

國立交通大學
統計學研究所

碩士論文

Copula 模式之下 雙變元存活資料之統計推論

Statistical Inference for Bivariate
Survival Data Based on Copula Models

研究生：花文妤

指導教授：王維菁 教授

中華民國九十六年六月

Statistical Inference for Bivariate Survival Data Based on
Copula Models

研 究 生：花文好

Student : Hua, Wen-Yu

指 導 教 授：王維菁

Advisor : Weijing Wang

國 立 交 通 大 學

統 計 學 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Statistics College of Science

National Chiao Tung University

in Partial fulfillment of the Requirements

for the Degree of

Master

in

Statistics

June 2007

Hsinchu, Taiwan, Republic of China

中 華 民 國 九 十 六 年 六 月

Copula 模式之下 雙變元存活資料之統計推論

學生：花文好

指導教授：王維菁 博士

國立交通大學統計研究所



摘要

本論文主要是考慮在 Copula 模式之下，雙變元存活資料受限於右設限之半母數推論。目前存在許多估計相關性參數之半母數推論方法。我們比較三種推論方法針對不存在解釋變數之同質性資料，並利用迴歸的概念將現有的方法延伸至處理邊際異質的資料。最後，藉由模擬結果檢視上述方法之有限樣本的表現。

關鍵字：雙變元存活資料；Cox 比例風險模式；相關存活時間；Clayton 模式；Copula 模式；半母數推論；二階段估計式； 2×2 表格

Statistical Inference for Bivariate Survival Data


Based on Copula Models

Student: Hua, Wen-Yu

Adivsor: Dr. Weijing Wang

Institute of Statistics Notional Chiao Tung University

Abstract



The thesis considers semi-parametric inference based on Copula models for bivariate survival data subject to right censoring. There exist several semi-parametric inference approaches to estimating the association parameter. We examine and compare three approaches developed for homogeneous data in absence of covariates. Then we extend these methods to a regression setting that accounts for marginal heterogeneity explained by the covariates. Finite-sample performances are examined by simulations.

Key words: Bivariate survival data; Cox proportional hazard model; Correlated failure times; Clayton model; Copula model; Semi-parametric inference; Two-stage estimation; Two-by-two tables.

致謝

首先誠摯的感謝指導教授王維菁博士，老師悉心的教導使我得以一窺統計領域的深奧，不時的討論並指點我正確的方向，使我在這兩年中獲益匪淺。老師對學問的嚴謹更是我輩學習的典範。

兩年裡的日子，研究室裡共同的生活點滴，學術上的討論、言不及義的閒扯、讓人又愛又怕的宵夜、趕作業的革命情感、因為睡太晚而遮遮掩掩閃進教室.....，要特別感謝我碩士班同學們，有你們的陪伴讓兩年的研生活變得絢麗多彩。

感謝 Takashi 學長、進見學長、秋婷學姐們不厭其煩的指出我研究中的缺失，且總能在我迷惘時為我解惑，也感謝侑侑、怡君、小逼的幫忙，讓我順利走過這兩年。益銘、建威、俊睿、永在以及吳志文，與你們一同共進晚餐，是我準備碩士畢業及考試最開心的時光。

我男朋友、老莊在背後的默默支持更是我前進的動力，沒有老莊的體諒、包容，相信這兩年的生活將是很不一樣的光景。

最後，謹以此文獻給我摯愛的家人，謝謝你們給我這個機會支持我體諒我，讓我順利的完成碩士的學歷。

感謝這一路上曾經幫助過我的人，你們都是我生命中無法取代的貴人。

文妤

2007年6月 交大統計所

Table of Content

Chapter 1	Introduction	1
	1.1 Background	1
	1.2 Notations and Outline of the Thesis	1
Chapter 2	Literature Review	3
	2.1 Copula Models	3
	2.2 Archimedean Copula Models	4
Chapter 3	Inference without Covariates	5
	3.1 Estimation based on Martingale Residuals	5
	3.1.1 Preliminary on Counting Processes and Martingales	5
	3.1.2 Estimation of α	6
	3.2 Two-Stage Estimation	8
	3.3 Estimation based on Two-by-Two Tables	10
	3.3.1 The Proposed Method	10
	3.4 Discussion	12
Chapter 4	Inference with Covariates	13
	4.1 Two-Stage Estimation	13
	4.2 Estimation based on Two-by-Two Tables	15
Chapter 5	Simulations	16
	5.1 Data Generation	16
	5.1.1 Data Generation without Covariates	16
	5.1.2 Date Generation with Covariates	17
	5.2 Simulation Results	18
	5.2.1 Results without Covariates	18
	5.2.2 The Method Proposed by Hsu & Prentice	20
	5.2.3 Results with Covariates	22

Chapter 6	Conclusion	23
References		30
Appendix	Checking the Validity of the Method by Hsu and Prentice	32



Chapter 1 Introduction

1.1 Background

Multivariate survival analysis has wide applications in various fields. Consider an application in the study of familial aggregation for age onset data. The outcomes of age-at-onset within families may be correlated due to genetic and shared environmental risk factors among family members. Understanding the pattern of association may shed light on the disease etiology and is an important step useful for further scientific investigation.

Let (T_1, \dots, T_n) be the failure time variables for k members in a family. Note that T_i and T_j for $i \neq j$ are often correlated. To simplify the analysis we let $k = 2$. In the thesis, we assume that (T_1, T_2) follows a copula model. The major goal is to estimate the association parameter of the imposed model assumption based on right censored data. The situation for homogeneous data (i.e. without covariate information) will be discussed first. We will review existing methods and present our proposal. Then we will discuss a more complicated situation that includes the covariate information. Extension of the former inference methods will be discussed.

1.2 Notations and Outline of the Thesis

The failure times, T_1, \dots, T_n , are subject to censoring by C_1, \dots, C_n . Denoted the observed data by $\{(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j}, Z_{1j}, Z_{2j}); j = 1, 2, \dots, n\}$, where, for an individual i , X_i is the observed time and δ_i indicates the corresponding censoring status. If $\delta_i = 1$, X_i is actually the observed failure time, T_i . Otherwise if $\delta_i = 0$, X_i is the censoring time, C_i . Let $Z_i : (p \times 1)$ be a vector of covariates for $i = 1, 2$. It is assumed that T_i and C_i are independent.

In Chapter 2, we briefly review properties of copula models and introduce Archimedean copula models which constitutes a useful sub-family of copula models. Chapter 3 discusses

semi-parametric inference approaches to estimating the association parameter of a copula model in absence of covariates. Three methods will be examined in Section 3.1~3.3. In Section 3.4, we compare the three approaches.

In Chapter 4, we consider a more general situation in presence of covariates. Assuming that the covariates only affect the marginal distributions, we will extend the previous methods to account for marginal heterogeneity. In Chapter 5, we examine finite-sample performances of the methods via simulation studies. In Chapter 6, we give some concluding remarks. The appendix contains our further analysis about the estimator proposed by Hsu and Prentice (1996). Since this estimator does not have satisfactory performance in our simulations, we examine its computation algorithm in more detail to investigate the problem.



Chapter 2 Literature Review

2.1 Copula Models

Suppose that C is the joint survival function of two correlated $U(0,1)$ random variables defined on the unit square $[0,1]^2$. Let (T_1, T_2) be a pair of positive failure variables with marginal survival functions (S_1, S_2) respectively. Note that if (T_1, T_2) are continuous, both of $S_i(T_i)$ for $i=1,2$ are distributed as $U(0,1)$. Hence, the joint survival distribution function of (T_1, T_2) can be expressed as:

$$\begin{aligned}
 S(t_1, t_2) &= \Pr(T_1 \geq t_1, T_2 \geq t_2) \\
 &= \Pr(S_1(T_1) \geq S_1(t_1), S_2(T_2) \geq S_2(t_2)) \\
 &= \Pr(U_1 \geq S_1(t_1), U_2 \geq S_2(t_2)) \\
 &= C\{S_1(t_1), S_2(t_2)\}.
 \end{aligned} \tag{2.1}$$

The above expression was named as a ‘‘copula model’’ by Sklar (1959). Notice that via the copula modeling, the space of joint analysis is transferred from $[0, \infty)^2$ to $[0,1]^2$. Hence copula models have been frequently used for describing the association between two failure variables which are often skewed. From the viewpoint of statistical inference, copula models have the nice feature that the dependence structure can be studied separately from the marginal distributions.

Usually the function $C(u_1, u_2)$ is parameterized and denoted by $C_\alpha(u_1, u_2)$, where the parameter α measures the association between (T_1, T_2) . Kendall’s τ is a rank-invariant association measurement, defined as

$$\tau = \Pr\{(T_1^* - T_1)(T_2^* - T_2) \geq 0\} - \Pr\{(T_1^* - T_1)(T_2^* - T_2) < 0\}, \tag{2.2}$$

where (T_1, T_2) and (T_1^*, T_2^*) are independent but have the same distribution. The copula parameter α is related to Kendall’s τ such that

$$\tau = 4 \int \int C_\alpha(u_1, u_2) \frac{\partial^2 C_\alpha(u_1, u_2)}{\partial u_1 \partial u_2} du_1 du_2 - 1. \tag{2.3}$$

By equation (2.3), we see that there is a direct relationship between α and τ . For many well-known copula models, such a relationship has a nice one-to-one correspondence.

