Analysis of complex survival data subject to semi-competing risks

Peng, Mengjiao

2019

Peng, M. (2019). Analysis of complex survival data subject to semi-competing risks. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/102662>

<https://doi.org/10.32657/10220/47771>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**ANALYSIS OF COMPLEX SURVIVAL DATA
SUBJECT TO SEMI-COMPETING RISKS**

PENG MENGJIAO

Division of Mathematical Sciences
School of Physical and Mathematical Sciences

2019

ANALYSIS OF COMPLEX SURVIVAL DATA SUBJECT TO SEMI-COMPETING RISKS

PENG MENGJIAO

Division of Mathematical Sciences
School of Physical and Mathematical Sciences

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

2019

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.

Mar 3, 2019

.....

Date

.....

PENG MENGJIAO

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.

Mar 3 , 2019

.....

Date

.....

XIANG LIMING

Authorship Attribution Statement

This thesis contains material from 1 paper published in the following peer-reviewed journal where I was the first author.

Chapter 2 is published as Peng et al. (2018): “M.J. Peng, L.M. Xiang, and S.S. Wang. Semiparametric regression analysis of clustered survival data with semi-competing risks. *Computational Statistics & Data Analysis*, 124:53–70, 2018. DOI: 10.1016/j.csda.2018.02.003.

The contributions of the co-authors are as follows:

Prof. Xiang suggested the materials area, proposed the interested topic, provided valuable guidance throughout the study of this topic, and edited the manuscript draft.

I wrote the draft of the manuscript. And the manuscript was revised together with Prof. Xiang.

I performed all the simulations with guidance from Prof. Xiang, and with fruitful discussions with Wang shanshan.

I proved the asymptotic properties with guidance from Prof. Xiang.

Mar 3, 2019

.....

Date

.....

PENG MENGJIAO

Abstract

This thesis is devoted to develop novel methods for the analysis of complex survival data subject to semi-competing risks. In many clinical trials, multiple time-to-event endpoints, including the primary endpoint (e.g., time to death) and secondary endpoints (e.g. progression-related endpoints) are commonly used to determine treatment efficacy. These endpoints are often biologically related. One secondary endpoint and one primary endpoint often arise in medical research and form semi-competing risks where the observation of the secondary endpoint will be censored by the primary endpoint when it occurs earlier than the secondary endpoint, but not vice versa. There has been a longstanding research interest in the analysis of semi-competing risks data.

We focus on the study of complex survival data including clustered data, data subject to two sets of semi-competing risks and data with ultrahigh dimensional covariates.

In many biomedical studies, subjects may also be nested within clusters, such as patients in a multi-center study, leading to possible association among event times due to unobserved shared factors across subjects. Analysis for clustered event times with semi-competing risks is needed. We propose a flexible semi-parametric modeling framework where a copula model is employed for the joint distribution of the nonterminal and terminal event times, and their marginal distributions are modeled by Cox PH models with random effects to incorporate association of event times within cluster. We develop a nonparametric maximum likelihood estimation procedure via a Monte Carlo EM algorithm, and establish

desirable asymptotic properties of the resulting estimators.

In addition, multiple endpoints including the primary endpoint and two secondary endpoints are often encountered in clinical trials. These give rise to data that are subject to two sets of semi-competing risks. We propose a novel statistical approach that jointly models such data via a pair of copulas to account for multiple dependence structures, while the marginal distribution of each endpoint is formulated by a Cox PH model. And we develop an estimation procedure based on pseudo-likelihood method.

Besides, we consider semi-competing risks data collected together with ultrahigh-dimensional gene features in modern cancer studies, where the number of gene features p is extremely larger than sample size n . Microarray-based expression studies have demonstrated that gene expression patterns are associated with outcomes of breast cancer patients. To identify the important gene features plays an essential role in disease classification, survival prediction and the development of precision medicine. We aim to identify such important features that greatly impact both the nonterminal and the terminal endpoints. We propose a model-free screening method based on the ranking of the correlation between the features and the joint survival function of two endpoints. The method incorporates the relationship between two endpoints and is simple and easy to compute. We establish the ranking consistency and sure independent screening properties for the proposed method.

Extensive simulation studies have been carried out and results show promising finite sample performance of all proposed methods. Analysis of real data sets from various medical studies demonstrate the practical utility of our proposed methods.

Acknowledgement

First and foremost, I would like to express my heartfelt gratitude to my supervisor Prof. Liming Xiang for her invaluable guidance and great help for my research work and my life. I have learned a lot from her rigorous attitude towards academic and shared life experience. Also I want to thank her for providing me the chance and privilege of working in this interesting area, and for her patient and meticulous instruction, encouragement, and support during my study in Nanyang Technological University. Without her priceless guidance and supervision over these years, this thesis could not have reached its present form.

I am grateful to the Ministry of Education in Singapore and Nanyang Technological University for awarding me the Research Scholarship. I am also benefited from the energetic and productive environment of the School of Physical and Mathematical Sciences. In particular, I appreciate the management and leadership of Prof. San Ling, Prof. Chan Song Heng, Professor Frédérique Oggier, and Prof. Huaxiong Wang. I am appreciative of the help from the IT and Admin. staff.

I would like to express my appreciation to Prof. Liming Xiang, Prof. Chaoping Xing, Prof. Huaxiong Wang and Prof. Guohua Wu for the interesting lectures and seminar courses. I am also indebted to the former and current members in our group including Zsolt Szabó, Rui Huang, Tonghui Yu, Xiaoyu Liu, for their friendship and help.

Last but not least, I wish to thank my parents, my husband Zhiguo Yang, my brother and sister, and my daughter for their love and support.

Contents

Abstract	1
Acknowledgement	3
List of Figures	11
List of Tables	18
1 Introduction	19
1.1 Survival data	19
1.2 Semi-competing risks data and dependent censoring mechanism .	21
1.3 Conventional methods for analyzing semi-competing risks data . .	22
1.4 Clustered survival data with semi-competing risks	24
1.5 Survival data in the presence of two sets of semi-competing risks	26
1.6 Screening for semi-competing risks data with ultrahigh dimensional gene features	29
1.7 Main contributions and outline	32
2 Semiparametric regression analysis of clustered semi-competing risks data	35
2.1 Model Formulation	36
2.2 Estimation	37

2.3	Implementation and algorithm	41
2.4	Theoretical result	45
2.5	Simulation Study	48
2.6	Real Data Analysis	68
3	Joint regression analysis for survival data in the presence of two sets of semi-competing risks	71
3.1	Proposed model	72
3.2	Pseudo-likelihood estimation	73
3.3	Implementation and algorithm	75
3.4	Cross ratio of two nonterminal event times	76
3.5	Simulation Study	77
3.6	Real Data Analysis	86
4	Joint correlation rank screening for semi-competing risks data	93
4.1	Screening procedures	94
4.1.1	Semi-competing risks Data	94
4.1.2	Joint correlation rank screening procedure	94
4.1.3	Estimation	95
4.1.4	Screening steps	96
4.2	Theoretical result	96
4.3	Simulation Study	98
4.4	Real Data Analysis	103
5	Discussion and Future Works	113
5.1	Conclusion and discussion for Chapters 2-4	113
5.2	Limitations and future works	115
	Bibliography	117
	Appendix A Proofs of Theorems 2.1-2.2	127

Appendix B Simulations for Weibull baseline hazard functions in Chapter 2	155
Appendix C Details of some notations used in Chapter 3	163
Appendix D Proofs of Theorems 4.1-4.2	165
Appendix E Simulations for Gumbel copula in Chapter 4	169

List of Figures

1.1	Disease process after bone marrow transplant	27
2.1	Simulation results for estimated cumulative baseline hazard function $\Lambda_{01}(\cdot)$ by the proposed method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{01}(t) = t$ (colored black) under the model with gamma frailty in Scenario 2.1 and $\tau = 0.3$	64
2.2	Simulation results for estimated cumulative baseline hazard function $\Lambda_{02}(\cdot)$ by the proposed method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{02}(t) = 0.4t$ (colored black) under the model with gamma frailty in Scenario 2.1 and $\tau = 0.3$	65
2.3	Simulation results for estimated cumulative baseline hazard function $\Lambda_{01}(\cdot)$ by Emura et al. (2017)'s method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{01}(t) = t$ (colored black) under the model with gamma frailty in Scenario 2.1 and $\tau = 0.3$	66

2.4	Simulation results for estimated cumulative baseline hazard function $\Lambda_{02}(\cdot)$ by Emura et al. (2017)'s method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{02}(t) = 0.4t$ (colored black) under the model with gamma frailty in Scenario 2.1 and $\tau = 0.3$	67
3.1	Estimated cumulative baseline hazard $\Lambda_{01}(t)$ based on the first 50 replications (colored red) with true cumulative baseline hazard $\Lambda_{01}(t_1) = \frac{2}{3}t_1^{3/2}$ (colored black) under Scenario 3.3 with $\tau_{10} = 0.3$ and $\tau_{20} = 0.3$	86
3.2	Estimated cumulative baseline hazard $\Lambda_{02}(t)$ based on the first 50 replications (colored red) with true cumulative baseline hazard $\Lambda_{02}(t_2) = \frac{1}{3}t_2$ (colored black) under Scenario 3.3 with $\tau_{10} = 0.3$ and $\tau_{20} = 0.3$	86
3.3	Estimated cumulative baseline hazard $\Lambda_{03}(t)$ based on the first 50 replications (colored red) with true cumulative baseline hazard $\Lambda_{03}(t_3) = \frac{2}{3}t_3^{3/2}$ (colored black) under Scenario 3.3 with $\tau_{10} = 0.3$ and $\tau_{20} = 0.3$	87
3.4	Estimated $\psi(t, t)$ based on the first 50 replications with $\tau_{10} = 0.3, \tau_{20} = 0.3$ (—); $\tau_{10} = 0.3, \tau_{20} = 0.7$ (--); $\tau_{10} = 0.7, \tau_{20} = 0.7$ ($\cdots$).	87
3.5	Estimated $\psi(t, t)$ for the bone marrow transplant data.	90
3.6	Estimated baseline survival functions using the proposed method (colored red) and the KM method (colored blue) based on the bone marrow transplant data. Dotted lines indicate the 95% confidence bounds of the KM estimator.	92
B.1	Simulation results for estimated cumulative baseline hazard function $\Lambda_{01}(\cdot)$ by the proposed method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{01}(t) = \frac{2}{3}t^{3/2}$ (colored black) under the model with normal random effects in Scenario 2.1 and $\tau = 0.3$	161

B.2	Simulation results for estimated cumulative baseline hazard function $\Lambda_{02}(\cdot)$ by the proposed method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{02}(t) = \frac{1}{2}t^{3/2}$ (colored black) under the model with normal random effects in Scenario 2.1 and $\tau = 0.3$.	162
-----	---	-----

List of Tables

2.1	Simulation results based on 200 replications under the model with normally distributed random effects in the intercept in Scenario 2.1, where parameters $\eta = 0.2$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6	53
2.2	Simulation results based on 200 replications under the model with normally distributed random effects in the intercept in Scenario 2.2, where parameters $\eta = 0.4$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6	54
2.3	Simulation results based on 200 replications under the model with normally distributed random effects in the slope in Scenario 2.3, where parameters $\eta = 0.2$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6	55
2.4	Simulation results based on 200 replications under the model with normally distributed random effects in the slope in Scenario 2.4, where parameters $\eta = 0.4$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6	56

-
- 2.5 Simulation results for the model with gamma frailty obtained by the proposed and Emura et al. (2017) 's methods based on 200 replications in Scenario 2.1, with parameters $\eta = 0.2$, $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ 57
- 2.6 Simulation results for the model with gamma frailty obtained by the proposed and Emura et al. (2017) 's methods based on 200 replications in Scenario 2.1, with parameters $\eta = 0.2$, $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.6$ 58
- 2.7 Simulation results for the model with gamma frailty obtained by the proposed and Emura et al. (2017)'s methods based on 200 replications in Scenario 2.2, with parameters $\eta = 0.4$, $\beta_1 = \beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ 59
- 2.8 Simulation results for the model with gamma frailty obtained by the proposed and Emura et al. (2017)'s methods based on 200 replications in Scenario 2.2, with parameters $\eta = 0.4$, $\beta_1 = \beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.6$ 60
- 2.9 Simulation results obtained by the proposed methods using a misspecified normal random effect model in Scenario 2.1 based on 200 replications when the true random effects distribution is gamma. The true parameters $\eta = 0.2$, $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6 61
- 2.10 Simulation results obtained by both the proposed and Emura et al. (2017)'s methods using a misspecified gamma frailty model in scenario 2.1 based on 200 replications when the true random effects distribution is a mixture of normals $0.5N(0, 0.2) + 0.5N(0, 0.4)$. The true parameters $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ 62

2.11	Simulation results obtained by both the proposed and Emura et al. (2017)'s methods using a misspecified gamma frailty model in scenario 2.1 based on 200 replications when the true random effects distribution is a mixture of normals $0.5N(0, 0.2) + 0.5N(0, 0.4)$. The true parameters $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.6$	63
2.12	Analysis of the breast cancer data	70
3.1	Biases, standard derivations(sd's in parentheses), and cp(%) for the proposed estimator based on 500 replications under Scenario 3.1.	81
3.2	Biases, standard derivations(sd's in parentheses), and cp(%) for the proposed estimator based on 500 replications under Scenario 3.2.	82
3.3	Biases, standard derivations(sd's in parentheses), and cp(%) for the proposed estimator based on 500 replications in Scenario 3.3.	83
3.4	Bootstrap standard errors and standard derivations (sd's in parentheses) for the proposed estimator based on 500 replications under Scenario 3.1.	84
3.5	Simulation results based on 500 replications under the proposed model and marginal models in Scenario 3.1 with $\tau_{10} = 0.3$ and $\tau_{20} = 0.7$	85
3.6	Results of estimates and standard errors (in parentheses) obtained under the proposed model and marginal models for the bone marrow transplant data.	91
4.1	Selection proportions for each active covariate, for all active covariates (i.e. $X_1, \dots, X_4$, denoted by $X_{1,\dots,4}$) associated with $T^{(2)}$ and for all active covariates (i.e. $X_1, \dots, X_6$, denoted by $X_{1,\dots,6}$) associated with $(T^{(1)}, T^{(2)})$ based on 500 replications under Scenario 4.2.	104

4.2	Selection proportions for each active covariate, for all active covariates (i.e. $X_1, \dots, X_4$, denoted by $X_{1,\dots,4}$) associated with $T^{(2)}$ and for all active covariates (i.e. $X_1, \dots, X_6$, denoted by $X_{1,\dots,6}$) associated with $(T^{(1)}, T^{(2)})$ based on 500 replications under Scenario 4.2.	105
4.3	Selection proportions for each active covariate, for all active covariates (i.e. $X_1, \dots, X_4$, denoted by $X_{1,\dots,4}$) associated with $T^{(2)}$ and for all active covariates (i.e. $X_1, \dots, X_6$, denoted by $X_{1,\dots,6}$) associated with $(T^{(1)}, T^{(2)})$ based on 500 replications under Scenario 4.3.	106
4.4	Quantiles of the minimum model size which includes all active covariates (i.e. X_1, X_2, X_3) associated with $T^{(2)}$ based on 500 replications under Scenario 4.1.	107
4.5	Quantiles of the minimum model size which includes all active covariates (i.e. $X_1, \dots, X_4$) associated with $T^{(2)}$ based on 500 replications under Scenario 4.2.	108
4.6	Quantiles of the minimum model size which includes all active covariates (i.e. $X_1, \dots, X_4$) associated with $T^{(2)}$ based on 500 replications under Scenario 4.3.	109
4.7	Quantiles of the minimum model size which includes all active covariates (i.e. $X_1, \dots, X_6$) associated with $(T^{(1)}, T^{(2)})$ obtained by the proposed JCR method, based on 500 replications under Scenario 4.1.	110
4.8	Quantiles of the minimum model size which includes all active covariates (i.e. $X_1, \dots, X_6$) associated with $(T^{(1)}, T^{(2)})$ obtained by the proposed JCR method, based on 500 replications under Scenario 4.2.	110

4.9	Quantiles of the minimum model size which includes all active covariates (i.e. $X_1, \dots, X_6$) associated with $(T^{(1)}, T^{(2)})$ obtained by the proposed JCR method, based on 500 replications under Scenario 4.3.	111
4.10	Gene selected by JCR and CR methods	112
B.1	Simulation results based on 200 replications under the model with normally distributed random effects in the intercept in Scenario 2.1, where parameters $\eta = 0.2$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = t^{1/2}$, $\lambda_{02}(t) = \frac{3}{4}t^{1/2}$, and Kendall's $\tau = 0.3$ and 0.6	157
B.2	Simulation results based on 200 replications under the model with normally distributed random effects in the intercept in Scenario 2.2, where parameters $\eta = 0.4$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = t^{1/2}$, $\lambda_{02}(t) = \frac{3}{4}t^{1/2}$ and Kendall's $\tau = 0.3$ and 0.6	158
B.3	Simulation results based on 200 replications under the model with normally distributed random effects in the slope in Scenario 2.3, where parameters $\eta = 0.2$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = t^{1/2}$, $\lambda_{02}(t) = \frac{3}{4}t^{1/2}$, and Kendall's $\tau = 0.3$ and 0.6	159
B.4	Simulation results based on 200 replications under the model with normally distributed random effects in the slope in Scenario 2.4, where parameters $\eta = 0.4$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = t^{1/2}$, $\lambda_{02}(t) = \frac{3}{4}t^{1/2}$ and Kendall's $\tau = 0.3$ and 0.6	160
E.1	Selection proportions for each active covariate, for all active covariates (i.e. $X_1, \dots, X_3$, denoted by $X_{1,2,3}$) associated with $T^{(2)}$ and for all active covariates (i.e. $X_1, \dots, X_6$, denoted by $X_{1,\dots,6}$) associated with $(T^{(1)}, T^{(2)})$ based on 500 replications under Scenario 4.1 and Gumbel copula.	170

E.2	Quantiles of the minimum model size which includes all active co- variates (i.e. X_1, X_2, X_3) associated with $T^{(2)}$ based on 500 repli- cations under Scenario 4.1 and Gumbel copula.	171
E.3	Quantiles of the minimum model size which includes all active covariates (i.e. $X_1, \dots, X_6$) associated with $(T^{(1)}, T^{(2)})$ obtained by the proposed JCR method, based on 500 replications under Scenario 4.1 and Gumbel copula.	172

Introduction

In this chapter, we elaborate on motivations and backgrounds of the topics to be studied in this thesis, and highlight the main contributions.

1.1 Survival data

Survival analysis is a collection of statistical procedures for analysis of survival data, which include observations of the time to an event of interest, such as disease incidence, relapse from remission, recovery and death. Survival time is a positive random variable. Although discrete survival time is possible, in this thesis we focus on survival time being a continuous random variable with continuous distribution functions, most likely to encounter in real life. The distribution of survival times can often be described in three different functions: probability distribution function (PDF), survival function and hazard function (or hazard rate). For survival time T , the cumulative distribution function (CDF) is defined as:

$$F(t) = P(T \leq t), \quad t \geq 0, \quad (1.1)$$

and the pdf is

$$f(t) = \frac{dF(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t}. \quad (1.2)$$

Here, $f(t)$ is the limit of the probability that the event occurs in the short time interval t to $t + \Delta t$ per unit time. It gives the rate of event occurrence at t .

In biomedical applications, it is often common to use the survival function defined as

$$S(t) = P(T > t) = 1 - F(t). \quad (1.3)$$

Here, $S(t)$ is the probability that a randomly selected individual will survive to time t or beyond against the event. It is monotone and non-increasing in t , and $S(0) = 1$ and $S(\infty) = 0$. Note that $S(t) = \int_t^\infty f(t)dt$, and $f(t) = -\frac{dS(t)}{dt}$.

The hazard function $\lambda(t)$ is defined as

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t}. \quad (1.4)$$

From the definition, $\lambda(t) \geq 0$, $\lambda(t)\Delta t$ can be viewed as the “approximate” probability of an individual of age t experiencing the event in the next instant. Thus the hazard rate is a failure rate that gives the risk of event occurrence per unit time during the aging process. Note that $\lambda(t) = \frac{f(t)}{S(t)} = \frac{-d \log[S(t)]}{dt}$.

The cumulative hazard function is defined by

$$\Lambda(t) = \int_0^t \lambda(x)dx = -\log[S(t)]. \quad (1.5)$$

There are 1 – 1 relationships between any two of the PDF, survival function and hazard rate, that is, given any one of the three functions, the other two can be easily derived.

Analysis of survival data is complicated due to censoring, that is the exact time to event is not observed for some individuals and is unavoidable in survival analysis. There are 3 major categories of censoring, which are right censoring, left censoring and interval censoring. Right censoring often occurs in medical studies. Right censoring happens if the event of interest occurs after a certain time but the exact event time is not known by the end of the study, for example, a medical study was terminated before some individuals experienced the event, or some individuals left the study before they experienced the event, or some

1.2 Semi-competing risks data and dependent censoring mechanism 21

individuals were lost to follow up in the study. Denote T and C as the time to event and external censoring time, respectively. Subject to right censoring, the observation of event time is defined by $Y = \min(T, C)$ and the indicator for observation is defined by $\delta = I(T \leq C)$. C may be the time to end of study. In this thesis, we focus on data subject to right censoring. And we assume that all the observations of survival time are distinct without loss of generality.

Most of the standard survival analysis methods give unbiased results under the independent censoring assumption in which T and C are statistically independent. For example, the Kaplan-Meier (KM) estimator (Kaplan and Meier, 1958) is a standard tool to estimate the survival function and is a decreasing step function with jumps only at those observed event times where the event occurs (Emura and Chen, 2018). Suppose there are n observations, the KM estimator is defined by

$$\hat{S}(t) = \prod_{i=1}^n \left\{ 1 - \frac{1}{n_i} \right\}^{\delta_i I(Y_i \leq t)}, \quad (1.6)$$

where $n_i = \sum_{j=1}^n I(Y_j \geq Y_i)$ is the number at risk at time Y_i , i.e., the number of subjects which have not experienced the event yet at time Y_i .

More generally, if event time of interest is censored by any mechanism related to the event, then dependent censoring is present in the data (Emura and Chen, 2018).

1.2 Semi-competing risks data and dependent censoring mechanism

Medical research for assessing treatment efficacy and disease progression often yields multiple survival endpoints, which include the primary endpoint (e.g. time to death) and some important secondary endpoints (e.g. times to intermediate disease progression and related disorders). The event death is generally referred to as a terminal event, the time to which can be observed even if intermediate disease

progressions occur earlier. Yet, the disease progression and related disorders are regarded as nonterminal events, which can be censored by death if it occurs before disease progressions. Both the nonterminal and the terminal events contribute “semi-competing risks” (Fine et al., 2001). As the times to the terminal and nonterminal events of an individual are usually correlated, the censoring of the nonterminal event by the terminal event forms a type of dependent/informative censoring.

1.3 Conventional methods for analyzing semi-competing risks data

There is a large literature on the study of semi-competing risks data (see e.g. Fine et al. (2001), Peng and Fine (2007), Xu et al. (2010), Emura et al. (2017) and Emura and Chen (2018)). The main interests lie in the effects of covariates on the nonterminal and terminal event times while incorporating dependency between them. For this purpose, regression analysis of such data has been studied generally through two kinds of models, copula models and multi-state models. Peng and Fine (2007), Hsieh et al. (2008) and Chen (2012) considered copula models by imposing separate marginal regression models on the two event times under dependent censoring scenario.

The copula models assumed that the joint survival function of the terminal event time $T^{(1)}$ and the nonterminal event time $T^{(2)}$ can be specified by a known copula subjects to an unknown parameter α

$$S(t^{(1)}, t^{(2)}) = \text{pr}(T^{(1)} \geq t^{(1)}, T^{(2)} \geq t^{(2)}) = c\{S_1(t^{(1)}), S_2(t^{(2)}); \alpha\}, \quad (1.7)$$

for $t^{(1)} \geq 0, t^{(2)} \geq 0$, where $c\{u, v; \alpha\}$ is a copula function, α is a constant association parameter in one-to-one correspondence with Kendall’s τ , and $S_k(\cdot)$ is the marginal survival function of $T^{(k)}$ ($k = 1, 2$). Kendall’s τ is defined as the probability of concordance minus the probability of discordance between the two

random variables, see more details in Section 5.1.1 of Nelsen (2007). The key advantage of model (1.7) is that it describes the joint distribution of the terminal and nonterminal event times through an explicit function of their marginal distributions and the corresponding association parameter. Below we give two common examples of copulas: the Clayton copula

$$c\{u, v; \alpha\} = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}, \quad \alpha > 0,$$

and the Gumbel copula

$$c\{u, v; \alpha\} = \exp \left[- \left\{ (-\log u)^{\alpha+1} + (-\log v)^{\alpha+1} \right\}^{1/(\alpha+1)} \right], \quad \alpha > 0.$$

The Clayton and Gumbel copulas above are two of the most important Archimedean copulas and are popular in practice because of their simple explicit expressions with only a single parameter controlling the strength of dependence. They can be used to capture asymmetry between the lower and upper tail dependence. The Clayton copula exhibits a higher degree of dependence in the lower tail than in the upper tail, while the Gumbel copula has strong upper-tail dependence with relatively weak lower-tail dependence. For more information on different types of copulas and their properties, and the generation of data with copula structure, we refer the reader to the book by Nelsen (2007), Rotolo et al. (2013) and the references therein.

Nowadays, copulas have been extensively applied for analysis of semi-competing risks (Fine et al., 2001; Wang, 2003; Peng and Fine, 2007; Hsieh et al., 2008; Chen, 2012). Copulas often offer a more transparent strategy for building a model of dependent censoring and estimating parameters from the methodological point of view. Moreover, copulas provide flexible bivariate survival models to analysis survival data subject to dependent censoring. One can choose from different kinds of copulas while without any restriction on the marginal survival models.

Alternatively, Xu et al. (2010) proposed a three-state illness-death model with shared gamma frailty for modeling the semi-competing risks data, where a subject can either transit directly to the terminal event or first to the nonterminal

event and then to the terminal event. The models could be specified by the transition intensity functions for the three distinct transitions with a shared frailty contributed to the dependence of both nonterminal and terminal events. Their model avoids a modeling problem of the restriction of triangle range to some extent. Their model was extended by Jiang and Haneuse (2016) to a class of transformation frailty models, allowing a wider range of possible frailty distributions.

1.4 Clustered survival data with semi-competing risks

Survival data arise with a cluster structure in many biomedical studies, such as the survival times of subjects observed in multicenter clinical trials, genetic and family studies, or group-randomized trials. The unique feature of clustered survival data is that event times obtained from subjects within clusters are expected to be correlated due to their exposure to the same environmental risk factors or similar life styles. Usually, the subjects within the same cluster share the common cluster effects. Such cluster effects lead to dependencies among survival times within each cluster. If the cluster effects are sufficient strong, they may exert a considerable influence on study outcomes. Evidence from the work in van Vliet et al. (2008) showed that data collected by different clinical institutions from the underlying breast cancer population produced different results in predicting disease outcome, implying potential cluster effects in the data. This motivates our work in Chapter 2 to analyze clustered semi-competing risks data by incorporating both the association within clusters and the dependence between the nonterminal and terminal endpoints.

A common approach to accommodate the intracluster dependence is to incorporate an unobserved frailty (or random effects) in a survival model, where

the frailty acts multiplicatively to the hazard function in order to model the correlation within a cluster. Incorporating random effects (or frailty) in survival models poses additional challenges to statistical inference. In particular, due to the random effects (may be multi-dimensional), the likelihood function includes an intractable integration, leading to great difficulties in theoretical justification of the consistency and asymptotic normality of the proposed estimator when using counting process techniques and the martingale central limit theorem. In the computational aspect, the involvement of random effects also requires extra computational efforts to handle numerical integration in the estimation procedure. Ripatti and Palmgren (2000), Vaida and Xu (2000) and Ripatti et al. (2002) generalized the frailty model to the proportional hazards model with random effects which allows random effects on arbitrary covariates. However, these works only considered one interested survival endpoint. When two or several different types of events are of interest, most of the existing works for clustered data deal with competing risks by modeling one type of event at a time and regarding the occurrence of other types of events as censoring based on independent censoring assumption. See for examples Gorfine and Hsu (2011), Christian et al. (2016), Lee et al. (2017) and references therein. This assumption of independent censoring is not true for semi-competing risks data. In addition, models by Peng and Fine (2007), Hsieh et al. (2008), Lakhal et al. (2008), Xu et al. (2010), Chen (2012) and Jiang and Haneuse (2016) discussed in Section 1.3, did not consider the dependencies among event times within clusters. The application of these methods to clustered semi-competing risk data may lead to a loss of efficiency. Failure to take into account such intra-cluster correlation in the analysis could lead to misleading inference. (Hougaard, 2012).

Despite of the large literature on analysis of semi-competing risks data, relatively few works are available for such data with clustered structure. Emura et al. (2017) introduced frailty to a joint copula model for semi-competing risks data arisen in a meta-analysis including several existing studies. They used splines

to approximate the baseline hazard functions, and estimated model parameters through a penalized maximum likelihood method. Their method requires the gamma frailty assumption so that estimation through maximizing a penalized log-likelihood can be mathematically tractable. However, the gamma frailty assumption is not necessary in practice. Their method would fail to produce consistent estimators if the gamma frailty distribution is misspecified. Additionally, there is a lack of theoretical justification for the asymptotic properties of the existing estimator. Using Bayesian hierarchical modeling approach, Lee et al. (2016) further extended the multistate illness-death model in Xu et al. (2010) to the clustered semi-competing risks data setting, allowing a parametric or non-parametric specification for the cluster-specific random effects distribution. Lee et al. (2016) performed estimation and inference within the Bayesian paradigm with advantages including the natural incorporation of prior information, and developed computationally efficient algorithms for the estimation, but did not establish asymptotic theory for the estimators.

Thus we are motivated to propose a novel and flexible model framework in Chapter 2 to deal with clustered semi-competing risks data, where the frailty assumption is relaxed and asymptotic properties are justified for the estimators.

1.5 Survival data in the presence of two sets of semi-competing risks

The conventional semicompeting risks setting consists of a pair of the primary endpoint and one secondary endpoint only. In many medical researches, multiple endpoints including one primary endpoint and some secondary endpoints are of interest, where the primary endpoint and each of secondary endpoints form a set of semicompeting risks. For example, in the study of a cohort of leukemia patients who received allogeneic bone marrow transplant (BMT) (Copelan et al., 1991),

Figure 1.1. Disease process after bone marrow transplant

the primary endpoint is the time to death after the BMT, namely progression-free survival, which is the most clinical relevant endpoint in leukemia studies. Despite the success of BMT, transplantation related complications, such as the acute Graft-Versus-host disease (GVHD) frequently occurs, which can affect the survival of patients (Shokouhi et al., 2015). Individuals may experience the acute GVHD, relapse of leukemia, and death after an allogeneic BMT, where the acute GVHD is associated with the relapse free survival, both the acute GVHD and relapse of leukemia are intermediate nonterminal events subject to dependent censoring by the informative terminal event death, but not vice versa. The relevant disease process is shown in Figure 1.1. In addition, some existing works in medical research found that the patients experienced the acute GVHD had a lower risk of relapse compared to patients who did not experience acute GVHD (Lee et al., 2005). These give rise to complex survival data that are subject to two sets of semicompeting risks. The aim of this study is to evaluate the associations between each intermediate event (acute GVHD or relapse) and the death, incorporating possible dependency between two intermediate events, and in the mean time to determine prognostic factors for all three endpoints.

The existing marginal estimation methods developed for data subject to one set of semi-competing risks mentioned in Section 1.3 to each pair of intermediate nonterminal and terminal endpoints separately ignores the dependency between

the two nonterminal events. Failure to account for such dependency may lead to invalid statistical inference. In addition, these models are unable to explain how both the nonterminal endpoints are associated. Although multi-state models have been used to analyze multiple endpoints as in the works of Keiding et al. (2001) and Siannis et al. (2007), the two types of relationships between the non-terminal and terminal events, and between the two nonterminal events still can not be well accounted for. An unobserved frailty term is used in the multi-state models to account for dependency among the events. However, the variance of the frailty cannot well explain to which extent the events are associated. To address these limitations, we seek to propose a novel modeling approach that jointly models complex survival data under two sets of semi-competing risks. Our method employs a pair of copulas to account for multiple dependence structures, while the marginal distribution of each endpoint is formulated by a Cox proportional hazards (PH) model.

To the best of our knowledge, our work in Chapter 3 is the first attempt to provide approach for analyzing data from multiple endpoints under complex dependent censoring caused by two sets of semi-competing risks. Such complexity among multiple endpoints brings challenges to obtain the full likelihood. Motivated by the pseudo-likelihood proposed in Varin and Vidoni (2005) and Varin et al. (2011), we consider a special form of pseudo-likelihood to accommodate complex interdependencies. The pseudo-likelihood method known as a kind of composite likelihood, has shown its high efficiency in the literature (see e.g. Renard et al. (2002) and Varin et al. (2011)). The pseudo-likelihood method is appealing because it requires only model assumptions on the lower dimensional marginal densities, instead of detailed specification of the full joint distribution, and thus naturally providing inference robust to misspecification of the full joint distribution.

1.6 Screening for semi-competing risks data with ultrahigh dimensional gene features

Microarray-based expression studies have demonstrated that breast cancer is a molecularly heterogeneous disease comprising subtypes with distinct gene expression patterns that are associated with overall and relapse-free survival (Sørlie et al., 2001; Sotiriou et al., 2003; Wirapati et al., 2008; Haibe-Kains et al., 2012). Gene expression data arising from such studies have the potential to aid accurate and individualized prognosis, and these data are assessed in terms of precise numerical predictions with ranges of probabilities of outcomes like metastasis and relapse (Huang et al., 2003). Some important genes are frequently used for tumor/disease classification. However, in such cancer studies, gene features involved are ultrahigh-dimensional, that is, the number of gene features p is extremely larger than sample size n . For example, a study of breast cancer by National Kanker Institute (the Netherlands) includes valid survival times of 295 patients and gene expression values of 14318 DNA elements for each patient. Statistical analysis of ultrahigh-dimensional data often starts with an initial screen step to reduce the number of covariates under consideration and then use one of the more sophisticated regularization techniques for variable selection in the second stage. Finally well-developed methods can then be adopted to analyze the data with the selected features. Therefore, identifying the important features plays an essential role in statistical procedures. In Chapter 4, we focus on the initial screen step, that is to identify genes that have great influence on patients' overall survival, which is of great importance for disease classification, precise survival predictions and valid assessments of risks specific for patients in precision medicine.

There are various regularization methods for feature selection, including Lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001), LARS (Efron et al., 2004) and adaptive Lasso (Zou, 2006). The regularization methods may not perform well

considering the challenges of computation cost, estimation accuracy and algorithm stability faced by ultrahigh dimensionality (Fan et al., 2009). To address the challenges of ultrahigh dimensionality, Fan and Lv (2008) proposed the sure independence screening (SIS) and iterated sure independence screening (ISIS) under an ordinary linear model with normal errors. The independence screening recruits the features having the best marginal utility, which corresponds to the largest marginal correlation with the response in least-squares regression (Fan and Lv, 2010). Later, Fan et al. (2010) extended the SIS and ISIS to generalized linear models. Zhu et al. (2011) proposed a model-free feature screening procedure under a general model framework, which covers many frequently used parametric and semiparametric models based on the marginal utility measure of the correlation between the features and the distribution of the univariate response variable to screen variables. A model-free SIS procedure was proposed in Li et al. (2012a) based on the distance correlation. However, all of the above methods were developed for complete data and thus can not be used to deal with censored survival data.

Song et al. (2014) proposed a censored rank independence sure screening based on a marginal rank correlation utility measure with inverse probability-of-censoring weights. Their method was shown to be robust against the potential outliers and to work for a general class of survival models. Zhou and Zhu (2017) also made use of the weight of the inverse probability of censoring to extend the earlier screening method for a general regression model framework by Zhu et al. (2011) to survival regression with censored data. In parallel, Zhang et al. (2017) also applied the idea used in Zhu et al. (2011) to define a correlation between cumulative density function of response and covariates, and estimate the survival function nonparametrically by using KM estimator. Numerical results of Zhang et al. (2017) showed that their method was better than that of Song et al. (2014), and comparable with the method of Zhou and Zhu (2017). However, all the screening methods only work for univariate survival data under independent

censoring mechanism.

In our motivating dataset from NKI cancer study, two types of endpoints subject to more complex censoring were observed, which were the time for distant metastasis-free survival (DMFS) and the time for overall survival (OS). DMFS is defined as the duration from the initial diagnosis to the time at which a distant metastasis was detected and OS is defined as the duration from initial diagnosis to death. Distant metastasis is a nonterminal event and death is a terminal event. The time to the nonterminal event may be censored if the terminal event occurs earlier but not vice versa. This leads to “semi-competing risks” and the censoring of the nonterminal event time by the terminal event time forms a type of dependent/informative censoring. Thus the aforementioned methods designed for univariate survival data under independent censoring can not be used to semi-competing risks data subject to dependent censoring.

Our aim of Chapter 4 lies in establishing an effective screening procedure for ultrahigh dimensional features associated with both the nonterminal event (e.g. Distant metastasis) and the terminal event (e.g. death), so that it is possible to jointly model the two types of events on the important gene features. Cai et al. (2005) proposed a model-based method, which is a penalized pseudo-partial likelihood method under the marginal Cox PH models of variable selection for multivariate failure time data subject to competing risks. However, the method of Cai et al. (2005) could not be applied to semi-competing risks data subject to dependent censoring because it only addressed independent censoring issue. Since it is quite challenging to specify appropriate model forms to the two event times and justify them in practice under such a complex censoring scheme; even if the model is specified, estimating model parameters is also challenging and computationally burdensome due to complex censoring scheme and ultrahigh dimensionality, it is desirable to develop a more flexible approach regardless the model structures, which is robust to model misspecification and in the meantime incorporates the relationship between two event times. Our proposal is a novel

model-free screening procedure, which utilizes the rank of correlation between the covariates and the joint survival distribution to measure the importance of the covariates contributed to the distribution of both event times. Unlike the model-based methods, the proposed screening method is more efficient as it only needs the computation of nonparametric KM estimator for the joint survival function.

Existing methods for marginally selecting variables in survival models have been applied to semi-competing risks data in the literature. For example, one could separately choose one set of covariates for predicting the nonterminal event times, and another set of covariates for predicting the terminal event times as previously proposed by Emura et al. (2018). However, this way ignores the fact of dependent censoring in the nonterminal event times and the influence of the nonterminal event to the terminal event in the semi-competing risks setting, and so their results of variable selection may be biased. This motivates us to propose a joint selection method in the thesis to identify important variables for both event times while incorporating the dependency between the two events.

1.7 Main contributions and outline

The main contents and contributions of this thesis are highlighted as follows.

In Chapter 2, we develop a flexible copula-based approach for analysis of clustered semi-competing risks data from a frequentist perspective. Compared with the existing multi-state modeling approach by Lee et al. (2016), the advantages of the proposed copula-based approach are three folded. First, it enables us to simultaneously and explicitly formulate both the marginal behaviors and dependence structure of two event times. Second, our method can directly measure the dependence between two event time variables using Kendall's tau. Third, it allows us to establish asymptotic theory for the resulting estimator. Particularly, our method relaxes the gamma distribution restriction on random effects, providing a more flexible modeling framework than the method of Emura et al.

(2017). We also prove the proposed estimators are consistent and asymptotically normal using the theory of martingale and counting process. Our numerical results demonstrate that the proposed estimation of parameters is robust against misspecification of the random effects distribution and the estimates of baseline hazard functions are more accurate than those obtained from the existing method.

