Survival Analysis for Censored Data under Referral Bias

Huan Wang

School of Computing, Engineering and Mathematics

University of Brighton

A thesis submitted for the degree of

Doctor of Philosophy

March, 2014

Declaration

I declare that the research contained in this thesis, unless otherwise formally indicated within the text, is the original work of the author. The thesis has not been previously submitted to this or any other university for a degree, and does not incorporate any material already submitted for a degree.

Signed

Date

I dedicate this thesis to

my families,

for their constant support and unconditional love.

I love you all dearly.

Acknowledgements

The completion of this Ph.D. thesis has been a long journey. It has been a truly valuable and life-changing experience for me. I would like to express my thanks to the people who have been very helpful to me during the past four years.

First and foremost, I would like to say a very big thank you to my Ph.D. supervisors, Dr. Hongsheng Dai, Dr. Roma Chakrabarti and Liz Cheek, for all the support you gave me, during both the time I spent undertaking my field work and also the long months I spent writing up my thesis. I have been very fortunate working with you all. Dear Hongsheng, my daily supervisor, thank you for your patient guidance during my entrance to this research field and all the things you have done for me. Without your help, I would never have published our paper and finished my thesis. Dear Liz and Roma, thank you very much for your guidance and support. The discussion with you made me think deeply and added vigorous components to this thesis. Without your constant feedback and correction of the manuscripts, my thesis would not have been achieved.

I also would like to express my sincerest appreciation to my colleagues and other staffs in our school. Thank you for your help and encouragement during the past four years.

A special thank to my girlfriend, Cong Zhang, who has been by my side during the past two years. Without you, I would not have had courage to persevere on this journey.

Finally, a warmest thank to my parents. Mom and dad, you are the ones who always support me, encourage me and love me. You helped me to overcome all the difficulties and made everything possible.

Thanks for life, giving me such valuable experience.

Brighton

March, 2014

Abstract

This work arises from a hepatitis C cohort study and focuses on estimating the effects of covariates on progression to cirrhosis. In hepatitis C cohort studies, patients may be recruited to the cohort with referral bias because clinically the patients with more rapid disease progression are preferentially referred to liver clinics. This referral bias can lead to significantly biased estimates of the effects of covariates on progression to cirrhosis.

A pair of correlated event times is observed for each patient, the time from infection to referral to the cohort (referral time) and the time from infection to the development of cirrhosis (cirrhosis time). Here the cirrhosis time may be subject to right censoring due to the last follow-up time. To take the referral bias into account, we assume that the referral time is right truncated. Then, based on a polar coordinate transformation, we propose a new non-parametric estimator for the bivariate survival function of truncation time and censoring time.

To handle such censored survival data with referral bias and to estimate the effects of covariates, we first consider an accelerated failure time model. We modify the estimating equation by including the estimated bivariate survival function of truncation time and censoring time. By applying the weighted least squares method to the modified estimating equation, unbiased estimates of the effects of covariates on the cirrhosis time are obtained. The large sample properties of the proposed coefficient estimator are also demonstrated. Simulation studies show that the proposed estimator and its covariance function estimator perform well. Through applying the proposed method to analyse the Edinburgh hepatitis C data, we find that our

proposed method can remove the referral bias and give more reliable estimates of the effects of covariates on the cirrhosis time.

The class of semi-parametric linear transformation models is also considered to study the effects of covariates on progression to cirrhosis. The commonly used proportional hazards model (Cox model) can be treated as a special case of this class. Once again we modify the estimating equation by incorporating the bivariate survival function of truncation time and censoring time. Heuristic proofs of the unbiasedness and asymptotic normality of the coefficient estimator are shown. Simulation studies for typical sample sizes show that our proposed methods perform well.

Contents

Chapter 1

Introduction

1.1 Background

Hepatitis C is an infectious liver disease caused by the hepatitis C virus (HCV). Clinically, it has been recognised as a leading cause of cirrhosis, a common cause of hepatocellular carcinoma (HCC) and consequently the most common reason for liver transplantation. In 1979, hepatitis C was first recognized as a separate disease and was categorized as "non-A, non-B hepatitis". Then in 1989, the hepatitis C virus was found and identified to account for the majority of those patients with "non-A, non-B hepatitis". According to the report of hepatitis C proposed by the World Health Organization, there are currently around 170 million people infected by HCV worldwide, comprising around 3% of the global population. About 3 - 4 million people are infected each year, and more than 350,000 people die from hepatitis C-related diseases every year (Fu et al., 2011).

The hepatitis C virus is most commonly transmitted through exposure to infectious blood. In the National Institutes of Health Consensus Statements (2002), it

was reported that infection can occur through intravenous drug use, blood transfusion, solid organ transplantation from infected donors, unsafe medical practices, occupational exposure to infected blood, birth to an infected mother, sex with an infected person, high-risk sexual practices and possibly intranasal cocaine use.

HCV infection has both acute and chronic forms. It has been reported by the World Health Organization that around 75-85% of newly infected people develop chronic disease. The incubation period for hepatitis C is 2 weeks to 6 months. Following initial infection, HCV progression is often clinically silent. Most infected people do not exhibit any symptoms until the development of complications after many years (Fu et al., 2011).

One of the most common complications of chronic HCV infection is liver cirrhosis. Among people who chronically infected, the risk for developing cirrhosis after 20 years varies between studies but has been estimated at around 10-15% for men and 1-5% for women (Freeman et al., 2001). Factors that have been reported to influence this risk include age at infection (older age at infection associated with more rapid progression), alcohol consumption (associated with an increased rate of disease progression), and HIV co-infection (associated with a markedly increased rate of disease progression). In hepatitis C natural history research, the primary focuses at present are estimation of the rate of progression to cirrhosis and assessment of risk factors associated with disease progression.

1.2 Motivation

1.2.1 Data Background

The motivation for this thesis arose from a hepatitis C cohort study in Fu et al. (2007), where the epidemiological interest is to study progression to liver cirrhosis in HCV-infected patients. The Edinburgh hepatitis C dataset studied in Fu et al. (2007) consists of 387 HCV-infected patients recruited to Edinburgh Royal Infirmary's liver clinic from early 1990 to the end of 1999. Data on individuals have been collected at registration including patient demographics, source of infection, calendar year of infection, referral year, risk factors, historical data and treatment history. The mean age at HCV-infection is around 22 years. A total of 63 of the patients (16%) were known to have progressed to liver cirrhosis diagnosed by liver biopsy examinations. For these 63 observed cases, 25% were diagnosed with cirrhosis within 18 years since infection; 50% within 24 years and 75% within 30 years. A total of 29 patients were known to have died and their deaths all occurred after diagnosis of cirrhosis. The demographic details are shown in Table 1.1.

In the Edinburgh hepatitis C dataset, the pair of event times (R, T) is recorded for each patient, where

- R, referral time: time period from HCV infection to referral to the clinic cohort;
- T, cirrhosis time: time period from HCV infection to the development of cirrhosis.

Here the cirrhosis time T may be subject to right censoring at the last diagnosis follow-up time. In survival analysis, censoring is a common form of missing data

Table 1.1: Demographic details for the Edinburgh liver clinic series

Cohort characteristics (name of covariate)	**Number (%)**
Mean age at HCV-infection (Age)	22.4
Standard deviation of age at HCV-infection	9.8
Number (%) known as ever injected drugs (IDU)	
Yes - 1	292 (75%)
No - 0	95 (25%)
Number (%) of patients with known HIV status (HIV)	
Positive - 1	41 (11%)
Negative - 0	346 (89%)
Number (%) of patients (Gender)	
Male - 1	250 (65%)
Female - 0	137 (35%)
Number (%) with excessive alcohol consumption (Alcohol)	
Excessive - 1	116 (30%)
Not Excessive - 0	271 (70%)

problem. It includes several different forms such as right censoring, left censoring, interval censoring and so on. In this hepatitis C cohort study, we only encounter right censoring. Let C denote the censoring time, which is the time period from HCV infection to the last diagnosis follow-up. As shown in Figure 1.1, if cirrhosis of an individual occurs before the last diagnosis follow-up, the cirrhosis time for this subject can be observed. However, if cirrhosis occurs after the last diagnosis follow-up, we cannot observe the exact cirrhosis time for this patient but only know that the cirrhosis time for him/her is at least equal to the censoring time C. Then the observed survival time and the corresponding censoring indicator can be denoted by (X, δ), where $X = \min(T, C)$, and $\delta = I[T \leqslant C]$. We will review some existing techniques which can handle the censored survival data in Chapter 2.

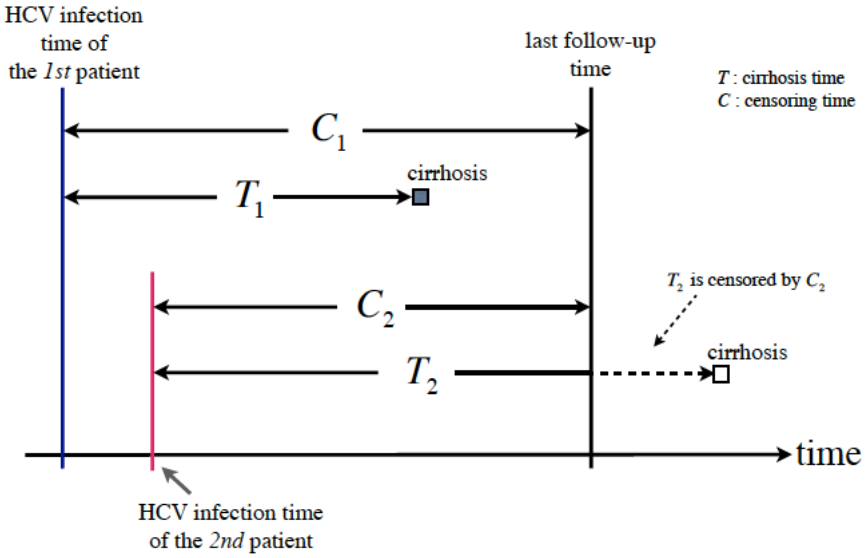

Figure 1.1: Right censoring

1.2.2 Referral bias

In the Edinburgh hepatitis C data set mentioned above, the cirrhosis time T and the referral time R are correlated since the patients with more rapid disease progression are preferentially referred to liver clinics (Dore et al., 2002). This is because the disease progression is often clinically silent after initial infection. Most infected people do not seek medical advice until severe symptoms exhibit. In practice, the closer a patient is to developing cirrhosis, the more likely he/she will go to hospital and subsequently be included into the study. Therefore, the time period from infection to cirrhosis observed in the clinics may be shorter than that for the whole HCV patients community (Wang et al., 2013). If so, the conventional analysis based on liver clinics may overestimate the progression rate among the HCV patients community. This is usually called referral bias in epidemiology.

Such a referral bias can explain why cirrhosis progression rates were estimated

very differently based on liver clinics assembled by different recruitment methods. For example, in a seminal systematic review of hepatitis C natural history studies, Freeman et al. (2001) estimated the rate of progression to cirrhosis among people with chronic hepatitis C and identified factors associated with more rapid disease progression. Their findings from 57 published studies over the period 1990-2000 demonstrated widely contrasting results. The estimated progression rate to cirrhosis after 20 years of HCV infection was 22% (95% CI, 18%-26%) for liver clinics. In contrast, among community-based cohorts, they obtained a much lower estimate of cirrhosis prevalence after 20 years which was 7% (95% CI, 4%-10%). Moreover, older age at HCV infection, male gender, and heavy alcohol intake were identified by Freeman et al. (2001) as risk factors associated with more rapid disease progression. Another example is given by Sweeting et al. (2006). With adjustment for factors including age at HCV infection and gender, Sweeting et al. (2006) estimated probability of progression to cirrhosis after 20 years of infection as 12% (95% CI, 6%-22%) for a hospital-based cohort, as 6% (95% CI, 3%-13%) for post-transfusion cohort (people with "non-A, non-B hepatitis", subsequently diagnosed as hepatitis C), and as 23% (95% CI, 14%-37%) for a cohort recruited from a tertiary referral centre.

A simulation was carried on by Fu et al. (2007) to investigate how event-biased referral to liver clinics can influence the progression rates for differently recruited cohorts. In particular, they compared the estimated progression rates for liver clinics with that for the whole community of HCV-infected patients. The estimated progression rate to cirrhosis after 20 years of infection was around 5% in the community, and about 20% for those who had been selectively referred to a liver clinic. Their studies demonstrated the wide variation in estimated progression rates for differ-

ent clinical cohorts and also indicated that the referral bias could produce severely biased estimates of progression rates.

In summary, all the studies to date (Freeman et al., 2001; Sweeting et al., 2006; Fu et al., 2007) have shown that liver clinics will overestimate the community-wide rate of progression to cirrhosis because the potential referral bias has considerable influence on the estimated progression rates. Therefore, to get an unbiased estimate of progression rate and to identify risk factors associated with hepatitis C progression, it is necessary to take the referral bias into account.

In this thesis, we are interested in the effects of risk factors on the disease progression to cirrhosis, such as age at infection, HIV co-infection, and alcohol intake. To take referral bias into account and to obtain unbiased estimates of the effects of covariates on the cirrhosis time, we consider the following truncation model. Let L denote the time period from HCV infection to the end of recruitment. As shown in Figure 1.2, a patient can be included into the study only if he/she is referred to the clinic before the end of recruitment, i.e. $R \leqslant L$. If a patient is still not referred to the clinic until the end of recruitment, i.e. $R > L$, the referral time R of this patient is right truncated by the corresponding truncation time L. Then no information of this individual is available. By setting up this kind of scheme, the patients with smaller values of R are more likely to be referred to the clinics. Consequently, the patients in the clinics are more likely to have more rapid disease progression because R and T are correlated. Then the referral bias can be taken into account by incorporating this kind of truncation.

Now we can see that, the referral time R may be right truncated by the truncation time L, and the cirrhosis time T may be right censored by the censoring time C. The pairwise event times (R, T) has one component censored and the other component

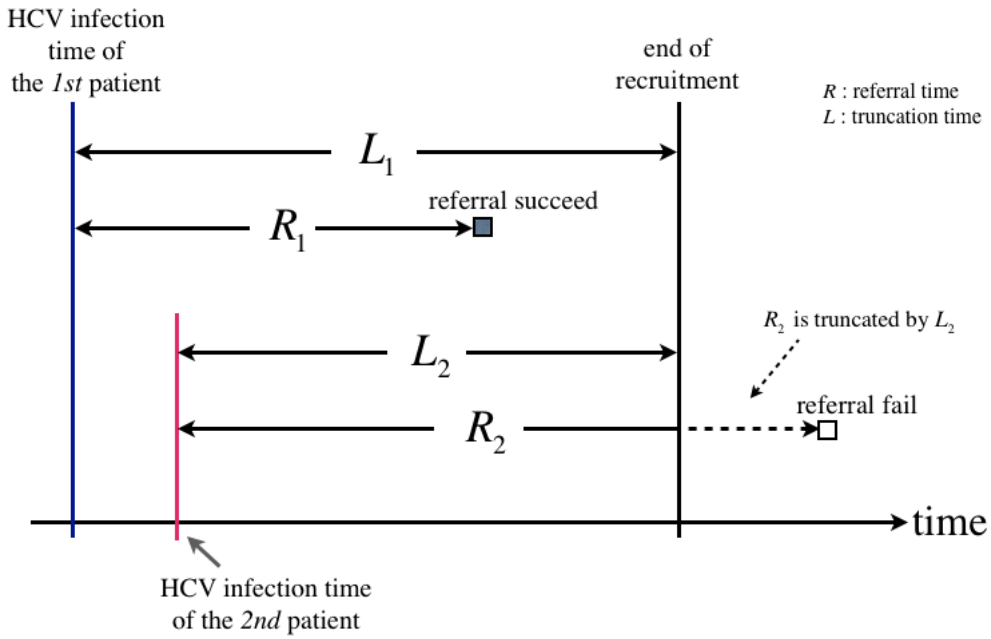

Figure 1.2: Right truncation

truncated. Existing methods in survival analysis are not readily available for such kind of bivariate censoring-truncation data. In this thesis we will develop new methodologies to handle such kind of bivariate survival data.

1.3 Contribution to existing knowledge

The potential referral bias has considerable influence on estimating progression rate to cirrhosis and on identifying risk factors which can affect the time period from HCV infection to cirrhosis (Freeman et al., 2001; Sweeting et al., 2006; Fu et al., 2007). To obtain unbiased estimates of the effects of covariates on the time period from infection to cirrhosis, we involve right truncation to take the referral bias into account. New methods for different regression models (the accelerated failure time model and the linear transformation model) are proposed to handle censored survival

data with referral bias. Our methods can remove the potential referral bias and then produce unbiased estimates of the effects of covariates on cirrhosis time. By using our methods, people can obtain more reliable estimates of progression rate based on data from clinic cohorts, and can also get more reasonable results in identifying risk factors which influence the progression to cirrhosis in chronic hepatitis C virus infection. Moreover, the developed methods can be directly applied to regression analysis of bivariate survival data with random censoring and truncation and can be generally extended to a general class of bivariate regression models with different types of truncation and censoring.

1.4 Thesis structure

The structure of this thesis is as follows. In Chapter 2, we review some existing methods which can handle the univariate censored data for the accelerated failure time (AFT) model and then give a short comparison and discussion for them. A brief summary of existing estimators of bivariate survival functions is also presented in this chapter. In Chapter 3, the AFT model is studied to explore the effects of covariates on cirrhosis time. An estimating procedure for the regression coefficient vector is introduced. We propose a new non-parametric estimator to take the referral bias into account so that we can handle a flexible bivariate distribution of the referral time and the cirrhosis time. The large sample properties of the proposed estimator are provided. Simulation studies and data analysis based on the Edinburgh hepatitis C data (Fu et al., 2007) are also motivated to demonstrate that the proposed estimator performs well. In Chapter 4, instead of the AFT model, a well-known class of linear transformation model is considered. New methodology which can deal with censored

survival data with referral bias is proposed. The proportional hazard model (Cox, 1972) is studied as a special case of the linear transformation model. Simulation studies for typical sample sizes and different censoring percentages are also provided. Chapter 5 gives the conclusion and proposals for future research.

Chapter 2

Literature review of regression estimators for univariate censoring data and estimators for bivariate survival functions

2.1 Introduction to the accelerated failure time model

Let T be the "failure time", the response variable, and $\boldsymbol{W}$ be the corresponding covariate vector. We are interested in making inferences about the effects of $\boldsymbol{W}$ on the response variable T. If there are no censored or truncated observations, we simply regress T or a transformation of it directly on the covariate vector $\boldsymbol{W}$. However, if there are censored or truncated observations in the data, conventional

regression approaches are not readily available to handle these kinds of data. We extensively utilise the proportional hazards model with partial likelihood inference to estimate the covariate effects. This model was proposed by Cox (1972) and has been known as the Cox model. For the Cox model, the hazard function is modelled and the effects of covariates act multiplicatively on the hazard function. In chapter 4, we will discuss how to apply the Cox model to our bivariate survival data.

A useful alternative to the Cox model is the accelerated failure time (AFT) model (Kalbfleisch and Prentice, 2002), which is given by

$$\log T_i = \boldsymbol{W}_i\boldsymbol{\beta} + \varepsilon_i, \tag{2.1}$$

where T_i denotes the failure time for individual i, $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{ip})$ is the p-dimensional covariate vector for the ith observation, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{tr}$ is the p-dimensional regression parameter vector, and ε_i is the corresponding error term. As we can see, the failure time (survival time) T or the transformation of it is modelled directly on the covariate vector $\boldsymbol{W}$ so that the linear regression technique can be used to handle the censored or truncated observations.

When T_i is only subject to random censoring, model (2.1) has been well studied (Miller, 1976; Buckley and James, 1979; Koul et al., 1981; Miller and Harplen, 1982; Stute, 1993, 1996; He and Wong, 2003; Jin et al., 2006). Moreover, comparison studies were given in Heller and Simonoff (1990) and Bao et al. (2007). When the response T_i itself is subject to truncation, model (2.1) can also be solved using the methods in Gross and Huber-Carol (1992); Gross and Lai (1996); He and Yang (2003). We review some key existing methods in the following sections.

2.2 Existing regression approaches for univariate censored survival data

2.2.1 Least squares estimator

For simplicity, let T be the logarithm of the actual survival time. Consider the standard linear regression model

$$T_i = \alpha + \boldsymbol{W}_i\boldsymbol{\beta} + \varepsilon_i, i = 1, \ldots, n, \tag{2.2}$$

where the error terms ε_i are assumed to be independent and identically distributed (i.i.d.) with unspecified distribution F, mean 0 and finite variance. In the absence of censoring, the usual least squares estimates $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ are those which minimize

$$\int u^2 d\hat{F}(u) = \frac{1}{n}\sum_{i=1}^{n}(T_i - \alpha - \boldsymbol{W}_i\boldsymbol{\beta})^2, \tag{2.3}$$

where $\hat{F}$ are the sample distribution estimates based on $\varepsilon_i = T_i - \alpha - \boldsymbol{W}_i\boldsymbol{\beta}$, $d\hat{F}(\varepsilon_i) = 1/n$, and $i = 1, \ldots, n$.

In the presence of censoring, T may be subjected to right censoring by the censoring variable C. Then the observed data for the ith subject will be

$$X_i = \min(T_i, C_i),\ \delta_i = I[T_i \leqslant C_i],\ \boldsymbol{W}_i = (W_{i1}, \ldots, W_{ip}),$$

where I is the indicator function. Given $\boldsymbol{W}_i$, we assume that T_i and C_i are independently distributed according to their respective distribution functions $H(t)$ and $G(t)$. The distribution estimates $\hat{F}$ are now based on the censored and uncensored residuals $\tilde{\varepsilon}_i = X_i - \alpha - \boldsymbol{W}_i\boldsymbol{\beta}$, $i = 1, \ldots, n$, and $d\hat{F}(\tilde{\varepsilon}_i)$ is no longer equal to $1/n$ for

each $\tilde{\varepsilon}_i$. Therefore, to estimate α and $\boldsymbol{\beta}$ by using the least squares technique, it is necessary to estimate the distribution function F for the data which contains both censored and uncensored observations.

Kaplan and Meier (1958) introduced the product-limit estimator of a distribution function which can extend conventional least squares approach to handle censored data. The product-limit estimator, hereafter referred to as the Kaplan-Meier (KM) estimator, of the distribution function F based on $\tilde{\varepsilon}_i = X_i - \alpha - \boldsymbol{W}_i\boldsymbol{\beta}$, $i = 1, \ldots, n$ is given by

$$\hat{F}^{\mathrm{KM}}(u) = 1 - \prod_{i;\tilde{\varepsilon}_{(i)} \leqslant u} \left(\frac{n-i}{n-i+1} \right)^{\delta_i}, \tag{2.4}$$

where $\tilde{\varepsilon}_{(1)} \leqslant \cdots \leqslant \tilde{\varepsilon}_{(n)}$ is the ordered sequence of the residual $\tilde{\varepsilon}$, and the index i refers to the rank for each $\tilde{\varepsilon}_{(i)}$.

Then for the combined sample with both censored and uncensored observations, we just choose α and β to minimize

$$\int u^2 d\hat{F}^{\mathrm{KM}}(u). \tag{2.5}$$

For fixed α and $\boldsymbol{\beta}$, the integral in (2.5) is equivalent to a weighted sum of squares

$$\sum_{i=1}^{n} \omega_i(\alpha, \boldsymbol{\beta})(X_i - \alpha - \boldsymbol{W}_i\boldsymbol{\beta})^2, \tag{2.6}$$

where the weights $\omega_i(\alpha, \boldsymbol{\beta}) = d\hat{F}^{\mathrm{KM}}(\tilde{\varepsilon}_i)$, $i = 1, \ldots, n$. The KM estimator $\hat{F}^{\mathrm{KM}}$ changes only by jumps at the uncensored points, so the size of the jump $d\hat{F}^{\mathrm{KM}} = 0$ at a censored point. Here if the largest $\tilde{\varepsilon}_{(n)}$ is censored, it will be reclassified as an uncensored point, and the corresponding $\hat{F}^{\mathrm{KM}}(\tilde{\varepsilon}_{(n)})$ is defined to be 1 (Miller, 1976). Therefore the summation in (2.6) is actually over all the uncensored observations

including the reclassified largest $\tilde{\varepsilon}_{(n)}$. In the absence of censoring, the KM estimator reduces to the usual sample distribution estimator which assigns weight $1/n$ to each observation.

For fixed $\boldsymbol{\beta}$, any change in the value of α shifts $\tilde{\varepsilon}_i = X_i - \alpha - \boldsymbol{W}_i\boldsymbol{\beta}$ without changing their order. Hence the weights would not be disturbed. Therefore, the minimum with respect to α can be computed directly without any iteration; the minimizing value of α is given by

$$\hat{\alpha}^{\mathrm{KM}} = \sum_{i=1}^{n} \omega_i(0, \boldsymbol{\beta})(X_i - \boldsymbol{W}_i\boldsymbol{\beta}). \tag{2.7}$$

Then $\hat{\boldsymbol{\beta}}^{\mathrm{KM}}$ can be obtained by minimizing the following function

$$f(\boldsymbol{\beta}) = \sum_{i=1}^{n} \omega_i(\boldsymbol{\beta})(X_i - \hat{\alpha}^{\mathrm{KM}} - \boldsymbol{W}_i\boldsymbol{\beta})^2, \tag{2.8}$$

where $\omega_i(\boldsymbol{\beta}) = \omega_i(0, \boldsymbol{\beta})$, and $\hat{\alpha}^{\mathrm{KM}}$ is given in (2.7). When obtaining $\hat{\boldsymbol{\beta}}^{\mathrm{KM}}$, we can substitute it into (2.7) to get the KM estimator $\hat{\alpha}^{\mathrm{KM}}$.

Unfortunately, the least squares estimates $\hat{\alpha}^{\mathrm{KM}}$ and $\hat{\boldsymbol{\beta}}^{\mathrm{KM}}$ from (2.7) and (2.8) do not always converge to the true value of α and $\boldsymbol{\beta}$ as $n \to \infty$. To guarantee the asymptotic consistency of the estimates, it is necessary that the censoring distribution functions $G(t; \boldsymbol{W}_i)$ satisfy

$$G(t; \boldsymbol{W}_i) = G(t - \alpha - \boldsymbol{W}_i\boldsymbol{\beta}), \tag{2.9}$$

which requires that, as $\boldsymbol{W}_i$ changes, the censoring distributions shift along the same line as the distributions of the survival time (Miller, 1976).

2.2.2 Miller estimator

A modification of the Kaplan-Meier least squares approach was introduced by Miller (1976). There are two reasons for his modification. (i) The asymptotic properties of α^{KM} and $\boldsymbol{\beta}^{\mathrm{KM}}$ are difficult to study since the sum of squares in (2.6) is a discontinuous function of $\boldsymbol{\beta}$. This will cause that the minimum of (2.8) can occur at a discontinuity point, so that the iterative procedures based on the power series expansions are not useful. (ii) It is computationally difficult to obtain the Kaplan-Meier least squares estimates with more than one regression variable since a grid search procedure is necessary when the minimum of (2.8) occurs at discontinuities.

Under the condition in (2.9), Miller (1976) suggested using an initial estimate $\hat{\boldsymbol{\beta}}^{(0)}$ to obtain the Kaplan-Meier weights and then minimizing the sum of squares (2.6) weighted by these estimated weights. The initial estimate $\hat{\boldsymbol{\beta}}^{(0)}$ is given by

$$\hat{\boldsymbol{\beta}}^{(0)} = \left\{ \left(\boldsymbol{W}^{\mathrm{uc}} - \bar{\boldsymbol{W}}^{\mathrm{uc}} \right)^{tr} \left(\boldsymbol{W}^{\mathrm{uc}} - \bar{\boldsymbol{W}}^{\mathrm{uc}} \right) \right\}^{-1} \left(\boldsymbol{W}^{\mathrm{uc}} - \bar{\boldsymbol{W}}^{\mathrm{uc}} \right)^{tr} \boldsymbol{T}^{\mathrm{uc}}, \tag{2.10}$$

where $\boldsymbol{W}^{\mathrm{uc}} = ((W_{ij}^{\mathrm{uc}}))$ is the matrix of the $\boldsymbol{W}_i$ associated with the uncensored T_i, $\bar{\boldsymbol{W}}^{\mathrm{uc}}$ has elements $n^{-1} \sum_i W_{ij}^{\mathrm{uc}}$, and $\boldsymbol{T}^{\mathrm{uc}}$ is the row vector of all uncensored T_i. Then this initial estimator $\hat{\boldsymbol{\beta}}^{(0)}$ is based only on the uncensored values.

