Copula Models for Multi-type Life History Processes

by

Liqun Diao

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics - Biostatistics

Waterloo, Ontario, Canada, 2013

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis considers statistical issues in the analysis of data in the studies of chronic diseases which involve modeling dependencies between life history processes using copula functions.

Many disease processes feature recurrent events which represent events arising from an underlying chronic condition; these are often modeled as point processes. In addition, however, there often exists a random variable which is realized upon the occurrence of each event, reflecting the severity of the event, the cost associated with its occurrence, or possibly a short term response indicating the effect of a therapeutic intervention, which is called a mark of the point process. When considered together, such processes are called marked point processes. Examples arise in diverse areas such as insurance, finance, biology, medicine, seismology, etc. In the existing literature, marks are often assumed to be independent of the points to enable one to model the marks and points separately, but this assumption is often questionable. A novel copula model for the marked point process is described here which uses copula functions to govern the association between marks and event times. Specifically, a copula function is used to link each mark with the next event time following the realization of that mark to reflect the pattern in the data wherein larger marks are often followed by longer time to the next event. Our model ensures that (1) a dependence structure between continuous marks and the event process is incorporated, (2) the marginal models for the events and the marks are compatible with standard models for recurrent event and longitudinal analyses respectively, (3) (conditional) independence assumptions made are weaker than those in the existing literature, and (4) efficiency is

gained compared to separate analysis of the marks and the events.

The extent of organ damage in an individual can often be characterized by ordered states, and interest frequently lies in modeling the rates at which individuals progress through these states. Risk factors can be studied and the effect of therapeutic interventions can be assessed based on relevant multistate models. When chronic diseases affect multiple organ systems, joint modeling of progression in several organ systems is also important. Disease progression in each organ can be characterized by a progressive multistate Markov process and an affected individual experiences multiple such processes. In contrast to common intensity-based or frailty-based approaches to modelling, this thesis considers a copula-based framework for modeling and analysis. Through decomposition of the density and by use of conditional independence assumptions, an appealing joint model is obtained by assuming that the joint survival function of absorption transition times is governed by a multivariate copula function. The copula formulation herein ensures that (1) a wide range of marginal processes can be specified, (2) the joint model will retain these marginal features to provide simple estimates and straightforward interpretation of transition rates and covariate effects for each component process, and (3) the scientific understanding regarding the relation between processes is facilitated. Different approaches to estimation and inference are discussed and compared including composite likelihood and two-stage estimation methods. Special attention is paid to the case of interval-censored data arising from intermittent assessment.

Attention is also directed to use of copula models for more general scenarios with a focus on semiparametric two-stage estimation procedures. In this approach nonparametric or semiparametric estimates of the marginal survivor functions are obtained in the

first stage and estimates of the association parameters are obtained in the second stage. Bivariate failure time models are considered for data under right-censoring and current status observation schemes, and right-censored multistate models. A new expression for the asymptotic variance of the second-stage estimator for the association parameter along with a way of estimating this for finite samples are presented under these models and observation schemes.

Acknowledgements

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me.

Foremost, I would like to express my deep and sincere gratitude to my supervisor Dr. Richard J. Cook for the continuous support of my Ph.D study and research, for his patience, kindness, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better mentor for my Ph.D. study.

I would also like to thank my thesis committee, Dr. Grace Y. Yi, Dr. Leilei Zeng, Dr. Carlos Rojas-Fernandez, and Dr. Joan Hu, for their helpful suggestions, insightful comments and inspiring questions. My special thanks go to Dr. Grace Y. Yi for sharing invaluable experiences with me and serving as a role model to me as a successful female researcher of academia.

It has been a great privilege to spend the last five years in the Department of Statistics and Actuarial Science at University of Waterloo. I would like to express my heartfelt thanks to all faculty members and staff. Especially, I would like to thank Ms. Mary Lou Dufton for addressing my numerous questions with patience and Ker-Ai Lee for help with the statistical computing.

My friends in Waterloo were sources of joy. Special thanks go to Min Chen, Lu Cheng, Celia Huang, Ruitong Huang, Zhiyue Huang, Kexin Ji, Meaghan Knight, Haocheng Li, Gracia Liang, Chris Men, Chao Qiu, Narges Nazeri Rad, Hua Shen, Longyang Wu, Lu Xin, Ying Yan, Zhou Ye, Hui Zhao, George Zhang, Yujie Zhong, Yayuan Zhu.

Last but not the least, I would like to thank my parents Shuren Diao and Jinghui Li for loving and supporting me unconditionally and being good role models for me. During my Ph.D. study, I met, fell in love and tied the knot with my husband, Chengguo Weng. I am blessed and grateful to have him standing by me, understanding me and making sacrifices for me.

Dedication

This thesis is dedicated to my grandparents Fengwu Li and Jianfan Zhao, who passed away in 1997 and 2003. I spent my happy childhood with them. They are the most generous and upright people. I will love and miss them forever.

Table of Contents

List of Tables **xiii**

List of Figures **xviii**

1 Introduction **1**

1.1 An Overview of Relevant Statistical Models 2

1.1.1 Marked Point Processes 2

1.1.2 Multistate Processes 3

1.1.3 Joint Modeling Event Times via Copulas 6

1.2 Motivating Studies in Health Research 8

1.2.1 Platelet Transfusions in Thrombocytopenic Patients 8

1.2.2 Joint Damage in Psoriatic Arthritis 10

1.2.3 Skeletal Complications from Bone Metastases 13

1.2.4 Seroconversion Following Anticoagulation Therapy 14

1.3 Outline of Thesis 15

2 A Copula Model for Marked Point Prcessess 17

2.1 Introduction 17

2.1.1 Overview 17

2.2 Model Formulation and Estimation Procedures 20

2.2.1 A Copula-Based Marked Point Process Model 20

2.2.2 Marginal Models for the Recurrent Events 25

2.2.3 Marginal Models for the Marks 27

2.2.4 Estimation and Statistical Inference 28

2.3 Simulation Studies 37

2.3.1 Empirical Performance under Correct Model Specification 37

2.3.2 Empirical Performance under Model Misspecification 47

2.4 Application to Data from the Mirasol Study 55

2.5 General Remarks 60

3 Correlated Markov Processes Under Interval Censoring 61

3.1 Introduction 61

3.1.1 Overview 61

3.1.2 Review of Composite Likelihood 64

3.2 Multiple Multistate Processes 66
3.3 Estimation and Inference 72
3.3.1 Notation for Interval-Censored Data 72
3.3.2 Composite Likelihood Construction 73
3.3.3 Two-stage Estimation 76
3.4 Simulation Studies and Illustration 77
3.4.1 Design and Analysis of Simulation Studies 77
3.4.2 Joint Damage Progression in Individuals with Arthritis 80
3.5 General Remarks 84

4 Semiparametric Two-stage Estimation for Copula Models in Survival Data 87

4.1 Introduction 87
4.2 Estimation and Inference 90
4.2.1 Preliminary Remarks 90
4.2.2 Data Structure 90
4.2.3 Stage I: Estimation of the Marginal Survivor Functions 91
4.2.4 Stage II: Estimation of the Association Parameter 94
4.2.5 Asymptotic Properties of the Estimator $\widetilde{\phi}$ 96
4.2.6 Regression Problem 97

4.2.7 Discrete Parameterization 100
4.3 Examples and Simulation Studies 100
4.3.1 Bivariate Right-censored Survival Data 101
4.3.2 Bivariate Current Status Survival Data 111
4.3.3 Absorption Times in Bivariate Multistate Processes 112
4.4 Application 115
4.4.1 Right-censored Data 115
4.4.2 Current Status Data 117
4.5 General Remarks 119
Appendix A: Asymptotic Normality of $\widetilde{\phi}$ 119
Appendix B: Consistency of the Asymptotic Variance Estimator 128

5 General Remarks and Future Research **133**

5.1 Overview 133
5.2 A Copula Model For Marked Point Processes 135
5.3 Multiple Multistate Processes Under Intermittent Inspection 136
5.4 Semiparametric Two-stage Estimation Procedure in Copula Models for Survival Data 138
5.5 Analysis of Recurrent Episodes in Chronic Disease with Vine Copula Models 139

References **141**

List of Tables

2.1 Frequency properties of parameter estimators for the recurrent event process through joint analysis of the marked point process with a Poisson margin under 40% random right censoring where $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; 500 simulated samples . . 40

2.2 Frequency properties of mean parameter estimators of the marks through joint analysis of the marked point process with a Poisson margin under 40% random right censoring where $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; $\gamma=0.5$; 500 simulated samples . . . 41

2.3 Frequency properties of dependence parameter estimators through joint analysis of the marked point process with a Poisson margin under 40% random right censoring where $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; $\gamma=0.5$; 500 simulated samples . . . 42

2.4 Frequency properties of parameter estimators for the recurrent event process through joint analysis of the marked point process with a negative binomial margin under 40% random right censoring where $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1 = 6.35$, $\alpha_2 = 0.75$, $\beta = \log 0.5$; $\eta_0 = 0$, $\eta_1 = 0.5$, $\sigma^2 = 1$; $\gamma = 0.5$; 500 simulated samples 44

2.5 Frequency properties of mean parameter estimators of the marks through joint analysis of the marked point process with a negative binomial margin under 40% random right censoring where $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1 = 6.35$, $\alpha_2 = 0.75$, $\beta = \log 0.5$; $\eta_0 = 0$, $\eta_1 = 0.5$, $\sigma^2 = 1$; $\gamma = 0.5$; 500 simulated samples 45

2.6 Frequency properties of parameter estimators for dependence parameters through joint analysis of the marked point process with a negative binomial margin under 40% random right censoring where $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1 = 6.35$, $\alpha_2 = 0.75$, $\beta = \log 0.5$; $\eta_0 = 0$, $\eta_1 = 0.5$, $\sigma^2 = 1$; $\gamma = 0.5$; 500 simulated samples 46

2.7 Sensitivity analysis on the effect of misspecification of the marginal recurrent event process generated from negative binomial process and analyzed by Poisson process showing the frequency properties of parameter estimators from the recurrent event process via joint analysis of the marked point process; 40% random right censoring with $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1 = 6.35$, $\alpha_2 = 0.75$, $\beta = \log 0.5$; $\eta_0 = 0$, $\eta_1 = 0.5$, $\sigma^2 = 1$; $\gamma = 0.5$; 500 simulated samples 48

2.8 Sensitivity analysis on the effect of misspecification of the marginal recurrent event process generated from negative binomial process and analyzed by Poisson process showing the frequency properties of mean parameter estimators for the marks via joint analysis of the marked point process; 40% random right censoring with $\mathrm{E}(N(1)|Z = 0) = 4$; $\alpha_1 = 6.35$, $\alpha_2 = 0.75$, $\beta = \log 0.5$; $\eta_0 = 0$, $\eta_1 = 0.5$, $\sigma^2 = 1$; $\gamma = 0.5$; 500 simulated samples . . . 49

2.9 Sensitivity analysis on the effect of misspecification of the marginal recurrent event process generated from negative binomial process and analyzed by Poisson process showing the frequency properties of dependence parameter estimators via joint analysis of the marked point process; 40% random right censoring with $\mathrm{E}(N(1)|Z = 0) = 4$; $\alpha_1 = 6.35$, $\alpha_2 = 0.75$, $\beta = \log 0.5$; $\eta_0 = 0$, $\eta_1 = 0.5$, $\sigma^2 = 1$; $\gamma = 0.5$; 500 simulated samples 50

2.10 Sensitivity analysis on the effect of copula misspecification showing the frequency properties of parameter estimators from the recurrent event process via joint analysis of the marked point process; marginal intensity is Poisson; 40% random right censoring with $\mathrm{E}(N(1)|Z = 0) = 4$; $\alpha_1 = 6.35$, $\alpha_2 = 0.75$, $\beta = \log 0.5$; $\eta_0 = 0$, $\eta_1 = 0.5$, $\sigma^2 = 1$; 200 subjects/sample; 500 simulated samples 52

2.11 Sensitivity analysis on the effect of copula misspecification showing the frequency properties of mean parameter estimators of the marks via joint analysis of the marked point process; marginal intensity is Poisson; 40% random right censoring with $\mathrm{E}(N(1)|Z = 0) = 4$; $\alpha_1 = 6.35$, $\alpha_2 = 0.75$, $\beta = \log 0.5$; $\eta_0 = 0$, $\eta_1 = 0.5$, $\sigma^2 = 1$; 200 subjects/sample; 500 simulated samples 53

2.12 Sensitivity analysis on the effect of copula misspecification showing the frequency properties of dependence parameter estimators via joint analysis of the marked point process; marginal intensity is Poisson; 40% random right censoring with $\mathrm{E}(N(1)|Z = 0) = 4$; $\alpha_1 = 6.35$, $\alpha_2 = 0.75$, $\beta = \log 0.5$; $\eta_0 = 0$, $\eta_1 = 0.5$, $\sigma^2 = 1$; 200 subjects/sample; 500 simulated samples . . . 54

2.13 Results from marginal and joint analyses of the transfusion events and the 24 hour corrected count increments (m^2/ℓ) from the Mirasol Study (Cazenave *and others*, 2010). 57

3.1 Frequency properties of estimators using three estimation methods (1000 individuals per sample; 2000 simulated samples) 79

3.2 Joint analysis of progression in the left and right sacroiliac joints in psoriatic arthritis (PsA) with the covariate HLA B27 and allowing different parameters in the two processes 82

4.1 Empirical frequency properties of the estimates and associated variance estimates for the association parameter of bivariate survival times with no-censoring; 2000 simulated samples 106

4.2 Empirical frequency properties of the estimates and associated variance estimates for the association parameter of bivariate survival times with 30% right-censoring; 2000 simulated samples 107

4.3 Empirical frequency properties of the estimates and associated variance estimates for the association parameter of bivariate survival times with covariate under 30% right-censoring; 2000 simulated samples 110

4.4 Empirical frequency properties of the estimates and associated variance estimates for the association parameter of bivariate current status data with 50% prevalence level; 2000 simulated samples 112

4.5 Empirical frequency properties of the estimates and associated variance estimates for the parameters of bivariate multistate processes with random right-censoring; $m = 500$; 2000 simulated samples 114

4.6 Results from semiparametric two-stage estimation of the association between the survival times for two types of events, fracture and need for radiotherapy 116

List of Figures

1.1 Profiles of four patients from the Mirasol study (Cazenave *and others*, 2010) showing the times of the transfusions, the corrected count increments and the duration of follow-up; the horizontal line at 4.5 $\times 10^9$ m^2/ℓ is the threshold for a successful transfusion 9

1.2 Multistate diagram for modeling damage in sacroiliac joints 11

1.3 Multivariate multistate processes observed at intermittent inspection times 12

2.1 Schematic diagram with notation for a marked point process 21

2.2 Plots of estimated baseline mean function obtained by integrating the marginal rate from a marginal analysis under a Poisson assumption with a parametric Weibull rate and semiparametric Andersen-Gill and from a joint analysis using (2.2.5) with a Poisson and negative binomial parametric marginal rate 59

3.1 State space diagram for multivariate multistate processes 66

3.2 A D-vine with four variables 70

3.3 Plots of the cumulative probability of ankylosing spondylitis by B27 status according to the composite likelihood (3.3.6) analysis from the joint model and based on nonparametric estimate of Gentleman and Vandal (2002); for the fitted parametric model the estimated joint probability is $P(\zeta_1(t) = \zeta_2(t) = 3|\zeta_1(0) = \zeta_2(0) = 1; \widehat{\psi})$ 83

3.4 Plots of the estimated conditional probability $P(\zeta_1(t) = 3|\zeta_1(0) = 1, \zeta_2(10) = 1, z; \widehat{\psi})$, $P(\zeta_1(t) = 3|\zeta_1(0) = 1, \zeta_2(10) = 2, z; \widehat{\psi})$ and $P(\zeta_1(t) = 3|\zeta_1(0) = 1, \zeta_2(10) = 3, z; \widehat{\psi})$ vs. time since disease onset (years) for the composite likelihood (3.3.6) analysis using the joint model 85

4.1 Plots of the cumulative probability of responses vs. time since surgery (days) 118

Chapter 1

Introduction

This thesis is concerned with the modeling and analysis of multifaceted life history processes. Innovative statistical models are developed which have appealing marginal properties and convenient parameterizations of association. Statistical inference is carried out based on likelihood and composite likelihood with both simultaneous and multistage estimation procedures. The relative efficiencies of joint versus marginal analyses are assessed and useful findings are made to help inform the planning of analyses.

We begin with a brief review of models and methods for life history analysis and copula models before describing the motivating studies and providing a brief overview of the remainder of the thesis.

1.1 An Overview of Relevant Statistical Models

1.1.1 Marked Point Processes

Recurrent event data can be analyzed by point process models (Cox and Isham, 1980; Snyder and Miller, 1991; Karr, 1991). For the purpose of both modelling and statistical analysis, the concepts of intensity functions and counting processes are particularly useful (Andersen *and others*, 1993). Let $Z = 1$ if the individual is randomized to receive the experimental intervention and $Z = 0$ otherwise. For a single recurrent event process starting for simplicity at $t = 0$, let $0 < T_1 < T_2 < \ldots$ denote the event times, where T_k is the time of the kth event. $N(t) = \sum_{k=1}^{\infty} I(T_k \leq t)$ is the number of events occurring over the time interval $[0, t]$, where $I(\cdot)$ is the indicator function such that $I(A) = 1$ if A is true and $I(A) = 0$ otherwise. The associated counting process $\{N(t), 0 < t\}$ is right-continuous and it records the cumulative number of events generated by the process and their respective times. More generally, $N(s, t) = N(t) - N(s)$ represents the number of events occurring over the interval $(s, t]$. We let t^- denote times that are infinitesimally smaller than t and let $\Delta N(t) = N(t + \Delta t) - N(t^-)$ denote the number of events in the interval $[t, t + \Delta t]$. We let $\mathcal{H}^N(t) = \{N(u), 0 < u < t\}$ denote the history of the process at time t. For events occurring in continuous time we make the mathematically convenient assumption that two events cannot occur simultaneously. The event intensity function (Cook and Lawless, 2007) gives the instantaneous probability of an event occurring at t,

conditional on the event process history, and is defined mathematically as

$$\lambda(t|\mathcal{H}^N(t)) = \lim_{\Delta t \downarrow 0} \frac{P\{\Delta N(t) = 1|\mathcal{H}^N(t)\}}{\Delta t} .$$

Models for recurrent events can be defined very generally by specifying the intensity function.

In many settings there are auxiliary data associated with an event that reflect the severity, importance, or implications of its occurrence. The associated random variable is called the mark of the event and the process as a whole is called a marked point process (Daley and Vere-Jones, 2008). Let Y_k be a possibly vector-valued random variable that denotes the mark realized at the occurrence of kth event. We let $Y(t) = \{Y_0, \ldots, Y_{N(t)}\}$ denote the set of marks realized over $[0, t]$. It is convenient to denote the history of the marks at t as $\mathcal{H}^Y(t) = \{Y(u), 0 < u < t\}$. The full history of a marked point process is then $\mathcal{H}(t) = \{N(u), Y(u), 0 < u < t, Z\}$ and the corresponding complete event intensity is

$$\lambda(t|\mathcal{H}(t)) = \lim_{\Delta t \downarrow 0} \frac{P\{\Delta N(t) = 1|\mathcal{H}(t)\}}{\Delta t} .$$

1.1.2 Multistate Processes

A multistate process is a stochastic process with a finite state space and a right-continuous sample path (Hougaard, 1999). Such processes can be used to describe how a disease leads to changes in conditions over time. With progressive disease processes, the extent of damage can be characterized by ordered states $1, 2, \ldots, K + 1$, where state 1 represents no

impairment and state $K+1$ represents the most severe degree of impairment or damage. In this setting, the only possible transition at any instant in time is to the state representing the next stage of damage (i.e. $k \to k+1$ transitions for $k = 1, 2, \ldots, K$), thus we use the term "progressive" multistate process.

Let $\zeta(t)$ represent the state occupied by the disease process at time t and $\{\zeta(t), 0 < t\}$ denote the associated stochastic process. A $p \times 1$ vector of time-independent covariates is denoted by Z and the history of the process is denoted by $\mathcal{H}(t) = \{\zeta(u), 0 < u < t, Z\}$. The term $I(\zeta(t^-) = k)$ indicates whether the individual is at risk of a $k \to k+1$ transition at time t, where $\zeta(t^-)$ is the left limit of $\zeta(u)$ for $u \uparrow t$. The intensity function governing transitions out of state k is defined as

$$\lambda_k(t|\mathcal{H}(t)) = I(\zeta(t^-) = k) \lim_{\Delta t \downarrow 0} \frac{P\{\zeta(t + \Delta t) = k + 1 | \mathcal{H}(t), \zeta(t^-) = k\}}{\Delta t} ,$$

$k = 1, \ldots, K$.

A Markov assumption (Andersen *and others*, 1993), common in studies of progressive conditions, gives

$$\lambda_k(t|\mathcal{H}(t)) = I(\zeta(t^-) = k)\lambda_k(t|z) ,$$

where

$$\lambda_k(t|z) = \lim_{\Delta t \downarrow 0} \frac{P\{\zeta(t + \Delta t) = k + 1 | \zeta(t^-) = k, Z = z\}}{\Delta t}$$

is the intensity rate of $k \to k + 1$ transition for an individual with covariate vector z. We

let $\Lambda_k(t|z) = \int_0^t \lambda_k(u|z)du$ denote the cumulative transition intensity, and let

$$\mathbb{A}(u|z) = \begin{bmatrix} -\Lambda_1(u|z) & \Lambda_1(u|z) & 0 & \dots & \dots & 0 \\ 0 & -\Lambda_2(u|z) & \Lambda_2(u|z) & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & -\Lambda_K(u|z) & \Lambda_K(u|z) \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

be the $(K+1)\times(K+1)$ cumulative transition intensity matrix. The transition probability matrix $\mathbb{P}(s,t|z)$ has $[k,\ell]$ entry

$$P_{k\ell}(s,t|z) = P(\zeta(t) = \ell|\zeta(s) = k, Z = z)\ ,\ \ell = k,\dots,K+1\ , s < t\ ,$$

and is obtained by product integration via

$$\mathbb{P}(s,t|z) = \prod_{u\in(s,t]} [\mathbb{I} + d\mathbb{A}(u|z)]\ , \tag{1.1.1}$$

where $\mathbb{I}$ is an identity matrix of size $K+1$ (Andersen *and others*, 1993).

We primarily consider Markov regression models with multiplicative effects of covariates on intensities of the form

$$\lambda_k(t|z;\alpha_k,\beta_k) = \lambda_k(t;\alpha_k)\exp(z'\beta_k)\ , \tag{1.1.2}$$

where $\lambda_k(t;\alpha_k)$ is a baseline intensity function indexed by a parameter vector α_k, and β_k is a $p \times 1$ vector of regression coefficients. When we conduct semiparametric analysis, we assume that $\lambda_k(t)$ does not bear any particular parametric form. The formulation (1.1.2) yields relative risks which are easy to interpret, and model diagnostics can be adopted from survival analysis for use here (Therneau and Grambsch, 2000). Asymptotic theory based on counting processes can be utilized to determine large sample properties of estimators (Fleming and Harrington, 1991; Andersen *and others*, 1993).

1.1.3 Joint Modeling Event Times via Copulas

If U_1 and U_2 are two uniformly-distributed random variables on $[0,1]$, the bivariate distribution

$$\mathcal{C}(u_1,u_2) = P(U_1 \leq u_1, U_2 \leq u_2) \ ,$$

corresponds to a copula function (Nelsen, 2006). For two continuous random variables V_1 and V_2 with marginal survivor functions $\mathcal{F}_j(\cdot)$, $j=1,2$, a joint survivor function $\mathcal{F}(v_1,v_2) = P(V_1 \geq v_1, V_2 \geq v_2)$ is constructed through a copula (Genest and MacKay, 1986) by defining

$$\mathcal{F}(v_1,v_2) = \mathcal{C}(\mathcal{F}_1(v_1),\mathcal{F}_2(v_2)) \ ;$$

Sklar's theorem (Sklar, 1959) ensures the existence and uniqueness of $\mathcal{C} : [0,1]^2 \to [0,1]$. Suppose $\mathcal{F}_j(v_j|z)$ is the survivor function of V_j given $Z=z$, $j=1,2$. Patton (2006) proved that for each $Z=z$ in the support of Z, the joint survivor function of V_1 and V_2 given

$Z = z$, $\mathcal{F}(v_1, v_2|z) = P(V_1 \geq v_1, V_2 \geq v_2|Z = z)$, is uniquely defined by

$$\mathcal{F}(v_1, v_2|z) = \mathcal{C}(\mathcal{F}_1(v_1|z), \mathcal{F}_2(v_2|z)|z) \ , \tag{1.1.3}$$

for all $(v_1, v_2) \in \mathbb{R}^2$.

If there exists a continuous, strictly decreasing and convex function $\Psi(u; \phi)$ such that $\Psi : [0, 1] \times \Phi \to [0, \infty)$ and $\Psi(1; \phi) = 0$, and if the copula function can be written as

$$\mathcal{C}(u_1, u_2; \phi) = \Psi^{[-1]}(\Psi(u_1; \phi) + \Psi(u_2; \phi); \phi) \ , \tag{1.1.4}$$

then the copula belongs to the Archimedean family. The vector of parameters ϕ has parameter space Φ. The function $\Psi(u; \phi)$ is called the generator (Nelsen, 2006) of the copula and $\Psi^{[-1]}$ is defined by

$$\Psi^{[-1]}(u; \phi) = \begin{cases} \Psi^{-1}(u; \phi) & \text{if } 0 \leq u \leq \Psi(0; \phi) \ , \\ 0 & \text{if } \Psi(0; \phi) \leq u \leq \infty \ . \end{cases}$$

If $(U_{i1}, U_{i2})'$ and $(U_{j1}, U_{j2})'$ are random variables drawn from (1.1.4), Kendall's τ is

$$\begin{aligned} \tau &= P\left[(U_{i1} - U_{j1})(U_{i2} - U_{j2}) > 0; \phi\right] - P\left[(U_{i1} - U_{j1})(U_{i2} - U_{j2}) < 0; \phi\right] \\ &= 1 + 4 \int_0^1 \frac{\Psi(u; \phi)}{\Psi'(u; \phi)} du \ . \end{aligned}$$

The Clayton copula is a widely-adopted member of the Archimedean family with

$\Psi(u;\phi) = u^{-\phi} - 1$ and the resulting form is

$$\mathcal{C}(u_1, u_2; \phi) = \left(u_1^{-\phi} + u_2^{-\phi} - 1\right)^{-1/\phi} , \qquad \phi \geq 0 . \tag{1.1.5}$$

For this copula, Kendall's τ is given by $\tau = \phi/(2+\phi)$.

1.2 Motivating Studies in Health Research

Here we briefly introduce several studies motivating the developments in this thesis.

1.2.1 Platelet Transfusions in Thrombocytopenic Patients

Mirasol is a pathogen inactivation technology which utilizes exposure to ultraviolet light to inhibit the proliferation of pathogens and inactivate white blood cell replication to produce pathogen-reduced platelets (PRT-PLT). The Cazenave *and others* (2010) report on a recent multicentre trial of 118 hematology/oncology patients with chemotherapy-induced thrombocytopenia who were randomized to receive either PRT-PLT or standard platelets over a 28-day treatment period. The primary outcome in this trial was the corrected count increment (CCI), which is the difference between the patients platelet count before and after the transfusion, adjusted for dose of platelets and body surface area of the patient (Davis *and others*, 1999). Patients received prophylatic transfusions whenever their platelet counts dropped below the defined threshold of $10 \times 10^9\ m^2/\ell$. As a result, one might expect an association between the CCI for a particular transfusion and the time to the next transfusion. The primary analysis is based on comparing the

probability of successful response between treatment arms under the implicit assumption that this probability does not change over time within treatment groups.

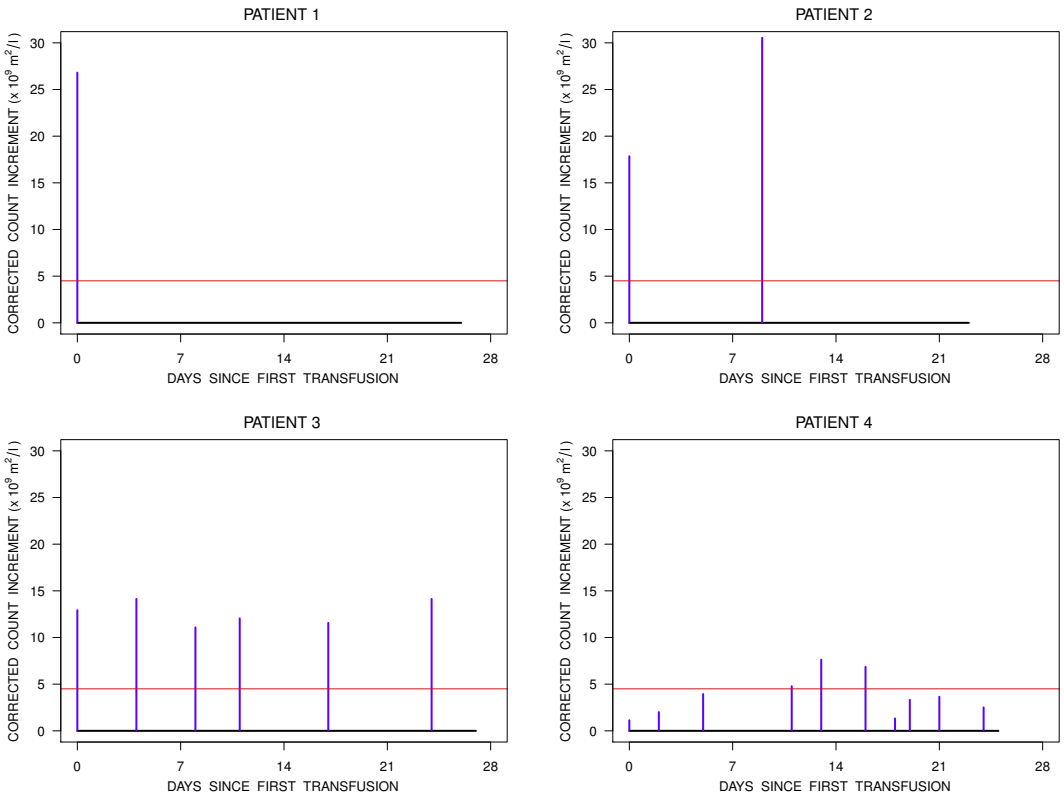

Figure 1.1: Profiles of four patients from the Mirasol study (Cazenave *and others*, 2010) showing the times of the transfusions, the corrected count increments and the duration of follow-up; the horizontal line at 4.5 $\times 10^9$ m^2/ℓ is the threshold for a successful transfusion

Figure 1.1 illustrates data for a sample of four patients from the platelet transfusion study. In each panel, the horizontal axis indicates the number of days since the first transfusion and the vertical axis is the corrected count increment based on the measurement 24 hours after the transfusion. The horizontal dashed line at 4.5 $\times 10^9$ m^2/ℓ corresponds to the threshold used to define a successful (CCI $> 4.5 \times 10^9$ m^2/ℓ) or unsuccessful (CCI ≤ 4.5

$\times 10^9$ m^2/ℓ) transfusion. Patient 1 required only one transfusion during the 28-day treatment period and experienced a large increase in their platelet count due to this transfusion. Patient 4 on the other hand, required 9 transfusions with only three of these leading to successful transfusions. Patients 2 and 3 had intermediate numbers of transfusions which generally led to successful responses.

The average duration of follow-up was 23.2 and 23.7 days for patients randomized to receive pathogen-inactived or standard platelets respectively with the mean number of platelet transfusions of 3.7 and 3.2. The average corrected count increment was 6.2 and 8.9 ($\times 10^9$ m^2/ℓ) respectively.

1.2.2 Joint Damage in Psoriatic Arthritis

Psoriatic arthritis (PsA) is an immunological disease associated with considerable joint pain, inflammation and destruction which can ultimately lead to serious disability and poor quality of life (Chandran *and others*, 2010). The Centre for Prognosis Studies in Rheumatic Disease is a tertiary referral center at the Toronto Western Hospital which treats patients with a variety of rheumatic diseases and maintains a clinic registry of patients with psoriatic arthritis. This cohort was established in 1976 and the patient have been recruited and followed since its formation; today it is one of the largest cohorts of patients with PsA in the world.

Upon entry to the clinic, the patients undergo a detailed clinical and radiological examination and provide serum samples which are subsequently stored. Follow-up clinical and radiological assessments are scheduled annually and biannually respectively in order

to track changes in joint damage, functional ability, and quality of life. Additional serum samples are taken at follow-up clinic visits to measure dynamic markers of inflammation and to store for future analysis of genetic data. To date, 1191 patients have been recruited and there is a median of 4.84 years of follow-up and a median of 6 clinical follow-up assessments. Given this data, disease progression can be modeled in a number of ways including the development of newly damaged joints (Gladman *and others*, 1995; Sutradhar and Cook, 2008), the involvement of particular types (e.g. spinal) of joints (Tolusso and Cook, 2009), or progression to a state of clinically important damage (Chandran *and others*, 2012).

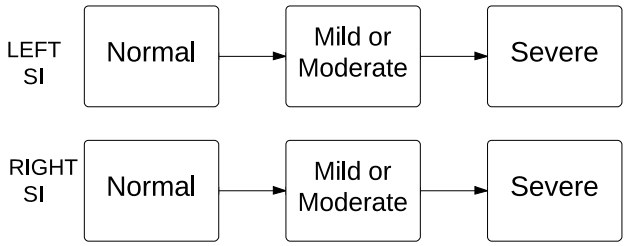

Figure 1.2: Multistate diagram for modeling damage in sacroiliac joints

We focus on the state of damage of the left and right sacroiliac (SI) joints since damage in these joints signifies the onset of spondyloarthritis, a condition associated with considerable disability. The modified Steinbrocker scale (Steinbrocker *and others*, 1949; Rahman *and others*, 1998) is a five-point scale used to formally record the extent of damage based on radiographic examination. The states are numbered $1-5$ with labels 1 = normal; 2 = equivocal; 3 = abnormal with erosions or sclerosis; 4 = unequivocally abnormal, moderate or advanced sacroiliitis showing one or more of erosions, sclerosis, widening, nar-

rowing or partial ankylosis; 5 = total ankylosis. In our analysis, we combine states 2 and 3 to form a state representing mild or moderate joint damage, and states 4 and 5 as a state denoting severe damage as illustrated in Figure 1.2.

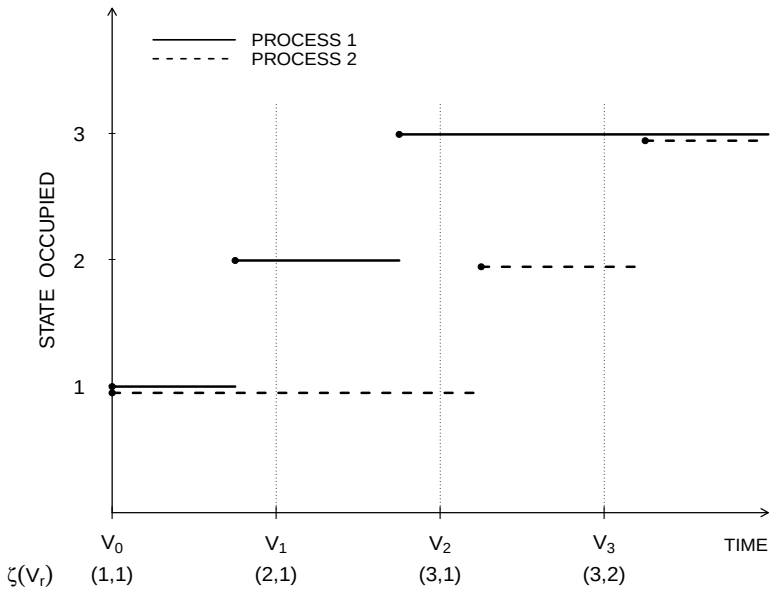

Figure 1.3: Multivariate multistate processes observed at intermittent inspection times

Figure 1.3 illustrates the process of disease development for one subject observed at a sequence of inspection times. The both processes of this individual, representing the left and right sacroiliac joints here, occupied state 1 at time V_0 meaning that the sacroiliac joints on his/her both sides of this subject are normal at the beginning of follow-up. At the first follow-up assessment at V_1, process 1 was in state 2 but process 2 was recorded to still be in state 1, which suggests that this subject had mild or moderate joint disease on the left side but normal joint on the right side at time V_1. At V_2, process 1 was in state 3 but process 2 remained in state 1. At the last assessment (at V_3), process 1 was in 3 (the

absorbing state) and process 2 was in state 2, which means the sacroiliac joint disease of this subject had developed into a severe state on the left side and a mild or moderate state on the right side at time V_3.

1.2.3 Skeletal Complications from Bone Metastases

Patients with skeletal metastases are at increased risk of pathological fractures and bone pain. Bisphosphonates are a class of compounds that are used to reduce the occurrence of these skeletal complications and hence improve quality of life. Hortobagyi *and others* (1996) report on an international multicenter trial of 382 women with stage IV breast cancer with at least one predominantly lytic bone lesion greater than or equal to one centimeter in diameter. Patients were accrued between January 1991 and March 1994 from 97 sites in the United States, Canada, Australia and New Zealand and randomized to receive pamidronate ($m = 185$) or a placebo control ($m = 197$). Two patients randomized to placebo did not have bone metastases and were therefore excluded from subsequent analyses. Patients randomized to the pamidronate arm received 90 mg of pamidronate disodium via a two-hour infusion every four weeks whereas patients randomized to the placebo received dextrose infusions. Patients were monitored closely and the occurrence of pathologic fractures and need for radiotherapy for the treatment of bone pain were recorded. After completion of the planned one-year follow-up, the observation was extended for an additional year and the results were published in Hortobagyi *and others* (1998). Each patient was followed until death, loss to follow-up, or February 1, 1996.

To avoid the complications arising from the terminal event of death, we restrict atten-

tion to patients who survived 12 months; note that patients who withdrew from the study early were followed for survival, so these patients were still censored over the 12-month period for the occurrence of these skeletal complications. Interest here lies in assessing the effect of pamidronate on the occurrence of these two types of skeletal complications and modeling the two types of skeletal complications jointly. We let $Z_i = 1$ for pamidronate-treated patients and $Z_i = 0$ for control patients.

1.2.4 Seroconversion Following Anticoagulation Therapy

Patients undergoing orthopedic surgery are at increased risk of developing thrombosis which is associated with increased morbidity and mortality (White *and others*, 1998). Prophylaxis with antithrombotic heparin-based therapies is highly effective at reducing the risk of thrombosis and is now standard practice in orthopedic surgery. Four large recent multicenter randomized trials were conducted to compare enoxaparin and fondaparinux for thromboprophylaxis. Two were set in Europe (Lassen *and others*, 2002; Eriksson *and others*, 2001) and two in North America (Bauer *and others*, 2001; Turpie *and others*, 2002). Some patients undergoing orthopedic surgery and exposed to antithrombotic drugs developed antibody responses and it is of interest to understand the correlation of the different types of antibody responses. The first injection of enoxaparin was given before surgery in 1011 (32.1%) of the patients and after surgery for the remainder.

We let the time of surgery be the time origin. Antibody formation can begin any time following surgery and we let T_i denote the time from surgery to the antibody response for individual i, and $\mathcal{F}(t) = P(T_i \geq t)$ the corresponding survivor function. The time from

surgery to recovery and discharge varied considerably from patient to patient, but upon discharge a blood sample was taken and the antibody status was determined.