In Chapter 3, we consider the situation in which one primary endpoint and two secondary endpoints are of interest. To the best of our knowledge, this work is the first attempt to provide approach for analyzing data from multiple endpoints under complex dependent censoring caused by two sets of semi-competing risks. Compared with the possible multi-state approach, the proposed joint analysis method has the following advantages. Firstly, it formulates both the marginal behaviors and dependence structure of three endpoints simultaneously and explicitly. Secondly, it directly measures the associations between two nonterminal endpoints and one terminal endpoint using two Kendall's taus. Thirdly, it accounts for potential association between two nonterminal endpoints by modeling the two pairs of endpoints with a shared terminal endpoint, but avoiding its explicit specification. Fourthly, it can be extended easily to deal with more complicated data structure consisting of one terminal endpoint and multiple non-terminal endpoints, and needs to fit $(2q-1)$ models for totally $q(q \geq 3)$ endpoints, while $(4q-5)$ models are required for multi-state models.

In Chapter 4, we propose a novel joint screening approach for the semi-competing risks data subject to dependent censoring mechanism with ultrahigh-dimensional covariates without any model assumption while incorporating the association between the nonterminal and the terminal endpoints. It is worth noting that the conventional screening methods (Song et al., 2014; Zhang et al., 2017; Zhou and Zhu, 2017) only addressed the cases with univariate complete/censored survival times or independent censoring. Our simulation results show that the proposed joint screening procedure performs satisfactorily regardless the type of model assumption.

We end this thesis with some concluding remarks and potential future works in the last chapter. Further theoretical details for Chapters 2-4 are provided in Appendices A, C, D. Appendix B and Appendix E include additional simulation results for Chapter 2 and Chapter 4, respectively.

Semiparametric regression analysis of clustered semi-competing risks data[†]

As addressed in Section 1.4, the main purpose of this chapter is the analysis of clustered semi-competing risks data. We propose a flexible semiparametric model and first formulate marginal distribution for each event time using the Cox PH model with shared random effects to account for association among event times within clusters. Through a copula model, it further allows us to connect the two marginal models together to form a joint model with an unspecified parameter for association between the nonterminal and terminal events. Particularly, our method does not restrict random effects following a gamma distribution, providing a more flexible modelling framework.

We first use a nonparametric maximum likelihood (NPML) approach to estimate the baseline cumulative hazard function by approximating it using piecewise constants, and then develop the maximum likelihood estimation (MLE) for the marginal regression parameters, variance-covariance of random effects and copula

[†]This chapter is based on the paper Peng et al. (2018): “M.J. Peng, L.M. Xiang, and S.S. Wang. Semiparametric regression analysis of clustered survival data with semi-competing risks. *Computational Statistics & Data Analysis*, 124:53–70, 2018.”

association parameter. The estimation procedure is implemented via an expectation maximization (EM) algorithm, which avoids the integral problem caused by random effects and approximates the required conditional expectation using the Markov Chain Monte Carlo (MCMC) method. Our numerical results demonstrate that the proposed estimation of parameters is robust against misspecification of the random effects distribution and the estimates of baseline hazard functions are more accurate than those obtained from the existing method.

We also establish asymptotic properties for the proposed estimators. Our estimators are shown to be consistent and asymptotically normal using techniques including martingale, the theory of counting process and martingale central limit theorem developed in Murphy (1995).

The rest of this chapter is organized as follows. Section 2.1 introduces the model, while Section 2.2 and Section 2.3 describe the proposed estimation method and a detailed algorithm for its implementation, respectively. Asymptotic properties of the proposed estimators are provided in Section 2.4. Results of simulation studies and of the real data applications are provided in Section 2.5 and Section 2.6, respectively. All technical details are relegated to Appendix A and some supplementary simulation results are provided in Appendix B.

2.1 Model Formulation

Let “event 1” be a nonterminal event, “event 2” be a terminal event, and $T^{(1)}$ and $T^{(2)}$ be the corresponding event times. We are interested in assessing the effects of covariates on the marginal distribution of $T^{(k)}$, which is modeled as

$$\lambda_k(t|b; X, Z) = \lambda_{0k}(t)e^{\beta'_k X + b'Z} \quad (2.1)$$

for $(k = 1, 2)$, where $\lambda_k(t|b; X, Z)$ is the conditional hazard function of $T^{(k)}$ given the event and covariate history before t , $\lambda_{0k}(\cdot)$ is the corresponding baseline hazard function, β_k is a vector of fixed effects, b is a vector of random effects with mean zero and density function $\psi(b; \Sigma)$ indexed by variance-covariance matrix

Σ , and X and Z are vectors of covariates associated with the fixed and random effects, respectively.

We further assume that the joint survival function of $T^{(1)}$ and $T^{(2)}$, given covariates X and Z , can be specified by a known copula subject to an unknown parameter α , which quantifies the association between $T^{(1)}$ and $T^{(2)}$:

$$S(t^{(1)}, t^{(2)}) = \Pr(T^{(1)} \geq t^{(1)}, T^{(2)} \geq t^{(2)}) = c\{S_1(t^{(1)}), S_2(t^{(2)}); \alpha\}, \quad (2.2)$$

for $t^{(1)} \geq 0, t^{(2)} \geq 0$, where $c\{u, v; \alpha\}$ is a copula function, α is a constant association parameter in one-to-one correspondence with Kendall's τ , and $S_k(\cdot)$ is the marginal survival function of $T^{(k)}$ ($k = 1, 2$). The key advantage of the copula model is that it allows us to describe the joint distribution of $T^{(1)}$ and $T^{(2)}$ through an explicit function of their marginal distributions and the corresponding association parameter.

2.2 Estimation

In this chapter, we assume that the data consist of possibly right-censored nonterminal and terminal event time observations from m clusters, with n_i observations in the cluster i for $i = 1, \dots, m$, where $N = \sum_{i=1}^m n_i$ denotes the total number of observations. To proceed with estimation, further notation is needed. Denote by $T_{ij}^{(1)}$ the nonterminal event time for subject j in cluster i and by $T_{ij}^{(2)}$ the corresponding terminal event time. The observations within a cluster are assumed to be dependent, but conditional on the cluster-specific random effect b_i the nonterminal and terminal event times $T_{ij}^{(k)}$ ($k = 1, 2$) are independent with the hazard functions following the proportional hazards model with random effects,

$$\lambda_k(t|X_{ij}, Z_{ij}, b_i) = \lambda_{0k}(t)e^{\beta'_k X_{ij} + b'_i Z_{ij}}. \quad (2.3)$$

The corresponding joint survival function of nonterminal and terminal event times for subject j in cluster i is specified as:

$$S(t_{ij}^{(1)}, t_{ij}^{(2)}) = \Pr(T_{ij}^{(1)} \geq t_{ij}^{(1)}, T_{ij}^{(2)} \geq t_{ij}^{(2)}) = c\{S_1(t_{ij}^{(1)}), S_2(t_{ij}^{(2)}); \alpha\}, t_{ij}^{(1)} \geq 0, t_{ij}^{(2)} \geq 0. \quad (2.4)$$

Let C^* denote the external censoring time that is independent of $(T^{(1)}, T^{(2)})$, and ς denote the duration of the study. Due to possible right-censoring, for subject j in cluster i we observe $\mathcal{O}_{ij} = \{T_{ij}, \tilde{T}_{ij}, X_{ij}, Z_{ij}, \delta_{ij}^{(1)}, \delta_{ij}^{(2)}\}$, where $T_{ij} = \min(T_{ij}^{(1)}, T_{ij}^{(2)}, C_{ij})$ with $C_{ij} = \min(C_{ij}^*, \varsigma)$, $\tilde{T}_{ij} = \min(T_{ij}^{(2)}, C_{ij})$, $\delta_{ij}^{(1)} = I(T_{ij} = T_{ij}^{(1)})$, $\delta_{ij}^{(2)} = I(\tilde{T}_{ij} = T_{ij}^{(2)})$, and $I(\cdot)$ is the indicator function. Note that there is no assumption that $T^{(2)} \geq T^{(1)}$ throughout the thesis. It is only assumed that $\tilde{T} \geq T$ for the observed data $(\tilde{T}, T)$, where $T = \min(T^{(1)}, T^{(2)}, C)$, $\tilde{T} = \min(T^{(2)}, C)$. Such observed data arises like that because of semi-competing risks and right censoring mechanism, in which the nonterminal event time might be censored by the terminal event time, but not vice versa. Thus for a random sample of m clusters, each of size n_i , the observed data consist of $\mathcal{O} = \{\mathcal{O}_i\}_{i=1}^m$, where $\mathcal{O}_i = \{\mathcal{O}_{ij}\}_{j=1}^{n_i}$. For each subject (i, j) , there are four possible cases that we can observe.

Case 1: $T_{ij}^{(1)} \leq C_{ij}, T_{ij}^{(2)} \leq C_{ij} \iff T_{ij}^{(1)} = T_{ij}, T_{ij}^{(2)} = \tilde{T}_{ij} \iff \delta_{ij}^{(1)}\delta_{ij}^{(2)} = 1$; which indicates that the nonterminal event occurs firstly and the terminal event occurs secondly before censoring.

Case 2: $T_{ij}^{(1)} \leq C_{ij}, T_{ij}^{(2)} > C_{ij} \iff T_{ij}^{(1)} = T_{ij}, T_{ij}^{(2)} > \tilde{T}_{ij} \iff \delta_{ij}^{(1)}(1 - \delta_{ij}^{(2)}) = 1$; which indicates that the nonterminal event occurs before censoring and the terminal event is right-censored.

Case 3: $C_{ij} < T_{ij}^{(1)}, C_{ij} < T_{ij}^{(2)} \iff T_{ij}^{(1)} > T_{ij}, T_{ij}^{(2)} > \tilde{T}_{ij} \iff (1 - \delta_{ij}^{(1)})(1 - \delta_{ij}^{(2)}) = 1$; which indicates that both the nonterminal and terminal events are right-censored.

Case 4: $T_{ij}^{(2)} \leq C_{ij}, T_{ij}^{(1)} > C_{ij} \iff T_{ij}^{(1)} > T_{ij}, T_{ij}^{(2)} = \tilde{T}_{ij} \iff (1 - \delta_{ij}^{(1)})\delta_{ij}^{(2)} = 1$; which indicates that the nonterminal event is right-censored and the terminal event occurs before censoring.

$\delta_{ij}^{(1)}\delta_{ij}^{(2)} = 1$; which indicates that the terminal events occurs firstly before censoring, and the nonterminal event is right-censored by the terminal event.

Therefore, let $\beta = (\beta_1, \beta_2)$, $\theta = (\alpha, \beta)$, $\lambda = (\lambda_{01}(\cdot), \lambda_{02}(\cdot))$, the likelihood function concerning the parameters $\omega = (\theta, \lambda, \Sigma)$ given observed data $\mathcal{O}$ is

$$L(\theta, \lambda, \Sigma|\mathcal{O}) = \prod_{i=1}^m \int_{b_i} L_i(\theta, \lambda|\mathcal{O}_i, b_i) \psi(b_i; \Sigma) db_i = \prod_{i=1}^m \int_{b_i} \prod_{j=1}^{n_i} L_{ij}(\theta, \lambda|\mathcal{O}_{ij}, b_i) \psi(b_i; \Sigma) db_i, \quad (2.5)$$

where

$$\begin{aligned} & L_{ij}(\theta, \lambda|b_i, \mathcal{O}_{ij}) \\ = & \Pr(T_{ij}^{(1)} = T_{ij}, T_{ij}^{(2)} = \tilde{T}_{ij}|b_i)^{\delta_{ij}^{(1)}\delta_{ij}^{(2)}} \cdot \Pr(T_{ij}^{(1)} = T_{ij}, T_{ij}^{(2)} > \tilde{T}_{ij}|b_i)^{\delta_{ij}^{(1)}(1-\delta_{ij}^{(2)})} \\ & \cdot \Pr(T_{ij}^{(1)} > T_{ij}, T_{ij}^{(2)} > \tilde{T}_{ij}|b_i)^{(1-\delta_{ij}^{(1)})(1-\delta_{ij}^{(2)})} \cdot \Pr(T_{ij}^{(1)} > T_{ij}, T_{ij}^{(2)} = \tilde{T}_{ij}|b_i)^{(1-\delta_{ij}^{(1)})\delta_{ij}^{(2)}}. \end{aligned}$$

Plugging models (2.3) and (2.4) in the likelihood, (2.5) can be written as:

$$\begin{aligned} L(\theta, \lambda, \Sigma|\mathcal{O}) = & \prod_{i=1}^m \int_{b_i} \prod_{j=1}^{n_i} \left[c_{uv}\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\} S_1(T_{ij}) \lambda_{01}(T_{ij}) \right. \\ & \cdot e^{\beta_1' x_{ij} + b_i' z_{ij}} S_2(\tilde{T}_{ij}) \lambda_{02}(\tilde{T}_{ij}) e^{\beta_2' x_{ij} + b_i' z_{ij}} \left. \right]^{\delta_{ij}^{(1)}\delta_{ij}^{(2)}} \\ & \cdot \left[c_u\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\} S_1(T_{ij}) \lambda_{01}(T_{ij}) e^{\beta_1' x_{ij} + b_i' z_{ij}} \right]^{\delta_{ij}^{(1)}(1-\delta_{ij}^{(2)})} \\ & \cdot \left[c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\} \right]^{(1-\delta_{ij}^{(1)})(1-\delta_{ij}^{(2)})} \\ & \cdot \left[c_v\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\} S_2(\tilde{T}_{ij}) \lambda_{02}(\tilde{T}_{ij}) e^{\beta_2' x_{ij} + b_i' z_{ij}} \right]^{(1-\delta_{ij}^{(1)})\delta_{ij}^{(2)}} \psi(b_i; \Sigma) db_i, \end{aligned} \quad (2.6)$$

where $c\{u, v\} = c\{u, v; \alpha\}$, $c_u\{u, v\} = \frac{\partial c\{u, v; \alpha\}}{\partial u}$, $c_v\{u, v\} = \frac{\partial c\{u, v; \alpha\}}{\partial v}$, $c_{uv}\{u, v\} = \frac{\partial^2 c\{u, v; \alpha\}}{\partial u \partial v}$.

Before deriving the maximum of $L(\theta, \lambda, \Sigma|\mathcal{O})$ in (2.6), we introduce some notations about counting processes here. Let $\Delta_{ij}(t) = I(T_{ij} > t)$, $\tilde{\Delta}_{ij}(t) =$

$I(\tilde{T}_{ij} > t)$, $N_{ij}^{(1)}(t) = \delta_{ij}^{(1)}(1 - \Delta_{ij}(t))$ be the counting process of event 1, $N_{ij}^{(2)}(t) = \delta_{ij}^{(2)}(1 - \tilde{\Delta}_{ij}(t))$ be the counting process of event 2. Also we assume that $\Pr(T_{ij}^{(1)} = T_{ij}^{(2)}) = 0$ for $i = 1, \dots, m$ and $j = 1, \dots, n_i$.

The maximum of $L(\theta, \lambda, \Sigma | \mathcal{O})$ in (2.6) does not exist if $\Lambda_{01}(\cdot)$ and $\Lambda_{02}(\cdot)$ are restricted to be absolutely continuous. By allowing $\Lambda_{01}(\cdot)$ and $\Lambda_{02}(\cdot)$ to be increasing right-continuous functions with the forms $\Lambda_{01}(t) = \int_0^t \lambda_{01}(t) dt$ and $\Lambda_{02}(t) = \int_0^t \lambda_{02}(t) dt$, we replace $\lambda_{01}(t)$ and $\lambda_{02}(t)$ in (2.6) with the jump sizes of Λ_{01} and Λ_{02} at time t denoted by $\Lambda_{01}\{t\}$ and $\Lambda_{02}\{t\}$, respectively. By letting the derivative of the modified complete likelihood with respect to $\Lambda_{01}\{T_{ij}\}$ and $\Lambda_{02}\{\tilde{T}_{ij}\}$ to 0, we obtain the nonparametric maximum likelihood estimates

$$\hat{\Lambda}_{01}\{T_{ij}\} = \frac{dN_{ij}^{(1)}(T_{ij})}{\sum_{(g,h) \in R(T_{ij})} e^{\beta_1' X_{gh}} E\{w_1(T_{gh}) e^{b_g' Z_{gh}}\}}, \quad (2.7)$$

$$\hat{\Lambda}_{02}\{\tilde{T}_{ij}\} = \frac{dN_{ij}^{(2)}(\tilde{T}_{ij})}{\sum_{(g,h) \in \tilde{R}(\tilde{T}_{ij})} e^{\beta_2' X_{gh}} E\{w_2(\tilde{T}_{gh}) e^{b_g' Z_{gh}}\}}, \quad (2.8)$$

where for $(g, h) \in R(T_{ij})$ or $(g, h) \in \tilde{R}(\tilde{T}_{ij})$, i.e., $T_{gh} \leq T_{ij}$ or $\tilde{T}_{gh} \leq \tilde{T}_{ij}$, $R(t)$ is the at-risk set at time t for the nonterminal event, $\tilde{R}(t)$ is the at-risk set at time t for the terminal event, $E\{\cdot\}$ is the expectation with respect to b_g ,

$$\begin{aligned} w_1(T_{gh}) &= \delta_{gh}^{(1)} \delta_{gh}^{(2)} \left\{ \frac{D_{121}(T_{gh}, \tilde{T}_{gh})}{D_{12}(T_{gh}, \tilde{T}_{gh})} S_1(T_{gh}) + 1 \right\} + \delta_{gh}^{(1)} (1 - \delta_{gh}^{(2)}) \left\{ \frac{D_{11}(T_{gh}, \tilde{T}_{gh})}{D_1(T_{gh}, \tilde{T}_{gh})} \right. \\ &\quad \cdot S_1(T_{gh}) + 1 \left. \right\} + (1 - \delta_{gh}^{(1)}) (1 - \delta_{gh}^{(2)}) \left\{ \frac{D_1(T_{gh}, \tilde{T}_{gh}) \cdot S_1(T_{gh})}{c\{S_1(T_{gh}), S_2(\tilde{T}_{gh})\}} \right\} \\ &\quad + (1 - \delta_{gh}^{(1)}) \delta_{gh}^{(2)} \left\{ \frac{D_{21}(T_{gh}, \tilde{T}_{gh})}{D_2(T_{gh}, \tilde{T}_{gh})} S_1(T_{gh}) \right\}, \\ w_2(\tilde{T}_{gh}) &= \delta_{gh}^{(1)} \delta_{gh}^{(2)} \left\{ \frac{D_{122}(T_{gh}, \tilde{T}_{gh})}{D_{12}(T_{gh}, \tilde{T}_{gh})} S_2(\tilde{T}_{gh}) + 1 \right\} + \delta_{gh}^{(1)} (1 - \delta_{gh}^{(2)}) \left\{ \frac{D_{12}(T_{gh}, \tilde{T}_{gh})}{D_1(T_{gh}, \tilde{T}_{gh})} \right. \\ &\quad \cdot S_2(\tilde{T}_{gh}) \left. \right\} + (1 - \delta_{gh}^{(1)}) (1 - \delta_{gh}^{(2)}) \left\{ \frac{D_2(T_{gh}, \tilde{T}_{gh}) \cdot S_2(\tilde{T}_{gh})}{c\{S_1(T_{gh}), S_2(\tilde{T}_{gh})\}} \right\} \\ &\quad + (1 - \delta_{gh}^{(1)}) \delta_{gh}^{(2)} \left\{ \frac{D_{22}(T_{gh}, \tilde{T}_{gh})}{D_2(T_{gh}, \tilde{T}_{gh})} S_2(\tilde{T}_{gh}) + 1 \right\}, \end{aligned}$$

and we resort to the notation

$$\begin{aligned}
D_1(t^{(1)}, t^{(2)}) &= \frac{\partial c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)})}, & D_2(t^{(1)}, t^{(2)}) &= \frac{\partial c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_2(t^{(2)})}, \\
D_{11}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial^2 S_1(t^{(1)})}, & D_{12}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)}) \partial S_2(t^{(2)})}, \\
D_{21}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_2(t^{(2)}) \partial S_1(t^{(1)})}, & D_{22}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial^2 S_2(t^{(2)})}, \\
D_{121}(t^{(1)}, t^{(2)}) &= \frac{\partial^3 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)}) \partial S_2(t^{(2)}) \partial S_1(t^{(1)})}, & D_{122}(t^{(1)}, t^{(2)}) &= \frac{\partial^3 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)}) \partial S_2(t^{(2)}) \partial S_2(t^{(2)})}.
\end{aligned} \tag{2.9}$$

Thus the cumulative hazard functions for the nonterminal and terminal events are estimated by the following piece-wise constant functions

$$\hat{\Lambda}_{01}(t) = \sum_{(i,j) \in R(t)} \hat{\Lambda}_{01}\{T_{ij}\}, \tag{2.10}$$

$$\hat{\Lambda}_{02}(t) = \sum_{(i,j) \in \tilde{R}(t)} \hat{\Lambda}_{02}\{\tilde{T}_{ij}\}, \tag{2.11}$$

respectively. Plugging $\lambda_{01}(T_{ij}), \lambda_{02}(\tilde{T}_{ij})$ by jump sizes $\hat{\Lambda}_{01}\{T_{ij}\}$ and $\hat{\Lambda}_{02}\{\tilde{T}_{ij}\}$, respectively, into the modified likelihood (2.6), the updated value of θ can be obtained through maximizing the likelihood.

2.3 Implementation and algorithm

To implement the proposed NPML estimation, we rely on the Monte Carlo Expectation Maximization (MCEM) algorithm (Wei and Tanner, 1990; Levine and Casella, 2001), in which random effects are treated as missing values to avoid complicated numerical integrations. Particularly, the expectation in the E-step of the algorithm is computed numerically through the MCMC method, such as the acceptance-rejection sampling, while the M-step maximizes the approximated expectation to update parameter estimates.

Let $\hat{\Lambda} = (\hat{\Lambda}_{01}, \hat{\Lambda}_{02})$. The complete likelihood corresponding to (2.6) is written as:

$$L(\theta|\mathcal{O}, \hat{\Lambda}, \hat{\Sigma}) = \prod_{i=1}^m L_i(\theta|\mathcal{O}_i, \hat{\Lambda}, \hat{\Sigma}) = \prod_{i=1}^m \int_{b_i} L_i(\theta|\mathcal{O}_i, \hat{\Lambda}, b_i) \psi(b_i; \hat{\Sigma}) db_i, \quad (2.12)$$

where

$$\begin{aligned} L_i(\theta|\mathcal{O}_i, \hat{\Lambda}, b_i) = & \prod_{j=1}^{n_i} \left[D_{12}(T_{ij}, \tilde{T}_{ij}) \hat{S}_1(T_{ij}) \cdot \hat{\Lambda}_{01}\{T_{ij}\} e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \hat{S}_2(\tilde{T}_{ij}) \hat{\Lambda}_{02}\{\tilde{T}_{ij}\} \right. \\ & \cdot e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \left. \right]^{\delta_{ij}^{(1)} \delta_{ij}^{(2)}} \cdot \left[D_1(T_{ij}, \tilde{T}_{ij}) \hat{S}_1(T_{ij}) \hat{\Lambda}_{01}\{T_{ij}\} \right. \\ & \cdot e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \left. \right]^{\delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)})} \cdot \left[c\{\hat{S}_1(T_{ij}), \hat{S}_2(\tilde{T}_{ij})\} \right]^{(1 - \delta_{ij}^{(1)}) (1 - \delta_{ij}^{(2)})} \\ & \cdot \left[D_2(T_{ij}, \tilde{T}_{ij}) \hat{S}_2(\tilde{T}_{ij}) \hat{\Lambda}_{02}\{\tilde{T}_{ij}\} e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \right]^{(1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)}}. \end{aligned}$$

In the E-step, we compute the expectation of the complete likelihood (2.12) given the observed data and current estimates $\hat{\omega}^{[k]} = (\hat{\theta}^{[k]}, \hat{\Lambda}^{[k]}, \hat{\Sigma}^{[k]})$, i.e.

$$E[L_i(\theta|\mathcal{O}_i, \hat{\Lambda}^{[k]})] = \int_{b_i} L_i(\theta|\mathcal{O}_i, \hat{\Lambda}^{[k]}, b_i) f(b_i|\mathcal{O}_i, \hat{\omega}^{[k]}) db_i. \quad (2.13)$$

The calculation is challenging as the expectation has no closed form expression in the semi-competing risks setting. We use the MCMC method to approximate the expectation. We first generate a MCMC sample of size Q , denoted by $\{b_i^q, q = 1, \dots, Q\}$, following density $f(b_i|\mathcal{O}_i, \hat{\omega}^{[k]})$ for $i = 1, \dots, m$ using the acceptance-rejection sampler, and then estimate the expectation (2.13) by empirical (sample) mean. Details are given as follows:

MCMC-step 1. Compute point $\tilde{b}_i$ through maximizing $f(b_i|\mathcal{O}_i, \hat{\omega}^{[k]})$ with respect to b_i ,

$$f(b_i|\mathcal{O}_i, \hat{\omega}^{[k]}) = \frac{f(b_i, \hat{\omega}^{[k]}|\mathcal{O}_i)}{f(\hat{\omega}^{[k]}|\mathcal{O}_i)} = \frac{f(\hat{\omega}^{[k]}|\mathcal{O}_i, b_i) \cdot \psi(b_i; \hat{\Sigma}^{[k]})}{\int_{b_i} f(\hat{\theta}^{[k]}, \hat{\Lambda}^{[k]}|\mathcal{O}_i, b_i) \cdot \psi(b_i; \hat{\Sigma}^{[k]}) db_i}.$$

It is equivalent to maximizing $f(\hat{\omega}^{[k]}|\mathcal{O}_i, b_i) \cdot \psi(b_i; \hat{\Sigma}^{[k]})$ with respect to b_i , where $f(\hat{\omega}^{[k]}|\mathcal{O}_i, b_i) = L_i(\hat{\theta}^{[k]}|\mathcal{O}_i, \hat{\Lambda}^{[k]}, b_i)$. By the Newton-Raphson method, we can obtain the value of $\tilde{b}_i$ and its inverse Hessian matrix $I^{-1}(\tilde{b}_i)$.

MCMC-step 2. Use the acceptance-rejection method to generate random numbers from a normal distribution with mean $\tilde{b}_i$ and variance $I^{-1}(\tilde{b}_i)$. That is, $b_i^q \sim N(\tilde{b}_i, I^{-1}(\tilde{b}_i))$ with pdf $\phi(\tilde{b}_i; I^{-1}(\tilde{b}_i))$ for $i = 1, \dots, m$ and $q = 1, \dots, Q$. More specifically,

$$\text{if } u_i \leq \frac{f(b_i^q | \mathcal{O}_i, \hat{\omega}^{[k]})}{M_i \cdot \phi(b_i^q; I^{-1}(\tilde{b}_i))}, \text{ then accept } b_i^q; \text{ else reject } b_i^q.$$

where u_i follows uniform $[0, 1]$ distribution and M_i is constant specified. We use

$$M_i = \max_q \left\{ f(b_i^q | \mathcal{O}_i, \hat{\omega}^{[k]}) / \phi(b_i^q; I^{-1}(\tilde{b}_i)) \right\}$$

in our simulation studies.

MCMC-step 3. Based on the MCMC sample, the expectation can be then approximated by

$$\hat{E}\{L_i(\theta | \mathcal{O}_i, \hat{\Lambda}^{[k]})\} = \frac{1}{Q} \sum_{q=1}^Q L_i(\theta | \mathcal{O}_i, \hat{\Lambda}^{[k]}, b_i^q). \quad (2.14)$$

MCMC-step 4. $\hat{\Sigma}^{[k+1]}$ can be computed by the sample variance or variance-covariance matrix of MCMC sample $\{b_i^q, q = 1, \dots, Q\}$ for $i = 1, \dots, m$.

In the M-step, we obtain updated parameter estimates $\hat{\theta}^{[k+1]}$ through maximizing the expected complete likelihood

$$\hat{E}\{L(\theta | \mathcal{O}_i, \hat{\Lambda}^{[k]})\} = \prod_{i=1}^n \hat{E}\{L_i(\theta | \mathcal{O}_i, \hat{\Lambda}^{[k]})\}$$

using numerical optimization algorithm. Given the estimates $\hat{\theta}^{[k+1]}$ we can consequently compute the updated estimate of Λ , $\hat{\Lambda}^{[k+1]}$, based on the NPML estimates in (2.10)-(2.11).

Starting with an initial set of estimates $\hat{\omega}^{[0]} = (\hat{\theta}^{[0]}, \hat{\Lambda}^{[0]}, \hat{\Sigma}^{[0]})$, the E-step and M-step of the algorithm are iterated until convergence is achieved. That is, the differences between subsequent estimates of (θ, Σ) are all less than or the score

function of θ is less than a small quantity ϵ . The resulting estimators at convergence are denoted by $\hat{\omega} = (\hat{\theta}, \hat{\Lambda}, \hat{\Sigma})$. We implement the proposed algorithm in R and use R function *optim* to perform numerical optimizations involved. Our simulation results show that the algorithm can find good solutions in reasonable computational time. In the case of $N = 200, m = 10$ and $n_i = 20$, for example, it takes 30 seconds for the calculation of all estimates.

It is worth mentioning that the performance of the MCEM algorithm proposed depends on how well the MCMC sample approximates to the target distribution. To ensure a good approximation, one can do the following: (1) choose accurate starting values; (2) check if multiple MCMC samples initialized from different initial values give similar results; (3) make sure to run MCMC samples with large size. In this chapter, we take the initial values of parameters β and Λ to be the corresponding estimates obtained in the two marginal models using R function *phmm*, while initial values of the variance component are set to be the average of the estimated frailty variances in both marginal models. α is initially set to be 1. Our experience in numerical studies shows that these initial values work well and the convergence of the proposed MCEM algorithm is not sensitive to the choice of starting value of α . In addition, we use the MCMC sample size as 100, which produces simulation results similar to those obtained with bigger MCMC samples of size, say, 200.

Another possible problem of using MCMC sampling to approximate the conditional expectation is that the likelihood based on the MCMC sample may not keep increasing in the EM algorithm. To overcome this problem, a data-driven strategy proposed in Caffo et al. (2005) can be adopted for recovering EM's ascent (i.e. likelihood increasing) property with high probability. For more details on the data-driven automated MCEM algorithm, we refer the interested reader to Caffo et al. (2005).

2.4 Theoretical result

We study the large sample properties of the proposed estimators in the setting where the random effects are assumed to follow a normal or log-gamma distribution with mean 0 and some variance. The normal and log-gamma distributions are two most popular assumptions for the random effects distribution in the literature though other parametric distribution assumptions about random effects are possible.

Before presenting the asymptotic properties of the proposed estimators $\hat{\omega} = (\hat{\theta}, \hat{\Lambda}, \hat{\Sigma})$, we first list the following conditions.

Condition 2.1. *Conditional on the covariates X_{ij} and Z_{ij} , the latent censoring times C_{ij} are independent of the event times $T_{ij}^{(k)}$ ($k = 1, 2$) and the random effects b_i .*

Condition 2.2. *The covariates X and Z are bounded.*

Condition 2.3. *The true values of θ and Σ are elements of the interior of a known compact set $\mathcal{K} = \{(\theta, \Sigma) : |\theta| \leq B \text{ for some constant } B, |\theta| \text{ is the } l_1\text{-norm of } \theta, \text{ and } \Sigma \text{ is symmetric and positive definite with eigenvalues bounded away from } 0 \text{ and } \infty.\}$*

Condition 2.4. *With probability 1, for $i = 1, \dots, m$ and $j = 1, \dots, n_i$, $\Pr(T_{ij} > \varsigma | X, Z) > 0$, $\Pr(\delta_{ij}^{(1)} = 0, \delta_{ij}^{(2)} = 0, T_{ij} = \varsigma | X, Z) > 0$, $\Pr(\tilde{T}_{ij} > \varsigma | X, Z) > 0$, $\Pr(\delta_{ij}^{(1)} = 0, \delta_{ij}^{(2)} = 0, \tilde{T}_{ij} = \varsigma | X, Z) > 0$.*

Condition 2.5. *The copula c has bounded partial derivatives in interval $[0, 1]$.*

Condition 2.6. *The function $\Lambda_{0k}(\cdot)$ ($k = 1, 2$) is strictly increasing and continuously differentiable.*

Condition 2.7. *If $a(n_i, \sum_{j=1}^{n_i} X'_{ij}) = 0$ exists for all $i = 1, \dots, m$, where a is a vector, then $a = 0$.*

Condition 2.8. *The number of clusters is larger than $1+p+q(q+1)/2$, and from all clusters, we can select $1+p+q(q+1)/2$ clusters denoted by $i = 1, \dots, (1+p+q(q+1)/2)$ such that the determinant of matrix $M_{(1+p+q(q+1)/2) \times (1+p+q(q+1)/2)}$ is not zero, where the i th row of M is $\left(n_i, \sum_{j=1}^{n_i} X_{ij}, \text{ the triangle elements of matrix } (\sum_{j=1}^{n_i} Z_{ij})'(\sum_{j=1}^{n_i} Z_{ij})\right)$, p is the dimension of X_{ij} , and q is the dimension of Z_{ij} .*

Condition 2.1 is standard and required for identifiability in the presence of censoring. Conditions 2.2-2.5 ensure the boundedness of $\hat{\Lambda}$ and the compactness of θ . Conditions 2.1-2.3 are used in Gamst et al. (2009). Condition 2.4 is used in Chen (2012). Condition 2.5 is an extra condition to ensure the boundedness of $\hat{\Lambda}$ in the presence of copula. Condition 2.6 is required to conduct nonparametric maximum likelihood estimation for Λ , which is used in Gamst et al. (2009) and Chen (2012). Conditions 2.7-2.8 are needed for identifiability in the proposed model, which involves two sets of regression parameters. Conditions 2.7-2.8 are similar to condition C7 in Gamst et al. (2009), which aim to avoid collinearity among the covariates, ensure that the semiparametric Fisher information operator is one-to-one and imply that the likelihood has a unique maximum. Overall, Conditions 2.1-2.8 are standard and realizable.

The consistency and the asymptotic normality of estimators $\hat{\theta}_m$, $\hat{\Lambda}_{mk}$ and $\hat{\Sigma}_m$ are established in the following two theorems, where $\hat{\theta}_m$, $\hat{\Lambda}_{mk}$, $\hat{\Sigma}_m$ are the proposed estimator for θ , Λ , Σ , respectively corresponding to the number of clusters m .

Theorem 2.1. *(Consistency) (1) Assume that Conditions 2.1-2.8 hold and random effects b_i follow a normal distribution $N(0, \Sigma_0)$ i.i.d. Then $|\hat{\theta}_m - \theta_0| \rightarrow 0$, $\|\hat{\Lambda}_{mk} - \Lambda_{0k}\|_{l^\infty[0, \varsigma]} \rightarrow 0$ ($k = 1, 2$) and $|\hat{\Sigma}_m - \Sigma_0| \rightarrow 0$ as $m \rightarrow \infty$, where $(\theta_0, \Lambda_{0k}, \Sigma_0)$ are true values of parameters and $\|\cdot\|_{l^\infty[0, \varsigma]}$ is the supremum norm in the interval $[0, \varsigma]$.*

(2) Assume that Conditions 2.1-2.7 hold, $Z_{ij} = 1$ and $u_i = \exp(b_i)$ follow a gamma distribution $\Gamma(1/\eta_0, \eta_0)$ i.i.d. with mean 1 and variance η_0 for $i = 1, \dots, m, j = 1, \dots, n_i$. Then $|\hat{\theta}_m - \theta_0| \rightarrow 0$, $\|\hat{\Lambda}_{mk} - \Lambda_{0k}\|_{l^\infty[0, \varsigma]} \rightarrow 0$ ($k = 1, 2$) and $|\hat{\eta}_m - \eta_0| \rightarrow 0$ as $m \rightarrow \infty$, where $(\theta_0, \Lambda_{0k}, \eta_0)$ are true values of parameters.

Theorem 2.2. (*Asymptotic Normality*) (1) Assume that Conditions 2.1-2.8 hold and random effects b_i follow a normal distribution $N(0, \Sigma_0)$ i.i.d. Then $\sqrt{m} \times (\hat{\theta}_m - \theta_0, \hat{\Lambda}_m - \Lambda_0, \hat{\Sigma}_m - \Sigma_0)$ converges weakly to a mean-0 Gaussian process as $m \rightarrow \infty$.

(2) Assume that Conditions 2.1-2.7 hold, $Z_{ij} = 1$ and $u_i = \exp(b_i)$ follow a gamma distribution $\Gamma(1/\eta_0, \eta_0)$ i.i.d. with mean 1 and variance η_0 for $i = 1, \dots, m, j = 1, \dots, n_i$. Then $\sqrt{m} \times (\hat{\theta}_m - \theta_0, \hat{\Lambda}_m - \Lambda_0, \hat{\eta}_m - \eta_0)$ converges weakly to a mean-0 Gaussian process as $m \rightarrow \infty$.

The detailed proofs of both theorems are provided in Appendix A. In particular, we apply the techniques of counting process and the strong law of large numbers to prove Theorem 2.1. We first invoke the compactness of the parameter space and Helly's selection theorem (Murphy, 1995) to conclude the existence of a convergent subsequence, which converges to $\omega^* = (\theta^*, \Lambda^*, \Sigma^*)$, of our estimators. By Glivenko-Cantelli Theorem (Gamst et al., 2009) and the property of Kullback-Leibler information, we have $L(\omega^*) = L(\omega_0)$. We then prove the identifiability, that is, $\omega^* = \omega_0$. Since our model involves two sets of regression parameters and three counting processes, the identifiability issue is much more challenging than the case with only one counting process. Given that the equality $L(\omega^*) = L(\omega_0)$ is satisfied for all the cases, it suffices to prove $\omega^* = \omega_0$ under some special cases. Therefore, we impose new Conditions 2.7-2.8 to conclude our estimators converge to true values.

To prove the asymptotic normality in Theorem 2.2, we first apply the martingale central limit theorem (Pollard (1984), chap. 8) to show that the derivative of $L_m(\omega_0)$, denoted by $\sqrt{m}S_m(\omega_0)$, converges in distribution to a mean-zero Gaussian process. Fréchet differentiability of score function $S(\omega)$ holds by its smoothness. Then we prove the derivative of the score function, denoted by $\dot{S}(\omega_0)$, is continuously invertible on its range under Conditions 2.7-2.8. By Theorem 2.2 in Murphy (1995), the asymptotic normality of our estimators follows from the verified four conditions on the score function: convergence to a tight Gaussian

process, Fréchet differentiability, invertibility, and the approximation condition.

It is noted that in the proofs of Theorems 2.1-2.2, we have provided a general way to characterize the distribution of random effects b_i and expressed the normal and log-gamma random effects as two special cases. In this way, the theoretical properties can be extended to incorporate other parametric distribution assumptions for random effects under conditions that their corresponding moments required in the proof procedures can be explicitly calculated and identifiability Conditions 2.7-2.8 need to be updated accordingly.

2.5 Simulation Study

In this section, we assess the performance of our proposed method via finite sample simulation studies and compare our proposed method with existing method.

We generate data for the two event times $T^{(1)}$ and $T^{(2)}$ from the proportional hazards models with random effects $\lambda_1(t) = \lambda_{01}(t)\exp\{\beta_1 X + bZ\}$ and $\lambda_2(t) = \lambda_{02}(t)\exp\{\beta_2 X + bZ\}$, respectively, where $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$, which correspond to two exponential marginal distributions. X follows a standard normal distribution truncated at ± 2 . We consider two different distributions for random effects: b follows a normal distribution with mean 0 and variance η and $u = \exp(b)$ follows a gamma distribution $\Gamma(1/\eta, \eta)$ with mean 1 and variance η . The joint distribution of $(T^{(1)}, T^{(2)})$ is specified by a Gumbel copula with association parameter $\alpha = 1/(1 - \tau) - 1$ where Kendall's tau is set to be $\tau = 0.3, 0.6$. We consider the Gumbel copula model only in the simulation studies. Literature show that the maximum likelihood estimators are not sensitive to the choice of copula. For example, Chen (2012) addressed this point when using the maximum likelihood approach to analyze semi-competing risks data, where the copula and/or marginal distribution of event time are misspecified. His results showed that the maximum likelihood analysis performs quite robust to the misspecification of the copula type and/or the marginal model. Heuchenne

et al. (2014) showed that the likelihood-based inference for semi-competing risks data are robust to different choice of copulas in simulation studies. We generate censoring times C from a uniform distribution $U[0, \varsigma]$, where the duration of the study is chosen to be $\varsigma = 4$. The observed nonterminal and terminal event times are obtained by $\min(T^{(1)}, T^{(2)}, C)$ and $\min(T^{(2)}, C)$, respectively. This results in 30% – 60% censoring for $T^{(1)}$ and 20% – 50% censoring for $T^{(2)}$.

