國 立 交 通 大 學

統計學研究所

博士論文

三種資料結構下雙變數存活時間之迴歸分析

Regression Analysis for

Bivariate Failure-Time Data

under Three Types of Data Structures

研究生：謝進見

指導教授: 王維菁 博士

中 華 民 國 九十六 年 六 月

三種資料結構下雙變數存活時間之迴歸分析

Regression Analysis for Bivariate Failure-Time Data under Three Types of Data Structures

研 究 生：謝進見　　Student：Jin-Jian Hsieh

指導教授：王維菁　　Advisor：Weijing Wang

國 立 交 通 大 學

統 計 學 研 究 所

博 士 論 文

A Dissertation

Submitted to Institute of Statistics

National Chiao Tung University

June 2007

Hsin-Chu, Taiwan, Republic of China

中 華 民 國 九 十 六 年 六 月

三種資料結構下雙變數存活時間之迴歸分析

學生：謝進見　　　　　　　　　　　　　　　　　　指導教授：王維菁

國立交通大學統計學研究所 博士班

摘　　要

雙變數存活分析已被廣泛應用於生物醫學的研究。早期的研究課題較偏向於探討兩個不同生物個體或器官組織存活時間之關連性。近年來的應用方向則拓廣到同一個體所發生不同事件的時間。後者在分析上，往往伴隨所探討的時間變數間彼此具有設限或截切關係，使得統計推論變得更為複雜。本論文包含兩個研究計劃，均在迴歸的架構下分析雙變數之存活時間。我們特別針對前面所述特殊的設限或截切資料，提出統計推論的方法。

第一個計劃針對半競爭風險資料，探討解釋變數對“中介事件發生時間”的影響。分析的難度在於所欲探討的時間長度受制於相關設限。大部分文獻所提出的方法均利用“人為設限”〔artificial censoring〕的技巧，以處理相關設限所造成的偏誤。不過這個方法因把部份觀測值捨棄而會產生估計效率上的損失，亦因添加了額外的模型假設而有缺乏穩健性的缺點。我們提出兩階段估計方法可改善前述方法的缺點。我們亦針對兩個所提出的假設，發展模型檢驗的方法。論文中並推導了大樣本性質，並且透過數值分析評估各推論方法在有限樣本下的表現。

在第二個計劃中，我們建構關聯性的迴歸模式，並且發展一套推論方法可以彈性的分析三種截然不同的資料結構。我們也針對此模式假設，提出模型檢驗的方法。論文中亦呈現大樣本分析與數值分析。

關鍵字：阿基米得關聯模式；雙變數存活時間；相關設限；相關截切；局部勝算比值；多重事件資料；半競爭風險資料；轉換模型。

Regression Analysis for Bivariate Failure-Time Data under Three Types of Data Structures

Student: Jin-Jian Hsieh Advisor: Weijing Wang

Institute of Statistics, National Chiao Tung University

ABSTRACT

Bivariate survival analysis has received substantial attentions due to its wide applications. The variables of interest may represent failure times occurred to two different biological units or different event times measured from the same subject. In the latter situation, the two failure times may have censoring or truncation relationship which complicates statistical analysis. The thesis contains two projects, both of which consider regression analysis for bivariate survival data.

The first project focuses on semi-competing risks data in which a terminal event censors a non-terminal event. In particular we investigate how covariates affect the marginal distribution of the time to a non-terminal event subject to dependent censoring. Most existing methods utilize the technique of artificial censoring to remove the sampling bias. However these approaches may result in efficiency loss and may not be robust under model mis-specification. We propose a two-stage procedure to tackle this problem. We also propose model selection methods to verify the two main assumptions. Large-sample properties are also proved. Numerical analysis is performed to evaluate finite-sample performances of the proposed methods.

In the second part of the thesis, we consider the situation that covariates may affect the level of association. We propose a flexible regression model and then develop a unified inference procedure which can be applied to three different types of data structures. For this part, we also present a model checking method for assessing the appropriateness of the Clayton assumption. Large-sample analysis and numerical studies are also presented.

Keywords: Archimedean copula; Bivariate failure-time; Clayton model; Dependent censoring; Dependent truncation; Kendall's tau; Local odds ratio; Multiple events data; Semi-competing risks data; Transformation model.

誌 謝

論文終於完成了！在喜悅之際，心中有滿滿的感恩、、、

首先，我要感謝我的恩師 王維菁 博士！感謝王老師悉心的指導，從論文題材的選取、研究方法的進行及模擬程式的撰寫，到今日論文的完成，皆歸功於王老師從旁循序漸進，不厭其煩的指導，才能順利完成。在老師身上我學到許多，不管在課業上、治學上、處世上老師給我許許多多的啓發與啓示；也感謝老師在生活上的關照，讓我覺得溫馨且愉快，這樣的經驗讓我永生難忘。老師 謝謝您！

再來，我要感謝合作指導老師 丁愛東 博士！感謝丁老師的悉心指導，特別是在數理方面的教導與協助，讓我學到許多。尤其是丁老師人在美國教書，常常不厭其煩的撥冗指導，實是感恩在心頭！我也要感謝所長 洪慧念 博士！感謝洪老師時常的提供研究上、生活上的指導與協助。並且感謝口試委員洪慧念老師、黃冠華老師、陳珍信老師及程毅豪老師給予本論文的指正與建議，使本論文更加充實，誠摯感謝！

在交大統計所一千四百多個日子，感謝所有老師的鼓勵與指導，在這大家庭中師長們的平易近人，學長姊、學弟妹、同學們(瑋華、建豪、秋婷、江村剛志、弘家、家鈴、信銘、資荃、家群、泰賓、維方、志浩、碩慧、清煜、俊達、進結、棟鴻、春火、穗碧)彼此關心照顧、相互扶持，我真心的感激與珍惜！

最後，僅以本文獻給我摯愛的家人與家葳，在人生的過程中，每當我低潮挫折時，家是我最佳的避風港，他們永遠的支持是我信心的來源，他們總是無私地給了我滿滿的愛，使我有信心、動力面對人生的種種考驗。感謝我的父母！感謝家葳與家葳的爸媽！感謝我的妹妹(琍敏、麗惠)！感謝上天的慈悲！感謝生命中的一切一切！

進見 于新竹

2007 年　夏

Table of Contents

Chapter 1. Introduction

Many biomedical studies involve analysis of multiple events. In recent years, the techniques of multivariate survival analysis have been applied to analyze various types of event-time data. Here we consider the bivariate case. Let (T_1, T_2) be a pair of failure times which are possibly correlated. Traditionally (T_1, T_2) refer to the failure times of different biological units such as twins, family members or paired organs on the same person. In the past few years, applications have been extended to include variables which have censoring or truncation relationship. Semi-competing risks data provide an example in which a terminal variable, such as death, may censor a non-terminal variable such as disease progression but not vice versa. The recent paper by Chaieb et al. (2006) discusses another situation in which one variable truncates the other. They consider an example of transfusion-related AIDS in which subjects were observed only if they had developed AIDS during the study period which lasted for 102 months. The example records the infection time (S), measured from the beginning of the study, and the induction time (T_1), from infection to AIDS, in months. Setting $T_2 = 102 - S$, subjects are observed only if $T_1 \leq T_2$. See Figure 1-1 for illustration. Hence the induction time T_1 is subject to right truncation by T_2. Although it seems that T_1 and T_2 should be independent, dependence between the two variables was discovered by Tsai (1990) and later researchers including Chaieb et al. (2006) and Emura et al. (2007). One possible explanation of the dependence is that the medical practice for HIV-infected patients may be different in different calendar years.

Statistical inference methods should account for the underlying data structures. However for analysis of multiple event data, the possibility of dependent censoring or truncation complicates subsequent statistical analysis. In the thesis, we are interested in studying the covariate effects on either the marginal distributions or the dependence structure based on multiple event data of various types. This article contains two dif-

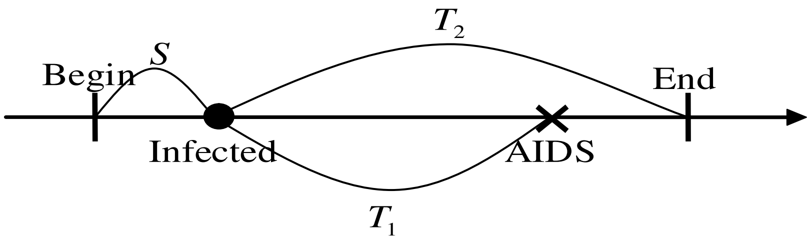

Figure 1-1: Illustration of AIDS data.

ferent projects. In the first part, we consider marginal regression analysis based on semi-competing risks data. The challenge of statistical inference is that the regression effect of interest can only be analyzed in presence of dependent censoring. In the second part, we propose a model which describes how covariates affect the level of association between the two failure times of interest. We aim to develop a unified inference procedure which can handle the three different data structures mentioned above.

Chapter 2 reviews the model assumptions and data structures that will be considered in both projects. Chapter 3 focuses on the first project on marginal regression analysis. Section 3.1∼3.3 review related literature and our main results are stated in Section 3.4. Chapter 4 considers the second project about the regression analysis for association. Section 4.1∼4.3 provides the background information and the review of related literature. The main results are summarized in Section 4.4. In Chapter 5, we give some concluding remarks and discuss possible future research directions.

Chapter 2. Literature Review

We first introduce useful model assumptions for describing the association between two failure time variables. Then, we examine three types of data, namely typical bivariate data, semi-competing risks data and truncation data, which will be considered in the thesis.

2.1. Copula Models and Archimedean Copula Models

Copula models are often used to describe the association between two failure time variables. In the book by Nelsen (2006), "An Introduction to Copulas": *From one point a view, copulas are functions that join or "copulas" multivariate distribution functions to their one-dimensional marginal distribution functions. Alternatively, copulas are multivariate distribution functions whose one-dimensional margins are uniform on the interval* $(0,1)$. For the bivariate case, a copula function can be written as $C(u,v)$, which may be parameterized as $C_\alpha(u,v)$ such that $C_\alpha(.,.) : [0,1]^2 \to [0,1]$ and
(i) $C_\alpha(0,u) = C_\alpha(u,0) = 0$ and $C_\alpha(1,u) = C_\alpha(u,1) = u$ for all $u \in [0,1]$; and
(ii) $C_\alpha(u_2,v_2) - C_\alpha(u_1,v_2) - C_\alpha(u_2,v_1) + C_\alpha(u_1,v_1) \geq 0$ for all u_1, u_2, v_1, v_2 in I such that $u_2 \leq u_1$ and $v_2 \leq v_1$.

The Archimedean copula (AC) family is a popular subclass of the copula family. A copula is said to be "Archimedean copula" (AC) if it can be expressed in the following form

$$C_\alpha(u,v) = \phi_\alpha^{-1}\{\phi_\alpha(u) + \phi_\alpha(v)\}, \tag{1}$$

where $\phi_\alpha : [0,1] \to [0,\infty]$ satisfying $\phi_\alpha(1) = 0$, $\phi'_\alpha(t) < 0$ and $\phi''_\alpha(t) > 0$. Note that the AC family simplifies the bivariate relationship via the univariate function $\phi_\alpha(\cdot)$. The function $\phi_\alpha(\cdot)$ is called the generator of the copula. Important properties of AC models have been derived in Genest and MacKay (1986), Oakes (1989) and Genest and Rivest (1993).

Now, we introduce some commonly-seen examples of Archimedean copulas.

Example (a): the Clayton copula (1978), where the generator function is $\phi_\alpha(t) = (t^{1-\alpha} - 1)/(\alpha - 1)$ for some $(\alpha > 1)$ and

$$C_\alpha(u, v) = \{u^{1-\alpha} + v^{1-\alpha} - 1\}^{1/(1-\alpha)}.$$

Example (b): the Frank copula (1979), where $\phi_\alpha(t) = \log[(1-\alpha)/(1-\alpha^t)]$ $(\alpha > 0)$ and

$$C_\alpha(u, v) = \log_\alpha\{1 + (\alpha^u - 1)(\alpha^v - 1)/(\alpha - 1)\}.$$

Example (c): the Gumbel model (1960), where $\phi_\alpha(t) = \{-\log(t)\}^{\alpha+1}$ $(\alpha > 0)$ and

$$C_\alpha(u, v) = \exp\{-[(-\log(u))^{\alpha+1} + (-\log(v))^{\alpha+1}]^{1/(\alpha+1)}\}.$$

When $C(u, v) = uv$, we have the product copula or independent copula.
Example (d): the Gumbel-Barnett copula, where $\phi_\alpha(t) = \log(1 - \alpha \log(t))$ $\alpha \in (0, 1]$.

$$C_\alpha(u, v) = uv \exp\{-\alpha \log(u) \log(v)\}.$$

In applications, the copula structure is usually imposed on the survival function of bivariate failure times (T_1, T_2) such that one can write

$$\Pr(T_1 > s, T_2 > t) = C_\alpha\{\Pr(T_1 > s), \Pr(T_2 > t)\}. \tag{2}$$

Accordingly an AC model defined on the joint survival function can be written as

$$\Pr(T_1 > s, T_2 > t) = \phi_\alpha^{-1}\{\phi_\alpha(\Pr(T_1 > s)) + \phi_\alpha(\Pr(T_2 > t))\}.$$

The AC family has nice analytic properties which are useful for further statistical inference.

Global association between (T_1, T_2) can be summarized by Kendall's tau, denoted as τ, which is defined as

$$\tau = \Pr\{(T_{1i} - T_{1j})(T_{2i} - T_{2j}) > 0\} - \Pr\{(T_{1i} - T_{1j})(T_{2i} - T_{2j}) < 0\}, \tag{3}$$

where (T_{1i}, T_{2i}) and (T_{1j}, T_{2j}) $(i \neq j)$ are two independent pair of (T_1, T_2). Note that $\tau = 0$, if T_1 and T_2 are independent; $\tau > 0$, if T_1 and T_2 are positively correlated; $\tau < 0$, if T_1 and T_2 are negatively correlated. Kendall's tau has a nice property of rank invariance. For a copula model indexed by $C_\alpha(\cdot,\cdot)$, the association parameter α is related to Kendall's τ such that

$$\tau = 4\int_0^1\int_0^1 C_\alpha(u,v)C_\alpha(du,dv) - 1. \tag{4}$$

Here we give some examples. For the Clayton model, $\tau = (\alpha-1)/(\alpha+1)$; for the Gumbel model, $\tau = \alpha/(\alpha+1)$; and for the Frank model, $\tau = 1 + 4\gamma^{-1}\{D_1(\gamma) - 1\}$, $\gamma = -\log(\alpha)$ and for integer $k \geq 1$, D_k is the Debye function defined by

$$D_k(\gamma) = \frac{k}{\gamma^k}\int_0^\gamma \frac{t^k}{e^t - 1}dt.$$

For describing local association, Oakes (1989) proposed the following cross ratio function

$$\theta(s,t) = \frac{\Pr(T_1 > s, T_2 > t) D_s D_t \Pr(T_1 > s, T_2 > t)}{\{D_s \Pr(T_1 > s, T_2 > t)\}\{D_t \Pr(T_1 > s, T_2 > t)\}}, \tag{5}$$

where D_s denotes the operator $-\partial/\partial s$. This function has an intuitive interpretation as the ratio of the hazard rate of the conditional distribution of T_1, given $T_2 = t$, to that of T_1, given $T_2 > t$. Specifically it follows that

$$\theta(s,t) = \frac{\lambda_1(s|T_2 = t)}{\lambda_1(s|T_2 > t)}, \tag{6}$$

where $\lambda_1(s|A)$ is the hazard of T_1 at time s given that event A occurs. The cross ratio function for an AC model possesses some nice analytic properties. Oakes (1989) showed that, for an AC model indexed by $\phi_\alpha(\cdot)$, $\theta(s,t)$ depends on (s,t) only through $\Pr(T_1 > s, T_2 > t)$ such that

$$\theta(s,t) = \tilde{\theta}_\alpha(\Pr(T_1 > s, T_2 > t)), \tag{7}$$

where $\tilde{\theta}_\alpha(v) = -v\phi''_\alpha(v)/\phi'_\alpha(v)$. The different expressions of $\theta(s,t)$ are useful in subsequent inference problems. For example equation (7) provides a semi-parametric expression of $\theta(s,t)$ in terms of α and $\Pr(T_1 > s, T_2 > t)$ which can be analyzed nonparametrically.

2.2. Different Types of Data Structures

In the thesis, we aim to study (T_1, T_2) which, besides their association, may have censoring or truncation relationship due to the restriction of the observational scheme. To simplify the analysis, we assume that T_k $(k = 1, 2)$ are continuous random variables.

Define the following two regions: $R_1 = \{(s,t) : 0 < s < \infty, 0 < t < \infty\}$ and $R_2 = \{(s,t) : 0 < s \le t < \infty\}$. The following three types of data structures will be considered in the thesis.

Data Structure 1 - Typical bivariate failure time data. The failure times (T_1, T_2) have no specific relationship and hence observations fall in the region of R_1;

Data Structure 2 - Semi-competing risks data. The time to the non-terminal event, T_1, is subject to censoring by the time to the terminal event T_2. One can observe $(T_1 \wedge T_2, T_2, I(T_1 \le T_2))$ and the first two variables fall in R_2;

Data Structure 3 - Truncation data. Let T_1 be subject to right truncation by T_2 or T_2 subject to left truncation by T_1 such that one can only observe replications of (T_1, T_2) with $T_1 \le T_2$ which are also located in R_2.

In Figure 2-1, hypothetical data from the three data types in absence of external censoring are plotted.

Examples of the first type may be found in studies of twins or paired organs on the same person. The study of association between the time of disease progression and survival usually encounters the second type. The example of transfusion-related AIDS mentioned earlier belongs to the third type.

Now we briefly discuss the difference of the latter two types of data which look similar

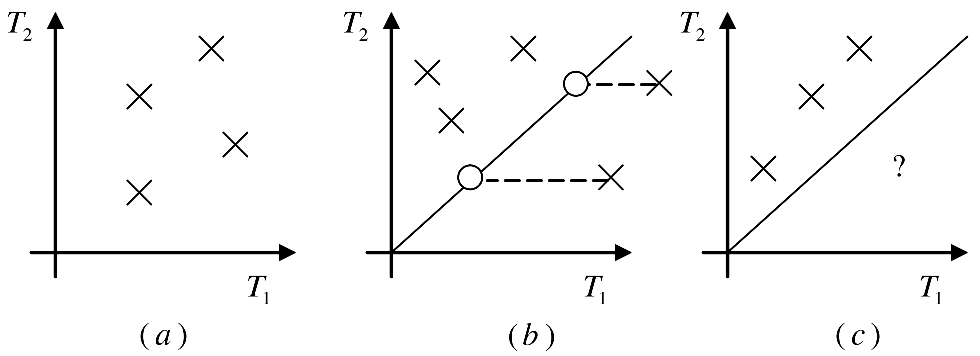

Figure 2-1: (a): Typical bivariate data; (b): Semi-competing risks data; (c): Truncation data.

at the first glance. For semi-competing risks data, observations with $T_1 > T_2$ can be identified and only their exact locations on the plane are uncertain. For truncation data, in contrast, we have no information for observations with $(T_1, T_2) \in \{(s,t) : 0 < t < s < \infty\}$ and even whether they exist is unknown.

In Chapter 3, we will consider marginal regression analysis based on semi-competing risks data in which T_1 is subject to dependent censoring by T_2. In Chapter 4, we consider association study under a general framework which includes all the three types of data structures discussed above. We aim to study the regression effect on the level of association by developing a unified inference procedure that can handle all the three data structures.

Chapter 3. Regression Analysis for Marginal Effect Based on Semi-Competing Risks Data

3.1. Model Framework and Data Structure

Let T_1 be the time to the non-terminal event of major interest, usually a status of disease progression, T_2 be the terminal event, such as death. The name of "semi-competing risk" proposed by Fine et al. (2001) explains the fact that T_2 is a competing risk for T_1 but not vice versa. Such a phenomenon has been analyzed by many researchers under the framework of a multi-state model. Figure 3-1 depicts a simple three-state model, also called as a "disability model" or an "illness-death" model. We refer *state 1* as the initial state; *state 2* as the state of disease procession (i.e. occurrence, recurrence, complications, metastases,..., etc.) and *state 3* as the absorbing state such as death.

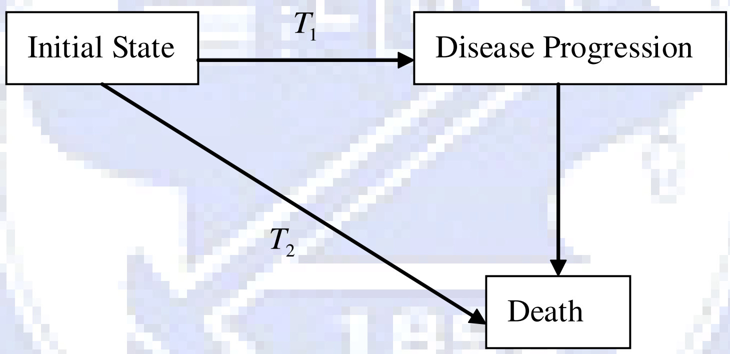

Figure 3-1: Illness-death model.

Since (T_1, T_2) usually represent two types of failure times on the same subject, we may assume that they are both subject to a common external censoring variable denoted as C. Assume that C is independent of (T_1, T_2). Under right censoring, semi-competing risks data (Fine et al., 2001) consist of $(X_i, \delta_{1i}, Y_i, \delta_{2i})$, where

$$X = T_1 \wedge T_2 \wedge C,\ Y = T_2 \wedge C,\ \delta_1 = I(T_1 \leq T_2 \wedge C) \text{ and } \delta_2 = I(T_2 \leq C). \qquad (8)$$

3.2. Marginal Regression Model on the Non-terminal Event

The major goal of the first project is to assess the covariate effect on the non-terminal event time, T_1, based on the data $(X_i, Y_i, \delta_{1i}, \delta_{2i}, Z_i)$. The following regression model will be considered:

$$h(T_1) = -Z^T\theta + \varepsilon, \tag{9}$$

where Z is the $p \times 1$ discrete covariate vector, θ is the $p \times 1$ parameter vector, $h(t)$ is a monotone function and ε is the error term. The parameter θ, which measures the covariate effect on T_1, is of major interest.

Model (9) can be classified into two classes. One class assumes that $h(t)$ is a known monotone function but leaves the distribution of ε to be unknown. For example, when $h(t) = t$, the model becomes a location-shift model with $T_1 = -Z^T\theta + \varepsilon$. When $h(t) = \log(t)$, the model follows an accelerated failure time Model such that $\log(T_1) = -Z^T\theta + \varepsilon$. The other class assumes that $h(t)$ is unknown but the distribution of ε is completely specified. For example, when ε is the extreme value distribution, the model becomes the Cox proportional hazard model such that

$$F_{1,z}(x) = F_{1,0}(x)^{\exp(z^T\theta)},$$

where $F_{1,z}(x) = \Pr(T_1 \geq x | Z = z)$. When ε is the standard logistic distribution, the model follows the proportional odds model with

$$logit(1 - F_{1,z}(x)) = logit(1 - F_{1,0}(x)) + z^T\theta.$$

The challenge of estimating θ comes from the fact that T_1 is subject to dependent censoring by T_2. In fact, the marginal distribution of T_1 is not identifiable nonparametrically. Therefore, besides the model assumption in (9), additional assumptions are needed for estimation of θ. A popular approach is to model marginal regression effects on both T_1 and T_2 and impose the assumption that the bivariate error variables, after removing the marginal effects, do not depend on the covariates. This approach has been adopted by many authors and the technique of artificial censoring is used to construct unbiased estimating functions. For example, under a two-sample setting, Lin et al. (1996) considered a bivariate location-shift model and Chang (2000) assumed a bivariate accelerated failure time model. This research direction has been further extended to general regression settings in which the non-terminal event is generalized to be recurrence events (Ghosh and Lin, 2003; Lin and Ying, 2003) while death still serves as a terminal event. The technique of artificial censoring is used in these papers to handle the problem of dependent censoring. Despite theoretically appealing, the degree of artificial censoring affects the efficiency of the resulting estimator. Peng & Fine (2006) extended the setting of Lin et al. (1996) to include multiple covariates, but proposed a new technique of artificial censoring which can improve the efficiency loss. Besides the drawback of the artificial censoring technique, the model assumptions are somewhat restrictive. Specifically these methods implicitly assume that the dependence structures for the two groups, or for all the levels of covariates, are the same. In other words, they do not account for the situation that covariates may affect the dependence structure. In Section 3.3, we review the aforementioned papers.

In Section 3.4, we present our proposed methodology for estimating θ in presence of dependent censoring. Our idea considers to model the dependence between T_1 and T_2 by an AC model for each covariate group. Then θ can be estimated without imposing additional regression model on T_2. We propose model checking methods to verify the appropriateness of the two types of model assumptions. Except for the modeling flexibility, our approach can make use of the available data, compared to the methods which require artificial censoring.

3.3. Inference Methods Based on Artificial Censoring

The definitions of T_1, T_2 and C have been given earlier. Let Z be the covariate. Usually it is assumed that C is independent of (T_1, T_2) conditional on Z. We review four papers which utilize the technique of artificial censoring to remove the bias due to dependent censoring.

3.3.1. Location-Shift Model

Lin, Robins & Wei (1996) considered a simplified case with binary Z. Specifically Z denotes the group indicator taking values of 0 or 1. The model assumed that there exist some unknown constants θ_0 and η_0 such that the bivariate random vectors $(T_{1i} - \theta_0 Z_i, T_{2i} - \eta_0 Z_i)$ $(i = 1, 2, ..., n)$ are independently and identically distributed from an unspecified joint distribution not depending on Z_i. In other words, (T_1, T_2) follow a bivariate location-shift model but the underlying dependence structure is un-specified.

The main objective of their paper is to draw inference about θ_0 but it is easier to obtain an estimator of η_0 first since T_2 is subject to independent censoring by C. In fact estimation of η_0 has been well studied (Louis, 1981; Wei & Gail, 1983). By constructing a log-rank statistic based on the transformed data $\{\check{Y}_i(\eta), \delta_{2i}, Z_i\}$ $(i = 1, 2, ..., n)$, where $\check{Y}_i(\eta) = Y_i - \eta Z_i$, an estimating function for η_0 is given by

$$U_1^L(\eta) = n^{-1/2} \sum_{i=1}^{n} \delta_{2i} [Z_i - \frac{\sum_{j=1}^{n} I\{\check{Y}_j(\eta) \geq \check{Y}_i(\eta)\} Z_j}{\sum_{j=1}^{n} I\{\check{Y}_j(\eta) \geq \check{Y}_i(\eta)\}}].$$

The estimator $\hat{\eta}$ is the solution to $U_1^L(\eta) = 0$. Because the random variable $(T_{2i} - \eta_0 Z_i)$ $(i = 1, 2, ..., n)$ have the same distribution and because the censoring time C_i is independent of Y_i in each group, the statistic $U_1^L(\eta_0)$ is asymptotically zero-mean normal (Fleming & Harrington, 1991). Therefore, $\hat{\eta}$ is consistent and asymptotically normal (Louis, 1981; Wei & Gail, 1983).

In order to construct a valid estimating function of θ_0, under dependent censoring, the authors suggested to transform X_i and δ_{1i} to $\check{X}_i(\eta, \theta)$ and $\check{\delta}_{1i}(\eta, \theta)$, where

$$\begin{aligned}
\check{X}_i(\theta, \eta) &= (T_{1i} - \theta Z_i) \wedge (T_{2i} - \eta Z_i - d) \wedge (C_i - \eta Z_i - d), \\
\check{\delta}_{1i}(\theta, \eta) &= I\{(T_{1i} - \theta Z_i) \leq (T_{2i} - \eta Z_i - d) \wedge (C_i - \eta Z_i - d)\},
\end{aligned}$$

with $d = 0$ if $\theta \le \eta$ and $d = \theta - \eta$ if $\theta > \eta$. Note that uncensored observations may become censored ones after the transformation. The hypothetical variable $(T_{1i} - \theta Z_i)$ can be "observed" as $\check{X}_i(\theta, \eta)$ in presence of two sources of censoring by $(T_{2i} - \eta Z_i - d)$ and $(C_i - \eta Z_i - d)$ with $\check{\delta}_{1i}(\theta, \eta)$ being the associated failure indicator. The transformed data $\{\check{X}_i(\theta, \eta), \check{\delta}_{1i}(\theta, \eta), Z_i\}$ $(i = 1, 2, ..., n)$ become homogeneous when $(\theta, \eta) = (\theta_0, \eta_0)$. In other words, the transformation makes the joint distribution of the two groups on the upper wedge be the same. Consequently the following log-rank type estimating function can be constructed

$$U_2^L(\theta, \eta) = n^{-1/2} \sum_{i=1}^{n} \check{\delta}_{1i}(\theta, \eta)[Z_i - \frac{\sum_{j=1}^{n} I\{\check{X}_j(\theta, \eta) \ge \check{X}_i(\theta, \eta)\} Z_j}{\sum_{j=1}^{n} I\{\check{X}_j(\theta, \eta) \ge \check{X}_i(\theta, \eta)\}}],$$

which is centred around 0 when $(\theta, \eta) = (\theta_0, \eta_0)$. Hence, $U_2^L(\theta, \eta)$ is a reasonable estimating function.

The authors proposed a two-stage estimation procedure instead of solving the equations $U_j^L(\theta, \eta) = 0$ $(j = 1, 2)$ jointly. Specifically they suggested to obtain $\hat{\eta}$ from $U_1^L(\eta)$ and then solve $U_2^L(\theta, \hat{\eta}) = 0$. Consistency and asymptotic normality of $\hat{\theta}$ are proved in the paper.

3.3.2. Accelerated Failure-Time Model

The methodology of Chang (2000) can be applied to recurrence data. Here, we consider the simple situation with no recurrence (i.e. $K = 1$ in Chang (2000)). The notations in Chang (2000) have been modified According to our setup. Let Z be the binary covariate which is coded as $Z_i = 0$ if subject i is in group 1 and $Z_i = 1$ if subject i is in group 2. For the bivariate accelerated failure time model, Chang assumed that there exist some unknown constants θ_0 and η_0 such that the bivariate random vectors $(e^{\theta_0 Z_i} T_{1i}, e^{\eta_0 Z_i} T_{2i})$ $(i = 1, 2, ..., n)$ are independently identically distributed with an unspecified joint distribution that is independent of Z_i.