1.3 Outline of Thesis

An innovative model for marked point processes is given in Chapter 2, which is motivated by problems arising in transfusion medicine where interest lies in the assessment of the effect of the platelet products on the corrected count increment, the mark, following a platelet transfusion, the recurrent event. A copula model is used to link the mark with the time to the next event. This formulation ensures that standard multivariate models can be formulated to link the marks over time and that any point process model can be used to model the events. The biases and relative efficiencies of separate versus joint analyses of the marks and events are assessed for a variety of estimation strategies including simultaneous estimation procedure, two-stage estimation procedure, asymmetric two-stage estimation procedure and three-stage estimation procedure. Robustness to misspecification of the model for recurrent events or the copula function is also explored. An application to the data from the motivating study in Section 1.2.1 is given to illustrate the model and methods.

In Chapter 3 we develop models for the joint analysis of multivariate multistate processes. Such models are useful for progressive disease processes which arise in clusters of individuals who may share observed, or more importantly unobserved, genetic or environmental exposures. They also appeal when a disease affects more than one feature in an individual but when the progression of each feature can be modeled in a similar

way. Attention is restricted primarily to Markov models, which are particularly attractive when dealing with interval-censored transition times arising from intermittent inspection of individuals. Estimation and statistical inference are based on composite likelihood with simultaneous or two-stage estimation procedures. Empirical biases and relative efficiencies are assessed and the proposed methods are found to perform well. An application to the motivating study of joint damage in patients with psoriatic arthritis in Section 1.2.2 is provided.

In Chapter 4 attention is directed to use of copula models for more general scenarios with a focus on semiparametric two-stage estimation procedures. In this approach, nonparametric or semiparametric estimates of the marginal survivor function are obtained in the first stage and estimates of the association parameters are obtained in the second stage. Bivariate failure time models are considered for data under right-censoring and current status observation schemes, and right-censored multistate models. Asymptotic variances are derived under these models and observation schemes and assessed empirically to compare with formula based on parametric ways of viewing nonparametric and semiparametric methods. Applications to motivating examples in Section 1.2.3 and 1.2.4 are given.

This thesis concludes with Chapter 5 where some avenues for further research are outlined, of which some topics are currently under investigation.

Chapter 2

A Copula Model for Marked Point Prcessess

2.1 Introduction

2.1.1 Overview

Many disease processes feature recurrent events which represent acute exacerbations of an underlying chronic condition. Examples include respiratory attacks in patients with asthma which can be associated with considerable disability and increased risk of death (Verona *and others*, 2003), flares of symptoms in patients with systemic lupus erythematosus (Petri *and others*, 1991; Fok *and others*, 2012), recurrent headaches among migraineurs (Pascual *and others*, 2000), and graft rejection episodes arising in transplant recipients (Cole *and others*, 1994).

Statistical methods for the analysis of recurrent events have seen considerable development in the last three decades. The three primary classes of methods are based on intensity-based models (Andersen *and others*, 1993; Aalen *and others*, 2008), random effect models (Lawless, 1987) and marginal models (Lawless and Nadeau, 1995). In the clinical trial arena, marginal methods based on rate functions (Andersen and Gill, 1982) have considerable appeal, and the development of methods for robust inference (Lawless and Nadeau, 1995) has led to their widespread use. Partially conditional models (Prentice *and others*, 1981) and marginal methods based on multivariate failure time data (Wei *and others*, 1989) are also used routinely.

The events in many conditions are severe enough to warrant therapeutic intervention for alleviation of symptoms and mitigation of risk for more serious complications. When such interventions are applied, there is typically a short-term response which reflects how effective the intervention was in alleviating symptoms and improving health. Studies of short-acting β_2-agonists for the treatment of asthma attacks (Sears *and others*, 1990), for example, aim to quickly improve lung function as measured by short-term change in forced expiratory volume.

The data resulting from such processes feature event times, with each event having an accompanying attribute realized upon event occurrence. Marked point processes are suitable for modeling such data, and have been used extensively in areas such as seismology (Holden *and others*, 2002), genetics (Robin, 2002), image analysis (Descombes and Zerubia, 2002), insurance (Grandell, 1997; Ch. 9), finance (Prigent, 2001), forestry (Penttinen *and others*, 1992), and management science (Chen and Zheng, 1997). The theory of marked point processes is given in several excellent books on stochastic processes including Cox

and Isham (1980) (Ch. 5), Snyder and Miller (1991) (Ch. 4), Karr (1991) (Sec. 1.4), and more recent Daley and Vere-Jones (2008) (Sec. 13.4). Semiparametric methods of analysis based on likelihood are discussed in Andersen *and others* (1993) (Sec. 2.4) and robust nonparametric marginal methods are considered in Cook *and others* (2003) and Cook and Lawless (2007). Goulard *and others* (1996) consider pseudo-likelihood methods and moment estimation is developed by Politis and Sherman (2001). In the existing literature, the marks are often assumed to be independent of the event times to facilitate their separate modeling using simple methods, but this assumption is often questionable. Tests of the independence between the marks and event times are developed by Schlather *and others* (2004), Schoenberg (2004), and Guan (2006).

We propose a novel model for a marked point process in which the marginal models for the events and the marks are compatible with standard models for recurrent event and longitudinal analyses. A copula function is used to link the "survival" function of each mark given the history of marks, with that of the next event time given the relevant event history. This appealing structure means that analyses of the incidence of the events and analysis of the marks are compatible with standard methods for recurrent event data and any continuous multivariate distribution; efficiency gains can therefore be explored for joint versus separate analyses.

The remainder of this chapter is organized as follows. In Section 2.2 we define notation, describe the formulation of a copula-based marked point process model, construct the likelihood and discuss multiple estimation procedures. Simulation studies and an illustration of the Mirasol platelet transfusion trial are presented in Section 2.3 and Section 2.4 respectively, and general remarks are given in Section 2.5.

2.2 Model Formulation and Estimation Procedures

2.2.1 A Copula-Based Marked Point Process Model

Recall that in Section 1.1.1 we defined $Z = 1$ if the individual is randomized to receive the experimental intervention and $Z = 0$ otherwise. We consider studies in which the primary purpose is to compare the effect of the experimental intervention to the standard intervention with respect to the distribution of the marks, and secondary interest lies in the effect of the randomized intervention on event occurrence.

We suppose the process begins with an initiating event at $T_0 = 0$, let T_k be the time of the kth event recurrence, and let $W_k = T_k - T_{k-1}$ denote the waiting time between the $(k-1)$st and kth event, $k = 1, 2, \ldots$. We let $N(t) = \sum_{k=1}^{\infty} I(T_k \leq t)$. The associated right-continuous counting process is $\{N(t), 0 < t\}$ and the associated history is $\mathcal{H}^N(t) = \{N(u), 0 < u < t\}$. Let $\Delta N(t) = N(t+\Delta t) - N(t^-)$ be the number of events over $[t, t+\Delta t]$, $dN(t) = \lim_{\Delta t \downarrow 0} \Delta N(t) = 1$ if an event occurs at time t, and $dN(t) = 0$ otherwise. The mark associated with the kth event is denoted by the random variable Y_k, $k = 0, 1, \ldots$, and we let $Y(t) = \{Y_0, \ldots, Y_{N(t)}\}$ denote the set of marks realized over $[0, t]$. It is convenient to denote the history of the marks at t as $\mathcal{H}^Y(t) = \{Y(u), 0 < u < t\}$, and the full history is then $\mathcal{H}(t) = \{N(u), Y(u), 0 < u < t, Z\}$. Let $[0, C_A]$ denote the planned period of observation, where $C_A > 0$ is an administrative right censoring time. Suppose further that $C_R > 0$ is a random right-censoring time (Kalbfleisch and Prentice, 2002; Lawless, 2003) giving a net duration of observation $C = \min(C_A, C_R)$. Figure 2.1 gives a schematic diagram relating key variables.

We let $C(t) = I(t \leq C)$ indicate whether the process is under observation at time t. Then if $d\bar{N}(t) = C(t)dN(t)$, $d\bar{N}(t) = 1$ implies that an event occurs and is observed at t. We define $\bar{N}(t) = \int_0^t d\bar{N}(u)$ and let $\bar{Y}(t) = \{Y_0, \ldots, Y_{\bar{N}(t)}\}$ denote the number of events and the respective marks observed over $[0, t]$. The corresponding histories at time t are then $H^N(t) = \{\bar{N}(u), C(u), 0 < u < t\}$ and $H^Y(t) = \{\bar{Y}(u), C(u), 0 < u < t\}$ and the full history is $H(t) = \{\bar{N}(u), \bar{Y}(u), C(u), 0 < u < t, Z\}$.

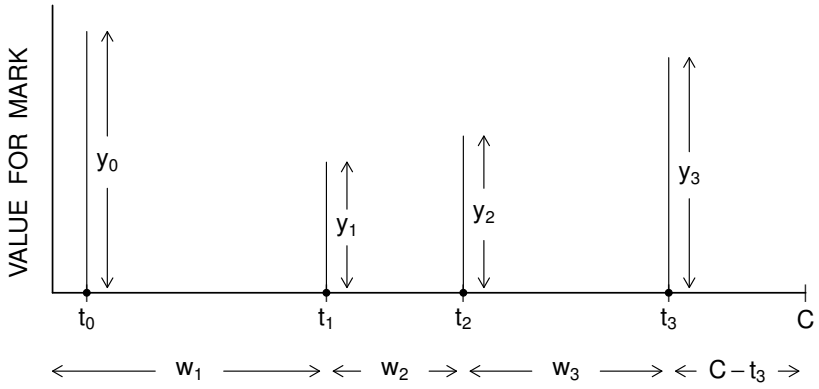

Figure 2.1: Schematic diagram with notation for a marked point process

The full likelihood for a randomly-censored marked point process over $[0, C_A]$ can be expressed in product integral notation (Cook and Lawless, 2007; Ch. 8) as

$$\prod_{u \in [0, C_A]} P(C(u)|H(u)) \Big[P(d\bar{N}(u)|H(u), C(u) = 1) P(Y_{\bar{N}(u)}|H(u), C(u) = 1, d\bar{N}(u) = 1)^{d\bar{N}(u)} \Big]^{C(u)},$$

where we informally let the conditional probability involving $Y_{\bar{N}(t)}$ represent a conditional density when the marks are continuous. Under non-informative censoring we omit the term

$P(C(u)|H(u))$ and work with the partial likelihood

$$\prod_{u\in[0,C_A]}\left[P(d\bar{N}(u)|H(u),C(u)=1)P(Y_{\bar{N}(u)}|H(u),C(u)=1,d\bar{N}(u)=1)^{d\bar{N}(u)}\right]^{C(u)} . \tag{2.2.1}$$

If censoring is conditionally independent (Lawless, 2003), (2.2.1) is equal to

$$\prod_{u\in[0,C_A]} P(dN(u)|\mathcal{H}(u))^{C(u)} \prod_{u\in[0,C_A]} P(Y_{N(u)}|\mathcal{H}(u),dN(u)=1)^{d\bar{N}(u)} , \tag{2.2.2}$$

and we can express the partial likelihood in terms of the model of interest.

We consider the case where the marks are continuous random variables and for convenience let $\mathcal{H}_k^N = \{T_1,\ldots,T_{k-1}\}$, $\mathcal{H}_k^Y = \{Y_0,\ldots,Y_{k-1}\}$ and $\mathcal{H}_k = \{T_1,\ldots,T_{k-1},Y_0,\ldots,Y_{k-1},Z\}$. If we let $\bar{N}(C)=n$, note that

$$\prod_{u\in[0,C_A]} P(dN(u)|\mathcal{H}(u))^{C(u)} = \prod_{k=1}^{n} f(t_k|\mathcal{H}_k)\cdot P(T_{n+1}>C|\mathcal{H}_{n+1}) . \tag{2.2.3}$$

We formulate the joint conditional "survival" distribution $\mathcal{F}(t_k,y_{k-1}|\mathcal{H}_k^N,\mathcal{H}_{k-1}^Y,Z)$ through a copula $\mathcal{C}$ as

$$\mathcal{C}(\mathcal{F}(t_k|\mathcal{H}_k^N,\mathcal{H}_{k-1}^Y,Z),\mathcal{F}(y_{k-1}|\mathcal{H}_k^N,\mathcal{H}_{k-1}^Y,Z);\phi) , \tag{2.2.4}$$

where ϕ is a vector of association parameters. The structure of the motivating problem in transfusion medicine motivates consideration of models which accommodate a dependence between the mark Y_{k-1} and the time to the kth transfusion. We also aim to specify a model

such that the event intensity is formulated so that a marginal analysis of the recurrent event process yields parameters compatible with a standard recurrent event model. Copula models can be used for this purpose (Joe, 1997; Nelsen, 2006). That is, larger realized values of Y_{k-1} are expected and empirically observed to be associated with larger gap times $W_k = T_k - T_{k-1}$. To retain the feature that the marginal models for the event times and marks have a standard form, we make the following two assumptions:

A.2.1 $\quad T_k \perp \mathcal{H}_{k-1}^Y | \mathcal{H}_k^N, Z,\ k = 1, 2, \ldots$

A.2.2 $\quad Y_{k-1} \perp \mathcal{H}_k^N | \mathcal{H}_{k-1}^Y, Z,\ k = 2, 3, \ldots$

Assumption A.2.1 means that given the history of the event times and the fixed covariate the marks at times prior to T_{k-1} are not associated with $W_k = T_k - T_{k-1}$. Assumption A.2.2 states that the mark at T_{k-1} is independent of the event times prior to T_k given the previous marks and the fixed covariate. This means that any models for longitudinal data can be used for the joint distribution of the marks. Under assumptions A.2.1 and A.2.2, (2.2.4) simplifies to involve $\mathcal{F}(t_k|\mathcal{H}_k^N, Z)$ (under A.2.1) and $\mathcal{F}(y_{k-1}|\mathcal{H}_{k-1}^Y, Z)$, the "survivor function" of the mark given the history of the marks (under A.2.2). By applying the copula constructed in (2.2.4) with assumptions A.2.1 and A.2.2 to $\prod_{k=1}^n f(t_k|\mathcal{H}_k)$ in (2.2.3), it becomes

$$\prod_{k=1}^{n} \left[f(t_k|\mathcal{H}_k^N, Z) \cdot c\left(\mathcal{F}(t_k|\mathcal{H}_k^N, Z), \mathcal{F}(y_{k-1}|\mathcal{H}_{k-1}^Y, Z)\right) \right],$$

where $c(u, v) = \partial^2 \mathcal{C}(u, v)/\partial u \partial v$, the density function of the copula $\mathcal{C}(u, v)$ in (2.2.4).

The term $P(Y_{N(u)}|\mathcal{H}(u), dN(u) = 1)$ in the second part of the product integrand in

(2.2.2) is the density of the mark at time u conditional on the full history $\mathcal{H}(u)$ and the fact that an event occurred at time u. It can be written in terms of event times and marks as $f(y_k|\mathcal{H}^N_{k+1}, \mathcal{H}^Y_k, Z)$ and by assumption A.2.2 this becomes $f(y_k|\mathcal{H}^Y_k, Z)$.

If n events are observed for one individual over $(0, C]$, the likelihood of the observed outcome "n events occur at times $t_1 < \cdots < t_n$ with respective marks $y_0, y_1, \ldots, y_n$ given covariate z" is proportional to

$$\begin{aligned} & \Big[\mathcal{C}^{(01)}\left(P(T_{n+1} > C|\mathcal{H}^N_{n+1}, Z), P(Y_n > y_n|\mathcal{H}^Y_n, Z)\right) \\ \cdot & \prod_{k=1}^{n} c\left(P(T_k > t_k|\mathcal{H}^N_k, Z), P(Y_{k-1} > y_{k-1}|\mathcal{H}^Y_{k-1}, Z)\right)\Big] \\ & \cdot \left[\prod_{k=1}^{n} P(T_k \in [t_k, t_k + dt_k)|\mathcal{H}^N_k, Z)\right] \cdot \left[\prod_{k=0}^{n} P(Y_k \in [y_k, y_k + dy_k)|\mathcal{H}^Y_k, Z)\right], \end{aligned} \tag{2.2.5}$$

where $\mathcal{C}^{(01)}(u, v) = \partial\mathcal{C}(u, v)/\partial v$. The likelihood (2.2.5) is in a very amenable form. The first $n+1$ components provide information about all parameters, the second n components provide information about the marginal model for the recurrent event, and the last $n + 1$ components relate to the marginal model for the marks.

The formulation of this model through the use of copula functions enables specification of the marginal and association models separately. Particular models for the recurrent events are mentioned in Section 2.2.2 and those for the marks in Section 2.2.3.

2.2.2 Marginal Models for the Recurrent Events

The "marginal" intensity function of the recurrent event process is defined as

$$\lambda(t|\mathcal{H}^N(t), Z) = \lim_{\Delta t \downarrow 0} \frac{P\left(\Delta N(t) = 1|\mathcal{H}^N(t), Z\right)}{\Delta t} ,$$

where we assume that two events cannot occur at the same time. We call this a marginal intensity since it does not incorporate the history of the marks and is concerned only with the event process. Specific intensity functions are indicated by making it explicit what aspects of the event history have bearing on the instantaneous risks of event occurrence.

Markov Models

One might adopt a Markov model, of which the marginal intensity function is of the form

$$\lambda(t|\mathcal{H}^N(t), Z) = \lambda_k(t|Z) ,$$

where $N(t^-) = k$, whereby the transition intensity depends on cumulative number of events over $(0, t]$; see Prentice *and others* (1981). The Poisson intensity (Lawless, 1987b) as a special case arises if we set $\lambda_k(t|Z) = \lambda(t|Z)$ for $k = 0, 1, \ldots$. Models with multiplicative covariate effects, of the form

$$\lambda(t|Z; \theta) = \lambda_0(t; \alpha) \exp(Z\beta) ,$$

are very common, where $\lambda_0(t;\alpha)$ is a baseline intensity (rate) function indexed by α, and β is a regression coefficient; $\theta = (\alpha, \beta)'$.

A Mixed Markov Model

To allow for extra-Poisson variation, one can consider a standard mixed Poisson model, in which conditional on a random effect $U > 0$ with $\mathrm{E}(U) = 1$ and $\mathrm{Var}(U) = \gamma$, the marginal intensity given U is of the form

$$\lambda(t|\mathcal{H}^N(t), Z, U) = \lim_{\Delta t \downarrow 0} \frac{P(\Delta N(t) = 1|Z, U)}{\Delta t} = U\lambda_0(t;\alpha)\exp(Z\beta) \ .$$

Under this mixed Poisson model, if we define $\mathrm{E}(N(t)|Z) = \mu(t|Z)$, then $\mathrm{Var}(N(t)|Z) = \mu(t|Z) + \mu(t|Z)^2\gamma$ and it is apparent that extra Poisson variation is accommodated (Lawless, 1987a). When U follows a gamma distribution, the marginal intensity function has a convenient closed-form

$$\left(\frac{1 + N(t^-)\gamma}{1 + \mu(t|Z)\gamma}\right)\lambda_0(t;\alpha)\exp(Z\beta) \ , \tag{2.2.6}$$

which is the intensity for a negative binomial process (Cook and Lawless, 2007).

A Semi-Markov Model

Semi-Markov models are useful when there is a renewal in the process upon event occurence. Such models feature intensities of the form (Andersen *and others*, 1993; Zhao and Hu, 2013)

$$\lambda(t|\mathcal{H}^N(t), Z) = h_k(B(t)|Z) \ ,$$

where $N(t^-) = k$ and $B(t) = t - T_{N(t^-)}$ is the time since the most recent event. A renewal process is obtained if $h_k(u) = h(u)$ for $k = 0, 1, \ldots$ in which case the waiting times are independent and identically distributed within subjects. Use of mixed semi-Markov models is likewise possible, where given a random effect U we may have

$$\lambda(t|\mathcal{H}^N(t), Z, U) = U h_k(B(t)|Z) \ .$$

2.2.3 Marginal Models for the Marks

Let $Y = (Y_0, Y_1, \ldots, Y_K)'$ denote the vector of $K + 1$ marks. We may assume, for example, that $Y|Z$ is a $(K + 1) \times 1$ multivariate normal random variable with mean $\mu(Z) = (\mu_0(Z), \ldots, \mu_K(Z))'$ and $(K+1)\times(K+1)$ covariance matrix Σ with diagonal events σ^2 and off-diagonal elements $\rho\sigma^2$. We may specify, for example, that $\mu_k(Z) = \eta_{0k} + \eta_{1k}Z$ and so the effect of treatment is to change the mean mark. Often one would set $\eta_{1k} = \eta_1$, $k = 1, 2, \ldots$, to obtain a parsimonious representation of the treatment effect, although tests of $H_0 : \eta_{1k} = \eta_1$ are often sensible. When marks are binary, analogous multivariate binary models may be adopted and often these would be most naturally formulated

with marginal specifications of treatment and other covariate effects, particularly for data arising in clinical trials. We do not consider the use of discrete marks in what follows.

2.2.4 Estimation and Statistical Inference

Let $L(\theta; D)$ denote the likelihood function (2.2.5), where D denotes the data comprised of the marks $Y_0, Y_1, \ldots, Y_n$ observed at time points $0, T_1, \ldots, T_n$, the right censoring time C, and the covariate Z; the vector $\theta = (\psi_1', \psi_2', \phi')'$ is the full vector of parameters where ψ_1 indexes the marginal recurrent event process, ψ_2 indexes the joint distribution of the marks, and ϕ characterizes the association between the marks and the event times. Suppose that there is a random sample of size m and the observed data for individual i are denoted by D_i so that all of the data in the sample are recorded in $D = (D_1', \ldots, D_m')'$. Estimation and statistical inference can be approached using one of the following four different estimation procedures.

The Simultaneous Estimation Procedure

The simultaneous estimation procedure is conducted by estimating the full vector of parameters θ simultaneously. Conditional on covariate Z, the maximum likelihood estimator (MLE) $\widehat{\theta}$ is the solution to the score equation $U(\theta) = \sum_{i=1}^{m} U_i(\theta) = 0$, where

$$U_i(\theta) = \frac{\partial \log L(\theta; D_i)}{\partial \theta} .$$

If the model is correctly specified, the MLE $\widehat{\theta}$ is consistent and $\sqrt{m}(\widehat{\theta}-\theta) \xrightarrow{d} N(0, \mathbb{A}_1^{-1}(\theta))$ as $m \to \infty$, where $\mathbb{A}_1(\theta) = -\mathrm{E}[\partial U_i(\theta)/\partial\theta']$. The estimator of the matrix $\mathbb{A}_1(\theta)$ is

$$\widehat{\mathbb{A}}_1(\widehat{\theta}) = -\frac{1}{m}\sum_{i=1}^{m} \left.\frac{\partial U_i(\theta)}{\partial\theta'}\right|_{\theta=\widehat{\theta}} . \tag{2.2.7}$$

The simultaneous estimation procedure produces the most efficient estimators for the full vector of parameters. However, misspecification of either the marginal models or the association model may lead to biased estimators.

The Two-stage Estimation Procedure

Instead of simultaneously estimating all the parameters in the full likelihood function (2.2.5), a two-stage estimation procedure can be adopted. Under this approach, the parameters for the marginal recurrent event process (ψ_1) and those for the marks (ψ_2) are estimated in the first stage; and the association parameters (ϕ) are estimated in the second stage by maximizing the full likelihood with respect to ϕ with the first-stage estimates plugged in. Specifically, the partial likelihood for the recurrent event process of one subject is

$$\begin{aligned} L(\psi_1; \mathcal{H}_{n+1}^N, C, Z) &= \prod_{k=1}^{n} f(t_k|\mathcal{H}_k^N, Z) \cdot P(T_{n+1} > C|\mathcal{H}_{n+1}^N, Z) \\ &= \prod_{k=1}^{n} \lambda(t_k|\mathcal{H}^N(t_k), Z) \exp\left[-\int_0^C \lambda(u|\mathcal{H}^N(u), Z)du\right] , \end{aligned}$$

and the partial likelihood for the marks is

$$L(\psi_2; \mathcal{H}_{n+1}^Y, Z) \;\; = \;\; \prod_{k=0}^{n} f(y_k|\mathcal{H}_k^Y, Z) = f(y_0, \cdots, y_n|Z) \; .$$

The stage-one estimators of ψ_1 and ψ_2, denoted by $\widetilde{\psi}_1$ and $\widetilde{\psi}_2$, are the solutions to the score equations

$$U_1(\psi_1) = \sum_{i=1}^{m} U_{i1}(\psi_1) = 0 \; , \tag{2.2.8}$$

and

$$U_2(\psi_2) = \sum_{i=1}^{m} U_{i2}(\psi_2) = 0 \; , \tag{2.2.9}$$

respectively, obtained from a sample of size m, where

$$U_{i1}(\psi_1) = \frac{\partial \log L(\psi_1; \mathcal{H}_{i,n+1}^N, C_i, Z_i)}{\partial \psi_1} \; ,$$

and

$$U_{i2}(\psi_2) = \frac{\partial \log L(\psi_2; \mathcal{H}_{i,n+1}^Y, Z_i)}{\partial \psi_2} \; .$$

The second-stage estimator of ϕ, denoted by $\widetilde{\phi}$, is the solution to the score equation

$$U_3(\phi; \widetilde{\psi}_1, \widetilde{\psi}_2) = \sum_{i=1}^{m} U_{i3}(\phi; \widetilde{\psi}_1, \widetilde{\psi}_2) = 0 \; , \tag{2.2.10}$$

where

$$U_{i3}(\phi; \widetilde{\psi}_1, \widetilde{\psi}_2) = \frac{\partial \log L(\widetilde{\psi}_1, \widetilde{\psi}_2, \phi; D_i)}{\partial \phi} \ .$$

If $U_{i1}(\psi_1)$, $U_{i2}(\psi_2)$ and $U_{i3}(\phi; \psi_1, \psi_2)$ are "stacked" to form

$$G_{i1}(\theta) = [U'_{i1}(\psi_1), U'_{i2}(\psi_2), U'_{i3}(\phi; \psi_1, \psi_2)]' \ ,$$

then equations (2.2.8), (2.2.9) and (2.2.10) are simply the three components of the joint estimating equation $\sum_{i=1}^{m} G_{i1}(\theta) = 0$. Since $G_{i1}(\theta)$ is unbiased, under certain regularity conditions (Godambe, 1960), $\widetilde{\theta} = (\widetilde{\psi}'_1, \widetilde{\psi}'_2, \widetilde{\phi}')'$ is consistent. The consistency of $\widetilde{\psi}_1$ is subject to correct model specification of the marginal model for the recurrent events, that of $\widetilde{\psi}_2$ is subject to correct specification of the marginal model for the marks, and that of $\widetilde{\phi}$ requires correct specification of the full model. We also have

$$\sqrt{m}(\widetilde{\theta} - \theta) \xrightarrow{d} N(0, \mathbb{A}_2^{-1}(\theta)\mathbb{B}_2(\theta) \left[\mathbb{A}_2^{-1}(\theta)\right]') \ ,$$

as $m \to \infty$, where $\mathbb{A}_2(\theta) = -\mathrm{E}[\partial G_{i1}(\theta)/\partial \theta']$ and is equal to

$$-\mathrm{E} \begin{pmatrix} \partial U_{i1}(\psi_1)/\partial \psi'_1 & 0 & 0 \\ 0 & \partial U_{i2}(\psi_2)/\partial \psi'_2 & 0 \\ \partial U_{i3}(\phi; \psi_1, \psi_2)/\partial \psi'_1 & \partial U_{i3}(\phi; \psi_1, \psi_2)/\partial \psi'_2 & \partial U_{i3}(\phi; \psi_1, \psi_2)/\partial \phi' \end{pmatrix} \ ,$$

$\mathbb{B}_2(\theta) = \mathrm{E}[G_{i1}(\theta)G'_{i1}(\theta)]$ and is equal to

$$\mathrm{E}\begin{pmatrix} U_{i1}(\psi_1)U'_{i1}(\psi_1) & U_{i1}(\psi_1)U'_{i2}(\psi_2) & U_{i1}(\psi_1)U'_{i3}(\phi;\psi_1,\psi_2) \\ U_{i2}(\psi_2)U'_{i1}(\psi_1) & U_{i2}(\psi_2)U'_{i2}(\psi_2) & U_{i2}(\psi_2)U'_{i3}(\phi;\psi_1,\psi_2) \\ U_{i3}(\phi;\psi_1,\psi_2)U'_{i1}(\psi_1) & U_{i3}(\phi;\psi_1,\psi_2)U'_{i2}(\psi_2) & U_{i3}(\phi;\psi_1,\psi_2)U'_{i3}(\phi;\psi_1,\psi_2) \end{pmatrix}.$$

The asymptotic covariance matrix is estimated by

$$\widehat{\mathbb{A}}_2^{-1}(\widetilde{\theta})\widehat{\mathbb{B}}_2(\widetilde{\theta})\left[\widehat{\mathbb{A}}_2^{-1}(\widetilde{\theta})\right]',$$

where

$$\widehat{\mathbb{A}}_2(\widetilde{\theta}) = -\frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} \partial U_{i1}(\psi_1)/\partial\psi'_1 & 0 & 0 \\ 0 & \partial U_{i2}(\psi_2)/\partial\psi'_2 & 0 \\ \partial U_{i3}(\phi;\psi_1,\psi_2)/\partial\psi'_1 & \partial U_{i3}(\phi;\psi_1,\psi_2)/\partial\psi'_2 & \partial U_{i3}(\phi;\psi_1,\psi_2)/\partial\phi' \end{pmatrix}\Bigg|_{\theta=\widetilde{\theta}},$$

and $\widehat{\mathbb{B}}_2(\widetilde{\theta})$ is equal to

$$\frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} U_{i1}(\psi_1)U'_{i1}(\psi_1) & U_{i1}(\psi_1)U'_{i2}(\psi_2) & U_{i1}(\psi_1)U'_{i3}(\phi;\psi_1,\psi_2) \\ U_{i2}(\psi_2)U'_{i1}(\psi_1) & U_{i2}(\psi_2)U'_{i2}(\psi_2) & U_{i2}(\psi_2)U'_{i3}(\phi;\psi_1,\psi_2) \\ U_{i3}(\phi;\psi_1,\psi_2)U'_{i1}(\psi_1) & U_{i3}(\phi;\psi_1,\psi_2)U'_{i2}(\psi_2) & U_{i3}(\phi;\psi_1,\psi_2)U'_{i3}(\phi;\psi_1,\psi_2) \end{pmatrix}\Bigg|_{\theta=\widetilde{\theta}}.$$

This usual two-stage estimation procedure provides some protection for the estimators for the marginal parameters from misspecification of the association model. However, the joint modelling does not enhance the efficiency of the marginal estimators since the first

stage estimation is exactly the same as marginal analyses under working independence assumptions.

Asymmetric Two-stage Estimation Procedure

Here we propose a novel two-stage estimation procedure. There is a tendency in the existing literature on two-stage procedures to treat each marginal distribution symmetrically, such that all of the marginal parameters are estimated in the first stage. Under the proposed asymmetric two-stage estimation procedure, only the marginal parameters for the recurrent event process are estimated in the first stage, and the parameters governing the marginal distribution of the marks and the association parameters are estimated in the second stage. So, $\widetilde{\psi}_1$ is estimated in the first stage as before, but in the second stage, $(\psi_2', \phi')'$ is estimated by solving the score equation $U_{23}(\psi_2, \phi; \widetilde{\psi}_1) = \sum_{i=1}^{m} U_{i23}(\psi_2, \phi; \widetilde{\psi}_1) = 0$, denoted by $(\bar{\psi}_2', \bar{\phi}')'$, where

$$U_{i23}(\psi_2, \phi; \widetilde{\psi}_1) = \frac{\partial \log L(\widetilde{\psi}_1, \psi_2, \phi; D_i)}{\partial (\psi_2', \phi')'} \ . \tag{2.2.11}$$

If $U_{i1}(\psi_1)$ and $U_{i23}(\psi_2, \phi; \psi_1)$ are "stacked" to form

$$G_{i2}(\theta) = [U_{i1}'(\psi_1), U_{i23}'(\psi_2, \phi; \psi_1)]' \ ,$$

then equations (2.2.8) and (2.2.11) are the two components of the joint estimating equation $\sum_{i=1}^{m} G_{i2}(\theta) = 0$. Since $G_{i2}(\theta)$ is unbiased, under the usual regularity conditions (Godambe, 1960), $\bar{\theta} = (\widetilde{\psi}_1', \bar{\psi}_2', \bar{\phi}')'$ is consistent. The consistency of $\widetilde{\psi}_1$ is subject to correct model

specification of the marginal model for the recurrent events and that of $(\bar{\psi}_2', \bar{\phi}')'$ requires correct specification of the full model. We also have

$$\sqrt{m}(\bar{\theta} - \theta) \xrightarrow{d} N(0, \mathbb{A}_3^{-1}(\theta)\mathbb{B}_3(\theta)\left[\mathbb{A}_3^{-1}(\theta)\right]') ,$$

as $m \to \infty$, where $\mathbb{A}_3(\theta) = -\mathrm{E}[\partial G_{i2}(\theta)/\partial\theta']$ and is equal to

$$-\mathrm{E}\begin{pmatrix} \partial U_{i1}(\psi_1)/\partial\psi_1' & 0 \\ \partial U_{i23}(\psi_2, \phi; \psi_1)/\partial\psi_1' & \partial U_{i23}(\psi_2, \phi; \psi_1)/\partial(\psi_2', \phi') \end{pmatrix} ,$$

$\mathbb{B}_3(\theta) = \mathrm{E}[G_{i2}(\theta)G_{i2}'(\theta)]$ and it is equal to

$$\mathrm{E}\begin{pmatrix} U_{i1}(\psi_1)U_{i1}'(\psi_1) & U_{i1}(\psi_1)U_{i23}'(\psi_2, \phi; \psi_1) \\ U_{i23}(\psi_2, \phi; \psi_1)U_{i1}'(\psi_1) & U_{i23}(\psi_2, \phi; \psi_1)U_{i23}'(\psi_2, \phi; \psi_1) \end{pmatrix} .$$

The asymptotic covariance matrix is estimated by

$$\widehat{\mathbb{A}}_3^{-1}(\bar{\theta})\widehat{\mathbb{B}}_3(\bar{\theta})\left[\widehat{\mathbb{A}}_3^{-1}(\bar{\theta})\right]' ,$$

where

$$\widehat{\mathbb{A}}_3(\bar{\theta}) = -\frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} \partial U_{i1}(\psi_1)/\partial\psi_1' & 0 \\ \partial U_{i23}(\psi_2, \phi; \psi_1)/\partial\psi_1' & \partial U_{i23}(\psi_2, \phi; \psi_1)/\partial(\psi_2', \phi') \end{pmatrix}\Bigg|_{\theta=\bar{\theta}} ,$$

and $\widehat{\mathbb{B}}_3(\bar{\theta})$ is equal to

$$\frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} U_{i1}(\psi_1)U'_{i1}(\psi_1) & U_{i1}(\psi_1)U'_{i23}(\psi_2,\phi;\psi_1) \\ U_{i23}(\psi_2,\phi;\psi_1)U'_{i1}(\psi_1) & U_{i23}(\psi_2,\phi;\psi_1)U'_{i23}(\psi_2,\phi;\psi_1) \end{pmatrix}\Bigg|_{\theta=\bar{\theta}} .$$

Asymmetric two-stage estimation procedure provides protection for the estimators of the parameters for the recurrent event process from misspecification of the model for the marks or the association model. The estimators for the parameters of the marks are more efficient than those using the usual two-stage estimation but less efficient than those using simultaneous estimation.

A Three-stage Estimation Procedure

The three-stage estimation procedure goes one-step further than the usual two-stage estimation procedure. Following the first two steps of the usual two-stage procedure, in the third stage, we plug the estimator for the association parameters $\widetilde{\phi}$ into the full likelihood (2.2.5) and re-estimate all the marginal parameters. In the third stage, $\breve{\psi} = (\breve{\psi}'_1, \breve{\psi}'_2)'$, the estimator of $\psi = (\psi'_1, \psi'_2)'$, is estimated by solving the score equation

$$U_{12}(\psi;\widetilde{\phi}) = \sum_{i=1}^{m} U_{i12}(\psi;\widetilde{\phi}) = 0 \ , \tag{2.2.12}$$

where

$$U_{i12}(\psi;\widetilde{\phi}) = \frac{\partial \log L(\psi,\widetilde{\phi};D_i)}{\partial\psi} \ .$$

The asymptotic properties of $\widetilde{\phi}$ is given in the section of the usual two-stage estimation procedure. Under regularity conditions (Godambe, 1960) and in the view of consistency of $\widetilde{\phi}$, $\breve{\psi}$ is consistent and

$$\sqrt{m}(\breve{\psi} - \psi) \xrightarrow{d} N\left(0, \mathbb{A}_4^{-1}(\psi)\mathbb{B}_4(\psi)\left[\mathbb{A}_4^{-1}(\psi)\right]'\right) ,$$

as $m \to \infty$, where $\mathbb{A}_4(\psi) = -\mathrm{E}[\partial U_{i12}(\psi;\phi)/\partial\psi']$ and

$$\mathbb{B}_4(\psi) \;=\; \mathbb{A}_4(\psi) + \mathbb{H}\mathrm{Cov}(\widetilde{\phi})\mathbb{H}'$$

and $\mathbb{H}$ is equal to

$$\mathrm{E}\left[\frac{\partial}{\partial\phi'}U_{i12}(\psi;\widetilde{\phi})\right] + \mathrm{E}\left[U_{i12}(\psi;\widetilde{\phi})U_{i3}(\phi;\psi)\right] .$$

The asymptotic covariance matrix is estimated by

$$\widehat{\mathbb{A}}_4^{-1}(\breve{\psi})\widehat{\mathbb{B}}_4(\breve{\psi})\left[\widehat{\mathbb{A}}_4^{-1}(\breve{\psi})\right]' ,$$

where

$$\widehat{\mathbb{A}}_4(\breve{\psi}) = -\frac{1}{m}\sum_{i=1}^{m}\left.\frac{\partial U_{i12}(\psi;\phi)}{\partial\psi'}\right|_{\psi=\breve{\psi}} ,$$

$\widehat{\mathbb{B}}_4(\breve{\psi}) = \widehat{\mathbb{A}}_4(\breve{\psi}) + \widehat{\mathbb{H}}\widehat{\mathrm{Cov}}(\widetilde{\phi})\widehat{\mathbb{H}}'$ and

$$\widehat{\mathbb{H}} = \frac{1}{m}\sum_{i=1}^{m}\left[\frac{\partial}{\partial\phi'}U_{i12}(\psi;\phi) + U_{i12}(\psi;\phi)U_{i3}(\phi;\psi)\right]\Bigg|_{\psi=\breve{\psi},\phi=\widetilde{\phi}}.$$

$\widehat{\mathrm{Cov}}(\widetilde{\phi})$ was given in the section introducing the usual two-stage estimation procedure.

The consistency of the estimators is subject to correct specification of the full model. The estimators using three-stage estimation are almost as efficient as those using simultaneous estimation. In our future research, we can actually conduct the iterative multistate estimation procedure. We may estimate ϕ and ψ iteratively by solving (2.2.10) and (2.2.12) with plugged-in estimators from the last iteration until the difference of estimates between two steps is smaller than some tolerance value. The resulting estimators should converge to the MLEs obtained from the simultaneous estimation procedure.

