# TIME-DEPENDENT CROSS-RATIO ESTIMATION FOR BIVARIATE FAILURE TIMES

by

Tianle Hu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2011

Doctoral Committee:

Professor Bin Nan, Chair
Professor John D. Kalbfleisch
Associate Professor Moulinath Banerjee
Associate Professor Douglas E. Schaubel

To my wife, my parents and my mother in law

# ACKNOWLEDGEMENTS

First and foremost, I wish to express my great gratitude to my advisors Dr. Bin Nan for guiding and supervising me throughout the last three years. To me, he was more than a supervisor. He is a friend, whose tutelage made what would have been an arduous academic odyssey quite smooth. I benefit so much from his insight into statistics, his constant encouragement and above all, his paramount mentorship. I also like to thank Dr. Xihong Lin and Dr. James M. Robins of Harvard School of Public Health, whose intellectual acuity gave birth to the blueprint of my thesis. Without them, this work would not have been possible.

My thanks go out to Dr. John D. Kalbfleisch, Dr. Moulinath Banerjee and Dr. Douglas E. Schaubel for serving on my dissertation committee and providing very insightful comments and suggestions on my research. I am deeply grateful to Dr. Trivellore Eachambadi Raghunathan and Dr. Michael Boehnke who give me financial support through their research grants. It has been a very pleasant experience working as a research assistant under their supervision. In addition, I would like to thank Dr. Sebastian Zöllner for being my academic advisor during my first year of study.

I also wish to express my deepest gratitude to all the knowledgeable faculty members of the Department of Biostatistics at the University of Michigan for their excellent lectures; all the friendly staff for their assistance; and smart fellow students for their support and empathy throughout graduate school.

Finally and most importantly, I must thank my wife Li and daughter Chloe, whose support and unwavering love sustained me; and my parents whose utmost unselfishness and absolute faith in me allow me to be as ambitious as I wanted. I appreciate the patience and support of my mother-in-law during the lengthy process of my dissertation. It is the knowledge that my family will always be there

for me that gives me the strength and ability to tackle challenges head on.

## TABLE OF CONTENTS

# LIST OF FIGURES

**Figure**

# LIST OF TABLES

**Table**

## CHAPTER I

# Estimation of Time-Dependent Cross-Ratio for Bivariate Failure Times

### 1.1 Introduction

Statistical methods for analyzing bivariate correlated failure time data are of increasing need in many medical investigations. In the analysis of such data, it is important to measure the strength of association among the correlated failure times. Several global dependence measures have been proposed, such as Kendall's $\tau$ and Spearman's correlation coefficient (Hougaard 2000, chap. 4), and weighted average reciprocal cross-ratio (Fan, Hsu, & Prentice 2000a; Fan, Prentice, & Hsu 2000b). Local dependence measures have also been proposed for bivariate survival data. One commonly used such measure is the cross-ratio that is formulated as the ratio of two conditional hazard functions and thus measures the relative hazard of one time component conditional on another time component at some time point and beyond (Kalbfleisch & Prentice 2002, chap. 10).

Global measures, though quantitatively simple, are not always desirable because they can mask important features of the data and often do not address scientific questions of interest when the dependence of two event times is time-dependent and modeling such dependence is of major interest (Nan, Lin, Lisabeth, & Harlow 2006). In such settings, the cross-ratio function is of particular interest because of its attractive hazard ratio interpretation. Clayton (1978) considered a constant cross-ratio that yields an explicit closed-form bivariate survival function, the Clayton copula model. Re-parameterizing the Clayton model, Oakes (1982, 1986, 1989) analyzed bivariate failure times in the frailty framework, where a common latent frailty variable induces the correlation between survival times. Though such bivariate distributions induced by frailty models generate a rich subclass of archimedean distributions, these models only provide restrictive approaches to modeling time dependent cross-ratio, because the time dependent cross-ratio in the archimedean distributions is completely dictated by the specification of the copula as well as the marginal distributions. For the Clayton model and the gamma frailty model, many likelihood based estimating methods have been developed for the parameter estimation, see e.g. Clayton (1978), Oakes (1986), Shih & Louis (1995), Glidden & Self (1999), and Glidden (2000), among others. Such methods may also be applied to other copula models.

To estimate the cross-ratio as a function of time, Nan et al. (2006) partitioned the sample space of the bivariate survival time into rectangular regions with edges parallel to the time axes and assumed that the cross-ratio is constant in each rectangular region. A Clayton type of model was established for the joint survival function, and a two-stage likelihood based method similar to Shih & Louis (1995) was used to estimate the piecewise constant cross-ratio. In the context of competing risks, Bandeen-Roche & Ning (2008) proposed a nonparametric method for estimating the piecewise constant time-varying cause-specific cross-ratio using the binned survival data based on the same partitioning idea for the sample space. In both methods, how to choose the partition is arguable. Time-varying cross-ratio can also be estimated from a copula model. Recently, Li & Lin (2006) and Li, Prentice, & Lin (2008) characterized the dependence of bivariate survival data through the correlation coefficient of normally transformed bivariate survival times. Such methods, however, require assumptions of specific copula models for the joint survival function, for which appropriate model checking techniques are lacking.

We are not aware of any method in the literature on estimating the cross-ratio as a flexible continuous function of both time components without modeling the joint survival function. The methods of estimating weighted reciprocal of cross-ratios proposed by Fan et al. (2000a,b) cannot be applied to the estimation of the cross-ratio itself. In this article, we consider a parametric model for the log transformed cross-ratio, in particular, a polynomial function of the two time components, for censored bivariate survival data. Other parametric models can also be considered similarly. Since a closed form of the joint survival function is not available for the cross-ratio with a general polynomial functional form of times, it is difficult to develop a likelihood based approach for the cross-ratio estimation. Instead we construct an objective function for the cross-ratio parameters, which we call the pseudo-partial likelihood, by mimicking the partial likelihood of the Cox proportional hazards model (Cox 1972). Specifically, we treat whether an event happens at a time point or beyond along one time axis as a binary covariate and the other time component as the survival outcome variable, and then construct the corresponding partial likelihood function. Such a construction does not need any model for either the joint or the marginal survival function, thus is robust against model misspecification. We obtain the parameter estimates by maximizing the pseudo-partial likelihood function. We show that the estimator is consistent and asymptotically normal. The proofs are heavily relying on the modern empirical process theory. The proposed methodology is readily extendable to the estimation of an arbitrary cross-ratio function by using the tensor product splines. The work in this chapter has been published recently in Biometrika (Hu et al. 2011).

## 1.2 The cross-ratio function

Let $(T_1, T_2)$ be a pair of absolutely continuous correlated failure times. The cross-ratio function of $T_1$ and $T_2$ (Clayton 1978; Oakes 1989) is defined as

$$\theta(t_1, t_2) = \frac{\lambda_2(t_2|T_1 = t_1)}{\lambda_2(t_2|T_1 > t_1)} = \frac{\lambda_1(t_1|T_2 = t_2)}{\lambda_1(t_1|T_2 > t_2)}, \tag{1.1}$$

where $\lambda_1$ and $\lambda_2$ are the conditional hazard functions of $T_1$ and $T_2$, respectively. The second equality in (1.1) can be verified via direct calculation using the Bayes rule. The two event times $T_1$ and $T_2$ are independent if $\theta(t_1, t_2) = 1$, positively correlated if $\theta(t_1, t_2) > 1$, and negatively correlated if $\theta(t_1, t_2) < 1$ (Kalbfleisch & Prentice 2002). Following Clayton (1978) and Oakes (1982, 1986, 1989), model (1.1) is equivalent to the following second-order partial differential equation:

$$\frac{\partial^2 h}{\partial t_1 \partial t_2} + (\theta - 1)\frac{\partial h}{\partial t_1}\frac{\partial h}{\partial t_2} = 0, \tag{1.2}$$

where $h(t_1, t_2) = -\log\{S(t_1, t_2)\}$, $S(t_1, t_2)$ is the joint survival function of $(T_1, T_2)$ at $(t_1, t_2)$.

When $\theta$ is constant, it can be shown that equation (1.2) has a unique solution (Clayton 1978), which is given by the following Clayton copula model:

$$S(t_1, t_2) = \begin{cases} \left\{S_1(t_1)^{-(\theta-1)} + S_2(t_2)^{-(\theta-1)} - 1\right\}^{-\frac{1}{\theta-1}}, & \theta > 1, \\ S_1(t_1)S_2(t_2), & \theta = 1, \\ \max\left(\left\{S_1(t_1)^{-(\theta-1)} + S_2(t_2)^{-(\theta-1)} - 1\right\}^{-\frac{1}{\theta-1}}, 0\right), & \theta < 1. \end{cases}$$

where $S_1$ and $S_2$ are the marginal survival functions of $T_1$ and $T_2$. When $\theta$ is piecewise constant on a grid of the sample space of $(T_1, T_2)$, equation (1.2) is also solvable (Nan et al. 2006). The solution is similar to the above Clayton model within a rectangular region $(t_1, t_2) \in [u_1, u_2) \times [v_1, v_2)$, but with left truncation at the point $(u_1, v_1)$:

$$\begin{aligned} &S(t_1, t_2) \\ = &\begin{cases} \left\{S(t_1, v_1)^{-(\theta-1)} + S(u_1, t_2)^{-(\theta-1)} - S(u_1, v_1)^{-(\theta-1)}\right\}^{-1/(\theta-1)}, & \theta > 1, \\ S(t_1, v_1)S(u_1, t_2)/S(u_1, v_1), & \theta = 1, \\ \max\left(\left\{S(t_1, v_1)^{-(\theta-1)} + S(u_1, t_2)^{-(\theta-1)} - S(u_1, v_1)^{-(\theta-1)}\right\}^{-1/(\theta-1)}, 0\right), & \theta < 1. \end{cases} \end{aligned}$$

It is clearly seen that all the pieces of the joint survival function are interconnected through survival functions on the edges of the grid. Once the analytical form of the joint survival function is available, it becomes possible to develop a likelihood based approach. For example, Nan et al. (2006) extended the two-stage approach of Shih & Louis (1995) to the estimation of the piecewise constant cross-ratio.

In fact the Clayton copula belongs to an important family of copulas known as Archimedean copulas which have a simple form with a variety of dependence structures. Archimedean copula model has the following representation:

$$H(u, v) = \phi^{-1}(\phi(u) + \phi(v)), (u, v) \in [0, 1]^2$$

where $\phi : [0, 1] \to [0, +\infty]$ is a function satisfying $\phi(1) = 0$, $\phi(0) = \infty$, $\phi'(x) < 0$ and $\phi''(x) > 0$.

Then $H(u, v)$ is a distribution function on $[0, 1]^2$ with uniform marginals.

Commonly used Archimedean copula models include:

- Clayton copula, where $\phi(u, \theta) = u^{-(\theta-1)} - 1$,
- Frank copula, where $\phi(u, \theta) = \log \frac{1-\theta}{1-\theta^u}$, and
- Gumbel copula, where $\phi(u, \theta) = (-\log u)^\theta$.

Like Clayton copula, Frank and Gumbel copula are parameterized by a single parameter which dictates the dependence structure.

On the one hand, likelihood based approach to estimating $\theta$ would require knowing the solution of (1.2). However, obtaining an explicit analytical solution of equation (1.2) for an arbitrary cross-ratio function $\theta$ is impossible, even when $\theta$ is a simple function of $t_1$ and $t_2$, for example, a linear function; On the other hand, computing the cross-ratio $\theta$ through directly modeling the joint survival function $S(t_1, t_2)$ using, for example, a copula model may not be desirable because the model assumption can be very sensitive and rigorous model checking tools are lacking. Alternatively, we propose a pseudo-partial likelihood approach with details given in the following section to estimate the cross-ratio $\theta$ that is a continuous function of $(t_1, t_2)$. In particular, we consider a parametric model:

$$\beta(t_1, t_2; \boldsymbol{\gamma}) = \log\{\theta(t_1, t_2; \boldsymbol{\gamma})\}, \tag{1.3}$$

where $\boldsymbol{\gamma}$ is a finite dimensional Euclidean parameter. It is straightforward to extend the parametric model to a nonparametric model using tensor product splines. We focus on parametric model in this article because it can approximate a smooth function arbitrarily well in practice and is advantageous in theoretical investigation. More discussions of nonparametric regression can be found in the discussion section.

### 1.3 The pseudo-partial likelihood method

Consider a pair of correlated continuous failure times $(T_1, T_2)$ that are subject to right censoring by a pair of censoring times $(C_1, C_2)$. Assume censoring times are independent of failure times. Suppose we observe $n$ independent and identically distributed copies of $(X_1, X_2, \Delta_1, \Delta_2)$, where $X_1 = \min(T_1, C_1)$, $X_2 = \min(T_2, C_2)$, $\Delta_1 = I(T_1 \leq C_1)$, and $\Delta_2 = I(T_2 \leq C_2)$. Here $I(\cdot)$ denotes the indicator function. We further assume that there is no ties among observed times for each of the two time components.

In view of the difficulty in directly solving the differential equation (1.2) analytically, we propose in this section a simple pseudo-partial likelihood method by introducing the Cox's partial likelihood idea into the cross-ratio regression framework by treating one time component as the covariate. To motivate the idea, we connect the cross-ratio definition in (1.1) with the Cox model partial likelihood for the two-group regression problem. Specifically, using the epidemiological terminology, if we treat $\{j : T_{1j} = t_1\}$ and $\{j : T_{1j} > t_1\}$ as the "exposure" and "non-exposure" groups, respectively, then from the first equality in (1.1), the cross-ratio $\theta(t_1, t_2)$ becomes the hazard ratio of $T_2$ between these two groups. Given $t_1 = X_{1i}$, for subjects $k$ with $X_{1k} \geq t_1$, his/her conditional hazard at $t_2 = X_{2j}$ is simply $\lambda_2(X_{2j}|X_{1j} > X_{1i})\theta(X_{1i}, X_{2j})^{I(X_{1k}=X_{1i})}$. By mimicking the partial likelihood idea, we can construct a similar objective function as follows based on these two groups categorized by $t_1 = X_{1i}$:

$$\prod_{j=1}^{n} \left[ \frac{\lambda_2(X_{2j}|X_{1j} > X_{1i})\theta(X_{1i}, X_{2j})^{I(X_{1j}=X_{1i})}}{\sum_{X_{2k} \geq X_{2j}} I(X_{1k} \geq X_{1i})\lambda_2(X_{2j}|X_{1j} > X_{1i})\theta(X_{1i}, X_{2j})^{I(X_{1k}=X_{1i})}} \right]^{I(X_{1j} \geq X_{1i})\Delta_{2j}\Delta_{1i}}$$
$$= \prod_{j=1}^{n} \left[ \frac{\theta(X_{1i}, X_{2j})^{I(X_{1j}=X_{1i})}}{\sum_{X_{2k} \geq X_{2j}} I(X_{1k} \geq X_{1i})\theta(X_{1i}, X_{2j})^{I(X_{1k}=X_{1i})}} \right]^{I(X_{1j} \geq X_{1i})\Delta_{2j}\Delta_{1i}}.$$

The indicators $I(X_{1j} \geq X_{1i})$ in the outer exponent and $I(X_{1k} \geq X_{1i})$ in the denominator exclude the subjects not belonging to either the "exposure" or the "non-exposure" group. Under the "no-tie" assumption, only when $k = i$ can the indicator $I(X_{1k} = X_{1i})$ take value $1$. The risk set condition $\{k : X_{2k} \geq X_{2j}\}$ together with $k = i$ is equivalent to $X_{2i} \geq X_{2j}$. Thus the denominator inside the bracket is equal to $N(X_{1i}, X_{2j})$, where $N(t_1, t_2) = \sum_{k=1}^{n} I(X_{1k} \geq t_1, X_{2k} \geq t_2)$, if $X_{2j} \leq X_{2i}$, which implies that all subjects in the risk set belong to the "non-exposure" group. If $I(X_{2j} \leq X_{2i}) = 1$, the denominator becomes $N(X_{1i}, X_{2j}) - 1 + \theta(X_{1i}, X_{2j})$. So we can re-write the above objective function as

$$\prod_{j=1}^{n} \left[ \frac{\theta(X_{1i}, X_{2j})^{I(X_{1j}=X_{1i})}}{N(X_{1i}, X_{2j}) - I(X_{2j} \leq X_{2i})(1 - \theta(X_{1i}, X_{2j}))} \right]^{I(X_{1j} \geq X_{1i})\Delta_{2j}\Delta_{1i}}. \tag{1.4}$$

Considering the symmetric structure of the definition of $\theta(t_1, t_2)$ determined by the second equality in (1.1), we can construct the same objective function as (1.4) by switching the roles of $X_1$ and $X_2$. By multiplying such constructed two objective functions over all possible ways of creating the "exposure" and "non-exposure" groups, i.e. all subjects, we obtain the following pseudo-partial likelihood function:

$$L_n = \prod_{i=1}^{n} L_i^{(1)} L_i^{(2)}, \tag{1.5}$$

where $L_i^{(1)}$ is given in (1.4) and $L_i^{(2)}$ is given in the following:

$$L_i^{(2)} = \prod_{j=1}^{n} \left[ \frac{\theta(X_{1j}, X_{2i})^{I(X_{2j}=X_{2i})}}{N(X_{1j}, X_{2i}) - I(X_{1j} \leq X_{1i})(1 - \theta(X_{1j}, X_{2i}))} \right]^{I(X_{2j} \geq X_{2i})\Delta_{1j}\Delta_{2i}}. \tag{1.6}$$

The maximizer of (1.5) is then called the pseudo-partial likelihood estimator.

To proceed, we replace $\theta$ by $\beta$ through model (1.3) and denote $l_n = n^{-1} \log L_n$. Let $\dot{\beta}_\gamma(t_1, t_2; \boldsymbol{\gamma}) = \partial\beta(t_1, t_2; \boldsymbol{\gamma})/\partial\boldsymbol{\gamma}$. Differentiating $l_n(\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$ and assuming no ties among observed times, we obtain the following estimating function for $\boldsymbol{\gamma}$:

$$\boldsymbol{U}_n(\boldsymbol{\gamma}) = \partial l_n(\boldsymbol{\gamma})/\partial\gamma = \boldsymbol{U}_n^{(1)}(\boldsymbol{\gamma}) - \boldsymbol{U}_n^{(2)}(\boldsymbol{\gamma}) + \boldsymbol{U}_n^{(3)}(\boldsymbol{\gamma}) - \boldsymbol{U}_n^{(4)}(\boldsymbol{\gamma}),$$

where

$$\begin{aligned}
\boldsymbol{U}_n^{(1)}(\boldsymbol{\gamma}) &= \boldsymbol{U}_n^{(3)}(\boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=1}^{n} \Delta_{1i}\Delta_{2i}\dot{\beta}_\gamma(X_{1i}, X_{2i}; \boldsymbol{\gamma}), \\
\boldsymbol{U}_n^{(2)}(\boldsymbol{\gamma}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\Delta_{1i}\Delta_{2j} I(X_{1j} \geq X_{1i}) I(X_{2j} \leq X_{2i}) e^{\beta(X_{1i}, X_{2j}; \boldsymbol{\gamma})}}{N(X_{1i}, X_{2j}) - I(X_{2j} \leq X_{2i})\{1 - e^{\beta(X_{1i}, X_{2j}; \boldsymbol{\gamma})}\}} \dot{\beta}_\gamma(X_{1i}, X_{2j}; \boldsymbol{\gamma}), \\
\boldsymbol{U}_n^{(4)}(\boldsymbol{\gamma}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\Delta_{1j}\Delta_{2i} I(X_{2j} \geq X_{2i}) I(X_{1j} \leq X_{1i}) e^{\beta(X_{1j}, X_{2i}; \boldsymbol{\gamma})}}{N(X_{1j}, X_{2i}) - I(X_{1j} \leq X_{1i})\{1 - e^{\beta(X_{1j}, X_{2i}; \boldsymbol{\gamma})}\}} \dot{\beta}_\gamma(X_{1j}, X_{2i}; \boldsymbol{\gamma}).
\end{aligned}$$

Then an estimator $\hat{\gamma}_n$ can be obtained by solving the equation $\boldsymbol{U}_n(\boldsymbol{\gamma}) = 0$ using the Newton-Raphson algorithm.

### 1.4 Asymptotic properties

We consider a polynomial parametric model with finite number of terms for $\beta(t_1, t_2; \boldsymbol{\gamma})$ in (1.3). In particular, we assume

$$\beta(t_1, t_2; \boldsymbol{\gamma}) = \sum_{k,l} \gamma_{kl} t_1^k t_2^l = \boldsymbol{z}(t_1, t_2)' \boldsymbol{\gamma}, \tag{1.7}$$

where $\boldsymbol{\gamma}$ is the vector of coefficients $\{\gamma_{kl}\}$. It is easily seen that $\dot{\beta}_\gamma(t_1, t_2, \boldsymbol{\gamma}) = \boldsymbol{z}(t_1, t_2)$ and is free of $\boldsymbol{\gamma}$. We have found that a cubic model with $0 \leq k + l \leq 3$ in (1.7) often yields satisfactory estimates for smooth cross-ratio functions. In this section, we provide asymptotic results for the estimation of $\boldsymbol{\gamma}$ in (1.7). Other parametric models can also be considered and theoretical calculations can be proceeded similarly with modified regularity conditions to guarantee that both $\beta(t_1, t_2; \boldsymbol{\gamma})$ and $\dot{\beta}_\gamma(t_1, t_2, \boldsymbol{\gamma})$ belong to Donsker classes. For model (1.7) we consider the following regularity conditions:

C1.1. The failure times are truncated at $(\tau_1, \tau_2)$, $0 < \tau_1, \tau_2 < \infty$, such that $\mathrm{pr}(T_1 > \tau_1, C_1 > \tau_1, T_2 > \tau_2, C_2 > \tau_2) > 0$.

C1.2. The parameter space $\Gamma$ is a compact set and the true value $\gamma_0$ is an interior point of $\Gamma$.

C1.3. The matrix $E\{\Delta_1 \Delta_2 \boldsymbol{z}(X_1, X_2)^{\otimes 2}\}$ is positive definite. Here $\boldsymbol{z}^{\otimes 2} = \boldsymbol{z}\boldsymbol{z}'$.

**Theorem I.1.** *Under Conditions C1.1-C1.3, the solution of* $\boldsymbol{U}_n(\boldsymbol{\gamma}) = 0$*, denoted by* $\hat{\gamma}_n$*, is a consistent estimator of* $\boldsymbol{\gamma}_0$.

The proof of Theorem I.1 proceeds in following steps. We first show that $\boldsymbol{U}_n(\boldsymbol{\gamma})$ converges to a deterministic function $\boldsymbol{u}(\boldsymbol{\gamma})$ uniformly, then show that $\boldsymbol{u}(\boldsymbol{\gamma})$ is a monotone function with a unique root at $\boldsymbol{\gamma}_0$. Then consistency follows easily. The calculation heavily involves the modern empirical process theory (see e.g. van der Vaart & Wellner 1996). Details are provided in Appendix.

**Theorem I.2.** *Under Conditions C1.1-C1.3, we have that* $n^{1/2}(\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0)$ *converges in distribution to a normal random variable with mean zero and variance* $\boldsymbol{I}(\boldsymbol{\gamma}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)\boldsymbol{I}(\boldsymbol{\gamma}_0)^{-1}$*, where* $\boldsymbol{I}(\boldsymbol{\gamma}_0) = 2E\{\Delta_1\Delta_2\boldsymbol{z}(X_1, X_2)^{\otimes 2}\}$ *and* $\boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)$*, which is given in the Appendix, is the asymptotic variance of* $\boldsymbol{U}_n(\boldsymbol{\gamma}_0)$.

The asymptotic normality in Theorem I.2 can be achieved by using the Taylor expansion of $\boldsymbol{U}_n(\hat{\boldsymbol{\gamma}}_n)$ around $\boldsymbol{\gamma}_0$. Again the detailed calculation involves empirical process theory and is deferred to Appendix. The asymptotic expression of $\boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)$ also provides a variance estimator of $n^{1/2}(\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0)$.

## 1.5 Numerical examples

### 1.5.1 Simulations

We conduct simulations to assess the performance of the proposed method. Directly generating data from a bivariate distribution with a cross-ratio function given in (1.7) is technically formidable because the second order nonlinear partial differential equation (1.2) does not have a closed form solution. Instead, we generate data from given joint distribution functions of $(T_1, T_2)$. Thus our method is an approximation to the true cross-ratio function.

We consider the Frank family as in Fan et al. (2000a,b). We begin with generating independent Uniform (0,1) random numbers $u_1$ and $u_2$. Then let $t_1 = -\log u_1$ so that $T_1$ follows unit exponential distribution. Finally let $t_2 = -\log\left(\log_\alpha[a/\{a + (1-\alpha)u_2\}]\right)$ where $a = \alpha^{u_1} + (\alpha - \alpha^{u_1})u_2$. Such generated $T_2$ also follows exponential distribution. The cross-ratio function is

$$\frac{(\alpha-1)\log(\alpha)\alpha^{2-e^{-t_1}-e^{-t_2}}}{(\alpha^{1-e^{-t_1}}-\alpha)\times(\alpha^{1-e^{-t_2}}-\alpha)}\Bigg[-1+e^{-t_1}+e^{-t_2} + \log_\alpha\left\{1+\frac{(\alpha^{1-e^{-t_1}}-1)(\alpha^{1-e^{-t_2}}-1)}{\alpha-1}\right\}\Bigg].$$

Following Fan et al. (2000a,b), we choose $\alpha = 0.0023$ and generate censoring times $C_1$ and $C_2$ from a Uniform $(0, 2.3)$ distribution, yielding a censoring rate of 40%. The estimated cross-ratio is obtained using the cubic polynomial model (1.7) with $0 \le k+l \le 3$ by maximizing the pseudo-partial likelihood function (1.5) with respect to coefficients $\boldsymbol{\gamma}$. Results are averaged over 1000 simulation runs, each with a sample size of 400.

We plot the estimated surface together with the true cross-ratio surface in the top panel of Figure 1.1. The bottom panel gives the cross-ratio as a function of one time component fixing the other time component from $t = 0.25$ to $t = 1.50$ with 0.25 time unit increment. Based on the empirical variance of $\hat{\boldsymbol{\gamma}}$, we calculate the confidence bands for $\beta$. Then by exponentiating, we obtain the empirical confidence bands for $\theta$. Figure 1.1 show that the proposed method estimates the true cross-ratio of the Frank family very well, despite the fact that model (1.7) is only an approximation of the true $\theta(t_1, t_2)$.

To check the performance of proposed variance estimator, we choose nine points based on the quartiles of the observed time distribution, and calculate the empirical standard error and the average of the model based standard error estimates at those points together with the coverage probabilities. Results given in Table 1.1 show that our proposed variance estimator works well with coverage probabilities close to 95% at most surveyed grid points. We use bootstrap to obtain variance estimators as well. Based on our results not shown here, they generally work well too.

Table 1.1: Comparison of empirical standard error and average model based standard error for the Frank family. The points on both margins are the quartiles of the marginal distributions of $X_1$ and $X_2$. The true log cross-ratio is $\beta$ and its estimator is $\hat{\beta}$. In the parentheses, $E.SE$ is the empirical standard error, $M.SE$ is the average model based standard error estimate, and $CP$ is the coverage probability of the 95% confidence interval.

| | $X_2$ 25% | | $X_2$ 50% | | $X_2$ 75% | |
|---|---|---|---|---|---|---|
| $X_1$ | $\beta$ | $\hat{\beta}(E.SE, M.SE, CP)$ | $\beta$ | $\hat{\beta}(E.SE, M.SE, CP)$ | $\beta$ | $\hat{\beta}(E.SE, M.SE, CP)$ |
| 25% | 1.51 | 1.51(0.12,0.13,96%) | 1.31 | 1.33(0.17,0.17,95%) | 0.98 | 1.06(0.35,0.31,91%) |
| 50% | 1.31 | 1.33(0.17,0.17,96%) | 1.20 | 1.19(0.16,0.16,96%) | 0.94 | 0.95(0.24,0.24,94%) |
| 75% | 0.98 | 1.04(0.34,0.31,92%) | 0.94 | 0.94(0.24,0.24,95%) | 0.79 | 0.76(0.27,0.26,93%) |

We also consider the following joint survival function to examine the proposed method for cross-ratios close to 1 with a different curvature than the Frank family:

$$S(t_1, t_2) = \exp\left\{c - at_1^b - at_2^b - c\cos(at_1^b + at_2^b)\right\}. \tag{1.8}$$

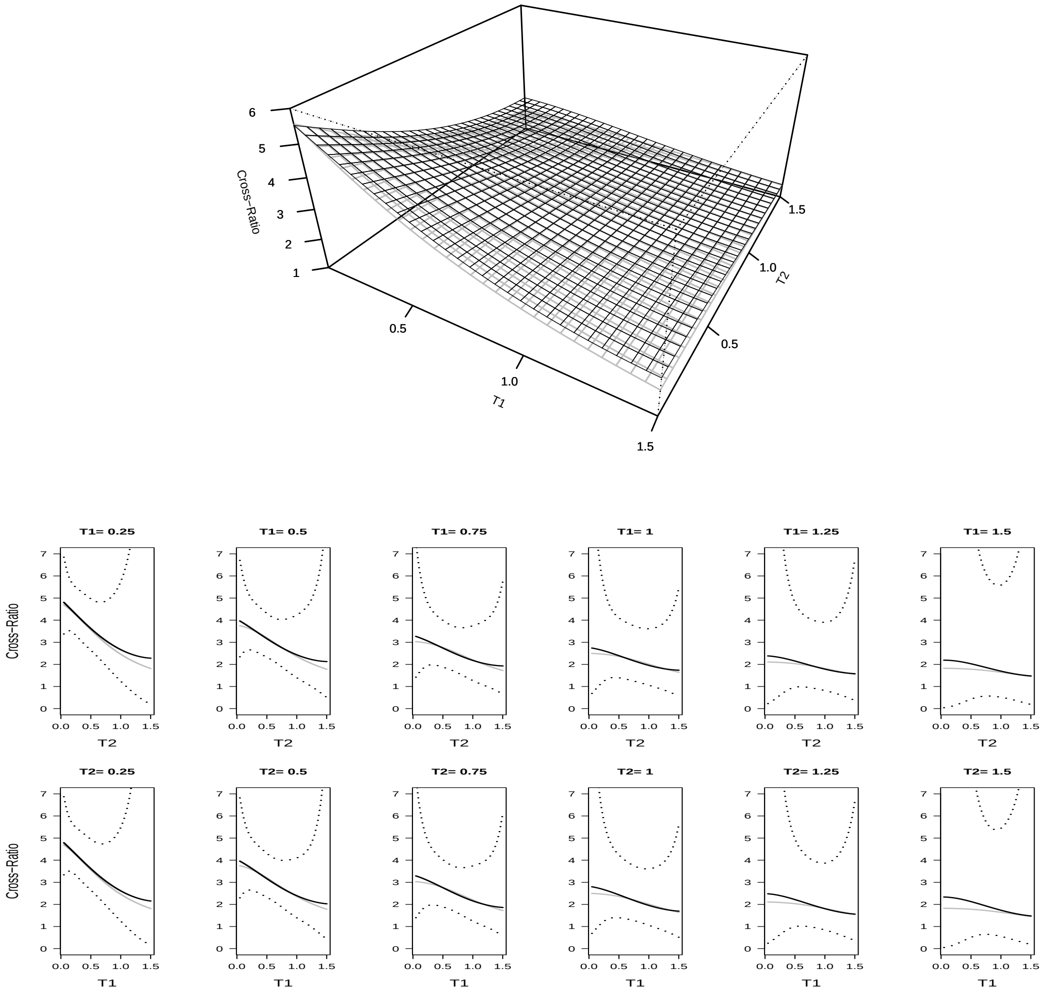


Figure 1.1: Cross-ratio for the Frank family. In the top panel, the surface in gray is the true cross-ratio and the surface in black is the estimated cross-ratio. In the bottom panel, gray curves are the true cross-ratio, black curves are the estimated cross-ratio, and dot curves are the empirical pointwise 95% confidence bands.

We call it the cosine model that has the following cross-ratio function

$$\theta(t_1, t_2) = \frac{c\cos(at_1^b + at_2^b)}{\{1 - c\sin(at_1^b + at_2^b)\}^2} + 1.$$

We choose $a = 0.7$, $b = 0.7$, and $c = 0.5$ in the following simulations. Random numbers of $T_1$ can be generated from its marginal survival function $S(t_1) = \exp\left\{c - at_1^b - c\cos(at_1^b)\right\}$. Conditioning on $T_1$, the conditional distribution function of $T_2$ is given by

$$1 - \exp\left\{-at_2^b - c\cos(at_1^b + at_2^b) + c\cos(at_1^b)\right\} \frac{1 - c\sin(at_1^b + at_2^b)}{1 - c\sin(at_1^b)},$$

from which we can generate random numbers of $T_2$.

From the cosine model (1.8), we generate 1000 pairs of bivariate survival times, close to the sample size of the Australian Twin study that we used for real data analysis. Independent censoring times for both components are generated from Uniform (0,3), which yield a censoring rate of 30%. We repeat all the calculations done for the Frank family and obtain very similar results. The true cross-ratio surface and its estimate are plotted in the top panel of Figure 1.2. Cross-ratio curves and their confidence bands as a function of one time component while fixing the other time component are plotted in the bottom panel of Figure 1.2. The comparison of model based standard error estimates to the empirical standard error is given in Table 1.2. The simulation results show that the proposed method works well for cross-ratios under or close to 1 that correspond to the cases of negative and weak dependence between $T_1$ and $T_2$.

