 50% are considered, to assess the performance of the proposed method. Simulation results based on 500 replications are reported in Table 2.6. Compared with the corresponding results in Table 2.1, the scenario of $q = 50$ in Table 2.6 shows that when the censoring rate increases to 50%, the Biases (or Rbiases) and SDs of the estimated α and β increase as expected. When the cluster size is increased to be $q = 100$, it can be seen that estimates of both α and β in this case have less Biases (or Rbiases) and lower SDs than those in the scenario of $q = 50$. In addition, we also find that the Rbiases and SDs of our primary interest parameter β tend to be reduced with increasing clusters. These findings indicate that a much larger sample size may be required to ensure the proposed estimator of our primary interest parameter β getting close to its true value as the censoring rate increases.

Table 2.1. Simulation results using hierarchical quasi-likelihood method based on frailty and conventional PMRL models: lognormal distributed frailty with true variance parameter $\alpha=0.0, 0.25, 0.5$.

Frailty PMRL														Conventional PMRL					
m	C%	α			β					β									
		%Bias	RMSE	SD	%Rbias	RMSE	SD	SE	CP%	%Rbias	RMSE	SD	SE	CP%					
		True value of $\alpha=0.0$																	
5	10	2.46	0.0310	0.0188	-0.61	0.1270	0.1264	0.1200	94.8	-0.36	0.1248	0.1246	0.1195	95.4					
5	20	4.00	0.0505	0.0309	-0.72	0.1413	0.1406	0.1281	95.4	-0.03	0.1373	0.1373	0.1278	95.4					
5	30	6.67	0.0848	0.0523	-0.97	0.1565	0.1553	0.1397	94.2	0.47	0.1506	0.1503	0.1401	94.5					
10	10	2.45	0.0275	0.0125	0.05	0.0877	0.0877	0.0855	95.3	0.25	0.0867	0.0866	0.0852	94.4					
10	20	3.83	0.0429	0.0193	-0.21	0.0934	0.0933	0.0910	94.5	0.39	0.0911	0.0908	0.0906	94.7					
10	30	6.35	0.0708	0.0314	-0.83	0.1096	0.1083	0.0993	95.1	0.72	0.1058	0.1048	0.0992	94.0					
20	10	2.39	0.0253	0.0082	-0.07	0.0629	0.0629	0.0608	95.1	0.12	0.0620	0.0620	0.0606	95.2					
20	20	3.58	0.0383	0.0135	-0.17	0.0651	0.0650	0.0646	94.8	0.42	0.0644	0.0639	0.0643	94.2					
20	30	6.29	0.0669	0.0229	-0.71	0.0763	0.0750	0.0705	94.6	0.85	0.0721	0.0701	0.0703	94.0					
True value of $\alpha=0.25$																			
5	10	0.91	0.1717	0.1717	0.38	0.1265	0.1263	0.1235	94.3	-2.91	0.1559	0.1446	0.1368	93.4					
5	20	2.32	0.1720	0.1719	0.37	0.1421	0.1419	0.1315	94.8	-2.23	0.1600	0.1537	0.1451	94.1					
5	30	5.76	0.1726	0.1720	-0.75	0.1674	0.1667	0.1437	95.2	-2.31	0.1818	0.1758	0.1588	93.8					
10	10	-0.69	0.1110	0.1110	0.11	0.0971	0.0971	0.0902	95.1	-3.66	0.1349	0.1133	0.1014	90.8					
10	20	2.19	0.1145	0.1144	-0.30	0.1055	0.1053	0.0968	94.3	-3.63	0.1426	0.1227	0.1088	91.9					
10	30	2.41	0.1086	0.1084	-0.86	0.1136	0.1123	0.1048	95.6	-2.82	0.1402	0.1284	0.1166	93.0					
20	10	0.13	0.0819	0.0819	0.03	0.0658	0.0658	0.0645	95.0	-4.09	0.1117	0.0760	0.0730	79.8					
20	20	1.83	0.0799	0.0798	-0.13	0.0714	0.0714	0.0698	95.1	-3.62	0.1093	0.0819	0.0788	85.6					
20	30	2.38	0.0801	0.0799	-0.33	0.0806	0.0803	0.0750	94.5	-2.45	0.1033	0.0909	0.0838	91.7					

(Continued on following page)

Table 2.1 (Continued)

Frailty PMRL														Conventional PMRL									
m	C%	α			β			β															
		%Bias	RMSE	SD	%Rbias	RMSE	SD	SE	CP%	%Rbias	RMSE	SD	SE	CP%									
True value of $\alpha=0.5$																							
5	10	-7.17	0.2980	0.2958	-0.47	0.1392	0.1389	0.1349	95.1	-6.45	0.2147	0.1716	0.1562	89.8									
5	20	-7.44	0.3102	0.3080	-1.22	0.1710	0.1693	0.1469	94.5	-6.36	0.2407	0.2044	0.1699	91.7									
5	30	-8.87	0.2986	0.2953	-4.70	0.2384	0.2191	0.1697	92.2	-7.79	0.2908	0.2455	0.1935	89.5									
10	10	-6.54	0.1994	0.1967	-0.23	0.0980	0.0979	0.0973	95.3	-7.09	0.1890	0.1250	0.1150	80.1									
10	20	-6.83	0.2071	0.2043	-0.67	0.1146	0.1138	0.1077	93.5	-5.98	0.1811	0.1360	0.1263	86.0									
10	30	-7.69	0.1912	0.1873	-2.99	0.1594	0.1478	0.1223	92.8	-6.02	0.2049	0.1658	0.1437	88.4									
20	10	-7.30	0.1406	0.1358	-0.29	0.0739	0.0737	0.0714	95.8	-7.14	0.1698	0.0918	0.0849	65.8									
20	20	-5.93	0.1416	0.1385	-1.26	0.0891	0.0855	0.0796	93.9	-7.22	0.1767	0.1019	0.0946	72.2									
20	30	-6.86	0.1374	0.1331	-4.09	0.1399	0.1135	0.0952	90.1	-6.64	0.1824	0.1250	0.1124	82.8									

C% represents censoring rate in percentage; %Bias (%Rbias), RMSE, SD, SE and CP% represent mean of bias (relative bias) in percentage, root of mean square error, sample standard deviation, standard error and empirical cover probability for a nominal 95% confidence interval in percentage, respectively.

Table 2.2. Simulation results using hierarchical quasi-likelihood method based on frailty and conventional PMRL models: gamma distributed frailty with true variance parameter $\alpha=0.0, 0.25, 0.5$.

Frailty PMRL														Conventional PMRL					
m	C%	α			β					β									
		%Bias	RMSE	SD	%Rbias	RMSE	SD	SE	CP%	%Rbias	RMSE	SD	SE	CP%					
		True value of $\alpha=0.0$																	
5	10	2.32	0.0291	0.0176	-0.25	0.1245	0.1244	0.1203	94.6	-0.11	0.1228	0.1228	0.1200	95.1					
5	20	4.31	0.0546	0.0336	-1.26	0.1446	0.1424	0.1306	95.2	-0.42	0.1385	0.1382	0.1305	94.7					
5	30	6.98	0.0859	0.0501	-1.38	0.1631	0.1607	0.1390	93.8	0.35	0.1538	0.1536	0.1396	94.9					
10	10	2.44	0.0271	0.0119	0.06	0.0918	0.0918	0.0857	95.9	0.28	0.0902	0.0900	0.0854	95.4					
10	20	4.22	0.0477	0.0223	-0.45	0.1014	0.1010	0.0928	95.3	0.31	0.0979	0.0977	0.0925	95.6					
10	30	6.49	0.0723	0.0318	-1.03	0.1075	0.1055	0.0995	94.3	0.68	0.1012	0.1003	0.0996	94.9					
20	10	2.31	0.0244	0.0078	-0.11	0.0626	0.0626	0.0611	94.8	0.09	0.0615	0.0615	0.0608	94.5					
20	20	3.56	0.0377	0.0125	-0.34	0.0675	0.0672	0.0651	95.0	0.27	0.0657	0.0655	0.0648	95.1					
20	30	6.05	0.0636	0.0195	-0.71	0.0766	0.0753	0.0703	95.2	0.83	0.0738	0.0719	0.0700	94.6					
True value of $\alpha=0.25$																			
5	10	5.66	0.1748	0.1742	0.51	0.1315	0.1311	0.1278	95.4	-3.02	0.1598	0.1479	0.1412	93.1					
5	20	6.12	0.1729	0.1722	0.28	0.1430	0.1429	0.1352	94.8	-2.60	0.1700	0.1618	0.1498	92.7					
5	30	6.93	0.1644	0.1635	-1.55	0.1755	0.1727	0.1497	94.0	-3.22	0.2031	0.1926	0.1670	93.6					
10	10	5.40	0.1160	0.1152	0.12	0.0930	0.0930	0.0924	95.4	-3.92	0.1370	0.1123	0.1038	89.4					
10	20	6.00	0.1131	0.1121	-0.16	0.1058	0.1058	0.0995	95.7	-3.51	0.1414	0.1228	0.1115	91.4					
10	30	6.72	0.1136	0.1124	-0.92	0.1219	0.1205	0.1094	94.8	-2.98	0.1457	0.1330	0.1228	93.2					
20	10	4.77	0.0825	0.0816	-0.01	0.0682	0.0682	0.0660	94.8	-4.47	0.1219	0.0828	0.0747	81.7					
20	20	4.77	0.0801	0.0792	-0.08	0.0738	0.0738	0.0713	94.3	-3.80	0.1157	0.0873	0.0804	85.9					
20	30	4.91	0.0759	0.0749	-0.87	0.0863	0.0845	0.0787	95.0	-2.93	0.1097	0.0927	0.0881	90.8					

(Continued on following page)

Table 2.2 (Continued)

Frailty PMRL														Conventional PMRL									
m	C%	α			β				β														
		%Bias	RMSE	SD	%Rbias	RMSE	SD	SE	CP%	%Rbias	RMSE	SD	SE	CP%									
True value of $\alpha=0.5$																							
5	10	-6.23	0.3039	0.3023	0.35	0.1395	0.1393	0.1330	95.5	-5.32	0.1991	0.1683	0.1525	90.0									
5	20	-8.13	0.2825	0.2796	-0.31	0.1595	0.1594	0.1451	95.2	-5.28	0.2174	0.1900	0.1672	91.7									
5	30	-9.42	0.2701	0.2660	-4.23	0.2323	0.2164	0.1698	93.8	-7.74	0.2955	0.2517	0.1960	90.7									
10	10	-6.16	0.2053	0.2030	-0.20	0.1049	0.1048	0.0973	95.3	-6.62	0.1854	0.1298	0.1138	82.1									
10	20	-6.83	0.2007	0.1978	-0.98	0.1212	0.1196	0.1077	94.3	-6.55	0.1931	0.1419	0.1263	84.6									
10	30	-7.74	0.1897	0.1857	-3.26	0.1693	0.1562	0.1233	92.2	-6.87	0.2292	0.1834	0.1461	88.8									
20	10	-5.41	0.1425	0.1399	-0.01	0.0700	0.0700	0.0693	94.9	-6.66	0.1596	0.0879	0.0814	68.6									
20	20	-6.55	0.1395	0.1356	-0.39	0.0811	0.0807	0.0767	94.8	-6.30	0.1587	0.0965	0.0898	75.1									
20	30	-7.12	0.1298	0.1248	-1.94	0.1084	0.1012	0.0865	93.0	-5.60	0.1584	0.1120	0.1015	85.7									

C% represents censoring rate in percentage; %Bias (%Rbias), RMSE, SD, SE and CP% represent mean of bias (relative bias) in percentage, root of mean square error, sample standard deviation, standard error and empirical cover probability for a nominal 95% confidence interval in percentage, respectively.

Table 2.3. Simulation results using hierarchical quasi-likelihood method based on frailty and conventional PMRL models: mixture normal distributed log-frailty with $2/3N(0, 0.25) + 1/3N(0, 1)$.

		Frailty PMRL						Conventional PMRL					
		α			β			β					
m	C%	%Rbias	RMSE	SD	%Rbias	RMSE	SD	%Rbias	RMSE	SD	SE	CP%	CP%
5	10	-2.35	0.3764	0.3762	-0.30	0.1404	0.1403	-0.30	0.1404	0.1403	0.1309	96.2	90.1
5	20	-4.82	0.3415	0.3406	-1.19	0.1609	0.1591	-1.19	0.1609	0.1591	0.1441	94.3	90.3
5	30	-6.01	0.3327	0.3313	-4.10	0.2299	0.2148	-4.10	0.2299	0.2148	0.1636	93.0	91.5
10	10	1.61	0.2784	0.2783	-0.26	0.0966	0.0965	-0.26	0.0966	0.0965	0.0945	94.5	84.9
10	20	1.94	0.2720	0.2718	-0.81	0.1131	0.1119	-0.81	0.1131	0.1119	0.1047	94.4	87.4
10	30	4.83	0.2662	0.2651	-3.12	0.1620	0.1495	-3.12	0.1620	0.1495	0.1216	93.0	91.3
20	10	0.72	0.2025	0.2025	-0.26	0.0712	0.0710	-0.26	0.0712	0.0710	0.0678	95.2	71.6
20	20	-1.38	0.1896	0.1895	-0.79	0.0781	0.0765	-0.79	0.0781	0.0765	0.0748	94.4	80.1
20	30	2.64	0.1823	0.1818	-2.66	0.1098	0.0961	-2.66	0.1098	0.0961	0.0868	92.3	84.1

C% represents censoring rate in percentage; %Rbias, RMSE, SD, SE and CP% represent mean of relative bias in percentage, root of mean square error, sample standard deviation, standard error and empirical cover probability for a nominal 95% confidence interval in percentage, respectively.

Table 2.4. Simulation results using hierarchical quasi-likelihood method: obtained by a misspecified gamma frailty PMRL model.

m	C%	α			β				
		%Rbias	RMSE	SD	%Rbias	RMSE	SD	SE	CP%
True distribution of log-frailty is $N(0, 0.25)$									
5	10	-9.97	0.1381	0.1358	0.38	0.1265	0.1263	0.1235	94.3
5	20	-9.31	0.1381	0.1361	0.37	0.1421	0.1419	0.1315	94.8
5	30	-6.36	0.1379	0.1370	-0.75	0.1674	0.1667	0.1437	95.2
10	10	-9.74	0.0937	0.0905	0.11	0.0971	0.0971	0.0902	95.1
10	20	-7.67	0.0941	0.0921	-0.30	0.1055	0.1053	0.0968	94.3
10	30	-7.71	0.0902	0.0881	-0.86	0.1136	0.1123	0.1048	95.6
20	10	-9.20	0.0687	0.0647	0.03	0.0658	0.0658	0.0645	95.0
20	20	-7.61	0.0672	0.0644	-0.13	0.0714	0.0714	0.0698	95.1
20	30	-8.00	0.0659	0.0628	-0.33	0.0806	0.0803	0.0750	94.5
True distribution of log-frailty is $2/3N(0, 0.25)+1/3N(0, 1)$									
5	10	-28.43	0.2748	0.2352	-0.30	0.1404	0.1403	0.1309	96.2
5	20	-28.69	0.2612	0.2183	-1.19	0.1609	0.1591	0.1441	94.2
5	30	-28.78	0.2585	0.2148	-4.10	0.2299	0.2148	0.1637	93.0
10	10	-29.56	0.2143	0.1552	-0.26	0.0966	0.0965	0.0945	94.5
10	20	-29.42	0.2083	0.1475	-0.81	0.1131	0.1119	0.1047	94.4
10	30	-26.12	0.1966	0.1470	-3.12	0.1620	0.1495	0.1216	93.0
20	10	-32.07	0.1925	0.1065	-0.26	0.0712	0.0710	0.0678	95.2
20	20	-33.33	0.1946	0.1004	-0.79	0.0781	0.0765	0.0748	94.5
20	30	-29.23	0.1769	0.0996	-2.66	0.1098	0.0961	0.0868	92.3

C% represents censoring rate in percentage; %Rbias, RMSE, SD, SE and CP% represent mean of relative bias in percentage, root of mean square error, sample standard deviation, standard error and empirical cover probability for a nominal 95% confidence interval in percentage, respectively.

Table 2.5. Simulation results using hierarchical quasi-likelihood method: obtained by a misspecified lognormal frailty PMRL model.

m	C%	α			β				
		%Rbias	RMSE	SD	%Rbias	RMSE	SD	SE	CP%
True distribution of log-frailty is $2/3N(0, 0.25)+1/3N(0, 1)$									
5	10	−10.83	0.3556	0.3515	−0.30	0.1404	0.1403	0.1309	96.2
5	20	−13.41	0.3220	0.3149	−1.19	0.1609	0.1591	0.1441	94.2
5	30	−14.63	0.3132	0.3045	−4.10	0.2299	0.2148	0.1637	93.0
10	10	−13.76	0.2483	0.2386	−0.26	0.0966	0.0965	0.0945	94.5
10	20	−13.45	0.2431	0.2336	−0.81	0.1131	0.1119	0.1047	94.4
10	30	−10.94	0.2351	0.2286	−3.12	0.1620	0.1495	0.1216	93.0
20	10	−17.07	0.1886	0.1682	−0.26	0.0712	0.0710	0.0678	95.2
20	20	−18.84	0.1825	0.1563	−0.79	0.0781	0.0765	0.0748	94.5
20	30	−15.53	0.1688	0.1499	−2.66	0.1098	0.0961	0.0868	92.3

C% represents censoring rate in percentage; %Rbias, RMSE, SD, SE and CP% represent mean of relative bias in percentage, root of mean square error, sample standard deviation, standard error and empirical cover probability for a nominal 95% confidence interval in percentage, respectively.