2.3 Simulation Studies

2.3.1 Empirical Performance under Correct Model Specification

Here we investigate the finite sample performance of the estimators based on the different estimation procedures mentioned in Section 2.2.4. We first consider events generated from a process where the marginal model for the recurrent events is a non-homogeneous Poisson process. The marginal rate function is of the form

$$\lambda(t|Z) = \alpha_1\alpha_2(\alpha_1 t)^{\alpha_2-1}\exp(Z\beta)$$

giving the corresponding marginal mean function

$$\mathrm{E}(N(t)|Z) = \int_0^t \lambda(u|Z)du = (\alpha_1 t)^{\alpha_2} \exp(Z\beta) \ .$$

We set the administrative censoring time C_A to be 1, $\alpha_2 = 0.75$ to induce mild trend, and set $\mathrm{E}(N(1)|Z = 0) = \alpha_1^{\alpha_2} = 4$ giving $\alpha_1 = 6.3496$. We set the coefficient $\beta = \log(0.5) = -0.6931$ to reflect a 50% reduction of the risk of events under the experimental treatment. We assume that the marks are multivariate normal with $\sigma^2 = 1$ and consider a moderate ($\rho = 0.3$) and strong correlation ($\rho = 0.6$) between marks under an exchangeable correlation structure. The means are assumed to be constant over time within individuals with $\eta_{0k} = \eta_0 = 0$ and $\eta_{1k} = \eta_1 = 0.5$, $k = 0, 1, \ldots$, corresponding to an experimental treatment giving a larger mean response. The events and the marks are linked through a Clayton copula (1.1.5). We consider scenarios with both moderate (Kendall's $\tau = 0.2$) and strong ($\tau = 0.6$) dependence between the marks and the recurrent event process leading to the association values $\phi = 2\tau/(1-\tau) = 0.5$ and 3 respectively (Nelsen, 2006). In addition, we suppose that $P(C_R \leq C_A) = 0.4$, i.e., 40% random right censoring, by assuming C_R to follow an exponential distribution with $P(C_R > C_A) = \exp(-\lambda_c) = 0.6$ and hence $\lambda_c = 0.5108$. A total of five hundred samples (nsim= 500) are simulated with $m = 200$ and 500 subjects per sample for each parameter configuration. For each parameter configuration, the empirical bias (BIAS), empirical standard error (ESE) and average asymptotic (large sample) standard error (ASE) are reported for each method of estimation. The ASE is the average of the 500 asymptotic sample standard errors, the ESE is the standard deviation of 500 parameter estimates, and the ECP is the proportion

of all trials for which the computed 95% confidence intervals contain the true respective parameter value.

The estimation methods include the simultaneous (ML), the two-stage (2S), the asymmetric two-stage (A2S) and the three-stage (3S) estimation procedures. The frequency properties of the estimates for the recurrent event process parameters are reported in Table 2.1 and those for the mark parameters and the association parameters are reported in Table 2.2 and 2.3 respectively. The empirical biases of all the estimators are negligible based on all the four estimation procedures under the correctly-specified model in Table 2.1, 2.2 and 2.3. Moreover, there is excellent agreement between the empirical standard errors and average model-based standard errors, indicating the validity of the four estimation methods. Estimators under three-stage estimation procedure are almost equally efficient as those under simultaneous estimation procedure. There is a marked decrease in the empirical and average model-based standard errors of all the estimators of the marginal parameters under the simultaneous estimation procedure compared to those under two-stage estimation procedure. The estimators of the parameters governing the distribution of the marks and the association parameters under the asymmetric two-stage procedure are more efficient than those under the usual two-stage estimation but less efficient than those under the simultaneous and three-stage procedure. The gain in efficiency is greatest when τ is largest; but there is little effect of the association between the marks on the relative efficiency.

In Diao *and others* (2013), comparisons of estimators from "joint analysis" (under simultaneous estimation procedure) and marginal analyses are made. Marginal analyses of the event times are conducted based on a parametric non-homogeneous Poisson model

Table 2.1: Frequency properties of parameter estimators for the recurrent event process through joint analysis of the marked point process with a Poisson margin under 40% random right censoring where $E(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; 500 simulated samples

			$\log(\alpha_1)$				$\log(\alpha_2)$				β			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S	ML	2S	A2S	3S
							Results based on 200 subjects/sample							
0.3	0.2	BIAS	0.0009	0.0001	0.0001	0.0012	0.0005	0.0008	0.0008	0.0007	-0.0060	-0.0050	-0.0050	-0.0061
		ESE	0.1022	0.1037	0.1037	0.1022	0.0420	0.0436	0.0436	0.0420	0.0965	0.0979	0.0979	0.0965
		ASE	0.1017	0.1027	0.1027	0.1008	0.0421	0.0429	0.0429	0.0417	0.0930	0.0956	0.0956	0.0924
	0.6	BIAS	0.0023	-0.0005	-0.0005	0.0095	-0.0003	0.0009	0.0009	-0.0004	-0.0033	-0.0031	-0.0031	-0.0037
		ESE	0.0882	0.1044	0.1044	0.0881	0.0314	0.0465	0.0465	0.0314	0.0748	0.0894	0.0894	0.0748
		ASE	0.0871	0.1030	0.1030	0.0876	0.0304	0.0429	0.0429	0.0304	0.0726	0.0957	0.0957	0.0727
0.6	0.2	BIAS	0.0006	0.0001	0.0001	0.0010	0.0005	0.0008	0.0008	0.0007	-0.0056	-0.0050	-0.0050	-0.0057
		ESE	0.1026	0.1037	0.1037	0.1026	0.0422	0.0436	0.0436	0.0422	0.0963	0.0979	0.0979	0.0963
		ASE	0.1021	0.1027	0.1027	0.1013	0.0422	0.0429	0.0429	0.0418	0.0927	0.0956	0.0956	0.0922
	0.6	BIAS	0.0011	-0.0005	-0.0005	0.0088	0.0001	0.0009	0.0009	-0.0004	-0.0016	-0.0031	-0.0031	-0.0016
		ESE	0.0925	0.1044	0.1044	0.0924	0.0339	0.0465	0.0465	0.0339	0.0725	0.0894	0.0894	0.0724
		ASE	0.0886	0.1030	0.1030	0.0898	0.0324	0.0429	0.0429	0.0325	0.0690	0.0957	0.0957	0.0692
							Results based on 500 subjects/sample							
0.3	0.2	BIAS	-0.0076	-0.0077	-0.0077	-0.0075	0.0021	0.0024	0.0024	0.0022	-0.0038	-0.0046	-0.0046	-0.0038
		ESE	0.0624	0.0621	0.0621	0.0623	0.0276	0.0286	0.0286	0.0276	0.0586	0.0595	0.0595	0.0586
		ASE	0.0645	0.0652	0.0652	0.0648	0.0267	0.0273	0.0273	0.0266	0.0590	0.0609	0.0609	0.0592
	0.6	BIAS	-0.0062	-0.0075	-0.0075	-0.0028	0.0006	0.0025	0.0025	0.0006	-0.0001	-0.0034	-0.0034	-0.0003
		ESE	0.0502	0.0621	0.0621	0.0502	0.0191	0.0279	0.0279	0.0191	0.0443	0.0616	0.0616	0.0443
		ASE	0.0552	0.0652	0.0652	0.0560	0.0193	0.0272	0.0272	0.0194	0.0460	0.0608	0.0608	0.0463
0.6	0.2	BIAS	-0.0078	-0.0077	-0.0077	-0.0076	0.0021	0.0023	0.0023	0.0022	-0.0035	-0.0046	-0.0046	-0.0035
		ESE	0.0625	0.0621	0.0621	0.0625	0.0275	0.0284	0.0284	0.0275	0.0586	0.0595	0.0595	0.0586
		ASE	0.0643	0.0652	0.0652	0.0643	0.0267	0.0273	0.0273	0.0266	0.0586	0.0609	0.0609	0.0586
	0.6	BIAS	-0.0070	-0.0075	-0.0075	-0.0036	0.0011	0.0025	0.0025	0.0009	0.0004	-0.0034	-0.0034	0.0003
		ESE	0.0516	0.0621	0.0621	0.0514	0.0200	0.0279	0.0279	0.0200	0.0428	0.0616	0.0616	0.0428
		ASE	0.0559	0.0652	0.0652	0.0561	0.0205	0.0272	0.0272	0.0205	0.0437	0.0608	0.0608	0.0437

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

Table 2.2: Frequency properties of mean parameter estimators of the marks through joint analysis of the marked point process with a Poisson margin under 40% random right censoring where $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; $\gamma=0.5$; 500 simulated samples

			η_0				η_1			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S
			Results based on 200 subjects/sample							
0.3	0.2	BIAS	-0.0015	-0.0024	-0.0014	-0.0019	0.0044	0.0049	0.0041	0.0045
		ESE	0.0685	0.0684	0.0685	0.0684	0.1003	0.1001	0.1000	0.1003
		ASE	0.0688	0.0702	0.0695	0.0701	0.1025	0.1048	0.1029	0.1032
	0.6	BIAS	-0.0002	-0.0031	0.0010	-0.0058	0.0023	0.0063	0.0014	0.0035
		ESE	0.0631	0.0674	0.0688	0.0631	0.0887	0.1000	0.0995	0.0887
		ASE	0.0627	0.0712	0.0716	0.0637	0.0865	0.1050	0.1043	0.0871
0.6	0.2	BIAS	-0.0010	-0.0028	-0.0010	-0.0016	0.0031	0.0043	0.0028	0.0032
		ESE	0.0825	0.0829	0.0826	0.0825	0.1179	0.1187	0.1177	0.1179
		ASE	0.0827	0.0842	0.0827	0.0843	0.1194	0.1226	0.1198	0.1209
	0.6	BIAS	0.0011	-0.0035	0.0019	-0.0052	-0.0006	0.0054	-0.0002	0.0004
		ESE	0.0740	0.0819	0.0797	0.0739	0.0993	0.1189	0.1113	0.0993
		ASE	0.0706	0.0866	0.0809	0.0724	0.0935	0.1245	0.1140	0.0948
			Results based on 500 subjects/sample							
0.3	0.2	BIAS	0.0002	-0.0002	0.0001	0.0001	0.0035	0.0037	0.0037	0.0035
		ESE	0.0419	0.0423	0.0419	0.0419	0.0637	0.0649	0.0640	0.0638
		ASE	0.0446	0.0445	0.0461	0.0461	0.0662	0.0670	0.0671	0.0676
	0.6	BIAS	0.0025	-0.0002	0.0020	-0.0001	0.0007	0.0043	0.0029	0.0012
		ESE	0.0369	0.0422	0.0416	0.0370	0.0511	0.0648	0.0641	0.0512
		ASE	0.0398	0.0469	0.0440	0.0405	0.0549	0.0686	0.0652	0.0556
0.6	0.2	BIAS	0.0011	0.0004	0.0010	0.0009	0.0025	0.0027	0.0028	0.0025
		ESE	0.0496	0.0504	0.0495	0.0496	0.0736	0.0757	0.0739	0.0736
		ASE	0.0513	0.0524	0.0528	0.0535	0.0752	0.0775	0.0767	0.0770
	0.6	BIAS	0.0036	0.0005	0.0030	0.0008	-0.0005	0.0034	0.0022	0.0000
		ESE	0.0418	0.0499	0.0474	0.0417	0.0542	0.0755	0.0699	0.0542
		ASE	0.0445	0.0557	0.0500	0.0447	0.0592	0.0802	0.0717	0.0594

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

Table 2.3: Frequency properties of dependence parameter estimators through joint analysis of the marked point process with a Poisson margin under 40% random right censoring where $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; $\gamma=0.5$; 500 simulated samples

			σ^2				ρ				Kendall's τ			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S	ML	2S	A2S	3S
Results based on 200 subjects/sample														
0.3	0.2	BIAS	-0.0018	-0.0024	-0.0019	-0.0024	-0.0026	-0.0030	-0.0026	-0.0026	0.0020	0.0010	0.0016	0.0010
		ESE	0.0605	0.0603	0.0605	0.0604	0.0431	0.0448	0.0431	0.0431	0.0254	0.0253	0.0254	0.0253
		ASE	0.0606	0.0610	0.0600	0.0599	0.0422	0.0440	0.0415	0.0416	0.0243	0.0242	0.0243	0.0242
	0.6	BIAS	0.0002	-0.0034	0.0000	-0.0037	-0.0004	-0.0035	-0.0006	0.0002	0.0025	-0.0015	0.0013	-0.0015
		ESE	0.0509	0.0610	0.0529	0.0506	0.0259	0.0460	0.0263	0.0259	0.0165	0.0172	0.0175	0.0172
		ASE	0.0500	0.0611	0.0525	0.0494	0.0261	0.0449	0.0263	0.0259	0.0165	0.0172	0.0173	0.0172
0.6	0.2	BIAS	-0.0040	-0.0051	-0.0041	-0.0047	-0.0038	-0.0045	-0.0039	-0.0039	0.0021	0.0009	0.0016	0.0009
		ESE	0.0752	0.0758	0.0752	0.0751	0.0353	0.0369	0.0353	0.0353	0.0254	0.0253	0.0254	0.0253
		ASE	0.0746	0.0764	0.0740	0.0739	0.0344	0.0361	0.0341	0.0341	0.0243	0.0242	0.0243	0.0242
	0.6	BIAS	0.0000	-0.0056	-0.0008	-0.0038	-0.0008	-0.0044	-0.0012	-0.0001	0.0027	-0.0019	0.0013	-0.001
		ESE	0.0580	0.0770	0.0601	0.0576	0.0222	0.0376	0.0228	0.0221	0.0160	0.0168	0.0173	0.0168
		ASE	0.0572	0.0771	0.0589	0.0567	0.0222	0.0369	0.0226	0.0221	0.0160	0.0171	0.0171	0.0171
Results based on 500 subjects/sample														
0.3	0.2	BIAS	0.0016	0.0017	0.0015	0.0013	-0.0011	-0.0011	-0.0011	-0.0011	0.0006	0.0002	0.0005	0.0002
		ESE	0.0362	0.0372	0.0362	0.0362	0.0275	0.0289	0.0275	0.0275	0.0156	0.0156	0.0156	0.0156
		ASE	0.0384	0.0390	0.0383	0.0382	0.0268	0.0283	0.0267	0.0268	0.0154	0.0153	0.0154	0.0153
	0.6	BIAS	0.0020	0.0004	0.0011	0.0002	-0.0008	-0.0023	-0.0006	-0.0005	0.0017	-0.0002	0.0009	-0.0002
		ESE	0.0298	0.0370	0.0308	0.0297	0.0167	0.0289	0.0168	0.0167	0.0099	0.0101	0.0103	0.0101
		ASE	0.0317	0.0391	0.0333	0.0316	0.0165	0.0288	0.0167	0.0165	0.0105	0.0108	0.0109	0.0108
0.6	0.2	BIAS	0.0012	0.0013	0.0011	0.0009	-0.0012	-0.0013	-0.0013	-0.0013	0.0007	0.0002	0.0005	0.0002
		ESE	0.0462	0.0476	0.0461	0.0462	0.0225	0.0235	0.0225	0.0225	0.0156	0.0156	0.0156	0.0156
		ASE	0.0475	0.0488	0.0472	0.0472	0.0217	0.0229	0.0217	0.0217	0.0153	0.0153	0.0153	0.0153
	0.6	BIAS	0.0017	0.0003	0.0008	0.0001	-0.0006	-0.0016	-0.0006	-0.0003	0.0016	-0.0003	0.0009	-0.0003
		ESE	0.0343	0.0467	0.0356	0.0343	0.0143	0.0231	0.0146	0.0143	0.0097	0.0099	0.0101	0.0099
		ASE	0.0362	0.0494	0.0374	0.0362	0.0140	0.0234	0.0143	0.0141	0.0101	0.0107	0.0108	0.0107

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

(NHPP) and a semiparametric Andersen-Gill model (AG). Marginal analysis of the marks are also carried out based on generalized estimating equations (GEE) under both an independence working correlation structure (WI) and an exchangeable working correlation structure (EXCH). Generally speaking, both the parametric and seimiparametric marginal analyses for the recurrent event process are valid but yield less efficient estimators compared to joint analysis; the efficacy loss is greater with larger values of Kendall's τ. The analysis of the marks based on generalized estimating equations with a working independence assumption yields heavily-biased estimators; use of an exchangeable working correlation structure provides some protection against the selection effects arising from the association between the marks and the event times.

We next repeat the simulation study with events generated according to a negative binomial process with intensity function given by (2.2.6) and the baseline intensity function is of the Weibull form as that in (2.3.1). Note that the Poisson model is a special case of this negative binomial model with $\gamma = 0$. The same parameter configuration is used as that for the Poisson model and γ is set to be 0.5. The frequency properties of the estimators are reported in Table 2.4, 2.5 and 2.6. The simulation results for the joint analysis with a negative binomial process for the events are similar to those obtained with a Poisson model. Small empirical biases of the estimators are observed except that those of γ are relatively large when sample size is 200. The empirical standard errors are close to average model-based standard errors. The efficiencies of estimators using the three-stage, the asymmetric two-stage and the usual two-stage estimation procedures are in a decreasing order compared to those using the simultaneous estimation procedure.

Table 2.4: Frequency properties of parameter estimators for the recurrent event process through joint analysis of the marked point process with a negative binomial margin under 40% random right censoring where $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; $\gamma=0.5$; 500 simulated samples

			$\log(\alpha_1)$				$\log(\alpha_2)$				β			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S	ML	2S	A2S	3S
							Results based on 200 subjects/sample							
0.3	0.2	BIAS	-0.0045	-0.0048	-0.0048	-0.0040	0.0012	0.0015	0.0015	0.0015	-0.0054	-0.0045	-0.0045	-0.0056
		ESE	0.1363	0.1394	0.1394	0.1363	0.0437	0.0452	0.0452	0.0437	0.1355	0.1390	0.1390	0.1355
		ASE	0.1374	0.1398	0.1398	0.1373	0.0429	0.0437	0.0437	0.0425	0.1366	0.1407	0.1407	0.1369
	0.6	BIAS	-0.0018	-0.0014	-0.0014	0.0068	0.0013	0.0008	0.0008	0.0029	-0.0041	-0.0102	-0.0102	-0.0063
		ESE	0.1147	0.1392	0.1392	0.1147	0.0321	0.0465	0.0465	0.0322	0.1114	0.1344	0.1344	0.1117
		ASE	0.1170	0.1401	0.1401	0.1162	0.0320	0.0437	0.0437	0.0317	0.1087	0.1412	0.1412	0.1091
0.6	0.2	BIAS	-0.0045	-0.0048	-0.0048	-0.0040	0.0012	0.0015	0.0015	0.0014	-0.0048	-0.0045	-0.0045	-0.0050
		ESE	0.1368	0.1394	0.1394	0.1368	0.0437	0.0452	0.0452	0.0437	0.1353	0.1390	0.1390	0.1353
		ASE	0.1375	0.1398	0.1398	0.1407	0.0429	0.0437	0.0437	0.0426	0.1360	0.1407	0.1407	0.1385
	0.6	BIAS	-0.0019	-0.0014	-0.0014	0.0069	0.0013	0.0008	0.0008	0.0026	-0.0019	-0.0102	-0.0102	-0.0039
		ESE	0.1182	0.1392	0.1392	0.1180	0.0323	0.0465	0.0465	0.0325	0.1066	0.1344	0.1344	0.1069
		ASE	0.1162	0.1401	0.1401	0.1167	0.0322	0.0437	0.0437	0.0320	0.1017	0.1412	0.1412	0.1025
							Results based on 500 subjects/sample							
0.3	0.2	BIAS	-0.0032	-0.0050	-0.0050	-0.0030	0.0018	0.0019	0.0019	0.0019	-0.0057	-0.0042	-0.0042	-0.0058
		ESE	0.0876	0.0913	0.0913	0.0876	0.0269	0.0279	0.0279	0.0269	0.0893	0.0932	0.0932	0.0894
		ASE	0.0874	0.0891	0.0891	0.0896	0.0271	0.0278	0.0278	0.0270	0.0871	0.0896	0.0896	0.0885
	0.6	BIAS	-0.0076	-0.0092	-0.0092	-0.0038	0.0010	0.0017	0.0017	0.0017	-0.0004	-0.0005	-0.0005	-0.0013
		ESE	0.0698	0.0845	0.0845	0.0697	0.0214	0.0291	0.0291	0.0214	0.0685	0.0927	0.0927	0.0685
		ASE	0.0734	0.0892	0.0892	0.0810	0.0202	0.0277	0.0277	0.0201	0.0684	0.0899	0.0899	0.0729
0.6	0.2	BIAS	-0.0106	-0.0103	-0.0103	-0.0103	0.0025	0.0027	0.0027	0.0026	0.0004	0.0001	0.0001	0.0003
		ESE	0.0860	0.0869	0.0869	0.0860	0.0281	0.0293	0.0293	0.0281	0.0833	0.0855	0.0855	0.0833
		ASE	0.0873	0.0887	0.0887	0.0875	0.0272	0.0277	0.0277	0.0270	0.0867	0.0896	0.0896	0.0868
	0.6	BIAS	-0.0089	-0.0092	-0.0092	-0.0050	0.0013	0.0017	0.0017	0.0019	0.0001	-0.0005	-0.0005	-0.0008
		ESE	0.0701	0.0845	0.0845	0.0700	0.0215	0.0291	0.0291	0.0215	0.0647	0.0927	0.0927	0.0647
		ASE	0.0734	0.0892	0.0892	0.0782	0.0204	0.0277	0.0277	0.0205	0.0641	0.0899	0.0899	0.0664

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

Table 2.5: Frequency properties of mean parameter estimators of the marks through joint analysis of the marked point process with a negative binomial margin under 40% random right censoring where $\mathrm{E}(N(1)|Z = 0) = 4$; $\alpha_1 = 6.35$, $\alpha_2 = 0.75$, $\beta = \log 0.5$; $\eta_0 = 0$, $\eta_1 = 0.5$, $\sigma^2 = 1$; $\gamma = 0.5$; 500 simulated samples

			γ				η_0				η_1			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S	ML	2S	A2S	3S
							Results based on 200 subjects/sample							
0.3	0.2	BIAS	-0.0195	-0.0192	-0.0192	-0.0199	-0.0028	-0.0020	-0.0027	-0.0032	0.0045	0.0039	0.0042	0.0046
		ESE	0.0936	0.0986	0.0986	0.0936	0.0711	0.0715	0.0713	0.0711	0.1008	0.1017	0.1007	0.1007
		ASE	0.0943	0.1000	0.1000	0.0940	0.0704	0.0739	0.0726	0.0727	0.1045	0.1084	0.1062	0.1061
	0.6	BIAS	-0.0102	-0.0122	-0.0122	-0.0128	-0.0019	-0.0026	-0.0023	-0.0080	0.0020	0.0058	0.0047	0.0035
		ESE	0.0652	0.1001	0.1001	0.0651	0.0666	0.0730	0.0747	0.0668	0.0912	0.1023	0.1016	0.0913
		ASE	0.0650	0.1008	0.1008	0.0647	0.0671	0.0764	0.0774	0.0668	0.0910	0.1091	0.1073	0.0908
0.6	0.2	BIAS	-0.0197	-0.0192	-0.0192	-0.0201	-0.0030	-0.0025	-0.0030	-0.0036	0.0033	0.0033	0.0031	0.0035
		ESE	0.0940	0.0986	0.0986	0.0940	0.0843	0.0852	0.0845	0.0843	0.1181	0.1202	0.1180	0.1181
		ASE	0.0947	0.1000	0.1000	0.0955	0.0838	0.0920	0.0923	0.0945	0.1207	0.1302	0.1286	0.1310
	0.6	BIAS	-0.0100	-0.0122	-0.0122	-0.0136	-0.0016	-0.0031	-0.0029	-0.0086	-0.0009	0.0049	0.0038	0.0009
		ESE	0.0659	0.1001	0.1001	0.0657	0.0772	0.0856	0.0857	0.0774	0.1016	0.1207	0.1141	0.1018
		ASE	0.0650	0.1008	0.1008	0.0649	0.0746	0.0870	0.0896	0.0758	0.0980	0.1250	0.1208	0.0987
							Results based on 500 subjects/sample							
0.3	0.2	BIAS	-0.0014	-0.0027	-0.0027	-0.0016	0.0031	0.0031	0.0033	0.0029	-0.0041	-0.0044	-0.0044	-0.0040
		ESE	0.0589	0.0641	0.0641	0.0589	0.0476	0.0484	0.0479	0.0476	0.0718	0.0730	0.0723	0.0718
		ASE	0.0609	0.0645	0.0645	0.0609	0.0444	0.0504	0.0477	0.0496	0.0662	0.0725	0.0690	0.0712
	0.6	BIAS	-0.0025	-0.0026	-0.0026	-0.0036	0.0013	0.0003	0.0016	-0.0013	0.0004	0.0016	0.0001	0.0011
		ESE	0.0429	0.0640	0.0640	0.0428	0.0391	0.0446	0.0434	0.0389	0.0562	0.0692	0.0683	0.0563
		ASE	0.0411	0.0646	0.0646	0.0427	0.0418	0.0463	0.0506	0.0482	0.0572	0.0680	0.0700	0.0625
0.6	0.2	BIAS	-0.0030	-0.0043	-0.0043	-0.0031	0.0017	0.0012	0.0014	0.0014	0.0007	0.0009	0.0008	0.0007
		ESE	0.0604	0.0647	0.0647	0.0604	0.0517	0.0524	0.0514	0.0517	0.0783	0.0803	0.0785	0.0783
		ASE	0.0610	0.0644	0.0644	0.0609	0.0526	0.1220	0.0584	0.0545	0.0765	0.1468	0.0820	0.0781
	0.6	BIAS	-0.0022	-0.0026	-0.0026	-0.0037	0.0026	0.0008	0.0022	-0.0005	-0.0003	0.0011	-0.0004	0.0004
		ESE	0.0424	0.0640	0.0640	0.0422	0.0444	0.0527	0.0494	0.0442	0.0609	0.0807	0.0752	0.0609
		ASE	0.0412	0.0646	0.0646	0.0422	0.0469	0.0552	0.0555	0.0518	0.0618	0.0795	0.0756	0.0655

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

Table 2.6: Frequency properties of parameter estimators for dependence parameters through joint analysis of the marked point process with a negative binomial margin under 40% random right censoring where $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; $\gamma=0.5$; 500 simulated samples

			σ^2				ρ				Kendall's τ			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S	ML	2S	A2S	3S
Results based on 200 subjects/sample														
0.3	0.2	BIAS	-0.0045	-0.0036	-0.0045	-0.0051	-0.0039	-0.0034	-0.0039	-0.0039	0.0012	0.0000	0.0004	0.0000
		ESE	0.0602	0.0610	0.0602	0.0602	0.0447	0.0466	0.0448	0.0447	0.0255	0.0254	0.0254	0.0254
		ASE	0.0617	0.0622	0.0611	0.0610	0.0436	0.0454	0.0429	0.0429	0.0244	0.0242	0.0243	0.0242
	0.6	BIAS	-0.0032	-0.0036	-0.0042	-0.0076	-0.0030	-0.0034	-0.0045	-0.0026	0.0014	-0.0027	-0.0006	-0.0027
		ESE	0.0540	0.0623	0.0567	0.0539	0.0321	0.0476	0.0401	0.0321	0.0169	0.0171	0.0175	0.0171
		ASE	0.0549	0.0623	0.0580	0.0544	0.0313	0.0459	0.0389	0.0311	0.0168	0.0174	0.0177	0.01674
0.6	0.2	BIAS	-0.0072	-0.0061	-0.0073	-0.0078	-0.0049	-0.0045	-0.0049	-0.0049	0.0012	0.0000	0.0004	0.0000
		ESE	0.0762	0.0771	0.0762	0.0761	0.0369	0.0378	0.0369	0.0369	0.0255	0.0254	0.0255	0.0254
		ASE	0.0765	0.0781	0.0757	0.0757	0.0356	0.0372	0.0353	0.0353	0.0243	0.0243	0.0243	0.0243
	0.6	BIAS	-0.0043	-0.0058	-0.0066	-0.0086	-0.0031	-0.0045	-0.0049	-0.0026	0.0015	-0.0030	-0.0008	-0.0030
		ESE	0.0665	0.0790	0.0708	0.0664	0.0277	0.0386	0.0338	0.0277	0.0165	0.0168	0.0173	0.0168
		ASE	0.0660	0.0787	0.0714	0.0656	0.0268	0.0378	0.0326	0.0266	0.0163	0.0172	0.0176	0.0172
Results based on 500 subjects/sample														
0.3	0.2	BIAS	-0.0015	-0.0009	-0.0015	-0.0017	-0.0019	-0.0013	-0.0020	-0.0019	0.0004	-0.0001	0.0001	-0.0001
		ESE	0.0386	0.0394	0.0386	0.0386	0.0280	0.0296	0.0280	0.0280	0.0149	0.0149	0.0149	0.0149
		ASE	0.0392	0.0398	0.0391	0.0392	0.0278	0.0293	0.0278	0.0279	0.0154	0.0154	0.0154	0.0154
	0.6	BIAS	0.0011	0.0019	0.0005	-0.0008	-0.0018	-0.0021	-0.0020	-0.0016	0.0016	-0.0002	0.0006	-0.0002
		ESE	0.0342	0.0402	0.0360	0.0342	0.0201	0.0289	0.0251	0.0201	0.0100	0.0102	0.0104	0.0102
		ASE	0.0348	0.0398	0.0372	0.0350	0.0198	0.0294	0.0251	0.0200	0.0106	0.0109	0.0111	0.0109
0.6	0.2	BIAS	0.0012	0.0011	0.0010	0.0009	-0.0021	-0.0024	-0.0022	-0.0021	0.0006	0.0001	0.0002	0.0001
		ESE	0.0489	0.0504	0.0489	0.0489	0.0228	0.0237	0.0229	0.0228	0.0155	0.0155	0.0154	0.0155
		ASE	0.0489	0.0521	0.0486	0.0486	0.0225	0.0251	0.0225	0.0224	0.0153	0.0169	0.0154	0.0169
	0.6	BIAS	-0.0004	-0.0002	-0.0014	-0.0023	-0.0021	-0.0029	-0.0025	-0.0019	0.0017	-0.0003	0.0005	-0.0003
		ESE	0.0412	0.0506	0.0446	0.0411	0.0172	0.0242	0.0213	0.0173	0.0099	0.0101	0.0103	0.0101
		ASE	0.0418	0.0502	0.0450	0.0416	0.0169	0.0240	0.0209	0.0169	0.0103	0.0107	0.0110	0.0107

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

2.3.2 Empirical Performance under Model Misspecification

Misspecification of the Marginal Models

In Table 2.7, 2.8 and 2.9, we report the results for the marginal event process when data are generated from a joint model with a negative binomial process for the events of which the intensity is of the form (2.2.6) with $\gamma = 0.5$ but analyzed based on a Poisson assumption. Not surprisingly the empirical biases of the marginal estimators under the two-stage estimation procedure and those of recurrent event parameters under asymmetric two-stage estimation procedure are ignorable. This is a consequence of the robustness of estimation of rate and mean function parameters in the context of a mixed Poisson model. Misspecification of the marginal model for the recurrent event process leads to biased estimators of the marginal parameters for the marks as well as estimation of the association parameter when using simultaneous, asymmetric and three-stage estimation procedures; this effect is stronger for larger τ.

In Diao *and others* (2013), the resulting data from a joint model with a negative binomial margin are analyzed marginally under a misspecified parametric non-homogeneous Poisson (NHPP), a semiparametric Anderson-Gill (AG) model, a semiparametric negative binomial model (SNB) for the recurrent events and GEE with a working independence correlation structure and an exchangeable correlation structure for the marks. The biases of the estimators under the marginal recurrent event models are negligible and this phenomenon reveals that misspecification of the marginal intensity is more of a concern in the joint analysis. The empirical biases of the estimators with a working independence correlation structure are larger than those with an exchangeable correlation structure.

Table 2.7: Sensitivity analysis on the effect of misspecification of the marginal recurrent event process generated from negative binomial process and analyzed by Poisson process showing the frequency properties of parameter estimators from the recurrent event process via joint analysis of the marked point process; 40% random right censoring with $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; $\gamma=0.5$; 500 simulated samples

			$\log(\alpha_1)$				$\log(\alpha_2)$				β			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S	ML	2S	A2S	3S
							Results based on 200 subjects/sample							
0.3	0.2	BIAS	-0.0182	-0.0062	-0.0062	-0.0181	0.0052	-0.0013	-0.0013	0.0052	0.0039	-0.0004	-0.0004	0.0038
		ESE	0.1369	0.1402	0.1402	0.1368	0.0441	0.0455	0.0455	0.0441	0.1358	0.1392	0.1392	0.1358
		ASE	0.1016	0.1401	0.1401	0.1371	0.0424	0.0441	0.0441	0.0430	0.0941	0.1408	0.1408	0.1368
	0.6	BIAS	-0.0868	-0.0037	-0.0037	-0.0461	0.0356	-0.0022	-0.0022	0.0370	0.0403	-0.0043	-0.0043	0.0338
		ESE	0.1188	0.1398	0.1398	0.1208	0.0351	0.0475	0.0475	0.0351	0.1146	0.1338	0.1338	0.1157
		ASE	0.0850	0.1403	0.1403	0.1225	0.0336	0.0442	0.0442	0.0339	0.0778	0.1412	0.1412	0.1193
0.6	0.2	BIAS	-0.0113	-0.0062	-0.0062	-0.0110	0.0033	-0.0013	-0.0013	0.0034	0.0039	-0.0004	-0.0004	0.0037
		ESE	0.1376	0.1402	0.1402	0.1376	0.0442	0.0455	0.0455	0.0443	0.1358	0.1392	0.1392	0.1358
		ASE	0.1021	0.1401	0.1401	0.1375	0.0426	0.0441	0.0441	0.0431	0.0940	0.1408	0.1408	0.1365
	0.6	BIAS	0.0062	-0.0037	-0.0037	0.0272	0.0206	-0.0022	-0.0022	0.0207	0.0164	-0.0043	-0.0043	0.0152
		ESE	0.1274	0.1398	0.1398	0.1287	0.0362	0.0475	0.0475	0.0364	0.1137	0.1338	0.1338	0.1141
		ASE	0.0909	0.1403	0.1403	0.1281	0.0348	0.0442	0.0442	0.0349	0.0771	0.1412	0.1412	0.1166
							Results based on 500 subjects/sample							
0.3	0.2	BIAS	-0.0179	-0.0065	-0.0065	-0.0179	0.0056	-0.0009	-0.0009	0.0056	0.0053	0.0005	0.0005	0.0052
		ESE	0.0886	0.0914	0.0914	0.0886	0.0276	0.0285	0.0285	0.0276	0.0903	0.0934	0.0934	0.0902
		ASE	0.0641	0.0893	0.0893	0.0879	0.0268	0.0280	0.0280	0.0273	0.0594	0.0896	0.0896	0.0874
	0.6	BIAS	-0.0944	-0.0105	-0.0105	-0.0532	0.0360	-0.0010	-0.0010	0.0376	0.0494	0.0035	0.0035	0.0427
		ESE	0.0716	0.0841	0.0841	0.0734	0.0226	0.0296	0.0296	0.0227	0.0741	0.0931	0.0931	0.0754
		ASE	0.0536	0.0893	0.0893	0.0777	0.0213	0.0280	0.0280	0.0214	0.0491	0.0899	0.0899	0.0754
0.6	0.2	BIAS	-0.0177	-0.0119	-0.0119	-0.0176	0.0046	0.0000	0.0000	0.0046	0.0102	0.0044	0.0044	0.0102
		ESE	0.0853	0.0867	0.0867	0.0854	0.0286	0.0297	0.0297	0.0286	0.0826	0.0855	0.0855	0.0826
		ASE	0.0644	0.0889	0.0889	0.0873	0.0269	0.0280	0.0280	0.0274	0.0593	0.0897	0.0897	0.0869
	0.6	BIAS	-0.0007	-0.0105	-0.0105	0.0201	0.0213	-0.0010	-0.0010	0.0214	0.0233	0.0035	0.0035	0.0220
		ESE	0.0757	0.0841	0.0841	0.0768	0.0232	0.0296	0.0296	0.0233	0.0733	0.0931	0.0931	0.0739
		ASE	0.0574	0.0893	0.0893	0.0815	0.0220	0.0280	0.0280	0.0221	0.0487	0.0899	0.0899	0.0739

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

Table 2.8: Sensitivity analysis on the effect of misspecification of the marginal recurrent event process generated from negative binomial process and analyzed by Poisson process showing the frequency properties of mean parameter estimators for the marks via joint analysis of the marked point process; 40% random right censoring with $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; $\gamma=0.5$; 500 simulated samples

			η_0				η_1			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S
			Results based on 200 subjects/sample							
0.3	0.2	BIAS	-0.0440	-0.0020	-0.0446	-0.0440	0.0198	0.0039	0.0210	0.0198
		ESE	0.0715	0.0715	0.0716	0.0715	0.1018	0.1017	0.1018	0.1018
		ASE	0.0681	0.0739	0.0720	0.0713	0.1021	0.1084	0.1068	0.1062
	0.6	BIAS	-0.1187	-0.0026	-0.1364	-0.1455	0.0036	0.0058	0.0338	0.0132
		ESE	0.0695	0.0730	0.0728	0.0716	0.0975	0.1023	0.1043	0.0986
		ASE	0.0521	0.0764	0.0747	0.0716	0.0752	0.1091	0.1084	0.1002
0.6	0.2	BIAS	-0.0709	-0.0025	-0.0707	-0.0709	0.0322	0.0033	0.0334	0.0322
		ESE	0.0855	0.0852	0.0857	0.0855	0.1198	0.1202	0.1199	0.1198
		ASE	0.0822	0.0920	0.0857	0.0847	0.1188	0.1302	0.1239	0.1229
	0.6	BIAS	-0.2032	-0.0031	-0.1790	-0.2181	0.0629	0.0049	0.0701	0.0671
		ESE	0.0833	0.0856	0.0873	0.0850	0.1128	0.1207	0.1212	0.1137
		ASE	0.0628	0.0870	0.0877	0.0837	0.0878	0.1250	0.1244	0.1137
			Results based on 500 subjects/sample							
0.3	0.2	BIAS	-0.0396	0.0031	-0.0402	-0.0396	0.0120	-0.0044	0.0133	0.0120
		ESE	0.0480	0.0484	0.0482	0.0480	0.0729	0.0730	0.0732	0.0729
		ASE	0.0433	0.0504	0.0455	0.0478	0.0648	0.0725	0.0676	0.0697
	0.6	BIAS	-0.1187	0.0003	-0.1377	-0.1460	0.0006	0.0016	0.0320	0.0105
		ESE	0.0426	0.0446	0.0445	0.0444	0.0623	0.0692	0.0703	0.0635
		ASE	0.0329	0.0463	0.0467	0.0451	0.0474	0.0680	0.0683	0.0631
0.6	0.2	BIAS	-0.0683	0.0012	-0.0684	-0.0683	0.0305	0.0009	0.0322	0.0305
		ESE	0.0522	0.0524	0.0521	0.0522	0.0797	0.0803	0.0799	0.0797
		ASE	0.0522	0.1220	0.0543	0.0542	0.0755	0.1468	0.0785	0.0782
	0.6	BIAS	-0.2038	0.0008	-0.1800	-0.2188	0.0630	0.0011	0.0697	0.0672
		ESE	0.0503	0.0527	0.0530	0.0515	0.0708	0.0807	0.0811	0.0717
		ASE	0.0396	0.0552	0.0558	0.0532	0.0554	0.0795	0.0791	0.0719

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

Table 2.9: Sensitivity analysis on the effect of misspecification of the marginal recurrent event process generated from negative binomial process and analyzed by Poisson process showing the frequency properties of dependence parameter estimators via joint analysis of the marked point process; 40% random right censoring with $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; $\gamma=0.5$; 500 simulated samples

			σ^2				ρ				Kendall's τ			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S	ML	2S	A2S	3S
Results based on 200 subjects/sample														
0.3	0.2	BIAS	-0.0410	-0.0036	-0.0408	-0.0410	-0.0329	-0.0034	-0.0330	-0.0328	-0.0332	-0.0335	-0.0333	-0.0335
		ESE	0.0564	0.0610	0.0565	0.0565	0.0440	0.0466	0.0440	0.0440	0.0231	0.0231	0.0231	0.0231
		ASE	0.0569	0.0622	0.0568	0.0567	0.0426	0.0454	0.0426	0.0426	0.0221	0.0226	0.0225	0.0226
	0.6	BIAS	-0.1764	-0.0036	-0.1732	-0.1892	-0.1735	-0.0034	-0.1750	-0.1631	-0.0703	-0.0997	-0.0717	-0.0997
		ESE	0.0427	0.0623	0.0441	0.0423	0.0278	0.0476	0.0276	0.0281	0.0210	0.0228	0.0206	0.0228
		ASE	0.0388	0.0623	0.0447	0.0422	0.0237	0.0459	0.0282	0.0281	0.0185	0.0241	0.0217	0.0241
0.6	0.2	BIAS	-0.0606	-0.0061	-0.0601	-0.0606	-0.0303	-0.0045	-0.0304	-0.0301	-0.0327	-0.0335	-0.0326	-0.0335
		ESE	0.0706	0.0771	0.0706	0.0706	0.0384	0.0378	0.0384	0.0384	0.0230	0.0231	0.0231	0.0231
		ASE	0.0695	0.0781	0.0698	0.0696	0.0368	0.0372	0.0368	0.0369	0.0219	0.0225	0.0223	0.0225
	0.6	BIAS	-0.2500	-0.0058	-0.2407	-0.2529	-0.1489	-0.0045	-0.1568	-0.1405	-0.0813	-0.0998	-0.0746	-0.0998
		ESE	0.0471	0.0790	0.0482	0.0475	0.0356	0.0386	0.0359	0.0345	0.0190	0.0216	0.0195	0.0216
		ASE	0.0399	0.0787	0.0479	0.0465	0.0299	0.0378	0.0353	0.0350	0.0174	0.0229	0.0204	0.0229
Results based on 500 subjects/sample														
0.3	0.2	BIAS	-0.0391	-0.0009	-0.0388	-0.0390	-0.0322	-0.0013	-0.0323	-0.0321	-0.0352	-0.0352	-0.0351	-0.0352
		ESE	0.0362	0.0394	0.0362	0.0362	0.0282	0.0296	0.0282	0.0282	0.0138	0.0140	0.0138	0.0140
		ASE	0.0360	0.0398	0.0363	0.0362	0.0271	0.0293	0.0275	0.0275	0.0139	0.0142	0.0141	0.0142
	0.6	BIAS	-0.1743	0.0019	-0.1720	-0.1874	-0.1768	-0.0021	-0.1782	-0.1664	-0.0715	-0.1012	-0.0727	-0.1012
		ESE	0.0274	0.0402	0.0283	0.0271	0.0174	0.0289	0.0173	0.0174	0.0132	0.0150	0.0132	0.0150
		ASE	0.0245	0.0398	0.0281	0.0268	0.0148	0.0294	0.0178	0.0178	0.0118	0.0154	0.0138	0.0154
0.6	0.2	BIAS	-0.0537	0.0011	-0.0533	-0.0537	-0.0281	-0.0024	-0.0282	-0.0280	-0.0344	-0.0348	-0.0342	-0.0348
		ESE	0.0450	0.0504	0.0451	0.0450	0.0239	0.0237	0.0240	0.0239	0.0136	0.0138	0.0136	0.0138
		ASE	0.0444	0.0521	0.0448	0.0447	0.0233	0.0251	0.0236	0.0236	0.0138	0.0148	0.0140	0.0148
	0.6	BIAS	-0.2503	-0.0002	-0.2415	-0.2535	-0.1519	-0.0029	-0.1598	-0.1439	-0.0831	-0.1012	-0.0759	-0.1012
		ESE	0.0286	0.0506	0.0295	0.0287	0.0211	0.0242	0.0212	0.0206	0.0122	0.0143	0.0125	0.0143
		ASE	0.0251	0.0502	0.0302	0.0294	0.0190	0.0240	0.0225	0.0223	0.0110	0.0147	0.0130	0.0147

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

Misspecification of the Copula Function

Prokhorov and Schmidt (2009) discuss asymptotic properties of parameter estimators under misspecification of the copula model and characterize settings in which inferences can be robust regarding features of the marginal distributions. Other authors have examined the issue of copula misspecification in special cases. Chatterjee *and others* (2006) investigate misspecification in the context of "parallel" bivariate survival data in response-dependent (i.e. case-control and case-only) designs and find misspecification of the copula function does not lead to appreciable bias in estimators of the marginal parameters for some types of misspecification. Craiu and Craiu (2008) empirically study the effect of copula misspecification on conditional means and variances and show there can be significant impact on inferences; careful consideration of the copula family is therefore required. He and Lawless (2005) study multivariate regression models whose marginals are of location-scale form and the copula is used to link the error terms. They find that, for complete data, the MLEs for regression coefficients are consistent even when the joint distribution is misspecified. In the case of censored responses, estimators of regression coefficients are no longer consistent under model misspecification, but bias is small in many practical situations given the simulation results.