Table 1.2: Comparison of empirical standard error and average model based standard error for the cosine model. The points on both margins are the quartiles of the marginal distributions of $X_1$ and $X_2$. The true log cross-ratio is $\beta$ and its estimator is $\hat{\beta}$. In the parentheses, $E.SE$ is the empirical standard error, $M.SE$ is the average model based standard error estimate, and $CP$ is the coverage probability of the 95% confidence interval.

| | $X_2$ 25% | | $X_2$ 50% | | $X_2$ 75% | |
|---|---|---|---|---|---|---|
| $X_1$ | $\beta$ | $\hat{\beta}(E.SE, M.SE, CP)$ | $\beta$ | $\hat{\beta}(E.SE, M.SE, CP)$ | $\beta$ | $\hat{\beta}(E.SE, M.SE, CP)$ |
| 25% | 0.55 | 0.54(0.09,0.09,94%) | 0.61 | 0.61(0.12,0.12,95%) | 0.53 | 0.53(0.21,0.21,95%) |
| 50% | 0.61 | 0.61(0.12,0.13,95%) | 0.58 | 0.57(0.15,0.16,96%) | 0.24 | 0.22(0.29,0.29,95%) |
| 75% | 0.53 | 0.51(0.22,0.21,93%) | 0.24 | 0.20(0.29,0.29,96%) | -0.63 | -0.80(0.65,0.54,93%) |

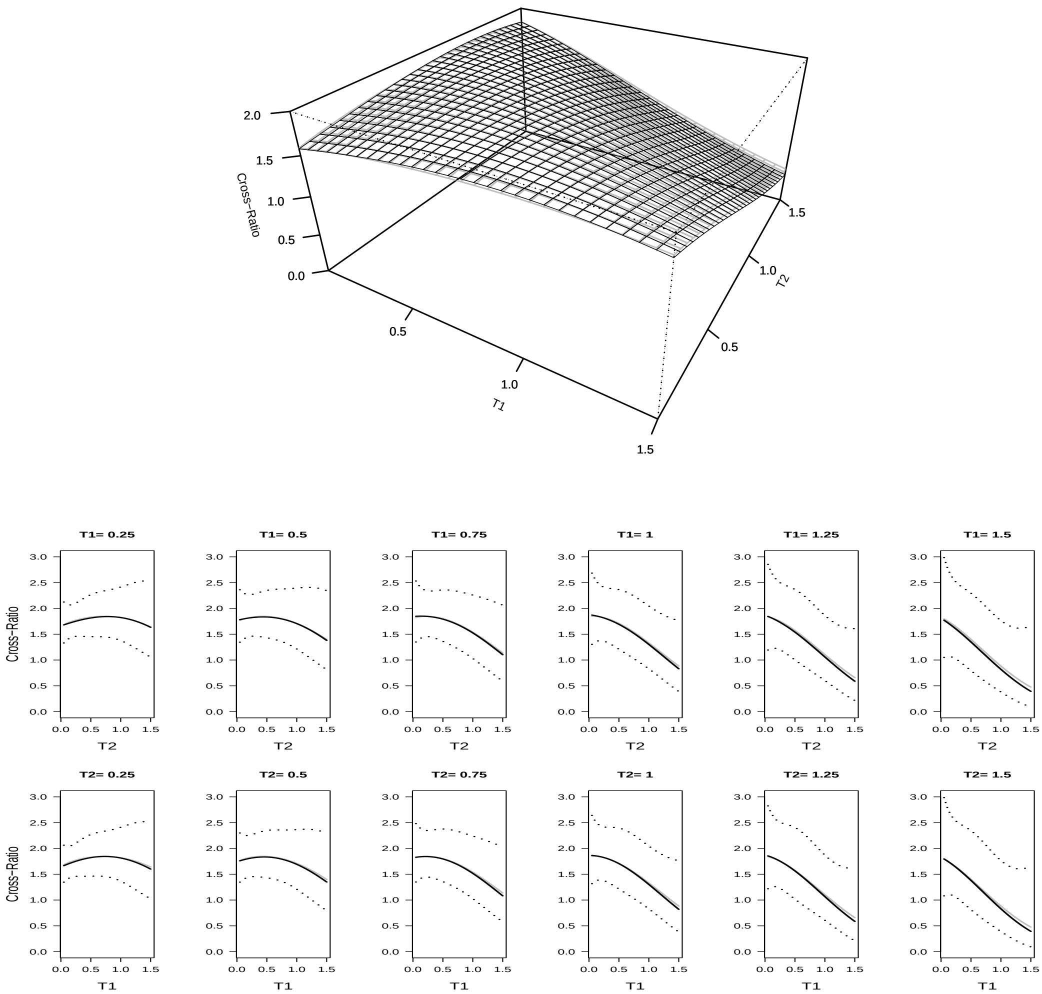


Figure 1.2: Cross-ratio for the Cosine model. In the top panel, the surface in gray is the true cross-ratio and the surface in black is the estimated cross-ratio. In the bottom panel, gray curves are the true cross-ratio, black curves are the estimated cross-ratio, and dot curves are the empirical pointwise 95% confidence bands.

To compare the efficiency of the proposed method with the two-stage method of Nan et al. (2006), we adopt the same simulation setup assuming that the cross-ratio $\theta(t_1)$ is piecewise constant over four intervals: $\theta(t_1) = .9$ when $t_1 \in [0, .25)$, $\theta(t_1) = 2.0$ when $t_1 \in [.25, .5)$, $\theta(t_1) = 4.0$ when $t_1 \in [.5, .75)$, and $\theta(t_1) = 1.5$ when $t_1 > .75$. The data are generated in the same mechanism as in Nan et al. (2006). The comparison results are shown in Table 1.3 with both bootstrap standard error and model based standard error and their respective coverage probabilities given for our method. Though the two stage method is more efficient, the loss in efficiency for our estimator is minor, especially at the beginning of the follow-up. On the other hand, our method is not limited to the piecewise assumption.

Table 1.3: Comparison of the pseudo-partial likelihood method with the two stage sequential method. The true values are $\theta = (.9, 2.0, 4.0, 1.5)$ when $t_1$ is in intervals $[0, .25), [.25, .5), [.5, .75)$, and above .75, the sample size is 500. The left panel is taken from Nan et al. (2006). $\hat{\theta}_{avg}$ is point estimate average; $E.SE$, empirical standard error; $B.SE$, average of the bootstrap standard error estimates using 100 bootstrap samples; $M.SE$, average of the model based standard error estimates; $B.CP$ and $M.CP$ are $95\%$ coverage probability using $B.SE$ and $M.SE$ respectively.

| | Two-Stage Piecewise | | | | Time Dependent | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\theta$ | $\hat{\theta}_{avg}$ | $E.SE$ | $B.SE$ | $B.CP$ | $\hat{\theta}_{avg}$ | $E.SE$ | $B.SE$ | $B.CP$ | $M.SE$ | $M.CP$ |
| 0.90 | 0.92 | 0.12 | 0.12 | 93% | 0.92 | 0.12 | 0.12 | 97% | 0.12 | 95% |
| 2.00 | 2.04 | 0.31 | 0.29 | 93% | 2.04 | 0.30 | 0.31 | 95% | 0.31 | 94% |
| 4.00 | 4.09 | 0.65 | 0.71 | 96% | 4.13 | 0.76 | 0.78 | 96% | 0.76 | 97% |
| 1.50 | 1.51 | 0.25 | 0.25 | 94% | 1.54 | 0.29 | 0.31 | 96% | 0.33 | 96% |

### 1.5.2 The Australian twin study

In this section, we present our real data analysis of ages at appendectomy for participating twin pairs in the Australian Twin Study (Duffy, Martin, & Mathews 1990). The same data were analyzed in Prentice & Hsu (1997) and Fan et al. (2000a,b). Primarily, the Australian Twin Study was conducted to compare monozygotic and dizygotic twins with respect to the strength of dependence in the risk for various diseases between twin pair members, because stronger dependence between monozygotic twin pair members would be indicative of genetic effect in the risk of disease of interest. Information was collected from twin pairs over the age of 17 on the occurrence, and the age at occurrence of disease related events, including the occurrence of vermiform appendectomy. Those who did not undergo appendectomy prior to survey, or were suspected of undergoing prophylactic appendectomy, gave rise to right censored failure times. For simplicity, the study was treated as a simple cohort study of twin pairs. Based on the descriptive analysis of Duffy et al. (1990), it is noted that females were more likely to undergo appendectomy across all ages and birth cohorts than males. Moreover, monozygotic female twins were found to be more concordant for appendectomy during their lifetime than their dizygotic counterparts. As the sample size for the females is twice as large as that for the males and the zygotic effect is more pronounced, we will focus on female twin pairs. By doing so, we also avoid the difficulties of modeling sex differences for the opposite sex dizygotic pairs.

As in Prentice & Hsu (1997), analyses presented here are confined to 1953 female twin pairs with available appendectomy information. The data are comprised of 1218 monozygotic twin pairs and 735 dizygotic twin pairs. Out of the monozygotic twin pairs, there are 144 pairs in which both twins were appendectomized, 304 pairs in which one twin underwent appendectomy and 770 pairs in which neither twin received the procedure. The corresponding numbers for the dizygotic twin pairs are 63, 208 and 464, respectively.

Since the order of twin one and twin two is arbitrary in the Australian Twin Study, we can take advantage of such symmetry to improve the estimation efficiency by using the following reduced model from (1.7):

$$\beta(t_1, t_2; \gamma) = \gamma_0 + \gamma_1(t_1 + t_2) + \gamma_2(t_1^2 + t_2^2) + \gamma_3 t_1 t_2 + \gamma_4(t_1^2 t_2 + t_1 t_2^2) + \gamma_5(t_1^3 + t_2^3). \tag{1.9}$$

We have observed that the analysis without such restriction yields similar results to the restricted analysis. Here we only report the restricted analysis. We conduct separate analyses for monozygotic and dizygotic twins.

The estimated cross-ratio surfaces for both monozygotic and dizygotic twins are plotted in the top panel of Figure 1.3. In the same figure, we also plot the difference between the two estimates. To clearly present the variability of the estimates, in the bottom panel of Figure 1.3 we also provide the pointwise 95% confidence bands for the estimated cross-ratio at three different values of $T_2$, which are close to the estimated quartiles of $T_2$. The confidence bands for the cross-ratio differences are obtained from separate analyses of 100 bootstrap samples from the pooled monozygotic and dizygotic data.

Figure 1.3 suggests that at a younger age the association for appendicitis risk between monozygotic twin pair members as well as dizygotic twins is strong, particularly in monozygotic twins, suggesting a genetic component to the disease. This finding is consistent with Fan et al. (2000a). Also both monozygotic and dizygotic twin pairs are more likely to undergo appendectomy around the same time. Top right plot in Figure 1.3 identifies the region shaded in black where the difference in the strength of association between monozygotic and dizygotic twins is statistically significant based on the pointwise confidence band. The plots on the top panel also show that such dependence diminishes over time. This suggests that later in life environmental causes may be more important in the development of the disease. From the pointwise confidence bands we can see that the association between monozygotic twins is significantly positive in a larger age range. In addition to testing the difference locally, we can also test whether the difference is significant globally. Existing methods had to assume that cross-ratio is a constant for a global test. Our method, without assuming the structure of cross-ratio surface, can test whether the cross-ratio surface is the same for monozygotic twins and dizygotic twins by testing $H_0 : \boldsymbol{\gamma}_{mz} = \boldsymbol{\gamma}_{dz}$, where $\boldsymbol{\gamma}_{mz}$ and $\boldsymbol{\gamma}_{dz}$ denote the coefficients in (1.9) for monozygotic twins and dizygotic twins respectively. Using a $\chi^2_6$ statistic $(\hat{\boldsymbol{\gamma}}_{mz} - \hat{\boldsymbol{\gamma}}_{dz})^T(\hat{Var}(\hat{\boldsymbol{\gamma}}_{mz}) + \hat{Var}(\hat{\boldsymbol{\gamma}}_{dz}))(\hat{\boldsymbol{\gamma}}_{mz} - \hat{\boldsymbol{\gamma}}_{dz})$, we obtain a insignificant p value of $0.13$ at $\alpha = 0.05$.

## 1.6 Discussion

Nonparametric estimation of the log cross-ratio $\beta(t_1, t_2)$ is of interest, particularly when it is a smooth function of $(t_1, t_2)$, for which the regression spline method using the tensor product splines can be implemented. When the number of knots are fixed, the model is essentially a parametric model and asymptotic properties can be derived in a similar way. If the number of knots is allowed to grow with the sample size, then a completely different approach needs to be developed for the proofs of asymptotic properties, which is usually a challenging problem. We have conducted the regression spline method in simulations under different combinations of number of knots and degree of smoothness, and observed slightly more variable results that are not presented in this article. Such an observation is likely due to the fact that the cross-ratio surfaces in both simulation settings only change with time gradually, i.e. they tend to be too flat to apply the regression spline method.

The proposed estimator is based on pseudo-partial likelihood method instead of the true likelihood, and thus likely not to be the most efficient estimator. If we were to construct the true likelihood as a function of arbitrary cross-ratio, we would need to solve for $h(t_1, t_2)$ in (1.2) when $\theta = \theta(t_1, t_2)$ to

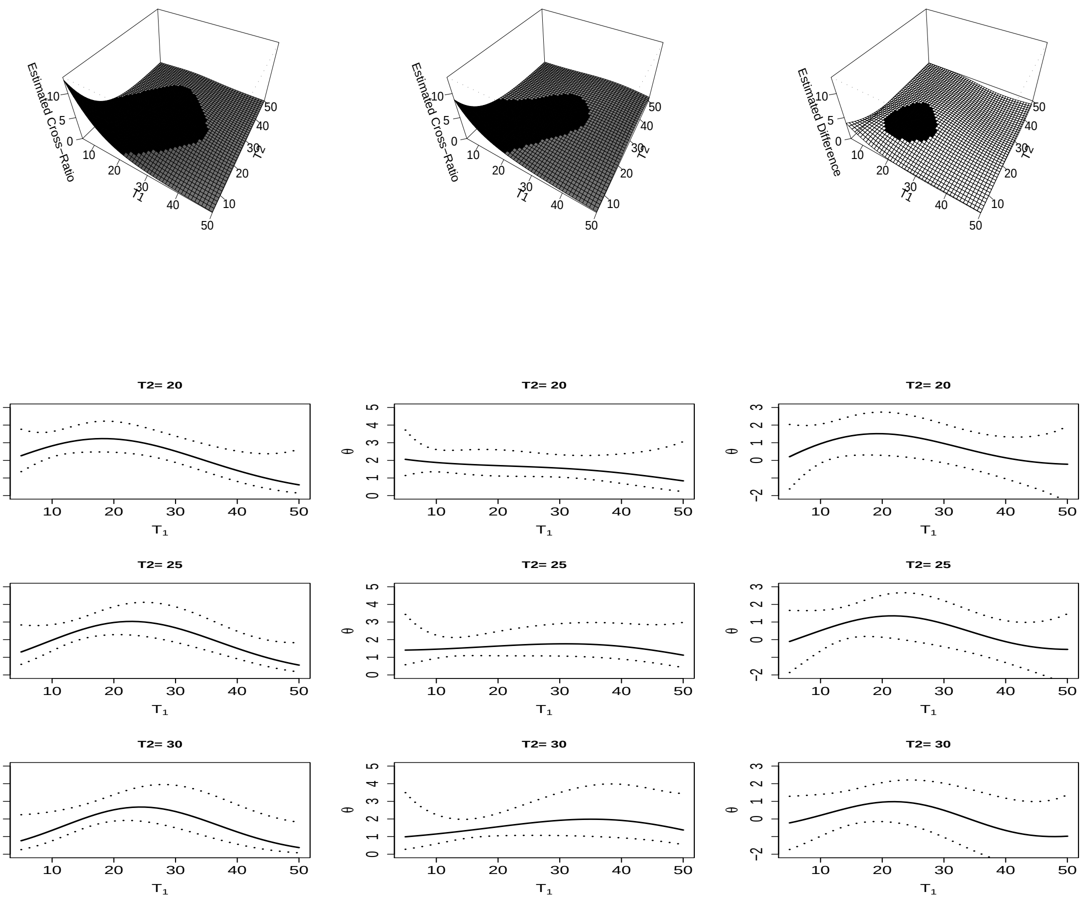


Figure 1.3: Estimated cross-ratio function for the Australian Twin Study. Figures on the left panel are for monozygotic twins, figures in the middle are for dizygotic twins, and figures on the right panel are for the difference between monozygotic twins and dizygotic twins. The shaded areas in the top left and top middle plots are regions where the association is statistically significant for monozygotic twins and dizygotic twins respectively. The shaded area in the top right plot is the region where the difference in the strength of association between monozygotic and dizygotic twins is significant based on the pointwise confidence band. In the bottom panel, the solid lines represent the cross-ratio estimate and dotted lines represent pointwise 95% confidence bands.

obtain the joint distribution $F(t_1, t_2)$. Unfortunately, closed form solution does not exist for even the simplest non-trial functional forms for $\theta(t_1, t_2)$, for example, linear functions. Our method, however, is robust because it bypasses modeling the joint and marginal distributions of the bivariate survival times.

Another interesting extension of the proposed approach is to allow the cross-ratio to vary with some covariate $W$. Prentice & Hsu (1997) proposed regression approach for the covariate effect on both the marginal hazard functions and cross ratio. Their cross-ratio, however, is not allowed to be dependent on time. The comparison of monozygotic and dizygotic twins in the Australian Twin Study presented in Figure 1.3 is equivalent to the estimation from an interaction model: $\beta(t_1, t_2, W) = \boldsymbol{z}(t_1, t_2)'\boldsymbol{\gamma}_1 + W\boldsymbol{z}(t_1, t_2)'\boldsymbol{\gamma}_2$, where $W$ is a binary $\{0, 1\}$ covariate. In contrast, if $W$ is not binary, we get a more parsimonious model than doing analysis separately at each level of $W$, which may be important for efficiency and for modeling the functional form of $W$. Yes, the most intuitive and straightforward model for the cross-ratio with covariates would be $\beta(t_1, t_2, \boldsymbol{W}) = \boldsymbol{z}(t_1, t_2)'\boldsymbol{\gamma} + \boldsymbol{W}'\boldsymbol{\alpha}$ for a covariate vector $\boldsymbol{W}$, which is simpler than the above model with interactions. Such extension will be investigated in Chapter II.

Due to the partial likelihood type of construction in (1.4), the proposed approach can be easily modified to handle left truncation in addition to right censoring. Let $(U_{1i}, U_{2i})$ be the truncation times for subject $i$, then (1.4) can be replaced by

$$\prod_{j=1}^{n}\left[\frac{\theta(X_{1i}, X_{2j})^{I(X_{1j}=X_{1i})\Delta_{1j}}}{N(X_{1i}, X_{2j}) - I(U_{2i} \le X_{2j} \le X_{2i})(1 - \theta(X_{1i}, X_{2j}))}\right]^{I(X_{1j}\ge X_{1i}\ge U_{1j})\Delta_{2j}\Delta_{1i}},$$

where $N(t_1, t_2) = \sum_{k=1}^{n} I(X_{1k} \ge t_1 \ge U_{1k}, X_{2k} \ge t_2 \ge U_{2k})$. This extension will be explored in Chapter III.

## 1.7 Appendix: Proofs

### 1.7.1 Proof of Theorem I.1

Let $(X_1^*, X_2^*, \Delta_1^*, \Delta_2^*)$ be an independent copy of $(X_1, X_2, \Delta_1, \Delta_2)$. Define the deterministic function $\boldsymbol{u}(\boldsymbol{\gamma}) = \boldsymbol{u}^{(1)}(\boldsymbol{\gamma}) - \boldsymbol{u}^{(2)}(\boldsymbol{\gamma}) + \boldsymbol{u}^{(3)}(\boldsymbol{\gamma}) - \boldsymbol{u}^{(4)}(\boldsymbol{\gamma})$, where

$$\begin{aligned}
\boldsymbol{u}^{(1)}(\boldsymbol{\gamma}) &= \boldsymbol{u}^{(3)}(\boldsymbol{\gamma}) = E\{\Delta_1\Delta_2\boldsymbol{z}(X_1, X_2)\},\\
\boldsymbol{u}^{(2)}(\boldsymbol{\gamma}) &= \boldsymbol{u}^{(4)}(\boldsymbol{\gamma}) = E\left[\Delta_1^*\Delta_2\boldsymbol{z}(X_1^*, X_2)\frac{I(X_1 \ge X_1^*)I(X_2 \le X_2^*)\exp\{\beta(X_1^*, X_2; \boldsymbol{\gamma})\}}{S_{X_1,X_2}(X_1^*, X_2)}\right],
\end{aligned}$$

and $S_{X_1,X_2}(\cdot,\cdot)$ is the bivariate survival function of the observation time $(X_1, X_2)$. We will first show that $\boldsymbol{U}_n^{(k)}(\boldsymbol{\gamma})$ converges uniformly to $\boldsymbol{u}^{(k)}(\boldsymbol{\gamma})$, $k = 1, \ldots, 4$, then show that $\boldsymbol{u}(\boldsymbol{\gamma}) = \boldsymbol{0}$ has the unique solution at $\boldsymbol{\gamma}_0$, and finally show the consistency of $\hat{\boldsymbol{\gamma}}_n$ that is the solution of $\boldsymbol{U}_n(\boldsymbol{\gamma}) = \boldsymbol{0}$. Due to the symmetric construction, we only need to show the convergence of $\boldsymbol{U}_n^{(1)}$ and $\boldsymbol{U}_n^{(2)}$.

Define the following simplified notation:

$$\partial_1 F(t_1,t_2) = \frac{\partial F(t_1,t_2)}{\partial t_1}, \quad \partial_2 F(t_1,t_2) = \frac{\partial F(t_1,t_2)}{\partial t_2}, \quad \partial_{1,2} F(t_1,t_2) = \frac{\partial^2 F(t_1,t_2)}{\partial t_1 \partial t_2},$$
$$\partial_1 G(t_1,t_2) = \frac{\partial G(t_1,t_2)}{\partial t_1}, \quad \partial_2 G(t_1,t_2) = \frac{\partial G(t_1,t_2)}{\partial t_2}, \quad \partial_{1,2} G(t_1,t_2) = \frac{\partial^2 G(t_1,t_2)}{\partial t_1 \partial t_2},$$

where $F$ and $G$ denote the joint survival functions of $(T_1, T_2)$ and $(C_1, C_2)$, respectively. Then the joint density function of $(X_1, X_2, \Delta_1, \Delta_2)$ can be written as

$$\begin{aligned} p(t_1,t_2,\delta_1,\delta_2) = \partial_{1,2}F(t_1,t_2)^{\delta_1\delta_2}\{-\partial_1 F(t_1,t_2)\}^{\delta_1(1-\delta_2)}\{-\partial_2 F(t_1,t_2)\}^{(1-\delta_1)\delta_2} \\ F(t_1,t_2)^{(1-\delta_1)(1-\delta_2)}\partial_{1,2}G(t_1,t_2)^{(1-\delta_1)(1-\delta_2)}\{-\partial_1 G(t_1,t_2)\}^{(1-\delta_1)\delta_2} \\ \{-\partial_2 G(t_1,t_2)\}^{\delta_1(1-\delta_2)}G(t_1,t_2)^{\delta_1\delta_2}. \end{aligned} \tag{1.10}$$

Following the notation of van der Vaart & Wellner (1996), we use $\mathbb{P}_n$ and $\mathbb{Q}_n$ to denote the empirical measures of $n$ independent copies of $(X_1^*, X_2^*, \Delta_1^*, \Delta_2^*)$ and $(X_1, X_2, \Delta_1, \Delta_2)$ that follow the distributions $P$ and $Q$, respectively. Although these two samples are in fact identical, i.e., $\mathbb{P}_n = \mathbb{Q}_n$ and $P = Q$, we use different letters to keep the notation tractable for the double summations, which will soon become clear in the following calculations.

For model (1.7), $\boldsymbol{U}_n^{(1)}(\boldsymbol{\gamma}) = \mathbb{Q}_n \Delta_1 \Delta_2 \boldsymbol{z}(X_1, X_2)$ that is free of $\boldsymbol{\gamma}$, and $\Delta_1$, $\Delta_2$ and $\boldsymbol{z}(X_1, X_2)$ are all bounded, hence by the law of large numbers, it is trivial to obtain

$$\sup_{\gamma} |\boldsymbol{U}_n^{(1)} - \boldsymbol{u}^{(1)}| = |(\mathbb{Q}_n - Q)\Delta_1\Delta_2 \boldsymbol{z}(X_1, X_2)| \to \mathbf{0}$$

either almost surely or in probability. Convergence in probability should be adequate here for the proof.

To show the uniform convergence of $\boldsymbol{U}_n^{(2)}(\boldsymbol{\gamma})$, we first define the following quantities:

$$\begin{aligned} g^{(n)}(\Delta_2, X_1, X_2, \Delta_1^*, X_1^*, X_2^*; \boldsymbol{\gamma}) &= \frac{\Delta_1^* \Delta_2 \boldsymbol{z}(X_1^*, X_2) I(X_1 \geq X_1^*) I(X_2 \leq X_2^*) e^{\beta(X_1^*, X_2; \gamma)}}{\frac{1}{n}(N(X_1^*, X_2) - I(X_2 \leq X_2^*)\{1 - e^{\beta(X_1^*, X_2; \gamma)}\})}, \\ \tilde{g}(\Delta_2, X_1, X_2, \Delta_1^*, X_1^*, X_2^*; \boldsymbol{\gamma}) &= \frac{\Delta_1^* \Delta_2 \boldsymbol{z}(X_1^*, X_2) I(X_1 \geq X_1^*) I(X_2 \leq X_2^*) e^{\beta(X_1^*, X_2; \gamma)}}{S_{X_1, X_2}(X_1^*, X_2)}. \end{aligned}$$

The only difference between the two expressions is in the denominators of the two fractions. By fixing

$(\Delta_1^*, X_1^*, X_2^*)$ at $(\delta_1, x_1, x_2)$, we also define

$$
\begin{aligned}
h^{(n)}_{\mathbb{Q}_n}(\delta_1, x_1, x_2; \boldsymbol{\gamma}) &= \mathbb{Q}_n g^{(n)}(\Delta_2, X_1, X_2, \delta_1, x_1, x_2; \boldsymbol{\gamma}), \\
\tilde{h}_Q(\delta_1, x_1, x_2; \boldsymbol{\gamma}) &= Q\tilde{g}(\Delta_2, X_1, X_2, \delta_1, x_1, x_2; \boldsymbol{\gamma}).
\end{aligned}
$$

Similarly, fixing $(\Delta_2, X_1, X_2)$ at $(\delta_2, x_1, x_2)$, define

$$
\begin{aligned}
h^{(n)}_{\mathbb{P}_n}(\delta_2, x_1, x_2; \boldsymbol{\gamma}) &= \mathbb{P}_n g^{(n)}(\delta_2, x_1, x_2, \Delta_1^*, X_1^*, X_2^*; \boldsymbol{\gamma}), \\
h^{(n)}_{P}(\delta_2, x_1, x_2; \boldsymbol{\gamma}) &= P g^{(n)}(\delta_2, x_1, x_2, \Delta_1^*, X_1^*, X_2^*; \boldsymbol{\gamma}), \\
\tilde{h}_P(\delta_2, x_1, x_2; \boldsymbol{\gamma}) &= P\tilde{g}(\delta_2, x_1, x_2, \Delta_1^*, X_1^*, X_2^*; \boldsymbol{\gamma}).
\end{aligned}
$$

Then we have $\boldsymbol{U}_n^{(2)}(\boldsymbol{\gamma}) = \mathbb{P}_n h^{(n)}_{\mathbb{Q}_n}$ and $\boldsymbol{u}^{(2)}(\boldsymbol{\gamma}) = P\tilde{h}_Q$. It is clear that, under Conditions AC1.1 and C1.2, the summation and integration are interchangeable, which yields $Ph^{(n)}_{\mathbb{Q}_n} = \mathbb{Q}_n h^{(n)}_P$. Thus by $P\tilde{h}_Q = PQ\tilde{g}$ and $Qh^{(n)}_P = QPg^{(n)}$, applying the triangle inequality we obtain

$$
\begin{aligned}
\sup_{\gamma} |\boldsymbol{U}_n^{(2)} - \boldsymbol{u}^{(2)}| &= \sup_{\gamma} |\mathbb{P}_n h^{(n)}_{\mathbb{Q}_n} - P\tilde{h}_Q| \\
&\leq \sup_{\gamma} |\mathbb{P}_n h^{(n)}_{\mathbb{Q}_n} - P h^{(n)}_{\mathbb{Q}_n}| + \sup_{\gamma} |P h^{(n)}_{\mathbb{Q}_n} - Q h^{(n)}_P| + \sup_{\gamma} |Q h^{(n)}_P - QP\tilde{g}| \\
&\leq \sup_{\gamma} |\mathbb{P}_n h^{(n)}_{\mathbb{Q}_n} - P h^{(n)}_{\mathbb{Q}_n}| + \sup_{\gamma} |\mathbb{Q}_n h^{(n)}_P - Q h^{(n)}_P| + \sup_{\omega,\gamma} |g^{(n)} - \tilde{g}|,
\end{aligned}
$$

where $\omega$ represents all the arguments of functions $g^{(n)}$ and $\tilde{g}$. For model (1.7), it is straightforward to argue that, under Conditions C1.1 and C1.2, functions $g^{(n)}$ and $\tilde{g}$ are Donsker. Then by Theorems 2.10.2 and 2.10.3 in van der Vaart & Wellner (1996), we know that $h^{(n)}_{\mathbb{Q}_n}$ and $h^{(n)}_P$ are Donsker, and hence Glivenco-Cantelli. Thus the first and second terms on the right hand side of the above last inequality converge to zero in probability. For the third term, it is easy to see that

$$
\frac{\Delta_1^* \Delta_2 \boldsymbol{z}(X_1^*, X_2) I(X_1 \geq X_1^*) I(X_2 \leq X_2^*) e^{\beta(X_1^*, X_2; \gamma)}}{\frac{1}{n}[N(X_1^*, X_2) - I(X_2 \leq X_2^*)\{1 - e^{\beta(X_1^*, X_2; \gamma)}\}] S_{X_1, X_2}(X_1^*, X_2)}
$$

is bounded, say by $K_n$ that has a finite limit, and $I(X_2 \leq X_2^*)\{1 - e^{\beta(X_1^*, X_2; \gamma)}\}$ is also bounded, say by $K^*$. Then

$$
\sup_{\omega,\gamma} |g^{(n)} - \tilde{g}| \leq K_n \sup_{t_1, t_2} |n^{-1} N(t_1, t_2) - S_{X_1, X_2}(t_1, t_2)| + \frac{K_n \cdot K^*}{n} \to 0
$$

in probability. Then $\boldsymbol{U}_n^{(2)}(\boldsymbol{\gamma})$ converges uniformly to $\boldsymbol{u}^{(2)}(\boldsymbol{\gamma})$ in probability. Thus we have shown that $\boldsymbol{U}_n(\boldsymbol{\gamma})$ converges uniformly to $\boldsymbol{u}(\boldsymbol{\gamma})$ in probability.

To show $\boldsymbol{u}(\boldsymbol{\gamma}_0) = 0$, it suffices to show $\boldsymbol{u}^{(1)}(\boldsymbol{\gamma}_0) = \boldsymbol{u}^{(2)}(\boldsymbol{\gamma}_0)$. We now calculate $\boldsymbol{u}^{(2)}(\boldsymbol{\gamma}_0)$ directly. Recall that $(X_1, X_2, \Delta_1, \Delta_2)$ and $(X_1^*, X_2^*, \Delta_1^*, \Delta_2^*)$ are independent and identically distributed with a density function given in ( 1.10). Thus

$$
\begin{aligned}
\boldsymbol{u}^{(2)}(\boldsymbol{\gamma}_0) &= PQ\tilde{g}(\Delta_2, X_1, X_2, \Delta_1^*, X_1^*, X_2^*; \boldsymbol{\gamma}_0) \\
&= P\left\{\int_0^\infty\int_0^\infty \sum_{\delta_1=0}^1\sum_{\delta_2=0}^1 \tilde{g}(\delta_2, t_1, t_2, \Delta_1^*, X_1^*, X_2^*; \boldsymbol{\gamma}_0)p(t_1, t_2, \delta_1, \delta_2)dt_1dt_2\right\} \\
&= P\Delta_1^*\left\{\int_0^{X_2^*} \frac{\boldsymbol{z}(X_1^*, t_2)e^{\beta(X_1^*, t_2; \boldsymbol{\gamma}_0)}}{S_{X_1,X_2}(X_1^*, t_2)}\left[\int_{X_1^*}^\infty d_1\{\partial_2 F(t_1, t_2)G(t_1, t_2)\}\right]dt_2\right\} \\
&= P\Delta_1^*\left\{-\int_0^{X_2^*} \frac{\boldsymbol{z}(X_1^*, t_2)e^{\beta(X_1^*, t_2; \boldsymbol{\gamma}_0)}}{S_{X_1,X_2}(X_1^*, t_2)}\partial_2 F(X_1^*, t_2)G(X_1^*, t_2)dt_2\right\}.
\end{aligned}
$$

Here we use $d_k$ to denote the infinitesimal change with respect to $t_k$, $k = 1, 2$. From definition (1) we can obtain that

$$
\theta(t_1, t_2; \boldsymbol{\gamma}_0) = e^{\beta(t_1, t_2; \boldsymbol{\gamma}_0)} = \frac{\partial_{1,2}F(t_1, t_2)F(t_1, t_2)}{\partial_1 F(t_1, t_2)\partial_2 F(t_1, t_2)}.
$$

Together with $S_{X_1,X_2}(t_1, t_2) = F(t_1, t_2)G(t_1, t_2)$ and integration by parts, we have

$$
\begin{aligned}
\boldsymbol{u}^{(2)}(\boldsymbol{\gamma}_0) &= P\Delta_1^*\left\{-\int_0^{X_2^*}\boldsymbol{z}(X_1^*, t_2)\frac{\partial_{1,2}F(X_1^*, t_2)}{\partial_1 F(X_1^*, t_2)}dt_2\right\} \\
&= \int_0^\infty\int_0^\infty\sum_{\delta_1=0}^1\sum_{\delta_2=0}^1\delta_1\left\{-\int_0^{s_2}\boldsymbol{z}(s_1, t_2)\frac{\partial_{1,2}F(s_1, t_2)}{\partial_1 F(s_1, t_2)}dt_2\right\}p(s_1, s_2, \delta_1, \delta_2)ds_2ds_1 \\
&= \int_0^\infty\int_0^\infty\left\{-\int_0^{s_2}\boldsymbol{z}(s_1, t_2)\frac{\partial_{1,2}F(s_1, t_2)}{\partial_1 F(s_1, t_2)}dt_2\right\}d_2\{\partial_1 F(s_1, s_2)G(s_1, s_2)\}ds_1 \\
&= \int_0^\infty\int_0^\infty \boldsymbol{z}(s_1, s_2)\partial_{1,2}F(s_1, s_2)G(s_1, s_2)ds_1ds_2 \\
&= E\{\Delta_1\Delta_2\boldsymbol{z}(X_1, X_2)\}.
\end{aligned}
$$

Note that $\boldsymbol{u}^{(1)}(\boldsymbol{\gamma})$ is in fact free of $\boldsymbol{\gamma}$ for model (1.7). We thus have shown $\boldsymbol{u}(\boldsymbol{\gamma}_0) = 0$.