With this initial estimate $\hat{\boldsymbol{\beta}}^{(0)}$, the weights $\omega_i(\hat{\boldsymbol{\beta}}^{(0)})$ are calculated by applying the Kaplan-Meier procedure to $X_i - \hat{\boldsymbol{\beta}}^{(0)} \boldsymbol{W}_i$, $i = 1, \ldots, n$. Then the modified estimate $\hat{\boldsymbol{\beta}}^{(1)}$ can be obtained by minimizing the weighted sum of squares

$$\sum_{i=1}^{n} \omega_i(\hat{\boldsymbol{\beta}}^{(0)})(X_i - \alpha - \boldsymbol{W}_i \boldsymbol{\beta})^2. \tag{2.11}$$

The convention in the Kaplan-Meier least squares approach is that, if the largest $\tilde{\varepsilon}_{(n)}$

is censored, it will be reclassified as uncensored and the corresponding $\hat{F}^{\mathrm{KM}}(\tilde{\varepsilon}_{(n)})$ is defined to be 1. If so, the mass $1-\hat{F}^{\mathrm{KM}}(\infty)$ is unassigned to any $\tilde{\varepsilon}_i$. However, it has been found that the estimate of $\boldsymbol{\beta}$ was unstable when this procedure was applied on the Stanford heart transplant data. Therefore, an alternative estimating procedure obtained by normalizing the weights so that they sum to be one was recommended by Miller (1976). To be specific, the modified weights are given by

$$\omega_i^*(\hat{\boldsymbol{\beta}}^{(0)}) = \frac{\omega_i(\hat{\boldsymbol{\beta}}^{(0)})}{\sum_{k=1}^{n} \omega_k(\hat{\boldsymbol{\beta}}^{(0)})}, \tag{2.12}$$

which assigns the remaining mass $1-\hat{F}^{\mathrm{KM}}(\infty)$ to the largest $\tilde{\varepsilon}_{(n)}$ (Miller, 1976). Then an iterative sequence to calculate the estimate of the regression coefficient vector $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \Big\{ \big(\boldsymbol{W} - \bar{\boldsymbol{W}}^*\big)^{tr} \boldsymbol{D}(\hat{\boldsymbol{\beta}}^{(k)}) \big(\boldsymbol{W} - \bar{\boldsymbol{W}}^*\big) \Big\}^{-1} \big(\boldsymbol{W} - \bar{\boldsymbol{W}}^*\big)^{tr} \boldsymbol{D}(\hat{\boldsymbol{\beta}}^{(k)}) \boldsymbol{X}, \tag{2.13}$$

where $i = 1, \ldots, n$; $j = 1, \ldots, p$; and

$$\boldsymbol{W} = ((W_{ij}))_{n\times p},\ \bar{\boldsymbol{W}}^* = \Big(\Big(\sum_{i=1}^{n} \omega_i^*(\hat{\boldsymbol{\beta}}^{(k)}) \Big) \Big)_{n\times p},$$

$$\boldsymbol{D}(\hat{\boldsymbol{\beta}}^{(k)}) = \mathrm{diag}\big\{\omega_i^*(\hat{\boldsymbol{\beta}}^{(k)})\big\}_{n\times n},\ \boldsymbol{X} = (X_1, X_2, \ldots, X_n)^{tr}.$$

The limit of the sequence in (2.13) for $k = 0, 1, \ldots,$ is the estimate of $\boldsymbol{\beta}$. Unfortunately, the sequence may become trapped in a loop because the modified weights in (2.12) is a discontinuous function of $\hat{\boldsymbol{\beta}}^{(k)}$. If the values in the loop are not far apart, Miller suggested that an average value over the loop can be used for the estimate $\hat{\boldsymbol{\beta}}$.

2.2.3 Buckley & James estimator

For the estimate $\hat{\boldsymbol{\beta}}$ to be consistent, the assumption given in (2.9) is necessary for the Kaplan-Meier least squares (KMLS) approach and the modified approach proposed by Miller (1976). However, this assumption will rarely be satisfied in practice. Therefore, an estimation method which does not require such restrictions was proposed by Buckley and James (1979). The difference between their method and that of Miller (1976) is that the least squares normal equations rather than the sum of squares of residuals are modified.

In the absence of censoring, the classical least squares estimator is obtained by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}(T_i - \alpha - \boldsymbol{W}_i\boldsymbol{\beta})^2 \tag{2.14}$$

with respect to α and $\boldsymbol{\beta}$. The estimate $\hat{\boldsymbol{\beta}}$ satisfies the following estimating equation

$$\sum_{i=1}^{n}(\boldsymbol{W}_i - \bar{\boldsymbol{W}})(T_i - \boldsymbol{W}_i\boldsymbol{\beta}) = 0, \tag{2.15}$$

where $\bar{\boldsymbol{W}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{W}_i$.

In the presence of censoring, the observed data are $(X_i, \delta_i, \boldsymbol{W}_i)$, which have been defined in Section 2.2.1. The values of the T_i associated with $\delta_i = 0$ are unknown, so that (2.15) cannot be used directly to estimate $\boldsymbol{\beta}$. In order to account for censoring, Buckley and James (1979) modified the classical least-squares procedure above by substituting the observed X_i by a new variable which is given by

$$X_i^* = \delta_i X_i + (1-\delta_i)E\big(T_i | T_i > X_i\big), \tag{2.16}$$

where $\delta_i = I[T_i \leqslant C_i]$ and $i = 1, \ldots, n$. Since $E(T_i|T_i > X_i)$ is unknown, Buckley and James suggested using a self-consistency approach to estimate it from the KM estimator $\hat{F}^{\mathrm{KM}}(\tilde{\varepsilon}_i)$ for $\tilde{\varepsilon}_i = X_i - \boldsymbol{W}_i\boldsymbol{\beta}$. Specifically, the replacing variable X_i^* can be estimated by

$$\hat{X}_i^*(\boldsymbol{\beta}) = \delta_i X_i + (1 - \delta_i)\left[\boldsymbol{W}_i\boldsymbol{\beta} + \frac{\sum_{j:\tilde{\varepsilon}_j > \tilde{\varepsilon}_i} \omega_j^*(\boldsymbol{\beta})\tilde{\varepsilon}_j}{1 - \hat{F}^{\mathrm{KM}}(\tilde{\varepsilon}_i)}\right], \tag{2.17}$$

where $\omega_j^*(\boldsymbol{\beta})$ are the modified weights defined in (2.12) and here $j = 1, \ldots, n$.

Treating all the observations as uncensored, the initial estimator $\hat{\boldsymbol{\beta}}^{(0)}$ is given by

$$\hat{\boldsymbol{\beta}}^{(0)} = \left\{(\boldsymbol{W} - \bar{\boldsymbol{W}})^{tr}(\boldsymbol{W} - \bar{\boldsymbol{W}})\right\}^{-1}(\boldsymbol{W} - \bar{\boldsymbol{W}})^{tr}\boldsymbol{X}, \tag{2.18}$$

where

$$\boldsymbol{W} = \big((W_{ij})\big)_{n\times p},\ \bar{\boldsymbol{W}} = \left(\left(n^{-1}\sum_{i=1}^{n} W_{ij}\right)\right)_{n\times p},\ \boldsymbol{X} = (X_1, X_2, \ldots, X_n)^{tr},$$

$$\text{and } i = 1, \ldots, n,\ j = 1, \ldots, p.$$

Then applying the usual least squares procedure iteratively, the estimator $\hat{\boldsymbol{\beta}}^{(k+1)}$ at the $k+1$ step is

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \left\{(\boldsymbol{W} - \bar{\boldsymbol{W}})^{tr}(\boldsymbol{W} - \bar{\boldsymbol{W}})\right\}^{-1}(\boldsymbol{W} - \bar{\boldsymbol{W}})^{tr}\hat{\boldsymbol{X}}^*(\hat{\boldsymbol{\beta}}^{(k)}), \tag{2.19}$$

where $\hat{\boldsymbol{X}}^*(\hat{\boldsymbol{\beta}}_k) = \big(\hat{X}_1^*(\hat{\boldsymbol{\beta}}^{(k)}), \ldots, \hat{X}_n^*(\hat{\boldsymbol{\beta}}^{(k)})\big)^{tr}$.

The limit of the sequence in (2.19) is the estimate of $\boldsymbol{\beta}$. However, the iterations may also settle down to oscillation between two values as that for Miller's estimator. According to Buckley and James (1979), these values are closer to each other than

for the Miller's estimator. In such cases the average of the two values can be used as the estimate of $\boldsymbol{\beta}$. Given the limiting value $\hat{\boldsymbol{\beta}}$, the corresponding estimate of α is

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{X}_i^*(\hat{\boldsymbol{\beta}}) - \boldsymbol{W}_i\hat{\boldsymbol{\beta}}\right\}. \tag{2.20}$$

The Buckley and James (BJ) estimator appears to overcome the inconsistency problems in Miller's approach, since the assumption (2.9) on the distribution of the censoring variable is not necessary here. As shown in Lai and Ying (1991), the estimating function of the BJ estimator is locally asymptotically linear. Using this result, Jin et al. (2006) showed that if the initial estimator is consistent, then for each fixed m, an m-step estimator must also be consistent. In addition, if the initial estimator is asymptotically normal, then so is the m-step estimator.

2.2.4 Koul, Susarla & Van Ryzin estimator

As we mentioned before, the Miller estimator and the BJ estimator are computed by using iterative procedures. In both cases, as the authors pointed out, the iterations may eventually settle down to oscillation between two values. Koul et al. (1981) proposed a new estimator of the parameter vector $\boldsymbol{\beta}$, hereafter called KSV estimator, which is explicitly defined and easily computable since no iterations are required.

Under the assumption that $G(t;\boldsymbol{W}) \equiv G(t)$, Koul et al. (1981) consider a linear relationship which is

$$E\left\{\frac{\delta_i X_i}{1-G(X_i-)}\middle|\boldsymbol{W}_i\right\} = \alpha + \boldsymbol{W}_i\boldsymbol{\beta}, \tag{2.21}$$

where $G(t-)$ denotes the left continuous version of the distribution function of the

censoring variable C. If $G(t)$ is substituted by an estimator $\hat{G}(t)$, say the KM estimator $\hat{G}^{\mathrm{KM}}(t)$, the synthetic data can be obtained by

$$\hat{X}_i^* = \frac{\delta_i X_i}{1 - \hat{G}^{\mathrm{KM}}(X_i-)}, i = 1, \ldots, n. \tag{2.22}$$

Once $\hat{X}_i^*$ are computed, the standard least squares procedure is processed to get the KSV estimator of $\boldsymbol{\beta}$. Specifically,

$$\hat{\boldsymbol{\beta}} = \left\{(\boldsymbol{W} - \bar{\boldsymbol{W}})^{tr}(\boldsymbol{W} - \bar{\boldsymbol{W}})\right\}^{-1}(\boldsymbol{W} - \bar{\boldsymbol{W}})^{tr}\hat{\boldsymbol{X}}^*, \tag{2.23}$$

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_i^* - \boldsymbol{W}_i\hat{\boldsymbol{\beta}}\right), \tag{2.24}$$

where $\bar{\boldsymbol{W}} = \left(\left(n^{-1}\sum_{i=1}^{n} W_{ij}\right)\right)_{n\times p}$, and $\hat{\boldsymbol{X}}^* = \left(\hat{X}_1^*, \ldots, \hat{X}_n^*\right)^{tr}$.

Comparing with the former methods, the KSV approach is the easiest to be carried out since no iterations are required and the standard least squares procedures can be used once the observations of the response are transformed by the censoring information. Denoting $\omega_i = \delta_i\left(1 - \hat{G}^{\mathrm{KM}}(X_i-)\right)^{-1}$, then we can see that the KSV estimator is based on the responses weighted by ω_i. Therefore when the censoring mechanism depends on covariates, the KSV estimator may lead to false results since only the responses are weighted by the censoring distribution estimates.

2.2.5 Weighted least squares estimator

Stute (1993, 1996) proposed a weighted least squares (WLS) estimator of $\boldsymbol{\beta}$ and proved its consistency and asymptotic normality. He and Wong (2003) obtained the WLS estimator of $\boldsymbol{\beta}$ and the variance estimator. They also proved the asymptotic normality for both estimators. The method in He and Wong (2003) requires sim-

pler conditions for the asymptotic normality of the estimators and provides a more concise form of limit covariance matrix comparing with that derived in Stute (1996).

Consider the linear regression model

$$T_i = \boldsymbol{W}_i\boldsymbol{\beta} + \varepsilon_i, i = 1, \ldots, n. \tag{2.25}$$

For the censoring mechanism it will be assumed that:

(i) T and C are independent;

(ii) $P(T \leqslant C|T, \boldsymbol{W}) = P(T \leqslant C|T)$.

The independence assumption (i) is a widely accepted assumption. The assumption (ii) was used in Stute (1993) and it makes the censored model (2.25) flexible enough to allow for a dependence between the covariates $\boldsymbol{W}$ and the censoring variable C. As pointed out by Stute (1996), the assumptions above can guarantee that the joint distribution of $(T, \boldsymbol{W})$ can be theoretically derived from that of $(X, \boldsymbol{W}, \delta)$ and consistently estimated from a sample $(X_i, \boldsymbol{W}_i, \delta_i)$, where $i = 1, \ldots, n$ (Bao et al., 2007).

Now we briefly describe the method proposed by He and Wong (2003). Let $\boldsymbol{W}_i = (W_{i1}, \cdots, W_{ip}), \boldsymbol{\Gamma} = E(\boldsymbol{W}_i^{tr}\boldsymbol{W}_i) = (\varsigma_{jk})_{j,k=1}^{p}, \boldsymbol{q} = E(\boldsymbol{W}_i^{tr}T_i) = (\varsigma_{01}, \cdots, \varsigma_{0p})^{tr}$, where

$$\varsigma_{0k} = E(W_{ik}T_i) = \int tw_k\tilde{F}(dt, d\boldsymbol{w}), 1 \leqslant k \leqslant p; \tag{2.26}$$

$$\varsigma_{jk} = E(W_{ij}W_{ik}) = \int w_jw_k\tilde{F}(dw_j, dw_k), 1 \leqslant j, k \leqslant p; \tag{2.27}$$

and $\tilde{F}(\cdot, \cdot)$ is the joint distribution function for T_i and $\boldsymbol{W}_i$. Multiplying the two sides

of (2.25) by $\boldsymbol{W}_i^{tr}$ and taking expectation yields

$$\boldsymbol{q} = \boldsymbol{\Gamma}\boldsymbol{\beta}, \tag{2.28}$$

where components of $\boldsymbol{q}$ and $\boldsymbol{\Gamma}$, ς_{0k} and ς_{jk}, can be estimated respectively by unbiased estimates

$$\hat{\varsigma}_{0k} = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i W_{ik}\delta_i}{1-\hat{G}^{\mathrm{KM}}(X_i-)}, \tag{2.29}$$

$$\hat{\varsigma}_{jk} = \frac{1}{n}\sum_{i=1}^{n}\frac{W_{ij}W_{2ik}\delta_i}{1-\hat{G}^{\mathrm{KM}}(X_i-)}, \tag{2.30}$$

where $j = 1, \ldots, p$, $k = 1, \ldots, p$, $\hat{G}^{\mathrm{KM}}(X_i-)$ is the left continuous version of the KM estimator for the distribution function of C_i.

Assuming that $\hat{\boldsymbol{\Gamma}}$ is invertible, the WLS estimator of $\boldsymbol{\beta}$ can be obtained by

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Gamma}}^{-1}\hat{\boldsymbol{q}}. \tag{2.31}$$

It can be easily proved that the estimator in (2.31) is a vector that minimizes

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n}\frac{\delta_i}{1-\hat{G}^{\mathrm{KM}}(X_i-)}(X_i - \boldsymbol{W}_i\boldsymbol{\beta})^2, \tag{2.32}$$

and hence is the same as the estimator proposed by Stute (1993).

Comparing this with the previous estimators, similar advantages to the KSV estimator are also achieved by the WLS estimator, which does not require iterative procedures. Furthermore, the WLS approach incorporates both weighted responses and weighted covariates. Therefore when the censoring mechanism depends on co-

variates, the false results from the KSV estimator may be avoided by the use of the WLS estimator (Bao et al., 2007).

2.2.6 Comparison & discussion

In this section, several existing estimators of regression parameter $\boldsymbol{\beta}$ for censored survival data are described. All approaches assume that the error terms ε_i are i.i.d random variables with mean 0 and an unknown distribution function, and the censoring variables C_i are i.i.d random variables which are independent of ε_i.

Both the KMLS estimator and the Miller estimator require the condition $G(t; \boldsymbol{W}) = G(t - \alpha - \boldsymbol{W}\boldsymbol{\beta})$ on the censoring distributions, which is necessary for the asymptotic consistency of the estimators. Since this assumption will rarely be satisfied in practice, a new estimator which does not need such restrictions was proposed by Buckley and James (1979). The BJ estimator does not depend on particular censoring patterns. The normal equations rather than the sum of squares of residuals are modified in BJ's approach and this can overcome the inconsistency problems in Miller's approach. The BJ estimator only assumes that the survival time T and the censoring time C are conditionally independent given the covariates $\boldsymbol{W}$, which also needs to be satisfied in Miller's approach. The KSV estimator is based on the assumption that the censoring distribution $G(t)$ does not vary with the covariates $\boldsymbol{W}$. However in practice censoring is frequently related to one of the covariates, such as age, therefore the basic assumption of the KSV estimator is not always valid. The WLS estimator assumes that $P(T \leqslant C|T, \boldsymbol{W}) = P(T \leqslant C|T)$, which makes the censored model (2.25) flexible enough to allow for a dependence between $\boldsymbol{W}$ and C. In other words, this assumption says that given the survival time, the covariates do

not provide any further information about whether censoring will happen or not. Note here both KSV and WLS approaches require that T and C should be independent, which is different from the conditional independence for the Miller estimator and BJ estimator.

Iterative procedures are needed in both the Buckley & James's approach and Miller's approach. The iterative sequences of the estimators can become trapped in loops and fail to converge for both methods. According to Buckley and James (1979), the loops are less frequent and less severe for the BJ estimator than for the Miller estimator. The advantage of the KSV estimator and the WLS estimator is that the estimates are in closed forms so that no iterations are required. The difference between these two methods is that the KSV estimator reweights only the responses, while the WLS estimator reweights both the responses and covariates. Therefore when the censoring mechanism depends on covariates, the covariates in the KSV approach remain unchanged which can produce a false result.

Miller and Harplen (1982) found that the BJ estimator was more reliable than the Miller estimator and the KSV estimator by applying these methods to the updated Stanford heart transplant data. Heller and Simonoff (1990) conducted an extensive simulation study to compare several methods of estimating parameters using linear regression when the response variable may be censored. They also concluded that the BJ estimator is preferred because the bias of the estimator is consistently smaller than that of the other methods, and its root mean squared error is usually smallest.

Bao et al. (2007) compared the estimator of Koul et al. and WLS estimator by using an extensive simulation study. Their simulation results indicated that the KSV estimator may be extremely biased when the censoring variable is dependent on the covariates. The same findings have also been reported by Fygenson and Zhou (1994).

The WLS approach, however, produced sensible estimates under various situations although the same weights as those in the KSV method are used. By carrying out an analysis of the Stanford heart transplant data, Bao et al. (2007) found that the WLS method yielded estimates as good as that by the BJ's method. They concluded that the WLS estimator performs much better than the KSV estimator, especially when the number of parameters in the model is large or the censoring is heavy.

2.3 Existing estimators for bivariate survival functions

In the above section, we have reviewed some existing regression techniques to handle the univariate censored survival data. When the response T itself is also subject to truncation, the univariate truncation models and the univariate truncation and censoring models have also been well studied (Gross and Huber-Carol, 1992; Gross and Lai, 1996; He and Yang, 2003). Besides univariate survival data, pairs of correlated event times are also often observed in survival analysis. For example in the hepatitis C cohort study in Fu et al. (2007), two recorded event times are the time from HCV infection to the development of cirrhosis, and the time from HCV infection to referral to the clinic cohort. The two correlated event times are referred to as bivariate survival data. To handle bivariate survival data, people usually utilize the KSV approach or the WLS approach since no iterations are required to estimate the regression parameter. Therefore, it is necessary to estimate the bivariate distribution function or the corresponding bivariate survival function of the pairwise event times.

Non-parametric estimators of bivariate distributions under right censoring have been proposed by Campbell (1981); Burke (1988); Dabrowska (1988); Tsai et al. (1990); Lin and Ying (1993); van der Laan (1996a); Akritas and Keilegom (2003); Prentice et al. (2004); Dai and Bao (2009). There has also been some work on estimating a bivariate distribution when observations are subject to truncation; see for example, Gürler (1996, 1997) for applications when a single component of the bivariate data is subject to truncation; van der Laan (1996b) and Huang et al. (2001) for bivariately truncated data.

However, there are few methods to estimate the bivariate survival function when both censoring and truncation occur. Gilbels and Gürler (1998) proposed a useful nonparametric estimator for bivariate survival data with both censoring and truncation. However, their method is confined to the special case where a single component of the bivariate data is subject to both censoring and truncation and the other component is fully observed.

For bivariate survival data where both components are under censoring and truncation, Shen (2006) proposed an inverse probability weighted (IPW) approach to estimate the joint survival function. Shen and Yan (2008) proposed two types of estimators as generalizations of Dabrowska (1988, 1989) estimator and Campbell and Földes (1982) estimator for the joint survival function under left truncation and right censoring. Their methods, however, require an iterative algorithm to calculate the distribution estimates, which is computationally heavy and may be impractical for data with a large sample size. Moreover, no analytic expressions of the asymptotic variance estimators are given, and their estimators are not illustrated by any real examples.

According to our literature search, the most recent work was done by Dai and

Fu (2012). They proposed a flexible estimator for paired survival times under both random truncation and random censoring. Their method is based on a polar coordinate transformation, which enables us to transform a bivariate survival function to a univariate survival function. A consistent estimator for the transformed univariate survival function is proposed. Then the univariate estimator can be transformed back to a bivariate estimator. Their estimator is in closed form, and it converges weakly to a zero-mean Gaussian process with an easily estimated covariance function. In this thesis, their method will be used to estimate the bivariate survival function we need.

Chapter 3

Accelerated failure time models

In this chapter, we will explore how the risk factors, such as age, HIV coinfection and heavy alcohol consumption, affect cirrhosis time. In our analysis, referral bias will be taken into account in order to obtain unbiased estimates for the covariates. We will focus on the accelerated failure time model (AFT) model in this chapter. Another commonly used model, the proportional hazards model (Cox model), will be discussed in the next chapter.

3.1 Statistical modelling and assumptions

3.1.1 Statistical modelling

We use T_i to denote the time from HCV infection to cirrhosis for patient i and $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{ip})$ to denote the p-dimensional covariate vector. The effect of covariates on incubation time from infection to cirrhosis is modelled by the accelerated failure time (AFT) model,

$$\log T_i = \boldsymbol{W}_i \boldsymbol{\beta} + \varepsilon_i, \tag{3.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{tr}$ is the regression coefficient vector and ε_i is the error term. Here we assume that the error terms $\{\varepsilon_i, i = 1, \cdots, n\}$ are independent and identically distributed (i.i.d.) with mean 0 and have an unknown distribution $F_\varepsilon(\cdot)$. We also assume that the error terms are independent of the covariates $\boldsymbol{W}$. Apart from these, we make no further assumptions on the distribution of the error terms, which allows our model to have wider applicability and increased robustness, compared to models with a parametric distribution assumption on the error terms.

In practice we observe a pair of event times for each patient, denoted by (R_i, T_i), where T_i is given above and R_i is the time from HCV infection to referral to the cohort. The value R_i is randomly right-truncated by L_i, which is the truncation time (from infection to the end of recruitment). The value T_i is subject to random right censoring by C_i, which is the censoring time (from infection to the last follow-up time). If $R_i > L_i$ then we cannot get any information about this patient. If $R_i \leqslant L_i$, then we can observe $\{R_i, L_i, X_i, \delta_i, \boldsymbol{W}_i, \}$, where $\delta_i = I[T_i \leqslant C_i]$, $X_i = \min\{T_i, C_i\}$. We denote the observed data as $\{R_i^*, L_i^*, X_i^*, \delta_i^*, \boldsymbol{W}_i^*\}$, $i = 1, \cdots, n$.

3.1.2 Independence assumption

Throughout this thesis, we make the following assumptions on $(R, T, \boldsymbol{W})$ and (L, C).

Assumption 3.1.1. *(i)* (L, C) *are independent of the covariate vector* $\boldsymbol{W}$.
(ii) (L, C) *are independent of the time pair* (R, T).

The assumption (i) is reasonable for our cirrhosis data since the end of recruitment was fixed before collecting the data so that the truncation time L can be viewed as independent of all patient information, and also the last follow-up time (censoring time) C can be usually any number of years after recruitment, which is

also independent of all patient information.

The assumption (ii) is a strong assumption and conditional independence which assumes that (L, C) and (R, T) are independent conditionally on $\boldsymbol{W}$ may be more appropriate in many other practical studies. Many existing methods for accelerated failure time models without referral bias only need this weaker assumption. These methods include Miller (1976) and Buckley and James (1979), which have been discussed in Chapter 2. When the weaker assumption, conditional independence on $\boldsymbol{W}$, is assumed, iteration algorithms are required to locate the estimates. This is because to find the model estimates we should consider the distribution of censoring variable C and under the assumption of conditional independence the distribution of C depends on the covariate vector $\boldsymbol{W}$. This further implies that the distribution of residual ε depends on the unknown parameter $\boldsymbol{\beta}$. Therefore, we usually give a staring value $\boldsymbol{\beta}^{(0)}$ to calculate the initial residual distribution. Then based on the residual distribution we estimate $\boldsymbol{\beta}^{(1)}$. The estimate of $\boldsymbol{\beta}$ is achieved when the iteration algorithm converges.

However, under the stronger assumption that $(R, T, \boldsymbol{W})$ and (L, C) are independent, we can work on the distribution of $(T, \boldsymbol{W})$ and C and no iteration algorithm is required. Examples of using stronger assumption, without using the iteration algorithms, include Koul et al. (1981) and Cheng et al. (1997).

3.1.3 Boundary conditions

Let $G(t_1, t_2) = P(L_i > t_1, C_i > t_2)$ be the continuous bivariate survival function for (L_i, C_i) and $\bar{F}(t_1, t_2) = P(R_i > t_1, T_i > t_2)$ be the continuous joint survival function

for (R_i, T_i). Define the boundaries of support for the density of $\bar{F}$ and G as

$$\boldsymbol{a}_{\bar{F}} = \inf\{(t_1, t_2) : \bar{F}(t_1, t_2) < 1\}, \boldsymbol{b}_{\bar{F}} = \sup\{(t_1, t_2) : \bar{F}(t_1, t_2) > 0\}, \tag{3.2}$$

$$\boldsymbol{a}_G = \inf\{(t_1, t_2) : \bar{G}(t_1, t_2) < 1\}, \boldsymbol{b}_G = \sup\{(t_1, t_2) : G(t_1, t_2) > 0\}. \tag{3.3}$$

Note that here $t_1 \geqslant 0$ and $t_2 \geqslant 0$ since $R \geqslant 0$ and $T \geqslant 0$. We also assume that the following conditions hold.

Condition 3.1.1. *(i) For* $\boldsymbol{s} = (s_1, s_2) \in \boldsymbol{b}_{\bar{F}}$ *and* $\boldsymbol{t} = (t_1, t_2) \in \boldsymbol{b}_G$, *we have* $\sqrt{s_1^2 + s_2^2} < \sqrt{t_1^2 + t_2^2}$, *given* $s_2/s_1 = t_2/t_1$.
(ii) For $\boldsymbol{s} = (s_1, s_2) \in \boldsymbol{a}_{\bar{F}}$ *and* $\boldsymbol{t} = (t_1, t_2) \in \boldsymbol{a}_G$, *we have* $\sqrt{s_1^2 + s_2^2} < \sqrt{t_1^2 + t_2^2}$, *given* $s_2/s_1 = t_2/t_1$.

The condition (i) requires that the upper bounds of the support regions of R and T are smaller than those for L and C respectively. It guarantees that the function $\bar{F}(t_1, t_2)$ is always identifiable in the whole support region from $\mathbf{0}$ to $\boldsymbol{b}_{\bar{F}}$. Basically, it requires that the maximum values that R and T can take are smaller than those for random variables L and C respectively. This is necessary because if the upper bound on the support of T is greater than that on C, no observation of T beyond the upper bound on the support of C can be observed. Thus the tail of the distribution of T cannot be identified. This is also reasonable in practice since our cirrhosis data were collected retrospectively (the data of HCV patients who were referred to hospital before 1999 were collected and the infection times were then identified by doctors). Therefore for any large values of R or T, it is always possible to collect data for an individual who was infected a long time ago with larger L or C. Our another basic assumption is that the observed data are from HCV-infected population who were

or will be eventually referred to hospital. We can not use clinic cohort data to make inference for those who are never referred to hospital.

Similarly, the condition (ii) guarantees that for any observed L there exists R such that $R < L$, which means all values of L can be possibly observed. Thus under this condition the bivariate survival function of (L, C), $G(t_1, t_2)$, is guaranteed to be identifiable. It is also realistic in our hepatitis C study. Even for a very small value of L, HCV-diagnosis and subsequent referral to liver clinics may happen shortly after infection because of regular HCV screening (more available in recent years), will give a very small value of $R \leqslant L$. Similarly, for a very small value of C, it is possible to have a patient who developed cirrhosis right after referral, which will give a small value of $T \leqslant C$.