In order to evaluate the finite sample performance of the proposed estimators with different variance component levels and show that it can handle the random effect in the intercept or slope, we consider the following scenarios:

Scenario 2.1: $\eta = 0.2, Z = 1$

Scenario 2.2: $\eta = 0.4, Z = 1$

Scenario 2.3: $\eta = 0.2, Z \sim U(-2, 2)$

Scenario 2.4: $\eta = 0.4, Z \sim U(-2, 2)$

Scenarios 2.1 and 2.2 present the random intercept cases, while Scenarios 2.3 and 2.4 allow the random effect on slope, where the variance component changes from 0.2 to 0.4. For each scenario, we conduct 200 simulation runs and summarize parameter estimates in terms of the mean bias (average of the difference between estimates and true values), the sample standard deviation of estimates (sd), the root mean square error (rmse) and the coverage percentage for the 95% confidence intervals (cp). Note that the bias and sd describe how far off the estimates is from the true value in average and how widely spread the estimates are, respectively, while the rmse measures the underlying estimator's performance by combining information of both bias and sd together. Simulation results are reported in Tables 2.1-2.4 for models with normally distributed random effects, i.e. $b_i \stackrel{\text{i.i.d}}{\sim} N(0, \eta)$, in Scenarios 2.1-2.4, respectively. In each scenario, the mean biases of the proposed estimators are close to zero with small sd's, small rmse's and corresponding cp nearly 95% for all settings. In general, when the cluster size n_i

for $i = 1, \dots, m$ or the number of clusters m increases, both the mean bias and sd of the proposed estimators reduce, indicating the improvements in accuracy and efficiency, respectively. For the estimate of τ , the reduction in relative bias (bias/true value) and sd tends to be more significant when the true value of τ becomes larger or association between both events becomes stronger. This implies that the proposed copula-based modeling approach can better capture the association between two events, which are highly dependent, and thus provides more accurate estimates of τ .

Comparing the performance between Table 2.1 and Table 2.2, or Table 2.3 and Table 2.4, for two different levels of the variance of random effects, we observe that with a greater variance component the aforementioned findings still hold, but resulting in relatively larger mean biases, sd's, rmse's as expected. However, when the sample size increases, for example from $N = 50$ to $N = 200$ in Table 2.2 and Table 2.4, the mean biases and sd's tend to be very small again and thus the differences in estimation induced by the change of the variance component can be ignored.

Next we consider models with gamma frailty $u_i = \exp(b_i) \sim \Gamma(1/\eta, \eta)$ i.i.d. For brief illustration, we here report simulation results in Tables 2.5-2.8 for Scenarios 2.1-2.2. Findings are similar to those with normal random effects shown in Tables 2.1-2.2. Results for Scenarios 2.3-2.4 are similar to those in Tables 2.3-2.4 and the corresponding results are not presented due to space limit. Since Emura et al. (2017)'s approach can be applied to the case with gamma frailty, we also present the corresponding results obtained by their approach using R package *joint.Cox* in Tables 2.5-2.8. The proposed method yields better estimates of variance component η and Kendall's τ than their approach in terms of both mean bias and sd, while estimates of regression coefficients β_1 and β_2 obtained from both methods are comparable in terms of rmse for some small sample sizes. When the total sample size increases to $N = 200$, our method outperforms theirs in terms of rmse for estimating all parameters in Scenario 2.1, regardless of association level.

With a large variance component in Scenario 2.2, we note from Tables 2.7-2.8 that the proposed method leads to slightly worse estimates of η and comparable estimates for the other parameters in comparison with the existing counterpart. As association τ increases, our estimates of η and τ significantly improve in terms of rmse.

To visualize the performance of estimated cumulative baseline hazard functions, we show curves of $\hat{\Lambda}_{01}(\cdot)$ and $\hat{\Lambda}_{02}(\cdot)$ obtained by the proposed method in Figures 2.1-2.2, respectively, based on the first 20 simulation runs in Scenario 2.1. The straight black lines in figures represent the true cumulative baseline hazard function and the red curves are the estimated cumulative baseline hazard functions. The corresponding curves estimated by the spline-based method of Emura et al. (2017) are also shown in Figures 2.3-2.4 for comparison purpose. Across all the settings, the proposed method provides the estimated cumulative baseline hazard functions close to the corresponding true function and the biases reduce as the number of clusters and/or the cluster size increases. Compared to the corresponding curves in Figures 2.3-2.4, the proposed method yields the estimates of the cumulative baseline hazards similar to those smooth estimates. When the sample size becomes large, our method slightly improves the performance of the cumulative baseline function estimation to some extent.

Moreover, we also examine the performance of the proposed estimators under a misspecified distribution of random effects. Table 2.9 reports estimation results under the proposed model with normally distributed random effects in Scenario 2.1 when the true distribution of random effects is log-gamma, that is $\exp(b_i)$ follows a gamma distribution. Comparing to the corresponding results in Table 2.5 and Table 2.6 with the correctly specified distribution of random effects, there is limited impact of such a misspecification on estimation efficiency of variance component η only especially when the value of τ is large, while estimation for the other parameters seems not sensitive to the misspecification. In addition, when the true distribution of random effects is a mixture of normal

distributions $0.5N(0, 0.2) + 0.5N(0, 0.4)$, we provide estimation results in Tables 2.10-2.11 obtained by both the proposed and existing methods with misspecified gamma distributed random effects. It is found that the existing method fails to provide valid estimates of all parameters, with over 30% divergent cases and unreasonably big biases and sd's for those convergent cases most likely due to some outliers caused by the mixture normal structure of the random effects. On the other hand, except for having slightly greater sd's for estimates of η , our method still works well in estimating parameters. These indicate that the proposed method is more robust against outliers and random effects distribution misspecification than the existing approach.

We have also considered the Weibull baseline hazard functions for the two marginal models. The corresponding simulation results are provided in Appendix B due to space limit, and exhibit similar trends to those seen in the exponential marginal model design above.

Table 2.1. Simulation results based on 200 replications under the model with normally distributed random effects in the intercept in Scenario 2.1, where parameters $\eta = 0.2$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6.

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	-0.0632	0.0196	0.0185	0.0402
	sd	0.1834	0.1063	0.2846	0.3062
	rmse	0.1935	0.1079	0.3086	0.2856
	cp(%)	95.0	94.0	94.5	95.0
(100, 5, 20)	bias	-0.0406	-0.0100	0.0206	0.0269
	sd	0.1613	0.0854	0.1818	0.2125
	rmse	0.1661	0.0856	0.1832	0.2147
	cp(%)	96.0	95.5	94.0	96.0
(100, 10, 10)	bias	-0.0054	0.0190	-0.0007	0.0200
	sd	0.1477	0.0799	0.1913	0.1846
	rmse	0.1481	0.0822	0.1910	0.1861
	cp(%)	96.5	95.5	93.5	94.5
(200, 10, 20)	bias	-0.0207	0.0182	-0.0046	0.0084
	sd	0.1250	0.0577	0.1200	0.1295
	rmse	0.1268	0.0607	0.1201	0.1302
	cp(%)	96.5	95.0	95.5	93.5
$\tau = 0.6$					
(50, 5, 10)	bias	-0.0102	-0.0046	0.0797	0.0498
	sd	0.2911	0.0963	0.6790	0.4672
	rmse	0.2912	0.0961	0.6837	0.4697
	cp(%)	98.5	96.5	99.0	98.5
(100, 5, 20)	bias	-0.0533	0.0023	0.0427	-0.0198
	sd	0.1495	0.0563	0.1860	0.1646
	rmse	0.1591	0.0560	0.1909	0.1662
	cp(%)	96.0	95.0	93.0	95.0
(100, 10, 10)	bias	-0.0270	-0.0016	0.0465	-0.0110
	sd	0.1607	0.0571	0.1908	0.1901
	rmse	0.1632	0.0570	0.1967	0.1903
	cp(%)	97.0	94.5	93.0	93.0
(200, 10, 20)	bias	-0.0251	0.0006	0.0385	-0.0214
	sd	0.1177	0.0388	0.1347	0.1246
	rmse	0.1206	0.0390	0.1405	0.1268
	cp(%)	96.0	94.0	94.0	94.5

Table 2.2. Simulation results based on 200 replications under the model with normally distributed random effects in the intercept in Scenario 2.2, where parameters $\eta = 0.4$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6.

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	-0.1181	0.0112	0.0541	0.1012
	sd	0.3187	0.1108	0.2553	0.4445
	rmse	0.3399	0.1114	0.2610	0.4559
	cp(%)	96.5	94.5	93.5	98.5
(100, 5, 20)	bias	-0.0751	0.0135	-0.0192	0.0319
	sd	0.2785	0.0846	0.1657	0.2039
	rmse	0.2884	0.0857	0.1668	0.2064
	cp(%)	95.0	96.0	95.5	95.0
(100, 10, 10)	bias	-0.0540	0.0205	-0.0149	0.0183
	sd	0.2242	0.0777	0.1835	0.2161
	rmse	0.2306	0.0804	0.1841	0.2169
	cp(%)	95.5	94.5	95.5	94.0
(200, 10, 20)	bias	-0.0334	0.0151	-0.0086	0.0114
	sd	0.2051	0.0594	0.1166	0.1344
	rmse	0.2078	0.0613	0.1169	0.1349
	cp(%)	97.0	93.5	96.5	96.0
$\tau = 0.6$					
(50, 5, 10)	bias	-0.0998	-0.0128	0.0738	0.0379
	sd	0.3216	0.0841	0.3082	0.4408
	rmse	0.3367	0.0851	0.3169	0.4424
	cp(%)	96.5	95.5	94.0	98.5
(100, 5, 20)	bias	-0.0997	-0.0012	0.0418	-0.0164
	sd	0.2264	0.0570	0.1860	0.2021
	rmse	0.2474	0.0570	0.1906	0.2028
	cp(%)	98.0	95.0	94.5	94.0
(100, 10, 10)	bias	-0.0620	-0.0018	0.0411	-0.0153
	sd	0.2341	0.0565	0.1889	0.1787
	rmse	0.2422	0.0565	0.1933	0.1794
	cp(%)	97.0	95.5	95.5	95.5
(200, 10, 20)	bias	-0.0763	-0.0062	0.0374	-0.0286
	sd	0.1796	0.0386	0.1304	0.1355
	rmse	0.1951	0.0391	0.1357	0.1385
	cp(%)	95.5	93.5	94.5	94.5

Table 2.3. Simulation results based on 200 replications under the model with normally distributed random effects in the slope in Scenario 2.3, where parameters $\eta = 0.2$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6 .

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	0.0243	0.0131	0.0117	0.0614
	sd	0.2091	0.1130	0.2506	0.3246
	rmse	0.2105	0.1138	0.2509	0.3304
	cp(%)	94.5	95.5	95.5	97.5
(100, 5, 20)	bias	-0.0004	0.0227	0.0031	0.0251
	sd	0.1535	0.0732	0.1732	0.2044
	rmse	0.1535	0.0766	0.1732	0.2059
	cp(%)	96.0	95.5	94.0	95.0
(100, 10, 10)	bias	0.0113	0.0104	0.0153	0.0337
	sd	0.1484	0.0883	0.1878	0.1989
	rmse	0.1488	0.0889	0.1884	0.2017
	cp(%)	95.0	96.5	96.0	95.0
(200, 10, 20)	bias	0.0095	0.0051	-0.0076	0.0102
	sd	0.1132	0.0607	0.1275	0.1367
	rmse	0.1136	0.0609	0.1277	0.1371
	cp(%)	95.0	96.5	94.0	96.0
$\tau = 0.6$					
(50, 5, 10)	bias	0.0223	-0.0248	0.0980	0.0250
	sd	0.2609	0.1052	0.2948	0.3034
	rmse	0.2619	0.1081	0.3107	0.3044
	cp(%)	95.0	95.5	92.0	98.0
(100, 5, 20)	bias	0.0414	-0.0188	0.0720	0.0143
	sd	0.2703	0.0773	0.1971	0.2781
	rmse	0.2735	0.0796	0.2098	0.2785
	cp(%)	98.0	97.0	93.5	97.5
(100, 10, 10)	bias	0.0185	-0.0089	0.0606	-0.0041
	sd	0.1811	0.0676	0.1774	0.1982
	rmse	0.1820	0.0682	0.1875	0.1982
	cp(%)	95.5	98.0	94.5	95.5
(200, 10, 20)	bias	0.0012	0.0009	0.0624	0.0035
	sd	0.1148	0.0390	0.1213	0.1301
	rmse	0.1148	0.0390	0.1364	0.1301
	cp(%)	96.5	94.5	93.5	95.0

Table 2.4. Simulation results based on 200 replications under the model with normally distributed random effects in the slope in Scenario 2.4, where parameters $\eta = 0.4$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6 .

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	0.0207	0.0134	0.0313	0.0788
	sd	0.3363	0.1120	0.2616	0.3429
	rmse	0.3369	0.1128	0.2635	0.3518
	cp(%)	94.5	94.0	94.5	96.0
(100, 5, 20)	bias	0.0096	0.0071	-0.0042	0.0359
	sd	0.3053	0.0855	0.1804	0.1787
	rmse	0.3055	0.0858	0.1804	0.1823
	cp(%)	94.0	95.5	97.0	95.0
(100, 10, 10)	bias	0.0095	0.0076	0.0163	0.0269
	sd	0.2546	0.0862	0.1935	0.1879
	rmse	0.2548	0.0865	0.1942	0.1898
	cp(%)	92.0	95.5	95.0	94.5
(200, 10, 20)	bias	-0.0045	0.0081	-0.0093	-0.0051
	sd	0.2141	0.0621	0.1259	0.1368
	rmse	0.2141	0.0626	0.1262	0.1369
	cp(%)	95.5	94.5	93.5	96.5
$\tau = 0.6$					
(50, 5, 10)	bias	0.0199	-0.0343	0.0914	0.0743
	sd	0.3347	0.1048	0.3806	0.5534
	rmse	0.3353	0.1103	0.3914	0.5584
	cp(%)	94.5	96.5	97.0	99.5
(100, 5, 20)	bias	0.0202	-0.0049	0.0858	0.0208
	sd	0.2761	0.0589	0.2541	0.2088
	rmse	0.2768	0.0591	0.2682	0.2098
	cp(%)	97.0	96.5	94.5	93.5
(100, 10, 10)	bias	0.0157	-0.0106	0.0453	-0.0099
	sd	0.2580	0.0722	0.2078	0.2064
	rmse	0.2585	0.0730	0.2127	0.2066
	cp(%)	96.5	97.0	95.5	94.5
(200, 10, 20)	bias	0.0112	-0.0107	0.0496	0.0103
	sd	0.2161	0.0594	0.1881	0.1349
	rmse	0.2164	0.0604	0.1945	0.1353
	cp(%)	95.0	98.0	96.5	96.5

Table 2.5. Simulation results for the model with gamma frailty obtained by the proposed and Emura et al. (2017)'s methods based on 200 replications in Scenario 2.1, with parameters $\eta = 0.2$, $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$.

(N, m, n_i)	Proposed method				Emura et al (2017)			
	η	τ	β_1	β_2	η	τ	β_1	β_2
(50, 5, 10)	bias	0.0030	0.0095	0.0301	-0.0034	0.1222	0.0374	0.0695
	sd	0.1918	0.1078	0.2938	0.3007	0.1756	0.2490	0.2878
	rmse	0.1919	0.1082	0.2953	0.3007	0.2139	0.2518	0.2961
	cp(%)	93.0	96.0	95.0	95.5	89.0	94.0	93.0
(100, 5, 20)	bias	-0.0267	0.0195	0.0353	0.0530	0.0416	0.0308	0.0392
	sd	0.1600	0.0827	0.1680	0.1988	0.1691	0.1868	0.1974
	rmse	0.1622	0.0850	0.1717	0.2058	0.1741	0.1893	0.2012
	cp(%)	96.5	95.5	94.0	94.0	97.5	95.0	96.5
(100, 10, 10)	bias	0.0322	0.0114	-0.0027	0.0244	0.0583	0.0363	0.0315
	sd	0.1611	0.0815	0.1776	0.2000	0.1706	0.1636	0.1867
	rmse	0.1643	0.0823	0.1776	0.2015	0.1803	0.1676	0.1894
	cp(%)	94.0	94.5	95.0	97.0	96.0	94.0	95.0
(200, 10, 20)	bias	0.0133	0.0074	0.0326	0.0384	0.0180	-0.0008	0.0011
	sd	0.1517	0.0640	0.1273	0.1396	0.1421	0.1485	0.1629
	rmse	0.1523	0.0644	0.1314	0.1448	0.1433	0.1485	0.1629
	cp(%)	93.0	93.5	92.5	92.5	97.5	97.0	99.0

Table 2.6. Simulation results for the model with gamma frailty obtained by the proposed and Emura et al. (2017)'s methods based on 200 replications in Scenario 2.1, with parameters $\eta = 0.2$, $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.6$.

(N, m, n_i)	Proposed method				Emura et al (2017)				
	η	τ	β_1	β_2	η	τ	β_1	β_2	
(50, 5, 10)	bias	-0.0407	0.0063	0.0669	-0.0172	-0.0644	-0.0035	0.0621	0.0687
	sd	0.0551	0.1007	0.3336	0.2644	0.1528	0.1664	0.2696	0.2907
	rmse	0.0685	0.1009	0.3402	0.2649	0.1659	0.1664	0.2767	0.2987
	cp(%)	93.0	95.0	94.0	94.5	96.5	95.0	94.5	94.0
(100, 5, 20)	bias	-0.0432	0.0161	0.0081	-0.0606	0.0510	-0.0274	0.0406	0.0476
	sd	0.0742	0.0596	0.1785	0.1816	0.2770	0.1373	0.1647	0.1948
	rmse	0.0859	0.0617	0.1787	0.1915	0.2816	0.1400	0.1696	0.2005
	cp(%)	97.5	93.5	96.5	94.0	92.0	93.5	94.0	95.5
(100, 10, 10)	bias	-0.0195	0.0149	-0.0013	-0.0432	0.0334	-0.0238	0.0663	0.0531
	sd	0.0512	0.0628	0.1843	0.1741	0.2354	0.1285	0.1864	0.2043
	rmse	0.0548	0.0646	0.1843	0.1794	0.2378	0.1307	0.1978	0.2111
	cp(%)	94.5	97.5	95.5	96.5	93.0	95.5	93.0	94.0
(200, 10, 20)	bias	-0.0275	0.0090	-0.0037	-0.0636	0.1210	-0.0361	0.0212	0.0266
	sd	0.0562	0.0387	0.1268	0.1117	0.3332	0.1167	0.1296	0.1464
	rmse	0.0626	0.0397	0.1269	0.1285	0.3545	0.1222	0.1313	0.1488
	cp(%)	94.5	94.5	95.5	93.5	93.5	94.5	94.5	96.5

Table 2.7. Simulation results for the model with gamma frailty obtained by the proposed and Emura et al. (2017)'s methods based on 200 replications in Scenario 2.2, with parameters $\eta = 0.4$, $\beta_1 = \beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$.

(N, m, n_i)	Proposed method				Emura et al (2017)			
	η	τ	β_1	β_2	η	τ	β_1	β_2
$\tau = 0.3$								
(50, 5, 10)	bias	-0.0229	0.0203	0.0411	0.0234	-0.0320	0.1281	0.0341
	sd	0.3617	0.1148	0.2719	0.3161	0.3354	0.1890	0.2842
	rmse	0.3624	0.1166	0.2750	0.3170	0.3369	0.2283	0.2862
	cp(%)	94.5	94.5	94.5	94.0	95.0	93.5	93.5
(100, 5, 20)	bias	0.0067	0.0157	0.0181	0.0298	-0.0684	0.0672	0.0181
	sd	0.4070	0.0827	0.1878	0.2131	0.2573	0.1719	0.1550
	rmse	0.4071	0.0842	0.1887	0.2152	0.2662	0.1846	0.1561
	cp(%)	94.5	94.5	91.5	95.0	97.0	94.5	93.5
(100, 10, 10)	bias	0.0821	0.0246	0.0255	0.0388	-0.0156	0.0478	0.0325
	sd	0.3455	0.0910	0.1863	0.2172	0.2350	0.1736	0.1823
	rmse	0.3551	0.0942	0.1880	0.2207	0.2355	0.1800	0.1852
	cp(%)	93.5	95.0	94.5	96.0	95.0	97.5	93.0
(200, 10, 20)	bias	0.0633	0.0146	0.0059	0.0131	-0.0155	-0.0001	0.0061
	sd	0.3089	0.0660	0.1218	0.1377	0.2143	0.1406	0.1205
	rmse	0.3154	0.0675	0.1219	0.1383	0.2149	0.1406	0.1206
	cp(%)	94.0	96.0	96.0	95.0	97.0	95.5	92.5

Table 2.8. Simulation results for the model with gamma frailty obtained by the proposed and Emura et al. (2017)'s methods based on 200 replications in Scenario 2.2, with parameters $\eta = 0.4$, $\beta_1 = \beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.6$.

(N, m, n_i)	Proposed method						Emura et al (2017)					
	η	τ	β_1	β_2	η	τ	β_1	β_2				
(50, 5, 10)	bias	-0.0577	0.0257	-0.0059	-0.0644	-0.0212	-0.0019	0.0709	0.0826			
	sd	0.1917	0.0772	0.2401	0.2645	0.3841	0.1563	0.3018	0.3248			
	rmse	0.2002	0.0814	0.2402	0.2722	0.3846	0.1563	0.3100	0.3351			
	cp(%)	95.5	95.5	97.0	95.5	96.0	95.0	92.5	95.0			
	bias	-0.0949	0.0163	-0.0196	-0.0661	-0.0258	-0.0129	0.0301	0.0404			
(100, 5, 20)	sd	0.2010	0.0506	0.1620	0.1952	0.3379	0.1294	0.1810	0.2054			
	rmse	0.2222	0.0532	0.1632	0.2061	0.3389	0.1300	0.1835	0.2093			
	cp(%)	97.0	94.0	96.0	94.5	95.0	94.0	93.0	95.5			
	bias	-0.0073	0.0245	-0.0509	-0.0890	-0.0050	-0.0086	0.0450	0.0624			
	sd	0.1393	0.0525	0.1910	0.2091	0.2562	0.1351	0.1891	0.2028			
(100, 10, 10)	rmse	0.1395	0.0579	0.1977	0.2273	0.2563	0.1353	0.1944	0.2122			
	cp(%)	96.5	92.0	95.0	93.5	96.0	95.0	96.0	94.0			
	bias	-0.0226	-0.0009	0.0254	-0.0313	-0.0111	-0.0172	0.0134	0.0213			
	sd	0.1679	0.0391	0.1215	0.1331	0.2393	0.1029	0.1162	0.1196			
	rmse	0.1695	0.0391	0.1241	0.1367	0.2395	0.1043	0.1170	0.1215			
(200, 10, 20)	cp(%)	97.5	95.0	94.5	94.0	96.0	94.5	94.5	96.0			

Table 2.9. Simulation results obtained by the proposed methods using a misspecified normal random effect model in Scenario 2.1 based on 200 replications when the true random effects distribution is gamma. The true parameters $\eta = 0.2$, $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6 .

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	-0.0434	0.0233	0.0439	0.0456
	sd	0.2035	0.1174	0.3131	0.3053
	cp(%)	97.0	95.5	93.5	93.5
(100, 10, 10)	bias	-0.0467	0.0191	0.0065	0.0400
	sd	0.1755	0.0897	0.1750	0.1967
	cp(%)	97.5	94.0	93.5	94.5
(100, 5, 20)	bias	-0.0125	0.0158	0.0258	0.0431
	sd	0.1796	0.0921	0.1971	0.2055
	cp(%)	95.0	95.5	95.5	96.0
(200, 10, 20)	bias	-0.0080	0.0174	-0.0120	0.0020
	sd	0.1412	0.0578	0.1162	0.1397
	cp(%)	95.5	95.0	95.0	94.0
$\tau = 0.6$					
(50, 5, 10)	bias	-0.0490	0.0128	0.0673	0.0439
	sd	0.1758	0.0836	0.3322	0.2857
	cp(%)	96.5	97.0	95.5	94.0
(100, 5, 20)	bias	-0.0327	0.0011	0.0522	-0.0110
	sd	0.1935	0.0581	0.1783	0.1718
	cp(%)	95.5	96.5	91.5	96.0
(100, 10, 10)	bias	-0.0226	-0.0097	0.0638	-0.0133
	sd	0.1602	0.0607	0.1892	0.1827
	cp(%)	95.5	96.0	94.5	95.0
(200, 10, 20)	bias	-0.0254	0.0089	0.0310	-0.0321
	sd	0.1198	0.0374	0.1088	0.1229
	cp(%)	98.0	96.0	92.5	94.5

Table 2.10. Simulation results obtained by both the proposed and Emura et al. (2017)'s methods using a misspecified gamma frailty model in scenario 2.1 based on 200 replications when the true random effects distribution is a mixture of normals $0.5N(0, 0.2) + 0.5N(0, 0.4)$. The true parameters $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$.

(N, m, n_i)	Proposed method				Emura et al (2017)				
	η	τ	β_1	β_2	η	τ	β_1	β_2	
(50, 5, 10)	bias	0.0807	0.0062	0.0380	0.1176	3.2862	0.6091	10.1897	7.4917
	sd	0.3997	0.1122	0.2528	0.4747	4.0565	0.1645	88.6486	53.0448
	rmse	0.4078	0.1124	0.2556	0.4890	5.2206	0.6309	89.2323	53.5713
	cp(%)	94.5	94.5	94.5	98.0	93.5	5.0	100.0	98.5
(100, 5, 20)	bias	-0.0197	0.0191	0.0108	0.0574	6.2254	0.6489	7.7824	6.5380
	sd	0.2419	0.0771	0.1819	0.2071	8.9309	0.1044	46.6844	31.4339
	rmse	0.2427	0.0795	0.1822	0.2149	10.8865	0.6572	47.3287	32.1066
	cp(%)	94.5	95.5	95.0	94.0	95.0	0.0	99.0	99.0
(100, 10, 10)	bias	-0.0052	0.0123	0.0149	0.0367	4.5206	0.6618	5.0968	5.4785
	sd	0.2263	0.0804	0.1912	0.1997	4.6840	0.0626	60.8759	61.0783
	rmse	0.2264	0.0813	0.1918	0.2031	6.5097	0.6647	61.0889	61.3235
	cp(%)	96.0	96.5	95.0	93.5	95.0	0.0	97.0	97.0
(200, 10, 20)	bias	-0.0338	0.0120	0.0183	0.0110	6.0779	0.6819	10.1392	9.9668
	sd	0.1611	0.0622	0.1304	0.1486	4.8037	0.0276	36.2865	36.0600
	rmse	0.1646	0.0633	0.1316	0.1491	7.7470	0.6824	37.6764	37.4120
	cp(%)	97.0	95.5	93.0	95.5	87.0	0.0	98.0	98.0

Table 2.11. Simulation results obtained by both the proposed and Emura et al. (2017)'s methods using a misspecified gamma frailty model in scenario 2.1 based on 200 replications when the true random effects distribution is a mixture of normals $0.5N(0, 0.2) + 0.5N(0, 0.4)$. The true parameters $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.6$.

(N, m, n_i)	Proposed method				Emura et al (2017)				
	η	τ	β_1	β_2	η	τ	β_1	β_2	
(50, 5, 10)	bias	-0.0133	-0.0202	0.0965	0.0283	4.3264	0.3464	8.2619	10.5323
	sd	0.3059	0.1135	0.7807	0.3358	7.7003	0.1153	59.3232	73.3230
	rmse	0.3062	0.1153	0.7866	0.3370	8.8325	0.3651	59.8957	74.0756
	cp(%)	95.5	96.5	97.5	96.0	97.0	0.0	96.0	96.0
(100, 5, 20)	bias	-0.0413	-0.0041	0.0642	0.0192	5.0458	0.3418	13.7960	11.7703
	sd	0.2271	0.0584	0.1788	0.1913	6.2825	0.1269	108.7297	104.3235
	rmse	0.2309	0.0585	0.1900	0.1923	8.0580	0.3646	109.6015	104.9854
	cp(%)	96.5	96.0	93.0	94.5	95.0	0.0	97.0	98.0
(100, 10, 10)	bias	-0.0375	-0.0017	0.1011	-0.0074	4.1429	0.3773	30.7685	28.0709
	sd	0.2050	0.0737	0.5055	0.2066	3.5439	0.0362	123.1173	113.0520
	rmse	0.2084	0.0737	0.5155	0.2067	5.4519	0.3790	126.9038	116.4849
	cp(%)	96.0	98.5	99.0	96.5	79.0	0.0	96.5	96.5
(200, 10, 20)	bias	-0.0642	-0.0001	0.0444	-0.0195	6.5153	0.3880	17.8160	18.2816
	sd	0.1365	0.0391	0.1338	0.1292	4.7287	0.0197	79.8196	79.7025
	rmse	0.1509	0.0391	0.1409	0.1306	8.0504	0.3885	81.7837	81.7723
	cp(%)	96.0	94.5	93.5	95.5	89.0	0.0	98.0	98.0

(a) $N = 50, m = 5, n_i = 10$ (b) $N = 100, m = 5, n_i = 20$ (c) $N = 100, m = 10, n_i = 10$ (d) $N = 200, m = 10, n_i = 20$

Figure 2.1. Simulation results for estimated cumulative baseline hazard function $\Lambda_{01}(\cdot)$ by the proposed method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{01}(t) = t$ (colored black) under the model with gamma frailty in Scenario 2.1 and $\tau = 0.3$.

(a) $N = 50, m = 5, n_i = 10$ (b) $N = 100, m = 5, n_i = 20$ (c) $N = 100, m = 10, n_i = 10$ (d) $N = 200, m = 10, n_i = 20$

Figure 2.2. Simulation results for estimated cumulative baseline hazard function $\Lambda_{02}(\cdot)$ by the proposed method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{02}(t) = 0.4t$ (colored black) under the model with gamma frailty in Scenario 2.1 and $\tau = 0.3$.

(a) $N = 50, m = 5, n_i = 10$ (b) $N = 100, m = 5, n_i = 20$ (c) $N = 100, m = 10, n_i = 10$ (d) $N = 200, m = 10, n_i = 20$

Figure 2.3. Simulation results for estimated cumulative baseline hazard function $\Lambda_{01}(\cdot)$ by Emura et al. (2017)'s method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{01}(t) = t$ (colored black) under the model with gamma frailty in Scenario 2.1 and $\tau = 0.3$.

(a) $N = 50, m = 5, n_i = 10$ (b) $N = 100, m = 5, n_i = 20$ (c) $N = 100, m = 10, n_i = 10$ (d) $N = 200, m = 10, n_i = 20$

Figure 2.4. Simulation results for estimated cumulative baseline hazard function $\Lambda_{02}(\cdot)$ by Emura et al. (2017)'s method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{02}(t) = 0.4t$ (colored black) under the model with gamma frailty in Scenario 2.1 and $\tau = 0.3$.

2.6 Real Data Analysis

For illustration purpose, we now apply the proposed method to analysis of data from the multi-institutional clinical studies of breast cancer, conducted by 36 cancer centers with total 5715 breast cancer patients in US and Europe (Haibe-Kains et al., 2012). The databases are available at <http://compbio.dfci.harvard.edu/pubs/sbtpaper/>. The aims are twofold: to evaluate the classification concordance based on genetic features and to assess the prognostic values of both clinical and genetic characteristics in predicting the disease outcome of patients. In the current analysis, we are interested in the later one. To assist treatment decisions for breast cancer, both clinical features and gene expression profiles of patients have also been used to predict patient survival outcome (Van De Vijver et al., 2002; Van't Veer et al., 2002a).

Data from 6 different patient cohorts are extracted from the available databases provided by 6 institutions including CAL dataset of breast cancer patients from the university of California, San Francisco, and the California Pacific Medical Center (United States); NKI dataset from National Kanker Institute (the Netherlands); STNO dataset from Stanford/Norway (United States and Norway); TRANSBIG dataset collected by the TransBIG consortium (Europe); UCSF dataset from University of California, San Francisco (United States); UNC dataset from University of North Carolina(United States). The survival outcomes of our interest are the time for DMFS and OS. DMFS and OS are considered under semi-competing risks, where distant metastasis is the nonterminal event and death is the terminal event. We are interested in how the clinical information of breast cancer patients affects both two events and the correlation between these two event times. The clinical features include estrogen receptor status (*er*, 1= positive and 0=negative), tumor size in cm (*size*, the longest length of invasive tumor measured by pathologists or surgeons, and grouped as small (< 2 cm) and large (≥ 2 cm).), lymph nodal status (*node*, 1=present and 0=absent), age at diagnosis (*age*, in years), histological grade (*grade*, 1=invasive ductal carcinoma, 2=other

invasive carcinoma, 3=noninvasive ductal carcinoma in situ and other noninvasive carcinoma; or 1=low, 2=intermediate, 3=high). There are only $n = 784$ patients who had complete information in all five clinical features. They were aged 24 to 89. Among them, 103 were from CAL dataset, 122 from NKI dataset, 88 from STNO dataset, 190 from TRANSBIG dataset, 93 from UCSF dataset and 188 from UNC dataset.

We analyze the data using the proposed method and two naive methods, marginal proportional hazards models with/without frailty for nonterminal and terminal events separately. Table 2.12 reports the estimates and their corresponding standard errors. The standard deviation is estimated by 500 randomly selected bootstrap samples using the stratified sampling method in Shao et al. (2003). For each cluster, the sample of original cluster size is first generated by with-replacement bootstrap, independently across clusters. These independent samples together form a bootstrap population sample. For comparison purpose, two naive methods are employed to fit the nonterminal DMFS and terminal OS separately: the usual Cox PH model with corresponding estimates obtained by R function *coxph* and the PH mixed-effects model $\lambda_k(t|b; X, Z) = \lambda_{0k}(t)e^{\beta'_k X + b_k}$ for ($k = 1, 2$) using R function *phmm*. All these three methods indicate that prognostic factors, estrogen receptor status and histological grade, are statistically significant (at the 5% level) in association with the DMFS and OS. However, after accounting for the dependence between the DMFS and OS, we find that covariate age appears insignificant in the proposed model, unlike it is both naive models. The estimated value of η is quite large and $\hat{\tau} = 0.5974$ with strong significance. Both results show strong evidence of significant cluster effects and relatively high correlation between the DMFS and OS in this study. Thus, the proposed method is a more suitable than the naive marginal methods for analysis of this data set.

Table 2.12. Analysis of the breast cancer data

parameter	proposed estimates ($\hat{se}$)	naive marginal estimate (sd)	cox estimate (sd)
regression for DMFS			
er	-0.2817(0.1388)*	-0.3299(0.1324)*	-0.0821(0.1275)
size	0.2081(0.0501)*	0.1980(0.0511)*	0.0836(0.0458)
node	0.1324(0.1055)	0.2831(0.1441)	0.5521(0.1237)*
age	-0.0111(0.0057)	-0.0150(0.0059)*	-0.0314(0.0053)*
grade 1	-0.4499(0.1866)*	-0.6317(0.2257)*	-0.4992(0.2153)*
grade 2	-0.1131(0.1190)	-0.0929(0.1319)	-0.1228(0.1276)
$\eta = 0.5985$			
regression for OS			
er	-0.6003(0.1594)*	-0.6072(0.1263)*	-0.5533(0.1260)*
size	0.2510(0.0560)*	0.2348(0.0477)*	0.2134(0.0454)*
node	0.2090(0.1276)	0.2026(0.1351)	0.7601(0.1261)*
age	0.0082(0.0047)	0.0104(0.0052)*	0.0098(0.0046)*
grade 1	-0.6054(0.1968)*	-0.6219(0.2149)*	-0.5125(0.2147)*
grade 2	-0.2419(0.1548)	-0.2670(0.1302)*	-0.2615(0.1314)
$\eta = 0.3074$			
τ	0.5974(0.0867)*	-	-
η	0.3787(0.0960)*	-	-

* significant at 5% significance level

Joint regression analysis for survival data in the presence of two sets of semi-competing risks

As already described in Section 1.5, we focus on the study of survival data subject to two sets of semi-competing risks in this chapter. We propose a novel statistical approach that jointly models such data via a pair of copulas to account for multiple dependence structures, while the marginal distribution of each endpoint is formulated by a Cox PH model. The rest of this chapter is organized as follows. Section 3.1 introduces the proposed model, while Section 3.2 and 3.3 develops the pseudo-likelihood estimation method with a detailed algorithm for its implementation. Section 3.4 measures the dependence between the two nonterminal events using the cross ratio. In Section 3.5, the performance of the proposed method is examined by finite sample numerical studies. We illustrate the practical utility of the proposed methodology in Section 3.6 with data from the motivating example on bone marrow transplant. And some technical details are provided in Appendix C.

3.1 Proposed model

Let “event 1” be a nonterminal event, “event 2” be a terminal event, and “event 3” be another nonterminal event. Event 1 and event 2 are termed semi-competing risks; event 2 and event 3 are also termed semi-competing risks. Event 2 potentially censors event 1 and event 3, however, an observation of event 1 or event 3 does not prevent a subsequent observation of event 2. Denote $T^{(1)}$, $T^{(2)}$ and $T^{(3)}$ be the underlying times to event 1, event 2 and event 3, respectively. We are interested in assessing the effects of covariates on the marginal distribution of $T^{(k)}$, which is modeled by the Cox PH model:

$$\lambda_k(t|X_k) = \lambda_{0k}(t)e^{\beta'_k X_k}, \quad (3.1)$$

for $k = 1, 2, 3$, where $\lambda_k(t|X_k)$ is the conditional hazard function of $T^{(k)}$ given event k and covariate vector X_k , $\lambda_{0k}(\cdot)$ is the baseline hazard function, β_k is a vector of unknown regression coefficients associated with X_k .

We further assume that the joint survival function of $T^{(1)}$ and $T^{(2)}$ can be specified by a known copula subject to an unknown parameter α_1 , which quantifies the association between $T^{(1)}$ and $T^{(2)}$:

$$S_{12}(t^{(1)}, t^{(2)}) = \text{pr}(T^{(1)} \geq t^{(1)}, T^{(2)} \geq t^{(2)}) = c_1\{S_1(t^{(1)}), S_2(t^{(2)}); \alpha_1\}, \quad (3.2)$$

for $t^{(1)} \geq 0, t^{(2)} \geq 0$, where $c_1\{u, v; \alpha_1\}$ is a copula function, α_1 is a constant association parameter one-to-one corresponding to Kendall's τ , and $S_k(\cdot)$ is the marginal survival function of $T^{(k)}$ for $k = 1, 2, 3$.

Similarly, we assume that the joint survival function of $T^{(2)}$ and $T^{(3)}$ can be specified by a known copula subject to an unknown parameter α_2 :

$$S_{23}(t^{(2)}, t^{(3)}) = \text{pr}(T^{(2)} \geq t^{(2)}, T^{(3)} \geq t^{(3)}) = c_2\{S_2(t^{(2)}), S_3(t^{(3)}); \alpha_2\}, \quad (3.3)$$

for $t^{(2)} \geq 0, t^{(3)} \geq 0$, where $c_2\{u, v; \alpha_2\}$ is a copula function.

