

SEMI-PARAMETRIC REGRESSION ANALYSIS OF INTERVAL-CENSORED FAILURE TIME DATA

A Dissertation presented to
the Faculty of the Graduate School
at the University of Missouri

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
LING MA
Dr. (Tony) Jianguo Sun, Dissertation Supervisor
JULY 2014

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

SEMI-PARAMETRIC REGRESSION ANALYSIS OF
INTERVAL-CENSORED FAILURE TIME DATA

presented by Ling Ma,

a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. (Tony) Jianguo Sun

Dr. Nancy Flournoy

Dr. Subharup Guha

Dr. Tieming Ji

Dr. (Shawn) Xiaoguang Ni

ACKNOWLEDGMENTS

I owe my deepest gratitude to my esteemed advisor Dr. (Tony) Jianguo Sun for his wonderful guidance, generous support and endless patience throughout the development of this dissertation. It was his continual encouragement and inspiring instruction that guaranteed the accomplishment of this work and led me to a good start of my research life.

I extend my gratitude to the members of my advisory committee members: Dr. Nancy Flournoy, Dr. Subharup Guha, Dr. Tieming Ji and Dr. (Shawn) Xiaoguang Ni for their continual support and insightful comments and suggestions on my work. Special thanks are due to Dr. Yanqin Feng and Dr. Tao Hu for their great academic support and help.

I truly appreciate all persons in our department that make my graduate life at Mizzou memorable. I want to express my deep gratitude to all faculty for providing so many great courses and assisted me in various ways during my course studies. I want to thank Dr. Ries for helping me to be a better instructor. I want to thank our great staff Judy, Kathleen and Tracy for their help. I would also like to take this opportunity to extend many thanks to my friends for being supportive and bringing so much fun to my life here. I appreciate all your help and will always treasure the friendship.

Finally, my deepest gratitude goes to my family, especially my parents, for their endless love and support. It is their care and dedication that make my life so happy and wonderful.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
LIST OF TABLES	vii
LIST OF FIGURES	ix
ABSTRACT	x
CHAPTER	
1 Introduction	1
1.1 Data Structure	1
1.1.1 Interval-Censored Data	1
1.1.2 Non-informative and Informative Censoring	3
1.1.3 Two Examples	5
1.2 Some Commonly Used Models	7
1.3 Regression Analysis of Interval-Censored Data	10
1.3.1 Regression Analysis of Current Status Data	10
1.3.2 Regression Analysis of Interval-Censored Data	12
1.4 Outline of The Dissertation	14
2 Empirical Analysis of Interval-Censored Failure Time Data with Linear Transformation Models	16
2.1 Introduction	16
2.2 Linear Transformation Models and Estimation	18
2.3 Empirical Likelihood Procedure	21

2.4	Numerical Studies	27
2.5	Discussion	29
3	Regression Analysis of Current Status Data under the Additive Hazards Model with Auxiliary Covariates	33
3.1	Introduction	33
3.2	Notation, Models and Assumptions	36
3.3	Estimation Procedures	38
3.4	Numerical Studies	44
3.5	Discussion and Concluding Remarks	48
4	Spline-Based Regression Analysis of Current Status Data with Dependent Censoring	59
4.1	Introduction	59
4.2	Cox Models and Likelihood	61
4.3	Spline-Based Sieve Semiparametric Maximum Likelihood Estimation	63
4.4	Simulation Studies	66
4.5	An Application	68
4.6	Concluding Remarks	70
5	Sieve Maximum Likelihood Regression Analysis of Dependent Interval-Censored Data	76
5.1	Introduction	76
5.2	Notation, Models and Likelihood Function	78
5.3	Sieve Semiparametric Maximum Likelihood Estimation	81
5.4	Simulation Studies	84

5.5	An Application	86
5.6	Concluding Remarks	87
6	Future Research	94
6.1	Multivariate Copula Models for Regression Analysis of Informatively Interval-Censored Data	94
6.2	Cure Rate Models for Interval-Censored Data with Informative Censoring	95
APPENDIX		
A	Theoretical proofs	96
A.1	Proofs of the Main Results in Chapter 2	96
A.1.1	Proof of Theorem 2.1	96
A.1.2	Proof of Theorem 2.2	101
A.2	Proofs of the Main Results in Chapter 3	103
A.2.1	Proof of Theorem 3.1	104
A.2.2	Proof of Theorem 3.2	107
A.3	Proofs of the Asymptotic Properties in Chapter 4	109
A.3.1	Proof of Theorem 4.1	110
A.3.2	Proof of Theorem 4.2	113
A.3.3	Proof of Theorem 4.3	114
A.4	Proofs of the Asymptotic Properties in Chapter 5	118
A.4.1	Proof of Theorem 5.1	119
A.4.2	Proof of Theorem 5.2	121
A.4.3	Proof of Theorem 5.3	122
BIBLIOGRAPHY		127

VITA 135

LIST OF TABLES

Table		Page
2.1	Empirical coverage probabilities based on simulated data with $\alpha = 0$.	31
2.2	Empirical coverage probabilities based on simulated data with $\alpha = 0.5$	32
2.3	Empirical coverage probabilities based on simulated data with $\alpha = 1$.	32
2.4	Analysis results of the breast cancer study	32
3.1	Estimation of β based on simulated data with $Z \sim B(1, 0.5)$	52
3.2	Estimation of β based on simulated data with $Z \sim U(0, \sqrt{3})$	53
3.3	Estimation of β based on simulated data with $Z \sim N(0, 0.25)$	54
3.4	Analysis results for the mice example	55
4.1	Simulation results for Clayton copula	71
4.2	Simulation results for Gumbel copula	72
4.3	Simulation results for FGM copula	73
4.4	Simulation results for Frank copula	74
4.5	Estimating results for $k_n = 5$	75
4.6	Estimating results for $k_n = 3$	75
5.1	Estimation of regression parameters under the Gumbel copula	89

5.2	Estimation of regression parameters under the FGM copula	90
5.3	Estimation of regression parameters under the Frank copula	91
5.4	Estimating results for $k_n = 3$	92
5.5	Estimating results for $k_n = 4$	93

LIST OF FIGURES

Figure	Page
3.1 QQ-plot of $\hat{\beta}_{EP}$ corresponding to Table 3.1	56
3.2 QQ-plot of $\hat{\beta}_{EP}$ corresponding to Table 3.2	57
3.3 QQ-plot of $\hat{\beta}_{EP}$ corresponding to Table 3.3	58

ABSTRACT

By interval-censored data, we mean that the failure time of interest is known only to lie within an interval instead of being observed exactly. Many clinical trials and longitudinal studies may generate interval-censored data. One common example occurs in medical or health studies that entail periodic follow-ups. An important special case of interval-censored data is the so called current status data when each subject is observed only once for the status of the occurrence of the event of interest. That is, instead of observing the survival endpoint directly, we only know the observation time and whether or not the event of interest has occurred at that time. Such data may occur in many fields, for example, cross-sectional studies and tumorigenicity experiments. Sometimes we also refer current status data to as case I interval-censored data and the general case as case II interval-censored data. In the following, for simplicity, we will refer current status data and interval-censored data to case I and case II interval-censored data, respectively.

The statistical analysis of both case I and case II interval-censored failure time data has recently attracted a great deal of attention and especially, many procedures have been proposed for their regression analysis under various models. However, due to the strict restrictions of existing regression analysis procedures and practical demands, new methodologies for regression analysis need to be developed.

For regression analysis of interval-censored data, many approaches have been proposed and for most of them, the inference is carried out based on the asymptotic normality. It's well known that the symmetric property implied by the normal distribution may not be appropriate sometimes and could underestimate the variance of

estimated parameters. In the first part of this dissertation, we adopt the linear transformation models for regression analysis of interval-censored data and propose an empirical likelihood-based procedure to address the underestimating problem from using symmetric property implied by the normal distribution of the parameter estimates. Simulation and analysis of a real data set are conducted to assess the performance of the procedure.

The second part of this dissertation discusses regression analysis of current status data under additive hazards models. In this part, we focus on the situation when some covariates could be missing or cannot be measured exactly due to various reasons. Furthermore, for missing covariates, there may exist some related information such as auxiliary covariates (Zhou and Pepe, 1995). We propose an estimated partial likelihood approach for estimation of regression parameters that make use of the available auxiliary information. To assess the finite sample performance of the proposed method, an extensive simulation study is conducted and indicates that the method works well in practical situations.

Several semi-parametric and non-parametric methods have been proposed for the analysis of current status data. However, most of these methods deal only with the situation where observation time is independent of the underlying survival time completely or given covariates. The third part of this dissertation discusses regression analysis of current status data when the observation time may be related to survival time. The correlation between observation time and survival time and the covariate effects are described by a copula model and the proportional hazards model, respectively. For estimation, a sieve maximum likelihood procedure with the use of monotone I -spline functions is proposed and the proposed method is examined

through a simulation study and illustrated with a real data set.

In the fourth part of this dissertation, we discuss the regression analysis of interval-censored data where the censoring mechanism could be related to the failure time. We consider a situation where the failure time depend on the censoring mechanism only through the length of the observed interval. The copula model and monotone I -splines are used and the asymptotic properties of the resulting estimates are established. In particular, the estimated regression parameters are shown to be semiparametrically efficient. An extensive simulation study and an illustrative example is provided.

Finally, we will talk about the directions for future research. One topic related the fourth part of this dissertation for future research could be to allow the failure time to depend on both the lower and upper bounds of the observation interval. Another possible future research topic could be to consider a cure rate model for interval-censored data with informative censoring.

Chapter 1

Introduction

1.1 Data Structure

1.1.1 Interval-Censored Data

In statistical literature, interval censoring usually represents a sampling scheme or an incomplete data structure. In the context of failure time data, interval censoring means that the failure time variable of interest is observed or known only to lie within some intervals or windows instead of being observed exactly. One field that often produces interval-censored data is medical or health studies that entail periodic follow-up. In this situation, an individual due for the pre-scheduled observations for a clinically observable change in disease or health status may miss some observations and return with a changed status. Accordingly, we only know that the true event time is greater than the last observation time at which the change has not occurred and less than or equal to the first observation time at which the change has been

observed to occur.

A more specific example of interval-censored data arises in the acquired immune deficiency syndrome (AIDS) trials (De Gruttola V and Lagakos S., 1989) that, for example, are interested in times to AIDS for human immunodeficiency virus (HIV) infected subjects. In these cases, the determination of AIDS onset is usually based on blood testing, which can be performed obviously only periodically but not continuously. In consequence, only interval-censored data may be available for AIDS diagnosis times.

In reality, interval censoring can occur in different forms and each form represents one type of interval-censored failure time data. Among them, one important special case of interval-censored data is the so called current status data. In this situation, each subject is observed only once for the status of the occurrence of the event of interest, and thus producing either a left- or a right-censored observation. That is, instead of observing the exact event time of interest, one only knows the observation time and whether or not the event has occurred at the time. One such example is given by the tumorigenicity study where the time to tumour onset is usually of interest but not directly observable. As a matter of fact, we only know the exact measurement of the observation time which is often the death or sacrifice time of the animal. If the animal dies at time t and if the tumour is present we have the interval $(0, t]$, while if the tumour is absent we have the interval $(t, \infty]$.

Sometimes, we also refer current status data to as case I interval-censored data and the general case as case II interval-censored data. For case II interval-censored data which is referred to as interval-censored data in the following for simplicity, a time interval $(L, R]$ instead of the exact failure time is observed for a study subject

such that the survival time T is contained in this interval.

1.1.2 Non-informative and Informative Censoring

A common assumption for the analysis of interval censored data is that the observation times are independent of the survival time of interest completely or given the covariates. Let T denote the survival time of interest, $F(t)$ and $S(t)$ denote the cumulative distribution function and survival function of T , respectively. If we observe current status data, which are usually denoted by

$$\{C_i, \delta_i = I(T_i \leq C_i)\}_{i=1}^n$$

where C_i represents the observation time on subject i , then the non-informative censoring means that T_i and C_i are independent completely or given covariates.

It is clear that this assumption cannot be directly generalized to interval censoring since in this case the endpoints of the interval, L and R , together with the survival time T , have a natural relationship $L < T \leq R$. Instead, for interval-censored data, the non-informative interval censoring assumption specified as

$$P(T \leq t | L = l, R = r, L < T \leq R) = P(T \leq t | l < T \leq r) \quad (1.1)$$

is usually used (Sun, 2006). This is saying that, except for the fact that T lies between l and r which are the realisations of L and R , the interval (L, R) does not provide any extra information for T . In other words, the probabilistic behaviour of T remains the same except that the original sample space $T \geq 0$ is now $l = L < T \leq R = r$. In

the existence of covariates Z , one can relax the assumption (1.1) to

$$P(T \leq t | L = l, R = r, L < T \leq R, Z = z) = P(T \leq t | l < T \leq r, Z = z). \quad (1.2)$$

Instead of the assumptions (1.1) or (1.2), another approach to define the non-informative interval censoring is through or using the stochastic process that yields interval censoring (Lawless and Babineau, 2006). For example, one can assume that there exists a sequence of observation times or an observation process and this process is independent of the survival time T .

Although non-informative censoring is a common assumption used in most existing literature for analysis of interval-censored data, there exist situations that this assumption may not hold. As mentioned earlier, an example of current status data is given by tumorigenicity experiments. In these studies, the time to tumour onset is usually of interest but not observable. Instead, only the death (natural death or sacrifice) time of animals at study and the status of tumour onset at the death time are observed. If the tumours are non-lethal, meaning that the tumour cannot cause death and does not alter the risk of death from other causes, it is obvious that we can reasonably assume that the tumour onset time and the death time are independent. If the tumours are lethal, meaning that the tumour onset kills animals right away, then it is reasonable to treat the death time equal to the tumour onset time. Most of the tumours are between non-lethal and lethal and thus the death time may depend on the tumour onset time.

For interval-censored data, there are also situations that may violate the non-informative censoring assumption. For example, in a periodic follow-up study, there could be some symptoms that occur first or together, making the a patient more

likely to visit the doctor. Another example is that there could exist some confounding variables influencing both the failure time and visiting times of patients (e.g. the general state of health). Note that the censoring mechanism behind interval censoring is usually much more complicated than that behind current status data. In consequence, it is usually difficult or impossible to generalize the methods developed for informatively current status data to them. To further see this, note that one can write the likelihood contribution from a single interval-censored observation as

$$Pr(L < T \leq R) = Pr(l < T \leq r | L = l, R = r) Pr(L = l, R = r).$$

This indicates that to conduct regression analysis, one would have to specify some joint models for $Pr(L < T \leq R)$ or for both $Pr(l < T \leq r | L = l, R = r)$ and $Pr(L = l, R = r)$. This is quite different from the case of current status data and usually not easy.

1.1.3 Two Examples

1.1.3.1 Current Status Data

The data about lung and bladder tumours from 617 animals had been produced from the ED_{10} study conducted at the National Center for Toxicological Research (NCTR). The study lasted 33 months and was about the development of tumors and time to tumour onset in relationship to the carcinogen 2-acetylaminofluorene (2-AAF). Originally, 24,000 female mice were involved in the study and were randomized to either a control group or one of the several dose levels of the carcinogen 2-AAF.

As the presence or absence of tumours was determined only through a pathologic

examination conducted at the time of death, which indicated the development of the tumours being before or after the death time. And each mouse was dead by either natural causes or the predetermined sacrifice. Thus we have current status, where the tumour onset time was the failure time of interest, T and the death time or sacrifice time was the observation time C . Note that the natural deaths may be related to the occurrence of tumours since the tumours may not be nonlethal.

1.1.3.2 Interval-Censored Data

A breast cosmesis data had been produced from a retrospective study on early breast cancer patients at Joint Center for Radiation in Boston between 1976 and 1980. One goal of this study was to compare early breast cancer patients who had been treated with primary radiation therapy and adjuvant chemotherapy to those treated with radiotherapy alone with respect to the cosmetic effects of their treatment.

It's known that adjuvant chemotherapy improves the relapse-free and overall survival in at least some subgroups of patients treated initially by mastectomy. However, there is experimental and clinical evidence which indicates that chemotherapy enhances the acute reponse of normal tissue to radiation treatment. Acute skin reactions are worse when adjuvant chemotherapy is administered in conjunction with either postoperative radiation or primary radiation treatment for breast cancer. The long-term impact of adjuvant chemotherapy on the radiation response of the breast is uncertain. Thus researchers want to compare the patients who received adjuvant chemotherapy following the initial radiation treatment to those who received only the radiation treatment, to determine whether the chemotherapy affects the rate of deterioration of the cosmetic state. Breast retraction was highly correlated with a

negative overall cosmetic result and was one of the least subjective of the endpoints that were followed. Thus the failure time of interest here was the time until breast retraction. The data consists of 94 early breast cancer patients, 46 of them were treated by radiotherapy only and the rest of them were treated with adjuvant chemotherapy in conjunction with primary radiation treatment. More details of this data set can be found in Finkelstein and Wolfe (1985).

1.2 Some Commonly Used Models

Regression analysis is usually performed if one is interested in quantifying the effect of some covariates on the survival time of interest or predicting survival probabilities for new individuals. Of course, the first step for regression analysis is to specify an appropriate regression model. In this dissertation, we will focus on semi-parametric models and some corresponding inference procedures for interval-censored data, which means that the baseline hazard function is treated as an infinite dimension parameter. Unlike most methods developed for right-censored data, estimating regression parameters under interval-censoring usually involves estimation of both the parametric and the non-parametric parts. In other words, for interval-censored data, one has to deal with estimation of some unknown baseline functions in order to estimate regression parameters.

The proportional hazards model (Cox, 1972) has been the most commonly used semi-parametric regression model for survival analysis for the past three decades.

Under the proportional hazards assumption, we have

$$\lambda(t|z) = \lambda_0(t) \exp(z'\beta). \quad (1.3)$$

where $\lambda_0(t)$ denotes the unknown baseline hazard function (the hazard function for subjects with $Z = 0$) and β the vector of unknown regression parameters. Then we could have

$$S(t) = S_0(t)^{\exp(z'\beta)}$$

Another commonly used semi-parametric model in survival analysis is the proportional odds model. This model can be expressed as

$$\log \left\{ \frac{F(t|z)}{1 - F(t|z)} \right\} = h(t) + z'\beta \quad (1.4)$$

where $h(t)$ is an unknown monotone-increasing function, also referred to as the baseline log odds. So that we know

$$S(t) = \frac{1}{1 + \exp\{h(t) + z'\beta\}}$$

As models (1.3) and (1.4), another attractive semi-parametric regression model is the additive hazards model given by

$$\lambda(t|z) = \lambda_0(t) + z'\beta \quad (1.5)$$

It specifies that the effects of the covariates are additive rather than multiplicative as in model (1.3).

In addition to the models discussed above, the accelerated failure time model is also widely used in survival analysis. It assumes that

$$\log(T) = Z'\beta + \epsilon \tag{1.6}$$

where ϵ is an error term whose distribution is usually unspecified.

The four semi-parametric models described above are all specific models in terms of the functional form of the effects of covariates. Sometimes one may prefer a model that gives more flexibility. One such model is the linear transformation model that specifies the relationship between the event time and the covariate as

$$h(T) = Z'\beta + \epsilon \tag{1.7}$$

where $h : \mathbb{R}^+ \rightarrow \mathbb{R}$ is an unknown strictly increasing function and the distribution of ϵ is assumed to be known.

Several other models or generalisations of the models discussed above are also available for regression analysis of interval-censored failure time data. For example, one may apply the partial linear model given by

$$\log(T) = Z_1'\beta + g(Z_2) + \epsilon,$$

a generalisation of the accelerated failure time model (1.6). Here both Z_1 and Z_2 are covariates which may or may not overlap, g is an unknown smooth function, and ϵ follows a pre-specified distribution.

1.3 Regression Analysis of Interval-Censored Data

In regression analysis, the primary objective is to estimate the covariate effects on the event time. Semi-parametric regression analysis of interval-censored data has recently attracted considerable attention and many procedures have been developed under various models.

1.3.1 Regression Analysis of Current Status Data

A number of methods have been developed for semi-parametric regression analysis of current status data. Huang (1996) developed the maximum likelihood approach for fitting the proportional hazards model to current status data and an ICM-type algorithm for estimation of unknown parameters. Chen et al. (2009) and Sun and Shen (2009) considered the same problem in the presence of clustering and competing risks, respectively. Instead of the proportional hazards model, among others, Huang (1995) and Rossini and Tsiatis (1996) considered the maximum likelihood approach under the proportional odds model but used different approximations for the baseline log odds function. Lin et al. (1998) and Chen and Sun (2009) discussed the fitting of the additive hazards model and developed some estimating equation-based and imputation-based procedures, respectively. In particular, the latter model describes a different aspect of the association between the failure time and covariates compared to the former model and could be more plausible in many applications. Martinussen and Scheike (2002) studies the same problem and provided an approach that can be more efficient than that given in Lin et al. (1998). However, it involves estimation of the baseline hazard function and can be much more complicated. Furthermore,

Sun and Sun (2005) investigated regression analysis of current status data arising from the linear transformation models. Note that all of the methods described above assume that all covariates are known or can be exactly observed. And also, most of them assume that the observation time and the survival time of interest are independent of each other.

In practice, there may be missing covariates or covariates cannot be measured exactly due to various reasons such as financial limitations or technical difficulties. In this case, the true covariates may only be measured precisely in a subset of the study cohort, which is often referred to as the validation set. Furthermore, for missing covariates, there may exist some related information such as auxiliary covariates (Zhou and Pepe, 1995). In epidemiological or genetic studies, for example, the measurements of some covariates could be expensive and to save the cost, sometimes one may instead collect information on some related covariates that can be obtained relatively cheaply. For statistical analysis of such data, a naive approach would be to base the analysis only on the subjects in the validation set. It's not hard to see that this would lose some efficiency. The authors who investigated approaches that take into account the available auxiliary covariates include Hu and Lin (2002), Huang and Wang (2000), Lin and Ying (1993), Liu et al. (2009), Wang et al. (1998), Zhou and Pepe (1995) and Zhou and Wang (2000).

To address the issue of informative censoring in the context of current status data, several approaches have been proposed. Dewanji and Kalbfleisch (1986), Dinse (1991) and Lindsey and Ryan (1993) addressed this problem for univariate current status data by assuming transition functions and mainly used an expectation-maximization (EM) algorithm for estimation. Dunson and Dinse (2002) proposed a Bayesian ap-

proach for multivariate current status data with informative censoring and characterized the tumor onset and death process in terms of transition probabilities. Instead of assuming a multistate model like the above procedures, Lagakos and Louis (1988) gave a different method that assumes that the tumour lethality is known. Zhang et al. (2005) postulated a frailty variable to characterize the correlation between the failure time and censoring time.

1.3.2 Regression Analysis of Interval-Censored Data

There also exist many methods for regression analysis of interval-censored failure time data. Finkelstein (1986) considered the fitting of proportional hazards model and proposed to apply the Newton-Raphson algorithm to determine the MLE of β and the baseline cumulative hazard function together. Sattner (1996) proposed a marginal likelihood method and Goggins et al. (1998) developed a Markov Chain Monte Carlo EM algorithm for the same problem. Huang and Rossni (1997) and Shen (1998) applied the sieve estimation procedures under the proportional odds model. The former employed a piecewise linear function, while the latter used a monotone spline to approximate the baseline log odds function. Rabinowitz et al. (2000) discussed the same problem with an alternative method which does not require estimation of the baseline log odds function but assumed that subjects are still under follow-up even after the failure event has occurred. For inference about the additive hazards model based on interval-censored data, Zeng et al. (2006) investigated the maximum likelihood approach. And Chen and Sun (2009) and Zhu et al. (2008) gave a multiple imputation-based procedure and a transformation approach, respectively. Rabinowitz et al. (1995) and Betensky et al. (2001) developed some procedures under

the accelerated failure time model and employed a class of score statistics and the estimating equation idea, respectively. Both of these two methods require estimation of the distribution of the error term ϵ . Li and Pu (2003) investigated the same problem and employed a rank-based estimating equation which does not require estimation of the distribution of ϵ but may not be efficient. Zhang et al. (2005) proposed an inference approach for fitting the linear transformation models to interval-censored data.

The inference under most of maximum likelihood methods and estimating equation-based approaches is performed based on the asymptotic normality. However, the symmetric property implied by the normal distribution may not be appropriate sometimes. One possible alternative could be the empirical likelihood-based procedure which does not impose the symmetric assumption on the distribution of the estimated regression parameters. The empirical likelihood has been widely applied in many different statistical areas over the last 20 or 30 years (Chen, 1994; Owen, 1988; Qin and Lawless, 1994).

In contrast with that for non-informative interval censoring, only limited literature exists for regression analysis of informatively interval-censored data, including Zhang et al. (2007) and Wang et al. (2010). They considered the cases where T follows proportional hazards model and additive hazards model, respectively, marginally and in both cases, the proportional hazards model was also used to model the censoring variables.

1.4 Outline of The Dissertation

The remainder of this dissertation is organized as follows. In Chapter 2, we discuss the regression analysis of interval-censored failure time data arising from the linear transformation models. It's well known that the symmetric property of the estimated regression parameters implied by the normal distribution may not be appropriate sometimes. For this problem, we propose an empirical likelihood-based procedure for deriving the confidence regions for regression parameters. Simulation studies and analysis of a real data set are conducted to assess the performance of the procedure.

In Chapter 3, we focus on regression analysis of current status data under the additive hazards model when there exist missing and auxiliary covariates. To address this problem, we propose an estimated partial likelihood approach for estimation of regression parameters that makes use of the available auxiliary information. The method can be easily implemented and the asymptotic properties of the resulting estimates are established. An extensive simulation study is conducted to assess the finite sample performance of the proposed method and it indicates that the method works well in practical situations.

In Chapter 4, we consider the regression analysis of current status data when the observation time may be related to the survival time. We employ a copula function to model the correlation between observation time and survival time and use the proportional hazards model for the covariate effects on both the observation time and the survival time. For inference, we develop a sieve maximum likelihood estimation procedure with the use of monotone I -spline functions. Asymptotic properties of the resulting estimates are established. Furthermore, the proposed method is examined through a simulation study and illustrated with a real data set.

In Chapter 5, we discuss the regression analysis of interval-censored data with informative censoring. We focus on the situation where the failure time depends on the censoring interval only through its length and model that dependence through a copula function. A sieve maximum likelihood approach is presented for the analysis of such data arising from the proportional hazards model. In the approach, monotone I -splines are used and the asymptotic properties of the resulting estimates are also established. We also provide estimation results of a simulations study as well as an illustrative example.

In Chapter 6, several directions for future research are discussed.

Chapter 2

Empirical Analysis of Interval-Censored Failure Time Data with Linear Transformation Models

2.1 Introduction

Interval-censored failure time data naturally arise in many studies including economic, epidemiological and medical follow-up studies as well as social sciences (Chen et al., 2012; Sun, 2006). In this case, the failure time of interest is observed only to belong to an interval or some windows instead of being observed exactly.

One example is for studies on HIV infection times. If a patient is HIV positive at the beginning of a study, then the HIV infection time is usually determined by a retrospective analysis of his or her medical history. Therefore, we are only able to obtain an interval given by the last HIV negative test date and the first HIV positive

test date for the HIV infection time.

Many methods have been developed for regression analysis of interval-censored failure time data under various models and this is especially the case for right-censored data, a special case of interval-censored data (Kalbfleisch and Prentice, 2002). For example, among others, Finkelstein (1986) and Huang (1999) considered the fitting of the proportional hazards model (Cox, 1972; Andersen and Gill, 1982) to interval-censored data and developed the maximum likelihood approaches. Other models that have been used for interval-censored data include the additive hazards model and the proportional odds model, and in these situations, a common tool used for the development of inference procedures is the estimating equation theory. For example, Wang et al. (2010) proposed an estimating equation-based approach for interval-censored data with the additive hazards model; Chen et al. (2007) investigated the proportional odds model for multivariate interval-censored failure time data.

The models mentioned above are all specific models and sometimes they could be too restrictive in practice. Corresponding to this, among others, Zhang et al. (2005) investigated the use of the linear transformation model for regression analysis of interval-censored data and proposed an estimating equation-based inference procedure. This latter model provides a class of flexible models and includes both the proportional hazards model and the proportional odds model as special cases (Chen et al., 2002). Note that the inference under most of both maximum likelihood methods and estimating equation-based approaches is performed based on the asymptotic normality. It is well-known that the symmetric property implied by the normal distribution may not be appropriate sometimes. Also the method given in Zhang et al. (2005) could underestimate the variance of the estimated regression parameters.

To address the issues above, in the following, we propose an empirical likelihood method for the regression problem with the linear transformation model. The empirical likelihood has been widely applied in many different statistical areas over last 20 or 30 years (Chen, 1994; Owen, 1988, 1990; Qin and Lawless, 1994). In particular, Zhao (2010) applied it to the fitting of the linear transformation model to right-censored failure time data. To present the proposed approach, we will first begin in Section 2.2 with introducing some notation and assumptions, especially the linear transformation model. Also we will briefly review the estimating equation procedure given by Zhang et al. (2005) in Section 2.2. An empirical likelihood method is then presented in Section 2.3 and the asymptotic distribution of the resulting empirical likelihood ratio is established. Some simulation results are presented in Section 2.4 and indicate that the presented approach works well for practical situations. An illustrative example is also provided in Section 2.4 and Section 2.5 contains some discussion and concluding remarks.

2.2 Linear Transformation Models and Estimation

Consider a survival study that consists of n independent subjects. For subject i , let T_i denote the failure time of interest and suppose that there exists a p -dimensional vector of covariates Z_i , $i = 1, \dots, n$. Define $S_{Z_i}(t) = P(T_i > t|Z_i)$, the survival function of T_i given Z_i . For the relationship between T_i and Z_i , suppose that the T_i 's follow the linear transformation model specified by

$$g\{S_{Z_i}(t)\} = h(t) + Z_i^T \beta. \quad (2.1)$$

Here $h(t)$ is a completely unknown strictly increasing function, $g(t)$ a known decreasing function, and β a vector of unknown regression coefficients. It is easy to see that model (2.1) can be equivalently expressed as

$$h(T_i) = -Z_i^T \beta + \epsilon_i, \quad (2.2)$$

where the ϵ_i 's are random errors following the distribution function $F(t) = 1 - g^{-1}(t)$.