To show $\boldsymbol{\gamma}_0$ is the unique solution of $u(\boldsymbol{\gamma}) = 0$, it suffices to show that (a) the matrix $\dot{\boldsymbol{u}}(\boldsymbol{\gamma}) \equiv d\boldsymbol{u}(\boldsymbol{\gamma})/d\boldsymbol{\gamma}$ is negative semidefinite for all $\boldsymbol{\gamma} \in \Gamma$, and (b) $\dot{\boldsymbol{u}}(\boldsymbol{\gamma})$ is negative definite at $\boldsymbol{\gamma}_0$. To see (a), let

$\boldsymbol{a}$ be an arbitrary vector with the same dimension as $\boldsymbol{\gamma}$, we have

$$\begin{aligned}
\boldsymbol{a}'\dot{\boldsymbol{u}}(\boldsymbol{\gamma})\boldsymbol{a} &= -2\boldsymbol{a}'\frac{d\boldsymbol{u}^{(2)}(\boldsymbol{\gamma})}{d\boldsymbol{\gamma}}\boldsymbol{a} \\
&= -2\boldsymbol{a}'E\left\{\Delta_1^*\Delta_2\boldsymbol{z}(X_1^*,X_2)^{\otimes 2}\frac{I(X_1\geq X_1^*)I(X_2\leq X_2^*)e^{\beta(X_1^*,X_2;\gamma)}}{S_{X_1,X_2}(X_1^*,X_2)}\right\}\boldsymbol{a} \\
&= -2E\left\{\Delta_1^*\Delta_2\{\boldsymbol{a}'\boldsymbol{z}(X_1^*,X_2)\}^2\frac{I(X_1\geq X_1^*)I(X_2\leq X_2^*)e^{\beta(X_1^*,X_2;\gamma)}}{S_{X_1,X_2}(X_1^*,X_2)}\right\} \\
&\leq 0.
\end{aligned}$$

To see (b), by going through a similar calculation for $\boldsymbol{u}^{(2)}(\boldsymbol{\gamma}_0)$ showing $\boldsymbol{u}^{(2)}(\boldsymbol{\gamma}_0)=\boldsymbol{u}^{(1)}(\boldsymbol{\gamma}_0)$, we have

$$\begin{aligned}
\dot{\boldsymbol{u}}(\boldsymbol{\gamma}_0) &= -2\frac{d\boldsymbol{u}^{(2)}(\boldsymbol{\gamma}_0)}{d\boldsymbol{\gamma}} \\
&= -2E\left\{\Delta_1^*\Delta_2\boldsymbol{z}(X_1^*,X_2)^{\otimes 2}\frac{I(X_1\geq X_1^*)I(X_2\leq X_2^*)e^{\beta(X_1^*,X_2;\gamma_0)}}{S_{X_1,X_2}(X_1^*,X_2)}\right\} \\
&= -2E\left\{\Delta_1\Delta_2 z(X_1,X_2)^{\otimes 2}\right\},
\end{aligned}$$

which is negative definite by Condition C1.3. Thus $\boldsymbol{\gamma}_0$ is the unique solution to $\boldsymbol{u}(\boldsymbol{\gamma})=0$.

We are now ready to show the consistency of $\hat{\boldsymbol{\gamma}}_n$. Given the fact that $\boldsymbol{U}_n(\hat{\boldsymbol{\gamma}}_n)=0$ and $\sup|\boldsymbol{U}_n(\boldsymbol{\gamma})-\boldsymbol{u}(\boldsymbol{\gamma})|=o_p(1)$, we have

$$|\boldsymbol{u}(\hat{\boldsymbol{\gamma}}_n)|=|\boldsymbol{U}_n(\hat{\boldsymbol{\gamma}}_n)-\boldsymbol{u}(\hat{\boldsymbol{\gamma}}_n)|\leq\sup_{\gamma}|\boldsymbol{U}_n(\boldsymbol{\gamma})-\boldsymbol{u}(\boldsymbol{\gamma})|=o_p(1).$$

Since $\boldsymbol{\gamma}_0$ is the unique solution to $\boldsymbol{u}(\boldsymbol{\gamma})=0$, for any fixed $\boldsymbol{\epsilon}>0$, there exists a $\boldsymbol{\delta}>0$ such that

$$\mathrm{pr}\left(|\hat{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_0|>\boldsymbol{\epsilon}\right)\leq\mathrm{pr}\left(|\boldsymbol{u}(\hat{\boldsymbol{\gamma}}_n)|>\boldsymbol{\delta}\right).$$

The consistency of $\hat{\boldsymbol{\gamma}}_n$ follows immediately.

#### 1.7.2 Proof of Theorem I.2

Define $\dot{\boldsymbol{U}}_n(\boldsymbol{\gamma})\equiv d\boldsymbol{U}_n(\boldsymbol{\gamma})/d\boldsymbol{\gamma}$. By the Taylor expansion of $\boldsymbol{U}_n(\hat{\boldsymbol{\gamma}}_n)$ around $\boldsymbol{\gamma}_0$, we have

$$n^{1/2}(\hat{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_0)=-\left\{\dot{\boldsymbol{U}}_n(\boldsymbol{\gamma}^*)\right\}^{-1}n^{1/2}\boldsymbol{U}_n(\boldsymbol{\gamma}_0), \tag{1.11}$$

where $\boldsymbol{\gamma}^*$ lies between $\hat{\boldsymbol{\gamma}}_n$ and $\boldsymbol{\gamma}_0$. By a similar calculation as in Appendix A showing the uniform consistency of $\boldsymbol{U}_n(\boldsymbol{\gamma})$, we can show that $\sup_{\boldsymbol{\gamma}} |\dot{\boldsymbol{U}}_n(\boldsymbol{\gamma}) - \dot{\boldsymbol{u}}(\boldsymbol{\gamma})| = o_p(1)$. Thus by the consistency of $\hat{\boldsymbol{\gamma}}_n$, which implies the consistency of $\boldsymbol{\gamma}^*$, and the continuity of $\dot{\boldsymbol{u}}(\boldsymbol{\gamma})$, we obtain $\dot{\boldsymbol{U}}_n(\boldsymbol{\gamma}^*) = \dot{\boldsymbol{u}}(\boldsymbol{\gamma}^*) + o_p(1) = \dot{\boldsymbol{u}}(\boldsymbol{\gamma}_0) + o_p(1)$, where $\dot{\boldsymbol{u}}(\boldsymbol{\gamma}_0) = -2E\{\Delta_1\Delta_2\boldsymbol{z}(X_1, X_2)^{\otimes 2}\} = -\boldsymbol{I}(\boldsymbol{\gamma}_0)$ is invertible by Condition C1.3. Hence based on the fact that continuity holds for the inverse operator, (1.11) can be written as

$$n^{1/2}(\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0) = \{\boldsymbol{I}(\boldsymbol{\gamma}_0)^{-1} + o_p(1)\} n^{1/2}\boldsymbol{U}_n(\boldsymbol{\gamma}_0). \tag{1.12}$$

We now need to find the asymptotic representation of $n^{1/2}\boldsymbol{U}_n(\boldsymbol{\gamma}_0)$. We only check it for $\boldsymbol{U}_n^{(1)}(\boldsymbol{\gamma}_0) - \boldsymbol{U}_n^{(2)}(\boldsymbol{\gamma}_0)$. The calculation for $\boldsymbol{U}_n^{(3)}(\boldsymbol{\gamma}_0) - \boldsymbol{U}_n^{(4)}(\boldsymbol{\gamma}_0)$ is virtually identical and yields the same asymptotic representation. It is easily seen that

$$n^{1/2}\left\{\boldsymbol{U}_n^{(1)}(\boldsymbol{\gamma}_0) - \boldsymbol{u}^{(1)}(\boldsymbol{\gamma}_0)\right\} = \mathbb{G}_n\{\Delta_1\Delta_2\boldsymbol{z}(X_1, X_2)\}, \tag{1.13}$$

where $\mathbb{G}_n = n^{1/2}(\mathbb{P}_n - P) = n^{1/2}(\mathbb{Q}_n - Q)$. We then focus on each term of the following decomposition:

$$\begin{aligned}
& n^{1/2}\left\{\boldsymbol{U}_n^{(2)}(\boldsymbol{\gamma}_0) - \boldsymbol{u}^{(2)}(\boldsymbol{\gamma}_0)\right\} \\
= \ & n^{1/2}\left(\mathbb{P}_n h_{\mathbb{Q}_n}^{(n)} - P\tilde{h}_Q\right) \\
= \ & \mathbb{G}_n\left(h_{\mathbb{Q}_n}^{(n)}\right) + n^{1/2}(P\mathbb{Q}_n g^{(n)} - PQ\tilde{g}) \\
= \ & \mathbb{G}_n\left(h_{\mathbb{Q}_n}^{(n)}\right) + \mathbb{G}_n\left(h_P^{(n)}\right) + n^{1/2}(PQg^{(n)} - PQ\tilde{g}).
\end{aligned} \tag{1.14}$$

For the first term on the right hand side of (1.14), we have $h_{\mathbb{Q}_n}^{(n)} = \left(h_{\mathbb{Q}_n}^{(n)} - \tilde{h}_{\mathbb{Q}_n}\right) + \left(\tilde{h}_{\mathbb{Q}_n} - \tilde{h}_Q\right) + \tilde{h}_Q$. It is straightforward to verify that

$$\begin{aligned}
& P\left(h_{\mathbb{Q}_n}^{(n)} - \tilde{h}_{\mathbb{Q}_n}\right)^2 = P\left(\mathbb{Q}_n g^{(n)} - \mathbb{Q}_n\tilde{g}\right)^2 \le \sup\left(g^{(n)} - \tilde{g}\right)^2 = o_p(1), \\
& P\left(\tilde{h}_{\mathbb{Q}_n} - \tilde{h}_Q\right)^2 = n^{-1}P\left\{\mathbb{G}_n(\tilde{g})\right\}^2 = o_p(1).
\end{aligned}$$

Together with the fact that $h_{\mathbb{Q}_n}^{(n)}$, $\tilde{h}_{\mathbb{Q}_n}$, and $\tilde{h}_Q$ are Donsker, we have

$$\mathbb{G}_n h_{\mathbb{Q}_n}^{(n)} = \mathbb{G}_n\left(h_{\mathbb{Q}_n}^{(n)} - \tilde{h}_{\mathbb{Q}_n}\right) + \mathbb{G}_n\left(\tilde{h}_{\mathbb{Q}_n} - \tilde{h}_Q\right) + \mathbb{G}_n\tilde{h}_Q = \mathbb{G}_n\tilde{h}_Q + o_p(1). \tag{1.15}$$

For the second term on the right hand side of (1.14), we write $h_P^{(n)} = \left(h_P^{(n)} - \tilde{h}_P\right) + \tilde{h}_P$. It can also be verified that

$$Q\left(h_P^{(n)} - \tilde{h}_P\right)^2 = Q\left\{P(g^{(n)} - \tilde{g})\right\}^2 \leq \sup_{\omega,\gamma}\left(g^{(n)} - \tilde{g}\right)^2 = o_p(1).$$

Together with the fact that $h_P^{(n)}$ and $\tilde{h}_P$ are Donsker, we obtain that

$$\mathbb{G}_n h_P^{(n)} = \mathbb{G}_n\left(h_P^{(n)} - \tilde{h}_P\right) + \mathbb{G}_n\tilde{h}_P = \mathbb{G}_n\tilde{h}_P + o_p(1). \tag{1.16}$$

We now calculate the third term on the right hand side of (1.14). First we define the following function:

$$f(\delta_1, x_1, x_2, \delta_2^*, x_1^*, x_2^*) = \frac{\delta_1^*\delta_2 z(x_1^*, x_2)I(x_1 \geq x_1^*)I(x_2 \leq x_2^*)e^{\beta(x_1^*, x_2;\gamma_0)}}{\{S_{X_1,X_2}(x_1^*, x_2)\}^2}.$$

Then we have

$$\begin{aligned}
&n^{1/2}\left(PQg^{(n)} - PQ\tilde{g}\right)\\
&= n^{1/2}\iint\left[\frac{1}{\frac{1}{n}\{N(x_1^*, x_2) - I(x_2 \leq x_2^*)(1 - e^{\beta(x_1^*, x_2;\gamma_0)})\}} - \frac{1}{S_{X_1,X_2}(x_1^*, x_2)}\right]\\
&\qquad f(\delta_1, x_1, x_2, \delta_2^*, x_1^*, x_2^*)\{S_{X_1,X_2}(x_1^*, x_2)\}^2 dP(\delta_1^*, \delta_2^*, x_1^*, x_2^*)dQ(\delta_1, \delta_2, x_1, x_2, )\\
&= -\, n^{1/2}\iint\left\{\frac{1}{n}N(x_1^*, x_2) - S_{X_1,X_2}(x_1^*, x_2)\right\}f(\delta_1, x_1, x_2, \delta_2^*, x_1^*, x_2^*)\\
&\qquad dP(\delta_1^*, \delta_2^*, x_1^*, x_2^*)dQ(\delta_1, \delta_2, x_1, x_2) + A + B\\
&= -\,\mathbb{G}_n\left\{\iint I(X_1 \geq x_1^*, X_2 \geq x_2)f(\delta_1, x_1, x_2, \delta_2^*, x_1^*, x_2^*)\right.\\
&\qquad \left. dP(\delta_1^*, \delta_2^*, x_1^*, x_2^*)dQ(\delta_1, \delta_2, x_1, x_2)\right\} + A + B,
\end{aligned} \tag{1.17}$$

where, by Conditions C1.1-C1.3,

$$\begin{aligned}
A &= n^{1/2}\iint\left(\frac{1}{\frac{1}{n}\left[N(x_1^*, x_2) - I(x_2 \leq x_2^*)\left\{1 - e^{\beta(x_1^*, x_2;\gamma_0)}\right\}\right]} - \frac{1}{\frac{1}{n}N(x_1^*, x_2)}\right)\\
&\qquad f(\delta_1, x_1, x_2, \delta_2^*, x_1^*, x_2^*)\{S_{X_1,X_2}(x_1^*, x_2)\}^2 dP(\delta_1^*, \delta_2^*, x_1^*, x_2^*)dQ(\delta_1, \delta_2, x_1, x_2)\\
&= \int\int\left(\frac{n^{-1/2}I(x_2 \leq x_2^*)\left\{1 - e^{\beta(x_1^*, x_2;\gamma_0)}\right\}}{\frac{1}{n}\left[N(x_1^*, x_2) - I(x_2 \leq x_2^*)\left\{1 - e^{\beta(x_1^*, x_2;\gamma_0)}\right\}\right]\frac{1}{n}N(x_1^*, x_2)}\right)\\
&\qquad f(\delta_1, x_1, x_2, \delta_2^*, x_1^*, x_2^*)\{S_{X_1,X_2}(x_1^*, x_2)\}^2 dP(\delta_1^*, \delta_2^*, x_1^*, x_2^*)dQ(\delta_1, \delta_2, x_1, x_2)\\
&= o_p(1),
\end{aligned} \tag{1.18}$$

and

$$\begin{aligned}
B &= n^{1/2}\iint\left\{\frac{1}{n}N(x_1^*,x_2)-S_{X_1,X_2}(x_1^*,x_2)\right\}^2\left[\frac{1}{\frac{1}{n}N(x_1^*,x_2)\{S_{X_1,X_2}(x_1^*,x_2)\}^2}\right] \\
&\quad f(\delta_1,x_1,x_2,\delta_2^*,x_1^*,x_2^*)\{S_{X_1,X_2}(x_1^*,x_2)\}^2 dP(\delta_1^*,\delta_2^*,x_1^*,x_2^*)dQ(\delta_1,\delta_2,x_1,x_2) \\
&= n^{-1/2}\iint\{\mathbb{G}_n I(X_1\geq x_1^*,X_2\geq x_2)\}^2\{\mathbb{P}_n I(X_1\geq x_1^*,X_2\geq x_2)\}^{-1} \\
&\quad f(\delta_1,x_1,x_2,\delta_2^*,x_1^*,x_2^*)dP(\delta_1^*,\delta_2^*,x_1^*,x_2^*)dQ(\delta_1,\delta_2,x_1,x_2) \\
&= o_p(1). \qquad (1.19)
\end{aligned}$$

Then by equations (1.13)-(1.19), we obtain

$$\begin{aligned}
n^{1/2}\boldsymbol{U}_n(\boldsymbol{\gamma}_0) &= n^{1/2}\{\boldsymbol{U}_n(\boldsymbol{\gamma}_0)-\boldsymbol{u}(\boldsymbol{\gamma}_0)\} \\
&= n^{1/2}\{\boldsymbol{U}_n^{(1)}(\boldsymbol{\gamma}_0)-\boldsymbol{u}^{(1)}(\boldsymbol{\gamma}_0)\}-n^{1/2}\{\boldsymbol{U}_n^{(2)}(\boldsymbol{\gamma}_0)-\boldsymbol{u}^{(2)}(\boldsymbol{\gamma}_0)\} \\
&\quad + n^{1/2}\{\boldsymbol{U}_n^{(3)}(\boldsymbol{\gamma}_0)-\boldsymbol{u}^{(3)}(\boldsymbol{\gamma}_0)\}-n^{1/2}\{\boldsymbol{U}_n^{(4)}(\boldsymbol{\gamma}_0)-\boldsymbol{u}^{(4)}(\boldsymbol{\gamma}_0)\} \\
&= 2\mathbb{G}_n\Bigg\{\Delta_1\Delta_2\boldsymbol{z}(X_1,X_2)-\tilde{h}_Q(\Delta_1,X_1,X_2;\boldsymbol{\gamma}_0)-\tilde{h}_P(\Delta_2,X_1,X_2;\boldsymbol{\gamma}_0) \\
&\quad + \iint I(X_1\geq x_1^*,X_2\geq x_2)f(\delta_1,x_1,x_2,\delta_2^*,x_1^*,x_2^*) \\
&\quad dP(\delta_1^*,\delta_2^*,x_1^*,x_2^*)dQ(\delta_1,\delta_2,x_1,x_2)\Bigg\}+o_p(1) \\
&\to N(\boldsymbol{0},\boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)) \qquad (1.20)
\end{aligned}$$

in distribution. Thus from (1.12) we obtain the desired asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_0)$. Now replacing $\tilde{h}_Q(\Delta_1,X_1,X_2;\boldsymbol{\gamma}_0)$ by $h^{(n)}_{\mathbb{Q}_n}(\Delta_1,X_1,X_2;\boldsymbol{\gamma}_0)$, $\tilde{h}_P(\Delta_2,X_1,X_2;\boldsymbol{\gamma}_0)$ by $h^{(n)}_{\mathbb{P}_n}(\Delta_2,X_1,X_2;\boldsymbol{\gamma}_0)$, $dP$ by $d\mathbb{P}_n$ and $dQ$ by $d\mathbb{Q}_n$ and plugging $\hat{\boldsymbol{\gamma}}_n$ for $\boldsymbol{\gamma}_0$ in (1.20), we can estimate $\boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)$ via sample variance. Together with $\dot{\boldsymbol{U}}_n(\hat{\boldsymbol{\gamma}}_n)$ or $\mathbb{P}_n\{\Delta_1\Delta_2\boldsymbol{z}(X_1,X_2)^{\otimes 2}\}$, both of which are an estimator for $\boldsymbol{I}(\boldsymbol{\gamma}_0)$, we can easily obtain a model based variance estimator for $\hat{\boldsymbol{\gamma}}_n$.

# CHAPTER II

# Cross-Ratio Regression

## 2.1 Introduction

In female reproductive aging research, there has been considerable interest in identifying marker events for the onset of menopausal transition and investigating their utility for predicting the age at menopause. Developing a staging system for female reproductive aging based on marker events is useful because it can help assess a woman's need for contraception and initiation of interventions such as bone density screening. One important component of a staging system is bleeding criteria, since bleeding patterns are readily observable.

In the Tremin study, conducted as part of the Menstrual and Reproductive Health Study (Treloar, Boynton, Behn, and Brown 1967), scientists are interested in understanding several bleeding pattern change criteria that have been proposed as potential marker events for the early stage of menopausal transition. For instance, it has been suggested that age at onset of experiencing a menstrual cycle length of at least 45 days might be a good marker for the early menopausal transition (Lisabeth, Harlow, Gillespie, Lin, and Sowers 2004). However, the validity of these proposed bleeding markers and their associations with age at menopause have not been adequately investigated, and sophisticated statistical analysis tools are lacking in this area.

Statistically, this problem can be formulated as estimating the dependence between censored bivariate survival times. However, formal analysis is challenging due to the fact that the 45-day cycle marker might not be useful before age 40 but might be a good marker after age 40, noted by Lisabeth et al. (2004). In other words, the dependence between these two event times varies with age at the 45-day cycle marker.

To formally assess the utility of a proposed bleeding marker, Nan et al. (2006) analyzed the association between age at a marker event (defined as age at onset of a specific bleeding pattern change) and age at natural menopause (defined as the final menstrual period (FMP), with FMP confirmed after at least 12 months of amenorrhea). They proposed using cross-ratio to measure their dependence by assuming the cross-ratio to be a piecewise constant function of age at onset of the marker event. They focused on the age at which a woman first experienced a menstrual cycle at least 45 days in length, which has been proposed as a marker event for entry into the early menopausal transition stage.

One advantage of using cross-ratio as the dependence measure is that it has an attractive hazard ratio interpretation comparing two groups of practical interest, which is simple to understand for practitioners and provides a convenient way to evaluate the marker. In particular, the cross-ratio can be interpreted as the relative hazard of menopause comparing women who have experienced the marker event at a certain age with women who have not yet experienced the marker event. However, in their proposed model, the piecewise constant assumption on the cross-ratio can be difficult to implement when prior knowledge in cut-off points is lacking.

This chapter is partially motivated by a direct application of the proposed method in Chapter I to the Tremin data. To bypass the difficulty in determining the cut-off points in the piecewise constant model, we estimate the cross-ratio as a smooth function of $t_1$ and $t_2$. Moreover, in the Tremin Trust data, since the cross-ratio of menopause and the 45-day cycle marker event are likely to be affected by age at menarche, we are interested in extending the proposed model to accommodate covariates in the cross-ratio function directly.

It is well known that when covariates exist, cross-ratio for the failure times of the two members of a pair should be estimated with some adjustment for known characteristics of the pair (Clayton 1978).

For example, in the Australian twin study of appendicitis, Duffy et al. (1990) discovered significant concordance rate with respect to appendicitis within twin pairs. It was also found that monozygotic twins exhibited higher concordance rate than dizygotic twins, likely due to shared genetic factors. Therefore, it is of interest to quantify this genetic effect on cross-ratio within twin pairs. It is also appropriate to characterize the dependence within twin pairs as a function of both times controlling for zygocity effect, since such dependence is partially due to shared environment and exposure to such shared environment is a time dependent process.

In the literature, the covariate effect is often modeled through marginal distributions. Shih & Louis (1995) proposed a model that incorporates covariates via marginal Cox regression model, assuming cross-ratio $\theta$ is constant. Likewise, when $\theta$ is piecewise constant on a grid of the sample space of $(T_1, T_2)$, Nan et al. (2006) proposed a sequential two stage method where covariates are modeled via marginal Cox regression model. Its estimation is similar to the two stage method of Shih & Louis (1995) for the Clayton copula model, but with left truncation at lower left corner of each strip. Fan and Prentice (2002) adjusted their previously proposed class of weighted dependence measures for bivariate failure times to accommodate covariate effects on marginal hazard rates as well.

However, when the cross-ratio function itself is of major interest, modeling the covariate effect via marginal models does not answer explicitly how covariates change the cross-ratio or by how much. Mimicking the Cox proportional hazards model, we propose an analogous model where the covariate effect is multiplicative on cross-ratio. One novelty of this model lies in linking the covariate effect to the cross-ratio explicitly.

For estimation, we construct an objective function, which we call the local pseudo-partial likelihood, by mimicking the partial likelihood of Cox proportional hazards model (Cox 1972). Specifically, when covariate is discrete with finite levels, we group observations into distinct strata by covariate values. We then treat whether an event happens at a time point or beyond along one time axis as a binary covariate and the other time component as the survival outcome variable, and construct the corresponding partial likelihood function. When covariate is continuous, kernel smoothing is applied to the estimating equations. We obtain the parameter estimates by maximizing the local pseudo-partial likelihood function. Such construction does not need any model for either the joint or the marginal survival function, and thus is robust against model misspecification. We show that the proposed parameter estimator is consistent. We also establish asymptotic normality of our estimator. The proposed methodology is readily extendable to the estimation of an arbitrary baseline cross-ratio function by using the tensor product splines.

### 2.2 The cross-ratio function

Let $(T_1, T_2)$ be a pair of absolutely continuous correlated failure times. In the Tremin Trust data, $T_1$ is time to the 45-day cycle marker and $T_2$ is time to menopause. Without covariate, the cross-ratio function of $T_1$ and $T_2$ (Clayton 1978; Oakes 1989) is defined as

$$\theta(t_1, t_2) = \frac{\lambda_2(t_2|T_1 = t_1)}{\lambda_2(t_2|T_1 > t_1)} = \frac{\lambda_1(t_1|T_2 = t_2)}{\lambda_1(t_1|T_2 > t_2)}, \tag{2.1}$$

where $\lambda_1$ and $\lambda_2$ are the conditional hazard functions of $T_1$ and $T_2$, respectively.

When covariate exists, e.g., age at menarche, cross-ratio is a quantity conditional on covariate. Specifically, the definition of cross-ratio becomes:

$$\theta(t_1, t_2, w) = \frac{\lambda_2(t_2|T_1 = t_1, W = w)}{\lambda_2(t_2|T_1 > t_1, W = w)} = \frac{\lambda_1(t_1|T_2 = t_2, W = w)}{\lambda_1(t_1|T_2 > t_2, W = w)}. \tag{2.2}$$

Mimicking the Cox proportional hazards model, we propose an analogous model where the effect of covariate is multiplicative on cross-ratio:

$$\theta(t_1, t_2, w) = \theta_0(t_1, t_2) \exp(w\alpha), \tag{2.3}$$

where $\theta_0(t_1, t_2)$ is the baseline cross-ratio, i.e.

$$\theta_0(t_1, t_2) = \frac{\lambda_2(t_2|T_1 = t_1, W = 0)}{\lambda_2(t_2|T_1 > t_1, W = 0)} = \frac{\lambda_1(t_1|T_2 = t_2, W = 0)}{\lambda_1(t_1|T_2 > t_2, W = 0)}.$$

Model (2.3) effectively separates the baseline cross-ratio function and the covariate effect, so that we can model each piece individually. We consider a parametric model for $\beta_0(t_1, t_2; \boldsymbol{\gamma}) \equiv \log \theta_0(t_1, t_2)$ parameterized by a finite–dimensional Euclidean parameter $\boldsymbol{\gamma}$. It is straightforward to extend the parametric model to a nonparametric model using tensor product splines. For covariate $W$, we consider a linear function parameterized by a finite–dimensional Euclidean parameter $\alpha$. We call model (2.3) the proportional cross-ratio model. We considered a parametric model for the $\beta_0(t_1, t_2; \boldsymbol{\gamma})$ without covariates in Chapter I. When there is no covariate, model (2.3) reduces to the proposed model in Chapter I. For notational simplicity, we consider one-dimensional covariate $W$ hereafter. Results developed in this article hold for any finite-dimensional discrete covariate $W$. However, due to the implementation of kernel smoothing, the developed asymptotic properties holds only for a single continuous covariate $W$.

### 2.3 Estimation

To estimate the baseline cross-ratio function and covariate effect jointly, we first focus on discrete covariate with a finite number of levels, by creating a dummy variable for each level or assuming a linear trend across levels. We then extend this method to continuous covariate using smoothing techniques, in particular, applying kernel smoothing to estimating equation obtained for discrete covariate.

Suppose we observe $n$ independent and identically distributed copies of $(X_1, X_2, \Delta_1, \Delta_2, W)$, where $X_1 = \min(T_1, C_1)$, $X_2 = \min(T_2, C_2)$, $\Delta_1 = I(T_1 \leq C_1)$, and $\Delta_2 = I(T_2 \leq C_2)$. Here $I(\cdot)$ denotes the indicator function. The pair of correlated continuous failure times $(T_1, T_2)$ are subject to right censoring by a pair of censoring times $(C_1, C_2)$. Assume censoring times are independent of failure times conditional on covariate $W$. We further assume that there is no ties among observed times for each of the two time components.

### 2.3.1 Discrete covariate with finite levels

To motivate our idea, we connect the cross-ratio definition in (2.2) with the Cox model partial likelihood for the two-group regression problem for observations within the stratum with covariate $W = w$. Using the epidemiological terminology, if we treat $\{j : T_{1j} = t_1\}$ and $\{j : T_{1j} > t_1\}$ as the "exposure" and "non-exposure" groups respectively, then from the first equality in (2.2), the cross-ratio $\theta(t_1, t_2, w)$ becomes the hazard ratio of $T_2$ between these two groups within the stratum $W = w$. Denote $\lambda_2(X_{2j}|X_{1j} > X_{1i}, W_k = W_i)$ by $A_{ij}^k$ and $\theta(X_{1i}, X_{2j}, W_i)^{I(X_{1k}=X_{1i})}$ by $B_{ij}^k$ respectively. By mimicking the partial likelihood idea, we can construct a similar objective function as follows based on these two groups categorized by $t_1 = X_{1i}$:

$$\prod_{j=1}^{n} \left[ \frac{A_{ij}^j B_{ij}^j}{\sum_{X_{2k} \geq X_{2j}} I(W_k = W_i) I(X_{1k} \geq X_{1i}) A_{ij}^k B_{ij}^k} \right]^{I(W_j = W_i) I(X_{1j} \geq X_{1i}) \Delta_{2j} \Delta_{1i}}$$

$$= \prod_{j=1}^{n} \left[ \frac{B_{ij}^j}{\sum_{X_{2k} \geq X_{2j}} I(W_k = W_i) I(X_{1k} \geq X_{1i}) B_{ij}^k} \right]^{I(W_j = W_i) I(X_{1j} \geq X_{1i}) \Delta_{2j} \Delta_{1i}},$$

where $A_{ij}^j$ cancels with$A_{ij}^k$ in the above equation because of the restriction $W = W_i$, which is achieved by indicators $I(W_j = W_i)$ in the outer exponent and $I(W_k = W_i)$ in the denominator. Following a similar argument as in Chapter I, the denominator in the bracket can be simplified as $N(X_{1i}, X_{2j}, W_i) - I(X_{2j} \leq X_{2i})(1 - \theta(X_{1i}, X_{2j}, W_i))$, where $N(t_1, t_2, w) = \sum_{k=1}^{n} I(X_{1k} \geq t_1, X_{2k} \geq t_2, W_k = w)$. So we can re-write the above objective function as

$$\prod_{j=1}^{n} \left[ \frac{\theta(X_{1i}, X_{2j}, W_i)^{I(X_{1j}=X_{1i})}}{N(X_{1i}, X_{2j}, W_i) - I(X_{2j} \leq X_{2i})(1 - \theta(X_{1i}, X_{2j}, W_i))} \right]^{I(W_i = W_j) I(X_{1j} \geq X_{1i}) \Delta_{1i} \Delta_{2j}}. \qquad (2.4)$$

Now denote (2.4) as $L_i^{(1)}$. Considering the symmetric structure of the definition of $\theta(t_1, t_2, w)$ determined by the second equality in (2.2), we can construct a similar objective function as (2.4) by switching the roles of $X_1$ and $X_2$, and denote it as $L_i^{(2)}$. By multiplying such constructed two objective functions over all possible ways of creating the "exposure" and "non-exposure" groups, i.e. all subjects, we obtain the following local pseudo-partial likelihood function:

$$L_n = \prod_{i=1}^{n} L_i^{(1)} L_i^{(2)}. \tag{2.5}$$

The estimator obtained by maximizing (2.5) is then called the maximum local pseudo-partial likelihood estimator.

Denote $l_n = n^{-1} \log L_n$, $\boldsymbol{\xi} = (\boldsymbol{\gamma}, \alpha)$ and $\dot{\beta}(t_1, t_2, w; \boldsymbol{\xi}) = \partial\beta(t_1, t_2, w; \boldsymbol{\xi})/\partial\boldsymbol{\xi}$. Differentiating $l_n(\boldsymbol{\xi})$ with respect to $\boldsymbol{\xi}$ and assuming no ties among observed times, we obtain the following estimating function for $\boldsymbol{\xi}$:

$$\boldsymbol{U}_n(\boldsymbol{\xi}) = \frac{\partial l_n(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \boldsymbol{U}_n^{(1)}(\boldsymbol{\xi}) - \boldsymbol{U}_n^{(2)}(\boldsymbol{\xi}) + \boldsymbol{U}_n^{(3)}(\boldsymbol{\xi}) - \boldsymbol{U}_n^{(4)}(\boldsymbol{\xi}),$$

where

$$\boldsymbol{U}_n^{(1)}(\boldsymbol{\xi}) = \boldsymbol{U}_n^{(3)}(\boldsymbol{\xi}) = \frac{1}{n}\sum_{i=1}^{n} \Delta_{1i}\Delta_{2i}\dot{\beta}(X_{1i}, X_{2i}, W_i; \boldsymbol{\xi}) \tag{2.6}$$

and

$$\begin{aligned}
\boldsymbol{U}_n^{(2)}(\boldsymbol{\xi}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{I(W_j = W_i)\Delta_{1i}\Delta_{2j} I(X_{1j} \geq X_{1i}) I(X_{2j} \leq X_{2i}) e^{\beta(X_{1i}, X_{2j}, W_i; \xi)}}{N(X_{1i}, X_{2j}, W_i) - I(X_{2j} \leq X_{2i})(1 - e^{\beta(X_{1i}, X_{2j}, W_i; \xi)})} \\
&\quad \times \dot{\beta}(X_{1i}, X_{2j}, W_i; \boldsymbol{\xi}), \qquad (2.7) \\
\boldsymbol{U}_n^{(4)}(\boldsymbol{\xi}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{I(W_j = W_i)\Delta_{1j}\Delta_{2i} I(X_{2j} \geq X_{2i}) I(X_{1j} \leq X_{1i}) e^{\beta(X_{1j}, X_{2i}, W_i; \xi)}}{N(X_{1j}, X_{2i}, W_i) - I(X_{1j} \leq X_{1i})(1 - e^{\beta(X_{1j}, X_{2i}, W_i; \xi)})} \\
&\quad \times \dot{\beta}(X_{1j}, X_{2i}, W_i; \boldsymbol{\xi}). \qquad (2.8)
\end{aligned}$$

Note that by switching indices $i$ and $j$, (2.7) and (2.8) only differ in the second term of their denominators, which is a negligible term asymptotically. Then an estimator $\hat{\boldsymbol{\xi}}_n$ can be obtained by solving the equation $\boldsymbol{U}_n(\boldsymbol{\xi}) = 0$ using the Newton-Raphson algorithm.