3.2 Pseudo-likelihood estimation

We consider possibly right-censored observations of times to event 1, event 2 and event 3, denoted by $T_i^{(1)}$, $T_i^{(2)}$ and $T_i^{(3)}$, respectively, for each subject i , $i = 1, \dots, n$. Let C be an external censoring time that is independent of the triplet $(T^{(1)}, T^{(2)}, T^{(3)})$. Suppose that observations $\mathcal{O}_i = \{(T_i, \tilde{T}_i, Y_i, X_{ki}, \delta_i^{(k)}), k = 1, 2, 3\}$ are obtained, where $T_i = \min(T_i^{(1)}, T_i^{(2)}, C_i)$, $\tilde{T}_i = \min(T_i^{(2)}, C_i)$, $Y_i = \min(T_i^{(2)}, T_i^{(3)}, C_i)$, $\delta_i^{(1)} = I(T_i = T_i^{(1)})$, $\delta_i^{(2)} = I(\tilde{T}_i = T_i^{(2)})$, $\delta_i^{(3)} = I(Y_i = T_i^{(3)})$, and $I(\cdot)$ is the indicator function. For a random sample of size n , the observed data consist of $\mathcal{O} = \{\mathcal{O}_i, i = 1, \dots, n\}$.

Since the data involve complex interdependencies, it is difficult to specify the full 3-dimensional distribution explicitly but it is possible to specify the 2-dimensional marginal distribution. Learnt from the pseudo-likelihood method in Varin and Vidoni (2005) and Varin et al. (2011), we construct a pseudo-likelihood based on bivariate marginal densities for two sets of semi-competing risks data. This pseudo-likelihood is an approximation to the full likelihood function. Its appealing advantage is that instead of fully specification, it only requires model assumptions on the lower dimensional marginal densities. Therefore, it provides convenient way to estimate model parameters and in the meantime its resulted estimator enjoys high efficiency and robustness shown in the literature (see e.g. Varin and Vidoni (2005) and Varin et al. (2011)). Let $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3)$ and $\boldsymbol{\lambda} = (\lambda_{01}, \lambda_{02}, \lambda_{03})$. The proposed pseudo-likelihood function of $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ is given by

$$LP(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathcal{O}) = \prod_{i=1}^n L_{12i}(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathcal{O}_i) \cdot L_{23i}(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathcal{O}_i), \quad (3.4)$$

where

$$\begin{aligned} & L_{12i}(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathcal{O}_{ij}) \\ &= \Pr(T_i^{(1)} = T_i, T_i^{(2)} = \tilde{T}_i)^{\delta_i^{(1)} \delta_i^{(2)}} \cdot \Pr(T_i^{(1)} = T_i, T_i^{(2)} > \tilde{T}_i)^{\delta_i^{(1)} (1 - \delta_i^{(2)})} \\ & \quad \cdot \Pr(T_i^{(1)} > T_i, T_i^{(2)} > \tilde{T}_i)^{(1 - \delta_i^{(1)}) (1 - \delta_i^{(2)})} \cdot \Pr(T_i^{(1)} > T_i, T_i^{(2)} = \tilde{T}_i)^{(1 - \delta_i^{(1)}) \delta_i^{(2)}}, \end{aligned}$$

$$\begin{aligned}
 & L_{23i}(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathcal{O}_{ij}) \\
 &= \Pr(T_i^{(2)} = \tilde{T}_i, T_i^{(3)} = Y_i)^{\delta_i^{(2)} \delta_i^{(3)}} \cdot \Pr(T_i^{(2)} = \tilde{T}_i, T_i^{(3)} > Y_i)^{\delta_i^{(2)} (1 - \delta_i^{(3)})} \\
 &\quad \cdot \Pr(T_i^{(2)} > \tilde{T}_i, T_i^{(2)} > Y_i)^{(1 - \delta_i^{(2)}) (1 - \delta_i^{(3)})} \cdot \Pr(T_i^{(2)} > \tilde{T}_i, T_i^{(3)} = Y_i)^{(1 - \delta_i^{(2)}) \delta_i^{(3)}}.
 \end{aligned}$$

We obtained the estimators through maximizing the pseudo-likelihood in (3.4). First we seek to estimate unknown functions $\lambda_{0k}(\cdot)$ for $k = 1, 2, 3$ by treating each of them as piecewise constants between distinct uncensored event times. The cumulative baseline hazard functions $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(t) dt$ are in turn allowed to be step functions. Let $\boldsymbol{\Lambda}(t) = (\Lambda_{01}(t), \Lambda_{02}(t), \Lambda_{03}(t))$. One can estimate $\Lambda_{0k}(\cdot)$ nonparametrically using Breslow's estimator and estimate $\lambda_{0k}(t)$ as the jump sizes of $\Lambda_{0k}(\cdot)$ at time t , denoted by $\Lambda_{0k}\{t\}$ for $k = 1, 2, 3$. In particular, based on models (3.1)-(3.3) and plugging the estimated $\lambda_{0k}(\cdot)$ in the pseudo-likelihood $LP(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathcal{O})$, we obtain a modified pseudo-likelihood:

$$\begin{aligned}
 LP^*(\boldsymbol{\theta}, \boldsymbol{\Lambda} | \mathcal{O}) &= \prod_{i=1}^n \left[c_{112}\{S_1(T_i), S_2(\tilde{T}_i); \alpha_1\} S_1(T_i) \Lambda_{01}\{T_i\} \cdot e^{\beta'_1 X_{1i}} S_2(\tilde{T}_i) \Lambda_{02}\{\tilde{T}_i\} \right. \\
 &\quad \cdot e^{\beta'_2 X_{2i}} \left. \right]^{\delta_i^{(1)} \delta_i^{(2)}} \cdot \left[c_{11}\{S_1(T_i), S_2(\tilde{T}_i); \alpha_1\} S_1(T_i) \Lambda_{01}\{T_i\} e^{\beta'_1 X_{1i}} \right]^{\delta_i^{(1)} (1 - \delta_i^{(2)})} \\
 &\quad \cdot \left[c_1\{S_1(T_i), S_2(\tilde{T}_i); \alpha_1\} \right]^{(1 - \delta_i^{(1)}) (1 - \delta_i^{(2)})} \\
 &\quad \cdot \left[c_{12}\{S_1(T_i), S_2(\tilde{T}_i); \alpha_1\} S_2(\tilde{T}_i) \Lambda_{02}\{\tilde{T}_i\} e^{\beta'_2 X_{2i}} \right]^{(1 - \delta_i^{(1)}) \delta_i^{(2)}} \\
 &\quad \left[c_{212}\{S_2(\tilde{T}_i), S_3(Y_i); \alpha_2\} S_2(\tilde{T}_i) \Lambda_{02}\{\tilde{T}_i\} \cdot e^{\beta'_2 X_{2i}} S_3(Y_i) \Lambda_{03}\{Y_i\} \right. \\
 &\quad \cdot e^{\beta'_3 X_{3i}} \left. \right]^{\delta_i^{(2)} \delta_i^{(3)}} \cdot \left[c_{21}\{S_2(\tilde{T}_i), S_3(Y_i); \alpha_2\} S_2(\tilde{T}_i) \Lambda_{02}\{\tilde{T}_i\} e^{\beta'_2 X_{2i}} \right]^{\delta_i^{(2)} (1 - \delta_i^{(3)})} \\
 &\quad \cdot \left[c_2\{S_2(\tilde{T}_i), S_3(Y_i); \alpha_2\} \right]^{(1 - \delta_i^{(2)}) (1 - \delta_i^{(3)})} \\
 &\quad \cdot \left[c_{22}\{S_2(\tilde{T}_i), S_3(Y_i); \alpha_2\} S_3(Y_i) \Lambda_{03}\{Y_i\} e^{\beta'_3 X_{3i}} \right]^{(1 - \delta_i^{(2)}) \delta_i^{(3)}}
 \end{aligned} \tag{3.5}$$

where $c_{k1}\{u, v; \alpha_k\} = \frac{\partial c_k(u, v; \alpha_k)}{\partial u}$, $c_{k2}\{u, v; \alpha_k\} = \frac{\partial c_k(u, v; \alpha_k)}{\partial v}$, $c_{k12}\{u, v; \alpha_k\} = \frac{\partial^2 c_k(u, v; \alpha_k)}{\partial u \partial v}$, $k = 1, 2$. Taking derivatives of the logarithm of the modified pseudo-likelihood function in (3.5) with respect to $\Lambda_{01}\{T_i\}$, $\Lambda_{02}\{\tilde{T}_i\}$, and $\Lambda_{03}\{Y_i\}$, respectively, and letting them equal to 0's, we obtain the nonparametric maximum likelihood estimates.

$$\hat{\Lambda}_{01}\{T_i\} = \frac{\delta_i^{(1)}}{\sum_{j \in R_1(T_i)} w_{1j} e^{\beta'_1 X_{1j}}}, \quad (3.6)$$

$$\hat{\Lambda}_{02}\{\tilde{T}_i\} = \frac{\delta_i^{(2)}}{\sum_{j \in R_2(\tilde{T}_i)} w_{2j} e^{\beta'_2 X_{2j}}}, \quad (3.7)$$

$$\hat{\Lambda}_{03}\{Y_i\} = \frac{\delta_i^{(3)}}{\sum_{j \in R_3(Y_i)} w_{3j} e^{\beta'_3 X_{3j}}}, \quad (3.8)$$

where $j \in R_1(T_i)$, $j \in R_2(\tilde{T}_i)$ and $j \in R_3(Y_i)$ stand for $T_j \leq T_i$, $\tilde{T}_j \leq \tilde{T}_i$ and $Y_j \leq Y_i$, respectively. Details about notation w_{1j} , w_{2j} and w_{3j} can be found in Appendix C. Thus the cumulative baseline hazard functions for the nonterminal and terminal events are estimated by the following step functions

$$\hat{\Lambda}_{01}(t) = \sum_{i \in R_1(t)} \hat{\Lambda}_{01}\{T_i\}, \quad (3.9)$$

$$\hat{\Lambda}_{02}(t) = \sum_{i \in R_2(t)} \hat{\Lambda}_{02}\{\tilde{T}_i\}, \quad (3.10)$$

$$\hat{\Lambda}_{03}(t) = \sum_{i \in R_3(t)} \hat{\Lambda}_{03}\{Y_i\}. \quad (3.11)$$

Plugging the estimated jumps and cumulative hazard functions given in (3.6)-(3.11) at observed event times into the modified pseudo-likelihood (3.5), the estimates of θ can be obtained by maximizing the modified pseudo-likelihood as described in Section 3.3.

3.3 Implementation and algorithm

We rely on the the following algorithm to implement the proposed pseudo-likelihood estimation.

1. Set initial estimates for parameters $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3)$ and cumulative baseline hazards $\boldsymbol{\Lambda} = (\Lambda_{01}, \Lambda_{02}, \Lambda_{03})$ as: $\alpha_1^{(0)} = \alpha_2^{(0)}$ being the value corresponding to $\tau_1 = \tau_2 = 0.5$, and $\beta_k^{(0)}$ and $\Lambda_{0k}^{(0)}$ obtained by fitting a marginal Cox PH model to observed data using R function *coxph*, $k = 1, 2, 3$.
2. With the estimates $\boldsymbol{\Lambda}^{(d)} = (\Lambda_{01}^{(d)}, \Lambda_{02}^{(d)}, \Lambda_{03}^{(d)})$, update parameter estimates $\boldsymbol{\theta}^{(d+1)}$ through maximizing the pseudo-likelihood $LP^*(\boldsymbol{\theta}, \boldsymbol{\Lambda}^{(d)}|\mathcal{O})$ using a numerical optimization algorithm, which can be done by R function *optim*.
3. With the estimates $\boldsymbol{\theta}^{(d+1)}$, compute the updated estimates $\boldsymbol{\Lambda}^{(d+1)}$ based on (3.9)-(3.11).

Repeat **2-3** until convergence is obtained. That is, the absolute differences between subsequent estimates of $\boldsymbol{\theta}$ are all less than or the score function of $\boldsymbol{\theta}$ is less than a small quantity ϵ , e.g., $\epsilon = 10^{-2}$.

In joint regression analysis of usual semi-competing risks data, Chen (2012) proposed to estimate standard error (SE) of the maximum likelihood estimator using the inverse of Hessian matrix. Based on our numerical experience, this method often underestimate SEs of the maximum psuedo-likelihood estimates of $\boldsymbol{\theta}$ obtained from the algorithm above, and thus cannot be directly applied to the case we considered.

Typically, the SE of a maximum pseudo likelihood estimator can be estimated either through the bootstrap method or the sandwich estimate of variance. Due to the complex structure of the psuedo-likelihood in the two sets of semicompeting risks data setting, the sandwich method seems not perform stably. Therefore, we suggest the bootstrap method with its performance shown in Section 3.5.

3.4 Cross ratio of two nonterminal event times

Although the association between two intermediate nonterminal events is not explicitly formulated in the proposed model, it can be implied by the model

construction in which both of the non-terminal events are associated with the same terminal event. To assess such association, we apply the cross ratio, a desirable time-dependent association measure (Anderson et al., 1992), defined by

$$\psi(t^{(1)}, t^{(3)}) = \frac{P(T^{(3)} > t^{(3)} | T^{(1)} > t^{(1)})}{P(T^{(3)} > t^{(3)})} = \frac{P(T^{(1)} > t^{(1)}, T^{(3)} > t^{(3)})}{P(T^{(1)} > t^{(1)})P(T^{(3)} > t^{(3)})}. \quad (3.12)$$

$\psi(t^{(1)}, t^{(3)}) = 1$ for independent events $T^{(1)}$ and $T^{(3)}$, while $\psi(t^{(1)}, t^{(3)}) \neq 1$ for dependent events. $\psi(t^{(1)}, t^{(3)}) > 1$ (or $\psi(t^{(1)}, t^{(3)}) < 1$) indicates positive (or negative) dependence between $T^{(1)}$ and $T^{(3)}$, similar to the property of the rate ratio proposed in Ning et al. (2015). A large value of $\psi(t^{(1)}, t^{(3)})$ shows strong dependence between $T^{(1)}$ and $T^{(3)}$. To estimate $\psi(t^{(1)}, t^{(3)})$, we first use the bivariate survival function developed in Lin and Ying (1993) to estimate the joint survival function of $T^{(1)}$ and $T^{(3)}$. It is defined by

$$\hat{S}_{13}(t^{(1)}, t^{(3)}) = \frac{\sum_{i=1}^n I(T_i \geq t^{(1)}, Y_i \geq t^{(3)})}{n\hat{G}(t^{(1)} \vee t^{(3)})}, \quad (3.13)$$

where $\hat{G}(\cdot)$ is the Kaplan-Meier estimator for the survival function of censoring time $C \vee T^{(2)}$, that is $G(t) = P((C \vee T^{(3)}) \geq t)$, and

$$\hat{G}(t^{(1)} \vee t^{(3)}) = \prod_{i=1}^n \left\{ 1 - \frac{1}{\sum_{j=1}^n I(\tilde{C}_j \geq \tilde{C}_i)} \right\}^{\delta_i^{\tilde{C}}}, \quad (3.14)$$

with $\tilde{C}_i = (C_i \vee T_i^{(2)}) \wedge (T_i^{(1)} \vee T_i^{(3)}) = T_i \vee Y_i$, $\delta_i^{\tilde{C}} = I\{(C_i \vee T_i^{(2)}) \leq (T_i^{(1)} \vee T_i^{(3)})\} = 1 - \delta_i^{(1)}\delta_i^{(3)}$. We adopt the nonparametric maximum likelihood estimators in (3.9)-(3.11) to compute the marginal survival functions. The performance of estimated $\psi(t^{(1)}, t^{(3)})$ is demonstrated in the following numerical studies.

3.5 Simulation Study

In this section, we examine the performance of the proposed method via finite sample simulation studies. $(T^{(1)}, T^{(2)}, T^{(3)})$ are generated by the quasi-inverse method in (Nelsen (2007), chap. 2) using the following steps:

Step 1. Generate u, s, t following uniform distribution $U(0, 1)$ independently.

Step 2. Set $v = c_{1u}^{-1}(s)$ and $w = c_{2u}^{-1}(t)$, where $c_{1u}(v) = \frac{\partial c_1\{u, v; \alpha_1\}}{\partial u}$ and $c_{2u}(w) = \frac{\partial c_2\{u, w; \alpha_2\}}{\partial u}$.

Step 3. $T^{(1)} = F_1^{-1}(v)$, $T^{(2)} = F_2^{-1}(u)$ and $T^{(3)} = F_3^{-1}(w)$, where $F_k(t) = 1 - S_k(t)$ and $F_k^{-1}(t) = \inf\{x | F_k(x) \geq t\} = \sup\{x | F_k(x) \leq t\}$ for $k = 1, 2, 3$.

This data generation process allows dependence between $T^{(1)}, T^{(2)}$ and $T^{(3)}$.

The three event times $T^{(1)}$, $T^{(2)}$ and $T^{(3)}$ are generated from the proportional hazards models $\lambda_1(t_1|X_1) = \lambda_{01}(t_1)\exp(\beta_1'X_1)$, $\lambda_2(t_2|X_2) = \lambda_{02}(t_2)\exp(\beta_2'X_2)$, and $\lambda_3(t_3|X_3) = \lambda_{03}(t_3)\exp(\beta_3'X_3)$, respectively. Assume that $X_1 = (Z_1, Z_2)$, $X_2 = (Z_1, Z_3)$ and $X_3 = (Z_1, Z_4)$, where Z_1 follows a uniform distribution $U(0, 1)$, Z_2 follows a standard normal distribution truncated at ± 2 , Z_3 follows a uniform distribution $U(0, 1)$ and Z_4 is generated from a Bernoulli distribution with success probability of 0.5. We take $\beta_1 = (1, 1)'$, $\beta_2 = (1, -1)'$ and $\beta_3 = (1, 0.5)'$. The joint distribution of $(T^{(1)}, T^{(2)})$ is specified by a Clayton copula $c_1\{u, v; \alpha_1\}$ with association parameter α_1 , and the joint distribution of $(T^{(2)}, T^{(3)})$ is specified by another Clayton copula $c_2\{u, v; \alpha_2\}$ with association parameter α_2 . Under the Clayton copulas, $\alpha_k = 2\tau_k/(1-\tau_k)$, where τ_k is the Kendall's tau one-to-one corresponding to α_k for $k = 1, 2$. We choose values of (τ_1, τ_2) as $(0.3, 0.3)$, $(0.3, 0.7)$, $(0.7, 0.7)$ to represent different degrees of associations between event 1 (or event 3) and event 2. Censoring times C are generated from a uniform distribution $U[0, C_0]$ so that observed event times can be obtained by $T = \min(T^{(1)}, T^{(2)}, C)$, $\tilde{T} = \min(T^{(2)}, C)$ and $Y = \min(T^{(2)}, T^{(3)}, C)$, respectively.

We consider the following three scenarios to access the finite sample performance of the proposed estimator with different types of baseline hazard functions.

Scenario 3.1: $\lambda_{01}(t_1) = 1$, $\lambda_{02}(t_2) = 0.2$, $\lambda_{03}(t_3) = 0.5$, $C_0 = 3$.

Scenario 3.2: $\lambda_{01}(t_1) = t_1^{1/2}$, $\lambda_{02}(t_2) = \frac{1}{2}t_2^{1/2}$, $\lambda_{03}(t_3) = t_3^{1/2}$, $C_0 = 2$.

Scenario 3.3: $\lambda_{01}(t_1) = t_1^{1/2}$, $\lambda_{02}(t_2) = \frac{1}{3}$, $\lambda_{03}(t_3) = t_3^{1/2}$, $C_0 = 2$.

Constant baseline hazards are assumed in Scenario 3.1 and Weibull baseline hazards are in Scenario 3.2, while a mixture of constant and Weibull baseline hazards is allowed in Scenario 3.3. We choose different values of C in different scenarios, leading to censoring rates of nearly 40% for $T^{(1)}$, 60% for $T^{(2)}$ and 40% for $T^{(3)}$ in each scenario.

We conduct 500 simulation runs for each scenario and summarize the performance of the proposed estimator in terms of mean bias and the sample standard deviation of estimates and the coverage percentage for the 95% confidence intervals (cp). Results obtained from Scenarios 3.1-3.3 are reported in Tables 3.1-3.3, respectively. In each scenario, all biases tend to be close to zero with small sd's and nearly 95% cp. In general, when the sample size increases, both biases and sd's of most estimated parameters reduce. Particularly, the relative reductions in bias and sd, corresponding to the improvements in accuracy and efficiency, tend to become great for all regression parameters as association measures τ_1 and τ_2 increase. It can be seen that the proposed estimator performs similarly over Scenarios 3.1-3.3, in which different types of the baseline hazard functions are assumed.

Table 3.4 compares simulation results of the bootstrap standard error(BSE) with sd's in Scenario 3.1, where the BSE is computed as the sample sd of estimates based on bootstrap resamples. We first draw $B = 200$ bootstrap samples $\mathcal{O}_i^{(1)}, \mathcal{O}_i^{(2)}, \dots, \mathcal{O}_i^{(B)}$ for each simulated sample $\mathcal{O}_i$ of size n , $i = 1, \dots, n$. More specifically, for each bootstrap sample we re-sample from the simulated observations of the true time-to-event variables $T_i^{(1)}, T_i^{(2)}$ and $T_i^{(3)}$, and then regenerate censoring times C_i according to the pre-specified censoring rates to get the censored sample data. The results in Table 3.4 show that the values of BSE are close to the corresponding sd's, implying the satisfactory performance of the BSE.

Table 3.5 summarizes simulation results in terms of bias, sd's and mean square error (MSE) obtained by using the proposed method and three marginal methods

in Scenario 3.1 with $\tau_{10} = 0.3$ and $\tau_{20} = 0.7$. The biases in both the proposed model and marginal Cox PH model are comparable and all close to 0, while estimates of most regression parameters in the marginal copula models are quite biased. Compared with marginal methods, the proposed method yields estimates with smaller sd's and MSEs, indicating its outperformance over those marginal ones in terms of estimation efficiency.

Figures 3.1-3.3 show the estimated cumulative baseline hazard curves obtained from the proposed method for $\hat{\Lambda}_{01}$, $\hat{\Lambda}_{02}$, and $\hat{\Lambda}_{03}$, respectively, which are based on the first 50 simulation runs under Scenario 3 with $\tau_{10} = 0.3$ and $\tau_{20} = 0.3$. The true cumulative baseline hazards are colored black in the figures, while the estimated cumulative baseline hazards are the red curves. The estimated cumulative baseline hazard curves are shown to be close to the corresponding true functions across all the settings and their biases reduce as the sample size increases.

Figure 3.4 shows the estimated curves of $\psi(t, t)$ under different settings. We can see that the value of $\psi(t, t)$ becomes greater when the association between $T^{(1)}$ and $T^{(3)}$ increases for all the settings. This indicates that the proposed model is able to incorporate different levels of the dependence between two nonterminal events by specifying different values of τ_1 and τ_2 .

Table 3.1. Biases, standard derivations(sd's in parentheses), and cp(%) for the proposed estimator based on 500 replications under Scenario 3.1.

$\tau_{10} = 0.3, \tau_{20} = 0.3$															
$n = 150$				$n = 300$				$n = 150$				$n = 300$			

Table 3.2. Biases, standard derivations(sd's in parentheses), and cp(%) for the proposed estimator based on 500 replications under Scenario 3.2.

	$\tau_{10} = 0.3, \tau_{20} = 0.3$				$\tau_{10} = 0.3, \tau_{20} = 0.7$				$\tau_{10} = 0.7, \tau_{20} = 0.7$			
	$n = 150$		$n = 300$		$n = 150$		$n = 300$		$n = 150$		$n = 300$	
	bias(sd)	cp(%)	bias(sd)	cp(%)	bias(sd)	cp(%)	bias(sd)	cp(%)	bias(sd)	cp(%)	bias(sd)	cp(%)
τ_{10}	0.024(0.116)	96.0	0.004(0.081)	95.4	0.026(0.114)	95.4	0.003(0.080)	95.6	0.011(0.056)	95.6	0.003(0.041)	94.2
τ_{20}	0.016(0.103)	95.4	0.008(0.073)	94.2	0.011(0.059)	94.2	0.004(0.038)	95.0	0.008(0.059)	93.6	0.004(0.038)	95.0
β_{11}	-0.041(0.356)	93.4	-0.032(0.274)	94.8	-0.041(0.359)	93.4	-0.041(0.277)	94.8	-0.055(0.331)	93.8	-0.034(0.250)	95.2
β_{12}	-0.003(0.153)	94.6	-0.005(0.108)	94.6	0.001(0.152)	94.8	-0.005(0.108)	95.0	-0.009(0.139)	95.4	0.002(0.092)	96.0
β_{21}	-0.092(0.334)	94.5	-0.053(0.259)	93.8	-0.046(0.299)	94.4	-0.020(0.244)	94.8	-0.055(0.318)	94.4	-0.024(0.235)	94.0
β_{22}	-0.051(0.210)	94.4	-0.018(0.143)	94.8	-0.025(0.167)	95.2	-0.003(0.111)	93.8	-0.022(0.169)	94.2	-0.003(0.113)	94.0
β_{31}	-0.005(0.477)	94.4	0.008(0.398)	96.4	-0.068(0.434)	94.8	-0.033(0.364)	94.8	-0.093(0.418)	95.8	-0.035(0.339)	95.2
β_{32}	0.007(0.510)	95.6	-0.037(0.396)	95.8	-0.045(0.444)	95.4	-0.029(0.313)	95.2	-0.038(0.402)	95.2	-0.026(0.279)	94.8

Table 3.3. Biases, standard derivations(sd's in parentheses), and cp(%) for the proposed estimator based on 500 replications in Scenario 3.3.

	$\tau_{10} = 0.3, \tau_{20} = 0.3$						$\tau_{10} = 0.3, \tau_{20} = 0.7$						$\tau_{10} = 0.7, \tau_{20} = 0.7$					
	$n = 150$		$n = 300$		$n = 150$		$n = 300$		$n = 150$		$n = 300$		$n = 150$		$n = 300$		$n = 150$	
	bias(sd)	cp(%)	bias(sd)	cp(%)	bias(sd)	cp(%)	bias(sd)	cp(%)	bias(sd)	cp(%)	bias(sd)	cp(%)	bias(sd)	cp(%)	bias(sd)	cp(%)	bias(sd)	cp(%)
τ_{10}	0.025(0.118)	94.0	0.004(0.082)	94.8	0.027(0.116)	94.2	0.007(0.081)	95.0	0.008(0.052)	94.0	0.003(0.040)	94.4	0.008(0.052)	94.0	0.003(0.040)	94.4	0.003(0.040)	94.4
τ_{20}	0.019(0.104)	95.2	0.010(0.073)	94.2	0.010(0.052)	93.4	0.004(0.037)	95.4	0.004(0.052)	95.4	0.003(0.036)	94.8	0.004(0.052)	95.0	0.003(0.036)	94.8	0.003(0.036)	94.8
β_{11}	-0.029(0.375)	94.2	-0.033(0.286)	94.6	-0.031(0.377)	94.4	-0.045(0.283)	95.0	-0.035(0.329)	95.2	-0.025(0.256)	94.0	-0.035(0.329)	95.2	-0.025(0.256)	94.0	-0.025(0.256)	94.0
β_{12}	0.000(0.160)	93.4	-0.007(0.113)	95.0	0.000(0.159)	93.6	-0.007(0.113)	94.8	-0.007(0.142)	94.8	0.000(0.094)	95.8	-0.007(0.142)	94.6	0.000(0.094)	95.8	0.000(0.094)	95.8
β_{21}	-0.072(0.358)	93.0	-0.049(0.265)	94.6	-0.035(0.311)	95.4	-0.021(0.240)	94.6	-0.034(0.327)	94.6	-0.027(0.240)	95.0	-0.034(0.327)	94.0	-0.027(0.240)	95.0	-0.027(0.240)	95.0
β_{22}	-0.042(0.218)	94.6	-0.019(0.146)	94.8	-0.024(0.163)	95.2	-0.003(0.109)	95.0	-0.016(0.166)	95.0	-0.001(0.110)	95.2	-0.016(0.166)	95.0	-0.001(0.110)	95.2	-0.001(0.110)	95.2
β_{31}	0.003(0.432)	94.6	0.014(0.370)	96.4	-0.056(0.399)	95.0	-0.021(0.333)	96.0	-0.056(0.376)	94.8	-0.028(0.309)	95.0	-0.056(0.376)	94.8	-0.028(0.309)	95.0	-0.028(0.309)	95.0
β_{32}	-0.054(0.482)	95.0	-0.028(0.356)	95.4	-0.055(0.382)	94.8	-0.018(0.279)	95.6	-0.047(0.350)	94.4	-0.015(0.245)	95.8	-0.047(0.350)	94.4	-0.015(0.245)	95.8	-0.015(0.245)	95.8

Table 3.4. Bootstrap standard errors and standard derivations (sd's in parentheses) for the proposed estimator based on 500 replications under Scenario 3.1.

	$\tau_{10} = 0.3, \tau_{20} = 0.3$		$\tau_{10} = 0.3, \tau_{20} = 0.7$		$\tau_{10} = 0.7, \tau_{20} = 0.7$	
	$n = 150$	$n = 300$	$n = 150$	$n = 300$	$n = 150$	$n = 300$
τ_{10}	0.098(0.100)	0.066(0.069)	0.097(0.099)	0.065(0.068)	0.053(0.050)	0.036(0.036)
τ_{20}	0.102(0.097)	0.069(0.067)	0.052(0.053)	0.037(0.036)	0.053(0.053)	0.036(0.035)
β_{11}	0.336(0.312)	0.239(0.238)	0.337(0.319)	0.239(0.243)	0.339(0.293)	0.221(0.221)
β_{12}	0.136(0.126)	0.094(0.095)	0.135(0.128)	0.092(0.095)	0.123(0.116)	0.079(0.079)
β_{21}	0.358(0.328)	0.253(0.252)	0.336(0.290)	0.230(0.227)	0.350(0.284)	0.229(0.228)
β_{22}	0.202(0.202)	0.140(0.142)	0.161(0.157)	0.109(0.105)	0.162(0.156)	0.108(0.106)
β_{31}	0.542(0.454)	0.380(0.373)	0.497(0.416)	0.334(0.338)	0.499(0.386)	0.320(0.316)
β_{32}	0.543(0.496)	0.372(0.375)	0.438(0.402)	0.289(0.294)	0.407(0.353)	0.261(0.257)

Table 3.5. Simulation results based on 500 replications under the proposed model and marginal models in Scenario 3.1 with $\tau_{10} = 0.3$ and $\tau_{20} = 0.7$.

	Proposed model				Marginal copula for $(T^{(1)}, T^{(2)})$				Marginal copula for $(T^{(2)}, T^{(3)})$				Marginal Cox PH model for each $T^{(k)}$			
	bias	sd	rmse		bias	sd	rmse		bias	sd	rmse		bias	sd	rmse	
n=150																
τ_{10}	0.020	0.099	0.101		0.300	0.000	0.300									
τ_{20}	0.007	0.053	0.053						0.041	0.051	0.065					
β_{11}	-0.037	0.319	0.321		0.318	0.388	0.502						0.008	0.364	0.364	
β_{12}	0.002	0.128	0.128		0.106	0.136	0.173						0.028	0.134	0.137	
β_{21}	-0.046	0.290	0.294						0.289	0.360	0.462		0.026	0.355	0.356	
β_{22}	-0.019	0.157	0.158						0.134	0.165	0.213		0.036	0.213	0.216	
β_{31}	-0.059	0.416	0.420		0.502	0.617	0.795		0.432	0.535	0.688		0.024	0.580	0.580	
β_{32}	-0.043	0.402	0.404		0.510	0.624	0.806		0.306	0.400	0.504		-0.044	0.594	0.596	
n=300																
τ_{10}	0.000	0.068	0.068		0.300	0.000	0.300									
τ_{20}	0.004	0.036	0.036						0.021	0.026	0.033					
β_{11}	-0.043	0.243	0.247		0.201	0.262	0.330						-0.024	0.256	0.257	
β_{12}	-0.002	0.095	0.095		0.064	0.076	0.099						0.020	0.100	0.102	
β_{21}	-0.018	0.227	0.228						0.213	0.281	0.353		0.041	0.257	0.260	
β_{22}	-0.003	0.105	0.105						0.098	0.124	0.158		0.036	0.145	0.149	
β_{31}	-0.020	0.338	0.339		0.287	0.358	0.459		0.317	0.402	0.512		0.042	0.420	0.422	
β_{32}	-0.019	0.294	0.295		0.281	0.365	0.461		0.181	0.219	0.285		-0.011	0.409	0.409

Figure 3.1. Estimated cumulative baseline hazard $\Lambda_{01}(t)$ based on the first 50 replications (colored red) with true cumulative baseline hazard $\Lambda_{01}(t_1) = \frac{2}{3}t_1^{3/2}$ (colored black) under Scenario 3.3 with $\tau_{10} = 0.3$ and $\tau_{20} = 0.3$.

Figure 3.2. Estimated cumulative baseline hazard $\Lambda_{02}(t)$ based on the first 50 replications (colored red) with true cumulative baseline hazard $\Lambda_{02}(t_2) = \frac{1}{3}t_2$ (colored black) under Scenario 3.3 with $\tau_{10} = 0.3$ and $\tau_{20} = 0.3$.

3.6 Real Data Analysis

We apply the proposed method to analyze the bone marrow transplant data from the motivating example for illustration purpose. Bone marrow transplants is a standard treatment option for many patients with acute leukemia. However, recovery after the transplantation is a complex process. Despite advances

Figure 3.3. Estimated cumulative baseline hazard $\Lambda_{03}(t)$ based on the first 50 replications (colored red) with true cumulative baseline hazard $\Lambda_{03}(t_3) = \frac{2}{3}t_3^{3/2}$ (colored black) under Scenario 3.3 with $\tau_{10} = 0.3$ and $\tau_{20} = 0.3$.

Figure 3.4. Estimated $\psi(t, t)$ based on the first 50 replications with $\tau_{10} = 0.3, \tau_{20} = 0.3$ (—); $\tau_{10} = 0.3, \tau_{20} = 0.7$ (--); $\tau_{10} = 0.7, \tau_{20} = 0.7$ ($\cdots$).

in treatment, relapse of acute leukemia after the BMT remains a significant therapeutic challenge and is the main cause of treatment failure. Allogeneic BMT also frequently cause transplantation related diseases such as acute GVHD, which can be severe and even life threatening for some people. It is important to assess dependencies between acute GVHD and death, relapse of Leukemia and death as well as acute GVHD and relapse, respectively, and in the meantime to identify risks factors for the occurrence of these three events, which can help understand the efficacy of the treatment and progression-free survival for leukemia patients.

Our study sample includes a cohort of 137 leukemia patients who received bone marrow transplants during the time period of 1984 to 1989 (Copelan et al., 1991). The data set collected from this cohort is available in the book of Klein and Moeschberger (2006) and can be downloaded from R package *KMsurv*. It contains observations of the death time, which might be censored by the study endpoint, the time to acute GVHD and the time to relapse of leukemia, which might be censored by the death time or the study endpoint. Two pairs of events: acute GVHD and death, and relapse of leukemia and death, are both subject to semi-competing risks. The acute GVHD and relapse of leukemia are considered as two nonterminal events, while death is the terminal event.

We focus on the effects of two risk factors: the disease-type having three levels, 1=acute lymphoblastic leukemia (ALL, 38 patients), 2=acute myelocytic leukemia (AML) low-risk first remission (54 patients) and 3=AML high-risk (45 patients); the donor age ranged from 2 to 56 years old at time of acute GVHD, relapse of leukemia or death. Among 137 patients in the study, 26 patients experienced acute GVHD and 16 of them dead within the study period; 42 patients experienced relapse of leukemia and 40 of them experienced death; 81 patients dead before the end of study. In addition, 5 patients experienced acute GVHD and relapse but dead by the end of study; 30 patients dead without experiencing acute GVHD and relapse. Therefore, the proposed joint modeling approach can be more appropriate for the analysis.

Table 3.6 presents the estimation results based on the proposed pseudo-likelihood estimation, in comparison with results obtained by two marginal copula models for times to acute GVHD and death and times to relapse and death, respectively, and marginal Cox PH models for three different events. The estimates of marginal copula models are obtained using the method developed by Chen (2012), while estimates of marginal Cox PH models are obtained by using R function *coxph*. The standard error by proposed method is estimated by the sample sd of estimates based on 200 randomly selected bootstrap samples using the with-replacement sampling method while controlling the sample censoring rates close to in the bmt data. According to the results of Kendall's τ from the proposed model and two marginal copula models, we have significant evidence that there is strong association between the acute GVHD/relapse and death. This finding indicates the joint proposed model is more suitable than the marginal models to analyzing this data set.

All these three methods indicate that the factor AML high-risk can significantly increase risks of leukemia relapse and death. Both two naive methods indicate that the donor age significantly affects the risk of acute GVHD. However, we find that donor age appears insignificant in the proposed model, after accounting for multiple associations between any pair of events acute GVHD, relapse and death. In the marginal Cox PH model, the factor ALL can significantly increase risks of leukemia relapse and death. While it becomes insignificant in the marginal copula and proposed models, after accounting for associations between any pair of events acute GVHD, relapse and death.

To further evaluate dependency between the two nonterminal events acute GVHD and relapse, we compute the estimated cross ratio $\psi(t, t)$ and present its curve over time interval $(0, 4)$ in Figure 3.5. As the observations of the acute GVHD mostly happens in time interval $(0, 3)$, we take a little longer time interval $(0, 4)$, to check the associations between acute GVHD and relapse during it. The curve differs from the reference line $\psi(t, t) = 1$, showing that the time to

Figure 3.5. Estimated $\psi(t, t)$ for the bone marrow transplant data.

acute GVHD and the time to relapse of leukemia are not independent. Thus, jointly analyzing the three endpoints using our proposed method is more reasonable than naive methods. In addition, from Figure 3.5, we find that acute GVHD is negatively related to relapse over a short period of time (about 1.5 months), during which all the observations of acute GVHD happens. This finding coincides with existing medical research work, which reported that some patients experienced GVHD actually had a lower risk of relapse than those patients who did not experience GVHD (Lee et al., 2005). It is also evident from the fact in this dataset that 20% of those patients experienced acute GVHD first and then experienced relapse, while 33% of the patients did not experienced acute GVHD but experienced relapse later.

Moreover, Figure 3.6 presents the estimated baseline survival curves obtained by the proposed method in comparison with the corresponding KM estimator, and implies that our proposed estimates of S_{k0} for $k = 1, 2, 3$ perform well and are close to the corresponding true curves to some extent.

Table 3.6. Results of estimates and standard errors (in parentheses) obtained under the proposed model and marginal models for the bone marrow transplant data.

Parameter	Proposed model	Marginal copula for $(T^{(1)}, T^{(2)})$	Marginal copula for $(T^{(2)}, T^{(3)})$	Marginal Cox PH for each $T^{(k)}$
Regression for acute GVHD				
AML high	-0.330(0.677)	-0.834(0.501)	-	-0.768(0.529)
ALL	0.383(0.512)	-0.124(0.405)	-	0.136(0.450)
Donor age	0.018(0.015)	0.022(0.009)*	-	0.056(0.022)*
Regression for leukemia relapse				
AML high	1.125(0.417)*	-	0.975(0.281)*	1.519(0.405)*
ALL	0.648(0.446)	-	0.541(0.283)	0.893(0.442)*
Donor age	-0.017(0.030)	-	0.005(0.007)	-0.004(0.017)
Regression for death				
AML high	0.766(0.272)*	0.777(0.241)*	0.604(0.197)*	0.979(0.275)*
ALL	0.488(0.261)	0.366(0.257)	0.228(0.225)	0.655(0.294)*
Donor age	-0.001(0.011)	-0.007(0.006)	-0.002(0.005)	0.015(0.012)
τ_{13} between acute GVHD and death	0.285(0.143)*	0.252(0.116)*	-	-
τ_{23} between leukemia relapse and death	0.674(0.035)*	-	0.707(0.047)*	-

* significant at 5% significance level

Figure 3.6. Estimated baseline survival functions using the proposed method (colored red) and the KM method (colored blue) based on the bone marrow transplant data. Dotted lines indicate the 95% confidence bounds of the KM estimator.