In this chapter, we deal with a setting involving life history data in which the copula plays a role in defining the conditional distribution of successive marks and event times. To provide insight into the effect of misspecification in this context, we carry out further simulations in which data are generated using a Frank copula or a Gumbel copula, both of which are in the Archimedian family (Nelsen, 2006); analyses, however, are conducted

Table 2.10: Sensitivity analysis on the effect of copula misspecification showing the frequency properties of parameter estimators from the recurrent event process via joint analysis of the marked point process; marginal intensity is Poisson; 40% random right censoring with $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; 200 subjects/sample; 500 simulated samples

			$\log(\alpha_1)$				$\log(\alpha_2)$				β			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S	ML	2S	A2S	3S
							Data are generated using Frank copula							
0.3	0.2	BIAS	0.0023	-0.0055	-0.0055	0.0026	0.0000	0.0038	0.0038	0.0002	0.0019	-0.0036	-0.0036	0.0017
		ESE	0.1022	0.1014	0.1014	0.1022	0.0448	0.0454	0.0454	0.0448	0.0946	0.0938	0.0938	0.0946
		ASE	0.1027	0.1030	0.1030	0.1041	0.0425	0.0427	0.0427	0.0420	0.0942	0.0957	0.0957	0.0964
	0.6	BIAS	0.0553	-0.0006	-0.0006	0.0665	-0.0285	0.0023	0.0023	-0.0272	0.0577	-0.0068	-0.0068	0.0552
		ESE	0.0978	0.1010	0.1010	0.0977	0.0357	0.0438	0.0438	0.0358	0.0887	0.0935	0.0935	0.0887
		ASE	0.0947	0.1032	0.1032	0.1064	0.0337	0.0428	0.0428	0.0349	0.0781	0.0956	0.0956	0.0928
0.6	0.2	BIAS	0.0019	-0.0055	-0.0055	0.0023	-0.0001	0.0038	0.0038	0.0002	0.0024	-0.0036	-0.0036	0.0022
		ESE	0.1024	0.1014	0.1014	0.1024	0.0450	0.0454	0.0454	0.0450	0.0943	0.0938	0.0938	0.0943
		ASE	0.1028	0.1030	0.1030	0.1034	0.0426	0.0427	0.0427	0.0422	0.0940	0.0957	0.0957	0.0955
	0.6	BIAS	0.0628	-0.0006	-0.0006	0.0742	-0.0329	0.0023	0.0023	-0.0317	0.0533	-0.0068	-0.0068	0.0513
		ESE	0.1003	0.1010	0.1010	0.1001	0.0377	0.0438	0.0438	0.0378	0.0846	0.0935	0.0935	0.0846
		ASE	0.0968	0.1032	0.1032	0.1034	0.0353	0.0428	0.0428	0.0365	0.0749	0.0956	0.0956	0.0858
							Data are generated using Gumbel copula							
0.3	0.2	BIAS	0.0089	-0.0005	-0.0005	0.0092	-0.0047	0.0006	0.0006	-0.0044	0.0025	-0.0033	-0.0033	0.0022
		ESE	0.1018	0.1016	0.1016	0.1018	0.0444	0.0454	0.0454	0.0444	0.0973	0.0952	0.0952	0.0973
		ASE	0.1035	0.1034	0.1034	0.1055	0.0427	0.0431	0.0431	0.0429	0.0946	0.0958	0.0958	0.0976
	0.6	BIAS	0.0659	-0.0008	-0.0008	0.0785	-0.0371	-0.0002	-0.0002	-0.0354	0.0588	-0.0026	-0.0026	0.0558
		ESE	0.1005	0.1007	0.1007	0.1005	0.0368	0.0445	0.0445	0.0370	0.0949	0.0925	0.0925	0.0949
		ASE	0.0956	0.1038	0.1038	0.1068	0.0343	0.0432	0.0432	0.0366	0.0788	0.0958	0.0958	0.0946
0.6	0.2	BIAS	0.0089	-0.0005	-0.0005	0.0093	-0.0048	0.0006	0.0006	-0.0045	0.0025	-0.0033	-0.0033	0.0022
		ESE	0.1025	0.1016	0.1016	0.1025	0.0447	0.0454	0.0454	0.0447	0.0968	0.0952	0.0952	0.0967
		ASE	0.1037	0.1034	0.1034	0.1051	0.0428	0.0431	0.0431	0.0431	0.0944	0.0958	0.0958	0.0967
	0.6	BIAS	0.0766	-0.0008	-0.0008	0.0888	-0.0425	-0.0002	-0.0002	-0.0408	0.0539	-0.0026	-0.0026	0.0515
		ESE	0.1051	0.1007	0.1007	0.1053	0.0398	0.0445	0.0445	0.0399	0.0904	0.0925	0.0925	0.0904
		ASE	0.0994	0.1038	0.1038	0.1151	0.0361	0.0432	0.0432	0.0405	0.0762	0.0958	0.0958	0.0912

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

Table 2.11: Sensitivity analysis on the effect of copula misspecification showing the frequency properties of mean parameter estimators of the marks via joint analysis of the marked point process; marginal intensity is Poisson; 40% random right censoring with $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; 200 subjects/sample; 500 simulated samples

			η_0				η_1			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S
			Data are generated using Frank copula							
0.3	0.2	BIAS	-0.0063	-0.0036	-0.0057	-0.0068	0.0040	0.0075	0.0050	0.0042
		ESE	0.0657	0.0649	0.0656	0.0656	0.0996	0.0986	0.0993	0.0995
		ASE	0.0686	0.0692	0.0719	0.0744	0.1035	0.1048	0.1067	0.1095
	0.6	BIAS	-0.0338	-0.0030	-0.0218	-0.0433	-0.0400	0.0074	0.0030	-0.0362
		ESE	0.0685	0.0663	0.0700	0.0675	0.1031	0.0983	0.1044	0.1028
		ASE	0.0684	0.0711	0.0744	0.0783	0.0976	0.1052	0.1104	0.1133
0.6	0.2	BIAS	-0.0061	-0.0037	-0.0054	-0.0067	0.0027	0.0062	0.0039	0.0029
		ESE	0.0814	0.0805	0.0811	0.0813	0.1189	0.1182	0.1185	0.1189
		ASE	0.0827	0.0861	0.1010	0.0846	0.1207	0.1244	0.1386	0.1231
	0.6	BIAS	-0.0453	-0.0034	-0.0237	-0.0552	-0.0333	0.0062	0.0046	-0.0299
		ESE	0.0838	0.0811	0.0836	0.0825	0.1182	0.1179	0.1191	0.1176
		ASE	0.0796	0.0849	0.0871	0.0866	0.1080	0.1233	0.1243	0.1205
			Data are generated using Gumbel copula							
0.3	0.2	BIAS	-0.0019	0.0020	-0.0014	-0.0024	-0.0036	-0.0005	-0.0026	-0.0034
		ESE	0.0751	0.0740	0.0748	0.0751	0.1084	0.1062	0.1077	0.1083
		ASE	0.0698	0.0702	0.0709	0.0761	0.1045	0.1056	0.1063	0.1116
	0.6	BIAS	-0.0373	0.0028	-0.0234	-0.0479	-0.0434	-0.0018	-0.0041	-0.0390
		ESE	0.0783	0.0749	0.0794	0.0773	0.1165	0.1082	0.1137	0.1161
		ASE	0.0684	0.0719	0.0749	0.0759	0.0982	0.1068	0.1107	0.1121
0.6	0.2	BIAS	-0.0026	0.0013	-0.0018	-0.0032	-0.0035	-0.0004	-0.0026	-0.0033
		ESE	0.0914	0.0900	0.0911	0.0914	0.1270	0.1246	0.1264	0.1270
		ASE	0.0834	0.0828	0.0890	0.0878	0.1217	0.1220	0.1272	0.1261
	0.6	BIAS	-0.0524	0.0024	-0.0262	-0.0632	-0.0350	-0.0024	-0.0026	-0.0312
		ESE	0.0935	0.0890	0.0939	0.0925	0.1308	0.1262	0.1296	0.1304
		ASE	0.0816	0.0855	0.0894	0.0966	0.1104	0.1243	0.1256	0.1296

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

Table 2.12: Sensitivity analysis on the effect of copula misspecification showing the frequency properties of dependence parameter estimators via joint analysis of the marked point process; marginal intensity is Poisson; 40% random right censoring with $\mathrm{E}(N(1)|Z=0)=4$; $\alpha_1=6.35$, $\alpha_2=0.75$, $\beta=\log 0.5$; $\eta_0=0$, $\eta_1=0.5$, $\sigma^2=1$; 200 subjects/sample; 500 simulated samples

			σ^2				ρ				Kendall's τ			
ρ	Kendall's τ		ML	2S	A2S	3S	ML	2S	A2S	3S	ML	2S	A2S	3S
							Data are generated using Frank copula							
0.3	0.2	BIAS	0.0058	-0.0057	0.0060	0.0049	-0.0022	-0.0033	-0.0022	-0.0023	-0.0410	-0.0427	-0.0414	-0.0427
		ESE	0.0624	0.0612	0.0625	0.0623	0.0429	0.0437	0.0429	0.0429	0.0253	0.0249	0.0252	0.0249
		ASE	0.0615	0.0602	0.0611	0.0611	0.0421	0.0426	0.0419	0.0420	0.0251	0.0251	0.0254	0.0251
	0.6	BIAS	0.0809	-0.0038	0.0794	0.0695	0.0078	-0.0015	0.0077	0.0085	-0.0780	-0.0876	-0.0776	-0.0876
		ESE	0.0641	0.0606	0.0657	0.0628	0.0355	0.0433	0.0357	0.0355	0.0216	0.0220	0.0214	0.0220
		ASE	0.0599	0.0605	0.0666	0.0649	0.0311	0.0432	0.0361	0.0362	0.0191	0.0217	0.0211	0.0217
0.6	0.2	BIAS	0.0058	-0.0073	0.0060	0.0047	-0.0022	-0.0037	-0.0022	-0.0023	-0.0409	-0.0427	-0.0412	-0.0427
		ESE	0.0768	0.0758	0.0769	0.0767	0.0351	0.0358	0.0351	0.0351	0.0253	0.0248	0.0252	0.0248
		ASE	0.0762	0.0753	0.0766	0.0764	0.0345	0.0353	0.0350	0.0349	0.0251	0.0251	0.0255	0.0251
	0.6	BIAS	0.0963	-0.0047	0.0889	0.0836	0.0104	-0.0026	0.0086	0.0110	-0.0768	-0.0876	-0.0773	-0.0876
		ESE	0.0806	0.0762	0.0806	0.0790	0.0308	0.0356	0.0312	0.0307	0.0217	0.0219	0.0214	0.0219
		ASE	0.0717	0.0761	0.0825	0.0826	0.0258	0.0358	0.0321	0.0317	0.0186	0.0216	0.0210	0.0216
							Data are generated using Gumbel copula							
0.3	0.2	BIAS	0.0083	-0.0054	0.0084	0.0072	-0.0011	-0.0022	-0.0010	-0.0011	-0.0520	-0.0540	-0.0525	-0.0540
		ESE	0.0620	0.0607	0.0620	0.0619	0.0417	0.0426	0.0417	0.0417	0.0269	0.0263	0.0267	0.0263
		ASE	0.0618	0.0600	0.0615	0.0614	0.0420	0.0424	0.0419	0.0419	0.0255	0.0260	0.0264	0.0260
	0.6	BIAS	0.0940	-0.0016	0.0903	0.0803	0.0069	-0.0015	0.0066	0.0077	-0.0872	-0.0985	-0.0876	-0.0985
		ESE	0.0666	0.0620	0.0684	0.0658	0.0382	0.0441	0.0382	0.0381	0.0217	0.0222	0.0216	0.0222
		ASE	0.0613	0.0609	0.0656	0.0641	0.0317	0.0435	0.0361	0.0362	0.0192	0.0221	0.0215	0.0221
0.6	0.2	BIAS	0.0085	-0.0069	0.0086	0.0073	-0.0017	-0.0033	-0.0017	-0.0018	-0.0518	-0.0540	-0.0523	-0.0540
		ESE	0.0781	0.0765	0.0781	0.0779	0.0347	0.0352	0.0347	0.0347	0.0270	0.0263	0.0268	0.0263
		ASE	0.0766	0.0750	0.0766	0.0765	0.0345	0.0351	0.0347	0.0347	0.0255	0.0260	0.0264	0.0260
	0.6	BIAS	0.1082	-0.0035	0.0978	0.0930	0.0087	-0.0033	0.0067	0.0094	-0.0860	-0.0986	-0.0872	-0.0986
		ESE	0.0838	0.0770	0.0841	0.0824	0.0327	0.0359	0.0330	0.0326	0.0220	0.0222	0.0216	0.0222
		ASE	0.0734	0.0764	0.0812	0.0819	0.0263	0.0360	0.0313	0.0312	0.0189	0.0222	0.0215	0.0222

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

based on a Clayton copula. We consider $\rho = 0.3$ and 0.6 for the correlations between the marks, and $\tau = 0.2$ and 0.6 for the dependence between the marks and the event times, and restrict attention to the setting in which the events are generated according to a marginal Poisson process and with a sample size of $m = 200$.

The results are reported in Table 2.10, 2.11 and 2.12 and reveal that biases arising from copula misspecification are negligible for the marginal parameters when Kendall's τ is modest, but can be appreciable when Kendall's τ is large. Not surprisingly the biases for Kendall's τ are generally quite large. We remark that in addition to biases that may arise due to copula misspecification in the context of "parallel" bivariate survival data, the limiting behaviour in our context depends on the validity of the likelihood in (2.2.5) and specifically the conditional independence assumptions A.2.1 and A.2.2, which are not satisfied under misspecification of the copula.

2.4 Application to Data from the Mirasol Study

The Mirasol Study (Cazenave *and others*, 2010) is a multicenter, open-label, parallel-group non-inferiority randomized controlled trial. As mentioned in Section 1.2.1, 118 haematology/oncology patients with thrombocytopenia were randomized to receive either a pathogen-reduced platelet product (PRT-PLT) ($Z = 1$) or standard reference platelets ($Z = 0$) over a 28-day treatment period (i.e. $C_A = 28$). Events of interest are blood transfusions. The primary outcome is transfusion-based and defined as the 24-hour corrected count increment (CCI). The count increment is the post-transfusion platelet count measured 24 hours after transfusion minus the pre-transfusion platelet count and the corrected

count increment adjusts this number by the dose of platelets transfused and the patients' body surface area (Davis *and others*, 1999).

In Table 2.13 we report on analyses of the Mirasol Study using the copula-based joint model introduced in Section 2.2.1 with Poisson and negative binomial marginal models respectively for the events and the estimation procedures discussed in Section 2.2.4. Generally speaking, the analysis results based on the four estimation procedures are in close agreement. We report results analyzed using the simultaneous estimation procedure in what follows since it is the most efficient one based on simulation studies in Section 2.3. Under a joint model with a Poisson model for the recurrent event process, the estimated coefficient $\widehat{\beta}$ and the corresponding standard error are 0.162 and 0.126 respectively. Recall that this standard error (0.126) is calculated by taking the square root on the diagonal of the matrix $[m\widehat{\mathbb{A}}_1(\widehat{\theta})]^{-1}$ where $\widehat{\mathbb{A}}_1(\widehat{\theta})$ is given in (2.2.7), smaller than its respective robust standard error 0.163, which is calculated by taking square root on the diagonal of the matrix

$$\frac{1}{m}\widehat{\mathbb{A}}_1^{-1}(\widehat{\theta})\widehat{\mathbb{B}}_1(\widehat{\theta})\left[\widehat{\mathbb{A}}_1^{-1}(\widehat{\theta})\right]' ,$$

where

$$\widehat{\mathbb{B}}_1(\widehat{\theta}) = \frac{1}{m}\sum_{i=1}^{m} U_i(\theta)U_i'(\theta)\bigg|_{\theta=\widehat{\theta}} .$$

It is in turn close to the model-based standard error of $\widehat{\beta}$ from the semiparametric negative binomial model (0.163) as reported in Diao *and others* (2013), suggesting the presence of extra-Poisson variation. The likelihood ratio test statistic of $H_0 : \gamma = 0$ (the marginal

Table 2.13: Results from marginal and joint analyses of the transfusion events and the 24 hour corrected count increments (m^2/ℓ) from the Mirasol Study (Cazenave *and others*, 2010).

		Joint Analysis (Poisson)				Joint Analysis (NB)			
		ML	2S	A2S	3S	ML	2S	A2S	3S
α_1	Est.	0.119	0.115	0.115	0.119	0.118	0.114	0.114	0.118
	S.E.	0.016	0.016	0.016	0.016	0.019	0.016	0.016	0.016
α_2	Est.	0.796	0.793	0.793	0.796	0.788	0.788	0.788	0.789
	S.E.	0.050	0.041	0.041	0.041	0.050	0.039	0.039	0.038
β	Est.	0.162	0.212	0.212	0.162	0.165	0.218	0.218	0.166
	S.E.	0.126	0.167	0.167	0.163	0.154	0.165	0.165	0.161
	p-value	0.200	0.205	0.205	0.322	0.283	0.187	0.187	0.302
γ	Est.	-	-	-	-	0.193	0.222	0.222	0.193
	S.E.	-	-	-	-	0.082	0.091	0.091	0.089
	p-value	-	-	-	-	0.019	0.015	0.015	0.031
η_0	Est.	10.072	9.995	10.109	10.072	10.138	9.995	10.172	10.132
	S.E.	0.720	0.881	0.917	0.906	0.728	0.881	0.916	0.906
η_1	Est.	-3.360	-3.355	-3.434	-3.360	-3.366	-3.355	-3.429	-3.364
	S.E.	0.999	1.029	1.067	1.045	1.007	1.029	1.059	1.038
	p-value	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
σ^2	Est.	33.537	33.470	33.548	33.535	34.016	33.470	34.052	33.972
	S.E.	3.766	5.097	5.091	5.090	3.874	5.097	5.209	5.190
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
ρ	Est.	0.628	0.616	0.628	0.628	0.635	0.616	0.636	0.635
	S.E.	0.050	0.068	0.066	0.066	0.049	0.068	0.065	0.065
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
τ	Est.	0.154	0.154	0.154	0.154	0.157	0.155	0.156	0.155
	S.E.	0.030	0.031	0.031	0.031	0.031	0.031	0.031	0.031
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

ML corresponds to the simultaneous estimation procedure; 2S corresponds to the two-stage estimation procedure; A2S corresponds to the asymmetric two-stage estimation procedure; 3S corresponds to the three-stage estimation procedure.

model for the events is a Poisson model) versus $H_A : \gamma > 0$ (the marginal model for the events is a negative binomial model) asymptotically follows a distribution which is a $50 : 50$ mixture of a point mass at zero and a χ^2_1 distribution (Self and Liang, 1987). Under joint model with negative binomial recurrent event process, the resulting test statistic is 10.344 which gives a p−value of $0.5P(\chi^2_1 > 10.344) = 0.0007$. This leads to rejection of the Poisson assumption for the marginal model and supports the use of the joint model with a negative binomial recurrent event process. Thus we report results with the negative binomial margin in what follows. The estimated treatment effect on the event rate is $RR = 1.18$ (95% CI: 0.86, 1.62; $p = 0.306$) reflecting a nonsignificant trend towards increased need for transfusion in the PRT-PLT arm. Regarding the marks, there is a significantly lower response in the PRT-PLT arm compared to the reference arm. We see a lower response by an average of 3.37 (95% CI: -5.40, -1.33; $p = 0.001$) $\times 10^9$ m^2/ℓ. Thus noninferiority of the PRT-PLT product is not successfully demonstrated based on the 24-hour CCI response. The correlation between the marks is quite strong at 0.635, suggesting the importance of accommodating this association. The estimate of Kendall's τ is 0.157 (95% CI: 0.094, 0.220; $p < 0.001$), so there is significant but modest dependence between the CCI at the kth transfusion and the time to the next transfusion. We therefore may not expect an appreciable gain in efficiency in this dataset from the joint modeling using simultaneous estimation procedure.

Figure 2.2 contains plots of the estimated mean number of transfusions from a parametric analysis with a nonhomogeneous baseline Poisson rate function, a semiparametric Andersen-Gill analysis, and a joint analysis based on a Poisson and negative binomial intensity function. There is excellent agreement between the three parametric analyses in

terms of the expected number of transfusions required over the 28 day treatment period. The estimated mean function based on the semiparametric Andersen-Gill model reflects a higher rate during the first two weeks followed by a lower rate in the latter two weeks; this is not exhibited in the parametric models suggesting the need for more flexible piecewise constant baseline rate functions, which will be studied in the future research.

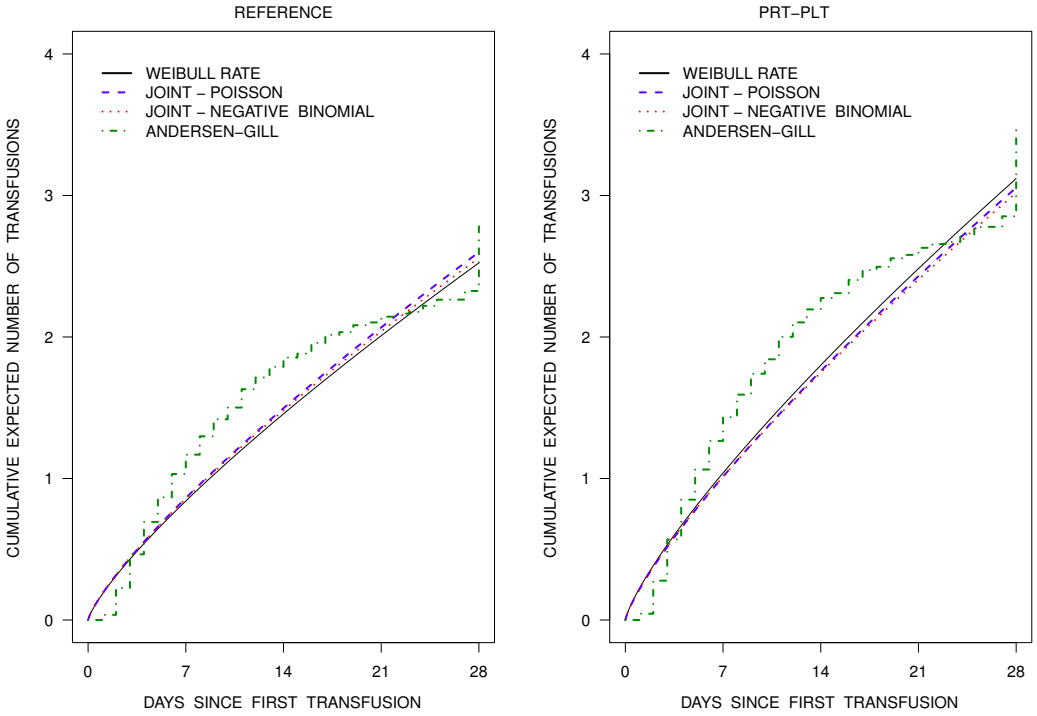

Figure 2.2: Plots of estimated baseline mean function obtained by integrating the marginal rate from a marginal analysis under a Poisson assumption with a parametric Weibull rate and semiparametric Andersen-Gill and from a joint analysis using (2.2.5) with a Poisson and negative binomial parametric marginal rate

2.5 General Remarks

We have described a novel model for marked point processes which incorporates a dependence between continuous marks and the event process through the use of a copula function. The idea of using copula functions to link the marks and the event times has been proposed in the ruin theory literature based on renewal processes (Albrecher and Teugels, 2006; Landriault *and others*, 2012) but these models have different formulations motivated by issues arising in actuarial science. This model is the first we are aware of that address an important problem prevalent in the health research.

We have empirically studied the finite sample behaviour of estimators from joint analyses using the simultaneous, the two-stage, the asymmetric two-stage and the three-stage estimation procedures. We have also studied the efficiency loss of the other three estimation procedures compared to using simultaneous estimation procedure, the effect of misspecifying the marginal intensity for event times under the assumption that the copula function and the marginal model for the marks are correctly specified. We examined the empirical properties of the estimators arising from misspecification of the copula function and correct specification of the marginal models and found, in our limited investigation, only modest impact when Kendall's τ is modest. When the association between the marks and subsequent event times is strong, the copula model should be checked carefully. One attractive avenue for doing this is through model expansion; see Yilmaz and Lawless (2011) for a discussion of this approach in the failure time context.

Chapter 3

Correlated Markov Processes Under Interval Censoring

3.1 Introduction

3.1.1 Overview

Multistate models are used routinely to characterize, identify risk factors for, and make predictions about chronic disease processes (e.g. Hougaard, 1999, 2000). Areas of application are diverse and include health promotion (Kalbfleisch and Lawless, 1985; Cook *and others*, 2002), research on dynamics of infectious disease (Gentleman *and others*, 1994; Sweeting *and others*, 2010), and studies of cognitive function (Tyas *and others*, 2007). Markov and semi-Markov processes are two fundamental classes of models with the former being most widely adopted in settings involving progressive conditions. The considerable

advances in counting process theory in recent years have led to a unification of survival and more general event history methods (Andersen *and others*, 1993; Therneau and Grambsch, 2000; Kalbfleisch and Prentice, 2002; Lawless, 2003; Cook and Lawless, 2007; Aalen *and others*, 2008) and it is now relatively easy to fit parametric and semiparametric models.

Chronic diseases frequently affect multiple organ systems or multiple locations in the body raising the need to accommodate dependencies between related disease processes. Diseases of the eyes are often a consequence of genetic abnormalities or environmental exposures and as a result, visual acuity deteriorates over time in both eyes (Marshall and Jones, 1995). Kidney function naturally declines with age but more rapidly among diabetic individuals (Al-Kateb *and others*, 2008), and rates of damage for the left and right kidneys may be associated within individuals. In psoriatic arthritis, interest lies in modeling changes in joint damage over time and the extent of damage in individual joints is measured on a five-point scale (Rahman *and others*, 1998). Models for the rates at which joints progress through these states are useful to characterize disease progression, identify risk factors, and evaluate therapeutic interventions; interest also lies in modelling progression in many joints simultaneously.

There are a variety of frameworks available for analysis of multiple multistate processes. First, models for two or more multistate processes may be constructed based on the complete intensity functions, which characterize the instantaneous risk of transition between disease states in terms of the full process history (Andersen *and others*, 1993). One may view this as working with an expanded state space defined by all combinations of states from the marginal processes (Ross, 1996). Second, mixed-effect models can be specified in which transitions for the different processes are made independently, conditional on ran-

dom effects (Satten, 1999; Cook *and others*, 2004; Sutradhar and Cook, 2008). Third, standard separate analysis of each process is justified under a working independence assumption (Lee and Kim, 1998) and a robust covariance matrix can be obtained to ensure valid simultaneous inference in the spirit of generalized estimating equations (Liang and Zeger, 1986) and marginal methods of multivariate survival data (Wei *and others*, 1989).

A natural goal in the analysis of multiple multistate processes is to provide simple estimates of transition rates and related covariate effects for each component process, which have a straightforward "marginal" interpretation. Estimates of this sort do not arise naturally from the aforementioned approaches to analysis of several processes, except the one based on a working independence assumption. It may, however, be important to parametrically model the association between processes to improve efficiency and advance scientific understanding about the relation between the processes under study. We develop a joint model for multiple multistate processes based on copula functions (Joe, 1997; Nelsen, 2006), which yields straightforward inferences regarding the marginal processes while parameterizing the association.

The remainder of this chapter is organized as follows. In Section 3.2 we define notation and formulate a joint model for multiple multistate processes. In Section 3.3 we discuss methods for estimation and statistical inference. We focus on the setting in which the transition times are interval-censored since disease processes are often only observed at periodic assessment times. Simulation studies and an application to data on joint damage in psoriatic arthritis are presented in Section 3.4, and general remarks are given in Section 3.5. We begin first, however, with a review of composite likelihood.

3.1.2 Review of Composite Likelihood

Composite likelihoods are based on partial specification of the full likelihood (Besag, 1974; Lindsay, 1988; Cox and Reid, 2004; Lindsay *and others*, 2011). Let $\{\mathcal{A}_1, \ldots, \mathcal{A}_Q\}$ denote a set of Q user-selected marginal or conditional events. Component likelihoods $L_q(\psi) \propto f(t \in \mathcal{A}_q; \psi)$, indexed by a vector of parameters ψ, are associated with $\mathcal{A}_q$ and a composite likelihood is simply a product of the component likelihoods,

$$CL(\psi) = \prod_{q=1}^{Q} L_q(\psi) \; . \tag{3.1.1}$$

When the selected events are not independent, we apply a "working independence assumption" and simply multiply the component likelihoods together as in (3.1.1).

Since each component likelihood is a true likelihood in some context, it has some of the features of an ordinary likelihood; see Lindsay (1988) and Molenberghs and Verbeke (2005) for the asymptotic theory. Under mild regularity conditions, the component score functions satisfy

$$\mathrm{E}(\partial \log L_q(\psi)/\partial\psi) = 0 \; , \quad q = 1, \ldots, Q,$$

and from (3.1.1) it is apparent that the composite score $\partial \log CL(\psi)/\partial\psi$ is simply the summation of the component score functions, implying under regularity conditions,

$$\mathrm{E}(\partial \log CL(\psi)/\partial\psi) = 0 \; .$$

We denote $CL_i(\psi)$ as the composite likelihood for the ith individual. With a sample of m

independent individuals, the overall composite likelihood is $\prod_{i=1}^{m} CL_i(\psi)$ and a consistent estimator $\widehat{\psi}$ is obtained by solving

$$\sum_{i=1}^{m} \partial \log CL_i(\psi)/\partial\psi = 0 \ .$$

Moreover,

$$\sqrt{m}(\widehat{\psi} - \psi) \xrightarrow{d} N(0, \mathbb{D}^{-1}(\psi)\mathbb{B}(\psi)[\mathbb{D}^{-1}(\psi)]') \ , \text{ as } m \to \infty, \tag{3.1.2}$$

where

$$\begin{aligned} \mathbb{D}(\psi) &= -\mathrm{E}\left[\frac{\partial^2 \log CL_i(\psi)}{\partial\psi\partial\psi'}\right], \\ \mathbb{B}(\psi) &= \mathrm{E}\left[\frac{\partial \log CL_i(\psi)}{\partial\psi}\frac{\partial \log CL_i(\psi)}{\partial\psi'}\right]. \end{aligned}$$

In the analysis of a particular dataset, standard errors are estimated based on this result by replacing the expectations in (3.1.2) with their empirical counterparts and evaluating at the estimate $\widehat{\psi}$.

The natural question is how to select $\{\mathcal{A}_1, \ldots, \mathcal{A}_Q\}$ to construct the composite likelihood. One approach is to construct the composite likelihood from low-dimensional marginal or conditional densities; this is called the "construction method". Alternatively, a composite likelihood can be constructed by omitting particular terms for a full likelihood; this is referred to as "omission method" (Varin, 2008). The general guideline for both the construction and the omission method is that the parts kept in the composite likelihood

should be informative, easily computed and contain parameters of interest; in contrast, the parts omitted are usually hard to evaluate, not very informative, or pose a significant computational burden. Both approaches invoke a series of working independence assumptions under which we can write down a new, more convenient composite likelihood.

3.2 Multiple Multistate Processes

Recall that we defined notation for a single multistate process in Section 1.1.2 and denoted $p \times 1$ vector of time-independent covariates by Z. Consider a disease process in which dam-

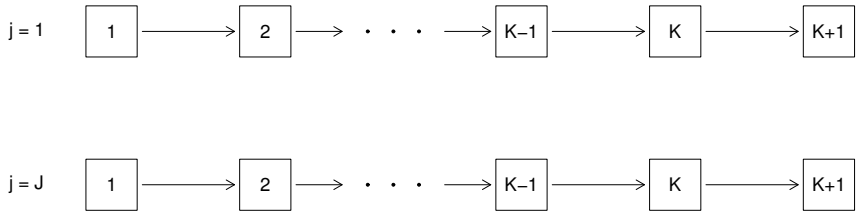

Figure 3.1: State space diagram for multivariate multistate processes

age may occur in J similar organs of the affected individuals as depicted in Figure 3.1. Examples involving paired organs include left and right eyes, kidneys, or lungs, where $J = 2$, or more generally joints at risk of damage in patients with arthritis. Let T_{jk} denote the time of a $k \to k+1$ transition for process j, $k = 1, \cdots, K$, where $0 < T_{j1} < T_{j2} < \cdots < T_{jK}$, $j = 1, \ldots, J$, $T_j = (T_{j1}, \ldots, T_{jK})'$, and $T = (T_1', \ldots, T_J')'$. Let $(T_{1K}, \ldots, T_{JK})'$ denote the vector of absorption times for the J processes. Define $T_{j,-K} = (T_{j1}, \ldots, T_{j,K-1})'$, $T_{-j,k} = (T_{1k}, \ldots, T_{j-1,k}, T_{j+1,k}, \ldots, T_{Jk})'$ and $T_{-j,-K} = (T'_{-j,1}, \ldots, T'_{-j,K-1})'$. We similarly define notation t_j, t, $t_{j,-K}$, $t_{-j,k}$ and $t_{-j,-K}$. A fully-specified multiple multistate model

requires a complete specification of the joint density of all transition times given covariate $Z = z$, $f(t|z)$. This can be written by decomposing the full density into a product of conditional and unconditional densities, and working (conditional) independence assumptions can be made to avoid specifying (conditional) dependence structures of secondary interest. These conditional independence assumptions lead to some simplifications and motivate our use of composite likelihood. There are many ways to decompose the joint density, with different decompositions and working independence assumptions serving to address different research objectives.