Estimation of η_0 is quite standard. Louis (1981) and Wei and Gail (1983) considered a log-rank-type estimating function for η_0 based on the transformed data $\{(e^{\eta Z_i}Y_i, \delta_{2i})\}$ $(i = 1, 2, ..., n)$ as follows:

$$U_1^C(\eta) = n^{-1/2}\sum_{i=1}^{n}\delta_{2i}[Z_i - \frac{\sum_{j=1}^{n} I\{e^{\eta Z_j}Y_j \geq e^{\eta Z_i}Y_i\}Z_j}{\sum_{j=1}^{n} I\{e^{\eta Z_j}Y_j \geq e^{\eta Z_i}Y_i\}}].$$

Again, the challenging part is in estimation of θ_0. Motivated by Lin et al. (1996), Chang (2000) suggested to re-scale the observed censoring time Y_i by multiplying by $e^{\eta Z_i - \nu}$, where $\nu = \max\{0, \eta - \theta\}$, in order to guarantee the value of $\eta Z_i - \nu$ no greater than the value of θZ_i for each Z_i. Therefore, the transformed data become

$$\check{X}_i(\theta, \eta) = e^{\theta Z_i}T_{1i} \wedge e^{\eta Z_i - \nu}Y_i,$$

$$\check{\delta}_{1i}(\theta, \eta) = I\{e^{\theta Z_i}T_{1i} \leq e^{\eta Z_i - \nu}Y_i\}.$$

For a censored observation, $e^{\theta Z_i}T_{1i}$ must be larger than $e^{\eta Z_i - \nu}Y_i$ and therefore $\check{\delta}_{1i}(\theta, \eta) = 0$. However, an uncensored observation may be artificially censored because it is likely that $e^{\theta Z_i}T_{1i}$ may exceed $e^{\eta Z_i - \nu}Y_i$ after the transformation. The proposed estimating function of θ based on the transformed data, $(\check{X}_i(\theta, \eta), \check{\delta}_{1i}(\theta, \eta), Z_i)$ $(i = 1, 2, ..., n)$, is given by

$$U_2^C(\theta, \eta) = n^{-1/2}\sum_{i=1}^{n}\check{\delta}_{1i}(\theta, \eta)[Z_i - \frac{\sum_{j=1}^{n} I\{\check{X}_j(\theta, \eta) \geq \check{X}_i(\theta, \eta)\}Z_j}{\sum_{j=1}^{n} I\{\check{X}_j(\theta, \eta) \geq \check{X}_i(\theta, \eta)\}}].$$

The resulting estimators denoted as $(\hat{\theta}, \hat{\eta})'$ which solve $U^C(\theta, \eta) = (U_1^C(\eta), U_2^C(\theta, \eta))' = 0$. In Chang (2000), she showed the corresponding estimators are consistent and asymptotic normal.

3.3.3. Location-Shift Model by Pairwise Artificial Censoring

Peng & Fine (2006) proposed a method for estimating the marginal effect on T_1 which uses a new artificial censoring technique by pairwise ranking. Following Lin et al. (1996), they assumed a bivariate model for (T_{1i}, T_{2i}), $i = 1, 2, ..., n$: $T_{1i} = \theta_0^T Z_i + \varepsilon_{1i}$ and

$T_{2i} = \eta_0^T Z_i + \varepsilon_{2i}$. They also made the assumption that $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i})^T$ $(i = 1, 2, ..., n)$ are independently and identically distributed with an unspecified joint survivor function not depending on Z_i.

Because T_2 is subject to independent right censoring by C, estimation of η_0 reduces to a classical estimation problem. A log-rank test (Wei et al. (1990); Tsiatis (1990)) has been constructed based on $\breve{\varepsilon}_{2i}(\eta) = Y_i - \eta^T Z_i$, which is the observed proxy of the true error ε_{2i} if $\eta = \eta_0$. It can be written as

$$U_1^P(\eta) = n^{-1/2} \sum_{i=1}^n \delta_{2i}[Z_i - \frac{\sum_{j=1}^n I\{\breve{\varepsilon}_{2j}(\eta) \geq \breve{\varepsilon}_{2i}(\eta)\} Z_j}{\sum_{j=1}^n I\{\breve{\varepsilon}_{2j}(\eta) \geq \breve{\varepsilon}_{2i}(\eta)\}}].$$

An estimator $\hat{\eta}$ is obtained as the solution of $U_1^P(\eta) = 0$ which is consistent and asymptotically normal.

Different from the artificial censoring technique of Lin et al. (1996) and Chang (2000), Peng & Fine (2006) proposed an alternative rank estimator which can reduce the level of artificial censoring. Specifically they suggested to trim separately within pairs of observations, say $\{(X_i, Y_i, \delta_{1i}, \delta_{2i}, Z_i), (X_j, Y_j, \delta_{1j}, \delta_{2j}, Z_j)\}$. The artificial censoring parameter is no longer fixed across pairs and does not involve the bounds on the support of the covariate distribution. A different value is determined for each pair using the covariate vectors Z_i and Z_j. The data transformation within the (i, j) pair is $\{\check{X}_{i(j)}(\beta), \check{\delta}_{1,i(j)}(\beta); \check{X}_{j(i)}(\beta), \check{\delta}_{1,j(i)}(\beta)\}$, where

$$\check{X}_{i(j)}(\beta) = (T_{1i} - \theta^T Z_i) \wedge \{T_{2i} - \eta^T Z_i - d_{ij}(\beta)\} \wedge \{C_i - \eta^T Z_i - d_{ij}(\beta)\},$$

$$\check{\delta}_{1,i(j)}(\beta) = I[(T_{1i} - \theta^T Z_i) \leq \{T_{2i} - \eta^T Z_i - d_{ij}(\beta)\} \wedge \{C_i - \eta^T Z_i - d_{ij}(\beta)\}],$$

$d_{ij}(\beta) = \max\{0, (\theta - \eta)^T Z_i, (\theta - \eta)^T Z_j\}$ and $\beta = (\theta^T, \eta^T)^T$. The choice of $d_{ij}(\beta)$ ensures that both $\check{X}_{i(j)}(\beta)$ and $\check{\delta}_{1,i(j)}(\beta)$ can be determined from the observed data. Because $d_{ij}(\beta)$ is always no greater than d defined in §3.3.1, the level of artificial censoring can be reduced the approach of Lin et al. (1996). Define $\psi_{ij}(\beta) = \check{\delta}_{1,i(j)}(\beta) I\{\check{X}_{i(j)}(\beta) \leq$

$\check{X}_{j(i)}(\beta)\} - \check{\delta}_{1,j(i)}(\beta)I\{\check{X}_{j(i)}(\beta) \leq \check{X}_{i(j)}(\beta)\}$ and notice that $(Z_i - Z_j)\psi_{ij}(\beta)$ is symmetric in i and j. Because $(T_{1i} - \theta_0^T Z_i, T_{2i} - \eta_0^T Z_i)$ and $(T_{1j} - \theta_0^T Z_j, T_{2j} - \eta_0^T Z_j)$ are independent and with a common survival function on the upper wedge, $\psi_{ij}(\beta)$ has mean 0 at the true value $\beta_0 = (\theta_0^T, \eta_0^T)^T$. This suggests the following U-statistic estimating equation

$$U_2^P(\beta) = 2\sqrt{n} \sum_{1 \leq i < j \leq n} \frac{(Z_i - Z_j)\psi_{ij}(\beta)}{n(n-1)}.$$

It follows that that $E\{U_2^P(\beta_0)\} = 0$. Thus a reasonable estimator for θ_0, denote by $\hat{\theta}$, can be obtained by solving $U_2^P\{(\theta^T, \hat{\eta}^T)^T\} = 0$. Furthermore, it is shown that $\hat{\beta}$ is consistent and asymptotic normal.

3.3.4. Artificial Censoring under Flexible Regression Models

The manuscript by Ding, Wang, Hsieh & Shi (2006) extend the artificial censoring technique to a flexible regression setting. They considered the following models:

$$\begin{aligned} h_1(T_1) &= \theta^T Z + \varepsilon_1 \\ h_2(T_2) &= \eta^T Z + \varepsilon_2, \end{aligned} \tag{10}$$

where $h_1(t)$ is a known monotone function, $h_2(t)$ is a monotone function which may be known or unknown, ε_j $(j = 1, 2)$ are the error terms. It is assumed that $(\varepsilon_{1i}, \varepsilon_{2i})^T$ $(i = 1, ..., n)$ are independently and identically distributed with an unspecified joint survivor function not depending on Z_i.

Consider the situation when $h_2(t)$ is known and the distribution of ε_2 is unknown. Estimation of η_0 is straightforward. Let $\check{T}_2(\eta) = h_2(T_2) - \eta^T Z$, $\check{C}(\eta) = h_2(C) - \eta^T Z$ and $\check{Y}_i(\eta) = \check{T}_{2i}(\eta) \wedge \check{C}_i(\eta)$. The estimating function for η can be constructed as

$$U_1^D(\eta) = n^{-1} \sum_{i=1}^{n} \delta_{2i}[Z_i - \frac{\sum_{j=1}^n I\{\check{Y}_j(\eta) \geq \check{Y}_i(\eta)\} Z_j}{\sum_{j=1}^n I\{\check{Y}_j(\eta) \geq \check{Y}_i(\eta)\}}].$$

An estimator $\hat{\eta}$ is the solution of $U_1^D(\eta) = 0$.

Estimation of θ faces a new challenge that involves specification of an appropriate transformation that can handle the more general model assumptions. Let $\check{T}_1(\theta) = h_1(T_1) - \theta^T Z$. For semi-competing risks data, (T_1, T_2) are observable only if $T_1 \le T_2$. Hence, $\check{T}_1(\theta) = h_1(T_1) - \theta^T Z$ is subject to censoring by

$$h_1(T_2) - \theta^T Z = h_1 \circ h_2^{-1}(\check{T}_2(\eta) + \eta^T Z) - \theta^T Z. \tag{11}$$

The goal is to find a transformation which makes the joint distribution of all covariate groups on the upper wedge to be the same. Hence, equation (11) has to be hold for all values of Z. Therefore, the authors suggested the following transformation

$$H_{\theta,\eta}(t) = \inf_{z\in\Omega} h_1 \circ h_2^{-1}(t + \eta^T z) - \theta^T z,$$

where Ω is the set of possible Z. The resulting transformed variables become

$$\begin{aligned} \check{X}_i(\theta,\eta) &= \check{T}_{1i}(\theta) \wedge H_{\theta,\eta}(\check{T}_{2i}(\eta) \wedge \check{C}_i(\eta)) \\ \check{\delta}_{1i}(\theta,\eta) &= I\{\check{T}_{1i}(\theta) \le H_{\theta,\eta}(\check{T}_{2i}(\eta) \wedge \check{C}_i(\eta))\}. \end{aligned}$$

Figure 3-2 provides a graphical illustration of the transformation based on a simple example with a binary Z, $h_1(t) = t$, $h_2(t) = \log(t)$, and $\theta_0 = \eta_0 = 1$. Accordingly, the estimating function based on the above transformation can be written as

$$U_2^D(\theta,\eta) = n^{-1}\sum_{i=1}^{n} \check{\delta}_{1i}(\theta,\eta)[Z_i - \frac{\sum_{j=1}^n I\{\check{X}_j(\theta,\eta) \ge \check{X}_i(\theta,\eta)\}Z_j}{\sum_{j=1}^n I\{\check{X}_j(\theta,\eta) \ge \check{X}_i(\theta,\eta)\}}].$$

The estimator $\hat{\theta}$ is the solution of $U_2^D(\theta,\hat{\eta}) = 0$. Denote $\hat{\beta} = (\hat{\theta}^T, \hat{\eta}^T)^T$ as the estimators and $\beta_0 = (\theta_0^T, \eta_0^T)^T$ as the true parameters. It has been showed that $n^{1/2}(\hat{\beta} - \beta_0)$ converges to a bivariate mean-zero normal distribution.

Consider the other situation that $h_2(t)$ is a unknown monotone function but the distribution of ε_2 is completely specified. Additional difficulty arises since $H_{\theta,\eta}(t)$ involves specification of $h_2(t)$ which is unknown. Because the distribution of ε_2 is known, hence

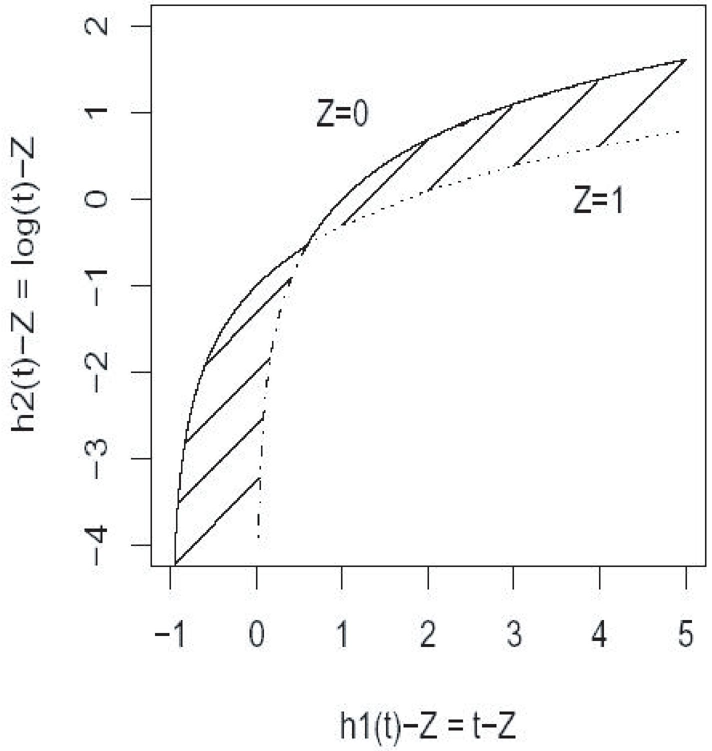

Figure 3-2: The censoring lines draws in the transformed scale with $h_1(t) = t$, $h_2(t) = \log(t)$, and $\theta_0 = \eta_0 = 1$; the shaded region is artificially censoring.

$\tilde{S}_2(t) = \Pr(\varepsilon_2 > t)$ is also a known function. Define the baseline survival function of T_2 as

$$S_2(t) = \Pr(T_2 > t | Z = 0).$$

Since $S_2(h_2^{-1}(t)) = \tilde{S}_2(t)$, it follows that $h_2^{-1}(t) = S_2^{-1} \circ \tilde{S}_2(t)$ and $h_2(t) = \tilde{S}_2^{-1} \circ S_2(t)$. Therefore,

$$H_{\theta,\eta}(t) = \inf_{z \in \Omega} h_1(S_2^{-1}(\tilde{S}_2(t + \eta^T z))) - \theta^T z,$$

which is still unknown since $S_2(t)$ is unknown. It is recommended to plug in a consistent estimator $\hat{S}_2(t)$ for $S_2(t)$. Therefore, $H_{\eta,\theta}(t)$ can be replaced by $\hat{H}_{\theta,\eta}(t) = \inf_{z\in\Omega} h_1(\hat{S}_2^{-1}(\tilde{S}_2(t+\eta^T z))) - \theta^T z$. By applying existing methods, a reasonable estimator of $S_2(t)$ can be obtained. For example, under the Cox Proportional hazard model, one may estimate $S_2(t)$ by the Nelson-Aalen estimator.

3.3.5. Discussions

All the methods discussed above are developed under two types of assumptions. The first assumption is about marginal regression models for both T_1 and T_2. The other assumption is that the dependence structure is the same for each covariate group. In practical applications, it may happen that different treatment plans change a patient's internal biological system. Therefore assuming a common dependence structure for different treatment groups may not be convincing. Furthermore when Z includes multiple covariates as in the last two papers, the degree of efficiency loss due to artificial censoring may be substantial. In Section 3.4, we consider another alternative to handle such a situation.

3.4. The Proposed Methodology for Marginal Regression Analysis

As in the previous papers we assume that

$$h(T_1) = -Z^T\theta + \varepsilon,$$

where T_1 is the time to a non-terminal event and is subject to censoring by T_2. Given semi-competing risks data $(X_i, \delta_{1i}, Y_i, \delta_{2i}, Z_i)$ $(i = 1, ..., n)$, the parameter θ is not identifiable without making additional assumptions. Unlike the aforementioned papers, we do not want to specify the marginal regression on T_2 as in (10). Instead, we model the dependence structure of $(T_1, T_2)|Z$ for each value of Z under an Archimedean copulas assumption. Note that Z is a discrete covariate vector. We assume that (T_1, T_2) jointly follow an Archimedean copulas in the upper wedge $\mathcal{P} = \{(x, y) : 0 < x \leq y < \infty\}$. Accounting for the possibility that the dependence structures for different covariate groups are different, we assume separate Archimedean copula (AC) models for each group such that

$$\begin{aligned} F_z(x, y) &= C_{z,\alpha_z}\{F_{1,z}(x), F_{2,z}(y)\} \\ &= \phi^{-1}_{z,\alpha_z}\{\phi_{z,\alpha_z}[F_{1,z}(x)] + \phi_{z,\alpha_z}[F_{2,z}(y)]\}, \end{aligned} \tag{12}$$

where $\phi_{z,\alpha_z}(\cdot) : [0, 1] \mapsto [0, \infty]$ is a generator function as described in Section 2.1, $F_z(x, y) = \Pr(T_1 \geq x, T_2 \geq y|Z = z)$, $F_{1,z}(x) = \Pr(T_1 \geq x|Z = z)$ and $F_{2,z}(y) = \Pr(T_2 \geq y|Z = z)$. Note that, for different groups, we allow not only different association parameter α_z but also different forms of $\phi_{z,\alpha_z}(\cdot)$.

3.4.1. A Two-Stage Inference Procedure

The p-dimensional covariate vector for subject i is denoted as Z_i which takes discrete values, say $z_1, \ldots, z_K$. Denote $n_k = \sum_{i=1}^n I(Z_i = z_k)$ as the number of observations for the kth sub-sample and $n = \sum_{k=1}^K n_k$. The proposed inference procedure contains two steps. The parameters in model (12), namely α_z, $F_{2,z}(y)$, $F_z(x, y)$ and $F_{1,z}(x)$, are estimated in the first stage. In the second stage, the proposed estimating function of θ is constructed based on the estimator of $F_{1,z}(x)$.

3.4.1.1. First-Stage: Estimating Nuisance Parameters

First we obtain the estimators of $F_z(x, y)$, $F_{2,z}(y)$, $F_{1,z}(x)$, $G(y) = \Pr(C \geq y)$ and α_z, denoted as $\hat{F}_z(x, y)$, $\hat{F}_{2,z}(y)$, $\hat{F}_{1,z}(x)$, $\hat{G}(y)$ and $\hat{\alpha}_z$ respectively, by applying existing methods in the literature to the sub-sample with $Z = z$.

For $x \leq y$, it follows that $F_z(x, y) = \Pr(X \geq x, Y \geq y|Z = z)/G(y)$. Hence using the plug-in approach, $F_z(x, y)$ can be estimated by

$$\hat{F}_z(x, y) = \hat{\Pr}\{X \geq x, Y \geq y|Z = z\}/\hat{G}(y) = \frac{\sum_{i=1}^{n} I(X_i \geq x, Y_i \geq y, Z_i = z)}{n_z \hat{G}(y)}, \quad (13)$$

where $\hat{G}(y) = \prod_{u<y}[1 - (\sum_{i=1}^{n} I(Y_i = u, \delta_{2i} = 0)/\sum_{i=1}^{n} I(Y_i \geq u))]$. This estimator is based on the assumption that covariates Z do not affect the distribution of censoring variable C. In the situation that the distribution of C depends on discrete covariate Z, $\hat{G}(y)$ can be modified by the corresponding K-M estimator $\hat{G}_z(y)$ which uses only those data points with $Z_i = z$. Similarly the estimator of $F_{2,z}(y)$ is given by

$$\hat{F}_{2,z}(y) = \frac{\sum_{i=1}^{n} I(Y_i \geq y, Z_i = z)}{n_z \hat{G}(y)}. \quad (14)$$

There exist several estimators of α_z based on semi-competing risks data. Assuming the Clayton model in the upper wedge, the estimating function proposed by Day et al. (1997) was constructed based on two-by-two tables and that proposed by Fine et al. (2001) utilized the concordant information for paired observations. Wang (2003) generalized the former approach to general AC models. In absence of covariates, her estimating function of α can be expressed as

$$L(\alpha, \hat{\eta}) = n^{-1} \int\int_{(x,y)\in\mathcal{P}} w(x, y)\{N_{11}(dx, dy) - \tilde{E}_{11}(dx, dy; \alpha, \hat{\eta})\}, \quad (15)$$

where $w(x, y)$ is a weight function,

$$\tilde{E}_{11}(dx, dy; \alpha, \eta) = \frac{\theta_{\alpha,\eta}(x, y)N_{10}(dx, y)N_{01}(x, dy)}{\theta_{\alpha,\eta}(x, y)N_{10}(dx, y) + R(x, y) - N_{10}(dx, y)},$$

$N_{11}(dx, dy) = \sum_{i=1}^n I(X_i = x, Y_i = y, \delta_{1i} = 1, \delta_{2i} = 1)$, $N_{10}(dx, y) = \sum_{i=1}^n I(X_i = x, \delta_{1i} = 1, Y_i \geq y)$, $N_{01}(x, dy) = \sum_{i=1}^n I(X_i \geq x, Y_i = y, \delta_{2i} = 1)$, $R(x, y) = \sum_{i=1}^n I(X_i \geq x, Y_i \geq y)$ and $\theta_{\alpha,\eta}(x, y) = \tilde{\theta}_\alpha\{F(x, y)\}$ with

$$\tilde{\theta}_\alpha(v) = -v \frac{\partial^2 \phi_\alpha(v)/\partial v^2}{\partial \phi_\alpha(v)/\partial v} = -v \frac{\phi_\alpha''(v)}{\phi_\alpha'(v)}$$

and $\eta = F(x, y)$ can be estimated by $\hat{\eta} = \hat{F}(x, y)$ using the formula in (13) without further partitioning by Z.

Here we modify Wang's method to estimate α_z by using only data points with $Z_i = z$. Then based on (12), one can derive $F_{1,z}(x)$ in terms of $\phi_{z,\alpha_z}(\cdot)$, $F_z(x, y)$ and $F_{2,z}(y)$. Fine et al. (2001) suggested to consider the relationship on the diagonal line $y = x$ and, by straightforward calculation, we get

$$F_{1,z}(x) = \phi_{z,\alpha_z}^{-1}\{\phi_{z,\alpha_z}[F_z(x, x)] - \phi_{z,\alpha_z}[F_{2,z}(x)]\} = H_z(F_z(x, x), F_{2,z}(x), \alpha_z). \quad (16)$$

The marginal function $F_{1,z}(x)$ can be estimated by

$$\hat{F}_{1,z}(x) = \phi_{z,\hat{\alpha}_z}^{-1}\{\phi_{z,\hat{\alpha}_z}[\hat{F}_z(x, x)] - \phi_{z,\hat{\alpha}_z}[\hat{F}_{2,z}(x)]\} = H_z(\hat{F}_z(x, x), \hat{F}_{2,z}(x), \hat{\alpha}_z). \quad (17)$$

3.4.1.2. Second-Stage: Estimating the Regression Parameter

The proposed estimating equation of θ is motivated by the following two-sample test statistic with $Z = 0, 1$. Specifically to test $F_{1,0}(t) = F_{1,1}(t)$ for every t within the range of the data, one can use

$$U_T = \sqrt{\frac{n_0 n_1}{n}} \int W(x) \left\{ \hat{F}_{1,0}(x) - \hat{F}_{1,1}(x) \right\} dx, \quad (18)$$

where $W(x)$ is a weight function.

Now we modify the test statistic U_T in (18) to construct an estimating equation for one-dimensional θ with $Z = 0, 1$. Let θ_0 be the true value of θ. Model (9) induces a functional transformation $\xi_\theta(\cdot)$ such that $\xi_{\theta_0}(F_{1,0}) = F_{1,1}$. When $h(\cdot)$ is known but the distribution of ε is unknown, $\xi_\theta(F)(t) = F[h^{-1}\{h(t) + \theta\}]$; when $h(\cdot)$ is unknown but the distribution of ε is known, $\xi_\theta(F)(t) = F_\varepsilon\left[F_\varepsilon^{-1}\{F(t)\} + \theta\right]$, where $F_\varepsilon(t) = \Pr(\varepsilon \geq t)$ denotes the survival function of ε. Now we can define a function $g(t, \theta)$ such that

$$g(t, \theta) = \xi_\theta(F_{1,0})(t) - F_{1,1}(t).$$

Then $g(t, \theta_0) = 0$ for all t. Since $\sqrt{\frac{n_0 n_1}{n}} \int W(x) g(x, \theta_0) dx = 0$, we can then estimate θ by solving the corresponding estimating equation

$$U(\theta) = \sqrt{\frac{n_0 n_1}{n}} \int W(x) \hat{g}(x, \theta) dx = 0,$$

where $\hat{g}(t, \theta) = \xi_\theta(\hat{F}_{1,0})(t) - \hat{F}_{1,1}(t)$.

The above idea can be modified to account for the situation that Z contains multiple covariates but all of them have finite discrete values. In such a case, let $\{z_k, k = 1, 2, ..., K\}$ denote the set of all possible Z values. Now z_k, θ and θ_0 are $p \times 1$ vectors. When model (9) is true, it follows that $\xi_{(z_j - z_k)^T \theta_0}(F_{1,z_k}) = F_{1,z_j}$. Define $g_{kj}(t, \theta) = \xi_{z_{kj}^T \theta}(F_{1,z_k})(t) - F_{1,z_j}(t)$ and $\hat{g}_{kj}(t, \theta) = \xi_{z_{kj}^T \theta}(\hat{F}_{1,z_k})(t) - \hat{F}_{1,z_j}(t)$, where $z_{kj} = z_j - z_k$ and $\hat{F}_{1,z_k}$ is the estimator (17) based on the sub-sample with $Z = z_k$. The estimating function then becomes

$$U(\theta) = \sum_{k<j} w_0(z_{kj}^T \theta) z_{kj} \sqrt{\frac{n_k n_j}{n_k + n_j}} \{\int_0^{t_{kj}} W_{kj}(t) \hat{g}_{kj}(t, \theta) dt\} \tag{19}$$

where $w_0(\cdot)$ and $W_{kj}(\cdot)$ are the weight functions, and t_{kj} is the largest value of X in the pooled sub-sample with $Z = z_k$ or $Z = z_j$. The proposed estimator of θ is the solution to $U(\theta) = 0$, denoted as $\hat{\theta}$.

Asymptotic properties of $\hat{\theta}$ which solves $U(\theta) = 0$ are given in the following theorem.
Theorem 1A: *Assume that models (9) and (12) hold. Under the regularity conditions stated in Appendix 1, $\hat{\theta}$ is a consistent estimator of θ_0 and $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically normal with mean-zero, where θ_0 is the true value.*
The proof of Theorem 1A is provided in Appendix 2. Since it is not easy to estimate the asymptotic variance of $\hat{\theta}$ by an analytic formula, we suggest to use a bootstrap or a jackknife method to estimate its variance.

In practice, the weight function may also be estimated. Replacing $W_{kj}(t)$ in (19) with $\hat{W}_{kj}(t)$, we have the following estimating function:

$$\hat{U}(\theta) = \sum_{k<j} w_0(z_{kj}^T\theta) z_{kj} \sqrt{\frac{n_k n_j}{n_k + n_j}} \{ \int_0^{t_{kj}} \hat{W}_{kj}(t) \hat{g}_{kj}(t, \theta) dt \}.$$

The Gehan-type weights are often used (page 230 of Klein and Moeschberger, 2003) which can be written as

$$\hat{W}_{kj}(x) = \frac{(n_k + n_j)\hat{G}_{z_k}(x)\hat{G}_{z_j}(x)}{n_k \hat{G}_{z_k}(x) + n_j \hat{G}_{z_j}(x)},$$

where $\hat{G}_{z_k}(x)$ is the Kaplan-Meier estimator of $G_{z_k}(x) = \Pr(C \geq x | Z = z_k)$. Note that $\hat{W}_{kj}(x)$ is an estimator of

$$W_{kj}(x) = \frac{(c_k + c_j) G_{z_k}(x) G_{z_j}(x)}{c_k G_{z_k}(x) + c_j G_{z_j}(x)},$$

where c_k and c_j are the constants defined in the first regularity condition (a) listed in Appendix 1. Let $\tilde{\theta}$ solves $\hat{U}(\theta) = 0$. Its asymptotic properties are stated in the following theorem. In Appendix 3, we present the proof.

Theorem 1B: *If $\hat{W}_{kj}(t)$ uniformly strongly converges to $W_{kj}(t)$, then under the conditions for Theorem 1A, the solution to the estimating equation $\hat{U}(\theta) = 0$ is also asymptotically normal. That is, let $\tilde{\theta}$ denote the solution to $\hat{U}(\theta) = 0$, then $\sqrt{n}(\tilde{\theta} - \theta_0)$ weakly converges to a mean-zero normal random variable, where θ_0 is the true value.*

For computation, we may use the fact that $\hat{F}_{1,0}(t)$ and $\hat{F}_{1,1}(t)$ are piecewise constant functions. Let $t_{(1)} \leq \ldots \leq t_{(n)}$ be the observed ordered times of X in the pooled sample and set $t_{(0)} = 0$. Then $\hat{F}_{1,0}(t)$ and $\hat{F}_{1,1}(t)$ are constants on the time intervals $(t_{(i-1)}, t_{(i)}]$. Usually, the estimated weight functions such as the Gehan-type weights can also be taken to be piecewise constant functions between $t_{(i-1)}$ and $t_{(i)}$ which would enable simplication for compuation. For example, with piecewise constant weight function $\hat{W}(t)$, the quantity corresponding to U_T in (18) can be rewritten as

$$\hat{U}_T = \sqrt{\frac{n_0 n_1}{n}} \sum_{i=1}^{n} W(t_{(i)}) \left\{ t_{(i)} - t_{(i-1)} \right\} \left\{ \hat{F}_{1,0}(t_{(i)}) - \hat{F}_{1,1}(t_{(i)}) \right\}. \qquad (20)$$

For illustration, we now derive the estimating equations under a two-sample setting for selected examples.