Note that model (2.1) or (2.2) gives a class of flexible models and include many commonly used models as special cases. For example, it gives the proportional hazards model (Kalbfleisch & Prentice, 2002) if we take $g(x) = \log(-\log(x))$ and one gets the proportional odds model if letting $g(x) = -\text{logit}(x)$ (Pettitt, 1982, 1984). Many authors have considered the inference about model (2.1), especially for the case of right-censored failure time data (Zhao, 2010; Yu et al., 2011).

In the following, we briefly describe the inference procedure given in Zhang et al. (2005) based on interval-censored data. For this, suppose that one observes interval-censored data given by $\{(L_i, R_i], Z_i \mid i = 1, \dots, n\}$, where $L_i < T_i \leq R_i$. Also suppose that the Z_i 's are discrete variables and the censoring is independent in the sense that

$$P\{T_i \leq t \mid L_i = l, R_i = r, L_i < T_i \leq R_i, Z_i\} = P\{T_i \leq t \mid l < T_i \leq r, Z_i\}.$$

This means the mechanism generating censoring intervals for T is independent of T given Z . Let H_{Z_i} denote the distribution function of the T_i 's given Z_i and \hat{H}_{Z_i} the maximum likelihood estimator of H_{Z_i} , which can be obtained by the self-consistency

algorithm among other algorithms (Sun, 2006). Then one can show that

$$E \left\{ (a_i a_j)^{-1} \int_{L_i}^{R_i} \int_{L_j}^{R_j} I(t_i \geq t_j) dH_{Z_i}(t_i) dH_{Z_j}(t_j) | Z_i, Z_j \right\} = \tau(Z_{ij}^T \beta_0), \quad (2.3)$$

where $Z_{ij} = Z_i - Z_j$, $a_i = \int_{L_i}^{R_i} dH_{Z_i}(t_i)$, β_0 is the true value of β and

$$\tau(t) = \int_{-\infty}^{\infty} \{1 - F(s+t)\} dF(s).$$

Motivated by the equation above, Zhang et al. (2005) proposed to use the estimation equation

$$U(\beta) = \sum_{i=1}^n \sum_{j=1}^n \omega_{ij}(\beta) \tau'(Z_{ij}^T \beta) \left\{ (\hat{a}_i \hat{a}_j)^{-1} \int_{L_i}^{R_i} \int_{L_j}^{R_j} I(t_i \geq t_j) d\hat{H}_{Z_i}(t_i) d\hat{H}_{Z_j}(t_j) - \tau(Z_{ij}^T \beta) \right\} Z_{ij} = 0 \quad (2.4)$$

for estimation of β . In the above, the $\omega_{ij}(\beta)$'s are some positive bounded weight functions and the \hat{a}_i 's are the a_i 's defined above with H_Z replaced by \hat{H}_Z . Let $\hat{\beta}$ denote the estimator of β given by the solution to the equation above. Zhang et al (2005) proved that under some regularity conditions, one can approximate the distribution of $\sqrt{n}(\hat{\beta} - \beta_0)$ by the normal distribution with mean zero and the covariance that can be estimated by $\hat{\Sigma}_Z = \hat{D}^{-1} \hat{\Gamma} \hat{D}^{-1}$. Here

$$\hat{D} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \omega_{ij}(\hat{\beta}) \{ \tau'(Z_{ij}^T \hat{\beta}) \}^2 Z_{ij}^{\otimes 2}$$

and

$$\hat{\Gamma} = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k \neq j}^n (\hat{e}_{ij} - \hat{e}_{ji}) (\hat{e}_{ik} - \hat{e}_{ki}) Z_{ij} Z_{ik}^T, \quad (2.5)$$

where $Z^{\otimes 2} = ZZ^T$ and \hat{e}_{ij} is equal to

$$e_{ij}(\beta) = \omega_{ij}(\beta)\tau'(Z_{ij}^T\beta) \left\{ (a_i a_j)^{-1} \int_{L_i}^{R_i} \int_{L_j}^{R_j} I(t_i \geq t_j) dH_{Z_i}(t_i) dH_{Z_j}(t_j) - \tau(Z_{ij}^T\beta) \right\} \quad (2.6)$$

with β and H_Z replaced by $\hat{\beta}$ and \hat{H}_Z , respectively. It follows that one can obtain the $100(1 - \alpha)\%$ confidence region for β as

$$\mathcal{R}_1 = \{ \beta : n(\hat{\beta} - \beta)^T \hat{\Sigma}_Z^{-1} (\hat{\beta} - \beta) \leq \chi_\alpha^2(p) \}$$

based on the normal approximation above, where $\chi_\alpha^2(p)$ denotes the upper α -quantile of the chi-square distribution with degrees of freedom p .

It is obvious that \mathcal{R}_1 is symmetric. However, the distribution of $\hat{\beta}$ may not be, especially for small n . Another issue related to \mathcal{R}_1 is that as will be seen below, it could lead to lower coverage probabilities since the method described tends to underestimate the variance of the estimated parameters. In the next section, to address these issues, a new empirical likelihood-based confidence region is developed.

2.3 Empirical Likelihood Procedure

In this section, we present an empirical likelihood procedure for obtaining the confidence region and making the inference about β in model (2.1) or (2.2). For this, let $U_i = (Z_i^T, L_i, R_i)^T$, and define $b(U_i, U_j, \beta) = \{Z_{ij}e_{ij}(\beta) + Z_{ji}e_{ji}(\beta)\}$ and $b_n(U_i, U_j, \beta) = \{Z_{ij}\hat{e}_{ij}(\beta) + Z_{ji}\hat{e}_{ji}(\beta)\}$, where $\hat{e}_{ij}(\beta)$ denote the quantity defined in (2.6) with replacing

H_Z by \hat{H}_Z . Also define

$$W_i(\beta) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n b(U_i, U_j, \beta)$$

and

$$W_{ni}(\beta) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n b_n(U_i, U_j, \beta),$$

$i = 1, \dots, n$. It is easy to see that the $W_i(\beta)$'s are identically distributed and

$$S_n(\beta) = \frac{1}{n} \sum_{i=1}^n W_i(\beta)$$

is a multivariate U -statistic for fixed β . It follows from (2.3) and (2.4) that $E\{W_i(\beta_0)\} = 0$. This suggests the following empirical likelihood function

$$L(\beta) = \sup \left\{ \prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i W_i(\beta) = 0, p_i \geq 0 \right\}$$

and thus the estimated empirical likelihood function

$$L_n(\beta) = \sup \left\{ \prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i W_{ni}(\beta) = 0, p_i \geq 0 \right\}$$

for β , where $\mathbf{p} = (p_1, \dots, p_n)^T$ denotes the probability vector.

Note that without any restriction, the natural estimator of \mathbf{p} is given by $(1/n, \dots, 1/n)^T$.

This leads to the following empirical likelihood ratio

$$R(\beta) = \sup \left\{ \prod_{i=1}^n np_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i W_{ni}(\beta) = 0, p_i \geq 0 \right\}$$

for β . With the use of the lagrange multiplier method, one can easily obtain

$$\hat{l}(\beta) = -2 \log R(\beta) = 2 \sum_{i=1}^n \log \{ 1 + \lambda^T W_{ni}(\beta) \} \quad (2.7)$$

with the p -dimensional vector λ given by the equation

$$\frac{1}{n} \sum_{i=1}^n \frac{W_{ni}(\beta)}{\{1 + \lambda^T W_{ni}(\beta)\}} = 0. \quad (2.8)$$

In practice, one can calculate $\hat{l}(\beta)$ directly through $R(\beta)$ using the restricted maximization instead of calculating λ first. It is apparent that one can easily develop a confidence region for β based on $\hat{l}(\beta)$ if its distribution is known.

To derive the asymptotic distribution of $\hat{l}(\beta)$, we need the following regularity conditions and two lemmas.

Condition 1. The covariate vectors Z_i 's are bounded such that $\| Z_i \| \leq C_1$ for some positive constant C_1 , where $\| \cdot \|$ denotes the Euclidean norm.

Condition 2. The functions $\omega(\cdot)$ and $\tau(\cdot)$ defined in the previous section are differentiable and whose derivatives $\omega'(\cdot)$ and $\tau'(\cdot)$ are continuous.

Condition 3. The following two matrices

$$D(\beta_0) = \lim_{n \rightarrow \infty} E \left(\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \omega_{ij}(\beta_0) \{ \tau'(Z_{ij}^T \beta_0) \}^2 Z_{ij}^{\otimes 2} \right)$$

and

$$\Gamma(\beta_0) = \lim_{n \rightarrow \infty} E \left(\frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k \neq j}^n (e_{ij}(\beta_0) - e_{ji}(\beta_0)) (e_{ik}(\beta_0) - e_{ki}(\beta_0)) Z_{ij} Z_{ik}^T, \right)$$

are positive definite.

Define $\Sigma(\beta) = E\{b(U_1, U_2, \beta)b^T(U_1, U_3, \beta)\}$, $\Sigma = \Sigma(\beta_0)$,

$$\Sigma_n = \frac{1}{n} \sum_{i=1}^n W_i(\beta_0) W_i^T(\beta_0), \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n W_{ni}(\hat{\beta}) W_{ni}^T(\hat{\beta}). \quad (2.9)$$

Note that by following Hoeffding (1948) on scalar U -statistics, we can obtain

$$\text{var}\{S_n(\beta_0)\} = \frac{4\Sigma}{n} + O\left(\frac{1}{n^2}\right), \text{ a.s.}$$

Also by following Hoeffding (1948) on the asymptotic normality for multivariate U -statistics of degree 2 and Zhao (2010), respectively, one can derive the following two lemmas.

Lemma 2.1 *Assume that $E\{b(U_1, U_2, \beta_0)b^T(U_1, U_2, \beta_0)\} < \infty$ and Σ is positive definite. Then as $n \rightarrow \infty$, we have*

$$\sqrt{n} S_n(\beta_0) \xrightarrow{\mathcal{D}} N(0, 4\Sigma).$$

Lemma 2.2 *Also assume that $E\{b(U_1, U_2, \beta_0)b^T(U_1, U_2, \beta_0)\} < \infty$. Then*

$$\Sigma_n = \Sigma + o(1), \text{ a.s.}$$

Now we are ready to describe the main results needed to derive the confidence region for β .

Theorem 2.1 *Assume that the conditions 1 and 2 described above hold. Then as*

$n \rightarrow \infty$, we have

$$(i). \frac{1}{n} \sum_{i=1}^n W_{ni}(\beta_0) W_{ni}^T(\beta_0) \xrightarrow{P} \Sigma, \quad (ii). \hat{\Sigma} \xrightarrow{P} \Sigma, \quad (iii). \hat{\Gamma} \xrightarrow{P} \Gamma(\beta_0).$$

Theorem 2.2 *Assume that the conditions 1-3 described above hold. Then the empirical log likelihood statistic $\hat{l}(\beta)/4$ converges in distribution to $\kappa_1 \chi_1^2(1) + \dots + \kappa_p \chi_p^2(1)$, where $\kappa_1, \dots, \kappa_p$ are the eigenvalues of $\Sigma^{-1} \Gamma(\beta_0)$ and the $\chi_j^2(1)$'s are independent chi-square random variables with 1 degree of freedom.*

The proofs for the two theorems above are given in the Appendix. Theorem 2.2 says that the limiting distribution of the empirical log likelihood ratio is a weighted sum of iid $\chi^2(1)$ random variables. To apply it, one needs to estimate the weights κ_j 's and it is obvious that a natural way is to use the eigenvalues of matrix $\hat{\Sigma}^{-1} \hat{\Gamma}$. Specifically, let $\hat{\kappa}_1, \dots, \hat{\kappa}_p$ denote such defined estimates of the κ_j 's. Then one can obtain an asymptotic $100(1 - \alpha)\%$ empirical likelihood confidence region of β by

$$\mathcal{R}_2 = \{ \beta : \hat{l}(\beta)/4 \leq C_\alpha \},$$

where C_α is the upper α -quantile of the distribution of $\hat{\kappa}_1 \chi_1^2(1) + \dots + \hat{\kappa}_p \chi_p^2(1)$.

It is well-known that in general, it is not easy to use the weighted χ^2 distribution and for this, one may want to apply the adjusted empirical log likelihood ratio (Wang and Rao, 2001). Specifically, define $\rho(\beta) = p/\text{tr}\{\Sigma^{-1}(\beta) \Gamma(\beta)\}$, where $\text{tr}(A)$ denotes the trace of a matrix A . Then it follows from Rao and Scott (1981) that one can approximate the distribution of $\rho(\beta) \{ \kappa_1 \chi_1^2(1) + \dots + \kappa_p \chi_p^2(1) \}$ by $\chi^2(p)$. In consequence, this suggests that one can asymptotically approximate the distribution of the

adjusted empirical log likelihood ratio statistic $\hat{\rho}(\beta)\hat{l}(\beta)/4$ by $\chi^2(p)$, where $\hat{\rho}(\beta)$ is defined as $\rho(\beta)$ but with $\Sigma(\beta)$ and $\Gamma(\beta)$ replaced by $\hat{\Sigma}(\beta)$ and $\hat{\Gamma}(\beta)$, given in (2.9) and (2.5) with $\hat{\beta}$ replaced by β , respectively. To establish this formally, define the adjusted empirical log likelihood ratio statistic as

$$\hat{l}_{ad}(\beta) = \hat{\gamma}(\beta)\hat{l}(\beta),$$

where

$$\hat{\gamma}(\beta) = \text{tr} \left\{ \hat{\Gamma}^{-1}(\beta) \hat{S}(\beta) \right\} \text{tr}^{-1} \left\{ \hat{\Sigma}^{-1}(\beta) \hat{S}(\beta) \right\}$$

with

$$\hat{S}(\beta) = \frac{1}{n^2} \left\{ \sum_{i=1}^n W_{ni}(\beta) \right\} \left\{ \sum_{i=1}^n W_{ni}^T(\beta) \right\}.$$

The following theorem gives the asymptotic distribution of $\hat{l}_{ad}(\beta)$.

Theorem 1.3 *Assume that the conditions 1-3 described above hold. Then as $n \rightarrow \infty$, the empirical log likelihood ratio statistic $\hat{l}_{ad}(\beta_0)/4$ converges in distribution to $\chi^2(p)$.*

The proof of the theorem above is similar to that of Theorem 2.2 of Zhao (2010) and thus omitted. It suggests that an asymptotic $100(1 - \alpha)\%$ confidence region for β can be obtained as

$$\mathcal{R}_3 = \{ \beta : \hat{l}_{ad}(\beta)/4 \leq \chi_\alpha^2(p) \}.$$

For the determination of \mathcal{R}_3 , note that the adjusted factor $\hat{\gamma}(\beta)$ involves the unknown regression parameter β . To deal with it, one can apply the profile analysis method and more specifically, by noting the convexity of the function $\hat{l}_{ad}(\beta)$, one can easily get left and right end points of the region \mathcal{R}_3 by solving $\hat{l}_{ad}(\beta)/4 = \chi_\alpha^2(p)$ for β .

2.4 Numerical Studies

An extensive simulation was conducted to assess the performance of the adjusted empirical likelihood procedure (AEL) for obtaining the confidence region \mathcal{R}_3 . For comparison, we also obtained the confidence region \mathcal{R}_1 given by the normality approximation-based procedure (NA) proposed in Zhang et al. (2005). In the study, we considered model (2.1) with

$$g(t) = \begin{cases} \log\{(1 - t^\alpha)(\alpha t^\alpha)^{-1}\} & \text{if } \alpha > 0, \\ \log\{-\log(t)\} & \text{if } \alpha = 0 \end{cases}$$

and $\alpha = 0, 0.5, 1$. Note that $\alpha = 0$ and $\alpha = 1$ correspond to the proportional hazards model and the proportional odds model, respectively. For covariates, we assumed that there exists one covariate with the Z_i 's generated from the Bernoulli distribution with the success probability of 0.5.

To generate the failure times T_i 's of interest, we took $h(t) = \log(t)$ and $h(t) = \log(0.5t)$ for the cases with $\alpha = 0$ and 0.5, respectively, and rounded off the generated survival times to their first decimal places. For the case with $\alpha = 1$, we used $h(t) = \log(0.08t)$ and rounded off the generated survival times to the nearest integers. For the generation of censoring intervals, for the first two cases, the L_i 's and R_i 's were defined to be the generated failure times minus and plus some random numbers generated from the uniform distribution $U\{0, 0.1, 0.2\}$, respectively. Note that as pointed in Zhang et al. (2005), this does not give completely independent observation times, but is more practical since it was motivated by and is equivalent to the usual set-up in follow-up studies. For the third case, the same method was used except that the uniform distribution $U\{0, 1, 2\}$ was used. For all situations, we assumed

that there existed some constant right-censoring times, which were chosen to give 10%, 20%, or 30% right-censored observations. The results given below are based on 1000 replications and the sample size $n = 60$ or 100 .

Table 2.1 presents the 95% empirical coverage probabilities for β given by the two methods, NA and AEL, under three different right censored ratios (CR) with $\beta_0 = -1, -0.5, 0, 0.5, 1$, respectively, and $\alpha = 0$. The results for the cases with $\alpha = 0.5$ and 1 are given in Tables 2.2 and 2.3, respectively. These results suggest that the proposed empirical procedure, AEL, seems to perform well for the situations considered here, especially for the cases with $\alpha = 0.5$ or 1 and $n = 100$. It is clear that the performance gets better when the sample size increases as expected. In comparison, as mentioned before, the NA procedure indeed tends to underestimate the coverage probability. In the study, we also compared the sizes of the confidence region given by the two procedures and although the proposed procedure performed better than the NA procedure sometimes, the improvement is not significant.

To illustrate the proposed inference procedure, we apply it to a set of well-known interval-censored failure time data arising from a breast cancer study (Finkelstein, 1986; Sun, 2006). The study consists of 94 early breast cancer patients, randomly assigned to two treatment groups, radiotherapy alone and radiotherapy plus adjuvant chemotherapy. Here the failure time of interest is the time to breast retraction and the patients were examined only at different time points. Thus only interval-censored data are available. One objective of the study is to compare the two treatments with respect to the time until the appearance of breast retraction.

For the analysis, define $Z_i = 0$ if the patient was given radiotherapy alone and $Z_i = 1$ otherwise. Table 2.4 gives the results obtained by the two methods AEL

and NA under the three models used for the simulation study. They include the confidence regions and the p -values for testing $\beta = 0$. Note that for the three cases of $\alpha = 0, 0.5, 1$, the point estimates based on the NA procedure are $\hat{\beta} = 0.697, 0.866$ and 1.041 , respectively. One can see from Table 2.4 that all p -values given by both AEL and NA procedures are very close to each other and less than 0.01 level. In other words, they all suggest that the patients given the radiotherapy plus adjuvant chemotherapy had significantly higher risks for breast retraction than those given only radiotherapy alone.

2.5 Discussion

This chapter discussed regression analysis of interval-censored failure time data arising from the linear transformation model. Corresponding to the estimating equation approach proposed in Zhang et al. (2005), we presented an empirical likelihood procedure for deriving the confidence regions for regression parameters. Under the procedure, one can either obtain an unadjusted empirical likelihood confidence region \mathcal{R}_2 or an adjusted empirical likelihood confidence region \mathcal{R}_3 . For the former, however, the simulation is needed to obtain the critical value C_α , while the latter can be implemented much easily.

As mentioned above, one advantage of the proposed empirical inference procedure is that it does not impose the symmetry assumption on the distribution of the estimated regression parameters. In contrast, the procedure given by Zhang et al. (2005) and other asymptotic normality-based procedures always yield symmetric confidence regions, which may not be true for finite sample situations. On the other hand, the

empirical likelihood method can be relatively complicated in computation compared to the other procedures as it involves the computation of Lagrange multipliers.

β	n	10%CR		20%CR		30%CR	
		NA	AEL	NA	AEL	NA	AEL
-1	60	0.916	0.932	0.906	0.925	0.913	0.936
	100	0.923	0.941	0.926	0.940	0.907	0.936
-0.5	60	0.919	0.942	0.917	0.951	0.917	0.948
	100	0.926	0.942	0.915	0.936	0.919	0.931
0	60	0.918	0.937	0.907	0.935	0.915	0.932
	100	0.929	0.935	0.927	0.940	0.925	0.932
0.5	60	0.911	0.930	0.911	0.931	0.908	0.927
	100	0.923	0.938	0.921	0.935	0.916	0.924
1	60	0.904	0.913	0.901	0.912	0.891	0.912
	100	0.915	0.928	0.907	0.929	0.903	0.930

Table 2.1: Empirical coverage probabilities based on simulated data with $\alpha = 0$.

β	n	10%CR		20%CR		30%CR	
		NA	AEL	NA	AEL	NA	AEL
-1	60	0.927	0.946	0.927	0.945	0.927	0.942
	100	0.932	0.949	0.932	0.949	0.935	0.951
-0.5	60	0.919	0.946	0.928	0.945	0.921	0.952
	100	0.924	0.953	0.929	0.949	0.929	0.948
0	60	0.919	0.948	0.900	0.944	0.904	0.935
	100	0.928	0.949	0.923	0.951	0.920	0.950
0.5	60	0.918	0.946	0.915	0.951	0.918	0.939
	100	0.930	0.945	0.929	0.945	0.928	0.940
1	60	0.918	0.938	0.920	0.938	0.913	0.945
	100	0.938	0.943	0.934	0.949	0.934	0.941

Table 2.2: Empirical coverage probabilities based on simulated data with $\alpha = 0.5$.

β	n	10%CR		20%CR		30%CR	
		NA	AEL	NA	AEL	NA	AEL
-1	60	0.923	0.943	0.923	0.945	0.925	0.947
	100	0.935	0.943	0.935	0.953	0.930	0.946
-0.5	60	0.925	0.949	0.928	0.938	0.928	0.941
	100	0.925	0.944	0.926	0.941	0.924	0.948
0	60	0.926	0.944	0.927	0.938	0.930	0.939
	100	0.937	0.950	0.933	0.947	0.931	0.953
0.5	60	0.928	0.939	0.914	0.936	0.918	0.938
	100	0.928	0.950	0.936	0.946	0.929	0.953
1	60	0.922	0.940	0.912	0.934	0.910	0.930
	100	0.925	0.947	0.919	0.936	0.920	0.938

Table 2.3: Empirical coverage probabilities based on simulated data with $\alpha = 1$.

α	$\hat{\beta}$	confidence regions		p -values	
		NA	AEL	NA	AEL
0	0.697	[0.2050, 1.1890]	[0.1703, 1.1458]	0.006	0.005
0.5	0.866	[0.2568, 1.4738]	[0.2119, 1.4179]	0.005	0.005
1	1.041	[0.3199, 1.7701]	[0.2543, 1.7026]	0.005	0.005

Table 2.4: Analysis results of the breast cancer study.

Chapter 3

Regression Analysis of Current Status Data under the Additive Hazards Model with Auxiliary Covariates

3.1 Introduction

This chapter discusses regression analysis of current status or case I interval-censored failure time data arising from the additive hazards model (Lin *et al.*, 1998; Martinussen & Scheike, 2002; Sun, 2006). As introduced earlier, in current status data, the failure time of interest is not exactly observed but is either left- or right-censored at the observation time. One can find current status data in many other areas too including cross-sectional studies, economics, demographical studies, medical studies, reliability studies and social sciences (Jewell & van der Laan, 1995; Huang, 1996; Rossini & Tsiatis, 1996).

For regression analysis of failure time data, it is apparent that one needs covariate information. However, in practice, there may be missing covariates or covariates cannot be measured exactly due to various reasons such as financial limitations or technical difficulties. In this case, the true covariates may only be measured precisely in a subset of the study cohort, which is often referred to as the validation set. Furthermore, for missing covariates, there may exist some related information such as auxiliary covariates (Zhou & Pepe, 1995). In epidemiological or genetic studies, for example, the measurements of some covariates could be expensive and to save the cost, sometimes one may instead collect information on some related covariates that can be obtained relatively cheaply. For statistical analysis of such data, it is apparent that a naive approach is to base the analysis only on the subjects in the validation set. On the other hand, it is not hard to see that this would lose some efficiency and one should employ some methods that take into account the available auxiliary covariates. The authors who investigated this include Hu & Lin (2002), Huang & Wang (2000), Lin & Ying (1993), Liu *et al.* (2009), Wang *et al.* (1998), Zhou & Pepe (1995) and Zhou & Wang (2000).

Many procedures have been developed for regression analysis of current status failure time data under various models. For example, Huang (1996) developed the maximum likelihood approach for fitting the proportional hazards model to current status data and established the asymptotic properties of the maximum likelihood estimates. Chen *et al.* (2009) and Sun & Shen (2009) considered the same problem in the presence of clustering and competing risks, respectively. Instead of the proportional hazards model, among others, Lin *et al.* (1998) and Chen & Sun (2009) discussed the fitting of the additive hazards model to current status data and developed some esti-

rating equation-based and imputation-based procedures, respectively. In particular, Lin *et al.* (1998) pointed out that the latter model describes a different aspect of the association between the failure time and covariates compared to the former model and could be more plausible than the former in many applications. Furthermore, Sun & Sun (2005) investigated regression analysis of current status data arising from the linear transformation model. Note that all of the methods described above assume that all covariates are known or can be exactly observed. It does not seem to exist an established approach for regression analysis of current status data in the presence of auxiliary covariates. In the following, we will discuss this problem under the additive hazards model.

The remainder of this chapter is organized as follows. We will begin in Section 3.2 with introducing some notation, models and assumptions that will be used throughout the chapter. Section 3.3 presents the estimation procedures for estimation of regression parameters in the additive hazards model based on current status data with auxiliary covariates. The methods can be easily implemented since they do not involve estimation of an unknown function as some existing ones for the analysis of current status data (Huang, 1996). Also the determination of the proposed estimates is straightforward. In addition, the asymptotic properties of the proposed estimates, including the consistency and the asymptotic normality, are established. In Section 3.4, an extensive simulation study is conducted to evaluate the finite sample performance of the proposed method and the results indicate that it works well for practical situations. An illustrative example is also given in Section 3.4 and Section 3.5 contains some discussion and concluding remarks.

3.2 Notation, Models and Assumptions

Consider a failure time study that consists of n independent subjects. For subject i , let T_i denote the failure time of interest and suppose that there exists a vector of covariates $Z_i(t)$, which may depend on time t . For the relationship between T_i and $Z_i(t)$, in the following, we assume that given the history of covariates up to time t , the hazard function of T_i has the form

$$\lambda_T(t|Z_i(s), s \leq t) = \lambda_0(t) + \beta_0' Z_i(t). \quad (3.1)$$

That is, T_i follows the additive hazards model (Lin & Ying, 1994). In the above, $\lambda_0(t)$ denotes an unknown baseline hazard function and β_0 a vector of unknown regression parameters.

In the following, we assume that for some subjects, covariates $Z_i(t)$ are missing or not observed but there exists a vector of auxiliary covariates denoted by $X_i(t)$ that are known or observed for all subjects. Let V denote the set of indices of the subjects whose true covariates $Z_i(t)$ are known, \bar{V} the complement of V , and n_v and $n_{\bar{v}}$ the sizes of V and \bar{V} , respectively. The set V is usually referred to as the validation set. Note that here for simplicity, we assume that all components of the covariates are either known or missing together and some comments will be given below for the situation where the missing happens only on some of the components. Also we assume that V is a simple random sub-sample of the whole set of study subjects. For the data on the failure times T_i 's of interest, it will be assumed that each subject is observed only once at time C_i and the observed information consists only of C_i and $\delta_i = I(T_i \geq C_i)$, $i = 1, \dots, n$. That is, we have current status data.

In practice, the observation times C_i 's may depend on covariates too. For this, we will assume that given $Z_i(t)$, the hazard function of C_i has the form

$$\lambda_c(t) = \lambda_{c0}(t) e^{\gamma'_0 Z_i(t)},$$

where $\lambda_{c0}(t)$ denotes an unknown baseline hazard function and γ_0 a vector of unknown regression parameters as β_0 . That is, the C_i 's follow the proportional hazards model. In the following, we will assume that given Z_i , T_i and C_i are independent.

For each i , define $N_i(t) = I(C_i \leq \min(t, T_i))$ and $Y_i(t) = I(C_i \geq t)$. Then $N_i(t)$ is a counting process with the intensity process

$$\lambda_i(t|Z_i(s), s \leq t) = \lambda_{c0}(t) e^{-\Lambda_0(t)} e^{-\beta'_0 Z_i^*(t) + \gamma'_0 Z_i(t)} \triangleq \lambda_0^c(t) e^{-\beta'_0 Z_i^*(t) + \gamma'_0 Z_i(t)} \quad (3.2)$$

(Lin *et al.*, 1998), where $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ and $Z_i^*(t) = \int_0^t Z_i(s) ds$. Note that the equation above says that $\lambda_i(t|Z_i(s), s \leq t)$ satisfies the Cox proportional hazards model. Based on this, if all true covariates $Z_i(t)$ were observed, Lin *et al.* (1998) suggested to estimate both β_0 and γ_0 by using the partial likelihood function

$$L_P(\beta) = \prod_{i=1}^n \left\{ \frac{\exp(-\beta'_0 Z_i^*(C_i) + \gamma'_0 Z_i(C_i))}{\sum_{j=1}^n Y_j(C_i) \exp(-\beta'_0 Z_j^*(C_i) + \gamma'_0 Z_j(C_i))} \right\}^{\delta_i}. \quad (3.3)$$

It is apparent that $L_P(\beta)$ is not available for our situation. In the next section, we will generalize it to the auxiliary covariate situation.

3.3 Estimation Procedures

Now we consider estimation of regression parameters β_0 as well as γ_0 . For the simplicity of presentation, first we will consider the situation where $\gamma_0 = 0$ and then discuss the general situation.

Assume that $\gamma_0 = 0$. That is, the observation times C_i 's do not depend on the covariates. To estimate β_0 , as mentioned above, the partial likelihood function L_P is not available as the right side of equation (3.2) is unknown for some i . To deal with this, by following Prentice (1982) and for such i , it is natural to consider the intensity conditional on $X_i(t)$ rather than on $Z_i(t)$ or the induced intensity process

$$\tilde{\lambda}_i(t) = \lambda_0^c(t) E \left\{ e^{-\beta_0' Z_i^*(t)} | Y_i(t) = 1, X_i(t) \right\}.$$

It is easy to see that this is still the proportional hazard model with the relative risk function

$$\phi_i(t, \beta_0) \triangleq E \left\{ e^{-\beta_0' Z_i^*(t)} | Y_i(t) = 1, X_i(t) \right\}.$$

This suggests that for estimation of β_0 , one could employ the induced partial likelihood function

$$L_{IP}(\beta) = \prod_{i=1}^n \left\{ \frac{\Phi_i(C_i, \beta)}{\sum_{j=1}^n Y_j(C_i) \Phi_j(C_i, \beta)} \right\}^{\delta_i},$$

where

$$\Phi_i(t, \beta) = \phi_i(t, \beta) I(i \in \bar{V}) + \varphi_i(t, \beta) I(i \in V)$$

with $\varphi_i(t, \beta) = e^{-\beta' Z_i^*(t)}$.