3.2 Estimation of $\boldsymbol{\beta}$

For simplicity we use (R_i, T_i) to denote the pair of log-event times for the ith subject, where R_i is the logarithm of time from hepatitis C virus infection to referral to the cohort, and T_i is the logarithm of time from infection to the development of cirrhosis. The value R_i, is randomly right-truncated by L_i, which is the logarithm of truncation time. The value T_i, is subject to random right censoring by C_i, which is the logarithm of the censoring time. If $R_i > L_i$ then we cannot get any information about this patient. If $R_i \leqslant L_i$, then we can observe $\{R_i, L_i, X_i, \delta_i, \boldsymbol{W}_i, \}$, where $\delta_i = I[T_i \leqslant C_i]$, $X_i = \min\{T_i, C_i\}$ and $\boldsymbol{W}_i$ is the covariate vector. We denote the observed data as $\{R_i^*, L_i^*, X_i^*, \delta_i^*, \boldsymbol{W}_i^*\}$, $i = 1, \cdots, n$. Then the AFT model in (3.1) can be written as

$$T_i = \boldsymbol{W}_i\boldsymbol{\beta} + \varepsilon_i. \tag{3.4}$$

Let $F(t_1, t_2, \boldsymbol{w})$ be the joint distribution function for R_i, T_i and $\boldsymbol{W}_i$. Let $F^*(t_1, t_2, \boldsymbol{w}) = P(R_i^* \leqslant t_1, X_i^* \leqslant t_2, \boldsymbol{W}_i^* \leqslant \boldsymbol{w}, \delta_i^* = 1)$ be the joint cumulative distribution function for the uncensored vector $(R_i^*, X_i^*, \boldsymbol{W}_i^*, \delta_i^* = 1)$. Then we have the following result,

$$\begin{aligned} F^*(t_1, t_2, \boldsymbol{w}) &= P(R_i \leqslant t_1, X_i \leqslant t_2, \boldsymbol{W}_i \leqslant \boldsymbol{w}, \delta_i = 1 \big| R_i \leqslant L_i) \\ &= \frac{P(R_i \leqslant t_1, T_i \leqslant t_2, \boldsymbol{W}_i \leqslant \boldsymbol{w}, T_i \leqslant C_i, R_i \leqslant L_i)}{P(R_i \leqslant L_i)} \\ &= \gamma^{-1} E\big\{ P(R_i \leqslant t_1, T_i \leqslant t_2, \boldsymbol{W}_i \leqslant \boldsymbol{w}, T_i \leqslant C_i, R_i \leqslant L_i \big| R_i, T_i, \boldsymbol{W}_i) \big\} \\ &= \gamma^{-1} E\Big\{ I[R_i \leqslant t_1, T_i \leqslant t_2, \boldsymbol{W}_i \leqslant \boldsymbol{w}] \cdot G(R_i-, T_i-) \Big\} \\ &= \gamma^{-1} \int_{s_1 \leqslant t_1} \int_{s_2 \leqslant t_2} \int_{\boldsymbol{u} \leqslant \boldsymbol{w}} G(s_1-, s_2-) F(ds_1, ds_2, d\boldsymbol{u}). \end{aligned} \tag{3.5}$$

Under (i) of Condition 3.1.1, for all $\boldsymbol{t} = (t_1, t_2)$, $\boldsymbol{0} \leqslant \boldsymbol{t} \leqslant \boldsymbol{b}_{\bar{F}}$, we have $G(t_1, t_2) = P(L_i > t_1, C_i > t_2) > 0$. Therefore by rearranging (3.5) we have the following relation,

$$F(t_1, t_2, \boldsymbol{w}) = \gamma \cdot \int_{s_1 \leqslant t_1} \int_{s_2 \leqslant t_2} \int_{\boldsymbol{u} \leqslant \boldsymbol{w}} \frac{1}{G(s_1-, s_2-)} F^*(ds_1, ds_2, d\boldsymbol{u}) \tag{3.6}$$

where $\gamma = P(R_i \leqslant L_i)$ is the truncation probability.

The following lemma gives an unbiased estimating equation for $\boldsymbol{\beta}$, provided $G(t_1, t_2)$ is known.

Lemma 3.2.1. *Parameter $\boldsymbol{\beta}$ satisfies*

$$\boldsymbol{q} = \boldsymbol{\Gamma}\boldsymbol{\beta} \tag{3.7}$$

where $\boldsymbol{\Gamma} = (\varsigma_{jk})_{j,k=1}^{p}$, $\varsigma_{jk} = E(W_{ij}W_{ik}) = \int w_j w_k F(dt_1, dt_2, d\boldsymbol{w})$ *and* $\boldsymbol{q} = (\varsigma_{01}, \cdots, \varsigma_{0p})^{tr}$, $\varsigma_{0k} = E(W_{ik}T_i) = \int t_2 w_k F(dt_1, dt_2, d\boldsymbol{w})$.

If G is known, ς_{0k} and ς_{jk}, can be estimated respectively by the unbiased estimates

$$\hat{\varsigma}_{0k}(G) = \frac{\hat{\gamma}}{n}\sum_{i=1}^{n}\frac{X_i^* W_{ik}^* \delta_i^*}{G(R_i^*-, X_i^*-)}, \tag{3.8}$$

$$\hat{\varsigma}_{jk}(G) = \frac{\hat{\gamma}}{n}\sum_{i=1}^{n}\frac{W_{ij}^* W_{ik}^* \delta_i^*}{G(R_i^*-, X_i^*-)}, \tag{3.9}$$

where $\hat{\gamma}$ is the truncation probability estimate. Then a consistent estimator

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Gamma}^{-1}(\hat{G})\boldsymbol{q}(\hat{G}), \tag{3.10}$$

can be obtained if $\hat{G}(t_1, t_2)$ is a consistent estimator for the function $G(t_1, t_2)$.

Proof. See Appendix A.1. □

The method in Lemma 3.2.1 is actually the weighted least squares (WLS) method. We may also use the KSV method (Koul et al., 1981) to estimate the parameters. The KSV method is also based on an inverse probability weighted approach, but it only re-weights the response variable. In contrast, the WLS method re-weights both the response variable and the predictors. Bao et al. (2007) provided detailed simulation studies which can indicate that the KSV estimator may be extremely biased when the censoring variable is dependent on the covariates as well as when the censoring times are i.i.d. and do not vary with the covariates. The WLS method, however, yields satisfactory estimates under various situations although the same weights as those in the KSV method are used. Therefore here in this chapter we did not consider the KSV estimator, although the proposed method can be extended to the KSV approach.

From Lemma 3.2.1, we can see that we only need a consistent estimate $\hat{G}$ to

evaluate the estimate for $\hat{\boldsymbol{\beta}}$. The truncation probability estimate $\hat{\gamma}$ is cancelled out in (3.10) and does not need to be estimated in advance. We will discuss how to estimate the bivariate survival function G in the next section.

3.3 Estimation of the bivariate survival function G

Following the idea of Dai and Fu (2012), we consider the following polar coordinate transformation

$$\begin{aligned} G(t_1, t_2) &= P(L > t_1, C > t_2) \\ &= P\left(\frac{L}{t_1}\sqrt{{t_1}^2 + {t_2}^2} > \sqrt{{t_1}^2 + {t_2}^2}, \frac{C}{t_2}\sqrt{{t_1}^2 + {t_2}^2} > \sqrt{{t_1}^2 + {t_2}^2}\right) \\ &= P(Z(\alpha) > z) \\ &:= G(z; \alpha), \end{aligned} \tag{3.11}$$

where $\alpha = t_2/t_1$, $z = \sqrt{{t_1}^2 + {t_2}^2}$ and $Z(\alpha) = \min\{L\sqrt{1+\alpha^2}, C\sqrt{1+\alpha^{-2}}\}$. The above transformation can be explained by Figure 3.1. Points $\boldsymbol{P}_1 : (L, \alpha L)$ and $\boldsymbol{P}_2 : (\alpha^{-1}C, C)$ correspond to vertical and horizontal projections of (L, C) onto the line $v_2 = \alpha v_1$. The value of $Z(\alpha)$ is the minimum of the distances of these two points from the origin. Then $G(z; \alpha) = P(Z(\alpha) > z)$ means the survival function of $Z(\alpha)$ is on the line $v_2 = \alpha v_1$. Note that if $t_1 = 0$, $t_2 > 0$ then $Z(\alpha) = C$ and if $t_1 > 0$, $t_2 = 0$ then $Z(\alpha) = L$. Therefore the above transformation exists for all $(t_1, t_2) \in [0, \infty] \times [0, \infty]$.

The advantage of such a transformation is that by fixing α the function $G(z; \alpha)$

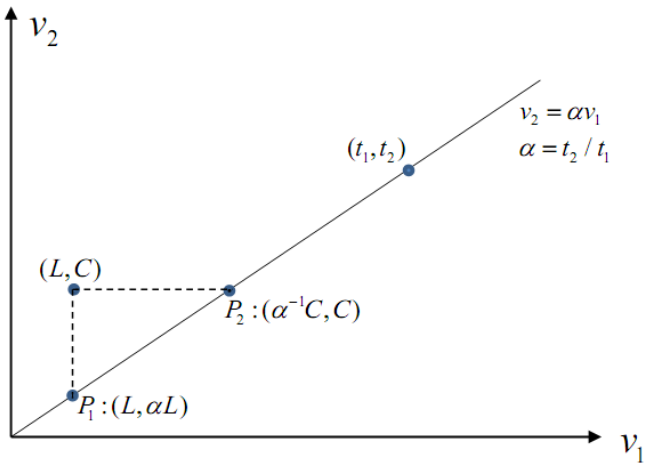

Figure 3.1: $\boldsymbol{P}_1 : (L, \alpha L)$ is obtained by vertical projection; $\boldsymbol{P}_2 : (\alpha^{-1}C, C)$ is obtained by horizontal projection.

is univariate and can be estimated using existing methods based on the transformed data $Z(\alpha)$. The transformed data $Z(\alpha)$ still have the information of both L and C. Therefore using such a transformation will make the estimation easier but it does not lose data information.

We only need to find a consistent estimator, $\hat{G}(z;\alpha)$, for the transformed univariate survival function $G(z;\alpha)$. Then, according to (3.11), the estimator $\hat{G}(z;\alpha)$ is also an estimate for $G(t_1, t_2)$.

In practice, due to censoring and truncation, the values of (L, C) may not be

obtained. Thus $Z(\alpha)$ may not be available. Define the transformed data as:

$$\begin{aligned}
\tilde{Z}(\alpha) &= \min\{\tilde{L}, \tilde{X}\}, \\
\Delta(\alpha) &= I[\tilde{L} \leqslant \tilde{X}] + (1-\delta)I[\tilde{L} > \tilde{X}], \\
V(\alpha) &= R\sqrt{1+\alpha^2},
\end{aligned}$$

where $\tilde{L} = L\sqrt{1+\alpha^2}$ and $\tilde{X} = X\sqrt{1+\alpha^{-2}}$. Then we observe $\big(\tilde{Z}(\alpha), \Delta(\alpha), V(\alpha)\big)$ if $R \leqslant L$ and nothing otherwise. Let $\tilde{L}^* = L^*\sqrt{1+\alpha^2}$ and $\tilde{X}^* = X^*\sqrt{1+\alpha^{-2}}$. The observed data after transformation are:

$$\begin{aligned}
\tilde{Z}^*(\alpha) &= \min\{\tilde{L}^*, \tilde{X}^*\}, \\
\Delta^*(\alpha) &= I[\tilde{L}^* \leqslant \tilde{X}^*] + (1-\delta^*)I[\tilde{L}^* > \tilde{X}^*], \\
V^*(\alpha) &= R^*\sqrt{1+\alpha^2}, \qquad (3.12)
\end{aligned}$$

where L^*, R^*, X^*, δ^* are the original observed data.

Lemma 3.3.1. *Based on the transformed data given in (3.12), we have*

$$\big\{\Delta^*(\alpha) = 1\big\} \Leftrightarrow \big\{\tilde{Z}^*(\alpha) = Z^*(\alpha)\big\}, \qquad (3.13)$$

$$\big\{\Delta^*(\alpha) = 0\big\} \Leftrightarrow \big\{\tilde{Z}^*(\alpha) \leqslant Z^*(\alpha)\big\}. \qquad (3.14)$$

Proof. See Appendix A.2. □

Lemma 3.3.1 indicates that $\tilde{Z}^*(\alpha)$ is an observed value for $Z^*(\alpha)$ and $\Delta^*(\alpha) = 0$ implies censoring (Wang et al., 2013). The transformation in (3.12) introduces artificial truncation and censoring. Truncation information is given by $V^*(\alpha)$ and

censoring information is given by $\Delta^*(\alpha)$. Then based on the transformed observations in (3.12), we can estimate $G(z;\alpha)$ using the following lemma.

Lemma 3.3.2. *For fixed α, the hazard rate function of $Z(\alpha)$ is denoted by $\Lambda(dz;\alpha) = -G(dz;\alpha)/G(z-;\alpha)$. Then we have*

$$\Lambda(dz;\alpha) = \frac{P(\tilde{Z}^*(\alpha) \in dz, z > V^*(\alpha), \Delta^*(\alpha) = 1)}{P(\tilde{Z}^*(\alpha) \geqslant z > V^*(\alpha))},$$

where $\tilde{Z}^*(\alpha) \in dz$ denotes $z \leqslant \tilde{Z}^*(\alpha) < z + dz$.

Proof. See Appendix A.3. □

Based on the transformed observations in (3.12), we also define

$$\begin{aligned}
N(ds;\alpha) &= n^{-1}\sum_{i=1}^{n} N_i(ds;\alpha), \\
&= n^{-1}\sum_{i=1}^{n} I[\tilde{Z}_i^*(\alpha) \in ds, s > V_i^*(\alpha), \Delta_i^*(\alpha) = 1], \\
H_{(n)}(s;\alpha) &= n^{-1}\sum_{i=1}^{n} H_i(s;\alpha), \\
&= n^{-1}\sum_{i=1}^{n} I[\tilde{Z}_i^*(\alpha) > s \geqslant V_i^*(\alpha)], \\
H_{(n)}(t_1,t_2) &= n^{-1}\sum_{i=1}^{n} H_i(t_1,t_2), \\
&= n^{-1}\sum_{i=1}^{n} I\left[L_i^* > t_1 \geqslant R_i^*, X_i^* > t_2\right]. \qquad (3.15)
\end{aligned}$$

Note that $H_{(n)}(t_1,t_2) = H_{(n)}(z;\alpha)$ and $H_i(t_1,t_2) = H_i(z;\alpha)$.

Hence Lemma 3.3.2 implies that an estimator for $\Lambda(dz;\alpha)$ is of the form $\hat{\Lambda}(dz;\alpha) =$

$N(dz;\alpha)/H_{(n)}(z-;\alpha)$. Then the product-limit estimator for $G(z;\alpha)$ is given by

$$\hat{G}(z;\alpha) = \prod_{s \leqslant z} \left\{ 1 - \frac{\Delta N(s;\alpha)}{H_{(n)}(s-;\alpha)} \right\}, \tag{3.16}$$

where $\Delta N(s;\alpha) = N(s;\alpha) - N(s-;\alpha)$. Since $G(z;\alpha) = G(t_1, t_2)$, $\hat{G}(z;\alpha)$ is also an estimator for $G(t_1, t_2)$.

3.4 Large sample properties of $\hat{G}$

In this section we first give necessary notations and a lemma in the subsection 3.4.1. Then we show the consistency of $\hat{G}$ given by (3.16) in the subsection 3.4.2 and show its asymptotic normality in the subsection 3.4.3.

3.4.1 Notation

Define the σ-field

$$\mathcal{F}^i_{z;\alpha} = \sigma\Big\{ I\big[\tilde{Z}^*_i(\alpha) \leqslant s\big], \Delta^*_i(\alpha), I\big[V^*_i(\alpha) \leqslant s\big], 0 \leqslant s \leqslant z \Big\}. \tag{3.17}$$

Following Dai and Fu (2012), if we equip the counting process $\{N_i(z;\alpha), z \geqslant 0\}$ with the filtration $\{\mathcal{F}^i_{z;\alpha}, z \geqslant 0\}$, we have

$$E\big[N_i(z;\alpha)\big|\mathcal{F}^i_{z-;\alpha}\big] = H_i(z-;\alpha)\Lambda(dz;\alpha). \tag{3.18}$$

Let

$$\begin{aligned} M_j(ds;\alpha) &= N_j(ds;\alpha) - H_j(s-;\alpha)\Lambda(ds;\alpha), \\ M(ds;\alpha) &= n^{-1}\sum_{j=1}^{n} M_j(ds;\alpha). \end{aligned} \tag{3.19}$$

Then for fixed α, $M_j(ds;\alpha)$ and $M(ds;\alpha)$ are martingales with respect to the filtration $\mathcal{F}^i_{z;\alpha}$ and $\mathcal{F}_{z;\alpha}$ respectively, where $\mathcal{F}_{z;\alpha} = \vee_{i=1}^{n}\mathcal{F}_{z;\alpha}$.

Let

$$H(s;\alpha) = E[H_i(s;\alpha)] = G(s_1,s_2)P[s_1 \geqslant R, T > s_2]/\gamma, \tag{3.20}$$

and for $(t_1,t_2) \in \boldsymbol{b}_{\bar{F}}, t_2/t_1 = \alpha$, define $\tau_\alpha = \sup_s\big\{s : H(s;\alpha) > 0\big\}$. Then $[0,\tau_\alpha] \times [0,\infty]$ can be viewed as the range for $(\tilde{Z}^*_i(\alpha^*_i),\alpha^*_i)$.

The following conditions and lemma are needed to prove the consistency of $\hat{G}(z;\alpha)$.

Condition 3.4.1. *(i) For $\alpha \in [0,\infty], z \in [0,\tau_\alpha]$, we have $\int_0^z \frac{1}{H(s-;\alpha)}\Lambda(ds;\alpha) < \infty$. (ii) For $\alpha \in [0,\infty], z \in [0,\tau_\alpha]$, we have $\int_0^z \Lambda(ds;\alpha) < \infty$.*

These conditions are common in survival analysis based on counting processes (Fleming and Harrington, 1991). The condition (i) guarantees that $\int_0^z \frac{1}{H(s-;\alpha)}\Lambda(ds;\alpha)$ is bounded for $\alpha \in [0,\infty], z \in [0,\tau_\alpha]$. The condition (ii) can be obtained from (i) immediately. We assume that Condition 3.4.1 hold throughout the later sections and chapters.

Lemma 3.4.1. *Under certain mild conditions, the following terms*
(i) $\sup_{\alpha,z} E\left\{\sqrt{n}\int_0^z I[H_{(n)}(s-;\alpha)=0]\Lambda(ds;\alpha)\right\}^2$
(ii) $\sup_{\alpha,z} E\left\{\sqrt{n}\int_0^z \frac{1}{H(s-;\alpha)} I[H_{(n)}(s-;\alpha)=0]\Lambda(ds;\alpha)\right\}^2$

converges to 0 *as* $n \to \infty$*, where* $\alpha \in [0, \infty]$*, and* $z \in [0, \tau_\alpha]$.

Proof. See Appendix B.1. □

3.4.2 Consistency of $\hat{G}(z; \alpha)$

For $z \in [0, \tau_\alpha]$, following the results in Fleming and Harrington (1991) and using integration by parts, we have the following martingale representation for $\hat{G}(z; \alpha)$,

$$\begin{aligned} & \hat{G}(z;\alpha) - G(z;\alpha) \\ = \; & -G(z;\alpha) \int_0^z \frac{\hat{G}(s-;\alpha)}{G(s;\alpha)} \left[\frac{N(ds;\alpha)}{H_{(n)}(s-;\alpha)} - \Lambda(ds;\alpha) \right] \\ = \; & -G(z;\alpha) \int_0^z \frac{\hat{G}(s-;\alpha)}{G(s;\alpha)} \frac{I[H_{(n)}(s-;\alpha) > 0]}{H_{(n)}(s-;\alpha)} M(ds;\alpha) + B(z;\alpha), \end{aligned}$$

where

$$B(z;\alpha) = G(z;\alpha) \int_0^z \frac{\hat{G}(s-;\alpha)}{G(s;\alpha)} I[H_{(n)}(s-;\alpha) = 0] \Lambda(ds;\alpha). \tag{3.21}$$

Following this martingale representation, we have that for all $z \in [0, \tau_\alpha]$,

$$\begin{aligned} & \sqrt{n} \left[\hat{G}(z;\alpha) - G(z;\alpha) \right] \\ = \; & -\sqrt{n} G(z;\alpha) \int_0^z \left[\frac{\hat{G}(s-;\alpha)}{G(s;\alpha)} \frac{I[H_{(n)}(s-;\alpha) > 0]}{H_{(n)}(s-;\alpha)} - \frac{1}{H(s-;\alpha)} \right] M(ds;\alpha) \\ & -\sqrt{n} G(z;\alpha) \int_0^z \frac{1}{H(s-;\alpha)} M(ds;\alpha) + \sqrt{n} B(z;\alpha) \\ := \; & \omega_n(z;\alpha) - \sqrt{n} G(z;\alpha) \int_0^z \frac{1}{H(s-;\alpha)} M(ds;\alpha) + \sqrt{n} B(z;\alpha). \end{aligned} \tag{3.22}$$

To prove the consistency of $\hat{G}(z; \alpha)$, we need to prove that each term in (3.22)

converges to 0 as $n \to \infty$. Since $G(z;\alpha) \leqslant G(s;\alpha)$ given $z \geqslant s$, we have

$$B(z;\alpha) \leqslant \int_0^z I[H_{(n)}(s-;\alpha) = 0]\Lambda(ds;\alpha), \, a.s..$$

Then using the results of Lemma 3.4.1, we have

$$\sup_{\alpha\in[0,\infty],z\in[0,\tau_\alpha]} E\big[\sqrt{n}B(z;\alpha)\big]^2 = o(1). \tag{3.23}$$

In addition, from (i) in Condition 3.4.1, we can easily have that

$$E\bigg[\int_0^z \frac{1}{H(s-;\alpha)} M(ds;\alpha)\bigg]^2 = \frac{1}{n}\int_0^z \frac{1}{H(s-;\alpha)}\Lambda(ds;\alpha) = O(n^{-1}). \tag{3.24}$$

The following lemma is also needed to prove the consistency of $\hat{G}(z;\alpha)$.

Lemma 3.4.2. *Define* $\tau_{n;\alpha} = \sup_s \big\{s : H(s;\alpha) > n^{-\kappa/2} \cdot \log n\big\}$ *for* $\kappa \in (0,1)$ *, so that* $\tau_{n;\alpha} \to \tau_\alpha$ *as* $n \to \infty$*. Then the first term in (3.22) satisfies*

$$\sup_{\alpha\in[0,\infty],z\in[0,\tau_{n;\alpha}]} E\big[\omega_n(z;\alpha)\big]^2 = o(1). \tag{3.25}$$

Proof. See in Appendix B.2. □

Using the results of Lemma 3.4.2, together with (3.22), (3.23) and (3.24), we have that for $\alpha \in [0,\infty]$, $z \in [0,\tau_{n;\alpha}]$,

$$E\big[\hat{G}(z;\alpha) - G(z;\alpha)\big]^2 = o(1), \tag{3.26}$$

which indicates the consistency of $\hat{G}(z;\alpha)$.

3.4.3 Asymptotic normality of $\hat{G}(z;\alpha)$

From (3.22), we have that

$$\begin{aligned}
& \sqrt{n}\left[\hat{G}(z;\alpha) - G(z;\alpha)\right] \\
= & -\sqrt{n}G(z;\alpha)\int_{s\leqslant z}\frac{1}{H(s-;\alpha)}M(ds;\alpha) + r_n(z;\alpha) \\
= & -\frac{1}{\sqrt{n}}G(z;\alpha)\sum_{j=1}^{n}\int_{s\leqslant z}\frac{1}{H(s-;\alpha)}M_j(ds;\alpha) + r_n(z;\alpha), \qquad (3.27)
\end{aligned}$$

where

$$\begin{aligned}
r_n(z;\alpha) = & -\sqrt{n}G(z;\alpha)\int_0^z\left[\frac{\hat{G}(s-;\alpha)}{G(s;\alpha)}\frac{I[H_{(n)}(s-;\alpha)>0]}{H_{(n)}(s-;\alpha)} - \frac{1}{H(s-;\alpha)}\right]M(ds;\alpha) \\
& +\sqrt{n}B(z;\alpha), \qquad (3.28)
\end{aligned}$$

and it is such that $\sup_{z\in[0,\tau_\alpha],\alpha\in[0,\infty]} E\left[r_n(z;\alpha)\right]^2 \to 0$.

The equation (3.27) implies that $\sqrt{n}\left[\hat{G}(z;\alpha) - G(z;\alpha)\right]$ can be written as sum of some i.i.d. terms. The asymptotic normality follows immediately. And further for $\alpha \in [0,\infty]$, $z \in [0,\tau_\alpha]$, we have that $\sqrt{n}\left[\hat{G}(z;\alpha) - G(z;\alpha)\right] \to N\left(0,\sigma^2(z;\alpha)\right)$, where

$$\sigma^2(z;\alpha) = G^2(z;\alpha)\int_{s\leqslant z}\frac{1}{H(s-;\alpha)}\Lambda(ds;\alpha). \qquad (3.29)$$

3.5 Large sample properties of $\hat{\boldsymbol{\beta}}$

3.5.1 Consistency of $\hat{\boldsymbol{\beta}}$

Let $\boldsymbol{\beta}^*$ be the true value of $\boldsymbol{\beta}$. Following the Lemma 8.3.1 in Fleming and Harrington (1991) (see in Appendix C.1), let a sequence of concave functions $F_n(\boldsymbol{\beta})$ be of the

form

$$\begin{aligned} F_n(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i^* I\big[\hat{G}(R_i^*-,X_i^*-)>0\big]}{\hat{G}(R_i^*-,X_i^*-)}(X_i^*-\boldsymbol{W}_i^*\boldsymbol{\beta})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i^* I\big[\hat{G}(R_i^*-,X_i^*-)>0\big]}{\hat{G}(R_i^*-,X_i^*-)}(X_i^*-\boldsymbol{W}_i^*\boldsymbol{\beta}^*)^2 \\ &\quad +\frac{2}{n}\sum_{i=1}^{n}\frac{\delta_i^* I\big[\hat{G}(R_i^*-,X_i^*-)>0\big]}{\hat{G}(R_i^*-,X_i^*-)}(X_i^*-\boldsymbol{W}_i^*\boldsymbol{\beta}^*)\,(\boldsymbol{W}_i^*\boldsymbol{\beta}^*-\boldsymbol{W}_i^*\boldsymbol{\beta}) \\ &\quad +\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i^* I\big[\hat{G}(R_i^*-,X_i^*-)>0\big]}{\hat{G}(R_i^*-,X_i^*-)}(\boldsymbol{W}_i^*\boldsymbol{\beta}^*-\boldsymbol{W}_i^*\boldsymbol{\beta})^2 \\ &:= I+II+III, \end{aligned} \tag{3.30}$$

where the term II can be written as

$$\begin{aligned} II &= \frac{2}{n}\sum_{i=1}^{n}\left[\frac{\delta_i^* I\big[\hat{G}(R_i^*-,X_i^*-)>0\big]}{\hat{G}(R_i^*-,X_i^*-)}-\frac{\delta_i^*}{G(R_i^*-,X_i^*-)}\right](X_i^*-\boldsymbol{W}_i^*\boldsymbol{\beta}^*)\,(\boldsymbol{W}_i^*\boldsymbol{\beta}^*-\boldsymbol{W}_i^*\boldsymbol{\beta}) \\ &\quad +\frac{2}{n}\sum_{i=1}^{n}\frac{\delta_i^*}{G(R_i^*-,X_i^*-)}(X_i^*-\boldsymbol{W}_i^*\boldsymbol{\beta}^*)\,(\boldsymbol{W}_i^*\boldsymbol{\beta}^*-\boldsymbol{W}_i^*\boldsymbol{\beta}) \\ &:= II_{(1)}+II_{(2)}. \end{aligned} \tag{3.31}$$

The term $II_{(1)}\to 0$ as $n\to\infty$, since

$$\frac{I\big[\hat{G}(R_i^*-,X_i^*-)>0\big]}{\hat{G}(R_i^*-,X_i^*-)}-\frac{1}{G(R_i^*-,X_i^*-)}\xrightarrow{p}0.$$

By using the similar idea in the proof of Lemma 3.2.1, the term $II_{(2)}$ can be written as

$$II_{(2)}=\int(t_2-\boldsymbol{W}_i\boldsymbol{\beta}^*)(\boldsymbol{W}_i\boldsymbol{\beta}^*-\boldsymbol{W}_i\boldsymbol{\beta})\hat{F}(dt_1,dt_2,d\boldsymbol{w}), \tag{3.32}$$

where $\hat{F}(t_1,t_2,\boldsymbol{w})$ is a consistent estimate of $F(t_1,t_2,\boldsymbol{w})$, the joint distribution func-

tion for $(R_i, T_i, \boldsymbol{W}_i)$ defined in section 3.2. Then when $n \to \infty$ we have that

$$\begin{aligned} II_{(2)} &\to E\big[(T_i - \boldsymbol{W}_i\boldsymbol{\beta}^*)(\boldsymbol{W}_i\boldsymbol{\beta}^* - \boldsymbol{W}_i\boldsymbol{\beta})\big] \\ &= E_{\boldsymbol{W}_i}\Big\{E\big[T_i - \boldsymbol{W}_i\boldsymbol{\beta}^*|\boldsymbol{W}_i\big] \cdot (\boldsymbol{W}_i\boldsymbol{\beta}^* - \boldsymbol{W}_i\boldsymbol{\beta})\Big\} \\ &= E_{\boldsymbol{W}_i}\Big\{E\big[\varepsilon_i|\boldsymbol{W}_i\big] \cdot (\boldsymbol{W}_i\boldsymbol{\beta}^* - \boldsymbol{W}_i\boldsymbol{\beta})\Big\} \\ &= 0, \end{aligned} \tag{3.33}$$

since the mean of ε_i is assumed to be 0. Therefore, we conclude that the term II in (3.30) converges to 0 as $n \to \infty$.