Example 1: Cox PH model

When ε has the extreme value distribution, model (9) becomes the Cox proportional hazard model. Then $F_\varepsilon(t) = \exp\{-\exp(t)\}$ and $\xi_\theta(F) = F^{\exp(\theta)}$. When θ equals its true value θ_0, it follows that

$$F_{1,1}(x) = \{F_{1,0}(x)\}^{\exp(\theta_0)}.$$

Therefore $g(t,\theta) = F_{1,0}(t)^{\exp(\theta)} - F_{1,1}(t)$, and the estimating equation is

$$\hat{U}(\theta) = \sqrt{\frac{n_0 n_1}{n}} \int_0^{t_{(n)}} \hat{W}(t) \left\{\hat{F}_{1,0}(t)^{\exp(\theta)} - \hat{F}_{1,1}(t)\right\} dt = 0.$$

Under the piecewise constant weight function, the resulting estimating equation becomes

$$\hat{U}(\theta) = \sqrt{\frac{n_0 n_1}{n}} \sum_{i=1}^{n} \hat{W}(t_{(i)})\{t_{(i)} - t_{(i-1)}\} \left\{\hat{F}_{1,0}(t_{(i)})^{\exp(\theta)} - \hat{F}_{1,1}(t_{(i)})\right\} = 0.$$

Example 2: The proportional odds model

When ε is the standard logistic distribution, model (9) becomes the proportional odds model, where $F_\varepsilon(t) = \frac{1}{1+\exp(t)}$ and $\xi_\theta(F) = \frac{F}{\exp(\theta) - F\exp(\theta) + F}$. When θ equals its true value θ_0, it follows that

$$F_{1,1}(t) = \frac{F_{1,0}(t)}{\exp(\theta_0) - F_{1,0}(t)\exp(\theta_0) + F_{1,0}(t)},$$

and

$$g(t,\theta) = \frac{F_{1,0}(t)}{\exp(\theta) - F_{1,0}(t)\exp(\theta) + F_{1,0}(t)} - F_{1,1}(t).$$

So the estimating equation is

$$\hat{U}(\theta) = \sqrt{\frac{n_0 n_1}{n}} \int_0^{t_{(n)}} \hat{W}(t) \left\{\frac{\hat{F}_{1,0}(t)}{\exp(\theta) - \hat{F}_{1,0}(t)\exp(\theta) + \hat{F}_{1,0}(t)} - \hat{F}_{1,1}(t)\right\} dt = 0.$$

Under the piecewise constant weight function, the resulting estimating equation becomes

$$\hat{U}(\theta) = \sqrt{\frac{n_0 n_1}{n}} \sum_{i=1}^{n} \hat{W}(t_{(i)})\{t_{(i)} - t_{(i-1)}\}[\frac{\hat{F}_{1,0}(t_{(i)})}{\exp(\theta) - \hat{F}_{1,0}(t_{(i)})\exp(\theta) + \hat{F}_{1,0}(t_{(i)})} - \hat{F}_{1,1}(t_{(i)})] = 0.$$

Example 3: The accelerated failure time model

When $h(t) = log(t)$, model (9) becomes the accelerated failure time model. Now $\xi_\theta(F)(t) = F(\exp(\theta)t)$. When θ equals its true value θ_0, it follows that

$$F_{1,1}(t) = F_{1,0}(\exp(\theta_0)t),$$

and

$$g(t,\theta) = F_{1,0}(\exp(\theta_0)t) - F_{1,1}(t).$$

So the estimating equation is

$$\hat{U}(\theta) = \sqrt{\frac{n_0 n_1}{n}} \int \hat{W}(t) \left\{ \hat{F}_{1,0}(\exp(\theta)t) - \hat{F}_{1,1}(t) \right\} dt = 0,$$

where $\hat{F}_{1,0}(\exp(\theta)t) = \hat{\Pr}\{T_1 \geq \exp(\theta)t | Z = 0\} = \hat{\Pr}\{\exp(-\theta)T_1 \geq t | Z = 0\}$ which is the estimator $\hat{F}_{1,0}(t)$ based on the transformed data $\{(\exp(-\theta)X_i, \exp(-\theta)Y_i, \delta_{1i}, \delta_{2i}) : i = 1, 2, ..., n_0\}$ and denote as $\hat{F}^*_{1,0}(t)$. Let $\tilde{t}_{(1)} \leq \ldots \leq \tilde{t}_{(n)}$ be the order times of the pooled sample $\{\exp(-\theta)X_i : i = 1, 2, ..., n_0\}$ and $\{X_j : j = 1, 2, ..., n_1\}$. Under the piecewise constant weight function, the resulting estimating equation becomes

$$\hat{U}(\theta) = \sqrt{\frac{n_0 n_1}{n}} \sum_{i=1}^{n} W(\tilde{t}_{(i)}) \{\tilde{t}_{(i)} - \tilde{t}_{(i-1)}\} [\hat{F}^*_{1,0}(\tilde{t}_{(i)}) - \hat{F}_{1,1}(\tilde{t}_{(i)})] = 0.$$

3.4.2. Model Selection

The proposed procedure is developed based on two assumptions: the dependence structure of an AC model characterized by $\phi_{z,\alpha_z}(\cdot)$ in (12) and the regression model in (9). By specifying the dependence relationship between T_1 and T_2 for each value of Z, we can avoid making unnecessary assumption about the covariate effect on T_2 as in Lin et al. (1996). Now we discuss how to justify the imposed assumptions.

3.4.2.1. Selection of a Copula Model

We first consider how to check whether a copula model $\phi_{z,\alpha_z}(\cdot)$ fits the data at hand for each covariate group. Without loss of generality and to simplify the presentation, the discussions here are based on a homogeneous sample $\{(X_i, \delta_{1i}, Y_i, \delta_{2i})\ (i = 1, 2, ..., n)\}$ such that (T_1, T_2) follows an AC model

$$F(x, y) = C_\alpha(F_1(x), F_2(y)) = \phi_\alpha^{-1}\{\phi_\alpha[F_1(x)] + \phi_\alpha[F_2(y)]\}. \tag{21}$$

We briefly summarize our ideas. Consider the function $F^{11}(t_1, t_2) = \Pr(T_1 \geq t_1, T_2 \geq t_2 | \delta_1 = 1, \delta_2 = 1)$ which is identifiable nonparametrically in the upper wedge $\{(t_1, t_2) : 0 < t_1 \leq t_2 < \infty\}$. By comparing the nonparametric estimator of $F^{11}(t_1, t_2)$ and its model-based estimator for $F^{11}(t_1, t_2)$ based on some distance measure, one can find the most plausible model which is the one that yields the smallest distance among the candidates. Furthermore a formal goodness-of-fit test can be constructed if the distribution of the distance measure under the null hypothesis can be derived. Since analytic derivations are complicated, we suggest using the bootstrap re-sampling method to obtain the cut-off value in the test.

The nonparametric estimator, denoted as $\hat{F}^{11}(t_1, t_2)\ (t_1 \leq t_2)$, is given by $\sum_{i=1}^n I(X_i \geq t_1, Y_i \geq t_2, \delta_{1i} = 1, \delta_{2i} = 1) / \sum_{i=1}^n I(\delta_{1i} = 1, \delta_{2i} = 1)$. Assume that there are K model candidates $C_\alpha^{(k)}(F_1(x), F_2(y))\ (k = 1, 2, ..., K)$, each of which can be characterized by $\phi_\alpha^{(k)}$. Note that the definition of α depends on the chosen model. For an AC model indexed by $\phi_\alpha^{(k)}$, the model-based estimator, denoted as $\tilde{F}_k^{11}(t_1, t_2)$, can be computed over the region $\{t_1 \leq t_2\}$ as follows:

$$\tilde{F}_k^{11}(t_1, t_2) = \frac{\int_{y=t_2}^{\infty} \int_{x=t_1}^{y} \tilde{F}_k(dx, dy)\hat{G}(y)}{\int_{y=0}^{\infty} \int_{x=0}^{y} \tilde{F}_k(dx, dy)\hat{G}(y)},$$

where $\tilde{F}_k(dx, dy) = \tilde{F}_k(x, y) - \tilde{F}_k(x+dx, y) - \tilde{F}_k(x, y+dy) + \tilde{F}_k(x+dx, y+dy)$, $\tilde{F}_k(x, y) = \phi_{\hat{\alpha}}^{(k)^{-1}}\{\phi_{\hat{\alpha}}^{(k)}[\hat{F}_1(x)] + \phi_{\hat{\alpha}}^{(k)}[\hat{F}_2(y)]\}$. To verify whether a copula model $\phi_\alpha^{(k)}$ fits the data, we can perform a formal testing procedure as follows. Consider testing $H_0 : \phi_\alpha = \phi_\alpha^{(k)}$ versus $H_a : \phi_\alpha \neq \phi_\alpha^{(k)}$. Define

$$D^k = \sup_{t_1 \leq t_2} |\hat{F}^{11}(t_1, t_2) - \tilde{F}_k^{11}(t_1, t_2)|. \tag{22}$$

We can reject H_0 if $D^k > c_k$, where c_k is the critical value satisfying $\Pr(D^k > c_k | H_0) = \gamma$, the pre-specified type-one error rate.

Because the distribution of D^k is difficult to derive analytically, we suggest using the bootstrap re-sampling method to obtain the cut-off value, p-value and power. Here we briefly describe the procedure. A bootstrap sample under model $\phi_\alpha^{(k)}$ can be generated as follows. Recall that given the original data, we have obtained $\hat{G}(c)$, $\hat{F}_2(y)$ and $\hat{F}_1(x)$ under the assumption of model $\phi_\alpha^{(k)}$. Then generate $(U_i^*, V_i^*) \sim$ copula model k with $U_i^* \sim U(0,1)$ and $V_i^* \sim U(0,1)$. Then set $T_{1i}^* = s$ if $\hat{F}_1(s^+) < 1 - U_i^* \leq \hat{F}_1(s)$, $T_{2i}^* = t$ if $\hat{F}_2(t^+) < 1 - V_i^* \leq \hat{F}_2(t)$ and $C_i^* \sim \hat{G}(c)$. Given $(T_{1i}^*, T_{2i}^*, C_i^*)$ $(i = 1, ..., n)$, we can construct a bootstrap sample $\{(X_i^*, \delta_{1i}^*, Y_i^*, \delta_{2i}^*)\ (i = 1, 2, ..., n)\}$, where $X_i^* = T_{1i}^* \wedge T_{2i}^* \wedge C_i^*$, $Y_i^* = T_{2i}^* \wedge C_i^*$, $\delta_{1i}^* = I(T_{1i}^* \leq T_{2i}^* \wedge C_i^*)$ and $\delta_{2i}^* = I(T_{2i}^* \leq C_i^*)$. With a bootstrapped sample, we can compute the corresponding values of D^k. Repeating the bootstrapping procedure many times, the distribution of D^k can be approximated by the empirical counterparts from the bootstrapped samples.

The above tests will reject the null hypothesis if the data obviously violates the copula model $\phi_\alpha^{(k)}$. In practice, we may be more interested in choosing the best fitted copula model from several candidates indexed by $k = 1, 2, ..., K$. For this purpose, we can select the model that yields the smallest D^k.

Now we derive theoretical properties of the proposed model selection procedure. In Appendix 4, we provide the proof of Theorem 2.

Theorem 2: *Assume that (T_1, T_2) follow model (21) and both variables are continuous and the independent censoring variable C has bigger support than the supports of T_1 and T_2. Suppose that there are K model candidates in the AC family. Let the kth model $C_\alpha^{(k)}(u, v)$ be characterized by $\phi_\alpha^{(k)}(t)$ which possesses regular analytic properties in t and is continuous in α whose parameter space is a closed set. If $\phi_\alpha^{(k)}$ is the true copula model, $D^k \xrightarrow{P} 0$ as $n \to \infty$. If $\phi_\alpha^{(k)}$ is not the true model, $Pr(\liminf_{n\to\infty} D^k > 0) = 1$. Furthermore let $\hat{k}$ denote the copula model that yields the smallest D^k among all the candidates. Then $\phi_\alpha^{(\hat{k})}$ is consistent if the true copula model is included in the list of candidates.*

3.4.2.2. Selection of the Covariate Model

After specifying the form of model (12), our procedure requires choosing an appropriate regression model in (9). If model (9) is correctly specified, $g_{kj}(t, \theta_0) = 0$ and it is reasonable to expect that $\hat{g}_{kj}(t, \hat{\theta})$ is closer to zero for the correct model than a wrong model for moderate sample sizes. This fact can be used to check the model assumption (9). Let $D_R = \max_{k,j,t} |\hat{g}_{kj}(t, \hat{\theta})|$. A formal model checking procedure can be formulated as testing the hypothesis H_0: the form of model (9) is correct versus H_a: the form of model (9) is not correct. The null hypothesis is rejected if D_R is too big. The cutoff value for the test can be calculated by applying the bootstrapped method which can also be used for model selection. Suppose that there are several choices for model (9), say model $k = 1, 2, ..., K$. To select the best fitted model, we can simply choose the one with smallest D_R^k, where D_R^k is calculated under model k.

3.4.3. Numerical Analysis

3.4.3.1. Simulation Results

We design several simulation settings to examine the validity and robustness of the proposed methods. Data generation algorithms for the Clayton model and the Frank model have been given in Prentice and Cai (1992) and Genest (1987), respectively. In the following analysis, we set the weight functions as $w_0(z_{ij}'\theta) = 1$ and $\hat{W}_{ij}(x) = (n_i + n_j)\hat{G}_{z_i}(x)\hat{G}_{z_j}(x)/(n_i\hat{G}_{z_i}(x) + n_j\hat{G}_{z_j}(x))$. For each estimator under evaluation, the average bias and the standard deviation based on 1000 runs are reported.

The first analysis compared our proposed estimator $\hat{\theta}$ and its competitor estimator $\hat{\theta}_L$ proposed by Lin et al. (1996). The results are summarized in Table 3-1 and 3-2. We set $(\varepsilon, \xi)|Z$ to follow an AC model with $Z = 0, 1$. Then based on (ε, ξ, Z), the value of (T_1, T_2) can be determined from the models $h_1(T_1) = -\theta_0 Z + \varepsilon$ and $h_2(T_2) = -\eta_0 Z + \xi$. Here we set $\theta_0 = \eta_0 = 0.5$ and $n_0 = n_1 = 150$. Note that all the assumptions are satisfied for $\hat{\theta}$. However in the evaluation of $\hat{\theta}_L$, the covariate model for T_1 is correct but their assumption about common dependence structures for the two groups or the extra assumption on a covariate model for T_2 may be mis-specified. In some settings, α_z or τ_z may be different for $z = 0, 1$. In the first four cases of Table 3-1, we generated the location-shift model with $h_1(t) = h_2(t) = t$, $-\theta_0+\varepsilon \sim \exp(0.8)$, $-\eta_0+\xi \sim \exp(1)$, $C|Z = 1 \sim U(0, 6)$, and $C|Z = 0 \sim U(0.5, 6.5)$ but the copula dependence structures for the two groups may vary. We will use the notation {Clayton(τ_0), Frank(τ_1)} to denote the situation that one group with $Z = 0$ follows the Clayton model with $\tau = \tau_0$ and the other with $Z = 1$ follows the Frank model with $\tau = \tau_1$. The dependence structures for the first four cases are *case 1:* {Clayton(0.5), Clayton (0.5)}, *case 2:* {Clayton(0.8), Clayton(0.1)}, *case 3:* {Frank(0.5), Clayton(0.5)} and *case 4:* {Frank(0.8), Clayton(0.1)}. In case 1 where the conditions for both estimators are valid, $\hat{\theta}_L$ slightly outperforms $\hat{\theta}$. However in the last three cases, $\hat{\theta}_L$ is biased. It seems that the bias of $\hat{\theta}_L$ is affected more by the discrepancy in the level of associations for the two groups than the difference in the dependence structures. The dependence structures in cases 5-8 of Table 3-2 follow the same patterns as in cases 1-4 of Table 3-1. Here we set $h_1(t) = t$ but $h_2(t) = \log(t)$, $-\theta_0+\varepsilon \sim \exp(0.8)$, $\exp(-\eta_0)\exp(\xi) \sim \exp(1)$, $C|Z = 1 \sim U(0, 6)$ and $C|Z = 0 \sim U(0.5, 6.5)$. Note that $h_2(t) \neq h_1(t)$ which is a condition that violates the assumption made by Lin et al. (1996). We see that $\hat{\theta}$ outperforms $\hat{\theta}_L$ even more since, for the latter, the two types of assumptions are both mis-specified.

Model	$\hat{\theta}$	$\hat{\theta}_L$
case 1:	-0.0026 (0.0934)	-0.0025 (0.0909)
case 2:	-0.0013 (0.1136)	0.0969 (0.0849)
case 3:	0.0022 (0.0950)	-0.0122 (0.0888)
case 4:	0.0008 (0.1100)	0.0982 (0.0840)

Table 3-1: Finite sample performance of two estimators evaluated under 4 situations: the correlation structures are the same for two covariate groups in the first case and different in the last three cases. The first number is the average bias of the estimator and the number in the parenthesis is the standard deviation based on 1000 replications.

Model	$\hat{\theta}$	$\hat{\theta}_L$
case 5:	-0.0041 (0.0974)	0.0890 (0.1175)
case 6:	-0.0067 (0.1135)	0.3387 (0.1127)
case 7:	0.0025 (0.1156)	0.0884 (0.1170)
case 8:	0.0125 (0.1152)	0.3793 (0.1081)

Table 3-2: Finite sample performance of two estimators evaluated under 4 situations with different covariate models for progression time and death time (thus invalides $\hat{\theta}_L$). The first number is the average bias of the estimator and the number in the parenthesis is the standard deviation based on 1000 replications.

The second analysis checks the validity of the proposed method for selecting an appropriate copula model. We generated $\{T_{1i}, T_{2i}, C_i\}$ $(i = 1, \cdots, 150)$, where $T_{1i} \sim \exp(0.8)$, $T_{2i} \sim \exp(1)$, and $C_i \sim U(0, 6)$ and $(T_{1i}, T_{2i}) \sim$ copula model $(\tau = 0.5)$. There are two copula models under comparison where model $k = 1$ is the Clayton model and model $k = 2$ is the Frank model. First we set the Clayton model as the true one. The mean and standard deviation (in parentheses) of D^1 and D^2 are 0.0780 (0.0187) and 0.1397 (0.0304) based on 1000 replications. The percentages of successfully selecting the Clayton model are 93.4% based on the order of D^j $(j = 1, 2)$. Then we set the Frank model as the true one. The mean and standard deviation (in parentheses) of D^1 and D^2 are 0.1398 (0.0330), 0.0819 (0.0206). The percentages of successfully selecting the Frank model are 92.3% based on the order of D^j $(j = 1, 2)$. Finally, we examine the proposed testing procedure using the re-sampling method. Under the Clayton model, we set up the goodness-of-fit test: H_0 : the data follows the Clayton model versus H_a : the data does not follow the Clayton model. By re-sampling 1000 times, we obtained $D^1 = 0.0511$ with p-value=0.909 and the cut-off value: $c_1 = 0.1004$ (at 0.05 significance level). Hence H_0 is accepted which is a correct decision. For the same data set, we run the analysis again with H_0 : the data follows the Frank model versus H_a : the data does not follow the Frank model. We obtained that $D^2 = 0.1247$ with p-value=0.012; the cut-off value $(\gamma = 0.05)$: $c_2 = 0.1058$. Accordingly we reject H_0 which is also a correct decision.

The purpose of the third analysis is to examine the proposed method for selecting an appropriate regression model. We generate data for $Z = 0$ or $Z = 1$ with equal sample sizes in each group. First we generated data according to margins $T_1|Z = 0 \sim \exp(0.8)$, $T_2|Z = 0 \sim \exp(1)$, $C|Z = 0 \sim U(0, 6)$, $T_1|Z = 1 \sim \exp(0.8)+\theta_0$, $T_2|Z = 1 \sim \exp(1)+\theta_0$, $C|Z = 1 \sim U(0, 6) + \theta_0$. The correlation structure follows the Clayton model with $\tau_0 = 0.5$, $\tau_1 = 0.6$. There are four regression models under consideration: the location shift model (LS), the accelerated failure time model (AFT), the Cox proportional hazard model (PH) and the proportional odds model (PO). Table 3-3 lists the proportions of each model being selected by the proposed method based on 500 simulation runs. The results shown that the correct model (LS) is chosen most of the times (96.2% when $n = 100$, 99.6% when $n = 200$ and 100% when $n = 400$). Secondly, we generated data according to margins $T_1|Z = 0 \sim \exp(0.8)$, $T_2|Z = 0 \sim \exp(1)$, $C|Z = 0 \sim U(0, 6)$, $T_1|Z = 1 \sim \exp(-\theta_0)\exp(0.8)$, $T_2|Z = 1 \sim \exp(-\theta_0)\exp(1)$, $C|Z = 1 \sim U(0, 6)$. The correlation structure again follows the Clayton model with $\tau_0 = 0.5$, $\tau_1 = 0.6$. Note that in this case both AFT and PH models are correct. Together, these two models are chosen most of the times. As the sample size increases, the proportion of a correct decision also increases (79% when $n = 100$, 82.8% when $n = 200$ and 92% when $n = 400$).

		Chosen model			
True	n	LS	AFT	PH	PO
LS	100	96.2%	3.8%	0%	0%
	200	99.6%	0.4%	0%	0%
	400	100%	0%	0%	0%
AFT and PH	100	0.8%	43.4%	35.6%	20.2%
	200	0.2%	39%	43.8%	17%
	400	0%	47.8%	44.2%	8%

Table 3-3: Proportion of the covariate models selected by the proposed method based on 500 replications. The first column lists the true covariate model; the second column lists the sample size; the last four columns contain the proportion of each of the four covariate models selected.

In the fourth analysis, we examine the finite-sample performance of $\hat{\theta}$ when Z contains multiple covariates. Under the model $h(T_1) = -Z'\theta + \varepsilon$, where $\theta' = (\theta_1, \theta_2)$, $Z' = (Z^{(1)}, Z^{(2)})$, $Z^{(1)}$=0 or 1, $Z^{(2)}$=0 or 1, we can partition the sample into four groups with $Z_1' = (0,0)$, $Z_2' = (0,1)$, $Z_3' = (1,0)$ and $Z_4' = (1,1)$. The sample sizes in the four groups are 75 with $\tau_{(0,0)} = 0.2$, $\tau_{(0,1)} = 0.3$, $\tau_{(1,0)} = 0.4$ and $\tau_{(1,1)} = 0.5$. The Clayton and Frank models are evaluated. Four regression models, namely LS, AFT, PH and PO are considered. The true parameter values are set to be $\theta_0' = (0.3, 0.3)$. The marginal distributions in the group $Z_1' = (0,0)$ follow $T_1 \sim \exp(0.8)$, $T_2 \sim \exp(1)$. The censoring distribution is $C \sim U(0,6)$. The average bias and the standard deviation based on 1000 simulation runs are reported in Table 3-4. The results show that the proposed method still performs well under the more general regression setting.

Model	LS	AFT	PH	PO
Clayton	-0.0024 (0.1113)	0.0029 (0.1736)	-0.0022 (0.1507)	0.0039 (0.2683)
	-0.0015 (0.1105)	0.0058 (0.1662)	-0.0032 (0.1514)	-0.0023 (0.2633)
Frank	-0.0042 (0.1016)	-0.0084 (0.1734)	0.0067 (0.1544)	-0.0028 (0.2573)
	0.0011 (0.0995)	-0.0094 (0.1661)	-0.0096 (0.1680)	0.0013 (0.2602)

Table 3-4: Finite sample performance of $\hat{\theta}'$. The first number is the average bias of $\hat{\theta}_1$, the second number in the parenthesis is the standard deviation of $\hat{\theta}_1$ based on 1000 replications; The third number is the average bias of $\hat{\theta}_2$ and the fourth number in the parenthesis is the standard deviation of $\hat{\theta}_2$ based on 1000 replications.

3.4.3.2. Real Data

The proposed methodology is applied to the bone marrow transplants data given in Klein and Moeschberger (2003, p.484). There were 137 leukemia patients receiving bone marrow transplants. Let T_1 be the time to relapse of leukemia, T_2 be the time to death and C be the time from transplant to the end of study. Let $\delta_1 = I(T_1 \leq T_2 \wedge C)$ be the relapse indicator and let $\delta_2 = I(T_2 \leq C)$ be the death indicator. The sample can be divided into three groups with $Z' = (0,0)$ indicating the AML low-risk group,

$Z' = (0, 1)$ indicating the ALL group and $Z' = (1, 0)$ indicating the AML high-risk group. The regression model of interest is $h(T_1) = -Z'\theta + \varepsilon$, where $\theta' = (\theta_1, \theta_2)$ which measures whether the disease type affects the relapse time.

For each covariate group, we test the hypothesis $H_0 : \phi_\alpha \sim$ the Clayton model versus H_a : not H_0. By bootstrapping 1000 times, the p-values of D^C for the AML high-risk group, the ALL group and the AML low-risk group are 0.752, 0.656 and 0.177, respectively. Hence the Clayton model is adopted for all the three groups. Using Day's method (or equivalently Wang's method) to estimate τ_z, we obtain $\hat{\tau}_{(0,0)}$=0.7485 (0.1176),$\hat{\tau}_{(0,1)}$=0.7894 (0.0853) and $\hat{\tau}_{(1,0)}$=0.7685 (0.0872), where the number in parentheses is the estimated standard derivation using the jackknife method. The above analysis implies that the dependence structures in the three groups are similar and the two events are highly correlated.

Then we choose a model for measuring the group effect on T_1. Figure 3-3 shows the fitted log-log plot of $\hat{F}_1(x)$ for the three groups. Since the three curves look parallel, we choose the proportional hazard (PH) model to measure the group effect. Based on the method described in Section 3.4.2.2, we can formally test the PH model assumption. By bootstrapping 1000 times, we obtain p-value=0.774 which implies that this model is appropriate. Figure 3-4 depicts the three survival curves of $\hat{F}_1(x)$. Under the PH regression model and the Clayton assumption for each covariate group, we obtain $\hat{\theta}_1$=1.3624 (0.3765) and $\hat{\theta}_2$=0.9503 (0.3984). The results show that the risk of relapse for the AML high-risk group is 3.9 times to the risk for the AML low-risk and the risk for the ALL group is 2.59 times to that for the AML low-risk group. The difference is statistically significant.

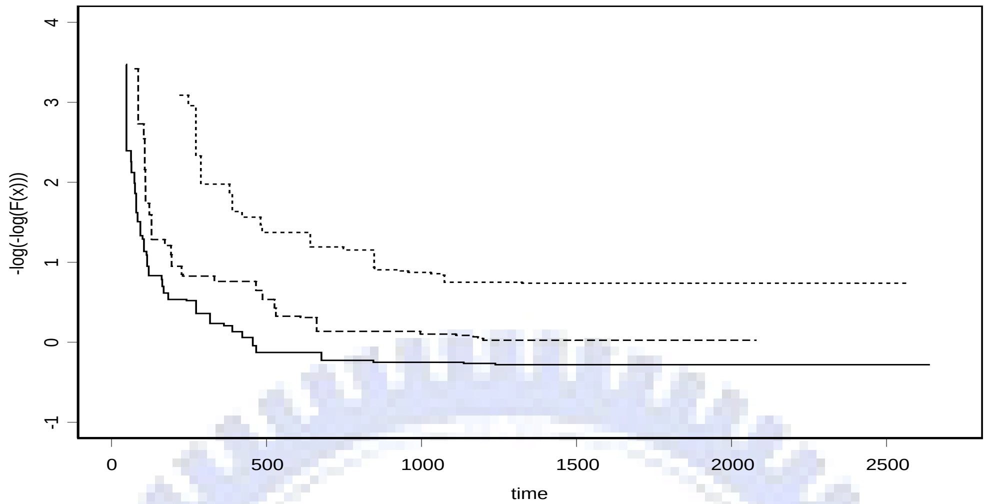

Figure 3-3: Log-Log plot for the three groups. Solid line: AML high risk group; dashed line: ALL group; dotted line: AML low risk group.

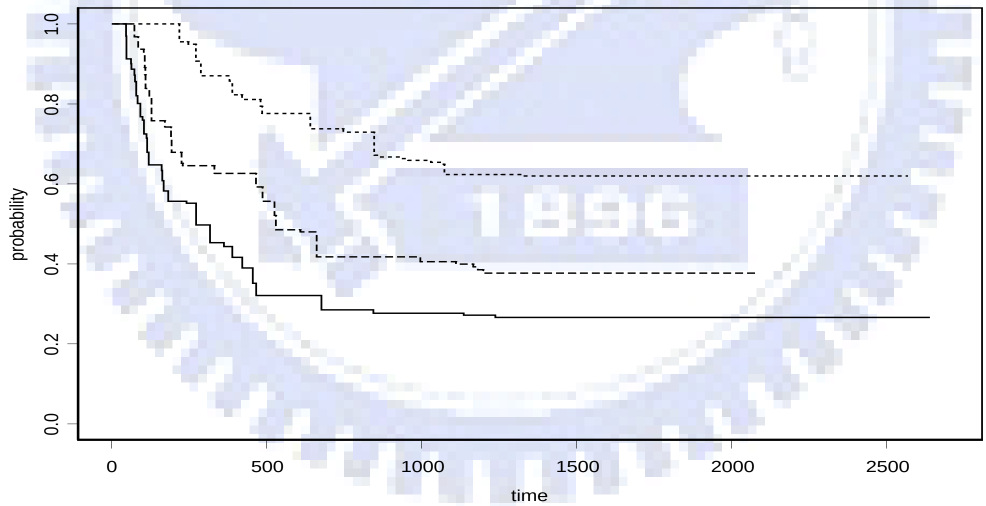

Figure 3-4: $\hat{F}_1(x)$ *for the three groups. solid line: AML high risk group; dashed line: ALL group; dotted line: AML low risk group.*

Chapter 4. Regression Analysis for Association Based on Three Types of Data

4.1. Preliminary

The second part of the thesis studies the effect of covariates on the level of association between T_1 and T_2 which may follow one type of the three data structures discussed in Section 2.2. In general, we may describe the association in terms of global association or local dependence. For measuring global association, the Pearson's correlation coefficient and Kendall's tau are more well-known. In particular, the rank invariance property of Kendall's τ makes it very useful for describing the relationship between lifetime variables which are often skewed. For describing local association, the function $\theta(s,t)$ in (5) proposed by Oakes (1989) is often used in applications of survival data.