Our first goal is to model each component process in a way that is similar to that of a single multistate process as described in Section 1.1.2. Specifically, we want each component process to be modeled by a Markov process with intensities of the form (1.1.2). Introducing a subscript j before every subscript k, the joint density of $T_j = (T_{j1}, \ldots, T_{jK})'$ given $Z = z$ has the form

$$f_j(t_j|z;\theta_j) \;=\; \prod_{k=1}^{K}\left[\lambda_{jk}(t_{jk}|z;\theta_{jk})\exp\left(-\int_{t_{j,k-1}}^{t_{jk}}\lambda_{jk}(u|z;\theta_{jk})du\right)\right], \qquad (3.2.1)$$

where $0 = t_{j0} < t_{j1} < \ldots < t_{jK}$ for $j = 1, \ldots, J$, $\theta_{jk} = (\alpha'_{jk}, \beta'_{jk})'$ and $\theta_j = (\theta'_{j1}, \ldots, \theta'_{jK})'$ (Andersen *and others*, 1993). In this case, each process preserves the Markov property and covariate effects retain their interpretation as multiplicative effects on marginal intensities. In what follows, an extra subscript is introduced to terms defined in Section 1.1.2 to denote the counterpart specific to process j; we restrict attention, however, to "cluster-level" covariates, $Z_j = Z$, common to all processes and representing, for example, a genetic marker, sex or treatment. Our second goal is to parameterize the association between

processes which we can do by representing the joint survivor function of the absorption times $(T_{1K}, \ldots, T_{JK})'$ given $Z = z$ as

$$P(T_{1K} \geq t_{1K}, \ldots, T_{JK} \geq t_{JK}|z;\psi) = \mathcal{C}(\mathcal{F}_{1K}(t_{1K}|z;\theta_1), \ldots, \mathcal{F}_{JK}(t_{JK}|z;\theta_J);\phi) \ , \quad (3.2.2)$$

(Nelsen, 2006; Patton, 2006), where $\mathcal{C}(\cdot;\phi)$ is a multivariate copula function with association parameters ϕ, $\mathcal{F}_{jK}(t_{jK}|z;\theta_j)$ is the marginal survivor function of the absorption time in process j, $\theta = (\theta_1', \ldots, \theta_J')'$ and $\psi = (\theta', \phi')'$. If process j is Markov, the survivor function for T_{jk} is obtained as the complement of the $[1, K+1]$ entry of (1.1.1) where $s = 0$.

To ensure the satisfication of these two goals, we decompose the joint density $f(t|z;\psi)$ in a particular way and make "working" conditional independence assumptions about the dependence relations of little interest. First, we decompose the full joint density $f(t|z;\psi)$ as

$$f(t|z;\psi) \ = \ f(t_{1,-K}, \ldots, t_{J,-K}|t_{1K}, \ldots, t_{JK}, z;\psi) \cdot f(t_{1K}, \ldots, t_{JK}|z;\psi) \ , \quad (3.2.3)$$

which can be rewritten as

$$f(t|z;\psi) \ = \ \prod_{j=1}^{J} f(t_{j,-K}|t_{1K}, \ldots, t_{JK}, z;\psi) \cdot f(t_{1K}, \ldots, t_{JK}|z;\psi) \ , \quad (3.2.4)$$

under the first working conditional independence assumption

A.3.1 $\quad T_{j,-K} \perp T_{-j,-K}|(T_{1K}, \ldots, T_{JK}, Z')'$,

where we write $Y_1 \perp Y_2 | Y_3$ to denote settings in which

$$f_{Y_1,Y_2|Y_3}(y_1, y_2|y_3) = f_{Y_1|Y_3}(y_1|y_3) f_{Y_2|Y_3}(y_2|y_3)$$

for random vectors Y_1, Y_2 and Y_3. This assumption states that the dependence between processes is governed by the associations between absorption times and intermediate transition times are independent between processes given the absorption times for all processes. Second, expression (3.2.4) can then be further simplified to

$$f(t|z;\psi) \;=\; \prod_{j=1}^{J} f(t_{j,-K}|t_{jK}, z; \theta_j) \cdot f(t_{1K}, \ldots, t_{JK}|z;\psi) \;, \tag{3.2.5}$$

by making the second working conditional independence assumption

A.3.2 $\quad T_{j,-K} \perp T_{-j,K} | (T_{jK}, Z')'$.

This assumption states that the intermediate transition times for a particular process are conditionally independent of the absorption times for other processes given its own absorption time. By (3.2.2) and (3.2.5), the joint density $f(t|z;\psi)$ can then be expressed as

$$f(t|z;\psi) \;=\; \prod_{j=1}^{J} f_j(t_j|z;\theta_j) \cdot c(\mathcal{F}_{1K}(t_{1K}|z;\theta_1), \ldots, \mathcal{F}_{JK}(t_{JK}|z;\theta_J);\phi) \;, \tag{3.2.6}$$

where the first J components are marginal density functions which correspond to marginal models (3.2.1), and the last component is a copula density function $c(\cdot)$ of the copula $\mathcal{C}(\cdot)$ in (3.2.2) governing the absorption time distribution.

Some conditional dependence structures are left unspecified by making the working conditional independence assumptions A.3.1 and A.3.2, so (3.2.6) is only a partial specification of the full likelihood (3.2.3). As such it can be characterized as a composite likelihood for a fully observed multiple multistate processes. The working independence approach of Lee and Kim (1998) involving separate marginal analyses can be cast in this framework. They require their multiple multistate model to have the first feature only, and do not model the dependence structure between processes. Thus (3.2.1) is a composite likelihood under working independence assumptions between processes. We also remark that, in the case $J = 2$ and $K = 2$, our model can be also justified by a vine copula decomposition (Joe, 1996; Bedford and Cooke, 2001; Bedford and Cooke, 2002; Aas and Berg, 2009; Aas *and others*, 2009). Figure 3.2 shows the decomposition specification of the joint density

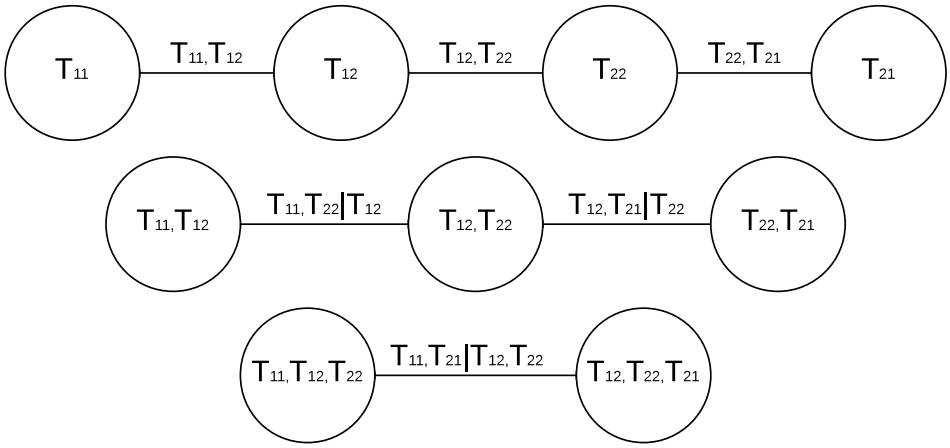

Figure 3.2: A D-vine with four variables

$f(t|z;\psi)$ according to a D-vine (Kurowicka and Cooke, 2005). Each edge in Figure 3.2

corresponds to a pair-copula (conditional) density, e.g., the edge $T_{11}, T_{22}|T_{12}$ corresponds to the conditional copula density $c(\mathcal{F}(t_{11}|t_{12}, z; \theta_1), \mathcal{F}(t_{22}|t_{12}, z; \theta, \phi); \phi_2)$. The joint density of T_{j1}, T_{j2} is given by (3.2.1), which is not induced by a copula function, for $j = 1, 2$. The joint density $f(t|z; \psi)$ corresponding to the D-vine illustrated in Figure 3.2 may be written as

$$\begin{aligned} f(t|z;\psi) &= f_1(t_{11}, t_{12}|z;\theta_1) \cdot c(\mathcal{F}_{12}(t_{12}|z;\theta_1), \mathcal{F}_{22}(t_{22}|z;\theta_2); \phi) \cdot f_2(t_{21}, t_{22}|z;\theta_2) \\ &\cdot\ c(\mathcal{F}(t_{11}|t_{12}, z;\theta_1), \mathcal{F}(t_{22}|t_{12}, z;\theta, \phi); \phi_2) \\ &\cdot\ c(\mathcal{F}(t_{12}|t_{22}, z;\theta, \phi), \mathcal{F}(t_{21}|t_{22}, z;\theta_2); \phi_3) \\ &\cdot\ c(\mathcal{F}(t_{11}|t_{12}, t_{22}, z;\theta, \phi, \phi_2), \mathcal{F}(t_{21}|t_{12}, t_{22}, z;\theta, \phi, \phi_3); \phi_4)\ . \end{aligned} \tag{3.2.7}$$

Conditional independence assumptions are commonly used in the vine copula framework to reduce the number of pair copulas in the decomposition and hence simplify model construction. The working conditional independence assumptions A.3.1 and A.3.2 lead one to set the last three terms of (3.2.7) equal to one, simplifying (3.2.7) to (3.2.6) with $J = 2$.

The marginal processes are compatible with those of a single multistate process and each component process in (3.2.6) yields parameters with a straightforward interpretation in terms of transition rates and covariate effects. However, our model features a parameterized association structure and hence a measure of association can be readily calculated based on the functional form of the copula $\mathcal{C}(\cdot)$ and association parameter ϕ (Genest and MacKay, 1986). In addition, our working assumptions are weaker than those of complete independence, and may lead to more efficient estimation. Under (3.2.6), one can separately

specify the marginal models for each process and the model for the association among the processes, thereby avoiding specification of the conditional dependence structures of little interest. Many options exist for specification of the marginal models and association models, making (3.2.6) quite flexible.

3.3 Estimation and Inference

3.3.1 Notation for Interval-Censored Data

When individuals are assessed intermittently, the times of transitions between states are subject to interval censoring. This is routinely the case when the processes relate to damage of internal organs. For notational convenience, we restrict attention to the case in which all processes are assessed at the same $M(>1)$ time points denoted by $v_0 < v_1 < \cdots < v_M < v_{M+1}$, where $v_0 = 0$, $v_{M+1} = \infty$. Let $V_1, \ldots, V_M$ be a sequence of corresponding random variables with joint density $f_{V_1,\ldots,V_M}(v_1, \ldots, v_M; \nu)$ indexed by ν. We assume that $\zeta_j(v_0) = 1$ with probability one, $j = 1, \ldots, J$. The data available then consist of the covariate vector, the inspection times, and the states occupied at those times for each process:

$$\{(v_m, \zeta_1(v_m), \ldots, \zeta_J(v_m));\ m = 0, 1, \ldots, M,\ Z\}\ . \tag{3.3.1}$$

It is also helpful to define random variables recording the number of "transitions" of a particular type. Let $N^m_{jk\ell} = I(\zeta_j(v_{m-1}) = k, \zeta_j(v_m) = \ell)$ indicate that process j occupies

state k at assessment time v_{m-1} and state ℓ at v_m, $k \leq \ell$. In this case,

$$\left\{(v_m, N^m_{jk\ell},\ \ell = k, \ldots, K+1,\ k = 1, \ldots, K,\ j = 1, \ldots, J);\ m = 1, \ldots, M,\ Z\right\}$$

contains exactly the same information as (3.3.1). The data can also be expressed as the left and right censoring time of transition times which leads to the third way of expressing data:

$$\{T_{jk} \in (L_{jk}, R_{jk}];\ k = 1, \ldots, K,\ j = 1, \ldots, J,\ Z\}\ ,$$

where $L_{jk} = v_{M(T_{jk})}$ and $R_{jk} = v_{M(T_{jk})+1}$. The observed history of process j at time t is denoted $H_j(t) = \{(v_m, \zeta_j(v_m));\ m = 1, \ldots, M(t), Z\}$ where $M(t) = argmax_m\{v_m < t\}$.

3.3.2 Composite Likelihood Construction

We assume that the parameter ν associated with the inspection time process in

$$f_{V_1,\ldots,V_M}(v_1, \ldots, v_M; \nu)$$

is functionally independent of the parameter of interest ψ. Under the conditions of Grüger *and others* (1991) we proceed to construct the full likelihood arising from intermittent inspection of a multiple multistate process as if the inspection times are fixed and hence in what follows we restrict attention to the likelihood

$$L(\psi) = P(T_{jk} \in (l_{jk}, r_{jk}];\ k = 1, \ldots, K;\ j = 1, \ldots, J | z, v_1, \ldots, v_M; \psi)\ . \qquad (3.3.2)$$

The likelihood in (3.3.2) is obtained by computing $J \times K$ dimensional integrals over the fully specified $f(t|z;\psi)$ in (3.2.3). For example, in the case $J = K = 2$, four dimensional integration on $f(t|z;\psi)$ in (3.2.7) is required. The likelihood involves computationally-demanding high-dimensional integration when J or K is large. Use of composite likelihood enables some simplification in model specification and increases robustness of the model structure.

Lee and Kim (1998) discuss the case when interest lies only in estimation of marginal parameters. If a working independence assumption among processes is reasonable, the estimation problem simplifies to one which has been addressed in the literature (Kalbfleisch and Lawless, 1985). Since process j is Markov, the composite likelihood of process j is

$$L_j(\theta_j) \;=\; \prod_{m=1}^{M}\prod_{k=1}^{K}\prod_{\ell=k}^{K+1} P_{jk\ell}(v_{m-1}, v_m|z;\theta_j)^{n^m_{jk\ell}}\,, \tag{3.3.3}$$

where the transition probability is defined in Section 1.1.2. A Fisher-scoring or Newton-Raphson algorithm can be used for estimation, and robust variance estimation is in the spirit of (3.1.2).

If both marginal and association parameters are of interest, we make the following working conditional independence assumptions:

A.3.3 $T_{j,-K} \perp T_{-j,-K} | (T_{1K} \in (L_{1K}, R_{1K}], \ldots, T_{JK} \in (L_{JK}, R_{JK}], Z')'$,

A.3.4 $T_{j,-K} \perp T_{-j,K} | (T_{jK} \in (L_{jK}, R_{jK}], Z')'$.

These are slightly different from assumptions A.3.1 and A.3.2, but enable one to write

down the composite likelihood arising from intermittent inspection:

$$\begin{aligned} CL_1(\psi) &= \prod_{j=1}^{J} P\left(T_{jk} \in (l_{jk}, r_{jk}], k = 1, \ldots, K-1 | T_{jK} \in (l_{jK}, r_{jK}], z; \theta_j\right) \\ &\times P\left(T_{jK} \in (l_{jK}, r_{jK}], j = 1, \ldots, J | z; \psi\right) \ . \end{aligned} \tag{3.3.4}$$

In (3.3.4),

$$\begin{aligned} &P(T_{jK} \in (l_{jK}, r_{jK}], j = 1, \ldots, J | z; \psi) \\ &\qquad = \sum_{a \in A} (-1)^{d_a} \cdot \mathcal{C}(\mathcal{F}_{1K}(a_{1K} | z; \theta_1), \ldots, \mathcal{F}_{JK}(a_{JK} | z; \theta_J); \phi) \ , \end{aligned} \tag{3.3.5}$$

where $a = (a_{1K}, \ldots, a_{JK})'$, $A = \{a : a_{jK} \in \{l_{jK}, r_{jK}\}, j = 1, \ldots, J\}$, $d_a = \sum_{j=1}^{J} I(a_{jK} = r_{jK})$, and (3.3.5) involves a summation of 2^K items. Note that since $\{T_{jk} \in (L_{jk}, R_{jk}];\ k = 1, \ldots, K,\ j = 1, \ldots, J,\ Z\}$ contains the same information as $\{(v_m, N^m_{jk\ell},\ \ell = k, \ldots, K+1,\ k = 1, \ldots, K,\ j = 1, \ldots, J);\ m = 1, \ldots, M,\ Z\}$, $P(T_{jk} \in (l_{jk}, r_{jk}], k = 1, \ldots, K | z; \theta_j)$ is equal to the marginal likelihood $L_j(\theta_j)$ of process j in (3.3.3). The composite likelihood (3.3.4) can therefore be written as

$$\begin{aligned} CL_1(\psi) = \prod_{j=1}^{J} &\frac{L_j(\theta_j)}{\mathcal{F}_{jK}(l_{jK} | z; \theta_j) - \mathcal{F}_{jK}(r_{jK} | z; \theta_j)} \\ &\times P\left(T_{jK} \in (l_{jK}, r_{jK}], j = 1, \ldots, J | z; \psi\right) \ . \end{aligned} \tag{3.3.6}$$

A composite likelihood can alternatively be built using the "construction method" by using J marginal likelihoods to obtain marginal estimates and using the joint probability of the J absorption times to estimate the association parameters. The composite likelihood is

then

$$CL_2(\psi) \;=\; \prod_{j=1}^{J} L_j(\theta_j) \times P\left(T_{jK} \in (l_{jK}, r_{jK}], j = 1, \ldots, J | z; \psi\right) \;, \qquad (3.3.7)$$

in which the $J + 1$ components are analogous to those in (3.2.6).

Composite likelihoods based on (3.3.3), (3.3.6) and (3.3.7) represent simplifications to the full likelihood (3.3.2) and so may lead to some loss of efficiency, but their use introduces robustness and significant computational advantages. The composite likelihood (3.3.3) is obtained under the strongest working independence assumption so it does not provide estimation of any association parameters and the estimators of the marginal parameters obtained based on (3.3.3) would be expected to be the least efficient. The composite likelihoods in (3.3.6) and (3.3.7) are constructed based on different ideas but have similar forms, and both successfully avoid the need for high-dimensional integration.

3.3.3 Two-stage Estimation

A two-stage estimation procedure (Shih and Louis, 1995; Newey and McFadden, 1994) is possible with the formulation described due to use of the copula function for the association model. In the first stage, an estimate of the marginal parameters θ_j is obtained for each process j using the marginal likelihood (3.3.3), $j = 1, \ldots, J$. In the second stage, the estimate $\widehat{\theta}$ is inserted into composite likelihood function $CL_1(\psi)$ in (3.3.6) or $CL_2(\psi)$ in (3.3.7), and the composite likelihood is then maximized with respect to ϕ to obtain an estimate $\widetilde{\phi}$. With regard to the two composite likelihoods (3.3.6) and (3.3.7), only

$P\left(T_{jK} \in (l_{jK}, r_{jK}], j = 1, \ldots, J | z; \psi\right)$ in (3.3.5) contains the association parameters, so this is the objective function in the second stage. Shih and Louis (1995) develop the asymptotic distribution for the case when the association parameter is a scalar. The corresponding asymptotic results for a vector of association parameters are given in Newey and McFadden (1994).

3.4 Simulation Studies and Illustration

3.4.1 Design and Analysis of Simulation Studies

The simulation studies conducted here are designed to assess the finite sample properties of estimators from the various composite likelihoods. We consider two processes ($J = 2$) with three states each ($K = 2$), where state 1 represents a "normal" condition, state 2 represents "abnormal", and state 3 represents the absorbing state of "organ damage"; we assume that all subjects start from state 1 for both processes. We consider one Bernoulli covariate Z, with $P(Z = 1) = 0.5$. We generate data from the joint density of the form (3.2.7) where the marginal progress adopts a model with progressive time-homogeneous Markov margins with transition intensity $\lambda_{jk}(t|z; \theta_{jk}) = \alpha_{jk} \exp(z\beta_{jk})$ for $j, k = 1, 2$ and bivariate Clayton copulas for all the conditional or unconditional densities. The Clayton copula, a member of the Archimedian family, has the form (1.1.5) and is frequently used in survival analysis due to its connection with the gamma frailty model (Joe, 1997). We assume here that there are $M = 10$ common inspection times evenly spaced over the interval $(0, 1]$, giving visit times $v_m = 0.1 \times m$ for $m = 1, \ldots, 10$.

We assume that the two processes have the same margins, as would be the case with clustered processes, so that $\alpha_{1k} = \alpha_{2k}$ and $\beta_{1k} = \beta_{2k}$ for $k = 1, 2$. We set $\beta_{j1} = \log(1.25)$ to reflect a mild increase of the risk of transition from state 1 to 2 when $Z = 1$ and set $\beta_{j2} = \log(1.4)$ to reflect a moderate effect on increasing the risk of transition from state 2 to 3 in both processes. The baseline transition intensities α_{jk} for $j, k = 1, 2$ are set under the following constraints: (i) the baseline transition rate out of state 2 is 1.5 times of that out of state 1, i.e. $\alpha_{j2} = 1.5\alpha_{j1}$ for $j = 1, 2$; (ii) the probability of both processes being in state 3 by time 1 is 0.4 in the control group, so

$$P\left(\zeta_1(1) = 3, \zeta_2(1) = 3 | Z = 0\right) = P\left(T_{12} \leq 1, T_{22} \leq 1 | Z = 0\right) = 0.4.$$

These constraints give $\alpha_{j1} = 1.8148$ and $\alpha_{j2} = 2.7221$. The association parameters of the Clayton copulas in (3.2.7) are set to $(\phi, \phi_2, \phi_3, \phi_4)' = (3, 8, 2, 4.6667)'$ giving Kendall's τ's of $(0.6, 0.8, 0.5, 0.7)'$ respectively (Nelsen, 2006). Two thousand samples of $m = 1000$ individuals each are simulated.

For each dataset, analyses are carried out based on the composite likelihoods (3.3.6) and (3.3.7), and two-stage estimation used the composite likelihood to estimate ψ. The empirical bias (BIAS), average asymptotic (large sample) standard error (ASE), empirical standard error (ESE) and empirical coverage probability (ECP) are evaluated for all parameter estimates and reported in Table 3.1. Analyses are carried out using `R`.

As expected from the asymptotic theory of composite likelihood, the empirical biases are all very small for both the estimates of marginal parameters and association parameters using all methods. The ASE and ESE are consistent with each other and the ECP

Table 3.1: Frequency properties of estimators using three estimation methods (1000 individuals per sample; 2000 simulated samples)

		Composite Likelihood (3.3.6)				Composite Likelihood (3.3.7)			
	True Value	BIAS	ASE	ESE	ECP	BIAS	ASE	ESE	ECP
$\log(\alpha_{11})$	0.5960	0.0006	0.0449	0.0451	0.9475	0.0007	0.0466	0.0469	0.9455
$\log(\alpha_{12})$	1.0014	-0.0007	0.0560	0.0553	0.9540	-0.0009	0.0558	0.0550	0.9535
$\log(\alpha_{21})$	0.5960	0.0001	0.0455	0.0456	0.9485	0.0002	0.0469	0.0471	0.9515
$\log(\alpha_{22})$	1.0014	0.0016	0.0547	0.0527	0.9545	0.0015	0.0552	0.0536	0.9495
β_{11}	0.2231	-0.0028	0.0559	0.0565	0.9485	-0.0032	0.0612	0.0618	0.9490
β_{12}	0.3365	0.0025	0.0775	0.0765	0.9550	0.0023	0.0760	0.0749	0.9530
β_{21}	0.2231	-0.0011	0.0582	0.0588	0.9495	-0.0014	0.0622	0.0626	0.9465
β_{22}	0.3365	-0.0032	0.0744	0.0726	0.9535	-0.0035	0.0747	0.0732	0.9495
$\log(\phi)$	1.0986	0.0022	0.0609	0.0618	0.9455	0.0017	0.0610	0.0620	0.9410
		Two-stage Estimation				Relative Efficiency			
	True Value	BIAS	ASE	ESE	ECP	RE^1	RE^2	RE^3	RE^4
$\log(\alpha_{11})$	0.5960	0.0006	0.0489	0.0495	0.9465	0.9639	0.9603	0.9184	0.9102
$\log(\alpha_{12})$	1.0014	-0.0007	0.0561	0.0553	0.9530	1.0040	1.0052	0.9985	1.0011
$\log(\alpha_{21})$	0.5960	0.0005	0.0489	0.0491	0.9520	0.9702	0.9690	0.9297	0.9298
$\log(\alpha_{22})$	1.0014	0.0019	0.0561	0.0545	0.9495	0.9894	0.9830	0.9747	0.9666
β_{11}	0.2231	-0.0031	0.0680	0.0689	0.9490	0.9140	0.9130	0.8219	0.8196
β_{12}	0.3365	0.0021	0.0761	0.0749	0.9485	1.0198	1.0219	1.0183	1.0212
β_{21}	0.2231	-0.0016	0.0680	0.0685	0.9495	0.9368	0.9384	0.8561	0.8577
β_{22}	0.3365	-0.0035	0.0760	0.0750	0.9495	0.9972	0.9916	0.9793	0.9677
$\log(\phi)$	1.0986	0.0001	0.0610	0.0619	0.9450	0.9977	0.9964	0.9977	0.9985

RE^1-RE^4 respectively denote relative efficiency of composite likelihood (3.3.7) v.s. composite likelihood (3.3.6) based on ASE, composite likelihood (3.3.7) v.s. composite likelihood (3.3.6) based on ESE, two-stage estimation v.s. composite likelihood (3.3.6) based on ASE, two-stage estimation v.s. composite likelihood (3.3.6) based on ESE.

are all very close to the nominal confidence level of 95%, suggesting that the methods proposed provide a valid basis for inference. The estimate of the association parameter by the two-stage procedure is of comparable efficiencies, however, there is more substantial loss for the estimates of the marginal parameters compared to those using simultaneous estimation procedure based on the composite likelihood (3.3.6) and (3.3.7). We remark as well that estimates of the marginal parameters with respect to the transition from mild to intermediate state obtained from the composite likelihood (3.3.6) are slightly more efficient than those from the composite likelihood (3.3.7).

3.4.2 Joint Damage Progression in Individuals with Arthritis

Patients with arthritic conditions are at risk of developing debilitating joint damage and it is common to use the total joint count as a global summary of damage. We consider data from the University of Toronto Psoriatic Arthritis (PsA) Clinic which are comprised of several hundred patients as mentioned in Section 1.2.2. We consider the Human Leukocyte Antigen (HLA) B27 as a covariate Z, since it is an inherited genetic marker associated with a number of related rheumatic diseases including ankylosing spondylosis. The severity of joint damage is recorded in a five-point modified Steinbrocker scale as mentioned earlier. The primary interest is to evaluate the covariate effect on the joint damage progression of the left and right sacroiliac (SI) joints. We restrict attention to data, as of December 1, 2007, for 640 patients with complete covariate information (HLA B27) and use data obtained at all assessment times that the modified Steinbrocker score could be assessed. In our analysis, we combine the original states 2 and 3 to form a state representing mild joint

damage, and states 4 and 5 as a state denoting moderate to severe damage. We allow the covariate HLA B27 to have different effects for the left and right SI joints, and also allow different baseline transition rates for both transition into the mild-moderate state and that into severe state. We conduct joint analysis of the PsA data to estimate the effect of the HLA B27 and association parameters between the two processes using the three methods we proposed in Section 3.2.

The results are summarized in Table 3.2. The upper third of the table gives estimates pertaining to baseline transition rates, the middle third is of the regression coefficients, and the lower third is for the association parameter. The estimated regression coefficients suggest that HLA B27 increases the risk of transition into the mild-moderate and severe damage states. Based on analysis using the composite likelihood (3.3.6), for example, individuals being HLA B27 positive have a significantly higher transition rate to mild-moderate damage on the left SI joint (RR = 1.28, 95% CI: 1.16–1.41, p-value < 0.001) and a significantly higher rate of progression to the state of severe damage on that side (RR = 1.68, 95% CI: 1.33–2.03, p-value < 0.001). On the right SI joint, being B27 positive is associated with an increased risk of mild-moderate damage (RR = 1.16, 95% CI: 1.11–1.21, p-value < 0.001) and there is evidence of a more rapid onset of severe damage (RR = 1.48, 95% CI: 1.07–1.90, p-value < 0.001). The estimate of Kendall's τ based on (3.3.6) was $\widehat{\tau} = 0.82$ (95% CI: 0.77–0.87, p-value < 0.001) corresponding to significant evidence of a very strong association in progression times to severe damage.

One of the New York criteria (Moll and Wright, 1973) for diagnosis of ankylosing spondylitis is satisfied if both the left and right SI joints are in state 3 (i.e. $(\zeta_1(t), \zeta_2(t)) = (3, 3)$). The joint model is particularly appealing here then, since it permits prediction of

Table 3.2: Joint analysis of progression in the left and right sacroiliac joints in psoriatic arthritis (PsA) with the covariate HLA B27 and allowing different parameters in the two processes

	Independence Assumption†			Composite (3.3.6)		Composite (3.3.7)	
	EST.	Naive S.E.	Robust S.E.	EST.	S.E.	EST.	S.E.
Baseline Intensity							
Left-side							
$\log(\alpha_{11})$	-0.215	0.057	0.035	-0.182	0.015	-0.196	0.028
$\log(\alpha_{12})$	-0.977	0.105	0.187	-0.788	0.027	-0.944	0.098
Right-side							
$\log(\alpha_{21})$	-0.005	0.007	0.003	0.009	0.001	0.019	0.003
$\log(\alpha_{22})$	-0.903	0.097	0.136	-0.828	0.049	-0.978	0.093
Coefficients							
Left-side							
β_{11}	0.265	0.131	0.440	0.249	0.049	0.291	0.081
β_{12}	0.649	0.191	0.835	0.519	0.107	0.568	0.251
Right-side							
β_{21}	0.176	0.106	0.306	0.149	0.022	0.173	0.211
β_{22}	0.398	0.192	0.728	0.395	0.143	0.428	0.419
Association Parameter							
$\log(\phi)$	-	-	-	2.188	0.161	2.288	0.137

† The marginal estimates under working independence assumption are plugged into the composite likelihood to obtain $\log(\widehat{\phi}) = 2.239$ (S.E. $= 0.246$).

time to the development of ankylosing spondylitis in PsA patients. Figure 3.3 gives plots

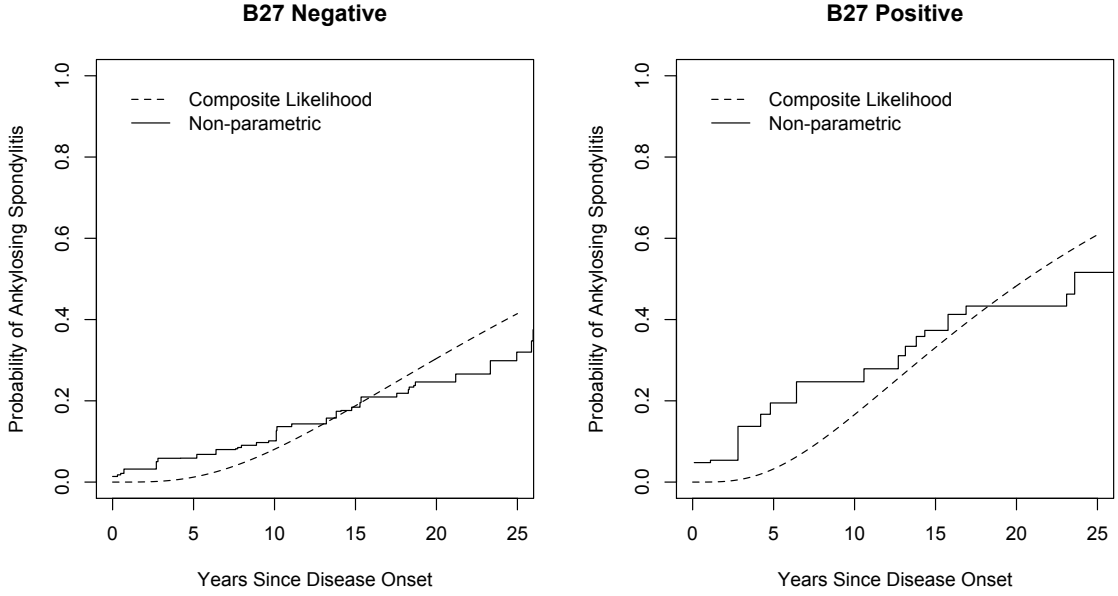

Figure 3.3: Plots of the cumulative probability of ankylosing spondylitis by B27 status according to the composite likelihood (3.3.6) analysis from the joint model and based on nonparametric estimate of Gentleman and Vandal (2002); for the fitted parametric model the estimated joint probability is $P(\zeta_1(t) = \zeta_2(t) = 3 | \zeta_1(0) = \zeta_2(0) = 1; \widehat{\psi})$

of the cumulative probability of ankylosing spondylitis by this criteria based on the fitted model using the composite likelihood (3.3.6) as an illustration. The left-hand panel shows this probability estimated for individuals who are B27 negative and the right-hand panel is for those who are B27 positive. Overlaid on these plots are estimates obtained by the graph theoretic approach to nonparametric estimation of bivariate failure time distributions with interval-censored data developed in Gentleman and Vandal (2002) and implemented in the `R` package `MLEcens` (Maathuis, 2013). There is reasonable agreement between the estimates obtained with the joint model and the nonparametric estimates.

The joint model is also useful for examining how risks of damage in a particular SI joint depend on the damage state of the contralateral SI joint. For example if we consider the risk of the left SI joint exhibiting severe damage since onset, we can consider three scenarios: the right SI joint has developed

i) no damage by 10 years,

ii) mild-moderate damage by 10 years, and

iii) severe damage by 10 years.

The fitted model yields estimates as $P(\zeta_1(t) = 3|\zeta_1(0) = 1, \zeta_2(10) = 1, z; \widehat{\psi})$, $P(\zeta_1(t) = 3|\zeta_1(0) = 1, \zeta_2(10) = 2, z; \widehat{\psi})$, and $P(\zeta_1(t) = 3|\zeta_1(0) = 1, \zeta_2(10) = 3, z; \widehat{\psi})$ respectively. These are plotted in Figure 3.4 and reveal that the appreciable estimate of Kendall's τ leads to a strong influence on the conditional probabilities and hence prediction in the course of disease.

3.5 General Remarks

In contrast to commonly-adopted intensity-based or frailty-based approaches, we have formulated a copula-based joint model for multiple multistate Markov processes and have paid special attention to the case that the processes are observed subject to intermittent inspections. Through decomposition of the density and conditional independence assumptions, an appealing joint model is obtained by assuming that the joint survival function of absorption transition times is governed by a multivariate copula function.

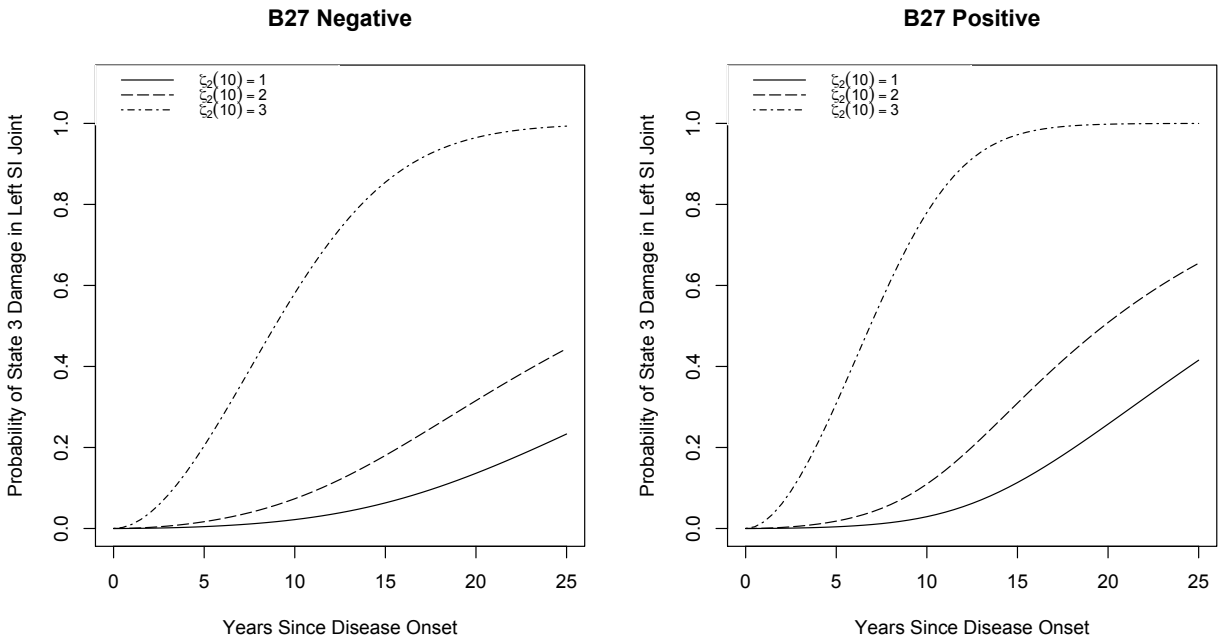

Figure 3.4: Plots of the estimated conditional probability $P(\zeta_1(t) = 3|\zeta_1(0) = 1, \zeta_2(10) = 1, z; \widehat{\psi})$, $P(\zeta_1(t) = 3|\zeta_1(0) = 1, \zeta_2(10) = 2, z; \widehat{\psi})$ and $P(\zeta_1(t) = 3|\zeta_1(0) = 1, \zeta_2(10) = 3, z; \widehat{\psi})$ vs. time since disease onset (years) for the composite likelihood (3.3.6) analysis using the joint model

This copula formulation has obvious advantages. It ensures comparability of marginal analyses to those of a single disease process and allows us to use the marginal methods in the existing literature. It guarantees an easy interpretation of marginal covariate effects which is not available if a frailty-based model is adopted. In addition to offering a number of options for estimation as discussed in Section 3.3, the copula formulation also ensures flexibility of model selection in that a wide range of marginal processes can be specified and copula functions can be selected from a rich family. Compared to the marginal method based on working independence assumption, our method facilitates scientific understanding regarding how the progression of one marginal disease process affects the other processes.

With interval-censored data arising from intermittent inspection, the full likelihood may involve high-dimensional integration when the marginal processes have a large number of states. Use of composite likelihood helps overcome such problems. The composite likelihood approaches and the two-stage methods offer considerable computational advantages and bring about increased robustness since they involve a lower degree of model specification at the price of some loss in efficiency. The robustness regarding consistency is similar in spirit to the robustness of generalized estimating equations (GEE) since both methods avoid specification of the higher-order dependencies (Xu and Reid, 2011). The computational advantages are based on the fact that the composite likelihood is integration-free and is easier to maximize (Varin *and others*, 2011). As is often the case, the computational convenience and robustness are gained by sacrificing statistical efficiency, so that the trade-off between those factors needs to taken into account when formulating a composite likelihood function.

Chapter 4

Semiparametric Two-stage Estimation for Copula Models in Survival Data

4.1 Introduction

Multivariate time-to-event data frequently arise in medical studies and are drawing increased attention. Appropriate multivariate models can be useful to identify risk factors, characterize their effects, and make predictions about multivariate event times. Moreover, it is often of interest to model the association between event times to advance scientific understanding about the relation between event times under study.

Two approaches are commonly used in modeling multivariate data: random effect (or

"frailty") models and marginal approaches. In random effect models, conditional independence assumptions are made between event times given a scalar or vector-valued nonnegative random variable. In marginal methods, the marginal distribution is modeled directly under a working independence assumption (Wei *and others*, 1989) and a robust covariance matrix can be obtained to ensure valid simultaneous inference in the spirit of generalized estimation equations (Liang and Zeger, 1986). A dependency structure can alternatively be imposed and this is often done through selection of a copula model (Joe, 1997; Nelsen, 2006), leading to a fully-specified model.