Joint correlation rank screening for semi-competing risks data

As already addressed in Section 1.6, we propose a model-free screening method allowing more flexibility, which utilizes the ranking of the correlation between the covariates and the joint survival distribution. Moreover, the proposed screening method is quite efficient in computation as it only needs estimation of the KM estimator for the joint survival function.

The rest of this chapter is organized as follows. Section 4.1 describes the semi-competing risks data and presents the development of the joint correlation rank screening procedure. Theoretical properties of the proposed method are established in Section 4.2. Simulation studies and the real data applications are demonstrated in Section 4.3 and Section 4.4, respectively. All technical details are relegated to Appendix D and some additional simulation results are provided in Appendix E.

4.1 Screening procedures

4.1.1 Semi-competing risks Data

Suppose that the sample in the analysis consists of observations of both the nonterminal and terminal event times from n subjects, and data of covariates with ultrahigh dimension p , $p \gg n$, and there may exist some $a_0 \in (0, 0.5)$ such that $\log p = O(n^{a_0})$. Let “event 1” be the nonterminal event, “event 2” be the terminal event, and $T^{(1)}$ and $T^{(2)}$ be their corresponding event times. Let C denote the external censoring time that is independent of $(T^{(1)}, T^{(2)})$. Denote the vector of ultrahigh-dimensional covariates as $X = (X_1, \dots, X_p)$. Due to possible right-censoring, for subject i we observe $T_i = \min(T_i^{(1)}, T_i^{(2)}, C_i)$, $\tilde{T}_i = \min(T_i^{(2)}, C_i)$, $X_i = (X_{i1}, \dots, X_{ip})$, and censoring indicators $\delta_i^{(1)} = I(T_i = T_i^{(1)})$, $\delta_i^{(2)} = I(\tilde{T}_i = T_i^{(2)})$, and $I(\cdot)$ is the indicator function. Thus the observed data are denoted by $\mathcal{O} = \{T_i, \tilde{T}_i, X_i, \delta_i^{(1)}, \delta_i^{(2)}\}_{i=1}^n$.

4.1.2 Joint correlation rank screening procedure

Let $H(t^{(1)}, t^{(2)}|X) = P(T^{(1)} \geq t^{(1)}, T^{(2)} \geq t^{(2)}|X)$ be the joint survival function of the nonterminal and terminal event times conditional on information of covariate X . To identify which covariates among the p ones contribute to the joint distribution of $T^{(1)}$ and $T^{(2)}$, we first define the active covariate index set ignoring model specifications:

$$\mathcal{A} = \{k : H(t^{(1)}, t^{(2)}|X) \text{ depends on } X_k, k = 1, \dots, p\}.$$

X_k is known as an active covariate if $k \in \mathcal{A}$. Otherwise, X_k is inactive and $k \in \mathcal{A}^c$, where $\mathcal{A}^c$ is the complement set of $\mathcal{A}$. Our goal is to identify the active set $\mathcal{A}$ as precisely as possible based on the observed sample $\mathcal{O}$.

Without loss of generality, we assume in this chapter that $E(X_k) = 0$ for $k = 1, \dots, p$. This assumption is valid in practice when all the covariates are centralized. Let $S(t^{(1)}, t^{(2)}) = P(T^{(1)} \geq t^{(1)}, T^{(2)} \geq t^{(2)})$ be the unconditional joint

survival function of two event times. Define $\mathbf{R}(T^{(1)}, T^{(2)}) = E\{XS(T^{(1)}, T^{(2)})\}$, which represents the vector of some utility measures for correlation between covariates and the distribution of $(T^{(1)}, T^{(2)})$, with the k th component $\mathbf{R}_k(T^{(1)}, T^{(2)}) = E\{X_k S(T^{(1)}, T^{(2)})\} = \text{cov}\{X_k, S(T^{(1)}, T^{(2)})\}$. We define a marginal utility measure for the k th covariate by

$$r_k = [\mathbf{R}_k(T^{(1)}, T^{(2)})]^2 = [E\{X_k S(T^{(1)}, T^{(2)})\}]^2, \quad (4.1)$$

for $k = 1, \dots, p$. As a population version, the utility r_k reflects the correlation between X_k and the distribution of $(T^{(1)}, T^{(2)})$. The definition in (4.1) is an extension of methods for univariate survival time (Song et al., 2014; Zhang et al., 2017) to bivariate data. Intuitively, the unconditional joint survival function of $(T^{(1)}, T^{(2)})$, $S(T^{(1)}, T^{(2)})$, is expected to contain the whole information of $(T^{(1)}, T^{(2)})$. If X_k and $(T^{(1)}, T^{(2)})$ are independent, X_k and $S(T^{(1)}, T^{(2)})$ should be independent; hence $r_k = 0$. Otherwise, it is reasonable to expect $r_k > 0$ if X_k and $(T^{(1)}, T^{(2)})$ are dependent.

4.1.3 Estimation

Define an auxiliary random variable $T_{12} = \min(T^{(1)}, T^{(2)})$. Then T_{12} is independent of C . In other words, T_{12} and $T^{(2)}$ are subject to independent censoring by C . Therefore, the joint survival function of T_{12} and $T^{(2)}$ can be estimated by the KM estimator. Note that $P(T_{12} \geq t, T^{(2)} \geq \tilde{t}) = P(T^{(1)} \geq t, T^{(2)} \geq t, T^{(2)} \geq \tilde{t}) = P(T^{(1)} \geq t, T^{(2)} \geq \tilde{t})$ when $0 \leq t \leq \tilde{t}$. Following Lin and Ying (1993) and Lakhal et al. (2008), we construct a nonparametric estimator for the joint survival function of $(T^{(1)}, T^{(2)})$ in the upper wedge as

$$\hat{S}(t, \tilde{t}) = \frac{\sum_{i=1}^n I(T_i \geq t, \tilde{T}_i \geq \tilde{t})}{n\hat{\pi}(\tilde{t})}, 0 \leq t \leq \tilde{t}, \quad (4.2)$$

where $\hat{\pi}(\cdot)$ is the KM estimator of the survival function of censoring time C , that is $\pi(t) = P(C \geq t)$, and

$$\hat{\pi}(\tilde{t}) = \prod_{i=1}^n \left\{ 1 - \frac{1}{\sum_{j=1}^n I(\tilde{T}_j \geq \tilde{T}_i)} \right\}^{(1-\delta_i^{(2)})I(\tilde{T}_i \leq \tilde{t})}. \quad (4.3)$$

Thus the estimator of r_k can be obtained by

$$\hat{r}_k = \left\{ \frac{1}{n} \sum_{i=1}^n X_{ik} \hat{S}(T_i, \tilde{T}_i) \right\}^2. \quad (4.4)$$

Consequently, the screening procedure can be proposed by ranking $\hat{r}_k$ from the largest to smallest and selecting the top ones and corresponding indexes as an active set. Specifically, we can define the estimated active set as

$$\hat{\mathcal{A}} = \{k : \hat{r}_k \geq cn^{-w}, k = 1, \dots, p\}, \quad (4.5)$$

where $c > 0$ and $w \in [0, 1/2)$ are two constants, which can be determined by a user-defined size d of the active set $\mathcal{A}$ and d is often chosen as the integer part of $n/\log(n)$ following the thresholding rule in Fan and Lv (2008).

4.1.4 Screening steps

We termed the proposed method a joint correlation rank screening procedure (JCR), and summarize the implementation of the JCR as follows.

1. Estimate the KM estimator of the survival function of C by equation (4.3) and the joint survival function of $(T^{(1)}, T^{(2)})$ by (4.2).
2. Obtain the marginal utility measure estimator $\hat{r}_k$ by (4.4).
3. Rank $\hat{r}_k$ from the largest to smallest and select the top d ones as the active covariates, where d is the chosen model size.

4.2 Theoretical result

Under the following regularity conditions, we establish the sure independence and ranking consistency properties for the proposed screening method. The rank consistency property of the JCR procedure refers to that the values of r_k for important (active) covariates are greater than those of unimportant (inactive)

ones. The sure independence property ensures that the screening procedure can reduce the covariate dimension to an extent determined by the number of truly important (active) covariates (Zhou and Zhu, 2017).

Condition 4.1. *There exists a constant $\varsigma > 0$ satisfying*

$$S(\varsigma, \varsigma)\pi(\varsigma) > 0.$$

Condition 4.2. *There exists a positive constant ξ such that*

$$\max_{1 \leq k \leq p} E(X_k^2) < \xi.$$

Condition 4.3. *It holds that*

$$\min_{k \in \mathcal{A}} r_k \geq c_0 n^{-w}$$

for some constants $c_0 > 0$ and $w \in [0, 1/2)$.

Condition 4.1, also used in Lin and Ying (1993), is required for the derivation of asymptotic property for the KM estimator of the joint survival function. Conditions 4.2 and 4.3 are often used in defining screening methods. See Fan and Lv (2008), Zhu et al. (2011) and Zhang et al. (2017) for examples. Particularly, Condition 4.2 requires that all covariates have finite variances, which can be satisfied by covariates under various commonly used distributions, including the normal, uniform, binomial and multinomial distributions and so forth. Condition 4.3 emphasizes the level of the minimal signal of active covariates has to be high enough, so that it controls the error rate in selection of important covariates.

Theorem 4.1. *(Ranking consistency and sure independence) Under Conditions 4.1 and 4.2, there exist positive constants c, c_1, c_2, c_3 such that*

$$\Pr\left(\max_{1 \leq k \leq p} |\hat{r}_k - r_k| \geq cn^{-w}\right) \leq p \left\{ \exp(-c_1 n^{-2w}) + \exp\left(-\frac{c_2 n^{1/2-w}}{(\log \log n)^{1/2}} + c_3\right) \right\}.$$

Under Conditions 4.1-4.3, it holds that

$$\Pr\left(\mathcal{A} \subseteq \hat{\mathcal{A}}\right) \geq 1 - s \left\{ \exp(-c_1 n^{-2w}) + \exp\left(-\frac{c_2 n^{1/2-w}}{(\log \log n)^{1/2}} + c_3\right) \right\},$$

where $s = |\mathcal{A}|$ is the number of variables in $\mathcal{A}$.

Denote $X_{\mathcal{A}} = \{X_j : j \in \mathcal{A}\}$ and $X_{\mathcal{A}^c} = \{X_j : j \in \mathcal{A}^c\}$.

Theorem 4.2. *Assume that (i) $(T^{(1)}, T^{(2)})$ and $X_{\mathcal{A}^c}$ are conditionally independent given $X_{\mathcal{A}}$; (ii) $X_{\mathcal{A}}$ is independent of $X_{\mathcal{A}^c}$. Under Condition 4.3, we have that*

$$\max_{k \in \mathcal{A}^c} r_k < \min_{k \in \mathcal{A}} r_k,$$

and $r_k = 0$ if and only if $k \in \mathcal{A}^c$. Furthermore, under Conditions 4.1-4.3, there exist positive constants c, c_1, c_2, c_3 such that

$$\Pr\left(\max_{k \in \mathcal{A}^c} \hat{r}_k < \min_{k \in \mathcal{A}} \hat{r}_k\right) \geq 1 - 2p\left\{\exp(-c_1 n^{-2w}) + \exp\left(-\frac{c_2 n^{1/2-w}}{(\log \log n)^{1/2}} + c_3\right)\right\}.$$

The above theorem ensures that the proposed JCR method tends to rank the active covariates above the inactive ones with high probability.

The detailed proofs of both theorems can be found in the Appendix D.

4.3 Simulation Study

In this section, we conduct simulation studies to assess the finite sample performance of the proposed screening procedure (JCR) for both the nonterminal and terminal event times in comparison with the existing method (CR) by Zhang et al. (2017) for the terminal event time only. The joint survival function for the nonterminal and terminal event times in the simulation is specified by a Clayton copula function with an unknown constant parameter α one-to-one corresponding to Kendall's τ and $\alpha = 2\tau/(1 - \tau)$ (Nelsen, 2007). We set the values of τ to be 0.2, 0.8, -0.2, -0.8 for different levels of associations between the two event times.

We consider the following scenarios to evaluate the performance of the proposed screening method under different marginal models for the terminal and nonterminal event times:

Scenario 4.1: $\lambda_1(t^{(1)}|X) = 2t^{(1)} \exp(\beta'_1 X)$, $\lambda_2(t^{(2)}|X) = t^{(2)} \exp(\beta'_2 X)$,

Scenario 4.2: $\lambda_1(t^{(1)}|X) = 2t^{(1)} \exp(\beta'_3 X)$, $\lambda_2(t^{(2)}|X) = t^{(2)} \exp(\beta'_4 X)$,

Scenario 4.3: $\log(t^{(1)}) = \beta'_5 X + \epsilon$, $\lambda_2(t^{(2)}|X) = \frac{1}{3} \exp(\beta'_6 X)$,

In Scenario 4.1-4.3, $T^{(1)}$, $T^{(2)}$ are generated by the models specified above, where $\lambda_{01}(t^{(1)}) = 2t^{(1)}$, $\lambda_{02}(t^{(2)}) = t^{(2)}$, ϵ follows a Logistic distribution with location 0 and scale 0.2, and covariate $X = (X_1, \dots, X_p)'$ is generated from a multivariate normal distribution with mean $\mathbf{0}$ and correlation matrix $\Sigma = (0.9^{|i-j|})$ for $i, j = 1, \dots, p$. Particularly, in Scenario 4.1, we set $\beta_1 = (0, 0, 0, 0.3, 0.3, 0.3, 0, \dots, 0)'$ and $\beta_2 = (0.3, 0.3, 0.3, 0, \dots, 0)'$. In this scenario, the active covariates for $T^{(1)}$ are completely different from those for $T^{(2)}$. In Scenario 4.2, $T^{(1)}$ and $T^{(2)}$ share two common important covariates (X_3, X_4) with $\beta_3 = (0, 0, 0.3, 0.3, 0.3, 0.3, 0, \dots, 0)'$, $\beta_4 = (0.3, 0.3, 0.3, 0.3, 0, \dots, 0)'$. In Scenario 4.3, there are two common active covariates (X_3, X_4) for $T^{(1)}$ and $T^{(2)}$ with $\beta_5 = (0, 0, 0.3, 0.3, 0.3, 0.3, 0, \dots, 0)'$ and $\beta_6 = (0.3, -0.3, -0.3, -0.3, \dots, 0)'$ so that X_3 and X_4 both have positive impacts on $T^{(1)}$ and $T^{(2)}$. Thus the signs of the effects of X_3 and X_4 are opposite in the marginal AFT model and the marginal Cox PH model. We take the overall censoring time C from a uniform distribution to yield nearly 60% censoring rates for both $T^{(1)}$ and $T^{(2)}$. The observed times to the nonterminal and terminal events for subject i ($i = 1, \dots, n$) are $T_i = \min(T_i^{(1)}, T_i^{(2)}, C_i)$ and $\tilde{T}_i = \min(T_i^{(2)}, C_i)$, respectively. We consider different values of sample size ranging from 100 to 300, the number of covariates $p = 2000$, and the chosen model size $d = \lceil n/\log n \rceil$, where $\lceil x \rceil$ represents the integer part of x . This choice of model size d is commonly adopted in the literature, see for examples, Fan and Lv (2008), Zhu et al. (2011), Zhou and Zhu (2017) and Zhang et al. (2017).

Based on 500 simulation runs, we evaluate the performance of two screening procedures using the following criteria (Li et al., 2012b): selection proportions among 500 replications for each active covariate, for all active covariates associated with $T^{(2)}$ and for all active covariates associated with $(T^{(1)}, T^{(2)})$, and quantiles of the minimum model size including all active covariates. The contrast between “selection proportion” and “the minimum model size including all

active covariates” in our analysis is similar to the contrast between “sensitivity” and “specificity” in Emura and Chen (2016). The selection proportion measures the performance under a fixed model size $d = \lceil n/\log n \rceil$, where $\lceil x \rceil$ represents the integer part of x . The minimum model size including all active covariates measures the performance under a fixed selection criterion that is to select all the active covariates.

The proposed JCR method is applied to the observed data of two event times in comparison with the marginal correlation ranking method (CR) of Zhang et al. (2017) applied to the observed data of the terminal event only. Note that the terminal event is subject to independent censoring, while the nonterminal event could be subject to a dependent censoring by the terminal event, thus the standard KM estimator can be applied to estimate the marginal survival function of the terminal event time, instead of the marginal survival function of the nonterminal event time (Lakhal et al., 2008). In turn, the KM estimator-based CR method can be biased when applying to screen variables for the nonterminal event.

Tables 4.1-4.3 summarize the screening results under Scenarios 4.1-4.3, respectively, where results in Table 4.1 are based on data with $\tau = 0.2, 0.8, -0.2, -0.8$ under Scenario 4.1, while for simplicity, Tables 4.2-4.3 include the results with $\tau = 0.2, 0.8$ only under Scenarios 4.2-4.3, respectively. We skip the results for $\tau = -0.2, -0.8$ due to their similar trends to those for $\tau = 0.2, 0.8$. It is observed that under all these three scenarios and regardless of levels of associations between the two event times, the proposed JCR method outperforms the CR method in terms of all three criteria for selection proportions. These observed trends are due to the marginal CR method ignores the observed nonterminal event times, which are informative to the terminal event times. As the sample size increases, the selection proportions increase, and the difference of selection proportions between two methods decreases. For large sample cases (e.g. $n = 300$), our proposed method can correctly select all active covariates with selection proportions that are all close to 1.

Comparing results in Tables 4.1-4.2, the difference between both methods is greater in Scenario 4.2 than that in Scenario 4.1, thus the proposed JCR method gains more efficiency when there are some overlapped active covariates associated with $T^{(1)}$ and $T^{(2)}$ under Scenario 4.2. The difference of two methods in Scenario 4.3 turned out to be the greatest among all three scenarios as the marginal models for two event times are quite different in Scenario 4.3 and the information of $T^{(1)}$ accounted by the JCR method helps us recover some valid information of $T^{(2)}$ under the complex dependent censoring. Therefore the proposed JCR method is overwhelmingly superior to the marginal CR method.

Results of five quantiles of the minimum model size which includes all active covariates associated with $T^{(2)}$ are summarized in Tables 4.4-4.6 under Scenarios 4.1-4.3, respectively. We find using the JCR method these quantiles are all smaller than those obtained by the CR method, especially in the cases with small sample sizes. When the sample size is large, say $n = 300$, we have chosen the model size $\lceil n/\log n \rceil = 52$. All quantiles of the minimum model size resulting from the JCR method are shown to be small (less than or equal to 17), indicating a high probability that the estimated active set includes all active covariates. Hence, the corresponding selection proportion is around 98% as given in Table 4.1-4.3. However, the 95% quantile of the counterpart minimum model size obtained by the CR method under Scenario 4.3 is much greater, showing that the CR method requires a much bigger model size to ensure the inclusion of all active covariates.

Tables 4.7-4.9 show the results of five quantiles of the minimum model size which includes all active covariates associated with $(T^{(1)}, T^{(2)})$ by the proposed JCR method under Scenarios 4.1-4.3, respectively. It is seen that the minimum model size is decreasing as sample size increases. The JCR screening for $(T^{(1)}, T^{(2)})$ also requires a large enough sample size (say $n \geq 300$) for successful selection of all active covariates with a very high probability ranging from 98% to 100% (as in Tables 4.1-4.3).

Our simulation experience shows that different choices of copulas do not influence the behavior of the proposed screening method. Take Gumbel copula for another example, the corresponding simulation results are provided in Appendix E due to space limit, and exhibit similar trends to those seen in the Clayton copula design above.

4.4 Real Data Analysis

We illustrate the practical utility of the proposed method by the NKI dataset of breast cancer patients from National Kanker Institute (the Netherlands). More details about the NKI data can be found in Haibe-Kains et al. (2012), Van De Vijver et al. (2002) and Van't Veer et al. (2002b). The semi-competing risks data of interest in this study include the time for DMFS and OS where the nonterminal event is the distant metastasis and the terminal event is death.

We consider a total of $n = 295$ patients with valid survival time observations and each patient had $p = 14318$ gene expression values. The data set is available at <http://compbio.dfci.harvard.edu/pubs/sbtpaper/>. We are interested in identifying genes that have great impacts on patients' OS (ranging from 0.02 to 18.34 years) and DMFS (ranging from 0.02 to 18.34 years). Out of all 295 studied patients, 101 of them experienced distant metastasis and 79 experienced death within the study period, corresponding to around 65% censoring rate for DMFS and 75% censoring rate for OS.

Given the sample size, we choose the model size to be $\lceil 295 / \log 295 \rceil = 51$. Applying the proposed JCR method for the times to DMFS and OS, we calculate $\hat{r}_k$ for $k = 1, \dots, p$, and then select the top 51 values of $\hat{r}_k$. The corresponding gene names and their unique NKI IDs are presented in Table 4.10.

Using the KM estimator for the survival function of the terminal event time and applying the marginal CR method to the observations of OS, we can also select 51 genes that have the greatest impacts on OS only. The corresponding gene names and unique NKI IDs are presented in Table 4.10. We find that 40 genes bolded in Table 4.10 are both selected by the proposed JCR method for the times to DMFS and OS, and the marginal CR method for OS, indicating that these 40 genes are not only strongly associated with DMFS and OS jointly, but also substantially associated with OS only.

The rest 11 genes are different between the JCR and CR methods due to whether the information of the nonterminal event is taken into account or not.

104 Semi-competing risks data with ultrahigh dimensional covariates

Table 4.1. Selection proportions for each active covariate, for all active covariates (i.e. $X_1, \dots, X_4$, denoted by $X_{1,\dots,4}$) associated with $T^{(2)}$ and for all active covariates (i.e. $X_1, \dots, X_6$, denoted by $X_{1,\dots,6}$) associated with $(T^{(1)}, T^{(2)})$ based on 500 replications under Scenario 4.2.

τ	n	Method	X_1	X_2	X_3	X_4	X_5	X_6	$X_{1,\dots,4}$	$X_{1,\dots,6}$
0.2	100	JCR	0.536	0.576	0.592	0.557	0.488	0.388	0.436	0.252
		CR	0.412	0.462	0.406	0.306	0.220	0.170	0.278	0.090
	200	JCR	0.920	0.952	0.960	0.942	0.912	0.852	0.904	0.786
		CR	0.868	0.894	0.886	0.792	0.658	0.570	0.818	0.486
	300	JCR	1.000	0.998	0.998	0.998	0.992	0.986	0.998	0.980
		CR	0.986	0.994	0.990	0.968	0.936	0.854	0.984	0.846
0.8	100	JCR	0.446	0.536	0.576	0.558	0.516	0.426	0.346	0.222
		CR	0.422	0.472	0.406	0.320	0.248	0.180	0.286	0.102
	200	JCR	0.890	0.936	0.958	0.950	0.918	0.862	0.870	0.774
		CR	0.888	0.900	0.894	0.800	0.660	0.518	0.834	0.466
	300	JCR	0.992	0.998	0.998	0.998	0.994	0.982	0.988	0.974
		CR	0.992	0.996	0.992	0.968	0.918	0.840	0.986	0.826
-0.2	100	JCR	0.576	0.654	0.662	0.606	0.550	0.452	0.492	0.298
		CR	0.412	0.488	0.418	0.332	0.224	0.158	0.304	0.084
	200	JCR	0.940	0.972	0.984	0.958	0.942	0.896	0.932	0.844
		CR	0.874	0.898	0.886	0.782	0.660	0.562	0.816	0.488
	300	JCR	1.000	1.000	1.000	1.000	0.998	0.994	1.000	0.994
		CR	0.990	0.990	0.988	0.982	0.930	0.850	0.986	0.836
-0.8	100	JCR	0.656	0.718	0.738	0.696	0.638	0.534	0.572	0.374
		CR	0.466	0.508	0.450	0.374	0.270	0.178	0.336	0.118
	200	JCR	0.964	0.972	0.980	0.984	0.960	0.924	0.952	0.894
		CR	0.886	0.908	0.878	0.786	0.680	0.548	0.820	0.480
	300	JCR	0.998	1.000	1.000	1.000	1.000	0.996	0.998	0.994
		CR	0.986	0.992	0.986	0.966	0.912	0.860	0.978	0.842

Table 4.2. Selection proportions for each active covariate, for all active covariates (i.e. $X_1, \dots, X_4$, denoted by $X_{1,\dots,4}$) associated with $T^{(2)}$ and for all active covariates (i.e. $X_1, \dots, X_6$, denoted by $X_{1,\dots,6}$) associated with $(T^{(1)}, T^{(2)})$ based on 500 replications under Scenario 4.2.

τ	n	Method	X_1	X_2	X_3	X_4	X_5	X_6	$X_{1,\dots,4}$	$X_{1,\dots,6}$
0.2	100	JCR	0.528	0.602	0.666	0.642	0.566	0.474	0.408	0.286
		CR	0.344	0.410	0.388	0.348	0.266	0.176	0.190	0.080
	200	JCR	0.918	0.950	0.982	0.980	0.946	0.886	0.906	0.826
		CR	0.822	0.864	0.874	0.836	0.714	0.626	0.728	0.526
	300	JCR	0.998	1.000	1.000	1.000	0.996	0.992	0.998	0.988
		CR	0.976	0.986	0.988	0.984	0.954	0.892	0.962	0.872
0.8	100	JCR	0.478	0.578	0.634	0.646	0.562	0.462	0.366	0.244
		CR	0.344	0.398	0.380	0.344	0.282	0.192	0.194	0.096
	200	JCR	0.900	0.944	0.974	0.976	0.942	0.890	0.884	0.808
		CR	0.822	0.864	0.878	0.834	0.716	0.580	0.714	0.494
	300	JCR	0.992	1.000	1.000	0.998	0.998	0.988	0.990	0.982
		CR	0.984	0.994	0.988	0.980	0.932	0.860	0.966	0.840

Table 4.3. Selection proportions for each active covariate, for all active covariates (i.e. $X_1, \dots, X_4$, denoted by $X_{1,\dots,4}$) associated with $T^{(2)}$ and for all active covariates (i.e. $X_1, \dots, X_6$, denoted by $X_{1,\dots,6}$) associated with $(T^{(1)}, T^{(2)})$ based on 500 replications under Scenario 4.3.

τ	n	Method	X_1	X_2	X_3	X_4	X_5	X_6	$X_{1,\dots,4}$	$X_{1,\dots,6}$
0.2	100	JCR	0.682	0.852	0.906	0.930	0.890	0.850	0.664	0.616
		CR	0.156	0.256	0.336	0.322	0.256	0.194	0.116	0.062
	200	JCR	0.980	1.000	1.000	1.000	1.000	1.000	0.980	0.980
		CR	0.418	0.656	0.780	0.732	0.628	0.544	0.384	0.312
	300	JCR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		CR	0.666	0.874	0.942	0.936	0.890	0.784	0.660	0.584
0.8	100	JCR	0.674	0.844	0.904	0.940	0.906	0.856	0.664	0.610
		CR	0.148	0.280	0.326	0.296	0.260	0.184	0.110	0.052
	200	JCR	0.974	0.998	1.000	1.000	1.000	1.000	0.972	0.972
		CR	0.388	0.662	0.756	0.760	0.658	0.558	0.364	0.298
	300	JCR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		CR	0.678	0.878	0.924	0.928	0.876	0.796	0.662	0.592

Table 4.4. Quantiles of the minimum model size which includes all active covariates (i.e. X_1, X_2, X_3) associated with $T^{(2)}$ based on 500 replications under Scenario 4.1.

τ	n	Method	5%	25%	50%	75%	95%
0.2	100	JCR	3.00	8.75	32.00	144.25	730.35
		CR	3.00	17.00	69.50	295.50	1149.25
	200	JCR	3.00	4.00	6.00	12.00	87.05
		CR	3.00	4.00	6.00	21.25	177.10
	300	JCR	3.00	4.00	5.00	6.25	12.00
		CR	3.00	4.00	6.00	7.10	18.05
0.8	100	JCR	4.00	11.00	41.00	169.25	768.65
		CR	3.00	16.00	73.00	254.25	1025.95
	200	JCR	3.00	5.00	7.00	15.25	118.65
		CR	3.00	4.00	7.00	22.00	196.00
	300	JCR	3.00	3.00	4.00	5.00	16.05
		CR	3.00	3.00	4.00	5.00	18.00
-0.2	100	JCR	3.00	7.00	22.00	114.25	605.50
		CR	3.95	15.00	71.00	276.50	1139.00
	200	JCR	3.00	3.00	5.50	9.00	56.00
		CR	3.00	3.00	6.00	21.50	161.25
	300	JCR	3.00	3.00	5.00	6.00	10.00
		CR	3.00	3.00	5.00	6.00	17.00
-0.8	100	JCR	3.00	6.00	15.00	66.25	401.50
		CR	3.00	12.00	63.00	238.00	1191.00
	200	JCR	3.00	4.00	5.00	8.00	35.00
		CR	3.00	4.00	6.00	21.25	189.25
	300	JCR	3.00	3.00	4.00	5.00	8.00
		CR	3.00	3.00	4.00	5.00	17.00

Table 4.5. Quantiles of the minimum model size which includes all active covariates (i.e. $X_1, \dots, X_4$) associated with $T^{(2)}$ based on 500 replications under Scenario 4.2.

τ	n	Method	5%	25%	50%	75%	95%
0.2	100	JCR	4.00	9.00	34.50	136.00	774.75
		CR	5.00	31.00	117.50	434.25	1451.15
	200	JCR	4.00	5.00	7.00	12.00	80.05
		CR	4.00	5.00	12.00	42.25	263.40
	300	JCR	4.00	4.00	6.00	7.00	12.00
		CR	4.00	4.00	5.00	8.00	42.10
0.8	100	JCR	4.95	13.00	40.00	172.50	804.80
		CR	5.00	33.75	118.50	478.75	1375.05
	200	JCR	4.00	5.00	8.00	16.00	119.10
		CR	4.00	5.00	12.00	45.00	230.40
	300	JCR	4.00	4.00	5.00	8.00	15.00
		CR	4.00	4.00	5.00	8.00	37.05

Table 4.6. Quantiles of the minimum model size which includes all active co-variates (i.e. $X_1, \dots, X_4$) associated with $T^{(2)}$ based on 500 replications under Scenario 4.3.

τ	n	Method	5%	25%	50%	75%	95%
0.2	100	JCR	4.00	7.00	11.00	34.50	323.45
		CR	8.00	70.75	287.50	818.25	1837.25
	200	JCR	4.95	6.00	7.00	10.00	16.00
		CR	5.00	17.00	67.00	268.75	954.45
	300	JCR	4.00	6.00	7.00	8.00	12.00
		CR	4.00	7.00	20.00	99.25	553.20
0.8	100	JCR	4.00	7.00	12.00	35.25	346.05
		CR	11.00	73.50	291.50	830.00	1897.35
	200	JCR	5.00	6.00	8.00	10.00	21.10
		CR	5.00	18.00	81.00	240.00	1122.40
	300	JCR	5.00	6.00	7.00	9.00	11.00
		CR	5.00	7.00	19.00	87.75	570.35

Table 4.7. Quantiles of the minimum model size which includes all active covariates (i.e. $X_1, \dots, X_6$) associated with $(T^{(1)}, T^{(2)})$ obtained by the proposed JCR method, based on 500 replications under Scenario 4.1.

τ	n	5%	25%	50%	75%	95%
0.2	100	3.00	8.75	32.00	144.25	730.35
	200	5.00	7.00	10.00	29.25	202.05
	300	6.00	6.00	7.00	9.00	22.10
0.8	100	7.00	25.75	85.00	309.75	996.00
	200	6.00	7.00	11.00	32.00	217.00
	300	6.00	6.00	7.00	9.00	21.00

Table 4.8. Quantiles of the minimum model size which includes all active covariates (i.e. $X_1, \dots, X_6$) associated with $(T^{(1)}, T^{(2)})$ obtained by the proposed JCR method, based on 500 replications under Scenario 4.2.

τ	n	5%	25%	50%	75%	95%
0.2	100	6.00	17.00	63.00	244.75	956.10
	200	6.00	6.00	9.00	21.25	138.10
	300	6.00	6.00	6.00	8.00	17.05
0.8	100	7.00	22.00	69.50	254.25	959.60
	200	6.00	7.00	10.00	27.00	158.05
	300	6.00	6.00	7.00	9.00	22.00

Table 4.9. Quantiles of the minimum model size which includes all active covariates (i.e. $X_1, \dots, X_6$) associated with $(T^{(1)}, T^{(2)})$ obtained by the proposed JCR method, based on 500 replications under Scenario 4.3.

τ	n	5%	25%	50%	75%	95%
0.2	100	6.00	8.00	13.00	42.50	367.60
	200	6.00	6.00	8.00	10.00	16.00
	300	6.00	6.00	7.00	8.00	12.00
0.8	100	6.00	8.00	14.00	44.25	359.10
	200	6.00	6.00	8.00	10.00	21.10
	300	6.00	6.00	7.00	9.00	11.00

By the method of Lakhal et al. (2008) and under the Clayton copula assumption for the joint survival function of DMFS and OS, the value of Kendall's τ is estimated to be 0.92. Based on our experience in simulation, the JCR method has higher selection proportions than those of the CR method in the case of $\tau = 0.8, n = 300$, similar to the current situation in this example.

The above selected genes by the JCR method can be used to build a prediction scheme for overall survival. In medical studies, it is a common practice to re-fit a multivariate regression model based on the selected genes (e.g., Lossos et al. (2004)), thus the JCR selected genes can be used to re-fit in a joint-copula regression model. Alternatively, one can use Tukey's compound covariate predictor (Tukey, 1993) that combines the results of univariate analyses without going through a multivariate analysis. The compound covariate has been successfully employed in many medical studies (e.g., Beer et al. (2002), Wang et al. (2005)) and biostatistical studies (Matsui et al. (2012); Emura and Michimae (2017)).

Table 4.10. Gene selected by JCR and CR methods

Genes selected			
Proposed joint method for $T^{(1)}$ and $T^{(2)}$ (JCR)		Marginal method for $T^{(2)}$ (CR)	
Gene name	Gene ID	Gene name	Gene ID
Contig56678_RC	18	Contig56678_RC	18
NM_000909	350	NM_002343	281
NM_000926	403	NM_000909	350
Contig18502_RC	898	NM_001647	379
NM_003258	1112	NM_001657	409
NM_002570	1164	Contig18502_RC	898
NM_002614	1577	NM_003258	1112
Contig57138_RC	1794	NM_002614	1577
NM_020130	1960	NM_020130	1960
NM_004143	2014	NM_004143	2014
NM_003462	2182	NM_003462	2182
Contig53307_RC	2708	Contig53307_RC	2708
AB040926	3004	AB040926	3004
NM_004496	3945	NM_004496	3945
AL133619	4121	AL133619	4121
NM_006096	4862	NM_006006	4602
NM_005375	4907	Contig46597_RC	4692
NM_003981	4995	NM_005375	4907
NM_006113	5174	NM_006115	5177
NM_006115	5177	Contig53968_RC	5179
Contig35629_RC	5217	NM_004701	5332
X82693	5257	Contig48328_RC	5818
NM_004701	5332	NM_020974	6240
Contig54913_RC	5726	Contig46937_RC	6466
Contig48328_RC	5818	NM_005733	6802
NM_020974	6240	NM_006551	7201
Contig46937_RC	6466	NM_014668	7676
NM_005733	6802	Contig47405_RC	8289
NM_014668	7676	NM_016267	8527
Contig47405_RC	8289	Contig51369_RC	8761
Contig38901_RC	8466	AL049265	8915
Contig51369_RC	8761	Contig55725_RC	8960
AL049265	8915	K02403	9216
Contig55725_RC	8960	Contig39226_RC	9953
K02403	9216	Contig45537_RC	10018
Contig39226_RC	9953	Contig41887_RC	10187
Contig45537_RC	10018	AB020689	10732
Contig41887_RC	10187	U79293	10826
Contig41413_RC	10279	AF007153	10927
AB020689	10732	NM_017680	11652
AF007153	10927	Contig24609_RC	11847
NM_017680	11652	NM_001185	12377
Contig24609_RC	11847	AL050227	13163
NM_001185	12377	NM_002051	13225
AL050227	13163	NM_001333	13246
NM_002051	13225	NM_000633	13325
NM_001333	13246	NM_000662	13388
NM_000633	13325	NM_000685	13414
NM_000662	13388	Contig14284_RC	13447
Contig14284_RC	13447	Contig58301_RC	13923
Contig58301_RC	13923	AF067420	14318

Genes selected by both methods are bolded.

Discussion and Future Works

In this chapter, we give a summary of the works done in this thesis and explore some relevant topics for further investigation.

5.1 Conclusion and discussion for Chapters 2-4

- (i) In Chapter 2, we have proposed a copula regression model with random effects for joint analysis of both non-terminal and terminal event times under semi-competing risks and clustered structure. The MLE is developed for regression parameters, while the baseline hazards are estimated using the nonparametric MLE. The proposed method extends the joint regression analysis studied in Chen (2012) to the clustered semi-competing risks data setting.

The appealing feature of our method is that it does not require gamma distribution assumption for random effects as is in the work of Emura et al. (2017), and thus allows for a flexible modeling framework and wider range of applications. We provide rigorous justification of the estimators through the counting process and martingale central limit theory. Our theoretical results show that the estimators have desired asymptotic properties, which

facilitate further inference. The proposed estimation procedure is implemented through an MCEM algorithm, which generally results in stable estimates in our experiment with both simulated and real data sets. Furthermore, our simulation results show that the proposed method performs satisfactorily in terms of the bias and standard error for either gamma or normally distributed random effects, and it is also robust against misspecification of the random effects distribution. In addition, under the model with random effects being gamma distributed, our method provides more favorable estimates of the baseline hazard functions than and comparable estimates of model parameters with the existing approach.

- (ii) In Chapter 3, we have extended the existing copula-based modeling approach for semi-competing risks data to a case in which two nonterminal and one terminal endpoints are of interest and subject to two sets of semi-competing risks, by proposing a novel regression model consisted by a pair of copulas. The regression parameters are estimated by the proposed maximum pseudo-likelihood estimation method, while the cumulative baseline hazards are computed using the nonparametric MLE. The simulation results show that the proposed joint analysis that particularly accounts for both two sets of semi-competing risks gains efficiency in estimation, in comparison with the marginal methods taking at most one of the two sets of semi-competing risks into account.

Multiple intermediate or surrogate endpoints are commonly observed after the start of treatment in cancer studies and are considered to be useful in assessing the effects of treatment and predicting the terminal endpoint. The method proposed in this chapter can be readily generalized to simultaneously analyze more than three endpoints consisting of one terminal endpoint (cancer/death) and multiple nonterminal endpoints. Thus, we anticipate that the method will provide a useful tool to evaluate disease

progression and overall survival with multiple endpoints subject to multiple sets of semi-competing risks.

- (iii) Unlike conventional screening methods, which only address univariate complete/censored survival times or independent censoring issues, in Chapter 4 we have proposed a novel joint screening approach for the semi-competing risks data subject to dependent censoring mechanism. Our proposed screening method deals with ultrahigh-dimensional covariates without any model assumption, while incorporating the dependency between the nonterminal and terminal endpoints simultaneously. Our results from numerical analysis show that the proposed joint screening procedure performs satisfactorily regardless the type of marginal model assumption.

5.2 Limitations and future works

There are several potential topics, which deserve for further investigation in the future.

In Chapter 2, our joint model for the nonterminal and terminal event times $T^{(1)}$ and $T^{(2)}$ is fitted based on observations $T^{(2)} \geq T^{(1)}$, as other joint analysis for semi-competing risks data. It limits in the uptriangle wedge of the $(T^{(1)}, T^{(2)})$. Such a limitation is presented when adopting copula approaches in Fine et al. (2001), Wang (2003), Peng and Fine (2007), Hsieh et al. (2008) and Chen (2012). This restriction may cause potential modeling problems as discussed by Xu et al. (2010), who proposed a three-state illness-death model with a common frailty to accommodate the dependence between $T^{(1)}$ and $T^{(2)}$. Their model avoids such a modeling problem to some extent but are difficult to be extended to handle clustered data. To alleviate such a restriction is definitely an important research and worthy of further studies.