By using the similar idea above, we have that when $n \to \infty$, the term $I \to E[\varepsilon_i^2] = Var(\varepsilon_i)$, which is a constant. The term III is a concave function which reaches the minimum at $\boldsymbol{\beta} = \boldsymbol{\beta}^*$.

Let $f(\boldsymbol{\beta})$ be the limit of $F_n(\boldsymbol{\beta})$ as $n \to \infty$. Then we have that the function $f(\boldsymbol{\beta})$ has a unique minimum value at $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. Moreover, the concave function $F_n(\boldsymbol{\beta})$ reaches its unique minimum value at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$. Therefore, from the results of the Lemma 8.3.1 in Fleming and Harrington (1991), we conclude that $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}^*$ in probability as $n \to \infty$.

3.5.2 Asymptotic normality of $\hat{\boldsymbol{\beta}}$

We have proved that the asymptotic normality for $G(z;\alpha)$. Since $G(z;\alpha) = G(t_1, t_2)$, $\hat{G}(z;\alpha)$ is also an estimator for the bivariate survival function $G(t_1, t_2)$. By substituting the $\hat{G}$ given in (3.16) into (3.10), we can get the estimator $\hat{\boldsymbol{\beta}}(\hat{G})$.

The estimator $\hat{\boldsymbol{\beta}}(\hat{G})$ given in (3.10) is equivalent to solving the estimating equa-

tion

$$\boldsymbol{Q}\left(\boldsymbol{\beta};\hat{G}\right)=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i^*I\left[\hat{G}(R_i^*-,X_i^*-)>0\right]}{\hat{G}(R_i^*-,X_i^*-)}(X_i^*-\boldsymbol{W}_i^*\boldsymbol{\beta})\boldsymbol{W}_i^{*tr}=\mathbf{0}, \tag{3.34}$$

since $\boldsymbol{Q}\left(\hat{\boldsymbol{\beta}};\hat{G}\right)=\boldsymbol{\Gamma}(\hat{G})\hat{\boldsymbol{\beta}}-\boldsymbol{q}(\hat{G})=\mathbf{0}$. The following theorem provides the results of asymptotic normality for $\hat{\boldsymbol{\beta}}$.

Theorem 3.5.1. *Let*

$$\boldsymbol{\eta}_i=\delta_i^*(X_i^*-\boldsymbol{W}_i^*\boldsymbol{\beta}^*)\boldsymbol{W}_i^{*tr},\ \ \xi_{ji}=\int_{s<\tilde{Z}_i^*(\alpha_i)}\frac{1}{H(s-;\alpha_i)}M_j(ds;\alpha_i) \tag{3.35}$$

where $\alpha_i=X_i^*/R_i^*$ *and the notation "tr" stands for matrix transpose. Let* $\boldsymbol{Q}'(\boldsymbol{\beta};G)=\partial\boldsymbol{Q}(\boldsymbol{\beta};G)/\partial\boldsymbol{\beta}$ *and* $\mathcal{D}_k=\{R_k^*,L_k^*,X_k^*,\delta_k^*,\boldsymbol{W}_k^*\}$ *denote the observed information of patient k. Then* $\sqrt{n}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\right)\sim N\left(\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right)$*, where* $\boldsymbol{\beta}^*$ *denotes the true value of* $\boldsymbol{\beta}$*, and* $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}=\left[\boldsymbol{Q}'(\boldsymbol{\beta};G)\right]^{-1}\boldsymbol{\Sigma}_{\boldsymbol{Q}}\left\{\left[\boldsymbol{Q}'(\boldsymbol{\beta};G)\right]^{-1}\right\}^{tr}$*. The matrix* $\boldsymbol{\Sigma}_{\boldsymbol{Q}}$ *is given by*

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{Q}} &= \mathrm{Var}\left[\frac{\boldsymbol{\eta}_k}{G(R_k^*-,X_k^*-)}+\boldsymbol{\Phi}(\mathcal{D}_k)\right], &(3.36)\\ \boldsymbol{\Phi}(\mathcal{D}_k) &= E\left\{G^{-1}(R_i^*-,X_i^*-)\xi_{ki}\boldsymbol{\eta}_i\,|\mathcal{D}_k\right\},\ i\neq k. &(3.37)\end{aligned}$$

Proof. See Appendix C.2. □

Let $\hat{\boldsymbol{\eta}}_i,\hat{\boldsymbol{\Phi}}(\mathcal{D}_i)$ to denote the estimates for $\boldsymbol{\eta}_i,\boldsymbol{\Phi}(\mathcal{D}_i)$ (replacing G, $\boldsymbol{\beta}$ with $\hat{G}$, $\hat{\boldsymbol{\beta}}$ respectively in $\boldsymbol{\eta}_i,\boldsymbol{\Phi}(\mathcal{D}_i)$). Then an estimate of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ is given by

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}=\left[\boldsymbol{Q}'\left(\hat{\boldsymbol{\beta}};\hat{G}\right)\right]^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{Q}}\left\{\left[\boldsymbol{Q}'\left(\hat{\boldsymbol{\beta}};\hat{G}\right)\right]^{-1}\right\}^{tr}, \tag{3.38}$$

where

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{Q}} = \widehat{\mathrm{V}ar}\left[\frac{\hat{\boldsymbol{\eta}}_i}{\hat{G}(R_i^*-, X_i^*-)} + \hat{\boldsymbol{\Phi}}(\mathcal{D}_i)\right]$$

is the sample covariance matrix based on $\dfrac{\hat{\boldsymbol{\eta}}_i}{\hat{G}(R_i^*-, X_i^*-)} + \hat{\boldsymbol{\Phi}}(\mathcal{D}_i)$, $i = 1, \cdots, n$. Here the notation $\widehat{\mathrm{V}ar}$ denotes the estimated covariance matrix.

3.6 Simulation studies and data analysis

3.6.1 Simulation studies

In this section, we conduct simulation studies to assess the performance of the proposed methods. Truncation times L and censoring times C are generated respectively from

$$\begin{aligned} C &= a\nu_1 + b\nu_2, \\ L &= c\nu_1 + d\nu_2 + U[0,1], \end{aligned}$$

where ν_1 and ν_2 are exponentially distributed with unit mean. We can change the values of (a, b, c, d) to adjust censoring/truncation probabilities and correlations of L and C. Pairs of survival times are generated from the following model to mimic the data analysis. The logarithm of survival time follows $T = \boldsymbol{W}\boldsymbol{\beta}^* + \varepsilon$, where the true value $\boldsymbol{\beta}^* = (3.7, -0.05, -0.3, -0.1)^{tr}$ and the covariate $\boldsymbol{W}$ is defined as: the intercept $W_1 = 1$ and the predictors $W_2 \sim U[20, 30]$, $W_3 \sim$ Bernoulli(0.5) and $W_4 \sim$ Bernoulli(0.5). The observed sample size is set at $n = 200$ and 100. The replication time is 500. The time R has mean $ER = 1.35$ and (R, T) are correlated through

a joint distribution of $(R - ER, \varepsilon)$. Once we simulate $(R - ER, \varepsilon)$ we then have the simulated values for R and $T = \boldsymbol{W}\boldsymbol{\beta}^* + \varepsilon$. We here consider two types of joint distribution functions of $(R - ER, \varepsilon)$.

Scenario 1: The two error terms $R - ER$ and ε are generated as follows,

$$\begin{aligned} R - ER &= 1.0\upsilon_1 + 0.5\upsilon_2, \\ \varepsilon &= 0.4\upsilon_1 + 0.35\upsilon_2, \end{aligned}$$

where υ_1 and υ_2 are two uniform random variables from $\mathbf{U}[-0.5, 0.5]$. The linear combinations above can make $R - ER$ and ε correlated. The simulation results are presented in Table 3.1 and Table 3.2. We choose different values for a, b, c, d to achieve different censoring and truncation percentages.

When sample size is large ($n = 200$), censoring percentage is low (about 20%) and truncation probability is high (about 0.85), the simulation results indicate that the biases are 0.001 for all parameters, which are very small. In this case, the mean standard deviation estimate and standard deviation for Monte Carlo estimates are very close. Even when the data are severely biased (80% censoring and 0.15 truncation probability), the biases for the predictor parameters $(\beta_2, \beta_3, \beta_4)$ are still very small, $(0.001, 0.010, 0.004)$ respectively, although the bias for intercept β_1 is larger. Therefore we can conclude that the proposed estimators work well for large sample sizes.

Note that when sample size is $n = 100$, censoring percentage is about 80% and truncation probability is about 0.15, the estimate for the intercept has the largest biases 0.064 and the estimates for other parameters are still good. In this case the standard error based on Monte Carlo simulations $\hat{s}_\beta$ and the mean of estimated

Table 3.1: Scenario 1. (e) estimate (bias in parenthesis).

	$n = 200$		c% = 20%			c% = 50%			c% = 80%		
	$\gamma =$	$\boldsymbol{\beta}^*$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$
(1)	0.85	3.700	3.699(0.001)	0.115	0.112	3.717(0.017)	0.160	0.156	3.749(0.049)	0.269	0.246
	0.85	-0.050	-0.051(0.001)	0.004	0.004	-0.051(0.001)	0.006	0.006	-0.051(0.001)	0.010	0.009
	0.85	-0.300	-0.301(0.001)	0.024	0.025	-0.305(0.005)	0.035	0.035	-0.304(0.004)	0.068	0.059
	0.85	-0.100	-0.099(0.001)	0.025	0.025	-0.103(0.003)	0.035	0.034	-0.102(0.002)	0.058	0.055
(2)	0.5	3.700	3.708(0.008)	0.126	0.131	3.748(0.048)	0.176	0.171	3.741(0.041)	0.317	0.272
	0.5	-0.050	-0.050(0.000)	0.005	0.005	-0.051(0.001)	0.007	0.007	-0.051(0.001)	0.012	0.010
	0.5	-0.300	-0.305(0.005)	0.029	0.028	-0.304(0.004)	0.038	0.037	-0.308(0.008)	0.068	0.058
	0.5	-0.100	-0.100(0.000)	0.029	0.028	-0.101(0.001)	0.036	0.035	-0.103(0.003)	0.058	0.055
(3)	0.15	3.700	3.649(0.051)	0.217	0.183	3.755(0.055)	0.255	0.198	3.759(0.059)	0.363	0.278
	0.15	-0.050	-0.049(0.001)	0.008	0.007	-0.050(0.001)	0.010	0.008	-0.051(0.001)	0.014	0.010
	0.15	-0.300	-0.287(0.013)	0.054	0.043	-0.298(0.002)	0.056	0.044	-0.290(0.010)	0.080	0.058
	0.15	-0.100	-0.099(0.001)	0.058	0.043	-0.102(0.002)	0.056	0.043	-0.096(0.004)	0.076	0.055

n: observed sample size; γ: truncation probability; $c\%$: censoring percentage; $\hat{s}_\beta$: means of standard deviation estimates obtained by (3.38); $\hat{\sigma}_\beta$: the standard deviation for $\hat{\beta}$ and $\hat{\beta}$ based on the 500 simulations. (1) $c = d = 1.1$ corresponding to truncation probability $\gamma \approx 0.85$ and $a = b = 1.8$, $a = b = 1.0$ and $a = b = 0.65$ approximately corresponding to 20% censoring, 50% censoring and 80% censoring, respectively; (2) $c = d = 0.45$ corresponding to $\gamma \approx 0.5$ and $a = b = 1.3$, $a = b = 0.8$ and $a = b = 0.55$ approximately corresponding to 20% censoring, 50% censoring and 80% censoring, respectively; (3) $c = d = 0.2$ corresponding to $\gamma \approx 0.15$ and $a = b = 1.1$, $a = b = 0.7$ and $a = b = 0.45$ approximately corresponding to 20% censoring, 50% censoring and 80% censoring, respectively.

Table 3.2: Scenario 1. (e) estimate (bias in parenthesis).

	$n = 100$		c% = 20%			c% = 50%			c% = 80%		
	$\gamma =$	$\boldsymbol{\beta}^*$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$
(1)	0.85	3.700	3.725(0.025)	0.162	0.161	3.726(0.026)	0.241	0.216	3.665(0.335)	0.446	0.325
	0.85	-0.050	-0.051(0.001)	0.006	0.006	-0.050(0.000)	0.009	0.008	-0.048(0.002)	0.017	0.012
	0.85	-0.300	-0.301(0.001)	0.039	0.037	-0.302(0.002)	0.055	0.051	-0.295(0.005)	0.102	0.073
	0.85	-0.100	-0.100(0.000)	0.037	0.035	-0.100(0.000)	0.049	0.048	-0.107(0.007)	0.094	0.071
(2)	0.5	3.700	3.710(0.010)	0.180	0.180	3.664(0.036)	0.250	0.229	3.726(0.026)	0.429	0.338
	0.5	-0.050	-0.051(0.001)	0.007	0.007	-0.052(0.002)	0.009	0.009	-0.051(0.001)	0.015	0.012
	0.5	-0.300	-0.304(0.004)	0.041	0.040	-0.308(0.008)	0.053	0.050	-0.292(0.008)	0.099	0.069
	0.5	-0.100	-0.098(0.002)	0.040	0.040	-0.102(0.002)	0.053	0.049	-0.101(0.001)	0.091	0.069
(3)	0.15	3.700	3.638(0.062)	0.266	0.202	3.655(0.045)	0.320	0.222	3.764(0.064)	0.528	0.343
	0.15	-0.050	-0.049(0.001)	0.010	0.008	-0.048(0.002)	0.012	0.009	-0.049(0.001)	0.020	0.013
	0.15	-0.300	-0.293(0.007)	0.063	0.048	-0.295(0.005)	0.067	0.050	-0.285(0.015)	0.109	0.067
	0.15	-0.100	-0.099(0.001)	0.067	0.047	-0.091(0.009)	0.069	0.051	-0.094(0.006)	0.097	0.062

n: observed sample size; γ: truncation probability; $c\%$: censoring percentage; $\hat{s}_\beta$: means of standard deviation estimates obtained by (3.38); $\hat{\sigma}_\beta$: the standard deviation for $\hat{\beta}$ and $\hat{\beta}$ based on the 500 simulations. (1) $c = d = 1.1$ corresponding to truncation probability $\gamma \approx 0.85$ and $a = b = 1.8$, $a = b = 1.0$ and $a = b = 0.65$ approximately corresponding to 20% censoring, 50% censoring and 80% censoring, respectively; (2) $c = d = 0.45$ corresponding to $\gamma \approx 0.5$ and $a = b = 1.3$, $a = b = 0.8$ and $a = b = 0.55$ approximately corresponding to 20% censoring, 50% censoring and 80% censoring, respectively; (3) $c = d = 0.2$ corresponding to $\gamma \approx 0.15$ and $a = b = 1.1$, $a = b = 0.7$ and $a = b = 0.45$ approximately corresponding to 20% censoring, 50% censoring and 80% censoring, respectively.

standard errors $\hat{\sigma}_\beta$ are not close. This is not surprising as the observed sample is severely biased (truncation probability is only about 0.15) and among the observed samples 80% are censored, i.e. only about 20 observations are fully observed.

Scenario 2: Instead of generating the error terms from linear combinations, we generate $R - ER$ and ε as follows. First generate (ω_1, ω_2) from the well-known bivariate parametric model in Clayton (1978), which has joint survival function $S_\epsilon(s_1, s_2) = (S_{\epsilon 1}(s_1)^{-\phi} + S_{\epsilon 2}(s_2)^{-\phi} - 1)^{-1/\phi}$. We take $\phi = 4$ (see for example Prentice et al. (2004)). The marginal survival functions $S_i(s_i)$ are from unit exponential distribution truncated at 1 and its mean is $E = (1 - 2e^{-1})/(1 - e^{-1})$. Then we let $R = ER + \omega_1 - E$ and $\varepsilon = \omega_2 - E$. We can also choose different values for a, b, c, d to achieve different censoring and truncation percentages. The simulation results are summarized in Table 3.3 and Table 3.4. We have similar findings and conclusions as for Scenario 1.

The boundary conditions in 3.1.1 are satisfied in both scenarios. Note that (R, T) has a finite upper boundary for their density support but (L, C) are from linear combinations of exponential distributions and can take values up to ∞. This guarantees that there is always a positive probability to observe all possible values for (R, T), which makes the model identifiable.

On the other hand, if the boundary condition is not satisfied, the estimates could be biased. For example, we consider **Scenario 3** by drawing L and C from $C = a\nu_1 + b\nu_2, \;\; L = c\nu_1 + d\nu_2 + U[0, 1]$, where ν_1 and ν_2 are from unit exponential distribution truncated at 2. The two error terms $R - ER$ and ε are generated as $R - ER = 0.5\upsilon_1 + 0.5\upsilon_2, \;\; \varepsilon = 0.4\upsilon_1 + 0.4\upsilon_2$, where $ER = 0.5$, υ_1 and υ_2 are from Normal$(0, 1.0)$. In this scenario, the upper boundary for the support of the density of L is $2(c+d)+1$. For all (R, T) with $R > 2(c+d)+1$ can never be observed. Also

Table 3.3: Scenario 2. (e) estimate (bias in parenthesis).

	$n = 200$		c% = 25%			c% = 50%			c% = 75%		
	$\gamma =$	$\boldsymbol{\beta}^*$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$
(1)	0.8	3.700	3.665(0.005)	0.220	0.208	3.621(0.079)	0.307	0.286	3.573(0.137)	0.427	0.377
	0.8	-0.050	-0.049(0.001)	0.008	0.008	-0.049(0.001)	0.012	0.011	-0.048(0.002)	0.016	0.015
	0.8	-0.300	-0.298(0.002)	0.048	0.047	-0.289(0.011)	0.062	0.061	-0.279(0.021)	0.096	0.087
	0.8	-0.100	-0.099(0.001)	0.047	0.046	-0.102(0.002)	0.063	0.061	-0.093(0.007)	0.095	0.086
(2)	0.5	3.700	3.646(0.054)	0.218	0.212	3.637(0.063)	0.262	0.248	3.561(0.139)	0.430	0.382
	0.5	-0.050	-0.050(0.000)	0.008	0.008	-0.049(0.001)	0.010	0.010	-0.048(0.002)	0.016	0.015
	0.5	-0.300	-0.290(0.010)	0.047	0.048	-0.296(0.004)	0.060	0.060	-0.274(0.026)	0.010	0.090
	0.5	-0.100	-0.096(0.004)	0.049	0.048	-0.105(0.005)	0.060	0.058	-0.095(0.005)	0.095	0.088
(3)	0.25	3.700	3.590(0.110)	0.319	0.296	3.547(0.153)	0.346	0.316	3.514(0.186)	0.441	0.391
	0.25	-0.050	-0.048(0.002)	0.013	0.012	-0.047(0.003)	0.014	0.013	-0.046(0.004)	0.017	0.015
	0.25	-0.300	-0.289(0.011)	0.070	0.068	-0.278(0.022)	0.080	0.076	-0.265(0.035)	0.111	0.089
	0.25	-0.100	-0.093(0.007)	0.073	0.068	-0.094(0.006)	0.082	0.074	-0.098(0.002)	0.105	0.089

n: observed sample size; γ: truncation probability; $\%c$: censoring percentage; $\hat{s}_\beta$: means of standard deviation estimates obtained by (3.38); $\hat{\sigma}_\beta$: the standard deviation for $\hat{\beta}$ and $\hat{\beta}$ based on the 500 simulations. (1) $c = d = 0.9$ corresponding to truncation probability $\gamma \approx 0.8$ and $a = b = 1.5$, $a = b = 1.0$ and $a = b = 0.7$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively; (2) $c = d = 0.5$ corresponding to $\gamma \approx 0.5$ and $a = b = 1.2$, $a = b = 0.9$ and $a = b = 0.6$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively; (3) $c = d = 0.2$ corresponding to $\gamma \approx 0.25$ and $a = b = 1.0$, $a = b = 0.7$ and $a = b = 0.5$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively.

Table 3.4: Scenario 2. (e) estimate (bias inparenthesis).

	$n = 100$		c% = 25%			c% = 50%			c% = 75%		
	$\gamma =$	$\boldsymbol{\beta}^*$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$
(1)	0.8	3.700	3.640(0.060)	0.296	0.295	3.603(0.097)	0.404	0.353	3.480(0.220)	0.538	0.472
	0.8	-0.050	-0.049(0.001)	0.011	0.011	-0.048(0.002)	0.016	0.014	-0.045(0.005)	0.022	0.019
	0.8	-0.300	-0.288(0.012)	0.066	0.066	-0.293(0.007)	0.089	0.085	-0.281(0.019)	0.140	0.112
	0.8	-0.100	-0.098(0.002)	0.064	0.066	-0.095(0.005)	0.089	0.085	-0.081(0.019)	0.133	0.112
(2)	0.5	3.700	3.627(0.073)	0.376	0.355	3.613(0.087)	0.397	0.359	3.496(0.204)	0.635	0.500
	0.5	-0.050	-0.049(0.001)	0.014	0.014	-0.048(0.002)	0.016	0.014	-0.046(0.004)	0.024	0.019
	0.5	-0.300	-0.283(0.017)	0.072	0.071	-0.293(0.007)	0.084	0.083	-0.271(0.029)	0.137	0.113
	0.5	-0.100	-0.098(0.001)	0.072	0.067	-0.099(0.001)	0.086	0.084	-0.082(0.018)	0.139	0.111
(3)	0.25	3.700	3.512(0.188)	0.384	0.326	3.494(0.206)	0.480	0.412	3.434(0.266)	0.647	0.537
	0.25	-0.050	-0.046(0.004)	0.015	0.013	-0.045(0.005)	0.018	0.016	-0.044(0.006)	0.025	0.021
	0.25	-0.300	-0.265(0.035)	0.089	0.079	-0.267(0.033)	0.109	0.089	-0.260(0.040)	0.141	0.108
	0.25	-0.100	-0.095(0.005)	0.093	0.081	-0.093(0.007)	0.105	0.090	-0.078(0.022)	0.136	0.110

n: observed sample size; γ: truncation probability; $\%c$: censoring percentage; $\hat{s}_\beta$: means of standard deviation estimates obtained by (3.38); $\hat{\sigma}_\beta$: the standard deviation for $\hat{\beta}$ and $\hat{\beta}$ based on the 500 simulations. (1) $c = d = 0.9$ corresponding to truncation probability $\gamma \approx 0.8$ and $a = b = 1.5$, $a = b = 1.0$ and $a = b = 0.7$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively; (2) $c = d = 0.5$ corresponding to $\gamma \approx 0.5$ and $a = b = 1.2$, $a = b = 0.9$ and $a = b = 0.6$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively; (3) $c = d = 0.2$ corresponding to $\gamma \approx 0.25$ and $a = b = 1.0$, $a = b = 0.7$ and $a = b = 0.5$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively.

larger value of T will always be censored. Therefore, if $c + d$ is small the estimates will be severely biased. The results are shown in Table 3.5 and Table 3.6.

From the results we can see that the intercept is estimated much smaller than the true value and the predictor parameter estimates are much less significant than that in Scenarios 1 and 2. This is because large values of R can never be observed and large values of T will always be censored. This will result in a much smaller intercept estimate and less significant predictor parameter estimates.

3.6.2 Comparison with KSV estimator

To see how well our proposed estimators perform when handle censored survival data with referral bias, we apply the KSV method to the same simulated datasets in Scenario 1 in section 3.6.1.

The KSV estimator $\hat{\boldsymbol{\beta}}_{\text{KSV}}$ are calculated by

$$\hat{\boldsymbol{\beta}}_{\text{KSV}} = \left\{(\boldsymbol{W} - \bar{\boldsymbol{W}})^{tr}(\boldsymbol{W} - \bar{\boldsymbol{W}})\right\}^{-1}(\boldsymbol{W} - \bar{\boldsymbol{W}})^{tr}\tilde{\boldsymbol{X}}, \tag{3.39}$$

where $\bar{\boldsymbol{W}} = \left(\left(n^{-1}\sum_{i=1}^{n} W_{ij}\right)\right)_{n\times p}$, $\tilde{\boldsymbol{X}} = \left(\tilde{X}_1, \ldots, \tilde{X}_n\right)^{tr}$, and

$$\tilde{X}_i = \frac{\delta_i X_i}{\hat{G}(R_i-, X_i-)}, i = 1, \ldots, n. \tag{3.40}$$

The function $\hat{G}(R_i-, X_i-)$ here is the bivariate survival function estimated by (3.16), and it is exactly the same as what we adopted in Lemma 3.2.1 to get our proposed estimates of $\boldsymbol{\beta}$. The results of 500 simulations are shown in Table 3.7 and Table 3.8.

Here we present the simulation results for different sample sizes ($n = 100$ and 200), different truncation probabilities ($\gamma = 0.85, 0.5$ and 0.15) and different cen-

Table 3.5: Scenario 3. (e) estimate (bias in parenthesis).

	$n = 200$		c% = 25%			c% = 50%			c% = 75%		
	$\gamma =$	$\boldsymbol{\beta}^*$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$
(1)	0.85	3.700	3.409(0.291)	0.520	0.512	3.132(0.568)	0.723	0.642	2.521(1.179)	0.971	0.745
	0.85	-0.050	-0.042(0.008)	0.020	0.020	-0.033(0.017)	0.027	0.025	-0.019(0.031)	0.038	0.029
	0.85	-0.300	-0.253(0.047)	0.126	0.117	-0.202(0.098)	0.155	0.137	-0.071(0.229)	0.245	0.180
	0.85	-0.100	-0.082(0.018)	0.120	0.117	-0.076(0.024)	0.150	0.138	-0.064(0.036)	0.234	0.177
(2)	0.5	3.700	2.878(0.822)	1.270	0.845	2.833(0.867)	1.400	0.970	2.272(1.418)	1.683	1.140
	0.5	-0.050	-0.041(0.009)	0.049	0.033	-0.043(0.007)	0.058	0.037	-0.024(0.026)	0.065	0.044
	0.5	-0.300	-0.266(0.034)	0.295	0.203	-0.289(0.011)	0.348	0.214	-0.175(0.125)	0.373	0.239
	0.5	-0.100	-0.103(0.003)	0.281	0.204	-0.108(0.008)	0.325	0.213	-0.066(0.034)	0.368	0.246
(3)	0.15	3.700	2.415(1.285)	1.636	1.375	2.301(1.399)	1.897	1.680	1.887(1.813)	2.104	1.953
	0.15	-0.050	-0.024(0.026)	0.073	0.057	-0.026(0.024)	0.091	0.067	-0.019(0.031)	0.125	0.089
	0.15	-0.300	-0.168(0.032)	0.414	0.358	-0.137(0.163)	0.629	0.392	-0.092(0.208)	0.730	0.629
	0.15	-0.100	-0.113(0.013)	0.402	0.333	-0.121(0.021)	0.597	0.389	-0.083(0.017)	0.764	0.702

n: observed sample size; γ: truncation probability; $\%c$: censoring percentage; $\hat{s}_\beta$: means of standard deviation estimates obtained by (3.38); $\hat{\sigma}_\beta$: the standard deviation for $\hat{\beta}$ and $\hat{\beta}$ based on the 500 simulations. (1) $c = d = 0.8$ corresponding to $\gamma \approx 0.85$ and $a = b = 1.9$, $a = b = 1.4$ and $a = b = 1.0$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively; (2) $c = d = 0.35$ corresponding to $\gamma \approx 0.5$ and $a = b = 1.8$, $a = b = 1.2$ and $a = b = 0.85$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively; (3) $c = d = 0.1$ corresponding to $\gamma \approx 0.15$ and $a = b = 1.6$, $a = b = 0.9$ and $a = b = 0.5$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively.

Table 3.6: Scenario 3. (e) estimate (bias in parenthesis).