Of course, $L_{IP}(\beta)$ is still not applicable since the relative risk function $\phi_i(t, \beta)$ is still unknown. On the other hand, Prentice (1982) argued that, if $\phi_i(t, \beta)$ could be

calculated or estimated as a function of β , then statistical inference could be based on the corresponding induced partial likelihood. To estimate $\phi_i(t, \beta)$, we will assume that the $X_i(t)$'s are discrete and take only finite values. Then an intuitive and unbiased estimate of $\phi_i(t, \beta)$ is given by

$$\hat{\phi}_i(t, \beta) = \frac{\sum_{j \in V} Y_j(t) I(X_j(t) = X_i(t)) \varphi_j(t, \beta)}{\sum_{j \in V} Y_j(t) I(X_j(t) = X_i(t))}$$

based on the validation set V . Some comments will be given below for the situation where the $X_i(t)$'s are continuous. Let $\hat{\Phi}_i(t, \beta)$ denote $\Phi_i(t, \beta)$ with $\phi_i(t, \beta)$ replaced by $\hat{\phi}_i(t, \beta)$. Then it is natural to base the estimation of β_0 on the following estimated partial likelihood function

$$L_{EP}(\beta) = \prod_{i=1}^n \left\{ \frac{\hat{\Phi}_i(C_i, \beta)}{\sum_{j=1}^n Y_j(C_i) \hat{\Phi}_j(C_i, \beta)} \right\}^{\delta_i}.$$

It is apparent that if all true covariates $Z_i(t)$'s are known or observed, $L_{EP}(\beta)$ reduces to $L_P(\beta)$.

Let $\hat{\beta}_{EP}$ denote the estimate of β_0 defined as the value of β that maximizes the estimated partial likelihood function $L_{EP}(\beta)$. For the determination and investigation of the asymptotic properties of $\hat{\beta}_{EP}$, we need some more notation. Let τ denote the longest follow-up time and define

$$\hat{S}_n^{(j)}(t, \beta) = \frac{1}{n} \sum_{i=1}^n Y_i(t) \hat{\Phi}_i^{(j)}(t, \beta), \quad j = 0, 1, 2,$$

$$D_{nl}^{(j)}(t, \beta) = \frac{1}{n} \sum_{i=1}^n Y_i(t) \left(\frac{\hat{\Phi}_i^{(j)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} \right)^{\otimes l} \Phi_i(t, \beta_0), \quad l = 1, 2; j = 1, 2,$$

$$s^{(j)}(t, \beta) = E \left\{ Y_i(t) \Phi_i^{(j)}(t, \beta) \right\}, \quad j = 0, 1, 2,$$

and

$$d_l^{(j)}(t, \beta) = E \left[Y_i(t) \left\{ \frac{\Phi_i^{(j)}(t, \beta)}{\Phi_i(t, \beta)} \right\}^{\otimes l} \Phi_i(t, \beta_0) \right], \quad l = 1, 2; j = 1, 2.$$

In the above, $f^{(j)}(t, \beta)$ denotes the j -order partial derivative of function $f(t, \beta)$ with respect to β , and $A^{\otimes l} = 1, A$ and AA' , respectively, for $l = 0, 1, 2$. Also define

$$\mathcal{F}_t = \sigma \{ C_i \leq t, X_i(u), Z_i(u), Y_i(u), i = 1, \dots, n, 0 \leq u \leq t \}$$

and

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) \Phi_i(u, \beta_0) \lambda_0^c(u) du, \quad i = 1, \dots, n,$$

which are martingales.

Using the notation above, we have

$$\begin{aligned} l_{EP}(\beta) &= \log \{ L_{EP}(\beta) \} = \sum_{i=1}^n \delta_i \left\{ \log \left(\hat{\Phi}_i(C_i, \beta) \right) - \log \left(\sum_{j=1}^n Y_j(C_i) \hat{\Phi}_j(C_i, \beta) \right) \right\}, \\ U_{EP}(\beta) &= \frac{\partial l_{EP}(\beta)}{\partial \beta} = \sum_{i=1}^n \int_0^\tau \frac{\hat{\Phi}_i^{(1)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} dN_i(t) - \sum_{i=1}^n \int_0^\tau \frac{\sum_{j=1}^n Y_j(t) \hat{\Phi}_j^{(1)}(t, \beta)}{\sum_{j=1}^n Y_j(t) \hat{\Phi}_j(t, \beta)} dN_i(t) \\ &= \sum_{i=1}^n \int_0^\tau \left(\frac{\hat{\Phi}_i^{(1)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} - \frac{\hat{S}_n^{(1)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} \right) dM_i(t) \\ &\quad + \sum_{i=1}^n \int_0^\tau \left(\frac{\hat{\Phi}_i^{(1)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} - \frac{\hat{S}_n^{(1)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} \right) \Phi_i(t, \beta_0) Y_i(t) \lambda_0^c(t) dt, \end{aligned}$$

and

$$\begin{aligned}
\frac{1}{n} \frac{\partial U_{EP}(\beta)}{\partial \beta} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \frac{\hat{\Phi}_i^{(2)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} - \left(\frac{\hat{\Phi}_i^{(1)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} \right)^{\otimes 2} - \left(\frac{\hat{S}_n^{(2)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} - \left(\frac{\hat{S}_n^{(1)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} \right)^{\otimes 2} \right) \right\} dM_i(t) \\
&+ \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \frac{\hat{\Phi}_i^{(2)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} - \left(\frac{\hat{\Phi}_i^{(1)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} \right)^{\otimes 2} - \left(\frac{\hat{S}_n^{(2)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} - \left(\frac{\hat{S}_n^{(1)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} \right)^{\otimes 2} \right) \right\} Y_i(t) d\Lambda^{\Phi_i}(t) \\
&\triangleq A_1(\tau, \beta) + A_2(\tau, \beta),
\end{aligned}$$

where $d\Lambda^{\Phi_i}(t) = \Phi_i(t, \beta_0) \lambda_0^c(t) dt$. It is easy to see that $l_{EP}(\beta)$ is a concave function with a unique maximum at $\hat{\beta}_{EP}$. That is, $\hat{\beta}_{EP}$ can be determined by solving the estimated partial likelihood score equation $U_{EP}(\beta) = 0$. The following theorems give the consistency and the asymptotic normality of $\hat{\beta}_{EP}$.

Theorem 3.1 *Suppose that the regularity conditions (1) - (4) given in the Appendix hold. Then as $n \rightarrow \infty$, we have $\hat{\beta}_{EP} - \beta_0 \xrightarrow{P} 0$.*

Theorem 3.2 *Suppose that the regularity conditions (1) - (6) given in the Appendix hold. Then as $n \rightarrow \infty$, $\sqrt{n}(\hat{\beta}_{EP} - \beta_0)$ is asymptotically normally distributed with mean zero and the covariance matrix*

$$\Sigma_{EP}(\beta_0) = \Sigma(\beta_0)^{-1} \{ (1 - \rho)\Sigma_1(\beta_0) + \rho\Sigma_2(\beta_0) \} \Sigma(\beta_0)^{-1}, \quad (3.4)$$

where ρ denotes the limit of n_v/n as $n \rightarrow \infty$,

$$\Sigma(\beta_0) = \int_0^\tau \left(d_2^{(1)}(t, \beta_0) - \frac{(s^{(1)}(t, \beta_0))^{\otimes 2}}{s^{(0)}(t, \beta_0)} \right) \lambda_0^c(t) dt,$$

$$\Sigma_1(\beta_0) = \int_0^\tau E \left\{ \left(Y_i(t) \frac{(\phi_i^{(1)}(t, \beta_0))^{\otimes 2}}{\phi_i(t, \beta_0)} \right) - \frac{(s^{(1)}(t, \beta_0))^{\otimes 2}}{s^{(0)}(t, \beta_0)} \right\} \lambda_0^c(t) dt,$$

and

$$\begin{aligned} \Sigma_2(\beta_0) &= E \left\{ \int_0^\tau \left(\frac{\varphi_j^{(1)}(t, \beta_0)}{\varphi_j(t, \beta_0)} - \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \right) dM_j(t) \right. \\ &\quad \left. - \frac{1-\rho}{\rho} \int_0^\tau \left(\frac{\phi_j^{(1)}(t, \beta_0)}{\phi_j(t, \beta_0)} - \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \right) (\varphi_j(t, \beta_0) - \phi_j(t, \beta_0)) Y_j(t) \lambda_0^c(t) dt \right\}^{\otimes 2}. \end{aligned}$$

The proofs of the two theorems above are sketched in the Appendix. Also in the Appendix, we show that the matrices $\Sigma(\beta_0)$, $\Sigma_1(\beta_0)$ and $\Sigma_2(\beta_0)$ can be consistently estimated by

$$\hat{\Sigma}(\beta) = -\frac{1}{n} \frac{\partial^2 l_{EP}(\beta)}{\partial \beta \partial \beta'}, \quad \hat{\Sigma}_1(\beta) = \frac{1}{n_{\bar{v}}} \sum_{i \in \bar{V}} \int_0^\tau \left(\frac{\hat{\phi}_i^{(1)}(t, \beta)}{\hat{\phi}_i(t, \beta)} - \frac{\hat{S}_n^{(1)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} \right)^{\otimes 2} dN_i(t),$$

and

$$\hat{\Sigma}_2(\beta) = \frac{1}{n_v} \sum_{i \in V} \left\{ \int_0^\tau \left(\frac{\varphi_i^{(1)}(t, \beta)}{\varphi_i(t, \beta)} - \frac{\hat{S}_n^{(1)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} \right) \left(dN_i(t) - Y_i(t) \varphi_i(t, \beta) d\hat{\Lambda}_0^c(t) \right) - \frac{n_{\bar{v}}}{n_v} \hat{Q}_i \right\}^{\otimes 2},$$

respectively, with β replaced by $\hat{\beta}_{EP}$. In the above,

$$\hat{Q}_i = \int_0^\tau \left(\frac{\hat{\phi}_i^{(1)}(t, \beta)}{\hat{\phi}_i(t, \beta)} - \frac{\hat{S}_n^{(1)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} \right) (\varphi_i(t, \beta) - \hat{\phi}_i(t, \beta)) Y_i(t) d\hat{\Lambda}_0^c(t)$$

and

$$\hat{\Lambda}_0^c(t) = \int_0^\tau \frac{\sum_{i=1}^n dN_i(t)}{\sum_{j=1}^n Y_j(t) \hat{\Phi}_j(t, \beta)}.$$

Thus the covariance matrix of $\hat{\beta}_{EP}$ can be consistently estimated by using (3.4) with $\Sigma(\beta_0)$, $\Sigma_1(\beta_0)$ and $\Sigma_2(\beta_0)$ replaced by their estimates given above and ρ by n_v/n .

Now we consider the general situation where γ_0 is unknown. That is, the observation times C_i 's may depend on the covariates through the proportional hazards model specified above. In this case, it is actually straightforward to generalize the estimation procedure described above. More specifically, one can easily show that the induced intensity process $\tilde{\lambda}_i(t)$ and partial likelihood function L_{IP} defined above now have the forms

$$\lambda_{e0}(t) e^{-\Lambda_0(t)} E \left\{ e^{-\beta'_0 Z_i^*(t) + \gamma'_0 Z_i(t)} | Y_i(t) = 1, X_i(t) \right\}, \quad (3.5)$$

and

$$\prod_{i=1}^n \left(\frac{\tilde{\Phi}_i(C_i, \beta, \gamma)}{\sum_{j=1}^n Y_j(C_i) \tilde{\Phi}_j(C_i, \beta, \gamma)} \right)^{\delta_i},$$

respectively. In the above

$$\tilde{\Phi}_i(t, \beta, \gamma) = \left\{ \tilde{\phi}_i(t, \beta, \gamma) I(i \in \bar{V}) + \tilde{\varphi}_i(t, \beta, \gamma) I(i \in V) \right\},$$

where $\tilde{\phi}_i(t, \beta, \gamma) = E\{e^{-\beta' Z_i^*(t) + \gamma' Z_i(t)} | Y_i(t) = 1, X_i(t)\}$ and $\tilde{\varphi}_i(t, \beta, \gamma) = e^{-\beta' Z_i^*(t) + \gamma' Z_i(t)}$.

For the simultaneous estimation of β_0 and γ_0 , motivated by $L_{EP}(\beta)$, we can employ the estimated partial likelihood function

$$L_{EP}^*(\beta, \gamma) = \prod_{i=1}^n \left(\frac{\hat{\tilde{\Phi}}_i(C_i, \beta, \gamma)}{\sum_{j=1}^n Y_j(C_i) \hat{\tilde{\Phi}}_j(C_i, \beta, \gamma)} \right)^{\delta_i}, \quad (3.6)$$

where $\hat{\Phi}_i(t, \beta, \gamma)$ is equal to $\tilde{\Phi}_i(t, \beta, \gamma)$ with $\tilde{\phi}_i(t, \beta, \gamma)$ replaced by

$$\hat{\phi}_i(t, \beta, \gamma) = \frac{\sum_{j \in V} Y_j(t) I(X_j(t) = X_i(t)) \tilde{\varphi}_j(t, \beta, \gamma)}{\sum_{j \in V} Y_j(t) I(X_j(t) = X_i(t))}.$$

It is easy to see that the only difference between $L_{EP}(\beta)$ and $L_{EP}^*(\beta, \gamma)$ is the dimension of the covariates involved.

Define $\theta = (\beta', \gamma')'$ and $\theta_0 = (\beta_0', \gamma_0')'$. Also define the estimate $\bar{\theta}_{EP} = (\bar{\beta}'_{EP}, \bar{\gamma}'_{EP})'$ of θ as the value of θ that maximizes the the estimated partial likelihood function $L_{EP}^*(\beta, \gamma)$. Then one can show that as with $\hat{\beta}_{EP}$, $\bar{\theta}_{EP}$ is consistent and the distribution of $\sqrt{n}(\bar{\theta}_{EP} - \theta_0)$ can be approximated by the normal distribution with mean zero and the covariance matrix

$$\hat{\Sigma}^{-1}(\bar{\theta}) \left\{ \left(1 - \frac{n_v}{n}\right) \hat{\Sigma}_1(\bar{\theta}) + \frac{n_v}{n} \hat{\Sigma}_2(\bar{\theta}) \right\} \hat{\Sigma}^{-1}(\bar{\theta}).$$

Here $\hat{\Sigma}(\bar{\theta})$, $\hat{\Sigma}_1(\bar{\theta})$ and $\hat{\Sigma}_2(\bar{\theta})$ are $\hat{\Sigma}(\beta)$, $\hat{\Sigma}_1(\beta)$ and $\hat{\Sigma}_2(\beta)$ with all $\varphi_i^{(j)}(t, \beta)$, $\hat{\phi}_i^{(j)}(t, \beta)$ and $\hat{\Phi}_i^{(j)}(t, \beta)$ replaced by $\tilde{\varphi}_i^{(j)}(t, \bar{\theta})$, $\hat{\phi}_i^{(j)}(t, \bar{\theta})$ and $\hat{\Phi}_i^{(j)}(t, \bar{\theta})$, respectively, $j = 0, 1$.

3.4 Numerical Studies

In this section, we conduct an extensive simulation study to evaluate the performance of the estimation procedures proposed in the previous section with the focus on the situation where $\gamma_0 = 0$. As mentioned above, the estimation procedure based on $L_{EP}^*(\beta, \gamma)$ is a simple generalization of that based on $L_{EP}(\beta)$ by incorporating a large vector of covariates. In the simulation study, we considered three situations for the true covariate $Z_i(t)$, generated from the Bernoulli distribution $B(1, 0.5)$, the

uniform distribution $U(0, \sqrt{3})$ or the normal distribution $N(0, 0.25)$. Given the Z_i 's, the failure times T_i 's of interest were assumed to follow model (3.1) with $\lambda_0(t) = 1$ and the observation times C_i 's were generated from the exponential distribution with the hazard rate $\lambda_c(t) = 0.5, 1.0, \text{ or } 1.5$. For the generation of the auxiliary covariate X_i , we first generated $W_i = Z_i + e_i$ with the e_i 's being the random sample from $N(0, \sigma^2)$ and then define $X_i = I(W_i \leq 0) + 2I(W_i > 0)$. Note that here the parameter σ controls the strength of the association between Z_i and W_i and also between Z_i and X_i and as σ increases, X_i becomes less informative about Z_i . The results given below are based on 1000 replications and the sample size $n = 200$.

Table 3.1 presents the results on estimation of β_0 with the Z_i 's generated from $B(1, 0.5)$, $\beta_0 = -0.5, 0, \text{ or } 0.5$, $\sigma = 0.1 \text{ or } 0.4$, and the fraction of the validation set being 50% of the total sample. The results include the estimated biases (Bias) given by the averages of the point estimates $\hat{\beta}_{EP}$ minus the true values, the averages of the standard error estimates (SEE), the sampling standard errors of the point estimates (SSE), and the 95% empirical coverage probabilities (CP). For comparison, we also obtained the naive estimates of β_0 , denoted by $\hat{\beta}_P$, given by maximizing $L_P(\beta)$ given in (3) based only on the validation sample, and include them in the table along with the relative efficiency (RE) given by $\text{SSE}(\hat{\beta}_P)/\text{SSE}(\hat{\beta}_{EP})$. The results indicate that the proposed estimate seems to be approximately unbiased and the proposed variance estimate also seems to be reasonable. It is clear that $\hat{\beta}_{EP}$ is more efficient, suggesting that the auxiliary information and the proposed estimation procedure should be used.

Tables 3.2 and 3.3 give the results obtained on estimation of β_0 with the Z_i 's generated from $U(0, \sqrt{3})$ and $N(0, 0.25)$, respectively. All other set-ups are the same as those used for Table 3.1. All results are similar to those given in Table 3.1 and yield

the same conclusions. Also they all indicate that as expected, the estimate $\hat{\beta}_{EP}$ is more accurate for smaller σ . In order to assess the normal distribution approximation to the distribution of $\hat{\beta}_{EP}$, we studied the quantile plots of the standardized $\hat{\beta}_{EP}$ against the standard normal random variable. Figures 3.1, 3.2 and 3.3 present the quantile plots corresponding to the simulated data used in Tables 3.1, 3.2 and 3.3 with $\beta_0 = -0.5, 0$ and 0.5 , $\lambda_c = 1$ and $\sigma = 0.1$, respectively. They suggest that the normal approximation is reasonable for the situations considered here. We also considered many other set-ups and obtained similar results.

To illustrate the estimation procedures proposed in the previous sections, we apply them to a set of current status data arising from a 2-year tumorigenicity study conducted by National Toxicology Program (Wang et al., 2008). In the study, the groups of 100 male and female F344/N rats and B6C3F₁ mice were exposed to chloroprene at concentrations of 0, 12.8, 32, or 80 ppm by inhalation for 2 years. Each animal was examined for various tumors at its death time. Some animals died naturally during the study and the others were sacrificed at the end of study. Since the tumor status was only examined at death time, thus only current status data are available for the tumor onset time, the variable of interest. In the following, we will focus on the animals in the control (0 ppm) and high dose (80 ppm) groups with respect to a specific type of lung tumor, the Alveolar/Bronchiolar Adenoma (A/B A). One objective of interest is to compare the tumor growth rates between the two groups and investigate the dose effect.

For the analysis, let T denote the onset time of A/B A tumor and C the death time in months. Define Z to be 1 for the mice in the high dose group and 0 otherwise. To generate auxiliary covariates, we randomly select the $1 - p$ percentage of the animals

for whom the covariate Z is assumed to be missing and generate the auxiliary covariate X in the same way as in the simulation study above for all animals. Note that here p represents the percentage of the subjects in the validation set. Table 3.4 gives the estimated dose effects given by the two estimation procedures proposed in the previous section and the naive estimation procedure described above with $p = 0.3, 0.5$ or 0.7 and $\sigma = 0.1, 0.2$ or 0.4 . In addition, the table includes the estimated standard errors and the p -values for testing no dose effect. The top part of the table is for the case where we assume $\gamma_0 = 0$, while the bottom part corresponds to the general situation. Note that for both parts, the true values or estimates of the dose effect, corresponding to $p = 100\%$ or no missing covariates, are also given for comparison.

One can see from Table 3.4 that for the case assuming $\gamma_0 = 0$, all of the estimates given by the proposed method suggest that there seems to exist moderate or significant difference between the tumor growth rates of the two dose groups, which is the same as the conclusion that one would obtain if there are no missing covariates. On the other hand, almost all estimates given by the naive method suggest that the two tumor rates are statistically same. For the situation that dose not assume $\gamma_0 = 0$, again all of the estimates given by the proposed method yielded the same conclusion as that one would obtain if there are no missing covariates. In contrast, the results given by the naive method that ignores the missing covariates depend on the values of σ , or the association between the true and auxiliary covariates. In summary, these results indicate that the use of auxiliary covariates is needed and necessary when there exist missing covariates.

3.5 Discussion and Concluding Remarks

This chapter discussed regression analysis of current status failure time data arising from the additive hazards model when there exist missing and auxiliary covariates. For the problem, two estimated partial likelihood approaches were developed for estimation of regression parameters. The approaches are generalizations of the method proposed in Lin *et al.* (1998) for the case where all covariates can be observed. The asymptotic properties of the proposed estimates have been established. The numerical results indicate that the proposed methodology works well for practical situations and performs better than the naive method that bases the estimation only on the subjects with known covariates.

As the method given in Lin *et al.* (1998), the presented estimation procedures have the advantage that they do not involve any complicated nonparametric estimation of the distribution of the failure time of interest and thus can be easily implemented. In particular, the maximization of the proposed partial likelihood functions or the determination of the proposed estimates can be easily carried out by any software packages that can maximize a partial likelihood function given by right-censored failure time data. For example, one such function, which was used in the numerical study above, is the Matlab function *fminsearch*. Note that in the preceding sections, we have assumed that auxiliary covariates are discrete. Sometimes this may not be true and they are continuous. For the latter case, one can apply the same idea discussed above, but needs to employ different approaches to estimate $\phi_i(t, \beta)$. One such approach could be the kernel nonparametric estimation discussed in Nadaraya (1964) and Watson (1964) among others.

Also in the preceding sections, we have assumed that all components of covariates

are either observed or missing together. Actually the estimation methods developed above can be used to the case where some components are observed and some are subject to missing with auxiliary information. In this case, we can rewrite the additive hazards model (3.1) as

$$\lambda_T(t|Z_i(s), \xi_i(s), s \leq t) = \lambda_0(t) + \beta'_1 Z_i(t) + \beta'_2 \xi_i(t),$$

where $Z_i(t)$ denotes the part of covariates that may be subject to missing and $\xi_i(t)$ the part that is always observed. It follows that for the case where $\gamma_0 = 0$, the relative risk function has the form

$$e^{-\beta'_2 \xi_i^*(t)} E \left\{ e^{-\beta'_1 Z_i^*(t)} | Y_i(t) = 1, X_i(t), \xi_i(t) \right\}$$

and we have

$$\Phi_i(t, \beta) = \left\{ \psi_i(t, \beta_1) I(i \in \bar{V}) + \varphi_i(t, \beta_1) I(i \in V) \right\} e^{-\beta'_2 \xi_i^*(t)}.$$

For the general situation, the relative risk function has the form

$$e^{-\beta'_2 \xi_i^*(t) + \gamma'_2 \xi_i(t)} E \left\{ e^{-\beta'_1 Z_i^*(t) + \gamma'_1 Z_i(t)} | Y_i(t) = 1, X_i(t), \xi_i(t) \right\}$$

and we have

$$\Phi_i(t, \beta, \gamma) = \left\{ \tilde{\psi}_i(t, \beta_1, \gamma_1) I(i \in \bar{V}) + \tilde{\varphi}_i(t, \beta_1, \gamma_1) I(i \in V) \right\} e^{-\beta'_2 \xi_i^*(t) + \gamma'_2 \xi_i(t)}.$$

In the above,

$\xi_i^*(t) = \int_0^t \xi_i(s) ds$, $\beta = (\beta_1', \beta_2')'$, $\gamma = (\gamma_1', \gamma_2')'$, $\psi_i(t, \beta) = E \{e^{-\beta' Z_i^*(t)} | Y_i(t) = 1, X_i(t), \xi_i(t)\}$, and $\tilde{\psi}_i(t, \beta, \gamma) = E \{e^{-\beta' Z_i^*(t) + \gamma' Z_i(t)} | Y_i(t) = 1, X_i(t), \xi_i(t)\}$. Similarly as before, one can develop an estimation procedure for regression parameters and establish their asymptotic properties.

There exist several directions for future research related to the problem discussed here. One is that although the proposed methods can be easily implemented, they may not be fully efficient. To develop a more efficient estimation procedure, by following the suggestion of Lin *et al.* (1998), one may consider a two-stage approach that estimates γ_0 first and then estimates β_0 . More specifically, define $r_i(u, \gamma) = e^{\gamma' Z_i(u)}$, $e_i(u, \gamma) = E(e^{\gamma' Z_i(u)} | Y_i(u) = 1, X_i(u))$ and $R_i(u, \gamma) = e_i(u, \gamma)I(i \in \bar{V}) + r_i(u, \gamma)I(i \in V)$. Then one can estimate γ_0 by the value of γ_0 that maximizes

$$L^c(\gamma) = \prod_{i=1}^n \frac{\hat{R}_i(C_i, \gamma)}{\sum_{j=1}^n Y_j(C_j) \hat{R}_j(C_j, \gamma)},$$

where

$$\begin{aligned} \hat{R}_i(t, \gamma) &= \hat{e}_i(t, \gamma)I(i \in \bar{V}) + r_i(t, \gamma)I(i \in V) \\ &= \frac{\sum_{j \in \bar{V}} Y_j(t) I(X_j(t) = X_i(t)) r_i(t, \gamma)}{\sum_{j \in \bar{V}} Y_j(t) I(X_j(t) = X_i(t))} I(i \in \bar{V}) + r_i(t, \gamma)I(i \in V). \end{aligned}$$

Let $\tilde{\gamma}$ denote the estimate of γ_0 defined above. Then β_0 can be estimated by maximizing

$$L(\beta, \tilde{\gamma}) = \prod_{i=1}^n \left(\frac{\hat{\Phi}_i(C_i, \beta, \hat{\gamma})}{\sum_{j=1}^n Y_j(C_j) \hat{\Phi}_j(C_j, \beta, \hat{\gamma})} \right)^{\delta_i}. \quad (3.7)$$

Another method for the development of an efficient estimation procedure is to generalize the efficient score estimation procedure given in Martinussen & Scheike (2002), who discussed the same problem as that considered in Lin *et al.* (1998). It is easy

to see that both of these two ideas are much more complicated than that discussed above.

Throughout the chapter, we have assumed that the failure times of interest follow the additive hazards model (3.1). Of course, sometimes this may not be true and it would be useful to develop similar estimation procedures for other regression models such as the proportional hazards model. Also the method proposed above only applies to case I interval-censored data and it would be helpful to generalize it to more general case II interval-censored data (Sun, 2006). The latter means that the failure time of interest is observed to belong to some intervals and reduces to case I interval-censored data if the interval contains either zero or infinity.

	β	σ	λ_c	Bias	SSE	SEE	CP	RE
$\hat{\beta}_{EP}$	-0.5	0.1	0.5	-0.0235	0.2482	0.2180	0.943	1.4142
$\hat{\beta}_{EP}$			1.0	-0.0062	0.2721	0.2545	0.942	1.4425
$\hat{\beta}_{EP}$			1.5	-0.0035	0.3281	0.3114	0.948	1.3977
$\hat{\beta}_{EP}$		0.4	0.5	-0.0152	0.2638	0.2420	0.942	1.3306
$\hat{\beta}_{EP}$			1.0	-0.0083	0.2918	0.2783	0.947	1.3451
$\hat{\beta}_{EP}$			1.5	-0.0119	0.3625	0.3388	0.944	1.2651
$\hat{\beta}_P$			0.5	-0.0449	0.3510	0.3359	0.953	
$\hat{\beta}_P$			1.0	-0.0471	0.3925	0.3739	0.955	
$\hat{\beta}_P$			1.5	-0.0127	0.4586	0.4416	0.944	
	0	0.1	0.5	-0.0065	0.2868	0.2687	0.951	1.5847
$\hat{\beta}_{EP}$			1.0	0.0072	0.3238	0.2976	0.943	1.4336
$\hat{\beta}_{EP}$			1.5	-0.0045	0.3512	0.3491	0.945	1.4846
$\hat{\beta}_{EP}$		0.4	0.5	-0.0013	0.3345	0.2979	0.953	1.3587
$\hat{\beta}_{EP}$			1.0	0.0011	0.3521	0.3275	0.948	1.3184
$\hat{\beta}_{EP}$			1.5	-0.0030	0.3983	0.3819	0.953	1.3091
$\hat{\beta}_P$			0.5	0.0248	0.4545	0.4238	0.972	
$\hat{\beta}_P$			1.0	0.0067	0.4642	0.4438	0.965	
$\hat{\beta}_P$			1.5	0.0264	0.5214	0.5049	0.957	
$\hat{\beta}_{EP}$	0.5	0.1	0.5	0.0260	0.3884	0.3557	0.946	1.5633
$\hat{\beta}_{EP}$			1.0	0.0225	0.3757	0.3557	0.949	1.5193
$\hat{\beta}_{EP}$			1.5	-0.0068	0.3989	0.3978	0.953	1.5275
$\hat{\beta}_{EP}$		0.4	0.5	0.0459	0.4527	0.3941	0.943	1.3413
$\hat{\beta}_{EP}$			1.0	0.0167	0.4188	0.3901	0.945	1.3629
$\hat{\beta}_{EP}$			1.5	0.0265	0.4716	0.4380	0.947	1.2920
$\hat{\beta}_P$			0.5	0.0578	0.6072	0.5650	0.961	
$\hat{\beta}_P$			1.0	0.0363	0.5708	0.5510	0.966	
$\hat{\beta}_P$			1.5	0.0319	0.6093	0.5919	0.965	

Table 3.1: Estimation of β based on simulated data with $Z \sim B(1, 0.5)$.