Table 2.6. Simulation results using hierarchical quasi-likelihood method based on the lognormal frailty PMRL model with the true value of the frailty component $\alpha = 0.25$ and censoring rate to be 50%.

m	α			β				
	%Bias	RMSE	SD	%Rbias	RMSE	SD	SE	CP%
The cluster size $q = 50$								
5	5.69	0.2305	0.2234	-27.11	0.6360	0.3324	0.2553	64.8
10	6.45	0.1595	0.1459	-25.61	0.5729	0.2567	0.1963	51.0
20	6.24	0.1124	0.0935	-26.26	0.5550	0.1796	0.1480	13.2
The cluster size $q = 100$								
5	0.21	0.1731	0.1731	-21.02	0.5091	0.2873	0.1972	70.4
10	1.98	0.1170	0.1153	-20.25	0.4533	0.2036	0.1505	53.2
20	2.33	0.0907	0.0877	-19.77	0.4201	0.1421	0.1126	22.8

%Bias (%Rbias), RMSE, SD, SE and CP% represent mean of bias (relative bias) in percentage, root of mean square error, sample standard deviation, standard error and empirical cover probability for a nominal 95% confidence interval in percentage, respectively.

2.3 Real Data Analysis

Breast cancer is a cancer that forms in the cells of the breasts, which becomes the most common cancer in women after skin cancer. Currently, there are more than 3 million women who have been diagnosed with breast cancer in the United States. However, the number of deaths caused by breast cancer has stably declined for the past decades own to early diagnosis and timely treatment. We apply the proposed hierarchical inference method to the analysis of data from the motivating example on a multi-institutional study of breast cancer. Thirty-six cancer centers in different geographic locations in Europe and North America and with a wide range of population groups were originally included in this study (Haibe-Kains et al., 2012). Their corresponding databases can be found at <http://compbio.dfci.harvard.edu/pubs/sbtpaper/>. Survival times of patients from the same cancer center (cluster) are typically expected to have some potential dependence because of the same environmental background and medical condition. The primary interest of our analysis is to assess the contributions of prognostic factors to the mean residual lifetime of breast cancer patients by accounting for institutional variation.

As a direct measure of clinical benefit to a patient, the primary endpoint in the study is overall survival measured in days and defined as the duration from initial diagnosis to death for the patient. Two of best-established prognostic factors in breast cancer are considered: tumor size (size) in centimeters and histological grade of breast cancer (grade, ternary values 1 = “low”, 2 = “intermediate”, and 3 = “high”). Tumor size is a very important index of describing the breast cancer stage and thus directly related to an increasing probability of regional metastasis and an increasing probability of recurrence and death. The histological grade is a representative of the “aggressive potential” of the tumor, “low grade” cancers tend to be less aggressive than “high grade” cancers. It provides important information related to the clinical behavior of breast cancers (Rakha et al., 2010). It is noticed that out of all thirty-six breast cancer centers, only seven centers

contain patients with complete information for both prognostic factors and valid observations (either censored or completed) of overall survival time and thus are retrieved from the original databases. There are total $n = 1,117$ patients followed by these seven centers, including 110 patients from CAL (University of California, San Francisco, and the California Pacific Medical Center, United States), 295 from NKI (National Kanker Institute, the Netherlands), 81 from PNC, 111 from STNO (Stanford/Norway, United States and Norway), 196 from TRANSBIG (TransBIG consortium, Europe), 123 from UCSF (University of California, San Francisco, United States), and 201 from UNC (University of North Carolina, United States). Among these 1,117 patients, 169 (15.1%) experienced histological grade 1 breast cancer, 402 (36.0%) and 546 (48.9%) did grade 2 and grade 3 breast cancers, respectively. Overall, the patients' average tumor size was about 2.6 centimeters. 68.22% of patients' overall survival times are right censored.

Before applying the proposed method, we first verify the assumption of independence between the censoring time and covariates. Figure 2.1 shows the Kaplan-Meier (KM) estimates of survival rates of the censoring time by covariates tumor size and disease grade, where the continuous variable tumor size is categorized into two levels ($\text{size} \leq 2$ and $\text{size} > 2$) following the work of Haibe-Kains et al. (2012). It can be seen that the KM survival curves of the censoring time are almost overlapped regardless different categories of size or grade, indicating that there is no evidence of assumption violation.

The proposed PMRL regression model (2.2) with lognormal (or gamma) frailty, most commonly used assumption for the frailty term, is now applied to analyze the data, where all times for overall survival are divided by 30 for the ease of interpretation. We present the results obtained from the lognormal frailty model only, while results from the gamma frailty model are similar as demonstrated in the simulation studies and thus omitted here. For the sake of comparison, we also consider three alternative models to fit the data: the conventional PMRL model without frailty, the existing frailty Cox PH model using R functions *coxph* and

frailty and the frailty AFT model using function *frailtyGAF*T in R package *spBayesSurv* (<https://cran.r-project.org/web/packages/spBayesSurv>).

Fitting results are reported in Table 2.7. The proposed frailty PMRL model and the other three models all identify the factor grade as a significant factor at level of 0.05. In the proposed model, both negative estimated coefficients of grade2 and grade3 indicate that a patient with intermediate (grade = 2) or high (grade = 3) grade of breast cancer has a shorter expected residual lifetime than that with low grade one (grade = 1) in the long run. This result coincides with those obtained under the frailty AFT model and the frailty Cox PH model in which the positive corresponding estimates imply higher risk of death in comparison with the low grade of cancer. It is interesting to note that the factor size is significant in two existing frailty models but nonsignificant under the proposed model when assessing its impact on the mean residual life. To visualize such a difference, we estimate nonparametrically the logarithm of cumulative hazard and the logarithm of mean residual life curves based on the KM estimates of the overall survival time. Their corresponding plots across different levels of size are presented in Figure 2.2, which shows that the effect of factor size from the perspective of the MRL is not as great as that from the perspective of the hazard. It is also noted that size becomes significant in the conventional PMRL when ignoring the dependence within clusters.

It is found that estimates of all regression coefficients in the proposed model have smaller SEs than their counterparts in the conventional PMRL model. This finding is similar to those shown in the simulation studies, implying that to incorporate frailty in the proposed model is reasonable. Note that the asymptotic covariance of estimated coefficients $\hat{\beta}$ is computed via the sandwich formula given at the end of Section 2.1.2. This is because a patient's tumor size and disease grade are often not associated with which cancer center the patient attends, that is, the independence assumption between covariates and random effects holds. We also observe that the additional contribution of the last three terms in the

right side of Equation (2.12) is quite minor to $\widehat{Cov}(\hat{\beta})$ and thus can be ignored.

All three frailty models produce estimates of variance component α quite different from zero, indicating the non-ignorable within-cluster dependence among data. Furthermore, we seek to estimate $SE(\hat{\alpha})$ using a bootstrap method. Unlike the usual hierarchical likelihood inference for frailty models as in Ha et al. (2017), the asymptotic variance of $\hat{\alpha}$ obtained by the proposed quasi-likelihood approach is never the inverse of the negative Hessian. Following the nonparametric bootstrapping strategy suggested by Field and Welsh (2007) and Ren et al. (2010) for clustered data, we resample data from original 1,117 overall survival observations nested within $m=7$ clusters by first randomly sampling m clusters with replacement and randomly sampling q_i overall survival observations within each selected cluster i without replacement, for $i = 1, \dots, m$. We consider 1,000 bootstrap replications and apply the proposed method to each bootstrap sample, producing the bootstrap estimate of α , denoted by $\hat{\alpha}^B$ for $B = 1, \dots, 1,000$. The bootstrap SE of $\hat{\alpha}$ is then computed as $\widehat{SE}(\hat{\alpha}) = 0.2010$ by the sample standard deviation of the bootstrap estimates $\{\hat{\alpha}^B : B = 1, \dots, 1,000\}$, and 95% bootstrap confidence interval of α as (0.0462, 0.9068) by their lower and upper 2.5% quantiles. Both results indicate that random effects can not be ignored in this example.

The baseline mean residual life curve is estimated based on the proposed model and presented in Figure 2.3-A, which shows somehow a linear function of t . Similar to Chen and Cheng (2005), we also calculate the marginal mean residual life function from the Kaplan-Meier estimate of the survival function without considering any covariate information and random effects. The comparison between the estimated logarithm of baseline and logarithm of marginal mean residual life curves is given in Figure 2.3-B. The parallel trend between both curves indicates that the assumption of proportionality is deemed reasonable for this dataset.

In view that the proposed estimation hierarchical quasi-likelihood approach offers inference for random effects v_i ($i = 1, \dots, m$), it is possible to work out

the prediction about the realized value of each individual random effect, shedding lights on the heterogeneity among clusters/centers. In particular, this helps in identifying which clusters/centers perform exceptionally well or exceptionally poorly in treating the disease, or prolonging mean residual lifetime. A rough 95% confidence interval for the i -th random effect is therefore computed by $\hat{v}_i \pm 1.96\sqrt{(\mathbf{H}_{vv}^{inv}\boldsymbol{\Sigma}_{zz}\mathbf{H}_{vv}^{inv})_{ii}}$, where $\boldsymbol{\Sigma}_{zz} = \sum_{ij}\hat{\boldsymbol{\xi}}_{ij}^{z\otimes 2}$ and the definitions of $\hat{\boldsymbol{\xi}}_{ij}^z$ for $i=1, \dots, m$ and $j=1, \dots, q_i$ are given in the Appendix. The 95% confidence intervals over all seven clinical centers are presented in Figure 2.4. The confidence intervals corresponding to cancer centers CAL, NKI, STNO, TRANSBIG and UNC do not cover zero. It is evident that there is apparent heterogeneity across clusters, implying the necessity of incorporating cluster-specific random effects in the model. This also confirms the significance of α concluded from the bootstrap confidence interval aforementioned. Particularly, cancer centers STNO and TRANSBIG exhibit the outstanding negative and positive cluster impacts, respectively, on a patient's mean residual lifetime.

Table 2.7. Real data analysis for breast cancer data with focus on the overall survival time: parameter estimates based on the proposed frailty PMRL, conventional PMRL, frailty AFT and frailty Cox PH models.

Parameter	Frailty PMRL		Conventional PMRL		Frailty AFT		Frailty Cox PH	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
intercept	.		.		5.8621*	0.1650	.	
size	-0.0500	0.0436	-0.1205*	0.0595	-0.1852*	0.0414	0.2075*	0.0385
grade2	-0.3843*	0.1004	-0.4966*	0.1267	-0.5921*	0.2085	0.5173*	0.2063
grade3	-0.2741*	0.1368	-0.3781*	0.1993	-0.9217*	0.2275	0.8846*	0.1990
α	0.5042		.		0.7179		0.3910	

* indicates significant at 0.05 level.

Figure 2.1. Real data analysis for breast cancer data with focus on the overall survival time: Kaplan-Meier survival curves of the censoring time by categorized tumor size and grade.

Figure 2.2. Real data analysis for breast cancer data with focus on the overall survival time: plots of log cumulative hazard functions (left panel) and log mean residual life functions (right panel) across two different levels of size.

Figure 2.3. Real data analysis for breast cancer data with focus on the overall survival time: plots of A estimated baseline mean residual life function based on the proposed model and B comparison between the log baseline mean residual life function and log marginal mean residual life function.

Figure 2.4. Real data analysis for breast cancer data with focus on the overall survival time: plot of 95% confidence intervals for random effects in the proposed model over different centers, where the center number 1 stands for CAL, 2 for NKI, 3 for PNC, 4 for STNO, 5 for TRANSBIG, 6 for UCSF and 7 for UNC.

Appendix

We provide the derivation of Equation (2.10) and detailed notation and derivatives used in the Newton-Raphson algorithm in Section 2.1.

Define

$$\begin{aligned}
N_{ij}^c(t) &= (1 - \delta_{ij})I(Y_{ij} \leq t), \\
\hat{\pi}(t) &= n^{-1} \sum_{ij} I(Y_{ij} \geq t), \quad \hat{\Lambda}^c(t) = n^{-1} \sum_{ij} \int_0^t \frac{dN_{ij}^c(\omega)}{\hat{\pi}(\omega)}, \\
\hat{M}_{ij}^c(t) &= N_{ij}^c(t) - \int_0^t I(Y_{ij} \geq \omega) d\hat{\Lambda}^c(\omega), \\
\hat{M}_{ij}(t) &= \frac{\delta_{ij}I(Y_{ij} > t)}{\hat{G}(Y_{ij})} [(Y_{ij} - t) - \hat{m}_0(t; \hat{\beta}, \hat{\mathbf{v}}) \exp(\hat{\eta}_{ij})], \\
\hat{Q}_{xx}(t) &= n^{-1} \sum_{ij} I(Y_{ij} \geq t) \int_0^t \hat{M}_{ij}(\omega) (\mathbf{X}_{ij} - \bar{\mathbf{X}}(\omega; \hat{\beta}, \hat{\mathbf{v}})) dN(\omega), \\
\hat{Q}_{zz}(t) &= n^{-1} \sum_{ij} I(Y_{ij} \geq t) \int_0^t \hat{M}_{ij}(\omega) (\mathbf{Z}_{ij} - \bar{\mathbf{Z}}(\omega; \hat{\beta}, \hat{\mathbf{v}})) dN(\omega),
\end{aligned}$$

for $i = 1, \dots, m$ and $j = 1, \dots, q_i$, and

$$\bar{\mathbf{X}}(t; \beta, \mathbf{v}) = \frac{\sum_{ij} \frac{\delta_{ij}I(Y_{ij} > t) \mathbf{X}_{ij} \exp(\eta_{ij})}{\hat{G}(Y_{ij})}}{\sum_{ij} \frac{\delta_{ij}I(Y_{ij} > t) \exp(\eta_{ij})}{\hat{G}(Y_{ij})}}, \quad \bar{\mathbf{Z}}(t; \beta, \mathbf{v}) = \frac{\sum_{ij} \frac{\delta_{ij}I(Y_{ij} > t) \mathbf{Z}_{ij} \exp(\eta_{ij})}{\hat{G}(Y_{ij})}}{\sum_{ij} \frac{\delta_{ij}I(Y_{ij} > t) \exp(\eta_{ij})}{\hat{G}(Y_{ij})}}.$$

Following the method proposed by Sun and Zhang (2009), we construct $\hat{\xi}_{ij}^x$ in (2.12) as follows

$$\hat{\xi}_{ij}^x = \int_0^\tau \hat{M}_{ij}(t) (\mathbf{X}_{ij} - \bar{\mathbf{X}}(t; \hat{\beta}, \hat{\mathbf{v}})) dN(t) + \int_0^\tau \frac{\hat{Q}_{xx}(t)}{\hat{\pi}(t)} d\hat{M}_{ij}^c(t).$$

$\hat{\xi}_{ij}^z$ can be obtained similarly while a penalty term was added based on the quasi score for $\mathbf{v}$ in (2.10)

$$\hat{\xi}_{ij}^z = \int_0^\tau \hat{M}_{ij}(t) (\mathbf{Z}_{ij} - \bar{\mathbf{Z}}(t; \hat{\beta}, \hat{\mathbf{v}})) dN(t) + \int_0^\tau \frac{\hat{Q}_{zz}(t)}{\hat{\pi}(t)} d\hat{M}_{ij}^c(t).$$

By the definition of $\hat{m}_0$ in (2.7), we have

$$\frac{\partial \hat{m}_0(t; \beta, \mathbf{v})}{\partial \beta} = - \frac{\left[\sum_{ij} \frac{\delta_{ij}I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t) \right] \left[\sum_{ij} \frac{\delta_{ij}I(Y_{ij} > t) \mathbf{X}_{ij} \exp(\eta_{ij})}{\hat{G}(Y_{ij})} \right]}{\left[\sum_{ij} \frac{\delta_{ij}I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \exp(\eta_{ij}) \right]^2}$$

$$= -\hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \cdot \bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v}),$$

and

$$\begin{aligned} \frac{\partial \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v})}{\partial \mathbf{v}} &= - \frac{\left[\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t) \right] \left[\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{Z}_{ij}}{\hat{G}(Y_{ij})} \exp(\eta_{ij}) \right]}{\left[\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \exp(\eta_{ij}) \right]^2} \\ &= -\hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \cdot \bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v}). \end{aligned}$$