	$n = 100$		c% = 25%			c% = 50%			c% = 75%		
	$\gamma =$	$\boldsymbol{\beta}^*$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$
(1)	0.85	3.700	3.379(0.321)	0.517	0.533	3.079(0.621)	0.830	0.737	2.498(1.202)	1.171	1.089
	0.85	-0.050	-0.040(0.010)	0.031	0.030	-0.030(0.020)	0.037	0.038	-0.013(0.037)	0.054	0.047
	0.85	-0.300	-0.247(0.053)	0.158	0.147	-0.197(0.103)	0.174	0.167	-0.068(0.232)	0.276	0.235
	0.85	-0.100	-0.078(0.022)	0.140	0.137	-0.072(0.028)	0.163	0.160	-0.061(0.039)	0.253	0.231
(2)	0.5	3.700	2.798(0.902)	1.281	1.095	2.814(0.886)	1.484	1.203	2.158(1.542)	1.698	1.347
	0.5	-0.050	-0.038(0.012)	0.057	0.050	-0.037(0.013)	0.069	0.053	-0.021(0.029)	0.073	0.061
	0.5	-0.300	-0.257(0.043)	0.312	0.261	-0.267(0.033)	0.363	0.309	-0.163(0.137)	0.397	0.334
	0.5	-0.100	-0.104(0.004)	0.294	0.247	-0.103(0.003)	0.352	0.278	-0.062(0.038)	0.385	0.326
(3)	0.15	3.700	2.408(1.292)	1.700	1.517	2.314(1.386)	1.998	1.693	1.536(2.164)	2.325	1.989
	0.15	-0.050	-0.021(0.029)	0.087	0.070	-0.021(0.029)	0.103	0.082	-0.012(0.038)	0.138	0.102
	0.15	-0.300	-0.163(0.137)	0.493	0.413	-0.135(0.165)	0.695	0.455	-0.087(0.213)	0.792	0.687
	0.15	-0.100	-0.115(0.015)	0.475	0.402	-0.117(0.017)	0.601	0.428	-0.063(0.037)	0.784	0.601

n: observed sample size; γ: truncation probability; $\%c$: censoring percentage; $\hat{s}_\beta$: means of standard deviation estimates obtained by (3.38); $\hat{\sigma}_\beta$: the standard deviation for $\hat{\beta}$ and $\hat{\beta}$ based on the 500 simulations. (1) $c = d = 0.8$ corresponding to $\gamma \approx 0.85$ and $a = b = 1.9$, $a = b = 1.4$ and $a = b = 1.0$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively; (2) $c = d = 0.35$ corresponding to $\gamma \approx 0.5$ and $a = b = 1.8$, $a = b = 1.2$ and $a = b = 0.85$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively; (3) $c = d = 0.1$ corresponding to $\gamma \approx 0.15$ and $a = b = 1.6$, $a = b = 0.9$ and $a = b = 0.5$ approximately corresponding to 25% censoring, 50% censoring and 75% censoring, respectively.

Table 3.7: Simulation results for KSV estimator (bias in parenthesis).

	$n = 200$		c% = 20%		c% = 50%	
	$\gamma =$	$\boldsymbol{\beta}^*$	$\hat{\boldsymbol{\beta}}_{\text{KSV}}$	$\hat{\sigma}_{\beta_{\text{KSV}}}$	$\hat{\boldsymbol{\beta}}_{\text{KSV}}$	$\hat{\sigma}_{\beta_{\text{KSV}}}$
(1)	0.85	3.700	2.524(1.176)	0.567	0.638(3.062)	0.714
	0.85	-0.050	-0.022(0.028)	0.021	0.015(0.065)	0.028
	0.85	-0.300	-0.134(0.166)	0.124	0.080(0.380)	0.160
	0.85	-0.100	-0.038(0.062)	0.130	0.025(0.125)	0.154
(2)	0.5	3.700	2.515(1.185)	0.586	0.537(3.163)	0.494
	0.5	-0.050	-0.022(0.028)	0.022	0.022(0.072)	0.028
	0.5	-0.300	-0.138(0.162)	0.132	0.144(0.444)	0.158
	0.5	-0.100	-0.044(0.056)	0.134	0.049(0.149)	0.149
(3)	0.15	3.700	2.387(1.313)	0.544	0.224(3.476)	0.685
	0.15	-0.050	-0.021(0.029)	0.021	0.027(0.077)	0.027
	0.15	-0.300	-0.139(0.161)	0.126	0.158(0.458)	0.154
	0.15	-0.100	-0.047(0.053)	0.128	0.055(0.155)	0.164

n: observed sample size; γ: truncation probability; $\%c$: censoring percentage; $\hat{\sigma}_{\beta_{\text{KSV}}}$: the standard deviation for $\hat{\boldsymbol{\beta}}_{\text{KSV}}$ and $\hat{\boldsymbol{\beta}}_{\text{KSV}}$ based on the 500 simulations.
(1) $c = d = 1.6$ corresponding to $\gamma \approx 0.85$ and $a = b = 2.7$, $a = b = 1.5$ approximately corresponding to 20% censoring and 50% censoring, respectively;
(2) $c = d = 0.7$ corresponding to $\gamma \approx 0.5$ and $a = b = 2.0$, $a = b = 1.3$ approximately corresponding to 20% censoring and 50% censoring, respectively;
(3) $c = d = 0.3$ corresponding to $\gamma \approx 0.2$ and $a = b = 1.8$, $a = b = 1.1$ approximately corresponding to 20% censoring and 50% censoring, respectively.

soring percentages (c%=20% and 50%). The case for heavy censoring percentage (c%=80%) is not included since the KSV method already produces false results of the predictor parameter estimates when the censoring percentage $c\% = 50\%$. Comparing with the simulation results shown in Table 3.1 and Table 3.2, we can see that the intercept is estimated much smaller than the true value and none of the predictor parameter estimates are statistically significant. This is because only the responses are weighted by the bivariate distribution function. However, our proposed method re-weights both the covariates and the response.

Table 3.8: Simulation results for KSV estimator. (e) estimate (bias in parenthesis).

	$n = 100$		c% = 20%		c% = 50%	
	$\gamma =$	$\boldsymbol{\beta}^*$	$\hat{\boldsymbol{\beta}}_{\text{KSV}}$	$\hat{\sigma}_{\beta_{\text{KSV}}}$	$\hat{\boldsymbol{\beta}}_{\text{KSV}}$	$\hat{\sigma}_{\beta_{\text{KSV}}}$
(1)	0.85	3.700	2.485(1.215)	0.777	0.659(3.041)	1.019
	0.85	-0.050	-0.021(0.029)	0.03	0.014(0.064)	0.040
	0.85	-0.300	-0.135(0.165)	0.166	0.083(0.383)	0.234
	0.85	-0.100	-0.030(0.070)	0.170	0.034(0.134)	0.225
(2)	0.5	3.700	2.468(1.232)	0.801	0.494(3.206)	1.044
	0.5	-0.050	-0.020(0.030)	0.030	0.024(0.074)	0.041
	0.5	-0.300	-0.140(0.160)	0.166	0.141(0.441)	0.228
	0.5	-0.100	-0.042(0.058)	0.172	0.047(0.147)	0.210
(3)	0.15	3.700	2.376(1.324)	0.822	0.173(3.527)	1.012
	0.15	-0.050	-0.020(0.030)	0.031	0.029(0.079)	0.040
	0.15	-0.300	-0.131(0.169)	0.184	0.177(0.477)	0.243
	0.15	-0.100	-0.041(0.059)	0.176	0.069(0.141)	0.228

n: observed sample size; γ: truncation probability; $\%c$: censoring percentage; $\hat{\sigma}_{\beta_{\text{KSV}}}$: the standard deviation for $\hat{\boldsymbol{\beta}}_{\text{KSV}}$ and $\hat{\boldsymbol{\beta}}_{\text{KSV}}$ based on the 500 simulations.
(1) $c = d = 1.6$ corresponding to $\gamma \approx 0.85$ and $a = b = 2.7$, $a = b = 1.5$ approximately corresponding to 20% censoring and 50% censoring, respectively;
(2) $c = d = 0.7$ corresponding to $\gamma \approx 0.5$ and $a = b = 2.0$, $a = b = 1.3$ approximately corresponding to 20% censoring and 50% censoring, respectively;
(3) $c = d = 0.3$ corresponding to $\gamma \approx 0.2$ and $a = b = 1.8$, $a = b = 1.1$ approximately corresponding to 20% censoring and 50% censoring, respectively.

Table 3.9: Estimation results (SE in parenthesis), * means the estimate is significant at 5% level.

	With truncation – $\hat{\boldsymbol{\beta}}$ (SE)	Without truncation – $\hat{\boldsymbol{\beta}}$ (SE)
Intercept	3.628* (0.068)	3.844* (0.133)
Age	-0.011* (0.003)	-0.031* (0.004)
HIV	-0.313* (0.047)	-0.380* (0.089)
Alcohol	-0.098* (0.038)	-0.077 (0.070)

3.6.3 Data analysis

We illustrate the proposed method with the Edinburgh hepatitis C data in Fu et al. (2007). The patients were studied retrospectively and followed prospectively for the development of HCV-related cirrhosis. HCV patients usually experience no symptoms or mild symptoms in the early stages and are often referred to hospital shortly before they develop cirrhosis or complications. Among these individuals, there is no cirrhosis event occurred prior to their referral time and 63 (16%) developed cirrhosis during follow-up. The median duration time from HCV infection to referral is 17.1 years and the median follow-up time from referral to cirrhosis or censoring is 2.4 years.

The purpose of the study was to determine how the progression to cirrhosis is affected by three covariates: age at infection, HIV co-infection (yes:1 or no:0), and heavy alcohol consumption (yes:1 or no:0). An individual with heavy alcohol intake was defined as one consuming more than 50 units alcohol per week for at least five years.

Table 3.9 summarizes the estimates of regression parameters obtained from our method. The results from the truncated model, where the referral bias is considered, show that age at infection, HIV co-infection and heavy alcohol in-take are significantly identified as risk factors associated with more rapid disease progression. If

we compare the results with those from a non-truncated model, ignorance of the referral bias has failed to identify heavy alcohol in-take as a significant risk factor. In medical literatures, older age at infection, HIV co-infection and heavy alcohol intake have all been identified as factors associated with more rapid hepatitis C disease progression (Sharma and Sherker, 2009).

In Table 3.9, although the proposed methods do make a difference in terms of significance of the covariate "Alcohol", the resulting estimated value and the standard error by the two models seem to be very similar. This is because in the Edinburgh hepatitis C data in Fu et al. (2007), 387 patients are observed. Among these individuals, only 63 (16%) developed cirrhosis during follow-up. Therefore the censoring percentage is around 85%, which is quite high. The inverse probability weighted (IPW) estimator of the bivariate survival function G used in our WLS method does struggle when the censoring percentage is high (Dai and Bao, 2009; Dai and Fu, 2012). Therefore the insignificance of the improvement when analysing the severely biased survival data (with high censoring percentage) is reasonable.

Based on our coefficient estimates shown in Table 3.9, we use $\hat{T} = \exp(\boldsymbol{W}\boldsymbol{\beta})$ to predict the time period from infection to cirrhosis for individuals with different values of covariates. The range of age at infection is taken to be from 10 to 70, since in the Edinburgh hepatitis C data only 5 of 387 patients were infected by HCV before 10 years old. Therefore the effects estimates in Table 3.9 may not be very appropriate to be used to predict the cirrhosis time for these patients. The prediction results and the corresponding values of covariates are shown in Figure 3.2. We can see that if we consider referral bias, when the age at infection is greater than 10, the predicted values of X are larger than those without considering referral bias. This is because that the patients with more rapid disease progression are preferentially referred to

the liver clinic cohort. Hence the time period from infection to cirrhosis observed in the clinic cohort may be shorter than that for the whole HCV patients community (Wang et al., 2013). If so, removing the referral bias, the predicted values should be larger comparing with the case only censoring is involved, as revealed by Fu et al. (2007).

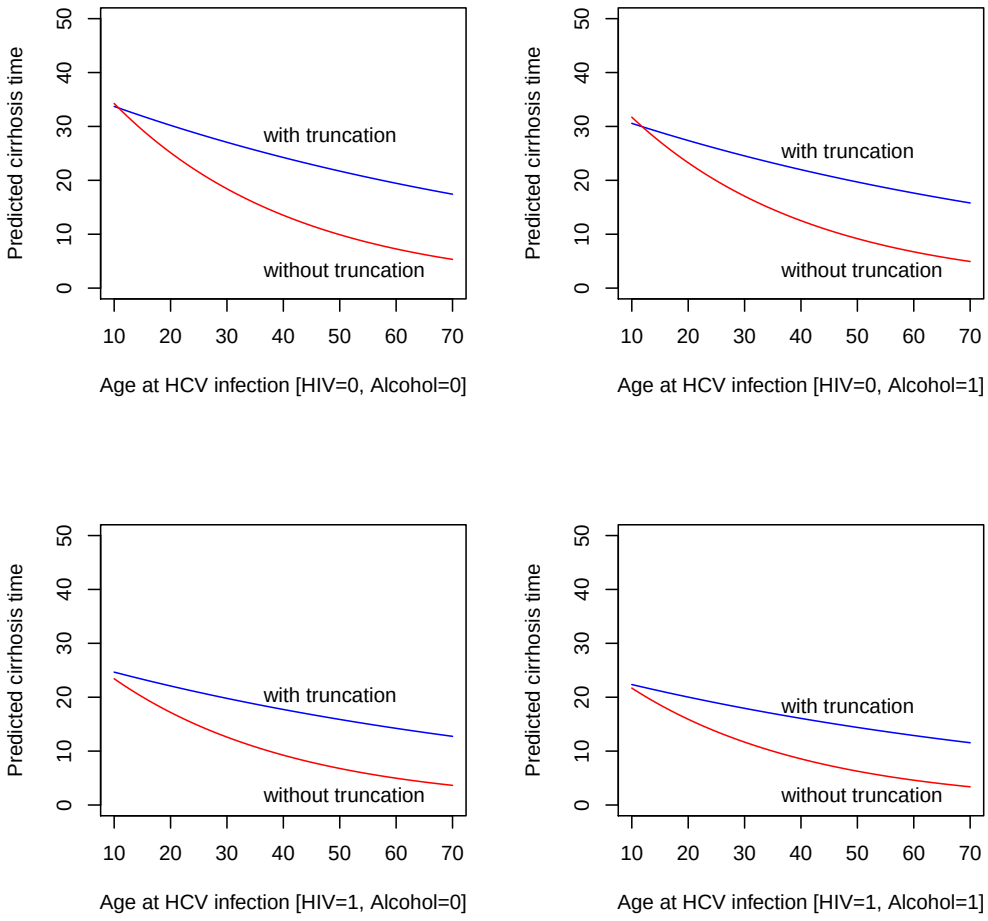

Figure 3.2: Prediction for cirrhosis time

3.7 Goodness of fit

To assess the validity of our linear model assumption, we consider to use the Chi-square goodness-of-fit test, where the null hypothesis is that the AFT model assumption is true. We follow Kim (1993)'s idea to use a revised Chi-square statistic for the goodness-of-fit test for models with incomplete data. To calculate the statistic, we partition the support of response variable, the logarithm of the observed cirrhosis time X, into $k+1$ subintervals and then work out the observed number of observations O_i, and the expected number of observations for each cell E_i, where $i = 1, \ldots, k+1$.

Let F^{KM} be the KM estimator based on the observed cirrhosis time X and the censoring indicator, and let $\tilde{F}^{\mathrm{KM}}$ be the KM estimator based on the error term $\varepsilon = X - \boldsymbol{W}\hat{\boldsymbol{\beta}}$. Then under the null hypothesis we calculate the Chi-square statistic

$$\sum_{i=1}^{k+1} \frac{(O_i - E_i)^2}{\text{variance}}, \tag{3.41}$$

where

$$O_i = N \cdot \left(F_i^{\mathrm{KM}} - F_{i-1}^{\mathrm{KM}}\right), \tag{3.42}$$

$$E_i = N \cdot \left(\tilde{F}_i^{\mathrm{KM}} - \tilde{F}_{i-1}^{\mathrm{KM}}\right), \tag{3.43}$$

and N is the sample size which is 387. The variance is calculated by

$$\text{variance} = \sum_i \int \frac{dF_i^{KM}}{(1 - H_i)(1 - F_i^{KM})}, \tag{3.44}$$

where $H(x)$ is the distribution function of $X = \min(T, C)$, and $1 - H(x)$ can be

estimated by the proportion that $X \geqslant x$.

Then if the AFT model assumption is true, the revised statistic in (3.41) should asymptotically follow a Chi-square distribution with degree of freedom $k - s - 1$, where s is the number of parameters to be estimated. Here s is taken to be 4 since we have three covariates and one intercept.

Following this method, we find that the revised statistic is 11.68 with $k = 12$ (d.f.=7). This statistic is less than the threshold which is 14.07. Therefore we do not have significant evidence to reject the null hypothesis. Thus we can conclude that the AFT model assumption is reasonable here. Note that this heuristic method needs further development since its theoretical proofs have not been done.

3.8 Conclusion and discussion

In this chapter, we considered the accelerated failure time model $\log T = \boldsymbol{W}\boldsymbol{\beta} + \varepsilon$ to study the relation between the cirrhosis time T and the covariates vector $\boldsymbol{W}$. A nonparametric method was proposed to allow a flexible bivariate distribution structure between T and another event time R. The dependency of the two event times and the referral bias in the sampling, incorporated by a right truncation on R, were both taken into account. We applied our proposed method to analyse the Edinburgh hepatitis C data in Fu et al. (2007). We concluded that when referral bias is considered, older age at HCV infection, HIV co-infection and heavy alcohol intake can all be identified as risk factors associated with more rapid HCV disease progression. Furthermore, the prediction results indicated that considering the referral bias will lead to longer time period from HCV infection to cirrhosis, which could reach an agreement with Fu et al. (2007). Our estimating method can remove the referral

bias to get more reliable estimates for $\hat{\boldsymbol{\beta}}$ (Wang et al., 2013).

In the developed theory, the cirrhosis time T can be either less than or more than the referral time R although in the data analysis we only observed $T > R$ (refer to liver clinics before development of cirrhosis) for all patients. In practice, $T > R$ should stand for majority of HCV patients in a clinic cohort since HCV patients will have a lot of symptoms and then go to see a doctor before a sever cirrhosis event is developed. For quite a few patients, cirrhosis may occur before recruitment to a clinic cohort ($T < R$). The method proposed in this chapter can be extended to deal with a case when T is both left censored and right censored and R is truncated, by some independent variables. However, the limit of applicability of our method indeed arises in the case when T is doubly censored by R and C since (R, T) is a correlated pair. New methodologies should be developed to deal with the case that T is also left censored by R. Although our method cannot deal with such kind of cases, it does not limit the applicability of our methodology in the HCV disease progression study or in general bivariate survival analysis with random censoring and truncation. This is because: (1) in most similar HCV examples the pair of event times (T, R) are correlated but they are not usually censored or truncated by each other; and (2) our theory does not require either $T < R$ or $T \geqslant R$, which means our methods can be applied in other bivariate survival analysis studies. The limit of applicability of the method indeed arises in the case when the cirrhosis time T is doubly censored by the referral time R and the censoring time C, and (T, R) is a correlated pair of event times.

Moreover, in our developed theory, we assumed that the upper bound of the support region of T should be less than that of C, which can guarantee that the model (the distribution of T) is identifiable (for the tails). This is a common condition

in survival analysis theory. Without this condition, many existing nonparametric survival estimating methods (including the method proposed in this chapter) may not provide unbiased estimates for tails. It is not because these methods are wrong, but because the model is not identifiable at tails. Although existing methods and the method proposed in this chapter can still be used as usual without this condition, they will provide estimates which are unbiased up to the upper bound of the support region of C. Of course some research work in survival analysis may not need this condition due to that the tail of the distribution of T is not important or due to employment of parametric methods.

An alternative to the AFT models is the proportional hazard model, which can be written as a transformation model (Chen et al., 2002), $h(T) = -\boldsymbol{W}\boldsymbol{\beta} + \epsilon$, where h is a monotone and unknown transformation function and ϵ is the residual with specified distribution. The methods proposed in this chapter can be extended to proportional hazards models with time-independent covariates since the parameter estimation problem for the transformation model can also be solved using the inverse probability weighted method. In next chapter, we will focus on the transformation model to estimate the parameter $\boldsymbol{\beta}$ and study the proportional hazard model as a special case.

Chapter 4

Linear transformation models

The proportional hazards model (Cox model) is another commonly used model which can be used to explore the effects of covariates on event times (Cox, 1972, 1975). The large sample properties of the proportional hazards model have been demonstrated using martingale theory (Anderson and Gill, 1982). Moreover, practitioners have easy access to statistical software for this model. Therefore, there is a temptation to use the proportional hazards model to analyse failure time observations, even though the model does not fit the data well (Cheng et al., 1995).

The proportional hazards model is given by

$$\lambda(t|\boldsymbol{W}) = \lambda_0(t)\exp(\boldsymbol{W}\boldsymbol{\beta}), \tag{4.1}$$

where $\boldsymbol{W} = (W_1, \ldots, W_p)$ is a p-dimensional row vector of covariates, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{tr}$ is a p-dimensional column vector of regression coefficients, $\lambda(t|\boldsymbol{W})$ is the hazard rate function, and $\lambda_0(t)$ is the baseline hazard rate which denotes the conditional hazard rate given $\boldsymbol{W} = \boldsymbol{0}$. Let T be the failure time, and $S(t|\boldsymbol{W})$ be the survival function

of T given $\boldsymbol{W}$. Then the Cox model in (4.1) can be written as

$$\log\big\{-\log[S(t|\boldsymbol{W})]\big\} = h(t) + \boldsymbol{W}\boldsymbol{\beta}, \tag{4.2}$$

where $h(t) = \log\int \lambda_0(t)dt$ is a completely unspecified strictly increasing function, which maps the positive half-line onto the whole real line. The model in (4.2) can be naturally generalized to

$$g\{S(t|\boldsymbol{W})\} = h(t) + \boldsymbol{W}\boldsymbol{\beta}, \tag{4.3}$$

where g is a known continuous and strictly decreasing function. In Appendix D we show that (4.3) is equivalent to the linear transformation model

$$h(T) = -\boldsymbol{W}\boldsymbol{\beta} + \varepsilon, \tag{4.4}$$

where ε is a random error with distribution function $F(t) = 1 - \exp\{-\exp(t)\}$.

If $F(.)$ is the standard logistic distribution, (4.4) is the proportional odds model (Dabrowska and Doksum, 1988). Hence, the proportional hazards model and the proportional odds model are members of a well-known class of semi-parametric linear transformation models, under which an unknown transformation of the survival time is linearly related to the covariates with completely specified error distributions (Cheng et al., 1997). For the above reasons, we consider the linear transformation models in this chapter.

4.1 Existing methods of linear transformation models with censored data

The parametric version of the linear transformation model in (4.4), with $h(.)$ specified up to a finite-dimensional parameter vector, has been discussed extensively by Box and Cox (1964). For the case of $h(.)$ completely unspecified, Cuzick (1988) proposed methods for analysing failure time data with (4.4) and suggested a way to extend his estimator to the censored case. Many existing inference procedures for $\boldsymbol{\beta}$ in (4.4) are derived from some types of nonparametric likelihood function with a large number of nuisance parameters to be estimated, and can be rather complicated to implement in practice (Pettitt, 1982, 1984; Bennett, 1983; Clayton and Cuzick, 1986; Dabrowska and Doksum, 1989).

Cheng et al. (1995) proposed a class of simple estimating equations for $\boldsymbol{\beta}$ in the linear transformation model (4.4) with possibly censored observations. They also showed that the resulting estimators for $\boldsymbol{\beta}$ are consistent and asymptotically normal under rather mild conditions. However, their estimators are asymptotically biased when the support of the censoring variable is shorter than that of the survival time. Fine et al. (1998) proposed simple modifications of Cheng's procedures to obtain consistent estimators for the regression parameters $\boldsymbol{\beta}$. A key step in the above approaches is the estimation of the survival function for the censoring variable by the Kaplan-Meier estimator. Its validity relies on the assumption that the censoring variable is independent of the covariates. Thus, unlike Cox's partial likelihood approach (Cox, 1975), this method of estimation fails when this assumption is violated. A unified estimation procedure was proposed by Chen et al. (2002) for the analysis of censored data using linear transformation model, whose validity does not rely on

the assumption of independence between the censoring variable and the covariates.

4.2 Statistical modelling and estimating equation

Let (R_i, T_i) denote the pair of event times for the ith subject, where R_i is the time period from infection to referral, and T_i is the time period from infection to the development of cirrhosis. For a strictly increasing function $h(\cdot)$, the linear transformation model is of the form

$$h(T_i) = -\boldsymbol{W}_i\boldsymbol{\beta} + \varepsilon_i, \tag{4.5}$$

where ε_i are the error terms which have the completely specified distribution function F_ε.

The boundary supports of R, T, L, C are the same as those in Condition 3.1.1. In section 3.1.2 we have made the reasonable assumption that $(R, T, \boldsymbol{W})$ and (L, C) are independent throughout this thesis. Then similar estimating procedures as that proposed by Cheng et al. (1995) can be used here to estimate $\boldsymbol{\beta}$ in the linear transformation model (4.4).

Consider the variables $\big\{I[T_i \geqslant T_j], i = 1, \ldots, n, j = 1, \ldots, n, i \neq j\big\}$, where I denotes an indicator function. Then we have

$$\begin{aligned} E\big\{I[T_i \geqslant T_j]|\boldsymbol{W}_i, \boldsymbol{W}_j\big\} &= P\big\{h(T_i) \geqslant h(T_j)|\boldsymbol{W}_i, \boldsymbol{W}_j\big\} \\ &= P\big\{\varepsilon_i - \varepsilon_j \geqslant \boldsymbol{W}_{ij}\boldsymbol{\beta}\big\} \\ &= S(\boldsymbol{W}_{ij}\boldsymbol{\beta}), \end{aligned} \tag{4.6}$$

where $\boldsymbol{W}_{ij} = \boldsymbol{W}_i - \boldsymbol{W}_j$, $i = 1, \ldots, n$, $j = 1, \ldots, n$, and $i \neq j$. Let $\eta = \varepsilon_i - \varepsilon_j$. Then the survival function of η can be written as

$$S_\eta(t) = \int [1 - F_\varepsilon(t + s)] F_\varepsilon(ds). \tag{4.7}$$

In practice, due to censoring and truncation, the indicators $I[T_i \geqslant T_j]$ in (4.6) may not be observed. We only observe $I[X_i^* \geqslant X_j^*]$, where $X_i = \min(T_i, C_i)$, and $X_j = \min(T_j, C_j)$. Let $\gamma = P(R \leqslant L)$ denote the truncation probability. Then we have that

$$\begin{aligned}
&E\left\{ \frac{\delta_j^* I[X_i^* \geqslant X_j^*]}{G(R_i^*-, X_j^*-)G(R_j^*-, X_j^*-)} - \frac{S_\eta(\boldsymbol{W}_{ij}^*\boldsymbol{\beta})}{G(R_i^*-, 0)G(R_j^*-, 0)} \middle| \boldsymbol{W}_i^*, \boldsymbol{W}_j^* \right\} \\
=&\gamma^{-2} E\left\{ \frac{I[R_i \leqslant L_i, R_j \leqslant L_j] I[T_j \leqslant C_j] I[X_i \geqslant X_j]}{G(R_i-, X_j-)G(R_j-, X_j-)} - \frac{I[R_i \leqslant L_i, R_j \leqslant L_j] S_\eta(\boldsymbol{W}_{ij}^*\boldsymbol{\beta})}{G(R_i-, 0)G(R_j-, 0)} \middle| \boldsymbol{W}_i, \boldsymbol{W}_j \right\} \\
=&\gamma^{-2} E\Big\{ E\Big[I[T_i \geqslant T_j] - S_\eta(\boldsymbol{W}_{ij}\boldsymbol{\beta}) \big| T_i, T_j, \boldsymbol{W}_i, \boldsymbol{W}_j \Big] \Big\} \\
=&\gamma^{-2} E\Big\{ I[T_i \geqslant T_j] - S_\eta(\boldsymbol{W}_{ij}\boldsymbol{\beta}) \big| \boldsymbol{W}_i, \boldsymbol{W}_j \Big\} \\
=&0.
\end{aligned} \tag{4.8}$$

Therefore, the following unbiased estimating equation can be used to find out a consistent estimate for $\boldsymbol{\beta}$,

$$\begin{aligned}
\boldsymbol{Q}(\boldsymbol{\beta}; G) &= n^{-2} \sum_{i,j=1}^{n} \boldsymbol{W}_{ij}^* \left\{ \frac{\delta_j^* I[X_i^* \geqslant X_j^*]}{G(R_i^*-, X_j^*-)G(R_j^*-, X_j^*-)} - \frac{S_\eta(\boldsymbol{W}_{ij}^*\boldsymbol{\beta})}{G(R_i^*-, 0)G(R_j^*-, 0)} \right\} \\
&= \mathbf{0},
\end{aligned} \tag{4.9}$$

where $i = 1, \ldots, n$, $j = 1, \ldots, n$, and $i \neq j$. The estimating equation in (4.9)

requires that the bivariate survival function G is known. When G is unknown, we can replace it by the consistent estimator $\hat{G}$ which is proposed in Section 3.3, and then work out $\hat{\boldsymbol{\beta}}$ by solving the following estimating equation

$$\begin{aligned}\boldsymbol{Q}(\boldsymbol{\beta};\hat{G}) =& n^{-2}\sum_{i,j=1}^{n}\boldsymbol{W}_{ij}^{*}\delta_{j}^{*}I[X_{i}^{*}\geqslant X_{j}^{*}]\frac{I[\hat{G}(R_{i}^{*}-,X_{j}^{*}-)>0]\cdot I[\hat{G}(R_{j}^{*}-,X_{j}^{*}-)>0]}{\hat{G}(R_{i}^{*}-,X_{j}^{*}-)\hat{G}(R_{j}^{*}-,X_{j}^{*}-)}\\ &- n^{-2}\sum_{i,j=1}^{n}\boldsymbol{W}_{ij}^{*}S_{\eta}(\boldsymbol{W}_{ij}^{*}\boldsymbol{\beta})\frac{I[\hat{G}(R_{i}^{*}-,0)>0]\cdot I[\hat{G}(R_{j}^{*}-,0)>0]}{\hat{G}(R_{i}^{*}-,0)\hat{G}(R_{j}^{*}-,0)}\\ =&\mathbf{0}. \end{aligned} \tag{4.10}$$

4.3 Large sample properties of $\hat{\boldsymbol{\beta}}$

4.3.1 Consistency of $\hat{\boldsymbol{\beta}}$

By using the similar ideas in Section 3.5.1, we can show that the estimating equation in (4.10) has a unique solution $\hat{\boldsymbol{\beta}}$ (Cheng et al., 1995), which converges to the true parameter value $\boldsymbol{\beta}^{*}$ in probability as $n\to\infty$. This implies the consistency of $\hat{\boldsymbol{\beta}}$.