	β	σ	λ_c	Bias	SSE	SEE	CP	RE
$\hat{\beta}_{EP}$	-0.5	0.1	0.5	-0.0162	0.1859	0.2272	0.958	1.3131
$\hat{\beta}_{EP}$			1.0	-0.0199	0.2545	0.2613	0.962	1.2570
$\hat{\beta}_{EP}$			1.5	-0.0159	0.3205	0.3208	0.961	1.2799
$\hat{\beta}_{EP}$		0.4	0.5	-0.0115	0.2053	0.3517	0.957	1.1890
$\hat{\beta}_{EP}$			1.0	-0.0070	0.2615	0.2862	0.963	1.2233
$\hat{\beta}_{EP}$			1.5	-0.0068	0.3423	0.3599	0.959	1.1984
$\hat{\beta}_P$			0.5	-0.0189	0.2441	0.2411	0.958	
$\hat{\beta}_P$			1.0	-0.0377	0.3199	0.3136	0.959	
$\hat{\beta}_P$			1.5	-0.0377	0.4102	0.3972	0.947	
$\hat{\beta}_{EP}$	0	0.1	0.5	0.0032	0.3199	0.3083	0.954	1.3298
$\hat{\beta}_{EP}$			1.0	0.0060	0.3288	0.3212	0.963	1.3209
$\hat{\beta}_{EP}$			1.5	0.0020	0.4013	0.3817	0.954	1.2781
$\hat{\beta}_{EP}$		0.4	0.5	-0.0014	0.3299	0.3322	0.959	1.2895
$\hat{\beta}_{EP}$			1.0	0.0023	0.3628	0.3604	0.958	1.1971
$\hat{\beta}_{EP}$			1.5	0.0023	0.4327	0.4205	0.960	1.1853
$\hat{\beta}_P$			0.5	0.0156	0.4254	0.4091	0.967	
$\hat{\beta}_P$			1.0	-0.0125	0.4343	0.4427	0.966	
$\hat{\beta}_P$			1.5	-0.0142	0.5129	0.5112	0.958	
$\hat{\beta}_{EP}$	0.5	0.1	0.5	0.0216	0.4493	0.4221	0.951	1.4407
$\hat{\beta}_{EP}$			1.0	0.0276	0.4428	0.4155	0.949	1.3957
$\hat{\beta}_{EP}$			1.5	0.0118	0.4967	0.4591	0.946	1.3511
$\hat{\beta}_{EP}$		0.4	0.5	0.0409	0.5077	0.4665	0.949	1.2750
$\hat{\beta}_{EP}$			1.0	0.0130	0.4775	0.4507	0.950	1.2942
$\hat{\beta}_{EP}$			1.5	0.0074	0.5327	0.5023	0.952	1.2598
$\hat{\beta}_P$			0.5	0.0366	0.6473	0.6064	0.970	
$\hat{\beta}_P$			1.0	0.0456	0.6180	0.5955	0.968	
$\hat{\beta}_P$			1.5	0.0338	0.6711	0.6419	0.953	

Table 3.2: Estimation of β based on simulated data with $Z \sim U(0, \sqrt{3})$.

	β	σ	λ_c	Bias	SSE	SEE	CP	RE
$\hat{\beta}_{EP}$	-0.5	0.1	0.5	-0.0192	0.3111	0.2925	0.948	1.3208
$\hat{\beta}_{EP}$			1.0	-0.0110	0.3448	0.3201	0.945	1.3736
$\hat{\beta}_{EP}$			1.5	-0.0051	0.3986	0.3748	0.950	1.3846
$\hat{\beta}_{EP}$		0.4	0.5	-0.0289	0.3184	0.3205	0.945	1.2905
$\hat{\beta}_{EP}$			1.0	-0.0203	0.3741	0.3415	0.950	1.2660
$\hat{\beta}_{EP}$			1.5	-0.0068	0.4457	0.4116	0.947	1.2383
$\hat{\beta}_P$			0.5	-0.0544	0.4109	0.3925	0.958	
$\hat{\beta}_P$			1.0	-0.0526	0.4736	0.4424	0.946	
$\hat{\beta}_P$			1.5	-0.0435	0.5519	0.5180	0.948	
$\hat{\beta}_{EP}$	0	0.1	0.5	-0.0030	0.3084	0.2849	0.941	1.3693
$\hat{\beta}_{EP}$			1.0	-0.0063	0.3308	0.3172	0.948	1.3736
$\hat{\beta}_{EP}$			1.5	-0.0012	0.3938	0.3788	0.947	1.4137
$\hat{\beta}_{EP}$		0.4	0.5	0.0015	0.3415	0.3114	0.942	1.2366
$\hat{\beta}_{EP}$			1.0	-0.0054	0.3798	0.3494	0.946	1.1964
$\hat{\beta}_{EP}$			1.5	0.0010	0.4266	0.4116	0.947	1.3050
$\hat{\beta}_P$			0.5	0.0063	0.4223	0.4115	0.958	
$\hat{\beta}_P$			1.0	-0.0087	0.4544	0.4478	0.952	
$\hat{\beta}_P$			1.5	0.0050	0.5567	0.5219	0.958	
$\hat{\beta}_{EP}$	0.5	0.1	0.5	0.0115	0.2954	0.2791	0.953	1.4320
$\hat{\beta}_{EP}$			1.0	0.0130	0.3377	0.3181	0.954	1.4516
$\hat{\beta}_{EP}$			1.5	0.0267	0.4115	0.3771	0.948	1.2994
$\hat{\beta}_{EP}$		0.4	0.5	0.0099	0.3282	0.3094	0.942	1.2888
$\hat{\beta}_{EP}$			1.0	0.0114	0.3614	0.3438	0.951	1.3564
$\hat{\beta}_{EP}$			1.5	-0.0112	0.4404	0.4113	0.953	1.2141
$\hat{\beta}_P$			0.5	0.0580	0.4230	0.3945	0.952	
$\hat{\beta}_P$			1.0	0.0381	0.4902	0.4420	0.957	
$\hat{\beta}_P$			1.5	0.0578	0.5347	0.5017	0.959	

Table 3.3: Estimation of β based on simulated data with $Z \sim N(0, 0.25)$.

Assuming $\gamma_0 = 0$							
		Proposed method			Naive method		
σ	p	$\hat{\beta}_{EP}$	SEE	p -values	$\hat{\beta}_P$	SEE	p -values
0.1	30%	-0.0168	0.0084	0.0455	-0.0243	0.0177	0.1698
	50%	-0.0168	0.0085	0.0481	-0.0160	0.0116	0.1678
	70%	-0.0168	0.0086	0.0508	-0.0208	0.0105	0.0476
0.2	30%	-0.0168	0.0084	0.0455	-0.0091	0.0154	0.5546
	50%	-0.0181	0.0085	0.0332	-0.0069	0.0116	0.5520
	70%	-0.0168	0.0086	0.0508	-0.0104	0.0101	0.3031
0.4	30%	-0.0161	0.0091	0.0769	-0.0062	0.0085	0.4657
	50%	-0.0157	0.0089	0.0777	-0.0105	0.0120	0.3816
	70%	-0.0207	0.0091	0.0229	-0.0130	0.0102	0.2025
	100%	-0.0163	0.0086	0.0580	-0.0163	0.0086	0.0580
Without assuming $\gamma_0 = 0$							
		Proposed method			Naive method		
σ	p	$\bar{\beta}_{EP}$	SEE	p -values	$\bar{\beta}_P$	SEE	p -values
0.1	30%	0.0776	0.0617	0.2086	0.1340	0.1989	0.0298
	50%	0.0776	0.0623	0.2086	0.3344	0.4253	0.0000
	70%	0.0776	0.0617	0.2084	0.1385	0.1057	0.0248
0.2	30%	0.0748	0.0607	0.2177	0.0209	0.0807	0.7301
	50%	0.0776	0.0619	0.2097	0.0749	0.0911	0.1995
	70%	0.0776	0.0619	0.2103	0.0563	0.0598	0.3636
0.4	30%	0.0887	0.0975	0.3628	0.3168	0.3184	0.0012
	50%	0.0413	0.0593	0.4860	0.0600	0.0795	0.3121
	70%	0.0569	0.0588	0.3336	0.0767	0.0742	0.1921
	100%	0.0776	0.0619	0.2103	0.0776	0.0619	0.2103

Table 3.4: Analysis results for the mice example.

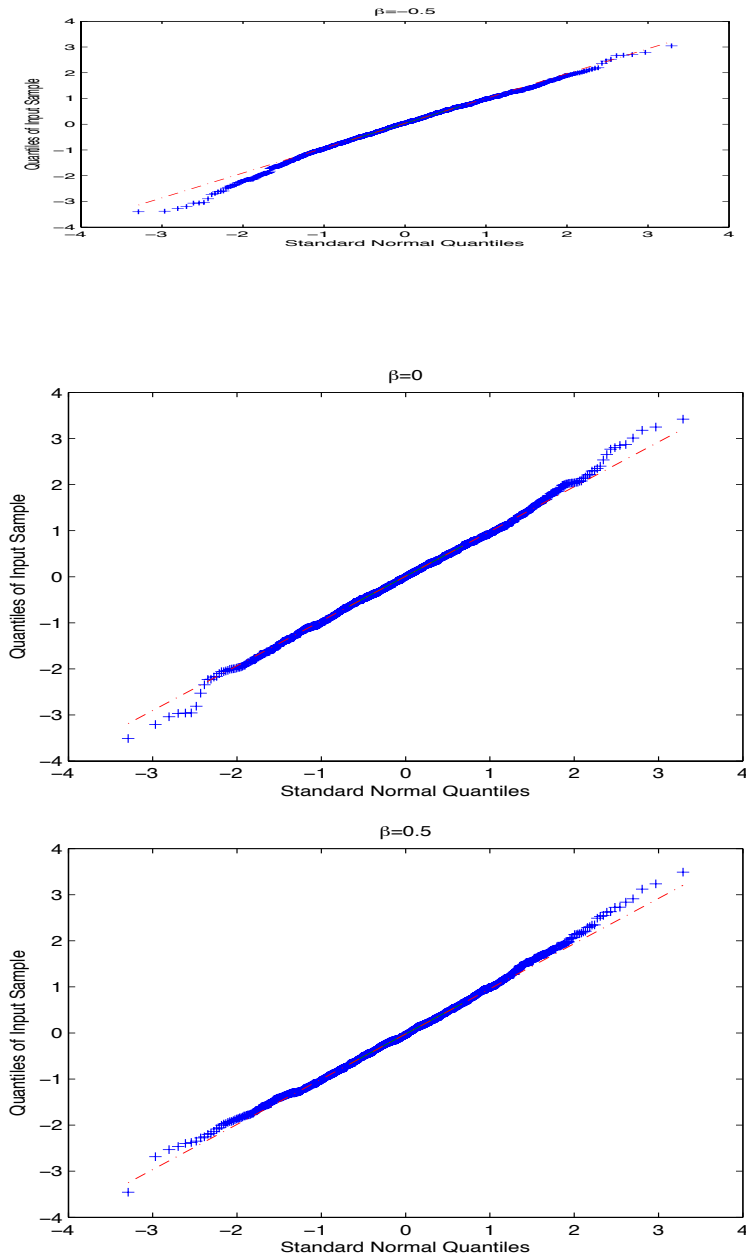


Figure 3.1: QQ-plot of $\hat{\beta}_{EP}$ corresponding to Table 3.1.

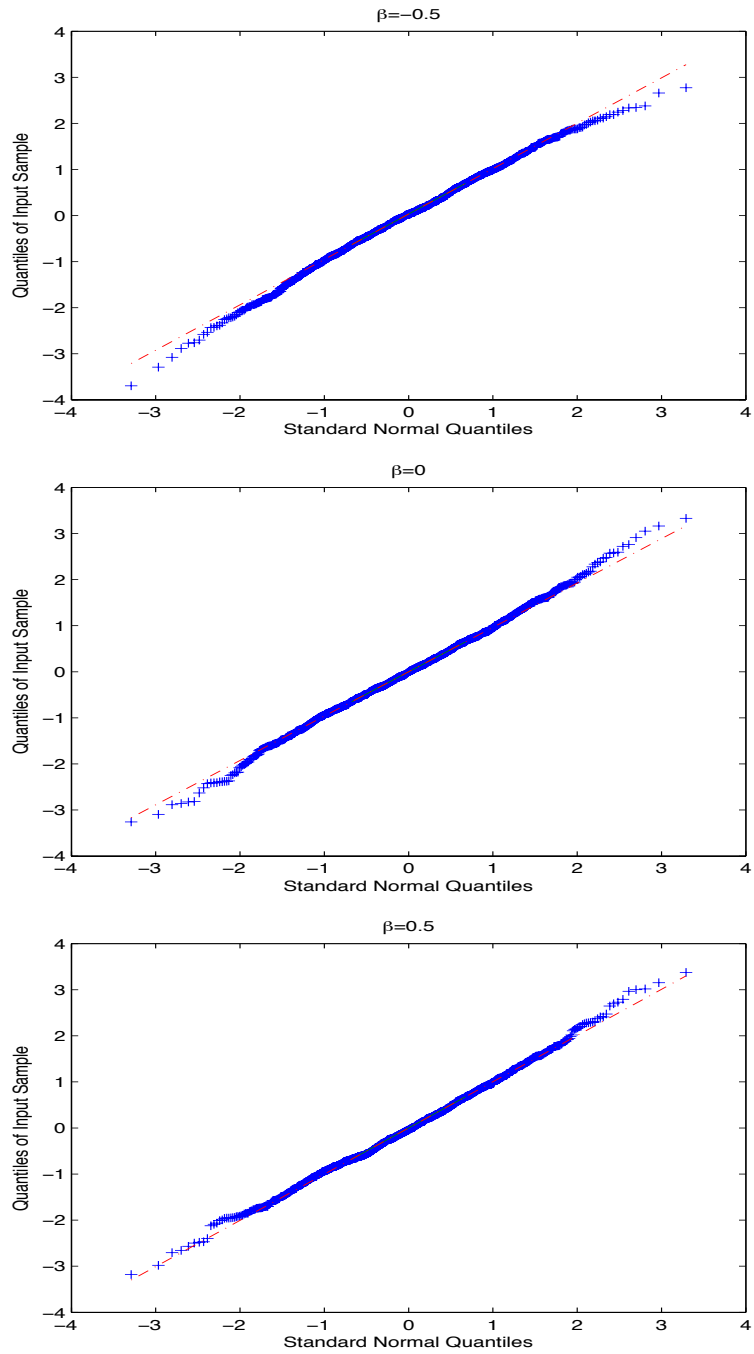


Figure 3.2: QQ-plot of $\hat{\beta}_{EP}$ corresponding to Table 3.2.

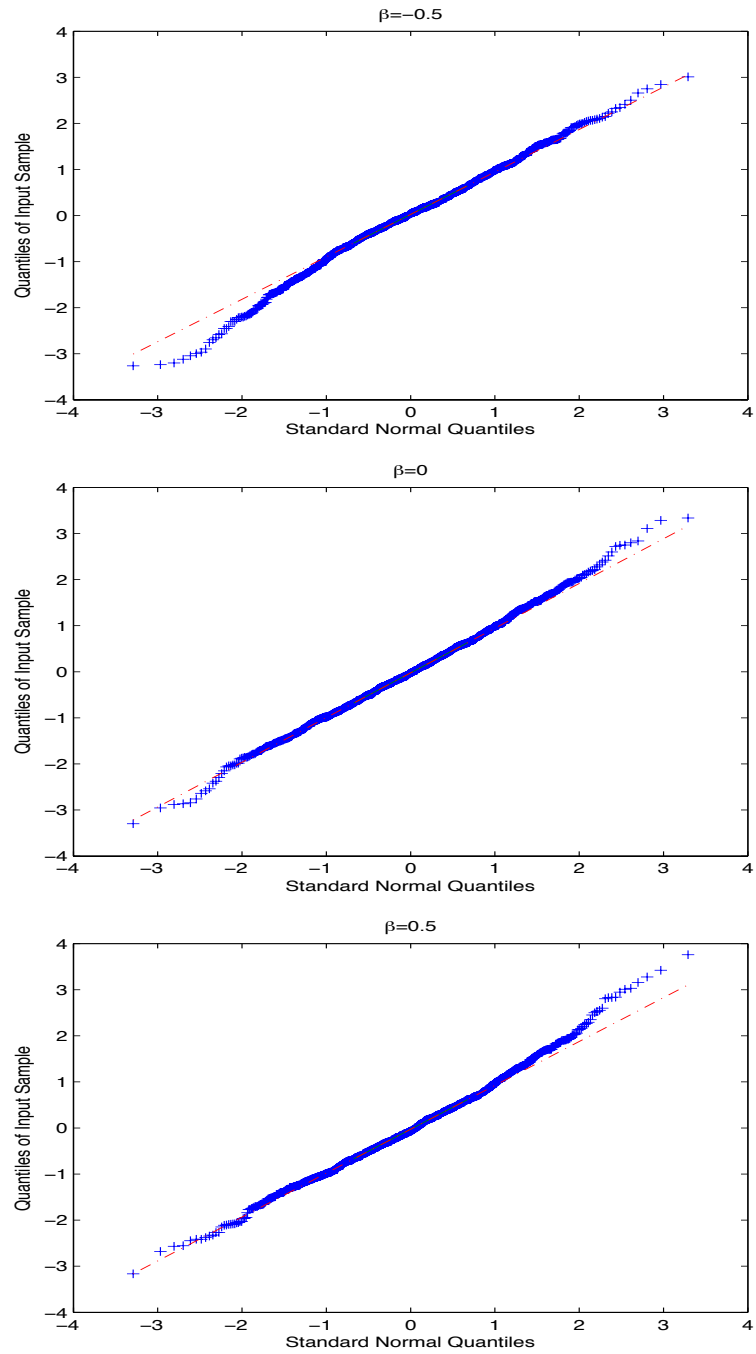


Figure 3.3: QQ-plot of $\hat{\beta}_{EP}$ corresponding to Table 3.3.

Chapter 4

Spline-Based Regression Analysis of Current Status Data with Dependent Censoring

4.1 Introduction

The statistical analysis of current status data has recently attracted considerable attention and a number of semi-parametric and non-parametric methods have been proposed. For example, Andersen and Ronn (1995) and Sun (1999) developed non-parametric tests for treatment comparison based on current status data. Among others, Huang (1996), Rossini and Tsiatis (1996), and Lin et al (1998) investigated regression analysis of current status data under the proportional hazards model, proportional odds model and additive hazards model, respectively. Most of these papers assumed that the observation time is independent of the survival time of interest completely or conditional on covariates. That is, the observation time is treated as a

noninformative censoring time.

However, the dependence between observation time and survival time may exist in practical examples, such as the tumorigenicity example we discussed earlier. For the current status data when the observation time could depend on the survival time, some approaches have been proposed. Most of the literature has focused on assuming a multistate model which is usually complicated (Dewanji and Kalbfleisch, 1986; Dinse, 1991; Lindsey and Ryan, 1993). Lagakos and Louis (1988) proposed another approach by assuming that the tumour lethality is known. Zhang et al. (2005) postulated a frailty variable to characterize the correlation between the observation time and survival time where estimating equations for both regression parameters about the survival time and the observation time were constructed. More recently, Chen et al. (2012) considered a general class of semiparametric transformation models with log-normal frailty to allow correlation between the two times and an EM algorithm is used for inference.

Note that in tumorigenicity experiments, the animals could die from natural causes or the predetermined sacrifice, that is, observation time can sometimes consist of two parts: time that depends on survival time (e.g. natural death time) and time that is independent of survival time (e.g. sacrifice time). In this case, the real observed time is the minimum of the two times. In the following, for convenience, the two times will be referred to as observation time and censoring time, respectively.

In this chapter, we discuss regression analysis of current status data under the proportional hazards model and takes into account the dependence between the observation time and the survival time. To develop the maximum likelihood estimation approach, following Zhang et al. (2010), we employ the sieve method by using I -

splines (Lu et al., 2007; Ramsay, 1988) to approximate the baseline cumulative hazard functions for both the observation time and the survival time. Note that unlike B-splines in Zhang et al. (2010), for which some order restrictions are needed to guarantee the monotonicity, the monotonicity of the I -spline functions is guaranteed by the nonnegativity constraints and thus no order restrictions is needed in optimization. For the correlation between the observation time and survival time, the copula model will be applied (Wang et al. 2012; Titman 2013).

The rest of this chapter is organized as follows. We will begin in Section 4.2 with describing the notation, assumptions and models that will be used throughout the chapter as well as the likelihood function. Section 4.3 presents the sieve semiparametric maximum likelihood estimation approach for all parameters. The resulting estimates are shown to be consistent and furthermore, we prove that the estimated regression parameters are asymptotically efficient and normally distributed. Section 4.4 gives some simulation results obtained from a simulation study performed to evaluate the finite sample performance of the proposed method and they suggest that the approach works well for the situations considered. In Section 4.5, we apply the method to a real data set and Section 4.6 contains some discussion and concluding remarks.

4.2 Cox Models and Likelihood

Consider a failure time study that involves n independent subjects. Let T denote the failure time of interest, C the observation time which may depend on T and Z a p -dimensional vector of covariates. In the following, we assume that given Z , T and

C both follow the proportional hazards model with

$$\lambda^{(T)}(t|Z) = \lambda_1(t) \exp(Z'\beta), \quad (4.1)$$

$$\lambda^{(C)}(c|Z) = \lambda_2(c) \exp(Z'\gamma). \quad (4.2)$$

where $\lambda_1(t)$ and $\lambda_2(c)$ denote the unknown baseline hazard functions for T and C , respectively. β and γ are p -dimensional vectors of regression parameters that denote the effect of covariates on T and C , respectively. Let F_T and F_C denote the marginal distributions of T and C , respectively, and F their joint distribution. Then it is well-known that there exists a copula function $C_\alpha(u, v)$ defined on $I^2 = [0, 1] \times [0, 1]$ such that $F(t, c) = C_\alpha(F_T(t), F_C(c))$ (Nelsen, 2006). Here α is often referred to as the association parameter representing the relationship between T and C , with $C_\alpha(u, 0) = C_\alpha(0, v) = 0$, $C_\alpha(u, 1) = u$ and $C_\alpha(1, v) = v$. It follows that

$$P(T \leq t|C = c, Z) = \frac{\partial C_\alpha(u, v)}{\partial v} \Big|_{u=F_T(t), v=F_C(c)} := m_\alpha(F_T(t), F_C(c)). \quad (4.3)$$

Note that the models and problems similar to these described above have been discussed for current status data in Wang et al., 2012 and Titman, 2013 among others and also other areas such as the competing risks analysis with dependent right-censored data (Zhang and Klein, 1995; Huang and Zhang, 2008; Chen, 2010). Especially, they showed that without prior or extra information, the association parameter α is not identifiable given the copula function. Among others, one use of the inference procedures developed under these situations is for sensitivity analysis. Following these arguments, we will assume that both the copula function and α are known. More discussion on this is given below.

Define $\tilde{C} = \min\{C, \zeta\}$ with ζ being an administrative censoring due to the end of study independent of T and C . Define $\Delta = I(C \leq \zeta)$ and $\delta = I(T \leq \tilde{C})$ and suppose that the observed data are i.i.d. replications of $X = (\Delta, \delta, \tilde{C}, Z)$, denoted by

$$\{X_i = (\Delta_i, \delta_i, \tilde{C}_i, Z_i), i = 1, \dots, n\}.$$

Then the likelihood function can be written as

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n \left\{ \left\{ m_\alpha(F_T(\tilde{c}_i), F_C(\tilde{c}_i)) \right\}^{\delta_i} \left\{ 1 - m_\alpha(F_T(\tilde{c}_i), F_C(\tilde{c}_i)) \right\}^{1-\delta_i} f_C(\tilde{c}_i) \right\}^{\Delta_i} \\ &\quad \times \left\{ \left\{ F_T(\tilde{c}_i) - C_\alpha(F_T(\tilde{c}_i), F_C(\tilde{c}_i)) \right\}^{\delta_i} \left\{ 1 - F_T(\tilde{c}_i) - F_C(\tilde{c}_i) + C_\alpha(F_T(\tilde{c}_i), F_C(\tilde{c}_i)) \right\}^{1-\delta_i} \right\}^{1-\Delta_i} \end{aligned}$$

where $\theta = (\beta, \gamma, \Lambda_T(\cdot), \Lambda_C(\cdot))$ and

$$F_T(t) = 1 - \exp\{-\Lambda_T(t) \exp(Z'\beta)\}$$

$$F_C(c) = 1 - \exp\{-\Lambda_C(c) \exp(Z'\gamma)\}$$

$$f_C(c) = \lambda_2(c) \exp(Z'\gamma) \exp\{-\Lambda_C(c) \exp(Z'\gamma)\}$$

with $\Lambda_T(t) = \int_0^t \lambda_1(s) ds$ and $\Lambda_C(c) = \int_0^c \lambda_2(s) ds$.

4.3 Spline-Based Sieve Semiparametric Maximum Likelihood Estimation

Now we discuss the estimation of all unknown parameters $\theta = (\beta, \gamma, \Lambda_T(\cdot), \Lambda_C(\cdot))$. For this, it is apparent that a natural method is to maximize the log likelihood function

$l(\theta) = \log L(\theta)$. On the other hand, it is easy to see that this is difficult due to the dimension of $\Lambda_T(\cdot)$ and $\Lambda_C(\cdot)$. To address this, following Huang and Rossini (1997) and others, we consider the sieve maximum likelihood approach by approximating $\Lambda_T(\cdot)$ and $\Lambda_C(\cdot)$ with I -spline functions.

More specifically, let M denote a positive constant and $\{I_j(t)\}_{j=1}^{m+k_n}$ the I -spline base functions with order m and number of interior knots k_n , where $k_n = o(n^\nu)$ with $0 < \nu < 0.5$. The selection of m and k_n will be discussed below. Let $\varphi = (\Lambda_T, \Lambda_C)$ and define

$$\Theta_n = \left\{ \theta_n = (\beta, \gamma, \varphi_n) : \varphi_n = (\Lambda_{Tn}, \Lambda_{Cn}) \right\} = \mathcal{B} \otimes \mathcal{M}_n^1 \otimes \mathcal{M}_n^2,$$

where $\mathcal{B} = \{(\beta, \gamma \in R^{2p}, \|\beta\| + \|\gamma\| \leq M\}$, $\mathcal{M}_n^1 = \{\Lambda_{Tn} : \Lambda_{Tn}(t) = \sum_{j=1}^{m+k_n} \xi_j I_j(t), \xi_j \geq 0, j = 1, \dots, m+k_n, t \in [0, u_c]\}$ and $\mathcal{M}_n^2 = \{\Lambda_{Cn} : \Lambda_{Cn}(t) = \sum_{j=1}^{m+k_n} \eta_j I_j(t), \eta_j \geq 0, j = 1, \dots, m+k_n, t \in [0, u_c]\}$ with u_c being the upper bounds of the finite observation times $\{\tilde{C}_i : i = 1, 2, \dots, n\}$. Note that we propose to use the same I -spline bases for Λ_{Tn} and Λ_{Cn} . This is because we only observe \tilde{C}_i 's in the data, it's reasonable to use the same knots for the I -spline bases.

Also denote by $\|a\|$ the Euclidean norm of a vector a , $\|f\|_\infty = \sup_t |f(t)|$ the supremum norm of a function f , and $\|f(X)\|_2 = (\int f^2 dP)^{1/2}$ the $L^2(P)$ norm of a function f with X being distributed according to the probability measure P . For $\theta^i = (\beta^i, \gamma^i, \Lambda_T^i, \Lambda_C^i)$, $i = 1, 2$, define the distance $d^2(\theta^1, \theta^2) = \|\beta^1 - \beta^2\|_2^2 + \|\gamma^1 - \gamma^2\|_2^2 + \|\Lambda_T^1 - \Lambda_T^2\|_2^2 + \|\Lambda_C^1 - \Lambda_C^2\|_2^2$. Let $\theta_0 = (\beta_0, \gamma_0, \Lambda_{T0}, \Lambda_{C0})$ denote the true values of θ . Then for any $\theta \in \Theta$, there exists $\theta_n = (\beta, \gamma, \Lambda_{Tn}, \Lambda_{Cn}) \in \Theta_n$ such that $d(\theta, \theta_n) \leq c(\|\Lambda_{T0} - \Lambda_{Tn}\|_\infty + \|\Lambda_{C0} - \Lambda_{Cn}\|_\infty) = O(n^{-r\nu})$ for some constant c with r defined in condition(A4) in the appendix. It thus follows from Lemma A1 of Lu et al. (2007)

that Θ_n can be used as a sieve space of Θ .

For estimation of $\theta = (\beta, \gamma, \Lambda_T(\cdot), \Lambda_C(\cdot))$, we define the estimate $\hat{\theta} = (\hat{\beta}, \hat{\gamma}, \hat{\Lambda}_{Tn}(\cdot), \hat{\Lambda}_{Cn}(\cdot))$ as the value of θ that maximizes the log likelihood function $l(\theta)$ over Θ_n . Let $\theta_0 = (\beta_0, \gamma_0, \Lambda_{T0}, \Lambda_{C0})$ denote the true values of θ . The following three theorems give the asymptotic properties of $\hat{\theta}$.

Theorem 4.1. *Assume that the conditions A1 - A4 given in the Appendix hold. Then $\hat{\beta}$ and $\hat{\gamma}$ are strongly consistent, and $\|\hat{\Lambda}_{Tn} - \Lambda_{T0}\|_2 \rightarrow 0$ and $\|\hat{\Lambda}_{Cn} - \Lambda_{C0}\|_2 \rightarrow 0$ almost surely as $n \rightarrow \infty$.*

Theorem 4.2. *Again assume that the conditions A1 - A4 given in the Appendix hold. Then as $n \rightarrow \infty$, we have*

$$\|\hat{\Lambda}_{Tn} - \Lambda_{T0}\|_2 + \|\hat{\Lambda}_{Cn} - \Lambda_{C0}\|_2 = O_p(n^{-(1-\nu)/2} + n^{-r\nu}).$$

Theorem 4.3. *Suppose that the conditions A1 - A5 described in the Appendix hold and $r > 2$ in the condition A4. Then as $n \rightarrow \infty$, we have*

$$n^{1/2} \left\{ (\hat{\beta} - \beta_0)', (\hat{\gamma} - \gamma_0)' \right\}' \rightarrow N(0, \Sigma)$$

in distribution, and furthermore, $(\hat{\beta}', \hat{\gamma}')$ is semiparametrically efficient, where Σ is defined in the appendix.