In Chapter 3, we have proposed a pseudo-likelihood estimation method for

survival data subject to two sets of semi-competing risks in which the full likelihood is difficult to obtain due to complex interdependencies involved. Given models (3.2) and (3.3) to model $(T^{(1)}, T^{(3)})$ would lead to an identifiability problem. Therefore, there is no true Kendall's tau for $(T^{(1)}, T^{(3)})$. We use the measure $\psi(t, t)$ defined by (3.12) to estimate such an indirect dependency between $T^{(1)}$ and $T^{(3)}$. We may further develop new models such that complete maximum likelihood can be explicitly formulated and theoretical properties can be established for the estimators consequently. More recently, Emura et al. (2017) and Peng et al. (2018) studied the copula-based joint modeling approaches for analysis of clustered semi-competing risks data. When multiple nonterminal endpoints appear in clustered studies or meta-analysis, the proposed method can be extended by incorporating frailty in models (3.2) and (3.3). This provides a more complicated model framework and difficulties in estimation, opening an important research question worth investigating.

In Chapter 4, we have proposed a novel joint screening approach for the semi-competing risks data subject to dependent censoring mechanism and with ultrahigh-dimensional covariates. This is the first attempt for screening important genes from ultrahigh dimensional data under semi-competing risks. Our proposed method is free of model assumption. Therefore it is flexible by avoiding model specifications, which are hard to be justified in practice due to ultrahigh dimensionality. Furthermore, besides joint screening, we may consider to develop screening procedure only for the terminal endpoint with ultrahigh-dimensional covariates while conditioning on the information of the nonterminal endpoint. We have done some initial studies on this topic and obtained promising numerical results. Further theoretical justifications are anticipated in the future investigations.

Bibliography

- Anderson, J. E., Louis, T. A., Holm, N. V., and Harvald, B. (1992). Time-dependent association measures for bivariate survival distributions. *Journal of the American Statistical Association*, 87(419):641–650.
- Azaïs, J.-M. and Wschebor, M. (2009). *Level sets and extrema of random processes and fields*. John Wiley & Sons.
- Beer, D. G., Kardia, S. L., Huang, C.-C., Giordano, T. J., Levin, A. M., Misek, D. E., Lin, L., Chen, G., Gharib, T. G., Thomas, D. G., et al. (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Medicine*, 8(8):816.
- Caffo, B. S., Jank, W., and Jones, G. L. (2005). Ascent-based monte carlo expectation–maximization. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):235–251.
- Cai, J., Fan, J., Li, R., and Zhou, H. (2005). Variable selection for multivariate failure time data. *Biometrika*, 92(2):303–316.
- Chen, Y. H. (2012). Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. *Lifetime Data Analysis*, 18(1):36–57.

- Christian, N. J., Ha, I. D., and Jeong, J.-H. (2016). Hierarchical likelihood inference on clustered competing risks data. *Statistics in Medicine*, 35(2):251–267.
- Copelan, E., Biggs, J., Thompson, J., et al. (1991). Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with bucy2. *Blood*, 78(3):838–843.
- Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004). Least angle regression. *The Annals of Statistics*, 32(2):407–499.
- Emura, T. and Chen, Y.-H. (2016). Gene selection for survival data under dependent censoring: a copula-based approach. *Statistical Methods in Medical Research*, 25(6):2840–2857.
- Emura, T. and Chen, Y.-H. (2018). *Analysis of Survival Data with Dependent Censoring: Copula-Based Approaches*. Springer.
- Emura, T. and Michimae, H. (2017). A copula-based inference to piecewise exponential models under dependent censoring, with application to time to metamorphosis of salamander larvae. *Environmental and Ecological Statistics*, 24(1):151–173.
- Emura, T., Nakatochi, M., Matsui, S., Michimae, H., and Rondeau, V. (2018). Personalized dynamic prediction of death according to tumour progression and high-dimensional genetic factors: meta-analysis with a joint model. *Statistical Methods in Medical Research*, 27(9):2842–2858.
- Emura, T., Nakatochi, M., Murotani, K., and Rondeau, V. (2017). A joint frailty-copula model between tumour progression and death for meta-analysis. *Statistical Methods in Medical Research*, 26(6):2649–2666.

Bibliography

- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.
- Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):849–911.
- Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1):101.
- Fan, J., Samworth, R., and Wu, Y. (2009). Ultrahigh dimensional feature selection: beyond the linear model. *Journal of Machine Learning Research*, 10:2013–2038.
- Fan, J., Song, R., et al. (2010). Sure independence screening in generalized linear models with np-dimensionality. *The Annals of Statistics*, 38(6):3567–3604.
- Fine, J. P., Jiang, H. Y., and Chappell, R. (2001). On semi-competing risks data. *Biometrika*, 88(4):907–919.
- Gamst, A., Donohue, M., and Xu, R. (2009). Asymptotic properties and empirical evaluation of the npml in the proportional hazards mixed-effects model. *Statistica Sinica*, 19(3):997–1011.
- Gorfine, M. and Hsu, L. (2011). Frailty-based competing risks model for multivariate survival data. *Biometrics*, 67(2):415–426.
- 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.

Bibliography

- Heuchenne, C., Laurent, S., Legrand, C., and Keilegom, I. V. (2014). Likelihood-based inference for semi-competing risks. *Communications in Statistics-Simulation and Computation*, 43(5):1112–1132.
- Hougaard, P. (2012). *Analysis of Multivariate Survival Data*. Springer Science & Business Media.
- Hsieh, J. J., Wang, W. J., and A., A. D. (2008). Regression analysis based on semicompeting risks data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):3–20.
- Huang, E., Cheng, S. H., Dressman, H., Pittman, J., Tsou, M. H., Horng, C. F., Bild, A., Iversen, E. S., Liao, M., Chen, C. M., et al. (2003). Gene expression predictors of breast cancer outcomes. *The Lancet*, 361(9369):1590–1596.
- Jiang, F. and Haneuse, S. (2016). A semi-parametric transformation frailty model for semi-competing risks survival data. *Scandinavian Journal of Statistics*, 44(1):112–129.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.
- Keiding, N., Klein, J. P., and Horowitz, M. M. (2001). Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in Medicine*, 20(12):1871–1885.
- Klein, J. P. and Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.
- Lakhal, L., Rivest, L., and Abdous, B. (2008). Estimating survival and association in a semicompeting risks model. *Biometrics*, 64(1):180–188.
- Lee, J.-H., Choi, S.-J., et al. (2005). Anti-leukemic effect of graft-versus-host disease on bone marrow and extramedullary relapses in acute leukemia. *Haematologica*, 90(10):1380–1388.

Bibliography

- Lee, K. H., Dominici, F., Schrag, D., and Haneuse, S. (2016). Hierarchical models for semicompeting risks data with application to quality of end-of-life care for pancreatic cancer. *Journal of the American Statistical Association*, 111(515):1075–1095.
- Lee, M., Ha, I. D., and Lee, Y. (2017). Frailty modeling for clustered competing risks data with missing cause of failure. *Statistical Methods in Medical Research*, 26(1):356–373.
- Levine, R. A. and Casella, G. (2001). Implementations of the monte carlo em algorithm. *Journal of Computational and Graphical Statistics*, 10(3):422–439.
- Li, G., Peng, H., Zhang, J., Zhu, L., et al. (2012a). Robust rank correlation based screening. *The Annals of Statistics*, 40(3):1846–1877.
- Li, R., Zhong, W., and Zhu, L. (2012b). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107(499):1129–1139.
- Lin, D. and Ying, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, 80(3):573–581.
- Lossos, I. S., Czerwinski, D. K., Alizadeh, A. A., Wechser, M. A., Tibshirani, R., Botstein, D., and Levy, R. (2004). Prediction of survival in diffuse large-b-cell lymphoma based on the expression of six genes. *New England Journal of Medicine*, 350(18):1828–1837.
- Matsui, S., Simon, R., Qu, P., Shaughnessy, J. D., Barlogie, B., and Crowley, J. J. (2012). Developing and validating continuous genomic signatures in randomized clinical trials for predictive medicine. *Clinical Cancer Research*, pages clincanres–1206.
- Murphy, S. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics*, 23:182–198.

- Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.
- Ning, J., Chen, Y., Cai, C., Huang, X., and Wang, M.-C. (2015). On the dependence structure of bivariate recurrent event processes: inference and estimation. *Biometrika*, 102(2):345–358.
- Peng, L. M. and Fine, J. P. (2007). Regression modeling of semicompeting risks data. *Biometrics*, 63(1):96–108.
- Peng, M., Xiang, L., and Wang, S. (2018). Semiparametric regression analysis of clustered survival data with semi-competing risks. *Computational Statistics & Data Analysis*, 124:53–70.
- Pollard, D. (1984). *Convergence of stochastic processes*. Springer, New York.
- Renard, D., Geys, H., Molenberghs, G., Burzykowski, T., and Buyse, M. (2002). Validation of surrogate endpoints in multiple randomized clinical trials with discrete outcomes. *Biometrical Journal*, 44(8):921–935.
- Ripatti, S., Larsen, K., and Palmgren, J. (2002). Maximum likelihood inference for multivariate frailty models using an automated monte carlo em algorithm. *Lifetime Data Analysis*, 8(4):349–360.
- Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4):1016–1022.
- Rotolo, F., Legrand, C., and Van Keilegom, I. (2013). A simulation procedure based on copulas to generate clustered multi-state survival data. *Computer methods and programs in biomedicine*, 109(3):305–312.
- Rudin, W. (1973). *Functional analysis*. *McGraw-Hill Series in Higher Mathematics*.

Bibliography

- Shao, J. et al. (2003). Impact of the bootstrap on sample surveys. *Statistical Science*, 18(2):191–198.
- Shokouhi, S., Bra, y. S., Bakhtiyari, S., Sayehmiri, K., Alimoghadam, K., and Ghavamzadeh, A. (2015). Effects of agvhd and cgvhhd on survival rate in patients with acute myeloid leukemia after allogeneic stem cell transplantation. *International Journal of Hematology-Oncology and Stem Cell Research*, 9(3):112–121.
- Siannis, F., Farewell, V., and Head, J. (2007). A multi-state model for joint modelling of terminal and non-terminal events with application to whitehall ii. *Statistics in Medicine*, 26(2):426–442.
- Song, R., Lu, W., Ma, S., and Jessie Jeng, X. (2014). Censored rank independence screening for high-dimensional survival data. *Biometrika*, 101(4):799–814.
- Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., Van De Rijn, M., Jeffrey, S. S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*, 98(19):10869–10874.
- Sotiriou, C., Neo, S.-Y., McShane, L. M., Korn, E. L., Long, P. M., Jazaeri, A., Martiat, P., Fox, S. B., Harris, A. L., and Liu, E. T. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences*, 100(18):10393–10398.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:267–288.
- Tukey, J. W. (1993). Tightening the clinical trial. *Controlled clinical trials*, 14(4):266–285.

- Vaida, F. and Xu, R. H. (2000). Proportional hazards model with random effects. *Statistics in Medicine*, 19(24):3309–3324.
- Van De Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine*, 347(25):1999–2009.
- van Vliet, M. H., Reyat, F., Horlings, H. M., van de Vijver, M. J., Reinders, M. J., and Wessels, L. F. (2008). Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. *BMC Genomics*, 9(1):375.
- Van't Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., et al. (2002a). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415(6871):530–536.
- Van't Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., Van Der Kooy, K., Marton, M. J., Witteveen, A. T., et al. (2002b). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415(6871):530–536.
- Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42.
- Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.
- Wang, W. (2003). Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):257–273.

Bibliography

- Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J., et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *The Lancet*, 365(9460):671–679.
- Wei, G. C. and Tanner, M. A. (1990). A monte carlo implementation of the em algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699–704.
- Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., Desmedt, C., Ignatiadis, M., Sengstag, T., Schütz, F., et al. (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Research*, 10(4):R65.
- Xu, J., Kalbfleisch, J. D., and Tai, B. (2010). Statistical analysis of illness-death processes and semicompeting risks data. *Biometrics*, 66(3):716–725.
- Zeng, D., Lin, D., and Yin, G. (2005). Maximum likelihood estimation for the proportional odds model with random effects. *Journal of the American Statistical Association*, 100(470):470–483.
- Zhang, J., Liu, Y., and Wu, Y. (2017). Correlation rank screening for ultrahigh-dimensional survival data. *Computational Statistics & Data Analysis*, 108:121–132.
- Zhou, T. and Zhu, L. (2017). Model-free feature screening for ultrahigh dimensional censored regression. *Statistics and Computing*, 27(4):947–961.
- Zhu, L.-P., Li, L., Li, R., and Zhu, L.-X. (2011). Model-free feature screening for ultrahigh-dimensional data. *Journal of the American Statistical Association*, 106(496):1464–1475.

Bibliography

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

Proofs of Theorems 2.1-2.2[†]

Proof of Theorem 2.1

We establish the result in Theorem 2.1 through the following three steps. We first prove $\hat{\Lambda}_m(\cdot) = (\hat{\Lambda}_{m1}(\cdot), \hat{\Lambda}_{m2}(\cdot))$ is bounded on $[0, \varsigma]$. Then we show the existence of a convergent subsequence of $\hat{\omega}_m = (\hat{\alpha}_m, \hat{\beta}_m, \hat{\Lambda}_m, \hat{\Sigma}_m)$ by the compactness of the parameter space and Helly's selection theorem. Finally we prove that the limit of this subsequence must be ω_0 . Let

$$\bar{\Lambda}_{m1}(t) = \sum_{i,j} \frac{\delta_{ij}^{(1)}(1 - \Delta_{ij}(t))}{\sum_{g,h} \Delta_{gh}(T_{ij}) e^{\beta'_{01} X_{gh}} E_{\omega} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\}},$$

$$\bar{\Lambda}_{m2}(t) = \sum_{i,j} \frac{\delta_{ij}^{(2)}(1 - \tilde{\Delta}_{ij}(t))}{\sum_{g,h} \tilde{\Delta}_{gh}(\tilde{T}_{ij}) e^{\beta'_{02} X_{gh}} E_{\omega} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\}},$$

$$a_i(t) = n_i^{-1} \sum_{j=1}^{n_i} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) - \Delta_{ij}(u) e^{\beta'_{01} X_{ij}} E_{\omega} \{w_1(T_{ij}) e^{b'_i Z_{ij}}\} d\Lambda_{01}(u) \right\},$$

[†]This chapter is based on the supplementary material related to the paper Peng et al. (2018): “M.J. Peng, L.M. Xiang, and S.S. Wang. Semiparametric regression analysis of clustered survival data with semi-competing risks. *Computational Statistics & Data Analysis*, 124:53–70, 2018.”

$$b_i(t) = n_i^{-1} \sum_{j=1}^{n_i} \int_{u=0}^t \left\{ dN_{ij}^{(2)}(u) - \tilde{\Delta}_{ij}(u) e^{\beta'_{02} X_{ij}} E_{\omega} \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} d\Lambda_{02}(u) \right\},$$

$$f_{m1}(u) = m^{-1} \sum_{i=1}^m n_i^{-1} \sum_{j=1}^{n_i} \Delta_{ij}(u) e^{\beta'_{01} X_{ij}} E_{\omega} \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \},$$

$$f_{m2}(u) = m^{-1} \sum_{i=1}^m n_i^{-1} \sum_{j=1}^{n_i} \tilde{\Delta}_{ij}(u) e^{\beta'_{02} X_{ij}} E_{\omega} \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \}.$$

We show $\sup_{t \in [0, \varsigma]} |\bar{\Lambda}_{m1}(t) - \Lambda_{01}(t)| \rightarrow 0$ and $\sup_{t \in [0, \varsigma]} |\bar{\Lambda}_{m2}(t) - \Lambda_{02}(t)| \rightarrow 0$ almost surely (a.s.). Note that $\{a_i(t) : i = 1, 2, \dots\}$ and $\{b_i(t) : i = 1, 2, \dots\}$ are two mean zero independent sequences for fixed t . By the strong law of large numbers (SLLN), $m^{-1} \sum_i a_i(t) \rightarrow 0$ a.s. and $m^{-1} \sum_i b_i(t) \rightarrow 0$ a.s.. Given that $S_1(T_{ij})$, $S_2(\tilde{T}_{ij})$ and $c\{S_1(T_{ij}), S_2(\tilde{T}_{ij}); \alpha\}$ are bounded above 0, by the bounded assumption on X_{ij} and Z_{ij} we have $E_{\omega}\{w_1(T_{ij})e^{b'_i Z_{ij}}\}$, $E_{\omega}\{w_2(\tilde{T}_{ij})e^{b'_i Z_{ij}}\}$, $e^{\beta'_{01} X_{ij}}$ and $e^{\beta'_{02} X_{ij}}$ are all bounded. Applying the *SLLN* again, we have

$$f_{m1}(u) \rightarrow E_{\omega} \left[\Delta_{ij}(u) e^{\beta'_{01} X_{ij}} E_{\omega} \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \} \right] \text{ a.s.},$$

$$f_{m2}(u) \rightarrow E_{\omega} \left[\tilde{\Delta}_{ij}(u) e^{\beta'_{02} X_{ij}} E_{\omega} \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} \right] \text{ a.s.},$$

Note that

$$\begin{aligned} & \sum_{i,j} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) - \Delta_{ij}(u) e^{\beta'_{01} X_{ij}} E_{\omega} \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \} d\bar{\Lambda}_{n1}(u) \right\} \\ &= \sum_{i,j} \left\{ \delta_{ij}^{(1)} (1 - \Delta_{ij}(t)) - \sum_{g,h} \frac{\Delta_{ij}(T_{gh}) e^{\beta'_{01} X_{ij}} E_{\omega} \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \} \delta_{gh}^{(1)} (1 - \Delta_{gh}(t))}{\sum_{s,r} \Delta_{sr}(T_{gh}) e^{\beta'_{01} X_{sr}} E_{\omega} \{ w_1(T_{sr}) e^{b'_s Z_{sr}} \}} \right\} \\ &= \sum_{i,j} \delta_{ij}^{(1)} (1 - \Delta_{ij}(t)) - \sum_{g,h} \left\{ \frac{\sum_{i,j} \Delta_{ij}(T_{gh}) e^{\beta'_{01} X_{ij}} E_{\omega} \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \} \delta_{gh}^{(1)} (1 - \Delta_{gh}(t))}{\sum_{s,r} \Delta_{sr}(T_{gh}) e^{\beta'_{01} X_{sr}} E_{\omega} \{ w_1(T_{sr}) e^{b'_s Z_{sr}} \}} \right\} \\ &= 0. \end{aligned}$$

$$\begin{aligned}
& \sum_{i,j} \int_{u=0}^t \left\{ dN_{ij}^{(2)}(u) - \tilde{\Delta}_{ij}(u) e^{\beta'_{02} X_{ij}} E_{\omega} \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} d\bar{\Lambda}_{n2}(u) \right\} \\
&= \sum_{i,j} \left\{ \delta_{ij}^{(2)}(1 - \tilde{\Delta}_{ij}(t)) - \sum_{g,h} \frac{\tilde{\Delta}_{ij}(\tilde{T}_{gh}) e^{\beta'_{02} X_{ij}} E_{\omega} \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} + \delta_{gh}^{(2)}(1 - \tilde{\Delta}_{gh}(t))}{\sum_{s,r} \tilde{\Delta}_{sr}(\tilde{T}_{gh}) e^{\beta'_{02} X_{sr}} E_{\omega} \{ w_2(\tilde{T}_{sr}) e^{b'_s Z_{sr}} \}} \right\} \\
&= \sum_{i,j} \delta_{ij}^{(2)}(1 - \tilde{\Delta}_{ij}(t)) - \sum_{g,h} \left\{ \frac{\sum_{i,j} \tilde{\Delta}_{ij}(\tilde{T}_{gh}) e^{\beta'_{02} X_{ij}} E_{\omega} \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} \delta_{gh}^{(2)}(1 - \tilde{\Delta}_{gh}(t))}{\sum_{s,r} \tilde{\Delta}_{sr}(\tilde{T}_{gh}) e^{\beta'_{02} X_{sr}} E_{\omega} \{ w_2(\tilde{T}_{sr}) e^{b'_s Z_{sr}} \}} \right\} \\
&= 0.
\end{aligned}$$

For fixed t , by the *SLLN* we have

$$\begin{aligned}
& \int_{u=0}^t f_{m1}(u) d(\bar{\Lambda}_{m1} - \Lambda_{01})(u) \\
&= m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) - \Delta_{ij}(u) e^{\beta'_{01} X_{ij}} E_{\omega} \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \} d\Lambda_{01}(u) \right\} \\
&\quad - m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) - \Delta_{ij}(u) e^{\beta'_{01} X_{ij}} E_{\omega} \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \} d\bar{\Lambda}_{m1}(u) \right\} \\
&= n^{-1} \sum_{i=1}^n a_i(t) \rightarrow 0 \text{ a.s.}, \text{ and} \\
& \int_{u=0}^t f_{m2}(u) d(\bar{\Lambda}_{m2} - \Lambda_{02})(u) \\
&= m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^t \left\{ dN_{ij}^{(2)}(u) - \tilde{\Delta}_{ij}(u) e^{\beta'_{02} X_{ij}} E_{\omega} \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} d\Lambda_{02}(u) \right\} \\
&\quad - m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^t \left\{ dN_{ij}^{(2)}(u) - \tilde{\Delta}_{ij}(u) e^{\beta'_{02} X_{ij}} E_{\omega} \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} d\bar{\Lambda}_{m2}(u) \right\} \\
&= m^{-1} \sum_{i=1}^n b_i(t) \rightarrow 0 \text{ a.s.}
\end{aligned}$$

Since $\lim_{m \rightarrow \infty} f_{m1}(u) = E[\Delta_{ij}(u) e^{\beta'_{01} X_{ij}} E_{\omega} \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \}]$, which is bounded away from zero, there exists some $c_1(u) > 0$ such that eventually $f_{m1}(u) > c_1(u)$ almost surely. Let $c_1 = \sup_{u \in [0, c]} c_1(u)$. For sufficiently large n , we can write

$$0 \leq c_1 \int_{u=0}^t d(\bar{\Lambda}_{m1} - \Lambda_{01})(u) \leq \int_{u=0}^t f_{m1}(u) d(\bar{\Lambda}_{m1} - \Lambda_{01})(u) \rightarrow 0 \text{ a.s.}$$

Using the squeeze theorem, we have $\int_{u=0}^t d(\bar{\Lambda}_{m1} - \Lambda_{01})(u) \rightarrow 0$ a.s.. It follows that $\bar{\Lambda}_{m1}(t) \rightarrow \Lambda_{01}(t)$ a.s. and $\bar{\Lambda}_{m2}(t) \rightarrow \Lambda_{02}(t)$ a.s. for all $t \in [0, \varsigma]$. Since $\bar{\Lambda}_{mk}(t)$ are non-decreasing functions converging pointwise to the continuous functions $\Lambda_{0k}(t)$ for $k = 1, 2$, hence $\bar{\Lambda}_{mk}(t)$ are locally (on $[0, \varsigma]$ in particular) uniformly convergent.

Since $\hat{\alpha}_m, \hat{\beta}_{m1}, \hat{\beta}_{m2}, \hat{\Sigma}_m, X_{gh}$ and Z_{gh} are in compact sets and $w_1(T_{gh})$ and $w_2(\tilde{T}_{gh})$ are bounded, there exist some finite and possibly negative constants c_2 and c_3 such that

$$\hat{\beta}'_{m1} X_{gh} + \log E_{\hat{\omega}_m} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\} \geq \hat{\beta}'_{01} X_{gh} + \log E_{\omega_0} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\} + c_2,$$

$$\hat{\beta}'_{m2} X_{gh} + \log E_{\hat{\omega}_m} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\} \geq \hat{\beta}'_{02} X_{gh} + \log E_{\omega_0} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\} + c_3.$$

Therefore

$$\begin{aligned} \hat{\Lambda}_{m1}(\varsigma) &= \sum_{i,j} \frac{\delta_{ij}^{(1)}(1 - \Delta_{ij}(\varsigma))}{\Delta_{gh}(T_{ij}) \exp[\hat{\beta}'_{m1} X_{gh} + \log E_{\hat{\omega}_m} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\}]} \\ &\leq \sum_{i,j} \frac{\delta_{ij}^{(1)}(1 - \Delta_{ij}(\varsigma))}{\Delta_{gh}(T_{ij}) \exp[\hat{\beta}'_{01} X_{gh} + \log E_{\omega_0} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\} + c_2]} \\ &= e^{-c_2} \bar{\Lambda}_{m1}(\varsigma) \rightarrow e^{-c_2} \Lambda_{01}(\varsigma), \end{aligned}$$

$$\begin{aligned} \hat{\Lambda}_{m2}(\varsigma) &= \sum_{i,j} \frac{\delta_{ij}^{(2)}(1 - \tilde{\Delta}_{ij}(\varsigma))}{\tilde{\Delta}_{gh}(\tilde{T}_{ij}) \exp[\hat{\beta}'_{m2} X_{gh} + \log E_{\hat{\omega}_m} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\}]} \\ &\leq \sum_{i,j} \frac{\delta_{ij}^{(2)}(1 - \tilde{\Delta}_{ij}(\varsigma))}{\tilde{\Delta}_{gh}(\tilde{T}_{ij}) \exp[\hat{\beta}'_{02} X_{gh} + \log E_{\omega_0} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\} + c_3]} \\ &= e^{-c_3} \bar{\Lambda}_{m2}(\varsigma) \rightarrow e^{-c_3} \Lambda_{02}(\varsigma). \end{aligned}$$

So far we have established that $\hat{\Lambda}_{m1}$ and $\hat{\Lambda}_{m2}$ have upper bounds almost surely, and $\hat{\alpha}_m, \hat{\beta}_m$ and $\hat{\Sigma}_m$ are in compact sets. Now we can apply Helly's selection theorem to infer the existence of a convergent subsequence of $\hat{\omega}_m = (\hat{\alpha}_m, \hat{\beta}_m, \hat{\Lambda}_m, \hat{\Sigma}_m)$ having limit ω^* , where $\hat{\beta}_m = (\hat{\beta}_{m1}, \hat{\beta}_{m2})$ and $\hat{\Lambda}_m = (\hat{\Lambda}_{m1}, \hat{\Lambda}_{m2})$.

Next we show $\omega^* = \omega_0$. Since

$$\hat{\Lambda}_{m1}(t) = \int_0^t \frac{\sum_{g,h} \Delta_{gh}(u) \exp[\beta'_{01} X_{gh} + \log E_{\omega_0} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\}]}{\sum_{g,h} \Delta_{gh}(u) \exp[\hat{\beta}'_{m1} X_{gh} + \log E_{\hat{\omega}_m} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\}]} d\bar{\Lambda}_{m1}(u) \rightarrow \Lambda_{01}^*(t), \quad (\text{A.1})$$

$$\hat{\Lambda}_{m2}(t) = \int_0^t \frac{\sum_{g,h} \tilde{\Delta}_{gh}(u) \exp[\beta'_{02} X_{gh} + \log E_{\omega_0} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\}]}{\sum_{g,h} \tilde{\Delta}_{gh}(u) \exp[\hat{\beta}'_{m2} X_{gh} + \log E_{\hat{\omega}_m} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\}]} d\bar{\Lambda}_{m2}(u) \rightarrow \Lambda_{02}^*(t), \quad (\text{A.2})$$

we can see that $\Lambda^* = (\Lambda_{01}^*, \Lambda_{02}^*)$ is absolutely continuous with respect to $\Lambda_0 = (\Lambda_{01}, \Lambda_{02})$. In addition, $\Lambda^*(t) = (\Lambda_{01}^*(t), \Lambda_{02}^*(t))$ is differentiable with respect to t and $d\hat{\Lambda}_m(t)/d\bar{\Lambda}_m(t)$ converges to $d\Lambda^*(t)/d\Lambda_0(t)$.

Note that the finite sample likelihood as expressed in (2.6) in Chapter 2 has no finite maximum since λ_{01} and λ_{02} are free to go to infinity at any T_{ij} and $\tilde{T}_{ij}$, respectively. We restrict Λ_{01} to be right continuous with jumps at T_{ij} and Λ_{02} to be right continuous with jumps at $\tilde{T}_{ij}$. We then have the log-likelihood based on the observed data

$$\begin{aligned} l_m(\alpha, \beta, \Lambda, \Sigma) = & \sum_{i=1}^m \log \left\{ \int_{b_i} R_{1i}(\alpha, \beta, \Lambda, b_i) \psi(b_i, \Sigma) db_i \right\} \\ & + \sum_{i=1}^m \sum_{j=1}^{n_i} \delta_{ij}^{(1)} \log(\Lambda_{01}\{T_{ij}\}) + \delta_{ij}^{(2)} \log(\Lambda_{02}\{\tilde{T}_{ij}\}). \end{aligned} \quad (\text{A.3})$$

where

$$\begin{aligned} R_{1i}(\alpha, \beta, \Lambda, b) = & \exp \left[\sum_{j=1}^{n_i} \left\{ \delta_{ij}^{(1)} \delta_{ij}^{(2)} \log(D_{12}(T_{ij}, \tilde{T}_{ij})) - \delta_{ij}^{(1)} \Lambda_{01}(T_{ij}) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \right. \right. \\ & + \delta_{ij}^{(1)} (\beta'_1 X_{ij} + b'_i Z_{ij}) + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) \log(D_1(T_{ij}, \tilde{T}_{ij})) \\ & + (1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)} \log(D_2(T_{ij}, \tilde{T}_{ij})) - \delta_{ij}^{(2)} \Lambda_{02}(\tilde{T}_{ij}) e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \\ & \left. \left. + \delta_{ij}^{(2)} (\beta'_2 X_{ij} + b'_i Z_{ij}) + (1 - \delta_{ij}^{(1)}) (1 - \delta_{ij}^{(2)}) \log \left[c \{S_1(T_{ij}), S_2(\tilde{T}_{ij})\} \right] \right\} \right]. \end{aligned}$$

Note that

$$\begin{aligned}
& 0 \leq m^{-1} \{l_m(\hat{\alpha}_m, \hat{\beta}_m, \hat{\Lambda}_m, \hat{\Sigma}_m) - l_m(\alpha_0, \beta_0, \bar{\Lambda}_m, \Sigma_0)\} \\
& = m^{-1} \sum_{i=1}^m \log \left\{ \int_{b_i} R_{1i}(\hat{\alpha}_m, \hat{\beta}_m, \hat{\Lambda}_m, b_i) \psi(b_i, \hat{\Sigma}_m) db_i \right\} \\
& \quad - m^{-1} \sum_{i=1}^m \log \left\{ \int_{b_i} R_{1i}(\alpha_0, \beta_0, \bar{\Lambda}_m, b_i) \psi(b_i, \Sigma_0) db_i \right\} \\
& \quad + m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \delta_{ij}^{(1)} \log(\hat{\Lambda}_m\{T_{ij}\} / \bar{\Lambda}_m\{T_{ij}\}) + m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \delta_{ij}^{(2)} \log(\hat{\Lambda}_{02}\{\tilde{T}_{ij}\} / \bar{\Lambda}_{02}\{\tilde{T}_{ij}\}) \\
& \rightarrow E \left[\log \left\{ \left(\int_{b_i} R_{1i}(\alpha^*, \beta^*, \Lambda^*, b_i) \psi(b_i, \Sigma^*) db_i \prod_{j=1}^{n_i} \left\{ \lambda_{01}^*(T_{ij})^{\delta_{ij}^{(1)}} \lambda_{02}^*(\tilde{T}_{ij})^{\delta_{ij}^{(2)}} \right\} \right) \right. \right. \\
& \quad \cdot \left. \left(\int_{b_i} R_{1i}(\alpha_0, \beta_0, \Lambda_0, b_i) \psi(b_i, \Sigma_0) db_i \prod_{j=1}^{n_i} \left\{ \lambda_{01}(T_{ij})^{\delta_{ij}^{(1)}} \lambda_{02}(\tilde{T}_{ij})^{\delta_{ij}^{(2)}} \right\} \right)^{-1} \right\} \right].
\end{aligned}$$

Because the Kullback-Leibler information between the density indexed by $(\alpha^*, \beta^*, \Lambda^*, \Sigma^*)$ and the true density is negative, we have that $l(\alpha^*, \beta^*, \Lambda^*, \Sigma^*) = l(\alpha_0, \beta_0, \Lambda_0, \Sigma_0)$ almost surely. That is almost surely

$$\begin{aligned}
& \int_{b_i} R_{1i}(\alpha^*, \beta^*, \Lambda^*, b_i) \psi(b_i, \Sigma^*) db_i \prod_{j=1}^{n_i} \left\{ \lambda_{01}^*(T_{ij})^{\delta_{ij}^{(1)}} \lambda_{02}^*(\tilde{T}_{ij})^{\delta_{ij}^{(2)}} \right\} \\
& = \int_{b_i} R_{1i}(\alpha_0, \beta_0, \Lambda_0, b_i) \psi(b_i, \Sigma_0) db_i \prod_{j=1}^{n_i} \left\{ \lambda_{01}(T_{ij})^{\delta_{ij}^{(1)}} \lambda_{02}(\tilde{T}_{ij})^{\delta_{ij}^{(2)}} \right\}.
\end{aligned} \tag{A.4}$$

Next we adapt the techniques used in the identifiability step of Zeng et al. (2005) and Gamst et al. (2009) to show that $(\alpha^*, \beta^*, \Lambda^*, \Sigma^*) = (\alpha_0, \beta_0, \Lambda_0, \Sigma_0)$. For notational convenience, we recall the notation in (2.9) of Chapter 2 and denote

that

$$\begin{aligned}
D_{1^*}(t^{(1)}, t^{(2)}) &= \frac{\partial c\{S_1^*(t^{(1)}), S_2^*(t^{(2)})\}}{\partial S_1^*(t^{(1)})}, & D_{12^*}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2^*(t^{(2)})\}}{\partial S_1(t^{(1)})\partial S_2^*(t^{(2)})}, \\
D_{1^*2^*}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1^*(t^{(1)}), S_2^*(t^{(2)})\}}{\partial S_1^*(t^{(1)})\partial S_2^*(t^{(2)})}, \\
D_\alpha(t^{(1)}, t^{(2)}) &= \frac{\partial c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial \alpha}, & D_{1\alpha}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)})\partial \alpha}, \\
D_{2\alpha}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_2(t^{(2)})\partial \alpha}, & D_{12\alpha}(t^{(1)}, t^{(2)}) &= \frac{\partial^3 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)})\partial S_2(t^{(2)})\partial \alpha}.
\end{aligned} \tag{A.5}$$

Let $\delta_{ij}^{(1)} = 1$ and $T_{ij} = 0$, for all $j = 1, \dots, n_i$. With some fix h , $1 \leq h \leq n_i$, set $\delta_{ij}^{(2)} = 1$ and $\tilde{T}_{ij} = 0$ for $j \leq h$. For $j > h$, we replace $\tilde{T}_{ij}$ with ς when $\delta_{ij}^{(2)} = 0$, while we integrate $\tilde{T}_{ij}$ from 0 to ς when $\delta_{ij}^{(2)} = 1$. Thus, according to (2.5), we obtain that

$$\begin{aligned}
&L_i(\theta, \lambda | \mathcal{O}_i, b_i) \\
&= \prod_{j=1}^h \left\{ D_{12}(0, 0) \lambda_{01}(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \lambda_{02}(0) e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \right\} \\
&\quad \cdot \prod_{j=h+1}^{n_i} \left\{ D_1(0, \varsigma) \lambda_{01}(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \right\}^{1 - \delta_{ij}^{(2)}} \\
&\quad \cdot \prod_{j=h+1}^{n_i} \left\{ \int_0^\varsigma D_{12}(0, y) \lambda_{01}(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} e^{-\Lambda_{02}(y) e^{\beta'_2 X_{ij} + b'_i Z_{ij}}} \lambda_{02}(y) e^{\beta'_2 X_{ij} + b'_i Z_{ij}} dy \right\}^{\delta_{ij}^{(2)}} \\
&= \prod_{j=1}^h \left\{ D_{12}(0, 0) \lambda_{01}(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \lambda_{02}(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \right\} \\
&\quad \cdot \prod_{j=h+1}^{n_i} \left[\left\{ (1 - \delta_{ij}^{(2)}) D_1(0, \varsigma) + \delta_{ij}^{(2)} (-D_1(0, \varsigma) + D_1(0, 0)) \right\} \cdot \lambda_{01}(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \right] \\
&= \{D_{12}(0, 0)\}^h \{D_1(0, 0)\}^{n_i - h} \prod_{j=1}^h \left\{ \lambda_{02}(0) e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \right\} \cdot \prod_{j=1}^{n_i} \left\{ \lambda_{01}(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \right\}.
\end{aligned}$$

Equation (A.4) can be written as

$$\begin{aligned}
& \{D_{1*2*}(0, 0)\}^h \{D_{1*}(0, 0)\}^{n_i-h} \\
& \cdot \int_b \prod_{j=1}^h \left\{ \lambda_{02}^*(0) e^{\beta_2^{*'} X_{ij} + b'_i Z_{ij}} \right\} \prod_{j=1}^{n_i} \left\{ \lambda_{01}^*(0) e^{\beta_1^{*'} X_{ij} + b'_i Z_{ij}} \right\} \psi(b, \Sigma^*) db \\
& = \{D_{12}(0, 0)\}^h \{D_1(0, 0)\}^{n_i-h} \\
& \cdot \int_b \prod_{j=1}^h \left\{ \lambda_{02}(0) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \right\} \prod_{j=1}^{n_i} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \right\} \psi(b, \Sigma_0) db.
\end{aligned} \tag{A.6}$$

Let $h = 0$, it follows from (A.6) that

$$\begin{aligned}
& \{D_{1*}(0, 0)\}^{n_i} \int_{b_i} \prod_{j=1}^{n_i} \left\{ \lambda_{01}^*(0) e^{\beta_1^{*'} X_{ij} + b'_i Z_{ij}} \right\} \psi(b_i, \Sigma^*) db_i \\
& = \{D_1(0, 0)\}^{n_i} \int_{b_i} \prod_{j=1}^{n_i} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \right\} \psi(b_i, \Sigma_0) db_i.
\end{aligned} \tag{A.7}$$

That is equivalent to

$$\begin{aligned}
& \{D_{1*}(0, 0) \lambda_{01}^*(0)\}^{n_i} e^{\sum_{j=1}^{n_i} \beta_1^{*'} X_{ij}} E_{b_i} (e^{\sum_{j=1}^{n_i} b'_i Z_{ij}} | \Sigma = \Sigma^*) \\
& = \{D_1(0, 0) \lambda_{01}(0)\}^{n_i} e^{\sum_{j=1}^{n_i} \beta'_{01} X_{ij}} E_{b_i} (e^{\sum_{j=1}^{n_i} b'_i Z_{ij}} | \Sigma = \Sigma_0).
\end{aligned} \tag{A.8}$$

Since $S_1^*(0) = S_1(0) = S_2^*(0) = S_2(0) = 1$ and $D_1(0, 0)$ is function only related to α , let $f(\alpha) = D_1(0, 0)$, (A.8) can be written as

$$\begin{aligned}
& n_i \left[(\log(f(\alpha^*)) - \log(f(\alpha_0))) + (\log \lambda_{01}^*(0) - \log \lambda_{01}(0)) \right] + \sum_{j=1}^{n_i} X'_{ij} (\beta_1^* - \beta_{01}) \\
& + \left\{ E_{b_i} (e^{\sum_{j=1}^{n_i} b'_i Z_{ij}} | \Sigma = \Sigma^*) - E_{b_i} (e^{\sum_{j=1}^{n_i} b'_i Z_{ij}} | \Sigma = \Sigma_0) \right\} = 0,
\end{aligned} \tag{A.9}$$

Specifically, when b_i is assumed to follow a normal distribution with mean 0 and variance Σ_0 , (A.9) can be written as

$$\begin{aligned}
& n_i \left\{ (\log(f(\alpha^*)) - \log(f(\alpha_0))) + (\log \lambda_{01}^*(0) - \log \lambda_{01}(0)) \right\} \\
& + \sum_{j=1}^{n_i} X'_{ij} (\beta_1^* - \beta_{01}) + \frac{(\sum_{j=1}^{n_i} Z_{ij})' (\Sigma^* - \Sigma_0) (\sum_{j=1}^{n_i} Z_{ij})}{2} = 0,
\end{aligned} \tag{A.10}$$

for all $i = 1, \dots, m$.