4.3.2 Asymptotic normality for $\hat{\boldsymbol{\beta}}$

Now we give a heuristic proof for the asymptotic normality of $\hat{\boldsymbol{\beta}}$. For $\boldsymbol{\beta}=\boldsymbol{\beta}^{*}$, the estimating equation in (4.10) can be written as

$$\begin{aligned}\sqrt{n}\big[\boldsymbol{Q}(\boldsymbol{\beta}^{*};\hat{G})\big] &= \sqrt{n}\big[\boldsymbol{Q}(\boldsymbol{\beta}^{*};G)\big]+\sqrt{n}\big[\boldsymbol{Q}(\boldsymbol{\beta}^{*};\hat{G})-\boldsymbol{Q}(\boldsymbol{\beta}^{*};G)\big]\\ &= I+II, \end{aligned} \tag{4.11}$$

where I is a sum of i.i.d. variables since

$$\sqrt{n}\big[\boldsymbol{Q}(\boldsymbol{\beta}^*;G)\big] = n^{-3/2}\sum_{i,j=1}^{n}\boldsymbol{W}_{ij}^*\left\{\frac{\delta_j^* I[X_i^* \geqslant X_j^*]}{G(R_i^*-,X_j^*-)G(R_j^*-,X_j^*-)} - \frac{S_\eta(\boldsymbol{W}_{ij}^*\boldsymbol{\beta}^*)}{G(R_i^*-,0)G(R_j^*-,0)}\right\}.$$

Now we focus on the second term II, which can be written as

$$\begin{aligned} II = {} & n^{-3/2}\sum_{i,j=1}^{n}\frac{\boldsymbol{W}_{ij}^*\delta_j^* I[X_i^* \geqslant X_j^*]}{G^{(1)}G^{(2)}}\cdot\left[\frac{G^{(1)}-\hat{G}^{(1)}}{G^{(1)}} + \frac{G^{(2)}-\hat{G}^{(2)}}{G^{(2)}}\right] \\ & - n^{-3/2}\sum_{i,j=1}^{n}\frac{\boldsymbol{W}_{ij}^* S_\eta(\boldsymbol{W}_{ij}^*\boldsymbol{\beta}^*)}{G^{(3)}G^{(4)}}\cdot\left[\frac{G^{(3)}-\hat{G}^{(3)}}{G^{(3)}} + \frac{G^{(4)}-\hat{G}^{(4)}}{G^{(4)}}\right] \\ & + o_p(1), \end{aligned} \tag{4.12}$$

where $G^{(1)} = G(R_j^*-,X_j^*-)$, $G^{(2)} = G(R_i^*-,X_j^*-)$, $G^{(3)} = G(R_j^*-,0)$, and $G^{(4)} = G(R_i^*-,0)$.

Following the notations in Chapter 3 and using the result in (3.27), we have that for $G^{(1)} = G(R_j^*-,X_j^*-)$,

$$\begin{aligned} & \sqrt{n}\left[\frac{G(R_j^*-,X_j^*-)-\hat{G}(R_j^*-,X_j^*-)}{G(R_j^*-,X_j^*-)}\right] \\ = {} & \frac{1}{\sqrt{n}}\left[\sum_{k=1}^{n}\int_{s<\tilde{Z}_j^*(\alpha_j)}\frac{1}{H(s-;\alpha_j)}M_k(ds;\alpha_j)\right] + r_n\big(\tilde{Z}_j^*(\alpha_j)-;\alpha_j\big) \\ = {} & \frac{1}{\sqrt{n}}\left[\sum_{k=1}^{n}\xi_k(R_j^*-,X_j^*-)\right] + r_n\big(\tilde{Z}_j^*(\alpha_j)-;\alpha_j\big). \end{aligned} \tag{4.13}$$

Similar results for $G^{(2)} = G(R_i^*-,X_j^*-)$, $G^{(3)} = G(R_j^*-,0)$, $G^{(4)} = G(R_i^*-,0)$ can

also be obtained. Therefore, II in (4.11) can be written as

$$
\begin{aligned}
II = n^{-5/2} \sum_{i,j,k=1}^{n} & \left\{ \frac{\boldsymbol{W}_{ij}^* \delta_j^* I[X_i^* \geqslant X_j^*]}{G(R_i^*-, X_j^*-) G(R_j^*-, X_j^*-)} \cdot \left[\xi_k(R_j^*-, X_j^*-) + \xi_k(R_i^*-, X_j^*-) \right] \right\} \\
& - n^{-5/2} \sum_{i,j,k=1}^{n} \left\{ \frac{\boldsymbol{W}_{ij}^* S_\eta(\boldsymbol{W}_{ij}^* \boldsymbol{\beta}^*)}{G(R_i^*-, 0) G(R_j^*-, 0)} \cdot \left[\xi_k(R_j^*-, 0) + \xi_k(R_i^*-, 0) \right] \right\} + o_p(1) \\
:= n^{-5/2} & \sum_{i,j,k=1}^{n} \boldsymbol{\varsigma}_{ijk} + o_p(1) \\
:= \boldsymbol{U_n} & + o_p(1). \qquad (4.14)
\end{aligned}
$$

According to the properties of U-statistics (Serfling, 1980) we have $\boldsymbol{U_n} = \widehat{\boldsymbol{U}}_{\boldsymbol{n}} + o\big(n^{-1}(\log n)^\rho\big)$ for some $\rho > 0$, where

$$
\begin{aligned}
\widehat{\boldsymbol{U}}_{\boldsymbol{n}} &= \sum_{k=1}^{n} E\left\{ \boldsymbol{U_n} \,|\mathcal{D}_l \right\} \\
&= \sum_{l=1}^{n} E\Big\{ n^{-5/2} \sum_{i,j,k=1}^{n} \boldsymbol{\varsigma}_{ijk} \,|\mathcal{D}_l \Big\}, \qquad (4.15)
\end{aligned}
$$

and $\mathcal{D}_l = \{R_l^*, L_l^*, X_l^*, \delta_l^*, \boldsymbol{W}_l^*\}$ denotes the observed information of patient l. The result given in (4.15), together with (4.14), imply that term II in (4.11) can be written as a sum of i.i.d. terms. Hence $\sqrt{n}\Big[\boldsymbol{Q}\big(\boldsymbol{\beta}^*; \hat{G}\big)\Big]$ is a sum of i.i.d. terms.

Therefore, the variance-covariance matrix of $\sqrt{n}\Big[\boldsymbol{Q}\big(\boldsymbol{\beta}^*; \hat{G}\big)\Big]$ is given by

$$
\begin{aligned}
\boldsymbol{\Sigma_Q} &= \mathrm{Var}\left\{ \sqrt{n}\Big[\boldsymbol{Q}\big(\boldsymbol{\beta}^*; \hat{G}\big)\Big] \right\} \\
&= \mathrm{Var}\left\{ \sqrt{n}\left[\boldsymbol{Q}(\boldsymbol{\beta}^*; G)\right] + \sqrt{n}\left[\boldsymbol{Q}\big(\boldsymbol{\beta}^*; \hat{G}\big) - \boldsymbol{Q}\left(\boldsymbol{\beta}^*; G\right)\right] \right\}. \qquad (4.16)
\end{aligned}
$$

Then following from the first-order Taylor expansion, we have that

$$\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big) = \boldsymbol{Q}\big(\hat{\boldsymbol{\beta}};\hat{G}\big) + \boldsymbol{Q}'\big(\hat{\boldsymbol{\beta}};\hat{G}\big)\big(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\big) + O\big((\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}})^2\big). \tag{4.17}$$

Since $\hat{\boldsymbol{\beta}}$ is a root of $\boldsymbol{Q}\big(\boldsymbol{\beta};\hat{G}\big) = \mathbf{0}$, we have $\boldsymbol{Q}\big(\hat{\boldsymbol{\beta}};\hat{G}\big) = \mathbf{0}$. Together with $\hat{\boldsymbol{\beta}}$ is a consistent estimate of $\boldsymbol{\beta}$, we have that

$$\sqrt{n}\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big) = \sqrt{n}\boldsymbol{Q}'\big(\hat{\boldsymbol{\beta}};\hat{G}\big)\big(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\big). \tag{4.18}$$

Hence,

$$\mathrm{Var}\Big\{\sqrt{n}\big(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\big)\Big\} = \mathrm{Var}\Big\{\sqrt{n}\big[\boldsymbol{Q}'\big(\hat{\boldsymbol{\beta}};\hat{G}\big)\big]^{-1}\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big)\Big\}. \tag{4.19}$$

Therefore we have $\sqrt{n}\big(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\big) \sim N\big(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\big)$, where

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \Big[\boldsymbol{Q}'(\hat{\boldsymbol{\beta}};\hat{G})\Big]^{-1}\boldsymbol{\Sigma}_{\boldsymbol{Q}}\left\{\Big[\boldsymbol{Q}'(\hat{\boldsymbol{\beta}};\hat{G})\Big]^{-1}\right\}^{tr}. \tag{4.20}$$

4.4 Simulation studies and data analysis

4.4.1 Simulation studies

In this section, we conduct simulation studies to assess the performance of the proposed methods. Truncation times L and censoring times C are generated respectively from

$$\begin{aligned} L &= a\nu_1 + b\nu_2 + U[0,1] \\ C &= c\nu_1 + d\nu_2 \end{aligned}$$

where $\nu_1 \sim U[1,6]$ and $\nu_2 \sim 5 \cdot Exp(1)$. We can change the values of (a,b,c,d) to adjust censoring/truncation probabilities and correlations of L and C. Pairs of survival times are generated from the following model to mimic the Edinburgh data in Fu et al. (2007). The cirrhosis time follows $h(T) = -\boldsymbol{W}\boldsymbol{\beta}^* + \varepsilon$, where the true value $\boldsymbol{\beta}^* = (1.5, 1.0, 0.8)^{tr}$ and the function h is taken to be $h(T) = 2(\log T - 3.5)$. The covariate $\boldsymbol{W}$ is defined as: $W_1 \sim U[1,4]$, $W_2 \sim$ Bernoulli(0.5) and $W_3 \sim$ Bernoulli(0.5). The referral time R follows $R = 0.6 \cdot T + U[0,1] + 1.0$. The observed sample size is set at $n = 200$ and 100. The replication time is 100. To mimic the proportional hazards model, the distribution function of ε is taken to be a standard extreme value distribution, which is

$$F_\varepsilon(t) = 1 - \exp[-\exp(t)].$$

The estimated regression coefficient vector $\hat{\boldsymbol{\beta}}$ is calculated iteratively by Newton's method. The simulation results are presented in Table 4.1 and 4.2. We choose different values for a, b, c, d to achieve different censoring and truncation percentages.

We can see that when sample size is large ($n = 200$), truncation probability is high (around 0.85) and censoring percentage is low (around 20%), the biases of the proposed estimators are (0.021, 0.064, 0.062) respectively, which are very small. In this case, the mean standard deviation estimates and standard deviation for Monte Carlo estimates are very close. When the data are more biased (with around 50% censoring), the biases for the predictor parameters are still acceptable, (0.279, 0.236, 0.207) respectively. We can conclude that when the observed sample is not severely biased (with truncation probability $\geqslant 0.5$ and censoring percentage $\leqslant 0.5$), the

Table 4.1: Simulation results. $n = 200$: observed sample size; (e) estimate (bias in parenthesis). γ: truncation probability; $\%c$: censoring percentage; $\hat{s}_\beta$: means of standard deviation estimates obtained by (4.20); $\hat{\sigma}_\beta$: the standard deviation for $\hat{\beta}$ and $\hat{\beta}$ based on the 100 simulations.

	$n = 200$		c% = 20%			c% = 50%			c% = 80%		
	$\gamma =$	$\boldsymbol{\beta}^*$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$
(1)	0.85	1.500	1.479(0.021)	0.279	0.287	1.221(0.279)	0.204	0.218	1.124(0.376)	0.201	0.225
	0.85	1.000	0.936(0.064)	0.226	0.239	0.764(0.236)	0.207	0.223	0.663(0.337)	0.214	0.237
	0.85	0.800	0.738(0.062)	0.239	0.242	0.593(0.207)	0.232	0.251	0.537(0.263)	0.232	0.261
(2)	0.5	1.500	1.273(0.227)	0.316	0.318	1.144(0.356)	0.282	0.296	1.038(0.462)	0.301	0.324
	0.5	1.000	0.815(0.185)	0.238	0.254	0.772(0.228)	0.286	0.309	0.712(0.288)	0.338	0.347
	0.5	0.800	0.709(0.091)	0.234	0.272	0.644(0.156)	0.251	0.268	0.659(0.141)	0.302	0.318
(3)	0.15	1.500	1.134(0.366)	0.308	0.319	0.925(0.575)	0.311	0.389	0.772(0.728)	0.401	0.429
	0.15	1.000	0.713(0.287)	0.307	0.325	0.614(0.386)	0.327	0.342	0.581(0.419)	0.437	0.453
	0.15	0.800	0.625(0.175)	0.310	0.323	0.598(0.202)	0.335	0.337	0.503(0.297)	0.465	0.484

Table 4.2: Simulation results. $n = 100$: observed sample size; (e) estimate (bias in parenthesis). γ: truncation probability; $\%c$: censoring percentage; $\hat{s}_\beta$: means of standard deviation estimates obtained by (4.20); $\hat{\sigma}_\beta$: the standard deviation for $\hat{\beta}$ and $\hat{\beta}$ based on the 100 simulations.

	$n = 100$		c% = 20%			c% = 50%			c% = 80%		
	$\gamma =$	$\boldsymbol{\beta}^*$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$	(e)	$\hat{s}_\beta$	$\hat{\sigma}_\beta$
(1)	0.85	1.500	1.398(0.102)	0.286	0.297	1.132(0.368)	0.257	0.261	1.003(0.497)	0.223	0.238
	0.85	1.000	0.898(0.102)	0.283	0.290	0.709(0.291)	0.308	0.315	0.587(0.413)	0.324	0.337
	0.85	0.800	0.671(0.129)	0.290	0.301	0.558(0.242)	0.329	0.337	0.496(0.304)	0.351	0.380
(2)	0.5	1.500	1.140(0.360)	0.289	0.298	1.010(0.490)	0.293	0.297	0.989(0.511)	0.342	0.351
	0.5	1.000	0.721(0.279)	0.287	0.298	0.627(0.373)	0.310	0.314	0.573(0.427)	0.338	0.349
	0.5	0.800	0.602(0.198)	0.293	0.305	0.569(0.231)	0.335	0.342	0.528(0.272)	0.337	0.385
(3)	0.15	1.500	1.009(0.491)	0.304	0.312	0.734(0.766)	0.389	0.453	0.658(0.842)	0.428	0.474
	0.15	1.000	0.620(0.380)	0.308	0.327	0.573(0.427)	0.392	0.447	0.527(0.473)	0.453	0.486
	0.15	0.800	0.593(0.207)	0.347	0.400	0.511(0.289)	0.447	0.461	0.302(0.498)	0.455	0.497

proposed estimators work well for large sample sizes.

When the truncation probability is about 0.15 and the censoring percentage is around 80%, the observed sample is severly biased (only 15% of the population can be observed and recruited successfully and among the observed sample 80% are censored, i.e. only about 20% are fully observed). Therefore the biases of the predictors, the mean standard deviation estimates and the standard deviation for Monte Carlo estimates are quite large.

Moreover, the simulation results indicate that for approximately the same observed sample size and approximately the same truncation probability, the biases will increase with the censoring percentage increasing. For approximately the same truncation probability and approximately the same censoring percentage, the standard deviations decrease with the increasing of the observed sample size. For approximately the same observed sample size and approximately the same censoring percentage, the biases will increase with the truncation probability decreasing.

Comparing with the simulation results from the AFT model in Chapter 3, the biases of the predictors from the linear transformation model are relatively larger. This is not surprsing as in AFT model there is an intercept term which has larger biases than the predictors. However, we do agree that the method proposed in this chapter may not be a good choice for severely biased data (with low truncation probability and high censoring percentage).

The predictor estimates are calculated iteratively by the Newton-Raphson's method. Comparing with the WLS method for AFT model in Chapter 3, this iterative method is much more time consuming. Moreover, the convergence of the estimates strongly depends on the choice of the initial estimator $\hat{\boldsymbol{\beta}}^{(0)}$. This problem is caused by the Newton-Raphson's method. When the initial estimator $\hat{\boldsymbol{\beta}}^{(0)}$ is not

close enough to the true value of $\hat{\boldsymbol{\beta}}$, the iteration procedure may not converge.

4.4.2 Comparison with the case ignoring the referral bias

To see how well our proposed estimators perform when handle censored survival data with referral bias, we apply the method proposed in (Cheng et al., 1995) to the same simulated datasets in section 4.4.1.

The estimating equation is given by

$$\boldsymbol{Q}(\boldsymbol{\beta}; \hat{G}) = n^{-2} \sum_{i,j=1}^{n} \boldsymbol{W}_{ij}^{*} \left\{ \frac{\delta_j^* I[X_i^* \geqslant X_j^*]}{\hat{G}^2(X_j^*)} - S_\eta\big(\boldsymbol{W}_{ij}^{*}\boldsymbol{\beta}\big) \right\} = \mathbf{0}. \tag{4.21}$$

Here $\hat{G}(X_j^*)$ is the Kaplan-Meier estimator for the univariate survival function of the censoring variable C based on the observed data pairs $\big(X_j^*, \delta_j^*\big)$ $(j = 1, \ldots, n)$. The results of 100 simulations are presented in Table 4.3.

Here we present the simulation results for observed sample size ($n = 200$), different truncation probabilities ($\gamma = 0.85, 0.5$ and 0.15) and different censoring percentages (c%=20% and 50%). The results for heavy censoring percentage (c%=80%) are not included since for our simulated dataset, the method without considering referral bias already produces severely biased results of the predictor parameter estimates when the censoring percentage $c\% = 50\%$.

Comparing with the simulation results shown in Table 4.1, we can see that the predictor parameter estimates are more biased when the referral bias is ignored. This also indicates that our proposed method, which can remove the referral bias, can produce more reasonable estimates for the effects of the covariates.

Table 4.3: Simulation results for the linear transformation model without considering referral bias.

	$n = 200$		c% = 20%		c% = 50%	
	$\gamma =$	$\boldsymbol{\beta}^*$	$\hat{\boldsymbol{\beta}}$	$\hat{\sigma}_{\boldsymbol{\beta}}$	$\hat{\boldsymbol{\beta}}$	$\hat{\sigma}_{\boldsymbol{\beta}}$
(1)	0.85	1.500	1.269(0.231)	0.335	0.748(0.752)	0.206
	0.85	1.000	0.839(0.161)	0.562	0.481(0.519)	0.372
	0.85	0.800	0.615(0.185)	0.242	0.420(0.380)	0.170
(2)	0.5	1.500	1.174(0.326)	0.455	0.670(0.830)	0.228
	0.5	1.000	0.793(0.207)	0.629	0.462(0.538)	0.374
	0.5	0.800	0.619(0.181)	0.278	0.357(0.443)	0.213
(3)	0.15	1.500	0.804(0.696)	0.387	0.569(0.931)	0.302
	0.15	1.000	0.505(0.495)	0.622	0.415(0.585)	0.482
	0.15	0.800	0.443(0.357)	0.254	0.276(0.524)	0.254

n: observed sample size; γ: truncation probability; $\%c$: censoring percentage; $\hat{\sigma}_{\boldsymbol{\beta}}$: the standard deviation for $\hat{\boldsymbol{\beta}}$ based on the 100 simulations.
(1) $c = d = 1.0$ corresponding to $\gamma \approx 0.85$ and $a = b = 0.7$, $a = b = 0.3$ approximately corresponding to 20% censoring and 50% censoring, respectively;
(2) $c = d = 0.4$ corresponding to $\gamma \approx 0.5$ and $a = b = 0.4$, $a = b = 0.2$ approximately corresponding to 20% censoring and 50% censoring, respectively;
(3) $c = d = 0.2$ corresponding to $\gamma \approx 0.2$ and $a = b = 0.2$, $a = b = 0.15$ approximately corresponding to 20% censoring and 50% censoring, respectively.

4.4.3 Data analysis

We illustrate the proposed method with the Edinburgh hepatitis C data in Fu et al. (2007). The purpose of the study is to determine how the progression to cirrhosis is affected by three covariates: age at infection, HIV co-infection (yes:1 or no:0), and heavy alcohol consumption (yes:1 or no:0). And individual with heavy alcohol intake was defined as one consuming more than 50 units alcohol per week for at least five years.

Table 4.4: Estimation results (SE in parenthesis)

	Considering referral bias	Without considering referral bias
	$\hat{\boldsymbol{\beta}}$ (SE)	$\hat{\boldsymbol{\beta}}$ (SE)
Age	1.201* (0.227)	1.668* (0.363)
HIV	0.901* (0.242)	1.081* (0.265)
Alcohol	0.577 (0.301)	0.881 (0.571)

Note: * means the estimate is significant at 5% level.

Table 4.4 summarizes the estimates of regression parameters obtained from our method. The coavariate heavy alcohol intake is failed to identified as a risk factor associated with more rapid disease progression no matter we consider the referral bias or not. This is not surprising since the Schoenfeld residuals checking (Collett, 2003) shown in Figure 4.1 implies that the proportional hazards model assumption may be violated for the variable heavy alcohol. Schoenfeld residuals can be thought of as the observed minus the expected values of the covariates at each failure time. Therefore Schoenfeld residuals are only defined for uncensored individuals. If the PH assumption holds, then we would expect the Schoenfeld residuals to display no systematic patterns. Hence in residual plots, we expect the slope of the Schoenfeld residuals with respect to time should be zero. However in Figure 4.1, the fitted smooth curve is not flat for time between 16 years and 30 years, which suggests

that as time passes, the effect of the covariate heavy alcohol is changing. Note that in Figure 4.1 the residual dots fall into two parts (top and bottom), because the covariate heavy alcohol is a 0-1 nominal variable. Even for the residual dots, those at the top of the graph are surely not flat.

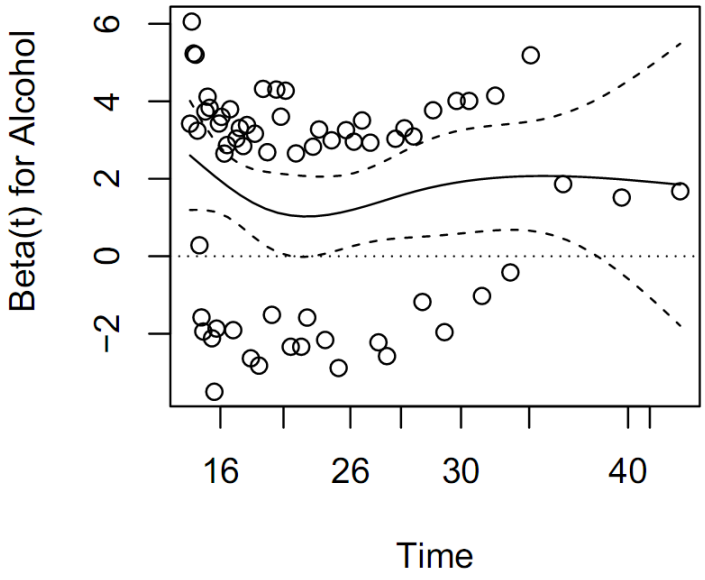

Figure 4.1: Plot of Schoenfeld residuals. (It implies that the parameter for Alcohol is not constant over time.)

Comparing the results from the truncated model (where the referral bias is considered) with the results from the non-truncation model, we can see that considering the referral bias will produce smaller values of the estimates and then lead to larger values of progression to cirrhosis. This is reasonable since we have explained before that the patients with more rapid disease progression are preferrentially referred to the liver clinics. Hence the patients within the clinics should have more rapid disease progression and shorter time period from infection to cirrhosis. Our proposed method can remove the referral bias and consequently produce smaller estimates for

the effects of the covariates.

4.5 Conclusion and discussion

In this chapter, we considered a well-known class of semi-parametric linear transformation models, under which an unknown transformation $h(\cdot)$ of the cirrhosis time T is linearly related to the covariates vector $\boldsymbol{W}$ with completely unspecified error distributions. An unbiased estimating equation for the regression coefficient vector $\boldsymbol{\beta}$ is proposed. The referral bias is taken into account by incorporating the bivariate survival function G of the truncation time and the censoring time into the estimating equation. A heuristic proof of the asymptotic normality of the coefficient estimator is shown. Simulation studies are proposed for the proportional hazards model since it can be treated as a special case of the linear transformation models. The results of the simulation studies show that our proposed methods perform well for the datasets which are not severely biased. We applied our proposed method to analyse the Edinburgh hepatitis C data in Fu et al. (2007). We concluded that older age at HCV infection and HIV co-infection can be identified as risk factors associated with more rapid HCV disease progression. Furthermore, the results indicated that considering the referral bias will lead to longer time period from HCV infection to cirrhosis, which could reach an agreement with Fu et al. (2007).

Comparing with the results of the data analysis for the AFT model, the difference is that the covariate heavy alcohol intake was failed to identified as a risk factor associated with more rapid disease progression although the referral bias has been considered and removed. This is because that the proportional hazards model assumption may be violated for this variable. Of course the Schoenfeld residuals

checking is just a method based on a graphical method.

Comparison of the values of the estimates in Table 3.9 and 4.4 seems impossible, since $h(.)$ is an nondecreasing function with an unknown form. It is possible to do the comparison in the simulation studies if the same simulated datasets are applied to these two models. Furthermore, a problem will arise when doing the data anlaysis, which is the convergence of the iterative sequence. In the simulation studies, we have mentioned that it is necessary to choose a starting value $\hat{\boldsymbol{\beta}}^{(0)}$ close to the true value of $\boldsymbol{\beta}$. However, it is not possible to know the true value of $\boldsymbol{\beta}$ when we analyse real datasets. Therefore for future work it is valuable that we develop the theory that how to choose the starting value as well as how to estimate the function $h(.)$.

Chapter 5

Conclusion and future works

The motivation for this thesis arose from a hepatitis C cohort study where the epidemiological interest was to study progression to liver cirrhosis in HCV-infected patients. Paired event times (R, T) was recorded for each patient in the Edinburgh hepatitis C data studied by Fu et al. (2007). Clinically, the patients with more rapid disease progression are preferentially referred to liver clinics (Dore et al., 2002). Conventional analysis based on liver clinics has been seen to overestimate the community-wide rate of progression to cirrhosis because of the ignorance of the potential referral bias. Therefore, to obtain more reasonable results in estimating the effects of covariates on the progression to cirrhosis, it is necessary to take the referral bias into account.

In this thesis, truncation was involved so that referral bias could be taken into account. The paired event times (R, T) can be treated as bivariate survival data since the referral time R may be right truncated by the truncation time L, and the cirrhosis time T may be right censored by the censoring time C. Existing methods in survival analysis are not readily available for such a bivariate censoring-truncation

model. Therefore, we developed new methodologies for different regression models to handle such kind of censored survival data with referral bias.

In Chapter 3, we considered the AFT model which simply regresses the logarithm of the cirrhosis time T on the covariates vector $\boldsymbol{W}$. A nonparametric method which based on a polar coordinate transformation was proposed to allow a flexible bivariate distribution structure between R and T. The dependency of the two event times and the referral bias in the sampling, incorporated by a right truncation on R, were both taken into account. Then the effects of covariates were estimated by employing a weighted least squares method for censored survival data. The estimating equation was modified by including the estimated bivariate survival function of truncation time and censoring time into the weights. The large sample properties of the estimated bivariate function and the estimated vector of covariates effects were also demonstrated.