#### 2.3.2 Continuous covariate

When covariate is continuous, the "grouping" idea by restricting observations with the same covariate values into distinct strata is no longer applicable. However, based on the estimating equations obtained for discrete covariate, we replace the grouping indicator function $I(W_j = W_i)$ by a kernel function $\mathbf{K}_h(W_j - W_i)$ in (2.7) and (2.8), where $\mathbf{K}_h(\cdot) = 1/h\mathbf{K}(\cdot/h)$ and $h$ is a bandwidth. Function $\mathbf{K}(\cdot)$ is usually chosen to be a symmetric probability density function. In the numerical study presented later, we use the standard normal kernel. Specifically, we propose the following estimating function for $\boldsymbol{\xi}$ when the covariate is continuous:

$$\boldsymbol{U}_n(\boldsymbol{\xi}) = \boldsymbol{U}_n^{(1)}(\boldsymbol{\xi}) - \boldsymbol{U}_n^{(2)}(\boldsymbol{\xi}) + \boldsymbol{U}_n^{(3)}(\boldsymbol{\xi}) - \boldsymbol{U}_n^{(4)}(\boldsymbol{\xi}),$$

where $\boldsymbol{U}_n^{(1)}$ and $\boldsymbol{U}_n^{(3)}$ are the same as in (2.6) and

$$\begin{aligned}
\boldsymbol{U}_n^{(2)}(\boldsymbol{\xi}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i)\Delta_{1i}\Delta_{2j} I(X_{1j} \geq X_{1i}) I(X_{2j} \leq X_{2i}) e^{\beta(X_{1i}, X_{2j}, W_i; \xi)}}{N(X_{1i}, X_{2j}, W_i) - \mathbf{K}_h(0) I(X_{2j} \leq X_{2i})(1 - e^{\beta(X_{1i}, X_{2j}, W_i; \xi)})} \\
&\quad \times \dot{\beta}(X_{1i}, X_{2j}, W_i; \boldsymbol{\xi}), \\
\boldsymbol{U}_n^{(4)}(\boldsymbol{\xi}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i)\Delta_{1j}\Delta_{2i} I(X_{2j} \geq X_{2i}) I(X_{1j} \leq X_{1i}) e^{\beta(X_{1j}, X_{2i}, W_i; \xi)}}{N(X_{1j}, X_{2i}, W_i) - \mathbf{K}_h(0) I(X_{1j} \leq X_{1i})(1 - e^{\beta(X_{1j}, X_{2i}, W_i; \xi)})} \\
&\quad \times \dot{\beta}(X_{1j}, X_{2i}, W_i; \boldsymbol{\xi}),
\end{aligned}$$

where $N(t_1, t_2, w) = \sum_{k=1}^{n} I(X_{1k} \geq t_1, X_{2k} \geq t_2)\mathbf{K}_h(W_k - w)$. Then an estimator $\hat{\boldsymbol{\xi}}_n$ can be obtained by solving the equation $\boldsymbol{U}_n(\boldsymbol{\xi}) = 0$ using the Newton-Raphson algorithm.

## 2.4 Asymptotic properties

We consider a parametric model for $\theta_0(t_1, t_2)$ in (2.3). In particular, we assume

$$\beta(t_1, t_2, w; \boldsymbol{\xi}) = \sum_{k,l} \gamma_{kl} b_{kl}(t_1, t_2) + w\alpha, \tag{2.9}$$

where $\boldsymbol{\xi}$ is the finite dimensional vector of coefficients $\{\gamma_{kl}\}$ and $\alpha$, and $\{b_{kl}\}$ are the basis functions of $t_1$ and $t_2$ that do not involve the parameter $\boldsymbol{\xi}$. In this section, we provide asymptotic results for the estimation of $\boldsymbol{\xi}$ in (2.9). In particular, we consider functions of bounded variations for $\{b_{kl}\}$ such that both $\beta(t_1, t_2, w; \boldsymbol{\xi})$ and $\dot{\beta}(t_1, t_2, w)$ belong to Donsker classes. Note that $\dot{\beta}$ for model (2.9) is free of $\boldsymbol{\xi}$. Such property plays an important role in the proofs of the following theorems. We consider the following regularity conditions for model (2.9):

C2.1. The covariate $W$ is either continuous or discrete with finite levels, whose sample space $\mathcal{W}$ is bounded with $0 < \inf_{w \in \mathcal{W}} f(w)$ and $\sup_{w \in \mathcal{W}} f(w) < \infty$. Here $f$ is the density function of $W$.

C2.2. The failure times are truncated at $(\tau_1, \tau_2)$, $0 < \tau_1, \tau_2 < \infty$, such that $\sup_{w \in \mathcal{W}} Pr(T_1 > \tau_1, C_1 > \tau_1, T_2 > \tau_2, C_2 > \tau_2 | W = w) > 0$.

C2.3. The parameter space of $\boldsymbol{\xi}$, denoted by $\Gamma$, is a compact set, and the true value $\boldsymbol{\xi}_0$ is an interior point of $\Gamma$.

C2.4. The matrix $E\{\Delta_1 \Delta_2 \dot{\beta}(X_1, X_2, W)^{\otimes 2}\}$ is positive definite. Here $\dot{\beta}^{\otimes 2} = \dot{\beta}\dot{\beta}'$.

C2.5. $(T_1, T_2)$ and $(C_1, C_2)$ are independent conditional on $W$.

In order for the kernel smoothing technique to work, following conditions are further warranted in addition to the regularity conditions previously specificized for continuous covariate with functions $\tilde{h}()$, $t()$ and $S()$ defined in the Appendix equations (2.17), (2.18) and (2.19):

C2.6. For some $\epsilon$, $0 < \epsilon \leq 1$, $\tilde{h}(V; \boldsymbol{\xi}) < \infty$ is uniformly locally Lipschitz of order $\epsilon$,

$$\sup_{x_1, x_2, \delta_1, \delta_2} \sup_{|W - W'| \leq \delta_\epsilon} |\tilde{h}(x_1, x_2, \delta_1, \delta_2, W; \boldsymbol{\xi}) - \tilde{h}(x_1, x_2, \delta_1, \delta_2, W'; \boldsymbol{\xi})| \leq M_\epsilon |W - W'|^\epsilon,$$

C2.7. $E\left(\left|\frac{t(V^*, V; \boldsymbol{\xi})}{S(X_1^*, X_2, W^*)}\right|^\lambda\right)^{1/\lambda} < \infty$ for some $\lambda$, $2 < \lambda \leq \infty$,

C2.8. Bandwidth $h$ satisfies (i) $0 \leq h \to 0$, (ii) $nh/\log n \to \infty$, (iii) $n^{1/4}h \to 0$, and (iv)$(n/\log n)^{1-2/\lambda} h \to \infty$,

C2.9. The kernel **K** is bounded and of bounded variation.

Details on conditions C2.6-C2.8 can be found in Härdle, Janssen and Serfling (1988), and conditions C2.8 (i), (ii) and C2.9 can be found in Nolan and Pollard (1987).

**Theorem II.1.** *Suppose Conditions C2.1-C2.5 hold for discrete $W$ and Conditions C2.1-C2.8 hold for continuous $W$, the solution of $\boldsymbol{U}_n(\boldsymbol{\xi}) = 0$, denoted by $\hat{\boldsymbol{\xi}}_n$, is a consistent estimator of $\boldsymbol{\xi}_0$.*

The proof of Theorem II.1 is treated separately for discrete $W$ with finite levels and continuous $W$, but follows similar steps. We first show that $\boldsymbol{U}_n(\boldsymbol{\xi})$ converges to a deterministic function $\boldsymbol{u}(\boldsymbol{\xi})$ uniformly, then show that $\boldsymbol{u}(\boldsymbol{\xi})$ is monotone and has a unique root at $\boldsymbol{\xi}_0$. Then consistency follows easily. Details are provided in the Appendix.

**Theorem II.2.** *Suppose Conditions C2.1-C2.5 hold for discrete $W$ and Conditions C2.1-C2.9 hold for continuous $W$, we have that $n^{1/2}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0)$ converges in distribution to a normal random variable with mean zero and variance $\boldsymbol{I}(\boldsymbol{\xi}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)\boldsymbol{I}(\boldsymbol{\xi}_0)^{-1}$, where $\boldsymbol{I}(\boldsymbol{\xi}_0) = 2E\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2, W)^{\otimes 2}\}$ and $\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)$ is the asymptotic variance of $\boldsymbol{U}_n(\boldsymbol{\xi}_0)$, which is given in the Appendix.*

The asymptotic normality in Theorem II.2 can be achieved by using the Taylor expansion of $\boldsymbol{U}_n(\hat{\boldsymbol{\xi}}_n)$ around $\boldsymbol{\xi}_0$. Again the detailed calculation which centers on the linearization of $U_n(\boldsymbol{\xi}_0) - u(\boldsymbol{\xi}_0)$ is deferred to the Appendix. The asymptotic expression of $\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)$ also provides a variance estimator of $n^{1/2}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0)$.

## 2.5 Simulations

### 2.5.1 Discrete covariate with finite levels

We conduct simulations to assess the performance of the proposed method. Generating data from a bivariate distribution with an arbitrary cross-ratio function is impossible because there is no corresponding closed form survival function in general. Moreover, unlike in Chapter I, we can no longer generate data from Frank family when we have covariate since Frank family does not accommodate multiplicative covariate effect. Nevertheless, we are able to generate data from the clayton model and piecewise constant cross-ratio model. For simplicity, we assume $W$ is a binary random variable from $Bernoulli(0.5)$. We generate data for $\beta_0(t_1, t_2; \boldsymbol{\gamma}) = 0.25$ and $\alpha = 0.5$. Such setup is equivalent to generating data from two Clayton models with $\theta = e^{0.25}$ when $W = 0$ and $\theta = e^{0.75}$ when $W = 1$. The basis functions used for the estimation are $1$, $X_1$ and $X_2$, though only the intercept term is needed in the true model. The results based on sample sizes of 400 and 800 are summarized in Table 2.1, where $\alpha$ is the true covariate effect and $\gamma$'s are the true coefficients for the basis functions $1$, $X_1$ and $X_2$ respectively. Simulation results based on 1000 replications show that our estimators work well. The model based variance estimator also works well since the empirical coverage probabilities are all close to the $95\%$ nominal value.

Table 2.1: Cross-ratio regression for discrete covariate with $\alpha = 0.5$ and constant baseline cross-ratio with $\beta_0 = 0.25$. $\hat{\alpha}$ and $\hat{\gamma}$, point estimate average; $E.SE$, the empirical standard error; $M.SE$, the average of the model based standard error estimates; $M.CP$, the $95\%$ coverage probability.

| | n=400 | | | | n=800 | | | |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\hat{\alpha}$ | $E.SE$ | $M.SE$ | $M.CP$ | $\hat{\alpha}$ | $E.SE$ | $M.SE$ | $M.CP$ |
| 0.50 | 0.51 | 0.19 | 0.18 | 95% | 0.50 | 0.12 | 0.13 | 96% |
| $\gamma$ | $\hat{\gamma}$ | $E.SE$ | $M.SE$ | $M.CP$ | $\hat{\gamma}$ | $E.SE$ | $M.SE$ | $M.CP$ |
| 0.25 | 0.24 | 0.19 | 0.19 | 96% | 0.25 | 0.13 | 0.13 | 96% |
| 0 | 0.02 | 0.23 | 0.22 | 96% | 0.01 | 0.15 | 0.15 | 96% |
| 0 | 0.02 | 0.22 | 0.22 | 95% | 0.01 | 0.14 | 0.15 | 95% |

Table 2.2: Cross-ratio regression for discrete covariate with $\alpha = 0.5$ and the piecewise constant baseline cross-ratio. $\hat{\alpha}$ and $\hat{\beta}$, point estimate average; $E.SE$, the empirical standard error; $M.SE$, the average of the model based standard error estimates; $M.CP$, the $95\%$ coverage probability.

| | | n=400 | | | | n=800 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | $\hat{\alpha}$ | $E.SE$ | $M.SE$ | $M.CP$ | $\hat{\alpha}$ | $E.SE$ | $M.SE$ | $M.CP$ |
| | 0.50 | 0.50 | 0.21 | 0.21 | 95% | 0.50 | 0.14 | 0.14 | 96% |
| $\theta$ | $\beta$ | $\hat{\beta}$ | $E.SE$ | $M.SE$ | $M.CP$ | $\hat{\beta}$ | $E.SE$ | $M.SE$ | $M.CP$ |
| 0.90 | -0.11 | -0.10 | 0.18 | 0.18 | 95% | -0.10 | 0.13 | 0.13 | 94% |
| 2.00 | 0.69 | 0.72 | 0.20 | 0.19 | 94% | 0.70 | 0.13 | 0.13 | 95% |
| 4.00 | 1.39 | 1.41 | 0.23 | 0.24 | 96% | 1.41 | 0.16 | 0.16 | 95% |
| 1.50 | 0.41 | 0.41 | 0.26 | 0.25 | 94% | 0.42 | 0.17 | 0.17 | 94% |

To mimic the analysis results of the Tremin Trust data, We also simulate data using algorithm in Nan et al. (2006) with a binary covariate $W \sim Bernoulli(0.5)$ and $\alpha = 0.5$. For $W = 0$, the cross-ratio is piecewise constant over four intervals: $\theta = .9$ when $t_1 \in [0, .25)$, $\theta = 2.0$ when $t_1 \in [.25, .5)$, $\theta = 4.0$ when $t_1 \in [.5, .75)$, and $\theta = 1.5$ when $t_1 > .75$. For $W = 1$, the cross-ratio $\theta$ is equal to $0.9 \times e^{0.5}$, $2.0 \times e^{0.5}$, $4.0 \times e^{0.5}$ and $1.5 \times e^{0.5}$ in the above intervals. The results in Table 2.2 show that our estimators as well as their model based variance estimators all work well.

### 2.5.2 Continuous covariate

Like the simulations for the discrete covariate, we simulate data with $W \sim unif(-0.5, 0.5)$ and $\alpha = 0.5$ assuming the same Clayton model and piecewise constant model for the baseline cross-ratio. To save computing cost, we simply choose $h = 100n^{-1/3}$. Simulation results with sample sizes equal to 400 and 800 are summarized in Table 2.3 and 2.4. We plot the normal Q-Q plot of $\hat{\boldsymbol{\xi}}$ in Figure 2.1 for the piecewise constant cross-ratio model with n=1600 based on 1000 replications. The plot supports our conclusion that $\hat{\boldsymbol{\xi}}$ is asymptotically normal.

Table 2.3: Cross-ratio regression for continuous covariate with $\alpha = 0.5$ and constant baseline cross-ratio with $\beta_0 = 0.25$. $\hat{\alpha}$ and $\hat{\gamma}$, point estimate average; $E.SE$, the empirical standard error; $M.SE$, the average of the model based standard error estimates; $M.CP$, the $95\%$ coverage probability.

| | n=400 | | | | n=800 | | | |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\hat{\alpha}$ | $E.SE$ | $M.SE$ | $M.CP$ | $\hat{\alpha}$ | $E.SE$ | $M.SE$ | $M.CP$ |
| 0.50 | 0.48 | 0.30 | 0.29 | 94% | 0.48 | 0.20 | 0.21 | 95% |
| $\gamma$ | $\hat{\gamma}$ | $E.SE$ | $M.SE$ | $M.CP$ | $\hat{\gamma}$ | $E.SE$ | $M.SE$ | $M.CP$ |
| 0.25 | 0.24 | 0.17 | 0.17 | 94% | 0.25 | 0.11 | 0.11 | 95% |
| 0 | 0.02 | 0.22 | 0.21 | 95% | 0.01 | 0.15 | 0.14 | 95% |
| 0 | 0.02 | 0.22 | 0.21 | 96% | 0.00 | 0.14 | 0.15 | 95% |

Table 2.4: Cross-ratio regression for continuous covariate with $\alpha = 0.5$ and the piecewise constant baseline cross-ratio. $\hat{\alpha}$ and $\hat{\beta}$, point estimate average; $E.SE$, the empirical standard error; $M.SE$, the average of the model based standard error estimates; $M.CP$, the $95\%$ coverage probability.

| | | n=400 | | | | n=800 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | $\hat{\alpha}$ | $E.SE$ | $M.SE$ | $M.CP$ | $\hat{\alpha}$ | $E.SE$ | $M.SE$ | $M.CP$ |
| | 0.50 | 0.46 | 0.33 | 0.34 | 96% | 0.46 | 0.24 | 0.24 | 94% |
| $\theta$ | $\beta$ | $\hat{\beta}$ | $E.SE$ | $M.SE$ | $M.CP$ | $\hat{\beta}$ | $E.SE$ | $M.SE$ | $M.CP$ |
| 0.90 | -0.11 | -0.10 | 0.15 | 0.15 | 95% | -0.10 | 0.11 | 0.10 | 94% |
| 2.00 | 0.69 | 0.71 | 0.17 | 0.17 | 95% | 0.70 | 0.12 | 0.12 | 95% |
| 4.00 | 1.39 | 1.40 | 0.21 | 0.23 | 96% | 1.40 | 0.15 | 0.15 | 96% |
| 1.50 | 0.41 | 0.41 | 0.23 | 0.23 | 95% | 0.41 | 0.15 | 0.16 | 96% |

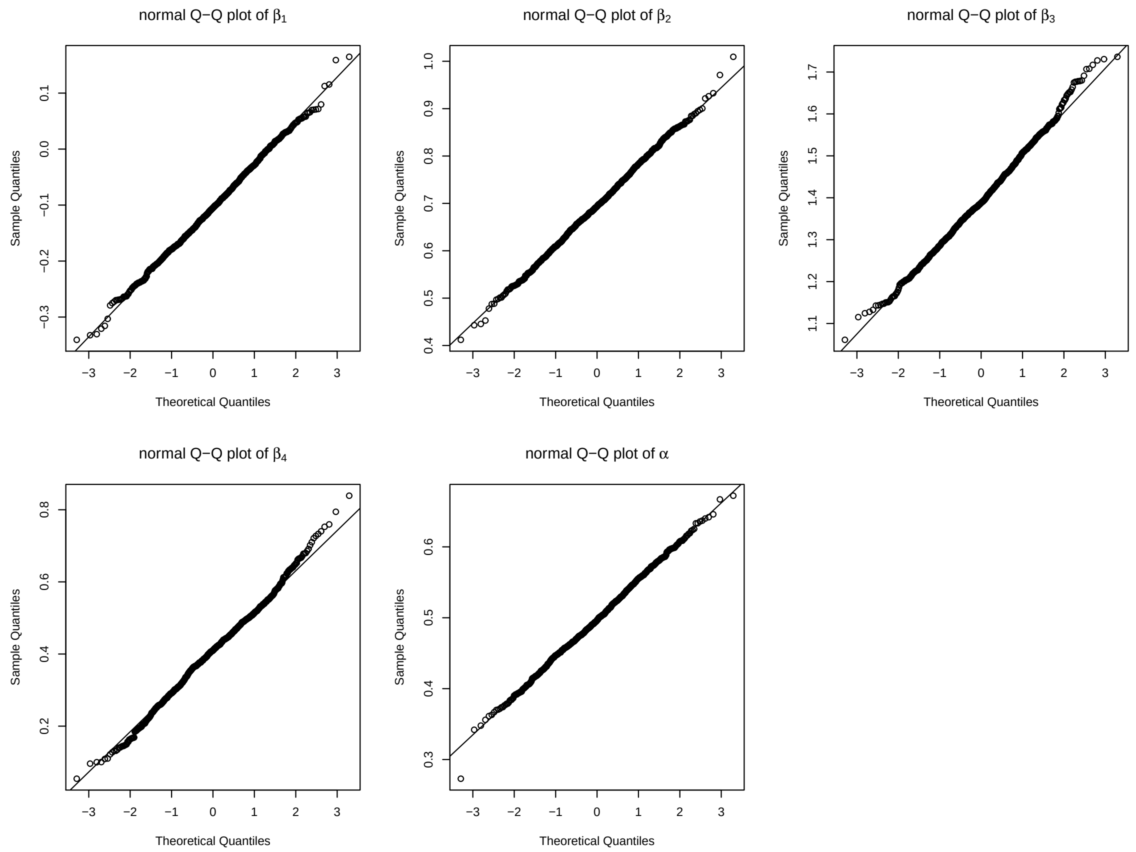


Figure 2.1: Normal Q-Q plot of parameter estimates $\hat{\boldsymbol{\xi}} = (\hat{\alpha}, \hat{\gamma})$ when the covariate $W$ is continuous for $n = 1600$ based on 1000 replications.

## 2.6 Data Analysis

### 2.6.1 The Tremin study

The Tremin Trust data were collected as part of the Menstrual and Reproductive Health Study (Treloar et al. 1967). This longitudinal cohort study followed participants throughout their reproductive life span. It provides a unique opportunity to investigate the process of female reproductive aging and menopausal transition. The study sample consisted of white college students enrolled at the University of Minnesota. Data collection started in 1935 and enrolled a sample of 1,997 women over 4 years. Study participants were followed for up to 40 years. Each woman was asked to use menstrual diary cards to record the days when bleeding was experienced. Some covariate information (e.g., age at menarche) was available.

Nan et al. (2006) used a subset of the Tremin Trust data to study age at onset of a 45-day cycle as the bleeding pattern change criteria for the early and late stages of menopausal transition. They estimated the cross-ratio as a piecewise constant function. Here we analyze the same subset that consisted of 562 women in the original study cohort who were age 25 or younger at enrollment, had information on age at menarche, and who were still participating in the study at age 35 (which they used as the baseline age in their study). Both time to a marker event and time to menopause were subject to right-censoring in the Tremin Trust data. For each individual, the censoring time was the same for both events. A total of 193 (34%) women were observed to experience natural menopause, and a total of 357 (64%) women were observed to experience a 45-day cycle marker. The median age at menopause was 51.7 years, the median age at the 45-day cycle marker was 42.7 years and the median age at menarche is 12.

To be able to compare the results with Nan et al. (2006) and for the ease of interpretation, we model the cross-ratio as a quadratic function of $t_1$ only, i.e. age at onset of 45-day cycle, based on the same data. Assuming a multiplicative effect of menarche on cross-ratio, we model the log cross-ratio as:

$$\beta(t_1, t_2, w; \gamma) = \gamma_0 + \gamma_1 t_1 + \gamma_2 t_1^2 + w\alpha, \tag{2.10}$$

where $w$ is age at menarche and compare model (2.10) with model

$$\beta(t_1, t_2, w; \gamma) = \gamma_0 + \gamma_1 t_1 + \gamma_2 t_1^2 \tag{2.11}$$

and the piecewise constant model in Nan et al. (2006). For model (2.10), we further consider three functional forms for the age at menarche: a continuous covariate with linear effect, ordinal covariate with 5 levels $(\leq 10 = 1, 11 = 2, 12 = 3, 13 = 4, \geq 14 = 5)$ with linear effect and nominal covariate with the same 5 levels. When age at menarche is treated as a continuous covariate, bandwidth $h$ is chosen to be $100n^{-1/3} \approx 10$.

We compare the results of (2.10) fitted at median age at menarche ($w = 12$) with (2.11). The general pattern of the estimated cross-ratio curvatures are open-down parabola for both model (2.10) and model (2.11). This finding is consistent with piecewise-constant result in Nan et al. (2006). However, we also caution that the age at menarche is not significant in any of the three covariate models, which yields similar results in terms of both covariate effect and baseline cross-ratio. So interpretation with respect to menarche effect should proceed with caution.

### 2.6.2 The Australian twin study revisited

In Chapter I, we analyzed the Australian twin study of appendicitis for monozygotic twin pairs and dizygotic twin pairs separately. It was found that monozygotic twins exhibited higher concordance rate than dizygotic twins. It is therefore of interest to quantify the disparity between the different type of twin pair. Additionally, it is desirable to characterize the dependence between twin pairs when the effect of zygocity is controlled for.

Since the order of twin one and twin two is arbitrary in the Australian Twin Study, we can take advantage of such symmetry to improve the estimation efficiency. Assuming a multiplicative effect of zygocity on cross-ratio, we model the log cross-ratio as:

$$\beta(t_1, t_2, w; \gamma) = \gamma_0 + \gamma_1(t_1 + t_2) + \gamma_2(t_1^2 + t_2^2) + \gamma_3 t_1 t_2 + \gamma_4(t_1^2 t_2 + t_1 t_2^2) + \gamma_5(t_1^3 + t_2^3) + w\alpha, \quad (2.12)$$

where $w$ is a binary variable that encodes monozygotic twins vs dizygotic twins.

Implementing our proposed estimating method, we obtain an estimator of $\alpha$ at $0.39$, (95% CI: 0.08 - 0.70), suggesting a genetic component to the disease. So the cross-ratio of monozygotic twins is estimated to be 1.47 times higher than that of dizygotic twins. Figure 2.2 suggests a stronger association for appendicitis risk between either monozygotic twins or dizygotic twins at a younger age. This finding is consistent with existing literature, e.g., Fan et al. (2000a). Also both monozygotic and dizygotic twin pairs are more likely to undergo appendectomy around the same time. The figure also shows that such dependence diminishes over time. This suggests that later in life environmental causes may be more important in the development of the disease.

In Figure 2.3, we compare the cross-ratio as a function of $T_1$ by fixing $T_2$ at 20, 25 and 30 estimated either by the proposed joint analysis or by the separate analysis conducted in Chapter I. The estimates

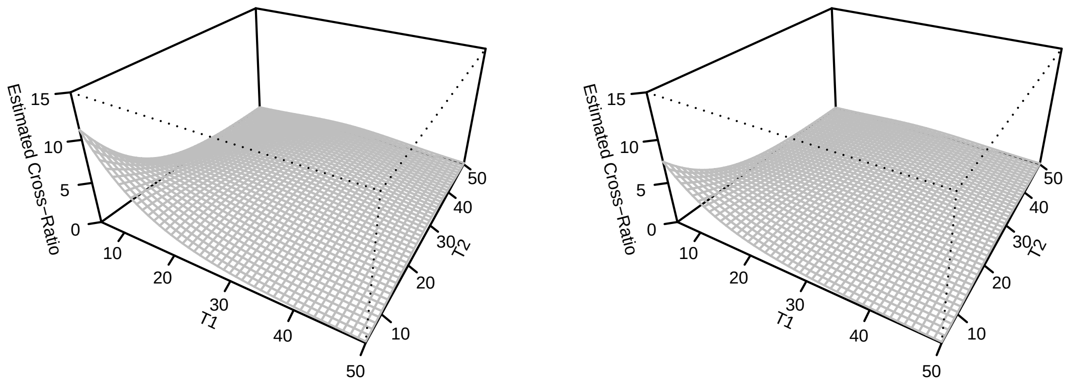


Figure 2.2: Comparison of estimated cross-ratio functions for monozygotic twins and dizygotic twins in the Australian Twin Study based on the joint analysis of model (2.12). The left panel is for female monozygotic twins and the right panel is for female dizygotic twins.

are similar, supporting the proportional cross-ratio assumption. The joint analysis, however, is more efficient.

## 2.7 Discussion

A question of significant interest in female reproductive aging is to identify bleeding criteria for menopausal transition. The Tremin Trust data provide a unique opportunity for evaluating the utility of a bleeding criterion-based marker event by assessing the association between age at onset of the bleeding marker and age at onset of menopause. Formal statistical analysis of this dependence is challenging due to the facts that both the marker event and menopause are subject to right-censoring and that their association depends on age at the marker event. In this chapter, we consider a cross-ratio regression model estimating the dependence between the marker event and the event of primary interest adjusting for age at menarche, where the log cross-ratio is assumed to be a smooth polynomial function of the marker event time and covariate.

In the cross-ratio regression, we essentially separated the baseline log cross-ratio and covariate effect

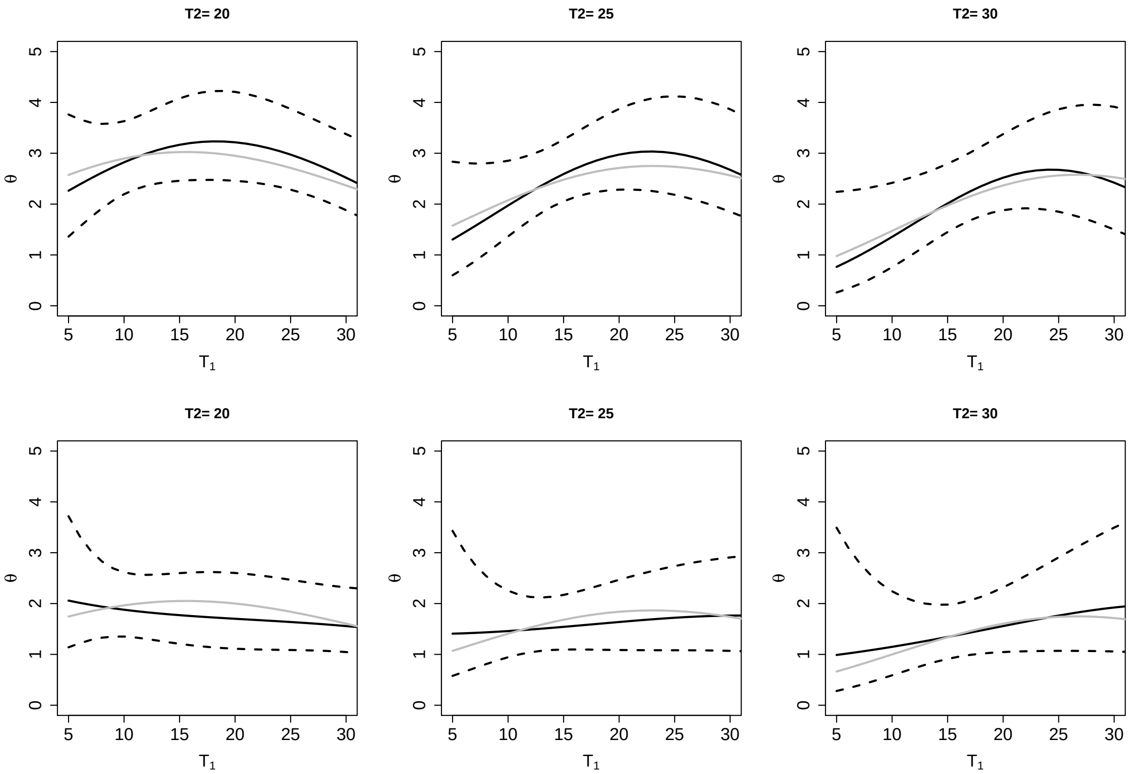


Figure 2.3: Comparison of estimated cross-ratio function for monozygotic twins and dizygotic twins in the Australian Twin Study based on the separate analysis and joint analysis. The plots on the top panel are for monozygotic twins, and on the bottom panel are for dizygotic twins. The black solid curves are the cross-ratio as a function of $T_1$ estimated from the separate analysis and grey curves are estimated from the joint analysis. The black dotted curves are the confidence bands for cross-ratio estimated from the separate analysis.

by mimicking the Cox proportional hazards model, where baseline log cross-ratio $\beta_0(t_1, t_2)$ is parameterized by a finite-dimensional Euclidean parameter. When the true baseline log cross-ratio is a smooth function of unknown functional form, regression spline method using the tensor product splines can be implemented. When the number of knots is fixed, the model is essentially a parametric model and asymptotic properties can be derived in a similar way. If the number of knots is allowed to grow with the sample size, the sieve M-estimation theory may be used for the development of asymptotic theory, see e.g., Shen and Wong (1994) and Shen (1997).

The proposed estimator is based on a pseudo-partial likelihood method instead of the true likelihood. If we were to construct the true likelihood for an arbitrary cross-ratio, we would need to solve for $h(t_1, t_2)$ in (1.2) to obtain the joint survival function $F(t_1, t_2)$ for every fixed $w$. Unfortunately, a closed-form solution does not exist for even the simplest non-trial functional forms for $\theta(t_1, t_2)$, for example, a linear function. Our method is robust because it bypasses modeling the joint and marginal distributions of the bivariate survival times.

The proposed model requires proportionality of covariate effect on cross-ratio, which is a rather strong an assumption. A richer class of model for covariate effect that does not rely on the proportionality assumption would be

$$\beta(t_1, t_2, W) = \boldsymbol{z}(t_1, t_2)'\boldsymbol{\gamma}_0 + W\boldsymbol{z}(t_1, t_2)'\boldsymbol{\gamma}. \tag{2.13}$$

If $W$ is a binary $\{0, 1\}$ variable then the model (2.13) is equivalent to just doing the analysis separately in those with $W = 1$ and $W = 0$ in terms of estimation; In contrast, if $W$ is not binary, we get a more parsimonious model than doing analysis separately at each level of $W$, which may be important for efficiency and for modeling the functional form of $W$. In general, $W$ can be a vector and we can further extend the above formulation to

$$\beta(t_1, t_2, W) = \boldsymbol{z}(t_1, t_2)'\boldsymbol{\gamma}_0 + \sum_{j=1}^{k} W_j\boldsymbol{z}(t_1, t_2)'\boldsymbol{\gamma}_j, \tag{2.14}$$

where $k$ is the dimension of $W$. However, model (2.14) involves a lot of parameters and required substantial amount of data for estimation. Instead, we could consider a reduced model of (2.13)

$$\beta(t_1, t_2, W) = \boldsymbol{z}(t_1, t_2)'\boldsymbol{\gamma}_0 + \sum_{j=1}^{k} W_j\boldsymbol{m}_j(t_1, t_2)'\boldsymbol{\gamma}_j, \tag{2.15}$$

where $\boldsymbol{m}$ is a sub-vector function of $\boldsymbol{z}$. In particular, if $\boldsymbol{m}_j = 1$ for all $j$, model (2.15) reduces to the proportional cross-ratio model being proposed.

## 2.8 Appendix: Proofs

### 2.8.1 Proof of Theorem II.1

Let $V^* \equiv (X_1^*, X_2^*, \Delta_1^*, \Delta_2^*, W^*)$ be an independent copy of $V \equiv (X_1, X_2, \Delta_1, \Delta_2, W)$. Define the deterministic function $\boldsymbol{u}(\boldsymbol{\xi}) = \boldsymbol{u}^{(1)}(\boldsymbol{\xi}) - \boldsymbol{u}^{(2)}(\boldsymbol{\xi}) + \boldsymbol{u}^{(3)}(\boldsymbol{\xi}) - \boldsymbol{u}^{(4)}(\boldsymbol{\xi})$, where

$$\boldsymbol{u}^{(1)}(\boldsymbol{\xi}) = \boldsymbol{u}^{(3)}(\boldsymbol{\xi}) = E\left\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2, W)\right\},$$
$$\boldsymbol{u}^{(2)}(\boldsymbol{\xi}) = \boldsymbol{u}^{(4)}(\boldsymbol{\xi}) = E\left\{\Delta_1^*\Delta_2\dot{\beta}(X_1^*, X_2, W)\frac{I(X_1 \geq X_1^*)I(X_2 \leq X_2^*)\theta(X_1^*, X_2, W; \boldsymbol{\xi})}{S(X_1^*, X_2|W^*)}\right\},$$

where $S(x_1, x_2|w) = Pr(X_1 > x_1, X_2 > x_2|W = w)$.