2.2 Archimedean Copula Models

The family of copula models has a useful sub-class, called the Archimedean copula (AC) family, in which $C_\alpha(u_1, u_2)$ can be further simplified as

$$C_\alpha\{u_1, u_2\} = \varphi_\alpha^{-1}\{\varphi_\alpha(u_1) + \varphi_\alpha(u_2)\}, \quad (2.4)$$

where $\varphi_\alpha(u_1) + \varphi_\alpha(u_2) \leq \varphi_\alpha(0)$ and $\varphi_\alpha(\cdot) : [0,1] \rightarrow [0, \infty]$ is the generating function satisfying $\varphi_\alpha(1) = 0$, $\varphi_\alpha'(t) = \frac{d\varphi_\alpha(t)}{dt} < 0$ and $\varphi_\alpha''(t) = \frac{d^2\varphi_\alpha(t)}{dt^2} > 0$. The AC family has nice analytic properties useful for more detailed investigation.

Three examples of AC models are presented below:

(i) *Clayton model (Clayton, 1978)*

$$C_\alpha(u_1, u_2) = \begin{cases} (u_1^{1-\alpha} + u_2^{1-\alpha} - 1)^{\frac{1}{1-\alpha}} & \text{if } \alpha > 1, \text{ and} \\ u_1 u_2 & \text{if } \alpha = 1 \end{cases}$$

$$\varphi_\alpha(v) = (v^{1-\alpha} - 1)/(1-\alpha).$$

(ii) *Frank model (Frank, 1979)*

$$C_\alpha(u_1, u_2) = \begin{cases} \log_\alpha \left\{ 1 + \frac{(\alpha^{u_1} - 1)(\alpha^{u_2} - 1)}{\alpha - 1} \right\}, & \text{if } 0 < \alpha < 1, \text{ and} \\ u_1 u_2 & \text{if } \alpha = 1 \end{cases}$$

$$\varphi_\alpha(v) = \log\{(1 - \alpha)/(1 - \alpha^v)\}.$$

(iii) *Positive stable frailty model (Hougaard, 1986)*

$$C_\alpha(u_1, u_2) = \begin{cases} \exp\{-[(-\log(u_1))^{\frac{1}{\alpha}} + (-\log(u_2))^{\frac{1}{\alpha}}]^\alpha\}, & \text{if } 0 < \alpha < 1, \text{ and} \\ u_1 u_2 & \text{if } \alpha = 1 \end{cases}$$

$$\varphi_\alpha(v) = \{-\log(v)\}^{\frac{1}{\alpha}}.$$

Chapter 3 Inference without Covariates

Consider a pair of failure times (T_1, T_2) which are assumed to be absolutely continuous. In presence of right censoring, we only observe $(X_1, X_2, \delta_1, \delta_2)$, where $X_i = T_i \wedge C_i$, $\delta_i = I(T_i \leq C_i)$ ($i = 1, 2$), and (C_1, C_2) are a pair of censoring times independent of (T_1, T_2) . Observed data can be denoted as $\{(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j}), (j = 1, 2, \dots, n)\}$. We assume that the joint survival function follows copula models:

$$\Pr(T_1 \geq t_1, T_2 \geq t_2) = C\{S_1(t_1), S_2(t_2)\} = \varphi_\alpha^{-1}\{\varphi_\alpha[S_1(t_1)] + \varphi_\alpha[S_2(t_2)]\} \quad (3.1)$$

The objective is to estimate α without specifying the marginal distributions.

In this chapter, we will review three semi-parametric inference approaches. In Section 3.1, we first review the concept of martingales which has been used by Hsu and Prentice (1996) to construct an estimating equation for estimating α . In Section 3.2, we review the paper of Shih and Louis (1995) who proposed a two-stage of estimation procedure. Specifically they suggest to estimate the marginal distribution first. Then these marginal estimates are treated as pseudo-observations of $\{S_1(T_1), S_2(T_2)\}$ in the likelihood based on $C_\alpha(\cdot, \cdot)$. In Section 3.3, we apply the idea of the Log-Rank statistic by constructing a series of two-by-two tables to estimate the association parameter. This idea has been used by Day et al. (1997) and Wang (2003) for analyzing semi-competing risks data.

3.1 Estimation based on Martingale Residuals

3.1.1 Preliminary on Counting Processes and Martingales

Based on observed variables, one can define $N(t) = I(X \leq t, \delta = 1)$ which is a counting process and the filtration $F_t = \sigma[N(u), I(T \leq u, \delta = 0); 0 \leq u \leq t]$, which describes the history of $N(t)$ prior to or at time t .

Define the cumulative intensity process:

$$\Lambda(t) = \int_0^t I(X \geq u) \lambda(u) du, \quad (3.2)$$

where

$$\lambda(u) = \lim_{\Delta \rightarrow 0} \frac{\Pr(T \in [u, u + \Delta) | T \geq u)}{\Delta}$$

is the hazard of T at time u . Then define

$$M(t) = I(X \leq t, \delta = 1) - \int_0^t I\{X \geq u\} \lambda(u) du = N(t) - \Lambda(t). \quad (3.3)$$

According to the Doob-Meyer decomposition (Fleming and Harrington, 1991), the expectation of $M(t)$ is a mean-zero martingale. Now we briefly verify the properties of $M(t)$:

(A) Given F_t which contains the history of the process prior the time of t , the expectation of $dN(t)$ be an intensity process, such as:

$$\begin{aligned} E(dN(t) | F_t) &= I(X \geq t) \Pr(X = t, \delta = 1 | T \geq t) + I(X \leq t) \Pr(X = t, \delta = 1 | T \leq t) \\ &= I(X \geq t) \frac{\Pr(X = t, \delta = 1)}{\Pr(X \geq t)} \\ &= I(X \geq t) \lambda(t) dt \\ &= d\Lambda(t). \end{aligned} \quad (3.4)$$

(B) By the Doob-Meyer decomposition, the expectation of $dM(t)$ equals to zero, such that:

$$E(dM(t) | F_t) = E(dN(t) - d\Lambda(t) | F_t) = 0. \quad (3.5)$$

3.1.2 Estimation of α

For the bivariate case, define

$$M_i(t) = I(X_i \leq t, \delta_i = 1) - \int_0^t I(X_i \geq u) \lambda_i(u) du = N_i(t) - \Lambda_i(t), \quad (3.6)$$

where $\lambda_i(u)$ is the hazard of T_i ($i=1,2$). When T_1 and T_2 are correlated, $M_1(t_1)$ and $M_2(t_2)$ are also correlated. Hsu and Prentice (1996) define the following cumulative covariance function, or the covariance rate, as:

$$\psi(\tau_1, \tau_2) = \int_0^{\tau_1} \int_0^{\tau_2} \psi(dt_1, dt_2), \quad (3.7)$$

where

$$\psi(dt_1, dt_2) = E\{M_1(dt_1)M_2(dt_2) | T_1 \geq t_1, T_2 \geq t_2\}.$$

The function $\psi(\tau_1, \tau_2)$ contains the information of association and can be estimated either non-parametrically or under a specified model.

A nonparametric estimator of $\psi(dt_1, dt_2)$ is given by

$$\hat{\psi}(dt_1, dt_2) = \frac{\sum_{j \in R(t_1, t_2)} \hat{M}_{1j}(dt_1) \hat{M}_{2j}(dt_2)}{r(t_1, t_2)}, \quad (3.8)$$

where $r(t_1, t_2) = \sum_{j=1}^n I(X_{1j} \geq t_1, X_{2j} \geq t_2)$ and $\hat{M}_i(t_i) = N_i(t_i) - \hat{\Lambda}_i(t_i \wedge T_i)$ ($i = 1, 2$).