A significant appeal of copula models is that they enable separate modeling and estimation of the marginal distribution and the association. A nonparametric two-stage estimation (Chen and Huang, 2007) can be conducted by estimating the two marginals by one kernel distribution estimator given by Bowman *and others* (1998) and estimating the copula function by another kernel estimator considered by Fermanian and Scaillet (2003) in the second stage based on the estimated marginals . Nonparametric approaches can provide robustness but are inefficient and, have difficulty in handling censoring. If interest primarily lies in estimation of the association parameters in the copula, a parametric two-stage estimation can be followed by modelling the marginal distribution parametrically and estimating the marginal parameters in the first stage, and the association parameter in the copula in the second stage. Use of parametric models can lead to highly-efficient inference but often with the price that there is little robustness to inferences. A semiparametric two-stage approach can combine the best features of both approaches and will be our focus in what follows. This approach is carried out by conducting marginal analyses nonparamerically when no covariates appear or based on semiparametric models when co-

variates are present in the first stage. The association parameter in the copula is then estimated in the second stage by optimizing an objective function (often a likelihood) exploiting the estimates of the marginal distribution functions obtained in the first stage. Genest *and others* (1995) study the estimation of the association parameter and asymptotic properties of the estimator obtained from such a semiparametric two-stage approach for complete data. In medical studies, the event times are usually subject to censoring. The asymptotic properties of the estimator in the semiparametric two-stage estimation have been discussed in the literature for right-censored data with random censoring time (Shih and Louis, 1995), right-censored data with dependent censoring time (Wang, 2003), current status data (Wang and Ding, 2000), and interval-censored data (Sun *and others*, 2006). The asymptotic variance is also estimated by discretizing parameterization (Bailey, 1984; Lawless and Yilmaz, 2011; Yilmaz and Lawless, 2011) and calculating the asymptotic covariance matrix in the spirit of parametric two-stage estimation procedure (Newey and McFadden, 1994; Shih and Louis, 1995)

This chapter presents a new expression for the asymptotic variance of the second-stage estimator for the association parameter along with a way of estimating this in finite samples. This estimator of the asymptotic variance can be applied to complete, right-censored, and current status data with or without covariates. Simulation studies are conducted to compare the behaviour of the new estimators of the asymptotic variance with those in the established literature listed above and the empirical variance estimated in simulation studies.

The remainder of this chapter is organized as follows. The proposed estimator for the asymptotic variance of the second-stage estimator for the association parameter is given in

Section 4.2. Simulation studies and analyses of data from motivating studies are in Section 4.3 and 4.4 respectively, and general remarks are in Section 4.5.

4.2 Estimation and Inference

4.2.1 Preliminary Remarks

We first consider a single-sample problem. Let (T_1, T_2) denote a pair of random survival times with continuous marginal survivor functions $(\mathcal{F}_1(\cdot), \mathcal{F}_2(\cdot))$, where $\mathcal{F}_j(t) = P(T_j \geq t)$, $j = 1, 2$, and density functions $(f_1(\cdot), f_2(\cdot))$, where $f_j(t) = -d\mathcal{F}_j(t)/dt$, $j = 1, 2$. If (T_1, T_2) is governed by a copula function $\mathcal{C}_\phi(\cdot, \cdot)$, where ϕ is an association parameter, then $\mathcal{F}(t_1, t_2) = P(T_1 \geq t_1, T_2 \geq t_2)$, the joint survivor function of (T_1, T_2), is of the form

$$\mathcal{F}(t_1, t_2) \;=\; \mathcal{C}_\phi(\mathcal{F}_1(t_1), \mathcal{F}_2(t_2)) \;, \quad t_1, t_2 \geq 0 \;. \tag{4.2.1}$$

We let $c_\phi(\cdot)$ denote the bivariate density corresponding to the copula $\mathcal{C}_\phi(\cdot)$.

4.2.2 Data Structure

Let (C_1, C_2) denote a pair of random times with continuous marginal survivor functions $(\mathcal{G}_1(\cdot), \mathcal{G}_2(\cdot))$ and density functions $(g_1(\cdot), g_2(\cdot))$, and suppose further that (C_1, C_2) are independent of (T_1, T_2). For right-censored data, let $X_j = \min(T_j, C_j)$ and $\Delta_j = I(T_j \leq C_j)$, $j = 1, 2$. Under a current status observation scheme, we observe $X_j = C_j$ and $\Delta_j = I(T_j \leq C_j)$, $j = 1, 2$. In an effort to unify the way that these two settings are handled,

we note that the observed data for these two cases can be expressed as times (X_1, X_2) and indicators (Δ_1, Δ_2). Here we let X_j has the survivor function $\mathcal{H}_j(\cdot)$ and density function $h_j(\cdot)$. Since X_1 and X_2 are continuous but Δ_1 and Δ_2 are discrete random variables, we define $(X_1, X_2, \Delta_1, \Delta_2)$ having the "mixed joint density"

$$h(x_1, x_2, \delta_1, \delta_2) = h_{X_1,X_2}(x_1, x_2|\Delta_1 = \delta_1, \Delta_2 = \delta_2)P(\Delta_1 = \delta_1, \Delta_2 = \delta_2) \ ,$$

and the joint cumulative distribution function (c.d.f.)

$$H(x_1, x_2, \delta_1, \delta_2) = \sum_{s_1 \leq \delta_1} \sum_{s_2 \leq \delta_2} \int_{t_2=-\infty}^{x_2} \int_{t_1=-\infty}^{x_1} h(t_1, t_2, s_1, s_2) dt_1 dt_2 \ .$$

Let $\{(X_{1i}, X_{2i}, \Delta_{1i}, \Delta_{2i}), i = 1, \ldots, m\}$ be a random sample from H and let H_m denote the empirical distribution function corresponding to H based on a sample of size m. For right-censored data, the marginal density function of X_j is

$$h_j(x_j) = f_j(x_j)\mathcal{G}_j(x_j) + g_j(x_j)\mathcal{F}_j(x_j) \ ,$$

and for current status data, $h_j(x_j) = g_j(x_j)$.

4.2.3 Stage I: Estimation of the Marginal Survivor Functions

In the first stage of the semiparametric two-stage estimation procedure, a working independence assumption is made and estimation of marginal survivor functions is carried out. The problem is then simplified to that of a univariate survival analysis. Let $F_j(t) = 1 - \mathcal{F}_j(t)$.

Complete Data

If the complete data are observed, the estimated marginal c.d.f. is taken to be $m/(m+1)$ times the marginal empirical distribution function, i.e.,

$$\widehat{F}_j(t) = \frac{1}{m}\sum_{i=1}^{m} I(X_{ji} < t) \ ,$$

and $\widehat{\mathcal{F}}_j(t) = 1 - \widehat{F}_j(t)$, $j = 1, 2$. This rescaling avoids difficulties arising from the potential unboundedness of $\log c_\phi(u_1, u_2)$ as u or v tend to one (Genest *and others*, 1995).

Right-censored Data

If the event times are subject to right-censoring, the marginal survivor function $\mathcal{F}_j(t)$ can be estimated by the Kaplan-Meier method (Kaplan and Meier, 1958) as

$$\widehat{\mathcal{F}}_j(t) = \prod_{i:X_{ji}<t} \left(1 - d\widehat{\Lambda}_j(X_{ji})\right) \ , \tag{4.2.2}$$

where

$$d\widehat{\Lambda}_j(X_{ji}) = \frac{\Delta_{ji}}{\sum_{k=1}^{m} I(X_{jk} \geq X_{ji})} \ , \tag{4.2.3}$$

if we assume there are no ties in the data.

Current Status Data

The estimation of the survival function in this context is thoroughly reviewed in Huang and Wellner (1997). The nonparametric maximum likelihood estimator (NPMLE) of F_j, denoted by $\widehat{F}_j$, maximizes the function

$$\ell(F_j) = \sum_{i=1}^{m} \left[\Delta_{ji} \log F_j(C_{j(i)}) + (1 - \Delta_{ji}) \log \mathcal{F}_j(C_{j(i)}) \right] ,$$

for $j = 1, 2$. The marginal estimator of the c.d.f. for events of type j, $\widehat{F}_j$, can be represented by the max-min formula with current status data,

$$\widehat{F}_j(C_{j(i)}) = \max_{\ell \leq i} \left[\min_{k \geq i} \left(\frac{\sum_{n=\ell}^{k} \Delta_{j(n)}}{k - \ell + 1} \right) \right] ,$$

where $C_{j(1)} \leq C_{j(2)} \leq \ldots \leq C_{j(m)}$ are ordered values of $(C_{j1}, \ldots, C_{jm})$ and $\Delta_{j(i)}$ are the associated indicators for $C_{j(i)}$, for $j = 1, 2$ and $i = 1, \ldots, m$. This expression is not optimal for actual computation of $\widehat{F}_j(C_{j(i)})$, and a popular algorithm for this setting is the pool adjacent violators algorithm (PAVA) (Ayer *and others*, 1955; van Eeden, 1956, 1957). No attempt is made to smooth the estimate of what is likely a continuous function and so this estimate has been criticized for its lack of smoothness, the presence of too many flat regions, and its slow ($n^{1/3}$) rate of convergence (Groeneboom and Wellner, 1992). The isotonized kernel estimator (Schimek, 2000) is defined by

$$\widehat{F}_j^*(t) = \frac{\sum_{i=1}^{m} \mathcal{K}\left((t - C_{j(i)})/b_m\right) \widehat{F}_j(C_{j(i)})}{\sum_{i=1}^{m} \mathcal{K}\left((t - C_{j(i)})/b_m\right)} , \tag{4.2.4}$$

for time t, where b_m is the bandwidth and $\mathcal{K}(\cdot)$ is a log-concave density (i.e., $\mathcal{K}(\cdot) = \exp(Q(\cdot))$, $Q(\cdot)$ is a concave function).

4.2.4 Stage II: Estimation of the Association Parameter

In the second stage of the two-stage estimation procedure, the association parameter is estimated by maximizing the likelihood with respect to ϕ when it is evaluated at the estimated marginal survivor functions at the observed times $(\widehat{\mathcal{F}}_1(X_{1i}), \widehat{\mathcal{F}}_2(X_{2i}))$ and the corresponding indicators $(\Delta_{1i}, \Delta_{2i})$ for $i = 1, \ldots, m$. This is denoted by

$$\prod_{i=1}^{m} L(\phi; \widehat{\mathcal{F}}_1(X_{1i}), \widehat{\mathcal{F}}_2(X_{2i}), \Delta_{1i}, \Delta_{2i}) \ , \tag{4.2.5}$$

and we let $\ell(\phi; u_1, u_2, \delta_1, \delta_2) = \log L(\phi; u_1, u_2, \delta_1, \delta_2)$, where $u_j \in [0, 1]$ and $\delta_j = \{0, 1\}$, $j = 1, 2$. The estimator of ϕ, denoted by $\widetilde{\phi}$, is the solution to the estimating equation

$$\sum_{i=1}^{m} U_\phi(\phi; \widehat{\mathcal{F}}_1(X_{1i}), \widehat{\mathcal{F}}_2(X_{2i}), \Delta_{1i}, \Delta_{2i}) = 0 \ , \tag{4.2.6}$$

where

$$U_\phi(\phi; u_1, u_2, \delta_1, \delta_2) \ = \ \frac{\partial \ell(\phi; u_1, u_2, \delta_1, \delta_2)}{\partial \phi} \ .$$

We also define some auxiliary quantities

$$\begin{aligned} V_{\phi}(\phi; u_1, u_2, \delta_1, \delta_2) &= \frac{\partial^2 \ell(\phi; u_1, u_2, \delta_1, \delta_2)}{\partial \phi^2} , \\ V_{\phi,1}(\phi; u_1, u_2, \delta_1, \delta_2) &= \frac{\partial^2 \ell(\phi; u_1, u_2, \delta_1, \delta_2)}{\partial \phi \partial u_1} , \\ V_{\phi,2}(\phi; u_1, u_2, \delta_1, \delta_2) &= \frac{\partial^2 \ell(\phi; u_1, u_2, \delta_1, \delta_2)}{\partial \phi \partial u_2} . \end{aligned}$$

Complete Data

For complete data (Genest *and others*, 1995), $L(\phi; u_1, u_2, \delta_1 = 1, \delta_2 = 1) = L(\phi; u_1, u_2)$ and

$$L(\phi; u_1, u_2) = c_{\phi}(u_1, u_2).$$

Right-censored Data

Under right censoring (Shih and Louis, 1995), we obtain

$$L(\phi; u_1, u_2, \delta_1, \delta_2) = c_{\phi}(u_1, u_2)^{\delta_1 \delta_2} \frac{\partial \mathcal{C}_{\phi}(u_1, u_2)^{\delta_1(1-\delta_2)}}{\partial u_1} \frac{\partial \mathcal{C}_{\phi}(u_1, u_2)^{\delta_2(1-\delta_1)}}{\partial u_2} \mathcal{C}_{\phi}(u_1, u_2)^{(1-\delta_1)(1-\delta_2)} .$$

Current Status Data

For current status data (Wang and Ding, 2000),

$$\begin{aligned} L(\phi; u_1, u_2, \delta_1, \delta_2) &= [1 - u_1 - u_2 + \mathcal{C}_{\phi}(u_1, u_2)]^{\delta_1 \delta_2} \cdot [u_2 - \mathcal{C}_{\phi}(u_1, u_2)]^{\delta_1(1-\delta_2)} \\ &\cdot \quad [u_1 - \mathcal{C}_{\phi}(u_1, u_2)]^{\delta_2(1-\delta_1)} \cdot \mathcal{C}_{\phi}(u_1, u_2)^{(1-\delta_1)(1-\delta_2)} . \end{aligned}$$

4.2.5 Asymptotic Properties of the Estimator $\widetilde{\phi}$

The consistency of $\widetilde{\phi}$ can be shown by applying Theorem 2.5 in Newey and McFadden (1994) or justified by asymptotical unbiasedness of (4.2.6) in the theory of estimating equations (Godambe, 1991). This section focuses on conditions for asymptotic normality.

Theorem 1. *Under regularity conditions (a)-(h) of Appendix A,* $m^{1/2}(\widetilde{\phi}-\phi)$ *converges to a normal distribution with mean zero and variance* $\sigma^2 = (\sigma_1^2+\sigma_2^2)/\sigma_1^4$*, as* $m \to \infty$*, where*

$$\begin{aligned} \sigma_1^2 &= E\left[-V_\phi(\phi; \mathcal{F}_1(X_1), \mathcal{F}_2(X_2), \Delta_1, \Delta_2)\right] , \\ \sigma_2^2 &= E\left[W_1(\phi, X_1) + W_2(\phi, X_2)\right]^2 , \end{aligned}$$

and

$$W_j(\phi, X_j) = \int_{\mathcal{A}} V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} I(X_j \geq x_j) dH(x_1, x_2, \delta_1, \delta_2) .$$

The proof of Theorem 1 is given in Appendix A along with the regularity conditions (a)-(h).

For finite samples, an estimator of the asymtotic variance of $\widetilde{\phi}$ may be obtained by replacing H by its empirical distribution function H_m, and $\mathcal{F}_j(\cdot)$, $f_j(\cdot)$, $h_j(\cdot)$ and ϕ by

their (uniformly) consistent estimators $\widehat{\mathcal{F}}_1(\cdot)$, $\widehat{f}_j(\cdot)$, $\widehat{h}_j(\cdot)$ and $\widetilde{\phi}$, for $j = 1, 2$. Specifically,

$$\widehat{\sigma}_1^2 = -\frac{1}{m}\sum_{i=1}^{m} V_\phi(\widetilde{\phi}; \widehat{\mathcal{F}}_1(X_{1i}), \widehat{\mathcal{F}}_2(X_{2i}), \Delta_{1i}, \Delta_{2i}) \ , \tag{4.2.7}$$

$$\widehat{\sigma}_2^2 = \frac{1}{m}\sum_{i=1}^{m} \left[\widehat{W}_1(\widetilde{\phi}, X_{1i}) + \widehat{W}_2(\widetilde{\phi}, X_{2i})\right]^2 \ , \tag{4.2.8}$$

where

$$\widehat{W}_j(\widetilde{\phi}, X_{ji}) = \frac{1}{m}\sum_{k=1}^{m} V_{\phi,j}(\widetilde{\phi}; \widehat{\mathcal{F}}_1(X_{1k}), \widehat{\mathcal{F}}_2(X_{2k}), \Delta_{1k}, \Delta_{2k}) \frac{\widehat{f}_j(X_{jk})}{\widehat{h}_j(X_{jk})} I(X_{ji} \geq X_{jk}) \ , \tag{4.2.9}$$

$j = 1, 2$.

Theorem 2. *Under regularity conditions (a)-(j) of Appendices A and B, $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ in* (4.2.7) *and* (4.2.8) *are consistent estimators of σ_1^2 and σ_2^2 respectively.*

The proof of Theorem 2 is given in Appendix B.

4.2.6 Regression Problem

Let Z be a vector of fixed discrete covariates, such as treatment indicators, or gender. The Cox model (Cox, 1972) specifies the hazard rate of failure for the survival time T_j given $Z = z$ to have the form

$$\lambda_j(t|z) = \lambda_{j0}(t) \exp(z'\beta_j) \ , \quad t \geq 0 \ ,$$

where the baseline intensity function $\lambda_{j0}(t)$ is not assumed to have any particular parametric form. The marginal survivor function of T_j given $Z = z$ has the form $\mathcal{F}_j(t_j|z) = \mathcal{F}_{j0}(t_j)^{\exp(z'\beta_j)}$ and the corresponding density function is denoted by $f_j(t_j|z)$; $\mathcal{F}_{j0}(t_j) = \exp\left[-\int_0^{t_j} \lambda_{j0}(u)du\right]$. We also assume that the joint survivor function of (T_1, T_2) conditional on $Z = z$ is induced by a conditional copula function (Patton, 2006) as introduced in (1.1.3) and bears the form

$$\mathcal{F}(t_1, t_2|z) = \mathcal{C}_\phi(\mathcal{F}_1(t_1|z), \mathcal{F}_2(t_2|z)), \quad t_1, t_2 \geq 0 \, .$$

Suppose that the censoring times (C_1, C_2) and (T_1, T_2) are conditional independent given covariate Z. Under right censoring, the marginal density function of X_j conditional on $Z = z$ is then

$$h_j(x_j|z) = f_j(x_j|z)\mathcal{G}_j(x_j|z) + g_j(x_j|z)\mathcal{F}_j(x_j|z) \, ,$$

and for current status data $h_j(x_j|z) = g_j(x_j|z)$.

In the first stage, we estimate $\widehat{\mathcal{F}}_j(\cdot|z)$, the marginal survivor function conditional on $Z = z$, and in the second stage, the likelihood (4.2.5) is maximized with respect to ϕ when it is evaluated at the estimated marginal survivor function $(\widehat{\mathcal{F}}_1(X_{1i}|Z_i), \widehat{\mathcal{F}}_2(X_{2i}|Z_i), \Delta_{1i}, \Delta_{2i})$ to obtain an estimate of the association parameter ϕ, denoted by $\widetilde{\phi}$. Asymptotic normality can be proven as in Theorem 1, except that here the relevant functionals are conditional on Z and so take the form

$$\sigma_1^2 \;=\; \mathrm{E}\left[-V_\phi(\phi; \mathcal{F}_1(X_1|Z), \mathcal{F}_2(X_2|Z), \Delta_1, \Delta_2)\right] \, ,$$

and

$$\sigma_2^2 \;=\; \mathrm{E}\left[W_1(\phi, X_1, Z) + W_2(\phi, X_2, Z)\right]^2 \;,$$

where

$$W_j(\phi, X_j, Z) = \int_{\mathcal{A}} V_{\phi,j}(\phi; \mathcal{F}_1(x_1|z), \mathcal{F}_2(x_2|z), \delta_1, \delta_2) \cdot \frac{f_j(x_j|z)}{h_j(x_j|z)} \cdot \frac{I(X_j \geq x_j) I(Z = z)}{P(Z = z)} dH(x_1, x_2, \delta_1, \delta_2|z) \;.$$

Consistent estimators of σ_1^2 and σ_2^2 are

$$\widehat{\sigma}_1^2 \;=\; -\frac{1}{m} \sum_{i=1}^{m} V_\phi(\widetilde{\phi}; \widehat{\mathcal{F}}_1(X_{1i}|Z_i), \widehat{\mathcal{F}}_2(X_{2i}|Z_i), \Delta_{1i}, \Delta_{2i}) \;, \tag{4.2.10}$$

and

$$\widehat{\sigma}_2^2 \;=\; \frac{1}{m} \sum_{i=1}^{m} \left[\widehat{W}_1(\widetilde{\phi}, X_{1i}, Z_i) + \widehat{W}_2(\widetilde{\phi}, X_{2i}, Z_i)\right]^2 \;, \tag{4.2.11}$$

where we define $m_i = \sum_{k=1}^{m} I(Z_k = Z_i)$ and

$$\widehat{W}_j(\widetilde{\phi}, X_{ji}, Z_i) = \frac{1}{m_i} \sum_{k=1}^{m} V_{\phi,j}(\widetilde{\phi}; \widehat{\mathcal{F}}_1(X_{1k}|Z_i), \widehat{\mathcal{F}}_2(X_{2k}|Z_i), \Delta_{1k}, \Delta_{2k}) \cdot \frac{\widehat{f}_j(X_{jk}|Z_i)}{\widehat{h}_j(X_{jk}|Z_i)} \cdot I(X_{ji} \geq X_{jk}) \cdot I(Z_k = Z_i) \;,$$

for $j = 1, 2$.

4.2.7 Discrete Parameterization

In this approach we drop the assumption that the baseline survivor functions $\mathcal{F}_{j0}(\cdot)$, $j = 1, 2$, are continuous functions, and instead assume them to be discrete survivor distributions which place mass only on observed failure times. Following this discrete parameterization, parametric maximum likelihood methods can be used for estimation and inference, making this virtually the same problem as that of parametric two-stage estimation procedures (Newey and McFadden, 1994; Shih and Louis, 1995). This *ad hoc* approach is frequently adopted in research because it by-passes the nonparametric or semiparametric nature of the problem by conducting estimation and inference in the parametric setting. Ackerberg *and others* (2012) derive a number of results demonstrating the numerical equivalence between nonparametric or semiparametric variance estimates and variance estimates based on a parametric treatment of the problem. The equivalence will be studied in some cases here through examination of the theory and by simulation.

4.3 Examples and Simulation Studies

In this section, we empirically assess the performance of the variance estimates of $\log \widetilde{\phi}$, the logarithm of the estimate of the dependence parameter in the copula, under different scenarios. We reparameterize and estimate $\log(\phi)$ instead of ϕ to remove issues with estimates near the boundary of the parameter space. A pair of survival times are generated by a Clayton copula (1.1.5).

4.3.1 Bivariate Right-censored Survival Data

We first consider bivariate right-censored survival data.

The One-sample Problem

We assume unit exponential margins and set Kendall's τ from the Clayton copula to be 0.25, 0.50 and 0.75 to represent mild, moderate and strong association respectively. We consider the cases of no-censoring and 30% right-censoring with the latter achieved by generating a pair of independent uniformly-distributed random censoring times over $[0, 3.197]$. For each parameter configuration, we generate 2000 samples with $m = 200$ or 400 subjects in each sample.

For each dataset, analyses are carried out based on a semiparametric two-stage estimation method, in which the first-stage estimation is conducted as in (4.2.2) and the second-stage estimation is carried out by plugging in the marginal distribution estimators as shown in (4.2.5). The empirical bias (BIAS), empirical standard error (ESE), average asymptotic (large sample) standard error (ASE) and empirical coverage probability (ECP) are evaluated for the association parameter estimate. Analyses are carried out using programs in `R`.

The ASE_1 is the average of the 2000 large sample standard errors, each of which is calculated based on the method proposed in Section 4.2.5, computed by $\sqrt{\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2}/\widehat{\sigma}_1^2$, where $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ are given in (4.2.7) and (4.2.8) respectively. In (4.2.9), the functions $f_j(t)$ and $h_j(t)$ are estimated by kernel function estimators $\widehat{f}_j(t)$ and $\widehat{h}_j(t)$. The kernel function

estimator of the marginal density function $f_j(t)$ is

$$\widehat{f}_j(t) = -b_m^{-1} \int \mathcal{K}\left(\frac{t-u}{b_m}\right) d\widehat{\mathcal{F}}_j(u) \ , \tag{4.3.1}$$

where $\mathcal{K}(\cdot)$ is a kernel probability density function with support over $[-1, 1]$. When $\mathcal{K}(u)$ is the uniform kernel function, i.e., $\mathcal{K}(u) = \frac{1}{2} I(-1 \leq u \leq 1)$, this amounts to estimating $f_j(t)$ by

$$\widehat{f}_j(t) = \frac{1}{2b_m} \left[\widehat{\mathcal{F}}_j(t - b_m) - \widehat{\mathcal{F}}_j(t + b_m)\right] \ ,$$

where b_m is a bandwidth. Similarly, $h_j(t)$ is estimated by

$$\widehat{h}_j(t) = \frac{1}{2b_m} \left[\widehat{\mathcal{H}}_j(t - b_m) - \widehat{\mathcal{H}}_j(t + b_m)\right] \ , \tag{4.3.2}$$

where $\widehat{\mathcal{H}}_j(t)$ is the Kaplan-Meier estimator for $\mathcal{H}_j(t)$ of the form

$$\widehat{\mathcal{H}}_j(t) = \prod_{i:X_{ji}<t} \left(1 - \frac{1}{\sum_{k=1}^{m} I(X_{jk} \geq X_{ji})}\right) \ ,$$

again assuming that there are no ties but this time the assumption is made for the X terms. Uniform convergence stands if

(i) $f_j(t)$ and $h_j(t)$ are continuous on $t \in [0, \infty]$,

(ii) $\mathcal{K}(u)$ is a continuous density function with support $u \in [-1, 1]$ and is of bounded variation,

(iii) $b_m \to 0$ and $(\log m)/(mb_m) \to 0$ as $m \to \infty$.

The selection of the bandwidth b_m is critical as it will affect the performance of the estimator for asymptotic variance of $\widetilde{\phi}$. This usually can be carried out by cross-validation method. The bandwidth selection for kernel estimator of $f_j(t)$ and $h_j(t)$ can be tackled in the same way, and we take $f_j(t)$ as an example. We follow Bowman (1984) and select a bandwidth by minimizing the following quantity

$$\frac{1}{m}\sum_{i=1}^{m}\int\left(\widehat{f}_j^{(-i)}(y)\right)^2 dy - \frac{2}{m}\sum_{i=1}^{m}\widehat{f}_j^{(-i)}(X_{ji})$$

where $\widehat{f}_j^{(-i)}(y)$ denotes the kernel estimator constructed from the data without the ith observation. This method is commonly referred to as least square cross-validation, since it is base on the so-called leave-one-out density estimator $\widehat{f}_j^{(-i)}(y)$. In our specific implementation, we choose candidate values from 0.001 to 0.1 in step of 0.001 and select the optimal value for bandwidth b_m by the least square cross-validation method. In principle, we can conduct cross-validation for each of the 2000 simulated samples. Due to the computational burden, we refrain from doing so and instead use a common bandwidth b_m for all of the 2000 samples selected by conducting cross-validation for a few trials on some randomly-generated samples.

The ASE_2 is the average of the 2000 large sample standard errors introduced in Shih and Louis (1995), of the form $\sqrt{\widehat{\sigma}_1^2 + \widetilde{\sigma}_2^2}/\widehat{\sigma}_1^2$, in which

$$\widetilde{\sigma}_2^2 = \frac{1}{m}\sum_{i=1}^{m}\left[\widetilde{W}_1(\widetilde{\phi}, X_{1i}, \Delta_{1i}) + \widetilde{W}_2(\widetilde{\phi}, X_{2i}, \Delta_{2i})\right]^2 ,$$

with

$$\widetilde{W}_j(\widetilde{\phi}, X_{ji}, \Delta_{ji}) = \frac{1}{m}\sum_{k=1}^{m} V_{\phi,j}(\widetilde{\phi}; \widehat{\mathcal{F}}_1(X_{1k}), \widehat{\mathcal{F}}_2(X_{2k}), \Delta_{1k}, \Delta_{2k})\widehat{I}_j(X_{ji}, \Delta_{ji}, X_{jk}) \ ,$$

$$\widehat{I}_j(X_{ji}, \Delta_{ji}, X_{jk}) = -\widehat{\mathcal{F}}_j(X_{jk}) \left\{ \frac{I(X_{ji} \leq X_{jk}, \Delta_{ji} = 1)}{\widehat{p}_{ji}} - \sum_{X_{j\ell} \leq \min(X_{ji}, X_{jk})} \frac{d\widehat{\Lambda}_j(X_{j\ell})}{\widehat{p}_{j\ell}} \right\} \ ,$$

and

$$\widehat{p}_{ji} = \frac{\sum_\ell \{X_{j\ell} \geq X_{ji}\}}{m} \ , \text{ for } j = 1, 2.$$

The ASE_3 is the average of the 2000 large sample standard errors calculated assuming a parametric model with a discrete hazard in which the estimates of $\mathcal{F}_1(\cdot)$ and $\mathcal{F}_2(\cdot)$ have jumps only at observed (i.e. uncensored) times (Bailey, 1984). Let $\lambda_j = (\lambda_{j1}, \ldots, \lambda_{jN_j})'$, $N_j = \sum_{i=1}^m \Delta_{ji}$ and $\psi = (\phi, \lambda_1', \lambda_2')'$ be the full vector of parameters. The two-stage "parametric" two-stage estimation procedure can be conducted in the way that marginal discrete hazard parameters λ_j are estimated in the first stage by maximizing the function

$$\ell_j(\lambda_j) = \sum_{i=1}^{m} \left[\sum_{k:X_{jk}<X_{ji}} \log(1 - \lambda_{jk}) + \Delta_{ji} \log \lambda_{ji} \right] \ .$$

We thus obtain

$$\widehat{\lambda}_{ji} = \frac{\Delta_{ji}}{\sum_{k=1}^m I(X_{jk} > X_{ji}) + \Delta_{ji}} \ ,$$

which is equal to $d\widehat{\Lambda}_j(X_{ji})$ in (4.2.3). The marginal survival function is estimated by

$$\widehat{\mathcal{F}}_j(t) = \prod_{i:X_{ji}<t} (1 - \widehat{\lambda}_{ji}) \ ,$$

which is equal to the Kaplan-Meier estimator (4.2.2) and ensures the equivalence between nonparameteric and parametric methods for estimation of the marginal distributions. In the second stage, $\widehat{\mathcal{F}}_j(t)$ is then plugged into a likelihood as in (4.2.5), of which the log-likelihood can be written as $\ell(\psi) = \sum_{i=1}^{m} \ell_i(\psi)$. We can get an estimate of ϕ by maximizing $\ell(\psi)$ with respect to ϕ given $(\widehat{\lambda}_1', \widehat{\lambda}_2')'$. The sample standard error is the square root of item $[1,1]$ of the matrix

$$\widehat{\mathbb{A}}^{-1}(\widehat{\psi})\widehat{\mathbb{B}}(\widehat{\psi}) \left[\widehat{\mathbb{A}}^{-1}(\widehat{\psi})\right]' \ ,$$

where

$$\widehat{\mathbb{A}}(\widehat{\psi}) = \sum_{i=1}^{m} \left(\begin{array}{ccc} \frac{\partial^2 \ell_i(\psi)}{\partial\phi\partial\phi'} & \frac{\partial^2 \ell_i(\psi)}{\partial\phi\partial\lambda_1'} & \frac{\partial^2 \ell_i(\psi)}{\partial\phi\partial\lambda_2'} \\ 0 & \frac{\partial^2 \ell_{i1}(\lambda_1)}{\partial\lambda_1\partial\lambda_1'} & 0 \\ 0 & 0 & \frac{\partial^2 \ell_{i2}(\lambda_2)}{\partial\lambda_2\partial\lambda_2'} \end{array} \right) \Bigg|_{\psi=\widehat{\psi}} \ ,$$

and

$$\widehat{\mathbb{B}}(\widehat{\psi}) = \sum_{i=1}^{m} \left(\begin{array}{ccc} \frac{\partial \ell_i(\psi)}{\partial\phi}\frac{\partial \ell_i(\psi)}{\partial\phi'} & \frac{\partial \ell_i(\psi)}{\partial\phi}\frac{\partial \ell_{i1}(\lambda_1)}{\partial\lambda_1'} & \frac{\partial \ell_i(\psi)}{\partial\phi}\frac{\partial \ell_{i2}(\lambda_2)}{\partial\lambda_2'} \\ \frac{\partial \ell_{i1}(\lambda_1)}{\partial\lambda_1}\frac{\partial \ell_i(\psi)}{\partial\phi'} & \frac{\partial \ell_{i1}(\lambda_1)}{\partial\lambda_1}\frac{\partial \ell_{i1}(\lambda_1)}{\partial\lambda_1'} & \frac{\partial \ell_{i1}(\lambda_1)}{\partial\lambda_1}\frac{\partial \ell_{i2}(\lambda_2)}{\partial\lambda_2'} \\ \frac{\partial \ell_{i2}(\lambda_2)}{\partial\lambda_2}\frac{\partial \ell_i(\psi)}{\partial\phi'} & \frac{\partial \ell_{i2}(\lambda_2)}{\partial\lambda_2}\frac{\partial \ell_{i1}(\lambda_1)}{\partial\lambda_1'} & \frac{\partial \ell_{i2}(\lambda_2)}{\partial\lambda_2}\frac{\partial \ell_{i2}(\lambda_2)}{\partial\lambda_2'} \end{array} \right) \Bigg|_{\psi=\widehat{\psi}} \ ,$$

Table 4.1: Empirical frequency properties of the estimates and associated variance estimates for the association parameter of bivariate survival times with no-censoring; 2000 simulated samples

n	Kendall's τ	BIAS	ESE	ASE_1	ECP_1	ASE_2	ECP_2	ASE_3	ECP_3
200	0.25	0.027	0.219	0.220	0.944	0.221	0.930	0.221	0.944
	0.50	0.003	0.139	0.139	0.951	0.137	0.944	0.141	0.952
	0.75	-0.023	0.111	0.113	0.944	0.111	0.943	0.114	0.946
400	0.25	0.017	0.154	0.155	0.946	0.155	0.939	0.155	0.947
	0.50	0.004	0.096	0.099	0.956	0.097	0.952	0.099	0.956
	0.75	-0.014	0.077	0.079	0.954	0.078	0.952	0.079	0.954

1. average large sample standard error estimate from proposed method in Section 4.2.5; 2. average large sample standard error estimate of Shih and Louis (1995); 3. average large sample standard error estimate by a parametric approach.

and $\widehat{\psi} = (\widetilde{\phi}, \widehat{\lambda}_1', \widehat{\lambda}_2')'$.

Table 4.1 reports the empirical frequency properties of the estimate of the association parameter and its corresponding standard error estimates when data are complete. The findings in Table 4.1 reveal, as expected from the asymptotic theory, that the empirical biases are all very small for the estimates of the association parameters with mild, moderate or strong dependence between survival times. The ASE and ESE are in close agreement and the empirical coverage probabilities are close to the nominal 95% level, suggesting that the methods proposed provide a valid basis for inference and behave equally well as the method proposed by Shih and Louis (1995) and the *ad hoc* parametric approach of Section 4.2.7.

Table 4.2 reports the same type of information as Table 4.1, but when the survival

Table 4.2: Empirical frequency properties of the estimates and associated variance estimates for the association parameter of bivariate survival times with 30% right-censoring; 2000 simulated samples

n	Kendall's τ	BIAS	ESE	ASE_1	ECP_1	ASE_2	ECP_2	ASE_3	ECP_3
200	0.25	-0.019	0.290	0.280	0.954	0.297	0.960	0.290	0.963
	0.50	-0.007	0.173	0.171	0.948	0.177	0.952	0.176	0.954
	0.75	-0.040	0.138	0.139	0.945	0.149	0.956	0.147	0.955
400	0.25	-0.010	0.199	0.192	0.946	0.203	0.953	0.200	0.950
	0.50	-0.007	0.116	0.114	0.946	0.123	0.962	0.123	0.964
	0.75	-0.026	0.093	0.092	0.942	0.101	0.962	0.101	0.960

1. average large sample standard error estimate from proposed method in Section 4.2.5; 2. average large sample standard error estimate of Shih and Louis (1995); 3. average large sample standard error estimate by a parametric approach.

times are subject to 30% right-censoring. The results for this case demonstrate that there is often much closer agreement between the empirical standard error and the proposed standard error estimates based on (4.2.7) and (4.2.8); as these estimates are often smaller, the resulting confidence intervals are also narrower.

The Regression Problem

In the first stage of the two-stage estimation procedure, we obtain the estimator of the marginal coefficient β_j and the marginal baseline hazard function $d\Lambda_j(\cdot)$. Inference about β_j can be carried out by constructing the marginal partial likelihood (Andersen and Gill,

1982)

$$L(\beta_j) = \prod_{i=1}^{m} \left[\frac{\exp(z_i'\beta_j)}{\sum_{k=1}^{m} I(X_{jk} \geq X_{ji}) \exp(z_k'\beta_j)} \right]^{\Delta_{ji}} . \tag{4.3.3}$$

Let $\widehat{\beta}_j$ be the value that maximizes (4.3.3) and the estimator for the corresponding baseline hazard function at time X_{ji} is

$$d\widehat{\Lambda}_{j0}(X_{ji}) = \frac{\Delta_{ji}}{\sum_{k=1}^{m} I(X_{jk} \geq X_{ji}) \exp(z_k'\widehat{\beta}_j)} , \tag{4.3.4}$$

if we assume there are no ties in the data. The marginal survivor function $\mathcal{F}_j(t|z)$ can be estimated by (Andersen *and others*, 1993)

$$\widehat{\mathcal{F}}_j(t|z) = \prod_{i:X_{ji}<t} \left[1 - d\widehat{\Lambda}_{j0}(X_{ji}) \exp(z_i'\widehat{\beta}_j) \right] . \tag{4.3.5}$$

In the second stage, an estimate of ϕ is obtained by maximizing (4.2.5) with respect to ϕ with plugged-in $\widehat{\mathcal{F}}_j(X_{ji}|Z_i)$, $j = 1, 2$.