To implement the estimation procedure above, one needs to choose m and k_n . The choice for m is associated with the smoothness of the true baseline cumulative hazard functions. Usually, quadratic or cubic spline functions should work sufficiently well.

A simple approach for selection is clearly to try different values of them and compare the obtained results. As an alternative, one can also apply the Akaike information criterion (AIC) to choose m and k_n that give the smallest AIC. Note that in the method developed above, we have to assume that both copula function $C_\alpha(\cdot, \cdot)$ and the association parameter α are known without prior or extra information. In practice, we actually could consider different possible choices for them too and apply the AIC for their selection together with m and k_n . More comments on this are given in Section 4.5.

To use the estimation procedure, it is apparent that one also needs to estimate the covariance matrix of $\hat{\beta}$ and $\hat{\gamma}$. For this, a natural and simple method is to employ the inverse of the observed information matrix by treating Λ_{T_n} and Λ_{C_n} as finite-dimensional nuisance parameters. Of course, this could be computationally intensive. On the other hand, if the functions Λ_T and Λ_C is smooth enough, a small value of knots is generally sufficient and the simulation results indicate that it seems to work reasonably well.

4.4 Simulation Studies

To evaluate the estimation procedure proposed above, a simulation study is performed. In the study, we assume that the covariate Z is a Bernoulli variable with the success probability of 0.5. For the copula function, four choices are investigated,

$$C_\alpha(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}, \quad \alpha > 0$$

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

$$C_\alpha(u, v) = \log_\alpha \left\{ 1 + \frac{(\alpha^u - 1)(\alpha^v - 1)}{\alpha - 1} \right\}, \quad \alpha \geq 0, \alpha \neq 1$$

and

$$C_\alpha(u, v) = uv + \alpha uv(1 - u)(1 - v), \quad -1 \leq \alpha \leq 1$$

The four models are usually referred to as Clayton, Gumbel, FGM and Frank models, respectively. Note that for different copula models, the range of the association parameter α is different. So in the following, we use the Kendall's τ to measure the association between T and C . For the Clayton copula, $\tau = \alpha/(\alpha + 2)$, for the Gumbel copula, we have $\tau = 1 - 1/\alpha$. The FGM copula gives $\tau = 2\alpha/9$, and the Frank copula has that $\tau = 1 + 4\rho^{-1}\{D_1(\rho) - 1\}$, where $\rho = -\log \alpha$ and $D_1(\rho) = \rho^{-1} \int_0^\rho t(e^t - 1)^{-1} dt$.

To generate the observed data, we first generated the T_i 's under the model (4.1) with $\Lambda_{T_0}(t) = t$ and then the C_i 's from its conditional distribution given the T_i 's. More specifically, with the random number a generated from the uniform distribution $U(0, 1)$, we solve the equation

$$P(C \leq c_i | T = t_i, Z) = \frac{\partial C_\alpha(u, v)}{\partial u} \Big|_{u=F_T(t_i), v=F_C(c_i)} = a$$

for c_i . Furthermore, given T_i , C_i and ζ , we defined $\tilde{C}_i = \min\{C_i, \zeta\}$, $\Delta_i = I(C_i \leq \zeta)$ and $\delta_i = I(T_i \leq \tilde{C}_i)$. The results given below are based on 500 replications with the sample size $n = 200$ or 400.

Table 4.1 presents the results on estimation of regression parameters β and γ based on the simulated data generated under the Clayton model with $\tau = 0.2, 0.5$, $\gamma = 0.5, 1$, $\beta_0 = -0.5, 0, 0.5$ for $\tau = 0.2$ and $\beta = 0.5, 1$ for $\tau = 0.5$. Here for the approximation of Λ_{T_0} and Λ_{C_0} , the quadratic splines were used with the 0.25, 0.5, 0.75 quantiles of the \tilde{C}_i 's chosen as the 3 interior knots. ζ was chosen to give about 16%

right censoring rates for C_i 's. The results include the estimated bias (Bias) given by the average of the estimates minus the true value, the sample standard deviation (SSE) of the estimates, the average of the estimated standard errors (SEE), and the 95% empirical coverage probability (CP). The results obtained under the Gumbel model are given in Table 4.2 with the same set-ups as above and the results obtained under the FGM model are given in Table 4.3 with $\tau = -0.1, 0.1$, $\beta = -0.5, 0, 0.5$. And similarly, the results under the Frank model are given in Table 4.4 with $\tau = -0.5, -0.2, 0.2, 0.5$, $\beta = -1, -0.5$ for $\tau = -0.5$, $\beta = -0.5, 0, 0.5$ for $\tau = \pm 0.2$ and $\beta = 0.5, 1$ for $\tau = 0.5$. These results indicate that the proposed estimate seems to be unbiased and the variance estimation method also seems to work well. In addition, the asymptotic distribution seems to perform well too in the situations considered here and as expected, the results become better as the sample size increases.

4.5 An Application

In this section, we apply the proposed methodology to a set of current status data arising from a 2-year tuorigenicity study conducted by National Toxicology Program which has been studied by Wang et. al. (2008). In the study, the groups of 50 male and female F344/N rats and B6C3F₁ mice were exposed to chloroprene at concentrations of 0, 12.8, 32 or 80 ppm by inhalation, 6h per day, 5 days per week, for 2 years. Each animal was examined for various tumors at its death time. Some animals died naturally during the study, and those who survived at the end of study were sacrificed for examinations. Since the tumor status was only examined at death time, only current status data are available for the tumor onset time, the variable of inter-

est. Among others, one objective of the study is to compare the tumor growth rates between the different dose groups and investigate the dose effect. In the following, we will focus on a specific type of lung tumor, the Alveolar/Bronchiolar Adenoma(A/B A), for male and female B6C3F₁ mice in the control (0 ppm) and high dose (80 ppm) groups.

For the analysis, let T denote the onset time of A/B A tumor and C the death time in months up to 25 months, the end of the study. Define $Z_i = 1$ if the mice is in high dose group and $Z_i = 0$ otherwise. In both groups, there were many mice that survived at the end of the study and were sacrificed for tumour examinations. Thus, we have right censored data available for the death time C . To apply the estimation procedure proposed in the previous section, we considered Frank and Gumbel copulas and $k_n = 3, 4, 5, 6$ based on the sample size here. Given the copula function, we considered several values of τ and applied the AIC to choose both k_n and τ together. Under the Frank copula, we found that the smallest AIC is given by $k_n = 5$ and $\tau = 0.25$, while under the Gumbel model, the choice is $k_n = 5$ and $\tau = 0.05$.

Table 4.5 gives the results obtained on the estimation of the dose effect on the tumor onset time with $k_n = 5$ and several values of τ under both models. One can easily see that the results are quite different and depends on the possible correlation between the tumor onset time and the observation time. All the p -values in this table, for both β and γ are significant with any reasonable level of significance. That is, mice in the high dose group have significantly shorter tumor onset time. This is consistent with the results obtained by other authors. For comparison, we also present the results obtained with $k_n = 3$ in Table 4.6 and they give similar conclusions.

4.6 Concluding Remarks

In this chapter, we discussed regression analysis of current status data using the proportional hazards model when observation time may be related to the survival time of interest. For inference, the sieve maximum likelihood estimation approach has been developed with the use of I -splines and the copula function. The asymptotic properties of the resulting estimates were established. Note that one advantage of I -spline functions over B-spline functions is that they do not need the order restriction on spline coefficients as required by the latter. The simulation study suggested that the approach works well in practical situations.

Note that in the developed methodology, it has been assumed that both the observation time and the survival time of interest follow the proportional hazards models. It is straightforward to generalize the proposed method to other models such as the linear transformation model (Zhang et al., 2005).

As pointed out above, among others, one use of the proposed approach is for sensitivity analysis and it is usually difficult or impossible to estimate the underlying copula model and the association parameter without other information. On the other hand, if prior or external information is available on the association between the observation time and the survival time, the approach can incorporate such information to yield valid analysis.

n	τ	β_0	Bias	SSE	SEE	CP	γ_0	Bias	SSE	SEE	CP		
200	0.2	-0.5	-0.013	0.224	0.233	0.968	0.5	0.002	0.167	0.156	0.948		
		0	0.013	0.228	0.227	0.959	0.5	0.003	0.167	0.156	0.944		
		0.5	0.044	0.255	0.239	0.946	0.5	0.005	0.167	0.156	0.946		
	0.5	0.5	0.009	0.213	0.225	0.940	0.5	0.009	0.149	0.156	0.958		
		1	0.032	0.263	0.250	0.954	0.5	-0.005	0.149	0.156	0.958		
		0.2	-0.5	-0.021	0.254	0.252	0.956	1	0.008	0.175	0.162	0.928	
	0.2	0	0.013	0.249	0.243	0.950	1	0.008	0.176	0.162	0.938		
		0.5	0.049	0.261	0.249	0.948	1	0.010	0.175	0.162	0.934		
		0.5	0.5	-0.004	0.224	0.256	0.953	1	0.012	0.173	0.163	0.938	
	0.5	1	0.020	0.236	0.249	0.936	1	0.016	0.175	0.163	0.942		
		400	0.2	-0.5	-0.020	0.148	0.160	0.966	0.5	-0.004	0.107	0.110	0.958
				0	-0.010	0.147	0.156	0.956	0.5	-0.004	0.108	0.110	0.956
0.5	0.010			0.164	0.164	0.962	0.5	-0.002	0.108	0.110	0.960		
0.5	0.5	0.010	0.149	0.157	0.950	0.5	0.002	0.108	0.110	0.956			
	1	0.004	0.174	0.168	0.956	0.5	0.003	0.114	0.110	0.954			
	0.2	-0.5	-0.032	0.183	0.175	0.944	1	0.006	0.125	0.114	0.924		
0		-0.016	0.164	0.168	0.958	1	-0.002	0.109	0.114	0.954			
0.5		0.006	0.170	0.172	0.956	1	0.001	0.110	0.114	0.952			
0.5	0.5	-0.030	0.177	0.178	0.956	1	-0.004	0.128	0.115	0.936			
	1	-0.005	0.179	0.159	0.962	1	-0.004	0.123	0.114	0.944			

Table 4.1: Simulation results for Clayton copula.

n	τ	β_0	Bias	SSE	SEE	CP	γ_0	Bias	SSE	SEE	CP		
200	0.2	-0.5	-0.057	0.221	0.239	0.970	0.5	0.008	0.167	0.156	0.936		
		0	-0.023	0.222	0.226	0.966	0.5	0.008	0.161	0.156	0.952		
		0.5	0.040	0.243	0.238	0.944	0.5	0.007	0.147	0.156	0.952		
	0.5	0.5	0.036	0.223	0.236	0.958	0.5	0.012	0.150	0.156	0.964		
		1	0.096	0.308	0.264	0.921	0.5	0.013	0.150	0.156	0.966		
		0.2	-0.5	-0.065	0.254	0.249	0.940	1	0.027	0.155	0.163	0.956	
	400	0.2	0	-0.063	0.230	0.233	0.952	1	0.028	0.154	0.163	0.958	
			0.5	0.007	0.234	0.241	0.954	1	0.023	0.154	0.163	0.952	
			0.5	0.5	-0.036	0.207	0.229	0.956	1	0.016	0.156	0.162	0.960
		0.5	1	0.066	0.252	0.248	0.941	1	0.019	0.155	0.163	0.954	
			0.2	-0.5	-0.055	0.156	0.165	0.954	0.5	0.010	0.116	0.110	0.940
			0	-0.031	0.149	0.156	0.964	0.5	0.007	0.116	0.110	0.940	
400	0.5	0.5	0.019	0.170	0.165	0.944	0.5	0.005	0.116	0.110	0.940		
		0.5	0.012	0.149	0.163	0.964	0.5	0.001	0.114	0.110	0.950		
		1	0.044	0.194	0.181	0.942	0.5	0.005	0.107	0.110	0.952		
	0.2	-0.5	-0.062	0.169	0.172	0.954	1	0.009	0.109	0.114	0.956		
		0	-0.063	0.151	0.161	0.944	1	0.011	0.109	0.114	0.958		
		0.5	-0.022	0.163	0.166	0.956	1	0.005	0.109	0.114	0.960		
0.5	0.5	-0.054	0.140	0.156	0.956	1	-0.002	0.113	0.114	0.952			
	1	0.032	0.165	0.171	0.945	1	0.005	0.113	0.114	0.952			

Table 4.2: Simulation results for Gumbel copula.

n	τ	β_0	Bias	SSE	SEE	CP	γ_0	Bias	SSE	SEE	CP
200	-0.1	-0.5	-0.032	0.264	0.242	0.934	0.5	0.006	0.162	0.156	0.940
		0	-0.001	0.268	0.243	0.940	0.5	0.006	0.162	0.156	0.938
		0.5	0.029	0.278	0.260	0.932	0.5	0.006	0.162	0.156	0.938
	0.1	-0.5	-0.034	0.247	0.234	0.942	0.5	0.008	0.161	0.156	0.938
		0	-0.013	0.250	0.230	0.934	0.5	0.009	0.161	0.156	0.940
		0.5	0.010	0.248	0.239	0.954	0.5	0.009	0.161	0.156	0.940
	-0.1	-0.5	-0.030	0.288	0.269	0.942	1	0.014	0.168	0.162	0.934
		0	0.008	0.273	0.262	0.948	1	0.014	0.168	0.162	0.936
		0.5	0.040	0.302	0.277	0.940	1	0.014	0.168	0.162	0.940
	0.1	-0.5	-0.041	0.272	0.253	0.934	1	0.014	0.165	0.162	0.942
		0	-0.004	0.264	0.245	0.942	1	0.014	0.164	0.162	0.940
		0.5	0.023	0.279	0.252	0.932	1	0.015	0.164	0.162	0.938
400	-0.1	-0.5	-0.017	0.171	0.165	0.946	0.5	-0.007	0.112	0.110	0.948
		0	0.003	0.172	0.166	0.936	0.5	-0.007	0.112	0.110	0.948
		0.5	0.025	0.176	0.179	0.958	0.5	-0.007	0.112	0.110	0.946
	0.1	-0.5	-0.018	0.160	0.160	0.963	0.5	-0.005	0.114	0.110	0.940
		0	-0.007	0.159	0.157	0.946	0.5	-0.005	0.114	0.110	0.940
		0.5	0.014	0.172	0.165	0.948	0.5	-0.005	0.114	0.110	0.940
	-0.1	-0.5	-0.015	0.185	0.184	0.946	1	0.001	0.118	0.114	0.946
		0	0.001	0.179	0.180	0.950	1	0.001	0.118	0.114	0.944
		0.5	0.023	0.191	0.189	0.936	1	0.001	0.118	0.114	0.942
	0.1	-0.5	-0.027	0.176	0.175	0.948	1	-0.004	0.119	0.114	0.940
		0	-0.011	0.172	0.169	0.942	1	-0.003	0.119	0.114	0.942
		0.5	0.015	0.180	0.174	0.938	1	-0.002	0.119	0.114	0.938

Table 4.3: Simulation results for FGM copula.

n	τ	β_0	Bias	SSE	SEE	CP	γ_0	Bias	SSE	SEE	CP
200	-0.5	-1	-0.046	0.306	0.307	0.939	0.5	0.007	0.155	0.156	0.956
		-0.5	-0.016	0.281	0.299	0.956	0.5	0.006	0.156	0.156	0.952
	-0.2	-0.5	-0.032	0.272	0.248	0.938	0.5	0.001	0.164	0.156	0.936
		0	0.005	0.277	0.255	0.931	0.5	-0.001	0.164	0.156	0.938
	0.2	0.5	0.049	0.286	0.281	0.947	0.5	-0.001	0.164	0.156	0.934
		-0.5	-0.050	0.234	0.232	0.950	0.5	0.002	0.158	0.156	0.944
	0.5	0	-0.029	0.214	0.225	0.960	0.5	-0.007	0.166	0.156	0.926
		0.5	0.010	0.232	0.233	0.956	0.5	-0.005	0.166	0.156	0.926
	-0.5	0.5	0.006	0.232	0.231	0.956	0.5	-0.006	0.161	0.156	0.942
		1	0.040	0.279	0.256	0.933	0.5	-0.005	0.162	0.156	0.938
	-0.2	-1	-0.022	0.304	0.358	0.948	1	0.006	0.161	0.162	0.942
		-0.5	0.030	0.303	0.358	0.950	1	0.013	0.160	0.162	0.956
	0.2	-0.5	-0.023	0.294	0.278	0.948	1	0.021	0.160	0.163	0.950
		0	0.014	0.297	0.275	0.939	1	0.007	0.167	0.162	0.930
	0.5	0.5	0.057	0.326	0.298	0.943	1	0.006	0.167	0.162	0.932
		-0.5	-0.062	0.254	0.247	0.952	1	-0.001	0.172	0.162	0.934
	-0.2	0	-0.036	0.236	0.238	0.942	1	0.001	0.172	0.162	0.934
		0.5	-0.001	0.245	0.243	0.958	1	0.002	0.172	0.162	0.930
	0.5	0.5	-0.049	0.207	0.222	0.942	1	-0.008	0.169	0.162	0.942
		1	0.005	0.231	0.233	0.955	1	-0.005	0.169	0.162	0.942
400	-0.5	-1	-0.026	0.206	0.209	0.943	0.5	-0.002	0.112	0.110	0.942
		-0.5	-0.020	0.202	0.208	0.936	0.5	0.012	0.115	0.110	0.940
	-0.2	-0.5	-0.018	0.176	0.171	0.944	0.5	0.008	0.112	0.110	0.944
		0	-0.002	0.175	0.172	0.944	0.5	0.007	0.112	0.110	0.942
	0.2	0.5	0.024	0.190	0.187	0.956	0.5	0.007	0.112	0.110	0.940
		-0.5	-0.034	0.144	0.160	0.962	0.5	-0.001	0.112	0.110	0.956
	0.5	0	-0.017	0.151	0.155	0.952	0.5	-0.001	0.112	0.110	0.954
		0.5	0.009	0.163	0.162	0.954	0.5	0.001	0.112	0.110	0.958
	-0.5	0.5	-0.010	0.151	0.157	0.956	0.5	-0.005	0.111	0.110	0.960
		1	0.042	0.193	0.174	0.924	0.5	-0.003	0.112	0.110	0.952
	-0.2	-1	-0.018	0.226	0.249	0.945	1	0.012	0.120	0.115	0.940
		-0.5	0.001	0.211	0.246	0.948	1	0.011	0.119	0.115	0.942
	0.2	-0.5	-0.011	0.202	0.193	0.954	1	0.012	0.119	0.115	0.942
		0	0.003	0.192	0.189	0.956	1	0.012	0.119	0.115	0.940
	0.5	0.5	0.018	0.201	0.200	0.946	1	0.011	0.119	0.115	0.942
		-0.5	-0.049	0.169	0.172	0.952	1	0.003	0.113	0.114	0.954
	-0.2	0	-0.032	0.157	0.165	0.948	1	0.003	0.113	0.114	0.956
		0.5	0.002	0.171	0.169	0.944	1	0.005	0.113	0.114	0.958
	0.5	0.5	-0.031	0.152	0.156	0.947	1	-0.011	0.115	0.114	0.944
		1	0.007	0.163	0.163	0.940	1	-0.003	0.115	0.114	0.952

Table 4.4: Simulation results for Frank copula.

copula	τ	$\hat{\beta}$	SEE	$p(\times 10^9)$	$\hat{\gamma}$	SEE	$p(\times 10^9)$	AIC
Frank	-0.5	1.9484	0.3264	2.3788	1.2004	0.2020	2.7882	1099.771
	-0.25	1.9998	0.3375	3.1172	1.3328	0.2002	0.0278	1066.116
	0	2.2785	0.3939	7.2524	1.3750	0.2003	0.0067	1057.002
	0.25	2.4722	0.3632	0.0100	1.3893	0.2005	0.0042	1055.845
	0.5	2.3357	0.3476	0.0182	1.3926	0.2003	0.0036	1057.551
Gumbel	0	2.2785	0.3939	7.2524	1.3750	0.2003	0.0067	1057.002
	0.05	2.3798	0.3798	0.3710	1.3762	0.2004	0.0065	1056.944
	0.25	2.6423	0.3724	0.0013	1.3823	0.2006	0.0056	1057.299
	0.5	2.5029	0.3584	0.0029	1.3911	0.2010	0.0045	1058.163

Table 4.5: Estimating results for $k_n = 5$ (p denotes the p -value).

copula	τ	$\hat{\beta}$	SEE	$p(\times 10^9)$	$\hat{\gamma}$	SEE	$p(\times 10^9)$	AIC
Frank	-0.5	1.9081	0.3211	2.7997	1.2773	0.2050	0.4616	1112.888
	-0.25	1.9945	0.3347	2.5510	1.3796	0.2011	0.0069	1084.576
	0	2.2795	0.3945	7.5175	1.4112	0.2004	0.0019	1077.819
	0.15	2.4494	0.3667	0.0239	1.4189	0.2004	0.0015	1076.529
	0.25	2.4837	0.3541	0.0023	1.4180	0.2005	0.0015	1076.769
	0.5	2.3625	0.3258	0.0004	1.4164	0.1996	0.0013	1078.927
Gumbel	0	2.2795	0.3945	7.5175	1.4112	0.2004	0.0019	1077.819
	0.15	2.5697	0.3724	0.0052	1.4125	0.2004	0.0018	1077.713
	0.25	2.6667	0.3692	0.0005	1.4141	0.2006	0.0018	1077.768
	0.5	2.5341	0.3440	0.0002	1.4240	0.2008	0.0013	1078.619

Table 4.6: Estimating results for $k_n = 3$ (p denotes the p -value).

Chapter 5

Sieve Maximum Likelihood Regression Analysis of Dependent Interval-Censored Data

5.1 Introduction

The development of regression analysis of interval-censored data has been very active in the last decade. And most of the work has been focusing on semiparametric models, notably the Cox regression model (Finkelstein 1986, Huang 1996, Goggins *et al.* 1998, Goggins and Finkelstein 2000, Seaman and Bird 2001, Zhang *et al.* 2005, Zhang *et al.* 2010). Most of these existing methods for regression analysis of interval-censored data require 'non-informative' censoring mechanism. However, as we discussed earlier in Chapter 1, there exist situations under which such assumption may not hold.

The literature on the topic of informative interval censoring is very limited due to the complexity of the dependence structure between the failure time and censoring.

van der Lann and Robins (1998) proposed a locally efficient estimation procedure by introducing an inverse probability of censoring weighted estimator of the distribution function of the single monitoring time. And their approach requires conditional independence between failure time and monitoring time given covariates. Zhang *et al.* (2005) also discussed the current status data when the dependence structure between the failure time and the monitoring time can be modeled through an unobservable random effect. Finkelstein *et al.* (2002) and Betensky and Finkelstein (2002) investigated one-sample problem of general interval-censored data with dependent interval censoring. Their approaches assume that the complete visit compliance process is known. Zhang *et al.* (2007) proposed to model the dependence between the censoring variables and the failure time through some latent variables. Their method requires discretization of the baseline function and asymptotic properties of parameter estimates are not provided. Given the fact that sometimes patients may tend to visit more frequently when they get sicker, we propose a method where the failure time depend on the observation interval only through its length. In stead of the latent variables, we employ the copula model approach to model this dependence jointly, which applies to much more general situations.

The rest of this chapter is organized as follows. We will begin in Section 5.2 with describing the notation, assumptions and models that will be used throughout the paper as well as the likelihood function. As mentioned above, we will jointly model the failure time and the observation interval length together and consider the marginal proportional hazards model for covariate effects. Section 5.3 presents the sieve semiparametric maximum likelihood estimation approach for estimation of unknown parameters and for this, we will employ I -splines (Lu, *et al.* 2007, Ramsay

1988) to approximate the baseline cumulative hazard function. The resulting estimators are shown to be consistent and furthermore, we prove that the estimated regression parameters are asymptotically efficient and normally distributed. In Section 5.4, we present some results obtained from an extensive simulation study, and they suggest that the approach works well for the situations considered. Section 5.5 applies the method to a motivating example from a breast cosmesis study and Section 5.6 contains some discussion and concluding remarks.

5.2 Notation, Models and Likelihood Function

Consider a failure time study that involves n independent subjects and in which each subject is observed periodically. For subject i , let T_i denote the failure time of interest, (L_i, R_i) be the pair of examination times bracketing the event time T_i , with $P(L_i < T_i \leq R_i) = 1$, and Z_i is a p -dimensional vector of covariates, $i = 1, \dots, n$. Note that L_i is the last examination time before and R_i is the first examination time after the event time T_i . In the following, we suppose that T_i may be related to (L_i, R_i) even given covariates and in this case, there usually exists a censoring time ζ_i such as the administrative stop time of observations. For example, in a cancer study with periodic follow-ups, T_i, L_i, R_i and ζ_i represent the time to cancer, last examination time before cancer, first examination time after cancer and study stop time, respectively. Also, we assume that T_i is related to the observation interval $(L_i, R_i]$ only through its length. That is, $P(T \leq t|L, R, Z) = P(T \leq t|W, Z)$ with $W = R - L$. This assumption is motivated by the fact that in clinical trials comparing different treatments in terms of efficacy or toxicity, patients who receive a less effective

or more toxic treatment may visit clinical centers for examination more frequently than other patients.

Define $\delta_i = I(R_i < \infty)$, then the observed data have the form

$$X_i = (L_i, R_i, \delta_i, Z_i), \quad i = 1, \dots, n.$$

To describe the effects of covariates, in the following, we assume that given Z_i , T_i and W_i follow the marginal proportional hazards models given by

$$\lambda_T(t|Z) = \lambda_1(t) \exp(Z'\beta) \tag{5.1}$$

and

$$\lambda_W(w|Z) = \lambda_2(w) \exp(Z'\gamma), \tag{5.2}$$

respectively. Here $\lambda_1(t)$ and $\lambda_2(w)$ denote unknown baseline hazard functions and β and γ are p -dimensional vectors of regression parameters. Let F_T and F_W denote the distribution of T, W given Z , respectively. Then under the models above, we have

$$F_T(t) = 1 - \exp(-\Lambda_1(t) \exp(Z'\beta))$$

and

$$F_W(w) = 1 - \exp(-\Lambda_2(w) \exp(Z'\gamma)),$$

respectively, where $\Lambda_1(t) = \int_0^t \lambda_1(s) ds$ and $\Lambda_2(w) = \int_0^w \lambda_2(s) ds$,

Let K denote the joint distribution of T_i and W_i . Then it is well known that there

exists a copula function $C_\alpha(u, v)$, defined on $I^2 = [0, 1] \times [0, 1]$ such that

$$K(t, w) = C_\alpha(F_T(t), F_W(w)), t \geq 0, w \geq 0 \quad (5.3)$$

(Nelsen, 2006). Here α is often referred to as the association parameter representing the relationship between T_i and W_i , and $C_\alpha(u, 0) = C_\alpha(0, v) = 0$, $C_\alpha(u, 1) = u$ and $C_\alpha(1, v) = v$. It follows that

$$P(T \leq t | W = w, Z) = \frac{\partial C_\alpha(u, v)}{\partial v} \Big|_{u=F_T(t), v=F_W(w)} := m_\alpha(F_T(t), F_W(w)). \quad (5.4)$$

and the resulting likelihood function has the form

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n \left\{ [m_\alpha(F_T(R_i), F_W(w_i)) - m_\alpha(F_T(L_i), F_W(w_i))] f_W(w_i) \right\}^{\delta_i} \\ &\times \left\{ 1 - F_W(\zeta_i - L_i) - F_T(L_i) + C_\alpha(F_T(L_i), F_W(\zeta_i - L_i)) \right\}^{1-\delta_i}. \end{aligned} \quad (5.5)$$

In the above, $\theta = (\beta, \gamma, \Lambda_1(\cdot), \Lambda_2(\cdot))$ and

$$f_W(w_i) = \lambda_2(w_i) \exp(Z_i' \gamma) \exp\{-\Lambda_2(w_i) \exp(Z_i' \gamma)\}.$$

Note that the formulation (5.3) or the copula model approach in general is commonly used in multivariate failure time data analysis (Hougaard, 2000). It has also been applied to the analysis of univariate failure time data in the presence of informative censoring. For example, by using the idea, Zheng and Klein (1995), Chen (2010) investigated the estimation of a marginal survival function based on right-censored data in the presence of dependent competing risks and the regression analysis of the

data with the same structure, respectively. Wang et. al. (2012) and Titman (2013) studied the nonparametric estimation of a survival function based on dependent current status data. In particular, they showed that without prior or extra information, the association parameter α is not identifiable given the copula function as pointed out before, and one application of the methods developed under the framework is for sensitivity analysis. In the following, following these authors, we will assume that both the copula function and α are known and our main goal is to estimate regression parameter β as well as γ .

5.3 Sieve Semiparametric Maximum Likelihood Estimation

Now we discuss the estimation of the regression parameters or θ . For this, it is apparent that a natural approach is to maximize the log likelihood function $l(\theta) = \log L(\theta)$. On the other hand, it is easy to see that this is difficult due to the dimension of Λ_1 and Λ_2 . To address this, following Huang and Rossini (1997) and others, we consider the sieve maximum likelihood approach by approximating $\Lambda_k (k = 1, 2)$ with I -spline functions.

More specifically, Let M denote a positive constant and $\{I_j(t)\}_{j=1}^{m+k_n}$ denote the I -spline base functions with order m and k_n interior knots, where $k_n = o(n^\nu)$ with $0 < \nu < 0.5$. The selection of m and k_n will be discussed below. Let $\varphi = (\Lambda_1, \Lambda_2)$ and define

$$\Theta_n = \left\{ \theta_n = (\beta, \gamma, \psi_n) : \varphi_n = (\Lambda_{1n}, \Lambda_{2n}) \right\} = \mathcal{B} \otimes \mathcal{M}_n^1 \otimes \mathcal{M}_n^2$$

where $\mathcal{B} = \{(\beta, \gamma \in R^{2p}, \|\beta\| + \|\gamma\| \leq M\}$, $\mathcal{M}_n^1 = \{\Lambda_{1n} : \Lambda_{1n}(t) = \sum_{j=1}^{m+k_n} \xi_j I_{j1}(t), \xi_j \geq 0, j = 1, \dots, m + k_n, t \in [l_t, u_t]\}$, $\mathcal{M}_n^2 = \{\Lambda_{2n} : \Lambda_{2n}(w) = \sum_{j=1}^{m+k_n} \eta_j I_{j2}(w), \eta_j \geq 0, j = 1, \dots, m + k_n, w \in [l_w, u_w]\}$, where $[l_t, u_t]$ and $[l_w, u_w]$ are the range of $\{L_i, R_i, i = 1, 2, \dots, n\}$ and the range of $\{\min(W_i, \zeta_i - L_i), i = 1, 2, \dots, n\}$, respectively.