We will first show that $\boldsymbol{U}_n^{(k)}(\boldsymbol{\xi})$ converges uniformly to $\boldsymbol{u}^{(k)}$, $k = 1, 2$, then show that $\boldsymbol{u}(\boldsymbol{\xi}) = \mathbf{0}$ has the unique solution at $\boldsymbol{\xi}_0$, and finally show the consistency of $\hat{\boldsymbol{\xi}}_n$ where $\boldsymbol{U}_n(\hat{\boldsymbol{\xi}}_n) = \mathbf{0}$.

Define the following simplified notation:

$$\partial_1 F(t_1, t_2|w) = \frac{\partial F(t_1, t_2|w)}{\partial t_1}, \quad \partial_1 G(t_1, t_2|w) = \frac{\partial G(t_1, t_2|w)}{\partial t_1},$$
$$\partial_2 F(t_1, t_2|w) = \frac{\partial F(t_1, t_2|w)}{\partial t_2}, \quad \partial_2 G(t_1, t_2|w) = \frac{\partial G(t_1, t_2|w)}{\partial t_2},$$
$$\partial_{1,2} F(t_1, t_2|w) = \frac{\partial^2 F(t_1, t_2|w)}{\partial t_1 \partial t_2}, \quad \partial_{1,2} G(t_1, t_2|w) = \frac{\partial^2 G(t_1, t_2|w)}{\partial t_1 \partial t_2}.$$

where $F$ and $G$ denote the survival functions of $(T_1, T_2)$ and $(C_1, C_2)$ conditional on $W = w$, respectively. Then the conditional density function of $(X_1, X_2, \Delta_1, \Delta_2)$ given $W = w$ can be written as

$$\begin{aligned} & q(t_1, t_2, \delta_1, \delta_2|w) \\ = \ & \partial_{1,2}F(t_1, t_2|w)^{\delta_1\delta_2}\{-\partial_1 F(t_1, t_2|w)\}^{\delta_1(1-\delta_2)}\{-\partial_2 F(t_1, t_2|w)\}^{(1-\delta_1)\delta_2} \\ & \quad F(t_1, t_2|w)^{(1-\delta_1)(1-\delta_2)}\partial_{1,2}G(t_1, t_2|w)^{(1-\delta_1)(1-\delta_2)}\{-\partial_1 G(t_1, t_2|w)\}^{(1-\delta_1)\delta_2} \\ & \quad \{-\partial_2 G(t_1, t_2|w)\}^{\delta_1(1-\delta_2)}G(t_1, t_2|w)^{\delta_1\delta_2}, \end{aligned}$$

and the joint density of $(X_1, X_2, \Delta_1, \Delta_2, W)$ is

$$p(t_1, t_2, \delta_1, \delta_2, w) = q(t_1, t_2, \delta_1, \delta_2|w)f_W(w) \tag{2.16}$$

where $f_W(w)$ denotes the distribution function of $W$.

Following van der Vaart and Wellner (1996), we use $\mathbb{P}_n$ and $\mathbb{Q}_n$ to denote the empirical measures of $n$ independent copies of $(X_1^*, X_2^*, \Delta_1^*, \Delta_2^*, W^*)$ and $(X_1, X_2, \Delta_1, \Delta_2, W)$ that follow the distributions $P$ and $Q$, respectively. Although these two samples are in fact identical, i.e., $\mathbb{P}_n \equiv \mathbb{Q}_n$ and $P \equiv Q$, we use different letters to keep the notation tractable for the double summations, which will soon become clear in the following calculations.

For model (2.9), $\boldsymbol{U}_n^{(1)}(\boldsymbol{\xi}) = \mathbb{Q}_n \Delta_1 \Delta_2 \dot{\beta}(X_1, X_2, W)$ is free of $\boldsymbol{\xi}$, and $\dot{\beta}(X_1, X_2, W)$ is bounded from Conditions C2.1-C2.2. Hence by the law of large numbers, we have

$$\sup_{\xi} |\boldsymbol{U}_n^{(1)}(\boldsymbol{\xi}) - \boldsymbol{u}^{(1)}(\boldsymbol{\xi})| = |(\mathbb{Q}_n - Q)\Delta_1 \Delta_2 \dot{\beta}(X_1, X_2, W)| \to \mathbf{0}$$

either almost surely or in probability. Convergence in probability should be adequate here for the proof.

To show the uniform convergence of $\boldsymbol{U}_n^{(2)}(\boldsymbol{\xi})$ to $\boldsymbol{u}^{(2)}(\boldsymbol{\xi})$, the proof is treated separately for discrete $W$ with finite levels and continuous $W$. When $W$ is discrete with finite levels, the proof is similar to that provided in Chapter I. Here we focus on continuous $W$.

For simplicity, define $V = (X_1, X_2, \Delta_1, \Delta_2, W)$ and also define the following quantities:

$$t(V_i, V_j; \boldsymbol{\xi}) = \Delta_{1i}\Delta_{2j} I(X_{1j} \geq X_{1i}) I(X_{2j} \leq X_{2i}) e^{\beta(X_{1i}, X_{2j}, W_i; \boldsymbol{\xi})} \dot{\boldsymbol{\beta}}(X_{1i}, X_{2j}, W_i), \tag{2.17}$$

$$\begin{aligned}
v(V_i, V_j; \boldsymbol{\xi}) &= \frac{1}{n}[N(X_{1i}, X_{2j}, W_i) - \mathbf{K}_h(0) I(X_{2j} \leq X_{2i})(1 - \theta(X_{1i}, X_{2j}, W_i; \boldsymbol{\xi}))], \\
g^{(n)}(V_i, V_j; \boldsymbol{\xi}) &= \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi})}{v(V_i, V_j; \boldsymbol{\xi})}, \\
g_h^{(n)}(V_i, V_j; \boldsymbol{\xi}) &= \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi})}{S_h(X_{1i}, X_{2j}, W_i)} \\
\tilde{g}^{(n)}(V_i, V_j; \boldsymbol{\xi}) &= \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi})}{S(X_{1i}, X_{2j}, W_i)} \\
\tilde{h}(V_i; \boldsymbol{\xi}) &= E_{X_{1j}, X_{2j}, \Delta_{1j}, \Delta_{2j} | W_j = W_i, X_{1i}, X_{2i}, \Delta_{1i}, \Delta_{2i}} \left[ \frac{t(V_i, V_j; \boldsymbol{\xi})}{S(X_{1i}, X_{2j} | W_i)} \right] \qquad (2.18) \\
\tilde{h}^*(V_j; \boldsymbol{\xi}) &= E_{X_{1i}, X_{2i}, \Delta_{1i}, \Delta_{2i} | W_i = W_j, X_{1j}, X_{2j}, \Delta_{1j}, \Delta_{2j}} \left[ \frac{t(V_i, V_j; \boldsymbol{\xi})}{S(X_{1i}, X_{2j} | W_j)} \right] \\
u^{(2)}(\boldsymbol{\xi}) &= E_{X_{1i}, X_{2i}, \Delta_{1i}, \Delta_{2i}, W_i} \tilde{h}(V_i; \boldsymbol{\xi}),
\end{aligned}$$

where

$$\begin{aligned} S(t_1, t_2, w) &= Pr(X_1 \geq t_1, X_2 \geq t_2 | W = w) f_W(w), \\ S_h(t_1, t_2, w) &= E\left[ I(X_1 \geq t_1, X_2 \geq t_2) \mathbf{K}_h(W - w) \right]. \end{aligned} \tag{2.19}$$

Clearly, we have

$$S(t_1, t_2, w) = \lim_{h \downarrow 0} S_h(t_1, t_2, w).$$

By Härdle, Janssen and Serfling (1988),

$$\begin{aligned} \frac{1}{n} N(t_1, t_2, w) &= \frac{1}{n} \sum_{k=1}^{n} I(X_{1k} \geq t_1, X_{2k} \geq t_2) \mathbf{K}_h(W_k - w) \\ &= \frac{\sum_{k=1}^{n} I(X_{1k} \geq t_1, X_{2k} \geq t_2) \mathbf{K}_h(W_k - w)}{\sum_{k=1}^{n} \mathbf{K}_h(W_k - w)} \times \frac{\sum_{k=1}^{n} \mathbf{K}_h(W_k - w)}{n} \\ &= E(I(X_1 \geq t_1, X_2 \geq t_2) | W = w) f(w) + o_p(1) \\ &= S(t_1, t_2, w) + o_p(1). \end{aligned}$$

Also note that the difference between $g^{(n)}$ and $\tilde{g}^{(n)}$ is their denominators wherein we replace the denominator of $g^{(n)}$ by its limit. We then have the following:

$$
\begin{aligned}
& \sup_{\xi} |\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} g^{(n)}(V_i,V_j;\boldsymbol{\xi}) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{g}^{(n)}(V_i,V_j;\boldsymbol{\xi})| \\
\leq\; & \frac{1}{n}\sum_{i=1}^{n}\sup_{\xi}|\frac{1}{n}\sum_{j=1}^{n} g^{(n)}(V_i,V_j;\boldsymbol{\xi}) - \frac{1}{n}\sum_{j=1}^{n}\tilde{g}^{(n)}(V_i,V_j;\boldsymbol{\xi})| \\
=\; & \frac{1}{n}\sum_{i=1}^{n}\sup_{\xi}\Big|\frac{1}{n}\sum_{j=1}^{n}\frac{\mathbf{K}_h(W_j-W_i)t(V_i,V_j;\boldsymbol{\xi})}{v(V_i,V_j;\boldsymbol{\xi})S(X_{1i},X_{2j},W_i)} \times \big(v(V_i,V_j;\boldsymbol{\xi}) - S(X_{1i},X_{2j},W_i)\big)\Big| \\
=\; & \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}_h(W_j-W_i)\sup_{\xi}\Big|\frac{t(V_i,V_j;\boldsymbol{\xi})}{v(V_i,V_j;\boldsymbol{\xi})S(X_{1i},X_{2j},W_i)} \\
& \times\big(v(V_i,V_j;\boldsymbol{\xi}) - S(X_{1i},X_{2j},W_i)\big)\Big| \\
\leq\; & \frac{1}{n}\sum_{i=1}^{n}\frac{\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}_h(W_j-W_i)}{\frac{1}{n}N(X_{1i},X_{2j},W_i)+o_p(1)}\sup_{\xi}\Big|\frac{t(V_i,V_j;\boldsymbol{\xi})}{S(X_{1i},X_{2j},W_i)}\Big|\Big(\sup\Big|(n^{-1}N(X_{1i},X_{2j},W_i) \\
& -S(X_{1i},X_{2j},W_i))\Big| + \sup_{\xi}\Big|n^{-1}\mathbf{K}_h(0)I(X_{2j}\leq X_{2i})(1-\theta(X_{1i},X_{2j},W_i;\boldsymbol{\xi}))\Big|\Big) \\
\leq\; & \frac{1}{n}\sum_{i=1}^{n} O_p(1)O_p(1)\Big(\sup|(n^{-1}N(X_{1i},X_{2j},W_i) - S(X_{1i},X_{2j},W_i))| + O_p((nh)^{-1})\Big) \\
\leq\; & O_p(1)\Big(O_p(\max\{(nh/\log n)^{-1/2},h^{\epsilon}\}) + O_p((nh)^{-1})\Big) \\
=\; & o_p(1).
\end{aligned}
$$

In the last inequality, we used the result of strong uniform consistency for conditional functional estimators of Härdle, Janssen and Serfling (1988).

Next, we want to show that difference between $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{g}_{ij}^{(n)}(\boldsymbol{\xi})$ and $\frac{1}{n}\sum_{i=1}^{n}\tilde{h}(V_i;\boldsymbol{\xi})$ is $o_p(1)$. Again using the result of Härdle, Janssen and Serfling (1988) for showing the second last equality in the following calculation, we have

$$\begin{aligned}
\sup_{\xi} \quad & \left| \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{g}^{(n)}(V_i, V_j; \boldsymbol{\xi}) - \frac{1}{n} \sum_{i=1}^{n} \tilde{h}(V_i; \boldsymbol{\xi}) \right| \\
\leq \quad & \frac{1}{n} \sum_{i=1}^{n} \sup_{\xi} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{g}^{(n)}(V_i, V_j; \boldsymbol{\xi}) - \tilde{h}(V_i; \boldsymbol{\xi}) \right| \\
\leq \quad & \frac{1}{n} \sum_{i=1}^{n} \sup_{\boldsymbol{\xi}, V_i} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{g}^{(n)}(V_i, V_j; \boldsymbol{\xi}) - \tilde{h}(V_i; \boldsymbol{\xi}) \right| \\
= \quad & \sup_{\boldsymbol{\xi}, V_i} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{g}^{(n)}(V_i, V_j; \boldsymbol{\xi}) - \tilde{h}(V_i; \boldsymbol{\xi}) \right| \\
= \quad & O_p(\max\{(nh/\log n)^{-1/2}, h^{\epsilon}\}) \\
= \quad & o_p(1).
\end{aligned}$$

Last, we want to show that difference between $\frac{1}{n}\sum_{i=1}^{n} \tilde{h}(V_i; \boldsymbol{\xi})$ and its deterministic limit $u^{(2)}(\boldsymbol{\xi})$ is $o_p(1)$ uniformly in $\boldsymbol{\xi}$. For model (2.9) under C2.1-C2.3, it is straightforward to see that all the component functions of $t(V_i, V_j; \boldsymbol{\xi})$ are Donsker. Thus $t(V_i, V_j; \boldsymbol{\xi})$ is Donsker. Then by Theorem 2.10.2 in van der Vaar and Wellner (1996), $\tilde{h}(V_i; \boldsymbol{\xi})$ is also Donsker. Hence, $\tilde{h}(V_i; \boldsymbol{\xi})$ is Glivenko-Cantelli. We then have

$$\sup_{\xi} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{h}(V_i; \boldsymbol{\xi}) - u^{(2)}(\boldsymbol{\xi}) \right| = o_p(1).$$

Thus we have shown that $\boldsymbol{U}_n(\boldsymbol{\xi})$ converges uniformly to $\boldsymbol{u}(\boldsymbol{\xi})$ in probability. To show $\boldsymbol{u}(\boldsymbol{\xi}_0) = 0$, it suffices to show that $\boldsymbol{u}^{(1)}(\boldsymbol{\xi}_0) = \boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)$. Recall that $(X_1, X_2, \Delta_1, \Delta_2, W)$ and $(X_1^*, X_2^*, \Delta_1^*, \Delta_2^*, W^*)$ are independent and identically distributed with a density function given in (2.16). Let $d_k$ denote the infinitesimal change with respect to $t_k$, $k = 1, 2$. Note that

$$\theta(t_1, t_2, w; \boldsymbol{\xi}_0) = e^{\beta(t_1, t_2, w; \xi_0)} = \frac{\partial_{1,2} F(t_1, t_2|w) F(t_1, t_2|w)}{\partial_1 F(t_1, t_2|w) \partial_2 F(t_1, t_2|w)}.$$

and $S(t_1, t_2|w) = F(t_1, t_2|w)G(t_1, t_2|w)$ by Condition C2.5. Then we have

$$
\begin{aligned}
&\boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\\
&= E_{V^*}E_{V|V^*}\left\{\frac{\Delta_1^*\Delta_2 I(X_1 \geq X_1^*)I(X_2 \leq X_2^*)e^{\beta(X_1^*, X_2, W^*;\xi_0)}\dot{\beta}(X_1^*, X_2, W^*)}{S(X_1^*, X_2|W^*)}\right\}\\
&= E_{V^*}\left\{\Delta_1^*\int_0^{X_2^*}\int_{X_1^*}^{\infty}\sum_{\delta_1=0}^{1}\sum_{\delta_2=0}^{1}\frac{\delta_2 e^{\beta(X_1^*, t_2, W^*;\xi_0)}\dot{\beta}(X_1^*, t_2, W)}{S(X_1^*, t_2|W^*)}q(t_1, t_2, \delta_1, \delta_2|W^*)dt_1dt_2\right\}\\
&= E_{V^*}\left\{\Delta_1^*\int_0^{X_2^*}\frac{\dot{\beta}(X_1^*, t_2, W^*)\theta(X_1^*, t_2, W^*;\boldsymbol{\xi}_0)}{S(X_1^*, t_2|W^*)}\right.\\
&\qquad\qquad\left.\times\left[\int_{X_1^*}^{\infty}d_1\{\partial_2 F(t_1, t_2|W^*)G(t_1, t_2|W^*)\}\right]dt_2\right\}\\
&= E_{V^*}\left\{-\Delta_1^*\int_0^{X_2^*}\frac{\dot{\beta}(X_1^*, t_2, W^*)\theta(X_1^*, t_2, W^*;\boldsymbol{\xi}_0)}{S(X_1^*, t_2|W^*)}\right.\\
&\qquad\qquad\left.\times\partial_2 F(X_1^*, t_2|W^*)G(X_1^*, t_2|W^*)dt_2\right\}\\
&= E_{V^*}\left\{-\Delta_1^*\int_0^{X_2^*}\frac{\dot{\beta}(X_1^*, t_2, W^*)}{\partial_1 F(X_1^*, t_2|W^*)}\partial_{12}F(X_1^*, t_2|W^*)dt_2\right\}\\
&= \int_{-\infty}^{\infty}\int_0^{\infty}\int_0^{\infty}\sum_{\delta_1^*=0}^{1}\sum_{\delta_2^*=0}^{1}\delta_1^*\left\{-\int_0^{t_2^*}\frac{\dot{\beta}(t_1^*, t_2, w^*)}{\partial_1 F(t_1^*, t_2, w^*)}\partial_{12}F(t_1^*, t_2|w^*)dt_2\right\}\\
&\qquad\qquad\times p(t_1^*, t_2^*, \delta_1^*, \delta_2^*, w^*)dt_2^*dt_1^*dw^*\\
&= \int_{-\infty}^{\infty}\int_0^{\infty}\int_0^{\infty}\left\{-\int_0^{t_2^*}\frac{\dot{\beta}(t_1^*, t_2, w^*)}{\partial_1 F(t_1^*, t_2|w^*)}\partial_{12}F(t_1^*, t_2|w^*)dt_2\right\}\\
&\qquad\qquad\times d_2\{\partial_1 F(t_1^*, t_2^*|w^*)G(t_1^*, t_2^*|w^*)\}f(w^*)dt_1^*dw^*\\
&= \int_{-\infty}^{\infty}\int_0^{\infty}\int_0^{\infty}\frac{\dot{\beta}(t_1^*, t_2^*, w^*)}{\partial_1 F(t_1^*, t_2^*|w^*)}\partial_{12}F(t_1^*, t_2^*|w^*)\partial_1 F(t_1^*, t_2^*|w^*)G(t_1^*, t_2^*|w^*)\\
&\qquad\qquad\times f(w^*)dt_2^*dt_1^*dw^*\\
&= \int_{-\infty}^{\infty}\int_0^{\infty}\int_0^{\infty}\dot{\beta}(t_1^*, t_2^*, w^*)\partial_{12}F(t_1^*, t_2^*|w^*)G(t_1^*, t_2^*|w^*)f(w^*)dt_1^*dt_2^*dw^*\\
&= E\left\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2, W)\right\}\\
&= \boldsymbol{u}^{(1)}(\boldsymbol{\xi}_0).
\end{aligned}
$$

To show $\boldsymbol{\xi}_0$ is the unique solution of $u(\boldsymbol{\xi}) = 0$, it suffices to show that (a) the matrix $\dot{\boldsymbol{u}}(\boldsymbol{\xi}) \equiv d\boldsymbol{u}(\boldsymbol{\xi})/d\boldsymbol{\xi}$ is negative semidefinite for all $\boldsymbol{\xi} \in \Gamma$, and (b) $\dot{\boldsymbol{u}}(\boldsymbol{\xi})$ is negative definite at $\boldsymbol{\xi}_0$. Similar to

Chapter I, it can be shown by straightforward calculations that $a'\dot{\boldsymbol{u}}(\boldsymbol{\xi})a \leq 0$ for any vector $a$ and $\dot{\boldsymbol{u}}(\boldsymbol{\xi}_0) = -2E\left\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2, W)^{\otimes 2}\right\}$ is negative definite by Condition C2.4. Then $\boldsymbol{\xi}_0$ is the unique solution to $\boldsymbol{u}(\boldsymbol{\xi}) = 0$. We are now ready to show the consistency of $\hat{\boldsymbol{\xi}}_n$. Given the fact that $\boldsymbol{U}_n(\hat{\boldsymbol{\xi}}_n) = 0$ and $\sup|\boldsymbol{U}_n(\boldsymbol{\xi}) - \boldsymbol{u}(\boldsymbol{\xi})| = o_p(1)$, we have

$$|\boldsymbol{u}(\hat{\boldsymbol{\xi}}_n)| = |\boldsymbol{U}_n(\hat{\boldsymbol{\xi}}_n) - \boldsymbol{u}(\hat{\boldsymbol{\xi}}_n)| \leq \sup|\boldsymbol{U}_n(\boldsymbol{\xi}) - \boldsymbol{u}(\boldsymbol{\xi})| = o_p(1).$$

Since $\boldsymbol{\xi}_0$ is the unique solution to $\boldsymbol{u}(\boldsymbol{\xi}) = 0$, for any fixed $\boldsymbol{\epsilon} > 0$, there exists a $\boldsymbol{\delta} > 0$ such that

$$Pr\left(|\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0| > \boldsymbol{\epsilon}\right) \leq Pr\left(|\boldsymbol{u}(\hat{\boldsymbol{\xi}}_n)| > \boldsymbol{\delta}\right).$$

The consistency of $\hat{\boldsymbol{\xi}}_n$ follows immediately.

### 2.8.2 Proof of Theorem II.2

Define $\dot{\boldsymbol{U}}_n(\boldsymbol{\xi}) \equiv d\boldsymbol{U}_n(\boldsymbol{\xi})/d\boldsymbol{\xi}$. By the Taylor expansion of $\boldsymbol{U}_n(\hat{\boldsymbol{\xi}}_n)$ around $\boldsymbol{\xi}_0$, we have

$$n^{1/2}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0) = -\left\{\dot{\boldsymbol{U}}_n(\boldsymbol{\xi}^*)\right\}^{-1} n^{1/2}\boldsymbol{U}_n(\boldsymbol{\xi}_0), \tag{2.20}$$

where $\boldsymbol{\xi}^*$ lies between $\hat{\boldsymbol{\xi}}_n$ and $\boldsymbol{\xi}_0$. By a similar calculation as in the proof of Theorem 1 showing the uniform consistency of $\boldsymbol{U}_n(\boldsymbol{\xi})$, we can show that $\sup|\dot{\boldsymbol{U}}_n(\boldsymbol{\xi}) - \dot{\boldsymbol{u}}(\boldsymbol{\xi})| = o_p(1)$. Thus by the consistency of $\hat{\boldsymbol{\xi}}_n$, which implies the consistency of $\boldsymbol{\xi}^*$, and the continuity of $\dot{\boldsymbol{u}}(\boldsymbol{\xi})$, we obtain $\dot{\boldsymbol{U}}_n(\boldsymbol{\xi}^*) = \dot{\boldsymbol{u}}(\boldsymbol{\xi}_0) + o_p(1)$, where $\dot{\boldsymbol{u}}(\boldsymbol{\xi}_0) = -2E\left\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2, W)^{\otimes 2}\right\} = -\boldsymbol{I}(\boldsymbol{\xi}_0)$ is invertible by Condition C2.4. Hence based on the fact that continuity holds for the inverse operator, (2.20) can be written as

$$n^{1/2}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0) = \left\{\boldsymbol{I}(\boldsymbol{\xi}_0)^{-1} + o_p(1)\right\} n^{1/2}\boldsymbol{U}_n(\boldsymbol{\xi}_0). \tag{2.21}$$

We now need to find the asymptotic representation of $n^{1/2}\boldsymbol{U}_n(\boldsymbol{\xi}_0)$. We only check it for $\boldsymbol{U}_n^{(1)}(\boldsymbol{\xi}_0) - \boldsymbol{U}_n^{(2)}(\boldsymbol{\xi}_0)$. The calculation for $\boldsymbol{U}_n^{(3)}(\boldsymbol{\xi}_0) - \boldsymbol{U}_n^{(4)}(\boldsymbol{\xi}_0)$ is virtually identical and yields the same asymptotic representation.

It is easily seen that

$$n^{1/2}\left(\boldsymbol{U}_n^{(1)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(1)}(\boldsymbol{\xi}_0)\right) = \mathbb{G}_n\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2, W)\}, \tag{2.22}$$

where $\mathbb{G}_n = n^{1/2}(\mathbb{P}_n - P)$. We then focus on $n^{1/2}\left(\boldsymbol{U}_n^{(2)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right)$, whose linearization differs vastly for discrete covariate and continuous continuous covariate, largely because we could no longer rely on Donsker Theorem for continuous covariate case when kernel functions are involved. Thus the two cases are treated separately in the proof.

**Linearization of $n^{1/2}\left(\boldsymbol{U}_n^{(2)}(V_i, V_j; \boldsymbol{\xi}_0) - \boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right)$ for Discrete Covariate**

First we introduce the following notation:

$$\begin{aligned}&\tilde{g}^d(\Delta_2, X_1, X_2, W, \Delta_1^*, X_1^*, X_2^*, W^*; \boldsymbol{\xi})\\&\quad = \frac{I(W = W^*)\Delta_1^*\Delta_2\dot{\beta}(X_1^*, X_2, W^*)I(X_1 \geq X_1^*)I(X_2 \leq X_2^*)\theta(X_1^*, X_2, W^*, \boldsymbol{\xi})}{S(X_1^*, X_2, W^*)}.\end{aligned}$$

By fixing $(\Delta_1^*, X_1^*, X_2^*, W^*)$ at $(\delta_1, x_1, x_2, w)$, we also define

$$\tilde{h}_Q^d(\delta_1, x_1, x_2, w; \boldsymbol{\xi}) \quad = \quad Q\tilde{g}^d(\Delta_2, X_1, X_2, W, \delta_1, x_1, x_2, w; \boldsymbol{\xi}).$$

Similarly, fixing $(\Delta_2, X_1, X_2, W)$ at $(\delta_2, x_1, x_2, w)$, define

$$\tilde{h}_P^d(\delta_2, x_1, x_2, w; \boldsymbol{\xi}) = P\tilde{g}^d(\delta_2, x_1, x_2, w, \Delta_1^*, X_1^*, X_2^*, W^*; \boldsymbol{\xi}).$$

Following similar calculation as in Chapter I, we can show that:

$$\begin{aligned}&n^{1/2}\left\{\boldsymbol{U}_n^{(2)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right\}\\&\quad = \mathbb{G}_n\Bigg\{\tilde{h}_Q^d(\Delta_1, X_1, X_2, W; \boldsymbol{\xi}_0) + \tilde{h}_P^d(\Delta_2, X_1, X_2, W; \boldsymbol{\xi}_0)\\&\qquad - \iint I(X_1 \geq x_1^*, X_2 \geq x_2, W = w^*)r(\delta_1, x_1, x_2, w, \delta_2^*, x_1^*, x_2^*, w^*)\\&\qquad\quad dP(\delta_1^*, \delta_2^*, x_1^*, x_2^*, w^*)dQ(\delta_1, \delta_2, x_1, x_2, w)\Bigg\} + o_p(1), \qquad (2.23)\end{aligned}$$

where

$$\begin{aligned}&r(\delta_1, x_1, x_2, w, \delta_2^*, x_1^*, x_2^*, w^*)\\&\quad = \frac{I(w = w^*)\delta_1^*\delta_2\dot{\beta}(x_1^*, x_2, w^*; \xi_0)I(x_1 \geq x_1^*)I(x_2 \leq x_2^*)e^{\beta(x_1^*, x_2, w^*; \xi_0)}}{\{S(x_1^*, x_2, w^*)\}^2}.\end{aligned}$$

Then we obtain

$$
\begin{aligned}
& n^{1/2}\boldsymbol{U}_n(\boldsymbol{\xi}_0) \\
& \quad = n^{1/2}\{\boldsymbol{U}_n(\boldsymbol{\xi}_0) - \boldsymbol{u}(\boldsymbol{\xi}_0)\} \\
& \quad = n^{1/2}\{\boldsymbol{U}_n^{(1)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(1)}(\boldsymbol{\xi}_0)\} - n^{1/2}\{\boldsymbol{U}_n^{(2)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\} \\
& \qquad + n^{1/2}\{\boldsymbol{U}_n^{(3)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(3)}(\boldsymbol{\xi}_0)\} - n^{1/2}\{\boldsymbol{U}_n^{(4)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(4)}(\boldsymbol{\xi}_0)\} \\
& \quad = 2\mathbb{G}_n \Bigg\{ \Delta_1\Delta_2\dot{\beta}(X_1, X_2, W) - \tilde{h}_Q^d(\Delta_1, X_1, X_2, W; \boldsymbol{\xi}_0) - \tilde{h}_P^d(\Delta_2, X_1, X_2, W; \boldsymbol{\xi}_0) \\
& \qquad + \iint I(X_1 \geq x_1^*, X_2 \geq x_2, W = w^*) r(\delta_1, x_1, x_2, w, \delta_2^*, x_1^*, x_2^*, w^*) \\
& \qquad\quad dP(\delta_1^*, \delta_2^*, x_1^*, x_2^*, w^*) dQ(\delta_1, \delta_2, x_1, x_2, w) \Bigg\} + o_p(1) \\
& \quad \to_d N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_0)).
\end{aligned}
$$

Thus from (2.21) we obtain the desired asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0)$.

**Linearization of $n^{1/2}\left(\boldsymbol{U}_n^{(2)}(V_i,V_j;\boldsymbol{\xi}_0)-\boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right)$ for Continuous Covariate**

We focus on $n^{1/2}\left(\boldsymbol{U}_n^{(2)}(V_i,V_j;\boldsymbol{\xi}_0)-\boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right)$ with following decomposition:

$$
\begin{aligned}
& n^{1/2}\left(\boldsymbol{U}_n^{(2)}(V_i,V_j;\boldsymbol{\xi}_0)-\boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right) \\
&= n^{1/2}\Big(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} g^{(n)}(V_i,V_j;\boldsymbol{\xi}_0)-\boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\Big) \\
&= n^{1/2}\Big(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} g^{(n)}(V_i,V_j;\boldsymbol{\xi}_0)-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{g}^{(n)}(V_i,V_j;\boldsymbol{\xi}_0)\Big) \\
&\qquad +n^{1/2}\Big(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{g}^{(n)}(V_i,V_j;\boldsymbol{\xi}_0)-\frac{1}{n}\sum_{i=1}^{n}\tilde{h}(V_i;\boldsymbol{\xi}_0)\Big) \\
&\qquad\qquad +n^{1/2}\Big(\frac{1}{n}\sum_{i=1}^{n}\tilde{h}(V_i;\boldsymbol{\xi}_0)-\boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\Big) \\
&= n^{1/2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(g^{(n)}(V_i,V_j;\boldsymbol{\xi}_0)-g_h^{(n)}(V_i,V_j;\boldsymbol{\xi}_0)\right) \\
&\qquad +n^{1/2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(g_h^{(n)}(V_i,V_j;\boldsymbol{\xi}_0)-\tilde{g}^{(n)}(V_i,V_j;\boldsymbol{\xi}_0)\right) \\
&\qquad\qquad +\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j=1}^{n}\tilde{g}^{(n)}(V_i,V_j;\boldsymbol{\xi}_0)-\tilde{h}(V_i;\boldsymbol{\xi}_0)\right) \\
&\qquad\qquad\quad +\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\tilde{h}(V_i;\boldsymbol{\xi}_0)-\boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right) \\
&= -A-B+C+D
\end{aligned}
$$

Now we will look at the four terms separately. Firstly, term $D$ is a sum of $iid$ items:

$$
\begin{aligned}
D &= \frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\tilde{h}_i(\boldsymbol{\xi}_0)-\boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right) \\
&= \mathbb{G}_n\left(\tilde{h}(\boldsymbol{\xi}_0)\right).
\end{aligned}
$$

Secondly, term $C$ can be decomposed as follows:

$$
\begin{aligned}
& C \\
= \;& \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left( \frac{1}{n} \sum_{j=1}^{n} \tilde{g}^{(n)}(V_i, V_j; \boldsymbol{\xi}_0) - \tilde{h}(V_i, \boldsymbol{\xi}_0) \right) \\
= \;& n^{1/2} \left( \mathbb{P}_n^* \mathbb{P}_n \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) - \mathbb{P}_n^* \tilde{h}(V^*, \boldsymbol{\xi}_0) \right) \\
= \;& n^{1/2} \left( \mathbb{P}_n \mathbb{P}_n^* \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) - P \mathbb{P}_n^* \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) + \mathbb{P}_n^* P \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) - \mathbb{P}_n^* \tilde{h}(V^*, \boldsymbol{\xi}_0) \right) \\
= \;& \mathbb{G}_n \left( \mathbb{P}_n^* \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) - P^* \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) + P^* \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) - \tilde{h}^*(V, \boldsymbol{\xi}_0) + \tilde{h}^*(V, \boldsymbol{\xi}_0) \right) \\
& \quad + n^{1/2} \mathbb{P}_n^* \left( P \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) - \tilde{h}(V^*, \boldsymbol{\xi}_0) \right) \\
= \;& \mathbb{G}_n \left( \mathbb{P}_n^* \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) - P^* \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) \right) + \mathbb{G}_n \left( P^* \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) - \tilde{h}^*(V, \boldsymbol{\xi}_0) \right) \\
& \quad + \mathbb{G}_n \left( \tilde{h}^*(V, \boldsymbol{\xi}_0) \right) + n^{1/2} \mathbb{P}_n^* \left( P \tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) - \tilde{h}(V^*, \boldsymbol{\xi}_0) \right) \\
= \;& C_1 + C_2 + C_3 + C_4
\end{aligned}
$$

For the last equality of the above equation, we want to show that $C_1 = o_p(1)$, $C_2 = o_p(1)$ and $C_4 = o_p(1)$, so that $C = C_3 + o_p(1)$. First, by lemma A.2 of Ichimura (1993) $P\tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) - \tilde{h}(V^*, \boldsymbol{\xi}_0) = O(h^2)$. Thus $C_4 = n^{1/2}O(h^2) = o_p(1)$ for $h$ satisfying C2.8.

Likewise, $P^*\tilde{g}^{(n)}(V, V^*, \boldsymbol{\xi}_0) - \tilde{h}^*(V, \boldsymbol{\xi}_0) = O(h^2)$, and therefore $C_2 = n^{1/2}O(h^2) = o_p(1)$ for $h$ satisfying C2.8.