Empirical frequency properties of the estimate of the association parameter and its standard error estimates are reported in Table 4.3 for the case of 30% right-censoring with a binary covariate. In Table 4.3, ASE_1 is the average of the 2000 large sample standard errors using the method proposed in Section 4.2.6, equal to $\sqrt{\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2}/\widehat{\sigma}_1^2$, where $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ are given in (4.2.10) and (4.2.11) respectively. The term ASE_2 is calculated by discretizing the hazard parameters and calculating parametric two-stage standard errors. In the first stage, the marginal discrete hazard parameters λ_j and coefficient β_j are estimated by

maximizing the function

$$\ell_j(\theta_j) = \sum_{i=1}^{m} \left\{ \sum_{k:X_{jk}<X_{ji}} \log\left[1 - \lambda_{jk}\exp(z_k'\beta_j)\right] + \Delta_{ji}\left[\log\lambda_{ji} + z_i'\beta_j\right] \right\} .$$

where $\theta_j = (\lambda_j', \beta_j)'$. Then we obtain

$$\widehat{\lambda}_{ji} = \frac{\Delta_{ji}}{\sum_{k=1}^{m} I(X_{jk} > X_{ji})\exp(z_k'\widehat{\beta}_j) + \Delta_{ji}\exp(z_i'\widehat{\beta}_j)} ,$$

which is equal to $d\widehat{\Lambda}_{j0}(X_{ji})$ (4.3.4). The marginal survival function is estimated by

$$\widehat{\mathcal{F}}_j(t|z) = \prod_{i:X_{ji}<t} \left[1 - \widehat{\lambda}_{ji}\exp(z_i'\widehat{\beta}_j)\right] ,$$

which is equal to (4.3.5). In the second stage, $\widehat{\mathcal{F}}_j(X_{ji}|Z_i)$ is plugged into (4.2.5), the log-likelihood of which can be written as $\ell(\psi)$ with $\psi = (\phi, \theta_1', \theta_2')'$. The sample standard error is square root of the $[1,1]$ element of the matrix

$$\widehat{\mathbb{A}}^{-1}(\widehat{\psi})\widehat{\mathbb{B}}(\widehat{\psi})\left[\widehat{\mathbb{A}}^{-1}(\widehat{\psi})\right]' ,$$

where

$$\widehat{\mathbb{A}}(\widehat{\psi}) = \sum_{i=1}^{m} \left(\begin{array}{ccc} \frac{\partial^2\ell_i(\psi)}{\partial\phi\partial\phi'} & \frac{\partial^2\ell_i(\psi)}{\partial\phi\partial\theta_1'} & \frac{\partial^2\ell_i(\psi)}{\partial\phi\partial\theta_2'} \\ 0 & \frac{\partial^2\ell_{i1}(\theta_1)}{\partial\theta_1\partial\theta_1'} & 0 \\ 0 & 0 & \frac{\partial^2\ell_{i2}(\theta_2)}{\partial\theta_2\partial\theta_2'} \end{array} \right)\Bigg|_{\psi=\widehat{\psi}} ,$$

Table 4.3: Empirical frequency properties of the estimates and associated variance estimates for the association parameter of bivariate survival times with covariate under 30% right-censoring; 2000 simulated samples

n	Kendall's τ	BIAS	ESE	ASE_1	ECP_1	ASE_2	ECP_2
200	0.25	-0.029	0.272	0.269	0.962	0.285	0.968
	0.50	-0.023	0.161	0.158	0.948	0.172	0.962
	0.75	-0.062	0.128	0.138	0.942	0.149	0.952
400	0.25	-0.014	0.188	0.182	0.946	0.194	0.956
	0.50	-0.011	0.118	0.116	0.942	0.119	0.957
	0.75	-0.033	0.092	0.099	0.951	0.099	0.956

1. average large sample standard error estimate from proposed method in Section 4.2.6; 2. average large sample standard error estimate by a parametric approach.

and

$$\widehat{\mathbb{B}}(\widehat{\psi}) = \sum_{i=1}^{m} \left(\begin{array}{ccc} \frac{\partial \ell_i(\psi)}{\partial \phi}\frac{\partial \ell_i(\psi)}{\partial \phi'} & \frac{\partial \ell_i(\psi)}{\partial \phi}\frac{\partial \ell_{i1}(\theta_1)}{\partial \theta_1'} & \frac{\partial \ell_i(\psi)}{\partial \phi}\frac{\partial \ell_{i2}(\theta_2)}{\partial \theta_2'} \\ \frac{\partial \ell_{i1}(\theta_1)}{\partial \theta_1}\frac{\partial \ell_i(\psi)}{\partial \phi'} & \frac{\partial \ell_{i1}(\theta_1)}{\partial \theta_1}\frac{\partial \ell_{i1}(\theta_1)}{\partial \theta_1'} & \frac{\partial \ell_{i1}(\theta_1)}{\partial \theta_1}\frac{\partial \ell_{i2}(\theta_2)}{\partial \theta_2'} \\ \frac{\partial \ell_{i2}(\theta_2)}{\partial \theta_2}\frac{\partial \ell_i(\psi)}{\partial \phi'} & \frac{\partial \ell_{i2}(\theta_2)}{\partial \theta_2}\frac{\partial \ell_{i1}(\theta_1)}{\partial \theta_1'} & \frac{\partial \ell_{i2}(\theta_2)}{\partial \theta_2}\frac{\partial \ell_{i2}(\theta_2)}{\partial \theta_2'} \end{array} \right) \Bigg|_{\psi=\widehat{\psi}},$$

and $\widehat{\psi} = (\widetilde{\phi}, \widehat{\theta}_1', \widehat{\theta}_2')'$.

In Table 4.3, the empirical biases are small for the estimates of the association parameter under different scenarios and there is close agreement between the empirical standard error and the standard errors proposed by us and estimated by (4.2.10) and (4.2.11). The standard error estimated by the parametric treatment of the problem tends to overestimate the empirical standard error, particularly with the smaller sample size and when the

dependence between survival times is weaker.

4.3.2 Bivariate Current Status Survival Data

We next consider the case with bivariate current status data where the times are generated by unit exponential margins. The performance of the proposed estimator is evaluated under three dependence levels as before wherein Kendall's τ is equal to 0.25, 0.50 or 0.75. We define $P(\Delta_j = 1)$ as the prevalence level and set it to be 0.5. To achieve this we generate the inspection times from a uniform distribution on $[0, 1.594]$. We generate 2000 samples with $m = 200$ or 400 subjects per sample for each dependence level.

The first-stage estimation $\widehat{F}_j^*(t)$ is given in (4.2.4) and the second stage estimation is conducted by plugging the marginal estimators $\widehat{F}_j^*(t)$ into (4.2.5). The ASE is calculated based on the method proposed in Section 4.2.5. The kernel function estimators of the marginal density function $f_j(t)$ and $h_j(t)$ are given as in (4.3.1) and (4.3.2). Bandwidth selection is similarly conducted as in Section 4.3.1.

Table 4.4 reports the empirical frequency properties of the proposed estimates for bivariate current status data. The empirical biases are, in general, very small. There is a close agreement between empirical standard errors and average large sample standard errors, and the empirical coverage probabilities are close to the nominal 95% level. This proposed approach to variance estimation for $\widetilde{\phi}$ appears to perform satisfactorily for bivariate current status data.

Table 4.4: Empirical frequency properties of the estimates and associated variance estimates for the association parameter of bivariate current status data with 50% prevalence level; 2000 simulated samples

n	Kendall's τ	BIAS	ESE	ASE	ECP
200	0.25	0.009	0.410	0.412	0.948
	0.50	0.057	0.275	0.271	0.944
	0.75	0.072	0.307	0.292	0.955
400	0.25	0.009	0.279	0.278	0.953
	0.50	0.031	0.185	0.183	0.947
	0.75	0.041	0.201	0.191	0.942

4.3.3 Absorption Times in Bivariate Multistate Processes

We consider bivariate multistate processes as formulated in Chapter 3. Under this set-up, we have two progressive processes with three states in each. We consider a single binary covariate Z with $P(Z = 1) = 0.5$. All transition times, T_{jk}, $j, k = 1, 2$, are generated from the joint density (3.2.7) where the marginal processes are progressive time-homogeneous Markov processes with transition intensity $\lambda_{jk}(t|z; \theta_{jk}) = \alpha_{jk} \exp(z\beta_{jk})$ for $j, k = 1, 2$, and a bivariate Clayton copula is adopted for all bivariate densities. The same parameter settings used in Section 3.4.1 are considered here. The left multistate process is observed subject to uniformly-distributed random censoring time on $[0.5, 1.5]$ and the right process is subject to uniformly-distributed random censoring time on $[0.5, 2]$. Two thousand samples of $m = 500$ individuals each are simulated.

The interest lies in estimating the coefficients in the marginal processes and the association parameter in the copula governing the absorption times of the two processes. We

adopt a semiparametric framework to analyze the marginal multistate process by assuming that the marginal intensity function is of multiplicative form

$$\lambda_{jk}(t|z;\beta_{jk}) = \lambda_{jk0}(t)\exp(z\beta_{jk}) \ ,$$

where the baseline intensity function $\lambda_{jk0}(t)$ is not assumed to have any particular parametric form. Under the working independence assumption between processes, inference regarding β_{jk} can be carried out by constructing the marginal partial likelihood (Andersen and Gill, 1982)

$$L(\beta_{jk}) = \prod_{i=1}^{m}\left[\frac{\exp(z_i\beta_{jk})}{\sum_{\ell=1}^{m}\bar{Y}_{jk\ell}(X_{jki})\exp(z_\ell\beta_{jk})}\right]^{\Delta_{jki}} \ , \tag{4.3.6}$$

where $X_{jki} = \min(T_{jki}, C_A)$, $\Delta_{jki} = I(T_{jki} \leq C_A)$, $i, = 1, \ldots, m$, $j, k = 1, 2$, $\bar{Y}_{j1\ell}(u) = I(u \leq X_{j1\ell})$ and $\bar{Y}_{j2\ell}(u) = I(X_{j1\ell} < u \leq X_{j2\ell})$. In the first stage, let $\widehat{\beta}_{jk}$ be the value that maximizes (4.3.6) and let the estimator for the corresponding baseline intensity be (Breslow, 1972)

$$d\widehat{\Lambda}_{jk0}(X_{jki}) = \frac{\Delta_{jki}}{\sum_{\ell=1}^{m}\bar{Y}_{jk\ell}(X_{jki})\exp(z_\ell\widehat{\beta}_{jk})} \ .$$

We estimate the marginal Markov transition probability matrix $\mathbb{P}_j(0,t|z)$ of process j by plugging in $d\widehat{\Lambda}_{jk}(\cdot|z) = d\widehat{\Lambda}_{jk0}(\cdot)\exp(z\widehat{\beta}_{jk})$ into (1.1.1) and the estimated survivor function of the absorption time given covariate $Z = z$, $\widehat{\mathcal{F}}_{j2}(t|z)$, is obtained as the complement of the $[1,3]$ entry of $\widehat{\mathbb{P}}_j(0,t|z)$. In the second stage, the estimates of the association parameters ϕ maximize the function (4.2.5) evaluated at $(\widehat{\mathcal{F}}_{12}(X_{12i}|Z_i), \widehat{\mathcal{F}}_{22}(X_{22i}|Z_i), \Delta_{12i}, \Delta_{22i})$.

Table 4.5: Empirical frequency properties of the estimates and associated variance estimates for the parameters of bivariate multistate processes with random right-censoring; $m = 500$; 2000 simulated samples

	BIAS	ESE	ASE_1	ECP_1	ASE_2	ECP_2	ASE_3	ECP_3
				Kendall's τ=0.2				
β_{11}	0.002	0.096	0.099	0.960	0.099	0.960	0.099	0.960
β_{12}	-0.003	0.113	0.112	0.952	0.112	0.952	0.113	0.953
β_{21}	0.001	0.096	0.096	0.957	0.096	0.957	0.097	0.957
β_{22}	0.002	0.108	0.106	0.948	0.106	0.948	0.106	0.950
$\log \phi$	-0.013	0.214	0.211	0.954	0.210	0.953	0.201	0.944
				Kendall's τ=0.4				
β_{11}	0.003	0.098	0.098	0.953	0.098	0.953	0.099	0.954
β_{12}	0.001	0.112	0.111	0.948	0.111	0.948	0.112	0.950
β_{21}	0.002	0.096	0.096	0.954	0.096	0.954	0.096	0.953
β_{22}	0.004	0.107	0.105	0.945	0.105	0.945	0.105	0.946
$\log \phi$	-0.004	0.124	0.126	0.951	0.121	0.942	0.106	0.904
				Kendall's τ=0.6				
β_{11}	0.003	0.097	0.098	0.954	0.098	0.954	0.098	0.954
β_{12}	0.001	0.110	0.110	0.950	0.110	0.950	0.110	0.953
β_{21}	0.002	0.096	0.096	0.954	0.096	0.954	0.096	0.957
β_{22}	0.004	0.106	0.104	0.947	0.104	0.947	0.104	0.951
$\log \phi$	-0.004	0.094	0.098	0.956	0.094	0.947	0.072	0.866

1. average large sample standard error estimate from proposed method in Section 4.2.6; 2. average large sample standard error estimate by a parametric approach; 3. naive average large sample standard error estimate.

Empirical frequency properties of both the marginal coefficients and the association parameter and their respective standard error estimates are reported in Table 4.5. The ASE_1 is calculated based on the method proposed in Section 4.2.6. The term ASE_2 is computed based on the *ad hoc* standard method as illustrated in Section 4.2.7 and 4.3.1, The ASE_3 is average naive standard error computed directly by taking the square root of inverse of the Fisher information function corresponding to the marginal likelihood $\prod_{i=1}^{m} L_i(\beta_{jk})$ and the second-stage likelihood (4.2.5) with respect to ϕ.

In Table 4.5, we observe small empirical biases for all the estimates. The naive ASE_3 is inappropriately smaller than the empirical standard error and the corresponding empirical coverage probability is much lower than 95% when Kendall's τ is large. The other two standard error estimates behave similarly well in the sense that they all have a reasonably close agreement to the empirical standard errors and the corresponding empirical coverage probabilities are around the nominal 95% level. There is slightly closer agreement between ASE_2 and the ESE when $\tau = 0.6$ but very little difference for the other values of Kendall's τ.

4.4 Application

4.4.1 Right-censored Data

Hortobagyi *and others* (1996, 1998) report on an international multicenter trial of 382 women with stage IV breast cancer with skeletal metastases in which patients were randomized to receive pamidronate ($Z = 1$) or a placebo control ($Z = 0$) as mentioned in

Table 4.6: Results from semiparametric two-stage estimation of the association between the survival times for two types of events, fracture and need for radiotherapy

Number of Samples	Setting	Parameter	EST	SE_1	SE_2	SE_3
One	-	$\log(\phi)$	-0.007	0.284	0.283	0.282
		τ	0.332	0.126	0.125	0.125
Two	Common β	$\log(\phi)$	-0.111	0.302	-	0.307
		τ	0.309	0.129	-	0.131
Two	Different β	$\log(\phi)$	-0.066	0.297	-	0.308
		τ	0.319	0.129	-	0.134
Two	Stratified-Treated	$\log(\phi)$	-0.200	0.550	0.551	0.524
		τ	0.290	0.227	0.227	0.216
Two	Stratified-Control	$\log(\phi)$	-0.026	0.347	0.357	0.374
		τ	0.328	0.153	0.157	0.165

1. standard error estimate from proposed method in Section 4.2.5 and 4.2.6; 2. standard error estimate of Shih and Louis (1995); 3. standard error estimate by a parametric approach.

Section 1.2.3. Patients were monitored closely and the occurrence of pathologic fractures and need for radiotherapy for the treatment of bone pain were recorded. Each patient was followed until death, loss of follow-up, or close of the study. Special attention is paid to patients who survived 12 months.

Semiparametric two-stage estimation for the survival times of the two types of events, fracture and need for radiotherapy, is conducted and results are summarized in Table 4.6. Here SE_1 is a standard error computed based on our proposed method, SE_2 is calculated by the method in Shih and Louis (1995) and SE_3 uses the *ad hoc* "parametric" method.

We first conduct a one-sample analysis ignoring the covariate effect for all the subjects. The estimated Kendall's τ is 0.332 (95% CI_1: 0.085, 0.579; $\text{p}-\text{value}_1$=0.008), indicating a moderate and significant dependence structure between the two types of events. We then consider a regression analysis. If we constrain the coefficients in the two marginal models to be the same and denote it by β, $\widehat{\beta} = -0.696$ (95% CI: -0.993, -0.399; p−value<0.001) and the estimated Kendall's τ is 0.309 (95% CI_1: 0.056, 0.562; $\text{p}-\text{value}_1$=0.017). If we allow the coefficients in the two marginal models to differ and denote them by β_1 and β_2, we obtain $\widehat{\beta}_1 = -0.475$ (95% CI: -0.872, -0.079; p−value=0.019), $\widehat{\beta}_2 = -0.964$ (95% CI: -1.393, -0.535; p−value<0.001), and the estimated Kendall's τ is 0.319 (95% CI_1: 0.066, 0.572; $\text{p}-\text{value}_1$=0.013). In a third analysis, we stratify based on the treatment covariate and conduct a one-sample analysis in each subsample. In the treatment group, the estimated Kendall's τ is 0.290 (95% CI_1: -0.155, 0.735; $\text{p}-\text{value}_1$=0.200) and not significant; in the control group, the estimated Kendall's τ is comparable at 0.328 (95% CI_1: 0.028, 0.628; $\text{p}-\text{value}_1$=0.032) and it is significant.

4.4.2 Current Status Data

As introduced in Section 1.2.4, four large recent multicenter randomized trials were conducted to compare enoxaparin and fondaparinux for thromboprophylaxis. Interest lies in understanding the factors associated with antibody responses caused by antithrombotic drugs (Warkentin *and others*, 2005). The two event times of interest are the time from surgery to the antibody response of EIA-A, denoted by T_1, and the time from surgery to the antibody response of HAM GTI, denoted by T_2. C is the time from surgery to the

blood test and the corresponding indicator is $\Delta_j = I(T_j < C)$, $j = 1, 2$. $\Delta_j = 1$ if antibody positive at C, and $\Delta_j = 0$ otherwise, for $j = 1, 2$.

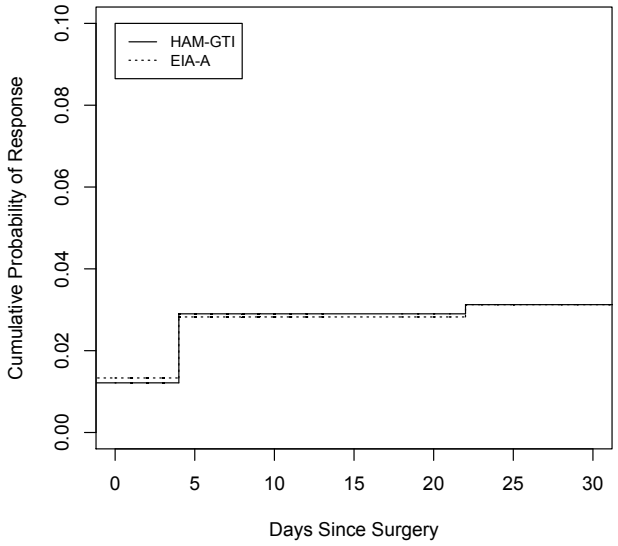

Figure 4.1: Plots of the cumulative probability of responses vs. time since surgery (days)

In Figure 4.1, we plot the cumulative probability of response versus time since surgery (days) for HAM GTI and EIA-A. Both curves are very close. Since the prevalence level for HAM GTI is 1.65% and that for EIA-A is 1.57%, the estimated cumulative probability is small. We conduct a one sample analysis for all the subjects. The estimated Kendall's τ is 0.914 (95% CI: 0.838, 0.990; p$-$value$<$ 0.001) indicating a very strong and significant association between the time to the two types of antibody responses.

4.5 General Remarks

We proposed a new estimator of the asymptotic variance of the second-stage estimator for the association parameter for bivariate event times. We proved the asymptotic properties and gave illustrations through simulation studies and the analysis of data from motivating studies under different scenarios. In the empirical studies, we found the proposed standard error estimator behaves very well in general in the sense that it often has closer agreement to the empirical standard error than those computed based on other methods, and this trend is more obvious when the sample size is smaller. The other advantage of our proposed estimator is that it bears a general form which applies to complete, right-censored, and current status data with or without covariate.

Kernel smoothing is required to obtain the estimated densify functions $\widehat{f}_j(\cdot)$ and $\widehat{h}_j(\cdot)$. Bandwidth selection is important and the ideal bandwidth should be selected by cross-validation.

Appendix A: Asymptotic Normality of $\widetilde{\phi}$

In deriving the limiting distribution of the two-stage estimator of ϕ, we assume that the following regularity conditions hold:

(a) Let $\{(X_{1i}, X_{2i}, \Delta_{1i}, \Delta_{2i}); i = 1, \ldots, m\}$ be independent and identically distributed (i.i.d.), each with density $h(x_1, x_2, \delta_1, \delta_2; \phi)$ depending on some real parameter $\phi \in \Phi$, where ϕ is a (non-empty) open interval in the real line.

(b) The set $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$ does not depend on the copula parameter ϕ, where

$$\mathcal{A}_j = \mathcal{B}_j \times \{0,1\} \text{ and } \mathcal{B}_j = \{x > 0 : h_j(x) > 0\}\,,\ j = 1,2.$$

(c) Define

$$\begin{aligned} V'_{\phi}(\phi; u_1, u_2, \delta_1, \delta_2) &= \frac{\partial^3 \ell(\phi; u_1, u_2, \delta_1, \delta_2)}{\partial \phi^3}\,, \\ V'_{\phi,1}(\phi; u_1, u_2, \delta_1, \delta_2) &= \frac{\partial^4 \ell(\phi; u_1, u_2, \delta_1, \delta_2)}{\partial \phi^3 \partial u_1}\,, \\ V'_{\phi,2}(\phi; u_1, u_2, \delta_1, \delta_2) &= \frac{\partial^4 \ell(\phi; u_1, u_2, \delta_1, \delta_2)}{\partial \phi^3 \partial u_2}\,. \end{aligned}$$

We assume that

(i) $U_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$, $V_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$, $V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)$, $V'_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$, and $V'_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)$ exist;

(ii) $U_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$, $V_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$ and $V'_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$ are continuous functions of u and v;

(iii) $U_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$, $V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)$, and $V'_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)$ are bounded functions for $(u_1, u_2, \delta_1, \delta_2) \in [0,1] \times [0,1] \times \{0,1\} \times \{0,1\}$ and all $\phi \in \Phi$, for $j = 1, 2$.

(d) It is possible to interchange derivatives and integrals so that

$$\frac{\partial^k}{\partial \phi^k} \int_{\mathcal{B}_2} \int_{\mathcal{B}_1} h(x_1, x_2, \delta_1, \delta_2; \phi) dx_1 dx_2 = \int_{\mathcal{B}_2} \int_{\mathcal{B}_1} \frac{\partial^k}{\partial \phi^k} h(x_1, x_2, \delta_1, \delta_2; \phi) dx_1 dx_2$$

for $k = 1, 2$.

(e) $\mathrm{E}\left[-V_\phi(\phi; u_1, u_2, \Delta_1, \Delta_2)\right]$ is strictly positive and finite for all $\phi \in \Phi$.

(f) There exists a real-valued function $b(x_1, x_2, \delta_1, \delta_2)$ which does not depend on ϕ, such that

$$\left|V'_\phi(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2)\right| \leq b(x_1, x_2, \delta_1, \delta_2)$$

for all $(x_1, x_2, \delta_1, \delta_2) \in \mathcal{A}$, and such that $\mathrm{E}[b(X_1, X_2, \Delta_1, \Delta_2)] < \infty$, for all $\phi \in \Phi$.

(g) $\widehat{\mathcal{F}}_j(x_j)$ converges in probability to $\mathcal{F}_j(x_j)$ uniformly in $x_j \in \mathcal{B}_j$, for $j = 1, 2$.

(h) There exists a differentiable function $q_j(\cdot)$ and its inverse function $q_j^{-1}(\cdot)$ such that $\mathcal{F}_j(x_j) = q_j(\mathcal{H}_j(x_j))$ for $x_j \in \mathcal{B}_j$.

Proof of Theorem 1: To simplify notation, we set

$$U_\phi(\phi; \mathcal{F}_1, \mathcal{F}_2, H_m) := \sum_{i=1}^{m} U_\phi(\phi; \mathcal{F}_1(X_{1i}), \mathcal{F}_2(X_{2i}), \Delta_{1i}, \Delta_{2i}) \ ,$$

and similarly define $V_\phi(\phi; \mathcal{F}_1, \mathcal{F}_2, H_m)$ and $V'_\phi(\phi; \mathcal{F}_1, \mathcal{F}_2, H_m)$. Expanding the score function $U_\phi(\phi; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m)$ in a Taylor series around ϕ and evaluating it at $\widetilde{\phi}$, we obtain

$$\begin{aligned} 0 &= U_\phi(\widetilde{\phi}; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m) \\ &= U_\phi(\phi; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m) + (\widetilde{\phi} - \phi) V_\phi(\phi; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m) + \frac{1}{2}(\widetilde{\phi} - \phi)^2 V'_\phi(\bar{\phi}; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m) \ , \end{aligned}$$

where $\bar{\phi}$ is some value between ϕ and $\widetilde{\phi}$. Thus

$$\sqrt{m}(\widetilde{\phi}-\phi)=\frac{U_\phi(\phi;\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)/\sqrt{m}}{-V_\phi(\phi;\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)/m-V'_\phi(\bar{\phi};\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)(\widetilde{\phi}-\phi)/(2m)}\ .$$

The desired result therefore follows from Slutsky's lemma (van der Vaart, 2000) if we can show as $m\to\infty$,

$$-V_\phi(\phi;\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)/m-V'_\phi(\bar{\phi};\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)(\widetilde{\phi}-\phi)/(2m)\xrightarrow{P}\sigma_1^2\ , \tag{4.A.1}$$

and

$$U_\phi(\phi;\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)/\sqrt{m}\xrightarrow{D}N(0,\sigma_1^2+\sigma_2^2)\ . \tag{4.A.2}$$

Proof of (4.A.1)

We first work on the asymptotic behaviour of $-V_\phi(\phi;\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)/m$. We begin with the proof of the asymptotic equivalence between $V_\phi(\phi;\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)/m$ and $V_\phi(\phi;\mathcal{F}_1,\mathcal{F}_2,H_m)/m$ in the sense that

$$V_\phi(\phi;\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)/m-V_\phi(\phi;\mathcal{F}_1,\mathcal{F}_2,H_m)/m\ ,$$

denoted by D_m, converges in probability to zero. We know that

$$\begin{aligned}
|D_m| &= \left| \int_{\mathcal{A}} V_\phi(\phi; \widehat{\mathcal{F}}_1(x_1), \widehat{\mathcal{F}}_2(x_2), \delta_1, \delta_2) - V_\phi(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) dH_m(x_1, x_2, \delta_1, \delta_2) \right| \\
&\leq \int_{\mathcal{A}} \left| V_\phi(\phi; \widehat{\mathcal{F}}_1(x_1), \widehat{\mathcal{F}}_2(x_2), \delta_1, \delta_2) - V_\phi(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \right| dH_m(x_1, x_2, \delta_1, \delta_2) \\
&\leq \int_{\mathcal{A}} \left| V_\phi(\phi; \widehat{\mathcal{F}}_1(x_1), \widehat{\mathcal{F}}_2(x_2), \delta_1, \delta_2) - V_\phi(\phi; \mathcal{F}_1(x_1), \widehat{\mathcal{F}}_2(x_2), \delta_1, \delta_2) \right| dH_m(x_1, x_2, \delta_1, \delta_2) \\
&+ \int_{\mathcal{A}} \left| V_\phi(\phi; \mathcal{F}_1(x_1), \widehat{\mathcal{F}}_2(x_2), \delta_1, \delta_2) - V_\phi(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \right| dH_m(x_1, x_2, \delta_1, \delta_2) \, .
\end{aligned}$$

From regularity condition (c), $V_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$ is continuous in u_1 and u_2 and there exists a constant M_j such that $|V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)| \leq M_j$ on $(u_1, u_2, \delta_1, \delta_2) \in [0,1] \times [0,1] \times \{0,1\} \times \{0,1\}$ and all $\phi \in \Phi$, $j = 1, 2$. We have then

$$\begin{aligned}
|D_m| &\leq \int_{\mathcal{A}} M_1 \left| \widehat{\mathcal{F}}_1(x_1) - \mathcal{F}_1(x_1) \right| dH_m(x_1, x_2, \delta_1, \delta_2) \\
&+ \int_{\mathcal{A}} M_2 \left| \widehat{\mathcal{F}}_2(x_2) - \mathcal{F}_2(x_2) \right| dH_m(x_1, x_2, \delta_1, \delta_2) \\
&\leq M_1 \sup_{x_1 \in \mathcal{B}_1} \left| \widehat{\mathcal{F}}_1(x_1) - \mathcal{F}_1(x_1) \right| + M_2 \sup_{x_2 \in \mathcal{B}_2} \left| \widehat{\mathcal{F}}_2(x_2) - \mathcal{F}_2(x_2) \right| \, .
\end{aligned}$$

In view of regularity condition (g), $|D_m| \xrightarrow{P} 0$, i.e.,

$$-\frac{1}{m} V_\phi(\phi; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m) + \frac{1}{m} V_\phi(\phi; \mathcal{F}_1, \mathcal{F}_2, H_m) \xrightarrow{P} 0 \, .$$

By regularity condition (a) and the law of large numbers (van der Vaart, 2000),

$$-\frac{1}{m} V_\phi(\phi; \mathcal{F}_1, \mathcal{F}_2, H_m) \xrightarrow{P} \sigma_1^2 \, ,$$

where σ_1^2 is strictly positive and finite by regularity condition (e). We obtain

$$-\frac{1}{m}V_\phi(\phi;\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)\xrightarrow{P}\sigma_1^2\ .$$

We next work on the asymptotic behaviour of $-V'_\phi(\phi;\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)/m$. By the continuity of V'_ϕ in u and v, and the boundedness of $V'_{\phi,j}$, $j=1,2$, in regularity condition (c) and condition (g), we can show that $V'_\phi(\bar{\phi};\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)/m$ is asymptotically equivalent to $V'_\phi(\bar{\phi};\mathcal{F}_1,\mathcal{F}_2,H_m)/m$. By regularity conditions (a), (f) and the law of large numbers, we have

$$\begin{aligned}\left|-\frac{1}{2m}V'_\phi(\bar{\phi};\mathcal{F}_1,\mathcal{F}_2,H_m)\right| &\leq \frac{1}{2m}\sum_{i=1}^m b(X_{1i},X_{2i},\Delta_{1i},\Delta_{2i})\\ &\xrightarrow{P} \frac{1}{2}\mathrm{E}[b(X_1,X_2,\Delta_1,\Delta_2)]<\infty\ .\end{aligned}$$

Since $\widetilde{\phi}\xrightarrow{P}\phi$, (4.A.1) has limit

$$-V_\phi(\phi;\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)/m-V'_\phi(\bar{\phi};\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m)(\widetilde{\phi}-\phi)/(2m)\xrightarrow{P}\sigma_1^2\ .$$

Proof of (4.A.2)

We first decompose

$$\frac{1}{\sqrt{m}}U_\phi(\phi;\widehat{\mathcal{F}}_1,\widehat{\mathcal{F}}_2,H_m) = \sqrt{m}\int_{\mathcal{A}}U_\phi(\phi;\widehat{\mathcal{F}}_1(x_1),\widehat{\mathcal{F}}_2(x_2),\delta_1,\delta_2)dH_m(x_1,x_2,\delta_1,\delta_2)$$

into the summation of three terms: $R_1(\phi)$, $R_2(\phi)$ and $R_3(\phi)$, where

$$\begin{aligned} R_1(\phi) &= \sqrt{m}\int_{\mathcal{A}} \left[U_\phi(\phi; \widehat{\mathcal{F}}_1(x_1), \widehat{\mathcal{F}}_2(x_2), \delta_1, \delta_2) - U_\phi(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2)\right] \\ &\quad d(H_m - H)(x_1, x_2, \delta_1, \delta_2) \ , \\ R_2(\phi) &= \sqrt{m}\int_{\mathcal{A}} U_\phi(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) d(H_m - H)(x_1, x_2, \delta_1, \delta_2) \ , \\ R_3(\phi) &= \sqrt{m}\int_{\mathcal{A}} U_\phi(\phi; \widehat{\mathcal{F}}_1(x_1), \widehat{\mathcal{F}}_2(x_2), \delta_1, \delta_2) dH(x_1, x_2, \delta_1, \delta_2) \ . \end{aligned}$$

In view of the assumption that $U_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$ is continuous and bounded by condition (c), condition (g), and the fact $\sqrt{m}(H_m - H) = O_p(1)$, $R_1(\phi)$ converges in probability to 0. Since regularity condition (d) allows for the interchanging of derivatives and integrals, we have

$$\mathrm{E}\left[U_\phi(\phi; \mathcal{F}_1(X_1), \mathcal{F}_2(X_2), \Delta_1, \Delta_2)\right] = 0 \ , \tag{4.A.3}$$

and

$$\mathrm{Var}\left[U_\phi(\phi; \mathcal{F}_1(X_1), \mathcal{F}_2(X_2), \Delta_1, \Delta_2)\right] = \sigma_1^2 \ .$$

Note that $R_2(\phi)$ is a sum of i.i.d. random variables of mean zero and variance σ_1^2. By regularity condition (h), $\mathcal{F}_j(x_j) = q_j(\mathcal{H}_j(x_j))$ on $x_j \in \mathcal{B}_j$, for $j = 1, 2$. Let $\widehat{\mathcal{H}}_j(x_j) = q_j^{-1}(\widehat{\mathcal{F}}_j(x_j))$. Using von Mises expansion (Fernholz, 1983) of $R_3(\phi) = R_3(\phi; q_1(\widehat{\mathcal{H}}_1), q_2(\widehat{\mathcal{H}}_2))$

around $\mathcal{H}_1, \mathcal{H}_2$, we get

$$\begin{aligned} R_3(\phi; q_1(\widehat{\mathcal{H}}_1), q_2(\widehat{\mathcal{H}}_2)) &= R_3(\phi; q_1(\mathcal{H}_1), q_2(\mathcal{H}_2)) - \int_{\mathcal{B}_1} \psi_{\mathcal{H}_1}(s_1) d(\widehat{\mathcal{H}}_1 - \mathcal{H}_1)(s_1) \\ &- \int_{\mathcal{B}_2} \psi_{\mathcal{H}_2}(s_2) d(\widehat{\mathcal{H}}_2 - \mathcal{H}_2)(s_2) + o_p(1) \ , \end{aligned} \tag{4.A.4}$$

where $\psi_{\mathcal{H}_j}$ is given as follows

$$\begin{aligned} \psi_{\mathcal{H}_j}(s_j) &= \frac{\partial}{\partial \epsilon_j} R_3(\phi; q_1(\mathcal{H}_1 + \epsilon_1(\eta_{s_1} - \mathcal{H}_1)), q_2(\mathcal{H}_2 + \epsilon_2(\eta_{s_2} - \mathcal{H}_2)))|_{\epsilon_1 = \epsilon_2 = 0} \\ &= \sqrt{m} \int_{\mathcal{A}} V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} (\eta_{s_j} - \mathcal{H}_j)(x_j) dH(x_1, x_2, \delta_1, \delta_2) \ , \end{aligned}$$

and $\eta_{s_j}(x_j) = I(s_j \geq x_j)$, the survivor function of the point mass one at s_j.

Then $R_3(\phi; q_1(\mathcal{H}_1), q_2(\mathcal{H}_2))$, the first term on the right of the equality sign in (4.A.4), is equal to zero. Next, we consider the integrals in (4.A.4). For $j = 1, 2$, we have

$$\begin{aligned} & \int_{\mathcal{B}_j} \psi_{\mathcal{H}_j}(s_j) d\mathcal{H}_j(s_j) \\ = \ & \sqrt{m} \int_{\mathcal{A}} V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} \left\{ \int_{\mathcal{B}_j} \left[I(s_j \geq x_j) - \mathcal{H}_j(x_j) \right] d\mathcal{H}_j(s_j) \right\} \\ & dH(x_1, x_2, \delta_1, \delta_2) \\ = \ & -\sqrt{m} \int_{\mathcal{A}} V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} \left[\mathcal{H}_j(x_j) - 1 \cdot \mathcal{H}_j(x_j) \right] dH(x_1, x_2, \delta_1, \delta_2) \\ = \ & 0 \ , \end{aligned}$$

and

$$
\begin{aligned}
- & \int_{\mathcal{B}_j} \psi_{\mathcal{H}_j}(s_j) d\widehat{\mathcal{H}}_j(s_j) = \frac{1}{m} \sum_{i=1}^{m} \psi_{\mathcal{H}_j}(X_{ji}) \\
= & \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \int_{\mathcal{A}} V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} \left[I(X_{ji} \geq x_j) - \mathcal{H}_j(x_j) \right] dH(x_1, x_2, \delta_1, \delta_2) \, .
\end{aligned}
\tag{4.A.5}
$$

Let $W_j(\phi, X_j) = \int_{\mathcal{A}} V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} I(X_j \geq x_j) dH(x_1, x_2, \delta_1, \delta_2)$, and then

$$
\begin{aligned}
\mathrm{E}\left[W_j(\phi, X_j) \right] &= - \int_{\mathcal{B}_j} W_j(\phi, s_j) d\mathcal{H}_j(s_j) \\
&= \int_{\mathcal{A}} V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} \mathcal{H}_j(x_j) dH(x_1, x_2, \delta_1, \delta_2) \, .
\end{aligned}
$$

Thus (4.A.5) is a sum of m i.i.d. random variables with mean zero and variance $\mathrm{E}\left[W_j(\phi, X_j) \right]^2$, and $R_3(\phi)$ is a sum of i.i.d. random variables of mean zero and variance

$$
\sigma_2^2 = \mathrm{E}\left[W_1(\phi, X_1) + W_2(\phi, X_2) \right]^2 .
$$

In the end, we show the covariance between $U_\phi(\phi; \mathcal{F}_1(X_1), \mathcal{F}_2(X_2), \Delta_1, \Delta_2)$ and $W_1(\phi, X_1) +$

$W_2(\phi, X_2)$ is equal to zero. Due to (4.A.3), it is sufficient to show

$$\begin{aligned}
& \mathrm{E}\left[W_1(\phi, X_1) \cdot U_\phi(\phi; \mathcal{F}_1(X_1), \mathcal{F}_2(X_2), \Delta_1, \Delta_2)\right] \\
= & \int_{\mathcal{A}} W_1(\phi, x_1) \cdot \frac{\partial h(x_1, x_2, \delta_1, \delta_2; \phi)/\partial\phi}{h(x_1, x_2, \delta_1, \delta_2; \phi)} \; dH(x_1, x_2, \delta_1, \delta_2) \\
= & \sum_{\delta_1=0}^{1} \sum_{\delta_2=0}^{1} \int_{\mathcal{B}_1} W_1(\phi, x_1) \left[\frac{\partial}{\partial\phi} \int_{\mathcal{B}_2} h(x_1, x_2, \delta_1, \delta_2; \phi) dx_2 \right] dx_1 \quad \text{(by condition (d))} \\
= & \sum_{\delta_1=0}^{1} \sum_{\delta_2=0}^{1} \int_{\mathcal{B}_1} W_1(\phi, x_1) \cdot 0 \; \; dx_1 = 0 \; ,
\end{aligned}$$

and $\mathrm{E}\left[W_2(\phi, X_2) \cdot U_\phi(\phi; \mathcal{F}_1(X_1), \mathcal{F}_2(X_2), \Delta_1, \Delta_2)\right] = 0$ can be similarly proved.

Ultimately then

$$\begin{aligned}
\frac{1}{\sqrt{m}} U_\phi(\phi; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m) \; = \; & \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \left\{ U_\phi(\phi; \mathcal{F}_1(X_{1i}), \mathcal{F}_2(X_{2i}), \Delta_{1i}, \Delta_{2i}) + W_1(\phi, X_{1i}) \right. \\
+ \; & \left. W_2(\phi, X_{2i}) - \mathrm{E}\left[W_1(\phi, X_{1i}) + W_2(\phi, X_{2i})\right] \right\} + o_p(1) \; ,
\end{aligned}$$

which, by the central limit theorem, converges to a normal distribution with mean zero and variance $\sigma_1^2 + \sigma_2^2$ as $m \to \infty$.

Appendix B: Consistency of the Asymptotic Variance Estimator

In addition to the regularity conditions (a)-(h) stated at the beginning of Appendix A here, we further assume that the following regularity conditions hold:

(i) $\widehat{f}_j(x_j)$ converges in probability to $f_j(x_j)$ uniformly and $\widehat{h}_j(x_j)$ converges in probability to $h_j(x_j)$ uniformly for $x_j \in \mathcal{B}_j$, for $j = 1, 2$.