For a vector a and a function f , let $\|a\|$ denote the Euclidean norm and $\|f\|_\infty = \sup_t |f(t)|$, the supremum norm. Also let $\|f(X)\|_2 = (\int f^2 dP)^{1/2}$, the $L^2(P)$ norm, with X being distributed according to the probability measure P . For $\theta^i = (\beta^i, \gamma^i, \Lambda_1^i, \Lambda_2^i)$, $i = 1, 2$, define the distance

$$d^2(\theta^1, \theta^2) = \|\beta^1 - \beta^2\|_2^2 + \|\gamma^1 - \gamma^2\|_2^2 + \|\Lambda_1^1 - \Lambda_1^2\|_2^2 + \|\Lambda_2^1 - \Lambda_2^2\|_2^2.$$

Then it is easy to see that for any $\theta \in \Theta$, there exists a sequence of $\theta_n = (\beta, \gamma, \Lambda_{1n}, \Lambda_{2n}) \in \Theta_n$ such that

$$d(\theta, \theta_n) \leq c(\|\Lambda_1 - \Lambda_{1n}\|_\infty + \|\Lambda_2 - \Lambda_{2n}\|_\infty) = O(n^{-r\nu})$$

for some constant c and $r = m + \eta$, where η is defined in the condition (A4) given in the Appendix. It thus follows from Lemma A1 of Lu et al. (2007) that Θ_n can be used as a sieve space of Θ .

For estimation of $\theta = (\beta, \gamma, \Lambda_1(\cdot), \Lambda_2(\cdot))$, we define the estimate $\hat{\theta} = (\hat{\beta}, \hat{\gamma}, \hat{\Lambda}_{1n}(\cdot), \hat{\Lambda}_{2n}(\cdot))$ as the value of θ that maximizes the log likelihood function $l(\theta)$ over Θ_n . Let $\theta_0 = (\beta_0, \gamma_0, \Lambda_{10}, \Lambda_{20})$ denote the true values of θ . The following three theorems give the asymptotic properties of $\hat{\theta}$.

Theorem 5.1. *Under Assumptions A1-A4 described in the Appendix, $\hat{\beta}$ and $\hat{\gamma}$ are strong consistent estimators of the true coefficients β_0, γ_0 and $\|\hat{\Lambda}_{1n} - \Lambda_{10}\|_2 \rightarrow 0$,*

$\|\hat{\Lambda}_{2n} - \Lambda_{20}\|_2 \rightarrow 0$, almost surely.

Theorem 5.2. *Under Assumptions A1-A4 described in the Appendix, it holds that*

$$\|\hat{\Lambda}_{1n} - \Lambda_{10}\|_2 + \|\hat{\Lambda}_{2n} - \Lambda_{20}\|_2 = O_p(n^{-(1-\nu)/2} + n^{-r\nu}).$$

Theorem 5.3. *Suppose that $r > 2$ and the conditions A1 - A5 described in the Appendix hold. Then we have*

$$n^{1/2} ((\hat{\beta} - \beta_0)', (\hat{\gamma} - \gamma_0)')' \rightarrow N(0, \Sigma)$$

in distribution and furthermore, $(\hat{\beta}', \hat{\gamma}')$ is semiparametrically efficient, where Σ is given in the appendix.

The proofs for the results above are sketched in the Appendix. Note that to implement the estimation procedure above, one needs to choose m and k_n and then maximize $l(\theta)$. In general, the degree m should be decided by the smoothness of the true baseline cumulative hazard functions and either quadratic or cubic spline functions usually work sufficiently well. Of course, one could try different values of them and compare the obtained results. As an alternative, one can apply the Akaike information criterion (AIC) to choose m and k_n that give the smallest AIC. Given m and k_n , the determination of $\hat{\theta}$ is relatively easy as one can simply employ the existing procedures such as the function `nlm` in R. Note that in the method developed above, it has been assumed that both the copula function $C_\alpha(\cdot, \cdot)$ and the association parameter α are known without prior or extra information. In practice, one could actually apply the AIC too for their selection. More comments on this are given in Section 5.

To implement the estimation procedure proposed above, it is apparent that one needs to estimate the covariance matrix of $\hat{\beta}$ and $\hat{\gamma}$ or Σ . For this, one approach is to apply the profile likelihood approach, which may be difficult. A natural and simple alternative method, which will be used below in numerical studies, is to employ the inverse of the observed information matrix by treating Λ_{1n} and Λ_{2n} as finite-dimensional nuisance parameters. The simulation results below suggest that it works well for practical situations.

5.4 Simulation Studies

An extensive simulation was conducted to assess the performance of the proposed methodology. In the study, we assumed that the covariate Z to be a Bernoulli variable with success probability 0.5 and considered three copula models:

$$C_\alpha(u, v) = \begin{cases} \exp\{-[(-\log u)^\alpha + (-\log v)^\alpha]^{1/\alpha}\} & \text{Gumbel, } \alpha \geq 1 \\ uv + \alpha uv(1-u)(1-v) & \text{FGM, } -1 \leq \alpha \leq 1 \\ \log_\alpha\{1 + (\alpha^u - 1)(\alpha^v - 1)/(\alpha - 1)\} & \text{Frank, } \alpha > 0, \alpha \neq 1 \end{cases}$$

Note that for different copula models as above, the range of the association parameter α is different, thus we use the Kendall's τ to measure the association between T and W in the following. Specifically, for the Gumbel copula, we have $\tau = 1 - 1/\alpha$. The FGM copula gives $\tau = 2\alpha/9$. And the Frank copula has $\tau = 1 + 4\rho^{-1}\{D_1(\rho) - 1\}$, where $\rho = -\log \alpha$ and $D_1(\rho) = \rho^{-1} \int_0^\rho t(e^t - 1)^{-1} dt$.

To generate the observed data, we first generate the failure time T_i 's under the proportional hazards model (5.1) with $\Lambda_{10}(t) = t$ and then W_i 's from their the con-

ditional distribution given T_i 's. More specifically, given T_i and a random number a generated from the uniform distribution $U(0, 1)$, we solve the equation

$$P(W \leq w_i | T = t_i, Z = z_i) = \frac{\partial C(u, v)}{\partial u} \Big|_{u=F_T(t_i), v=F_W(w_i)} = a$$

for w_i , where $\Lambda_{20}(w) = 2w$. The constant censoring times $\zeta_i = \zeta$'s were then chosen to give approximate right censoring rates (ARC) of 20%, or 40%. The situation with no right-censoring subjects (ARC=0%) is also considered in the simulation study.

To generate the observation interval, we set L_i as the largest number in $\{0, W_i, 2W_i, 3W_i, \dots\}$ that is smaller than the minimum of T_i and ζ and let $R_i = L_i + W_i$. The subject with $R_i \geq \zeta$ is considered to be right censored. And W_i for a right-censored subject is only known to be greater than $\zeta - L_i$. The results given below are based on 400 replications with the sample size $n = 200$.

Table 1 presents the results on estimation of regression parameters β and γ based on the simulated data generated under the Gumbel model with $\beta_0 = 0, 0.2$, $\gamma_0 = -0.2, 0, 0.2$ and $\tau = 0.5, 0.7$. Quadratic splines with 4 interior knots are used to estimate $\Lambda_{10}(t)$ and $\Lambda_{20}(w)$. For estimating $\Lambda_{10}(t)$, we consider 0.2, 0.4, 0.6, 0.8 quantiles of the pooled set of all L_i 's and the uncensored R_i 's, as the 4 interior knots. Correspondingly, we consider the 0.2, 0.4, 0.6, 0.8 quantiles of the pooled set of W_i 's of non-right-censored subjects and $(\zeta - L_i)$'s of right censored subjects, as the interior knots for estimation of $\Lambda_{20}(w)$. The results in the table include the estimated bias (Bias) given by the average of the estimates minus the true value, the sample standard deviation (SSE) of the estimates, the average of the estimated standard error (SEE), and the 95% empirical coverage probability (CP). Table 2 gives the results obtained under the FGM model with $\tau = -0.1, 0.1$, and the results obtained under

the Frank model with $\tau = -0.5, 0.5$ are displayed in Table 3. Other set-ups for Table 2 and Table 3 are the same as for Table 1. Note that different values of parameters were used for different copula models because the parameter range and the correlation described in different models are different. One can see that all results indicate that the proposed estimation procedure seems to work well for the situations considered here. In particular, the estimator seems to be unbiased and the variance estimation appears to be reasonable. The CP and the Q-Q plots, which are not shown here, also suggest that the normal approximation to the distribution of the estimated regression parameters is appropriate. We also considered other set-ups and obtained similar results.

5.5 An Application

In this section, we apply the proposed methodology to the breast cosmesis study conducted by Beadle *et al.* (1984). The study has also been analyzed by Zhang, Hua and Huang (2010). The trial consists of 94 early breast cancer patients, randomly assigned to radiotherapy alone (46 patients) or radiotherapy plus chemotherapy (48 patients). The failure time of interest is the time to breast retraction and only interval-censored data are available. One objective of the study is to compare the two treatments with respect to time until the appearance of breast retraction.

For the analysis, define $Z_i = 0$ if the patient was given radiotherapy alone and $Z_i = 1$ otherwise. In this study, the last possible examination time or the study end time is not available. For patients with right-censored event time, we assume ζ to be the largest examination time among all the patients which is 60. For the

application of the proposed estimation procedure, as in the simulation study, we employed the quadratic splines and considered the three copulas investigated there with $k_n = 1, 2, 3, 4$ and different values of τ . For each set-up, we calculated the AIC and found that the smallest AIC is given by the Frank copula with $k_n = 3$ and $\tau = 0.6$. Actually many AIC values are quite close.

Table 4 presents the estimated treatment effect, $\hat{\beta}$, obtained with $k_n = 3$ and several values of τ under the three copula models, along with the estimated standard error (SEE) and the p -values for testing no treatment effect. For comparison, the results obtained with $k_n = 4$ are included in Table 5. One can easily see that the marginal estimation results are quite different and depend on the possible correlation between the time to breast retraction and the length of observation interval. If we ignore the correlation between them by taking $\tau = 0$, we would get a very significant treatment effect. This is consistent to the results in the previously mentioned papers. However, if there exist moderate positive correlation as suggested by AIC, the estimated results would suggest that the adjuvant chemotherapy doesn't significantly increase the hazard of breast retraction.

5.6 Concluding Remarks

In this chapter, we proposed an approach for regression analysis of interval-censored failure time data under the Cox proportional hazards model when censoring mechanism may be dependent on the failure time of interest. We assume the failure time depend on the censoring interval only through its length and propose to model the dependence via a copula function. For estimation of the baseline cumulative hazard

functions, we propose to use an I -spline function which avoids the order restriction of the spline coefficients compared to using a monotone B-spline function. Also, under the copula model of the dependence, one can use techniques such as AIC to determine the degree of correlation between the failure time and interval length.

Although our method is proposed under Cox model, it should be straightforward to apply this method to other semiparametric regression models with interval-censored data. Specifically, it can be easily extended to the linear transformation models studied by Zhang *et al.* (2005). Another topic related to the proposed method for future research is to allow the failure time to depend on both the lower and upper bounds of the observation interval.

n	ARC	τ	β_0	γ_0	$\hat{\beta}$				$\hat{\gamma}$			
					Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
200	0%	0.5	0	-0.2	0.004	0.134	0.131	0.950	-0.005	0.136	.130	0.942
				0	0.002	0.133	0.130	0.948	0.001	0.135	0.129	0.955
				0.2	0.002	0.134	0.129	0.943	0.006	0.132	0.128	0.940
			0.2	-0.2	0.013	0.140	0.134	0.950	-0.006	0.134	0.130	0.932
				0	0.010	0.134	0.132	0.957	-0.001	0.134	0.130	0.952
				0.2	0.009	0.135	0.130	0.943	0.008	0.133	0.129	0.943
		0.7	0	-0.2	0.003	0.132	0.128	0.953	-0.009	0.135	0.128	0.938
				0	0.001	0.129	0.127	0.942	-0.001	0.132	0.126	0.944
				0.2	-0.001	0.130	0.127	0.950	0.003	0.127	0.126	0.957
			0.2	-0.2	0.003	0.137	0.134	0.935	-0.007	0.132	0.128	0.937
				0	0.010	0.132	0.128	0.955	-0.002	0.133	0.127	0.947
				0.2	0.005	0.131	0.128	0.945	0.003	0.133	0.127	0.935
	20%	0.5	0	-0.2	0.007	0.155	0.153	0.945	0.013	0.162	0.154	0.945
				0	0.002	0.151	0.152	0.945	0.005	0.158	0.152	0.945
				0.2	-0.005	0.153	0.151	0.943	-0.001	0.157	0.151	0.952
			0.2	-0.2	0.015	0.150	0.154	0.958	0.051	0.152	0.155	0.937
				0	0.004	0.155	0.153	0.958	0.041	0.151	0.153	0.937
				0.2	-0.003	0.155	0.152	0.960	0.031	0.150	0.152	0.950
		0.7	0	-0.2	-0.009	0.152	0.148	0.947	0.011	0.158	0.148	0.937
				0	-0.003	0.147	0.146	0.945	0.003	0.156	0.146	0.932
				0.2	-0.020	0.151	0.149	0.945	-0.015	0.153	0.149	0.940
			0.2	-0.2	-0.005	0.153	0.154	0.965	0.031	0.153	0.152	0.942
				0	0.005	0.153	0.148	0.950	0.030	0.153	0.148	0.932
				0.2	0.002	0.152	0.146	0.938	0.018	0.157	0.146	0.940
40%	0.5	0	-0.2	0.022	0.178	0.177	0.948	0.030	0.179	0.179	0.950	
			0	-0.007	0.184	0.175	0.950	0.004	0.183	0.177	0.945	
			0.2	-0.024	0.181	0.174	0.945	-0.016	0.180	0.176	0.955	
		0.2	-0.2	0.016	0.176	0.171	0.950	0.077	0.181	0.171	0.933	
			0	0.007	0.176	0.170	0.940	0.052	0.174	0.170	0.932	
			0.2	-0.003	0.177	0.168	0.935	0.030	0.174	0.169	0.938	
	0.7	0	-0.2	0.008	0.173	0.163	0.947	0.025	0.170	0.163	0.942	
			0	-0.007	0.182	0.169	0.940	0.001	0.176	0.170	0.947	
			0.2	-0.027	0.177	0.167	0.938	-0.029	0.172	0.168	0.947	
		0.2	-0.2	0.036	0.175	0.175	0.950	0.077	0.177	0.175	0.927	
			0	0.009	0.180	0.172	0.937	0.047	0.173	0.171	0.935	
			0.2	-0.013	0.178	0.169	0.935	0.029	0.176	0.170	0.942	

Table 5.1: Estimation of regression parameters under the Gumbel copula.

n	ARC	τ	β_0	γ_0	$\hat{\beta}$				$\hat{\gamma}$			
					Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
200	0%	-0.1	0	-0.2	0.010	0.143	0.146	0.950	-0.018	0.136	0.143	0.953
				0	0.011	0.145	0.146	0.940	-0.013	0.136	0.142	0.953
				0.2	0.012	0.143	0.145	0.950	-0.010	0.137	0.143	0.953
			0.2	-0.2	0.017	0.146	0.148	0.948	-0.018	0.136	0.143	0.950
				0	0.015	0.142	0.147	0.950	-0.014	0.136	0.142	0.952
				0.2	0.017	0.144	0.147	0.945	-0.010	0.137	0.143	0.950
		0.1	0	-0.2	0.013	0.148	0.146	0.935	-0.015	0.139	0.143	0.953
				0	0.011	0.145	0.145	0.938	-0.011	0.138	0.142	0.948
				0.2	0.013	0.145	0.144	0.943	-0.008	0.140	0.143	0.945
			0.2	-0.2	0.019	0.148	0.147	0.943	-0.015	0.139	0.143	0.955
				0	0.019	0.147	0.146	0.938	-0.011	0.139	0.142	0.950
				0.2	0.017	0.145	0.146	0.935	-0.007	0.139	0.143	0.948
	20%	-0.1	0	-0.2	0.006	0.159	0.163	0.948	-0.025	0.156	0.159	0.948
				0	0.008	0.160	0.162	0.950	-0.013	0.155	0.158	0.955
				0.2	0.009	0.157	0.161	0.958	0.001	0.154	0.157	0.945
			0.2	-0.2	0.010	0.164	0.165	0.943	0.013	0.156	0.160	0.955
				0	0.009	0.159	0.163	0.943	0.027	0.156	0.159	0.948
				0.2	0.011	0.160	0.162	0.940	0.037	0.156	0.158	0.940
		0.1	0	-0.2	0.010	0.168	0.164	0.950	-0.023	0.164	0.165	0.945
				0	0.008	0.164	0.162	0.935	-0.013	0.163	0.163	0.945
				0.2	0.010	0.164	0.161	0.940	-0.002	0.165	0.163	0.948
			0.2	-0.2	0.016	0.171	0.166	0.938	0.027	0.168	0.166	0.955
				0	0.015	0.168	0.164	0.953	0.036	0.166	0.164	0.953
				0.2	0.014	0.164	0.163	0.940	0.044	0.167	0.164	0.945
40%	-0.1	0	-0.2	0.008	0.190	0.191	0.965	-0.048	0.185	0.187	0.960	
			0	0.014	0.184	0.187	0.968	-0.019	0.183	0.183	0.960	
			0.2	0.016	0.178	0.185	0.963	0.013	0.179	0.181	0.955	
		0.2	-0.2	0.009	0.190	0.194	0.948	0.027	0.190	0.190	0.948	
			0	0.010	0.184	0.191	0.960	0.057	0.183	0.186	0.935	
			0.2	0.007	0.186	0.188	0.948	0.089	0.175	0.184	0.943	
	0.1	0	-0.2	0.017	0.193	0.191	0.935	-0.023	0.198	0.193	0.948	
			0	0.013	0.188	0.188	0.948	-0.009	0.193	0.190	0.945	
			0.2	0.013	0.178	0.186	0.950	0.001	0.189	0.189	0.955	
		0.2	-0.2	0.015	0.196	0.195	0.955	0.058	0.193	0.195	0.940	
			0	0.003	0.188	0.191	0.953	0.069	0.191	0.192	0.938	
			0.2	0.003	0.185	0.188	0.953	0.085	0.190	0.190	0.933	

Table 5.2: Estimation of regression parameters under the FGM copula.

n	ARC	τ	β_0	γ_0	$\hat{\beta}$				$\hat{\gamma}$			
					Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
200	0%	-0.5	0	-0.2	-0.005	0.131	0.131	0.955	0.002	0.135	0.132	0.948
				0	-0.005	0.131	0.131	0.958	0.007	0.134	0.131	0.950
				0.2	-0.007	0.131	0.131	0.960	0.012	0.135	0.132	0.958
			0.2	-0.2	-0.001	0.133	0.132	0.958	0.001	0.134	0.133	0.958
				0	-0.002	0.132	0.132	0.958	0.006	0.134	0.132	0.948
				0.2	-0.003	0.133	0.132	0.958	0.010	0.135	0.132	0.950
		0.5	0	-0.2	0.003	0.138	0.137	0.935	-0.004	0.139	0.135	0.955
				0	-0.006	0.144	0.136	0.938	0.001	0.130	0.134	0.958
				0.2	0.003	0.134	0.136	0.950	0.005	0.136	0.134	0.963
			0.2	-0.2	0.006	0.140	0.138	0.925	-0.003	0.139	0.135	0.953
				0	-0.001	0.145	0.137	0.945	0.002	0.130	0.134	0.955
				0.2	0.008	0.136	0.137	0.940	0.008	0.137	0.134	0.960
	20%	-0.5	0	-0.2	-0.013	0.137	0.141	0.958	-0.008	0.140	0.144	0.953
				0	-0.009	0.137	0.141	0.955	0.007	0.142	0.142	0.948
				0.2	-0.004	0.136	0.140	0.963	0.022	0.140	0.142	0.948
			0.2	-0.2	-0.004	0.139	0.142	0.958	0.011	0.140	0.145	0.955
				0	0.001	0.138	0.142	0.960	0.027	0.140	0.143	0.955
				0.2	0.005	0.138	0.142	0.953	0.040	0.140	0.143	0.945
		0.5	0	-0.2	0.011	0.162	0.159	0.943	0.008	0.170	0.164	0.948
				0	0.012	0.158	0.157	0.960	0.008	0.166	0.162	0.965
				0.2	0.007	0.160	0.156	0.950	-0.002	0.164	0.162	0.958
			0.2	-0.2	0.010	0.175	0.161	0.940	0.048	0.172	0.165	0.940
				0	0.008	0.175	0.159	0.938	0.042	0.172	0.163	0.940
				0.2	0.022	0.161	0.157	0.948	0.040	0.165	0.163	0.945
40%	-0.5	0	-0.2	-0.001	0.160	0.166	0.965	-0.044	0.161	0.164	0.960	
			0	-0.014	0.167	0.163	0.945	0.009	0.157	0.159	0.950	
			0.2	0.001	0.167	0.162	0.945	0.037	0.156	0.157	0.943	
		0.2	-0.2	-0.007	0.176	0.168	0.935	0.008	0.163	0.167	0.960	
			0	0.009	0.173	0.166	0.938	0.038	0.158	0.163	0.945	
			0.2	0.023	0.171	0.165	0.945	0.066	0.158	0.161	0.935	
	0.5	0	-0.2	0.004	0.190	0.178	0.938	0.017	0.181	0.181	0.948	
			0	-0.017	0.188	0.176	0.935	-0.010	0.186	0.180	0.950	
			0.2	-0.001	0.177	0.175	0.948	-0.020	0.172	0.179	0.970	
		0.2	-0.2	0.020	0.186	0.182	0.945	0.072	0.187	0.184	0.940	
			0	-0.010	0.186	0.179	0.943	0.038	0.186	0.182	0.943	
			0.2	0.011	0.175	0.177	0.945	0.043	0.171	0.181	0.960	

Table 5.3: Estimation of regression parameters under the Frank copula.

copula	τ	$\hat{\beta}$	SEE	p -value	$\hat{\gamma}$	SEE	p -value	AIC
Gumbel	0	0.8958	0.2870	0.0018	0.6773	0.2781	0.0149	720.0236
	0.25	0.8053	0.2853	0.0048	0.7126	0.2799	0.0109	683.4300
	0.5	0.6330	0.2793	0.0235	0.6586	0.2767	0.0173	660.2347
	0.6	0.5187	0.2761	0.0603	0.5983	0.2735	0.0287	656.2529
	0.75	0.3847	0.2734	0.1594	0.5065	0.2704	0.0610	663.4234
FGM	-2/9	0.9738	0.3107	0.0018	0.4208	0.2878	0.1438	772.3420
	-1/9	0.9239	0.2911	0.0015	0.6147	0.2824	0.0295	742.4282
	0	0.8958	0.2870	0.0018	0.6773	0.2781	0.0149	720.0236
	1/9	0.8531	0.2875	0.0030	0.6885	0.2760	0.0126	702.5803
Frank	2/9	0.7716	0.2884	0.0075	0.6685	0.2768	0.0157	688.7241
	-0.5	1.0324	0.3325	0.0019	0.1037	0.3019	0.7314	858.0332
	-0.25	0.9519	0.2937	0.0012	0.5321	0.2876	0.0644	773.7172
	0	0.8958	0.2870	0.0018	0.6773	0.2781	0.0149	720.0236
	0.25	0.7970	0.2856	0.0053	0.6923	0.2708	0.0106	682.5733
	0.5	0.6203	0.2823	0.0280	0.6042	0.2637	0.0219	658.2527
	0.6	0.5128	0.2794	0.0665	0.5280	0.2620	0.0438	654.6860
0.75	0.2776	0.2725	0.3083	0.3283	0.2599	0.2065	665.2190	

Table 5.4: Estimating results for $k_n = 3$.

copula	τ	$\hat{\beta}$	SEE	p -value	$\hat{\gamma}$	SEE	p -value	AIC
Gumbel	0	0.8967	0.2872	0.0018	0.6856	0.2782	0.0137	725.9535
	0.25	0.8095	0.2861	0.0047	0.7281	0.2790	0.0093	688.9773
	0.5	0.6456	0.2808	0.0215	0.6890	0.2775	0.0130	665.3454
	0.6	0.5372	0.2773	0.0527	0.6318	0.2741	0.0212	661.0255
	0.75	0.4223	0.2749	0.1244	0.5519	0.2714	0.0420	667.4047
FGM	-2/9	0.9800	0.3114	0.0016	0.4190	0.2874	0.1449	777.7934
	-1/9	0.9247	0.2915	0.0015	0.6188	0.2829	0.0287	748.3515
	0	0.8967	0.2872	0.0018	0.6856	0.2782	0.0137	725.9535
	1/9	0.8520	0.2877	0.0031	0.6965	0.2758	0.0116	708.3712
Frank	2/9	0.7666	0.2885	0.0079	0.6738	0.2765	0.0148	694.3282
	-0.5	1.0606	0.3304	0.0013	0.0662	0.2971	0.8236	860.8833
	-0.25	0.9513	0.2944	0.0012	0.5251	0.2878	0.0680	779.1396
	0	0.8967	0.2872	0.0018	0.6856	0.2782	0.0137	725.9535
	0.25	0.7929	0.2858	0.0055	0.6990	0.2704	0.0097	688.1050
	0.5	0.6166	0.2823	0.0289	0.6132	0.2637	0.0200	663.3364
	0.6	0.5158	0.2793	0.0648	0.5437	0.2621	0.0380	659.4415
0.75	0.2999	0.2728	0.2715	0.3626	0.2604	0.1637	669.0695	

Table 5.5: Estimating results for $k_n = 4$.

Chapter 6

Future Research

There exist many open questions in the analysis of interval-censored data. In this chapter, we will discuss several potential directions for future research that related to the questions investigated in the previous chapters.

6.1 Multivariate Copula Models for Regression Analysis of Informatively Interval-Censored Data

In Chapter 5, we assumed that the failure time of interest depends on the censoring interval only through its length which could be restrictive in some cases. An alternative would be to allow the failure time to depend on both the lower and upper bounds of the censoring interval and use a multivariate copula model for the correlation between them.

6.2 Cure Rate Models for Interval-Censored Data with Informative Censoring

In medical and health studies with period follow-ups where interval-censored data arises naturally, it often involves the analysis of time to a specific event where some individuals under study are highly susceptible to the event while others are at much lower risk, for example, some of the individuals get treated. Such situations motivate another possible future topic, which is to consider cure rate models for interval-censored data when the censoring mechanism could depend on the failure time of interest.

Appendix A

Theoretical proofs

A.1 Proofs of the Main Results in Chapter 2

A.1.1 Proof of Theorem 2.1

Let

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n W_{ni}(\beta_0) W_{ni}^T(\beta_0).$$

In order to prove (i), it is sufficient if we can show that $\hat{\Sigma}_n = \Sigma_n + o_p(1)$. For any $a \in \mathbb{R}^p$, we have

$$\begin{aligned} a^T(\hat{\Sigma}_n - \Sigma_n)a &= \frac{1}{n} \sum_{i=1}^n (a^T(W_{ni}(\beta_0) - W_i(\beta_0)))^2 + \frac{2}{n} \sum_{i=1}^n (a^T W_i(\beta_0))(a^T(W_{ni}(\beta_0) - W_i(\beta_0))) \\ &\triangleq D_1 + D_2. \end{aligned} \tag{A.1}$$

Denote

$$\begin{aligned}
\psi_{ij}(\beta_0) &= \frac{Z_{ij}\omega_{ij}(\beta_0)\tau'(Z_{ij}^T\beta_0) \int_{L_i}^{R_i} \int_{L_j}^{R_j} I(t_i \geq t_j) dH_{Z_i}(t_i) dH_{Z_j}(t_j)}{(H_{Z_i}(R_i) - H_{Z_i}(L_i))(H_{Z_j}(R_j) - H_{Z_j}(L_j))}, \\
|a^T \psi_{ij}(\beta_0)| &= \frac{|a^T Z_{ij}\omega_{ij}(\beta_0)\tau'(Z_{ij}^T\beta_0)| \int_{L_i}^{R_i} \int_{L_j}^{R_j} I(t_i \geq t_j) dH_{Z_i}(t_i) dH_{Z_j}(t_j)}{(H_{Z_i}(R_i) - H_{Z_i}(L_i))(H_{Z_j}(R_j) - H_{Z_j}(L_j))} \\
&\leq \frac{|a^T Z_{ij}\omega_{ij}(\beta_0)\tau'(Z_{ij}^T\beta_0)| \int_{L_i}^{R_i} (H_{Z_j}(R_j) - H_{Z_j}(L_j)) dH_{Z_i}(t_i)}{(H_{Z_i}(R_i) - H_{Z_i}(L_i))(H_{Z_j}(R_j) - H_{Z_j}(L_j))} \\
&= |a^T Z_{ij}\omega_{ij}(\beta_0)\tau'(Z_{ij}^T\beta_0)| \triangleq |\alpha_{ij}(\beta_0)|.
\end{aligned}$$

Let $\hat{\psi}_{ij}(\beta)$ be $\psi_{ij}(\beta)$ with H replaced by \hat{H} . Let t_i^* denote $((t_i \wedge R_j) \vee L_j)$. We have

$$|a^T(\hat{\psi}_{ij}(\beta_0) - \psi_{ij}(\beta_0))| = \frac{|\alpha_{ij}(\beta_0)(I_1 + I_2)|}{B(R_{ij}, L_{ij})},$$

where $B(R_{ij}, L_{ij}) = (\hat{H}_{Z_i}(R_i) - \hat{H}_{Z_i}(L_i))(\hat{H}_{Z_j}(R_j) - \hat{H}_{Z_j}(L_j))(H_{Z_i}(R_i) - H_{Z_i}(L_i))(H_{Z_j}(R_j) - H_{Z_j}(L_j))$,

$$\begin{aligned}
I_1 &= (H_{Z_i}(R_i) - H_{Z_i}(L_i))(H_{Z_j}(R_j) - H_{Z_j}(L_j)) \int_{L_i}^{R_i} \hat{H}_{Z_j}(t_i^*) d\hat{H}_{Z_i}(t_i) \\
&\quad - (\hat{H}_{Z_i}(R_i) - \hat{H}_{Z_i}(L_i))(\hat{H}_{Z_j}(R_j) - \hat{H}_{Z_j}(L_j)) \int_{L_i}^{R_i} H_{Z_j}(t_i^*) dH_{Z_i}(t_i), \\
I_2 &= (H_{Z_i}(R_i) - H_{Z_i}(L_i)) \left(\hat{H}_{Z_i}(R_i) - \hat{H}_{Z_i}(L_i) \right) S(R_j, L_j)
\end{aligned}$$

with $S(R_j, L_j) = (H_{Z_j}(R_j) - H_{Z_j}(L_j)) \hat{H}_{Z_j}(L_j) - (\hat{H}_{Z_j}(R_j) - \hat{H}_{Z_j}(L_j)) H_{Z_j}(L_j)$.