With covariates, most existing methods focus on studying the covariates effects on the marginal distributions. We have studied one of such application in Chapter 3. Now, we consider another application which investigates how covariates affect the degree of association. Our major goal is to develop a unified inference approach which can handle the three types of data structures discussed in Section 2.2. Here is the outline of our discussions. In Section 4.2, we review the paper on association under Clayton assumption with time transformations (Fine et al., 2000). In Section 4.3, we review a related paper on testing constancy of association across different covariate strata (Ghosh, 2006). Then we present our proposed inference procedure in Section 4.4.

4.2. Association in a Copula with Time Transformations

Fine & Jiang (2000) considered estimation of the cross ratio in Clayton's copula in which covariates are incorporated into the marginal distributions via semi-parametric accelerated life regression models. Let T_1 and T_2 be survival times with absolutely continuous joint distribution. Under Clayton model, the predictive hazard is constant, i.e.

$$\frac{\lambda_1(s|T_2 = t)}{\lambda_1(s|T_2 > t)} = \frac{\lambda_2(t|T_1 = s)}{\lambda_2(t|T_1 > s)} = \alpha, \tag{23}$$

where $\lambda_j(t|A)$ is the hazard of T_j at time t given that event A has occurred. (23) holds if and only if T_1 and T_2 have joint survival function

$$\Pr(T_1 > s, T_2 > t) = \{F_1(s)^{1-\alpha} + F_2(t)^{1-\alpha} - 1\}^{1/(1-\alpha)}, \tag{24}$$

where $F_1(s) = \Pr(T_1 > s)$ and $F_2(t) = \Pr(T_2 > t)$. The distribution (24) originates in a gamma frailty model (Hougaard, 1986). This is

$$\Pr(T_1 > s, T_2 > t) = \int_0^\infty \{F_1(s)F_2(t)\}^\omega f(\omega)d\omega,$$

where $f(\omega) \propto \exp[-\omega + \{(1-\alpha)^{-1} - 1\}\log(\omega)]$.

With covariates, a bivariate marginal effect model is considered which can be written as $h_1(T_1) = \beta_1^T Z_1 + \varepsilon_1$ and $h_2(T_2) = \beta_2^T Z_2 + \varepsilon_2$, where (β_1, β_2) and (Z_1, Z_2) are regression parameters and covariate vectors. $h_k(.)$ $(k = 1, 2)$ are known monotone functions. At the true h_k and β_k, let $T_k^* = \Psi_k(T_k, Z_k)$, where $\Psi_k(x, z) = h_k(x) - \beta_k^T z$ $(k = 1, 2)$ which is a known function, monotone in x for fixed z, and

$$\Pr(T_k^* \le u|Z_k) = \Pr(T_k^* \le u) \quad (k = 1, 2).$$

This structure is convenient when analyzing correlation between accelerated lifetimes. Consider the pairs

$$\{T_{11}^* = \Psi_1(T_{11}, Z_{11}), T_{21}^* = \Psi_2(T_{21}, Z_{21})\},\ \{T_{12}^* = \Psi_1(T_{12}, Z_{12}), T_{22}^* = \Psi_2(T_{22}, Z_{22})\},$$

which are independent and satisfy (24) conditional on covariates. It can be shown that

$$\Pr\{(T_{11}^* - T_{12}^*)(T_{21}^* - T_{22}^*) > 0|Z_{11}, Z_{12}, Z_{21}, Z_{22}\} = \alpha(1+\alpha)^{-1},$$

but the probability for (T_{11}, T_{21}) and (T_{12}, T_{22}) may not be the same.

Their ideas can be applied to multivariate failure times data and clustered failure times data. Here, we consider the simple case, bivariate failure times data. With censoring, the data consist of n replications of $(X, Y, \delta_1, \delta_2, \tilde{Z})$, where $X = T_1 \wedge C_1$, $Y = T_2 \wedge C_2$, $\delta_1 = I(T_1 \leq C_1)$, $\delta_2 = I(T_2 \leq C_2)$, and $\tilde{Z} = (Z_1, Z_2)$ is a $p \times 2$ matrix of bounded covariate vectors. Assume (C_1, C_2) is independent of (T_1, T_2) conditional on $\tilde{Z}$ and denote the observations by $(X_i, Y_i, \delta_{1i}, \delta_{2i}, \tilde{Z}_i)$ $(i = 1, 2, ..., n)$. Marginally, the failure times satisfy linear regression models. That is,

$$h_j(T_{ji}) = \beta_j^T Z_{ji} + \varepsilon_{ji},\ (j = 1, 2), \tag{25}$$

where h_j is a known function, β_j is an unknown $p \times 1$ parameter vector, and ε_{ji} $(i = 1, 2, ..., n)$ are independent and identically distributed with unknown $\Pr(\varepsilon_{ji} > x) = F_j(x)$ $(j = 1, 2)$. Fine et al. (2000) assumed that $\tilde{\varepsilon}_i = (\varepsilon_{1i}, \varepsilon_{2i})$, for $i = 1, ..., n$, have a independent and common, but completely unspecified, joint distribution. The pairwise model for ε_{1i} and ε_{2i} satisfies (24) with parameter α. One can estimate the parameter β_j in (25) using the methodology proposed by Lin & Wei (1992). Let $e_{1i}(\beta) = h_1(X_i) - \beta^T Z_{1i}$ and $e_{2i}(\beta) = h_2(Y_i) - \beta^T Z_{2i}$ $(i = 1, 2, ..., n)$. A rank-based estimating function for β_j $(j = 1, 2)$ is given by

$$U_j(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \hat{w}_j(\beta, u)[Z_{ji} - S_{j1}(\beta, u)\{S_{j0}(\beta, u)\}^{-1}]dN_{ji}(\beta, u),$$

where $\hat{w}_j(\beta, u)$ is a weight function, $N_{ji}(\beta, u) = I\{e_{ji}(\beta) \leq u, \delta_{ji} = 1\}$, $S_{jl}(\beta, u) = n^{-1} \sum_{i=1}^{n} I\{e_{ji}(\beta) \geq u\} Z_{ji}^{\otimes l}$, and for a vector v, $v^{\otimes 0} = 1$ and $v^{\otimes 1} = v$. When $\hat{w}_j(\beta, u) = S_{j0}(\beta, u)$, $U_j(\beta)$ is Gehan's statistic, which is monotone in each of the components of β. A popular class of weights is $\hat{w}_j(\beta, u) = w\{\hat{F}_j(\beta, u)\}$, where $w(.)$ is a twice continuously differentiable function on $[0, 1]$ and $\hat{F}_j(\beta, u)$ is the left-continuous version of the Kaplan-Meier estimator based on the pairs $\{e_{ji}(\beta), \delta_{ji}, i = 1, ..., n\}$ (Wei, Ying & Lin, 1990).

Next, consider the estimation of α. The term $\Delta_{lm}(\beta_1, \beta_2)$ indexes whether the pairs $\{h_1(T_{1l}) - \beta_1^T Z_{1l}, h_2(T_{2l}) - \beta_2^T Z_{2l}\}$ and $\{h_1(T_{1m}) - \beta_1^T Z_{1m}, h_2(T_{2m}) - \beta_2^T Z_{2m}\}$ are concordant or discordant. Formally, $\Delta_{lm}(\beta_1, \beta_2)$ equals

$$I[\{h_1(T_{1l}) - h_1(T_{1m}) - \beta_1^T(Z_{1l} - Z_{1m})\}\{h_2(T_{2l}) - h_2(T_{2m}) - \beta_2^T(Z_{2l} - Z_{2m})\} > 0].$$

Conditional on covariates, $\Delta_{lm}(\beta_{10}, \beta_{20})$ has mean $\alpha_0(1+\alpha_0)^{-1}$, where α_0 is the true value of α and (β_{10}, β_{20}) are the true value of (β_1, β_2). With censoring, $\Delta_{lm}(\beta_1, \beta_2)$ can be determined when

$$\min\{h_1(T_{1l}) - \beta_1^T Z_{1l}, h_1(T_{1m}) - \beta_1^T Z_{1m}\} \quad < \quad \min\{h_1(C_{1l}) - \beta_1^T Z_{1l}, h_1(C_{1m}) - \beta_1^T Z_{1m}\},$$

$$\min\{h_2(T_{2l}) - \beta_2^T Z_{2l}, h_2(T_{2m}) - \beta_2^T Z_{2m}\} \quad < \quad \min\{h_2(C_{2l}) - \beta_2^T Z_{2l}, h_2(C_{2m}) - \beta_2^T Z_{2m}\}.$$

Let $D_{lm}(\beta_1, \beta_2)$ equal 1 when this occurs and 0 otherwise. With $\beta_1 = \beta_{10}$ and $\beta_2 = \beta_{20}$, it is natural to estimate α by the ratio of the numbers of concordant pairs to discordant pairs, among all pairs where concordance status is determinable (Oakes, 1982). An estimating function can be constructed as follows. Suppose $W(u, v)$ is random positive function converging uniformly over u and v to a deterministic limit. Assume that the limit is finite over the support of $E_{1lm}(\beta_{10})$ and $E_{2lm}(\beta_{20})$, where $E_{jlm}(\beta) = \min\{e_{jl}(\beta), e_{jm}(\beta)\}$, for $j = 1, 2$. Define $\alpha(\beta_1, \beta_2)$ as the solution to $U(\beta_1, \beta_2, \alpha) = 0$, where

$$U(\beta_1, \beta_2, \alpha) = \sum_{l<m} W\{E_{1lm}(\beta_1), E_{2lm}(\beta_2)\} D_{lm}(\beta_1, \beta_2)\{\Delta_{lm}(\beta_1, \beta_2) - \alpha(1+\alpha)^{-1}\}.$$

The profile estimator $\alpha(\beta_1, \beta_2)$ has the following closed form expression:

$$\frac{\sum_{l<m} W\{E_{1lm}(\beta_1), E_{2lm}(\beta_2)\} D_{lm}(\beta_1, \beta_2) \Delta_{lm}(\beta_1, \beta_2)}{\sum_{l<m} W\{E_{1lm}(\beta_1), E_{2lm}(\beta_2)\} D_{lm}(\beta_1, \beta_2)\{1 - \Delta_{lm}(\beta_1, \beta_2)\}}.$$

A useful weight function is

$$W_{a,b}^{-1}(x, y) = n^{-1} \sum_{i=1}^{n} I\{e_{1i}(\hat{\beta}_1) \geq \min(a, x), e_{2i}(\hat{\beta}_2) \geq \min(b, y)\},$$

where a and b are constants. In addition, the authors showed that $(\hat{\beta}_1, \hat{\beta}_2, \hat{\alpha})$ are consistent and asymptotic normal.

4.3. Testing Constancy of Association across Strata

Ghosh (2006) considered semi-competing risks data with covariates: $\{(X_i, \delta_{1i}, Y_i, \delta_{2i}, Z_i)$ $(i = 1, ..., n)\}$, where $X_i = T_{1i} \wedge T_{2i} \wedge C_i$, $Y_i = T_{2i} \wedge C_i$, $\delta_{1i} = I(T_{1i} \leq T_{2i} \wedge C_i)$, and $\delta_{2i} =$

$I(T_{2i} \le C_i)$. If the covariate takes discrete values, it is assumed that $\theta(s,t|Z=z) = \alpha_z$, where $\theta(s,t|Z=z) = \lambda_1(s|T_2=t, Z=z)/\lambda_1(s|T_2>t, Z=z)$ and $\lambda_1(s|A, Z=z)$ is the hazard function of T_1 at time s for a subject with $Z=z$ given that event A also occurs. Thus, a Clayton-Oakes frailty model is assumed for each stratum defined by Z. The interest is in testing the null hypothesis that the predictive hazard ratio does not depend on Z (i.e. $H_0 : \alpha_z = \alpha$).

For testing the null hypothesis $H_0 : \theta(s,t|Z) = \alpha$, Ghosh (2006) proposed using the following class of test statistics

$$U_1 = \sum_{z=1}^{K} \sum_{i<j} W_z(\tilde{X}_{ij}, \tilde{Y}_{ij})(D_{ijz} - D_{ij})\Delta_{ij}, \tag{26}$$

where $\Delta_{ij} = I\{(X_i - X_j)(Y_i - Y_j) > 0\}$, $\tilde{X}_{ij} = X_i \wedge X_j$, $\tilde{Y}_{ij} = Y_i \wedge T_j$, $\tilde{C}_{ij} = C_i \wedge C_j$, $D_{ij} = I(\tilde{X}_{ij} < \tilde{Y}_{ij} < \tilde{C}_{ij})$, $D_{ijz} = I(\tilde{X}_{ij} < \tilde{Y}_{ij} < \tilde{C}_{ij}, Z_i = Z_j = z)$, and $W_z(u,v)$ is a weight function that converges uniformly to $w(u,v)$, a bounded deterministic function. Note that the statistics U_1 provides a measure of the difference between stratified and unstratified analysis. Ghosh proved that under the null hypothesis, $n^{-3/2}U_1$ has a limit normal distribution with mean zero.

However, it is not easy to derive an analytic expression for the variance of the limiting distribution of U_1. Ghosh suggested a re-sampling method originally proposed by Parzen, Wei & Ying (1994) for variance estimation. Equation (26) can be written as the following form:

$$U_1 = \sum_{z=1}^{K} \sum_{i<j} U_{ijz}, \tag{27}$$

where $U_{ijz} = W_z(\tilde{X}_{ij}, \tilde{Y}_{ij})(D_{ijz} - D_{ij})\Delta_{ij}$. The first step is to generate n standard normal random variables $(G_1, ..., G_n)$ and calculate the perturbations of (27) as follows

$$U_1^* = \sum_{z=1}^{K} \sum_{i<j} U_{ijz} G_i G_j. \tag{28}$$

Under the null hypothesis, $n^{-3/2}U_1^*$ and $n^{-3/2}U_1$ have the same limiting distribution. The algorithm can be summarized as follows:

1. *Generate n i.i.d.* $N(0,1)$ *random variables* $(G_1, ..., G_n)$ *and calculate* U_1^*.

2. *Repeat step 1 M times.*

There are two ways of constructing a 95% confidence interval for the limit of $n^{-3/2}U_1$. The first is to calculate the standard error based on empirical replications of U_1^*. The other way is to take 2.5th and 97.5th percentiles of the empirical distribution of U_1^*.

4.4. The Proposed Inference Procedure for the Association Model

We aim to develop a regression model that describes the effect of covariate on the dependence structure. We also want to propose a unified inference approach which can handle the three data structures discussed in Section 2.2.

To achieve the second objective, we need to find an appropriate dependence measure for each data structure. In Section 4.4.1, a flexible way of model formulation is presented. In Section 4.4.2, we describe the proposed regression model and in Section 4.4.3, we include external censoring in the three data structures. The proposed inference method is discussed in Section 4.4.4. In Section 4.4.5, we present a model checking method for Clayton assumption. We also present the numerical analysis for the proposed inference methods in Section 4.4.6.

4.4.1. Model Formulation

Most methods developed for typical bivariate survival data analyze the joint survival function

$$F_a(s,t) = \Pr(T_1 > s, T_2 > t),$$

which seems to be a straightforward extension from the univariate analysis. Mathematically the joint behavior between T_1 and T_2 can also be described other functions such as

$$\begin{aligned}
F_b(s,t) &= \Pr(T_1 \leq s, T_2 > t) \\
F_c(s,t) &= \Pr(T_1 \leq s, T_2 \leq t) \\
F_d(s,t) &= \Pr(T_1 > s, T_2 \leq t).
\end{aligned}$$

Choosing an appropriate function for further analysis depends not only on its interpretation but also the mathematical applicability. Note that $\theta(s,t)$ in (5) is defined based on the joint survival function $F_a(s,t)$. Now we denote $\theta(s,t) = \theta_a(s,t)$. Oakes (1989) derived another expression of $\theta_a(s,t)$ which is useful for further extensions and statistical inference. Specifically one can write

$$\theta_a(s,t) = \frac{\Pr(\Delta_{ij} = 1 | \tilde{T}_{1,ij} = s, \tilde{T}_{2,ij} = t)}{\Pr(\Delta_{ij} = 0 | \tilde{T}_{1,ij} = s, \tilde{T}_{2,ij} = t)},$$

where $\Delta_{ij} = I[(T_{1i} - T_{1j})(T_{2i} - T_{2j}) > 0]$ and $\tilde{T}_{k,ij} = T_{ki} \wedge T_{kj}$ $(k = 1,2)$. Based on this representation, $\theta_a(s,t)$ can be viewed as the odds ratio of concordance given the corner value $(\tilde{T}_{1,ij}, \tilde{T}_{2,ij}) = (s,t)$.

We can extend the ideas of $\theta_a(s,t)$ as follows. The two pairs, (T_{1i}, T_{2i}) and (T_{1j}, T_{2j}), can form four grid points (a,b,c,d) shown in Figure 4-1 with $a = (\tilde{T}_{1,ij}, \tilde{T}_{2,ij})$, $b = (\breve{T}_{1,ij}, \tilde{T}_{2,ij})$, $c = (\breve{T}_{1,ij}, \breve{T}_{2,ij})$, and $d = (\tilde{T}_{1,ij}, \breve{T}_{2,ij})$, where $\tilde{T}_{k,ij} = T_{ki} \wedge T_{kj}$ and $\breve{T}_{k,ij} = T_{ki} \vee T_{kj}$ $(k = 1,2)$. Selecting a different corner point gives a new odds ratio of concordance. Define

$$\theta_*(s,t) = \frac{\Pr(\Delta_{ij} = 1 | \text{corner} = *)}{\Pr(\Delta_{ij} = 0 | \text{corner} = *)} \tag{29}$$

$$= \frac{F_*(s,t) D_s D_t F_*(s,t)}{\{D_s F_*(s,t)\}\{D_t F_*(s,t)\}} \quad \text{if } * = a, c$$

$$= \frac{\{D_s F_*(s,t)\}\{D_t F_*(s,t)\}}{F_*(s,t) D_s D_t F_*(s,t)} \quad \text{if } * = b, d. \tag{30}$$

When T_1 and T_2 are independent, $\theta_*(s,t) = 1$. In general $\theta_*(s,t)$ describes local dependence at (s,t) such that $\theta_*(s,t) > 1$ indicates positive association and $\theta_*(s,t) < 1$ indicates negative association. From the above derivations, we see that imposing a structure on a version of $\theta_*(s,t)$ is associated with model specification on the corresponding joint function $F_*(s,t)$.

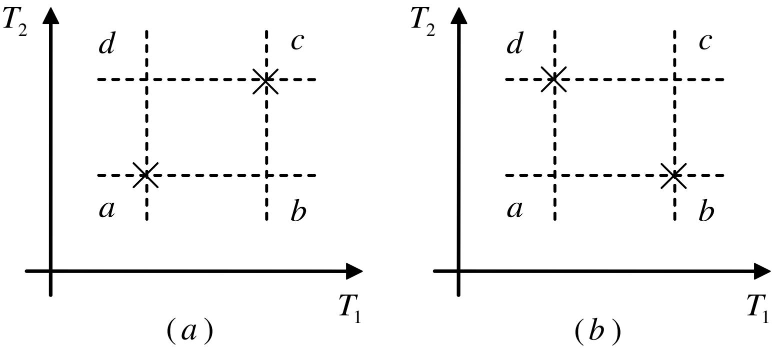

Figure 4-1: Four grid points formed by two pairs of observation.
(a): (i, j) pairs are concordant; (b): (i, j) pairs are discordant.

The two representations of $\theta_*(s,t)$ in (29) and (30) provide useful insight for further statistical inference. Specifically the first identity is equivalent to

$$\Pr(\Delta_{ij} = 1 | \text{corner} * = (s,t)) = \frac{\theta_*(s,t)}{1+\theta_*(s,t)} \quad (* = \text{a, b, c, d}).$$

This implies that the inference of $\theta_*(s,t)$ can be made by applying the method of moment based on data replications of Δ_{ij}. This approach has been adopted by Fine (2001) based on $\theta_a(s,t)$ for semi-competing risks data and by Chaieb et al. (2006) based on $\theta_b(s,t)$ for truncation data. The second identity of $\theta_*(s,t)$ suggests that one can construct the log-rank type of statistics based on a series of two-by-two tables. Figure 4-2 shows four versions of the table construction. This approach has been taken by Day et al. (1997) and Wang (2003) for analyzing semi-competing risks data based on the table in *Figure 4-2(a)* and by Emura, Wang and Hung (2006) based on *Figure 4-2(b)* for analyzing truncation data.

To choose an appropriate version of $\theta_*(s,t)$, one should examine whether the biological meaning is reasonable as well as whether the data provide enough information for the purpose of inference. As mentioned earlier, most existing methods specify the model assumption based on $F_a(s,t)$ which has a direct relationship with $\theta_a(s,t)$. For the first

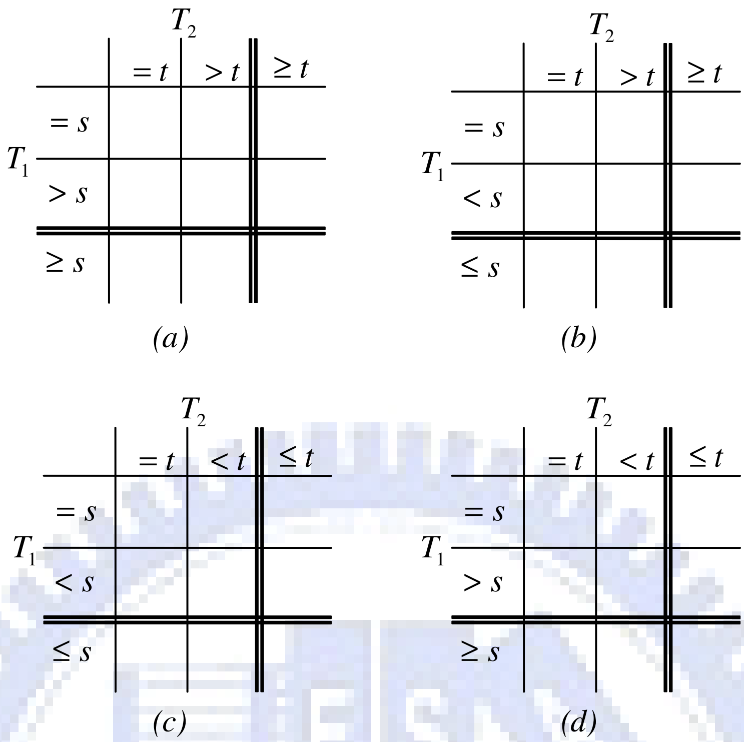

Figure 4-2: Four versions of table construction.

data structure, it seems that all four versions are mathematically suitable and corner a is usually adopted due to its straightforward interpretation. For semi-competing risks data, $\theta_b(s,t)$ is an obvious suitable candidate. However $\theta_a(s,t)$ is still appropriate since, from the second identity, the information provided by $(T_1 \wedge T_2, T_2, I(T_1 \leq T_2))$ is enough for recovering every component of $\theta_a(s,t)$ for $(s,t) \in R_2$. However for truncation data, only $\theta_b(s,t)$ is suitable. Note that for modeling dependent truncation data, Chaieb et al (2006) proposes a related measure

$$\check{\theta}(s,t) = \frac{F_b(s,t)D_sD_tF_b(s,t)}{\{D_sF_b(s,t)\}\{D_tF_b(s,t)\}} = 1/\theta_b(s,t).$$

Setting $\theta_b(s,t) = \alpha$ is equivalent to assuming the Clayton copula on $F_b(s,t)$:

$$\Pr(T_1 \leq s, T_2 > t) = \{\Pr(T_1 \leq s)^{1-\alpha} + \Pr(T_2 > t)^{1-\alpha} - 1\}^{1/(1-\alpha)}.$$

4.4.2. The Proposed Regression Model

Consider the situation that the levels of association between (T_1, T_2) vary in different covariate groups. Let $Z = (1, Z_1, ..., Z_p)^T$ be a vector of common discrete covariates for (T_1, T_2). We assume that for a chosen corner point $*$ $(* = a, b, c, d)$,

$$\begin{aligned} \theta_*^Z(s,t) &= \frac{\Pr(\Delta_{ij} = 1|\text{corner} * = (s,t), Z_i = Z_j = Z)}{\Pr(\Delta_{ij} = 0|\text{corner} * = (s,t), Z_i = Z_j = Z)} \\ &= \exp(Z^T\beta), \end{aligned} \tag{31}$$

where $\beta = (\beta_0, \beta_1, ..., \beta_p)^T$. Equivalently the above model assumes that

$$\Pr(\Delta_{ij} = 1|\text{corner} * = (s,t), Z_i = Z_j = Z) = \frac{\exp(Z^T\beta)}{1 + \exp(Z^T\beta)} \equiv \eta(Z^T\beta). \tag{32}$$

Note that β_0 is the log odds of concordance for the baseline group with $Z_1 = Z_2 = ... = Z_p = 0$. The slope parameter β_k $(k = 1, 2, ..., p)$ can be viewed as the difference of log odds by increasing one unit of Z_k with the rest of $Z_i's$ being fixed. The above model assumption is equivalent to assuming that $(T_1, T_2)|Z$ follows Clayton's model with

$$F_*(s,t|Z) = \{F_{1,*}(s|Z)^{1-\exp(Z^T\beta)} + F_{2,*}(t|Z)^{1-\exp(Z^T\beta)} - 1\}^{1/(1-\exp(Z^T\beta))}, \tag{33}$$

where $F_{1,*}(s|Z) = \Pr(T_1 > s|Z)$ for $* = a, d$; $F_{1,*}(s|Z) = \Pr(T_1 \le s|Z)$ for $* = b, c$, $F_{2,*}(t|Z) = \Pr(T_2 > t|Z)$ for $* = a, b$; $F_{2,*}(t|Z) = \Pr(T_2 \le t|Z)$ for $* = c, d$. The main purpose is to estimate β. From previous discussions, we consider the model $\theta_a^Z(s,t) = \exp(Z^T\beta)$ for typical bivariate failure-time data and semi-competing risks data. For the truncation data, we consider the model $\theta_b^Z(s,t) = \exp(Z^T\beta)$.

4.4.3. Three Data Structures with External Censoring

Now we incorporate external censoring. To simplify the presentation, we may use the same notation with different definitions under different data structures. We also discuss the condition under which the value of Δ_{ij} is certain for pair (i, j).

Data Structure 1: Typical bivariate data subject to right censoring

Assume that (T_1, T_2) is subject to independent censoring by (C_1, C_2) such that observed variables become $X = T_1 \wedge C_1$, $Y = T_2 \wedge C_2$, $\delta_1 = I(T_1 \leq C_1)$, and $\delta_2 = I(T_2 \leq C_2)$. Based on observed variables (X_i, δ_{1i}) and (X_j, δ_{1j}), to know the order of T_{1i} and T_{1j}, the smaller variable has to be uncensored. Similar properties can be applied for determination of the order of T_{2i} and T_{2j} based on (Y_i, δ_{2i}) and (Y_j, δ_{2j}). Formally define $\tilde{T}_{k,ij} = T_{ki} \wedge T_{kj}$ and $\tilde{C}_{k,ij} = C_{ki} \wedge C_{kj}$ $(k = 1, 2)$. As long as $\tilde{T}_{1,ij} < \tilde{C}_{1,ij}$ and $\tilde{T}_{2,ij} < \tilde{C}_{2,ij}$, the value of Δ_{ij} is known for sure which means that the (i, j) pair is orderable on the plane.

Data Structure 2: Semi-competing risks data subject to censoring

If often happens that (T_1, T_2) are subject to a common external censoring variable C. Observed variables are denoted as $X = T_1 \wedge T_2 \wedge C$, $Y = T_2 \wedge C$, $\delta_1 = I(T_1 \leq T_2 \wedge C)$, and $\delta_2 = I(T_2 \leq C)$. Applying previous arguments, the order of T_{1i} and T_{1j} can be known as long as $\tilde{T}_{1,ij} < \tilde{T}_{2,ij}$ and $\tilde{T}_{1,ij} < \tilde{C}_{ij}$, where $\tilde{C}_{ij} = C_i \wedge C_j$. The order of T_{2i} and T_{2j} can be known as long as $\tilde{T}_{2,ij} < \tilde{C}_{ij}$. Combining both conditions, the orderable condition for semi-competing risks data can be defined as $\tilde{T}_{1,ij} < \tilde{T}_{2,ij} < \tilde{C}_{ij}$.

Data Structure 3: Truncation data subject to censoring

Recall that T_2 is subject to left-truncation by T_1 or T_1 is subject to right-truncation by T_2 so that (T_1, T_2) can be observed only if $T_1 < T_2$. Now we assume that T_2 is subject to right censoring by C. Hence, the observed variables become $X = T_1$, $Y = T_2 \wedge C$ and $\delta_2 = I(T_2 \leq C)$. We can set $\delta_1 = 1$ which means that T_1 is always uncensored. The order of T_{2i} and T_{2j} can be known as long as $\tilde{T}_{2,ij} < \tilde{C}_{ij}$.

In absence of covariates, observed variables can be denoted as $(X, Y, \delta_1, \delta_2)$ for the three data structures. For each data type, statisticians have developed inference procedures for investigating the dependent relationship between T_1 and T_2 based on a random sample of $(X, Y, \delta_1, \delta_2)$. There are two approaches which turn out to be applicable to all the three data structures. The first approach utilizes the moment condition of Δ_{ij} in (29). The second approach is developed via constructing a series of two-by-two tables based on equation (30). This paper extends the first approach to a regression setting in which the dimension of regression parameters may exceed 1 and hence the method of moment is not directly applicable.