(j) $V_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$ is continuous of ϕ and $V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)$ are continuous in u_1, u_2 and ϕ; $V'_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$, $\partial V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)/\partial\phi$, $\partial V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)/\partial u_1$, and $\partial V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)/\partial u_2$ are bounded functions for $(u_1, u_2, \delta_1, \delta_2) \in [0,1] \times [0,1] \times \{0,1\} \times \{0,1\}$, and $f_j(x_j)/h_j(x_j)$ is bounded on $x_j \in \mathcal{B}_j$ and all $\phi \in \phi$, for $j = 1, 2$.

Proof of the Convergence of $\widehat{\sigma}_1^2 \xrightarrow{P} \sigma_1^2$

The asymptotic equivalence between $\widehat{\sigma}_1^2 = V_\phi(\widetilde{\phi}; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m)/m$ and $V_\phi(\phi; \mathcal{F}_1, \mathcal{F}_2, H_m)/m$ can be justified in the same way the asymptotic equivalence of $V_\phi(\phi; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m)/m$ and $V_\phi(\phi; \mathcal{F}_1, \mathcal{F}_2, H_m)/m$ was demonstrated in the proof of Theorem 1, with additional requirements in regularity conditions (j) that $\widetilde{\phi} \xrightarrow{P} \phi$ and $V_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$ is continuous in ϕ and $V'_\phi(\phi; u_1, u_2, \delta_1, \delta_2)$ is bounded, for $j = 1, 2$. We then have

$$-\frac{1}{m} V_\phi(\widetilde{\phi}; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m)/m + \frac{1}{m} V_\phi(\phi; \mathcal{F}_1, \mathcal{F}_2, H_m) \xrightarrow{P} 0 \ ,$$

and by condition (a) and the law of large numbers,

$$-\frac{1}{m} V_\phi(\phi; \mathcal{F}_1, \mathcal{F}_2, H_m) \xrightarrow{P} \sigma_1^2 \ .$$

Thus, we obtain $\widehat{\sigma}_1^2 \xrightarrow{P} \sigma_1^2$.

Proof of the Convergence of $\widehat{\sigma}_2^2 \xrightarrow{P} \sigma_2^2$

Let $H(x_1, x_2)$ denote the joint c.d.f. of (X_1, X_2) and $H_m(x_1, x_2)$ is its empirical distribution function. We write

$$\begin{aligned}
\left|\widehat{\sigma}_2^2 - \sigma_2^2\right| &= \left| \int_{\mathcal{B}_1 \times \mathcal{B}_2} \left[\widehat{W}_1(\widetilde{\phi}, s_1) + \widehat{W}_2(\widetilde{\phi}, s_2)\right]^2 dH_m(s_1, s_2) \right. \\
&- \left. \int_{\mathcal{B}_1 \times \mathcal{B}_2} \left[W_1(\phi, s_1) + W_2(\phi, s_2)\right]^2 H(s_1, s_2) \right| \\
&\leq \int_{\mathcal{B}_1 \times \mathcal{B}_2} \left| \left\{ \left[\widehat{W}_1(\widetilde{\phi}, s_1) + \widehat{W}_2(\widetilde{\phi}, s_2)\right]^2 - \left[W_1(\phi, s_1) + W_2(\phi, s_2)\right]^2 \right\} \right| dH_m(s_1, s_2) \qquad (4.B.1) \\
&+ \int_{\mathcal{B}_1 \times \mathcal{B}_2} \left| \left[W_1(\phi, s_1) + W_2(\phi, s_2)\right]^2 \right| d(H_m - H)(s_1, s_2) \; . \qquad (4.B.2)
\end{aligned}$$

By condition (a) and the law of large numbers, (4.B.2) converges to zero in probability. For (4.B.1), we need to prove the asymptotic equivalence between

$$\int_{\mathcal{B}_1 \times \mathcal{B}_2} \left[\widehat{W}_1(\widetilde{\phi}, s_1) + \widehat{W}_2(\widetilde{\phi}, s_2)\right]^2 dH_m(s_1, s_2)$$

and

$$\int_{\mathcal{B}_1 \times \mathcal{B}_2} \left[W_1(\phi, s_1) + W_2(\phi, s_2)\right]^2 dH_m(s_1, s_2) \; ,$$

which can be justified in the same way we justified the asymptotic equivalence between $V_\phi(\phi; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m)/m$ and $V_\phi(\phi; \mathcal{F}_1, \mathcal{F}_2, H_m)/m$ in the proof of Theorem 1. We require the

conditions (1) $W_1(\phi, s_1) + W_2(\phi, s_2)$ is bounded on $(s_1, s_2) \in \mathcal{B}_1 \times \mathcal{B}_2$ and (2)

$$\sup_{s_j \in \mathcal{B}_j} \left| \widehat{W}_j(\widetilde{\phi}_j, s_j) - W_j(\phi_j, s_j) \right| \xrightarrow{P} 0 \ ,$$

for $j = 1, 2$. Condition (1) is satisfied since $V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)$ and $f_j(x_j)/h_j(x_j)$ are bounded as assumed in condition (c) and (j). Thus, to show the first term in (4.B.1) converges to zero in probability, it is sufficient to show condition (2):

$$\begin{aligned}
& \sup_{s_j \in \mathcal{B}_j} \left| \widehat{W}_j(\widetilde{\phi}_j, s_j) - W_j(\phi_j, s_j) \right| \\
\leq \ & \sup_{s_j \in \mathcal{B}_j} \int_{\mathcal{A}} \Big| \big[V_{\phi,j}(\widetilde{\phi}; \widehat{\mathcal{F}}_1(x_1), \widehat{\mathcal{F}}_2(x_2), \delta_{1k}, \delta_{2k}) \frac{\widehat{f}_j(x_j)}{\widehat{h}_j(x_j)} I(s_j \geq x_j) \\
- \ & V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} I(s_j \geq x_j) \big] \Big| dH_m(x_1, x_2, \delta_1, \delta_2) \\
+ \ & \sup_{s_j \in \mathcal{B}_j} \int_{\mathcal{A}} \left| V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} I(s_j \geq x_j) \right| d(H_m - H)(x_1, x_2, \delta_1, \delta_2) \\
\leq \ & \int_{\mathcal{A}} \left| \left[V_{\phi,j}(\widetilde{\phi}; \widehat{\mathcal{F}}_1(x_1), \widehat{\mathcal{F}}_2(x_2), \delta_{1k}, \delta_{2k}) \frac{\widehat{f}_j(x_j)}{\widehat{h}_j(x_j)} - V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} \right] \right| \\
& dH_m(x_1, x_2, \delta_1, \delta_2) \qquad (4.B.3) \\
+ \ & \int_{\mathcal{A}} \left| V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} \right| d(H_m - H)(x_1, x_2, \delta_1, \delta_2) \ . \qquad (4.B.4)
\end{aligned}$$

Note that (4.B.4) converges in probability to zero by condition (a) and the law of large numbers. For (4.B.3), the asymptotic equivalence between

$$\int_{\mathcal{A}} V_{\phi,j}(\widetilde{\phi}; \widehat{\mathcal{F}}_1(x_1), \widehat{\mathcal{F}}_2(x_2), \delta_{1k}, \delta_{2k}) \frac{\widehat{f}_j(x_j)}{\widehat{h}_j(x_j)} dH_m(x_1, x_2, \delta_1, \delta_2)$$

and

$$\int_{\mathcal{A}} V_{\phi,j}(\phi; \mathcal{F}_1(x_1), \mathcal{F}_2(x_2), \delta_1, \delta_2) \frac{f_j(x_j)}{h_j(x_j)} dH_m(x_1, x_2, \delta_1, \delta_2) \ ,$$

can be proven using a similar argument by showing asymptotic equivalence between $V_\phi(\phi; \widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2, H_m)/m$ and $V_\phi(\phi; \mathcal{F}_1, \mathcal{F}_2, H_m)/m$ in the proof of Theorem 1, which requires (1) consistency of $\widetilde{\phi}$ and uniform consistency of $\widehat{\mathcal{F}}_j(\cdot)$, $\widehat{f}_j(\cdot)$ and $\widehat{h}_j(\cdot)$, $j = 1, 2$, given in condition (g) and (i), (2) continuity of $V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)$ in ϕ, u_1 and u_2 given in condition (c) and (3) boundedness of $V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)$, $f_j(x_j)/h_j(x_j)$, $\partial V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)/\partial\phi$, $\partial V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)/\partial u_1$, and $\partial V_{\phi,j}(\phi; u_1, u_2, \delta_1, \delta_2)/\partial u_2$ for $(u_1, u_2, \delta_1, \delta_2) \in [0, 1] \times [0, 1] \times \{0, 1\} \times \{0, 1\}$ and all $\phi \in \phi$, for $j = 1, 2$, given in condition (c) and (j). Taken together these results demonstrate $\widehat{\sigma}_2^2 \xrightarrow{P} \sigma_2^2$.

Chapter 5

General Remarks and Future Research

5.1 Overview

The three research projects address different problems in the field of life history analysis but the techniques to handle the complex dependence structures have a common theme. All involve decomposing a multivariate density and modeling the conditional or unconditional pairwise distributions using bivariate copula functions. In fact, the models in Chapter 2 and 3 along with many other interesting topics on dependence modeling in lifetime history analysis, can be cast into the framework of vine copulas. Robust estimation and model misspecification are also part of an underlying theme of this research.

It would be interesting to explore issues of model misspecification in more details. Two

types model assumptions are usually made in modeling dependence structure to simplify the model, reduce computational burden, and lower the number of parameters. The first is the so-called conditional independence assumption (Aas *and others*, 2009). As is customary in hierarchical modeling, a model simplifies only if we avoid specifying all or some of the conditional pairwise densities. The second simplifying assumption is that the copulas corresponding to conditional distributions are constant in the sense that the the copula function and the association parameters are the same irrespective of the values of the variables that they are conditioned on (Stöber *and others*, 2012). This assumption is usually made to keep model selection and inference tractable. While these assumptions may seem restrictive at the first glance, if it is not true, the conditional independence assumption will simply lead to less efficient marginal estimates; this can be justified from a composite likelihood argument. Haff *and others* (2010) comment that the simplified pair-copula construction is a rather good approximation even when the simplifying assumption is far from the correct. It is of interest to study the effect of those assumptions on inferences for lifetime data which feature censoring or truncation. Another interesting question is regarding model selection. Vine copula selection requires two steps: one is selection of the vine model and the other is choice of bivariate copula functions for each pair of variables selected in the vine model. There are a number of different ways to decompose a multivariate joint density when the dimension is high. Selection of a reasonable vine model becomes an important topic and it has been discussed in two recent papers (Dißmann *and others*, 2013; Czado *and others*, 2013). Genest *and others* (2009) give an excellent review of methods on the assessment of fitting copula functions. Most work in this area is based on financial data, however, the challenges arise when dealing with censored lifetime data

and it is important to explore this topic further.

5.2 A Copula Model For Marked Point Processes

In Chapter 2 we described a novel model for marked point processes which incorporates a dependence between continuous marks and the event process through the use of a copula function. The model proposed in Section 2.2 is a fully parametric model and it would be desirable to relax the parametric assumption for the baseline rate within the class of mixed Poisson marginal processes to obtain some robustness. A simple first step would be to assume a piecewise constant baseline rate which would require specification of break points at which the rate can change. This approach has the advantage that it admits a more flexible form for the baseline rate which can be shown to give good approximations to results from semiparametric models. A disadvantage is the need to specify the break points and the arbitrary nature of any choices one might make for them. Recent work (Lawless and Yilmaz, 2011a; Lawless and Yilmaz, 2011b) on fitting bivariate failure time data with marginal Cox models and a specified copula suggests that some progress could be made in semiparametric analysis. The two-stage semiparametric estimator of the copula parameter is about as good as the simultaneous semiparametric estimator of it (Lawless and Yilmaz, 2011a). Since there are considerable computational challenges that arise in a joint analysis, two-stage estimation procedures (Shih and Louis, 1995) can be applied in principle, however, there will always be a trade-off between the need to make greater assumptions in the joint analysis with the robustness to dependent observation schemes that arise from joint analysis.

Several other extensions are possible to this model. We assumed that the dependence between the recurrent event process and the marks is the same for each consecutive pair of marks and waiting times. We also assumed that the association between the mark and the subsequent waiting time was the same in the two treatment arms (4.2.1). One could generalize the model to allow different copula parameters or even different copula functions for successive pairs of marks and waiting times, and these could even differ between the two treatment arms.

5.3 Multiple Multistate Processes Under Intermittent Inspection

In Chapter 3, we formulated a copula-based joint model for multiple multistate Markov processes by assuming that the joint survival function for the absorption times is governed by a multivariate copula function. In Section 3.2, different baseline transition intensities and regression coefficients are accommodated between processes. In settings where processes are clustered, one may wish to constrain the parameters in marginal intensities to be the same (e.g. when the processes represent progression in damage in paired organs). In such cases, we can write $\alpha_{jk} = \alpha_k$ and $\beta_{jk} = \beta_k, j = 1, 2, \ldots, J$, $\alpha = (\alpha_1, \ldots, \alpha_K)'$, $\beta = (\beta_1, \ldots, \beta_K)'$ and $\theta = (\alpha', \beta')'$ (Lee *and others*, 1992).

In Section 3.3, we restricted attention to the case in which all of the processes were inspected at the same time. In studies of organ damage in diabetic patients, interest may lie in modeling diabetic retinopathy and nephropathy (Cook and Lawless, 2013). The extent

of damage in the eyes, assessed by a detailed clinical examination, and kidneys, assessed by blood tests or imaging, would routinely be measured at different times. Adaptation of the proposed methods are relatively straightforward to handle this case by allowing process j to be assessed at M_j time points $v_{j0} < v_{j1} < \cdots < v_{j,M_j} < v_{j,M_j+1}$ where $v_{j0} = v_0 = 0$, $v_{j,M_j+1} = v_{M_j+1} = \infty$ for $j = 1, \ldots, J$.

The use of a copula function to model the association enables separate modelling and estimation of the marginal parameters and the association parameters. Different copula functions yield different dependence structures and so there are many choices of copula functions from copula families one can make for the association model. The marginal processes may correspond to more general, non-Markov, intensity-based models (2.2.6). Multiple ways of devising estimation strategies in this paper point to the flexibility of estimation. We have focused on parametric estimation, but weakly-parametric piecewise constant transition rates, or even more robust semiparametric analysis should be explored for estimation of marginal parameters. Estimation and inference can be conducted using generalized estimating equations based on working independence assumptions, first order generalized estimating equations (Liang and Zeger, 1986), and second order generalized estimating equations (Prentice, 1988; Zhao and Prentice, 1990).

Several extensions are possible to the association model. First, we again assumed the dependence between the absorption transition times is the same whether $X = 1$ and $X = 0$; see (3.2.2). One could allow different association parameters for different covariate values; indeed entirely different copula functions could be adopted. Second, we modeled

the association between absorption times via a copula, but one could set,

$$u_{jk} = \exp\left(- \int_{t_{j,k-1}}^{t_{jk}} \lambda_{jk}(s|x;\theta_{jk})ds \right) \ , \ j = 1, \ldots, J,$$

and use a copula function to model the association between u_{jk} and $u_{j'k}$, and hence the transition times T_{jk} and $T_{j'k}$. If a semi-Markov model is adopted for the marginal processes, the association between sojourn times is then modeled, as is routinely done in survival analysis.

5.4 Semiparametric Two-stage Estimation Procedure in Copula Models for Survival Data

Recall that a new estimator of the asymptotic variance of the second-stage estimator for the association parameter is proposed in Chapter 4. This proposed standard error estimator applies to complete, right-censored and current status data with or without covariate. However, we assume the censoring times are independent of survival times. In our future study, we will consider more general framework including interval-censored data and dependent censoring.

For the scenario with right-censored event times, we assumed that there are no ties in our data. We will consider a more general situation when the ties appear and we will compare the proposed asymptotic variance estimator with those in the established literature. For current status data, we studied the case with 50% prevalence level. We will

further study the behaviour of the estimates under 5% and 30% prevalence level.

Other estimating methods can be considered for right-censored multiple multistate processes. Alternating logistic regression estimating equations (Carey *and others*, 1993) or the estimating equations constructed based on marginal martingale and correlation between the marginal martingales corresponding to the absorption times (Prentice and Hsu, 1997) could be adopted.

5.5 Analysis of Recurrent Episodes in Chronic Disease with Vine Copula Models

In some chronic disease settings, processes are not progressive but rather exhibit a continual risk of periodic episodic conditions. Examples include chronic respiratory diseases such as asthma, infectious disease, and psychotic disorders. For each of these examples, the recurrent events have non-ignorable durations associated with them and are better characterized as recurrent episodes. We will formulate models in which the onset times are generated according to a Markov time scale and the durations of the episodes are governed by a semi-Markov process. To reflect multiple dependencies between the episode onset times and the duration of the episodes, we will consider a construction based on vine copula model (Joe, 1996; Bedford and Cooke, 2001, 2002; Berg and Aas, 2009; Aas *and others*, 2009). The idea of a vine copula model is to decompose the multivariate joint density into a cascade of densities of the original variables and their conditional or unconditional pairwise density functions, and to use bivariate copula functions for joint

distributions in this cascade. For high-dimensional multivariate distributions, there are a significant number of possible pair-copula decompositions. The vine, a graphical model for dependent random variables, has been introduced in this context to help organize those structures. The class of regular vines is general enough to provide a very rich set of options for possible pair-copula decompositions. In future work we will propose several candidate models, study properties of the models, assess treatment effects, and examine the biases arising from simple analyses which ignore some dependencies. Options for dealing with the biases induced by dependent censoring will be also explored.

References

Aalen, O. O., Borgan, O. and Gjessing, H. K. (2008). *Survival and Event History Analysis: A Point Process Point of View*. New York: Springer.

Aas, K. and Berg, D. (2009). Models for construction of multivariate dependence - a comparison study. *The European Journal of Finance* **15**, 639–659.

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009*a*). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* **44**, 182–198.

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009*b*). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* **44**, 182–198.

Ackerberg, D., Chen, X. and Hahn, J. (2012). A practical asymptotic variance estimator for two-step semiparametric estimators. *Review of Economics and Statistics* **94**, 481–498.

Al-Kateb, H., Boright, A. P., Mirea, L., Xie, X., Sutradhar, R., Mowjoodi, A., Bharaj, B., Liu, M., Bucksca, J. M., Arends, V. L., Steffes, M. W., Cleary, P. A., Sun, W., Lachin, J. M., Thorner, P. S., Ho, M., McKnight, A. J., Maxwell, A. P., Savage, D. A., Kidd, K. K., Kidd, J. R., Speed, W. C., Orchard, T. J., Miller, R. G., Sun, L., Bull,

S. B., Paterson, A. D. *and others*. (2008). Multiple superoxide dismutase 1/splicing factor serine alanine 15 variants are associated with the development and progression of diabetic nephropathy: The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Genetics study. *Diabetes* **57**, 218–228.

Albrecher, H. and Teugels, J. L. (2006). Exponential behavior in the presence of dependence in risk theory. *Journal of Applied Probability* **43**, 257–273.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* **10**, 1100–1120.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics* **26**, 641–647.

Bailey, K. R. (1984). Asymptotic equivalence between the Cox estimator and the general ML estimators of regression and survival parameters in the Cox model. *The Annals of Statistics* **12**, 730–736.

Bauer, K. A., Eriksson, B. I., Lassen, M. R. and Turpie, A. G. G. (2001). Fondaparinux compared to enoxaparin for the prevention of venous thromboembolism after elective major knee surgery. *The New England Journal of Medicine* **345**, 1305–1310.

Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally

dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence* **32**, 245–268.

Bedford, T. and Cooke, R. M. (2002). Vines - a new graphical model for dependent random variables. *Annals of Statistics* **30**, 1031–1068.

Berg, D. and Aas, K. (2009). Models for construction of multivariate dependence – a comparison study. *European Journal of Finance* **15**, 639–659.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society - Series B* **36**, 192–236.

Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **27**, 353–360.

Bowman, A., Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika* **85**, 799–808.

Breslow, N. E. (1972). Discussion of the paper by d. r. cox. *Journal of the Royal Statistical Society - Series B* **34**, 216–217.

Carey, V., Zeger, S. L. and Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika* **80**, 517–526.

Cazenave, J.-P., Folléa, G., Bardiaux, L., Boiron, J.-M., Lafeuillade, B., Debost, M., Lioure, B., Harousseau, J.-L., Tabrizi, R., Cahn, J.-Y., Michallet, M., Ambruso, D., Schots, R., Tissot, J.-D., Sensebé, L., Kondo, T., McCullough, J., Rebulla, P., Escolar, G., Mintz, P., Heddle, N. M., Goodrich, R. P., Bruhwyler, J., Le, C., Cook, R. J.,

Stouch, B. *and others*. (2010). A randomized controlled clinical trial evaluating the performance and safety of platelets treated with MIRASOL pathogen reduction technology. *Transfusion* **50**, 2362–2375.

Chandran, V., Cook, R. J., Edwin, J., Shen, H., Pellett, F. J., Shanmugarajah, S., Rosen, C. F. and Gladman, D. D. (2010). Soluble biomarkers differentiate patients with psoriatic arthritis from those with psoriasis without arthritis. *Rheumatology* **49**(7), 139–1405.

Chandran, V., Cook, R. J., Thavaneswaran, A., Lee, K.-A., Pellett, F. J. and Gladman, D. D. (2012). Parametric survival analysis as well as multi-state analysis conrms the association between human leukocyte antigen alleles and the development of arthritis mutilans in patients with psoriatic arthritis. *The Journal of Rheumatology* **39**(8), 1723.

Chatterjee, N., Kalaylioglu, Z., Shih, J. H. and Gail, M. H. (2006). Case-control and case-only designs with genotype and family history data: estimating relative risk, residual familial aggregation, and cumulative risk. *Biometrics* **62**, 36–48.

Chen, F. and Zheng, Y.-S. (1997). One-warehouse multiretailer systems with centralized stock information. *Operations Research* **45**, 275–287.

Chen, S. X. and Huang, T.-M. (2007). Nonparametric estimation of copula functions for dependence modelling. *The Canadian Journal of Statistics* **35**, 265–282.

Cole, E. H., Cattran, D. C., Farewell, V. T., Aprile, M., Bear, R. A., Pei, Y. P., Fenton, S. S., Tober, J. A. and Cardella, C. J. (1994). A comparison of rabbit antithymocyte serum and OKT3 as prophylaxis against renal allograft rejection. *Transplantation* **57**, 60–67.

Cook, R. J., Kalbfleisch, J. D. and Yi, G. Y. (2002). A generalized mover-stayer model for panel data. *Biostatistics* **3**, 407–420.

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer.

Cook, R. J. and Lawless, J. F. (2013). Concepts and tests for trend in recurrent event processes. *Iranian Journal of Statistics* **12**, 35–69.

Cook, R. J., Lawless, J. F. and Lee, K.-A. (2003). Cumulative processes related to event histories. *SORT* **27**, 13–30.

Cook, R. J., Yi, G. Y., Lee, K.-A. and Gladman, D. D. (2004). A conditional Markov model for clustered progressive multistate processes under incomplete observation. *Biometrics* **60**, 436–443.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society - Series B* **34**, 187–220.

Cox, D. R. and Isham, V. (1980). *Point Processes*. London: Chapman and Hall.

Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91**, 729–737.

Craiu, M. and Craiu, R. V. (2008). Choice of parametric families of copulas. *Advances and Applications in Statistics* **10**, 25–40.

Czado, C., Brechmann, E. C. and Gruber, L. (2013). Selection of vine copulas. Springer To appear.

Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes: Volumn II: General Theory and Structure*, 2nd edition. New York: Springer.

Davis, K. B., Slichter, S. J. and Corash, L. (1999). Corrected count increment and percent platelet recovery as measures of post-transfusion platelet response: problems and a solution. *Transfusion* **39**, 586–592.

Descombes, X. and Zerubia, J. (2002). Marked point process in image analysis. *Signal Processing Magazine, IEEE* **19**, 77–84.

Diao, L., Cook, R. J. and Lee, K.-A. (2013). A copula model for marked point processes. *Lifetime Data Analysis* DOI: 10.1007/s10985-013-9259-3.

Dißmann, J., Brechmann, E.C., Czado, C. and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics and Data Analysis* **59**, 52–69.

Eriksson, B. I., Bauer, K. A., Lassen, M. R. and Turpie, A. G. G. for the Steering Committee of the Pentasaccharide in Hip Fracture Surgery Study. (2001). Fondaparinux compared with enoxaparin for the prevention of venous thromboembolism after hip-fracture surgery. *The New England Journal of Medicine* **345**, 129–1394.

Fermanian, J.-D. and Scaillet, O. (2003). Nonparametric estimate of copulas for time series. *Journal of Risk* **5**, 25–54.

Fernholz, L. T. (1983). *Von Mises Calculus for Statistical Functionals*. New York: Springer-Verlag.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: John Wiley and Sons.

Fok, C. C. T., Ramsay, J. O., Abrahamowicz, M. and Fortin, P. (2012). A functional marked point process model for lupus data. *Canadian Journal of Statistics* **40**, 517–529.

Genest, C., Ghoudi, K. and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–552.

Genest, C. and MacKay, J. (1986). The joy of copulas: bivariate distributions with uniform marginals. *The American Statistician* **40**, 280–283.

Genest, C., Rémillard, B. and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: a review and a power study. *Insurance: Mathematics and Economics* **44**, 199–213.

Gentleman, R. C., Lawless, J. F., Lindsey, J. and Yan, P. (1994). Multistate Markov models for analysing incomplete disease history data, with illustrations for HIV disease. *Statistics in Medicine* **13**, 805–821.

Gentleman, R. C. and Vandal, A. C. (2002). Nonparametric estimation of the bivariate CDF for arbitrarily censored data. *Canadian Journal of Statistics* **30**, 557–571.

Gladman, D. D., Farewell, V. T. and Nadeau, C. (1995). Clinical indicators of progression in psoriatic arthritis: multivariate relative risk model. *The Journal of Rheumatology* **22**(4), 675–679.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* **31**, 1028–1211.

Godambe, V. P. (1991). *Estimating Functions*. Oxford: Clarendon Press.

Goulard, M., Säkkä, A. and Grabarnik, P. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. *Scandinavian Journal of Statistics* **23**, 365–379.

Grandell, J. (1997). *Mixed Poisson Processes*. London: Chapman and Hall.

Groeneboom, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution. *Technical Report* 378, Department of Statistics, Stanford University.

Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and non-parametric maximum likelihood estimation*. Boston: Birkhäuser.

Grüger, J., Kay, R. and Schumacher, M. (1991). The validity of inferences based on incomplete observations in disease state models. *Biometrics* **47**, 595–605.

Guan, Y. (2006). Tests for independence between marks and points of a marked point process. *Biometrics* **62**, 126–134.

Haff, I. H., Aas, K. and Frigessi, A. (2010). On the simplified pair-copula construction – simply useful or too simplistic? *Journal of Multivariate Analysis* **101**, 1296–1310.

He, W. and Lawless, J. F. (2005). Bivariate location-scale models for regression analysis, with applications to lifetime data. *Journal of the Royal Statistical Society Series B* **67**, 63–78.

Holden, L., Sannan, S. and Bungum, H. (2002). A stochastic marked point process model for earthquakes. *Natural Hazards and Earth System Science* **3**, 95–101.

Hortobagyi, G. N., Theriault, R. L., Lipton, A., Porter, L., Blayney, D., Sinoff, C., Wheeler, H., Simeone, J. F., Seaman, J., Knight, R. D., Hefferman, M., Mellars, K. *and others.* (1998). Long-term prevention of skeletal complications of metastatic breast cancer with pamidronate. *Journal of Clinical Oncology* **16**, 2038–2044.

Hortobagyi, G. N., Theriault, R. L., Porter, L., Blayney, D., Lipton, A., Sinoff, C., Wheeler, H., Simeone, J. F., Seaman, J., Knight, R. D., Hefferman, M., Reitsma, D. J., Kennedy, I., Allan, S. G. *and others.* (1996). Efficacy of pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases. *Journal of Clinical Oncology* **335**, 1785–1791.

Hougaard, P. (1999). Multi-state models: a review. *Lifetime Data Analysis* **5**, 239–264.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.

Hsu, L. and Prentice, R. L. (1996). On assessing the strength of dependency between failure time variates. *Biometrika* **83**, 491–506.

Huang, J. and Wellner, J. A. (1997). Interval censored survival data: a review of recent progress. *Proceedings of the First Seattle Symposium in Biostatistics* **123**, 123–169.

Javid, A. A. (2009). Copulas with truncation-invariance property. *Communications in Statistics - Theory and Methods* **38**, 3756–3771.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *IMS Lecture Notes - Monograph Series* **28**, 120–141.

Joe, H. (1997). *Multivariate Dependence Concepts*. London: Chapman and Hall.

Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* **80**, 863–871.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edition. New York: John Wiley and Sons.

Kaplan, E. L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.

Karr, A. F. (1991). *Point Processes and Their Statistical Inference 2nd edition*. New York: Dekker.

Kurowicka, D. and Cooke, R. M. (2005). Distribution-free continuous Bayesian belief nets. In: Wilson, A., Limnios, N., Keller-McNulty, S. and Armijo, Y. (editors), *Modern Statistical and Mathematical Methods in Reliability*. Singapore: World Scientific Publishing Co. Pte. Ltd.

Landriault, D., Lee, W. Y., Willmot, G. E. and Woo, J.-K. (2012). A note on deficit analysis in dependency models involving Coxian claim amounts. *Scandinavian Journal of Statistics* DOI: 10.1080/03461238.2012.723044.

Lassen, M. R., Bauer, K. A., Eriksson, B. I. and Turpie, A. G. G. for the European Pentasaccharide Hip Elective Surgery Study (EPHESIS) Steering Committee. (2002). Post-

operative fondaparinux versus preoperative enoxaparin for prevention of venous thromboembolism in elective hip-replacement surgery: a randomized double-blind comparison. *The Lancet* **359**, 171–1720.

Lawless, J. F. (1987*a*). Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics* **15**, 209–225.

Lawless, J. F. (1987*b*). Regression methods for Poisson process data. *Journal of the American Statistical Association* **82**, 808–815.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd edition. Hoboken, NJ: Wiley.

Lawless, J. F. and Nadeau, J. C. (1995). Nonparametric estimation of cumulative mean functions for recurrent events. *Technometrics* **37**, 158–168.

Lawless, J. F. and Yilmaz, Y. E. (2011*a*). Comparison of semiparametric maximum likelihood estimation and two-stage semiparametric estimation in copula models. *Computational Statistics and Data Analysis* **55**, 2446–2455.

Lawless, J. F. and Yilmaz, Y. E. (2011*b*). Semiparametric estimation in copula models for bivariate sequential survival times. *Biometrical Journal* **53**, 779–796.

Lee, E. W. and Kim, M. Y. (1998). The analysis of correlated panel data using a continuous-time Markov model. *Biometrics* **54**, 1638–1644.

Lee, E. W., Wei, L. J. and Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In: Klein, J. P. and

Goel, P. K. (editors), *Survival Analysis: State of the Art*. Dordrecht: Kluwer Academic Publishers, pp. 237–247.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics* **80**, 221–239.

Lindsay, B. G., Yi, G. Y. and Sun, J. (2011). Issues and strategies in the selection of composite likelihoods. *Statistica Sinica* **21**, 71–105.

Maathuis, M. (2013). *MLEcens: computation of the MLE for bivariate (interval) censored data*. http://CRAN.R-project.org/package=MLEcens/R package version 0.1-4.

Marshall, G. and Jones, R. H. (1995). Multistate models and diabetic retinopathy. *Statistics in Medicine* **14**, 1975–1983.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

Moll, J. M. H. and Wright, V. (1973). New York clinical criteria for ankylosing spondylitis. A statistical evaluation. *Annual of Rheumatic Diseases* **32**, 354–363.

Nelsen, R. B. (2006). *An Introduction to Copulas*. New York: Springer.

Newey, W. K. and McFadden, D. (1994). *Large Sample Estimation and Hypothesis Testing*. North-Holland: Handbook of Econometrics.

Oakes, D. (2005). On the preservation of copula structure under truncation. *The Canadian Journal of Statistics* **33**, 465–468.

Pascual, J., Falk, R. M., Piessens, F., Prusinski, A., Docekal, P., Robert, M., Ferrer, P., Luria, X., Segarra, R. and Zayas, J. M. (2000). Consistent efficacy and tolerability of almotriptan in the acute treatment of multiple migraine attacks: results of a large, randomized, double-blind, placebo-controlled study. *Cephelegia* **20**, 588–596.

Patton, A. J. (2006). Modeling asymmetric exchange rate dependence. *International Economic Review* **47**, 527–556.

Penttinen, A., Stoyan, D. and Henttonen, H. M. (1992). Marked point processes in forest statistics. *Forest Science* **38**, 806–824.

Petri, M., Genovese, M., Engle, E. and Hochberg, M. (1991). Definition, incidence, and clinical description of flare in systemic lupus erythematosus. A prospective cohort study. *Arthritis & Rheumatism* **34**, 937–944.

Politis, D. N. and Sherman, M. (2001). Moment estimation for statistics from marked point processes. *Journal of the Royal Statistical Society - Series B* **63**, 261–275.

Prentice, R. L. (1988). Correlated binary regression with covariate specific to each binary observation. *Biometrics* **55**, 1228–1231.

Prentice, R. L. and Hsu, L. (1997). Regression on hazard ratios and cross ratios in multivariate failure time analysis. *Biometrika* **84**, 349–363.

Prentice, R. L., Williams, B. J. and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373–379.

Prigent, J. L. (2001). Option pricing with a general marked point process. *Mathematics of Operations Research* **26**, 50–66.

Prokhorov, A. and Schmidt, P. (2009). Likelihood-based estimation in a panel setting: robustness, redundancy and validity of copulas. *Journal of Econometrics* **153**, 93–104.

Rahman, P., Gladman, D. D., Cook, R. J., Zhou, Y., Young, G. and Salonen, D. (1998). Radiological assessment in psoriatic arthritis. *British Journal of Rheumatology* **37**, 760–765.

Robin, S. (2002). A compound Poisson model for word occurrences in DNA sequences. *Journal of the Royal Statistical Society - Series C* **51**, 437–451.

Ross, S. M. (1996). *Stochastic Processes*, 2nd edition. New York: Wiley.

Sánchez-Sellero, C., González-Manteiga, W. and Cao, R. (1999). Bandwidth selection in density estimation with truncated and censored data. *Annals of the Institute of Statistical Mathematics* **51**, 51–70.

Satten, G. A. (1999). Estimating the extent of tracking in interval-censored chain-of-events data. *Biometrics* **55**, 1228–1231.

Schimek, M. G. (2000). *Smoothing and Regression: Approaches, Computation, and Application*. New York: John Wiley & Sons.

Schlather, M., Ribeiro, P. J. Jr. and Diggle, P. J. (2004). Detecting dependence between marks and locations of marked point processes. *Journal of the Royal Statistical Society - Series B* **66**, 79–93.

Schoenberg, F. P. (2004). Testing separability in spatial temporal marked point processes. *Biometrics* **60**, 471–481.

Sears, M. R., Taylor, D. R., Print, C. G., Lake, D. C., Li, Q., Flannery, M., Yates, D. M., Lucas, M. K. and Herbison, G. P. (1990). Regular inhaled beta-agonist treatment in bronchial asthma. *Lancet* **336**, 1391–1396.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

Sheather, S. J. (2004). Density estimation. *Statistical Science* **19**, 588–597.

Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameters in copula models for bivariate survival data. *Biometrics* **51**, 1384–1399.

Sklar, A. (1959). Fonctions de répartition á n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* **8**, 229–231.

Snyder, D. L. and Miller, M. I. (1991). *Random Point Processes in Time and Space*. New York: Springer-Verlag.

Steinbrocker, O., Traeger, C. H. and Batterman, R. C. (1949). Therapeutic criteria in rheumatoid arthritis. *Journal of the American Medical Association* **140**, 659–662.

Stöber, J., Joe, H. and Czado, C. (2012). Simplified pair copula constructions – limits and extensions. submitted.

Sun, L., Wang, L. and Sun, J. (2006). Estimation of the association for bivariate interval-censored failure time data. *Scandinavian Journal of Statistics* **33**, 637–649.

Sutradhar, R. and Cook, R. J. (2008). Analysis of interval-censored data from clustered multi-state processes: application to joint damage in psoriatic arthritis. *Journal of the Royal Statistical Society - Series C* **57**, 553–566.

Sweeting, M. J., Farewell, V. T. and Angelis, D. De. (2010). Multistate Markov models for disease progression in the presence of informative examination times: an application to hepatitis C. *Statistics in Medicine* **29**, 1161–1174.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

Tolusso, D. and Cook, R. J. (2009). Robust estimation of state occupancy probabilities for interval-censored multistate data: an application involving spondylitis in psoriatic arthritis. *Communications in Statistics - Theory and Methods* **38**(18), 3307–3325.

Turpie, A. G. G., Bauer, K. A., Eriksson, B. I. and Lassen, M. R. (2002). Postoperative fondaparinux versus postoperative enoxaparin for prevention of venous thromboembolism after elective hip-replacement surgery: a randomized double-blind trial. *The Lancet* **359**, 1721–1726.

Tyas, S. L., Salazar, J. C., Snowdon, D. A., Desrosiers, M. F., Riley, K. P., Mendiondo,

M.S. and Kryscio, R. J. (2007). Transitions to mild cognitive impairments, dementia and death: findings from the Nun Study. *American Journal of Epidemiology* **165**, 1231–1238.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. New York: Cambridge University Press.

van Eeden, C. (1956). Maximum likelihood estimation of ordered probabilities. *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen A* **A59**, 444–455.

van Eeden, C. (1957). Maximum likelihood estimation of partially ordered or completely ordered probabilities. *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen A* **A60**, 128–136.

Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis* **92**, 1–28.

Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statistics Sinica* **21**, 5–42.

Verona, E., Petrov, D., Cserhati, E., Hofman, J., Geppe, N., Medley, H. and Hughes, S. (2003). Fluticasone propionate in asthma: a long term dose comparison study. *Archives of Disease in Childhood* **88**, 503–509.

Wang, W. (2003). Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society - Series B* **65**, 257–273.

Wang, W. and Ding, A. A. (2000). On assessing the association for bivariate current status data. *Biometrika* **87**, 879–893.

Warkentin, T. E., Cook, R. J., Marder, V. J., Sheppard, J. I., Moore, J. C., Eriksson, B. I., Greinacher, A. and Kelton, J. G. (2005). Anti-platelet factor 4/heparin antibodies in orthopedic surgery patients receiving antithrombotic prophylaxis with fondaparinux or enoxaparin. *Blood* **106**, 3791–3796.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065–1073.

White, R. H., Romano, P. S., Zhou, H., Rodrigo, J. and Bargar, W. (1998). Incidence and time course of thromboembolic outcomes following total hip or knee arthroplasty. *Journal of the American Medical Association* **158**, 1525–1531.

Xu, X. and Reid, N. (2011). On the robustness of maximum composite likelihood estimates. *Journal of Statistical Planning and Inferences* **141**, 3047–3054.

Yilmaz, Y. E. and Lawless, J. F. (2011). Likelihood ratio procedures and tests of fit in parametric and semiparametric copula models with censored data. *Lifetime Data Analysis* **17**, 386–408.

Zhao, L. and Hu, X. J. (2013). Estimation with right-censored observations under a semi-markov model. *Canadian Journal of Statistics* **41**, 237–256.

Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrics* **77**, 642–648.