By a straightforward computation, we have

$$\begin{aligned}
|S(R_j, L_j)| &= |H_{Z_j}(R_j)(\hat{H}_{Z_j}(L_j) - H_{Z_j}(L_j)) + H_{Z_j}(L_j)(H_{Z_j}(R_j) - \hat{H}_{Z_j}(R_j))| \\
&\leq |\hat{H}_{Z_j}(R_j) - H_{Z_j}(R_j)| + |\hat{H}_{Z_j}(L_j) - H_{Z_j}(L_j)|.
\end{aligned}$$

Hence we have

$$|I_2| \leq |\hat{H}_{Z_j}(R_j) - H_{Z_j}(R_j)| + |\hat{H}_{Z_j}(L_j) - H_{Z_j}(L_j)|.$$

Similarly, we can obtain that

$$\begin{aligned} |I_1| &\leq \left| \int_{L_i}^{R_i} \hat{H}_{Z_j}(t_i^*) d\hat{H}_{Z_i}(t_i) - \int_{L_i}^{R_i} H_{Z_j}(t_i^*) dH_{Z_i}(t_i) \right| + |\hat{H}_{Z_i}(R_i) - H_{Z_i}(R_i)| \\ &\quad + |\hat{H}_{Z_i}(L_i) - H_{Z_i}(L_i)| + |\hat{H}_{Z_j}(R_j) - H_{Z_j}(R_j)| + |\hat{H}_{Z_j}(L_j) - H_{Z_j}(L_j)| \\ &\triangleq |A_1| + |\hat{H}_{Z_i}(R_i) - H_{Z_i}(R_i)| + |\hat{H}_{Z_i}(L_i) - H_{Z_i}(L_i)| \\ &\quad + |\hat{H}_{Z_j}(R_j) - H_{Z_j}(R_j)| + |\hat{H}_{Z_j}(L_j) - H_{Z_j}(L_j)|, \end{aligned}$$

where

$$\begin{aligned} A_1 &= \int_{L_i}^{R_i} (\hat{H}_{Z_j}(t_i^*) - H_{Z_j}(t_i^*)) d\hat{H}_{Z_i}(t_i) + \int_{L_i}^{R_i} H_{Z_j}(t_i^*) d\hat{H}_{Z_i}(t_i) - \int_{L_i}^{R_i} H_{Z_j}(t_i^*) dH_{Z_i}(t_i) \\ &= \int_{L_i}^{R_i} (\hat{H}_{Z_j}(t_i^*) - H_{Z_j}(t_i^*)) d\hat{H}_{Z_i}(t_i) - \int_{L_i}^{R_i} (\hat{H}_{Z_j}(t_i) - H_{Z_j}(t_i)) dH_{Z_i}(t_i^*) \\ &\quad + H_{Z_j}((R_i \wedge R_j) \vee L_j) (\hat{H}_{Z_i}(R_i) - H_{Z_i}(R_i)) - H_{Z_j}((L_i \wedge R_j) \vee L_j) (\hat{H}_{Z_i}(L_i) - H_{Z_i}(L_i)). \end{aligned}$$

By Groeneboom & Wellner (1992), we have that

$$k_1 = \sup_{t \in \mathbb{R}} |\hat{H}_Z(t) - H_Z(t)| = o_p(1).$$

Next note that

$$\begin{aligned}
|a^T(W_{ni}(\beta_0) - W_i(\beta_0))| &= \frac{1}{n-1} \sum_{j=1, j \neq i}^n |a^T(\hat{\psi}_{ij}(\beta_0) - \psi_{ij}(\beta_0)) + a^T(\hat{\psi}_{ji}(\beta_0) - \psi_{ji}(\beta_0))| \\
&\leq \frac{1}{n-1} \sum_{j=1, j \neq i}^n \left(\frac{8|\alpha_{ij}(\beta_0)|k_1}{B(R_{ij}, L_{ij})} + \frac{8|\alpha_{ji}(\beta_0)|k_1}{B(R_{ji}, L_{ji})} \right) \\
&\leq \frac{8k_1}{B_1}(\alpha_{ij}(\beta_0) + \alpha_{ji}(\beta_0)),
\end{aligned}$$

where $B_1 = \min\{B(R_{ij}, L_{ij}) | i, j = 1, \dots, n\}$.

We also have

$$\begin{aligned}
|a^T W_i(\beta_0)| &\leq \frac{1}{n-1} \sum_{j=1, j \neq i}^n (|a^T \psi_{ij}(\beta_0)| + |a^T \psi_{ji}(\beta_0)|) \\
&\quad + \frac{1}{n-1} \sum_{j=1, j \neq i}^n (|a^T Z_{ij} \omega_{ij}(\beta_0) \tau'(Z_{ij}^T \beta_0) \tau(Z_{ij}^T \beta_0)| + |a^T Z_{ji} \omega_{ji}(\beta_0) \tau'(Z_{ji}^T \beta_0) \tau(Z_{ji}^T \beta_0)|).
\end{aligned}$$

By the conditions 1 and 2, thus $|a^T Z_{ij} \omega_{ij}(\beta_0) \tau'(Z_{ij}^T \beta_0) \tau(Z_{ij}^T \beta_0)|$ and $|a^T \psi_{ij}(\beta)|$, i.e, $|\alpha_{ij}(\beta_0)|$ are bounded and denoted by $C_1/2$ and $C_2/2$ for constants C_1 and C_2 , respectively. So we have

$$|D_2| \leq \frac{16C_2(C_1 + C_2)k_1}{B_1} = o_p(1). \quad (\text{A.2})$$

Similarly, we can show that $|D_1| = o_p(1)$. Combining (A.1) and (A.2), we obtain that (i) holds.

In order to prove (ii), we only need to show that $\hat{\Sigma} = \hat{\Sigma}_n + o_p(1)$. For any $a \in \mathbb{R}^p$,

we have

$$a^T(\hat{\Sigma}_n - \Sigma_n)a = \frac{1}{n} \sum_{i=1}^n (a^T(W_{ni}(\hat{\beta}) - W_{ni}(\beta_0)))^2 + \frac{2}{n} \sum_{i=1}^n (a^T W_{ni}(\beta_0))(a^T(W_{ni}(\hat{\beta}) - W_{ni}(\beta_0))).$$

By some straightforward computation, we have

$$\begin{aligned} & a^T(W_{ni}(\hat{\beta}) - W_{ni}(\beta_0)) \\ &= \frac{1}{n} \sum_{j=1, j \neq i}^n \left\{ a^T(\hat{\psi}_{ij}(\hat{\beta}) - \hat{\psi}_{ij}(\beta_0)) + a^T(\eta_{ij}(\hat{\beta}) - \eta_{ij}(\beta_0)) \right. \\ & \left. + a^T(\hat{\psi}_{ji}(\hat{\beta}) - \hat{\psi}_{ji}(\beta_0)) + a^T(\eta_{ji}(\hat{\beta}) - \eta_{ji}(\beta_0)) \right\}, \end{aligned}$$

where $\eta_{ij}(\beta) = Z_{ij}\omega_{ij}(\beta)\tau'(Z_{ij}^T\beta)\tau(Z_{ij}^T\beta)$. Under conditions 1 and 2 applying the differential mean value theorem, and using the consistency and the asymptotic normality of $\hat{\beta}$ obtained by Zhang et al. (2005), we have

$$|a^T(\hat{\psi}_{ij}(\hat{\beta}) - \hat{\psi}_{ij}(\beta_0))| = O_p(n^{-\frac{1}{2}}) \quad \text{and} \quad |a^T(\eta_{ji}(\hat{\beta}) - \eta_{ji}(\beta_0))| = O_p(n^{-\frac{1}{2}}) \quad \text{for } 1 \leq i, j \leq n.$$

Thus we have

$$|a^T(W_{ni}(\hat{\beta}) - W_{ni}(\beta_0))| = O_p(n^{-\frac{1}{2}}).$$

Seen from the proof for the result (i), it can be seen that $|a^T W_{ni}(\beta_0)| \leq C_3$ for a constant C_3 . Therefore, we have

$$|a^T(\hat{\Sigma}_n - \Sigma_n)a| = O_p(n^{-\frac{1}{2}}).$$

The result (ii) follows. By using the same proof method as above, we can show the result (iii).

This completes the proof of Theorem 2.1.

A.1.2 Proof of Theorem 2.2

Let $Y_n = \max_{1 \leq i \neq j \leq n} \|b(U_i, U_j, \beta_0)\|$. Under conditions 1 and 2, it is easy to see that $E \|b(U_1, U_2, \beta_0)\|^2 < \infty$. We have $Y_n = o(n^{\frac{1}{2}})$, a.s. (refer to the result of Jing et al (2008) on page 605). Note that

$$\|W_i(\beta_0)\| \leq \frac{1}{n-1} \sum_{j=1, j \neq i}^n \|b(U_i, U_j, \beta_0)\| \leq Y_n, \quad \text{for } i = 1, \dots, n.$$

Thus,

$$\max_{1 \leq i \leq n} \|W_i(\beta_0)\| = o(n^{\frac{1}{2}}), \quad \text{a.s.} \quad (\text{A.3})$$

Note that

$$W_{ni}(\beta_0) - W_i(\beta_0) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n \left\{ (\hat{\psi}_{ij}(\beta_0) - \psi_{ij}(\beta_0)) + (\hat{\psi}_{ji}(\beta_0) - \psi_{ji}(\beta_0)) \right\},$$

and

$$\begin{aligned}
& B(R_{ij}, L_{ij})(\hat{\psi}_{ji}(\beta_0) - \psi_{ji}(\beta_0)) \\
&= \kappa_{ij}(\beta_0) \left(\hat{H}_{Z_j}(R_j) - \hat{H}_{Z_j}(L_j) \right) \left\{ \int_{L_i}^{R_i} \frac{\hat{H}_{Z_j}(t_i^*) - \hat{H}_{Z_j}(L_j)}{\hat{H}_{Z_j}(R_j) - \hat{H}_{Z_j}(L_j)} d\hat{H}_{Z_i}(t_i) \right. \\
&\quad \left. \times \int_{L_i}^{R_i} \int_{L_j}^{R_j} dH_{Z_i}(t_i) dH_{Z_j}(t_j) - \int_{L_i}^{R_i} d\hat{H}_{Z_i}(t_i) \times \int_{L_i}^{R_i} \int_{L_j}^{R_j} I(t_i \geq t_j) dH_{Z_i}(t_i) dH_{Z_j}(t_j) \right\} \\
&= \theta_{ij}(\beta_0) \left\{ \int_{L_i}^{R_i} \left(\frac{\hat{H}_{Z_j}(t_i^*) - \hat{H}_{Z_j}(L_j)}{\hat{H}_{Z_j}(R_j) - \hat{H}_{Z_j}(L_j)} - 1 \right) d\hat{H}_{Z_i}(t_i) \int_{L_i}^{R_i} \int_{L_j}^{R_j} I(t_i \geq t_j) dH_{Z_i}(t_i) dH_{Z_j}(t_j) \right\} \\
&\quad + \kappa_{ij}(\beta_0) \int_{L_i}^{R_i} \int_{L_j}^{R_j} I(t_i \geq t_j) d\hat{H}_{Z_i}(t_i) d\hat{H}_{Z_j}(t_j) \times \int_{L_i}^{R_i} \int_{L_j}^{R_j} I(t_i < t_j) dH_{Z_i}(t_i) dH_{Z_j}(t_j),
\end{aligned}$$

where $\kappa_{ij}(\beta_0) = Z_{ij}\omega_{ij}(\beta_0)\tau'(Z_{ij}(\beta_0))$ and $\theta_{ij}(\beta_0) = \kappa_{ij}(\beta_0) \left(\hat{H}_{Z_j}(R_j) - \hat{H}_{Z_j}(L_j) \right)$. So

we have

$$\begin{aligned}
\| \hat{\psi}_{ji}(\beta_0) - \psi_{ji}(\beta_0) \| &\leq \frac{1}{\hat{H}_{Z_i}(R_i) - \hat{H}_{Z_i}(L_i)} \\
&\quad \times \frac{\| Z_{ij}\omega_{ij}(\beta_0)\tau'(Z_{ij}^T\beta_0) \| \int_{L_i}^{R_i} \int_{L_j}^{R_j} I(t_i \geq t_j) dH_{Z_i}(t_i) dH_{Z_j}(t_j)}{(H_{Z_i}(R_i) - H_{Z_i}(L_i))(H_{Z_j}(R_j) - H_{Z_j}(L_j))} \\
&\leq C_4 \| \psi_{ij}(\beta_0) \|
\end{aligned}$$

for a constant C_4 . It follows from (A.3) that

$$\max_{1 \leq i \leq n} \| W_{ni}(\beta_0) - W_i(\beta_0) \| \leq 2C_4 \max_{1 \leq i \neq j \leq n} \| \psi_{ij}(\beta_0) \| = o_p(n^{\frac{1}{2}}).$$

Hence we have

$$\max_{1 \leq i \leq n} \| W_{ni}(\beta_0) \| \leq \max_{1 \leq i \leq n} \| W_{ni}(\beta_0) - W_i(\beta_0) \| + \max_{1 \leq i \leq n} \| W_i(\beta_0) \| = o_p(n^{\frac{1}{2}}). \quad (\text{A.4})$$

From the Appendix of Zhang et al (2005), we obtain that $2U(\beta_0) = (n-1) \sum_{i=1}^n W_{ni}(\beta_0) = O_p(n^{\frac{3}{2}})$. It follows from (2.8) and the arguments as those in Owen (1990) that

$$\|\lambda\| = O_p(n^{-\frac{1}{2}}). \quad (\text{A.5})$$

Then by (A.4) and (A.5), we have

$$\max_{1 \leq i \leq n} |\lambda^T W_{ni}(\beta_0)| = o_p(1). \quad (\text{A.6})$$

It follows (A.6) and the arguments as those in Owen (1990), by using the Taylor's expansions to (2.7) and (2.8), refer to the Taylor's expansions of the functions $1/(1+x)$ and $\log(1+x)$ as $x \rightarrow 0$, we obtain that

$$\begin{aligned} \frac{\hat{l}(\beta_0)}{4} &= \frac{1}{4} \sum_{i=1}^n \lambda^T W_{ni}(\beta_0) + o_p(1) \\ &= \frac{1}{4} \left(n^{-\frac{3}{2}} 2U(\beta_0) \right)^T \left(\frac{1}{n} \sum_{i=1}^n W_{ni}(\beta_0) W_{ni}^T(\beta_0) \right)^{-1} \left(n^{-\frac{3}{2}} 2U(\beta_0) \right) + o_p(1) \\ &= \left(\Gamma^{-\frac{1}{2}}(\beta_0) n^{-\frac{3}{2}} U(\beta_0) \right)^T \left(\Gamma^{\frac{1}{2}}(\beta_0) \Sigma^{-1} \Gamma^{\frac{1}{2}}(\beta_0) \right) \left(\Gamma^{-\frac{1}{2}}(\beta_0) n^{-\frac{3}{2}} U(\beta_0) \right) + o_p(1). \end{aligned}$$

From the Zhang et al (2005), we have that $\Gamma^{-\frac{1}{2}}(\beta_0) n^{-\frac{3}{2}} U(\beta_0) \xrightarrow{D} N(0, I_p)$. Since $\Gamma^{\frac{1}{2}}(\beta_0) \Sigma^{-1} \Gamma^{\frac{1}{2}}(\beta_0)$ is a real symmetric and invertible matrix, what's more, the eigenvalues of this matrix is the same as those of matrix $\Sigma^{-1} \Gamma(\beta_0)$. Theorem 2.2 follows.

A.2 Proofs of the Main Results in Chapter 3

For the asymptotic properties of $\hat{\beta}_{EP}$, we need the following regularity conditions.

(1) $Z(t)$ is bounded.

(2) $P\{Y_i(t) = 1\} > 0$, for all $t \in [0, \tau]$ and $i = 1, \dots, n$.

(3) $\int_0^\tau \lambda_0(t) dt < \infty$.

(4) There exists a neighborhood B of the true parameter β_0 such that $(\partial^2/\partial\beta_i\partial\beta_j)\varphi(t, \beta)$ exists and is uniformly continuous on B ; the function $\phi(t, \beta)$ is bounded away from 0 on $[0, \tau] \times B$. The matrix $\Sigma(\beta_0)$ is positive definite.

(5) Assume that

$$\begin{aligned} E \left\{ \sup_{[0, \tau] \times B} |Y(t)\Phi^{(j)}(t, \beta)| \right\} &< \infty, \quad j = 0, 1, 2, \\ E \left\{ \sup_{[0, \tau] \times B} \left| Y(t) \left\| \left(\frac{\Phi^{(1)}(t, \beta)}{\Phi(t, \beta)} \right)^{\otimes 2} \right\|^j \Phi(t, \beta_0) \right| \right\} &< \infty, \quad j = 1, 2, \\ E \left\{ \sup_{[0, \tau] \times B} \left| Y(t) \left\| \frac{\Phi^{(2)}(t, \beta)}{\Phi(t, \beta)} \right\|^j \Phi(t, \beta_0) \right| \right\} &< \infty, \quad j = 1, 2. \end{aligned}$$

(6) $\sup_{t \in [0, \tau]} |H_X^{(j)}(t)| = O_p(1)$, where

$$H_X^{(j)}(t) = \sqrt{n_v} \left\{ \frac{1}{n_v} \sum_{i \in V} I(Y_i = 1, X_i = x) \varphi_i^{(j)}(t, \beta) - E(I(Y_i = 1, X_i = x) \varphi_i^{(j)}(t, \beta)) \right\}, \quad j = 0, 1.$$

A.2.1 Proof of Theorem 3.1

To prove the consistency of $\hat{\beta}_{EP}$, first note the fact that $n^{-1}U_{EP}(\hat{\beta}_{EP}) = 0$. Thus it follows from the arguments of Foutz (1977) that it is sufficient to show that the following four statements hold.

(a) $n^{-1}(\partial U_{EP}(\beta)/\partial\beta)$ exists and is continuous in an open neighborhood of β_0 .

(b) $n^{-1}(\partial U_{EP}(\beta_0)/\partial\beta_0)$ is negative definite in probability.

(c) $n^{-1}(\partial U_{EP}(\beta)/\partial\beta)$ converges in probability to a fixed function, say, $\Sigma(\beta)$, uniformly in an open neighborhood of β_0 .

(d) $n^{-1}U_{EP}(\beta_0)$ converges to 0 in probability.

First (a) holds obviously. For (b) and (c), using a similar method to that used in the proof of Lemma A.1 of Fan et al (2006), under conditions (1) and (2), we can obtain that

$$\sup_{B \times [0, \tau]} \|\hat{S}_n^{(j)} - s^{(j)}\| \longrightarrow 0, \text{ a.s. for } j = 0, 1, 2, \quad (\text{A.7})$$

$$\sup_{B \times [0, \tau]} \|D_{n2}^{(1)} - d_2^{(1)}\| \longrightarrow 0, \text{ a.s., and } \|D_{n1}^{(2)} - d_1^{(2)}\| \longrightarrow 0, \text{ a.s.} \quad (\text{A.8})$$

Also by the conditions (1) and (3), we can prove that

$$\frac{1}{n} \sum_{i=1}^n \int_0^\tau \left(\frac{\hat{\Phi}_i^{(2)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} - \left(\frac{\hat{\Phi}_i^{(1)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} \right)^{\otimes 2} \right) dM_i(t) \longrightarrow 0, \text{ in probability in } B.$$

Furthermore, note that

$$\frac{1}{n} \sum_{i=1}^n \int_0^\tau \left(\frac{S_n^{(2)}(t, \beta)}{S_n^{(0)}(t, \beta)} - \left(\frac{\hat{S}_n^{(1)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} \right)^{\otimes 2} \right) dM_i(t) \quad (\text{A.9})$$

is asymptotically equivalent to a local square integrable martingale. Therefore by the inequality of Lenglart (Andersen and Gill, 1982), we have that the function (A.9) converges to 0 in probability uniformly in B . Thus we get $A_1(\tau, \beta) \rightarrow 0$ in probability

in B. For $A_2(\tau, \beta)$, note that

$$\begin{aligned} A_2(\tau, \beta) &= \int_0^\tau \left(D_{n1}^{(2)}(t, \beta) - \frac{\hat{S}_n^{(2)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} S_n^{(0)}(t, \beta_0) \right) \lambda_0^c(t) dt \\ &\quad - \int_0^\tau \left(D_{n2}^{(1)}(t, \beta) - \left(\frac{\hat{S}_n^{(1)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} \right)^{\otimes 2} S_n^{(0)}(t, \beta_0) \right) \lambda_0^c(t) dt, \end{aligned}$$

which converges in probability to

$$\int_0^\tau \left[\left(d_1^{(2)}(t, \beta) - \frac{s^{(2)}(t, \beta)}{s^{(0)}(t, \beta)} s^{(0)}(t, \beta_0) \right) - \left(d_2^{(1)}(t, \beta) - \left(\frac{s^{(1)}(t, \beta)}{s^{(0)}(t, \beta)} \right)^{\otimes 2} s^{(0)}(t, \beta_0) \right) \right] \lambda_0^c(t) dt. \quad (\text{A.10})$$

So (c) holds. By replacing β with β_0 in equation (A.10), (b) follows from the condition (4) immediately.

For (d), using some similar arguments used as above, we can obtain that $n^{-1}U_{EP}(\beta_0)$ converges to the same limit as that of

$$D(\tau, \beta_0) \triangleq \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left(\frac{\hat{\Phi}_i^{(1)}(t, \beta_0)}{\hat{\Phi}_i(t, \beta_0)} - \frac{\hat{S}_n^{(1)}(t, \beta_0)}{\hat{S}_n^{(0)}(t, \beta_0)} \right) \Phi_i(t, \beta_0) Y_i(t) \lambda_0^c(t) dt.$$

Thus it is sufficient to show that $D(\tau, \beta_0)$ converges to 0 in probability. Note that

$$\frac{1}{n} \sum_{i=1}^n \frac{\hat{\Phi}_i^{(1)}(t, \beta_0)}{\hat{\Phi}_i(t, \beta_0)} \Phi_i(t, \beta_0) Y_i(t) = \frac{1}{n} \sum_{i \in \bar{V}} \frac{\hat{\phi}_i^{(1)}(t, \beta_0)}{\hat{\phi}_i(t, \beta_0)} Y_i(t) \phi_i(t, \beta_0) - \frac{1}{n} \sum_{i \in V} Z_i^*(t) Y_i(t) e^{-\beta_0 Z_i^*(t)} \triangleq L - \tilde{L}.$$

By the condition (5) and the law of large numbers, we have

$$L = \frac{n_{\bar{v}}}{n} \frac{1}{n_{\bar{v}}} \sum_{i \in \bar{V}} \frac{\hat{\phi}_i^{(1)}(t, \beta_0)}{\hat{\phi}_i(t, \beta_0)} Y_i(t) \phi_i(t, \beta_0) \xrightarrow{P} (1 - \rho) \sum_{l=1}^m \phi_{il}^{(1)}(t, \beta_0) E\{Y_i(t) I(X_i(t) = x_l)\},$$

where $\phi_{il}(t, \beta) = E(e^{-\beta Z_i^*(t)} | Y_i = 1, X_i(t) = x_l)$. On the other hand, by the law of large numbers, $\tilde{L} \xrightarrow{P} \rho E(Z_i^*(t) Y_i(t) e^{-\beta_0 Z_i^*(t)})$. Therefore, we have that

$$\frac{1}{n} \sum_{i=1}^n \frac{\hat{\Phi}_i^{(1)}(t, \beta_0)}{\hat{\Phi}_i(t, \beta_0)} \Phi_i(t, \beta_0) Y_i(t) \xrightarrow{P} s^{(1)}(t, \beta_0). \quad (\text{A.11})$$

Hence it follows from the law of large numbers, equations (A.7) and (A.11) that $D(\tau, \beta_0)$ converges to 0 in probability. That is, (d) is true and we complete the proof.

A.2.2 Proof of Theorem 3.2

First note that

$$\begin{aligned} \frac{1}{\sqrt{n}} U_{EP}(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \left(\frac{\hat{\Phi}_i^{(1)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} - \frac{S_n^{(1)}(t, \beta)}{S_n^{(0)}(t, \beta)} \right) dM_i(t) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \left(\frac{\hat{\Phi}_i^{(1)}(t, \beta)}{\hat{\Phi}_i(t, \beta)} - \frac{\hat{S}_n^{(1)}(t, \beta)}{\hat{S}_n^{(0)}(t, \beta)} \right) \Phi_i(t, \beta_0) Y_i(t) \lambda_0^c(t) dt, \end{aligned}$$

and

$$\frac{1}{\sqrt{n}} U_{EP}(\beta_0) = -\frac{1}{\sqrt{n}} (U_{EP}(\hat{\beta}_{EP}) - U_{EP}(\beta_0)) = \left\{ -\frac{1}{n} (\partial/\partial\beta^*) U_{EP}(\beta^*) \right\} \sqrt{n} (\hat{\beta}_{EP} - \beta_0), \quad (\text{A.12})$$

where β^* is between $\hat{\beta}_{EP}$ and β_0 . To prove the asymptotic normality, it suffices to prove that $n^{-1/2} U_{EP}(\beta_0)$ converges to a normal random variable in distribution and that $n^{-1} (\partial/\partial\beta^*) U_{EP}(\beta^*)$ converges to an invertible matrix. The latter is straightforward by the consistency of $\hat{\beta}_{EP}$ and the convergence proof of $n^{-1} (\partial/\partial\beta) U_{EP}(\beta)$ for (c).

That is,

$$-n^{-1}(\partial/\partial\beta^*)U_{EP}(\beta^*) \xrightarrow{P} \Sigma(\beta_0). \quad (\text{A.13})$$

For the asymptotic normality of $n^{-1/2}U_{EP}(\beta_0)$, note that by applying the first-order expansion (Zhou and Pepe, 1995),

$$\frac{x}{y} = \frac{x_0}{y_0} + \frac{x - x_0}{y_0} - \frac{(y - y_0)x_0}{y_0^2} + o\{(x - x_0)^2 + (y - y_0)^2\},$$

to $\hat{\Phi}_i^{(1)}/\hat{\Phi}_i^{(0)}$ and $\hat{S}_n^{(1)}/\hat{S}_n^{(0)}$ at $\Phi_i^{(1)}/\Phi_i^{(0)}$ and $s^{(1)}/s^{(0)}$, respectively. We can show that the second term of $n^{-1/2}U_{EP}(\beta_0)$ is equal to

$$\begin{aligned} & -\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \left(\frac{\Phi_i^{(1)}(t, \beta_0)}{\Phi_i(t, \beta_0)} - \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \right) Y_i(t) (\hat{\Phi}_i(t, \beta_0) - \Phi_i(t, \beta_0)) \lambda_0^c(t) dt + o_p(1) \\ & = -\frac{1}{\sqrt{n}} \sum_{i \in \bar{V}} \int_0^\tau \left(\frac{\phi_i^{(1)}(t, \beta_0)}{\phi_i(t, \beta_0)} - \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \right) Y_i(t) (\hat{\phi}_i(t, \beta_0) - \phi_i(t, \beta_0)) \lambda_0^c(t) dt + o_p(1) \\ & \triangleq \Gamma(\tau, \beta_0) + o_p(1). \end{aligned}$$

By the conditions (1) - (6), we have that

$$\begin{aligned} \Gamma(\tau, \beta_0) & = -\frac{1}{\sqrt{n}} \sum_{i \in \bar{V}} \int_0^\tau \left(\frac{\phi_i^{(1)}(t, \beta_0)}{\phi_i(t, \beta_0)} - \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \right) \frac{1}{P(X = X_i, C \geq t)} Y_i(t) \\ & \quad \times \left(\frac{1}{n_v} \sum_{j \in V} I(X_j = X_i) Y_j(t) (\varphi_j(t, \beta_0) - \phi_i(t, \beta_0)) \right) \lambda_0^c(t) dt + o_p(1) \\ & = -\frac{1}{\sqrt{n}} \frac{n_{\bar{v}}}{n_v} \sum_{j \in V} \int_0^\tau \left(\frac{\phi_j^{(1)}(t, \beta_0)}{\phi_j(t, \beta_0)} - \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \right) (\varphi_j(t, \beta_0) - \phi_j(t, \beta_0)) Y_j(t) \lambda_0^c(t) dt + o_p(1) \\ & \triangleq -\frac{1}{\sqrt{n}} \frac{n_{\bar{v}}}{n_v} \sum_{j \in V} Q_j + o_p(1). \end{aligned}$$

Therefore $n^{-1/2}U_{EP}(\beta_0)$ can be rewritten as

$$\begin{aligned} & \frac{1}{\sqrt{n}} \int_0^\tau \sum_{i \in \bar{V}} \left(\frac{\phi_i^{(1)}(t, \beta_0)}{\phi_i(t, \beta_0)} - \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \right) dM_i(t) \\ & + \frac{1}{\sqrt{n}} \sum_{j \in V} \left\{ \int_0^\tau \left(\frac{\varphi_j^{(1)}(t, \beta_0)}{\varphi_j(t, \beta_0)} - \frac{s^{(1)}(t, \beta_0)}{s^{(0)}(t, \beta_0)} \right) dM_j(t) - \frac{n_{\bar{v}}}{n_v} Q_j \right\} + o_p(1) \\ & \triangleq I_1 + I_2 + o_p(1). \end{aligned}$$

By the martingale central limit theorem (Fleming and Harrington, 1991), I_1 converges weakly to the normal variable with mean zero and the covariance $(1 - \rho)\Sigma_1(\beta_0)$. Note that I_2 is a summation of independent identically distributed terms from the subjects in the validation sample set, and the terms in I_2 have mean zero since that $M_j(t)$ is a martingale and $E(Q_j) = 0$. Through the simple central limit theorem, we obtain that I_2 is asymptotically normal with mean zero and covariance $\rho\Sigma_2(\beta_0)$. It thus follows from the independence between I_1 and I_2 that $n^{-1/2}U_{EP}(\beta_0)$ converges to a mean zero normal random vector with covariance $(1 - \rho)\Sigma_1(\beta_0) + \rho\Sigma_2(\beta_0)$. The theorem thus follows from combining equations (A.12), (A.13) and the asymptotic normality of $I_1 + I_2$.