If (T_1, T_2) follow a copula model, the model-restricted version of $\psi(t_1, t_2)$ can be expressed as:

$$\psi(\tau_1, \tau_2; \alpha) = \int_0^{\tau_1} \int_0^{\tau_2} \psi(dt_1, dt_2; \alpha), \quad (3.9)$$

where $\psi(dt_1, dt_2; \alpha) = \psi_0\{S_1(t_1), S_2(t_2); \alpha\} \Lambda_1(dt_1) \Lambda_2(dt_2)$ and

$$\psi_0\{u_1, u_2; \alpha\} = \{C_\alpha^{(11)}(u_1, u_2) - C_\alpha^{(10)}(u_1, u_2) - C_\alpha^{(01)}(u_1, u_2) + C_\alpha(u_1, u_2)\} C_\alpha(u_1, u_2)^{-1},$$

$$C_\alpha^{(11)}(u_1, u_2) = \frac{\partial^2 C_\alpha(u_1, u_2)}{\partial u_1 \partial u_2}, \quad C_\alpha^{(10)}(u_1, u_2) = \frac{\partial C_\alpha(u_1, u_2)}{\partial u_1} \quad \text{and}$$

$$C_\alpha^{(01)}(u_1, u_2) = \frac{\partial C_\alpha(u_1, u_2)}{\partial u_2}.$$

We have seen that, besides α , the model-based expression of the covariance function contains nuisance parameters. If the marginal functions $S_i(t)$ and $\Lambda_i(dt)$ ($i = 1, 2$) can be estimated, say by the Kaplan-Meier and Aalen's estimators, respectively. One can estimate $\psi(dt_1, dt_2)$ by

$$\psi(dt_1, dt_2; \alpha) = \psi_0\{\hat{S}_1(t_1), \hat{S}_2(t_2); \alpha\} \hat{\Lambda}_1(dt_1) \hat{\Lambda}_2(dt_2). \quad (3.10)$$

Hsu and Prentice (1996) suggested the following estimating equation based on the weighted difference between the nonparametric estimator of $\psi(t_1, t_2)$ and its model-based estimator:

$$\hat{U}_n(\alpha) = n^{-1} \int_0^{\tau_1} \int_0^{\tau_2} r(t_1, t_2) \{\hat{\psi}(dt_1, dt_2) - \hat{\psi}(dt_1, dt_2; \alpha)\} = 0. \quad (3.11)$$

Example: the Clayton model

$$C_\alpha(u_1, u_2) = (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{\alpha-1}} \quad (3.12)$$

The model-based expression of $\psi(t_1, t_2)$ is

$$\psi(t_1, t_2; \alpha) = \int_0^{t_1} \int_0^{t_2} \psi(du_1, du_2; \alpha), \quad (3.13)$$

where $\psi(du_1, du_2; \alpha) = \psi_0\{S_1(u_1), S_2(u_2); \alpha\} \Lambda_1(du_1) \Lambda_2(du_2)$ and

$$\psi_0\{u_1, u_2; \alpha\} = \alpha u_1^{-\alpha+1} u_2^{-\alpha+1} (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{-2} - (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{-1}.$$

3.2 Two-Stage Estimation

We can view $C_\alpha(u_1, u_2)$ as the joint survival function of $(U_1, U_2) = (S_1(T_1), S_2(T_2))$.

If one can obtain a random sample of (T_1, T_2) , denoted as $\{(T_{1j}, T_{2j}), (j=1, 2, \dots, n)\}$, and

$S_i(\cdot)$ ($i=1, 2$) are completely specified, the likelihood of α can be written as

$$\prod_{j=1}^n C_\alpha^{(11)}(u_{1j}, u_{2j}), \quad (3.14)$$

where $u_{ij} = S_i(t_{ij})$ ($i=1, 2, j=1, 2, \dots, n$). When (T_1, T_2) are subject to right censoring, the

likelihood of α can be modified as:

$$\prod_{j=1}^n C_\alpha^{10}(u_{1j}, u_{2j})^{\delta_{1j}(1-\delta_{2j})} \times C_\alpha^{01}(u_{1j}, u_{2j})^{(1-\delta_{1j})\delta_{2j}} \times C_\alpha(u_{1j}, u_{2j})^{(1-\delta_{1j})(1-\delta_{2j})}. \quad (3.15)$$

However, $(S_1(T_1), S_2(T_2))$ cannot be observed directly in practice. Shih and Louis (1995)

propose a straightforward method to estimate α . The ideas are to estimate $(S_1(T_1), S_2(T_2))$

first and then plug in the likelihood function of α . There are two methods for estimating the

marginal distribution. One is the parametric approach in which the marginal distributions are

specified up to some unknown parameters. By applying the maximum likelihood approach for

estimating the marginal parameters, one can estimate $(S_1(X_1), S_2(X_2))$ by

$(\tilde{S}_1(X_1), \tilde{S}_2(X_2))$. The other approach does not assume the form of the marginal distributions,

the Kaplan-Meier method can be applied for estimating $S_i(T_i)$ ($i=1, 2$). We illustrate this

approach using the Clayton model as an example.

Example: the Clayton model

$$C_\alpha(u_1, u_2) = (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{\alpha-1}} \quad (3.16)$$

And the likelihood function of α is:

$$\prod_{j=1}^n C_{\alpha}^{(11)}(u_{1j}, u_{2j})^{\delta_{1j}\delta_{2j}} C_{\alpha}^{(10)}(u_{1j}, u_{2j})^{\delta_{1j}(1-\delta_{2j})} \\ \times C_{\alpha}^{(01)}(u_{1j}, u_{2j})^{\delta_{1j}(1-\delta_{2j})} C_{\alpha}(u_{1j}, u_{2j})^{(1-\delta_{1j})(1-\delta_{2j})}, \quad (3.17)$$

where

$$C_{\alpha}^{(11)}(u_1, u_2) = \alpha \times u_1^{-\alpha} u_2^{-\alpha} \times (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{1-\alpha}-2},$$

$$C_{\alpha}^{(10)}(u_1, u_2) = u_1^{-\alpha} \times (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{1-\alpha}-1}$$

and

$$C_{\alpha}^{(01)}(u_1, u_2) = u_2^{-\alpha} \times (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{1-\alpha}-1}.$$

In presence of censoring, we observe $\{(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j}) (j = 1, 2, \dots, n)\}$. If the marginal distributions are specified, say $\exp(\lambda = 1)$, the maximum likelihood estimator of $S_i(X_{ij})$ is

$\tilde{S}_i(X_{ij})$, where

$$\tilde{S}_i(X_{ij}) = \exp\left(-\frac{\sum_{j=1}^n \delta_{ij}}{\sum_{j=1}^n X_{ij}} X_{ij}\right); \text{ for } i = 1, 2 \text{ and } j = 1, 2, \dots, n. \quad (3.18)$$

If the marginal form is unknown, we estimate $S_i(X_i)$ by Kaplan-Meier estimator:

$$\hat{S}_i(X_{ij}) = \prod_{u \leq X_{ij}} \left\{ 1 - \frac{\sum_{j=1}^n I(X_{ij} = u, \delta_{ij} = 1)}{\sum_{j=1}^n I(X_{ij} \geq u)} \right\}; \text{ for } i = 1, 2 \text{ and } j = 1, 2, \dots, n. \quad (3.19)$$

3.3 Estimation based on Two-by-Two Tables

3.3.1 The Proposed Method

In this section, we propose an estimator of α for an Archimedean Copula Models of the form:

$$\Pr(T_1 \geq t_1, T_2 \geq t_2) = C\{S_1(t_1), S_2(t_2)\} = \varphi_\alpha^{-1}\{\varphi_\alpha[S_1(t_1)] + \varphi_\alpha[S_2(t_2)]\}. \quad (3.20)$$

This idea is actually an application of the papers of Day et al. (1997) and Wang (2003) who considered semi-competing risks data. In presence of censoring, we observe $\{(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j}) (j = 1, \dots, n)\}$. The proposed estimating procedure is related to the Log-Rank statistic which can be constructed based on a series of two-by-two tables. At an observed failure points (t_1, t_2) , we can construct the following two-by-two tables as follows:

	$X_2 = t_2, \delta_2 = 1$	$X_2 > t_2$	
$X_1 = t_1, \delta_1 = 1$	$N_{11}(dt_1, dt_2)$		$N_{10}(dt_1, t_2)$
$X_1 > t_1$			
	$N_{01}(t_1, dt_2)$		$R(t_1, t_2)$

Table: Two-by-Two Tables at time (t_1, t_2)

The cell counts are defined as follows. Let

$$N_{11}(dt_1, dt_2) = \sum_{j=1}^n I(X_{1j} = t_1, \delta_{1j} = 1, X_{2j} = t_2, \delta_{2j} = 1),$$

$$N_{10}(dt_1, t_2) = \sum_{j=1}^n I(X_{1j} = t_1, \delta_{1j} = 1, X_{2j} \geq t_2),$$

$$N_{01}(t_1, dt_2) = \sum_{j=1}^n I(X_{1j} \geq t_1, X_{2j} = t_2, \delta_{2j} = 1)$$

and

$$R(t_1, t_2) = \sum_{j=1}^n I(X_{1j} \geq t_1, X_{2j} \geq t_2).$$

Notice that the odds ratio of the table converges to

$$\theta(t_1, t_2) = \frac{\frac{\partial^2}{\partial t_1 \partial t_2} \Pr(T_1 \geq t_1, T_2 \geq t_2) \Pr(T_1 \geq t_1, T_2 \geq t_2)}{\frac{\partial}{\partial t_1} \Pr(T_1 \geq t_1, T_2 \geq t_2) \frac{\partial}{\partial t_2} \Pr(T_1 \geq t_1, T_2 \geq t_2)}. \quad (3.21)$$

For models in the AC family, $\theta(t_1, t_2)$ can be simplified as

$$\theta(t_1, t_2) = \tilde{\theta}_\alpha (\Pr(T_1 \geq t_1, T_2 \geq t_2)), \quad (3.22)$$

where $\tilde{\theta}_\alpha(\cdot)$ is an univariate function satisfying

$$\tilde{\theta}_\alpha(v) = -v \times \frac{\varphi_\alpha''(v)}{\varphi_\alpha'(v)}. \quad (3.23)$$