By Condition 2.8 in Chapter 2, we have that $(\log(f(\alpha^*)) - \log(f(\alpha_0))) + (\log \lambda_{02}^*(0) - \log \lambda_{02}(0)) = 0$, $\beta_1^* = \beta_{01}$ and $\Sigma^* = \Sigma_0$. Since the above equations applied for different copulas with different forms of function $f(\cdot)$, it follows that $\alpha^* = \alpha_0$ and $\lambda_{01}^*(0) = \lambda_{01}(0)$.

When it is assumed that $Z_{ij} = 1$ and $u_i = \exp(b_i)$ follows a gamma distribution with mean 1 and variance η_0 , (A.9) can be rewritten as

$$\{f(\alpha^*)\lambda_{01}^*(0)\}^{n_i} e^{\sum_{j=1}^{n_i} \beta_1^{*'} X_{ij}} g(\eta^*, n_i) = \{f(\alpha_0)\lambda_{01}(0)\}^{n_i} e^{\sum_{j=1}^{n_i} \beta_{01}' X_{ij}} g(\eta_0, n_i), \quad (\text{A.11})$$

where $g(\eta, n) = (\eta + 1)(2\eta + 1) \cdots ((n - 1)\eta + 1)$. By Condition 2.7 in Chapter 2 and the above equations applied for different sample sizes, different types of $\lambda(\cdot)$ and copulas with different forms of function $f(\cdot)$, we have that $\alpha^* = \alpha_0$, $\lambda_{01}^*(0) = \lambda_{01}(0)$, $\beta_1^* = \beta_{01}$ and $\eta^* = \eta_0$.

It therefore follows from (A.6) that

$$(D_{12^*}(0, 0))^h \prod_{j=1}^h \left\{ \lambda_{02}^*(0) e^{\beta_2^{*'} X_{ij} + b_i' Z_{ij}} \right\} = (D_{12}(0, 0))^h \prod_{j=1}^h \left\{ \lambda_{02}(0) e^{\beta_{02}' X_{ij} + b_i' Z_{ij}} \right\}. \quad (\text{A.12})$$

As $S_1^*(0) = S_1(0) = S_2^*(0) = S_2(0) = 1$ and $\alpha^* = \alpha_0$, we have that $D_{12^*}(0, 0) = D_{12}(0, 0)$. The above equation can be written as

$$h \left(\log \lambda_{02}^*(0) - \log \lambda_{02}(0) \right) + (\beta_2^* - \beta_{02})' \sum_{j=1}^h X_{ij} = 0. \quad (\text{A.13})$$

Setting $h = n_i$, we can obtain that

$$n_i \left(\log \lambda_{02}^*(0) - \log \lambda_{02}(0) \right) + (\beta_2^* - \beta_{02})' \sum_{j=1}^{n_i} X_{ij} = 0, \quad (\text{A.14})$$

for all $i = 1, \dots, m$. By Condition 2.7 in Chapter 2, we have $\lambda_{02}^*(0) = \lambda_{02}(0)$ and $\beta_2^* = \beta_{02}$.

Next, we prove $\Lambda_{02}^*(t) = \Lambda_{02}(t)$. Let $\delta_{ij}^{(1)} = 1$ and $T_{ij} = 0$ for all $j = 1, \dots, n_i$. When $j = 1$ and $\delta_{i1}^{(2)} = 1$, we integrate $\tilde{T}_{i1}$ from 0 to t . For $j = 2, \dots, n_i$, we

replace $\tilde{T}_{ij}$ with ς if $\delta_{ij}^{(2)} = 0$ and integrate $\tilde{T}_{ij}$ from 0 to ς if $\delta_{ij}^{(2)} = 1$. Summing the two sides of (A.4) over all possible values of $\delta_{ij}^{(2)}$ for $j > 1$, we obtain that

$$\begin{aligned} & \int_{b_i} \int_0^t D_{1*2*}(0, y) e^{-\Lambda_{02}^*(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}}} \lambda_{02}^*(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} dy \cdot \psi(b_i, \Sigma_0) db_i \\ &= \int_{b_i} \int_0^t D_{12}(0, y) e^{-\Lambda_{02}(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}}} \lambda_{02}(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} dy \cdot \psi(b_i, \Sigma_0) db_i, \end{aligned}$$

which is equivalent to

$$\int_{b_i} -c_1\{S_1^*(0), S_2^*(t)\} \cdot \psi(b_i, \Sigma_0) db_i = \int_{b_i} -c_1\{S_1^*(0), S_2(t)\} \cdot \psi(b_i, \Sigma_0) db_i.$$

Therefore,

$$c_1\{S_1^*(0), S_2^*(t)\} = c_1\{S_1(0)^*, S_2(t)\}. \quad (\text{A.15})$$

As $\alpha^* = \alpha_0$, $S_1^*(0) = S_1(0)$ and the above equation holds for all kinds of copulas, we obtain that $S_2^*(t) = S_2(t)$ and $\Lambda_{02}^*(t) = \Lambda_{02}(t)$.

Similarly, we can show that $\Lambda_{01}^*(t) = \Lambda_{01}(t)$. Let $\delta_{ij}^{(2)} = 0$ and $\tilde{T}_{ij} = \varsigma$ for all $j = 1, \dots, n_i$. Choose k from index set $\{1, \dots, n_i\}$. For $j \leq k$, we integrate T_{ij} from 0 to t if $\delta_{ij}^{(1)} = 1$. On the other hand, for $j > k$, we replace T_{ij} with ς if $\delta_{ij}^{(1)} = 0$, otherwise we integrate T_{ij} from 0 to ς .

We choose $k = 1$ and then obtain that

$$\begin{aligned} & L_i(\alpha_0, \beta_0, \Lambda_0, \Sigma_0) \\ &= \int_{b_i} \int_0^t D_1(y, \varsigma) e^{-\Lambda_{01}(y) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}}} \lambda_{01}(y) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} dy \cdot \prod_{j=2}^{n_i} \left[\left\{ c\{S_1(\varsigma), S_2(\varsigma)\} \right\}^{1-\delta_{ij}^{(1)}} \right. \\ & \quad \cdot \left. \left\{ \int_0^\varsigma D_1(y, \varsigma) e^{-\Lambda_{01}(y) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}}} \lambda_{01}(y) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} dy \right\}^{\delta_{ij}^{(1)}} \right] \\ &= \int_{b_i} \left(\frac{c\{S_1(0), S_2(\varsigma)\} - c\{S_1(t), S_2(\varsigma)\}}{S_2(\varsigma)} \right) \cdot \prod_{j=1}^{n_i} S_2(\varsigma) \psi(b_i, \Sigma_0) db_i. \end{aligned} \quad (\text{A.16})$$

Following from (A.4) that

$$c\{S_1^*(0), S_2(\varsigma)\} - c\{S_1^*(t), S_2(\varsigma)\} = c\{S_1(0), S_2(\varsigma)\} - c\{S_1(t), S_2(\varsigma)\}. \quad (\text{A.17})$$

Since $\alpha^* = \alpha$ and $S_1^*(0) = S_1(0)$, $c\{S_1^*(t), S_2(\varsigma)\} = c\{S_1(t), S_2(\varsigma)\}$ for any copula c . This implies that $S_1^*(t) = S_1(t)$. Therefore, $\Lambda_{01}^*(t) = \Lambda_{01}(t)$ and the proof of Theorem 2.1 is complete.

Proof of Theorem 2.2

Let $\mathcal{H} = \{(h_1, h_2, \dots, h_6) : h_1 \in R^1, h_2 \in R^{d_1}, h_3 \in R^{d_2}, (h_4, h_5) \text{ are functions on } [0, \varsigma], h_6 \in R^{d_3(d_3+1)/2} \text{ and } \|h_1\|, \|h_2\|, \|h_3\|, \|h_4\|_v, \|h_5\|_v, \|h_6\| \leq 1\}$, where $\|h_4\|_v$ and $\|h_5\|_v$ denote the total variations of $h_4(\cdot)$ and $h_5(\cdot)$ on $[0, \varsigma]$, respectively. Let S_m be a sequence of maps from $\mathcal{U}$, a neighborhood of $(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0)$, to $l^\infty(\mathcal{H})$.

$$\begin{aligned} & S_m(\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma)[h_1, h_2, h_3, h_4, h_5, h_6] \\ &= m^{-1} \frac{d}{d\epsilon} l_m(\alpha + \epsilon h_1, \beta_1 + \epsilon h_2, \beta_2 + \epsilon h_3, \Lambda_1 + \epsilon h_4, \Lambda_2 + \epsilon h_5, \Sigma + \epsilon h_6)|_{\epsilon=0} \quad (\text{A.18}) \\ &= : A_{m_1}[h_1] + A_{m_2}[h_2] + A_{m_3}[h_3] + A_{m_4}[h_4] + A_{m_5}[h_5] + A_{m_6}[h_6], \end{aligned}$$

where we treat Σ as an extended column vector consisting of the upper triangle elements of the covariance matrix. The terms A_{mp} for $p = 1, 2, \dots, 6$ are linear functionals on $R^1, R^{d_1}, R^{d_2}, BV[0, \varsigma], BV[0, \varsigma]$ and $R^{d_3(d_3+1)/2}$, respectively. Here $BV[0, \varsigma]$ is the space of functions with finite variation on $[0, \varsigma]$. For a single cluster, let $l_\alpha, l_{\beta_1}, l_{\beta_2}, l_{\Lambda_1}, l_{\Lambda_2}, l_\Sigma$ be the score functions for $\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma$, respectively. Then $A_{m_1}[h_1] = \mathcal{P}_m[h'_1 l_\alpha]$, $A_{m_2}[h_2] = \mathcal{P}_m[h'_2 l_{\beta_1}]$, $A_{m_3}[h_3] = \mathcal{P}_m[h'_3 l_{\beta_2}]$, $A_{m_4}[h_4] = \mathcal{P}_m[h'_4 l_{\Lambda_1}]$, $A_{m_5}[h_5] = \mathcal{P}_m[h'_5 l_{\Lambda_2}]$ and $A_{m_6}[h_6] = \mathcal{P}_m[h'_6 l_\Sigma]$, where $\mathcal{P}_m$ denotes the empirical measure based on m independent clusters. We now seek for explicit expressions for A_{mp} .

Recall the log-likelihood in (A.3) and the notation defined in (A.5) in the proof of Theorem 2.1. For simplicity of notation, we further denote

$$d\mu_i(b_i) = \frac{R_{1i}(\omega, b_i)\psi(b_i, \Sigma)db_i}{\int_{b_i} R_{1i}(\omega, b_i)\psi(b_i, \Sigma)db_i}.$$

Then we have

$$\begin{aligned}
A_{m_1}[h_1] &= m^{-1} \frac{d}{d\epsilon} l_m(\alpha + \epsilon h_1) \Big|_{\epsilon=0} \\
&= m^{-1} \sum_{i=1}^m \int_{b_i} \sum_{j=1}^{n_i} h_1 \left\{ \delta_{ij}^{(1)} \delta_{ij}^{(2)} \frac{D_{12\alpha}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) \frac{D_{1\alpha}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} \right. \\
&\quad \left. + (1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)} \frac{D_{2\alpha}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} + (1 - \delta_{ij}^{(1)}) (1 - \delta_{ij}^{(2)}) \frac{D_\alpha(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} \right\} d\mu_i(b_i),
\end{aligned}$$

$$\begin{aligned}
A_{m_2}[h_2] &= m^{-1} \frac{d}{d\epsilon} l_m(\beta_1 + \epsilon h_2) \Big|_{\epsilon=0} \\
&= m^{-1} \sum_{i=1}^m \int_{b_i} \sum_{j=1}^{n_i} h'_2 X_{ij} \left\{ \delta_{ij}^{(1)} - \Lambda_1(T_{ij}) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \delta_{ij}^{(2)} \frac{D_{121}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) \right. \right. \\
&\quad \left. + \delta_{ij}^{(1)} + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) \frac{D_{11}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) + (1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)} \frac{D_{21}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) \right. \\
&\quad \left. \left. + (1 - \delta_{ij}^{(1)}) (1 - \delta_{ij}^{(2)}) \frac{D_1(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_1(T_{ij}) \right) \right\} d\mu_i(b_i),
\end{aligned}$$

$$\begin{aligned}
A_{m_3}[h_3] &= m^{-1} \frac{d}{d\epsilon} l_m(\beta_2 + \epsilon h_3) \Big|_{\epsilon=0} \\
&= m^{-1} \sum_{i=1}^m \int_{b_i} \sum_{j=1}^{n_i} h'_3 X_{ij} \left\{ \delta_{ij}^{(2)} - \Lambda_2(\tilde{T}_{ij}) e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \delta_{ij}^{(2)} \frac{D_{122}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \right. \right. \\
&\quad \left. + \delta_{ij}^{(2)} + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) + (1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)} \frac{D_{22}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \right. \\
&\quad \left. \left. + (1 - \delta_{ij}^{(1)}) (1 - \delta_{ij}^{(2)}) \frac{D_2(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_2(\tilde{T}_{ij}) \right) \right\} d\mu_i(b_i),
\end{aligned}$$

$$\begin{aligned}
& A_{m_4}[h_4] \\
&= m^{-1} \frac{d}{d\epsilon} l_m(\Lambda_1 + \epsilon h_4) \Big|_{\epsilon=0} \\
&= m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \left\{ \delta_{ij}^{(1)} h_4(T_{ij}) - \int_0^{T_{ij}} h_4(s) d\Lambda_1(s) \int_{b_i} e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \delta_{ij}^{(2)} \frac{D_{121}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} \right. \right. \\
&\quad \cdot S_1(T_{ij}) + \delta_{ij}^{(1)} + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) \frac{D_{11}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} \cdot S_1(T_{ij}) + (1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)} \frac{D_{21}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} \\
&\quad \left. \left. \cdot S_1(T_{ij}) + (1 - \delta_{ij}^{(1)}) (1 - \delta_{ij}^{(2)}) \frac{D_1(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_1(T_{ij}) \right) d\mu_i(b_i) \right\},
\end{aligned}$$

$$\begin{aligned}
& A_{m_5}[h_5] \\
&= m^{-1} \frac{d}{d\epsilon} l_m(\Lambda_2 + \epsilon h_5) \Big|_{\epsilon=0} \\
&= m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \left\{ \delta_{ij}^{(2)} h_5(\tilde{T}_{ij}) - \int_0^{\tilde{T}_{ij}} h_5(s) d\Lambda_2(s) \int_{b_i} e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \delta_{ij}^{(2)} \frac{D_{122}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} \right. \right. \\
&\quad \cdot S_2(\tilde{T}_{ij}) + \delta_{ij}^{(2)} + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) + (1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)} \frac{D_{22}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} \\
&\quad \left. \left. \cdot S_2(\tilde{T}_{ij}) + (1 - \delta_{ij}^{(1)}) (1 - \delta_{ij}^{(2)}) \frac{D_2(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_2(\tilde{T}_{ij}) \right) d\mu_i(b_i) \right\},
\end{aligned}$$

$$A_{m_6}[h_6] = m^{-1} \frac{d}{d\epsilon} l_m(\Sigma + \epsilon h_6) \Big|_{\epsilon=0} = m^{-1} \sum_{i=1}^m \int_{b_i} \frac{d}{d\epsilon} \psi(b_i, \Sigma + \epsilon h_6) \Big|_{\epsilon=0} d\mu_i(b_i).$$

Specifically, when b_i is assumed to follow a normal distribution with mean 0 and variance Σ ,

$$A_{m_6}[h_6] = m^{-1} \sum_{i=1}^m \int_{b_i} \{b'_i \Sigma^{-1} H_6 \Sigma^{-1} b_i / 2 - \text{trace}(\Sigma^{-1} H_6) / 2\} d\mu_i(b_i).$$

Or when it is assumed that $Z_{ij} = 1$ and $u_i = \exp(b_i)$ follows a gamma distribution with mean 1 and variance η ,

$$A_{m_6}[h_6] = m^{-1} \sum_{i=1}^m \int_{u_i} \left\{ -\frac{1}{u_i} \left(\frac{1}{\eta} - 1 \right) - u_i - 1 + \log(\eta) + \varphi_0\left(\frac{1}{\eta}\right) \right\} \cdot \frac{h_6}{\eta^2} d\mu_i(u_i),$$

where $\varphi_0(\cdot)$ is the polygamma function. By Conditions 2.2 and 2.5, and the result that $\Lambda_1(\cdot)$ and $\Lambda_2(\cdot)$ are bounded on $[0, \varsigma]$, which has been justified in the proof of Theorem 2.1, we have A_{mp} for $p = 1, 2, \dots, 6$ all have bounded derivations on $[0, \varsigma]$. Therefore, A_{mp} are all bounded Lipschitz functions.

Define the limit map $S : (\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma)[h_1, \dots, h_6] \rightarrow l^\infty(\mathcal{H})$ by

$$S(\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma)[h_1, \dots, h_6] = A_1[h_1] + A_2[h_2] + \dots + A_6[h_6],$$

where the linear functionals A_p for $p = 1, \dots, 6$ are defined similar to A_{mp} , but replacing the empirical sum with the expectation. By definition, we have that $S_m(\hat{\alpha}_m, \hat{\beta}_{m1}, \hat{\beta}_{m2}, \hat{\Lambda}_{01}, \hat{\Lambda}_{02}, \hat{\Sigma}_m) = 0$ and $S(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0) = 0$. Thus $\sqrt{m}\{S_m(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0) - S(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0)\}$ weakly converges to a tight Gaussian process on $l^\infty(\mathcal{H})$ since $\mathcal{H}$ is a Donsker class and the functionals A_{mp} are bounded Lipschitz functionals with respect to $\mathcal{H}$ (Pollard, 1984).

In fact, it is easy to show that $S(\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma)$ is Frechet differentiable because of its smoothness. In a manner similar to Lemma 1 in the appendix of Murphy (1995), we can show an approximation condition that

$$\begin{aligned} & \sup_{(h_1, \dots, h_6)} |(S_m - S)(\hat{\alpha}_m, \hat{\beta}_{m1}, \hat{\beta}_{m2}, \hat{\Lambda}_{m1}, \hat{\Lambda}_{m2}, \hat{\Sigma}_m) - (S_m - S)(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0)| \\ &= O_p\left(m^{-1/2} \vee \left\{ \|\hat{\alpha}_m - \alpha_0\| + \|\hat{\beta}_{m1} - \beta_{01}\| + \|\hat{\beta}_{m2} - \beta_{02}\| + \sup_{t \in [0, \varsigma]} |\hat{\Lambda}_{m1}(t) - \Lambda_{01}(t)| \right. \right. \\ & \left. \left. + \sup_{t \in [0, \varsigma]} |\hat{\Lambda}_{m2}(t) - \Lambda_{02}(t)| + \|\hat{\Sigma}_m - \Sigma_0\| \right\}\right). \end{aligned}$$

based on the functional central limit theorem (Gamst et al., 2009).

Next, we prove that the derivative, denoted by $\dot{S}(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0)$, is continuously invertible on its range. We consider $\dot{S}(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0)$ as a linear map, simply denoted by T , from the space $\{(\alpha - \alpha_0, \beta_1 - \beta_{01}, \beta_2 - \beta_{02}, \Lambda_1 - \Lambda_{01}, \Lambda_2 - \Lambda_{02}, \Sigma - \Sigma_0) : (\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma) \text{ is in the neighborhood } \mathcal{U} \text{ of}$

$(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0)\}$ to $l^\infty(\mathcal{H})$. That is

$$\begin{aligned} & T(\alpha - \alpha_0, \beta_1 - \beta_{01}, \beta_2 - \beta_{02}, \Lambda_1 - \Lambda_{01}, \Lambda_2 - \Lambda_{02}, \Sigma - \Sigma_0) \\ &= (\alpha - \alpha_0)'Q_1(h_1, \dots, h_6) + (\beta_1 - \beta_{01})'Q_2(h_1, \dots, h_6) + (\beta_2 - \beta_{02})'Q_3(h_1, \dots, h_6) \\ &+ \int_0^\varsigma Q_4(h_1, \dots, h_6)d(\Lambda_1 - \Lambda_{01}) + \int_0^\varsigma Q_5(h_1, \dots, h_6)d(\Lambda_2 - \Lambda_{02}) \\ &+ (\Sigma - \Sigma_0)'Q_6(h_1, \dots, h_6), \end{aligned}$$

where Q_i for $i = 1, \dots, 6$ are the respective partial derivatives of S with respect to $\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2$ and Σ . Specifically, Q_i ' have the following forms:

$$\begin{aligned} Q_1(h_1, \dots, h_6) &= B_1(h_1, h_2, h_3, h_6)' + \int_0^\varsigma h_4(t)D_1(t) + \int_0^\varsigma h_5(t)D_2(t), \\ Q_2(h_1, \dots, h_6) &= B_2(h_1, h_2, h_3, h_6)' + \int_0^\varsigma h_4(t)D_3(t) + \int_0^\varsigma h_5(t)D_4(t), \\ Q_3(h_1, \dots, h_6) &= B_3(h_1, h_2, h_3, h_6)' + \int_0^\varsigma h_4(t)D_5(t) + \int_0^\varsigma h_5(t)D_6(t), \\ Q_4(h_1, \dots, h_6) &= B_4(h_1, h_2, h_3, h_6)' + b_4h_4(t) + \int_0^\varsigma h_4(t)D_7(t) + \int_0^\varsigma h_5(t)D_8(t), \\ Q_5(h_1, \dots, h_6) &= B_5(h_1, h_2, h_3, h_6)' + b_5h_5(t) + \int_0^\varsigma h_5(t)D_9(t) + \int_0^\varsigma h_5(t)D_{10}(t), \\ Q_6(h_1, \dots, h_6) &= B_6(h_1, h_2, h_3, h_6)' + \int_0^\varsigma h_4(t)D_{11}(t) + \int_0^\varsigma h_5(t)D_{12}(t), \end{aligned}$$

where $B_1, \dots, B_6$ are constant matrices, $D_1(t), \dots, D_{12}(t)$ are continuously differentiable functions and $b_4 > 0, b_5 > 0$. Let $Q = (Q_1, \dots, Q_6)$. To prove T is invertible, it suffices to show the invertibility of the linear operator $Q(h_1, \dots, h_6)$ (Rudin, 1973). Suppose $Q(h_1, \dots, h_6) = 0$, then $T(\alpha - \alpha_0, \beta_1 - \beta_{01}, \beta_2 - \beta_{02}, \Lambda_1 - \Lambda_{01}, \Lambda_2 - \Lambda_{02}, \Sigma - \Sigma_0)[h_1, \dots, h_6] = 0$ for any $(\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma)$ in the neighborhood $\mathcal{U}$. For a small constant ϵ , let $\alpha = \alpha_0 + \epsilon h_1, \beta_1 = \beta_{01} + \epsilon h_2, \beta_2 = \beta_{02} + \epsilon h_3, \Lambda_1(t) = \Lambda_{01}(t) + \epsilon \int_0^t h_4(s)d\Lambda_{01}(s), \Lambda_2(t) = \Lambda_{02}(t) + \epsilon \int_0^t h_5(s)d\Lambda_{02}(s), \Sigma = \Sigma_0 + \epsilon h_6$. It follows, by the definition of T , that

$$\begin{aligned} 0 &= T(\alpha - \alpha_0, \beta_1 - \beta_{01}, \beta_2 - \beta_{02}, \Lambda_1 - \Lambda_{01}, \Lambda_2 - \Lambda_{02}, \Sigma - \Sigma_0)[h_1, \dots, h_6] \\ &= \epsilon E\{(l_{\alpha_0}[h_1] + l_{\beta_{01}}[h_2] + l_{\beta_{02}}[h_3] + l_{\Lambda_{01}}[h_4] + l_{\Lambda_{02}}[h_5] + l_{\Sigma_0}[h_6])^2\}, \end{aligned}$$

where l_{α_0} denotes the score function of α evaluated at α_0 , $l_{\beta_{0k}}$ the score function of β_k evaluated at β_{0k} , $l_{\Lambda_{0k}}$ the score function of Λ_k evaluated at Λ_{0k} , for $k = 1, 2$, and l_{Σ_0} the score function of Σ evaluated at Σ_0 . Then we have

$$l_{\alpha_0}[h_1] + l_{\beta_{01}}[h_2] + l_{\beta_{02}}[h_3] + l_{\Lambda_{01}}[h_4] + l_{\Lambda_{02}}[h_5] + l_{\Sigma_0}[h_6] = 0 \quad (\text{A.19})$$

almost surely, where

$$R_{2i}(\omega_0, b_i) = R_{1i}(\omega_0, b_i) \prod_{j=1}^{n_i} \left\{ \lambda_{01}(T_{ij})^{\delta_{ij}^{(1)}} \lambda_{02}(\tilde{T}_{ij})^{\delta_{ij}^{(2)}} \right\}, \quad (\text{A.20})$$

$$\begin{aligned} & l_{\alpha_0}[h_1] \\ &= \sum_{j=1}^{n_i} \int_{b_i} l_{\alpha_0,j}[h_1] \psi(b_i, \Sigma_0) db_i \\ &= \sum_{j=1}^{n_i} \int_{b_i} h_1 \left\{ \delta_{ij}^{(1)} \delta_{ij}^{(2)} \frac{D_{12\alpha_0}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) \frac{D_{1\alpha_0}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} \right. \\ & \quad \left. + (1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)} \frac{D_{2\alpha_0}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} + (1 - \delta_{ij}^{(1)}) (1 - \delta_{ij}^{(2)}) \frac{D_{\alpha_0}(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} \right\} \\ & \quad \cdot R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i, \end{aligned} \quad (\text{A.21})$$

$$\begin{aligned} & l_{\beta_{01}}[h_2] \\ &= \sum_{j=1}^{n_i} \int_{b_i} l_{\beta_{01},j}[h_2] \psi(b_i, \Sigma_0) db_i \\ &= \sum_{j=1}^{n_i} \int_{b_i} h_2' X_{ij} \left\{ \delta_{ij}^{(1)} - \Lambda_{01}(T_{ij}) e^{\beta_{01}' X_{ij} + b_i' Z_{ij}} \left(\delta_{ij}^{(1)} \delta_{ij}^{(2)} \frac{D_{121}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) + \delta_{ij}^{(1)} \right. \right. \\ & \quad \left. \left. + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) \frac{D_{11}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) + (1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)} \frac{D_{21}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) \right. \right. \\ & \quad \left. \left. + (1 - \delta_{ij}^{(1)}) (1 - \delta_{ij}^{(2)}) \frac{D_1(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_1(T_{ij}) \right) \right\} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i, \end{aligned} \quad (\text{A.22})$$

$$\begin{aligned}
& l_{\beta_{02}}[h_3] \\
&= \sum_{j=1}^{n_i} \int_{b_i} l_{\beta_{02},j}[h_3] \psi(b_i, \Sigma_0) db_i \\
&= \sum_{j=1}^{n_1} \int_{b_i} h'_3 X_{ij} \left\{ \delta_{ij}^{(2)} - \Lambda_{02}(\tilde{T}_{ij}) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \delta_{ij}^{(2)} \frac{D_{122}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \right. \right. \\
&\quad + \delta_{ij}^{(2)} + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) + (1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)} \frac{D_{22}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \\
&\quad \left. \left. + (1 - \delta_{ij}^{(2)}) (1 - \delta_{ij}^{(2)}) \frac{D_2(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_2(\tilde{T}_{ij}) \right) \right\} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i,
\end{aligned} \tag{A.23}$$

$$\begin{aligned}
& l_{\Lambda_{01}}[h_4] \\
&= \sum_{j=1}^{n_i} \int_{b_i} l_{\Lambda_{01},j}[h_4] \psi(b_i, \Sigma_0) db_i \\
&= \sum_{j=1}^{n_i} \int_{b_i} \left\{ \delta_{ij}^{(1)} h_4(T_{ij}) - \int_0^{T_{ij}} h_4(s) d\Lambda_{01}(s) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \delta_{ij}^{(2)} \frac{D_{121}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) \right. \right. \\
&\quad + \delta_{ij}^{(1)} + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) \frac{D_{11}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) + (1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)} \frac{D_{21}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) \\
&\quad \left. \left. + (1 - \delta_{ij}^{(1)}) (1 - \delta_{ij}^{(2)}) \frac{D_1(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_1(T_{ij}) \right) \right\} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i,
\end{aligned} \tag{A.24}$$

$$\begin{aligned}
& l_{\Lambda_{02}}[h_5] \\
&= \sum_{j=1}^{n_i} \int_{b_i} l_{\Lambda_{02},j}[h_5] \psi(b_i, \Sigma_0) db_i \\
&= \sum_{j=1}^{n_i} \int_{b_i} \left\{ \delta_{ij}^{(2)} h_5(\tilde{T}_{ij}) - \int_0^{\tilde{T}_{ij}} h_5(s) d\Lambda_{02}(s) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \delta_{ij}^{(2)} \frac{D_{122}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \right. \right. \\
&\quad + \delta_{ij}^{(2)} + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) + (1 - \delta_{ij}^{(1)}) \delta_{ij}^{(2)} \frac{D_{22}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \\
&\quad \left. \left. + (1 - \delta_{ij}^{(1)}) (1 - \delta_{ij}^{(2)}) \frac{D_2(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_2(\tilde{T}_{ij}) \right) \right\} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i,
\end{aligned} \tag{A.25}$$

$$l_{\Sigma_0}[h_6] = \int_{b_i} \frac{d}{d\epsilon} \psi(b_i, \Sigma_0 + \epsilon h_6) \Big|_{\epsilon=0} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i. \quad (\text{A.26})$$

In the following, we show that equation (A.19) implies that $h_1 = 0$, $h_2 = 0$, $h_3 = 0$, $h_4(\cdot) = 0$, $h_5(\cdot) = 0$, $h_6 = 0$. Let X_{ij} and Z_{ij} be fixed. Set $\delta_{ij}^{(1)} = 1$, and $T_{ij} = 0$ for all $j = 1, \dots, n_i$. For fixed integer h in $1, \dots, n_i$, let $\delta_{ij}^{(2)} = 1$, $\tilde{T}_{ij} = 0$ when $j \leq h$. When $j > h$, replace $\tilde{T}_{ij}$ with ς if $\delta_{ij}^{(2)} = 0$ and integrate $\tilde{T}_{ij}$ from 0 to ς otherwise. Then,

$$\begin{aligned} R_{2i}(\omega_0, b) &= \prod_{j=1}^{n_i} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_{12}(0, \tilde{T}_{ij})^{\delta_{ij}^{(2)}} D_1(0, \tilde{T}_{ij})^{1-\delta_{ij}^{(2)}} \\ &\quad \cdot \exp \left\{ \delta_{ij}^{(2)} (\beta'_{02} X_{ij} + b'_i Z_{ij}) - \delta_{ij}^{(2)} \Lambda_{02}(\tilde{T}_{ij}) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \right\} \lambda_{02}(\tilde{T}_{ij})^{\delta_{ij}^{(2)}}. \end{aligned} \quad (\text{A.27})$$

For fixed integer h in $1, \dots, n_i$, we define measures $\mu_1, \dots, \mu_{n_i}$ on the set $\{0, 1\} \times [0, \varsigma]$ as follows:

$$\mu_r(\{0\} \times A) = 0, \mu_r(\{1\} \times A) = I(0 \in A), r \leq h$$

and

$$\mu_r(\{0\} \times A) = I(\varsigma \in A), \mu_r(\{1\} \times A) = \int I_A dt, r > h,$$

where A is any Borel set in $[0, \varsigma]$.

We integrate both sides of equation (A.19) with respect to $\{(\delta_{i1}^{(2)}, \tilde{T}_{i1}), (\delta_{i2}^{(2)}, \tilde{T}_{i2}), \dots, (\delta_{in_i}^{(2)}, \tilde{T}_{in_i})\}$ and product measure $d\mu_1, \dots, d\mu_{n_i}$, denoted by $d\tilde{\mu}$ for simplicity.

Note that, for any b ,

$$\begin{aligned}
& \int R_{2i}(\omega_0, b) d\tilde{\mu} \\
&= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
&\quad \cdot \prod_{\substack{\delta_{ir}^{(2)} \in \{0,1\} \\ r > h}} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, \varsigma) \right\}^{1 - \delta_{ir}^{(2)}} \\
&\quad \cdot \left\{ \int_{y=0}^{\varsigma} \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, y) S_2(y) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(y) dy \right\}^{\delta_{ir}^{(2)}} \\
&= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
&\quad \cdot \prod_{\substack{\delta_{ir}^{(2)} \in \{0,1\} \\ r > h}} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, \varsigma) \right\}^{1 - \delta_{ir}^{(2)}} \\
&\quad \cdot \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} (-D_1(0, \varsigma) + D_1(0, 0)) \right\}^{\delta_{ir}^{(2)}} \\
&= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\}.
\end{aligned} \tag{A.28}$$

Consider the first term $l_{\alpha_0}[h_1]$ on the left side of equation (A.19). If $j \leq h$, from (A.21) we have

$$\begin{aligned}
& l_{\alpha_0, j}[h_1] \\
&= \int h_1 \frac{D_{12\alpha_0}(0, 0)}{D_{12}(0, 0)} R_{2i}(\omega_0, b) d\tilde{\mu} \\
&= h_1 \frac{D_{12\alpha_0}(0, 0)}{D_{12}(0, 0)} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
&\quad \cdot \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\};
\end{aligned}$$

Otherwise if $j > h$,

$$\begin{aligned}
& l_{\alpha_0, j}[h_1] \\
&= h_1 \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
&\quad \cdot \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\
&\quad \cdot \sum_{\delta_{ij}^{(2)} \in \{0, 1\}} \left\{ (1 - \delta_{ij}^{(2)}) D_{1\alpha_0}(0, \varsigma) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} + \delta_{ij}^{(2)} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \right. \\
&\quad \left. \cdot \int_0^\varsigma -D_{12\alpha_0}(0, y) dy \right\} \\
&= h_1 \frac{D_{1\alpha_0}(0, 0)}{D_1(0, 0)} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
&\quad \cdot \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\}.
\end{aligned}$$

Then the contribution of $l_{\alpha_0}[h_1]$ in (A.19) reduces to

$$\begin{aligned}
& h_1 \left\{ \sum_{j \leq h} \frac{D_{12\alpha_0}(0, 0)}{D_{12}(0, 0)} + \sum_{j > h} \frac{D_{1\alpha_0}(0, 0)}{D_1(0, 0)} \right\} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir}} \lambda_{02}(0) \right\} \\
&\quad \cdot \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_1(0, 0) \right\} \cdot E(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}}).
\end{aligned} \tag{A.29}$$

Consider the second term $l_{\beta_{01}}[h_2]$ on the left side of (A.19). If $j \leq h$, by (A.22) we have

$$\begin{aligned}
& l_{\beta_{01}, j}[h_2] \\
&= \int h'_2 X_{ij} (1 - 0) \times R_{2i}(\omega_0, b) d\tilde{\mu} \\
&= h'_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
&\quad \cdot \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\};
\end{aligned}$$

Otherwise,

$$\begin{aligned}
& l_{\beta_{01},j}[h_2] \\
&= h'_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
&\quad \cdot \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\
&\quad \cdot \sum_{\delta_{ij}^{(2)} \in \{0,1\}} \left[(1 - \delta_{ij}^{(2)}) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_1(0, \varsigma) \right. \\
&\quad \left. + \delta_{ij}^{(2)} \int_{y=0}^{\varsigma} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_{12}(0, y) S_2(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{02}(y) dy \right] \\
&= h'_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
&\quad \cdot \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\
&\quad \cdot \sum_{\delta_{ij}^{(2)} \in \{0,1\}} \left[(1 - \delta_{ij}^{(2)}) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_1(0, \varsigma) \right. \\
&\quad \left. + \delta_{ij}^{(2)} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \left(-D_1(0, \varsigma) + D_1(0, 0) \right) \right] \\
&= h'_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\}.
\end{aligned}$$

Then the contribution of $l_{\beta_{01}}[h_2]$ in (A.19) becomes

$$\begin{aligned}
& \sum_{j=1}^{n_i} h'_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\
& \cdot E(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}}).
\end{aligned} \tag{A.30}$$

Consider the third term $l_{\beta_{02}}[h_3]$ on the left side of (A.19). If $j \leq h$, by (A.23) we have

$$\begin{aligned}
& l_{\beta_{02},j}[h_3] \\
&= \int h'_3 X_{ij} (1 - 0) R_{2i}(\omega_0, b) d\tilde{\mu} \\
&= h'_3 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
&\quad \cdot \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\};
\end{aligned}$$

otherwise,

$$\begin{aligned}
& l_{\beta_{02},j}[h_3] \\
&= h'_3 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
& \quad \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\
& \quad \cdot \sum_{\delta_{ij}^{(2)} \in \{0,1\}} \left[(1 - \delta_{ij}^{(2)}) \left\{ -\Lambda_{02}(\varsigma) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} D_{12}(0, \varsigma) S_2(\varsigma) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \right\} \right. \\
& \quad + \delta_{ij}^{(2)} \int_{y=0}^{\varsigma} \left\{ 1 - \Lambda_{02}(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \frac{D_{122}(0, y)}{D_{12}(0, y)} \right. \\
& \quad \cdot S_2(y) \left. \right\} D_{12}(0, y) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} S_2(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{02}(y) dy \left. \right] \\
&= h'_3 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} \right. \\
& \quad \cdot D_1(0, 0) \left. \right\} \cdot \sum_{\delta_{ij}^{(2)} \in \{0,1\}} \left[(1 - \delta_{ij}^{(2)}) \left\{ -\Lambda_{02}(\varsigma) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} D_{12}(0, \varsigma) S_2(\varsigma) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \right\} \right. \\
& \quad + \delta_{ij}^{(2)} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \int_{y=0}^{\varsigma} \frac{\partial \left\{ \frac{\partial D_1(0, y)}{\partial \lambda_{02}(y)} \Lambda_{02}(y) \right\}}{\partial y} dy \left. \right] \\
&= h'_3 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
& \quad \cdot \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \cdot 0 \\
&= 0.
\end{aligned}$$

Therefore, the contribution of $l_{\beta_{02}}[h_3]$ in (A.19) turns to be

$$\begin{aligned}
& \sum_{j \leq h} h'_3 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_1(0, 0) \right\} \\
& \cdot E \left(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}} \right).
\end{aligned} \tag{A.31}$$

We consider the fourth term $l_{\Lambda_{01}}[h_4]$ of (A.19). If $j \leq h$, by (A.24)

$$\begin{aligned}
l_{\Lambda_{01},j}[h_4] &= h_4(0) \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
& \quad \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\};
\end{aligned}$$

Otherwise,

$$\begin{aligned}
& l_{\Lambda_{01},j}[h_4] \\
&= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
& \quad \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0,0) \right\} \\
& \quad \cdot \sum_{\delta_{ij}^{(2)} \in \{0,1\}} \left[(1 - \delta_{ij}^{(2)}) h_4(0) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_1(0, \varsigma) \right. \\
& \quad + \delta_{ij}^{(2)} h_4(0) \int_{y=0}^{\varsigma} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_{12}(0, y) \\
& \quad \cdot S_2(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{02}(y) dy \left. \right] \\
&= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} \right. \\
& \quad \cdot D_1(0,0) \left. \right\} \cdot \sum_{\delta_{ij}^{(2)} \in \{0,1\}} \left[(1 - \delta_{ij}^{(2)}) h_4(0) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_1(0, \varsigma) \right. \\
& \quad + \delta_{ij}^{(2)} h_4(0) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \left(-D_1(0, \varsigma) + D_1(0,0) \right) \left. \right] \\
&= h_4(0) \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\
& \quad \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0,0) \right\}.
\end{aligned}$$

Therefore, the contribution of $l_{\Lambda_{01}}[h_4]$ in (A.19) becomes

$$\begin{aligned}
& \sum_{j=1}^{n_i} h_4(0) \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_1(0,0) \right\} \\
& \cdot E\left(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}}\right).
\end{aligned} \tag{A.32}$$

For the fifth term $l_{\Lambda_{02}}[h_5]$ of (A.19), if $j \leq h$, it follows from (A.25) that for any

b

$$l_{\Lambda_{02},j}[h_5] = h_5(0) \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\ \cdot \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\}.$$

For the fifth term of (A.19), when $j > h$,

$$l_{\Lambda_{02},j}[h_5] \\ = \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\ \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\ \cdot \sum_{\delta_{ij}^{(2)} \in \{0,1\}} \left[- (1 - \delta_{ij}^{(2)}) \int_0^\varsigma h_5(y) d\Lambda_{02}(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_{12}(0, \varsigma) S_2(\varsigma) \right. \\ \left. + \delta_{ij}^{(2)} \int_{y=0}^\varsigma \left\{ h_5(y) - \int_{s=0}^y h_5(s) d\Lambda_{02}(s) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \left(1 + \frac{D_{122}(0, y)}{D_{12}(0, y)} S_2(y) \right) \right\} \right. \\ \left. \cdot D_{12}(0, y) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} S_2(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{02}(y) dy \right] \\ = \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\ \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\ \cdot \sum_{\delta_{ij}^{(2)} \in \{0,1\}} \left[- (1 - \delta_{ij}^{(2)}) \int_0^\varsigma h_5(y) d\Lambda_{02}(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_{12}(0, \varsigma) S_2(\varsigma) \right. \\ \left. + \delta_{ij}^{(2)} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \int_{y=0}^\varsigma \frac{\partial \left\{ \frac{\partial D_1(0, y)}{\partial \Lambda_{02}(y)} \right\}}{\partial y} dy \right] \\ = \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \\ \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \cdot 0 \\ = 0.$$

Thus the contribution of $l_{\Lambda_{02}}[h_5]$ in (A.19) becomes

$$\begin{aligned} & \sum_{j \leq h} h_5(0) \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_1(0, 0) \right\} \\ & \cdot E \left(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}} \right). \end{aligned} \quad (\text{A.33})$$

Through similar calculations, the sixth term $l_{\Sigma_0}[h_6]$ of (A.19) reduces to:

$$\begin{aligned} & \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_1(0, 0) \right\} \\ & \cdot E \left\{ \left[\frac{\partial \psi(b_i, \Sigma_0 + \epsilon h_6)}{\partial \epsilon} \right]_{\epsilon=0} e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}} \right\}. \end{aligned} \quad (\text{A.34})$$

Combining (A.29), (A.30), (A.31), (A.32), (A.33) and (A.34), equation (A.19) is rewritten as:

$$\begin{aligned} & \left[h_1 \left\{ \sum_{j \leq h} \frac{D_{12\alpha_0}(0, 0)}{D_{12}(0, 0)} + \sum_{j > h} \frac{D_{1\alpha_0}(0, 0)}{D_1(0, 0)} \right\} + \sum_{j=1}^{n_i} h'_2 X_{ij} + \sum_{j \leq h} h'_3 X_{ij} + \sum_{j=1}^{n_i} h_4(0) + \sum_{j \leq h} h_5(0) \right] \\ & \cdot E \left(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}} \right) + E \left\{ \left[\frac{\partial \psi(b_i, \Sigma_0 + \epsilon h_6)}{\partial \epsilon} \right]_{\epsilon=0} e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}} \right\} = 0. \end{aligned} \quad (\text{A.35})$$

The following is to prove (A.35) implies that $h_1 = h_2 = h_3 = 0$, $h_4(\cdot) = 0$, $h_5(\cdot) = 0$, $h_6 = 0$.