Finally, we need to show that $C_1 = \mathbb{G}_n \left( (\mathbb{P}_n^* - P^*) \mathbf{K}_h (W^* - W) \frac{t(V^*, V; \boldsymbol{\xi}_0)}{S(X_1^*, X_2, W^*; \boldsymbol{\xi}_0)} \right) = o_p(1)$. First set

$$
r_h(V, V^*) = \mathbf{K} \left( \frac{W^* - W}{h} \right) \frac{t(V^*, V; \boldsymbol{\xi}_0)}{S(X_1^*, X_2, W^*; \boldsymbol{\xi}_0)},
$$

and

$$
\tilde{r}_h(V, V^*) = r_h(V, V^*) - P r_h(V, V^*) - P^* r_h(V, V^*) + P P^* r_h(V, V^*),
$$

and

$$
T_n(\tilde{r}_h) = \sum_{1 \le i \ne j \le n} \tilde{r}_h(V_i, V_j).
$$

Then we have

$$\begin{aligned} C_1 &= h^{-1}\mathbb{G}_n\left((\mathbb{P}_n^* - P^*)r_h(V,V^*)\right) \\ &= h^{-1}\sqrt{n}\left((\mathbb{P}_n - P)(\mathbb{P}_n^* - P^*)r_h(V,V^*)\right) \\ &= h^{-1}\sqrt{n}\frac{1}{n^2}\left(T_n(\tilde{r}_h) + \sum_{i=1}^{n}\tilde{r}_h(V_i,V_i)\right) \\ &= \frac{1}{\sqrt{n}hn}T_n(\tilde{r}_h) + \frac{1}{nh\sqrt{n}}\sum_{i=1}^{n}\tilde{r}_h(V_i,V_i) \\ &= C_{11} + C_{12} \end{aligned}$$

Applying the central limit theorem, it is easy to see that $C_{12} = o_p(1)$. To show that $C_{11} = o_p(1)$, we need the following defintion and theorem from Nolan and Pollard (1987). We keep the same numbering for the definition and theorem as in the original paper for the ease of reference.

Definition 8. Call a class of functions $\mathcal{F}$ Euclidean for the envelop F if there exist constants A and V such that

$$N_1(\epsilon, Q, \mathcal{F}, F) \leq A\epsilon^{-V}, \text{ for } 0 < \epsilon \leq 1,$$

whenever $0 < QF < \infty$.

*Thoerem 9. Let $\mathcal{F}$ be a Euclidean class of P-degenerate functions with envelope 1. Let $W(n,x)$ be a bounded weight function that is decreasing in both arguments and satisfies*

$$\sum_{n=1}^{\infty}\int_0^1 n^{-1}W(n,x)(1+\log(1/x))dx < \infty$$

*If $v(\cdot)$ is a function on $\mathcal{F}$ for which $v(f) \geq \sup_x P|f(x,\cdot)|$, then*

$$n^{-1}||W(n, v(f)^{1/2})T_n(f)|| \to 0$$

In our case, each $\tilde{r}_h$ is P-degenerate; that is $P\tilde{r}_h(V,\cdot) = 0$. The class of all $\tilde{r}_h$ is a candidate for the above theorem. Following Nolan and Pollard (1987) page 795, it is easy to check that there exist a constant $C$ for which

$$\sup_{x,y,h} |\tilde{r}_h(x,y)| \le C,$$
$$\sup_x P^* |\tilde{r}_h(x,\cdot)| \le C(1 \wedge h), \text{ for all } h > 0.$$

We can rescale to make $C$ equal to 1.

If kernel $\mathbf{K}$ is of bounded variation, e.g. standard normal density, then $\{\tilde{r}_h\}$ is a Euclidean class. For details of establishing Euclidean property in a particular class, please refer to Section 5 of Nolan and Pollard (1987).

Invoking Thoerem 9 of Nolan and Pollard (1987), we obtain

$$n^{-1}||W(n, v(f)^{1/2})T_n(f)|| = o_p(1),$$

where $v(\tilde{r}_h) = 1 \wedge h$ and $W(n,x) = (1 + nx^{10})^{-1}$. Since $W$ is bounded by 1 and

$$\int_0^1 W(n,x)(1 + \log(1/x))dx = O(n^{-1/10} \log n)$$

the conditions of Thoerem 9 are satisfied.

Returning to the calculation for $C_{11}$,

$$\begin{aligned} C_{11} &= \frac{1}{\sqrt{n}hn} T_n(\tilde{r}_h) \\ &\le \frac{1}{\sqrt{n}hW(n, v(f)^{1/2})} ||n^{-1} W(n, v(f)^{1/2}) T_n(\tilde{r}_h)|| \\ &= \frac{1 + n(1 \wedge h)^5}{\sqrt{n}h} o_p(1) \\ &\le \frac{1 + nh^5}{\sqrt{n}h} o_p(1) \\ &= o_p(1) + \sqrt{n}h^4 o_p(1). \end{aligned}$$

Thus $C_{11} = o_p(1)$ for $h$ satisfying C2.8.

Thus, we obtain $C_1 = C_{11} + C_{12} = o_p(1)$ and $C = \mathbb{G}_n\left(\tilde{h}^*(V, \boldsymbol{\xi}_0)\right) + o_p(1)$.

Thirdly, term $B$ is $o_p(1)$ and hence negligible because

$$
\begin{aligned}
& B \\
= \;& n^{\frac{1}{2}} n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi}_0)}{S(X_{1i}, X_{2j}, W_i) S_h(X_{1i}, X_{2j}, W_i)} \left( S_h(X_{1i}, X_{2j}, W_i) - S(X_{1i}, X_{2j}, W_i) \right) \\
= \;& n^{\frac{1}{2}} n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi}_0)}{S(X_{1i}, X_{2j}, W_i) S_h(X_{1i}, X_{2j}, W_i)} O(h^2).
\end{aligned}
$$

The inner summation divided by $n$ is bounded by the density of $W$ at $W_i$ times $O(h^2)$, because

$$
\begin{aligned}
& n^{-1} \sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi}_0)}{S(X_{1i}, X_{2j}, W_i) S_h(X_{1i}, X_{2j}, W_i)} O(h^2) \\
= \;& n^{-1} \sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi}_0)}{S(X_{1i}, X_{2j}, W_i)(S(X_{1i}, X_{2j}, W_i) + o(1))} O(h^2) \\
\lesssim \;& O(h^2) n^{-1} \sum_{j=1}^{n} \mathbf{K}_h(W_j - W_i) \\
\approx \;& f(W_i) O(h^2) \\
= \;& O(h^2)
\end{aligned}
$$

where "$\lesssim$" denotes "less than up to some constant". Therefore, $B = n^{1/2} O(h^2) = o_p(1)$, for $h$ satisfying C2.8.

Lastly, term $A$ can be decomposed as

$$
\begin{aligned}
& A \\
= \;& n^{\frac{1}{2}} n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi}_0)}{v(V_i, V_j; \boldsymbol{\xi}_0) S_h(X_{1i}, X_{2j}, W_i)} \left( v(V_i, V_j; \boldsymbol{\xi}_0) - S_h(X_{1i}, X_{2j}, W_i) \right) \\
= \;& n^{\frac{1}{2}} n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi}_0)}{v(V_i, V_j; \boldsymbol{\xi}_0) S_h(X_{1i}, X_{2j}, W_i)} \left( \frac{1}{n} N(X_{1i}, X_{2j}, W_i) - S_h(X_{1i}, X_{2j}, W_i) \right) \\
+ \;& n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi}_0)}{v(V_i, V_j; \boldsymbol{\xi}_0) S_h(X_{1i}, X_{2j}, W_i)} \frac{\mathbf{K}_h(0)}{n^{\frac{1}{2}}} I(X_{2j} \le X_{2i})(1 - \theta(X_{1i}, X_{2j}, W_i; \boldsymbol{\xi}_0)) \\
= \;& n^{\frac{1}{2}} n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi}_0)}{S_h(X_{1i}, X_{2j}, W_i)^2} \\
& \left( \frac{1}{n} N(X_{1i}, X_{2j}, W_i) - S_h(X_{1i}, X_{2j}, W_i) \right) + o_p(1) \\
= \;& n^{\frac{1}{2}} n^{-3} \sum_{k=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i) t(V_i, V_j; \boldsymbol{\xi}_0)}{S_h(X_{1i}, X_{2j}, W_i)^2} \\
& \left( I(X_{1k} \ge X_{1i}, X_{2k} \ge X_{2j}) \mathbf{K}_h(W_k - W_i) - S_h(X_{1i}, X_{2j}, W_i) \right) + o_p(1) \\
= \;& n^{\frac{1}{2}} \Bigg( \mathbb{P}_n^{\dagger} \mathbb{P}_n^{*} \mathbf{K}_h(W^{\dagger} - W^{*}) \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) t(V^{*}, V; \boldsymbol{\xi}_0)}{S_h(X_1^{*}, X_2, W^{*})^2} \times I(X_1^{\dagger} \ge X_1^{*}, X_2^{\dagger} \ge X_2) \\
& - P^{\dagger} \mathbb{P}_n^{*} \mathbf{K}_h(W^{\dagger} - W^{*}) \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) t(V^{*}, V; \boldsymbol{\xi}_0)}{S_h(X_1^{*}, X_2, W^{*})^2} \times I(X_1^{\dagger} \ge X_1^{*}, X_2^{\dagger} \ge X_2) \\
& + P^{\dagger} \mathbb{P}_n^{*} \mathbf{K}_h(W^{\dagger} - W^{*}) \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) t(V^{*}, V; \boldsymbol{\xi}_0)}{S_h(X_1^{*}, X_2, W^{*})^2} \times I(X_1^{\dagger} \ge X_1^{*}, X_2^{\dagger} \ge X_2) \\
& - \mathbb{P}_n^{*} \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) t(V^{*}, V; \boldsymbol{\xi}_0)}{S_h(X_1^{*}, X_2, W^{*})} \Bigg) + o_p(1) \\
= \;& \mathbb{G}_n^{\dagger} \left( \mathbb{P}_n^{*} \mathbf{K}_h(W^{*} - W^{\dagger}) \mathbb{P}_n \mathbf{K}_h(W - W^{*}) \frac{t(V^{*}, V; \boldsymbol{\xi}_0)}{S_h(X_1^{*}, X_2, W^{*})^2} I(X_1^{\dagger} \ge X_1^{*}, X_2^{\dagger} \ge X_2) \right) \\
& + n^{\frac{1}{2}} P^{\dagger} \mathbb{P}_n^{*} \mathbf{K}_h(W^{\dagger} - W^{*}) \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) t(V^{*}, V; \boldsymbol{\xi}_0)}{S_h(X_1^{*}, X_2, W^{*})^2} \times I(X_1^{\dagger} \ge X_1^{*}, X_2^{\dagger} \ge X_2) \\
& - n^{\frac{1}{2}} \mathbb{P}_n^{*} \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) t(V^{*}, V; \boldsymbol{\xi}_0)}{S_h(X_1^{*}, X_2, W^{*})} + o_p(1) \\
& = A_1 + A_2 - A_3 + o_p(1).
\end{aligned}
$$

Term $A_1$ can be further decomposed as

$$
\begin{aligned}
& A_1 \\
= \;& \mathbb{G}_n^\dagger\Big(\mathbb{P}_n^*\mathbf{K}_h(W^*-W^\dagger)\mathbb{P}_n\mathbf{K}_h(W-W^*)\frac{t(V^*,V;\boldsymbol{\xi}_0)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2)\Big) \\
= \;& \mathbb{G}_n^\dagger\Big(\mathbb{P}_n^*\mathbf{K}_h(W^*-W^\dagger)\mathbb{P}_n\mathbf{K}_h(W-W^*)\frac{t(V^*,V;\boldsymbol{\xi}_0)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2) \\
& -P^*\mathbf{K}_h(W^*-W^\dagger)P\mathbf{K}_h(W-W^*)\frac{t(V^*,V;\boldsymbol{\xi}_0)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2)\Big) \\
+ \;& \mathbb{G}_n^\dagger\Big(P^*\mathbf{K}_h(W^*-W^\dagger)P\mathbf{K}_h(W-W^*)\frac{t(V^*,V;\boldsymbol{\xi}_0)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2) \\
& -E_{V^*|W^*=W^\dagger}E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2)f(W^\dagger)\Big) \\
+ \;& \mathbb{G}_n^\dagger\Big(E_{V^*|W^*=W^\dagger}E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2)f(W^\dagger)\Big) \\
= \;& A_{11}+A_{12}+A_{13}.
\end{aligned}
$$

We will show that $A_{12}=o_p(1)$ and $A_{11}=o_p(1)$ separately. First of all,

$$
\begin{aligned}
& A_{12} \\
= \;& \mathbb{G}_n^\dagger\Big(P^*\mathbf{K}_h(W^*-W^\dagger)P\mathbf{K}_h(W-W^*)\frac{t(V^*,V;\boldsymbol{\xi}_0)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2) \\
& -E_{V^*|W^*=W^\dagger}E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2)f(W^\dagger)\Big) \\
= \;& \mathbb{G}_n^\dagger\Big(P^*\mathbf{K}_h(W^*-W^\dagger)P\mathbf{K}_h(W-W^*)\frac{t(V^*,V;\boldsymbol{\xi}_0)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2) \\
& -P^*\mathbf{K}_h(W^*-W^\dagger)E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2) \\
& +P^*\mathbf{K}_h(W^*-W^\dagger)E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2) \\
& -E_{V^*|W^*=W^\dagger}E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2)f(W^\dagger)\Big) \\
= \;& \mathbb{G}_n^\dagger\Big(P^*\mathbf{K}_h(W^*-W^\dagger)\Big\{P\mathbf{K}_h(W-W^*)\frac{t(V^*,V;\boldsymbol{\xi}_0)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2) \\
& \qquad -E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2)\Big\} \\
& +P^*\mathbf{K}_h(W^*-W^\dagger)E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2) \\
& -E_{V^*|W^*=W^\dagger}E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger\geq X_1^*,X_2^\dagger\geq X_2)f(W^\dagger)\Big)
\end{aligned}
$$

Note that by Lemma A.2 of Ichimura (1993),

$$\begin{aligned}
& P\mathbf{K}_h(W-W^*)\frac{t(V^*,V;\boldsymbol{\xi}_0)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2) \\
= \quad & E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2) + O(h^2)
\end{aligned}$$

and

$$\begin{aligned}
& P^*\mathbf{K}_h(W^*-W^\dagger)E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2) \\
= \quad & E_{V^*|W^*=W^\dagger}E_{V|W=W^*}\frac{t(V^*,V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2)f(W^\dagger)\Big) + O(h^2).
\end{aligned}$$

So term $A_{12} = n^{1/2}O(h^2) = o_p(1)$ for $h$ satisfying C2.8.

To show term $A_{11} = o_p(1)$, first for fixed $V^\dagger$ set

$$\begin{aligned}
& m_h(V,V^*,V^\dagger) \\
= \quad & h^{-1}\mathbf{K}\left(\frac{W^*-W^\dagger}{h}\right)\mathbf{K}\left(\frac{W-W^*}{h}\right)\frac{t(V^*,V;\boldsymbol{\xi}_0)}{S_h(X_1^*,X_2,W^*)^2}I(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2)
\end{aligned}$$

and

$$\tilde{m}_h(V,V^*,V^\dagger) = m_h(V,V^*,V^\dagger) - Pm_h(V,V^*,V^\dagger) - P^*m_h(V,V^*,V^\dagger) + PP^*m_h(V,V^*,V^\dagger),$$

and

$$T_n(\tilde{m}_h) = \sum_{1\leq i\neq j\leq n}\tilde{m}_h(V_i,V_j,V^\dagger).$$

Then term $A_{11}$ can be decomposed into:

$$\begin{aligned}
& \mathbb{G}_n^\dagger(h^{-1}\mathbb{P}_n^*\mathbb{P}_n m_h - h^{-1}P^*Pm_h) \\
= \quad & \mathbb{G}_n^\dagger\left(h^{-1}\mathbb{P}_n^*\mathbb{P}_n\tilde{m}_h + (\mathbb{P}_n^* - P^*)h^{-1}Pm_h + (\mathbb{P}_n - P)h^{-1}P^*m_h\right).
\end{aligned}$$

Note that $\mathbb{P}_n^*\mathbb{P}_n\tilde{m}_h$ is again a U-process. Using proof similar to one that shows $C_1 = o_p(1)$, we have

$$\mathbb{G}_n^\dagger\left(h^{-1}\mathbb{P}_n^*\mathbb{P}_n\tilde{m}_h\right) = o_p(1)$$

for $h$ satisfying C2.8.

Similarly, it can can be argued that $\mathbb{G}_n^\dagger(\mathbb{P}_n^* - P^*)h^{-1}Pm_h = o_p(1)$ and $\mathbb{G}_n^\dagger(\mathbb{P}_n - P)h^{-1}P^*m_h = o_p(1)$ following a calculation for showing $C_1 = o_p(1)$.

Now focusing on $X_1^\dagger, X_2^\dagger, W^\dagger$ and their probability measure $P^\dagger$, we have

$$\begin{aligned}
& A_2 \\
= \;& P^\dagger \mathbb{P}_n^* \mathbf{K}_h(W^\dagger - W^*)\mathbb{P}_n \frac{\mathbf{K}_h(W - W^*)t(V^*, V; \boldsymbol{\xi}_0)}{S_h(X_1^*, X_2, W^*)^2} \times I(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2) \\
= \;& \mathbb{P}_n^* \mathbb{P}_n \frac{\mathbf{K}_h(W - W^*)t(V^*, V; \boldsymbol{\xi}_0)}{S_h(X_1^*, X_2, W^*)^2} \times (S_h(X_1^*, X_2, W^*) + O(h^2)) \\
= \;& \mathbb{P}_n^* \mathbb{P}_n \frac{\mathbf{K}_h(W - W^*)t(V^*, V; \boldsymbol{\xi}_0)}{S_h(X_1^*, X_2, W^*)} + \mathbb{P}_n^* \mathbb{P}_n \frac{\mathbf{K}_h(W - W^*)t(V^*, V; \boldsymbol{\xi}_0)}{S_h(X_1^*, X_2, W^*)^2} O(h^2) \\
= \;& \mathbb{P}_n^* \mathbb{P}_n \frac{\mathbf{K}_h(W - W^*)t(V^*, V; \boldsymbol{\xi}_0)}{S_h(X_1^*, X_2, W^*)} + O_p(1)O(h^2) \\
= \;& A_3 + o_p(1)
\end{aligned}$$

By combining, we have

$$\begin{aligned}
& A_1 \\
= \;& A_{13} + o_p(1) \\
= \;& \mathbb{G}_n^\dagger \Big( E_{V^*|W^* = W^\dagger} E_{V|W=W^*} \frac{t(V^*, V; \boldsymbol{\xi}_0) f(W^*)}{S_h(X_1^*, X_2, W^*)^2} I(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2) f(W^\dagger) \Big) + o_p(1).
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
& n^{1/2}\boldsymbol{U}_n(\boldsymbol{\gamma}_0) \\
= \;& n^{1/2}\{\boldsymbol{U}_n(\boldsymbol{\gamma}_0) - \boldsymbol{u}(\boldsymbol{\gamma}_0)\} \\
= \;& n^{1/2}\{\boldsymbol{U}_n^{(1)}(\boldsymbol{\gamma}_0) - \boldsymbol{u}^{(1)}(\boldsymbol{\gamma}_0)\} - n^{1/2}\{\boldsymbol{U}_n^{(2)}(\boldsymbol{\gamma}_0) - \boldsymbol{u}^{(2)}(\boldsymbol{\gamma}_0)\} \\
& \quad + n^{1/2}\{\boldsymbol{U}_n^{(3)}(\boldsymbol{\gamma}_0) - \boldsymbol{u}^{(3)}(\boldsymbol{\gamma}_0)\} - n^{1/2}\{\boldsymbol{U}_n^{(4)}(\boldsymbol{\gamma}_0) - \boldsymbol{u}^{(4)}(\boldsymbol{\gamma}_0)\} \\
= \;& 2\mathbb{G}_n \Bigg\{ \Delta_1 \Delta_2 \dot{\beta}(V) - \tilde{h}^\star(V; \boldsymbol{\xi}_0) - \tilde{h}(V; \boldsymbol{\xi}_0) \\
& + E_{V^*|W^* = W^\dagger} E_{V|W=W^*} \frac{t(V^*, V; \boldsymbol{\xi}_0) f(W^*)}{S_h(X_1^*, X_2, W^*)^2} I(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2) f(W^\dagger) \Bigg\} + o_p(1) \\
\to_d \;& N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_0)).
\end{aligned}$$

Thus from (2.21) we obtain the desired asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0)$.

# CHAPTER III

# Cross-Ratio Regression for Bivariate Failure Times with Random Truncation

## 3.1 Introduction

Like in univariate survival analysis, observations in bivariate survival applications are sometimes subject to delayed entry also known as left truncation. We only observe pairs in which both failure events occur after the corresponding left-truncation events. Such data are very common in the cohort studies, and random truncation models have gained great interest in recent years.

Van der Laan (1996) described data comprising regular hospital visits of hemophilia patients between 1978 and 1995. Suppose that $T_1$ is lag time between human immunodeficiency virus (HIV) infection and manifestation of acquired immune deficiency syndrome (AIDS) and $T_2$ is lag time between HIV infection and death. Then $T_2$ is left-truncated if HIV infection occurs before 1978. In other word, we observe $(T_1, T_2, U_2)$, only if $T_2 \geq U_2$, where $U_2 = \max$(0, 1978-time at infection). One might be interested in the dependence between $T_1$ and $T_2$, because high dependence would suggest the utility of $T_1$ in predicting $T_2$, the overall survival time of AIDS patients. The failure time of interest, $T_1$, is not left-truncated. Taking its corresponding truncation time to be $0$, these data are a special case of bivariate left truncated data. Clearly, this is a case of biased sampling since patients with a short survival time $T_2$ are less likely to be part of the dataset than patients with a long $T_2$.

Like left truncation, right truncation could also arise in AIDS incubation cohort study of HIV-positive subjects without AIDS. One well known such dataset is Transfusion related AIDS (TR AIDS) data (Lagakos et al. 1988; Kalbfleisch and Lawless 1989; Gürler 1996). For patients who were thought to be infected with HIV by blood or blood–product transfusion, the data record the age of the patient $T_1$, the date the patient is infected with HIV $t$ and the date of AIDS diagnosis $s$. The observation period is terminated at $u$, same for all patients. Only those individuals for whom the incubation time (the time from HIV infection to the manifestation of AIDS) $T_2 = s - t < U_2 = u - t$ can be observed, where $T_2$ is left truncated by $U_2$. The primary interest here is in the distribution of incubation time and its dependence on age at infection, because one would suspect that disease progression differs across different age groups. The established inference procedures for a right truncated sample focus on the reverse-time transformation. Specifically, let $\tau = \max(U_2)$ be the largest observed time in the truncated sample. The transformed variable $T_2^* = \tau - T_2$ is left truncated by $U_2^* = \tau - U_2$. Methods developed in the context of left truncation can be readily extended to right truncation with reverse-time transformation. For the TR AIDS data, where $T_2$ is of primary interest, Lagakos et al. (1988) studied the distribution function on the reverse time axis. Kalbfleisch and Lawless (1989) discussed the Cox regression analysis on the reverse-time hazard.

Thus, in this chapter, we focus on the models for bivariate left truncated data. Without censoring, bivariate left-truncated data are of the form $(T, U)|T \geq U$, where $T = (T_1, T_2)$ is a bivariate failure time, $U = (U_1, U_2)$ is a bivariate left truncation time, and $T \geq U$ is a coordinatewise inequality. We consider the more general setting of bivariate failure time data, subject to both left truncation and right censoring, where each observation is of the form $(U_1, U_2, X_1, X_2, \Delta_1, \Delta_2)$, where $X_1 = \min(T_1, C_1)$, $X_2 = \min(T_2, C_2)$, $\Delta_1 = I(T_1 \leq C_1)$, and $\Delta_2 = I(T_2 \leq C_2)$. Each pair of correlated continuous failure times $(T_1, T_2)$ are subject to left truncation and right censoring by a pair of truncation time $(U_1, U_2)$ and censoring times $(C_1, C_2)$ respectively. For instance, consider a bivariate left-truncated, right-censored dataset (Ino et al. 2001) comprising survival time pairs for 50 brain tumor patients. Time from diagnosis to initiation of radiation therapy and time from diagnosis to tumor progression are left-truncated by time from diagnosis to chemotherapy and right-censored by the time of last follow-up.

To be precise, let $T_1$ be the time from diagnosis to initiation of radiation therapy, $T_2$ be the time from diagnosis to tumor progression and $U$ be the time from diagnosis to chemotherapy. The failure times $(T_1, T_2)$ are observed only if $T_1 \geq U$ and $T_2 \geq U$. Several authors (van der Laan (1996), Gürler (1997), Quale and van der Laan (2000)) have investigated models for the bivariate distributions function when observed data is bivariate random truncated. Martin and Betensky (2005) proposed a test statistic for Quasi-Independence of bivariate failure and truncation times via conditional Kendall's tau.

This chapter is partially motivated by TR AIDS data and the need to study the dependence of incubation time for AIDS on the age at infection of HIV in the presence of right truncation. This is a scenario where cross-ratio is extremely useful. We first develop models for left truncated data without covariate. For completeness, we extend the method to handle covariate. When covariate is discrete with finite levels, we group observations into stratum by covariate values. We treat whether an event happens at a time point or beyond along one time axis as a binary covariate and the other time component as the survival outcome variable, and then construct the corresponding partial likelihood function. We modify the risk set and relevant indicators to handle left truncations. When covariate is continuous, kernel smoothing is applied to the estimating equations. We obtain the parameter estimates by maximizing the pseudo-partial likelihood function. Such construction does not need any model for either the joint or the marginal survival function, and thus is robust against model misspecification. We show the proposed parameter estimator is consistent. We also established asymptotic normality of our estimator . The proposed methodology is readily extendable to the estimation of an arbitrary baseline cross-ratio function by using the tensor product splines. We assume that truncation times, censoring times and failure times are mutually independent and that there are no ties among observed times for each of the two time components.

## 3.2 Estimation

### 3.2.1 Cross-ratio with left truncation

The problem here is to estimate the cross-ratio function $\theta(t_1, t_2)$ based on $n$ $i.i.d.$ random draws from the conditional distribution of $(U_1, U_2, X_1, X_2, \Delta_1, \Delta_2)$ given that $(X_1 \geq U_1, X_2 \geq U_2)$. In other words, we only observe $(U_1, U_2, X_1, X_2, \Delta_1, \Delta_2)$ if $(X_1 \geq U_1, X_2 \geq U_2)$. Our idea is motivated by the connection between cross-ratio definition in (1.1) and the Cox model partial likelihood for the two-group regression problem, treating $\{j : T_{1j} = t_1\}$ and $\{j : T_{1j} > t_1\}$ as the "exposure" and "non-exposure" groups and $T_{2j}$ as the survival outcome. The cross-ratio $\theta(t_1, t_2)$ becomes the hazard ratio of $T_2$ between these two groups. It is also well know that incomplete data induced by left truncation yield bias in estimation. A common technique to account for left truncation is to redefine the at risk process taking into account the truncation information. By mimicking the partial likelihood idea and accounting for left truncation through modified at risk set and the relevant indicators, we can construct a similar objective function as follows:

$$L_n = \prod_{i=1}^{n} L_i^{(1)} L_i^{(2)}, \tag{3.1}$$

with

$$\begin{aligned} L_i^{(1)} &= \prod_{j=1}^{n} \left[ \frac{\theta(X_{1i}, X_{2j})^{I(X_{1j} = X_{1i} \geq U_{1j})}}{N(X_{1i}, X_{2j}) + (\theta(X_{1i}, X_{2j}) - 1) I(U_{2i} \leq X_{2j} \leq X_{2i})} \right]^{I(X_{1j} \geq X_{1i} \geq U_{1j}) \Delta_{2j} \Delta_{1i}} \\ L_i^{(2)} &= \prod_{j=1}^{n} \left[ \frac{\theta(X_{1j}, X_{2i})^{I(X_{2j} = X_{2i} \geq U_{2j})}}{N(X_{1j}, X_{2i}) + (\theta(X_{1j}, X_{2i}) - 1) I(U_{1i} \leq X_{1j} \leq X_{1i})} \right]^{I(X_{2j} \geq X_{2i} \geq U_{2j}) \Delta_{1j} \Delta_{2i}}, \end{aligned}$$

where $N(t_1, t_2) = \sum_{k=1}^{n} I(X_{1k} \geq t_1 \geq U_{1k}, X_{2k} \geq t_2 \geq U_{2k})$. The estimator obtained by maximizing (3.1) is then called the pseudo-partial likelihood estimator.

To proceed, we replace $\theta$ by $\beta$ and denote $\dot{\beta}_\gamma(t_1, t_2; \boldsymbol{\gamma}) = \partial \beta(t_1, t_2; \boldsymbol{\gamma}) / \partial \boldsymbol{\gamma}$. Taking logarithm of the objective function $L_n$ and differentiating it with respect to $\boldsymbol{\gamma}$, we obtain the following estimating function for $\boldsymbol{\gamma}$:

$$\boldsymbol{U}_n(\boldsymbol{\gamma}) = \nabla_\gamma \log L_n(\boldsymbol{\gamma}) = \boldsymbol{U}_n^{(1)}(\boldsymbol{\gamma}) - \boldsymbol{U}_n^{(2)}(\boldsymbol{\gamma}) + \boldsymbol{U}_n^{(3)}(\boldsymbol{\gamma}) - \boldsymbol{U}_n^{(4)}(\boldsymbol{\gamma}),$$

where

$$\boldsymbol{U}_n^{(1)}(\boldsymbol{\gamma}) = \boldsymbol{U}_n^{(3)}(\boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=1}^{n} \Delta_{1i}\Delta_{2i}\dot{\beta}_{\gamma}(X_{1i}, X_{2i}; \boldsymbol{\gamma}),$$

and

$$\begin{aligned}\boldsymbol{U}_n^{(2)}(\boldsymbol{\gamma}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\Delta_{1i}\Delta_{2j} I(X_{1j} \geq X_{1i} \geq U_{1j}) I(U_{2i} \leq X_{2j} \leq X_{2i})}{N(X_{1i}, X_{2j}) - I(U_{2i} \leq X_{2j} \leq X_{2i})(1 - e^{\beta(X_{1i}, X_{2j}; \boldsymbol{\gamma})})} \\ &\quad \times e^{\beta(X_{1i}, X_{2j}; \boldsymbol{\gamma})} \dot{\beta}_{\gamma}(X_{1i}, X_{2j}, \boldsymbol{\gamma}), \\ \boldsymbol{U}_n^{(4)}(\boldsymbol{\gamma}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\Delta_{1j}\Delta_{2i} I(X_{2j} \geq X_{2i} \geq U_{2j}) I(U_{1i} \leq X_{1j} \leq X_{1i})}{N(X_{1j}, X_{2i}) - I(U_{1i} \leq X_{1j} \leq X_{1i})(1 - e^{\beta(X_{1j}, X_{2i})}; \boldsymbol{\gamma})} \\ &\quad \times e^{\beta(X_{1j}, X_{2i}, \boldsymbol{\gamma})} \dot{\beta}_{\gamma}(X_{1j}, X_{2i}; \boldsymbol{\gamma}).\end{aligned}$$

Then an estimator $\hat{\boldsymbol{\gamma}}_n$ can be obtained by solving the equation $\boldsymbol{U}_n(\boldsymbol{\gamma}) = 0$ using the Newton-Raphson algorithm.

#### 3.2.2 Cross-ratio regression with covariate and left truncation

When covariates exist, cross-ratio is a quantity conditional on covariates. Specifically, the definition of cross-ratio becomes:

$$\theta(t_1, t_2, w) = \frac{\lambda_2(t_2|T_1 = t_1, W = w)}{\lambda_2(t_2|T_1 > t_1, W = w)} = \frac{\lambda_1(t_1|T_2 = t_2, W = w)}{\lambda_1(t_1|T_2 > t_2, W = w)}, \tag{3.2}$$

where $w$ denotes the covariate. Mimicking the Cox proportional hazard model where the effect of covariates is assumed to be multiplicative, we propose an analogous model where the effect of covariate is multiplicative on cross-ratio:

$$\theta(t_1, t_2, w) = \theta_0(t_1, t_2) exp(w'\boldsymbol{\alpha}), \tag{3.3}$$

where $\theta_0(t_1, t_2)$ is the baseline cross-ratio for $w = 0$. Here we consider a parametric model for $\beta_0(t_1, t_2; \boldsymbol{\gamma}) \equiv \log\theta_0(t_1, t_2)$ parameterized by a finite–dimensional Euclidean parameter $\boldsymbol{\gamma}$. Mimicking the estimating equations for the case without covariate, we propose the following for covariate with finite levels:

$$\boldsymbol{U}_n^*(\boldsymbol{\xi}) = \boldsymbol{U}_n^{*(1)}(\boldsymbol{\xi}) - \boldsymbol{U}_n^{*(2)}(\boldsymbol{\xi}) + \boldsymbol{U}_n^{*(3)}(\boldsymbol{\xi}) - \boldsymbol{U}_n^{*(4)}(\boldsymbol{\xi}),$$

where

$$\boldsymbol{U}_n^{*(1)}(\boldsymbol{\xi}) = \boldsymbol{U}_n^{*(3)}(\boldsymbol{\xi}) = \frac{1}{n}\sum_{i=1}^{n} \Delta_{1i}\Delta_{2i}\dot{\beta}_{\boldsymbol{\xi}}(X_{1i}, X_{2i}, W_i; \boldsymbol{\xi}),$$

and

$$\begin{aligned}
\boldsymbol{U}_n^{*(2)}(\boldsymbol{\xi}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{I(W_i = W_j)\Delta_{1i}\Delta_{2j}I(X_{1j} \geq X_{1i} \geq U_{1j})I(U_{2i} \leq X_{2j} \leq X_{2i})}{n \times N^*(X_{1i}, X_{2j}, W_i) - I(U_{2i} \leq X_{2j} \leq X_{2i})(1 - e^{\beta(X_{1i}, X_{2j}, W_i)})} \\
&\quad \times e^{\beta(X_{1i}, X_{2j}, W_i)}\dot{\beta}_{\boldsymbol{\xi}}(X_{1i}, X_{2j}, W_i; \boldsymbol{\xi}), \\
\boldsymbol{U}_n^{*(4)}(\boldsymbol{\xi}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{I(W_i = W_j)\Delta_{1j}\Delta_{2i}I(X_{2j} \geq X_{2i} \geq U_{2j})I(U_{1i} \leq X_{1j} \leq X_{1i})}{n \times N^*(X_{1j}, X_{2i}W_i) - I(U_{1i} \leq X_{1j} \leq X_{1i})(1 - e^{\beta(X_{1j}, X_{2i}, W_i)})} \\
&\quad \times e^{\beta(X_{1j}, X_{2i}, W_i)}\dot{\beta}_{\boldsymbol{\xi}}(X_{1j}, X_{2i}, W_i; \boldsymbol{\xi}),
\end{aligned}$$

where $N^*(t_1, t_2, w) = \sum_{k=1}^{n} I(X_{1k} \geq t_1 \geq U_{1k}, X_{2k} \geq t_2 \geq U_{2k}, W_k = w)$.