Given values of $\boldsymbol{\beta}$ and α , differentiating h_{q1} in (2.9) with respect to $\mathbf{v}$ gives

$$\begin{aligned} \partial h_{q1} / \partial \mathbf{v} &= \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \mathbf{Z}_{ij} (Y_{ij} - t) dN(t) \\ &\quad - \int_0^\tau \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t) \frac{\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{Z}_{ij} \exp(\eta_{ij})}{\hat{G}(Y_{ij})}}{\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t) \exp(\eta_{ij})}{\hat{G}(Y_{ij})}} dN(t) \\ &= \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \mathbf{Z}_{ij} (Y_{ij} - t) dN(t) \\ &\quad - \int_0^\tau \frac{\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t) (Y_{ij} - t)}{\hat{G}(Y_{ij})}}{\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t) \exp(\eta_{ij})}{\hat{G}(Y_{ij})}} \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{Z}_{ij} \exp(\eta_{ij})}{\hat{G}(Y_{ij})} dN(t) \\ &= \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \mathbf{Z}_{ij} (Y_{ij} - t) dN(t) \\ &\quad - \sum_{ij} \int_0^\tau \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{Z}_{ij} \exp(\eta_{ij})}{\hat{G}(Y_{ij})} dN(t) \\ &= \sum_{ij} \int_0^\tau \left[\frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \mathbf{Z}_{ij} (Y_{ij} - t) - \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{Z}_{ij} \exp(\eta_{ij})}{\hat{G}(Y_{ij})} \right] dN(t) \\ &= \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{Z}_{ij}}{\hat{G}(Y_{ij})} [(Y_{ij} - t) - \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \exp(\eta_{ij})] dN(t). \end{aligned}$$

Thus Equation (2.10) can be obtained. Based on quasi scores in (2.8) and (2.10), the matrix $\mathbf{H}$ has the following entries:

$$\mathbf{H}_{bb} = - \frac{\partial \mathbf{D}(\boldsymbol{\beta} | \mathbf{v})}{\partial \boldsymbol{\beta}^T}$$

$$\begin{aligned}
&= \int_0^\tau \left[\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (\mathbf{X}_{ij} - \bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v}))^{\otimes 2} \exp(\eta_{ij}) \right] \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) dN(t), \\
\mathbf{H}_{bv} &= \mathbf{H}_{vb}^T \\
&= - \frac{\partial \mathbf{D}^T(\mathbf{v} | \boldsymbol{\beta}, \alpha)}{\partial \boldsymbol{\beta}} \\
&= \int_0^\tau \left[\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \mathbf{X}_{ij} (\mathbf{Z}_{ij} - \bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v}))^T \exp(\eta_{ij}) \right] \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) dN(t) \\
&= \int_0^\tau \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (\mathbf{X}_{ij} - \bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v})) (\mathbf{Z}_{ij} - \bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v}))^T \\
&\quad \cdot \exp(\eta_{ij}) \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) dN(t), \\
\mathbf{H}_{vv} &= - \frac{\partial \mathbf{D}(\mathbf{v} | \boldsymbol{\beta}, \alpha)}{\partial \mathbf{v}^T} \\
&= \int_0^\tau \left[\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (\mathbf{Z}_{ij} - \bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v}))^{\otimes 2} \exp(\eta_{ij}) \right] \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) dN(t) \\
&\quad - \frac{\partial^2 h_2(\alpha; \mathbf{v})}{\partial \mathbf{v} \partial \mathbf{v}^T}.
\end{aligned}$$

The justification of the positive definiteness of $\mathbf{H}$ is given as follows.

For any $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{v}^T)^T \in \mathbb{R}^{p+m}$, we have that

$$\begin{aligned}
\boldsymbol{\theta}^T \mathbf{H} \boldsymbol{\theta} &= (\boldsymbol{\beta}^T, \mathbf{v}^T) \begin{bmatrix} \mathbf{H}_{bb} & \mathbf{H}_{bv} \\ \mathbf{H}_{vb} & \mathbf{H}_{vv} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{v} \end{pmatrix} \\
&= \boldsymbol{\beta}^T \mathbf{H}_{bb} \boldsymbol{\beta} + 2\boldsymbol{\beta}^T \mathbf{H}_{bv} \mathbf{v} + \mathbf{v}^T \mathbf{H}_{vv} \mathbf{v}.
\end{aligned}$$

It follows from the definitions of $\mathbf{H}_{bb}$, $\mathbf{H}_{vv}$ and $\mathbf{H}_{bv}$ that

$$\begin{aligned}
\boldsymbol{\beta}^T \mathbf{H}_{bb} \boldsymbol{\beta} &= \int_0^\tau \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (\boldsymbol{\beta}^T \mathbf{X}_{ij} - \boldsymbol{\beta}^T \bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v}))^2 \exp(\eta_{ij}) \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) dN(t), \\
\mathbf{v}^T \mathbf{H}_{vv} \mathbf{v} &= \int_0^\tau \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (\mathbf{v}^T \mathbf{Z}_{ij} - \mathbf{v}^T \bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v}))^2 \exp(\eta_{ij}) \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) dN(t) \\
&\quad - \mathbf{v}^T \frac{\partial^2 h_2(\alpha; \mathbf{v})}{\partial \mathbf{v} \partial \mathbf{v}^T} \mathbf{v},
\end{aligned}$$

and

$$\boldsymbol{\beta}^T \mathbf{H}_{bv} \mathbf{v} = \int_0^\tau \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (\boldsymbol{\beta}^T \mathbf{X}_{ij} - \boldsymbol{\beta}^T \bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v})) (\mathbf{v}^T \mathbf{Z}_{ij} - \mathbf{v}^T \bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v}))$$

$$\cdot \exp(\eta_{ij}) \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \, dN(t).$$

Therefore,

$$\begin{aligned} \boldsymbol{\theta}^T \mathbf{H} \boldsymbol{\theta} &= \int_0^\tau \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \left[(\boldsymbol{\beta}^T \mathbf{X}_{ij} - \boldsymbol{\beta}^T \bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v}))^2 + 2(\boldsymbol{\beta}^T \mathbf{X}_{ij} - \boldsymbol{\beta}^T \bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v})) \right. \\ &\quad \cdot (\mathbf{v}^T \mathbf{Z}_{ij} - \mathbf{v}^T \bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v})) + (\mathbf{v}^T \mathbf{Z}_{ij} - \mathbf{v}^T \bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v}))^2 \left. \right] \exp(\eta_{ij}) \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \\ &\quad \cdot dN(t) - \mathbf{v}^T \frac{\partial^2 h_2(\alpha; \mathbf{v})}{\partial \mathbf{v} \partial \mathbf{v}^T} \mathbf{v} \\ &= \int_0^\tau \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \left[(\boldsymbol{\beta}^T \mathbf{X}_{ij} - \boldsymbol{\beta}^T \bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v})) + (\mathbf{v}^T \mathbf{Z}_{ij} - \mathbf{v}^T \bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v})) \right]^2 \\ &\quad \cdot \exp(\eta_{ij}) \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \, dN(t) - \mathbf{v}^T \frac{\partial^2 h_2(\alpha; \mathbf{v})}{\partial \mathbf{v} \partial \mathbf{v}^T} \mathbf{v} > 0 \end{aligned}$$

for all $\boldsymbol{\theta} \in \mathbb{R}^{p+m} \setminus \{\mathbf{0}\}$. The last inequality holds because the first term

$$\begin{aligned} &\int_0^\tau \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \left[(\boldsymbol{\beta}^T \mathbf{X}_{ij} - \boldsymbol{\beta}^T \bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v})) + (\mathbf{v}^T \mathbf{Z}_{ij} - \mathbf{v}^T \bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v})) \right]^2 \\ &\quad \cdot \exp(\eta_{ij}) \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \, dN(t) \end{aligned}$$

is an integral of a nonnegative sum, and the second term $-\mathbf{v}^T \frac{\partial^2 h_2(\alpha; \mathbf{v})}{\partial \mathbf{v} \partial \mathbf{v}^T} \mathbf{v}$ is a positive quadratic form where the observed fisher information of h_2 , i.e., $-\frac{\partial^2 h_2(\alpha; \mathbf{v})}{\partial \mathbf{v} \partial \mathbf{v}^T}$, is often assumed to be positive definite to ensure the solution of MLE exists especially when using Newton-Raphson algorithm.

Thus, $\mathbf{H}$ is positive definite.

Conditional Inference for Clustered Survival Data Based on Mean Residual Life Models

The hierarchical quasi-likelihood method proposed in Chapter 2 still requires a parametric distribution assumption for random effects though such a distribution is not necessary to be gamma or lognormal. Note that there is little information about the true distribution of random effects in practice and thus the distributional assumption is hard to be verified. Although simulation results obtained in Chapter 2 exhibit that the estimated fixed effects are quite robust when the distributional assumption of random effects is altered or violated, it is necessary from the methodological context to further investigate a more flexible model with an unspecified distribution of random effects.

As reviewed in Section 1.5, the conditional inference without parametric distribution assumption for random effects has been used for the analysis of clustered data in generalized linear mixed models. For example, Jiang (1999) proposed the penalized conditional quasi-log-likelihood function (1.11) and developed a penalized generalized weighted least square estimation procedure for the estimation of fixed and random effects; based on Jiang (1999)'s method, the penalized quadratic

inference function, a hybrid between quasi scores derived from penalized conditional quasi-log-likelihood and quadratic inference functions, was developed by Wang et al. (2012). However, these works focus on the analysis of exactly observed data and are incapable of handling survival data with censoring of our interest in this thesis. In this chapter, we propose a more flexible frailty proportional mean residual life (PMRL) model for survival analysis with clustered data. Based on the conditional quasi-likelihood proposed in Chapter 2, a penalized conditional quasi-likelihood (PCQL) function is developed and estimating equations are then derived for both fixed and random effects based on penalized quasi scores. The conditional inference method makes the frailty free of the parametric distribution assumption and the large sample properties of the proposed estimator are established.

3.1 Modeling Methodologies

3.1.1 Model elaboration

Similar to the notation used in Chapter 2, let $\{(Y_{ij}, \delta_{ij}, \mathbf{X}_{ij}, \mathbf{Z}_{ij}) : i = 1, \dots, m, j = 1, \dots, q_i\}$ be observations from all individuals in the study, where $Y_{ij} = \min(T_{ij}, C_{ij})$ is the observation of survival outcome T_{ij} with the censoring indicator $\delta_{ij} = I(T_{ij} \leq C_{ij})$, C_{ij} is the potential censoring time. $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ are covariate vectors associated with fixed effects and random effects, respectively. Let v_i be the cluster-specific random effect and $\mathbf{v} = (v_1, \dots, v_m)^T$ be the vector of random effects. The conditional MRL function of T_{ij} given covariates $\mathbf{X}_{ij}, \mathbf{Z}_{ij}$ and $\mathbf{v}$ is assumed by a proportional MRL (PMRL) regression in the following form:

$$m(t|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}) = m_0(t) \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_{ij}^T \mathbf{v}), \quad (3.1)$$

where $m_0(t)$ is the unknown baseline MRL function and $\boldsymbol{\beta}$ is the fixed-effects coefficient vector with dimension p . For simplicity, we denote $\eta_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_{ij}^T \mathbf{v}$ hereinafter.

Although model (3.1) has the same form as model (2.2), it differs from (2.2) by allowing non-intercept random effects in the model and letting random effects free of the distributional assumption except $E(\mathbf{v}) = \mathbf{0}$. Thus, model (3.1) extends (2.2) to a more general model framework in the analysis of clustered survival data.

3.1.2 Penalized conditional quasi-likelihood estimation

Sun and Zhang (2009) developed estimating equations for the analysis of MRL regression models with independent survival data. In presence of random effects $\mathbf{v}$, Chapter 2 supposes a conditional quasi-likelihood function of $\boldsymbol{\beta}$ given $\mathbf{v}$, which is formed from integrating the quasi score $\mathbf{D}(\boldsymbol{\beta}|\mathbf{v})$ in (2.8). Following this line, we construct a similar conditional quasi-likelihood function of $\boldsymbol{\beta}$ given $\mathbf{v}$ based on observations $\{(Y_{ij}, \delta_{ij}, \mathbf{X}_{ij}, \mathbf{Z}_{ij}) : i=1, \dots, m, j=1, \dots, q_i\}$ in the form of

$$\begin{aligned} l_c(\boldsymbol{\beta}|\mathbf{v}) &= \sum_{ij} l_{cij}(\boldsymbol{\beta}|\mathbf{v}) \\ &= \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t) \left[\eta_{ij} - \log \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \exp(\eta_{ij}) \right] dN(t). \end{aligned} \quad (3.2)$$

Similar to Chapter 2, we take the inverse probability of censoring weighting (IPCW) as $\frac{\delta_{ij}}{\hat{G}(Y_{ij})}$ with the Kaplan-Meier estimator $\hat{G}(t) = \prod_{k: C_{(k)} \leq t} \left(1 - \frac{d_{(k)}^c}{n_{(k)}^c} \right)$ for the survival function of censoring times, the estimator of the baseline MRL function $m_0(t)$ as that given in (2.7), and the same counting process $N(t) = \sum_{ij} \delta_{ij} I(Y_{ij} \leq t)$ as the increasing weight on $[0, \tau]$ with $\tau = \inf \{t : P(T \geq t) = 0\}$.

Note that the conditional quasi-likelihood function (3.2) depends on both fixed effects $\boldsymbol{\beta}$ and random effects $\mathbf{v}$ through $\eta_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_{ij}^T \mathbf{v}$, where $\boldsymbol{\beta}$ and $\mathbf{v}$ may not be identifiable without additional information. The constraint $E(\mathbf{v}) = \mathbf{0}$ has to be taken into account to avoid the parameter identifiability issue in the model. Along the similar line of conditional inference used in GLMMs (Jiang, 1999; Liu and Xiang, 2019; Wang et al., 2012), we incorporate a penalty term, which corresponds to placing the zero-mean constraint on random effects, in the

conditional quasi-likelihood function (3.2) and define the penalized conditional quasi-likelihood (PCQL) function as follows:

$$l_p(\boldsymbol{\beta}, \mathbf{v}, \gamma) = l_c(\boldsymbol{\beta}|\mathbf{v}) - \gamma \sum_i \frac{v_i^2}{2}, \quad (3.3)$$

where γ is a Lagrangian multiplier, which is also known as a tuning parameter and tunes the overall impact of the penalty term. As argued in Jiang (1999), the choice of the tuning parameter would not affect the consistency of the estimators but might affect the asymptotic efficiency and the small sample behavior.

The proposed PCQL function accommodates both fixed effects and random effects based on the frailty PMRL model (3.1) with the constraint $E(\mathbf{v}) = \mathbf{0}$. Given a value of $\mathbf{v}$, its derivative with respect to $\boldsymbol{\beta}$ leads to the following penalized quasi score of fixed effects:

$$\mathbf{U}(\boldsymbol{\beta}|\mathbf{v}) = \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{X}_{ij}}{\hat{G}(Y_{ij})} [(Y_{ij} - t) - \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \exp(\eta_{ij})] dN(t). \quad (3.4)$$

This penalized quasi score has an equivalent form as the quasi score $\mathbf{D}(\boldsymbol{\beta}|\mathbf{v})$ for $\boldsymbol{\beta}$ given $\mathbf{v}$ in (2.8) due to the fact that the MRL conditional on random effects does not depend on random effects themselves.

For a fixed tuning parameter γ and a given $\boldsymbol{\beta}$, differentiating the PCQL function (3.3) with respect to $\mathbf{v}$ produces the penalized quasi score of random effects:

$$\mathbf{U}(\mathbf{v}|\boldsymbol{\beta}) = \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{Z}_{ij}}{\hat{G}(Y_{ij})} [(Y_{ij} - t) - \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \exp(\eta_{ij})] dN(t) - \gamma \mathbf{v}. \quad (3.5)$$

Then, estimators of $\boldsymbol{\beta}$ and $\mathbf{v}$, denoting by $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{v}}$, can be obtained by solving the penalized quasi-score estimating equations $\mathbf{U}(\boldsymbol{\beta}|\mathbf{v}) = \mathbf{0}$ and $\mathbf{U}(\mathbf{v}|\boldsymbol{\beta}) = \mathbf{0}$ iteratively. Letting $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{v}^T)^T$, we define $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\mathbf{v}}^T)^T$ as the maximum penalized conditional quasi-likelihood estimator (MPCQLE) of $\boldsymbol{\theta}$. In this estimation procedure, random effects are treated as nuisance parameters to be estimated. The variance component associated with random effects can be simply estimated by the sample variance of $\hat{\mathbf{v}}$.

Like other regularization methods, the choice of the value of γ is important in the proposed estimation method. Following Wang et al. (2012), we choose $\log(n)$ as the tuning parameter γ , where $n = \sum_{i=1}^m q_i$ is the total sample size. It works very well in our numerical studies. The tuning parameter γ can also be determined using a BIC-like model selection criterion associated with the penalized quasi-likelihood function.

3.2 Implementation

Given a selected γ , the estimating procedure is easy to be implemented with any iterative algorithm. In this section, the Newton-Raphson method is adopted to obtain the estimates of fixed and random effects. The corresponding algorithm comprises the following steps:

Step 0: Set initial values $\hat{\beta}^{(0)}, \hat{v}^{(0)}$.

Step 1: In the k -th step, denote the MPCQLE as $\hat{\theta}^{(k)} = (\hat{\beta}^{(k)T}, \hat{v}^{(k)T})^T$ and update the penalized quasi scores of fixed and random effects as $U(\hat{\beta}^{(k)}|\hat{v}^{(k)})$ and $U(\hat{v}^{(k)}|\hat{\beta}^{(k)})$, respectively.