By carrying out different scenarios of simulations, we found out that our proposed estimator for the effects of covariates and its covariance estimator performed well. We also applied our proposed method to analyse the Edinburgh hepatitis C data in Fu et al. (2007). The results from the truncated model, where the referral bias is considered, implied that age at infection, HIV co-infection and heavy alcohol in-take are significantly identified as risk factors associated with more rapid disease progression. We compared the results with those from a non-truncated model. Ignorance of the referral bias has failed to identify heavy alcohol intake as a significant risk factor. However in medical literatures, older age at infection, HIV co-infection and heavy alcohol intake have all been identified as factors associated with more rapid hepatitis C disease progression (Sharma and Sherker, 2009). Furthermore, based on the estimates of regression parameters for both truncated and non-truncated models,

we predicted the time period from infection to cirrhosis for individuals with different values of covariates. The prediction results indicated that considering the referral bias would lead to longer time period from HCV infection to cirrhosis, which could reach an agreement with Fu et al. (2007). This also implied that our estimating method can remove the referral bias and get more reliable estimates for the effects of covariates.

In Chapter 4, we considered the semi-parametric linear transformation model which regresses the transformation of cirrhosis time on the covariates vector. By using similar ideas in Chapter 3, we obtained the asymptotically unbiased estimating equation for the regression coefficient vector. Heuristic proofs of the unbiasedness and asymptotic normality of the proposed coefficient estimator were shown. Since the proportional hazards model (Cox, 1972) is a special case of the transformation model, we carried on simulation studies for the proportional hazards model under different sample sizes, different truncation probabilities and different censoring percentages. The results of the simulation studies show that our proposed methods perform well for the datasets which are not severely biased. We applied our proposed method to analyse the Edinburgh hepatitis C data in Fu et al. (2007). We concluded that older age at HCV infection and HIV co-infection can be identified as risk factors associated with more rapid HCV disease progression. Furthermore, the results indicated that considering the referral bias will lead to longer time period from HCV infection to cirrhosis, which could reach an agreement with Fu et al. (2007).

In summary, both the AFT model and the linear transformation model, coupled with the corresponding newly developed theories, are useful techniques to handle censoring survival data with referral bias . Here in this thesis, the AFT model and

the method proposed in Chapter 3 is recommended. This is because:

- The method proposed in Chapter 3 is a non-parametric method, which does not need any assumptions on the distribution of the error terms. This allows our AFT model to have wider applicability and increased robustnees, compared to the semi-parametric linear transformation model in Chapter 4.

- By using the WLS method based on the AFT model, we can directly get the closed-form solution of $\boldsymbol{\beta}$. However, for the linear transformation model in Chapter 4, we have to use iterative algorithm such as Newton-Raphson's method to get the estimate of $\boldsymbol{\beta}$. The iterative algorithm is much more time-consuming and the convergence of the iterative sequence depends strongly on how to choose the starting value.

- From the results of the simulation studies, we see that the AFT model can produce estimates with much smaller biases and standard errors, even when the data are severely biased (quite low truncation probability and heavy censoring percentage). This will be helpful and can give more reliable results when handling the real dataset.

For future work, we may extend the KSV method (Koul et al., 1981) for the AFT model by using the polar coordinate transformation method. The KSV method is also based on an inverse probability weighted approach, but it only re-weights the response variable, whereas the weighted least squares method re-weights both the response variable and the predictors. It is worth making further comparisons for these two methods under both censoring and truncation. The independent assumption between (C, L) and $\boldsymbol{W}$, required by the KSV or WLS estimator, may not be

appropriate in other studies. It could be possible to adapt our method under a weaker assumption, which assumes that (L, C) and (R, T) are independent conditionally on the covariate vector $\boldsymbol{W}$. Moreover, another direction of future work is that we may extend our proposed method to the proportional hazards model or the proportional odds model with time-varying coefficients.

Appendix A

Proofs in Section 3.3

A.1 Proof of Lemma 3.2.1

Multiplying both sides of the AFT model equation $T_i = \boldsymbol{W}_i\boldsymbol{\beta}+\epsilon_i$ by $\boldsymbol{W}_i^{tr}$ and taking expectation yield (3.7). Let

$$\hat{F}^*(t_1, t_2, \boldsymbol{w}) = n^{-1}\sum_{i=1}^{n} I[R_i^* \leqslant t_1, X_i^* \leqslant t_2, \boldsymbol{W}_i^* \leqslant \boldsymbol{w}, \delta_i^* = 1]. \tag{A.1}$$

Then from equation (3.6) we know that $F(t_1, t_2, \boldsymbol{w})$ can be estimated by

$$\hat{F}(t_1, t_2, \boldsymbol{w}) \;=\; \hat{\gamma}\int_{s_1\leqslant t_1}\int_{s_2\leqslant t_2}\int_{\boldsymbol{u}\leqslant\boldsymbol{w}} \frac{1}{G(s_1-, s_2-)}\hat{F}^*(ds_1, ds_2, d\boldsymbol{u}). \tag{A.2}$$

Therefore if G is known, ς_{0k} and ς_{jk}, can be estimated respectively by the unbiased estimates

$$\begin{aligned}\hat{\varsigma}_{0k}(G) &= \int t_2 w_k \hat{F}(dt_1, dt_2, d\boldsymbol{w}) = \frac{\hat{\gamma}}{n}\sum_{i=1}^{n}\frac{X_i^* W_{ik}^* \delta_i^*}{G(R_i^*-, X_i^*-)},\\ \hat{\varsigma}_{jk}(G) &= \int w_j w_k F(dt_1, dt_2, d\boldsymbol{w}) = \frac{\hat{\gamma}}{n}\sum_{i=1}^{n}\frac{W_{ij}^* W_{ik}^* \delta_i^*}{G(R_i^*-, X_i^*-)},\end{aligned}$$

where $1 \leqslant j, k \leqslant p$, and p is also defined in Lemma 3.2.1.

A.2 Proof of Lemma 3.3.1

Let $Z^*(\alpha) = \min\{L^*\sqrt{1+\alpha^2}, C^*\sqrt{1+\alpha^{-2}}\}$. Then based on the observed data $(L^*, R^*, X^*, \delta^*)$, we have the following cases for the observed transformed data given in (3.12).

1. If $\delta^* = 0$, then $X^* = C^* < T^*$; and

   - if $\tilde{L}^* > \tilde{X}^*$, we have $\Delta^*(\alpha) = 1$,
     $L^*\sqrt{1+\alpha^2} > X^*\sqrt{1+\alpha^{-2}} = C^*\sqrt{1+\alpha^{-2}} < T^*\sqrt{1+\alpha^{-2}}$,
     and $\tilde{Z}^*(\alpha) = Z^*(\alpha) = C^*\sqrt{1+\alpha^{-2}}$;

   - if $\tilde{L}^* \leqslant \tilde{X}^*$, we have $\Delta^*(\alpha) = 1$,
     $L^*\sqrt{1+\alpha^2} \leqslant X^*\sqrt{1+\alpha^{-2}} = C^*\sqrt{1+\alpha^{-2}} < T^*\sqrt{1+\alpha^{-2}}$,
     and $\tilde{Z}^*(\alpha) = Z^*(\alpha) = L^*\sqrt{1+\alpha^2}$.

2. If $\delta^* = 1$, then $X^* = T^* \leqslant C^*$; and

   - if $\tilde{L}^* \leqslant \tilde{X}^*$, we have $\Delta^*(\alpha) = 1$,
     $L^*\sqrt{1+\alpha^2} \leqslant X^*\sqrt{1+\alpha^{-2}} = T^*\sqrt{1+\alpha^{-2}} \leqslant C^*\sqrt{1+\alpha^{-2}}$,

and $\tilde{Z}^*(\alpha) = Z^*(\alpha) = L^*\sqrt{1+\alpha^2}$;

- if $\tilde{L}^* > \tilde{X}^*$, we have $\Delta^*(\alpha) = 0$,
  $L^*\sqrt{1+\alpha^2} > X^*\sqrt{1+\alpha^{-2}} = T^*\sqrt{1+\alpha^{-2}}$, $T^*\sqrt{1+\alpha^{-2}} \leqslant C^*\sqrt{1+\alpha^{-2}}$,
  and $\tilde{Z}^*(\alpha) = T^*\sqrt{1+\alpha^{-2}} \leqslant \min\{L^*\sqrt{1+\alpha^2}, C^*\sqrt{1+\alpha^{-2}}\} = Z^*(\alpha)$.

In summary, if $\Delta^*(\alpha) = 1$, then $\tilde{Z}^*(\alpha) = Z^*(\alpha)$ and if $\Delta^*(\alpha) = 0$, then $\tilde{Z}^*(\alpha) \leqslant Z^*(\alpha)$. This implies that $\tilde{Z}^*(\alpha)$ is an observed value for $Z^*(\alpha)$ and $\Delta^*(\alpha) = 0$ implies censoring.

A.3 Proof of Lemma 3.3.2

From which the lemma follows, we have

$$\begin{aligned}
& \frac{P(\tilde{Z}^*(\alpha) \in dz, z > V^*(\alpha), \Delta^*(\alpha) = 1)}{P(\tilde{Z}^*(\alpha) \geqslant z > V^*(\alpha))} \\
= \; & \frac{P(\tilde{Z}(\alpha) \in dz, z > V(\alpha), \Delta(\alpha) = 1 | R \leqslant L)}{P(\tilde{Z}(\alpha) \geqslant z > V(\alpha) | R \leqslant L)} \\
= \; & \frac{P(z \leqslant \tilde{Z}(\alpha) < z + dz, z > V(\alpha), \Delta(\alpha) = 1, R \leqslant L)}{P(\tilde{Z}(\alpha) \geqslant z > V(\alpha), R \leqslant L)} \qquad \text{(A.3)} \\
= \; & \frac{P(z \leqslant \tilde{Z}(\alpha) < z + dz, \Delta(\alpha) = 1)}{P(\tilde{Z}(\alpha) \geqslant z)}. \qquad \text{(A.4)}
\end{aligned}$$

Here $\tilde{Z}(\alpha) = \min\{L\sqrt{1+\alpha^2}, X\sqrt{1+\alpha^{-2}}\}$ and $V(\alpha) = R\sqrt{1+\alpha^2}$. The term $R \leqslant L$ in (A.3) can be omitted since $\tilde{Z}(\alpha) \geqslant z > V(\alpha)$ already implies that. Together with the assumption that $(R, T, \boldsymbol{W})$ and (L, C) are independent, the last equality sign can be obtained.

From the proof of Lemma 3.3.1 we know that $\{\Delta(\alpha) = 1\} \Leftrightarrow \{\tilde{Z}(\alpha) = Z(\alpha) \leqslant T\sqrt{1+\alpha^{-2}}\}$ and $\{\Delta(\alpha) = 0\} \Leftrightarrow \{\tilde{Z}(\alpha) = T\sqrt{1+\alpha^{-2}} \leqslant Z(\alpha)\}$. Hence we have

that

$$
\begin{aligned}
(A.4) \quad &= \frac{P\big(z \leqslant Z(\alpha) < z + dz\big)}{P\Big(\min\big\{L\sqrt{1+\alpha^2}, X\sqrt{1+\alpha^{-2}}\big\} \geqslant z\Big)} \\
&= \frac{P\big(z \leqslant Z(\alpha) < z + dz, z \leqslant T\sqrt{1+\alpha^{-2}}\big)}{P\Big(\min\big\{L\sqrt{1+\alpha^2}, T\sqrt{1+\alpha^{-2}}, C\sqrt{1+\alpha^{-2}}\big\} \geqslant z\Big)} \\
&= \frac{P\big(z \leqslant Z(\alpha) < z + dz, z \leqslant T\sqrt{1+\alpha^{-2}}\big)}{P\big(Z(\alpha) \geqslant z, z \leqslant T\sqrt{1+\alpha^{-2}}\big)} \\
&= P\big(Z(\alpha) \in dz \big| Z(\alpha) \geqslant z\big) \\
&= \Lambda(dz; \alpha).
\end{aligned}
$$

Therefore the lemma is proved.

Appendix B

Proofs in Section 3.4

B.1 Proof of Lemma 3.4.1

From the definition of $H(s;\alpha)$ we know that $H(0;\alpha) = H(\tau_\alpha;\alpha) = 0$ and $H(s;\alpha) > 0$ for $0 < s < \tau_\alpha$. Since we assume that H is a continuous function, for $0 < \kappa < 1/2$, we can find the curves

$$\boldsymbol{c}_1 := \left\{(y_{n;\alpha};\alpha) : H(s;\alpha) \leqslant n^{\kappa-1}, \forall 0 \leqslant s \leqslant y_{n,\alpha}\right\}$$

and

$$\boldsymbol{c}_2 := \left\{(x_{n;\alpha};\alpha) : H(s;\alpha) \leqslant n^{\kappa-1}, \forall x_{n,\alpha} \leqslant s \leqslant \tau_\alpha\right\}$$

such that $H(s;\alpha) \geqslant n^{\kappa-1}$ for $y_{n;\alpha} \leqslant s \leqslant x_{n;\alpha}$.

Under the mild conditions that H is differentiable and bounded in $[0,\tau_\alpha]\times[0,\infty]$,

we have that for any fixed $\alpha \in [0, \infty]$, there exists $x^*_{n;\alpha}$ and $y^*_{n;\alpha}$ such that

$$\begin{aligned} H(x_{n;\alpha};\alpha) - H(\tau_\alpha;\alpha) &= H^{'}(x^*_{n;\alpha};\alpha)\big|x_{n;\alpha} - \tau_\alpha\big|, \\ H(y_{n;\alpha};\alpha) - H(0;\alpha) &= H^{'}(y^*_{n;\alpha};\alpha)\big|y_{n;\alpha} - 0\big|, \end{aligned}$$

and also $|x_{n;\alpha} - \tau_\alpha| = |y_{n;\alpha}| = O(n^{\kappa-1})$. Similarly, under the mild conditions that Λ is differentiable and bounded in $[0, \tau_\alpha] \times [0, \infty]$, it is easy to derive that $\Lambda(\tau_\alpha;\alpha) - \Lambda(x_{n;\alpha};\alpha) = \Lambda(y_{n;\alpha};\alpha) = O(n^{\kappa-1})$.

Then for $\alpha \in [0, \infty], z \in [0, \tau_\alpha]$, we have that

$$\begin{aligned} & \sup_{\alpha,z} E\bigg\{\sqrt{n}\int_0^z I[H_{(n)}(s-;\alpha) = 0]\Lambda(ds;\alpha)\bigg\}^2 \\ \leqslant\ & n \cdot \text{const} \cdot \sup_\alpha E\bigg\{\int_0^{y_{n;\alpha}} I[H_{(n)}(s-;\alpha) = 0]\Lambda(ds;\alpha)\bigg\}^2 \\ & + n \cdot \text{const} \cdot \sup_\alpha E\bigg\{\int_{y_{n;\alpha}}^{x_{n;\alpha}} I[H_{(n)}(s-;\alpha) = 0]\Lambda(ds;\alpha)\bigg\}^2 \\ & + n \cdot \text{const} \cdot \sup_\alpha E\bigg\{\int_{x_{n;\alpha}}^{\tau_\alpha} I[H_{(n)}(s-;\alpha) = 0]\Lambda(ds;\alpha)\bigg\}^2 \\ \leqslant\ & O(n^{2\kappa-1}) + n \cdot \text{const} \cdot \sup_\alpha E\bigg\{\int_{y_{n;\alpha}}^{x_{n;\alpha}} I[H_{(n)}(s-;\alpha) = 0]\Lambda(ds;\alpha)\bigg\}^2 \\ :=\ & I + II. \end{aligned}$$

Note that

$$
\begin{aligned}
II &\leqslant n \cdot \text{const} \cdot \sup_{\alpha} E\Big\{ \int_{y_{n;\alpha}}^{x_{n;\alpha}} I[H_{(n)}(s-;\alpha)=0]\Lambda(ds;\alpha) \int_{y_{n;\alpha}}^{x_{n;\alpha}} \Lambda(ds;\alpha) \Big\} \\
&= n \cdot \text{const} \cdot \sup_{\alpha} \Big\{ \int_{y_{n;\alpha}}^{x_{n;\alpha}} P\Big[\frac{1}{n}\sum_{i=1}^{n} H_i(s-;\alpha)=0\Big]\Lambda(ds;\alpha) \int_{y_{n;\alpha}}^{x_{n;\alpha}} \Lambda(ds;\alpha) \Big\} \\
&= n \cdot \text{const} \cdot \sup_{\alpha} \Big\{ \int_{y_{n;\alpha}}^{x_{n;\alpha}} \Big[P\big(H_i(s-;\alpha)=0\big)\Big]^n \Lambda(ds;\alpha) \int_{y_{n;\alpha}}^{x_{n;\alpha}} \Lambda(ds;\alpha) \Big\} \\
&= n \cdot \text{const} \cdot \sup_{\alpha} \Big\{ \int_{y_{n;\alpha}}^{x_{n;\alpha}} \Big[1-P\big(H_i(s-;\alpha)=1\big)\Big]^n \Lambda(ds;\alpha) \int_{y_{n;\alpha}}^{x_{n;\alpha}} \Lambda(ds;\alpha) \Big\} \\
&= n \cdot \text{const} \cdot \sup_{\alpha} \Big\{ \int_{y_{n;\alpha}}^{x_{n;\alpha}} \Big(1-E\big[H_i(s-;\alpha)\big]\Big)^n \Lambda(ds;\alpha) \int_{y_{n;\alpha}}^{x_{n;\alpha}} \Lambda(ds;\alpha) \Big\} \\
&= n \cdot \text{const} \cdot \sup_{\alpha} \Big\{ \int_{y_{n;\alpha}}^{x_{n;\alpha}} \big[1-H(s-;\alpha)\big]^n \Lambda(ds;\alpha) \int_{y_{n;\alpha}}^{x_{n;\alpha}} \Lambda(ds;\alpha) \Big\} \\
&\leqslant \text{const} \cdot \sup_{s\in[y_{n;\alpha},x_{n;\alpha}],\alpha\in[0,\infty]} \Big\{ n\big(1-n^{\kappa-1}\big)^n \cdot \big[\Lambda(x_{n;\alpha};\alpha)-\Lambda(y_{n;\alpha};\alpha)\big]^2 \Big\} \\
&= O\Big(n\big(1-n^{\kappa-1}\big)^n\Big) \to 0,
\end{aligned}
$$

when $n \to \infty$. The convergence rate is $n \cdot e^{-n^{\kappa}}$ which does not depend on α.

Therefore, the term (i) in the lemma converges to 0 since $I + II \to 0$ as $n \to \infty$. The convergence of term (ii) follows similarly.

B.2 Proof of Lemma 3.4.2

We need the following lemma to prove Lemma 3.4.2.

Lemma B.2.1. *Define* $\tau_{n;\alpha} = \sup_s \big\{ s : H(s;\alpha) > n^{-\kappa/2} \cdot \log n \big\}$ *for* $\kappa \in (0,1)$ *, so that* $\tau_{n;\alpha} \to \tau_{\alpha}$ *as* $n \to \infty$*. Then we have that*

$$
\sup_{\alpha\in[0,\infty],s\in[0,\tau_{n;\alpha}]} E\Big[\frac{I\big[H_{(n)}(s-;\alpha)>0\big]}{H_{(n)}^2(s-;\alpha)}\Big] \leqslant O(n^{\kappa}). \tag{B.1}
$$

Proof. Since $H_{(n)}(s;\alpha) > 0$ implies $H_{(n)}(s;\alpha) \geqslant n^{-1}$ as $n \to \infty$, we have that for $\kappa \in (0,1)$,

$$\begin{aligned} & E\left[\frac{I\big[H_{(n)}(s-;\alpha) > 0\big]}{H^2_{(n)}(s;\alpha)}\right] \\ = \;& E\left[\frac{I\big[H_{(n)}(s-;\alpha) > n^{-\kappa/2}\big]}{H^2_{(n)}(s-;\alpha)}\right] + E\left[\frac{I\big[n^{-\kappa/2} \geqslant H_{(n)}(s-;\alpha) \geqslant n^{-1}\big]}{H^2_{(n)}(s-;\alpha)}\right] \\ \leqslant \;& n^{\kappa} + n^2 P\big(H_{(n)}(s-;\alpha) \leqslant n^{-\kappa/2}\big). \end{aligned} \tag{B.2}$$

Note that $nH_{(n)}$ is a sum of n Bernoulli variables, since

$$H_{(n)}(s-;\alpha) = n^{-1}\sum_{i=1}^{n} I[\tilde{Z}^*_i(\alpha) \geqslant s > V^*_i(\alpha)].$$

Therefore $H_{(n)}(s-;\alpha) \leqslant n^{-\kappa/2}$ implies that among the Bernoulli variables $I[\tilde{Z}^*_i(\alpha) \geqslant s > V^*_i(\alpha)]$, $i = 1,\ldots,n$, there are at most $n^{1-\kappa/2}$ variables equal to 1. Hence we have

$$P\big(H_{(n)}(s-;\alpha) \leqslant n^{-\kappa/2}\big) = \sum_{m=1}^{n^{1-\kappa/2}} C^m_n q^m (1-q)^{n-m},$$

where

$$q = P[\tilde{Z}^*_i(\alpha) \geqslant s > V^*_i(\alpha)] = H(s-;\alpha) \geqslant n^{-\kappa/2} \cdot \log n.$$

Let $r = \kappa/2$. For $\alpha \in [0, \infty], s \in [0, \tau_{n;\alpha}]$, we have that

$$
\begin{aligned}
& n^2 \sum_{m=1}^{n^{1-\kappa/2}} C_n^m q^m (1-q)^{n-m} \\
= \;& n^2 \sum_{m=1}^{n^{1-r}} C_n^m q^m (1-q)^{n-m} \\
\leqslant \;& n^2 \sum_{m=1}^{n^{1-r}} C_n^m \Big(\frac{\log n}{n^r}\Big)^m \Big(1 - \frac{\log n}{n^r}\Big)^{n-m} \\
= \;& n^2 \cdot \Big(1 - \frac{\log n}{n^r}\Big)^n \cdot \sum_{m=1}^{n^{1-r}} C_n^m \cdot \left(\frac{n^{-r} \cdot \log n}{1 - n^{-r} \cdot \log n}\right)^m \\
= \;& n^2 \cdot \Big(1 - \frac{\log n}{n^r}\Big)^n \cdot \sum_{m=1}^{n^{1-r}} \frac{n \cdots (n-m+1)}{m!} \cdot \frac{1}{\big(\frac{n^r}{\log n} - 1\big)^m} \\
\leqslant \;& n^2 \cdot \Big(1 - \frac{\log n}{n^r}\Big)^n \cdot \sum_{m=1}^{n^{1-r}} \frac{n \cdots (n-m+1)}{\big(\frac{n^r}{\log n} - 1\big)^m} \\
\leqslant \;& n^2 \cdot \Big(1 - \frac{\log n}{n^r}\Big)^n \cdot \sum_{m=1}^{n^{1-r}} \left(\frac{n}{\frac{n^r}{\log n} - 1}\right)^m \\
\leqslant \;& n^2 \cdot \Big(1 - \frac{\log n}{n^r}\Big)^n \cdot \sum_{m=1}^{n^{1-r}} \left(\frac{n}{\frac{n^r}{\log n} - 1}\right)^{n^{1-r}} \\
\leqslant \;& n^2 \cdot e^{-n^{1-r} \cdot \log n} \cdot n^{1-r} \cdot n^{n^{1-r}} \cdot \Big(\frac{n^r}{\log n} - 1\Big)^{-n^{1-r}}. \qquad \text{(B.3)}
\end{aligned}
$$

Take logarithm of (B.3), then as $n \to \infty$,

$$
\begin{aligned}
\log (B.3) \;& = \; 2 \log n - n^{1-r} \cdot \log n + (1-r) \log n + n^{1-r} \cdot \log n - n^{1-r} \cdot \log \Big(\frac{n^r}{\log n} - 1\Big) \\
& \to \; -\infty.
\end{aligned}
$$

Therefore we have that when $n \to \infty$,

$$\sup_{\alpha\in[0,\infty],s\in[0,\tau_{n;\alpha}]} n^2 P\big(H_{(n)}(s-;\alpha) \leqslant n^{-\kappa/2}\big) \to 0. \tag{B.4}$$

This result together with (B.2) proves the lemma. □

Now we can prove Lemma 3.4.2. Following the definition of $\omega_n(z;\alpha)$ in (3.22), martingale central limit theorem implies

$$E\big[\omega_n(z;\alpha)\big]^2 = G^2(z;\alpha)\int_0^z E\left[\frac{\hat{G}(s-;\alpha)}{G(s;\alpha)}\frac{I[H_{(n)}(s-;\alpha)>0]}{H_{(n)}(s-;\alpha)} - \frac{1}{H(s-;\alpha)}\right]^2 H(s-;\alpha)\Lambda(ds;\alpha). \tag{B.5}$$

We also have that

$$\begin{aligned}
& \int_0^z E\left[\frac{\hat{G}(s-;\alpha)}{G(s;\alpha)}\frac{I[H_{(n)}(s-;\alpha)>0]}{H_{(n)}(s-;\alpha)} - \frac{1}{H(s-;\alpha)}\right]^2 H(s-;\alpha)\Lambda(ds;\alpha) \\
\leqslant\ & \int_0^z 2E\left[\frac{\hat{G}(s-;\alpha)}{G(s;\alpha)}\frac{I[H_{(n)}>0]}{H_{(n)}} - \frac{I[H_{(n)}>0]}{H_{(n)}}\right]^2 H(s-;\alpha)\Lambda(ds;\alpha) \\
& + \int_0^z 2E\left[\frac{I[H_{(n)}>0]}{H_{(n)}} - \frac{1}{H}\right]^2 H(s-;\alpha)\Lambda(ds;\alpha) \\
\leqslant\ & \text{const}\cdot\int_0^z E\left[\frac{\hat{G}(s-;\alpha)-G(s;\alpha)}{G(s;\alpha)}\cdot\frac{I[H_{(n)}>0]}{H_{(n)}}\right]^2 H(s-;\alpha)\Lambda(ds;\alpha) \\
& + \text{const}\cdot\int_0^z E\left[\frac{I[H_{(n)}>0]}{H_{(n)}}\cdot\frac{H-H_{(n)}}{H}\right]^2 H(s-;\alpha)\Lambda(ds;\alpha) \\
& + \text{const}\cdot\int_0^z E\left[\frac{I[H_{(n)}>0]-1}{H}\right]^2 H(s-;\alpha)\Lambda(ds;\alpha) \\
:=\ & I(\alpha;z) + II(\alpha;z) + III(\alpha;z).
\end{aligned} \tag{B.6}$$

For the term $III(\alpha; z)$ in (B.6), we have that

$$\sup_{\alpha,z} III(\alpha; z) \leqslant \text{const} \cdot \sup_{\alpha,z} \int_0^z \frac{1}{H(s-;\alpha)} \cdot E\big\{I[H_{(n)}(s-;\alpha) = 0]\big\}\Lambda(ds;\alpha).$$

Then from Lemma 3.4.1, we have $\sup_{\alpha\in[0,\infty],z\in[0,\tau_{n;\alpha}]} III(\alpha; z) = o(1)$.