When the covariate is continuous, grouping observations into distinct groups does not work anymore. However, replacing the grouping indicators with kernel smoothing functions, we obtain the following kernel smoothed estimating equations:

$$\boldsymbol{U}_n^{\dagger}(\boldsymbol{\xi}) = \boldsymbol{U}_n^{\dagger(1)}(\boldsymbol{\xi}) - \boldsymbol{U}_n^{\dagger(2)}(\boldsymbol{\xi}) + \boldsymbol{U}_n^{\dagger(3)}(\boldsymbol{\xi}) - \boldsymbol{U}_n^{\dagger(4)}(\boldsymbol{\xi}),$$

where

$$\boldsymbol{U}_n^{\dagger(1)}(\boldsymbol{\xi}) = \boldsymbol{U}_n^{\dagger(3)}(\boldsymbol{\xi}) = \frac{1}{n}\sum_{i=1}^{n} \Delta_{1i}\Delta_{2i}\dot{\beta}_{\boldsymbol{\xi}}(X_{1i}, X_{2i}, W_i; \boldsymbol{\xi}),$$

and

$$\begin{aligned}
\boldsymbol{U}_n^{\dagger(2)}(\boldsymbol{\xi}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i)\Delta_{1i}\Delta_{2j}I(X_{1j} \geq X_{1i} \geq U_{1j})I(U_{2i} \leq X_{2j} \leq X_{2i})}{n \times N^{\dagger}(X_{1i}, X_{2j}, W_i) - I(U_{2i} \leq X_{2j} \leq X_{2i})(1 - e^{\beta(X_{1i}, X_{2j}, W_i)})} \\
&\quad \times e^{\beta(X_{1i}, X_{2j}, W_i)}\dot{\beta}_{\boldsymbol{\xi}}(X_{1i}, X_{2j}, W_i; \boldsymbol{\xi}), \\
\boldsymbol{U}_n^{\dagger(4)}(\boldsymbol{\xi}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\mathbf{K}_h(W_j - W_i)\Delta_{1j}\Delta_{2i}I(X_{2j} \geq X_{2i} \geq U_{2j})I(U_{1i} \leq X_{1j} \leq X_{1i})}{n \times N^{\dagger}(X_{1j}, X_{2i}, W_i) - I(U_{1i} \leq X_{1j} \leq X_{1i})(1 - e^{\beta(X_{1j}, X_{2i}, W_i)})} \\
&\quad \times e^{\beta(X_{1j}, X_{2i}, W_i)}\dot{\beta}_{\boldsymbol{\xi}}(X_{1j}, X_{2i}, W_i; \boldsymbol{\xi}),
\end{aligned}$$

where $N^{\dagger}(t_1, t_2, w) = \sum_{k=1}^{n} I(X_{1k} \geq t_1 \geq U_{1k}, X_{2k} \geq t_2 \geq U_{2k})\mathbf{K}_h(W_k - w)$.

## 3.3 Asymptotic properties

Here we discuss the asymptotic properties for the with-covariate case, because without-covariate case is equivalent to the with-covariate case where the covariate has only one level. We consider a parametric model with finite number of terms for $\beta_0(t_1, t_2; \boldsymbol{\gamma})$. In particular, we assume

$$\beta(t_1, t_2, w; \boldsymbol{\xi}) = \sum_{k,l} \gamma_{kl} b_{kl}(t_1, t_2) + w' \boldsymbol{\alpha}, \tag{3.4}$$

where $\boldsymbol{\xi}$ is the vector of coefficients $\{\gamma_{kl}\}$ and $\boldsymbol{\alpha}$, and $\{b_{kl}\}$ are the basis functions of $t_1$ and $t_2$ that do not involve the parameter $\boldsymbol{\xi}$. In this section, we provide asymptotic results for the estimation of $\boldsymbol{\xi}$ in (3.4). In particular, we consider functions of bounded variations for $\{b_{kl}\}$, e.g. indicators functions or polynomial functions on a compact set. Such regularity conditions guarantee that both $\beta(t_1, t_2, w; \boldsymbol{\xi})$ and $\dot{\beta}_{\boldsymbol{\xi}}(t_1, t_2, w)$ belong to Donsker classes. For model (3.4) we consider the following regularity conditions:

C3.1. The covariate $W$ is either continuous or discrete with finite levels, whose sample space $\mathcal{W}$ is bounded with $0 < \inf_{w \in \mathcal{W}} f(w)$ and $\sup_{w \in \mathcal{W}} f(w) < \infty$. Here $f$ is the density function of $W$.

C3.2. The failure times will be truncated at $(\tau_1, \tau_2)$, $0 < \tau_1, \tau_2 < \infty$, such that $Pr(T_1 > \tau_1 > U_1, C_1 > \tau_1 > U_1, T_2 > \tau_2 > U_2, C_2 > \tau_2 > U_2 | W = w) > 0$ for any $w \in \mathcal{W}$.

C3.3. The parameter space of $\boldsymbol{\xi}$, $\Gamma$, is a compact set, and the true value $\boldsymbol{\xi}_0$ is an interior point of $\Gamma$.

C3.4. The matrix $E\{\Delta_1 \Delta_2 \dot{\boldsymbol{\beta}}(X_1, X_2, W)^{\otimes 2}\}$ is positive definite. Here $\dot{\boldsymbol{\beta}}^{\otimes 2} = \dot{\boldsymbol{\beta}} \dot{\boldsymbol{\beta}}'$.

C3.5. $(T_1, T_2)$ $(C_1, C_2)$ and $(U_1, U_2)$ are mutually independent conditional on $W$.

**Theorem III.1.** *Under Conditions C1-C5, the solution of $\boldsymbol{U}_n(\boldsymbol{\xi}) = 0$, denoted by $\hat{\boldsymbol{\xi}}_n$, is a consistent estimator of $\boldsymbol{\xi}_0$.*

The proof of Theorem III.1 is treated separately for discrete $W$ with finite levels and continuous $W$, but follows the same steps. We first show that $\boldsymbol{U}_n(\boldsymbol{\xi})$ converges to a deterministic function $\boldsymbol{u}(\boldsymbol{\xi})$

uniformly, then show that $\boldsymbol{u}(\boldsymbol{\xi})$ is monotone and has a unique root at $\boldsymbol{\xi}_0$. The consistency will follow easily. Details are provided in the Appendix.

**Theorem III.2.** *Under Conditions C1-C5, we have that* $n^{-1/2}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0)$ *converges in distribution to a normal random variable with mean zero and variance* $\boldsymbol{I}(\boldsymbol{\xi}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)\boldsymbol{I}(\boldsymbol{\xi}_0)^{-1}$, *where* $\boldsymbol{I}(\boldsymbol{\xi}_0) = 2E\{\Delta_1\Delta_2\dot{\boldsymbol{\beta}}(X_1, X_2, W)^{\otimes 2}\}$ *and* $\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)$ *is the asymptotic variance of* $\boldsymbol{U}_n(\boldsymbol{\xi}_0)$, *which is given in the Appendix.*

The asymptotic normality in Theorem III.2 can be achieved by using the Taylor expansion of $\boldsymbol{U}_n(\hat{\boldsymbol{\xi}}_n)$ around $\boldsymbol{\xi}_0$. Again the detailed calculation is deferred to the Appendix. The asymptotic expression of $\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)$ also provides an variance estimator of $n^{1/2}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0)$.

### 3.4 Simulations

We consider the Frank family as in Fan et al. (2000a,b). We begin by generating independent Uniform (0,1) random numbers $u_1$ and $u_2$. Then let $t_1 = -\log u_1$ so that $T_1$ follows unit exponential distribution. Finally let $t_2 = -\log\big(\log_\alpha[a/\{a + (1-\alpha)u_2\}]\big)$ where $a = \alpha^{u_1} + (\alpha - \alpha^{u_1})u_2$. Such generated $T_2$ also follows exponential distribution. The cross-ratio function is

$$\frac{(\alpha-1)\log(\alpha)\alpha^{2-e^{-t_1}-e^{-t_2}}}{(\alpha^{1-e^{-t_1}}-\alpha)\times(\alpha^{1-e^{-t_2}}-\alpha)}\Bigg[-1+e^{-t_1}+e^{-t_2} + \log_\alpha\left\{1+\frac{(\alpha^{1-e^{-t_1}}-1)(\alpha^{1-e^{-t_2}}-1)}{\alpha-1}\right\}\Bigg].$$

We follow the same simulation setup as in Chapter I. Additionally, we generate truncation times $U_1$ and $U_2$ from Uniform $(0,1)$ distribution. The estimated cross-ratio is obtained using the cubic polynomial model by maximizing the pseudo-partial likelihood function (3.1) with respect to coefficients $\boldsymbol{\gamma}$. Results are averaged over 1000 simulation runs, each with a sample size of 400.

The top panel of Figures 3.1 compares our estimator with naive estimator which ignores the truncation. It can be seen that our estimator corrects the bias properly. The bottom panel gives the cross-ratio

Table 3.1: Comparison of empirical variance and model based variance for the Frank family. The points on both margins are the quartiles of the marginal distributions of $X_1$ and $X_2$, which are different from the quartiles in Table 1.1 of Chapter I due to left truncation. The true log cross-ratio is $\beta$ and its estimator is $\hat{\beta}$. The first value in the parentheses is the empirical standard error, the second value is the mode based standard error estimator, and the third value is the coverage rate of the 95% confidence interval.

| | $X_2$ | | | | | |
|---|---|---|---|---|---|---|
| | 25% | | 50% | | 75% | |
| $X_1$ | $\beta$ | $\hat{\beta}$ | $\beta$ | $\hat{\beta}$ | $\beta$ | $\hat{\beta}$ |
| 25% | 1.01 | 0.99(0.19,0.19,96%) | 0.82 | 0.82(0.20,0.20,95%) | 0.63 | 0.67(0.30,0.30,94%) |
| 50% | 0.82 | 0.81(0.20,0.20,96%) | 0.70 | 0.67(0.18,0.18,94%) | 0.56 | 0.55(0.26,0.27,96%) |
| 75% | 0.63 | 0.67(0.30,0.30,95%) | 0.56 | 0.55(0.25,0.26,97%) | 0.46 | 0.46(0.28,0.30,96%) |

as a function of one time component fixing the other time component from $t = 0.25$ to $t = 1.50$ with 0.25 time unit increment. Based on the empirical variance of $\hat{\gamma}$, we calculate the confidence bands for $\beta$. Then by exponentiating, we obtain the empirical confidence bands for $\theta$. Figure 3.1 shows that the proposed method estimates the true cross-ratio of the Frank family very well, despite the fact that the working model is only an approximation of the true $\theta(t_1, t_2)$.

To check the performance of the model based variance estimator, we choose nine points based on the percentiles of the failure time distribution, and calculate the empirical variances based on 1000 replications and the average of the model based variance estimates. Results are given in Table 3.1, showing that the model based variance estimator works well.

We also simulated data using algorithm in Nan et al. (2006) with a binary $W \sim Bernoulli(0.5)$ and $\alpha = 0.5$. For $W = 0$, first we assume that the cross-ratio $\theta(t_1)$ is piecewise constant over four intervals: $\theta(t_1) = .9$ when $t_1 \in [0, .25)$, $\theta(t_1) = 2.0$ when $t_1 \in [.25, .5)$, $\theta(t_1) = 4.0$ when $t_1 \in [.5, .75)$, and $\theta(t_1) = 1.5$ when $t_1 > .75$. For $W = 1$, the cross-ratio $\theta(t_1)$ is equal to $0.9 \times e^{0.5}$, $2.0 \times e^{0.5}$, $4.0 \times e^{0.5}$ and $1.5 \times e^{0.5}$ in the above intervals. Additional, we generated truncation times $U_1$ and $U_2$ from Uniform (0,1). The results based on sample size of 400 and 800 summarized in table 3.2 show that our estimators work well. The proposed variance model based variance estimator also works well since the coverage probabilities are all close to the nominal coverage probability, $95\%$.

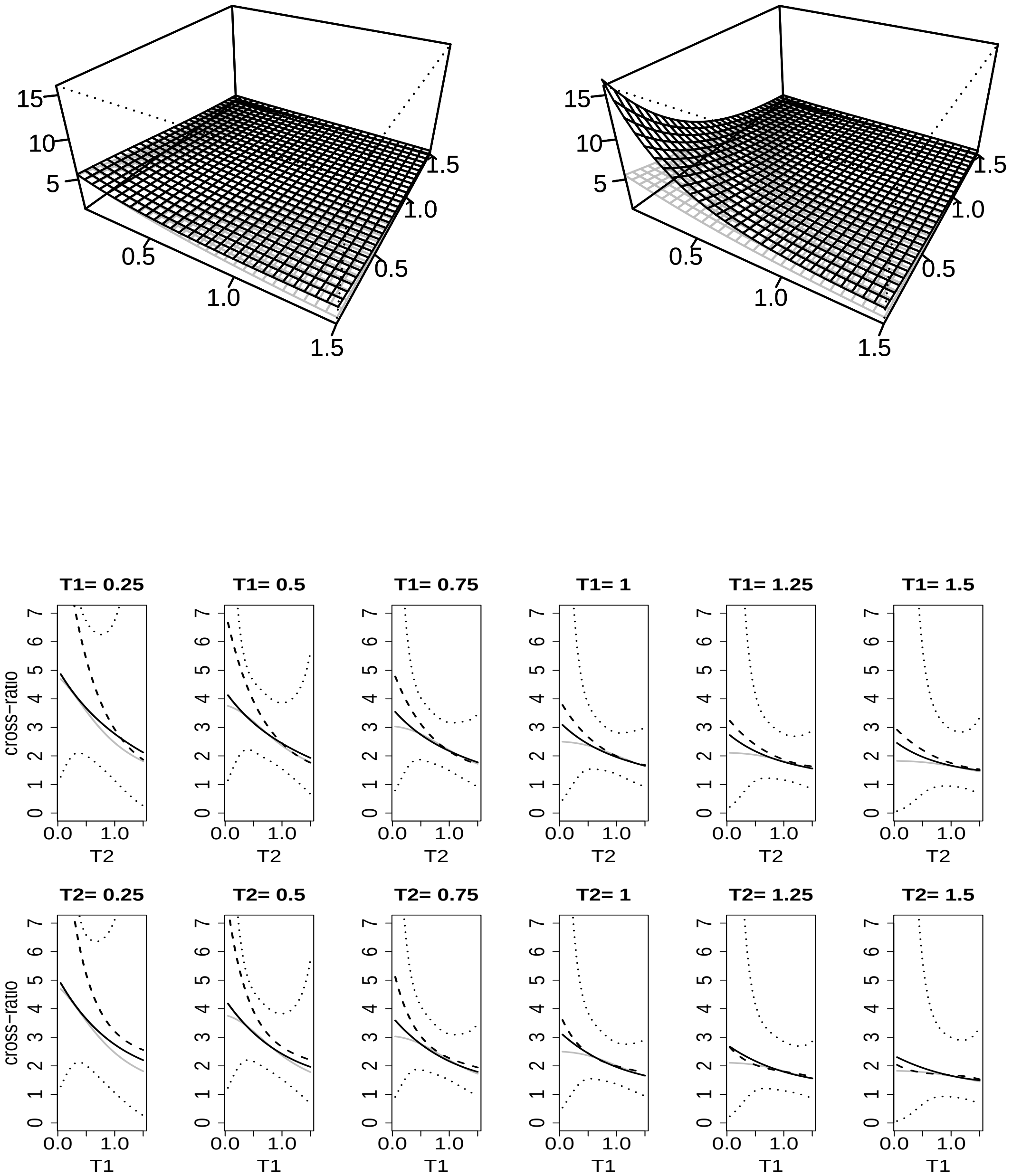


Figure 3.1: Cross-ratio for the Frank family. In the top panel, the surface in gray is the true cross-ratio and the surface in black is the estimated cross-ratio. The figure on the left is the estimator accounting for truncation. The figure on the right is the naive estimator, ignoring truncation. In the bottom panel, gray lines are the true cross-ratio, the black solid lines are the estimated cross-ratio, the dotted lines are their point-wise 95% confidence bands and the black dashed lines are the naive estimates, ignoring truncation.

Table 3.2: The true baseline cross ratios are $\theta = (.9, 2.0, 4.0, 1.5)$ when $t_1$ is in intervals $[0, .25)$, $[.25, .5)$, $[.5, .75)$, and above .75. The true $\alpha$ is 0.5 and $W \sim$ Bernoulli(0.5). The sample size is 400 and 800. $\hat{\beta}$, point estimate average; $E.SE$, the empirical standard error; $M.SE$, the average of the model based standard error estimates; $M.CP$ the $95\%$ coverage probability using $M.SE$.

| | | n=400 | | | | n=800 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | $\hat{\alpha}$ | $E.SE$ | $M.SE$ | $M.CP$ | $\hat{\alpha}$ | $E.SE$ | $M.SE$ | $M.CP$ |
| | 0.50 | 0.50 | 0.23 | 0.25 | 96% | 0.50 | 0.16 | 0.17 | 96% |
| $\theta$ | $\beta$ | $\hat{\beta}$ | $E.SE$ | $M.SE$ | $M.CP$ | $\hat{\beta}$ | $E.SE$ | $M.SE$ | $M.CP$ |
| 0.90 | -0.11 | -0.10 | 0.41 | 0.40 | 95% | -0.11 | 0.26 | 0.26 | 95% |
| 2.00 | 0.69 | 0.71 | 0.28 | 0.28 | 95% | 0.70 | 0.19 | 0.19 | 95% |
| 4.00 | 1.39 | 1.42 | 0.28 | 0.29 | 96% | 1.41 | 0.19 | 0.19 | 96% |
| 1.50 | 0.41 | 0.41 | 0.20 | 0.21 | 96% | 0.41 | 0.14 | 0.14 | 96% |

## 3.5 Data Application

### 3.5.1 Transfusion related AIDS

First we give a brief example of the implementing our methods for bivariate truncated TR AIDS data, which is uncensored. In this dataset, we observe 295 i.i.d. copies of recorded age at transfusion and time from transfusion to AIDS. However, only those subjects who had received a diagnosis of AIDS prior to July 1986, the end of the study, were included in the study. Thus, if we define $T_1$ as age at transfusion, $T_2$ as time from transfusion to AIDS and $U_2$ as time from transfusion to July 1986, we are only able to observe $(T_1, T_2, U_2)$ if $T_2 \leq U_2$. This is a special case of randomly right truncated bivariate survival data, by taking $U_1 = \infty$.

We discarded one questionable observations in this dataset. The observation we discarded had a $T_2$ value of zero, which means that this subject most likely contracted AIDS from a source other than the blood transfusion and is thus not of interest here. In order to apply the method developed for left truncated data, we reverse the time axis of $T_2$ by subtracting it from the maximum $U_2$. Specifically, let $\tau = \max(U_2)$ be the largest observed time in the truncated sample. The transformed variable $T_2^* = \tau - T_2$ is left truncated by $U_2^* = \tau - U_2$ (Figure 3.2).

With this transformation and applying the proposed method, we fit the following model and obtain

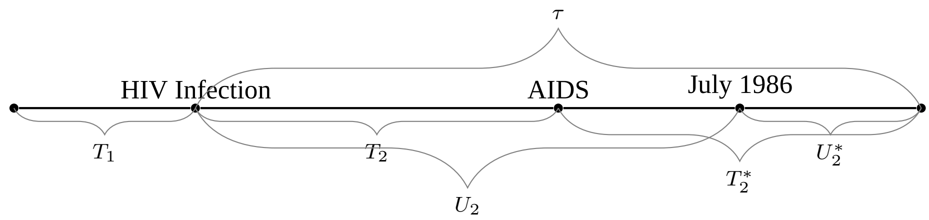


Figure 3.2: Illustration of TR AIDS data. $T_1$ = Age at HIV infection, $T_2$ = Incubation time, and $U_2$ = Time from HIV infection to July 1986, $\tau = \max(U_2)$, $U_2^* = \tau - U_2$ and $T_2^* = \tau - T_2$.

an estimate of the cross-ratio for age and reversed incubation time:

$$\beta(t_1, t_2^*; \gamma) = \gamma_0 + \gamma_1 t_1 + \gamma_2 t_2^* + \gamma_3 t_1^2 + \gamma_4 t_2^{*2} + \gamma_5 t_1 t_2^*.$$

Intuitively, the dependence between $T_1$ and $T_2$ is in the opposite direction of the dependence of $T_1$ and $T_2^*$. For the ease of interpretation, we further take the reciprocal of the estimated cross-ratio so that, with respect to $T_1$, the result can be interpreted as the dependence for $T_1$ and $T_2$. Based on the estimated cross-ratio surface, we see that age and the disease incubation are highly associated at a young age meaning that for children the disease incubation is much faster than adults. Such dependence diminishes quickly as the patient gets older, indicating that incubation does not differ much between the adults and the elderly. This finding is consistent with previous finding based on discrete age groups (children (1–4), adults (5–59) and elderly (60+)) determined according to immuno competence (Lagakos et al. 1988; Kalbfleisch and Lawless 1989; Gürler 1996).

#### 3.5.2 Anaplastic Oligodendroglioma Study

We then illustrate the the proposed method with left truncated, right-censored anaplastic oligodendroglioma dataset (Ino et al. 2001). The dataset contains 50 patients treated at the London Regional Cancer Center with histologically confirmed anaplastic oligodendrogliomas in whom chemotherapy was used as an integral part of an overall patient management strategy from diagnosis. Thirty-four of these patients received radiation therapy after chemotherapy (as consolidation or at recurrence), 5 received radiation therapy concurrent with chemotherapy, and 11 were not irradiated. Median follow-up time from diagnosis was 107 months (minimum, 7 months).

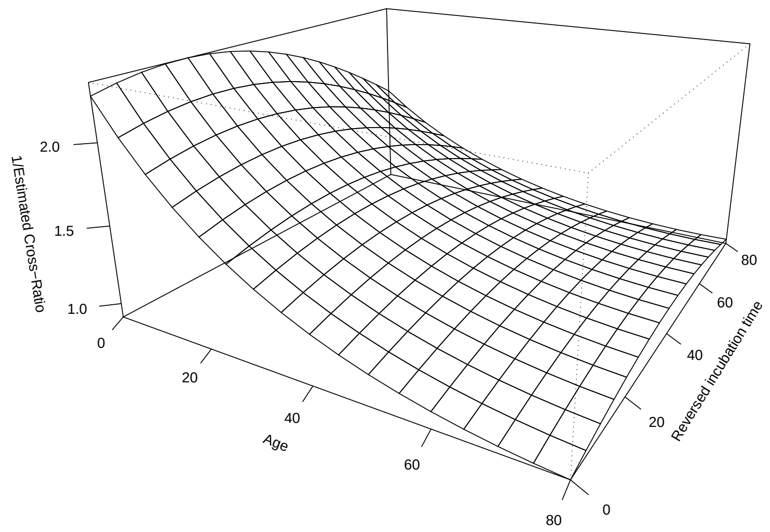


Figure 3.3: Reciprocal of estimated cross-ratio for age at transfusion $T_1$ and reverse transformed incubation time $T_2^*$.

Time from diagnosis to initiation of radiation therapy $T_1$ and time from diagnosis to tumor progression $T_2$ are left-truncated by time from diagnosis to chemotherapy and right-censored by time from diagnosis to end of study. Since the number of observations for the study is small, we start with a quadratic polynomial basis functions for $\beta(t_1, t_2; \gamma)$, i.e.,

$$\beta(t_1, t_2; \gamma) = \gamma_0 + \gamma_1(t_1 + t_2) + \gamma_2(t_1^2 + t_2^2) + \gamma_3 t_1 t_2$$

and use backward selection for variable selection. Table (3.3) illustrates the selection history. The final selected model is $\hat{\theta}(t_1, t_2) = \exp(2.86 - 1.72 \cdot 10^{-1} \cdot t_2 + 2.10 \cdot 10^{-3} \cdot t_2^2)$ (Figure 3.4). It is interesting to see that the dependence between $T_1$ and $T_2$ is a function of $T_2$ only. Intuitively, the result means for fast tumor progression (short tumor progression time), time from diagnosis to initiation of radiation therapy is highly predictive of time from diagnosis to tumor progression because of the strong correlation between the two time components; when time from diagnosis to tumor progression is long, time from diagnosis to initiation of radiation therapy is less predictive because of the weak correlation between the two time components.

Table 3.3: Estimated cross-ratio for Time from diagnosis to initiation of radiation therapy $T_1$ and time from diagnosis to tumor progression $T_2$, both left-truncated by time from diagnosis to chemotherapy and right-censored by time from diagnosis to end of study

| | Step 1 | | Step 2 | | Step 3 | | Step 4 | |
|---|---|---|---|---|---|---|---|---|
| coefficient | Estimate | $P$ | Estimate | $P$ | Estimate | $P$ | Estimate | $P$ |
| $\gamma_0(1)$ | $2.59$ | 0.67 | $2.65$ | $< 0.05$ | $2.87$ | $< 0.05$ | $2.86$ | $< 0.05$ |
| $\gamma_1(t_1)$ | $3.90 \cdot 10^{-1}$ | 0.93 | $3.46 \cdot 10^{-1}$ | 0.51 | | | | |
| $\gamma_2(t_2)$ | $-4.84 \cdot 10^{-1}$ | 0.88 | $-4.53 \cdot 10^{-1}$ | 0.21 | $-1.73 \cdot 10^{-1}$ | $< 0.05$ | $-1.72 \cdot 10^{-1}$ | $< 0.05$ |
| $\gamma_3(t_1^2)$ | $-2.60 \cdot 10^{-3}$ | 0.90 | $-3.61 \cdot 10^{-3}$ | 0.51 | $-4.23 \cdot 10^{-4}$ | 0.16 | | |
| $\gamma_4(t_2^2)$ | $5.88 \cdot 10^{-3}$ | 0.92 | $5.03 \cdot 10^{-3}$ | 0.13 | $1.85 \cdot 10^{-3}$ | $< 0.05$ | $2.10 \cdot 10^{-3}$ | $< 0.05$ |
| $\gamma_5(t_1 t_2)$ | $2.07 \cdot 10^{-3}$ | 0.98 | | | | | | |

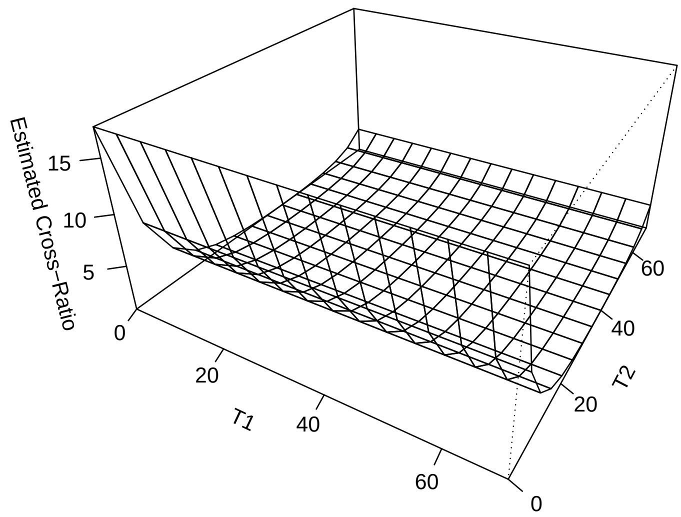


Figure 3.4: Estimated cross-ratio for time from diagnosis to initiation of radiation therapy $T_1$ and time from diagnosis to tumor progression $T_2$, both left-truncated by time from diagnosis to chemotherapy and right-censored by time from diagnosis to end of study.

### 3.5.3 Australian Twin Study

The last dataset we analyze is again the Australian Twin Study dataset for illustration purpose. The original dataset is not left truncated, but we artificially generate bivariate truncation times to illustrate how our proposed method can be applied to correct bias. Moreover since we have results based on complete data with which we can compare the new results, we can assess the performance of the proposed method.

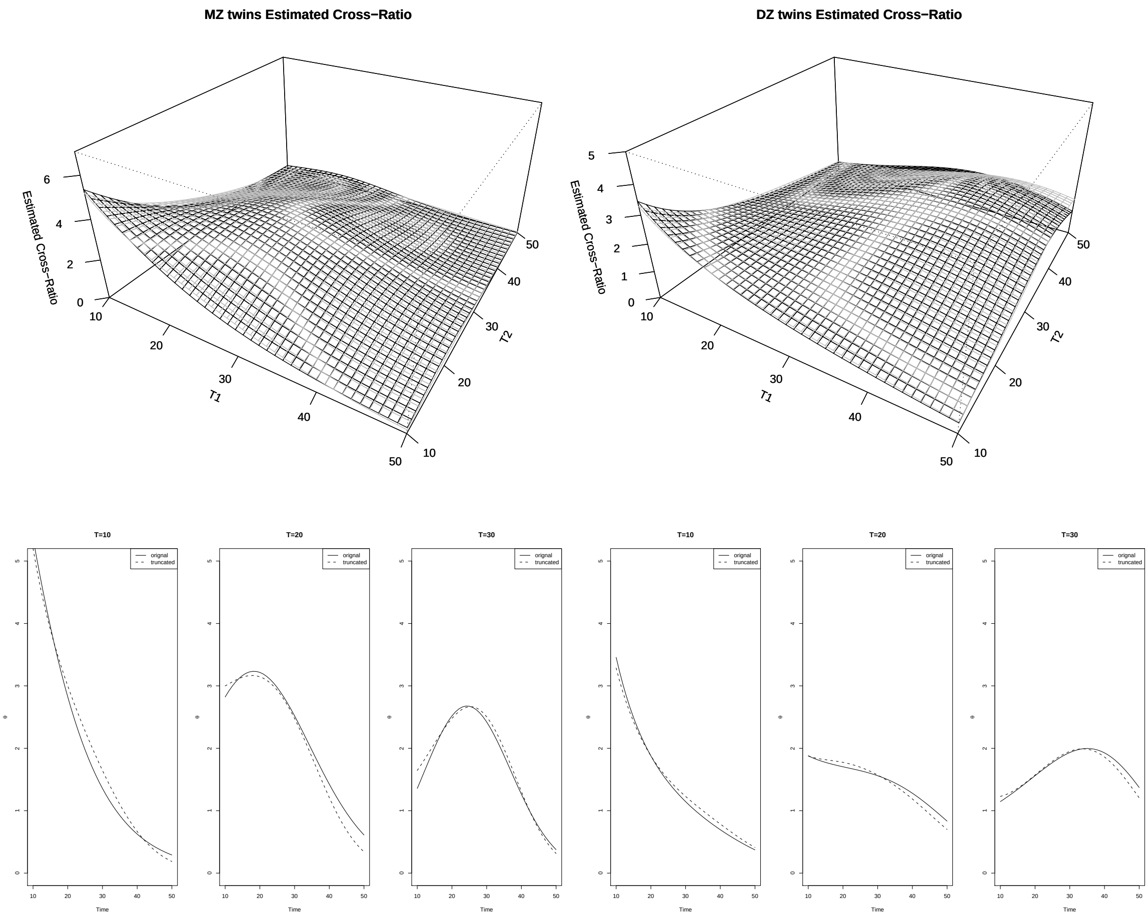


Figure 3.5: The top left figure shows the estimated cross-ratio of monozygotic twins. The black mesh is the result of complete data and grey mesh is based on 100 truncated datasets, each with 20% truncation rate. Fixing $T_1$ at 10, 20 and 30, in the three plots on the bottom left, solid curves are the cross-ratio estimated from complete data and grey curves are the average of estimated cross-ratio based on 100 truncated datasets. On the right, the plots show the results of dizygotic twins.

Applying our method, we first compare the average of results from 100 truncated datasets with 20% truncation rate with the complete data analysis for monozygotic twins and dizygotic twins separately (Figure 3.5). Fixing one time component at 10, 20 and 30, it is clearly seen that our proposed method successfully recovers the cross-ratio estimated from the complete data.

Pooling monozygotic twins and dizygotic twins and treating zygocity as binary covariate, we also perform the same comparison of the estimated baseline cross-ratio (dizygotic twins) estimated from complete data and the average of estimated cross-ratio. Figure 3.6 shows that our proposed method satisfactorily corrects the bias resulting from random truncation.

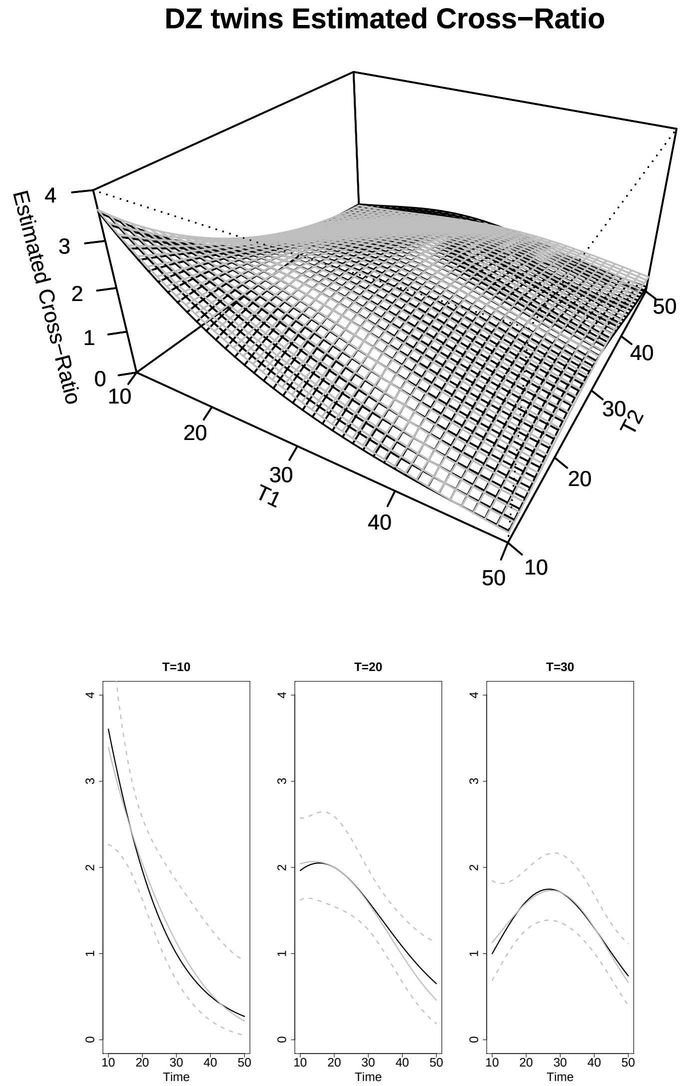


Figure 3.6: The black mesh is the result of complete data and grey mesh is based on 100 truncated datasets, each with 20% truncation rate. Fixing $T_1$ at 10, 20 and 30, in the three plots on the right, solid curves are the cross-ratio estimated from complete data and grey curves are the average of estimated cross-ratio based on 100 truncated datasets.