4.4.4. The Proposed Inference Procedure

Let (T_{1i}, T_{2i}, Z_i) $(i = 1, 2, ..., n)$ be a random sample following the model assumption in (31) or its equivalent versions in (32) or (33). Note that (T_1, T_2) may be any of the three data types introduced earlier. In presence of external censoring, the observed data are denoted as $(X_i, Y_i, \delta_{1i}, \delta_{2i}, Z_i)$ $(i = 1, 2, ..., n)$ which are random replications of $(X, Y, \delta_1, \delta_2, Z)$ described in Section 4.3.3.

When the covariates are discrete, we can partition the sample according to distinct values of Z. For a pair of observations in each sub-sample, they need to satisfy two criterion in order to be used in the analysis. Specifically we select a pair (i, j) with $Z_i = Z_j = z$ such that the corresponding value of Δ_{ij} is known and the chosen corner value is located in the model region. For typical bivariate data and semi-competing risks data, we choose $* = a$. For truncation data, we set $* = b$. Since the first type of data falls in R_1, we don't have to impose any restriction. For semi-competing risks data, the restriction for making corner a to fall in R_2 is $\tilde{T}_{1,ij} \leq \tilde{T}_{2,ij}$. For truncation data, we should set $\breve{T}_{1,ij} < \tilde{T}_{2,ij}$ for making corner b to fall in R_2. Let $D_{ij}(z)$ be the orderable indicator that shows whether pair (i, j) with $Z_i = Z_j = z$ can be selected in the analysis. For the three types of data structure, $D_{ij}(z)$ is defined as follows. For typical bivariate data, $D_{ij}(z) = I(\tilde{T}_{1,ij} < \tilde{C}_{1,ij}, \tilde{T}_{2,ij} < \tilde{C}_{2,ij}, Z_i = Z_j = z)$; for semi-competing risks data, $D_{ij}(z) = I(\tilde{T}_{1,ij} < \tilde{T}_{2,ij} < \tilde{C}_{ij}, Z_i = Z_j = z)$ and, for truncation data, $D_{ij}(z) = I(\breve{T}_{1,ij} < \tilde{T}_{2,ij} < \tilde{C}_{ij}, Z_i = Z_j = z)$.

Now we discuss estimation of β. Since the dimension of β usually exceeds 1, we can not directly apply the method of moment based on equation (32) as in existing methods developed for homogeneous data (Fine et al., 2001; Chaieb et al., 2006). Instead, we apply the least-squares principle to minimize

$$U(\beta) = \sum_{z}\sum_{i<j} W_z(\tilde{X}_{ij}, \tilde{Y}_{ij}) D_{ij}(z)[\Delta_{ij} - \eta(z^T\beta)]^2, \tag{34}$$

where $D_{ij}(z)$ is defined above, W_z is the weight function and the definition of $(\tilde{X}_{ij}, \tilde{Y}_{ij})$ depends on the data type which is given below. For typical bivariate data under right censoring, $\tilde{X}_{ij} = X_i \wedge X_j$ and $\tilde{Y}_{ij} = Y_i \wedge Y_j$; for semi-competing risks data, $\tilde{X}_{ij} = \tilde{T}_{1,ij} \wedge \tilde{T}_{2,ij} \wedge \tilde{C}_{ij}$ and $\tilde{Y}_{ij} = \tilde{T}_{2,ij} \wedge \tilde{C}_{ij}$; and, for truncation data, $\tilde{X}_{ij} = \breve{T}_{1,ij}$ and $\tilde{Y}_{ij} = Y_i \wedge Y_j$. The proposed estimator, denoted as $\hat{\beta}$, is the one such that $U(\beta)$ is minimized which can be obtained by solving $u(\beta) = 0$, where $u(\beta) = \partial U(\beta)/\partial\beta$. In the simulations, we will evaluate the weight function of the form,

$$W_{z,a,b}(x, y) = \frac{n_z}{\sum_{i=1}^{n} I\{X_i \geq \min(a, x), Y_i \geq \min(b, y), Z_i = z\}},$$

where n_z is the sample size of $Z = z$; a and b are constants. With $a = b = 0$, the function reduces to 1 which is the un-weighted case. With $a = b = \infty$, the weight function becomes $n_z / \sum_{i=1}^{n} I\{X_i \geq x, Y_i \geq y, Z_i = z\}$. The following theorem provides the asymptotic properties of $\hat{\beta}$. The proof of Theorem 3 is provided in Appendix 5.

Theorem 3: *Let $\hat{\beta}$ denote the solution minimizing (34). We make the following regularity assumptions: (a) The list of possible covariate values is $\mathcal{Z} = \{z_1, ..., z_K\}$ which spans a non-degenerate linear space. That is, the dimensionaltiy of linear space spanned by $\mathcal{Z}$ equals p, the dimensionality of β. (b) n_{z_k}/n converge to constants $0 < c_k < 1$ for $k = 1, ..., K$. (c) The weight function $W_z(u, v)$ has a uniform bounded limit $\tilde{W}_z(u, v)$. That is, $\sup_{z,u,v} |W_z(u, v) - \tilde{W}_z(u, v)| \to 0$ in probability, where $\tilde{W}_z$ is deterministic and bounded for (u, v) in the support of $(\tilde{X}_{ij}, \tilde{Y}_{ij})$. Let β^* be the true value of β. Then $\hat{\beta}$ is a consistent estimator and $\sqrt{n}(\hat{\beta} - \beta^*)$ converges in distribution to a multivariate normal distribution with variance Σ which is consistently estimated by $\hat{\Sigma} = \hat{I}^{-1}\hat{J}(\hat{I}^{-1})'$, where*

$$\hat{I} = \left(-\frac{1}{n^2}\frac{\partial^2 U(\beta)}{\partial\beta_k \partial\beta_l}|_{\beta=\hat{\beta}}\right)_{(p+1)\times(p+1)}, \hat{J} = \left(\hat{J}_{ij}\right)_{(p+1)\times(p+1)},$$

$$\hat{J}_{ij} = n^{-3}\sum_z [2\sum_{k<l<m}(\hat{Q}^{(i)}_{klz}\hat{Q}^{(j)}_{kmz} + \hat{Q}^{(i)}_{klz}\hat{Q}^{(j)}_{lmz} + \hat{Q}^{(i)}_{lmz}\hat{Q}^{(j)}_{kmz}) + \sum_{k<l}(\hat{Q}^{(i)}_{klz}\hat{Q}^{(j)}_{klz})],$$

$$\hat{Q}^{(k)}_{ijz} = 2W_z(\tilde{X}_{ij},\tilde{Y}_{ij})D_{ij}(z)[\Delta_{ij} - \eta(z'\hat{\beta})](-\frac{\exp(\hat{\beta}_0 + ... + \hat{\beta}_p Z_p)Z_k}{(1+\exp(\hat{\beta}_0 + ... + \hat{\beta}_p Z_p))^2}).$$

4.4.5. Checking the Clayton Assumption

Shih (1998) proposed a testing procedure to verify the Clayton assumption for typical bivariate right censored data. The test statistic is expressed as the difference of two estimators of the association parameter which converges to zero when the Clayton assumption holds but converges to a non-zero value when the model assumption is violated. This idea has been applied to semi-competing risks data by Fine et al. (2001). Now under the current regressing setting, we develop a unified approach of model checking which can handle the three data structures. Note that our result is the first application in the literature which can deal with dependent truncation data.

Let $U_1(\beta)$ and $U_2(\beta)$ follow the same form as $U(\beta)$ with the weight function W_z being specified as $W_{z,1}$ and $W_{z,2}$ respectively. The following weight functions are suggested. For typical bivariate right censored data and semi-competing risks data, we can set

$$W_{z,1}(x,y) = 1,\ W_{z,2}(x,y) = \frac{n_z}{\sum_{i=1}^n I(X_i \geq x, Y_i \geq y, Z_i = z)},$$

and for truncation data,

$$W_{z,1}(x,y) = 1,\ W_{z,2}(x,y) = \frac{n_z}{\sum_{i=1}^n I(X_i \leq x, Y_i \geq y, Z_i = z)}.$$

Let $\hat{\beta}_{W_{z,i}}$ be the solution to $u_i(\beta) = 0$ $(i = 1, 2)$. In principle, different weight functions can also be applied and the choice would affect the power of the corresponding test. Shih (1998) suggested to choose two weight functions such that, under the assumed model, one produces a more efficient estimator while the other results in a less efficient estimator.

The proposed test statistic can be expressed as

$$T = n(\hat{\beta}_{W_{z,1}} - \hat{\beta}_{W_{z,2}})'\hat{\Gamma}^{-1}(\hat{\beta}_{W_{z,1}} - \hat{\beta}_{W_{z,2}}),$$

where $\hat{\Gamma} = \left(\hat{\Gamma}_{ij}\right)_{(p+1)\times(p+1)}$,

$$\hat{\Gamma}_{ij} = n^{-3}\sum_{z}[2\sum_{k<l<m}(\hat{Q}^{*(i)}_{klz}\hat{Q}^{*(j)}_{kmz} + \hat{Q}^{*(i)}_{klz}\hat{Q}^{*(j)}_{lmz} + \hat{Q}^{*(i)}_{lmz}\hat{Q}^{*(j)}_{kmz}) + \sum_{k<l}(\hat{Q}^{*(i)}_{klz}\hat{Q}^{*(j)}_{klz})],$$

and $\hat{Q}^{*(i)}_{klz}$ is defined in Appendix. In Appendix 6, we show that when the regression assumption $\theta^{Z}_{*}(s,t) = \exp(Z'\beta)$ holds, T converges in distribution to χ^2_{p+1}. That is, for a γ-level test, we reject the null hypothesis if $T > \chi^2_{p+1,\gamma}$, where $\Pr(\chi^2_{p+1} > \chi^2_{p+1,\gamma}) = \gamma$.

4.4.6. Numerical Analysis

4.4.6.1. Simulations Results

We performed simulations to assess finite-sample performances of the proposed methods. Three regression settings were examined: *case 1 (two groups)*: $\theta^{Z}_{*}(s,t) = \exp(\beta_0 + \beta_1 Z_1)$, where $Z = (1,0)'$ or $(1,1)'$; *case 2 (three groups)*: $\theta^{Z}_{*}(s,t) = \exp(\beta_0 + \beta_1 Z_1)$, where $Z = (1,0)'$, $(1,1)'$ or $(1,2)'$ and *case 3 (three groups)*: $\theta^{Z}_{*}(s,t) = \exp(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2)$, where $Z = (1,0,0)'$, $(1,1,0)'$ or $(1,0,1)'$. The values of parameters were set as follows. For case 1 and case 2, $(\beta_0, \beta_1) = (0.5, 0.5)$ and $(1,1)$; and for case 3, $(\beta_0, \beta_1, \beta_2) = (0.5, 0.5, 0.5)$ and $(1,1,1)$. For each group, we generated (T_1, T_2) which follow model (32). For typical bivariate data and semi-competing risks data, we set $* = a$ and, for truncation data, we set $* = b$. The marginal distributions were generated from $T_1 \sim \exp(0.8)$ and $T_2 \sim \exp(1)$. Right censoring is incorporated in the three data structures. For all the cases, the censoring distribution was generated from $U(0,6)$. For bivariate censored data, we set C_1 to be independent of C_2 and the censoring proportion of T_j $(j = 1,2)$ is around 0.15. For semi-competing risks data, the censoring rate for T_1 which is subject to dependent censoring by T_2 varies from 0.35 $(\tau = 0.76)$ to 0.48 $(\tau = 0.25)$. For truncation data, the missing proportion $\Pr(T_1 > T_2)$ and the censoring rate $\Pr(T_j > C|T_1 \leq T_2)$ $(j = 1, 2)$ vary with τ. When $\tau = 0.25$, $\Pr(T_1 > T_2) \approx 0.43$, $\Pr(T_1 > C|T_1 \leq T_2) \approx 0.09$ and $\Pr(T_2 > C|T_1 \leq T_2) \approx 0.23$. When $\tau = 0.76$, $\Pr(T_1 > T_2) \approx 0.27$, $\Pr(T_1 > C|T_1 \leq T_2) \approx 0.11$ and $\Pr(T_2 > C|T_1 \leq T_2) \approx 0.18$.

The sample size was chosen to be 150 and 300. Two weight functions with $(a, b) = (0, 0)$ and $(a, b) = (\infty, \infty)$ were evaluated. The results for the three regression settings are summarized in Table 4-1 $\sim$ 4-3 respectively. Based on 1000 replications, we computed $\sum_{B=1}^{1000} \hat{\beta}_i^{(B)}/1000 - \beta_i^*$ (bias), and the empirical standard deviation of $\hat{\beta}_i$ $(\bar{\sigma}_i)$ and the estimated standard deviation $\sqrt{n^{-1}\hat{\Sigma}_{ii}}$ ($\hat{\sigma}_i$), and the coverage probability of the nominal 0.95 confidence interval for $\hat{\beta}_i$ (Cov95).

In all the cases, the proposed estimator $\hat{\beta}$ performs well and the variance estimator $\hat{\sigma}$ produces confidence intervals with reasonable coverage probabilities. For typical bivariate right-censored data and semi-competing risks data, the estimator with weight function $(a, b) = (\infty, \infty)$ performs better than $(a, b) = (0, 0)$ but, for truncation data, we get the opposite conclusion since there is no information in the wedge $T_1 > T_2$. Therefore, we evaluated another weight function,

$$W_z^*(x, y) = n_z / \sum_{i=1}^{n} I\{X_i \leq x, Y_i \geq y, Z_i = z\}, \tag{35}$$

and the results are presented in Table 4-4. We see that the new weight function does improve the performances.

			Bias	$\bar{\sigma}$	$\hat{\sigma}$	Cov95	Bias	$\bar{\sigma}$	$\hat{\sigma}$	Cov95
Data	(a, b)	β_0 β_1	$n = 150$				$n = 300$			
Data 1	$(0, 0)$	0.5	0.0065	0.1855	0.1821	0.957	0.0015	0.1259	0.1283	0.945
		0.5	-0.0025	0.2619	0.2593	0.948	-0.0007	0.1935	0.1835	0.953
		1	0.0117	0.1960	0.1856	0.957	0.0059	0.1332	0.1308	0.942
		1	0.0085	0.2858	0.2693	0.954	0.0067	0.1841	0.1890	0.956
	(∞, ∞)	0.5	0.0066	0.1751	0.1656	0.947	-0.0003	0.1127	0.1150	0.951
		0.5	0.0050	0.2578	0.2400	0.951	0.0096	0.1669	0.1677	0.945
		1	0.0083	0.1787	0.1728	0.953	0.0054	0.1278	0.1222	0.946
		1	-0.0079	0.2751	0.2553	0.942	0.0032	0.1862	0.1775	0.949
Data 2	$(0, 0)$	0.5	-0.0003	0.2343	0.2259	0.954	-0.0015	0.1600	0.1586	0.949
		0.5	-0.0094	0.3372	0.3137	0.952	0.0020	0.2229	0.2208	0.954
		1	0.0106	0.2282	0.2167	0.945	-0.0070	0.1592	0.1540	0.954
		1	-0.0093	0.3030	0.3006	0.955	0.0016	0.2132	0.2121	0.950
	(∞, ∞)	0.5	0.0071	0.1996	0.1959	0.953	0.0051	0.1399	0.1382	0.950
		0.5	0.0073	0.2760	0.2793	0.956	-0.0028	0.1923	0.1956	0.948
		1	0.0034	0.1975	0.1971	0.941	0.0026	0.1346	0.1386	0.946
		1	0.0045	0.2889	0.2838	0.947	-0.0011	0.1927	0.1956	0.944
Data 3	$(0, 0)$	0.5	-0.0015	0.1518	0.1469	0.959	0.0033	0.0999	0.1012	0.956
		0.5	0.0078	0.2341	0.2236	0.952	-0.0013	0.1565	0.1540	0.961
		1	0.0060	0.1745	0.1691	0.950	-0.0034	0.1180	0.1160	0.945
		1	-0.0041	0.3370	0.3205	0.951	0.0023	0.2192	0.2189	0.955
	(∞, ∞)	0.5	0.0134	0.2330	0.2155	0.952	0.0065	0.1664	0.1641	0.951
		0.5	-0.0012	0.3456	0.3259	0.956	-0.0055	0.2508	0.2372	0.953
		1	0.0216	0.2667	0.2437	0.950	0.0091	0.1923	0.1766	0.950
		1	0.0160	0.4942	0.4362	0.942	0.0026	0.3575	0.3125	0.945

Table 4-1: Simulation results for case 1. Data 1: Typical bivariate right-censored data; Data 2: Semi-competing risks data; Data 3: Truncation data.

Data	(a,b)	β_0 β_1	Bias ($n=150$)	$\bar{\sigma}$ ($n=150$)	$\hat{\sigma}$ ($n=150$)	Cov95 ($n=150$)	Bias ($n=300$)	$\bar{\sigma}$ ($n=300$)	$\hat{\sigma}$ ($n=300$)	Cov95 ($n=300$)
Data 1	$(0,0)$	0.5	0.0036	0.2096	0.2076	0.947	0.0084	0.1456	0.1459	0.946
		0.5	0.0061	0.1761	0.1711	0.955	-0.0005	0.1187	0.1192	0.949
		1	0.0137	0.2228	0.2197	0.954	0.0079	0.1649	0.1556	0.958
		1	0.0095	0.2308	0.2311	0.953	0.0025	0.1709	0.1597	0.953
	(∞,∞)	0.5	0.0060	0.1974	0.1895	0.952	0.0086	0.1353	0.1330	0.950
		0.5	0.0046	0.1656	0.1606	0.950	-0.0007	0.1131	0.1109	0.950
		1	0.0162	0.2071	0.2083	0.960	0.0053	0.1489	0.1459	0.949
		1	-0.0030	0.2271	0.2267	0.960	0.0049	0.1578	0.1543	0.960
Data 2	$(0,0)$	0.5	-0.0023	0.2612	0.2531	0.941	-0.0036	0.1868	0.1794	0.951
		0.5	0.0102	0.2115	0.2035	0.945	0.0035	0.1418	0.1407	0.948
		1	0.0091	0.2832	0.2589	0.946	-0.0067	0.1865	0.1819	0.947
		1	0.0077	0.2665	0.2543	0.950	0.0076	0.1838	0.1741	0.948
	(∞,∞)	0.5	-0.0054	0.2315	0.2256	0.951	0.0018	0.1622	0.1565	0.951
		0.5	0.0052	0.1874	0.1844	0.946	0.0058	0.1285	0.1273	0.950
		1	0.0121	0.2557	0.2385	0.942	0.0068	0.1712	0.1664	0.945
		1	-0.0054	0.2587	0.2466	0.948	-0.0058	0.1649	0.1643	0.945
Data 3	$(0,0)$	0.5	-0.0045	0.1776	0.1742	0.956	0.0043	0.1204	0.1187	0.948
		0.5	0.0070	0.1753	0.1668	0.950	-0.0028	0.1180	0.1114	0.951
		1	-0.0078	0.2225	0.2079	0.943	0.0023	0.1421	0.1418	0.947
		1	0.0298	0.3234	0.3207	0.948	0.0080	0.2081	0.2064	0.951
	(∞,∞)	0.5	0.0087	0.2668	0.2447	0.947	0.0037	0.1947	0.1825	0.952
		0.5	0.0086	0.2560	0.2416	0.943	-0.0005	0.1833	0.1700	0.952
		1	0.0138	0.3305	0.2901	0.953	0.0076	0.2313	0.2076	0.949
		1	0.0562	0.4820	0.4359	0.940	0.0337	0.3192	0.2919	0.948

Table 4-2: Simulation results for case 2. Data 1: Typical bivariate right-censored data; Data 2: Semi-competing risks data; Data 3: Truncation data.

			Bias	$\bar{\sigma}$	$\hat{\sigma}$	Cov95	Bias	$\bar{\sigma}$	$\hat{\sigma}$	Cov95
Data	(a,b)	β_0 β_1 β_2	$n=150$				$n=300$			
Data 1	$(0,0)$	0.5	0.0133	0.2351	0.2237	0.945	-0.0012	0.1543	0.1572	0.955
		0.5	0.0062	0.3405	0.3194	0.950	0.0089	0.2283	0.2245	0.950
		0.5	-0.0068	0.3353	0.3186	0.956	0.0035	0.2232	0.2247	0.948
		1	0.0155	0.2379	0.2263	0.950	0.0087	0.1715	0.1603	0.948
		1	-0.0065	0.3490	0.3283	0.949	-0.0067	0.2486	0.2319	0.947
		1	0.0083	0.3525	0.3268	0.954	0.0001	0.2487	0.2322	0.954
	(∞,∞)	0.5	0.0142	0.2182	0.2020	0.945	0.0062	0.1442	0.1432	0.941
		0.5	0.0052	0.3104	0.2943	0.949	0.0047	0.2102	0.2076	0.953
		0.5	0.0071	0.3099	0.2949	0.952	0.0044	0.2133	0.2071	0.946
		1	0.0194	0.2190	0.2145	0.939	0.0047	0.1526	0.1497	0.956
		1	-0.0052	0.3164	0.3178	0.959	0.0022	0.2209	0.2189	0.949
		1	-0.0047	0.3247	0.3185	0.954	0.0093	0.2173	0.2193	0.961
Data 2	$(0,0)$	0.5	-0.0054	0.2823	0.2780	0.952	-0.0009	0.1995	0.1944	0.949
		0.5	0.0101	0.3972	0.3855	0.960	-0.0014	0.2844	0.2706	0.956
		0.5	0.0027	0.3956	0.3843	0.945	0.0059	0.2700	0.2704	0.952
		1	-0.0012	0.2799	0.2670	0.945	0.0070	0.1926	0.1879	0.947
		1	0.0056	0.3881	0.3696	0.953	-0.0025	0.2665	0.2596	0.952
		1	0.0123	0.3994	0.3708	0.953	0.0028	0.2623	0.2594	0.949
	(∞,∞)	0.5	0.0153	0.2572	0.2428	0.949	0.0057	0.1744	0.1697	0.947
		0.5	-0.0017	0.3676	0.3456	0.949	0.0088	0.2565	0.2404	0.952
		0.5	0.0077	0.3674	0.3451	0.946	-0.0025	0.2408	0.2417	0.947
		1	0.0149	0.2604	0.2441	0.947	0.0036	0.1732	0.1718	0.961
		1	0.0063	0.3724	0.3541	0.953	-0.0024	0.2489	0.2435	0.946
		1	0.0092	0.3925	0.3539	0.950	0.0058	0.2544	0.2437	0.963
Data 3	$(0,0)$	0.5	-0.0072	0.1863	0.1823	0.940	-0.0005	0.1253	0.1257	0.944
		0.5	0.0039	0.3009	0.2807	0.949	-0.0004	0.1900	0.1910	0.959
		0.5	0.0060	0.2916	0.2802	0.953	0.0065	0.1939	0.1913	0.955
		1	0.0052	0.2242	0.2134	0.943	-0.0039	0.1503	0.1446	0.945
		1	0.0358	0.4464	0.4092	0.942	0.0050	0.2768	0.2731	0.941
		1	0.0319	0.4357	0.4054	0.948	0.0015	0.2924	0.2729	0.948
	(∞,∞)	0.5	0.0161	0.2934	0.2561	0.947	0.0022	0.2142	0.1908	0.952
		0.5	0.0101	0.4348	0.3866	0.949	0.0071	0.3036	0.2834	0.945
		0.5	0.0011	0.4551	0.3864	0.947	0.0032	0.3103	0.2841	0.950
		1	0.0326	0.3289	0.2856	0.945	0.0064	0.2289	0.2100	0.950
		1	0.0443	0.6228	0.5166	0.945	0.0171	0.4321	0.3755	0.952
		1	0.0321	0.6232	0.5109	0.953	0.0145	0.4346	0.3750	0.948

Table 4-3: Simulation results for case 3. Data 1: Typical bivariate right-censored data; Data 2: Semi-competing risks data; Data 3: Truncation data.

		Bias	$\bar{\sigma}$	$\hat{\sigma}$	Cov95	Bias	$\bar{\sigma}$	$\hat{\sigma}$	Cov95
Case	β	$n = 150$				$n = 300$			
Case 1	0.5	-0.0036	0.1103	0.1020	0.955	-0.0038	0.0663	0.0658	0.953
	0.5	0.0007	0.1632	0.1600	0.947	0.0015	0.1075	0.1021	0.954
	1	-0.0017	0.1322	0.1226	0.947	0.0015	0.0802	0.0783	0.957
	1	0.0085	0.2529	0.2463	0.946	-0.0071	0.1611	0.1553	0.945
Case 2	0.5	0.0073	0.1369	0.1267	0.947	-0.0043	0.0819	0.0806	0.957
	0.5	0.0044	0.1311	0.1226	0.953	0.0008	0.0781	0.0768	0.956
	1	-0.0085	0.1284	0.1592	0.948	-0.0019	0.1025	0.0995	0.947
	1	0.0167	0.2517	0.2465	0.954	0.0055	0.1453	0.1469	0.955
Case 3	0.5	-0.0088	0.1396	0.1346	0.949	-0.0024	0.0883	0.0848	0.936
	0.5	0.0033	0.2239	0.2118	0.949	0.0011	0.1389	0.1323	0.955
	0.5	0.0086	0.2111	0.2109	0.956	-0.0063	0.1345	0.1319	0.951
	1	-0.0030	0.1686	0.1610	0.950	-0.0038	0.1025	0.1015	0.945
	1	-0.0011	0.3502	0.3273	0.945	0.0063	0.2112	0.2035	0.954
	1	0.0152	0.3539	0.3313	0.947	-0.0027	0.2114	0.2022	0.945

Table 4-4: Simulation results for truncation data with weight function $W_z^*(x,y)$ in (35).

4.4.6.2. Robustness in Presence of Marginal Heterogeneity

The proposed methodology requires that observations in each subgroup with the same value of Z are identically distributed. In practice, there may exist covariates which may also affect the marginal distributions. If these covariates are discrete, they can be included in the list of Z and the proposed methods are still valid but may lose some efficiency due to extra grouping. We design two regression settings to examine this effect.

In the first setting, let Z_j =0 or 1 ($j = 1, 2$) such that Z_1 affects association and Z_2 affects the marginal distribution. We compare two model specifications with model 1: $\theta_*^Z(s,t) = \exp(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2)$ and model 2: $\theta_*^Z(s,t) = \exp(\beta_0 + \beta_1 Z_1)$. The results are summarized in Table 4-5. The results of model 1 indicate that our method is valid by extra grouping. The results of model 2 show that, if the marginal heterogeneity is ignored, the estimator of the intercept term β_0 is biased while the estimator of β_1 still has reasonable performance. The purpose of the second setting is to evaluate the loss in efficiency due to unnecessary grouping. Let Z_1 affects association and Z_2 is a redundant covariate. Table 4-6 shows that model 1 which includes Z_2 produces slightly larger bias and variation.

For the simulation settings, set $\beta_0 = 0.5$, $\beta_1 = 0.5$ and $\beta_2 = 0$, this is Z_1 affects the association, Z_2 doesn't affect the association. Let (T_1, T_2) follows the Clayton copula and the sample size and replications are 200 and 1000, respectively. For the first setting, Z_2 affects the marginal distribution. Let Z_2 affect the marginal distribution by the proportional hazard relationship, i.e. $S_i(t|Z_2) = S_{i0}(t)^{\exp(\eta Z_2)}$ $(i = 1, 2)$, where $S_{i0}(t)$ is the baseline survival function, $S_i(t|Z_2) = \Pr(T_i > t|Z_2)$ and set η=0.5 and 1. For the baseline survival function, $S_{i0}(t)$, the distribution of T_1 follows exp(0.8) and the distribution of T_2 follows exp(1). For the second setting, Z_2 is a redundant covariate. Let T_1 follow exp(0.8) and T_2 follow exp(1). The censoring variables are generated from $U(0, 6)$. For typical bivariate data, we set C_1 to be independent of C_2.

When there exist continuous covariates that affect the marginal distributions, we can group them into distinct classes. On the other hand, if the marginal covariate effect is ignored, we may suspect that the proposed method can estimate the slope parameters more accurately than the intercept parameter. A possible solution is to mimic the idea of Fine et al. (2000) who construct an association model based on error terms in which is the marginal effects have been removed. Since this is not a straightforward extension for more complicated data structures, we will leave it as future work.

			Model 1: with grouping				Model 2: without grouping			
Data	PH effect	β_0 β_1 β_2	Bias	$\bar{\sigma}$	$\hat{\sigma}$	Cov95	Bias	$\bar{\sigma}$	$\hat{\sigma}$	Cov95
Data 1	$\eta = 0.5$	0.5	0.0010	0.2059	0.1918	0.950	0.0384	0.1640	0.1533	0.949
		0.5	0.0089	0.2381	0.2188	0.954	-0.0098	0.2322	0.2200	0.952
		0	0.0004	0.2339	0.2265	0.954				
	$\eta = 1$	0.5	-0.0057	0.2037	0.1922	0.957	0.1466	0.1575	0.1507	0.837
		0.5	0.0137	0.2255	0.2048	0.948	-0.0357	0.2216	0.2174	0.947
		0	0.0045	0.2291	0.2239	0.942				
Data 2	$\eta = 0.5$	0.5	0.0028	0.2421	0.2389	0.940	0.0364	0.1918	0.1910	0.952
		0.5	0.0065	0.2784	0.2625	0.953	-0.0123	0.2675	0.2666	0.950
		0	-0.0142	0.2795	0.2802	0.940				
	$\eta = 1$	0.5	0.0025	0.2407	0.2376	0.953	0.1524	0.1878	0.1865	0.877
		0.5	-0.0059	0.2677	0.2564	0.948	-0.0533	0.2648	0.2631	0.942
		0	-0.0123	0.2869	0.2756	0.948				
Data 3	$\eta = 0.5$	0.5	0.0015	0.1717	0.1606	0.953	0.0035	0.1253	0.1233	0.956
		0.5	0.0089	0.2054	0.1878	0.956	-0.0122	0.1934	0.1879	0.958
		0	0.0022	0.2059	0.1985	0.948				
	$\eta = 1$	0.5	-0.0085	0.1644	0.1588	0.951	-0.0066	0.1194	0.1224	0.946
		0.5	0.0164	0.1984	0.1789	0.948	-0.0354	0.1827	0.1843	0.944
		0	0.0011	0.2052	0.1958	0.950				

Table 4-5: Simulations with marginal heterogeneity. Data 1: Typical bivariate right-censored data; Data 2: Semi-competing risks data; Data 3: Truncation data.