For $\tilde{\kappa} \in (0, 1/4)$, the term $II(\alpha; z)$ in (B.6) can be written as

$$\begin{aligned} II(\alpha; z) &= \text{const} \cdot \int_0^z E\left\{\left[\frac{H - H_{(n)}}{H}\right]^2 \cdot \frac{I[H_{(n)} > n^{-\tilde{\kappa}/2}]}{H_{(n)}^2}\right\} H(s-;\alpha)\Lambda(ds;\alpha) \\ &\quad + \text{const} \cdot \int_0^z E\left\{\left[\frac{H - H_{(n)}}{H}\right]^2 \cdot \frac{I[n^{-1} \leqslant H_{(n)} \leqslant n^{-\tilde{\kappa}/2}]}{H_{(n)}^2}\right\} H(s-;\alpha)\Lambda(ds;\alpha) \\ &:= II(\alpha; z)_{(1)} + II(\alpha; z)_{(2)}. \end{aligned} \tag{B.7}$$

For the term $II(\alpha; z)_{(1)}$, we have that for $\alpha \in [0, \infty], z \in [0, \tau_{n;\alpha}]$,

$$\begin{aligned} \sup_{\alpha,z} II(\alpha; z)_{(1)} &\leqslant \text{const} \cdot \sup_{\alpha,z} \int_0^z E\left[\frac{H - H_{(n)}}{H} \cdot n^{\tilde{\kappa}/2}\right]^2 H(s-;\alpha)\Lambda(ds;\alpha) \\ &= \text{const} \cdot \sup_{\alpha,z} \int_0^z n^{\tilde{\kappa}-1} \cdot \left[\frac{1}{H(s-;\alpha)} - 1\right] H(s-;\alpha)\Lambda(ds;\alpha) \\ &= o(1). \end{aligned} \tag{B.8}$$

In addition, for the term $II(\alpha;z)_{(2)}$, we have that

$$\begin{aligned}
&\sup_{\alpha\in[0,\infty],z\in[0,\tau_\alpha]} II(\alpha;z)_{(2)} \\
\leqslant\ & \text{const}\cdot\sup_{\alpha,z} n^2\int_0^z E\left\{\left[\frac{H-H_{(n)}}{H}\right]^2\cdot I\left[n^{-1}\leqslant H_{(n)}\leqslant n^{-\tilde{\kappa}/2}\right]\right\}H(s-;\alpha)\Lambda(ds;\alpha) \\
\leqslant\ & \text{const}\cdot\sup_{\alpha,z} n^2\int_0^z H\cdot E\left\{I[n^{-1}\leqslant H_{(n)}\leqslant n^{-\tilde{\kappa}/2}]\right\}\Lambda(ds;\alpha) \\
& +\text{const}\cdot\sup_{\alpha,z} n^2\int_0^z 2E\left\{H_{(n)}\cdot I[n^{-1}\leqslant H_{(n)}\leqslant n^{-\tilde{\kappa}/2}]\right\}\Lambda(ds;\alpha) \\
& +\text{const}\cdot\sup_{\alpha,z} n^2\int_0^z \frac{1}{H}E\left\{H_{(n)}^2\cdot I[n^{-1}\leqslant H_{(n)}\leqslant n^{-\tilde{\kappa}/2}]\right\}\Lambda(ds;\alpha) \\
\leqslant\ & \text{const}\cdot\sup_{\alpha,z} n^2\int_0^z H\cdot P\left(n^{-1}\leqslant H_{(n)}\leqslant n^{-\tilde{\kappa}/2}\right)\Lambda(ds;\alpha) \\
& +\text{const}\cdot\sup_{\alpha,z} n^2\int_0^z 2n^{-\tilde{\kappa}/2}\cdot P\left(n^{-1}\leqslant H_{(n)}\leqslant n^{-\tilde{\kappa}/2}\right)\Lambda(ds;\alpha) \\
& +\text{const}\cdot\sup_{\alpha,z} n^2\int_0^z \frac{1}{H}\cdot n^{-\tilde{\kappa}}\cdot P\left(n^{-1}\leqslant H_{(n)}\leqslant n^{-\tilde{\kappa}/2}\right)\Lambda(ds;\alpha). \qquad \text{(B.9)}
\end{aligned}$$

Using similar arguments as Lemma B.2.1, $\sup_{\alpha\in[0,\infty],z\in[0,\tau_{n;\alpha}]} II(\alpha;z)_{(2)} = o(1)$. Together with (B.8), we have that $\sup_{\alpha\in[0,\infty],z\in[0,\tau_{n;\alpha}]} II(\alpha;z) = o(1)$.

The term $I(\alpha;z)$ in (B.6) can be written as

$$\begin{aligned}
I(\alpha;z) &= \text{const}\cdot\int_0^z E\left\{\left[\frac{\hat{G}-G}{G}\right]^2\cdot\frac{I[H_{(n)}>n^{-\tilde{\kappa}/2}]}{H_{(n)}^2}\right\}H(s-;\alpha)\Lambda(ds;\alpha) \\
&\quad +\text{const}\cdot\int_0^z E\left\{\left[\frac{\hat{G}-G}{G}\right]^2\cdot\frac{I[n^{-1}\leqslant H_{(n)}\leqslant n^{-\tilde{\kappa}/2}]}{H_{(n)}^2}\right\}H(s-;\alpha)\Lambda(ds;\alpha) \\
&:= I(\alpha;z)_{(1)}+I(\alpha;z)_{(2)}. \qquad \text{(B.10)}
\end{aligned}$$

Since $\frac{\hat{G}(s-;\alpha)-G(s;\alpha)}{G(s;\alpha)}$ is bounded, we have $\sup_{\alpha\in[0,\infty],z\in[0,\tau_{n;\alpha}]} I(\alpha;z)_{(2)} = o(1)$. More-

over, we have that

$$\sup_{\alpha,z} I(\alpha;z)_{(1)} \leqslant \text{const} \cdot \int_0^z n^{\tilde{\kappa}} \cdot \sup_{\alpha,s} E\left[\frac{\hat{G}-G}{G}\right]^2 H(s-;\alpha)\Lambda(ds;\alpha), \tag{B.11}$$

where

$$\begin{aligned}
& \sup_{\alpha,s} E\left[\frac{\hat{G}(s-;\alpha)-G(s;\alpha)}{G(s;\alpha)}\right]^2 \\
= & \sup_{\alpha,s} E\left\{\int_0^s \frac{\hat{G}(u-;\alpha)}{G(u;\alpha)} \frac{I[H_{(n)}(u-;\alpha)>0]}{H_{(n)}(u-;\alpha)} M(ds;\alpha) + B(s;\alpha)\right\}^2 \\
\leqslant & 2\sup_{\alpha,s} E\left\{\int_0^s \frac{\hat{G}(u-;\alpha)}{G(u;\alpha)} \frac{I[H_{(n)}(u-;\alpha)>0]}{H_{(n)}(u-;\alpha)} M(ds;\alpha)\right\}^2 + 2\sup_{\alpha,s} E\big[B(s;\alpha)\big]^2 \\
= & \frac{2}{n} \cdot \sup_{\alpha,s} \int_0^s E\left\{\frac{\hat{G}(u-;\alpha)}{G(u;\alpha)} \frac{I[H_{(n)}(u-;\alpha)>0]}{H_{(n)}(u-;\alpha)}\right\}^2 H(u-;\alpha)\Lambda(du;\alpha) \\
& +2\sup_{\alpha,s} E\big[B(s;\alpha)\big]^2.
\end{aligned} \tag{B.12}$$

Then using the result in Lemma B.2.1, together with $\hat{G}/G$ is bounded and

$$\sup_{\alpha\in[0,\infty],z\in[0,\tau_{n;\alpha}]} I(\alpha;z)_{(1)} = o(1),$$

we have that $\sup_{\alpha\in[0,\infty],z\in[0,\tau_{n;\alpha}]} I(\alpha;z) \leqslant O(n^{2\tilde{\kappa}-1})$. Therefore, for $\alpha \in [0,\infty], z \in [0,\tau_{n;\alpha}]$,

$$\sup_{\alpha,z} E\big[\omega_n(z;\alpha)\big]^2 = G^2(z;\alpha) \cdot \sup_{\alpha,z}\big[I(\alpha;z) + II(\alpha;z) + III(\alpha;z)\big] = o(1).$$

The lemma is proved.

Appendix C

Proofs in Section 3.5

C.1 Lemma in Fleming and Harrington (1991)

Lemma C.1.1. *Let E be an open convex subset of R^p, and let $F_1, F_2, \ldots,$ be a sequence of random concave functions on E and f a real-valued function on E such that, for all $x \in E$,*

$$\lim_{n\to\infty} F_n(x) = f(x)$$

in probability. Then:

1. *The function f is concave.*

2. *For all compact subsets A of E,*

$$\sup_{x\in A} |F_n(x) - f(x)| \to 0$$

   in probability, as $n \to \infty$.

3. *If F_n has a unique maximum at X_n and f has one at x, then $X_n \to x$ in*

probability as $n \to \infty$.

C.2 Proof of Theorem 3.5.1

Let $\boldsymbol{\beta}^*$ be the true value of $\boldsymbol{\beta}$, then

$$\sqrt{n}\left[\boldsymbol{Q}(\boldsymbol{\beta}^*;\hat{G})\right] = \sqrt{n}\left[\boldsymbol{Q}\left(\boldsymbol{\beta}^*;G\right)\right] + \sqrt{n}\left[\boldsymbol{Q}(\boldsymbol{\beta}^*;\hat{G}) - \boldsymbol{Q}\left(\boldsymbol{\beta}^*;G\right)\right]. \tag{C.1}$$

The first term of the right side of (C.1) is a sum of i.i.d. variables since

$$\sqrt{n}\left[\boldsymbol{Q}\left(\boldsymbol{\beta}^*;G\right)\right] = n^{-1/2}\sum_{i=1}^{n}\frac{\delta_i^*}{G(R_i^*-,X_i^*-)}(X_i^* - \boldsymbol{W}_i^*\boldsymbol{\beta}^*)\boldsymbol{W}_i^{*tr}.$$

Now we focus on the second term of the right side of (C.1), which can be written as

$$\begin{aligned}
& \sqrt{n}\left[\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big) - \boldsymbol{Q}\left(\boldsymbol{\beta}^*;G\right)\right] \\
= \;& n^{-1/2}\left\{\sum_{i=1}^{n}\left[\frac{I\big[\hat{G}(R_i^*-,X_i^*-)>0\big]}{\hat{G}(R_i^*-,X_i^*-)} - \frac{1}{G(R_i^*-,X_i^*-)}\right]\delta_i^*(X_i^* - \boldsymbol{W}_i^*\boldsymbol{\beta}^*)\boldsymbol{W}_i^{*tr}\right\} \\
= \;& n^{-1/2}\sum_{i=1}^{n}\left[\frac{G(R_i^*-,X_i^*-) - \hat{G}(R_i^*-,X_i^*-)}{G^2(R_i^*-,X_i^*-)}\right]\boldsymbol{\eta}_i \\
+ \;& n^{-1/2}\sum_{i=1}^{n}\frac{G(R_i^*-,X_i^*-) - \hat{G}(R_i^*-,X_i^*-)}{G(R_i^*-,X_i^*-)}\left[\frac{I\big[\hat{G}(R_i^*-,X_i^*-)>0\big]}{\hat{G}(R_i^*-,X_i^*-)} - \frac{1}{G(R_i^*-,X_i^*-)}\right]\boldsymbol{\eta}_i \\
:= \;& I + II.
\end{aligned} \tag{C.2}$$

Using the result in (3.27) we have that as $n \to \infty$, II in (C.2) should converge to 0 in probability. On the other hand, based on the observed transformed data

defined in (3.12) and the result in (3.27), term I in (C.2) can be written as

$$
\begin{aligned}
& n^{-3/2}\sum_{i=1}^{n}\frac{1}{G\Big(\tilde{Z}_i^*(\alpha_i)-;\alpha_i\Big)}\left[\sum_{j=1}^{n}\int_{s<\tilde{Z}_i^*(\alpha_i)}\frac{1}{H(s-;\alpha_i)}M_j(ds;\alpha_i)+r_n\Big(\tilde{Z}_i^*(\alpha_i)-;\alpha_i\Big)\right]\boldsymbol{\eta}_i \\
&= n^{-3/2}\sum_{i=1}^{n}\frac{1}{G\Big(\tilde{Z}_i^*(\alpha_i)-;\alpha_i\Big)}\left(\sum_{j=1}^{n}\xi_{ji}\right)\boldsymbol{\eta}_i, \\
&= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\xi_{ji}\boldsymbol{\eta}_i}{G\Big(\tilde{Z}_i^*(\alpha_i)-;\alpha_i\Big)}, \\
&= n^{-3/2}\sum_{i,j=1}^{n}G^{-1}(R_i^*-,X_i^*-)\xi_{ji}\boldsymbol{\eta}_i, \\
&:= \boldsymbol{U_n}, \qquad\qquad (C.3)
\end{aligned}
$$

where $\boldsymbol{\eta}_i = \delta_i^*(X_i^* - \boldsymbol{W}_i^*\boldsymbol{\beta}^*)\boldsymbol{W}_i^{*tr}$, $\xi_{ji} = \int_{s<\tilde{Z}_i^*(\alpha_i)}\frac{1}{H(s-;\alpha_i)}M_j(ds;\alpha_i)$, and $\boldsymbol{U_n} = n^{-3/2}\sum_{i,j=1}^{n}G^{-1}(R_i^*-,X_i^*-)\xi_{ji}\boldsymbol{\eta}_i$.

According to the properties of U-statistics Serfling (1980) we have $\boldsymbol{U_n} = \widehat{\boldsymbol{U}}_{\boldsymbol{n}} + o\big(n^{-1}(\log n)^{\rho}\big)$ for some $\rho > 0$, where

$$
\begin{aligned}
\widehat{\boldsymbol{U}}_{\boldsymbol{n}} &= \sum_{k=1}^{n}E\left\{\boldsymbol{U_n}\,|\mathcal{D}_k\right\} \\
&= \sum_{k=1}^{n}E\Big\{n^{-3/2}\sum_{i,j=1}^{n}G^{-1}(R_i^*-,X_i^*-)\xi_{ji}\boldsymbol{\eta}_i\,|\mathcal{D}_k\Big\} \\
&= n^{-3/2}\sum_{i,j,k=1}^{n}E\big\{G^{-1}(R_i^*-,X_i^*-)\xi_{ji}\boldsymbol{\eta}_i\big|\mathcal{D}_k\big\}. \qquad (C.4)
\end{aligned}
$$

Since ξ_{ji}, defined in (3.35), is a zero-mean martingale with $\mathcal{D}_i$ given, we have that $E\{\xi_{ji}|\mathcal{D}_i\} = 0$ (Dai and Fu, 2012).

Thus if $i = k, j = k$, then

$$E\big\{G^{-1}(R_i^*-, X_i^*-)\xi_{ji}\boldsymbol{\eta}_i\big|\mathcal{D}_k\big\} = G^{-1}(R_k^*-, X_k^*-)\boldsymbol{\eta}_k E\{\xi_{kk}|\mathcal{D}_k\} = \mathbf{0},$$

since $E\{\xi_{kk}|\mathcal{D}_k\} = 0$.

If $i = k, j \neq k$, then

$$E\big\{G^{-1}(R_i^*-, X_i^*-)\xi_{ji}\boldsymbol{\eta}_i\big|\mathcal{D}_k\big\} = G^{-1}(R_k^*-, X_k^*-)\boldsymbol{\eta}_k E\{\xi_{jk}|\mathcal{D}_k\} = \mathbf{0},$$

since $E\{\xi_{jk}|\mathcal{D}_k\} = 0$.

If $i \neq k, j \neq k$, then

$$\begin{aligned}
& E\big\{G^{-1}(R_i^*-, X_i^*-)\xi_{ji}\boldsymbol{\eta}_i\big|\mathcal{D}_k\big\} \\
= \;& E\big\{G^{-1}(R_i^*-, X_i^*-)\xi_{ji}\boldsymbol{\eta}_i\big\} \\
= \;& E\Big\{E\big[G^{-1}(R_i^*-, X_i^*-)\xi_{ji}\boldsymbol{\eta}_i\big|\mathcal{D}_i\big]\Big\} \\
= \;& E\Big\{G^{-1}(R_i^*-, X_i^*-)\boldsymbol{\eta}_i E\big[\xi_{ji}\big|\mathcal{D}_i\big]\Big\} \\
= \;& \mathbf{0},
\end{aligned}$$

since $E\{\xi_{ji}|\mathcal{D}_i\} = 0$.

Then we have that,

$$\begin{aligned}
\widehat{\boldsymbol{U}}_{\boldsymbol{n}} &= n^{-1/2}\sum_{k=1}^{n}\frac{n-1}{n}E\left\{G^{-1}(R_i^*-, X_i^*-)\xi_{ki}\boldsymbol{\eta}_i\,|\mathcal{D}_k\right\} \\
&= n^{-1/2}\sum_{k=1}^{n}E\left\{G^{-1}(R_i^*-, X_i^*-)\xi_{ki}\boldsymbol{\eta}_i\,|\mathcal{D}_k\right\} + o_p(1) \\
&= n^{-1/2}\sum_{k=1}^{n}\boldsymbol{\Phi}(\mathcal{D}_k) + o_p(1). \qquad \text{(C.5)}
\end{aligned}$$

Hence in (C.2), term $I = n^{-1/2}\sum_{k=1}^{n}\boldsymbol{\Phi}(\mathcal{D}_k) + o_p(1)$. This, together with $II \xrightarrow{p} 0$, implies that $\sqrt{n}\left[\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big) - \boldsymbol{Q}\left(\boldsymbol{\beta}^*;G\right)\right] = n^{-1/2}\sum_{k=1}^{n}\boldsymbol{\Phi}(\mathcal{D}_k) + o_p(1)$, which is a sum of i.i.d. terms. Therefore, we have that

$$\sqrt{n}\Big[\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big)\Big] = n^{-1/2}\sum_{k=1}^{n}\left[\frac{\boldsymbol{\eta}_i}{G(R_k^*-,X_k^*-)} + \boldsymbol{\Phi}(\mathcal{D}_k)\right]. \tag{C.6}$$

Then the variance-covariance matrix of $\sqrt{n}\Big[\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big)\Big]$ is given by

$$\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{Q}} &= \mathrm{Var}\left\{\sqrt{n}\Big[\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big)\Big]\right\},\\
&= \mathrm{Var}\left\{\sqrt{n}\left[\boldsymbol{Q}(\boldsymbol{\beta}^*;G)\right] + \sqrt{n}\left[\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big) - \boldsymbol{Q}\left(\boldsymbol{\beta}^*;G\right)\right]\right\},\\
&= \mathrm{Var}\left[\frac{\boldsymbol{\eta}_i}{G(R_i^*-,X_i^*-)} + \boldsymbol{\Phi}(\mathcal{D}_i)\right],
\end{aligned}$$

where $i = 1,\ldots,n$.

Then following from the first-order Taylor expansion,

$$\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big) = \boldsymbol{Q}\big(\hat{\boldsymbol{\beta}};\hat{G}\big) + \boldsymbol{Q}'\big(\hat{\boldsymbol{\beta}};\hat{G}\big)\big(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\big) + O\big((\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}})^2\big), \tag{C.7}$$

where $\boldsymbol{\beta}^*$ denotes the true value of $\boldsymbol{\beta}$. Since $\hat{\boldsymbol{\beta}}$ is a root of $\boldsymbol{Q}\big(\boldsymbol{\beta};\hat{G}\big) = 0$, we have $\boldsymbol{Q}\big(\hat{\boldsymbol{\beta}};\hat{G}\big) = 0$. Together with $\hat{\boldsymbol{\beta}}$ is a consistent estimate of $\boldsymbol{\beta}$, we have that

$$\sqrt{n}\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big) = \sqrt{n}\boldsymbol{Q}'\big(\hat{\boldsymbol{\beta}};\hat{G}\big)\big(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\big). \tag{C.8}$$

Hence

$$\mathrm{Var}\Big\{\sqrt{n}\big(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\big)\Big\} = \mathrm{Var}\Big\{\sqrt{n}\big[\boldsymbol{Q}'\big(\hat{\boldsymbol{\beta}};\hat{G}\big)\big]^{-1}\boldsymbol{Q}\big(\boldsymbol{\beta}^*;\hat{G}\big)\Big\}. \tag{C.9}$$

Therefore we have $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \left[\boldsymbol{Q}'(\hat{\boldsymbol{\beta}}; \hat{G})\right]^{-1} \boldsymbol{\Sigma}_{\boldsymbol{Q}} \left\{ \left[\boldsymbol{Q}'(\hat{\boldsymbol{\beta}}; \hat{G})\right]^{-1} \right\}^{tr}$, where

$$\boldsymbol{\Sigma}_{\boldsymbol{Q}} = \mathrm{Var}\left[\frac{\boldsymbol{\eta}_k}{G(R_k^*-, X_k^*-)} + \boldsymbol{\Phi}(\mathcal{D}_k) \right], \tag{C.10}$$

$$\boldsymbol{\Phi}(\mathcal{D}_k) = E\big\{ G^{-1}(R_i^*-, X_i^*-) \xi_{ki} \boldsymbol{\eta}_i \,|\mathcal{D}_k \big\},\ i \neq k. \tag{C.11}$$

The theorem is proved.

Appendix D

Proof in Chapter 4

From (4.3) we have that

$$g\left\{1 - F_T(t|\boldsymbol{W})\right\} = h(t) + \boldsymbol{W}\boldsymbol{\beta},$$

and then

$$1 - g^{-1}\big(h(t) + \boldsymbol{W}\boldsymbol{\beta}\big) = F_T(t|\boldsymbol{W}) = Pr(T \leqslant t|\boldsymbol{W}).$$

Since $h(\cdot)$ is a strictly increasing function, we have that

$$1 - g^{-1}\big(h(t) + \boldsymbol{W}\boldsymbol{\beta}\big) = Pr\big(h(T) + \boldsymbol{W}\boldsymbol{\beta} \leqslant h(t) + \boldsymbol{W}\boldsymbol{\beta}\big|\boldsymbol{W}\big). \tag{D.1}$$

If we let $\varepsilon = h(T) + \boldsymbol{W}\boldsymbol{\beta}$, (D.1) can be written as

$$F_\varepsilon\big(h(t) + \boldsymbol{W}\boldsymbol{\beta}\big) = 1 - g^{-1}\big(h(t) + \boldsymbol{W}\boldsymbol{\beta}\big), \tag{D.2}$$

which leads to the model in (4.4).

References

Akritas, M. G. and Keilegom, I. V. (2003). Estimation of bivariate and marginal distributions with censored data. *Journal of Royal Statistical Society*, 65:457–471.

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

Bao, Y., He, S., and Mei, C. (2007). The koul-susarla-van ryzin and weighted least squares estimates for censored linear regression model: A comparative study. *Computational Statistics & Data Analysis*, 51:6488–6497.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2:273–277.

Box, G. and Cox, D. (1964). An analysis of transforamtions (with discussion). *Journal of Royal Statistical Society*, B 26:211–252.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66:429–436.

Burke, M. D. (1988). Estimation of a bivariate distribution function under random censorship. *Biometrika*, 75:379–382.

Campbell, G. (1981). Nonparametric bivariate estimation with randomly censored data. *Biometrika*, 68:417–422.

Campbell, G. and Földes, A. (1982). Large-sample properties of nonparametric bivariate estimatiors with censored data. *Nonparametric Statistical Inference Colloquia Mathematica Societatis Janos Bolyai*, 32:103–121.

Chen, K., Jin, Z., and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika*, 89:659–668.

Cheng, S., Wei, L., and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, 82:659–668.

Cheng, S., Wei, L., and Ying, Z. (1997). Predicting survival probabilities with semiparametric transformation models. *Journal of the American Statistical Association*, 92:227–235.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiology studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151.

Clayton, D. and Cuzick, J. (1986). The semiparametric pareto model for regression analysis of survival times. *In Proceedings of the 45th Session of the International Statistical Institute*, 23.3:1–18.

Collett, D. (2003). *Modelling Survival Data in Medical Research*. Chapman & Hall, second edition.

Cox, D. (1972). Regression models and life-tables. *Journal of Royal Statistical Society*, 34:187–202.

Cox, D. (1975). Partial likelihood. *Biometrika*, 62:269–276.

Cuzick, J. (1988). Rank regression. *The Annals of Statistics*, 16:1369–1389.

Dabrowska, D. (1988). Kaplan-meier estimate on the plane. *The Annals of Statistics*, 16:1475–1489.

Dabrowska, D. (1989). Kaplan-meier estimate on the plane: weak convergence, lil, and the bootstrap. *Journal of Multivariate Analysis*, 29:308–325.

Dabrowska, D. and Doksum, K. (1988). Estimation and testing in the two-sample generalized odds-rate model. *Journal of the American Statistical Association*, 83:744–749.

Dabrowska, D. and Doksum, K. (1989). Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics*, 15:1–23.

Dai, H. and Bao, Y. (2009). An inverse probability weighted estimator for the bivariate distribution function under right censoring. *Statistics and Probability Letters*, 79:1789–1797.

Dai, H. and Fu, B. (2012). A polar coordinate transformation for estimating bivariate survival functions with randomly censored and truncated data. *Journal of Statistical Planning and Inference*, 142:248–262.

Dore, G. J., Freeman, A. J., Law, M., and Kaldor, J. M. (2002). Is severe liver disease a common outcome for people with chronic hepatitis c. *Journal of Gastroenterology and Hepatology*, 17:423–430.

Fine, J., Ying, Z., and Wei, L. (1998). On the linear transformation model with censored data. *Biometrika*, 85:980–986.

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*. John Wiley & Sons.

Freeman, A., Dore, G., Law, M., Thorpe, M., Overbeck, J., Lloyd, A., Marinos, G., and Kaldor, J. M. (2001). Estimating progression to cirrhosis in chronic hepatitis c virus infection. *Hepatology*, 34:809–816.

Fu, B., Tom, B. D., Delahooke, T., Alexander, G. J., and Bird, S. M. (2007). Event-biased referral can distort estimation of hepatitis c virus progression rate to cirrhosis and of prognostic influences. *Journal of Clinical Epidemiology*, 60:1140–1148.

Fu, B., Wang, W., and Shi, X. (2011). Impact of delayed diagnosis time in estimating progression rates to hepatitis c virus-related cirrhosis and death. *Statistical Methods in Medical Research*, 0:1–15.

Fygenson, M. and Zhou, M. (1994). On using stratification in the analysis of linear regression models with right censoring. *The Annals of Statistics*, 22:747–762.

Gilbels, I. and Gürler, U. (1998). Covariance function of a bivariate distribution function estimator for left truncated and right censored data. *Statistica Sinica*, 1998:1219–1232.

Gross, S. T. and Huber-Carol, C. (1992). Regression models for truncated survival data. *Scandinavian Journal of Statistics*, 19:193–213.

Gross, S. T. and Lai, T. L. (1996). Nonparametric estimation and regression analysis with left-truncated and right-censored data. *Journal of American Statistical Association*, 91:1166–1180.

Gürler, U. (1996). Bivariate estimation with right truncated data. *Journal of American Statistical Association*, 91:1152–1165.

Gürler, U. (1997). Bivariate distribution and hazard functions when a component is randomly truncated. *Journal of Multivariate Analysis*, 60:20–47.

He, S. and Wong, X. (2003). The central limit theorem of linear regression model under right censorship. *Science in China Series A*, 46:600–610.

He, S. and Yang, G. (2003). Estimation of regression parameters with left truncated data. *Journal of Statistical Planning and Inference*, 117:99–122.

Heller, G. and Simonoff, J. (1990). A comparison of estimators for regression with a censored response variable. *Biometrics*, 48:101–115.

Huang, J., Vieland, V., and Wang, K. (2001). Nonparametric estimation of marginal distributions under bivariate truncation with application to testing for age-of-onset anticipation. *Statistica Sinica*, 11:1047–1068.

Jin, Z., Lin, D., and Ying, Z. (2006). On least-squares regression with censored data. *Biometrika*, 93:147–161.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York, second edition.

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

Kim, J. H. (1993). Chi-square goodness-of-fit tests for randomly censored data. *The Annals of Statistics*, 21.

Koul, H., Susarla, V., and Van, R. (1981). Regression analysis with randomly right censored data. *The Annals of Statistics*, 9:1276–1288.

Lai, T. and Ying, Z. (1991). Estimating a distribution function with truncated and censored data. *The Annals of Statistics*, 19(1):417–442.

Lin, D. and Ying, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, 80:572–581.

Miller, R. G. (1976). Least square regression with censored data. *Biometrika*, 63:449–464.

Miller, R. G. and Harplen, J. (1982). Regression with censored data. *Biometrika*, 69:521–531.

Pettitt, A. (1982). Inference for the linear model using a likelihood based on ranks. *Journal of Royal Statistical Society*, B 44:234–43.

Pettitt, A. (1984). Proportional odds model for survival data and estimates using ranks. *Applied Statistics*, 33:169–175.

Prentice, R., Moodie, F., and Wu, J. (2004). Hazard-based nonparametric survivor function estimation. *Journal of Royal Statistical Society*, B(66):305–319.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.

Sharma, N. K. and Sherker, A. H. (2009). Epidemiology, risk factors, and natural history of chronic hepatitis c. *Clinical Gastroenterology*, pages 33–70. edited by K. Shetty and G.Y. Wu.

Shen, P. (2006). An inverse-probability-weighted approach to estimation of the bivariate survival function under left-truncation and right-censoring. *Journal of Statistical Planning and Inference*, 136:4365–4384.

Shen, P. and Yan, Y. (2008). Nonparametric estimation of the bivariate survival function with left-truncated and right-censored data. *Journal of Statistical Planning and Inference*, 138:4041–4054.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, 45:89–103.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian Journal of Statistics*, 23(4):461–471.

Sweeting, M., Angelis, D. D., Neal, K., Ramsay, M., Irving, W., and et al., M. W. (2006). Estimated progression rates in three united kingdom hepatitis c cohorts differed according to method of recruitment. *Journal of Clinical Epidemiology*, 59:144–152.

Tsai, W., Leurgan, S., and Crowley, J. (1990). Nonparametric estimation of a bivariate survival function in the presence of censoring. *The Annals of Statistics*, 14:1351–1365.

van der Laan, M. (1996a). Efficient estimation in the bivariate censoring model and repairing npmle. *The Annals of Statistics*, 24:596–627.

van der Laan, M. (1996b). Nonparametric estimation of the bivariate survival function with truncated data. *Journal of Multivariate Analysis*, 58:107–131.

Wang, H., Dai, H., and Fu, B. (2013). Accelerated failure time models for censored survival data under referral bias. *Biostatistics*, 14:313–326.