Conditioning on the marginal counts, $N_{11}(dt_1, dt_2)$ follows a hypergeometric distribution with mean:

$$\begin{aligned} E_{11}(dt_1, dt_2, \alpha) &= \frac{\theta(t_1, t_2) N_{10}(dt_1, t_2) N_{01}(t_1, dt_2)}{\theta(t_1, t_2) N_{10}(dt_1, t_2) + R(t_1, t_2) - N_{10}(dt_1, t_2)} \\ &= \frac{\tilde{\theta}_\alpha(\Pr(X_1 \geq t_1, X_2 \geq t_2)) N_{10}(dt_1, t_2) N_{01}(t_1, dt_2)}{\tilde{\theta}_\alpha(\Pr(X_1 \geq t_1, X_2 \geq t_2)) N_{10}(dt_1, t_2) + R(t_1, t_2) - N_{10}(dt_1, t_2)}. \end{aligned} \quad (3.24)$$

By equating the empirical count with its model-based expected value and combining the tables with different (t_1, t_2) , we can construct the following estimating equation:

$$L(\alpha) = n^{-1} \int_0^{t_1} \int_0^{t_2} W(t_1, t_2) \{N_{11}(dt_1, dt_2) - E(dt_1, dt_2, \alpha)\} = 0, \quad (3.25)$$

where $W(t_1, t_2)$ is the weight function. It has been suggested to use

$$W(t_1, t_2) = \frac{\dot{\theta}_\alpha(S(t_1, t_2))}{\theta_\alpha(S(t_1, t_2))}, \quad (3.26)$$

where $S(t_1, t_2) = \Pr(T_1 \geq t_1, T_2 \geq t_2)$ and $\dot{\theta}_\alpha(v) = \frac{\partial \theta_\alpha(v)}{\partial \alpha}$.

Example: the Clayton model

$$C_\alpha(u_1, u_2) = (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{\alpha-1}} \quad (3.27)$$

For Clayton model with $\varphi_\alpha(t) = t^{-(\alpha-1)} - 1$, ($\alpha > 1$), we can obtain the model based expectation

of $N_{11}(dt_1, dt_2)$:

$$E_{11}(dt_1, dt_2, \alpha) = \frac{\alpha \times N_{10}(dt_1, t_2) N_{01}(t_1, dt_2)}{\alpha \times N_{10}(dt_1, t_2) + R(t_1, t_2) - N_{10}(dt_1, t_2)}. \quad (3.28)$$

3.4 Discussion

Shih and Louis (1995) proposed a two-stage estimation procedure. This approach is semi-parametric in the sense that the first stage can be estimated non-parametrically. However in some complicated data structures, such as semi-competing risk, nonparametric estimation in the first stage is not applicable. Hsu and Prentice (1996) constructed their estimating function based on martingale residuals. However the model-based expression involves too many high-dimensional nuisance parameters and therefore the resulting estimating equation becomes very complicated. Practical performance of this estimator heavily depends on the accuracy of the plug-in estimates in all data range. In our simulations, we have found that estimation in the tail region is not satisfactory.

The latter two approaches only use the conditions of the moments. It seems that the approach based on two-by-two tables is a more natural way for describing the dependence structure for AC models. Specifically for the Clayton model $E_{11}(dt_1, dt_2, \alpha)$ does not even contain any nuisance parameters. The key is that the theoretical odds ratio of the tables can well capture the association information for AC models. In contrast, $\psi(t_1, t_2; \alpha)$ is much less natural. It involves high-dimensional nuisance parameters that affect the subsequent inference procedure.

Chapter 4 Inference with Covariates

In this chapter, we discuss the estimation of α when there exist covariates that may affect the marginal distributions. Let (T_1, T_2) and (C_1, C_2) be failure times and censoring times, respectively. Let $Z_i : p \times 1 (i=1,2)$ be the covariate vectors. In presence of right censoring, observed data can be expressed as $\{(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j}, Z_{1j}, Z_{2j}), (j=1,2,\dots,n)\}$, where $X_{ij} = T_{ij} \wedge C_{ij}$ and $\delta_{ij} = I(T_{ij} < C_{ij})$, for $i=1,2$ and $j=1,2,\dots,n$.

Here we assume that Z_i only affects T_i ($i=1,2$), marginally. Let

$$U_i = S_i(t | Z_i) = \Pr(T_i \geq t | Z_i) \quad (i=1,2). \quad (4.1)$$

Furthermore, we assume that

$$\Pr(U_1 \geq u_1, U_2 \geq u_2) = C_\alpha(u_1, u_2). \quad (4.2)$$

In the following analysis, we assume that marginally the covariate effect follow the Cox Proportional Hazard model, such that

$$\lambda_i(t) = \lambda_{i,0}(t) \exp(Z_i' \beta_i) \quad (i=1,2), \quad (4.3)$$

where $\lambda_{i,0}(t)$ is the baseline hazard function. In Section 4.1, we modify the two-stage estimation approach. In Section 4.2, we apply our idea based on two-by-two tables to handle this more generalized situation.

4.1 Two-Stage Estimation

The first stage involves estimating pseudo-observations of (U_1, U_2) . Note that, under the Cox PH model, $U_i = S_i(t_i | Z_i) = S_{i,0}(t_i)^{\exp(Z_i' \beta_i)}$ ($i=1,2$). This implies that we need to estimate β_i and $S_{i,0}(t_i)$ first. The regression parameter β_i can be estimated by maximizing the following likelihood

$$L(\beta_i) = \prod_{\text{failure } u} \frac{\sum_{j=1}^n I(X_{ij} = u, \delta_j = 1) \times \exp(Z_{ij}' \beta_i)}{\sum_{j=1}^n I(X_{ij} \geq u) \times \exp(Z_{ij}' \beta_i)} \quad (i=1,2). \quad (4.4)$$

The estimator of the baseline function can be expressed by Breslow's estimator:

$$\hat{S}_{i,0}(t_i) = \prod_{t_{(i)} \leq t_i} \left(1 - \frac{\sum_{t_{(j)} \in R} I(X_i = t_{(j)}, \delta_i = 1) \exp(Z_{i(j)}' \hat{\beta}_i)}{\sum_{t_{(j)} \in R} \exp(Z_{i(j)}' \hat{\beta}_i)} \right) \quad (i = 1, 2). \quad (4.5)$$

Then we can obtain $U_1 = \hat{S}_{1,0}(T_1)^{\exp(Z_1' \hat{\beta}_1)}$ and $U_2 = \hat{S}_{2,0}(T_2)^{\exp(Z_2' \hat{\beta}_2)}$. Based on

$\{(U_{1j}, U_{2j}, \delta_{1j}, \delta_{2j}) \quad (j = 1, 2, \dots, n)\}$, we can estimate α by the following likelihood

function:

$$\begin{aligned} \prod_j L(\alpha) &= \prod_j C_\alpha^{(11)}(u_{1j}, u_{2j})^{\delta_{1j}\delta_{2j}} C_\alpha^{(10)}(u_{1j}, u_{2j})^{\delta_{1j}(1-\delta_{2j})} \\ &\quad \times C_\alpha^{(01)}(u_{1j}, u_{2j})^{(1-\delta_{1j})(1-\delta_{2j})} C_\alpha^{(00)}(u_{1j}, u_{2j})^{(1-\delta_{1j})(1-\delta_{2j})} \quad (j = 1, \dots, n). \quad (4.6) \end{aligned}$$

For Clayton's model with $C_\alpha(u_1, u_2) = (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{\alpha-1}}$, the above likelihood equals

$$\begin{aligned} \prod_j L(\alpha) &= \prod_j C_\alpha^{(11)}(u_{1j}, u_{2j})^{\delta_{1j}\delta_{2j}} C_\alpha^{(10)}(u_{1j}, u_{2j})^{\delta_{1j}(1-\delta_{2j})} \\ &\quad \times C_\alpha^{(01)}(u_{1j}, u_{2j})^{(1-\delta_{1j})(1-\delta_{2j})} C_\alpha^{(00)}(u_{1j}, u_{2j})^{(1-\delta_{1j})(1-\delta_{2j})} \quad (j = 1, \dots, n). \quad (4.7) \end{aligned}$$

where $C_\alpha^{(11)}(u_1, u_2) = \alpha \times u_1^{-\alpha} u_2^{-\alpha} \times (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{1-\alpha}-2}$,

$$C_\alpha^{(10)}(u_1, u_2) = u_1^{-\alpha} \times (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{1-\alpha}-1}$$

and

$$C_\alpha^{(01)}(u_1, u_2) = u_2^{-\alpha} \times (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{1-\alpha}-1}.$$

4.2 Estimation based on Two-by-Two Tables

Based on pseudo-observations $\{(U_{1j}, U_{2j}, \delta_{1j}, \delta_{2j}) (j=1, \dots, n)\}$, we can construct the following two-by-two table:

	$U_2 = u_2, \delta_2 = 1$	$U_2 > u_2$	
$U_1 = u_1, \delta_1 = 1$	$N_{11}(du_1, du_2)$		$N_{10}(du_1, u_2)$
$U_1 > u_1$			$R(u_1, u_2)$
	$N_{01}(u_1, du_2)$		

Table: Two-by-Two Table based on pseudo-observations

The cell counts are defined as follows. Let

$$N_{11}(du_1, du_2) = \sum_{j=1}^n I(U_{1j} = u_1, \delta_{1j} = 1, U_{2j} = u_2, \delta_{2j} = 1),$$

$$N_{10}(du_1, u_2) = \sum_{j=1}^n I(U_{1j} = u_1, \delta_{1j} = 1, U_{2j} \geq u_2),$$

$$N_{01}(u_1, du_2) = \sum_{j=1}^n I(U_{1j} \geq u_1, U_{2j} = u_2, \delta_{2j} = 1)$$

and

$$R(u_1, u_2) = \sum_{j=1}^n I(U_{1j} \geq u_1, U_{2j} \geq u_2).$$

Accordingly the estimating function becomes:

$$L(\alpha) = n^{-1} \int_0^{\tau_1} \int_0^{\tau_2} W(u_1, u_2) \{N_{11}(du_1, du_2) - E(du_1, du_2, \alpha)\} = 0, \quad (4.8)$$

where

$$E_{11}(du_1, du_2, \alpha) = \frac{\theta(u_1, u_2) N_{10}(du_1, u_2) N_{01}(u_1, du_2)}{\theta(u_1, u_2) N_{10}(du_1, u_2) + R(u_1, u_2) - N_{10}(du_1, u_2)}.$$

Consider the Clayton model:

$$C_\alpha(u_1, u_2) = (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{\alpha-1}}. \quad (4.9)$$

The above likelihood function of α can be simplified using the fact that $\theta(u_1, u_2) = \alpha$.

Chapter 5 Simulations

5.1 Data Generation

Via simulations, we will examine finite-sample performances of several estimators of α without and with covariates. We consider generating (T_1, T_2) from the Clayton model of the form:

$$C_\alpha(u_1, u_2) = \begin{cases} (u_1^{-\alpha+1} + u_2^{-\alpha+1} - 1)^{\frac{1}{\alpha-1}} & (\alpha > 1) \\ u_1 u_2 & (\alpha = 1). \end{cases} \quad (5.1)$$

Note that $\alpha = (1-\tau)/2\tau$, where τ is Kendall's τ . In particular, we will adopt the data generation algorithm for the Clayton model proposed by Prentice and Cai (1992).

5.1.1 Data Generation without Covariates

Step (i) Specify the value of τ and compute $\alpha = \frac{1+\tau}{1-\tau}$.

Step (ii) Generate independent variables $\{(U_{1j}, U_{2j}) (j = 1, \dots, n)\}$, where $U_{ij} \sim U(0,1)$ ($i = 1, 2$).

Step (iii) Generate $\{(T_{1j}, T_{2j}) (j = 1, 2, \dots, n)\}$, such that:

$$\gamma_j = (1 - U_{2j})^{\frac{1}{\alpha}},$$

$$T_{1j} = \alpha \times \log[1 - \gamma_j + \gamma_j(1 - U_{1j})^{\frac{1}{1+\alpha}}],$$

$$T_{2j} = (1 - U_{2j})^{\frac{1}{\alpha}}.$$

Then (T_{1j}, T_{2j}) follows Clayton(α) and marginally T_{ij} ($i = 1, 2$) $\sim \exp(1)$.

Step (iv) Generate $\{(C_{1j}, C_{2j}) (j = 1, 2, \dots, n)\}$, both of which are uniformly distributed.

Finally with $\{(T_{1j}, T_{2j}, C_{1j}, C_{2j}) (j = 1, \dots, n)\}$, we can create observed data $\{(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j}) (j = 1, 2, \dots, n)\}$ such that $X_{ij} = T_{ij} \wedge C_{ij}$ and $\delta_{ij} = I(T_{ij} \leq C_{ij})$ for $i = 1, 2$.

5.1.2 Date Generation with Covariates

Now we generate data which include a binary covariate. We assume the marginal effect follows the Cox proportional hazard model (1972), such as:

$$S_i(t) = S_{i,0}(t)^{\exp(Z_i\beta_i)}, \quad (5.2)$$

where Z_i is the covariate and $S_{i,0}(t)$ is the baseline survival function at time t ($i = 1, 2$).

The general procedure can be stated as follows. Let $S(T) = U$, where $U \sim U(0,1)$, and under the Cox proportional hazard model, we have $S_0(T)^{\exp(Z'\beta)} = U$. Hence it follows that

$$\log(S_0(T)) = \frac{\log(U)}{\exp(Z'\beta)}, \quad (5.3)$$

which implies that

$$T = S_0^{-1} \left[\exp\left(\frac{\log(U)}{\exp(Z'\beta)}\right) \right]. \quad (5.4)$$

We still need to specify the form of $S_0(t)$. For most distributions, the inverse of $S_0(t)$ may not have an explicit expression which increases the numerical difficulty in the analysis. In our simulations, we specify the baseline survival function to be $S_0(t) = \exp(-t)$ and obtain the following explicit expression:

$$T = \frac{-\log(U)}{\exp(Z'\beta)}, \quad (5.5)$$

where $-\log(U)$ which follows $\exp(1)$.

The data generation procedure is summarized below.

Step (i) Generate Z_{ij} from Bernoulli(0.5) for $i = 1, 2$.

Step (ii) Generate failure times (T_{1j}^*, T_{2j}^*) for the baseline group with

$Z_{ij} = 0$ using the above algorithm, where $T_{ij}^* \sim \exp(1)$ ($i = 1, 2$).

Step (iii) Given the value of β_i , for those with $Z_{ij} = 1$, we set

$$T_{ij} = \frac{T_{ij}^*}{\exp(Z_{ij}'\beta_i)} \quad (i = 1, 2).$$

Step (iv) Generate $\{(C_{1j}, C_{2j}) (j = 1, 2, \dots, n)\}$, both of which are exponential distributed.

Finally with $\{(T_{1j}, T_{2j}, C_{1j}, C_{2j}, Z_{1j}, Z_{2j}) (j = 1, \dots, n)\}$, we can create observed data

$\{(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j}, Z_{1j}, Z_{2j}) (j = 1, 2, \dots, n)\}$ such that $X_{ij} = T_{ij} \wedge C_{ij}$ and $\delta_{ij} = I(T_{ij} \leq C_{ij})$.

5.2 Simulation Results

5.2.1 Results without Covariates

In this section, we evaluate two approaches based on two-stage estimation and the construction of two-by-two tables. Two sample sizes with $n = 100$ and $n = 500$ are considered. The parameter α ranges from 1.2 to 19 which correspond to τ from 0.1 to 0.9 . For each estimator, the average bias and standard deviation of α are reported based on 500 replications. To achieve the targeted censoring rates, 30 % and 60 %, we set C_i to follow $U(0, 5.5)$ and $U(0, 2.5)$, respectively for $i = 1, 2$.

Table 1.1 summarizes the results in absence of censoring. Note that for the approach of two-stage estimation, we also present the results when the first stage of estimation is performed parametrically. Recall that we assume the marginal distribution of failure time X is $\exp(\lambda = 1)$. Hence, for the parametric two-stage procedure, we use

$$\hat{\lambda} = \frac{\sum_{j=1}^n \delta_j}{\sum_{j=1}^n x_j} \quad (\text{in complete data } \hat{\lambda} = \frac{1}{\bar{x}}) \quad (5.6)$$

to plug in the second stage for estimating α . This approach yields better results (smaller bias and smaller variation) than the semi-parametric two-stage procedure. As for the approach constructed based on two-by-two tables, it produces fairly nice results despite that it makes no assumption on the marginal distributions. Specifically it is fairly unbiased and the variance is only slightly larger than the parametric two-stage procedure. Note that the variation of all the

estimators becomes larger when α increases.

Table 1.2 and 1.3 are the results in presence of external censoring. Although variation of the estimators are larger than those without external censoring, the estimators still perform well. For Table 1.1 ~ 1.3, the variation is close to zero with the increasing of sample size (from 100 to 500). Hence, we can conclude that all of the estimations satisfy the property of consistency.



5.2.2 The Method Proposed by Hsu & Prentice

In this section, we examine the performance of the estimator proposed by Hsu and Prentice (1996) under the same settings. The results based on complete data are summarized below.

α	n=500	
	bias $\times 10^{-2}$ (st.error $\times 10^{-2}$)	
1.2	-2.283	(5.861)
1.5	-6.701	(7.396)
1.85714	-15.267	(8.659)
2.3	-31.961	(8.996)
3.0	-66.637	(7.981)
4.0	-138.223	(5.786)
5.6	-284.485	(3.156)
9.0	-606.891	(1.162)
19.0	-1602.436	(0.860)

Table A: Original version of Hsu & Prentice's estimator with no censoring

Strangely, based on Table A, we found that $\hat{\alpha}$ seems to be not consistent when $\alpha \geq 2.3$. To investigate what caused the problem, we checked several things. The details are summarized in the Appendix. We suspected that the problem may be attributed to the estimation instability of $\psi(t_1, t_2; \alpha)$ in (3.10) in some region of (t_1, t_2) . The plugged-in marginal estimators are usually unstable in the tail area. Hence we trimmed the integration area from $[0, \infty) \times [0, \infty)$ to a bounded region. Because this modification has no theoretical justification, we only evaluate the case without censoring.