## 3.6 Discussion

Most approaches to estimating cross-ratio as a measure of dependence among correlated failures times have been based on the assumption that cross-ratio function is a constant or piece-wise constant over time. With such assumption, it is possible to come up with a likelihood based estimation method for cross-ratio. However, it is not clear how flexible those approaches are when left truncation arises, which is very common in observational studies. In this chapter, we show how a pseudo-partial likelihood approach motivated by the Cox model can be applied to doubly truncated data or bivariate left-truncated, right-censored data, that cannot be easily handled in existing method for estimating cross-ratio. The use of risk sets as in the Cox partial likelihood function allows simple modification to accommodate left truncation.

We analyzed TR AIDS data which is right truncated. A right truncated variable can be converted to a left truncated variable if one reverses the time axis. Using model developed for left truncation with reverse-time technique, we estimate the cross-ratio for age and reversed incubation time. However, interpretation of such result is awkward, and therefore, inferences on the regular forward-time cross-ratio for right truncated data warrants further investigation.

## 3.7 Appendix

### 3.7.1 Proof of Theorem III.1

Here we focus on the proofs for discrete covariate case only. The proofs for continuous covariate is analogous to the proof in Chapter II, i.e. using results Härdle (1988) to show consistency and using U-process results in Nolan (1987) to obtain asymptotic normality of the parameter estimates.

First define the following simplified notation: Let $\alpha_w \equiv Pr(X_1 \geq U_1, X_2 \geq U_2 | W = w)$ and $h(u_1, u_2|w)$, $H(u_1, u_2|w)$ be the bivariate pdf and cdf of $(U_1, U_2|W)$ conditional on $W = w$, respectively. Then the joint conditional density function of $(U_1, U_2, X_1, X_2, \Delta_1, \Delta_2)$ conditional on $W = w$

can be written as

$$
\begin{aligned}
& p(u_1, u_2, x_1, x_2, \delta_1, \delta_2 | x_1 \geq u_1, x_2 \geq u_2, w) \\
= \;& h(u_1, u_2|w) I(u_1 \leq x_1, u_2 \leq x_2)/\alpha \partial_{1,2} F(x_1, x_2|w)^{\delta_1 \delta_2} \{-\partial_1 F(x_1, x_2|w)\}^{\delta_1(1-\delta_2)} \\
& \{-\partial_2 F(x_1, x_2|w)\}^{(1-\delta_1)\delta_2} F(x_1, x_2|w)^{(1-\delta_1)(1-\delta_2)} \partial_{1,2} G(x_1, x_2|w)^{(1-\delta_1)(1-\delta_2)} \\
& \quad \{-\partial_1 G(x_1, x_2|w)\}^{(1-\delta_1)\delta_2} \{-\partial_2 G(x_1, x_2|w)\}^{\delta_1(1-\delta_2)} G(x_1, x_2|w)^{\delta_1 \delta_2}. \qquad (3.5)
\end{aligned}
$$

Let $V^* = (U_1^*, U_2^*, X_1^*, X_2^*, \Delta_1^*, \Delta_2^*, W^*)$ be an independent identically distributed copy of $V = (U_1, U_2, X_1, X_2, \Delta_1, \Delta_2, W)$. Define the deterministic function $\boldsymbol{u}^*(\boldsymbol{\xi}) = \boldsymbol{u}^{*(1)}(\boldsymbol{\xi}) - \boldsymbol{u}^{*(2)}(\boldsymbol{\xi}) + \boldsymbol{u}^{*(3)}(\boldsymbol{\xi}) - \boldsymbol{u}^{*(4)}(\boldsymbol{\xi})$, where

$$
\begin{aligned}
\boldsymbol{u}^{*(1)}(\boldsymbol{\xi}) = \boldsymbol{u}^{*(3)}(\boldsymbol{\xi}) &= E\{\Delta_1 \Delta_2 \dot{\beta}(X_1, X_2, W)\}, \\
\boldsymbol{u}^{*(2)}(\boldsymbol{\xi}) = \boldsymbol{u}^{*(2)}(\boldsymbol{\xi}) &= E\Big\{\Delta_1^* \Delta_2 \dot{\beta}(X_1^*, X_2, W^*) \\
& \quad \frac{I(X_1 \geq X_1^* \geq U_1) I(U_2^* \leq X_2 \leq X_2^*) \exp\{\beta(X_1^*, X_2, W^*; \boldsymbol{\xi})\}}{S(X_1^*, X_2|W^*)}\Big\},
\end{aligned}
$$

where

$$
\begin{aligned}
& S(t_1, t_2|w) = Pr(X_1 \geq t_1 \geq U_1, X_2 \geq t_2 \geq U_2 | X_1 \geq U_1, X_2 \geq U_2, W = w) \\
= \;& Pr(X_1 \geq t_1 \geq U_1, X_2 \geq t_2 \geq U_2 | W = w)/\alpha_w \\
= \;& Pr(X_1 \geq t_1, X_2 \geq t_2 | W = w) Pr(t_1 \geq U_1, t_2 \geq U_2 | W = w)/\alpha_w \\
= \;& F(t_1, t_2|w) G(t_1, t_2|w) H(t_1, t_2|w)/\alpha_w \qquad (3.6)
\end{aligned}
$$

We will first show that $\boldsymbol{U}_n^{*(k)}(\boldsymbol{\xi})$ converges uniformly to $\boldsymbol{u}^{*(k)}$, $k = 1, \ldots, 4$, then show that $\boldsymbol{u}^*(\boldsymbol{\xi}) = \mathbf{0}$ has the unique solution at $\boldsymbol{\xi}_0$, and finally show the consistency of $\hat{\boldsymbol{\xi}}_n$ that is the solution of $\boldsymbol{U}_n^*(\boldsymbol{\xi}) = \mathbf{0}$.

Following van der Vaart and Wellner (1996), we use $\mathbb{P}_n$ and $\mathbb{Q}_n$ to denote the empirical measures of $n$ independent copies of $V^*$ and $V$ that follow the distributions $P$ and $Q$, respectively. Although these two samples are in fact identical, i.e., $\mathbb{P}_n \equiv \mathbb{Q}_n$ and $P \equiv Q$, we use different letters to keep the notation tractable for the double summations, which will soon become clear in the following calculations.

For model (1.7), $\boldsymbol{U}_n^{*(1)}(\boldsymbol{\xi}) = \mathbb{Q}_n \Delta_1 \Delta_2 \dot{\beta}(X_1, X_2, W)$ that is free of $\boldsymbol{\xi}$, and $\Delta_1$, $\Delta_2$ and $\dot{\beta}(X_1, X_2, W)$ are all bounded ($X_1$, $X_2$ and $W$ are bounded by Condition C3.1 and C3.2), hence by the law of large numbers, we have

$$\sup |\boldsymbol{U}_n^{*(1)}(\boldsymbol{\xi}) - \boldsymbol{u}^{*(1)}(\boldsymbol{\xi})| = |(\mathbb{Q}_n - Q)\Delta_1 \Delta_2 \dot{\beta}(X_1, X_2, W)| \to \mathbf{0}$$

either almost surely or in probability. Convergence in probability should be adequate here for the proof.

To show the uniform convergence of $\boldsymbol{U}_n^{*(2)}(\boldsymbol{\xi})$, we first define the following quantities:

$$\begin{aligned} & g^{(n)}(\Delta_2, U_1, U_2, X_1, X_2, \Delta_1^*, U_1^*, U_2^*, X_1^*, X_2^*; \boldsymbol{\xi}) \\ & \quad = \Delta_1^* \Delta_2 \dot{\beta}(X_1^*, X_2) \frac{I(X_1 \geq X_1^* \geq U_1) I(U_2^* \leq X_2 \leq X_2^*) e^{\beta(X_1^*, X_2; \xi)}}{\frac{1}{n}(N(X_1^*, X_2) - I(U_2^* \leq X_2 \leq X_2^*)(1 - e^{\beta(X_1^*, X_2; \xi)}))}, \end{aligned}$$

and

$$\begin{aligned} & \tilde{g}(\Delta_2, U_1, U_2, X_1, X_2, \Delta_1^*, U_1^*, U_2^*, X_1^*, X_2^*; \boldsymbol{\xi}) \\ & \quad = \Delta_1^* \Delta_2 \dot{\beta}(X_1^*, X_2) \frac{I(X_1 \geq X_1^* \geq U_1) I(U_2^* \leq X_2 \leq X_2^*) e^{\beta(X_1^*, X_2; \xi)}}{S(X_1^*, X_2)}. \end{aligned}$$

The only difference between the two expressions is in the denominators of the two fractions. By fixing $(\Delta_1^*, U_1, U_2, X_1^*, X_2^*)$ at $(\delta_1, u_1, u_2, x_1, x_2)$, we also define

$$\begin{aligned} h_{\mathbb{Q}_n}^{(n)}(\delta_1, u_1, u_2, x_1, x_2; \boldsymbol{\xi}) &= \mathbb{Q}_n g^{(n)}(\Delta_2, U_1, U_2, X_1, X_2, \delta_1, u_1, u_2, x_1, x_2; \boldsymbol{\xi}), \\ \tilde{h}_{\mathbb{Q}_n}(\delta_1, u_1, u_2, x_1, x_2; \boldsymbol{\xi}) &= \mathbb{Q}_n \tilde{g}(\Delta_2, U_1, U_2, X_1, X_2, \delta_1, u_1, u_2, x_1, x_2; \boldsymbol{\xi}), \\ \tilde{h}_Q(\delta_1, u_1, u_2, x_1, x_2; \boldsymbol{\xi}) &= Q \tilde{g}(\Delta_2, U_1, U_2, X_1, X_2, \delta_1, u_1, u_2, x_1, x_2; \boldsymbol{\xi}). \end{aligned}$$

Similarly, fixing $(\Delta_2, U_1, U_2, X_1, X_2)$ at $(\delta_2, u_1, u_2, x_1, x_2)$, define

$$\begin{aligned} h_{\mathbb{P}_n}^{(n)}(\delta_2, u_1, u_2, x_1, x_2; \boldsymbol{\xi}) &= \mathbb{P}_n g^{(n)}(\delta_2, u_1, u_2, x_1, x_2, \Delta_1^*, U_1^*, U_2^*, X_1^*, X_2^*; \boldsymbol{\xi}), \\ h_P^{(n)}(\delta_2, u_1, u_2, x_1, x_2; \boldsymbol{\xi}) &= P g^{(n)}(\delta_2, u_1, u_2, x_1, x_2, \Delta_1^*, U_1^*, U_2^*, X_1^*, X_2^*; \boldsymbol{\xi}), \\ \tilde{h}_P(\delta_2, u_1, u_2, x_1, x_2; \boldsymbol{\xi}) &= P \tilde{g}(\delta_2, u_1, u_2, x_1, x_2, \Delta_1^*, U_1^*, U_2^*, X_1^*, X_2^*; \boldsymbol{\xi}). \end{aligned}$$

Then we have $\boldsymbol{U}_n^{*(2)}(\boldsymbol{\xi}) = \mathbb{P}_n h_{\mathbb{Q}_n}^{(n)}$ and $\boldsymbol{u}^{*(2)}(\boldsymbol{\xi}) = P\tilde{h}_Q$. It is clear that under Conditions C3.1 and C3.2 we can easily interchange summations and integrations. We thus have

$$\begin{aligned}
&\sup |\boldsymbol{U}_n^{*(2)}(\boldsymbol{\xi}) - \boldsymbol{u}^{*(2)}(\boldsymbol{\xi})| \\
&\quad = \sup |\mathbb{P}_n h_{\mathbb{Q}_n}^{(n)} - P\tilde{h}_Q| \\
&\quad = \sup |\mathbb{P}_n h_{\mathbb{Q}_n}^{(n)} - P h_{\mathbb{Q}_n}^{(n)} + P h_{\mathbb{Q}_n}^{(n)} - P\tilde{h}_Q| \\
&\quad \leq \sup |\mathbb{P}_n h_{\mathbb{Q}_n}^{(n)} - P h_{\mathbb{Q}_n}^{(n)}| + \sup |P\mathbb{Q}_n g^{(n)} - PQ\tilde{g}| \\
&\quad = \sup |\mathbb{P}_n h_{\mathbb{Q}_n}^{(n)} - P h_{\mathbb{Q}_n}^{(n)}| + \sup |\mathbb{Q}_n P g^{(n)} - QPg^{(n)} + QPg^{(n)} - QP\tilde{g}| \\
&\quad \leq \sup |\mathbb{P}_n h_{\mathbb{Q}_n}^{(n)} - P h_{\mathbb{Q}_n}^{(n)}| + \sup |\mathbb{Q}_n h_p^{(n)} - Q h_p^{(n)}| + \sup |QPg^{(n)} - QP\tilde{g}| \\
&\quad \leq \sup |\mathbb{P}_n h_{\mathbb{Q}_n}^{(n)} - P h_{\mathbb{Q}_n}^{(n)}| + \sup |\mathbb{Q}_n h_p^{(n)} - Q h_p^{(n)}| + \sup |g^{(n)} - \tilde{g}|,
\end{aligned}$$

which converges to $0$ in probability following a similar proof in Chapter I. Then we have $\boldsymbol{U}_n^{*(2)}(\boldsymbol{\gamma})$ converges uniformly to $\boldsymbol{u}^{*(2)}(\boldsymbol{\gamma})$ in probability. By the same calculation we can also show that $\boldsymbol{U}_n^{*(4)}(\boldsymbol{\gamma})$ converges uniformly to $\boldsymbol{u}^{*(4)}(\boldsymbol{\gamma})$. Thus we have shown that $\boldsymbol{U}_n^{*}(\boldsymbol{\gamma})$ converges uniformly to $\boldsymbol{u}^{*}(\boldsymbol{\gamma})$ in probability.

To show $\boldsymbol{u}^*(\boldsymbol{\xi}_0) = 0$, it is sufficient to show $\boldsymbol{u}^{*(1)}(\boldsymbol{\xi}_0) = \boldsymbol{u}^{*(2)}(\boldsymbol{\xi}_0)$. We now calculate $\boldsymbol{u}^{*(2)}(\boldsymbol{\xi}_0)$ directly. Recall that $(U_1, U_2, X_1, X_2, \Delta_1, \Delta_2)$ and $(U_1^*, U_2^*, X_1^*, X_2^*, \Delta_1^*, \Delta_2^*)$ are independent and identically distributed with a density function given in (3.5). Thus

$$\begin{aligned}
&\boldsymbol{u}^{*(2)}(\boldsymbol{\xi}_0)\\
=\ & PQ\tilde{g}(\Delta_2, U_1, U_2, X_1, X_2, \Delta_1^*, W, U_1^*, U_2^*, X_1^*, X_2^*, W^*; \boldsymbol{\xi}_0)\\
=\ & P\Big\{\sum_w \int_0^\infty\!\!\int_0^\infty \sum_{\delta_1=0}^1\sum_{\delta_2=0}^1 \int_0^\infty\!\!\int_0^\infty \tilde{g}(\delta_2, u_1, u_2, t_1, t_2, w, \Delta_1^*, U_1^*, U_2^*, X_1^*, X_2^*, W^*; \boldsymbol{\xi}_0)\\
&\qquad p(u_1, u_2, t_1, t_2, \delta_1, \delta_2, w)du_1du_2dt_1dt_2\Big\}\\
=\ & P\Delta_1^*\Big\{\int_{U_2^*}^{X_2^*} \frac{\dot{\beta}(X_1^*, t_2, W^*)e^{\beta(X_1^*, t_2, W^*;\xi_0)}}{S(X_1^*, t_2, W^*)\alpha} f(W^*)\\
&\qquad \left[\int_{X_1^*}^\infty d_1\{H(t_1, t_2|W^*)\partial_2 F(t_1, t_2|W^*)G(t_1, t_2|W^*)\}\right] dt_2\Big\}\\
=\ & P\Delta_1^*\Big\{-\int_{U_2^*}^{X_2^*} \frac{\dot{\beta}(X_1^*, t_2, W^*)e^{\beta(X_1^*, t_2, W^*;\xi_0)}}{S(X_1^*, t_2|W^*)\alpha} f(W^*)\\
&\qquad H(X_1^*, t_2|W^*)\partial_2 F(X_1^*, t_2|W^*)G(X_1^*, t_2|W^*)dt_2\Big\}.
\end{aligned}$$

Here we use $d_k$ to denote the infinitesimal change with respect to $t_k$, $k = 1, 2$. From definition (1.1) we can obtain that

$$\theta(t_1, t_2, w; \boldsymbol{\xi}_0) = e^{\beta(t_1, t_2, w;\xi_0)} = \frac{\partial_{1,2}F(t_1, t_2|w)F(t_1, t_2|w)}{\partial_1 F(t_1, t_2|w)\partial_2 F(t_1, t_2|w)}.$$

Together with (3.6) and integration by parts, we have

$$
\begin{aligned}
& \boldsymbol{u}^{*(2)}(\boldsymbol{\xi}_0) \\
&= P\Delta_1^* \left\{ -\int_{U_2^*}^{X_2^*} \dot{\beta}(X_1^*, t_2, W^*) f(W^*) \frac{\partial_{1,2} F(X_1^*, t_2|W^*)}{\partial_1 F(X_1^*, t_2|W^*)} dt_2 \right\} \\
&= \sum_{w^*} \int_0^\infty \int_0^\infty \sum_{\delta_1=0}^{1} \sum_{\delta_2=0}^{1} \delta_1 \left\{ -\int_{u_2^*}^{s_2} \dot{\beta}(s_1, t_2, w^*) f(w^*) \frac{\partial_{1,2} F(s_1, t_2|w^*)}{\partial_1 F(s_1, t_2|w^*)} dt_2 \right\} \\
&\qquad p(u_1^*, u_2^*, s_1, s_2, \delta_1, \delta_2, w^*) du_1^* du_2^* ds_2 ds_1 \\
&= \sum_{w^*} \int_0^\infty \int_0^\infty \int_{u_2^*}^\infty \int_{u_1^*}^\infty \left\{ -\int_{u_2^*}^{s_2} \dot{\beta}(s_1, t_2, w^*) f(w^*) \frac{\partial_{1,2} F(s_1, t_2|w^*)}{\partial_1 F(s_1, t_2|w^*)} dt_2 \right\} \\
&\qquad h(u_1^*, u_2^*|w^*) d_2\{\partial_1 F(s_1, s_2|w^*) G(s_1, s_2|w^*)/\alpha\} ds_1 du_1^* du_2^* \\
&= \sum_{w^*} \int_0^\infty \int_0^\infty \int_{u_2}^\infty \int_{u_1}^\infty \dot{\beta}(s_1, s_2, w^*) h(u_1, u_2|w^*) f(w^*) \\
&\qquad \partial_{1,2} F(s_1, s_2|w^*) G(s_1, s_2|w^*)/\alpha_w ds_1 ds_2 du_1 du_2 \\
&= E\{\Delta_1 \Delta_2 \dot{\beta}(X_1, X_2, W) | (X_1 \geq U_1, X_2 \geq U_2)\} \\
&= \boldsymbol{u}^{*(1)}(\boldsymbol{\xi}_0).
\end{aligned}
$$

Note that $\boldsymbol{u}^{*(1)}(\boldsymbol{\xi})$ is in fact free of $\boldsymbol{\xi}$ for model (1.7). We thus have shown $\boldsymbol{u}^*(\boldsymbol{\xi}_0) = 0$.

To show $\boldsymbol{\xi}_0$ is the unique solution of $u(\boldsymbol{\xi}) = 0$, it suffices to show that (a) the matrix $\dot{\boldsymbol{u}}^*(\boldsymbol{\xi}) \equiv d\boldsymbol{u}^*(\boldsymbol{\xi})/d\boldsymbol{\xi}$ is negative semidefinite for all $\boldsymbol{\xi} \in \Gamma$, and (b) $\dot{\boldsymbol{u}}^*(\boldsymbol{\xi})$ negative definite at $\boldsymbol{\xi}_0$. Both (a) and (b) are satisfied following a similar argument in the proof of Chapter I.

We are now ready to show the consistency of $\hat{\boldsymbol{\xi}}_n$. Given the fact that $\boldsymbol{U}_n^*(\hat{\boldsymbol{\xi}}_n) = 0$ and $\sup |\boldsymbol{U}_n^*(\boldsymbol{\xi}) - \boldsymbol{u}^*(\boldsymbol{\xi})| = o_p(1)$, we have

$$
|\boldsymbol{u}^*(\hat{\boldsymbol{\xi}}_n)| = |\boldsymbol{U}_n^*(\hat{\boldsymbol{\xi}}_n) - \boldsymbol{u}^*(\hat{\boldsymbol{\xi}}_n)| \leq \sup |\boldsymbol{U}_n^*(\boldsymbol{\xi}) - \boldsymbol{u}^*(\boldsymbol{\xi})| = o_p(1).
$$

Since $\boldsymbol{\xi}_0$ is the unique solution to $\boldsymbol{u}^*(\boldsymbol{\xi}) = 0$, for any fixed $\boldsymbol{\epsilon} > 0$, there exists a $\boldsymbol{\delta} > 0$ such that

$$
Pr\left(|\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0| > \boldsymbol{\epsilon}\right) \leq Pr\left(|\boldsymbol{u}^*(\hat{\boldsymbol{\xi}}_n)| > \boldsymbol{\delta}\right).
$$

The consistency of $\hat{\boldsymbol{\xi}}_n$ follows immediately.

### 3.7.2 Proof of Theorem III.2

Define $\dot{\boldsymbol{U}}_n^*(\boldsymbol{\xi}) \equiv d\boldsymbol{U}_n^*(\boldsymbol{\xi})/d\boldsymbol{\xi}$. By the Taylor expansion of $\boldsymbol{U}_n^*(\hat{\boldsymbol{\xi}}_n)$ around $\boldsymbol{\xi}_0$, we have

$$n^{1/2}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0) = -\left\{\dot{\boldsymbol{U}}_n^*(\boldsymbol{\xi}^*)\right\}^{-1} n^{1/2}\boldsymbol{U}_n^*(\boldsymbol{\xi}_0), \tag{3.7}$$

where $\boldsymbol{\xi}^*$ lies between $\hat{\boldsymbol{\xi}}_n$ and $\boldsymbol{\xi}_0$. By a similar calculation as in Appendix A showing the uniform consistency of $\boldsymbol{U}_n^*(\boldsymbol{\xi})$, we can show that $\sup|\dot{\boldsymbol{U}}_n^*(\boldsymbol{\xi}) - \dot{\boldsymbol{u}}^*(\boldsymbol{\xi})| = o_p(1)$. Thus by the consistency of $\hat{\boldsymbol{\xi}}_n$, which implies the consistency of $\boldsymbol{\xi}^*$, and the continuity of $\dot{\boldsymbol{u}}^*(\boldsymbol{\xi})$, we obtain $\dot{\boldsymbol{U}}_n^*(\boldsymbol{\xi}^*) = \dot{\boldsymbol{u}}^*(\boldsymbol{\xi}^*) + o_p(1) = \dot{\boldsymbol{u}}^*(\boldsymbol{\xi}_0) + o_p(1)$, where $\dot{\boldsymbol{u}}^*(\boldsymbol{\xi}_0) = -2E\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2)^{\otimes 2}\} = -\boldsymbol{I}(\boldsymbol{\xi}_0)$ is invertible by Condition C3.5. Hence based on the fact that continuity holds for the inverse operator, (3.7) can be written as

$$n^{1/2}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0) = \{\boldsymbol{I}(\boldsymbol{\xi}_0)^{-1} + o_p(1)\}n^{1/2}\boldsymbol{U}_n^*(\boldsymbol{\xi}_0). \tag{3.8}$$

We now need to find the asymptotic representation of $n^{1/2}\boldsymbol{U}_n^*(\boldsymbol{\xi}_0)$. We only check it for $\boldsymbol{U}_n^{*(1)}(\boldsymbol{\xi}_0) - \boldsymbol{U}_n^{*(2)}(\boldsymbol{\xi}_0)$. The calculation for $\boldsymbol{U}_n^{*(3)}(\boldsymbol{\xi}_0) - \boldsymbol{U}_n^{*(4)}(\boldsymbol{\xi}_0)$ is virtually identical and yields the same asymptotic representation.

It is easily seen that

$$n^{1/2}\left(\boldsymbol{U}_n^{*(1)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{*(1)}(\boldsymbol{\xi}_0)\right) = \mathbb{G}_n\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2)\}, \tag{3.9}$$

where $\mathbb{G}_n = n^{1/2}(\mathbb{P}_n - P)$.

It can also be shown that

$$\begin{aligned}
& n^{1/2}\left(\boldsymbol{U}_n^{*(2)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{*(2)}(\boldsymbol{\xi}_0)\right) \\
= \; & \mathbb{G}_n\Bigg\{\tilde{h}_Q(\Delta_1, U_1, U_2, X_1, X_2, W; \boldsymbol{\xi}_0) + \tilde{h}_P(\Delta_2, U_1, U_2, X_1, X_2, W; \boldsymbol{\xi}_0) \\
& - \iint I(X_1 \geq x_1^* \geq U_1, X_2 \geq x_2 \geq U_2, W = w^*) \\
& \quad f(\delta_1, u_1, u_2, x_1, x_2, w, \delta_2^*, u_1^*, u_2^*, x_1^*, x_2^*, w^*) \\
& \quad dP(\delta_1^*, \delta_2^*, u_1^*, u_2^*, x_1^*, x_2^*, w^*)dQ(\delta_1, \delta_2, u_1, u_2, x_1, x_2, w)\Bigg\} + o_p(1) \\
\to_d \; & N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_0)),
\end{aligned}$$

where
$$f(\delta_1, u_1, u_2, x_1, x_2, w, \delta_2^*, u_1^*, u_2^*, x_1^*, x_2^*, w^*)$$
$$= \frac{\delta_1^*\delta_2\dot{\beta}(x_1^*, x_2, w^*)I(w = w^*)I(x_1 \geq x_1^* \geq u_1)I(u_2^* \leq x_2 \leq x_2^*)e^{\beta(x_1^*, x_2, w^*;\xi_0)}}{S(x_1^*, x_2, w^*)^2}.$$

Then we obtain the asymptotic linear representation of $n^{1/2}\boldsymbol{U}_n^*(\boldsymbol{\xi}_0)$:

$$\begin{aligned}
& n^{1/2}\boldsymbol{U}_n^*(\boldsymbol{\xi}_0) \\
= \; & n^{1/2}\{\boldsymbol{U}_n^*(\boldsymbol{\xi}_0) - \boldsymbol{u}^*(\boldsymbol{\xi}_0)\} \\
= \; & n^{1/2}\{\boldsymbol{U}_n^{*(1)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{*(1)}(\boldsymbol{\xi}_0)\} - n^{1/2}\{\boldsymbol{U}_n^{*(2)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{*(2)}(\boldsymbol{\xi}_0)\} \\
& \quad + n^{1/2}\{\boldsymbol{U}_n^{*(3)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{*(3)}(\boldsymbol{\xi}_0)\} - n^{1/2}\{\boldsymbol{U}_n^{*(4)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{*(4)}(\boldsymbol{\xi}_0)\} \\
= \; & 2\mathbb{G}_n\Bigg\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2, W) \\
& -\tilde{h}_Q(\Delta_1, U_1, U_2, X_1, X_2, W; \boldsymbol{\xi}_0) - \tilde{h}_P(\Delta_2, U_1, U_2, X_1, X_2, W; \boldsymbol{\xi}_0) \\
& \quad + \iint I(X_1 \geq x_1^* \geq U_1, X_2 \geq x_2 \geq U_2, W = w^*) \\
& \quad\quad f(\delta_1, u_1, u_2, x_1, x_2, w, \delta_2^*, u_1^*, u_2^*, x_1^*, x_2^*, w^*) \\
& \quad\quad\quad dP(\delta_1^*, \delta_2^*, u_1^*, u_2^*, x_1^*, x_2^*, w^*)dQ(\delta_1, \delta_2, u_1, u_2, x_1, x_2, w)\Bigg\} + o_p(1) \\
\to_d \; & N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_0)).
\end{aligned}$$

Thus from (3.8) we obtain the desired asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0)$.

# References


Bandeen-Rache, K. and Ning, J. (2008). Nonparametric Estimation of Bivariate Failure Time Association in the Presence of a Competing Risk. *Biometrika* **95**, 221–232.

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.

Cox, D. R. (1972). Regression Models and Life-Tables (with discussion). *J. R. Statist. Soc.* **B 34**, 187–220.

Duffy, D. L., Martin, N. G. and Matthews, J. D. (1990). Appendectomy in Australian Twins. *Am. J. Hum. Genet.* **47**, 590–592.

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

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* London: Chapman and Hall.

Fan, J., Hsu, L., and Prentice, R. L. (2000a). Dependence Estimation over a Finite Bivariate Failure Time Region. *Lifetime Data Anal.* **6**, 343–355.

Fan, J. and Prentice, R. L. (2002). Covariate-Adjusted Dependence Estimation on a Finite Bivariate Failure Time Region. *Statistica Sinica* **12**, 689–705.

Fan, J., Prentice, R. L., and Hsu, L. (2000b). A Class of Weighted Dependence Measures for Bivariate Failure Time Data. *J. R. Statist. Soc.* **B 62**, 181–190.

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

Glidden, D. V. (2000). A Two-Stage Estimator of the Dependence Parameter for the Clayton-Oakes Model. *Lifetime Data Anal.* **6**, 141–156.

Glidden, D. V. and Self, S. G. (1999). Semiparametric Likelihood Estimation in the Clayton–Oakes Failure Time Model. *Scand. J. Statist.* **26**, 363–372.

Gürler, Ü. (1996). Bivariate Estimation With Right Truncated Data. *J. Am. Statist. Assoc.* **91**, 1152–1165.

Gürler, Ü. (1997). Bivariate Distribution and Hazard Functions When a Component Is Randomly Truncated. *J. Multivariate Anal.* **60**, 20–47.

Härdle, P., Janssen P. and Serfling R. (1988), Strong Uniform Consistency Rates for Estimators of Conditional Functionals, *The Annals of Statistics*, **16**, 1428–1449.

Hu, T., Nan, B., Lin, X. and Robins, J. (2011). Time-Dependent Cross-Ratio Estimation for Bivariate Failure Times, *Biometrika* **98(2)**, 341–354.

Hsu, L. and Prentice, R. L. (1996). On Assessing the Strength of Dependency between Failure Time Variates. *Biometrika* **83**, 491–506.

Ichimura, H. (1993). Semiparamatric Least Square (SLS) and Weighted SLS Estimation of Single-Index Models. *Journal of Econometrics* **58**, 71–120.

Ino, Y., Betensky, R. A., Zlatescu, M. C., Sasaki, H., Macdonald, D. R., Stemmer-Rachamimov, A. O., Ramsay, D. A., Cairncross, J. G., and Louis, D. N. (2001). Molecular Subtypes of Anaplastic Oligodendroglioma: Implications for Patient Management at Diagnosis. *Clinical Cancer Research*

**7**, 839–845.

Kalbfleisch, J. D., and Lawless, J. F. (1989). Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS. *J. Am. Statist. Assoc.* **84**, 360–372.

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

Lagakos, S. W., Barraj, L. M., and De Gruttola, V. (1988), Nonparametric Analysis of Truncated Survival Data With Applications to AIDS. *Biometrika* **75**, 515–523.

Li, Y. and Lin, X. (2006). Semiparametric Normal Transformation Models for Spatially Correlated Survival Data. *J. Am. Statist. Assoc.* **101**, 591–603.

Li, Y., Prentice, R. L., and Lin, X. (2008). Semiparametric Maximum Likelihood Estimation in Normal Transformation Models for Bivariate Survival Data. *Biometrika* **95**, 947–960.

Lisabeth, L. D., Harlow, S. D., Gillespie, B., Lin, X., and Sowers, M. F. (2004). Staging Reproductive Aging: A Comparison of Proposed Bleeding Criteria for the Menopausal Transition. *Menopause* **11**, 186–197.

Martin, E.C. and Betensky R.A. (2005). Testing Quasi-Independence of Failure and Truncation Times via Conditional Kendall's Tau. *J. Am. Statist. Assoc.* **100**, 484–492.

Nan, B., Lin, X., Lisabeth, L. D., and Harlow, S. (2006). Piecewise Constant Cross-Ratio Estimation for Association of Age at a Marker Event and Age at Menopause. *J. Am. Statist. Assoc.* **101**, 65–77.

Nolan, D. and Pollard D. (1987). U-Process: Rates of Convergence. *Annals of Statistics* **15(2)**, 780–799.

Oakes, D. (1982). A Model for Association in Bivariate Survival Data. *J. R. Statist. Soc.* **B 44**, 414–422.

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

——(1989) Bivariate Survival Models Induced by Frailties. *J. Am. Statist. Assoc.* **84**, 487–493.

Prentice, R. L. and Hsu, L. (1997). Regression on Hazard Ratios and Cross Ratios in Multivariate Failure Time Analysis. *Biometrika* **84**, 349–363.

Quale, C.M. and van der Laan, M.J. (2000). Inference with Bivariate Truncated Data. *Lifetime Data Analysis* **6**, 391–408.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: John Wiley and Sons.

Shen, X. (1997). On Methods of Sieves and Penalization. *The Annals of Statistics* **25**, 2555–2591.

Shen, X. and Wong, W. H. (1994). Convergence Rate of Sieve Estimates. *The Annals of Statistics,* **22**, 580–615.

Shih, J. and Louis, T. A. (1995). Inference on the Association Parameter in Copula Models for Bivariate Survival Data. *Biometrics* **51**, 1384–1399.

Treloar, A. E., Boynton, R. E., Behn, B. G., and Brown, B. W. (1967). Variation of the Human Menstrual Cycle Through Reproductive Life. *International Journal of Fertility* **12**, 77–126.

Van der Laan, MJ. (1996). Nonparametric Estimation of the Bivariate Survival Function with Truncated Data. *Journal of Multivariate Analysis* **58**, 107–131.

Van der Vaart, A. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.