		Model 1: with gouping				Model 2: without grouping			
Data	β_0 β_1 β_2	Bias	$\bar{\sigma}$	$\hat{\sigma}$	Cov95	Bias	$\bar{\sigma}$	$\hat{\sigma}$	Cov95
Data 1	0.5	0.0062	0.2012	0.1934	0.950	0.0041	0.1572	0.1572	0.951
	0.5	0.0074	0.2371	0.2149	0.956	0.0061	0.2291	0.2245	0.952
	0	0.0011	0.2408	0.2322	0.951				
Data 2	0.5	-0.0043	0.2571	0.2394	0.951	-0.0018	0.2025	0.1944	0.950
	0.5	0.0089	0.2929	0.2766	0.952	0.0025	0.2849	0.2709	0.949
	0	0.0003	0.2961	0.2848	0.956				
Data 3	0.5	0.0045	0.1709	0.1606	0.946	-0.0014	0.1276	0.1253	0.946
	0.5	0.0068	0.2009	0.1968	0.946	0.0081	0.1894	0.1908	0.951
	0	-0.0086	0.2037	0.2021	0.959				

Table 4-6: Simulations with marginal homogeneity. Data 1: Typical bivariate right-censored data; Data 2: Semi-competing risks data; Data 3: Truncation data.

4.4.6.3. Real Datas

The proposed methods are applied to analyze two data sets: the bone marrow transplantation data (Klein and Moeschberger, 2003) which belong to semi-competing risks data and the transfusion-related AIDS data (Kalbfleisch and Lawless, 1989) which belong to truncation data.

For the first data set, the main objective is to investigate how the disease type (ALL, AML low risk, AML high risk) affects the level of association between the survival time (T_2) and the time to develop the chronic graft-versus-host disease (T_1). In presence of right censoring, we only observe $(X, Y, \delta_1, \delta_2, Z)$, where X is the time to chronic graft-versus-host disease or death or the end of study, Y is the time to death or the end of study and δ_j indicates whether T_j is observed ($j = 1, 2$). The covariate vector $Z = (Z_1, Z_2)^T$ is coded as follows: $(Z_1, Z_2) = (0, 0)$ if the disease type is the ALL group; $(Z_1, Z_2) = (1, 0)$ for the AML low risk group and $(Z_1, Z_2) = (0, 1)$ for the AML high risk group.

The regression model can be written as: $\theta_a^Z(s, t) = \exp(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2)$, where β_0 is the log odds of concordance for the baseline (ALL) group, β_1 represents the difference of the log odds by comparing the AML low risk group with the baseline group, β_2 is the difference of the log odds by comparing the AML high risk group with the baseline group. Hence $\beta_1 - \beta_2$ measures the difference of the log odds between the AML low risk group and the AML high risk group.

Applying the testing procedure discussed in Section 4.4.5, we obtain that $T = 2.9034$ with P-value=0.407 which implies that the Clayton model is suitable for this data. We run two analyses with different weight functions. When $a = b = \infty$, the estimators and the corresponding standard errors given in the parentheses for β and $\beta_1 - \beta_2$ are $\hat{\beta}_0$=-0.5355 (0.2756), $\hat{\beta}_1$=1.2188 (0.4770), $\hat{\beta}_2$=0.5629 (0.3666) and $\hat{\beta}_1 - \hat{\beta}_2$=0.6559 (0.4582). Accordingly the odds ratios for the above three sets of comparison are $\exp(\hat{\beta}_1)$=3.3831,

$\exp(\hat{\beta}_2)$=1.7557 and $\exp(\hat{\beta}_1 - \hat{\beta}_2)$=1.9268. Although there seems to be positive association between the two failure times for each risk group, only the result by comparing the AML low risk group and the baseline group is statistically significant in which the former reveals larger association. With the weight function $a = b = 0$, the estimators and the corresponding standard errors in the parentheses for β and $\beta_1 - \beta_2$ are $\hat{\beta}_0$=-0.4480 (0.3107), $\hat{\beta}_1$=1.2573 (0.5075), $\hat{\beta}_2$=0.4875 (0.3952) and $\hat{\beta}_1 - \hat{\beta}_2$=0.7698 (0.4697). The results of the two analyses are similar.

The second data set is from a study of blood-transfusion related AIDS (Kalbfleisch and Lawless, 1989). There are 293 subjects who infected HIV by contaminated blood transfusions and developed AIDS between January 1, 1978 and July 1, 1986. The data include the infection time (S) measured from the beginning of the study and induction time (T_1) in months, and the age in years at the time of transfusion. Subjects are included in the sample only if they developed AIDS within the study period which lasted 102 months. Let $T_2 = 102 - S$. Hence the sampling criteria is $T_1 \leq T_2$. Tsai (1990) analyzed this data and rejected the null hypothesis of quasi-independence. Now we want to investigate whether age affects the level of association. For the age variable, we classify it into three groups: 0-4 years, 5-59 years and $\geq$ 60 years. Hence, the covariate $Z = (Z_1, Z_2)^T$ is coded as $(Z_1, Z_2) = (0, 0)$ if the age of subject is in the $\geq$ 60 years group; $(Z_1, Z_2) = (1, 0)$ for the 5-59 years group and $(Z_1, Z_2) = (0, 1)$ for the 0-4 years group. Therefore, we consider the regression model: $\theta_b^Z(s, t) = \exp(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2)$.

Since the data contain many ties, we modify the data by adding a uniform random variable between $(-0.4, 0.4)$ to each subject. Applying the testing procedure in Section 4.4.5, we find $T = 1.395$ with p-value=0.707. Therefore, the Clayton model is suitable for this data. We run two analyses with different weight functions. When $a = b = 0$, the estimators and the corresponding standard errors given in the parentheses for β and $\beta_1 - \beta_2$ are $\hat{\beta}_0$ =0.2168 (0.0982), $\hat{\beta}_1$ =-0.0435 (0.1464), $\hat{\beta}_2$ =-0.0336 (0.1938), $\hat{\beta}_1 - \hat{\beta}_2$ =-0.0099 (0.1994). Accordingly the odds ratios for the above three groups are $\exp(\hat{\beta}_1)$ =0.9574, $\exp(\hat{\beta}_2)$ =0.9669 and $\exp(\hat{\beta}_1 - \hat{\beta}_2)$ =0.9901. From the results, the difference of the association for the three age groups is not significantly. From the results: $\hat{\beta}_0$ =0.2168 (0.0982), $\hat{\beta}_0 + \hat{\beta}_1$ =0.1733 (0.1087) and $\hat{\beta}_0 + \hat{\beta}_2$ =0.1832 (0.1671), it implies that the quasi-independent assumption is not appropriate for the ≥ 60 years group. The conclusion coincides with Tsai (1990). With the weight function (35), the estimators and the corresponding standard errors in the parentheses for β and $\beta_1 - \beta_2$ are $\hat{\beta}_0$ =0.2128 (0.0693), $\hat{\beta}_1$ =0.0023 (0.0993), $\hat{\beta}_2$ =-0.0496 (0.0989), $\hat{\beta}_1 - \hat{\beta}_2$ =0.0519 (0.1002).

Chapter 5. Conclusions

In this article, we consider two projects: regression analysis on marginal effect and association. In the first project, we model the failure time to a non-terminal event by a flexible transformation model under semi-competing risks data. To handle the problem of dependent censoring, we make an additional assumption that, for each covariate group, failure times of the two types of events follow a copula model in the identifiable region. Model checking procedures are also proposed to examine the appropriateness of these two model assumptions. To select an appropriate regression model for the non-terminal event, a formal model checking procedure is also proposed using the bootstrap method. The proposed strategy for checking the copula assumption is to compare the non-parametric estimator with its model-based estimator of a chosen reference function, say $F^{11}(t_1, t_2)$. Compared to existing methods such as that proposed by Lin et al. (1996), our approach allows for different dependence structures in each group, avoids making additional model assumption on the terminal event and utilizes all the data without paying the price for artificial censoring. The simulation analysis confirms our conjecture that the estimator proposed by Lin et al. (1996) becomes unreliable if the dependence structures in the two groups are different.

For possible future research of marginal regression analysis, one may examine how to choose such a function or a combination of several functions that contain most of the model information characterized by $\phi(\cdot)$ so that the corresponding test procedure would detect the departure from the null hypothesis better and hence gives higher power. The proposed regression method can handle multiple covariates with discrete values. Extension to continuous covariates has to face the challenge of imposing additional regression assumptions on model (12) or adopting some nonparametric techniques like smoothing. This goes beyond the scope of the current article but may deserve further investigation.

In the second project, we develop a unified inference approach to assessing the covariate effects on the level of association between two failure times which may be typical bivariate variables or have censoring/truncation relationship. We also develop a methodology to check the Clayton assumption for the three types of data.

Although we assume the Clayton copula, the proposed regression model can easily be extended to allow for dependency of local variation. Note that the form of $\theta_*^Z(s,t)$ in (32) can be extended to allow for the dependency on (s,t). For example, we may assume that $\theta_*^Z(s,t) = \exp(Z'\beta) + \eta \cdot (F_*(s,t) - v_0)$, where $v_0 \in [0,1]$ is a pre-determined constant. The interpretation of β depends on the chosen value v_0. For example if we set $v_0 = 0.5$, β represents the covariate effect for the covariate group at the "median" time region with $F_*(s,t) = 0.5$. Since such an extension may involve additional parameters, say $F_*(s,t)$ we leave it as a future research topic. Currently the proposed method can only be applied to the discrete covariates. When there exists continuous covariates, we can group the covariates into several classes or apply smoothing techniques.

Appendix

Appendix 1: *Regularity Conditions of Theorem 1A and 1B*

Let

$$\bar{U}(\theta) = \sum_{k<j} w_0(z_{kj}^T\theta) z_{kj} \sqrt{\frac{c_k c_j}{c_k + c_j}} \{\int_0^{T_{kj}} W_{kj}(t) g_{kj}(t,\theta) dt\} \tag{A.1}$$

where $c_k = \lim_{n\to\infty} n_k/n$, and $(0, T_{kj})$ is the support of X in the subgroup with $Z = z_k$ or $Z = z_j$. First we state the regularity conditions:

(a). As $n \to \infty$, $c_k = \lim_{n\to\infty} n_k/n > 0$ for all k values;

(b). For each $Z = z_k$, the $H_{z_k}(u, v, \alpha)$ has bounded partial derivatives with respect to u, v and α, where $H_z(u, v, \alpha) = \phi_{z,\alpha}^{-1}\{\phi_{z,\alpha}(u) - \phi_{z,\alpha}(v)\}$ is defined in (16).

(c). For each $Z = z_k$, the standard regularity conditions hold for estimating $F_{z_k}(x, x)$ and $F_{2,z_k}(x)$ (e.g. conditions for Theorem 6.3.2 in Flemingm and Harrington, 1991) so that $\sqrt{n_k}\{\hat{F}_{z_k}(x, x) - F_{z_k}(x, x)\}$ and $\sqrt{n_k}\{\hat{F}_{2,z_k}(x) - F_{2,z_k}(x)\}$ converge weakly to Gaussian processes;

(d). The weight functions $w_0(x)$ and $W_{kj}(t)$ are positive and bounded and $w_0(x)$ is differentiable with continuous derivatives;

(e). For each of the two classes of model (9), we impose the following assumptions:

- (e1) for the first case, $h(t)$ is differentiable, $h'(t) \neq 0$ and is continuous, $\tilde{W}_{kj}(t) = W_{kj}(t)/h'(t)$ is differentiable and $\int |\tilde{W}_{kj}(t)| dt < \infty$;
- (e2) for the second case, the distribution of ε has a density $f_\varepsilon(t)$ which is differentiable with bounded derivative;

(f). The function $\bar{U}(\theta)$ defined in (A.1) is differentiable with respect to θ and the matrix $(\frac{\partial}{\partial\theta_1}\bar{U}(\theta_0), ..., \frac{\partial}{\partial\theta_p}\bar{U}(\theta_0))$ is nonsingular. Furthermore $\bar{U}(\theta) \neq 0$ for $\theta \neq \theta_0$ and $\liminf_{\|\theta\|\to\infty} |\bar{U}(\theta)| > 0$.

Appendix 2: *Proof of Theorem 1A*

By the proof of (A.12) in the Appendix 7, it follows that $U(\theta)/\sqrt{n}$ converges (in probability) to

$$\bar{U}(\theta) = \sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{c_kc_j}{c_k+c_j}}\{\int_0^{T_{kj}} W_{kj}(t)g_{kj}(t,\theta)dt\},$$

in which the convergence is uniform in θ. Consider a compact set $D_r = \{\|\theta-\theta_0\| \leq r\}$ where r is a positive constant. By assumption (f) $\bar{U}(\theta) \neq 0$ for $\theta \neq \theta_0$ and $\liminf_{\|\theta\|\to\infty} |\bar{U}(\theta)| > 0$, then the continuity of $\bar{U}(\theta)$ implies that $\inf_{\|\theta-\theta_0\|>r} |\bar{U}(\theta)| > 0$. The (uniform) convergence of $U(\theta)/\sqrt{n}$ to $\bar{U}(\theta)$ implies that there will be no solution for $U(\theta) = 0$ outside the compact set D_r when n is large. Since this is true for every $r > 0$, $\hat{\theta}$ is consistent.

By Taylor expansion we get

$$U(\hat{\theta}) = 0 = U(\theta_0) + \sum_{l=1}^{p} \frac{\partial}{\partial\theta_l}U(\check{\theta})(\hat{\theta}_l - \theta_{l,0}), \tag{A.2}$$

where $\check{\theta}$ is an intermediate value between $\theta_0 = (\theta_{1,0}, ..., \theta_{p,0})^T$ and $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_p)^T$. Hence we have the following expression:

$$\frac{1}{\sqrt{n}}(\frac{\partial}{\partial\theta_1}U(\check{\theta}), ..., \frac{\partial}{\partial\theta_p}U(\check{\theta}))\sqrt{n}(\hat{\theta}-\theta_0) = -U(\theta_0). \tag{A.3}$$

From (A.13) about the convergence of $\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta_l}U(\theta)$ to $\frac{\partial}{\partial\theta_l}\bar{U}(\theta)$ locally uniformly at $\theta = \theta_0$ and the consistency of $\hat{\theta}$, we can show that $\frac{1}{\sqrt{n}}(\frac{\partial}{\partial\theta_1}U(\check{\theta}), ..., \frac{\partial}{\partial\theta_p}U(\check{\theta})) \xrightarrow{P} (\frac{\partial}{\partial\theta_1}\bar{U}(\theta_0), ..., \frac{\partial}{\partial\theta_p}\bar{U}(\theta_0))$ which, by assumption (f) is a non-singular constant matrix. By (A.14), $U(\theta_0)$ is asymptotic normal with mean zero. Therefore $\sqrt{n}(\hat{\theta}-\theta_0)$ is asymptotic normal with mean zero because it has the same asymptotic distribution as $-(\frac{\partial}{\partial\theta_1}\bar{U}(\theta_0), ..., \frac{\partial}{\partial\theta_p}\bar{U}(\theta_0))^{-1}U(\theta_0)$. This completes the proof.

Now we consider the situation that the weight function $W_{kj}(x)$ is estimated by $\hat{W}_{kj}(x)$. An example is the Gehan-type weight suggested in Klein and Moeschberger

(p.230)

$$\hat{W}_{kj}(x) = \frac{(n_k + n_j)\hat{G}_{z_k}(x)\hat{G}_{z_j}(x)}{n_k\hat{G}_{z_k}(x) + n_j\hat{G}_{z_j}(x)},$$

where $\hat{G}_{z_k}(x)$ is the Kaplan-Meier estimator of $G_{z_k}(x) = \Pr(C \geq x|Z = z_k)$. Replacing $W_{kj}(t)$ in (19) with $\hat{W}_{kj}(t)$, we have

$$\hat{U}(\theta) = \sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{n_k n_j}{n_k + n_j}}\{\int_0^{t_{kj}} \hat{W}_{kj}(t)\hat{g}_{kj}(t,\theta)dt\}.$$

Appendix 3: *Proof of Theorem 1B*

Compared with the previous proof, we only need to show that (A) $[\hat{U}(\theta) - U(\theta)]/\sqrt{n}$ uniformly strongly converges to zero; (B) $\frac{\partial}{\partial\theta_l}[\hat{U}(\theta) - U(\theta)]/\sqrt{n}$ strongly converges to zero which takes place locally uniformly at $\theta = \theta_0$ and (C) $\hat{U}(\theta_0) - U(\theta_0)$ strongly converges to zero.

Firstly,

$$[\hat{U}(\theta) - U(\theta)]/\sqrt{n} = \sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n) + (n_j/n)}}\{\int_0^{t_{kj}} [\hat{W}_{kj}(t) - W_{kj}(t)]\hat{g}_{kj}(t,\theta)dt\}. \tag{A.4}$$

We have $\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}} \to \sqrt{\frac{c_k c_j}{c_k + c_j}}$, $w_0(z_{kj}^T\theta_0)z_{kj}$ is bounded and t_{kj} is bounded by $T < \infty$.

These facts together with the uniform strong convergence of $\hat{W}_{kj}(t)$ to $W_{kj}(t)$ implies the uniform strong convergence of (A.4) to zero. So (A) holds.

Secondly,

$$\hat{U}(\theta_0) - U(\theta_0) = \sum_{k<j} w_0(z_{kj}^T\theta_0)z_{kj}\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n) + (n_j/n)}}\{\int_0^{t_{kj}} [\hat{W}_{kj}(t) - W_{kj}(t)]\sqrt{n}\hat{g}_{kj}(t,\theta_0)dt\}. \tag{A.5}$$

By the facts that $\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}} \to \sqrt{\frac{c_k c_j}{c_k + c_j}}$, $w_0(z_{kj}^T\theta_0)z_{kj}$ is bounded and t_{kj} is bounded by $T < \infty$ as well as $\sqrt{n}\hat{g}_{kj}(t,\theta_0) = O_p(1)$ for all t and $\hat{W}_{kj}(t) - W_{kj}(t) = o_p(1)$, we can show strong convergence of (A.5) to zero. So (C) holds.

Finally

$$\begin{aligned} & \frac{\partial}{\partial\theta_l}[\hat{U}(\theta)-U(\theta)]/\sqrt{n} \\ = & \sum_{k<j} w_0'(z_{kj}^T\theta)z_{kj}z_{kj,l}\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}}\{\int_0^{t_{kj}}[\hat{W}_{kj}(t)-W_{kj}(t)]\hat{g}_{kj}(t,\theta)dt\} \\ & +\sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}}\frac{\partial}{\partial\theta_l}\{\int_0^{t_{kj}}[\hat{W}_{kj}(t)-W_{kj}(t)]\hat{g}_{kj}(t,\theta)dt\}. \end{aligned} \tag{A.6}$$

$\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}} \to \sqrt{\frac{c_kc_j}{c_k+c_j}}$, $w_0(z_{kj}^T\theta)z_{kj}$ and $\hat{g}_{kj}(t,\theta)$ are bounded and t_{kj} is bounded by $T<\infty$. From assumption (d) that $w_0'(t)$ is continuous, thus $w_0(z_{kj}^T\theta)$ is locally bounded. Also $\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}} \to \sqrt{\frac{c_kc_j}{c_k+c_j}}$, $|\hat{g}_{kj}(t,\theta)| \le 2$ and t_{kj} is bounded by $T<\infty$. These facts plus the uniform strong convergence of $\hat{W}_{kj}(t)$ to $W_{kj}(t)$ implies that the first term in (A.6) converges uniformly (in θ) and strongly to zero. To prove the second term, we need to consider the two regression classes separately.

For the first case that $\xi_\theta(F)(t)=F[h^{-1}\{h(t)+\theta\}]$,

$$\begin{aligned} & \sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}}\frac{\partial}{\partial\theta_l}\{\int_0^{t_{kj}}[\hat{W}_{kj}(t)-W_{kj}(t)]\hat{g}_{kj}(t,\theta)dt\} \\ = & \sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}} \\ & \int_{h^{-1}\{h(0)+z_{kj}^T\theta\}}^{h^{-1}\{h(t_{kj})+z_{kj}^T\theta\}} \frac{[\hat{W}_{kj}(h^{-1}\{h(t*)-z_{kj}^T\theta\})-W_{kj}(h^{-1}\{h(t*)-z_{kj}^T\theta\})]z_{kj,l}}{h'(h^{-1}\{h(t*)-z_{kj}^T\theta\})}d\hat{F}_{1,z_k}(t*). \end{aligned}$$

(See the proof of (A.11) in Appendix 7.) This quantity locally (at θ_0) uniformly strongly converges to zero due to the boundedness of $w_0(z_{kj}^T\theta)z_{kj}$, $\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}} \to \sqrt{\frac{c_kc_j}{c_k+c_j}}$, the uniform strong convergence of $\hat{W}_{kj}(t)$ to $W_{kj}(t)$, the fact that t_{kj} is bounded by $T<\infty$, local boundedness of $1/h'(h^{-1}\{h(t*)-z_{kj}^T\theta\})$ due to the continuity of $h'(t)$, and the boundedness of $\hat{F}_{1,z_k}$. That is, the second term in (A.6) locally uniformly strongly converges to zero.

For the second case that $\xi_\theta(F)(t)=F_\varepsilon[F_\varepsilon^{-1}\{F(t)\}+\theta]$, then $\xi_\theta'(F)(t)=f_\varepsilon[F_\varepsilon^{-1}\{F(t)\}+\theta]$. Hence

$$\begin{aligned} & \sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}}\frac{\partial}{\partial\theta_l}\{\int_0^{t_{kj}}[\hat{W}_{kj}(t)-W_{kj}(t)]\hat{g}_{kj}(t,\theta)dt\} \\ = & \sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}}\int_0^{t_{kj}}[\hat{W}_{kj}(t)-W_{kj}(t)]z_{kj,l}f_\varepsilon[F_\varepsilon^{-1}\{\hat{F}(t)\}+z_{kj}^T\theta]dt, \end{aligned}$$

which converges uniformly to zero, due to the boundedness of $w_0(z_{kj}^T\theta)z_{kj}$, $\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}} \to \sqrt{\frac{c_kc_j}{c_k+c_j}}$, uniform strong convergence of $\hat{W}_{kj}(t)$ to $W_{kj}(t)$, t_{kj} is bounded by $T<\infty$ and the boundedness of $f_\varepsilon[F_\varepsilon^{-1}\{\hat{F}(t)\} + z_{kj}^T\theta]$. That is, the second term in (A.6) locally uniformly strongly converges to zero.

In summary (A.6) locally uniformly strongly converges to zero. That is, (B) holds. This completes the proof.

Discussion of the regularity conditions

Here we examine the plausibility of some assumptions.

Assumption (b) requires that the imposed AC model has reasonable analytic properties. For most commonly used models in the family, explicit expressions of $H(u, v, \alpha)$ are available and have nice analytic properties such as differentiability.

Assumption (d) is related to the weight functions $w_0(x)$ and $W_{kj}(t)$. It is easy to find reasonable weight functions which satisfy the above conditions. Examples of $w_0(x)$ include the unweighted version with $w_0(x) \equiv 1$ or any positive analytic functions such as power functions. Examples of $W_{kj}(t)$ may be the unweighted one, the Gehan-type function with $W_{kj}(t) = G(t)$.

Now we examine assumption (e1) which involve both $W_{kj}(t)$ and the first class of model (9). The requirement that $h'(t) \neq 0$ and is continuous is easily met by commonly seen transformations including the location-shift model with $h(t) = t$, the AFT model with $h(t) = log(t)$ or other power transforms. The condition on $\tilde{W}_{kj}(t) = W_{kj}(t)/h'(t)$ looks complicated at the first glance. However, if we let both $W_{kj}(t)$ and $h(t)$ be analytic, then obviously $\tilde{W}_{kj}(t)$ is differentiable and $\tilde{W}'_{kj}(t)$ is absolutely integrable so that $\tilde{W}_{kj}(t) = \int_0^t \tilde{W}'_{kj}(u)du$.

Assumption (e2) is also easy to attain by most continuous distributions. In particular, the proportional hazards model (with ϵ following the extreme value distribution) and the proportional odds model (with ϵ following the logistic distribution) clearly satisfy this assumption.

The condition (f) is related to properties of $\bar{U}(\theta)$ defined in (A.1). The differentiability property is easy to achieve. The derivations in the proof of (A.11) provide a concrete example. Non-singularity of $(\frac{\partial}{\partial\theta_1}\bar{U}(\theta_0), ..., \frac{\partial}{\partial\theta_p}\bar{U}(\theta_0))$ is harder to verify but is a standard regularity condition for this type of estimator in solving an estimating equation. For the second statement in (f), notice that as $\|\theta\| \to \infty$, $\int_0^{T_{kj}} W_{kj}(t)[\xi_{z_{kj}^T\theta}(F_{1,z_k})(t) - F_{1,z_j}(t)]dt$ converges to either $\int_0^{T_{kj}} W_{kj}(t)[1 - F_{1,z_j}(t)]dt$ or $\int_0^{T_{kj}} W_{kj}(t)[0 - F_{1,z_j}(t)]dt$ except along the direction $z_{kj}^T\theta = 0$. When there are enough $z_1, z_2, ..., z_K$ values so that they do not fall into a degenerate $p-1$ dimensional linear subspace, then $\liminf_{\|\theta\|\to\infty} |\int_0^{T_{kj}} W_{kj}(t)[\xi_{z_{kj}^T\theta}(F_{1,z_k})(t) - F_{1,z_j}(t)]dt| > 0$ for some k and j along every direction. If we take the weight function $w_0(x)$ such that $\liminf_{|x|\to\infty} w_0(x) > 0$, then as $\|\theta\| \to \infty$, some terms in the sum for $\bar{U}(\theta)$ in (A.1) is bounded away from zero for every direction. Hence $\liminf_{\|\theta\|\to\infty} |\bar{U}(\theta)| > 0$ unless the terms in sum of (A.1) happen to cancel each other out. Such a cancellation rarely happens. Considering the weight function $W_{kj}(t)$ in a metric functional space, the set of weight functions that allow the canceling effect has measure zero. In other words, if a weight function is selected at random, the second statement of (f) holds almost surely.

Appendix 4: *Proof of Theorem 2*

First, the empirical distribution function is uniformly consistent. That is, as $n \to \infty$,

$$\sup_{t_1 \le t_2} |\hat{F}^{11}(t_1, t_2) - F^{11}(t_1, t_2)| \xrightarrow{P} 0.$$

By Theorem 3.4.2 in Fleming & Harrington (1991), $\hat{G}(t)$ and $\hat{F}_2(t)$ are uniformly consistent for $G(t)$ and $F_2(t)$, respectively. In Fine et al. (2001), they show that $\hat{F}_1(t)$ uniformly and strongly converges to $F_1(t)$ for the Clayton family. Their result can be extended to all members in the AC family, see the proof of (A.8). These facts together with the continuous mapping theorem implies that $\tilde{F}_k^{11}(t_1, t_2)$ uniformly converges to

$$\bar{F}_k^{11}(t_1, t_2, \hat{\alpha}) = \frac{\int_{y=t_2}^{\infty} \int_{x=t_1}^{y} \bar{F}_k(dx, dy, \hat{\alpha})G(y)}{\int_{y=0}^{\infty} \int_{x=0}^{y} \bar{F}_k(dx, dy, \hat{\alpha})G(y)},$$

for $t_1 \le t_2$, where $\bar{F}_k(dx, dy, \hat{\alpha}) = \bar{F}_k(x, y, \hat{\alpha}) - \bar{F}_k(x+dx, y, \hat{\alpha}) - \bar{F}_k(x, y+dy, \hat{\alpha}) + \bar{F}_k(x+dx, y+dy, \hat{\alpha})$, $\bar{F}_k(x, y, \hat{\alpha}) = {\phi_{\hat{\alpha}}^{(k)}}^{-1}\{\phi_{\hat{\alpha}}^{(k)}[F_1(x)] + \phi_{\hat{\alpha}}^{(k)}[F_2(y)]\}$.