Table B below contains the results for the modified estimator which is analyzed in a bounded region. Specifically, for each margin, we trim 50 % of the tail region. Based on Table B, we can know that both of bias and standard deviation of $\hat{\alpha}$ have significant improvement. With the sample size increases from 100 to 500, the standard deviation of $\hat{\alpha}$ have less

variation. Despite of the improvement, the method still performs not as well as the previous two approaches. One possible reason is that the model-based expectation $\psi(t_1, t_2; \alpha)$ in (3.10) contains more nuisance parameters than the other two approaches.

α	n=500	n=100
	bias $\times 10^{-2}$ (st.error $\times 10^{-2}$)	
1.2	0.229 (10.741)	4.398 (24.211)
1.5	-0.162 (13.234)	3.805 (31.675)
1.85714	-0.145 (16.388)	3.804 (39.416)
2.3	-0.489 (20.883)	4.530 (49.133)
3.0	-0.161 (28.358)	5.530 (64.276)
4.0	0.922 (39.522)	9.705 (89.874)
5.6	2.931 (63.757)	12.754 (149.819)
9.0	13.172 (147.715)	43.050 (351.349)
19.0	104.310 (678.372)	322.146 (1652.888)

Table B: Modified version of Hsu and Prentice's estimator with no censoring

5.2.3 Results with Covariates

We have proposed to extend two inference approaches to a more complex situation that covariates affect the marginal distribution. Now we check the validity of the extension by simulations. Here we assume the Cox Proportional Hazard model to describe marginal heterogeneity. For the two-stage estimation approach, we only report the results that the marginal distributions are estimated non-parametrically. The parameter of α ranges from $1.\bar{2}$ to 19 and $\beta_1 = \beta_2 = 0.8$. We also evaluate the situation in presence of censoring with 30 % and 60 % censoring. To achieve the targeted censoring rates, we let C_i follow $\exp(\lambda = 3.5)$ and $\exp(\lambda = 1.5)$, respectively ($i=1,2$). Two sample sizes with $n = 100$ and $n = 500$ are evaluated. For each estimator, the average bias and standard deviation are reported based on 500 replications.

Table 2.1 summaries the result with covariates in absence of censoring. Our focus is on comparing the two methods after adjustment for the effects of β_1 and β_2 . The variation of the two approaches is close. However, the two-by-two table approach seems to produce less biased estimates. Table 2.2 and 2.3 are the results in presence of right censoring. The estimators of α have larger variation but still perform well. All of the estimators are consistent when the sample size increases. In the simulations not reported here, we find that the estimators of α become invalid if the marginal heterogeneity is ignored.

Chapter 6 Conclusion

In the thesis, we review three inference approaches for estimating the association parameter for copula models. The existing methods are originally developed for analyzing homogeneous data. Here we extend these methods to account for marginal heterogeneity explained by covariates.

The two-stage estimation procedure proposed by Shih and Louis (1995) is easy to implement but not applicable under more complicated data structures such as semi-competing risks data that involves dependent censoring. The proposed approach based on two-by-two tables is motivated by the Log-Rank statistics. In comparison, it is a simple procedure from both aspects of analytic derivations and computation. It also has nice performance in simulations. Since this approach only utilizes some moment conditions, it can be easily modified for different data structures. The estimator of Hsu and Prentice (1996) has poor performance in our simulations. If our numerical algorithm is correct, the poor performance may be caused by the plugged-in estimators of the nuisance functions.

The proposed method and the method by Hsu and Prentice (1996) are both moment-based procedures but their performances are very different. We have found that, for AC models, the odds ratio of the two-by-two table provides a better descriptive measure for the association. In contrast, the covariance function of martingale residuals proposed by Hsu and Prentice is much less natural. That is why it produces an estimating function that involves many nuisance parameters.

Table 1.1 Comparison of two approaches without external censoring.

α	<u>bias $\times 10^{-2}$ (st.error $\times 10^{-2}$)</u>					
	n=500			n=100		
	Two-Stage		Two-by-Two Table	Two-Stage		Two-by-Two Table
	<i>parametric</i>	<i>semi-parametric</i>		<i>Parametric</i>	<i>semi-parametric</i>	
1.2	-0.213 (8.114)	0.669 (8.382)	-0.026 (8.275)	0.968 (14.767)	3.8 (16.271)	0.383 (15.232)
1.5	-0.191 (10.047)	1.179 (10.531)	0.065 (10.536)	1.464 (19.040)	6.054 (20.966)	0.28 (20.749)
1.85714	-0.252 (12.514)	1.192 (13.326)	0.031 (13.401)	2.138 (23.801)	7.592 (26.739)	-0.074 (27.063)
2.3	-0.393 (15.841)	0.844 (17.039)	-0.042 (17.170)	2.793 (30.204)	7.993 (33.793)	-0.028 (34.737)
3.0	-0.578 (20.526)	0.018 (22.080)	-0.208 (22.244)	3.637 (39.343)	7.128 (43.565)	-1.11 (45.182)
4.0	-0.882 (27.588)	-1.561 (29.449)	-0.306 (29.807)	4.775 (53.258)	4.507 (57.379)	-2.252 (60.220)
5.6	-1.438 (39.439)	-5.398 (41.709)	-0.683 (42.362)	6.37 (76.751)	-2.423 (81.810)	-0.963 (85.493)
9.0	-2.532 (63.241)	-16.53 (65.670)	-1.721 (66.599)	9.213 (123.879)	-29.453 (128.361)	0.159 (136.928)
19.0	-5.370 (134.733)	-72.600 (138.037)	-2.178 (142.870)	16.84 (264.620)	-173.33 (271.930)	9.597 (303.381)

Table 1.2 Comparison of two approaches with censoring rate 0.3.

α	<u>bias $\times 10^{-2}$ (st.error $\times 10^{-2}$)</u>					
	n=500			n=100		
	Two-Stage		Two-by-Two Table	Two-Stage		Two-by-Two Table
	<i>parametric</i>	<i>semi-parametric</i>		<i>Parametric</i>	<i>semi-parametric</i>	
1.2	0.073 (9.088)	0.566 (9.182)	0.142 (9.068)	2.061 (16.984)	4.258 (18.502)	1.702 (17.562)
1.5	0.234 (11.437)	1.024 (11.641)	0.34 (11.635)	2.747 (22.384)	5.737 (23.998)	1.252 (23.918)
1.85714	0.19 (14.198)	0.965 (14.721)	0.334 (14.909)	3.487 (27.112)	6.941 (29.197)	1.423 (30.365)
2.3	0.024 (17.916)	0.395 (18.709)	0.244 (19.105)	4.139 (34.322)	7.02 (36.550)	1.254 (39.061)
3.0	-0.314 (22.888)	-0.743 (24.042)	0.255 (24.793)	4.853 (44.807)	5.003 (47.158)	0.051 (51.647)
4.0	-1.224 (30.484)	-3.377 (31.640)	0.078 (33.127)	3.853 (60.379)	-2.213 (61.451)	0.297 (69.924)
5.6	-2.872 (43.457)	-9.567 (44.718)	0.114 (47.279)	0.406 (86.207)	-21.197 (83.557)	5.469 (99.999)
9.0	-9.223 (68.992)	-31.33 (69.585)	0.634 (74.918)	-21.92 (141.180)	-88.543 (128.043)	16.75 (165.275)
19.0	-70.26 (158.490)	-189.043 (152.155)	3.405 (161.384)	-212.22 (348.587)	-481.44 (260.105)	33.59 (369.187)

Table 1.3 Comparison of two approaches with censoring rate 0.6.

α	bias $\times 10^{-2}$ (st.error $\times 10^{-2}$)					
	n=500			n=100		
	Two-Stage		Two-by-Two Table	Two-Stage		Two-by-Two Table
	<i>parametric</i>	<i>semi-parametric</i>		<i>parametric</i>	<i>semi-parametric</i>	
1.2	0.44 (10.714)	0.584 (10.759)	0.53 (10.736)	4.018 (21.149)	5.067 (22.596)	0.846 (21.655)
1.5	0.323 (13.202)	0.703 (13.357)	0.442 (13.401)	4.091 (27.924)	6.032 (29.777)	0.639 (29.280)
1.85714	0.317 (16.350)	0.685 (16.582)	0.409 (16.822)	5.072 (34.088)	7.577 (36.085)	1.563 (36.934)
2.3	0.199 (20.272)	0.409 (20.862)	0.416 (21.371)	6.378 (42.449)	8.277 (43.924)	0.798 (46.024)
3.0	0.073 (26.114)	-0.205 (27.040)	0.577 (28.097)	7.54 (54.756)	7.428 (55.933)	0.861 (60.081)
4.0	-0.299 (34.732)	-1.645 (36.026)	0.928 (37.965)	7.942 (72.487)	2.874 (73.717)	2.214 (83.575)
5.6	-1.692 (49.035)	-5.835 (50.423)	1.059 (53.934)	6.253 (104.281)	-10.142 (103.610)	7.152 (119.437)
9.0	-7.103 (78.580)	-19.496 (79.399)	2.336 (86.260)	-6.881 (166.990)	-53.666 (159.488)	25.38 (196.765)
19.0	-54.14 (175.675)	-124.01 (168.489)	6.457 (184.836)	-165.24 (373.784)	-377.02 (315.480)	32.51 (453.817)

Table 2.1 Comparison of two approaches under marginal heterogeneity without external censoring.