Step 2: Solve the following equations iteratively to update $\hat{\beta}^{(k+1)}$ and $\hat{v}^{(k+1)}$:

$$\begin{cases} \hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + \left[\hat{A}_{bb}^{inv}(\hat{\beta}^{(k)}, \hat{v}^{(k)})U(\hat{\beta}^{(k)}|\hat{v}^{(k)}) + \hat{A}_{bv}^{inv}(\hat{\beta}^{(k)}, \hat{v}^{(k)})U(\hat{v}^{(k)}|\hat{\beta}^{(k)}) \right], \\ \hat{v}^{(k+1)} = \hat{v}^{(k)} + \left[\hat{A}_{vb}^{inv}(\hat{\beta}^{(k)}, \hat{v}^{(k)})U(\hat{\beta}^{(k)}|\hat{v}^{(k)}) + \hat{A}_{vv}^{inv}(\hat{\beta}^{(k)}, \hat{v}^{(k)})U(\hat{v}^{(k)}|\hat{\beta}^{(k)}) \right]. \end{cases}$$

Details of matrices $\hat{A}_{bb}^{inv}(\beta, v)$, $\hat{A}_{bv}^{inv}(\beta, v)$, $\hat{A}_{vb}^{inv}(\beta, v)$ and $\hat{A}_{vv}^{inv}(\beta, v)$ are given in Section 3.3.1.

Step 3: Repeat **Steps 1-2** until the convergence is achieved.

Step 4: Estimate the variance component of random effects by the sample variance of the estimated random effects.

Regarding to the choice of initial values, we take $\hat{\beta}^{(0)} = \mathbf{0}$ and $\hat{\mathbf{v}}^{(0)} = \mathbf{0}$ in our numerical studies. As discussed in Chapter 2, an alternative choice for $\hat{\beta}^{(0)}$ is a rough estimate of β obtained from the conventional PMRL model. A stopping rule $\max \{|\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}|\} < 10^{-4}$ is employed to monitor the convergence in the numerical studies.

3.3 Notation and Conditions

In this section, we state details of notation used in this chapter and regularity conditions required in the theoretical justification.

3.3.1 Derivatives and notation

Based on the penalized quasi scores given in (3.4) and (3.5), the following derivatives are obtained:

$$\begin{aligned}\hat{A}_{bb}(\beta, \mathbf{v}) &= -\frac{1}{n} \frac{\partial \mathbf{U}(\beta | \mathbf{v})}{\partial \beta^T} \\ &= \frac{1}{n} \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} [\mathbf{X}_{ij} - \bar{\mathbf{X}}(t; \beta, \mathbf{v})]^{\otimes 2} \hat{m}_0(t; \beta, \mathbf{v}) \exp(\eta_{ij}) dN(t), \\ \hat{A}_{bv}(\beta, \mathbf{v}) &= \hat{A}_{vb}^T(\beta, \mathbf{v}) \\ &= -\frac{1}{n} \frac{\partial \mathbf{U}^T(\mathbf{v} | \beta)}{\partial \beta} \\ &= \frac{1}{n} \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \mathbf{X}_{ij} [\mathbf{Z}_{ij} - \bar{\mathbf{Z}}(t; \beta, \mathbf{v})]^T \hat{m}_0(t; \beta, \mathbf{v}) \exp(\eta_{ij}) dN(t), \\ \hat{A}_{vv}(\beta, \mathbf{v}) &= -\frac{1}{n} \frac{\partial \mathbf{U}(\mathbf{v} | \beta)}{\partial \mathbf{v}^T} \\ &= \frac{1}{n} \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} [\mathbf{Z}_{ij} - \bar{\mathbf{Z}}(t; \beta, \mathbf{v})]^{\otimes 2} \hat{m}_0(t; \beta, \mathbf{v}) \exp(\eta_{ij}) dN(t) + \frac{\gamma}{n} \mathbf{I}_m,\end{aligned}$$

where $\mathbf{I}_m$ is the identity matrix of size m ,

$$\bar{\mathbf{X}}(t; \beta, \mathbf{v}) = \frac{\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{X}_{ij} \exp(\eta_{ij})}{\hat{G}(Y_{ij})}}{\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t) \exp(\eta_{ij})}{\hat{G}(Y_{ij})}}, \quad \bar{\mathbf{Z}}(t; \beta, \mathbf{v}) = \frac{\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{Z}_{ij} \exp(\eta_{ij})}{\hat{G}(Y_{ij})}}{\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t) \exp(\eta_{ij})}{\hat{G}(Y_{ij})}}$$

and

$$\hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) = \frac{\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t)}{\sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \exp(\eta_{ij})}.$$

By the uniform strong law of large numbers, it can be shown that $\hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v})$ converges to $m_0(t)$ almost surely and uniformly in t . In a similar way, we can also prove that $\bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v})$ and $\bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v})$ are convergent, that is, there exist $\bar{\mathbf{x}}(t; \boldsymbol{\beta}, \mathbf{v})$ and $\bar{\mathbf{z}}(t; \boldsymbol{\beta}, \mathbf{v})$ such that

$$\bar{\mathbf{x}}(t; \boldsymbol{\beta}, \mathbf{v}) = \lim_{n \rightarrow \infty} \bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v}), \quad \bar{\mathbf{z}}(t; \boldsymbol{\beta}, \mathbf{v}) = \lim_{n \rightarrow \infty} \bar{\mathbf{Z}}(t; \boldsymbol{\beta}, \mathbf{v}).$$

Thus, $\hat{\mathbf{A}}_{bb}(\boldsymbol{\beta}, \mathbf{v})$ and $\hat{\mathbf{A}}_{bv}(\boldsymbol{\beta}, \mathbf{v})$ converge to the following nonrandom matrices respectively:

$$\begin{aligned} \mathbf{A}_{bb}(\boldsymbol{\beta}, \mathbf{v}) &= E \left\{ \int_0^\tau S(t | \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}) [\mathbf{X}_{ij} - \bar{\mathbf{x}}(t; \boldsymbol{\beta}, \mathbf{v})]^{\otimes 2} m(t | \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}) d\tilde{N}(t) \right\}, \\ \mathbf{A}_{bv}(\boldsymbol{\beta}, \mathbf{v}) &= E \left\{ \int_0^\tau S(t | \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}) \mathbf{X}_{ij} [\mathbf{Z}_{ij} - \bar{\mathbf{z}}(t; \boldsymbol{\beta}, \mathbf{v})]^T m(t | \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}) d\tilde{N}(t) \right\}, \end{aligned}$$

where $S(t | \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v})$ is the conditional survival function of T_{ij} given covariates $\mathbf{X}_{ij}, \mathbf{Z}_{ij}$ and random effects $\mathbf{v}$.

Furthermore, the matrices used in Section 3.2 are defined as follows:

$$\begin{aligned} \hat{\mathbf{A}}_{bb}^{inv}(\boldsymbol{\beta}, \mathbf{v}) &= n^{-1} [\hat{\mathbf{A}}_{bb}(\boldsymbol{\beta}, \mathbf{v}) - \hat{\mathbf{A}}_{bv}(\boldsymbol{\beta}, \mathbf{v}) \hat{\mathbf{A}}_{vv}^{-1}(\boldsymbol{\beta}, \mathbf{v}) \hat{\mathbf{A}}_{vb}(\boldsymbol{\beta}, \mathbf{v})]^{-1}, \\ \hat{\mathbf{A}}_{bv}^{inv}(\boldsymbol{\beta}, \mathbf{v}) &= -\hat{\mathbf{A}}_{bb}^{inv}(\boldsymbol{\beta}, \mathbf{v}) \hat{\mathbf{A}}_{bv}(\boldsymbol{\beta}, \mathbf{v}) \hat{\mathbf{A}}_{vv}^{-1}(\boldsymbol{\beta}, \mathbf{v}), \\ \hat{\mathbf{A}}_{vb}^{inv}(\boldsymbol{\beta}, \mathbf{v}) &= [\hat{\mathbf{A}}_{bv}^{inv}(\boldsymbol{\beta}, \mathbf{v})]^T, \\ \hat{\mathbf{A}}_{vv}^{inv}(\boldsymbol{\beta}, \mathbf{v}) &= n^{-1} \hat{\mathbf{A}}_{vv}^{-1}(\boldsymbol{\beta}, \mathbf{v}) + \hat{\mathbf{A}}_{vv}^{-1}(\boldsymbol{\beta}, \mathbf{v}) \hat{\mathbf{A}}_{vb}(\boldsymbol{\beta}, \mathbf{v}) \hat{\mathbf{A}}_{bb}^{inv}(\boldsymbol{\beta}, \mathbf{v}) \hat{\mathbf{A}}_{bv}(\boldsymbol{\beta}, \mathbf{v}) \hat{\mathbf{A}}_{vv}^{-1}(\boldsymbol{\beta}, \mathbf{v}). \end{aligned}$$

By substituting $m_0(t)$ and $G(Y_{ij})$ with their estimates $\hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v})$ and $\hat{G}(Y_{ij})$ in the zero-mean stochastic process (2.5), we define

$$\hat{M}_{ij}(t; \boldsymbol{\beta}, \mathbf{v}) = \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} [(Y_{ij} - t) - \hat{m}(t; \boldsymbol{\beta}, \mathbf{v})],$$

where $\hat{m}(t; \boldsymbol{\beta}, \mathbf{v}) = \hat{m}_0(t; \boldsymbol{\beta}, \mathbf{v}) \exp(\eta_{ij})$. For the convenience, let $\eta_{0ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta}_0 + \mathbf{Z}_{ij}^T \mathbf{v}_0$, where $\boldsymbol{\beta}_0$ and $\mathbf{v}_0$ are the true values of $\boldsymbol{\beta}$ and $\mathbf{v}$, and $m(t) = m_0(t) \exp(\eta_{0ij})$.

Denote the estimated fixed effects β given the true random effects $\mathbf{v}_0$ as $\hat{\beta}^0$ and that given the estimated random effects $\hat{\mathbf{v}}$ as $\hat{\beta}$, respectively. To further estimate the asymptotic covariance matrix of the estimated fixed effects, let

$$\hat{\mathbf{Q}}(t; \beta, \mathbf{v}) = n^{-1} \sum_{ij} I(Y_{ij} \geq t) \int_0^\tau \hat{M}_{ij}(w; \beta, \mathbf{v}) (\mathbf{X}_{ij} - \bar{\mathbf{X}}(w; \beta, \mathbf{v})) dN(w).$$

3.3.2 Regularity conditions and lemma

Before establishing the theoretical properties, some regularity conditions and Lemma for preliminary results are needed.

(C1) The increasing weight function $N(t)$ converges almost surely to a nonrandom and bounded function $\tilde{N}(t)$ on $[0, \tau]$.

(C2) The covariates $\mathbf{X}$ and $\mathbf{Z}$ are bounded.

(C3) The baseline MRL function $m_0(t)$ is continuously differentiable and the survival function of censoring time $G(t)$ is continuous on $[0, \tau]$.

(C4) $\mathbf{A}_{bb} = \mathbf{A}_{bb}(\beta_0, \mathbf{v}_0)$ is nonsingular and bounded in the sense that the maximum eigenvalue of the matrix $\mathbf{A}_{bb}$ is bounded, and $\hat{\mathbf{A}}_{bb}(\beta_0, \hat{\mathbf{v}})$ is uniformly bounded in the sense that the maximum eigenvalue of the matrix $\hat{\mathbf{A}}_{bb}(\beta_0, \hat{\mathbf{v}})$ is bounded for any value of $\hat{\mathbf{v}}$.

(C5) $\mathbf{A}_{bv} = \mathbf{A}_{bv}(\beta_0, \mathbf{v}_0)$ and $\|\mathbf{A}_{bv}\|_F$ uniformly approaches 0, where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

(C6) Given the true values of random effects $\mathbf{v}_0$, there is a unique β_0 which satisfies

$$E \left\{ \int_0^\tau m(t|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}_0) [\mathbf{X}_{ij} - \bar{\mathbf{x}}(t; \beta_0, \mathbf{v}_0)] d\tilde{N}(t) \right\} = \mathcal{O}(n^{-1/2}).$$

(C7) The expectation of the integral for the difference between $m(t|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}_0) \cdot \bar{\mathbf{x}}(t; \beta_0, \mathbf{v}_0)$ and $m(t|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}) \bar{\mathbf{x}}(t; \beta_0, \mathbf{v})$ converges to 0 in probability with the rate $1/\sqrt{n}$, that is,

$$E \left\{ \int_0^\tau [m(t|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}_0) \bar{\mathbf{x}}(t; \beta_0, \mathbf{v}_0) - m(t|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}) \bar{\mathbf{x}}(t; \beta_0, \mathbf{v})] d\tilde{N}(t) \right\} = \mathcal{O}(n^{-1/2}).$$

(C1)-(C4) are some regularity conditions required in proofs of the consistency and asymptotic normality properties. Condition (C5) indicates that random effects are independent of fixed effects. Combined with condition (C2), condition (C6) implies that the parameter vector β is identifiable conditional on the true values of random effects $\mathbf{v}_0$, that is, there exists a unique β_0 satisfying that

$$U(\beta_0|\mathbf{v}_0) = \mathcal{O}_p(\sqrt{n}).$$

Based on condition (C7), it can be derived that

$$U(\beta_0|\mathbf{v}) - U(\beta_0|\mathbf{v}_0) = \mathcal{O}_p(\sqrt{n})$$

uniformly in $\mathbf{v}$. It should be pointed out that the estimated random effects $\hat{\mathbf{v}}$ is not required to be consistent with $\mathbf{v}_0$.

Lemma 3.1. *Under the regularity conditions (C1)-(C4), we have that*

- (i) *there exists a unique and strongly consistent estimator $\hat{\beta}^0$ such that*

$$U(\hat{\beta}^0|\mathbf{v}_0) = \mathbf{0}.$$
- (ii) *$\sqrt{n}(\hat{\beta}^0 - \beta_0)$ is asymptotically normally distributed with mean zero and a covariance matrix Ω_0 .*

Remarks:

1. Lemma 3.1 tells us that the estimated fixed effects given the true values of random effects are consistent and asymptotically normally distributed. It provides initial results in the subsequent proofs of Theorems 3.1 and 3.2.
2. The asymptotic covariance matrix Ω_0 of $\sqrt{n}\hat{\beta}^0$ can be consistently estimated by $\hat{\mathbf{A}}_{bb}^{-1}(\hat{\beta}^0, \mathbf{v}_0)\Sigma^0\hat{\mathbf{A}}_{bb}^{-1}(\hat{\beta}^0, \mathbf{v}_0)$, where $\Sigma^0 = \frac{1}{n} \sum_{ij} \hat{\xi}_{ij}^{0\otimes 2}$ is constructed based on the outer product of vector $\hat{\xi}_{ij}^0$ and $\hat{\xi}_{ij}^0$ is defined as

$$\hat{\xi}_{ij}^0 = \int_0^\tau \hat{M}_{ij}(t; \hat{\beta}^0, \mathbf{v}_0) [\mathbf{X}_{ij} - \bar{\mathbf{X}}(t; \hat{\beta}^0, \mathbf{v}_0)] dN(t) + \int_0^\tau \frac{\hat{Q}(t; \hat{\beta}^0, \mathbf{v}_0)}{\hat{\pi}(t)} d\hat{M}_{ij}^c(t),$$

where $\hat{M}_{ij}^c(t)$ is a martingale-like process associated with censoring times and has the same form as that defined in Chapter 2.

Proof of Lemma 3.1. When the true value of the random-effects vector $\mathbf{v}_0$ is given, our model degenerates to the conventional PMRL regression model without random effects, which is also a special case of the transformed MRL model studied by Sun and Zhang (2009). As a consequence, the unique existence and the consistency of $\hat{\beta}^0$ and its asymptotic normality can be immediately obtained from Theorem 1(i) and (ii) in Sun and Zhang (2009), respectively. $\square$

3.4 Theoretical Results

Given conditions (C1)-(C7) and Lemma 3.1, asymptotic results including the consistency and the asymptotic normality of the proposed estimator of the fixed-effects vector $\hat{\beta}$ obtained by solving $\mathbf{U}(\beta|\hat{\mathbf{v}}) = \mathbf{0}$ are summarized in Theorems 3.1 and 3.2 and their proofs are provided subsequently.

Theorem 3.1 (Consistency). *The proposed estimator $\hat{\beta}$ is $\sqrt{n}$ -consistent, which means in some neighbor of β_0 , $\hat{\beta}$ satisfies that*

$$\sqrt{n}\|\hat{\beta} - \beta_0\|_2 = \mathcal{O}_p(1),$$

where $\|\cdot\|_2$ is $\mathbf{L}_2$ norm for a vector.

Theorem 3.2 (Asymptotic Normality). *The proposed estimator $\hat{\beta}$ is asymptotic normality as $n \rightarrow \infty$, that is,*

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(\mathbf{0}, \Omega).$$

In particular, $\Omega = \Omega_0$ if $\hat{\mathbf{v}}$ is consistent.