If $\phi_\alpha^{(k)}$ is the true copula model, then $\hat\alpha \xrightarrow{P} \alpha$ and $\bar F_k^{11}(t_1,t_2,\hat\alpha)$ uniformly converges to $F^{11}(t_1,t_2)$. Therefore,

$$\begin{aligned} & D^k \\ \le\ & \sup_{t_1\le t_2} |\hat F^{11}(t_1,t_2) - F^{11}(t_1,t_2)| + \sup_{t_1\le t_2} |\tilde F_k^{11}(t_1,t_2) - \bar F_k^{11}(t_1,t_2,\hat\alpha)| \\ & + \sup_{t_1\le t_2} |\bar F_k^{11}(t_1,t_2,\hat\alpha) - F^{11}(t_1,t_2)| \\ \xrightarrow{P}\ & 0 \end{aligned}$$

If $\phi_\alpha^{(k)}$ is not the true copula model, let $d_k(\alpha^*) = \sup_{t_1\le t_2} |\bar F_k^{11}(t_1,t_2,\alpha^*) - F^{11}(t_1,t_2)|$. Then $d_k(\alpha^*) > 0$ for all $\alpha^* \in A_k$. To see this, let us consider when $d_k(\alpha^*) = 0$. That is,

$$\frac{\int_{y=t_2}^\infty \int_{x=t_1}^y \bar F_k(dx,dy,\alpha^*)G(y)}{\int_{y=0}^\infty \int_{x=0}^y \bar F_k(dx,dy,\alpha^*)G(y)} = \frac{\int_{y=t_2}^\infty \int_{x=t_1}^y F(dx,dy)G(y)}{\int_{y=0}^\infty \int_{x=0}^y F(dx,dy)G(y)}$$

for all $t_1 \le t_2$, where $F(dx,dy) = F(x,y) - F(x+dx,y) - F(x,y+dy) + F(x+dx,y+dy)$, $F(x,y) = \phi_\alpha{}^{-1}\{\phi_\alpha[F_1(x)] + \phi_\alpha[F_2(y)]\}$. Let $p^* = \int_{y=0}^\infty \int_{x=0}^y \bar F_k(dx,dy,\alpha^*)G(y)$ and $p = \int_{y=0}^\infty \int_{x=0}^y F(dx,dy)G(y)$. Then p and p^* are constants independent of t_1 and t_2. Hence the above equation becomes

$$\int_{y=t_2}^\infty \int_{x=t_1}^y [\bar F_k(dx,dy,\alpha^*)/p^* - F(dx,dy)/p]G(y) = 0$$

for all $t_1 \le t_2$. Therefore,

$$\bar F_k(dx,dy,\alpha^*)/p^* - F(dx,dy)/p = 0$$

on the region $\{(x,y) : x \le y \text{ and } G(y) > 0\}$. Consider the variables $u = F_1(x)$ and $v = F_2(y)$, it is easy to see that $F(dx,dy) = C_\alpha(du,dv)$ and $\bar F_k(dx,dy,\alpha^*) = C_{\alpha^*}^{(k)}(du,dv)$. By the assumption that C has bigger support than the supports of T_1 and T_2, both $F_1(x)$ and $F_2(y)$ change from 1 to 0 on the region $\{(x,y) : x \le y \text{ and } G(y) > 0\}$. Therefore we have

$$C_{\alpha^*}^{(k)}(du,dv)/p^* - C_\alpha(du,dv)/p = 0$$

on the region $\{(u,v) : 0 \le F_1[F_2^{-1}(v)] \le u \le 1\}$. In other words,

$$p\frac{\partial^2}{\partial u \partial v} C_{\alpha^*}^{(k)}(u,v) = p^* \frac{\partial^2}{\partial u \partial v} C_\alpha(u,v)$$

on the region $\{(u,v) : 0 \le F_1[F_2^{-1}(v)] \le u \le 1\}$ which clearly contains a non-empty open set. Therefore, the analyticity of $\phi_{\alpha^*}^{(k)}$ and ϕ_α implies that

$$p\frac{\partial^2}{\partial u \partial v} C_{\alpha^*}^{(k)}(u,v) = p^* \frac{\partial^2}{\partial u \partial v} C_\alpha(u,v)$$

on the region $\{(u,v) : 0 \le u \le 1,\ 0 \le v \le 1\}$. This together with the fact that $C_{\alpha^*}^{(k)}(0,0) = C_\alpha(0,0) = 0$ implies

$$pC_{\alpha^*}^{(k)}(u,v) = p^* C_\alpha(u,v)$$

on the region $\{(u,v) : 0 \le u \le 1,\ 0 \le v \le 1\}$. Since $C_{\alpha^*}^{(k)}(1,1) = C_\alpha(1,1) = 1$, $p = p^*$. Now, $C_{\alpha^*}^{(k)}(u,v) = C_\alpha(u,v)$ on the region $\{(u,v) : 0 \le u \le 1,\ 0 \le v \le 1\}$. This contradicts with the fact that $C_{\alpha^*}^{(k)}(u,v)$ is not the true copula model. Hence $d_k(\alpha^*) > 0$ for all $\alpha^* \in A_k$. This fact together with the closedness of A_k and the continuity in α^* implies that $d_k = \inf_{\alpha^* \in A_k} d_k(\alpha^*) > 0$. Therefore,

$$\begin{aligned}
& D^k \\
\ge\ & \sup_{t_1 \le t_2} |\bar{F}_k^{11}(t_1,t_2,\hat{\alpha}) - F^{11}(t_1,t_2)| - \sup_{t_1 \le t_2} |\hat{F}^{11}(t_1,t_2) - F^{11}(t_1,t_2)| \\
& - \sup_{t_1 \le t_2} |\tilde{F}_k^{11}(t_1,t_2) - \bar{F}_k^{11}(t_1,t_2,\hat{\alpha})| \\
\ge\ & d_k - \sup_{t_1 \le t_2} |\hat{F}^{11}(t_1,t_2) - F^{11}(t_1,t_2)| - \sup_{t_1 \le t_2} |\tilde{F}_k^{11}(t_1,t_2) - \bar{F}_k^{11}(t_1,t_2,\hat{\alpha})| \\
\xrightarrow{P}\ & d_k > 0.
\end{aligned}$$

Now we have proved that, as $n \to \infty$, $D^k \xrightarrow{P} 0$ if $\phi_\alpha^{(k)}$ is the true copula model; and $\Pr(\liminf_{n\to\infty} D^k \ge d_k > 0) = 1$ if $\phi_\alpha^{(k)}$ is not the true copula model. So if there are K candidate models and let

$$d = \inf_{\{k: 1 \le k \le K,\ \phi_\alpha^{(k)} \text{ is not true copula model}\}} d_k.$$

Then $d > 0$. And as $n \to \infty$, $\Pr(D^k > d/2) \to 1$ if model k is wrong while $\Pr(D^k < d/2) \to 1$ if model k is correct. Therefore $\hat{k}$ is consistent.

<u>Discussion of Theorem 2</u>

Notice that there may be more than one correct candidate models. For example, $\phi_\alpha = \phi_{\alpha_1}^{(k_1)} = \phi_{\alpha_2}^{(k_2)}$. In such cases, Theorem 2 states that with probability one, for large n,

$\hat{k} = k_1$ or $\hat{k} = k_2$. That is, the selected model will be one of the correct models for large n, but do not specify which one.

Theorem 2 requires that the true and candidate AC models are all analytic in t and continuous in α. This condition can be easily verified for all commonly used AC models. The assumption requires a closed parameter space for α. This condition is satisfied for most commonly used AC model by including the limiting copula. For example, the Gumbel family is

$$C_\alpha(u,v) = e^{-[(-logu)^\alpha + (-logv)^\alpha]^{1/\alpha}} \quad (\alpha \geq 1).$$

The parameter space can be defined as $[1,\infty]$ by including the case $\alpha = \infty$ corresponding to the copula of Fréchet upper bound $C_\infty(u,v) = \min\{u,v\}$.

We have proved Theorem 2 for continuous random variables T_1 and T_2. From the proof, we can see that this condition can be relaxed to the situation that T_1 and T_2 have positive density on intervals $(t_{1,x}, t_{2,x})$ and $(t_{1,y}, t_{2,y})$ respectively, where $t_{1,x} < t_{2,x}$ and $t_{1,y} < t_{2,y}$. Then the region $\{(u,v) : u = F_1(t_x), v = F_2(t_y), t_{1,x} \leq t_x \leq t_{2,x}, t_{1,y} \leq t_y \leq t_{2,y}, t_x \leq t_y\}$ contains a non-empty open set and the previous proof still works.

Appendix 5: *Proof of Theorem 3*

We first show that the statistics $n^{-2}U(\beta)$ in (34) has a positive limiting deterministic function in β that is minimized at $\beta = \beta^*$. Let $\tilde{U}(\beta)$ denote the statistic $U(\beta)$ in (34) with the weight function W_z replaced by its deterministic limit $\tilde{W}_z$. That is,

$$\tilde{U}(\beta) = \sum_z \sum_{i<j} \tilde{W}_z(\tilde{X}_{ij}, \tilde{Y}_{ij}) D_{ij}(z)[\Delta_{ij} - \eta(z'\beta)]^2.$$

Then

$$\begin{aligned} n^{-2}|U(\beta) - \tilde{U}(\beta)| &\leq n^{-2} \textstyle\sum_z \sum_{i<j} |\tilde{W}_z(\tilde{X}_{ij}, \tilde{Y}_{ij}) - W_z(\tilde{X}_{ij}, \tilde{Y}_{ij})| D_{ij}(z) \\ &\leq \tfrac{1}{2} \sup_{z,u,v} |\tilde{W}_z(u,v) - W_z(u,v)| \end{aligned}$$

which by assumption (c) converges to zero in probability. Therefore we only need to show that the limit of $\tilde{U}(\beta)$ in probability is positive and is minimized at $\beta = \beta^*$. Under the model assumptions, for those $Z_i = Z_j = z$ and $D_{ij}(z) = 1$, Δ_{ij} is a Bernoulli random variable with $\Pr(\Delta_{ij} = 1) = \eta(z'\beta^*)$. Hence $E\{[\Delta_{ij} - \eta(z'\beta)]^2 | Z_i = Z_j = z, D_{ij}(z) = 1\} = \eta(z'\beta^*)[1 - \eta(z'\beta^*)] + [\eta(z'\beta^*) - \eta(z'\beta)]^2$. So

$$\begin{aligned} & n_z^{-2} E\{\textstyle\sum_{i<j} \tilde{W}_z(\tilde{X}_{ij}, \tilde{Y}_{ij}) D_{ij}(z)[\Delta_{ij} - \eta(z'\beta)]^2\} \\ = \; & \{\eta(z'\beta^*)[1 - \eta(z'\beta^*)] + [\eta(z'\beta^*) - \eta(z'\beta)]^2\} n_z^{-2} \textstyle\sum_{i<j} \tilde{W}_z(\tilde{X}_{ij}, \tilde{Y}_{ij}) E[D_{ij}(z)]. \end{aligned}$$

By the law of large numbers, $n_z^{-2} \sum_{i<j} \tilde{W}_z(\tilde{X}_{ij}, \tilde{Y}_{ij}) E[D_{ij}(z)]$ converges to a positive constant. Let d_z denote the limit. Then

$$\begin{aligned} & E[n^{-2}\tilde{U}(\beta)] \\ = \; & \textstyle\sum_z (n/n_z)^{-2} n_z^{-2} E\{\sum_{i<j} \tilde{W}_z(\tilde{X}_{ij}, \tilde{Y}_{ij}) D_{ij}(z)[\Delta_{ij} - \eta(z'\beta)]^2\} \\ \to \; & \textstyle\sum_z c_z^2 \{\eta(z'\beta^*)[1 - \eta(z'\beta^*)] + [\eta(z'\beta^*) - \eta(z'\beta)]^2\} d_z. \end{aligned}$$

Let $LU(\beta)$ denote the limiting function $\sum_z c_z^2 \{\eta(z'\beta^*)[1-\eta(z'\beta^*)] + [\eta(z'\beta^*) - \eta(z'\beta)]^2\} d_z$. This expectation $LU(\beta)$ is minimized if and only if $\eta(z'\beta^*) = \eta(z'\beta)$ for all $z \in \mathcal{Z}$. By condition (a), that means $\beta = \beta^*$. By the law of large numbers, we can see that $n^{-2}\tilde{U}(\beta)$ converges to its expectation. Hence $n^{-2}U(\beta)$ also converges to $LU(\beta)$ which is uniquely minimized by $\beta = \beta^*$.

We notice that when $|\beta| \to \infty$, $\eta(z'\beta) \to 0$ or 1 for each non-zero z value. Therefore $\sum_z c_z^2 [\eta(z'\beta^*) - \eta(z'\beta)]^2 d_z$ is uniformly bounded above an constant ϵ outside a neighborhood of $\beta = \beta^*$ (when $|\beta|$ is big). Therefore the minimizer to $U(\beta)$ in (34) can only occurs in the neighborhood of $\beta = \beta^*$ in probability for large n. Then the uniform convergence of $n^{-2}U(\beta)$ within this neighborhood implies the consistency of $\hat{\beta}$.

Let $u(\beta)$ denote the gradient of $U(\beta)$. That is, $u(\beta) = \frac{\partial U(\beta)}{\partial \beta} = \left(\frac{\partial U(\beta)}{\partial \beta_k}\right)_{(p+1)\times 1}$, where

$$\frac{\partial U(\beta)}{\partial \beta_k} = \sum_z \sum_{i<j} 2W_z(\tilde{X}_{ij}, \tilde{Y}_{ij}) D_{ij}(z)[\Delta_{ij} - \eta(z'\beta)](-\frac{exp(\beta_0 + ... + \beta_p Z_p) Z_k}{(1 + exp(\beta_0 + ... + \beta_p Z_p))^2}).$$

The local minimizer of $U(\beta)$ also solves $u(\beta) = 0$. So without loss of generality, we can take $\hat{\beta}$ as a consistent root of $u(\beta) = 0$. Let $\tilde{u}(\beta)$ denote $u(\beta)$ with W_z replaced by $\tilde{W}_z$. Under model (31), $E[\tilde{u}(\beta^*)] = 0$. By Taylor expansion:

$$\sqrt{n}(\hat{\beta} - \beta^*) \;\; = \;\; (-\frac{1}{n^2}\frac{\partial u(\beta)}{\partial \beta}|_{\beta=\beta^*})^{-1} n^{-3/2} u(\beta^*) + o_p(1)$$

$$= \quad I^{-1}\{n^{-3/2}u(\beta^*)\} + o_p(1). \qquad \text{(A.7)}$$

Straightforward calculations similar to those in Fine and Jiang (2000) show that

$$n^{-3/2}u(\beta^*) = n^{-3/2}\tilde{u}(\beta^*) + o_p(1) = \left(n^{-3/2}\sum_z\sum_{i<j} Q^{(k)}_{ijz}\right)_{(p+1)\times 1} + o_p(1),$$

where $Q^{(k)}_{ijz} = 2\tilde{W}_z(\tilde{X}_{ij}, \tilde{Y}_{ij})D_{ij}(z)[\Delta_{ij} - \eta(z'\beta^*)](-\frac{exp(\beta^*_0+...+\beta^*_p Z_p)Z_k}{(1+exp(\beta^*_0+...+\beta^*_p Z_p))^2})$. By the central limit theorem for U-statistic and Slutsky's theorem: $\sqrt{n}(\hat{\beta}-\beta^*)$ converges in distribution to a multivariate normal distribution with variance Σ which is consistently estimated by $\hat{\Sigma} = \hat{I}^{-1}\hat{J}(\hat{I}^{-1})'$, where

$$\hat{I} = \left(-\frac{1}{n^2}\frac{\partial^2 U(\beta)}{\partial\beta_k\partial\beta_l}|_{\beta=\hat{\beta}}\right)_{(p+1)\times(p+1)}, \hat{J} = \left(\hat{J}_{ij}\right)_{(p+1)\times(p+1)},$$

$$\hat{J}_{ij} = n^{-3}\sum_z[2\sum_{k<l<m}(\hat{Q}^{(i)}_{klz}\hat{Q}^{(j)}_{kmz} + \hat{Q}^{(i)}_{klz}\hat{Q}^{(j)}_{lmz} + \hat{Q}^{(i)}_{lmz}\hat{Q}^{(j)}_{kmz}) + \sum_{k<l}(\hat{Q}^{(i)}_{klz}\hat{Q}^{(j)}_{klz})],$$

$$\frac{\partial^2 U(\beta)}{\partial\beta_k\partial\beta_l} = \sum_z\sum_{i<j} 2W_z(\tilde{X}_{ij}, \tilde{Y}_{ij})D_{ij}(z)\frac{exp(2z'\beta)Z_kZ_l + (\Delta_{ij} - \eta(z'\beta))(exp(3z'\beta) - exp(z'\beta))Z_kZ_l}{(1+exp(z'\beta))^4}.$$

Appendix 6: *Asymptotic properties of the Test Statistic T*

Under the assumption, $\theta^Z_*(s,t) = \exp(Z'\beta)$, the distributions of $\hat{\beta}_{W_{z,1}}$ and $\hat{\beta}_{W_{z,2}}$ are centered around the same β^*. By the results of Appendix 5, we have

$$\sqrt{n}(\hat{\beta}_{W_{z,1}} - \hat{\beta}_{W_{z,2}}) \quad = \quad n^{-3/2}\sum_z\sum_{i<j}(I_1^{-1}\tilde{Q}_{1,ijz} - I_2^{-1}\tilde{Q}_{2,ijz}) + o_p(1)$$

$$= \quad n^{-3/2}\sum_z\sum_{i<j}\tilde{Q}^*_{ijz} + o_p(1),$$

where $\tilde{Q}^*_{ijz} = \left(Q^{*(k)}_{ijz}\right)_{(p+1)\times 1} = I_1^{-1}\tilde{Q}_{1,ijz} - I_2^{-1}\tilde{Q}_{2,ijz}$, $\tilde{Q}_{m,ijz} = \left(Q^{(k)}_{m,ijz}\right)_{(p+1)\times 1}$, I_m is I with W_z replaced by $W_{z,m}$, and $Q^{(k)}_{m,ijz}$ is $Q^{(k)}_{ijz}$ with $\tilde{W}_z$ replaced by $\tilde{W}_{z,m}$ $(m = 1, 2)$. By the central limit theorem for U-statistic and Slutsky's theorem: $\sqrt{n}(\hat{\beta}_{W_{z,1}} - \hat{\beta}_{W_{z,2}})$

converges in distribution to a mean-zero multivariate normal distribution with variance Γ which can be consistently estimated by $\hat{\Gamma} = \left(\hat{\Gamma}_{ij}\right)_{(p+1)\times(p+1)}$, where

$$\hat{\Gamma}_{ij} = n^{-3}\sum_{z}[2\sum_{k<l<m}(\hat{Q}^{*(i)}_{klz}\hat{Q}^{*(j)}_{kmz} + \hat{Q}^{*(i)}_{klz}\hat{Q}^{*(j)}_{lmz} + \hat{Q}^{*(i)}_{lmz}\hat{Q}^{*(j)}_{kmz}) + \sum_{k<l}(\hat{Q}^{*(i)}_{klz}\hat{Q}^{*(j)}_{klz})].$$

Therefore, $T = n(\hat{\beta}_{W_{z,1}} - \hat{\beta}_{W_{z,2}})'\hat{\Gamma}^{-1}(\hat{\beta}_{W_{z,1}} - \hat{\beta}_{W_{z,2}})$ converges in distribution to χ^2_{p+1}.

Appendix 7: *Proofs of Intermediate Results*

To establish large-sample properties of θ, we need to prove the following intermediate results.

$\hat{F}_{1,z_k}(t)$ is uniformly and strongly consistent for $F_{1,z_k}(t)$ over $t \in [0, T_k]$ and $\sqrt{n}\{\hat{F}_{1,z_k}(t) - F_{1,z_k}(t)\}$ converges weakly to a mean-zero Gaussian process. (A.8)

$\xi_\theta(\hat{F}_{1,z_k}(t))$ is uniformly strongly consistent for $\xi_\theta(F_{1,z_k}(t))$, over both t and θ. (A.9)

$$\int_0^{t_{kj}} W_{kj}(t)\hat{g}(t,\theta)dt \xrightarrow{P} \int_0^{T_{kj}} W_{kj}(t)g(t,\theta)dt \quad \text{uniformly in} \quad \theta. \tag{A.10}$$

$$\frac{\partial}{\partial\theta_l}\int_0^{t_{kj}} W_{kj}(t)\hat{g}(t,\theta)dt \xrightarrow{P} \frac{\partial}{\partial\theta_l}\int_0^{T_{kj}} W_{kj}(t)g(t,\theta)dt \quad \text{locally uniformly at} \quad \theta = \theta_0. \tag{A.11}$$

$$U(\theta)/\sqrt{n} \xrightarrow{P} \bar{U}(\theta) \quad \text{uniformly in} \quad \theta. \tag{A.12}$$

$$\frac{\partial}{\partial\theta_l}U(\theta)/\sqrt{n} \xrightarrow{P} \frac{\partial}{\partial\theta_l}\bar{U}(\theta) \quad \text{locally uniformly at} \quad \theta = \theta_0. \tag{A.13}$$

$U(\theta_0)$ converges in distribution to a mean-zero normal random variable. (A.14)

Proof of (A.8):

Fine et al. (2001) proved this condition for the Clayton family and here we extend their proof to the AC family. By applying the results in Chapter 6 of Fleming and Harrington (1991), uniform and strong consistency of $\hat{F}_{z_k}(x,x)$ and $\hat{F}_{2,z_k}(x)$ can be shown under standard regularity conditions. By assumption (b) that $H_{z_k}(u,v,\alpha)$ has bounded derivatives, a continuous mapping theorem gives uniform and strong consistency of $\hat{F}_{1,z_k}(t) = H_{z_k}(\hat{F}_{z_k}(x,x), \hat{F}_{2,z_k}(x), \hat{\alpha}_k)$ to $F_{1,z_k}(t)$.

To establish weak convergence of $\sqrt{n_k}\{\hat{F}_{1,z_k}(t) - F_{1,z_k}(t)\}$, consider the following martingale representation of the one-dimensional processes:

$$\sqrt{n_k}\{\hat{F}_{z_k}(t,t) - F_{z_k}(t,t)\} = -F_{z_k}(t,t)\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n}\int_0^t \frac{1}{\pi_{z_k}(u)}dM_{z_k,i}(u) + o_p(1), \quad (A.15)$$

$$\sqrt{n_k}\{\hat{F}_{2,z_k}(t) - F_{2,z_k}(t)\} = -F_{2,z_k}(t)\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n}\int_0^t \frac{1}{\pi_{2,z_k}(u)}dM_{2,z_k,i}(u) + o_p(1), \quad (A.16)$$

where $\pi_{z_k}(t)$ and $\pi_{2,z_k}(t)$ are limits of

$$\hat{\pi}_{z_k}(t) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \geq t, Z_i = z_k),\ \hat{\pi}_{2,z_k}(t) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \geq t, Z_i = z_k), \quad (A.17)$$

respectively,

$$M_{z_k,i}(t) = I(X_i \leq t, \tilde{\delta}_i = 1, Z_i = z_k) - \int_0^t I(X_i \geq u, Z_i = z_k)d\Lambda_{z_k}(u),$$
$$M_{2,z_k,i}(t) = I(Y_i \leq t, \delta_{2,i} = 1, Z_i = z_k) - \int_0^t I(Y_i \geq u, Z_i = z_k)d\Lambda_{2,z_k}(u)$$

which are martingales defined with respect to appropriate filtrations, $\tilde{\delta}_i = \delta_{1,i} + \delta_{2,i} - \delta_{1,i}\delta_{2,i}$, $\Lambda_{z_k}(u)$ and $\Lambda_{2,z_k}(u)$ are the cumulative hazard functions of $T_1 \wedge T_2$ and T_2, respectively. Applying the functional and finite-dimensional delta methods, we can show that $\sqrt{n_k}\{\hat{F}_{1,z_k}(t) - F_{1,z_k}(t)\}$ is asymptotically equivalent to

$$\begin{aligned} J(t) = \quad & H_1(F_{z_k}(t,t), F_{2,z_k}(t), \alpha_k)\sqrt{n_k}\{\hat{F}_{z_k}(t) - F_{z_k}(t)\} \\ & + H_2(F_{z_k}(t,t), F_{2,z_k}(t), \alpha_k)\sqrt{n_k}\{\hat{F}_{2,z_k}(t) - F_{2,z_k}(t)\} \\ & + H_3(F_{z_k}(t,t), F_{2,z_k}(t), \alpha_k)\sqrt{n_k}\{\hat{\alpha}_{z_k} - \alpha_{z_k}\}, \end{aligned} \quad (A.18)$$

where

$$H_1(u,v,\alpha) = \partial H_{z_k}(u,v,\alpha)/\partial u = \phi'_{z_k,\alpha}(u)/\phi'_{z_k,\alpha}[\phi^{-1}_{z_k,\alpha}\{\phi_{z_k,\alpha}(u) - \phi_{z_k,\alpha}(v)\}], \quad (A.19)$$

$$H_2(u,v,\alpha) = \partial H_{z_k}(u,v,\alpha)/\partial v = \phi'_{z_k,\alpha}(v)/\phi'_{z_k,\alpha}[\phi^{-1}_{z_k,\alpha}\{\phi_{z_k,\alpha}(u) - \phi_{z_k,\alpha}(v)\}], \quad (A.20)$$

$$H_3(u,v,\alpha) = \partial H_{z_k}(u,v,\alpha)/\partial \alpha = [\phi^{(1)}_{z_k,\alpha}(u) + \phi^{(1)}_{z_k,\alpha}(v)]/\phi^{(1)}_{z_k,\alpha}[\phi^{-1}_{z_k,\alpha}\{\phi_{z_k,\alpha}(u) - \phi_{z_k,\alpha}(v)\}]. \quad (A.21)$$

Here $\phi'_{z_k,\alpha}(t) = \partial\phi_{z_k,\alpha}(t)/\partial t$ and $\phi^{(1)}_{z_k,\alpha}(t) = \partial\phi_{z_k,\alpha}(t)/\partial\alpha$. The third term in (A.18) is asymptotically normal (Wang, 2003), and is naturally tight because it is time-independent.

Using (A.15) and (A.16), the first two terms in (A.18) are asymptotically equivalent to sum of tight martingale integrals. The multivariate central limit theorem gives joint asymptotic normality of $\{J(t_1), ..., J(t_s)\}$ for any finite collection of times $(t_1, ..., t_s)$. Since $J(t)$ is sum of tight processes, it is also tight. By finite-dimensional convergence and the tightness property, $\sqrt{n_k}\{\hat{F}_{1,z_k}(t) - F_{1,z_k}(t)\}$ converges weakly to a sum of mean-zero Gaussian processes which is also a mean-zero Gaussian process.

Proof of (A.9): The result in (A.9) is a corollary of (A.8). For the first case that $\xi_\theta(F)(t) = F[h^{-1}\{h(t) + \theta\}]$, it follows that

$$\begin{aligned} & \sup_{-\infty<t<\infty} |\xi_\theta(\hat{F}_{1,z_k})(t) - \xi_\theta(F_{1,z_k})(t)| \\ = & \sup_{-\infty<t<\infty} |\hat{F}_{1,z_k}(h^{-1}\{h(t)+\theta\}) - F_{1,z_k}(h^{-1}\{h(t)+\theta\})| \\ = & \sup_{-\infty<t<\infty} |\hat{F}_{1,z_k}(t) - F_{1,z_k}(t)|. \end{aligned}$$

For the second case that $\xi_\theta(F)(t) = F_\varepsilon\,[F_\varepsilon^{-1}\{F(t)\} + \theta]$, we have $\xi'_\theta(F)(t) = f_\varepsilon\,[F_\varepsilon^{-1}\{F(t)\} + \theta]$ where f_ε is the density for F_ε. So the boundedness of f_ε in assumption (e2) gives a uniform bound for $\xi'_\theta(F)(t)$. Hence the uniform strong convergence of $\hat{F}_{1,z_k}(t)$ to $F_{1,z_k}(t)$ implies the uniform strong convergence of $\xi_\theta(\hat{F}_{1,z_k}(t))$ to $\xi_\theta(F_{1,z_k}(t))$. In both cases, the uniform (in t) and strong convergence of $\hat{F}_{1,z_k}(t)$ to $F_{1,z_k}(t)$ implies the strong convergence of $\xi_\theta(\hat{F}_{1,z_k}(t))$ to $\xi_\theta(F_{1,z_k}(t))$ uniformly in both t and θ.

Proof of (A.10): It follows that

$$\begin{aligned} & |\int_0^{t_{kj}} W_{kj}(t)[\xi_{z_{kj}^T\theta}(\hat{F}_{1,z_k})(t) - \hat{F}_{1,z_j}(t)]dt - \int_0^{T_{kj}} W_{kj}(t)[\xi_{z_{kj}^T\theta}(F_{1,z_k})(t) - F_{1,z_j}(t)]dt| \\ \leq & \int_0^{T_{kj}} W_{kj}(t)|\xi_{z_{kj}^T\theta}(\hat{F}_{1,z_k})(t) - \xi_{z_{kj}^T\theta}(F_{1,z_k})(t)|dt + \int_0^{T_{kj}} W_{kj}(t)|\hat{F}_{1,z_j}(t) - F_{1,z_j}(t)|dt \\ & + \int_{t_{kj}}^{T_{kj}} W_{kj}(t)|\xi_{z_{kj}^T\theta}(\hat{F}_{1,z_k})(t) - \hat{F}_{1,z_j}(t)|dt. \end{aligned}$$

By (A.8) and (A.9), $|\hat{F}_{1,z_j}(t) - F_{1,z_j}(t)|$ and $|\xi_{z_{kj}^T\theta}(\hat{F}_{1,z_k})(t) - \xi_{z_{kj}^T\theta}(F_{1,z_k})(t)|$ converge uniformly and strongly to zero. Together with the boundedness of $W_{kj}(t)$ in assumption (d) and $T_{kj} < \infty$, we can show that the first two terms on the right-hand side converge uniformly and strongly to zero. Since $W_{kj}(t)$ is bounded, $|W_{kj}(t)[\xi_\theta(\hat{F}_{1,z_k}(t)) - \hat{F}_{1,z_j}(t)]| \leq 2|W_{kj}(t)|$ is also bounded uniformly in θ. This together with the fact that $t_{kj} \xrightarrow{P} T_{kj}$ implies the uniform strong convergence of the third term to zero. The above results prove (A.10).