Under the normal random-effects assumption, (A.35) is rewritten as:

$$\begin{aligned} & h_1 \left\{ \sum_{j \leq h} \frac{D_{12\alpha_0}(0, 0)}{D_{12}(0, 0)} + \sum_{j > h} \frac{D_{1\alpha_0}(0, 0)}{D_1(0, 0)} \right\} + \sum_{j=1}^{n_i} h'_2 X_{ij} + \sum_{j \leq h} h'_3 X_{ij} \\ & + \sum_{j=1}^{n_i} h_4(0) + \sum_{j \leq h} h_5(0) + \frac{1}{2} \left(\sum_{j=1}^{n_i} Z_{ij} \right)' H_6 \left(\sum_{j=1}^{n_i} Z_{ij} \right) = 0. \end{aligned} \quad (\text{A.36})$$

In (A.36), we can see that only the term multiplied with h_1 is a function of α_0 , other terms are unrelated with α_0 . (A.36) exists over different values of α_0 , and thus $h_1 = 0$. Since the choice of h is arbitrary, we conclude that $\sum_{j=k_1}^{k_2} h'_3 X_{ij} + (k_2 - k_1)h_5(0) = 0$ for any $1 \leq k_1 \leq k_2 \leq n_i$, leading to $h_3 = 0$ and $h_5(0) = 0$. Then (A.36) is simplified as

$$\sum_{j=1}^{n_i} h'_2 X_{ij} + \sum_{j=1}^{n_i} h_4(0) + \frac{1}{2} \left(\sum_{j=1}^{n_i} Z_{ij} \right)' H_6 \left(\sum_{j=1}^{n_i} Z_{ij} \right) = 0. \quad (\text{A.37})$$

Applying Condition 2.8 in Chapter 2, we obtain $h_2 = 0$, $h_4(0) = 0$, $H_6 = 0$.

Setting $\delta_{ij}^{(1)} = \delta_{ij}^{(2)} = 1$ and $T_{ij} = 0$ for $j = 1, \dots, n_i$, $\tilde{T}_{ij} = 0$ for $j = 2, \dots, n_i$, and $c(u, v) = uv$ in (A.19) gives that

$$h_5(\tilde{T}_{i1}) = \frac{\int_0^{\tilde{T}_{i1}} h_5(s) d\Lambda_{02}(s) \left\{ \int_b e^{\beta'_{02} X_{i1} + b'_i Z_{i1}} R_{2i}(\omega_0, b) \phi(0, \Sigma_0) db \right\}}{\int_b R_{2i}(\omega_0, b) \phi(0, \Sigma_0) db}.$$

Therefore, $g(y) \equiv \int_0^y h_5(t) d\Lambda_{02}(t)$ satisfies the homogeneous equation

$$\frac{g'(y)}{\lambda_{02}(y)} = \left[g(y) \left\{ \int_b e^{\beta'_{02} X_{i1} + b'_i Z_{i1}} R_{2i}(\omega_0, b) \phi(0, \Sigma_0) db \right\} \right] / \left\{ \int_b R_{2i}(\omega_0, b) \phi(0, \Sigma_0) db \right\}.$$

With boundary condition $g(0) = 0$, we conclude $g(y) = 0$, $h_5(y) = 0$ and thus $h_5(\cdot) = 0$. Similarly, setting $\delta_{ij}^{(1)} = 1$ for $j = 1, \dots, n_i$, $\delta_{ij}^{(2)} = 1$ and $T_{ij} = \tilde{T}_{ij} = 0$ for $j = 2, \dots, n_i$, $\delta_{i1}^{(2)} = 0$ and $\tilde{T}_{i1} = \varsigma$, $c(u, v) = uv$ in (A.19) finally yield that $h_4(\cdot) = 0$.

Under the gamma frailty assumption, (A.35) is rewritten as:

$$\begin{aligned} h_1 \left\{ \sum_{j \leq h} \frac{D_{12\alpha_0}(0, 0)}{D_{12}(0, 0)} + \sum_{j > h} \frac{D_{1\alpha_0}(0, 0)}{D_1(0, 0)} \right\} + \sum_{j=1}^{n_i} h'_2 X_{ij} + \sum_{j \leq h} h'_3 X_{ij} + \sum_{j=1}^{n_i} h_4(0) + \sum_{j \leq h} h_5(0) \\ + \left\{ -\left(\frac{1}{\eta_0} - 1\right) \frac{1}{(n_i + h - 2)\eta_0 + 1} - (n_i + h)\eta_0 - 2 + \log(\eta_0) + \varphi_0\left(\frac{1}{\eta_0}\right) \right\} \cdot \frac{h_6}{\eta_0^2} = 0, \end{aligned} \quad (\text{A.38})$$

where $\varphi_0(\cdot)$ is the polygamma function.

Note that we can see that only the term multiplied with h_1 is a function of α_0 , other terms are unrelated with α_0 . (A.38) exists over different values of α_0 , and thus $h_1 = 0$. In (A.38), we can see only the term multiplied with h_6 is a function of η_0 , while the other terms are unrelated with η_0 . (A.38) exists over different values of η_0 , thus $h_6 = 0$. Since the choice of h is arbitrary, we conclude that $\sum_{j=k_1}^{k_2} h'_3 X_{ij} + (k_2 - k_1)h_5(0) = 0$ for any $1 \leq k_1 \leq k_2 \leq n_i$, yielding $h_3 = 0$ and $h_5(0) = 0$. Now (A.38) is changed to

$$\sum_{j=1}^{n_i} h'_2 X_{ij} + \sum_{j=1}^{n_i} h_4(0) = 0. \quad (\text{A.39})$$

Applying Condition 2.7 in Chapter 2, we have $h_2 = 0$, $h_4(0) = 0$. Set $\delta_{ij}^{(1)} = \delta_{ij}^{(2)} = 1$ and $T_{ij} = 0$ for $j = 1, \dots, n_i$, $\tilde{T}_{ij} = 0$ for $j = 2, \dots, n_i$ and $c(u, v) = uv$ in (A.19). It follows that

$$h_5(\tilde{T}_{i1}) = \frac{\int_0^{\tilde{T}_{i1}} h_5(s) d\Lambda_{02}(s) \left\{ \int_u e^{\beta'_{02} X_{i1}} R_{2i}(\omega_0, b) d_u \Gamma(1/\eta_0, \eta_0) \right\}}{\int_u R_{2i}(\omega_0, b) d_u \Gamma(1/\eta_0, \eta_0)}.$$

Therefore $g(y) \equiv \int_0^y h_5(t) d\Lambda_{02}(t)$ satisfies the homogeneous equation

$$\frac{g'(y)}{\lambda_{02}(y)} = \left[g(y) \left\{ \int_u e^{\beta'_{02} X_{i1}} R_{2i}(\omega_0, b) d_u \Gamma(1/\eta_0, \eta_0) \right\} \right] / \left\{ \int_u R_{2i}(\omega_0, b) d_u \Gamma(1/\eta, \eta_0) \right\}.$$

With boundary condition $g(0) = 0$, we conclude $g(y) = 0$, $h_5(y) = 0$, thus $h_5(\cdot) = 0$. Similarly, set $\delta_{ij}^{(1)} = 1$ for $j = 1, \dots, n_i$, $\delta_{ij}^{(2)} = 1$ and $T_{ij} = \tilde{T}_{ij} = 0$ for $j = 2, \dots, n_i$, $\delta_{i1}^{(2)} = 0$ and $\tilde{T}_{i1} = \varsigma$, $c(u, v) = uv$ in (A.19). it follows that $h_4(\cdot) = 0$.

Now we have proved that $h_1 = h_2 = h_3 = 0$, $h_4(\cdot) = 0$, $h_5(\cdot) = 0$, $h_6 = 0$. This implies that Q is one-to-one and thus $\dot{S}(\omega_0)$ is invertible. Thus the asymptotic normality of $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\Lambda}_1, \hat{\Lambda}_2$ and $\hat{\Sigma}$ follows from the martingale central limit theorem in Theorem 2 of Murphy (1995) under the normal random-effects or gamma frailty assumption.

Simulations for Weibull baseline hazard functions in Chapter 2[†]

In addition to constant baseline hazard functions, we hereby examine the performance of the proposed method when Weibull baseline hazard functions are assumed in both marginal models. Data of $T^{(1)}$ and $T^{(2)}$ are generated from a Gumbel copula model with marginal proportional hazards models

$$\lambda_1(t) = \lambda_{01}(t)\exp\{\beta_1 X + bZ\} \text{ and } \lambda_2(t) = \lambda_{02}(t)\exp\{\beta_2 X + bZ\},$$

respectively, where $Z = 1$ in Scenarios 2.1-2.2 and $Z \sim U(-2, 2)$ in Scenarios 2.3-2.4. We set $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = t^{1/2}$ and $\lambda_{02}(t) = \frac{3}{4}t^{1/2}$. The covariate X , censoring times and other parameters are set in the same manner as that in Section 2.5 of Chapter 2. Tables B.1-B.4 report simulation results for the model with normally distributed random effects, i.e. $b \sim N(0, \eta)$, in Scenarios 2.1-2.4, respectively. The overall results are very similar to those given in Tables 2.1-2.4

[†]This chapter is based on the supplementary material related to the paper Peng et al. (2018): “M.J. Peng, L.M. Xiang, and S.S. Wang. Semiparametric regression analysis of clustered survival data with semi-competing risks. *Computational Statistics & Data Analysis*, 124:53–70, 2018.”

for constant baseline hazard functions in Chapter 2. In each scenario, it can be seen that mean biases of the estimates are also close to zero with small sd's and corresponding cp nearly 95%. When the cluster size or the number of clusters m increases, both the mean bias and sd reduce, indicating the improvements in accuracy and efficiency, respectively.

For illustration purpose, Figures B.1-B.2 show the estimated cumulative baseline hazard curves in Scenario 2.1 and $\tau = 0.3$. They exhibit satisfactory performance of the proposed method in estimating unknown cumulative baseline hazard functions in both marginal models. We also have created figures for estimated cumulative baseline hazard curves in the rest Scenarios 2.2-2.4. Their overall trends are very similar to those seen in the figures for Scenario 2.1.

Table B.1. Simulation results based on 200 replications under the model with normally distributed random effects in the intercept in Scenario 2.1, where parameters $\eta = 0.2$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = t^{1/2}$, $\lambda_{02}(t) = \frac{3}{4}t^{1/2}$, and Kendall's $\tau = 0.3$ and 0.6.

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	-0.0551	0.0007	0.0018	-0.0189
	sd	0.1776	0.1337	0.2943	0.2873
	rmse	0.1860	0.1337	0.2943	0.2879
	cp(%)	95.5	96.0	96.5	96.0
(100, 5, 20)	bias	-0.0462	0.0111	0.0400	0.0634
	sd	0.1543	0.0674	0.1962	0.1862
	rmse	0.1611	0.0683	0.2002	0.2062
	cp(%)	95.0	96.0	93.5	94.0
(100, 10, 10)	bias	-0.0146	0.0022	0.0257	0.0376
	sd	0.1416	0.0729	0.2166	0.2001
	rmse	0.1424	0.0729	0.2181	0.2036
	cp(%)	96.5	95.5	93.5	94.0
(200, 10, 20)	bias	-0.0217	0.0027	0.0052	0.0185
	sd	0.1079	0.0516	0.1361	0.1208
	rmse	0.1101	0.0517	0.1362	0.1222
	cp(%)	96.5	95.0	94.5	95.5
$\tau = 0.6$					
(50, 5, 10)	bias	-0.0416	-0.0211	0.0661	0.0164
	sd	0.2027	0.0924	0.2776	0.2605
	rmse	0.2069	0.0948	0.2854	0.2610
	cp(%)	94.5	96.5	94.0	95.0
(100, 5, 20)	bias	-0.0548	-0.0296	0.0705	0.0263
	sd	0.1699	0.1011	0.2050	0.2092
	rmse	0.1785	0.1053	0.2168	0.2108
	cp(%)	96.5	96.5	93.5	94.5
(100, 10, 10)	bias	-0.0415	-0.0268	0.0554	0.0024
	sd	0.1518	0.0973	0.2007	0.2247
	rmse	0.1574	0.1009	0.2082	0.2247
	cp(%)	97.0	95.5	95.0	94.5
(200, 10, 20)	bias	-0.0281	-0.0253	0.0493	0.0078
	sd	0.1049	0.0842	0.1596	0.1928
	rmse	0.1086	0.0879	0.1670	0.1930
	cp(%)	97.5	94.0	94.0	94.0

Table B.2. Simulation results based on 200 replications under the model with normally distributed random effects in the intercept in Scenario 2.2, where parameters $\eta = 0.4$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = t^{1/2}$, $\lambda_{02}(t) = \frac{3}{4}t^{1/2}$ and Kendall's $\tau = 0.3$ and 0.6 .

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	-0.0972	0.0217	0.0257	0.0023
	sd	0.3174	0.1377	0.3114	0.2816
	rmse	0.3319	0.1394	0.3125	0.2816
	cp(%)	97.5	97.0	93.5	95.5
(100, 5, 20)	bias	-0.0837	0.0071	0.0092	0.0358
	sd	0.2564	0.0797	0.1879	0.2118
	rmse	0.2697	0.0800	0.1881	0.2148
	cp(%)	96.0	95.0	93.5	95.5
(100, 10, 10)	bias	-0.0167	0.0029	0.0107	0.0493
	sd	0.2633	0.0742	0.2057	0.1882
	rmse	0.2638	0.0743	0.2060	0.1946
	cp(%)	95.0	95.0	93.5	93.0
(200, 10, 20)	bias	-0.0270	0.0072	0.0044	0.0257
	sd	0.1933	0.0527	0.1354	0.1258
	rmse	0.1952	0.0532	0.1355	0.1284
	cp(%)	94.5	96.0	94.5	94.0
$\tau = 0.6$					
(50, 5, 10)	bias	-0.1040	-0.0105	0.0766	0.0100
	sd	0.3246	0.0839	0.2922	0.2830
	rmse	0.3409	0.0846	0.3021	0.2832
	cp(%)	96.5	96.5	93.5	94.5
(100, 5, 20)	bias	-0.1027	-0.0267	0.0830	0.0204
	sd	0.2646	0.0921	0.2099	0.2183
	rmse	0.2838	0.0959	0.2257	0.2193
	cp(%)	97.5	94.0	93.5	97.5
(100, 10, 10)	bias	-0.0859	-0.0296	0.0910	0.0491
	sd	0.2351	0.0845	0.1896	0.2213
	rmse	0.2503	0.0895	0.2103	0.2267
	cp(%)	99.0	94.5	91.0	96.5
(200, 10, 20)	bias	-0.0503	-0.0232	0.0695	0.0581
	sd	0.1782	0.0793	0.1565	0.1851
	rmse	0.1852	0.0826	0.1712	0.1940
	cp(%)	95.0	94.5	93.5	93.0

Table B.3. Simulation results based on 200 replications under the model with normally distributed random effects in the slope in Scenario 2.3, where parameters $\eta = 0.2$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = t^{1/2}$, $\lambda_{02}(t) = \frac{3}{4}t^{1/2}$, and Kendall's $\tau = 0.3$ and 0.6.

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	-0.0535	0.0035	0.0497	0.0581
	sd	0.1784	0.0951	0.3154	0.3138
	rmse	0.1862	0.0952	0.3193	0.3191
	cp(%)	97.5	93.0	94.5	95.0
(100, 5, 20)	bias	-0.0437	-0.0070	0.0175	0.0433
	sd	0.1559	0.0801	0.1931	0.1990
	rmse	0.1619	0.0804	0.1939	0.2037
	cp(%)	98.0	95.0	96.5	92.0
(100, 10, 10)	bias	-0.0114	0.0059	0.0275	0.0502
	sd	0.1497	0.0742	0.1989	0.2048
	rmse	0.1501	0.0744	0.2008	0.2109
	cp(%)	96.0	96.0	96.0	95.0
(200, 10, 20)	bias	-0.0182	0.0052	0.0045	0.0309
	sd	0.1111	0.0582	0.1222	0.1272
	rmse	0.1126	0.0584	0.1223	0.1309
	cp(%)	97.5	96.0	95.5	94.5
$\tau = 0.6$					
(50, 5, 10)	bias	-0.0672	-0.0396	0.0861	0.0802
	sd	0.1617	0.1151	0.2738	0.2962
	rmse	0.1751	0.1217	0.2870	0.3069
	cp(%)	96.5	93.0	93.0	94.0
(100, 5, 20)	bias	-0.0624	-0.0137	0.0413	-0.0082
	sd	0.1356	0.0785	0.1911	0.1895
	rmse	0.1493	0.0797	0.1955	0.1897
	cp(%)	96.5	96.5	95.5	94.0
(100, 10, 10)	bias	-0.0493	-0.0287	0.0578	0.0326
	sd	0.1155	0.0902	0.2189	0.2079
	rmse	0.1256	0.0947	0.2264	0.2104
	cp(%)	97.0	95.5	93.0	95.0
(200, 10, 20)	bias	-0.0538	-0.0266	0.0488	0.0250
	sd	0.1030	0.0896	0.1502	0.1648
	rmse	0.1162	0.0935	0.1579	0.1667
	cp(%)	96.5	94.5	93.5	94.0

Table B.4. Simulation results based on 200 replications under the model with normally distributed random effects in the slope in Scenario 2.4, where parameters $\eta = 0.4$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = t^{1/2}$, $\lambda_{02}(t) = \frac{3}{4}t^{1/2}$ and Kendall's $\tau = 0.3$ and 0.6.

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	-0.1053	0.0048	0.0276	0.0805
	sd	0.3260	0.1007	0.3338	0.3158
	rmse	0.3426	0.1008	0.3349	0.3259
	cp(%)	97.5	94.0	97.5	97.0
(100, 5, 20)	bias	-0.0939	0.0038	0.0054	0.0446
	sd	0.2472	0.0779	0.1974	0.1948
	rmse	0.2644	0.0780	0.1975	0.1998
	cp(%)	96.0	94.0	95.5	95.5
(100, 10, 10)	bias	-0.0679	0.0039	-0.0042	0.0319
	sd	0.2451	0.0730	0.1928	0.2052
	rmse	0.2543	0.0731	0.1928	0.2077
	cp(%)	96.5	96.5	96.5	95.5
(200, 10, 20)	bias	-0.0529	0.0013	0.0008	0.0293
	sd	0.2019	0.0625	0.1353	0.1317
	rmse	0.2087	0.0625	0.1353	0.1349
	cp(%)	97.0	96.5	94.0	93.5
$\tau = 0.6$					
(50, 5, 10)	bias	-0.1193	-0.0305	0.1283	0.0502
	sd	0.2397	0.1089	0.3293	0.2913
	rmse	0.2677	0.1131	0.3534	0.2956
	cp(%)	97.0	95.5	94.0	94.0
(100, 5, 20)	bias	-0.1041	-0.0329	0.0869	0.0747
	sd	0.2459	0.0870	0.2003	0.3173
	rmse	0.2670	0.0930	0.2183	0.3260
	cp(%)	95.5	97.0	99.5	97.5
(100, 10, 10)	bias	-0.0900	-0.0208	0.0683	0.0235
	sd	0.2028	0.0810	0.2048	0.2186
	rmse	0.2219	0.0836	0.2159	0.2199
	cp(%)	96.5	95.5	92.5	95.0
(200, 10, 20)	bias	-0.0548	-0.0227	0.0482	0.0438
	sd	0.1954	0.0816	0.1510	0.1654
	rmse	0.2029	0.0847	0.1585	0.1711
	cp(%)	95.0	94.5	94.0	95.0

(a) $N = 50, m = 5, n_i = 10$ (b) $N = 100, m = 5, n_i = 20$ (c) $N = 100, m = 10, n_i = 10$ (d) $N = 200, m = 10, n_i = 20$

Figure B.1. Simulation results for estimated cumulative baseline hazard function $\Lambda_{01}(\cdot)$ by the proposed method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{01}(t) = \frac{2}{3}t^{3/2}$ (colored black) under the model with normal random effects in Scenario 2.1 and $\tau = 0.3$.

(a) $N = 50, m = 5, n_i = 10$ (b) $N = 100, m = 5, n_i = 20$ (c) $N = 100, m = 10, n_i = 10$ (d) $N = 200, m = 10, n_i = 20$

Figure B.2. Simulation results for estimated cumulative baseline hazard function $\Lambda_{02}(\cdot)$ by the proposed method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{02}(t) = \frac{1}{2}t^{3/2}$ (colored black) under the model with normal random effects in Scenario 2.1 and $\tau = 0.3$.

Some notations used in Chapter 3

$$\begin{aligned}
w_{1j} &= \delta_i^{(1)} \delta_i^{(2)} \left[\frac{c_{1121}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}}{c_{112}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}} S_1(T_j) + 1 \right] \\
&\quad + \delta_i^{(1)} (1 - \delta_i^{(2)}) \left[\frac{c_{111}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}}{c_{11}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}} S_1(T_j) + 1 \right] \\
&\quad + (1 - \delta_i^{(1)}) (1 - \delta_i^{(2)}) \left[\frac{c_{11}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\} \cdot S_1(T_j)}{c_1\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}} \right] \\
&\quad + (1 - \delta_i^{(1)}) \delta_i^{(2)} \left[\frac{c_{121}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}}{c_{12}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}} S_1(T_j) \right], \\
w_{2j} &= \delta_i^{(1)} \delta_i^{(2)} \left[\frac{c_{1122}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}}{c_{112}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}} S_2(\tilde{T}_j) + 1 \right] \\
&\quad + \delta_i^{(1)} (1 - \delta_i^{(2)}) \left[\frac{c_{112}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}}{c_{11}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}} S_2(\tilde{T}_j) \right] \\
&\quad + (1 - \delta_i^{(1)}) (1 - \delta_i^{(2)}) \left[\frac{c_{12}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\} \cdot S_2(\tilde{T}_j)}{c_1\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}} \right] \\
&\quad + (1 - \delta_i^{(1)}) \delta_i^{(2)} \left[\frac{c_{122}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}}{c_{12}\{S_1(T_j), S_2(\tilde{T}_j); \alpha_1\}} S_2(\tilde{T}_j) + 1 \right], \\
&\quad + \delta_i^{(2)} \delta_i^{(3)} \left[\frac{c_{2122}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}}{c_{212}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}} S_2(\tilde{T}_j) + 1 \right] \\
&\quad + \delta_i^{(2)} (1 - \delta_i^{(3)}) \left[\frac{c_{212}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}}{c_{21}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}} S_2(\tilde{T}_j) \right] \\
&\quad + (1 - \delta_i^{(2)}) (1 - \delta_i^{(3)}) \left[\frac{c_{22}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\} \cdot S_2(\tilde{T}_j)}{c_2\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}} \right] \\
&\quad + (1 - \delta_i^{(2)}) \delta_i^{(3)} \left[\frac{c_{222}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}}{c_{22}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}} S_2(\tilde{T}_j) + 1 \right],
\end{aligned}$$

$$\begin{aligned}
w_{3j} = & \delta_i^{(2)} \delta_i^{(3)} \left[\frac{c_{2121}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}}{c_{212}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}} S_3(Y_j) + 1 \right] \\
& + \delta_i^{(2)} (1 - \delta_i^{(3)}) \left[\frac{c_{211}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}}{c_{21}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}} S_3(Y_j) + 1 \right] \\
& + (1 - \delta_i^{(2)}) (1 - \delta_i^{(3)}) \left[\frac{c_{21}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\} \cdot S_3(Y_j)}{c_2\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}} \right] \\
& + (1 - \delta_i^{(2)}) \delta_i^{(3)} \left[\frac{c_{221}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}}{c_{22}\{S_2(\tilde{T}_j), S_3(Y_j); \alpha_2\}} S_3(Y_j) \right],
\end{aligned}$$

where

$$\begin{aligned}
c_{11}(u, v; \alpha_1) &= \frac{\partial c_1(u, v; \alpha_1)}{\partial u}, & c_{12}(u, v; \alpha_1) &= \frac{\partial c_1(u, v; \alpha_1)}{\partial v}, \\
c_{111}(u, v; \alpha_1) &= \frac{\partial^2 c_1(u, v; \alpha_1)}{\partial^2 u}, & c_{112}(u, v; \alpha_1) &= \frac{\partial^2 c_1(u, v; \alpha_1)}{\partial u \partial v}, \\
c_{121}(u, v; \alpha_1) &= \frac{\partial^2 c_1(u, v; \alpha_1)}{\partial v \partial u}, & c_{122}(u, v; \alpha_1) &= \frac{\partial^2 c_1(u, v; \alpha_1)}{\partial^2 v}, \\
c_{1121}(u, v; \alpha_1) &= \frac{\partial^3 c_1(u, v; \alpha_1)}{\partial u \partial v \partial u}, & c_{1122}(u, v; \alpha_1) &= \frac{\partial^3 c_1(u, v; \alpha_1)}{\partial u \partial v \partial v}, \\
c_{21}(u, v; \alpha_2) &= \frac{\partial c_2(u, v; \alpha_2)}{\partial u}, & c_{22}(u, v; \alpha_2) &= \frac{\partial c_2(u, v; \alpha_2)}{\partial v}, \\
c_{211}(u, v; \alpha_2) &= \frac{\partial^2 c_2(u, v; \alpha_2)}{\partial^2 u}, & c_{212}(u, v; \alpha_2) &= \frac{\partial^2 c_2(u, v; \alpha_2)}{\partial u \partial v}, \\
c_{221}(u, v; \alpha_2) &= \frac{\partial^2 c_2(u, v; \alpha_2)}{\partial v \partial u}, & c_{222}(u, v; \alpha_2) &= \frac{\partial^2 c_2(u, v; \alpha_2)}{\partial^2 v}, \\
c_{2121}(u, v; \alpha_2) &= \frac{\partial^3 c_2(u, v; \alpha_2)}{\partial u \partial v \partial u}, & c_{2122}(u, v; \alpha_2) &= \frac{\partial^3 c_2(u, v; \alpha_2)}{\partial u \partial v \partial v}.
\end{aligned}$$

Proofs of Theorems 4.1-4.2

Proof of Theorem 4.1

Let

$$r_k^* = \left\{ \frac{1}{n} \sum_{i=1}^n X_{ik} S(T_i, \tilde{T}_i) \right\}^2.$$

We prove this theorem via two steps.

First, we derive the exponential tail probability bound of $\Pr(|\hat{r}_k - r_k^*| \geq c_0 n^{-w})$ for any positive constants c_0 and $0 \leq w < 1/2$. Following the idea of proof in the Appendix of Zhang et al. (2017), under Condition 4.2, there exists a positive constant c_4 such that $\frac{1}{n} \sum_{i=1}^n X_{ik}^2 \leq c_4^2$, $\left| \frac{1}{n} \sum_{i=1}^n X_{ik} \hat{S}(T_i, \tilde{T}_i) \right| \leq c_4$, and $\left| \frac{1}{n} \sum_{i=1}^n X_{ik} S(T_i, \tilde{T}_i) \right| \leq c_4$. Using the above inequalities, we have

$$\begin{aligned} |\hat{r}_k - r_k^*| &\leq c_5 \left| \frac{1}{n} \sum_{i=1}^n X_{ik} \hat{S}(T_i, \tilde{T}_i) - \frac{1}{n} \sum_{i=1}^n X_{ik} S(T_i, \tilde{T}_i) \right| \\ &\leq c_5 c_4 \left[\frac{1}{n} \sum_{i=1}^n \{ \hat{S}(T_i, \tilde{T}_i) - S(T_i, \tilde{T}_i) \}^2 \right]^{1/2} \\ &\leq c_6 \|\hat{S} - S\|, \end{aligned}$$

where $c_5 = 2c_4$, $c_6 = c_5 c_4$, $\|\hat{S} - S\|$ is the maximum of $|\hat{S}(\cdot, \cdot) - S(\cdot, \cdot)|$ in the region $[0, \varsigma]^2$.

Under Condition 4.1, it has been proved in Lin and Ying (1993) that $n^{1/2}\{\hat{S}(\cdot, \cdot) - S(\cdot, \cdot)\}$ converges weakly to a zero-mean Gaussian process with finite variance on $[0, \varsigma]^2$. Refer to the Theorem 2.8 (Azaïs and Wschebor (2009), chap. 2), we have that for every $u > 0$,

$$\Pr\left(n^{1/2}\|\hat{S} - S\| \geq u\right) \leq \exp(-c_7 u^2), \quad (\text{D.1})$$

where c_7 is a positive constant. Thus

$$\begin{aligned} \Pr(|\hat{r}_k - r_k^*| \geq c_0 n^{-w}) &\leq \Pr\left(c_6 \|\hat{S} - S\| \geq c_0 n^{-w}\right) \\ &= \Pr\left(n^{1/2} \|\hat{S} - S\| \geq c_6^{-1} c_0 n^{1/2-w}\right) \\ &\leq \exp(-c_7 c_6^{-2} c_0^2 n^{1-2w}). \end{aligned} \quad (\text{D.2})$$

Second, we derive the exponential tail probability bound of $\Pr(|r_k^* - r_k| \geq c_0 n^{-w})$ for any positive constants c_0 and $0 \leq w < 1/2$. Following the idea of proof in the Appendix of Zhang et al. (2017), we can get

$$\Pr(|r_k^* - r_k| \geq c_0 n^{-w}) \leq \exp\left\{-\frac{n^{1/2-w}}{(\log \log n)^{1/2}} c_5^{-1} c_0 + c_3\right\}, \quad (\text{D.3})$$

where c_3 is a positive constant.

Combining (D.2) and (D.3), we have

$$\begin{aligned} \Pr(|\hat{r}_k - r_k| \geq c n^{-w}) &\leq \Pr(|\hat{r}_k - r_k^*| \geq c_0 n^{-w}) + \Pr(|r_k^* - r_k| \geq c_0 n^{-w}) \\ &\leq \exp(-c_7 c_6^{-2} c_0^2 n^{1-2w}) + \exp\left(-\frac{n^{1/2-w}}{(\log \log n)^{1/2}} c_5^{-1} c_0 + c_3\right), \\ &= \exp(-c_1 n^{-2w}) + \exp\left(-\frac{c_2 n^{1/2-w}}{(\log \log n)^{1/2}} + c_3\right), \end{aligned} \quad (\text{D.4})$$

where $c = 2c_0$, $c_1 = c_7 c_6^{-2} c_0^2$, $c_2 = c_5^{-1} c_0$. Immediately, we have

$$\Pr\left(\max_{1 \leq k \leq p} |\hat{r}_k - r_k| \geq c n^{-w}\right) \leq p \left\{ \exp(-c_1 n^{-2w}) + \exp\left(-\frac{c_2 n^{1/2-w}}{(\log \log n)^{1/2}} + c_3\right) \right\}. \quad (\text{D.5})$$

which proves the first part of Theorem 4.1.

Following the idea of proof in the Appendix of Zhang et al. (2017), based on (D.4), we have

$$\begin{aligned}\Pr(\mathcal{A} \subseteq \hat{\mathcal{A}}) &\geq \Pr(\max_{k \in \mathcal{A}} |\hat{r}_k - r_k| \leq cn^{-w}) \\ &\geq 1 - s \left\{ \exp(-c_1 n^{-2w}) + \exp\left(-\frac{c_2 n^{1/2-w}}{(\log \log n)^{1/2}} + c_3\right) \right\},\end{aligned}$$

where $s = |\mathcal{A}|$. Thus, we have proved Theorem 4.1.

Proof of Theorem 4.2

Following the idea of proof in the Appendix of Zhang et al. (2017), under Condition 4.3 and assumption (i) and (ii), coupled with (D.5), we have

$$\begin{aligned}\Pr(\min_{k \in \mathcal{A}} \hat{r}_k \leq \max_{k \in \mathcal{A}^c} \hat{r}_k) &\leq 2\Pr\left(\max_{1 \leq k \leq p} |\hat{r}_k - r_k| \geq cn^{-w}\right) \\ &\leq 2p \left\{ \exp(-c_1 n^{-2w}) + \exp\left(-\frac{c_2 n^{1/2-w}}{(\log \log n)^{1/2}} + c_3\right) \right\}.\end{aligned}$$

Thus, the proof of Theorem 4.2 is completed.

Simulations for Gumbel copula in Chapter 4

In addition to Clayton copula, we hereby examine the performance of the proposed method when the joint survival function for the nonterminal and terminal event times in the simulation is specified by a Gumbel copula function with an unknown constant parameter α one-to-one corresponding to Kendall's τ and $\alpha = 1/(1 - \tau)$ (Nelsen, 2007). We only include the results with $\tau = 0.2, 0.8$ under Scenarios 4.1 for simplicity.

Tables E.1-E.3 report the screening results under Scenario 4.1 for Gumbel copula. The overall results are very similar to those given in Tables 4.1, Table 4.4 and Table 4.7 for Clayton copula in Chapter 4. The proposed JCR method outperforms the CR method in terms of all three criteria for selection proportions. As the sample size increases, the selection proportions increase, and the difference of selection proportions between two methods decreases. The five quantiles of the minimum model size obtained by JCR method are all smaller than those obtained by the CR method, especially in the cases with small sample sizes. It is seen that the minimum model size is decreasing as sample size increases. The JCR screening for $(T^{(1)}, T^{(2)})$ requires a large enough sample size (say $n \geq 300$) for successful selection of all active covariates with a very high probability ranging from 97% to 100%.

Table E.1. Selection proportions for each active covariate, for all active covariates (i.e. $X_1, \dots, X_3$, denoted by $X_{1,2,3}$) associated with $T^{(2)}$ and for all active covariates (i.e. $X_1, \dots, X_6$, denoted by $X_{1,\dots,6}$) associated with $(T^{(1)}, T^{(2)})$ based on 500 replications under Scenario 4.1 and Gumbel copula.

τ	n	Method	X_1	X_2	X_3	X_4	X_5	X_6	$X_{1,\dots,4}$	$X_{1,\dots,6}$
0.2	100	JCR	0.546	0.594	0.632	0.592	0.560	0.454	0.456	0.276
		CR	0.458	0.472	0.446	0.374	0.292	0.192	0.332	0.114
	200	JCR	0.956	0.974	0.978	0.964	0.938	0.882	0.940	0.840
		CR	0.920	0.948	0.916	0.812	0.726	0.624	0.860	0.566
	300	JCR	0.990	0.996	1.000	0.996	0.990	0.986	0.988	0.976
		CR	0.986	0.986	0.988	0.958	0.894	0.822	0.978	0.798
0.8	100	JCR	0.496	0.534	0.572	0.588	0.558	0.462	0.412	0.270
		CR	0.462	0.460	0.460	0.372	0.264	0.176	0.324	0.110
	200	JCR	0.908	0.956	0.970	0.964	0.962	0.906	0.894	0.824
		CR	0.906	0.918	0.906	0.822	0.714	0.626	0.838	0.542
	300	JCR	0.982	0.992	0.996	0.996	0.994	0.994	0.978	0.974
		CR	0.984	0.984	0.990	0.958	0.904	0.830	0.976	0.812

Table E.2. Quantiles of the minimum model size which includes all active co-variates (i.e. X_1, X_2, X_3) associated with $T^{(2)}$ based on 500 replications under Scenario 4.1 and Gumbel copula.

τ	n	Method	5%	25%	50%	75%	95%
0.2	100	JCR	3.00	8.00	27.00	123.00	735.65
		CR	4.00	13.00	57.50	209.00	1085.25
	200	JCR	3.00	4.00	6.00	10.00	44.35
		CR	3.00	4.00	6.00	18.00	98.45
	300	JCR	3.00	3.00	4.00	5.00	14.00
		CR	3.00	3.00	4.00	5.00	25.00
0.8	100	JCR	4.00	10.00	40.00	174.25	856.55
		CR	3.00	13.00	66.00	215.25	992.95
	200	JCR	3.00	5.00	7.00	14.00	76.10
		CR	3.00	3.00	6.00	18.00	101.05
	300	JCR	3.00	3.00	4.00	6.00	24.00
		CR	3.00	3.00	4.00	6.00	27.10

Table E.3. Quantiles of the minimum model size which includes all active covariates (i.e. $X_1, \dots, X_6$) associated with $(T^{(1)}, T^{(2)})$ obtained by the proposed JCR method, based on 500 replications under Scenario 4.1 and Gumbel copula.

τ	n	5%	25%	50%	75%	95%
0.2	100	6.00	18.00	60.00	226.25	1146.75
	200	6.00	6.00	9.00	19.00	106.00
	300	6.00	6.00	7.00	8.00	28.00
0.8	100	6.00	19.00	75.50	271.00	1107.20
	200	6.00	7.00	9.00	22.00	125.10
	300	6.00	6.00	7.00	9.00	28.05