Similar to the construction of the estimate of Ω_0 given in the remark 2 of Lemma 3.1, Ω can be approximated by $\hat{\mathbf{A}}_{bb}^{-1}(\hat{\beta}, \hat{\mathbf{v}}) \Sigma \hat{\mathbf{A}}_{bb}^{-1}(\hat{\beta}, \hat{\mathbf{v}})$, where $\Sigma = \frac{1}{n} \sum_{ij} \hat{\xi}_{ij}^{\otimes 2}$ and $\hat{\xi}_{ij}$ is in the form of

$$\hat{\xi}_{ij} = \int_0^\tau \hat{M}_{ij}(t; \hat{\beta}, \hat{\mathbf{v}}) [\mathbf{X}_{ij} - \bar{\mathbf{X}}(t; \hat{\beta}, \hat{\mathbf{v}})] dN(t) + \int_0^\tau \frac{\hat{Q}(t; \hat{\beta}, \hat{\mathbf{v}})}{\hat{\pi}(t)} d\hat{M}_{ij}^c(t).$$

Proof of Theorem 3.1. Define $\mathcal{B}_n(c, \beta_0) = \{\beta : \|\beta - \beta_0\|_2 = c/\sqrt{n}\}$, where c is a fixed constant. To show the consistency of $\hat{\beta}$, we only need to prove that there exists a unique solution to $\mathbf{U}(\beta|\hat{\mathbf{v}}) = \mathbf{0}$ on $\mathcal{B}_n(c, \beta_0)$. It suffices to prove that for any $\epsilon > 0$,

$$P\left\{\sup_{\beta \in \mathcal{B}_n(c, \beta_0)} (\beta - \beta_0)^T \mathbf{U}(\beta|\hat{\mathbf{v}}) \leq 0\right\} > 1 - \epsilon. \quad (3.6)$$

We determine the order of $\frac{1}{\sqrt{n}}\mathbf{U}(\beta|\hat{\mathbf{v}})$ first. Note that $\hat{\mathbf{A}}_{bb}(\beta, \mathbf{v}) = -\frac{1}{n} \frac{\partial \mathbf{U}(\beta|\mathbf{v})}{\partial \beta^T}$, it follows from the Taylor expansion that

$$\begin{aligned} \frac{1}{\sqrt{n}}\mathbf{U}(\beta|\hat{\mathbf{v}}) &= \frac{1}{\sqrt{n}}\mathbf{U}(\beta_0|\hat{\mathbf{v}}) - \sqrt{n}\hat{\mathbf{A}}_{bb}(\beta_0, \hat{\mathbf{v}})(\beta - \beta_0) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\mathbf{U}(\beta_0|\mathbf{v}_0) + \frac{1}{\sqrt{n}}[\mathbf{U}(\beta_0|\hat{\mathbf{v}}) - \mathbf{U}(\beta_0|\mathbf{v}_0)] \\ &\quad - \sqrt{n}\hat{\mathbf{A}}_{bb}(\beta_0, \hat{\mathbf{v}})(\beta - \beta_0) + o_p(1). \end{aligned} \quad (3.7)$$

Let $f_Y(\cdot|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v})$ and $f_T(\cdot|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v})$ be the conditional probability density functions of Y_{ij} and T_{ij} , respectively, given random effects $\mathbf{v}$ and covariates $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$, and $f_C(\cdot)$ be the probability density function of the censoring time C_{ij} . Then

$$f_Y(y|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}) = f_T(y|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v})G(y) + S(y|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v})f_C(y).$$

By the law of large numbers, we have

$$\begin{aligned} &\frac{1}{n}\mathbf{U}(\beta_0|\mathbf{v}_0) \\ &= \frac{1}{n} \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{X}_{ij}}{\hat{G}(Y_{ij})} [(Y_{ij} - t) - \hat{m}(t; \beta_0, \mathbf{v}_0)] dN(t) \\ &= \frac{1}{n} \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) (Y_{ij} - t)}{\hat{G}(Y_{ij})} [\mathbf{X}_{ij} - \bar{\mathbf{X}}(t; \beta_0, \mathbf{v}_0)] dN(t) \\ &\xrightarrow{p} E\left\{\int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) (Y_{ij} - t)}{\hat{G}(Y_{ij})} [\mathbf{X}_{ij} - \bar{\mathbf{x}}(t; \beta_0, \mathbf{v}_0)] d\tilde{N}(t)\right\} \\ &= E\left\{E\left\{\int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) (Y_{ij} - t)}{\hat{G}(Y_{ij})} [\mathbf{X}_{ij} - \bar{\mathbf{x}}(t; \beta_0, \mathbf{v}_0)] d\tilde{N}(t) \middle| \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}_0\right\}\right\} \\ &= E\left\{\int_0^\infty \int_0^\tau \frac{\delta_{ij} I(y > t) (y - t)}{G(y)} [\mathbf{X}_{ij} - \bar{\mathbf{x}}(t; \beta_0, \mathbf{v}_0)] d\tilde{N}(t) f_Y(y|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}_0) dy\right\} \end{aligned}$$

$$\begin{aligned}
&= E \left\{ \int_0^\tau \int_0^\infty \frac{\delta_{ij} I(y > t)(y - t)}{G(y)} [f_T(y | \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}_0) G(y) \right. \\
&\quad \left. + S(y | \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}_0) f_C(y)] dy \times [\mathbf{X}_{ij} - \bar{\mathbf{x}}(t; \boldsymbol{\beta}_0, \mathbf{v}_0)] d\tilde{N}(t) \right\} \\
&= E \left\{ \int_0^\tau \int_t^\infty (y - t) f_T(y | \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}_0) dy [\mathbf{X}_{ij} - \bar{\mathbf{x}}(t; \boldsymbol{\beta}_0, \mathbf{v}_0)] d\tilde{N}(t) \right\} \\
&= E \left\{ \int_0^\tau m(t | \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}_0) [\mathbf{X}_{ij} - \bar{\mathbf{x}}(t; \boldsymbol{\beta}_0, \mathbf{v}_0)] d\tilde{N}(t) \right\}. \tag{3.8}
\end{aligned}$$

Similarly, it can be derived that

$$\begin{aligned}
&\frac{1}{n} [\mathbf{U}(\boldsymbol{\beta}_0 | \hat{\mathbf{v}}) - \mathbf{U}(\boldsymbol{\beta}_0 | \mathbf{v}_0)] \\
&= \frac{1}{n} \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{X}_{ij}}{\hat{G}(Y_{ij})} [\hat{m}(t; \boldsymbol{\beta}_0, \mathbf{v}_0) - \hat{m}(t; \boldsymbol{\beta}_0, \hat{\mathbf{v}})] dN(t) \\
&= \frac{1}{n} \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t)(Y_{ij} - t)}{\hat{G}(Y_{ij})} [\bar{\mathbf{X}}(t; \boldsymbol{\beta}_0, \mathbf{v}_0) - \bar{\mathbf{X}}(t; \boldsymbol{\beta}_0, \hat{\mathbf{v}})] dN(t) \\
&\xrightarrow{p} E \left\{ \int_0^\tau [m(t | \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}_0) \bar{\mathbf{x}}(t; \boldsymbol{\beta}_0, \mathbf{v}_0) - m(t | \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{v}) \bar{\mathbf{x}}(t; \boldsymbol{\beta}_0, \mathbf{v})] d\tilde{N}(t) \right\}. \tag{3.9}
\end{aligned}$$

For any $\boldsymbol{\beta} \in \mathcal{B}_n(c, \boldsymbol{\beta}_0)$, it follows from the regularity conditions (C4), (C6), (C7), and (3.8)-(3.9) that

$$(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{U}(\boldsymbol{\beta}_0 | \mathbf{v}_0) = \mathcal{O}_p\left(\frac{c}{\sqrt{n}} \cdot \sqrt{n}\right) = \mathcal{O}_p(c), \tag{3.10}$$

$$(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T [\mathbf{U}(\boldsymbol{\beta}_0 | \hat{\mathbf{v}}) - \mathbf{U}(\boldsymbol{\beta}_0 | \mathbf{v}_0)] = \mathcal{O}_p\left(\frac{c}{\sqrt{n}} \cdot \sqrt{n}\right) = \mathcal{O}_p(c), \tag{3.11}$$

and

$$n(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \hat{\mathbf{A}}_{bb}(\boldsymbol{\beta}_0, \hat{\mathbf{v}})(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = \mathcal{O}_p\left(n \cdot \frac{c^2}{n}\right) = \mathcal{O}_p(c^2). \tag{3.12}$$

Combining (3.10)-(3.12) with (3.7), we then have $(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{U}(\boldsymbol{\beta} | \hat{\mathbf{v}}) < 0$ that is because $\mathcal{O}_p(c^2)$ dominates $\mathcal{O}_p(c)$ uniformly for a sufficiently large c . This shows the truth of the inequality (3.6).

Since the penalized quasi score $\mathbf{U}(\boldsymbol{\beta} | \hat{\mathbf{v}})$ is continuous with respect to $\boldsymbol{\beta}$, by theorem (6.3.4) of Ortega and Rheinboldt (1970), there exists a solution $\hat{\boldsymbol{\beta}} \in \mathcal{B}_n(c, \boldsymbol{\beta}_0)$ such that $\mathbf{U}(\hat{\boldsymbol{\beta}} | \hat{\mathbf{v}}) = \mathbf{0}$, that is, $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$ -consistent to $\boldsymbol{\beta}_0$.

The proof of Theorem 3.1 is completed. $\square$

Proof of Theorem 3.2. Since $\hat{\mathbf{A}}_{bb}(\boldsymbol{\beta}, \mathbf{v}_0)$ converges almost surely to $\mathbf{A}_{bb}(\boldsymbol{\beta}, \mathbf{v}_0)$ uniformly in $\boldsymbol{\beta}$, together with the boundedness of $\mathbf{A}_{bb}$ in condition (C4) and the continuity of $\mathbf{A}_{bb}(\boldsymbol{\beta}, \mathbf{v}_0)$, it implies that $\hat{\mathbf{A}}_{bb}(\boldsymbol{\beta}, \mathbf{v}_0)$ is nonsingular and uniformly bounded on $\mathcal{B}_n(c, \boldsymbol{\beta}_0)$ when n is large enough.

By the Taylor expansion of $\mathbf{U}(\hat{\boldsymbol{\beta}}|\mathbf{v}_0)$ at $\hat{\boldsymbol{\beta}}^0$:

$$\begin{aligned}\mathbf{U}(\hat{\boldsymbol{\beta}}|\mathbf{v}_0) &= \mathbf{U}(\hat{\boldsymbol{\beta}}^0|\mathbf{v}_0) - n\hat{\mathbf{A}}_{bb}(\tilde{\boldsymbol{\beta}}_0^*, \mathbf{v}_0)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^0) \\ &= -n\hat{\mathbf{A}}_{bb}(\tilde{\boldsymbol{\beta}}_0^*, \mathbf{v}_0)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^0),\end{aligned}$$

where $\tilde{\boldsymbol{\beta}}_0^*$ lies between $\hat{\boldsymbol{\beta}}^0$ and $\hat{\boldsymbol{\beta}}$. It is easy to obtain that

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^0) = -\hat{\mathbf{A}}_{bb}^{-1}(\tilde{\boldsymbol{\beta}}_0^*, \mathbf{v}_0)n^{-1/2}\mathbf{U}(\hat{\boldsymbol{\beta}}|\mathbf{v}_0).$$

Similarly, it follows from the Taylor expansion of $\mathbf{U}(\boldsymbol{\beta}_0|\mathbf{v}_0)$ at $\hat{\boldsymbol{\beta}}^0$ that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^0 - \boldsymbol{\beta}_0) = \hat{\mathbf{A}}_{bb}^{-1}(\tilde{\boldsymbol{\beta}}_0^{**}, \mathbf{v}_0)n^{-1/2}\mathbf{U}(\boldsymbol{\beta}_0|\mathbf{v}_0),$$

where $\tilde{\boldsymbol{\beta}}_0^{**}$ lies between $\hat{\boldsymbol{\beta}}^0$ and $\boldsymbol{\beta}_0$.

Therefore,

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= \sqrt{n}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^0) + \sqrt{n}(\hat{\boldsymbol{\beta}}^0 - \boldsymbol{\beta}_0) \\ &= -\hat{\mathbf{A}}_{bb}^{-1}(\tilde{\boldsymbol{\beta}}_0^*, \mathbf{v}_0)n^{-1/2}\mathbf{U}(\hat{\boldsymbol{\beta}}|\mathbf{v}_0) + \hat{\mathbf{A}}_{bb}^{-1}(\tilde{\boldsymbol{\beta}}_0^{**}, \mathbf{v}_0)n^{-1/2}\mathbf{U}(\boldsymbol{\beta}_0|\mathbf{v}_0).\end{aligned}\quad (3.13)$$

Both $\hat{\mathbf{A}}_{bb}^{-1}(\tilde{\boldsymbol{\beta}}_0^*, \mathbf{v}_0)$ and $\hat{\mathbf{A}}_{bb}^{-1}(\tilde{\boldsymbol{\beta}}_0^{**}, \mathbf{v}_0)$ converge almost surely to $\mathbf{A}_{bb}^{-1}$. Using the penalized quasi score given in (3.4), we have

$$\begin{aligned}& \frac{1}{n}[\mathbf{U}(\boldsymbol{\beta}_0|\mathbf{v}_0) - \mathbf{U}(\hat{\boldsymbol{\beta}}|\mathbf{v}_0)] \\ &= \frac{1}{n} \left\{ \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{X}_{ij}}{\hat{G}(Y_{ij})} [(Y_{ij} - t) - \hat{m}(t)] dN(t) \right. \\ & \quad \left. - \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{X}_{ij}}{\hat{G}(Y_{ij})} [(Y_{ij} - t) - \hat{m}(t; \hat{\boldsymbol{\beta}}, \mathbf{v}_0)] dN(t) \right\} \\ &= \frac{1}{n} \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t) \mathbf{X}_{ij}}{\hat{G}(Y_{ij})} [\hat{m}(t; \hat{\boldsymbol{\beta}}, \mathbf{v}_0) - \hat{m}(t)] dN(t)\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{ij} \int_0^\tau \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t) [\bar{\mathbf{X}}(t; \hat{\boldsymbol{\beta}}, \mathbf{v}_0) - \bar{\mathbf{X}}(t; \boldsymbol{\beta}_0, \mathbf{v}_0)] dN(t) \\
&= \int_0^\tau \left[\frac{1}{n} \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t) \right] \cdot [\bar{\mathbf{X}}(t; \hat{\boldsymbol{\beta}}, \mathbf{v}_0) - \bar{\mathbf{X}}(t; \boldsymbol{\beta}_0, \mathbf{v}_0)] dN(t). \quad (3.14)
\end{aligned}$$

Using the consistency of $\hat{G}$ and the law of large numbers, it can be easily derived that

$$\frac{1}{n} \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t) \xrightarrow{p} E \left[\frac{\delta_{ij} I(Y_{ij} > t)}{G(Y_{ij})} (Y_{ij} - t) \right].$$

By the definition of $\bar{\mathbf{X}}(t; \boldsymbol{\beta}, \mathbf{v})$, we have

$$\bar{\mathbf{X}}(t; \hat{\boldsymbol{\beta}}, \mathbf{v}_0) = \frac{1}{n} \sum_{ij} \frac{\frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \mathbf{X}_{ij} \exp(\mathbf{X}_{ij}^T \hat{\boldsymbol{\beta}} + \mathbf{Z}_{ij}^T \mathbf{v}_0)}{\frac{1}{n} \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \exp(\mathbf{X}_{ij}^T \hat{\boldsymbol{\beta}} + \mathbf{Z}_{ij}^T \mathbf{v}_0)}. \quad (3.15)$$

From the consistency of $\hat{\boldsymbol{\beta}}$, the numerator of each term in the right-hand side of Equation (3.15) can be written as

$$\frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \mathbf{X}_{ij} \exp(\mathbf{X}_{ij}^T \hat{\boldsymbol{\beta}} + \mathbf{Z}_{ij}^T \mathbf{v}_0) = \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \mathbf{X}_{ij} \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0 + \mathbf{Z}_{ij}^T \mathbf{v}_0) + o_p(1).$$

Note that $\frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \mathbf{X}_{ij} \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0 + \mathbf{Z}_{ij}^T \mathbf{v}_0)$ ($i = 1, \dots, m, j = 1, \dots, q_i$) form i.i.d. random variables because each of them is a continuous function of $\mathbf{X}_{ij}$ and $\mathbf{X}_{ij}$ ($i = 1, \dots, m, j = 1, \dots, q_i$) are i.i.d. random variables.

For the denominator of the right-hand side of Equation (3.15), by the law of large numbers and the fact that $\hat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}_0$ in probability, we have

$$\frac{1}{n} \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \exp(\mathbf{X}_{ij}^T \hat{\boldsymbol{\beta}} + \mathbf{Z}_{ij}^T \mathbf{v}_0) \xrightarrow{p} E \left[\frac{\delta_{ij} I(Y_{ij} > t)}{G(Y_{ij})} \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0 + \mathbf{Z}_{ij}^T \mathbf{v}_0) \right].$$

In addition, the integral

$$\int_0^\tau \frac{E \left[\frac{\delta_{ij} I(Y_{ij} > t)}{G(Y_{ij})} (Y_{ij} - t) \right]}{E \left[\frac{\delta_{ij} I(Y_{ij} > t)}{G(Y_{ij})} \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0 + \mathbf{Z}_{ij}^T \mathbf{v}_0) \right]} \cdot \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} dN(t)$$

is bounded because $\mathbf{X}$ and $\mathbf{Z}$ are bounded from condition (C2). Using the central limit theorem and Slutsky's theorem, we can derive immediately that

$$\sqrt{n} \int_0^\tau \left[\frac{1}{n} \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t) \right] \bar{\mathbf{X}}(t; \hat{\boldsymbol{\beta}}, \mathbf{v}_0) dN(t)$$

$$= \frac{1}{\sqrt{n}} \sum_{ij} \left[\int_0^\tau \frac{\frac{1}{n} \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t)}{\frac{1}{n} \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \exp(\mathbf{X}_{ij}^T \hat{\boldsymbol{\beta}} + \mathbf{Z}_{ij}^T \mathbf{v}_0)} \cdot \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} dN(t) \right] \\ \cdot \mathbf{X}_{ij} \exp(\mathbf{X}_{ij}^T \hat{\boldsymbol{\beta}} + \mathbf{Z}_{ij}^T \mathbf{v}_0)$$

is asymptotically normally distributed.