Proof of (A.11): We can write

$$\frac{\partial}{\partial\theta_l}\int_0^{t_{kj}} W_{kj}(t)\hat{g}(t,\theta)dt = \frac{\partial}{\partial\theta_l}\int_0^{t_{kj}} W_{kj}(t)\xi_{z_{kj}^T\theta}(\hat{F}_{1,z_k})(t)dt$$

and

$$\frac{\partial}{\partial\theta_l}\int_0^{T_{kj}} W_{kj}(t)g(t,\theta)dt = \frac{\partial}{\partial\theta_l}\int_0^{T_{kj}} W_{kj}(t)\xi_{z_{kj}^T\theta}(F_{1,z_k})(t)dt.$$

For the first case that $\xi_\theta(F)(t) = F[h^{-1}\{h(t)+\theta\}]$, let $f_{1,z_k}(t)$ denote the density of $F_{1,z_k}(t)$ and let $z_{kj}^T = (z_{kj,1}, ...z_{kj,p})$. Then

$$\begin{aligned} & \tfrac{\partial}{\partial\theta_l}\textstyle\int_0^{T_{kj}} W_{kj}(t)\xi_{z_{kj}^T\theta}(F_{1,z_k})(t)dt \\ = & \tfrac{\partial}{\partial\theta_l}\textstyle\int_0^{T_{kj}} W_{kj}(t)F_{1,z_k}(h^{-1}\{h(t)+z_{kj}^T\theta\})dt \\ = & \textstyle\int_0^{T_{kj}} W_{kj}(t)f_{1,z_k}(h^{-1}\{h(t)+z_{kj}^T\theta\})/h'\{h(t)+z_{kj}^T\theta\}z_{kj,l}dt \\ = & \textstyle\int_0^{T_{kj}} \tilde{W}_{kj}(t)z_{kj,l}dF_{1,z_k}(h^{-1}\{h(t)+z_{kj}^T\theta\}). \end{aligned}$$

Performing a change of variable for $t* = h^{-1}\{h(t)+z_{kj}^T\theta\}$, the above partial derivative equals

$$\int_{h^{-1}\{h(0)+z_{kj}^T\theta\}}^{h^{-1}\{h(T_{kj})+z_{kj}^T\theta\}} \tilde{W}_{kj}(h^{-1}\{h(t*)-z_{kj}^T\theta\})z_{kj,l}dF_{1,z_k}(t*).$$

The next goal is to derive a similar expression for $\frac{\partial}{\partial\theta_l}\int_0^{t_{kj}} W_{kj}(t)\hat{F}_{1,z_k}(h^{-1}\{h(t)+z_{kj}^T\theta\})dt$. The challenge is that $\hat{F}_{1,z_k}(t)$ is a step function and hence not differentiable. Here we adopt a different approach to show the convergence. Let $t_{(1)} \le \ldots \le t_{(n_{kj})}$ be the observed ordered times of X in the pooled sample of $Z = z_k$ or $Z = z_j$, and set $t_{(0)} = 0$. Also let m_1 denote the smallest integer such that $t_{(m_1)} \ge h^{-1}\{h(0)+z_{kj}^T\theta\}$, and let m_2 denote the largest integer such that $t_{(m_2)} \le h^{-1}\{h(t_{ij})+z_{kj}^T\theta\}$. Then using the fact that $\hat{F}_{1,z_k}(t)$ is a step function, we have

$$\begin{aligned} & \tfrac{\partial}{\partial\theta_l}\textstyle\int_0^{t_{kj}} W_{kj}(t)\hat{F}_{1,z_k}(h^{-1}\{h(t)+z_{kj}^T\theta\})dt \\ = & \textstyle\sum_{i=m_1+1}^{m_2}\tfrac{\partial}{\partial\theta_l}\int_{h^{-1}\{h(t_{(i-1)})-z_{kj}^T\theta\}}^{h^{-1}\{h(t_{(i)})-z_{kj}^T\theta\}} W_{kj}(t)\hat{F}_{1,z_k}(t_{(i)})dt \\ & +\tfrac{\partial}{\partial\theta_l}\textstyle\int_0^{h^{-1}\{h(t_{(m_1)})-z_{kj}^T\theta\}} W_{kj}(t)\hat{F}_{1,z_k}(t_{(m_1)})dt + \tfrac{\partial}{\partial\theta_l}\int_{h^{-1}\{h(t_{(m_2)})-z_{kj}^T\theta\}}^{t_{kj}} W_{kj}(t)\hat{F}_{1,z_k}(t_{(m_2+1)})dt \\ = & \textstyle\sum_{i=m_1+1}^{m_2}\hat{F}_{1,z_k}(t_{(i)})[-\frac{W_{kj}(h^{-1}\{h(t_{(i)})-z_{kj}^T\theta\})z_{kj,l}}{h'\{h^{-1}\{h(t_{(i)})-z_{kj}^T\theta\}\}} + \frac{W_{kj}(h^{-1}\{h(t_{(i-1)})-z_{kj}^T\theta\})z_{kj,l}}{h'\{h^{-1}\{h(t_{(i-1)})-z_{kj}^T\theta\}\}}] \\ & -\hat{F}_{1,z_k}(t_{(m_1)})\tfrac{W_{kj}(h^{-1}\{h(t_{(m_1)})-z_{kj}^T\theta\})z_{kj,l}}{h'\{h^{-1}\{h(t_{(m_1)})-z_{kj}^T\theta\}\}} + \hat{F}_{1,z_k}(t_{(m_2+1)})\tfrac{W_{kj}(h^{-1}\{h(t_{(m_2)})-z_{kj}^T\theta\})z_{kj,l}}{h'\{h^{-1}\{h(t_{(m_2)})-z_{kj}^T\theta\}\}} \\ = & \textstyle\sum_{i=m_1}^{m_2}[\hat{F}_{1,z_k}(t_{(i+1)})-\hat{F}_{1,z_k}(t_{(i)})]\frac{W_{kj}(h^{-1}\{h(t_{(i)})-z_{kj}^T\theta\})z_{kj,l}}{h'\{h^{-1}\{h(t_{(i)})-z_{kj}^T\theta\}\}} \\ = & \textstyle\int_{h^{-1}\{h(0)+z_{kj}^T\theta\}}^{h^{-1}\{h(t_{kj})+z_{kj}^T\theta\}} \tilde{W}_{kj}(h^{-1}\{h(t*)-z_{kj}^T\theta\})z_{kj,l}d\hat{F}_{1,z_k}(t*). \end{aligned}$$

In the above derivation, if $m_2 = n_{kj}$, $t_{(m_2+1)}$ can be any time greater than $t_{(n_{kj})}$ and $\hat{F}_{1,z_k}(t_{(m_2+1)}) = 0$.

Hence we obtain the inequality

$$\begin{aligned} & |\tfrac{\partial}{\partial\theta_l}\int_0^{T_{kj}} W_{kj}(t)\xi_{z_{kj}^T\theta}(F_{1,z_k})(t)dt - \tfrac{\partial}{\partial\theta_l}\int_0^{t_{kj}} W_{kj}(t)\hat{F}_{1,z_k}(h^{-1}\{h(t)+z_{kj}^T\theta\})dt| \\ \leq \; & |\int_{h^{-1}\{h(0)+z_{kj}^T\theta\}}^{h^{-1}\{h(T_{kj})+z_{kj}^T\theta\}} \tilde{W}_{kj}(h^{-1}\{h(t*)-z_{kj}^T\theta\})z_{kj,l}d[F_{1,z_k}(t*)-\hat{F}_{1,z_k}(t*)]| \\ & +|\int_{h^{-1}\{h(t_{kj})+z_{kj}^T\theta\}}^{h^{-1}\{h(T_{kj})+z_{kj}^T\theta\}} \tilde{W}_{kj}(h^{-1}\{h(t*)-z_{kj}^T\theta\})z_{kj,l}d\hat{F}_{1,z_k}(t*)|. \end{aligned} \tag{A.22}$$

Now we are ready to show the convergence. Consider the convergence over a compact neighborhood $D = \{\theta : \|\theta - \theta_0\| \leq r\}$. Then $D^* = \{s : \; s = h^{-1}\{h(t) - z_{kj}^T\theta\}, \; t \in [0, T_{kj}], \theta \in D\}$ is also a compact set. Because $\tilde{W}_{kj}(s)$ is continuous (since it is differentiable by assumption (e1)), $\tilde{W}_{kj}(s)$ is bounded on D^*. This local boundedness of $\tilde{W}_{kj}(h^{-1}\{h(t*) - z_{kj}^T\theta\})z_{kj,l}$ as well as the fact that $t_{kj} \xrightarrow{P} T_{kj}$ implies the local uniform convergence to zero of the second term on the right-hand side of (A.22). The first term on the right-hand side of (A.22) is

$$\begin{aligned} & |z_{kj,l}\int_{h^{-1}\{h(0)+z_{kj}^T\theta\}}^{h^{-1}\{h(T_{kj})+z_{kj}^T\theta\}} \tilde{W}_{kj}(h^{-1}\{h(t*)-z_{kj}^T\theta\})d[F_{1,z_k}(t*)-\hat{F}_{1,z_k}(t*)]| \\ = \; & |z_{kj,l}\int_{t*=h^{-1}\{h(0)+z_{kj}^T\theta\}}^{h^{-1}\{h(T_{kj})+z_{kj}^T\theta\}} \int_{u=0}^{h^{-1}\{h(t*)-z_{kj}^T\theta\}} d\tilde{W}_{kj}(u)d[F_{1,z_k}(t*)-\hat{F}_{1,z_k}(t*)]| \\ = \; & |z_{kj,l}\int_{u=0}^{T_{kj}}\{[F_{1,z_k}(h^{-1}\{h(T_{kj})+z_{kj}^T\theta\})-\hat{F}_{1,z_k}(h^{-1}\{h(T_{kj})+z_{kj}^T\theta\})] \\ & \quad -[F_{1,z_k}(h^{-1}\{h(u)+z_{kj}^T\theta\})-\hat{F}_{1,z_k}(h^{-1}\{h(u)+z_{kj}^T\theta\})]\}d\tilde{W}_{kj}(u)| \\ \leq \; & 2|z_{kj,l}|\sup_t|F_{1,z_k}(t)-\hat{F}_{1,z_k}(t)|\int_{u=0}^{T_{kj}}|d\tilde{W}_{kj}(u)|. \end{aligned}$$

The last quantity has a finite upper bound for large n due to the absolute integrability of $\tilde{W}'_{kj}(u)$ by assumption (e1) and the uniform strong convergence of $\hat{F}_{1,z_k}(t)$ to $F_{1,z_k}(t)$. This finite bound also shows that the integral in the next to last expression above is in fact absolutely integrable, and justifies the changing the order of integration in the double integration. Also, this upper bound converges to zero uniformly. Therefore local uniform convergence (at $\theta = \theta_0$) of $\frac{\partial}{\partial\theta_l}\int_0^{t_{kj}} W_{kj}(t)\hat{g}(t,\theta)dt$ to $\frac{\partial}{\partial\theta_l}\int_0^{T_{kj}} W_{kj}(t)g(t,\theta)dt$ can be established for the first case.

For the second case that $\xi_\theta(F)(t) = F_\varepsilon\left[F_\varepsilon^{-1}\{F(t)\}+\theta\right]$, then $\xi'_\theta(F)(t) = f_\varepsilon\left[F_\varepsilon^{-1}\{F(t)\}+\theta\right]$ where f_ε is the density for F_ε. Applying Taylor series expansions and using the fact that

$t_{kj} \xrightarrow{P} T_{kj}$, it follows that

$$\begin{aligned} & \frac{\partial}{\partial\theta_l}\int_0^{t_{kj}} W_{kj}(t)\xi_{z_{kj}^T\theta}(\hat{F}_{1,z_k})dt - \frac{\partial}{\partial\theta_l}\int_0^{T_{kj}} W_{kj}(t)\xi_{z_{kj}^T\theta}(F_{1,z_k})(t)dt \\ = \; & \int_0^{T_{kj}} W_{kj}(t)z_{kj,l}f_\varepsilon'[F_\varepsilon^{-1}\{F_{1,z_k}(t)\} + z_{kj}^T\theta][\hat{F}_{1,z_k}(t) - F_{1,z_k}(t)]dt + o_p(1). \end{aligned}$$

The above term converges uniformly strongly to zero due to the boundedness of $W_{kj}(t)$ and $f_\varepsilon'(t)$ (see assumptions (d) and (e2)) and the uniform strong convergence of $\hat{F}_{1,z_k}(t)$ to $F_{1,z_k}(t)$. Therefore, this completes the proof for locally uniformly convergence (at $\theta = \theta_0$) of $\frac{\partial}{\partial\theta_l}\int_0^{t_{kj}} W_{kj}(t)\hat{g}(t,\theta)dt$ to $\frac{\partial}{\partial\theta_l}\int_0^{T_{kj}} W_{kj}(t)g(t,\theta)dt$ in the second case.

Proof of (A.12): Recall that

$$U(\theta)/\sqrt{n} = \sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}}\int_0^{t_{kj}} W_{kj}(t)\hat{g}_{kj}(t,\theta)dt.$$

By (A.10), $\int_0^{t_{kj}} W_{kj}(t)\hat{g}(t,\theta)dt$ converges uniformly (in θ) and strongly to $\int_0^{T_{kj}} W_{kj}(t)g(t,\theta)dt$. By assumption (a), $\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}}$ converges to $\sqrt{\frac{c_kc_j}{c_k+c_j}}$. These together with the boundedness of $w_0(z_{kj}^T\theta)z_{kj}$ in assumption (d) implies that the above expression converges uniformly (in θ) and strongly to

$$\bar{U}(\theta) = \sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{c_kc_j}{c_k+c_j}}\int_0^{T_{kj}} W_{kj}(t)g_{kj}(t,\theta)dt.$$

Proof of (A.13): One can write

$$\begin{aligned} \frac{\partial}{\partial\theta_l}U(\theta)/\sqrt{n} = \; & \sum_{k<j} w_0'(z_{kj}^T\theta)z_{kj,l}z_{kj}\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}}\int_0^{t_{kj}} W_{kj}(t)\hat{g}_{kj}(t,\theta)dt \\ & + \sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{(n_k/n)(n_j/n)}{(n_k/n)+(n_j/n)}}\frac{\partial}{\partial\theta_l}\int_0^{t_{kj}} W_{kj}(t)\hat{g}_{kj}(t,\theta)dt. \end{aligned}$$

From assumption (d) $w_0'(t)$ is continuous, thus $w_0(z_{kj}^T\theta)$ is locally bounded at $\theta = \theta_0$. Similar to the proof of (A.12), using (A.10), assumption (b) and the local boundedness of $w_0'(z_{kj}^T\theta)$, the first term converges uniformly (locally at $\theta = \theta_0$) to

$$\sum_{k<j} w_0'(z_{kj}^T\theta)z_{kj,l}z_{kj}\sqrt{\frac{c_kc_j}{c_k+c_j}}\int_0^{T_{kj}} W_{kj}(t)g_{kj}(t,\theta)dt.$$

For the second term, note that $\frac{\partial}{\partial\theta_l}\int_0^{t_{kj}} W_{kj}(t)\hat{g}(t,\theta)dt$ converges strongly and locally uniformly (at $\theta = \theta_0$) to $\frac{\partial}{\partial\theta_l}\int_0^{T_{kj}} W_{kj}(t)g(t,\theta)dt$ by (A.11). This fact together with assumption (b) and the boundedness of $w_0(z_{kj}^T\theta)z_{kj}$ from assumption (e) implies that the second term locally uniformly strongly converges to

$$\sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\frac{c_kc_j}{c_k+c_j}}\frac{\partial}{\partial\theta_l}\int_0^{T_{kj}} W_{kj}(t)g_{kj}(t,\theta)dt.$$

Hence $\frac{\partial}{\partial\theta_l}U(\theta)/\sqrt{n}$ locally uniformly strongly converges to

$$\begin{aligned}\frac{\partial}{\partial\theta_l}\bar{U}(\theta) = &\sum_{k<j} w_0'(z_{kj}^T\theta)z_{kj,l}z_{kj}\sqrt{\tfrac{c_kc_j}{c_k+c_j}}\int_0^{T_{kj}} W_{kj}(t)g_{kj}(t,\theta)dt\\ &+\sum_{k<j} w_0(z_{kj}^T\theta)z_{kj}\sqrt{\tfrac{c_kc_j}{c_k+c_j}}\frac{\partial}{\partial\theta_l}\int_0^{T_{kj}} W_{kj}(t)g_{kj}(t,\theta)dt.\end{aligned}$$

Proof of (A.14):

We want to show that $U(\theta_0)$ is asymptotic normal with mean zero. As in (19),

$$U(\theta_0) = \sum_{k<j} w_0(z_{kj}^T\theta_0)z_{kj}\sqrt{\frac{n_kn_j}{n_k+n_j}}\{\int_0^{t_{kj}} W_{kj}(t)\hat{g}_{kj}(t,\theta_0)dt\}.$$

Since $g_{kj}(t,\theta_0)\equiv 0$, it follows that

$$\begin{aligned}&U(\theta_0)\\ =\ &\sum_{k<j} w_0(z_{kj}^T\theta_0)z_{kj}\sqrt{\tfrac{n_kn_j}{n_k+n_j}}\int_0^{t_{kj}} W_{kj}(t)[\hat{g}_{kj}(t,\theta_0)-g_{kj}(t,\theta_0)]dt\\ =\ &\sum_{k<j} w_0(z_{kj}^T\theta_0)z_{kj}\sqrt{\tfrac{n_kn_j}{n_k+n_j}}\int_0^{t_{kj}} W_{kj}(t)[\xi_{z_{kj}^T\theta_0}(\hat{F}_{1,z_k})(t)-\xi_{z_{kj}^T\theta_0}(F_{1,z_k})(t)]dt\\ &-\sum_{k<j} w_0(z_{kj}^T\theta_0)z_{kj}\sqrt{\tfrac{n_kn_j}{n_k+n_j}}\int_0^{t_{kj}} W_{kj}(t)[\hat{F}_{1,z_j}(t)-F_{1,z_j}(t)]dt\\ =\ &\sum_{k<j} w_0(z_{kj}^T\theta_0)z_{kj}\sqrt{\tfrac{c_j}{c_k+c_j}}\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_k}[\xi_{z_{kj}^T\theta_0}(\hat{F}_{1,z_k})(t)-\xi_{z_{kj}^T\theta_0}(F_{1,z_k})(t)]dt\\ &-\sum_{k<j} w_0(z_{kj}^T\theta_0)z_{kj}\sqrt{\tfrac{c_k}{c_k+c_j}}\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_j}[\hat{F}_{1,z_j}(t)-F_{1,z_j}(t)]dt.\end{aligned} \tag{A.23}$$

Due to the boundedness of $w_0(z_{kj}^T\theta_0)z_{kj}$, it suffices to prove that each of $\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_j}[\hat{F}_{1,z_j}(t)-F_{1,z_j}(t)]dt$ and $\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_k}[\xi_{z_{kj}^T\theta_0}(\hat{F}_{1,z_k})(t)-\xi_{z_{kj}^T\theta_0}(F_{1,z_k})(t)]dt$ converges in distribution to a mean-zero normal random variable. Then each term in the sums on the right-hand side of (A.23) weakly in distribution to a mean-zero normal random variable, and hence $U(\theta_0)$ converges in distribution to a mean-zero normal random variable.

Now let us prove $\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_j}[\hat{F}_{1,z_j}(t)-F_{1,z_j}(t)]dt$ converges in distribution to a mean-zero normal random variable. As shown in the proof of (A.8), $\sqrt{n_j}[\hat{F}_{1,z_j}(t)-$

$F_{1,z_j}(t)]dt = J(t) + o_p(1)$ where $J(t)$ in (A.18) was shown to be a tight process. This together with the boundedness of $W_{kj}(t)$ and $t_{kj} \xrightarrow{P} T_{kj}$ implies that $\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_j}[\hat{F}_{1,z_j}(t) - F_{1,z_j}(t)]dt = \int_0^{T_{kj}} W_{kj}(t)J(t)dt$. Notice that

$$\int_0^{T_{kj}} W_{kj}(t)J(t)dt = \int_{t=0}^{T_{kj}} W_{kj}(t)\int_{u=0}^{t} dJ(u)dt = \int_{u=0}^{T_{kj}} W^*_{kj}(u)dJ(u)$$

where $W^*_{kj}(u) = \int_{t=u}^{T_{kj}} W_{kj}(t)dt$. Because $W_{kj}(t)$ is bounded and positive and $T_{kj} < \infty$, $W^*_{kj}(u)$ is also positive and bounded. Therefore, $\int_{u=0}^{T_{kj}} W^*_{kj}(u)dJ(u)$ converges in distribution to a mean-zero normal random variable.

The remaining part is to prove that the convergence of $\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_k}[\xi_{z_{kj}^T\theta_0}(\hat{F}_{1,z_k})(t) - \xi_{z_{kj}^T\theta_0}(F_{1,z_k})(t)]dt$ to a mean-zero normal random variable.

For the first case that $\xi_\theta(F)(t) = F[h^{-1}\{h(t) + \theta\}]$, by a change of variable for $t^* = h^{-1}\{h(t) + z_{kj}^T\theta_0\}$, we have

$$\begin{aligned} &\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_k}[\xi_{z_{kj}^T\theta_0}(\hat{F}_{1,z_k})(t) - \xi_{z_{kj}^T\theta_0}(F_{1,z_k})(t)]dt \\ = &\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_k}[\hat{F}_{1,z_k}[h^{-1}\{h(t) + z_{kj}^T\theta_0\}] - F_{1,z_k}[h^{-1}\{h(t) + z_{kj}^T\theta_0\}]]dt \\ = &\int_{h^{-1}\{h(0)+z_{kj}^T\theta_0\}}^{h^{-1}\{h(t_{kj})+z_{kj}^T\theta_0\}} \tilde{W}_{kj}(h^{-1}\{h(t^*) - z_{kj}^T\theta_0\})h'(t^*)\sqrt{n_k}[\hat{F}_{1,z_k}(t^*) - F_{1,z_k}(t^*)]dt^*. \end{aligned}$$

From assumption (e1) that $\tilde{W}_{kj}(t)$ and $h'(t)$ are continuous, $\tilde{W}_{kj}(t)$ is bounded on the compact set $[0, T_{kj}]$, and $h'(t)$ is bounded on the compact set $[h(0) + z_{kj}^T\theta_0, h(T_{kj}) + z_{kj}^T\theta_0]$. Using the boundedness of $\tilde{W}_{kj}(t)$, the boundedness of $h'(t)$ and the monotonicity of $h(t)$ (hence $h'(t)$ always positive or always negative), the new weight function $\tilde{W}_{kj}(h^{-1}\{h(t^*) - z_{kj}^T\theta_0\})h'(t^*) = \frac{W_{kj}(h^{-1}\{h(t^*)-z_{kj}^T\theta_0\})h'(t^*)}{h'(h^{-1}\{h(t^*)-z_{kj}^T\theta_0\})}$ in the above integral is also positive and bounded. Therefore, the same reasoning for proving the convergence of $\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_j}[\hat{F}_{1,z_j}(t) - F_{1,z_j}(t)]dt$ shows that $\int_{h^{-1}\{h(0)+z_{kj}^T\theta_0\}}^{h^{-1}\{h(t_{kj})+z_{kj}^T\theta_0\}} \tilde{W}_{kj}(h^{-1}\{h(t^*) - z_{kj}^T\theta_0\})h'(t^*)\sqrt{n_k}[\hat{F}_{1,z_k}(t^*) - F_{1,z_k}(t^*)]dt^*$ converges in distribution to a mean-zero normal random variable.

For the second case that $\xi_\theta(F)(t) = F_\varepsilon[F_\varepsilon^{-1}\{F(t)\} + \theta]$, then $\xi'_\theta(F)(t) = f_\varepsilon[F_\varepsilon^{-1}\{F(t)\} + \theta]$ where f_ε is the density for F_ε. Applying Taylor series expansions, it follows that

$$\sqrt{n_k}\{\xi_{z_{kj}^T\theta_0}(\hat{F}_{1,z_k}(t)) - \xi_{z_{kj}^T\theta_0}(F_{1,z_k}(t))\} \overset{a}{=} \xi'_{z_{kj}^T\theta_0}(F_{1,z_k})(t)\sqrt{n_k}\{\hat{F}_{1,z_k}(t) - F_{1,z_k}(t)\}.$$

Therefore

$$\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_k}[\xi_{z_{kj}^T\theta_0}(\hat{F}_{1,z_k})(t) - \xi_{z_{kj}^T\theta_0}(F_{1,z_k})(t)]dt$$

is asymptotically equivalent to

$$\int_0^{t_{kj}} W_{kj}(t)\xi'_{z_{kj}^T\theta_0}(F_{1,z_k})(t)\sqrt{n_k}[\hat{F}_{1,z_k}(t) - F_{1,z_k}(t)]dt.$$

Consider the new weight function $W_{kj}(t)\xi'_{z_{kj}^T\theta_0}(F_{1,z_k})(t)$ which is still positive and bounded. Hence the same reasoning for proving the convergence of $\int_0^{t_{kj}} W_{kj}(t)\sqrt{n_j}[\hat{F}_{1,z_j}(t)-F_{1,z_j}(t)]dt$ shows that $\int_0^{t_{kj}} W_{kj}(t)\xi'_{z_{kj}^T\theta_0}(F_{1,z_k})(t)\sqrt{n_k}[\hat{F}_{1,z_k}(t) - F_{1,z_k}(t)]dt$ converges in distribution to a mean-zero normal random variable. This completes the proof.

References

Anderson, J. E., Louis, T. A., Holm, N. V. and Harvald, B. (1992). Time-dependent Association Measures for Bivariate Survival Distributions. *Journal of the American Statistical Association* **87**, 641-650.

Chaieb, L., Rivest, L.-P. and Abdous, B. (2006). Estimating Survival under a Dependent Truncation. *Biometrika* **93**, 655-669.

Chang, S. H. (2000). A Two-Sample Comparison for Multiple Ordered Event Data. *Biometrics* **56**, 183-189.

Cheng, S. C., Wei, L. J. and Ying, Z. (1995). Analysis of Transformation Models with Censored Data. *Biometrika* **82**, 835-845.

Clayton, D. G. (1978). A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence *Biometrika* **65**, 141-151.

Day, R., Bryant, J. and Lefkopoulou, M. (1997). Adaptation of Bivariate Frailty Models for Prediction, with Application to Biological Markers as Prognostic Indicators. *Biometrika* **84**, 45-56.

Fine, J. P. and Jiang, H. (2000). On Association in a Copula with Time Transformations. *Biometrika* **87**, 559-571.

Fine, J. P., Jiang, H. and Chappell, R. (2001). On Semi-Competing Risks Data. *Biometrika* **88**, 4, 907-919.

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

Frank, M. J. (1979). On the Simultaneous Associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematicae* **19**, 194-226.

Genest, C. and Mackay, J. (1986). The Joy of Copulas: Bivariate Distributions With Uniform Marginals. *The American Statistician* vol. **40**, No. 4.

Genest, C. (1987). Frank's Family of Bivariate Distributions. *Biometrika* **74**, 3, 549-555.

Genest, C. and Rivest L. P. (1993). Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association* vol. **88**, No. 423.

Ghosh D. and Lin. D. Y. (2003). Semiparametric Analysis of Recurrent Events Data in the Presence of Dependent Censoring. *Biometrics* **59**, 877-885.

Ghosh, D. (2006). Semiparametric Inferences for Association with Semi-Competing Risks Data. *Statist. Med.* **25**, 2059-2070.

Gumbel, E. J. (1960). Bivariate Exponential Distributions. *Journal of the American Statistical Association* Vol. **55**, 698-707.

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

Kalbfleisch, J. D. and Lawless, J. F. (1989). Inference Based on Retrospective Ascertainment: An Analysis of the data on Transfusion-Related AIDS *Journal of the American Statistical Association* **84**, 360-372.

Klien, J. P. and Moeschberger, M. L. (2003) *Survival Analysis, Techniques for Censored and Truncated Data* 2nd Ed. New York: Springer.

Lee, A. J. (1990) *U-statistics : Theory and Practice.* New York & Basel/ Marcel Dekker.

Lehman, E.L. and Casella, G. (1998) *Theory of Point Estimation* 2nd Ed. New York: Springer.

Lin, D. Y., Robins, J. M. and Wei, L. J. (1996). Comparing Two Failure Time Distribution in The Presence of Dependent Censoring. *Biometrika* **83**, 2, 381-393.

Lin, D. Y. and Ying, Z. (2003). Semiparametric Regression Analysis of Longitudinal Data with Informative Drop-outs. *Biostatistics* **4**, 385-398.

Lin, J. S. and Wei, L. J. (1992). Linear Regression Analysis for Mulivariate Failure Time Observations. *Journal of the American Statistical Association* **92**, 1091-1098.

Louis, T. A. (1981). Non-Parametric Analysis of An Accelerated Failure Time Model. *Biometrika* **68**, 381-390.

Nelsen, Roger B. (2006) *An Introduction to Copulas* 2nd Ed. New York: Springer.

Oakes, D. (1982). A Model for Association in Bivariate Survival Data. *Journal of the Royal Statistical Society, Series B* **44**, 414-422.

Oakes, D. (1986). Semiparametric Inference in a Model for Association in Bivariate Survival Data. *Biometrika* **73**, 353-361.

Oakes, D. (1989). Bivariate Survival Models Induced by Frailties. *Journal of the American Statistical Association* Vol. **84**, No. 406.

Oakes, D. and Ritz, J. (2000). Regression in a Bivariate Copula Model. *Biometrika* **87**, 345-352.

Parzen, M.I., Wei L.J. and Ying Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika*, **81**, 341-350.

Peng, L. and Fine, J. P. (2006). Rank Estimation of Accelerated Lifetime Models with Dependent Censoring. *Journal of the American Statistical Association* Vol. **101**, 1085-1093.

Prentice, R. L. and Cai, J. (1992). Covariance and Survivor Function Estimation Using Censored Multivariate Failure Time Data. *Biometrika* **79**, 3, 495-512.

Shih, J. H. (1998). A Goodness-of Fit Test for Association in a Bivariate Survival Model. *Biometrika* **85**, 189-200.

Tsai, W.-Y. (1990). Testing the Assumption of Independence of Truncation Time and Failure Time. *Biometrika* **77**, 1, 169-177.

Tsai, W.-Y. and Crowley, J. (1998). A Note on Nonparametric Estimators of the Bivariate Survival Function under Univariate Censoring. *Biometrika* **85**, 3, 573-580.

Wang, W. and Wells, M. T. (2000). Model Selection and Semiparametric Inference for Bivariate Failure-Time Data. *Journal of the American Statistical Association* Vol. **95**, No. 449.

Wang, W. (2003). Estimating the Association Parameter for Copula Models under Dependent Censoring. *Journal of the Royal Statistical Society, Series B* **65**, 257-273.

Wei, L. J. and Gail, M. H. (1983). Nonparametric Estimation for a scale-Change with censored observations. *Journal of the American Statistical Association* Vol. **78**, 382-388.

Wei, L. J., Ying, Z. and Lin, D. Y. (1990). Linear Regression Analysis of Censored Survival Data Based on Rank Tests. *Biometrika* **77**, 845-851.