α	$\beta_1 = \beta_2$	<u>bias $\times 10^{-2}$ (st.error $\times 10^{-2}$)</u>							
		n=500				n=100			
		Two-Stage	Two-by-Two	$\hat{\beta}_1$	$\hat{\beta}_2$	Two-Stage	Two-by-Two	$\hat{\beta}_1$	$\hat{\beta}_2$
			$\hat{\alpha}$				$\hat{\alpha}$		
1.2	0.8	-0.501 (6.391)	0.059 (6.288)	-0.385 (9.188)	-0.188 (9.640)	-0.694 (14.900)	-0.396 (14.132)	0.022 (21.557)	-0.758 (21.390)
1.5		-1.983 (7.925)	-1.230 (7.934)	-0.257 (9.150)	-0.345 (9.759)	-3.612 (17.381)	-2.002 (17.066)	-0.526 (20.297)	1.460 (21.300)
1.85714		-3.055 (9.437)	-1.924 (9.747)	0.729 (9.110)	0.194 (9.501)	-7.404 (21.619)	-5.029 (21.922)	1.015 (21.017)	0.766 (22.258)
2.3		-4.229 (11.624)	-2.016 (11.763)	-0.265 (8.924)	0.400 (9.970)	-11.885 (25.475)	-6.569 (26.423)	1.496 (21.028)	1.476 (21.540)
3.0		-5.615 (15.486)	-2.360 (15.760)	0.626 (9.976)	0.494 (9.766)	-15.534 (34.260)	-6.149 (36.794)	1.836 (21.810)	2.830 (22.377)
4.0		-7.863 (20.590)	-2.752 (21.118)	1.845 (9.404)	1.580 (9.646)	-25.774 (47.262)	-10.616 (50.432)	1.717 (22.160)	2.244 (23.109)
5.6		-14.769 (26.894)	-5.651 (27.678)	-0.111 (9.259)	0.027 (9.180)	-42.934 (59.398)	-14.262 (66.131)	-0.876 (21.050)	-0.829 (21.554)
9.0		-31.589 (42.410)	-11.676 (43.023)	0.402 (9.888)	0.364 (9.788)	-98.841 (99.761)	-34.205 (105.528)	0.519 (21.985)	0.929 (21.968)
19.0		-114.648 (89.175)	-41.539 (93.940)	0.494 (9.542)	0.473 (9.435)	-341.087 (219.164)	-127.069 (245.475)	0.420 (23.016)	0.205 (22.920)

Table 2.2 Comparison of two approaches under marginal heterogeneity with censoring rate 0.3.

		<u>bias $\times 10^{-2}$ (st.error $\times 10^{-2}$)</u>							
		n=500				n=100			
α	$\beta_1 = \beta_2$	Two-Stage	Two-by-Two	$\hat{\beta}_1$	$\hat{\beta}_2$	Two-Stage	Two-by-Two	$\hat{\beta}_1$	$\hat{\beta}_2$
		$\hat{\alpha}$				$\hat{\alpha}$			
1.2		-0.267 (7.589)	-0.250 (8.413)	-0.345 (10.011)	0.080 (10.170)	0.638 (16.577)	1.978 (18.085)	-0.614 (22.271)	0.593 (23.457)
1.5		-1.858 (9.418)	-1.567 (10.379)	0.083 (9.424)	-0.124 (10.435)	-2.140 (21.365)	-1.309 (23.301)	-0.466 (22.783)	1.363 (25.114)
1.85714		-1.963 (10.775)	-1.007 (11.722)	0.337 (9.852)	1.079 (10.411)	-4.958 (23.924)	-0.703 (28.945)	0.933 (22.150)	0.658 (23.185)
2.3	0.8	-3.272 (14.306)	-0.576 (16.869)	0.232 (10.123)	0.704 (10.714)	-10.582 (31.164)	-3.821 (36.515)	-0.872 (23.256)	-0.840 (24.870)
3.0		-6.165 (17.556)	-1.408 (20.790)	0.327 (10.057)	0.704 (10.849)	-21.709 (37.096)	-10.014 (45.931)	1.367 (24.254)	1.187 (23.103)
4.0		-7.834 (23.841)	0.285 (27.205)	-0.431 (11.057)	-0.273 (10.434)	-24.520 (52.442)	3.311 (69.246)	1.304 (24.693)	1.672 (24.308)
5.6		-17.583 (31.227)	-1.378 (36.732)	0.574 (10.675)	0.352 (10.524)	-57.033 (69.257)	-2.362 (98.199)	1.592 (23.854)	1.033 (23.943)
9.0		-109.124 (55.087)	-58.938 (66.582)	0.586 (10.134)	0.349 (10.517)	-259.518 (111.386)	-135.602 (152.304)	2.719 (23.804)	0.992 (24.344)
19.0		-292.511 (127.019)	-97.013 (136.070)	0.397 (11.115)	0.562 (11.293)	-668.705 (210.578)	-183.206 (335.635)	2.082 (24.015)	2.155 (24.233)

Table 2.3 Comparison of two approaches under marginal heterogeneity with censoring rate 0.6.

		$\text{bias} \times 10^{-2}$ (st.error $\times 10^{-2}$)							
α	$\beta_1 = \beta_2$	n=500				n=100			
		Two-Stage	Two-by-Two	$\hat{\beta}_1$	$\hat{\beta}_2$	Two-Stage	Two-by-Two	$\hat{\beta}_1$	$\hat{\beta}_2$
		$\hat{\alpha}$				$\hat{\alpha}$			
1.2		-0.022 (9.175)	0.647 (11.317)	0.480 (11.516)	0.484 (12.338)	0.535 (19.467)	5.426 (27.392)	0.403 (27.579)	1.314 (28.132)
1.5		-0.700 (11.438)	0.997 (14.396)	-0.311 (11.200)	0.597 (12.072)	0.269 (27.556)	2.762 (34.610)	-0.788 (25.720)	0.609 (27.642)
1.85714		-1.103 (13.657)	0.187 (16.934)	-0.092 (11.824)	0.619 (11.884)	-2.658 (29.413)	1.710 (39.691)	1.143 (26.587)	4.003 (29.461)
2.3	0.8	-2.726 (18.460)	0.039 (23.483)	-0.025 (11.456)	0.230 (12.124)	-4.620 (40.159)	5.159 (57.367)	-0.679 (27.484)	-0.800 (28.109)
3.0		-5.260 (21.654)	-1.091 (27.524)	-0.243 (11.857)	0.166 (11.525)	-14.801 (51.198)	-1.219 (74.896)	1.004 (28.668)	2.123 (27.966)
4.0		-13.133 (28.383)	-5.477 (37.248)	0.005 (12.170)	0.090 (11.847)	-34.456 (62.953)	2.777 (97.260)	1.595 (27.887)	-0.243 (26.747)
5.6		-22.755 (42.830)	-6.343 (56.938)	-0.314 (11.934)	1.115 (13.131)	-70.193 (94.419)	-13.234 (133.771)	0.490 (27.494)	0.838 (26.873)
9.0		-70.929 (68.613)	-24.824 (92.450)	-0.423 (12.367)	0.509 (11.945)	-196.595 (142.673)	-53.852 (234.286)	0.169 (29.724)	-1.050 (26.876)
19.0		-395.364 (166.251)	-185.900 (228.931)	0.254 (12.735)	0.040 (12.080)	-840.543 (281.685)	-326.036 (561.663)	2.275 (29.854)	0.901 (28.127)