Similarly, $\bar{\mathbf{X}}(t; \boldsymbol{\beta}_0, \mathbf{v}_0)$ can be written as

$$\bar{\mathbf{X}}(t; \boldsymbol{\beta}_0, \mathbf{v}_0) = \frac{1}{n} \sum_{ij} \frac{\frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \mathbf{X}_{ij} \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0 + \mathbf{Z}_{ij}^T \mathbf{v}_0)}{\frac{1}{n} \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0 + \mathbf{Z}_{ij}^T \mathbf{v}_0)},$$

where

$$\frac{1}{n} \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0 + \mathbf{Z}_{ij}^T \mathbf{v}_0) \xrightarrow{p} E \left[\frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}_0 + \mathbf{Z}_{ij}^T \mathbf{v}_0) \right].$$

Thus, the asymptotic normality of $\sqrt{n} \int_0^\tau \left[\frac{1}{n} \sum_{ij} \frac{\delta_{ij} I(Y_{ij} > t)}{\hat{G}(Y_{ij})} (Y_{ij} - t) \right] \bar{\mathbf{X}}(t; \boldsymbol{\beta}_0, \mathbf{v}_0) dN(t)$ can be obtained likewise by the central limit theorem and Slutsky's theorem.

Therefore, there exists a matrix $\tilde{\boldsymbol{\Sigma}}$ such that $n^{-1/2} [\mathbf{U}(\boldsymbol{\beta}_0 | \mathbf{v}_0) - \mathbf{U}(\hat{\boldsymbol{\beta}} | \mathbf{v}_0)] \xrightarrow{d} N(\mathbf{0}, \tilde{\boldsymbol{\Sigma}})$. Assume that the limit of the sequence $\{Cov(n^{-1/2} [\mathbf{U}(\boldsymbol{\beta}_0 | \mathbf{v}_0) - \mathbf{U}(\hat{\boldsymbol{\beta}} | \mathbf{v}_0)])\}$ exists, and then $\tilde{\boldsymbol{\Sigma}}$ can be determined by

$$\begin{aligned} \tilde{\boldsymbol{\Sigma}} &= \lim_{n \rightarrow \infty} Cov(n^{-1/2} [\mathbf{U}(\boldsymbol{\beta}_0 | \mathbf{v}_0) - \mathbf{U}(\hat{\boldsymbol{\beta}} | \mathbf{v}_0)]) \\ &= \lim_{n \rightarrow \infty} E \{ n^{-1} [\mathbf{U}(\boldsymbol{\beta}_0 | \mathbf{v}_0) - \mathbf{U}(\hat{\boldsymbol{\beta}} | \mathbf{v}_0)] [\mathbf{U}(\boldsymbol{\beta}_0 | \mathbf{v}_0) - \mathbf{U}(\hat{\boldsymbol{\beta}} | \mathbf{v}_0)]^T \}. \end{aligned}$$

Thus, the asymptotic covariance of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ can be obtained by

$$\boldsymbol{\Omega} = \mathbf{A}_{bb}^{-1} \tilde{\boldsymbol{\Sigma}} \mathbf{A}_{bb}^{-1}. \quad (3.16)$$

Similar to the estimation of $\boldsymbol{\Omega}_0$ given in the remark 2 of Lemma 3.1, it follows from (3.16) that $\boldsymbol{\Omega}$ can be estimated by $\hat{\mathbf{A}}_{bb}^{-1}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}}) \boldsymbol{\Sigma} \hat{\mathbf{A}}_{bb}^{-1}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}})$, where $\boldsymbol{\Sigma} = \frac{1}{n} \sum_{ij} \hat{\boldsymbol{\xi}}_{ij}^{\otimes 2}$.

Next, we consider the case when the estimator of random effects $\hat{\mathbf{v}}$ is $\sqrt{n}$ -consistent in norm. Based on the Taylor expansion of $\mathbf{U}(\hat{\boldsymbol{\beta}} | \hat{\mathbf{v}})$ at $\mathbf{v}_0$ and condition (C5), we have

$$n^{-1/2} \mathbf{U}(\hat{\boldsymbol{\beta}} | \mathbf{v}_0) = \sqrt{n} \hat{\mathbf{A}}_{bv}(\hat{\boldsymbol{\beta}}, \mathbf{v}_0)(\hat{\mathbf{v}} - \mathbf{v}_0) + o_p(1)$$

$$= \mathcal{O}_p(\sqrt{n} \cdot \frac{1}{\sqrt{n}}) \cdot o_p(1) + o_p(1) = o_p(1).$$

Then Equation (3.13) can be rewritten as

$$\sqrt{n}(\hat{\beta} - \beta_0) = \sqrt{n}(\hat{\beta}^0 - \beta_0) + o_p(1).$$

It follows from Lemma 3.1(ii) that

$$\Omega = \Omega_0,$$

which completes the proof of Theorem 3.2. $\square$

3.5 Simulation Studies

We evaluate the finite sample performance of the proposed conditional inference method using simulation studies under different distributional assumptions for the frailty term. The performance of the proposed method is also compared with that of the conventional PMRL regression which ignores the potential dependence between individuals within the same cluster.

Data of survival outcomes are generated from the following PMRL regression model with random effects:

$$m(t|X_{ij}, \mathbf{v}) = m_0(t) \exp(X_{ij}\beta + \mathbf{Z}_{ij}^T \mathbf{v}), \quad (3.17)$$

where only random intercepts are considered, that is, $\mathbf{Z}_{ij}$ is an m -dimensional vector with the i -th element being 1 and all the other elements being 0. We set the number of clusters $m = 5, 10, 20$ and the size of each cluster $q = 50$. The baseline MRL function is taken as $m_0(t) = -0.05t + 0.1$. Same as in Section 2.2, we set $\beta = 2.0$ and generate covariate X_{ij} from a Bernoulli distribution with success probability 0.5. In the simulation, we consider two scenarios for random effects $\mathbf{v}$: i) v_i is generated from a normal distribution with mean 0 and variance α ; ii) $u_i = \exp(v_i)$ is generated from a gamma distribution with shape $1/\alpha$ and scale α . The variance component α of random effects or log random effects varies over

three different levels: $\alpha=0.0$, 0.25 or 0.5, representing a none, medium or strong association within clusters. Censoring times are generated independently from an exponential distribution with the rate parameter determined by the desired censoring rate of 10%, 20% or 30%. Finally, the tuning parameter in the penalized conditional quasi-likelihood function is chosen as $\gamma=\log(n)$ with the total sample size $n=m \times q$.

1,000 replications are carried out for each setting. Estimation results obtained from both the proposed frailty PMRL model and the conventional PMRL model are reported in Tables 3.1 and 3.2 for scenarios i) and ii), respectively, in terms of the root of mean square errors (RMSE), sample standard deviation (SD), average of estimated standard errors (SE) and empirical coverage probability in percentage (CP%) of 95% confidence intervals, where the SE of the estimated fixed effects can be computed based on the approximated asymptotic variance $\hat{\mathbf{A}}_{bb}^{-1}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}})\boldsymbol{\Sigma}\hat{\mathbf{A}}_{bb}^{-1}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}})$ given in Theorem 3.2.

Table 3.1 shows that both of the proposed and the conventional methods perform similarly well for the estimation of fixed effects β when there is no cluster-specific correlation among data because both methods lead to small RMSEs, SEs, SDs and reasonable CPs in all simulation cases. As the correlation among clusters increases to medium, that is, $\alpha = 0.25$, compared with the conventional PMRL model, all the RMSEs, SDs and SEs achieved from the proposed frailty PMRL model decrease and CPs are more reasonable, which suggest a better performance of our method. When the influence of random effects becomes even stronger, says $\alpha=0.5$, all evaluation indexes for the estimates of β based on the proposed method outperform those based on the conventional method as expected. For all three levels of α , the proposed method shows a good performance in estimating α with small RMSEs and SDs for the estimates of α . Comparing the estimates of β from the proposed frailty PMRL model in different cases in Table 3.1, we can see, unsurprisingly, that the RMSEs, SDs and SEs of the estimated β increase as the variance component of random effects α increases, that is, the variation of

data becomes greater. It can also be found that these evaluation indexes of the estimated β decrease as the number of clusters increases under a fixed censoring rate. When the number of clusters is given, the RMSEs, SDs and SEs of the estimated β increase along with the increasing of the censoring rate. Table 3.1 also shows that SEs of the estimated β get closer to the corresponding SDs as the number of clusters becomes larger, indicating that the sandwich estimated covariance of the estimated fixed effects works well in the simulation.

In scenario ii), the frailty (log random effects) is assumed to be gamma distributed. Similar to Table 3.1, Table 3.2 reveals that the estimated β based on the proposed PMRL model performs competitively or better than that based on the conventional PMRL model in terms of all presented indexes. When the variance component α varies from 0.0 to 0.5, the results suggest that the clustered structure can not be ignored as the association within clusters becomes stronger. Under both situations (decreased censoring rate with a fixed number of clusters and increased number of clusters with a given censoring rate), all RMSEs and SDs of the proposed estimated β decrease, the differences between SDs and SEs become smaller as the sample size increases, while those of estimates from the conventional method do not have such favorable performance. In addition, the proposed method also exhibits more reasonable CPs of the estimates of fixed effects than the conventional method.

Tables 3.1 and 3.2 indicate the efficiency of the proposed conditional inference method in the analysis of clustered survival data. The simulation results of the proposed estimated β are consistent with the theoretical results that we have established in Theorems 3.1 and 3.2.

Table 3.1. Simulation results using penalized conditional quasi-likelihood method based on frailty and conventional PMRL models: lognormal frailty with variance component $\alpha = 0.0, 0.25, 0.5$.

		Frailty PMRL				Conventional PMRL					
		α		β		β					
m	C%	RMSE	SD	RMSE	SD	SE	CP%	RMSE	SD	SE	CP%
$\alpha=0.0$											
5	10	0.0296	0.0178	0.1254	0.1254	0.1196	94.7	0.1241	0.1241	0.1192	94.4
5	20	0.0499	0.0301	0.1383	0.1376	0.1287	95.3	0.1342	0.1342	0.1284	94.4
5	30	0.1002	0.0580	0.1671	0.1600	0.1432	93.0	0.1539	0.1536	0.1442	94.3
10	10	0.0264	0.0121	0.0867	0.0867	0.0855	94.5	0.0860	0.0859	0.0852	94.2
10	20	0.0464	0.0215	0.0987	0.0980	0.0922	95.0	0.0941	0.0941	0.0918	95.0
10	30	0.0891	0.0377	0.1228	0.1132	0.1031	92.9	0.1061	0.1061	0.1037	95.2
20	10	0.0256	0.0086	0.0620	0.0620	0.0607	94.9	0.0616	0.0615	0.0605	94.9
20	20	0.0437	0.0154	0.0683	0.0673	0.0656	94.1	0.0658	0.0658	0.0653	94.7
20	30	0.0880	0.0289	0.0909	0.0797	0.0735	91.1	0.0756	0.0756	0.0737	95.7
$\alpha=0.25$											
5	10	0.1912	0.1893	0.1275	0.1275	0.1374	94.9	0.2189	0.1930	0.1394	91.8
5	20	0.1985	0.1940	0.1500	0.1499	0.1502	94.3	0.2366	0.2134	0.1524	91.3
5	30	0.2108	0.2015	0.1796	0.1738	0.1712	93.3	0.2585	0.2321	0.1747	92.1
10	10	0.1330	0.1301	0.0942	0.0942	0.1007	94.7	0.2137	0.1688	0.1040	89.7
10	20	0.1410	0.1347	0.1062	0.1057	0.1108	94.8	0.2149	0.1736	0.1148	89.4
10	30	0.1562	0.1409	0.1369	0.1297	0.1266	93.3	0.2327	0.1935	0.1324	89.9
20	10	0.0932	0.0893	0.0644	0.0643	0.0733	94.4	0.1893	0.1206	0.0767	80.8
20	20	0.0992	0.0916	0.0726	0.0716	0.0809	94.8	0.1877	0.1220	0.0850	81.0
20	30	0.1162	0.0972	0.0976	0.0869	0.0934	91.7	0.1938	0.1342	0.0989	84.9

(Continued on following page)

Table 3.1 (Continued)

		Frailty PMRL						Conventional PMRL					
		α			β			β					
m	C%	RMSE	SD	RMSE	SD	SE	CP%	RMSE	SD	SE	CP%		
$\alpha=0.5$													
5	10	0.3649	0.3640	0.1350	0.1350	0.1982	94.5	0.4460	0.3455	0.1528	85.7		
5	20	0.3614	0.3596	0.1554	0.1552	0.2191	95.6	0.4509	0.3578	0.1706	87.2		
5	30	0.3693	0.3662	0.2052	0.1947	0.2616	94.2	0.4999	0.3957	0.2040	87.0		
10	10	0.2453	0.2436	0.0957	0.0956	0.1402	95.4	0.4327	0.2864	0.1167	81.3		
10	20	0.2507	0.2475	0.1092	0.1088	0.1583	94.8	0.4253	0.2874	0.1333	81.8		
10	30	0.2526	0.2464	0.1552	0.1430	0.1879	94.1	0.4500	0.3128	0.1623	83.3		
20	10	0.1703	0.1686	0.0680	0.0679	0.1007	94.1	0.4036	0.2056	0.0878	63.0		
20	20	0.1738	0.1703	0.0782	0.0770	0.1152	94.9	0.4048	0.2119	0.1019	66.2		
20	30	0.1783	0.1709	0.1125	0.0984	0.1385	89.8	0.4007	0.2225	0.1262	71.8		

C% represents censoring rate in percentage; RMSE, SD, SE and CP% represent root of mean square error, sample standard deviation, standard errors and empirical coverage probability in percentage for estimates, respectively.

Table 3.2. Simulation results using penalized conditional quasi-likelihood method based on frailty and conventional PMRL models: gamma frailty with variance component $\alpha=0.0, 0.25, 0.5$.

Frailty PMRL								Conventional PMRL			
m	C%	α		β			β				
		RMSE	SD	RMSE	SD	SE	CP%	RMSE	SD	SE	CP%
$\alpha=0.0$											
5	10	0.0288	0.0171	0.1246	0.1246	0.1194	94.6	0.1236	0.1236	0.1190	94.6
5	20	0.0486	0.0298	0.1374	0.1368	0.1282	95.2	0.1336	0.1336	0.1279	94.8
5	30	0.0983	0.0573	0.1658	0.1594	0.1424	93.0	0.1528	0.1526	0.1434	94.2
10	10	0.0259	0.0118	0.0867	0.0867	0.0853	94.4	0.0861	0.0860	0.0850	94.6
10	20	0.0449	0.0206	0.0978	0.0972	0.0918	94.5	0.0937	0.0936	0.0914	95.1
10	30	0.0877	0.0376	0.1213	0.1120	0.1027	93.1	0.1060	0.1059	0.1032	95.7
20	10	0.0253	0.0084	0.0620	0.0620	0.0606	95.0	0.0616	0.0615	0.0603	95.1
20	20	0.0424	0.0150	0.0675	0.0666	0.0653	94.0	0.0653	0.0653	0.0650	95.5
20	30	0.0859	0.0283	0.0897	0.0791	0.0732	90.9	0.0750	0.0750	0.0734	95.6
$\alpha=0.25$											
5	10	0.2400	0.2300	0.1325	0.1324	0.1304	94.3	0.1913	0.1715	0.1411	92.1
5	20	0.2490	0.2346	0.1456	0.1455	0.1405	95.2	0.1982	0.1807	0.1530	93.1
5	30	0.2693	0.2471	0.1958	0.1886	0.1623	94.3	0.2485	0.2252	0.1765	93.3
10	10	0.1783	0.1642	0.0956	0.0956	0.0951	94.7	0.1623	0.1261	0.1051	87.8
10	20	0.1868	0.1677	0.1066	0.1063	0.1040	95.3	0.1690	0.1375	0.1148	89.7
10	30	0.1999	0.1692	0.1434	0.1345	0.1191	94.0	0.1940	0.1626	0.1334	91.3
20	10	0.1377	0.1176	0.0689	0.0688	0.0687	95.5	0.1425	0.0925	0.0763	78.5
20	20	0.1463	0.1189	0.0798	0.0788	0.0758	94.0	0.1481	0.1029	0.0843	82.8
20	30	0.1595	0.1171	0.1066	0.0943	0.0880	93.0	0.1555	0.1140	0.0985	86.5

(Continued on following page)

Table 3.2 (Continued)

		Frailty PMRL						Conventional PMRL					
		α			β			β					
m	C%	RMSE	SD	RMSE	SD	SE	CP%	RMSE	SD	SE	CP%		
$\alpha=0.5$													
5	10	0.5504	0.5233	0.1407	0.1401	0.1587	94.9	0.3097	0.2583	0.1549	91.3		
5	20	0.5503	0.5214	0.1610	0.1610	0.1746	95.3	0.3232	0.2706	0.1724	90.8		
5	30	0.5373	0.5080	0.2386	0.2245	0.2077	93.4	0.4074	0.3419	0.2072	91.2		
10	10	0.3684	0.3366	0.0998	0.0998	0.1145	94.6	0.2820	0.1947	0.1171	83.8		
10	20	0.3691	0.3338	0.1172	0.1163	0.1272	94.8	0.2865	0.2064	0.1315	86.2		
10	30	0.3675	0.3294	0.1761	0.1583	0.1553	93.3	0.3171	0.2385	0.1636	86.6		
20	10	0.3105	0.2590	0.0729	0.0728	0.0818	94.8	0.2633	0.1420	0.0863	70.4		
20	20	0.3136	0.2570	0.0849	0.0837	0.0925	94.0	0.2644	0.1544	0.0981	75.6		
20	30	0.3095	0.2512	0.1408	0.1146	0.1148	88.2	0.2825	0.1736	0.1244	79.0		

C% represents censoring rate in percentage; RMSE, SD, SE and CP% represent root of mean square error, sample standard deviation, standard errors and empirical coverage probability in percentage for estimates, respectively.

3.6 Application to Breast Cancer Data

For indicative purposes, we apply the proposed conditional inference method to analyze clustered data on distant metastasis-free survival of breast cancer, where patients are from an international multi-center study (Haibe-Kains et al., 2012). The distant metastasis is a severe condition without curative treatment for breast cancer patients, it refers to that cancer has been spread from the breast or auxiliary lymph nodes to distant organs or distant lymph nodes. It is an important signal for doctors to diagnose which stage a patient is in and closely associated to deaths from breast cancer. There are probably 20 ~ 30% of breast cancer patients in the early stage deteriorated distant metastasis (Song et al., 2017). Time to distant metastasis is also known as distant metastasis-free survival, which has a significant impact on breast cancer mortality. As thus, the study of the distant metastasis-free survival (DMFS) defined by the duration between the diagnosis of breast cancer and distant metastasis is important for more accurately predict the risk of tumor dissemination in breast cancer research.

In the databases used in Haibe-Kains et al. (2012), the DMFS time recorded in the breast cancer data is measured in days. We denote it as “t.dmfs” after dividing by 30 for the easy interpretation in this study. The DMFS time is censored if there is no distant metastasis with the corresponding censored indicator “e.dmfs” being 0, else $e.dmfs = 1$. The development of distant metastasis is related to different prognostic features of cancer. We aim to undertake a comprehensive analysis of how the prognostic factors and cluster-specific association affect the survival time of a patient without breast cancer recurrence after distant metastasis was diagnosed. These factors include size and grade which have been defined in Section 2.3, estrogen receptor denoted by “er” which is used to determine sensitivity of breast cancer lesions to the therapy ($er = 1$ denotes the therapy can prevent the expression of estrogen receptors and 0 means can not), and the nodal status “node” which describes that breast cancer has spread to the lymph nodes or not ($node = 1$ denotes spread and 0 means not yet).

Fifteen out of thirty-six breast cancer centers contain patients with complete information about the above four clinical prognostic factors and valid observations of the DMFS time are extracted from the original databases. There are total 2,668 patients included, where 109 out of them from CAL dataset (University of California, San Francisco, and the California Pacific Medical Center, United States); 115 from DFHCC (Dana-Farber Harvard Cancer Center, United States); 200 from MAINZ (Mainz hospital, Germany); 99 from NCI (National Cancer Institute, United States); 319 from NKI (National Kanker Instituut, the Netherlands); 77 from PNC; 88 from STK (Stockholm, Karolinska University Hospital, Sweden); 507 from SUPERTAM_HGU133A; 128 from SUPERTAM_HGU133PLUS2; 196 from TRANSBIG (dataset collected by the TransBIG consortium, Europe); 120 from UCSF (University of California, San Francisco, United States); 189 from UNC (University of North Carolina, United States); 104 from UNT (cohort of untreated breast cancer patients from the Oxford Radcliffe, United Kingdom and Karolinska, Sweden hospitals); 220 from UPP (Uppsala Hospital, Sweden); 197 from VDX (Veridex, the Netherlands). These fifteen breast cancer centers are indexed as clusters 1-15, respectively, in the following analysis.

In the dataset considered, 71.5% of the DMFS times are right censored. The tumor size ranges from 0 to 11 centimeters with the average around 2.3 centimeters. Information about the other categorical variables are summarized in Table 3.3.

Before the application of the proposed method, we first check empirically if the required independence assumption between the censored time and covariates holds. By applying the R function *Surv* to observations *t.dmfs* and $1 - e.dmfs$, we create a survival object for censored times. The function *survfit* is then applied to the above survival object and each covariate in order to compute estimates of the Kaplan-Meier (KM) curves for censored times over different covariates. Plots of the KM curves over different covariate categories are presented in Figure

3.1. A two-level categorical variable ($\text{size} \leq 2$ and $\text{size} > 2$) is used to replace the continuous variable size. For each covariate, we can see that the KM curves across different levels are almost overlapped, providing clear evidence to support the independence assumption between the censored time and covariates.

We applied the proposed conditional inference method to fit the subset of the breast cancer data on DMFS times with the frailty PMRL regression model. Results are compared with those from the conventional PMRL model ignoring the potential dependence among data, the Cox proportional hazards (PH) model using the partial likelihood method implemented by employing R functions *coxph* and *frailty.gaussian* in package *survival*, and the frailty Cox PH model using the hierarchical likelihood method carried out through the function *frailtyHL* in R package *frailtyHL* (<https://cran.r-project.org/web/packages/frailtyHL/>). Both likelihood-based estimation methods for the frailty Cox PH model require the distributional assumption for the frailty term and the lognormal is employed in this study.

The fitting results are reported in Table 3.4. All four methods identify the covariate *er* as a significant factor at level of 0.05. The signs of estimated coefficients are different for the frailty Cox PH model and PMRL models with or without frailty, that's because the decreased hazard usually means an increase of the expected residual life to some extent. However, the effects of other three covariates are very different for hazard-based and MRL-based models in terms of statistical significance and sign of the estimation coefficient.

It shows that the factor size is significant under the frailty Cox PH model with likelihood estimation methods while nonsignificant under both the proposed and the conventional PMRL models. Only the estimated factor node from the proposed estimation method is nonsignificant and has an inconsistent sign compared with estimates from the other three methods. For the categorical variable grade, grade1 (grade = 1) is chosen as the baseline level. The significance of grade2 (grade = 2) in the proposed frailty PMRL model is the same as in the

existing hazard-based models but with different interpretations in term of sign. The factor grade3 (grade = 3) is significant under the frailty Cox PH model but nonsignificant in PMRL models. To explain these different effects, we visualize the nonparametric estimation of the cumulative hazard and MRL functions for each of these three covariates using plots of the logarithm of estimation curves in Figures 3.2-3.4.

Comparing the left and right panels of Figure 3.2, it can be found that the gap in hazards between two levels of the tumor size is clearly greater than that in MRLs, in accordance with the finding revealed in Table 3.4. Figure 3.3 is used to explain the different estimates of the factor node obtained from PMRL models and the hazard-based model. The corresponding picture on the right panel shows a big gap of the logarithm of MRL estimates for the two levels of node, which is consistent with the estimation of the regression coefficient under the conventional model based on PMRL but inconsistent with the result under the proposed frailty PMRL model. The right panel plot in Figure 3.4 provides strong evidences to support contrasting estimation results for effects of grade from different models in Table 3.4. Comparing with the low histological grade of breast cancer, the intermediate histological grade seems to increase the expectancy life of patients. Additional investigations are needed for such an unreasonable phenomenon.

First of all, the potential association between covariates has been checked by a stepwise procedure and the results are shown in Table 3.5, where steps 1-4 represent that four covariates (er, size, node and grade) are stepwise added into the proposed frailty PMRL regression model. Table 3.5 indicates that there is no strong association between covariates since the estimates in each step have no much difference.

Further, the logarithm of MRL is nonparametrically estimated for each cluster (center). Figures 3.5 and 3.6 present the corresponding curves for different

levels of factors, node and grade, over all 15 clusters. It is shown that clusters 4, 5, 7, 11, 12 in Figure 3.5 indicate that patients experienced distant metastasis have longer expected residual life times than those who did not; while the situation is quite the reverse for clusters 1, 2, 6, 8, 9, 14, where the expected residual life of patients without distant metastasis is longer; the rest clusters don't have patients with distant metastasis and thus are not comparable. Similarly, a support for the estimation result of grade2 under the proposed model in Table 3.4 can be found. Comparing with the low grade, Figure 3.6 shows that the intermediate grade has a positive effect on the mean residual life of patients in clusters 4, 5, 8, 10, 13, 14. All these provide a strong evidence for the estimation results from the proposed conditional inference method under the frailty PMRL model.

To evaluate the estimated baseline mean residual life function obtained from Equation (2.7), the estimated curve at distinct failure time points is presented in Figure 3.7-A. The linear function of survival time in the figure shows that the baseline mean residual is likely to follow the Hall-Wellner family. Both the curves of logarithm of the marginal mean residual life and the logarithm of the baseline mean residual life in Figure 3.7-B run parallel, indicating that the proportional mean residual life assumed in the model is reasonable.

The fitting results of the frailty Cox PH model using hierarchical likelihood method provided in Table 3.4 show that the frailty variance component α is significant. To check the significance of the estimated α from the proposed frailty PMRL model for DMFS times in this study, we visualize the dispersion of the predicted random effects in Figure 3.8 using 95% bootstrap confidence intervals of $\hat{v}_i$ ($i = 1, \dots, 15$) based on 1,000 bootstrap replications. In each bootstrap repeat, 15 clusters are randomly sampled with replacement from original clusters 1-15 followed by sampling of q_i survival outcomes in the corresponding i -th cluster without replacement, leading to a bootstrap data used to calculate confidence intervals, and finally the proposed conditional inference method is applied to make

inference about random effects based on this bootstrap data. For the cluster i , we denote the bootstrap estimates of v_i as $\hat{v}_i^b$ for $b = 1, \dots, 1,000$. Thereafter a 95% bootstrap confidence interval of $\hat{v}_i$ can be constructed based on the lower and upper 2.5% quantiles of a valid series of the bootstrap estimates, which excludes those absent entries $\hat{v}_i^b$ if cluster i is not selected in the b -th bootstrap repeat. It is clearly seen from the figure that many interval bars over different clusters do not overlap and most of them do not cover 0, thus the significance of $\hat{\alpha}$ is implicitly confirmed.

Table 3.3. Application to distant metastasis-free survival: covariate information for total 2,668 patients in the valid dataset, the number (percentage) of observations for different covariates over 15 clusters.

Total/ Cluster No.	er		node		grade		
	1(%)	0(%)	1(%)	0(%)	1(%)	2(%)	3(%)
Total	2,034(76.2)	634(23.8)	1,813(32.0)	855(68.0)	440(16.5)	1,079(40.4)	1,149(43.1)
1	71(65.1)	38(34.9)	62(56.9)	47(43.1)	10(9.2)	42(38.5)	57(52.3)
2	70(60.9)	45(39.1)	53(46.1)	62(53.9)	23(20.0)	28(24.3)	64(55.7)
3	162(81.0)	38(19.0)	0(0.0)	200(100)	29(14.5)	136(68.0)	35(17.5)
4	65(65.7)	34(34.3)	53(53.5)	46(46.5)	16(16.2)	38(38.4)	45(45.4)
5	245(76.8)	74(23.2)	144(45.1)	175(54.9)	78(24.4)	108(33.9)	133(41.7)
6	39(50.6)	38(49.4)	36(46.8)	41(53.2)	11(14.3)	3(3.9)	63(81.8)
7	65(73.9)	23(26.1)	61(69.3)	27(30.7)	10(11.4)	39(44.3)	39(44.3)
8	447(88.2)	60(11.8)	142(28.0)	365(72.0)	76(15.0)	222(43.8)	209(41.2)
9	128(100)	0(0.0)	74(57.8)	54(42.2)	31(24.2)	57(44.5)	40(31.3)
10	132(67.3)	64(32.7)	0(0.0)	196(100)	30(15.4)	83(42.3)	83(42.3)
11	82(68.3)	38(31.7)	66(55.0)	54(45.0)	9(7.5)	52(43.3)	59(49.2)
12	122(64.6)	67(35.4)	89(47.1)	100(52.9)	22(11.7)	66(34.9)	101(53.4)
13	72(69.2)	32(30.8)	0(0.0)	104(100)	30(28.8)	47(45.2)	27(26.0)
14	192(87.3)	28(12.7)	75(34.1)	145(65.9)	58(26.4)	116(52.7)	46(20.9)
15	142(72.1)	55(27.9)	0(0.0)	197(100)	7(3.6)	42(21.3)	148(75.1)

Table 3.4. Application to distant metastasis-free survival: parameter estimates based on the proposed frailty PMRL, the conventional PMRL and the frailty Cox PH models.

Parameter	Frailty PMRL		Conventional PMRL		Frailty Cox PH			
	Estimate	SE	Estimate	SE	Partial Likelihood		Hierarchical Likelihood	
					Estimate	SE	Estimate	SE
er	0.5389*	0.1150	0.7880*	0.1678	-0.3070*	0.0876	-0.2886*	0.0876
size	-0.0068	0.0324	-0.0576	0.0672	0.1524*	0.0272	0.1564*	0.0270
node	0.1038	0.2158	-0.7668*	0.1691	0.4542*	0.0900	0.4591*	0.0896
grade2	0.1844*	0.0809	0.3307	0.2011	0.6936*	0.1445	0.7466*	0.1475
grade3	-0.0321	0.1097	-0.2139	0.3006	1.0948*	0.1451	1.1531*	0.1480
α	0.4844		.		0.1518		0.1705*	0.0797

* indicates significant at 0.05 level.

Table 3.5. Application to distant metastasis-free survival: adding covariates stepwise.

Parameter	Step 1		Step 2		Step 3		Step 4	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
er	0.5962*	0.1164	0.5909*	0.1191	0.5834*	0.1217	0.5389*	0.1150
size	.		-0.0208	0.0367	-0.0257	0.0365	-0.0068	0.0324
node	.		.		0.1284	0.2247	0.1038	0.2158
grade2	.		.		.		0.1844*	0.0809
grade3	.		.		.		-0.0321	0.1097
α	0.4889		0.4838		0.5138		0.4844	

* indicates significant at 0.05 level.

Figure 3.1. Application to distant metastasis-free survival: survival curves based on the KM estimates for censoring times over different covariates.

Figure 3.2. Application to distant metastasis-free survival: curves of log cumulative hazards (left panel) and log mean residual life functions (right panel) over different classes of categorized size.

Figure 3.3. Application to distant metastasis-free survival: curves of log cumulative hazards (left panel) and log mean residual life functions (right panel) over different classes of node.

Figure 3.4. Application to distant metastasis-free survival: curves of log cumulative hazards (left panel) and log mean residual life functions (right panel) over different classes of grade.

Figure 3.5. Application to distant metastasis-free survival: curves of log mean residual life functions over different classes of node across 15 clusters.

Figure 3.6. Application to distant metastasis-free survival: curves of log mean residual life functions over different classes of grade across 15 clusters.

Figure 3.7. Application to distant metastasis-free survival: plots of the proposed estimated baseline mean residual life function (left panel) and the estimated log baseline and marginal mean residual life functions (right panel).

Figure 3.8. Application to distant metastasis-free survival: plot of 95% bootstrap confidence intervals for the predicted random effects across 15 clusters, where 1 denotes for CAL, 2 for DFHCC, 3 for MAINZ, 4 for NCI, 5 for NKI, 6 for PNC, 7 for STK, 8 for SUPERTAM_HGU133A; 9 for SUPERTAM_HGU133PLUS2, 10 for TRANSBIG, 11 for UCSF, 12 for UNC, 13 for UNT, 14 for UPP and 15 for VDX.

Conclusion and Future Work

In this chapter, we summarize the main work and contributions in the present thesis. Some extended topics along the line of mean residual life are discussed for further investigations.

4.1 Main Contributions

- (i) In Chapter 2, we have proposed a frailty proportional mean residual life (PMRL) regression model for the analysis of clustered survival data subject to right censoring collected from multi-center clinical trials using a novel hierarchical quasi-likelihood method for estimation of fixed effects, random effects and the frailty variance component.

This is the first attempt to incorporate the frailty into mean residual life models. Compared with the conventional PMRL model, the proposed frailty PMRL model is able to address potential correlations in clustered survival data. Unlike the frailty Cox PH model, the mean residual life used in the proposed model is much easier to be understood by nonstatisticians than the hazard function and can directly reflect a patient's remaining life or remission time (until he/she dies or the disease recurs) on average under the condition that he/she has survived or lived without disease recurrence

currently. The proposed frailty PMRL regression explicates the relationship between the expectation of remaining life and related covariates and is very important for researchers who are interested in how the prognostic factors and the within-cluster association affect the remaining life rather than a more abstract term, the relative risk of a patient.