References

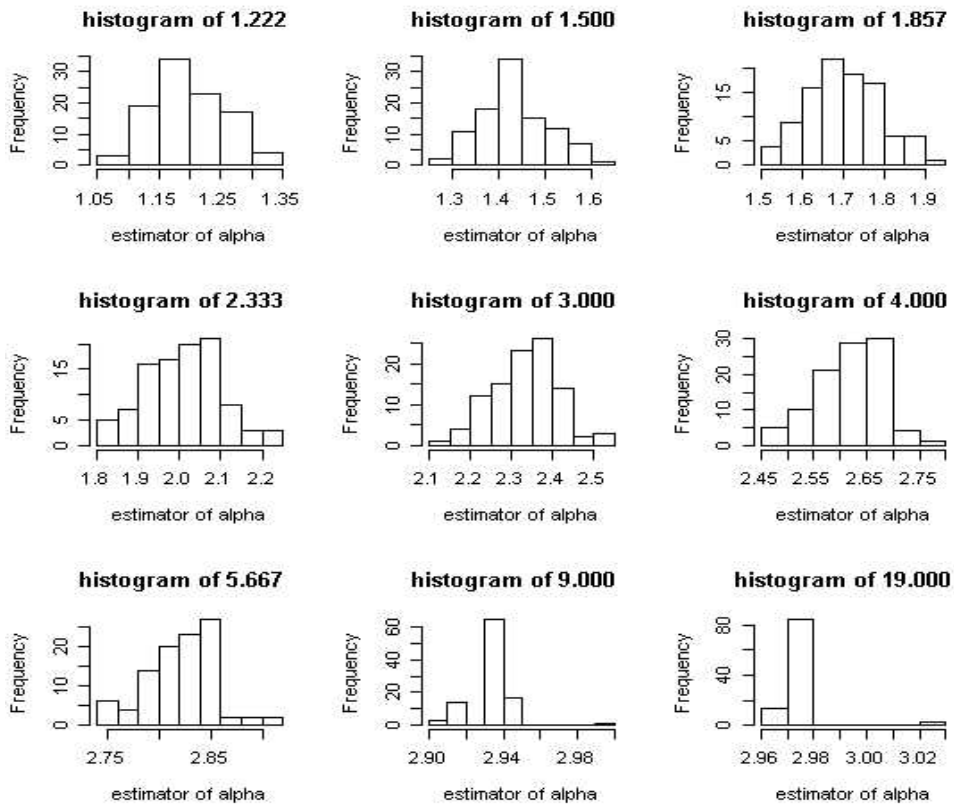
- Clayton, D. G. (1978). A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrics*, **65**, 141-151.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley.
- Genest, C. and Mackay, J. (1986). The Joy of Copulas: Bivariate Distributions with Uniform Marginals. *The American Statistician*, **40**, No. 4.
- Genest, C. and Rivest, L. P. (1993). Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association*, **88**, No. 423.
- Hsu, L. and Prentice, R. L. (1996). On Assessing the Strength of Dependency Between Failure Time Variates. *Biometrika*, **83**, 491-506.
- Hsu, L. and Zhao, L. P. (1996). Assessing Familial Aggregation of Age at Onset, by Using Estimating Equations, with Application to Breast Cancer. *Am. J. Hum. Genet.*, **58**, 1057-1071.
- Nelsen, R. B. (1997). Dependence and Order in Families of Archimedean Copulas. *Journal of Multivariate Analysis*, **60**, 111-122.
- Oakes, D. (1989). Bivariate Survival Models Induced by Frailties. *Journal of the American Statistical Association*, **84**, 487-493.
- Prentice, R. L. and Cai, J. (1992). Covariance and Survival Function Estimation Using Censored Multivariate Failure Time Data. *Biometrika*, **79**, 495-512.
- Shih, J. H. and Louis, T. A. (1995). Inference on the Association Parameter in Copula Models for Bivariate Survival Data. *Biometrics*, **51**, 1384-1399.
- Wang, W. (2003). Estimating the Association Parameter for Copula Models under Dependent Censoring. *Journal of the Royal Statistical Society: Series B*, **65**, 257-73.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of the American Statistical Association*, **84**, 1065-1073.



Appendix: Checking the Validity of the Method by Hsu and Prentice

Investigation #1: Is the distribution of $\hat{\alpha}$ reasonable?



Finding: There seems to be a bound on $\hat{\alpha}$.

Investigation #2: Whether the above problem is caused by the root-finding procedure?

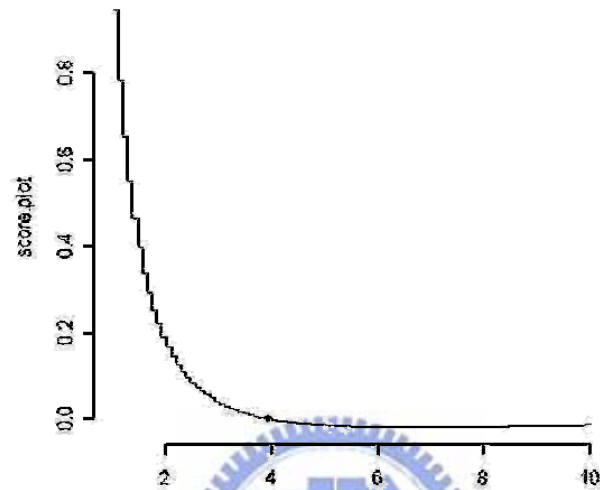


Figure A.2 plot of $\alpha = 9$

Finding: The estimating equation has a unique but wrong solution in some situation.

Investigation #3: Whether the plug-in estimators for the nuisance functions are not accurate?

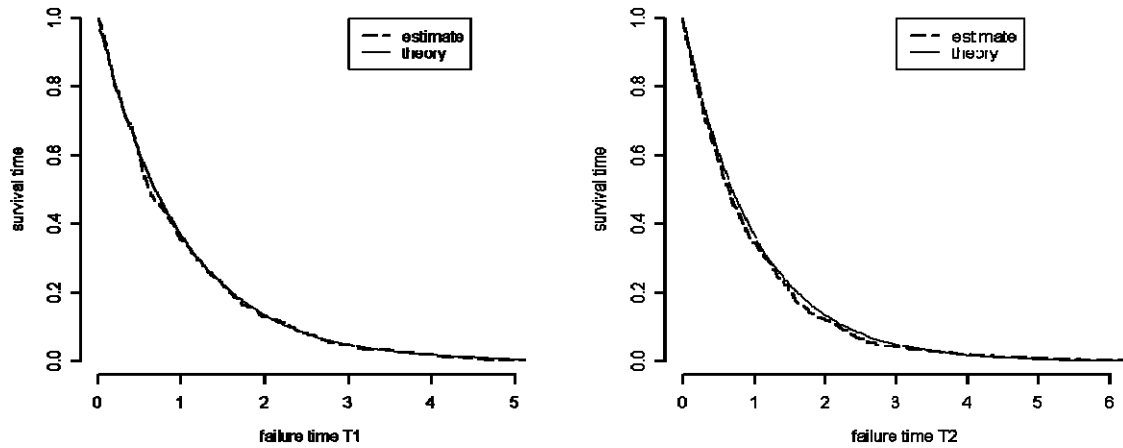


Figure A.3 the marginal survival function and its estimator

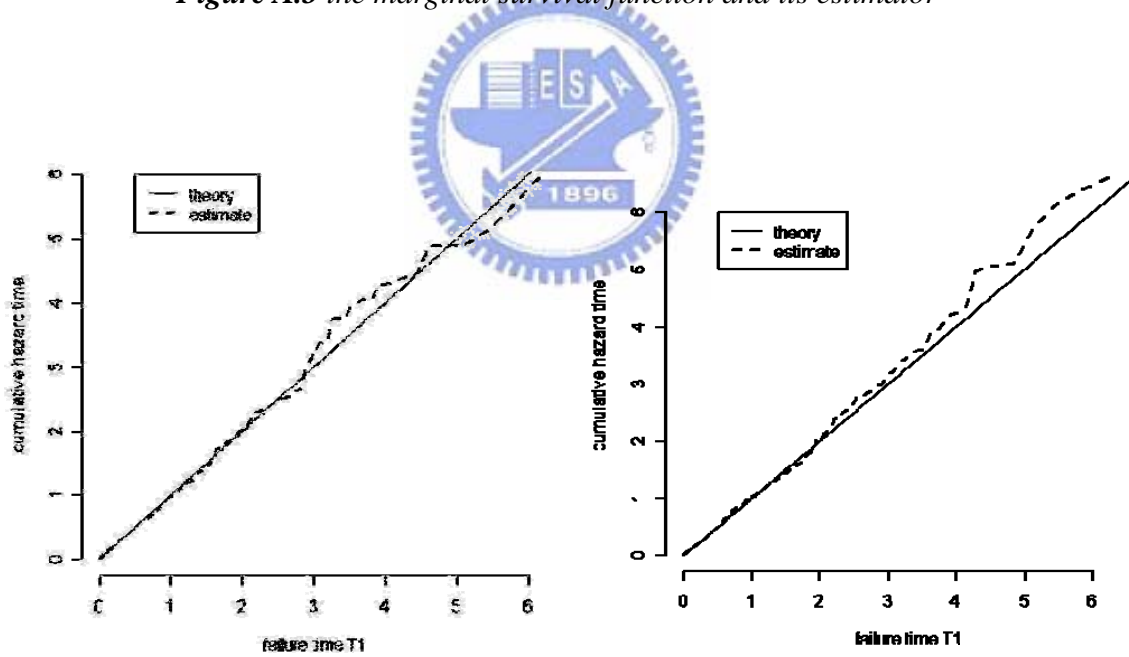


Figure A.4 the cumulative hazard function and its estimator

Finding: The plugged-in estimator have reasonable performance only in some region.