By using the inverse probability of censoring weighting (IPCW) estimating equations and the idea of hierarchical likelihood, the proposed hierarchical quasi-likelihood method reduces the computational complexity involved in conventional estimation procedures based on the marginal likelihood or the standard hierarchical likelihood, allowing any parametric distribution assumption for random effects, not limited to lognormal and gamma distributions which are often assumed in the literature.

Favorable performances of the proposed hierarchical inference method have been shown in the numerical studies regardless of the frailty distribution. When the distribution of frailty is misspecified, simulation results exhibit the robust performance of the proposed estimates of fixed effects.

- (ii) Chapter 3 extends the frailty model proposed in Chapter 2 to a more flexible model framework for clustered survival data analysis. A penalized conditional quasi-likelihood estimation procedure is proposed for estimating both fixed and random effects in the model.

In this extended frailty PMRL model, random effects are not necessary imposed on the intercept and free of the distributional assumption. By applying conditional inference to the quasi-score estimating equations, a penalized conditional quasi-likelihood function has been proposed, where the resulting estimator of fixed effects possesses the consistency and the asymptotic normality properties under some regularity conditions.

The proposed conditional inference procedure can be implemented easily via an iterative algorithm. Both simulation studies and real data analysis

show the satisfactory performance of the proposed conditional inference approach and its practical utility.

- (iii) The penalty term considered in the penalized conditional quasi-likelihood function (3.3) has a similar form to the log-likelihood of $\boldsymbol{v}$ in the hierarchical quasi-likelihood function (2.9) under the lognormal frailty assumption. In the application, when the true distribution of the frailty is close to a lognormal, similar results would be expected by using either the conditional inference method or the hierarchical method. In general, the proposed conditional inference method is more appealing for non-lognormal frailty though it requires a larger sample size to compensate the loss of information caused by lacking of a distributional assumption.

4.2 Future Work

The frailty PMRL model has been proposed to describe the unobservable random effects for independent censoring, it can also be applied to investigate the dependent censoring cases, such as competing risks data (Beyersmann et al., 2009; Elashoff et al., 2007; Larson and Dinse, 1985) and semicompeting risks data (Chen, 2012; Fine et al., 2001; Lee et al., 2015; Xu et al., 2010). In fact, we have done some preliminary work on applying the proposed frailty PMRL model to the analysis of survival data subject to semicompeting risks, where each individual may experience both nonterminal (such as recurrence of a disease) and terminal (such as death) events and the nonterminal event could be censored by the terminal event but not vice versa.

Chen (2012) proposed marginal regression and joint association analyses based on semiparametric transformation models with a copula for semicompeting risks data. Usually the joint distribution of semicompeting risks data is nonparametrically identifiable limited in the upper wedge, to breakthrough the limitation, Fine et al. (2001) applied the plug-in technique for the estimation of the nonterminal

event's marginal distribution. Xu et al. (2010) imposed a shared frailty on the illness-death model and developed an inference procedure by using nonparametric maximum likelihood estimation. Based on the shared frailty model proposed by Xu et al. (2010), Lee et al. (2015) developed a Bayesian inference method to estimate regression parameters in the model, depict the dependent structure and make a prediction.

To the best of our knowledge, there is no research for the analysis of semicompeting risks data based on MRL or modelling the three stages of the illness-death model using the PMRL regression in the existing literature. Particularly, let T_1 , T_2 and C represent the nonterminal time, the terminal time and the potential censoring time for an individual, respectively, where T_1 can be censored by T_2 but not vice versa. Defined $Y_1 = \min(T_1, Y_2)$, $\delta_1 = I(T_1 \leq Y_2)$ and $Y_2 = \min(T_2, C)$, $\delta_2 = I(T_2 \leq C)$, where $I(\cdot)$ is indicator function. The illness-death model with transition functions defined by MRL can be written as:

$$m_1(t) = E(T_1 - t | T_2 \geq T_1 > t), \quad (4.1)$$

$$m_2(t) = E(T_2 - t | T_1 > T_2 > t), \quad (4.2)$$

$$m_3(t) = E(T_2 - t | T_2 > t \geq T_1). \quad (4.3)$$

Suppose that $\mathbf{X}$ is a p -dimensional vector of fixed covariates. Researchers and clinicians are interested in studying how the related covariates affect the expected remaining life of a patient in different transition stages. Let u be the shared frailty random variable between nonterminal and terminal events. Conditional on the shared frailty u and covariates $\mathbf{X}$, we assume that the transition MRL functions can be modelled by proportional MRL (PMRL) regression models, that is, the conditional PMRL functions of T_j ($j = 1, 2$) in different transitions are constructed in the form of

$$m_1(t | \mathbf{X}, u) = m_{01}(t) \exp(\mathbf{X}^T \boldsymbol{\beta}_1) u, \quad (4.4)$$

$$m_2(t | \mathbf{X}, u) = m_{02}(t) \exp(\mathbf{X}^T \boldsymbol{\beta}_2) u, \quad (4.5)$$

$$m_3(t | \mathbf{X}, u) = m_{03}(t) \exp(\mathbf{X}^T \boldsymbol{\beta}_3) u, \quad (4.6)$$

for all $t > 0$, where $m_{0j}(\cdot)$ is the unknown baseline MRL function and β_j is p -dimensional vector of regression parameters associated with fixed effects.

Along the similar line of the proposed methods for one event in this thesis, the inference procedure based on quasi-score estimating equations can be extended to estimate regression parameters $\beta_1, \beta_2, \beta_3$ and the shared frailty u in models (4.4)-(4.6). We are ongoing on this topic. Both numerical studies and theoretical justifications are under current investigation.

In the thesis, we have extended the conventional PMRL regression model to frailty models for the analysis of clustered survival data. Two novel methods based on the hierarchical quasi-likelihood and the penalized conditional quasi-likelihood are proposed for estimation. A main assumption in the PMRL regression model is that mean residual life functions are supposed to be proportional. It is tenable in many cases though may not be true sometimes and so various alternative MRL-based models have been developed in the literature. For instance, the additive MRL model studied by Chen and Cheng (2006) and Chen (2007); the transformation MRL models by Sun and Zhang (2009) and Chen (2012); and the additive-multiplicative MRL model developed by Mansourvar et al. (2016) and Cai et al. (2017). Analogous to our work on the PMRL model in the thesis, it is possible to extend the aforementioned MRL-based models to corresponding frailty models.

Prediction analysis of a patient's residual life expectancy after a specific treatment is often interesting for researchers and clinicians in practice. With both the baseline and post-baseline covariates, a dynamic prediction of the quantile residual life was studied by Li et al. (2016) based on a flexible quantile residual life regression model. Lin et al. (2018) extracted the major features of longitudinal biomarker data using functional principal component analysis for dynamically predicting the MRL in a transformed MRL regression model. When time-dependent covariates are considered in the proposed frailty PMRL model, it seems to be possible to make predictions about the expected residual lifetimes of

individuals nested within clusters. Topics along this line deserve more dedicated and in-depth investigation.

In addition, methods for assessing the adequacy of the proposed survival model is essential. Although methods for goodness-of-fit of the conventional Cox PH model and various frailty models have been well developed in the literature, see Lin (1991), Bagdonavičius and Nikulin (2000) and Hanagal (2019) for example, related studies for MRL-based models are limited due to more complicated features of the MRL function than the hazard. Without considering censoring in the data, Yuen et al. (2003) proposed a goodness-of-fit test to check the conventional PMRL model using an asymptotical Gaussian process, both the bootstrap and the random symmetrization methods had been explored to construct the rejection region of the test. Zhang et al. (2010) developed two test statistics to check the additive MRL model for right censored data based on stochastic processes and approximated the corresponding critical values by adopting resampling methods. Price and Jones (2017) constructed a Kolmogorov-Smirnov test statistic based on a general martingale residual. Incorporated with the weighted zero-mean stochastic processes constructed in this thesis, the idea of the general martingale residual used in Price and Jones (2017) can be applied to perform goodness-of-fit tests for the frailty PMRL model either theoretically or graphically by choosing an appropriate predictable process. It is definitely worthy of further study.

Another interesting topic is variable selection based on the hierarchical likelihood method. Considering three different penalty functions (LASSO, SCAD and HL), Ha et al. (2014) proposed the penalized hierarchical likelihood method for variable selection in general frailty models. Park and Ha (2019) applied the penalized hierarchical likelihood method to variable selection for AFT models. The extension of their methods to the frailty PMRL model by imposing an appropriate penalty on the proposed hierarchical quasi-likelihood function is worth consideration in the future.

Bibliography

- Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726.
- Andersen, P. K. and Gill, R. D. (1982). Cox’s regression model for counting processes: a large sample study. *The Annals of Statistics*, pages 1100–1120.
- Bagdonavičius, V. and Nikulin, M. (2000). On goodness-of-fit for the linear transformation and frailty models. *Statistics & Probability Letters*, 47(2):177–188.
- Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2(2):273–277.
- Beyersmann, J., Latouche, A., Buchholz, A., and Schumacher, M. (2009). Simulating competing risks data in survival analysis. *Statistics in Medicine*, 28(6):956–971.
- Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66(3):429–436.
- Cai, J., He, H., Song, X., and Sun, L. (2017). An additive-multiplicative mean residual life model for right-censored data. *Biometrical Journal*, 59(3):579–592.

- Chen, Y., Jewell, N., Lei, X., and Cheng, S. (2005). Semiparametric estimation of proportional mean residual life model in presence of censoring. *Biometrics*, 61(1):170–178.
- Chen, Y. H. (2012). Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. *Lifetime Data Analysis*, 18(1):36–57.
- Chen, Y. Q. (2007). Additive expectancy regression. *Journal of the American Statistical Association*, 102(477):153–166.
- Chen, Y. Q. and Cheng, S. (2005). Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika*, 92(1):19–29.
- Chen, Y. Q. and Cheng, S. (2006). Linear life expectancy regression with censored data. *Biometrika*, 93(2):303–313.
- Cho, H., Wang, P., and Qu, A. (2017). Personalize treatment for longitudinal data using unspecified random-effects model. *Statistica Sinica*, pages 187–206.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.
- Cox, D. R. and Oakes, D. (1984). *Analysis of survival data*, volume 21. CRC Press.
- Elashoff, R. M., Li, G., and Li, N. (2007). An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in Medicine*, 26(14):2813–2835.
- Field, C. A. and Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3):369–390.
- Fine, J. P., Jiang, H., and Chappell, R. (2001). On semi-competing risks data. *Biometrika*, 88(4):907–919.

- Fleming, T. R. and Harrington, D. P. (2011). *Counting processes and survival analysis*, volume 169. John Wiley & Sons.
- Ha, I. D., Jeong, J. H., and Lee, Y. (2017). *Statistical Modelling of Survival Data with Random Effects: H-likelihood Approach*. Springer.
- Ha, I. D. and Lee, Y. (2003). Estimating frailty models via poisson hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*, 12(3):663–681.
- Ha, I. D., Lee, Y., and Song, J. (2001). Hierarchical likelihood approach for frailty models. *Biometrika*, 88(1):233–243.
- Ha, I. D., Pan, J., Oh, S., and Lee, Y. (2014). Variable selection in general frailty models using penalized h-likelihood. *Journal of Computational and Graphical Statistics*, 23(4):1044–1060.
- Ha, I. D., Sylvester, R., Legrand, C., and MacKenzie, G. (2011). Frailty modelling for survival data from multi-centre clinical trials. *Statistics in Medicine*, 30(17):2144–2159.
- Haibe-Kains, B., Desmedt, C., Loi, S., Culhane, A. C., Bontempi, G., Quackenbush, J., and Sotiriou, C. (2012). A three-gene model to robustly identify breast cancer molecular subtypes. *Journal of the National Cancer Institute*, 104(4):311–325.
- Hanagal, D. D. (2019). Tests of hypotheses in frailty models. In *Modeling Survival Data Using Frailty Models*, pages 153–163. Springer.
- Huang, R., Xiang, L., and Ha, I. D. (2019). Frailty proportional mean residual life regression for clustered survival data: A hierarchical quasi-likelihood method. *Statistics in Medicine*, 38(24):4854–4870.
- Jiang, J. (1999). Conditional inference about generalized linear mixed models. *The Annals of Statistics*, 27(6):1974–2007.

- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.
- Klein, J. P. (1991). Small sample moments of some estimators of the variance of the kaplan-meier and nelson-aalen estimators. *Scandinavian Journal of Statistics*, pages 333–340.
- Klein, J. P. (1992). Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics*, pages 795–806.
- Larson, M. G. and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 34(3):201–211.
- Lee, K. H., Haneuse, S., Schrag, D., and Dominici, F. (2015). Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(2):253–273.
- Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(4):619–656.
- Lee, Y., Nelder, J. A., and Pawitanand, Y. (2017). *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*. Chapman and Hall/CRC, 2 edition.
- Li, R., Huang, X., and Cortes, J. (2016). Quantile residual life regression with longitudinal biomarker measurements for dynamic prediction. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(5):755–773.
- Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

- Lin, D. (1991). Goodness-of-fit analysis for the cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association*, 86(415):725–728.
- Lin, D., Ying, Z., et al. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The Annals of Statistics*, 23(5):1712–1734.
- Lin, X., Lu, T., Yan, F., Li, R., and Huang, X. (2018). Mean residual life regression with functional principal component analysis on longitudinal data for dynamic prediction. *Biometrics*, 74(4):1482–1491.
- Lin, Y. and Chen, K. (2012). Efficient estimation of the censored linear regression model. *Biometrika*, 100(2):525–530.
- Liu, L. and Xiang, L. (2019). Missing covariate data in generalized linear mixed models with distribution-free random effects. *Computational Statistics & Data Analysis*, 134:1–16.
- Maguluri, G. and Zhang, C.-H. (1994). Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3):477–489.
- Mansourvar, Z., Martinussen, T., and Scheike, T. H. (2016). An additive-multiplicative restricted mean residual life model. *Scandinavian Journal of Statistics*, 43(2):487–504.
- Murphy, S., Rossini, A., and van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, 92(439):968–976.
- Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4):945–966.

- Neuhaus, J. M. and McCulloch, C. E. (2011). Estimation of covariate effects in generalized linear mixed models with informative cluster sizes. *Biometrika*, 98(1):147–162.
- Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sørensen, T. I. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, pages 25–43.
- Oakes, D. and Dasu, T. (1990). A note on residual life. *Biometrika*, 77(2):409–410.
- Ortega, J. M. and Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables*, volume 30. SIAM.
- Park, E. and Ha, I. D. (2019). Penalized variable selection for accelerated failure time models with random effects. *Statistics in Medicine*, 38(5):878–892.
- Parner, E. et al. (1998). Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics*, 26(1):183–214.
- Price, M. H. and Jones, J. H. (2017). A general goodness-of-fit test for survival analysis. *bioRxiv*, pages 104–406.
- Qu, A., Lindsay, B. G., and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika*, 87(4):823–836.
- Rakha, E. A., Reis-Filho, J. S., Baehner, F., Dabbs, D. J., Decker, T., Eusebi, V., Fox, S. B., Ichihara, S., Jacquemier, J., Lakhani, S. R., et al. (2010). Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Research*, 12(4):1–12.
- Ren, S., Lai, H., Tong, W., Aminzadeh, M., Hou, X., and Lai, S. (2010). Non-parametric bootstrapping for hierarchical data. *Journal of Applied Statistics*, 37(9):1487–1498.

- Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4):1016–1022.
- Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS epidemiology*, pages 297–331. Springer.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866.
- Song, S. E., Shin, S. U., Moon, H.-G., Ryu, H. S., Kim, K., and Moon, W. K. (2017). Mr imaging features associated with distant metastasis-free survival of patients with invasive breast cancer: a case-control study. *Breast Cancer Research and Treatment*, 162(3):559–569.
- Sun, L. and Zhang, Z. (2009). A class of transformed mean residual life models with censored survival data. *Journal of the American Statistical Association*, 104(486):803–815.
- Verbeke, G., Spiessens, B., and Lesaffre, E. (2001). Conditional linear mixed models. *The American Statistician*, 55(1):25–34.
- Wang, P., Tsai, G.-f., and Qu, A. (2012). Conditional inference functions for mixed-effects models with unspecified random-effects distribution. *Journal of the American Statistical Association*, 107(498):725–736.
- Xing, Y., Xu, L., Ma, W., and Zhu, Z. (2018). Conditional mix-gee models for longitudinal data with unspecified random-effects distributions. *Communications in Statistics-Theory and Methods*, 47(4):862–876.
- Xu, J., Kalbfleisch, J. D., and Tai, B. (2010). Statistical analysis of illness-death processes and semicompeting risks data. *Biometrics*, 66(3):716–725.

- Yang, G. L. et al. (1978). Estimation of a biometric function. *The Annals of Statistics*, 6(1):112–116.
- Yu, D., Zhang, X., and Yau, K. K. (2018). Asymptotic properties and information criteria for misspecified generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):817–836.
- Yuen, K., Zhu, L., and Tang, N. (2003). On the mean residual life regression model. *Journal of Statistical Planning and Inference*, 113(2):685–698.
- Zhang, Z., Zhao, X., and Sun, L. (2010). Goodness-of-fit tests for additive mean residual life model under right censoring. *Lifetime Data Analysis*, 16(3):385–408.