A.3 Proofs of the Asymptotic Properties in Chapter 4

Assumptions

(A1) Z has a bounded support in R^p where p is the dimension of Z .

(A2) $C(\cdot, \cdot)$ has a bounded first order partial derivatives for C on $(0, 1)^2$. $\frac{\partial C(u, v)}{\partial u}$ and $\frac{\partial C(u, v)}{\partial v}$ be Lipschitz.

(A3) $\sup_{d(\theta, \theta_0) > \epsilon} Pl(\theta, X) < Pl(\theta_0, X)$.

(A4) The m th derivatives of $\Lambda_T(\cdot)$, $\Lambda_C(\cdot)$, denoted by $\Lambda_T^{(m)}(\cdot)$, $\Lambda_C^{(m)}(\cdot)$. $\Lambda_T(\cdot)$, $\Lambda_C(\cdot)$ belong to $\mathcal{A} = \{\Lambda : |\Lambda^{(m)}(t_1) - \Lambda^{(m)}(t_2)| \leq M|t_1 - t_2|^\eta\}$ for some $\eta \in (0, 1]$ and for all $t_1, t_2 \in (l, u)$, $k = 1, 2$ and M is some constant. Denote $r = m + \eta$.

(A5) $E(S_\vartheta S'_\vartheta)$ is finite and positive definite with $\vartheta = (\beta', \gamma')'$ where S_ϑ defined in the proof of Theorem 4.3.

Note that the conditions above are mild and usually satisfied in practical situations (e.g., Huang and Rossini, 1997; Zhang et al. 2010). Condition (A3) states that the true parameters should be well separated from the point of maximum in terms of the Kullback-Leiber divergence, which is usually imposed for the consistency of M-estimators (van der Vaart and Wellner, 1996, page 286). The proofs of Theorems 4.1 - 4.3 involve the theory of empirical processes and some techniques commonly used in nonparametric literature. For ease of exposition, we denote $Pg = \int g(x)dP(x)$ and $P_n g = n^{-1} \sum_{i=1}^n g(X_i)$ with the empirical process indexed by function $g(X)$.

A.3.1 Proof of Theorem 4.1

To establish the consistency using empirical process theory, we consider a class of functions \mathcal{L}_n defined by $\{l(\theta, X) : \theta \in \Theta_n\}$. For any $\theta^1 = (\beta^1, \gamma^1, \Lambda_T^1, \Lambda_C^1)$, $\theta^2 = (\beta^2, \gamma^2, \Lambda_T^2, \Lambda_C^2) \in \Theta_n$,

$$\begin{aligned}
& |l(\theta^1, X) - l(\theta^2, X)| \\
&= |l_\beta(\tilde{\theta}, X)(\beta^1 - \beta^2)| + |l_\gamma(\theta, X)(\gamma^1 - \gamma^2)| \\
&\quad + |l_{\Lambda_T}(\tilde{\theta}, X)(\Lambda_T^1 - \Lambda_T^2)| + \|l_{\Lambda_C}(\theta, X)(\Lambda_C^1 - \Lambda_C^2)\| \\
&\leq K(\|\beta^1 - \beta^2\| + \|\gamma^1 - \gamma^2\|) + K\|\Lambda_T^1 - \Lambda_T^2\|_\infty + K\|\Lambda_C^1 - \Lambda_C^2\|_\infty \\
&\leq K(\|\beta^1 - \beta^2\| + \|\gamma^1 - \gamma^2\| + \|\Lambda_T^1 - \Lambda_T^2\|_\infty + \|\Lambda_C^1 - \Lambda_C^2\|_\infty)
\end{aligned}$$

where $\tilde{\theta}$ lies between θ^1 and θ^2 ,

$$\begin{aligned}
l_\beta(\theta, X) &= \Delta \left\{ \frac{\delta}{m_\alpha(F_T, F_C)} + \frac{1 - \delta}{1 - m_\alpha(F_T, F_C)} \right\} \frac{\partial F_T}{\partial \beta} \\
&\quad + (1 - \Delta) \left\{ \frac{\delta}{F_T - C_\alpha(F_T, F_C)} + \frac{1 - \delta}{1 - F_T - F_C - C_\alpha(F_T, F_C)} \right\} \frac{\partial F_T}{\partial \beta}, \\
l_\gamma(\theta, X) &= \Delta \left\{ \frac{\delta}{m_\alpha(F_T, F_C)} + \frac{1 - \delta}{1 - m_\alpha(F_T, F_C)} \right\} \frac{\partial F_T}{\partial \gamma} \\
&\quad + (1 - \Delta) \left\{ \frac{\delta}{F_T - C_\alpha(F_T, F_C)} + \frac{1 - \delta}{1 - F_T - F_C - C_\alpha(F_T, F_C)} \right\} \frac{\partial F_C}{\partial \gamma} + \frac{1 - \Delta}{f_C} \frac{\partial f_C}{\partial \gamma}, \\
l_{\Lambda_T}(\theta, X) &= \Delta \left\{ \frac{\delta}{m_\alpha(F_T, F_C)} + \frac{1 - \delta}{1 - m_\alpha(F_T, F_C)} \right\} \frac{\partial m_\alpha}{\partial F_T} \frac{\partial F_T}{\partial \Lambda_T}, \\
&\quad + (1 - \Delta) \left\{ \frac{\delta}{F_T - C_\alpha(F_T, F_C)} \left[\frac{\partial F_T}{\partial \Lambda_T} - \frac{\partial C_\alpha}{\partial F_T} \frac{\partial F_T}{\partial \Lambda_T} \right], \right. \\
&\quad \left. + \frac{1 - \delta}{1 - F_T - F_C - C_\alpha(F_T, F_C)} \left[-\frac{\partial F_T}{\partial \Lambda_T} - \frac{\partial C_\alpha}{\partial F_T} \frac{\partial F_T}{\partial \Lambda_T} \right] \right\}, \\
l_{\Lambda_C}(\theta, X) &= \Delta \left\{ \frac{\delta}{m_\alpha(F_T, F_C)} + \frac{1 - \delta}{1 - m_\alpha(F_T, F_C)} \right\} \frac{\partial m_\alpha}{\partial F_C} \frac{\partial F_C}{\partial \Lambda_C} + \frac{\Delta}{f_C} \frac{\partial f_C}{\partial \Lambda_C} \\
&\quad + (1 - \Delta) \left\{ \frac{\delta}{F_T - C_\alpha(F_T, F_C)} \left[-\frac{\partial C_\alpha}{\partial F_C} \frac{\partial F_C}{\partial \Lambda_C} \right] + \frac{1 - \delta}{1 - F_T - F_C - C_\alpha(F_T, F_C)} \left[\frac{\partial F_C}{\partial \Lambda_C} - \frac{\partial C_\alpha}{\partial F_C} \frac{\partial F_C}{\partial \Lambda_T} \right] \right\}.
\end{aligned}$$

Denote $p_m = 2p + 2(m + k_n)$. By the calculation of van der Vart and Well-

ner(1996)(p.94), we have

$$\begin{aligned}
& N\left(\epsilon, \mathcal{L}_n, L_1(P_n)\right) \\
& \leq N\left(\frac{\epsilon}{3M}, \mathcal{B}, \|\cdot\|\right) \cdot N\left(\frac{\epsilon}{3M_n}, \mathcal{M}_n^1, L_\infty\right) \cdot N\left(\frac{\epsilon}{3M_n}, \mathcal{M}_n^2, L_\infty\right) \\
& \leq \left(\frac{9M}{\epsilon}\right)^{2p} \cdot \left(\frac{9M_n^2}{\epsilon}\right)^{m+k_n} \cdot \left(\frac{9M_n^2}{\epsilon}\right)^{m+k_n} \\
& \leq KM^{2p}M_n^{4(m+k_n)}\epsilon^{-pm}.
\end{aligned}$$

Applying the inequality (31) of Pollard (1984), we have

$$\sup_{\theta \in \Theta_n} |P_n l(\theta, X) - Pl(\theta, X)| \rightarrow 0 \quad (\text{A.14})$$

almost surely.

Let

$$M(\theta, X) = -l(\theta, X),$$

$$\zeta_{1n} = \sup_{\theta \in \Theta_n} |P_n M(\theta, X) - PM(\theta, X)|,$$

$$\zeta_{2n} = P_n M(\theta_0, X) - PM(\theta_0, X).$$

Denote $K_\epsilon = \{\theta : d(\theta, \theta_0) \geq \epsilon, \theta \in \Theta_n\}$.

$$\begin{aligned}
\inf_{K_\epsilon} PM(\theta, X) &= \inf_{K_\epsilon} \left\{ PM(\theta, X) - P_n M(\theta, X) + P_n M(\theta, X) \right\} \\
&\leq \inf_{K_\epsilon} \left\{ |PM(\theta, X) - P_n M(\theta, X)| + P_n M(\theta, X) \right\}
\end{aligned}$$

$$\leq \zeta_{1n} + \inf_{K_\epsilon} P_n M(\theta, X). \quad (\text{A.15})$$

If $\hat{\theta}_n \in K_\epsilon$, we have

$$\begin{aligned} \inf_{K_\epsilon} P_n M(\theta, X) &= P_n M(\hat{\theta}_n, X) \leq P_n M(\theta_0, X) = P_n M(\theta_0, X) - PM(\theta_0, X) + PM(\theta_0, X) \\ &= \zeta_{2n} + PM(\theta_0, X). \end{aligned} \quad (\text{A.16})$$

By condition (A3), we obtain that $\inf_{K_\epsilon} PM(\theta, X) - PM(\theta_0, X) = \delta_\epsilon > 0$.

By (A.15) and (A.16), we have

$$\inf_{K_\epsilon} PM(\theta, X) \leq \zeta_{1n} + \zeta_{2n} + PM(\theta_0, X) = \zeta_n + PM(\theta_0, X)$$

with $\zeta_n = \zeta_{1n} + \zeta_{2n}$. Hence, we can get that $\zeta_n \geq \delta_\epsilon$. Furthermore, we have $\{\hat{\theta}_n \in K_\epsilon\} \subseteq \{\zeta_n \geq \delta_\epsilon\}$. By (A.14) and Strong Law of Large Numbers, we have $\zeta_{1n} = o(1)$ almost surely, $\zeta_{2n} = o(1)$ almost surely. Therefore, by $\cup_{k=1}^\infty \cap_{n=k}^\infty \{\hat{\theta}_n \in K_\epsilon\} \subseteq \cup_{k=1}^\infty \cap_{n=k}^\infty \{\zeta_n \geq \delta_\epsilon\}$, we complete the proof.

A.3.2 Proof of Theorem 4.2

To show the convergence rate.

For any $\eta > 0$, define the class $\mathcal{F}_\eta = \{l(\theta_{n0}, X) - l(\theta, X) : \theta \in \Theta_n, d(\theta, \theta_{n0}) \leq \eta\}$ with $\theta_{n0} = (\beta_0, \gamma_0, \Lambda_{Tn0}, \Lambda_{Cn0})$. Following the calculation of Shen and Wong (1994) (p.597), we can establish that $\log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) \leq CN \log(\eta/\varepsilon)$ with $N = 2(m + k_n)$. Moreover, some algebraic calculations lead to $\|l(\theta_{n0}, X) - l(\theta, X)\|_2^2 \leq C\eta^2$ for any $l(\theta_{n0}, X) - l(\theta, X) \in \mathcal{F}_\eta$.

Therefore, by Lemma 3.4.2 of van der Vaart and Wellner (1996), we obtain

$$E_P \|n^{1/2}(P_n - P)\|_{\mathcal{F}_\eta} \leq C J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) \left\{ 1 + \frac{J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)}{\eta^2 n^{1/2}} \right\}, \quad (\text{A.17})$$

where $J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) = \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)\}^{1/2} d\varepsilon \leq CN^{1/2}\eta$. The right-hand side of (A.17) yields $\phi_n(\eta) = C(N^{1/2}\eta + N/n^{1/2})$. It is easy to see that $\phi_n(\eta)/\eta$ decreasing in η , and $r_n^2\phi_n(1/r_n) = r_n N^{1/2} + r_n^2 N/n^{1/2} < 2n^{1/2}$, where $r_n = N^{-1/2}n^{1/2} = n^{(1-\nu)/2}$ with $0 < \nu < 0.5$. Hence $n^{(1-\nu)/2}d(\hat{\theta}, \theta_{n0}) = O_P(1)$ by Theorem 3.2.5 of van der Vaart and Wellner (1996). This, together with $d(\theta_{n0}, \theta_0) = O_p(n^{-r\nu})$ (Lemma A1 in Lu et al. 2007) yields that $d(\hat{\theta}, \theta_0) = O_p(n^{-(1-\nu)/2} + n^{-r\nu})$. The choice of $\nu = 1/(1+2r)$ yields the rate of convergence of $d(\hat{\theta}_n, \theta_0) = O_p(n^{-\frac{r}{1+2r}})$.

A.3.3 Proof of Theorem 4.3

Denote V as the linear span of $\Theta_0 - \theta_0$ where θ_0 denote the true value of $\theta = (\theta, \gamma, \psi)$ and Θ_0 denote the true parameter space. Let $l(\theta, W)$ be the log-likelihood for a sample of size one and $\delta_n = (n^{-(1-\nu)/2} + n^{-r\nu})$. For any $\theta \in \{\theta \in \Theta_0 : \|\theta - \theta_0\| = O(\delta_n)\}$, define the first order directional derivative of $l(\theta, X)$ at the direction $v \in V$ as

$$\dot{l}(\theta, X)[v] = \left. \frac{dl(\theta + sv, X)}{ds} \right|_{s=0},$$

and the second order directional derivative as

$$\ddot{l}(\theta, X)[v, \tilde{v}] = \left. \frac{d^2 l(\theta + sv + \tilde{s}\tilde{v}, X)}{d\tilde{s}ds} \right|_{s=0} \Big|_{\tilde{s}=0} = \left. \frac{d\dot{l}(\theta + \tilde{s}\tilde{v}, X)}{d\tilde{s}} \right|_{\tilde{s}=0}.$$

Define the Fisher inner product on the space V as

$$\langle v, \tilde{v} \rangle = P \left\{ \dot{l}(\theta, X)[v] \dot{l}(\theta, X)[\tilde{v}] \right\}$$

and the Fisher norm for $v \in V$ as $\|v\|^{1/2} = \langle v, v \rangle$. Let \bar{V} be the closed linear span of V under the Fisher norm. Then $(\bar{V}, \|\cdot\|)$ is a Hilbert space.

Define the smooth functional of θ as

$$\gamma(\theta) = b'_1 \beta + b'_2 \gamma,$$

where $b = (b'_1, b'_2)'$ is any vector of $2p$ dimension with $\|b\| \leq 1$. For any $v \in V$, we denote

$$\dot{\gamma}(\theta_0)[v] = \left. \frac{d\gamma(\theta_0 + sv)}{ds} \right|_{s=0} = r(v).$$

Note that $\gamma(\theta) - \gamma(\theta_0) = \dot{\gamma}(\theta_0)[\theta - \theta_0]$. It follows by the Riesz representation theorem that, there exists $v^* \in \bar{V}$ such that $\dot{\gamma}(\theta_0)[v] = \langle v^*, v \rangle$ for all $v \in \bar{V}$ and $\|v^*\|^2 = \|\dot{\gamma}(\theta_0)\|$.

Let ε_n be any positive sequence satisfying $\varepsilon_n = o(n^{-1/2})$. For any $v^* \in \Theta_0$, by (A4), Corollary 6.21 of Schumaker (1981) (p.227), there exists $\Pi_n v^* \in \Theta_n$ such that $\|\Pi_n v^* - v^*\| = o(1)$ and $\delta_n \|\Pi_n v^* - v^*\| = o(n^{-1/2})$. Also define $g[\theta - \theta_0, X] = l(\theta, X) -$

$l(\theta_0, X) - \dot{l}(\theta, X)[\theta - \theta_0]$. Then by definition of $\hat{\theta}$, we have

$$\begin{aligned}
0 &\leq P_n[l(\hat{\theta}, W) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, W)] \\
&= (P_n - P)[l(\hat{\theta}, W) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, W)] + P[l(\hat{\theta}, W) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, W)] \\
&= \pm \varepsilon_n P_n \dot{l}(\theta, W)[\Pi_n v^*] + (P_n - P) \left\{ g[\hat{\theta} - \theta_0, W] - g[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, W] \right\} \\
&\quad + P \left\{ g[\hat{\theta} - \theta_0, W] - g[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, W] \right\} \\
&= \mp \varepsilon_n P_n \dot{l}(\theta; W)[v^*] \pm \varepsilon_n P_n \dot{l}(\theta, W)[\Pi_n v^* - v^*] + (P_n - P) \left\{ g[\hat{\theta} - \theta_0, W] \right. \\
&\quad \left. - g[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, W] \right\} + P \left\{ g[\hat{\theta} - \theta_0, W] - g[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, W] \right\} \\
&:= \mp \varepsilon_n P_n \dot{l}(\theta, W)[v^*] + I_1 + I_2 + I_3.
\end{aligned}$$

For I_1 , it follows from Conditions (A1)-(A2), Chebyshev inequality and $\|\Pi_n v^* - v^*\| = o(1)$ that $I_1 = \varepsilon_n \times o_p(n^{-1/2})$.

For I_2 , we have

$$\begin{aligned}
I_2 &= (P_n - P) \left\{ l(\hat{\theta}, W) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, W) \pm \varepsilon_n \dot{l}(\theta_0, W)[\Pi_n v^*] \right\} \\
&= \mp \varepsilon_n (P_n - P) \left\{ \dot{l}(\tilde{\theta}, W) - \dot{l}(\theta_0, W)[\Pi_n v^*] \right\},
\end{aligned}$$

where $\tilde{\theta}$ lies between $\hat{\theta}$ and $\hat{\theta} \pm \varepsilon_n \Pi_n v^*$. By Theorem 2.8.3 in of van der Vaart and Wellner (1996), we know that $\{\dot{l}(\theta; W)[\Pi_n v^*] : \|\theta - \theta_0\| = O(\delta_n)\}$ is Donsker class. Therefore, by Theorem 2.11.23 of van der Vaart and Wellner (1996), we have $I_2 = \varepsilon_n \times o_p(n^{-1/2})$.

For I_3 , note that

$$\begin{aligned}
P(g[\theta - \theta_0, W]) &= P\{l(\theta, W) - l(\theta_0, W) - \dot{l}(\theta_0, W[\theta - \theta_0])\} \\
&= 2^{-1}P\{\ddot{l}(\tilde{\theta}, W)[\theta - \theta_0, \theta - \theta_0] - \ddot{l}(\theta_0, W)[\theta - \theta_0, \theta - \theta_0]\} \\
&\quad + 2^{-1}P\{\ddot{l}(\theta_0, W)[\theta - \theta_0, \theta - \theta_0]\} \\
&= 2^{-1}P\{\ddot{l}(\theta_0, W)[\theta - \theta_0, \theta - \theta_0]\} + \varepsilon_n \times o_p(n^{-1/2})
\end{aligned}$$

where $\tilde{\theta}$ lies between θ_0 and θ and the last equation is due to Taylor expansion, (A1)-(A2) and $r > 2$. Therefore,

$$\begin{aligned}
I_3 &= -2^{-1}\{\|\hat{\theta} - \theta_0\|^2 - \|\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0\|^2\} + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm \varepsilon_n \langle \hat{\theta} - \theta_0, \Pi_n v^* \rangle + 2^{-1}\|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + 2^{-1}\|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + \varepsilon_n \times o_p(n^{-1/2})
\end{aligned}$$

where the last equality holds since $\delta_n \|\Pi_n v^* - v^*\| = o(n^{-1/2})$, Cauchy-Schwartz inequality, and $\|\Pi_n v^*\|^2 \rightarrow \|v^*\|^2$. Combing the above facts, together with $P\dot{l}(\theta_0, W[v^*]) = 0$, we can establish that

$$\begin{aligned}
0 &\leq P_n\{l(\hat{\theta}, W) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, W)\} \\
&= \mp \varepsilon_n P_n \dot{l}(\theta_0, W)[v^*] \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \mp \varepsilon_n (P_n - P)\{\dot{l}(\theta_0, W)[v^*]\} \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + \varepsilon_n \times o_p(n^{-1/2}).
\end{aligned}$$

Therefore, we obtain $\sqrt{n} \langle \hat{\theta} - \theta_0, v^* \rangle = \sqrt{n}(P_n - P)\{\dot{l}(\theta_0, W)[v^*]\} + o_p(1) \rightarrow$

$N(0, \|v^*\|^2)$, where the asymptotic normality is guaranteed by Central limits Theorem and the asymptotic variance being equal to $\|v^*\|^2 = \|\dot{l}(\theta_0, W)[v^*]\|^2$. This implies $n^{1/2}(\gamma(\hat{\theta}) - \gamma(\theta_0)) = n^{1/2} \langle \hat{\theta} - \theta_0, v^* \rangle + o_p(1) \rightarrow N(0, \|v^*\|^2)$ in distribution. The semiparametric efficiency can be established by applying the result of Bickel and Kwon (2001) or Theorem 4 in Shen (1997).

For each component $\vartheta_q, q = 1, 2, \dots, 2p$, we denote by $\psi_q^* = (b_{1q}^*, b_{2q}^*)$ the solution to

$$\inf_{\psi_q^*} E \left\{ l_{\vartheta} \cdot e_q - l_{b_1^*} \cdot b_{1q}^* - l_{b_2^*} \cdot b_{2q}^* \right\}^2. \quad (\text{A.18})$$

where $l_{\vartheta} = (l'_{\beta}, l'_{\gamma})'$, $l_{b_1^*}$ and $l_{b_2^*}$ can be calculated as $l_{\Lambda_T}(\theta, X)$ and $l_{\Lambda_C}(\theta, X)$ in the proof of Theorem 4.1. Now let $\psi^* = (\psi_1^*, \dots, \psi_q^*)$. By the calculations of Chen et al. (2006), we have $\|v^*\|^2 = \|\dot{\gamma}(\theta_0)\|^2 = \sup_{v \in \bar{V}: \|v\| > 0} \frac{|\dot{\gamma}(\theta_0)[v]|}{\|v\|} = b' \Sigma b$, where $\Sigma = E(S_{\vartheta} S'_{\vartheta})$, $S_{\vartheta} = \{l_{\vartheta} - l_{b_1^*} b_1^* - l_{b_2^*} b_2^*\}$. Now, since $b'((\hat{\beta} - \beta_0)', (\hat{\gamma} - \gamma_0)') = \langle \hat{\theta} - \theta_0, v^* \rangle$, the conclusion of the theorem follows by the Cramér-Wold device.

A.4 Proofs of the Asymptotic Properties in Chapter 5

Assumptions

(A1) Z has a bounded support in R^p where p is the dimension of Z .

(A2) $C(\cdot, \cdot)$ has a bounded first order partial derivatives for C on $(0, 1)^2$. $\frac{\partial C(u, v)}{\partial u}$ and $\frac{\partial C(u, v)}{\partial v}$ be Lipschitz.

(A3) For any positive number ϵ , it holds that

$$\sup_{d(\theta, \theta_0) > \epsilon} Pl(\theta, X) < Pl(\theta_0, X).$$

(A4) The m th derivative of $\Lambda_k(\cdot)$, denoted by $\Lambda_k^{(m)}(\cdot)$, is Holder continuous with the exponent η , i.e., $|\Lambda_k^{(m)}(t_1) - \Lambda_k^{(m)}(t_2)| \leq M|t_1 - t_2|^\eta$ for some $\eta \in (0, 1]$ and for all $t_1, t_2 \in (l, u)$, $k = 1, 2$ and M is some constant. Denote $r = m + \eta$.

(A5) $E(S_\vartheta S_\vartheta')$ is finite and positive definite with $\vartheta = (\beta', \gamma')$ where S_ϑ defined in the proof of Theorem 5.3.

A.4.1 Proof of Theorem 5.1

To establish the consistency using empirical process theory, we consider a class of functions \mathcal{L}_n defined by

$$\{l(\theta, X) : \theta \in \Theta_n\}.$$

For any $\theta^1 = (\beta^1, \gamma^1, \Lambda_1^1, \Lambda_2^1)$, $\theta^2 = (\beta^2, \gamma^2, \Lambda_1^2, \Lambda_2^2) \in \Theta_n$, by Taylor's expansion, we have

$$\begin{aligned} & |l(\theta^1, X) - l(\theta^2, X)| \\ &= |l_\beta(\tilde{\theta}, X)(\beta^1 - \beta^2)| + |l_\gamma(\theta, X)(\gamma^1 - \gamma^2)| \\ &\quad + |l_{\Lambda_1}(\tilde{\theta}, X)(\Lambda_1^1 - \Lambda_1^2)| + |l_{\Lambda_2}(\theta, X)(\Lambda_2^1 - \Lambda_2^2)| \\ &\leq K(\|\beta^1 - \beta^2\| + \|\gamma^1 - \gamma^2\|) + K\|\Lambda_1^1 - \Lambda_1^2\|_\infty + K\|\Lambda_2^1 - \Lambda_2^2\|_\infty \\ &\leq K(\|\beta^1 - \beta^2\| + \|\gamma^1 - \gamma^2\| + \|\Lambda_1^1 - \Lambda_1^2\|_\infty + \|\Lambda_2^1 - \Lambda_2^2\|_\infty) \end{aligned}$$

for some constant K . Denote $p_m = 2p + 2(m + k_n)$. Then we have

$$\begin{aligned}
& N\left(\epsilon, \mathcal{L}_n, L_1(P_n)\right) \\
& \leq N\left(\frac{\epsilon}{3M}, \mathcal{B}, \|\cdot\|\right) \cdot N\left(\frac{\epsilon}{3M_n}, \mathcal{M}_n^1, L_\infty\right) \cdot N\left(\frac{\epsilon}{3M_n}, \mathcal{M}_n^2, L_\infty\right) \\
& \leq \left(\frac{9M}{\epsilon}\right)^{2p} \cdot \left(\frac{9M_n^2}{\epsilon}\right)^{m+k_n} \cdot \left(\frac{9M_n^2}{\epsilon}\right)^{m+k_n} \\
& \leq KM^{2p}M_n^{4(m+k_n)}\epsilon^{-p_m}.
\end{aligned}$$

where the second inequality is based on calculation of van der Vart and Wellner(1996)(p.94).

Applying the inequality (31) in Pollard (1984) (p.31), we have

$$\sup_{\theta \in \Theta_n} |P_n l(\theta, X) - Pl(\theta, X)| \rightarrow 0 \quad (\text{A.15})$$

almost surely.

Let $M(\theta, X) = -l(\theta, X)$ and

$$\zeta_{1n} = \sup_{\theta \in \Theta_n} |P_n M(\theta, X) - PM(\theta, X)|, \quad \zeta_{2n} = P_n M(\theta_0, X) - PM(\theta_0, X).$$

Denote $K_\epsilon = \{\theta : d(\theta, \theta_0) \geq \epsilon, \theta \in \Theta_n\}$.

$$\begin{aligned}
\inf_{K_\epsilon} PM(\theta, X) &= \inf_{K_\epsilon} \left\{ PM(\theta, X) - P_n M(\theta, X) + P_n M(\theta, X) \right\} \\
&\leq \zeta_{1n} + \inf_{K_\epsilon} P_n M(\theta, X).
\end{aligned} \quad (\text{A.16})$$

If $\hat{\theta}_n \in K_\epsilon$, we have

$$\inf_{K_\epsilon} P_n M(\theta, X) = P_n M(\hat{\theta}_n, X) \leq P_n M(\theta_0, X) = \zeta_{2n} + PM(\theta_0, X). \quad (\text{A.17})$$

By condition (A3), we obtain that $\inf_{K_\epsilon} PM(\theta, X) - PM(\theta_0, X) = \delta_\epsilon > 0$.

By (A.16) and (A.17), we have

$$\inf_{K_\epsilon} PM(\theta, X) \leq \zeta_{1n} + \zeta_{2n} + PM(\theta_0, X) = \zeta_n + PM(\theta_0, X)$$

with $\zeta_n = \zeta_{1n} + \zeta_{2n}$. Hence, we can get that $\zeta_n \geq \delta_\epsilon$. Furthermore, we have $\{\hat{\theta}_n \in K_\epsilon\} \subseteq \{\zeta_n \geq \delta_\epsilon\}$. By (A.15) and Strong Law of Large Numbers, we have $\zeta_{1n} = o(1)$ almost surely, $\zeta_{2n} = o(1)$ almost surely. Therefore, by $\cup_{k=1}^\infty \cap_{n=k}^\infty \{\hat{\theta}_n \in K_\epsilon\} \subseteq \cup_{k=1}^\infty \cap_{n=k}^\infty \{\zeta_n \geq \delta_\epsilon\}$, we complete the proof.

A.4.2 Proof of Theorem 5.2

To show the convergence rate.

For any $\eta > 0$, define the class $\mathcal{F}_\eta = \{l(\theta_{n0}, X) - l(\theta, X) : \theta \in \Theta_n, d(\theta, \theta_{n0}) \leq \eta\}$ with $\theta_{n0} = (\beta_0, \gamma_0, \Lambda_{1n0}, \Lambda_{2n0})$. Following the calculation of Shen and Wong (1994) (p.597), we can establish that $\log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) \leq CN \log(\eta/\varepsilon)$ with $N = 2(m + k_n)$. Moreover, some algebraic calculations lead to $\|l(\theta_{n0}, X) - l(\theta, X)\|_2^2 \leq C\eta^2$ for any $l(\theta_{n0}, X) - l(\theta, X) \in \mathcal{F}_\eta$.

Therefore, by Lemma 3.4.2 of van der Vaart and Wellner (1996), we obtain

$$E_P \|n^{1/2}(P_n - P)\|_{\mathcal{F}_\eta} \leq C J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) \left\{ 1 + \frac{J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)}{\eta^2 n^{1/2}} \right\}, \quad (\text{A.18})$$

where $J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) = \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)\}^{1/2} d\varepsilon \leq CN^{1/2}\eta$. The right-hand side of (A.18) yields $\phi_n(\eta) = C(N^{1/2}\eta + N/n^{1/2})$. It is easy to see that $\phi_n(\eta)/\eta$ decreasing in η , and $r_n^2 \phi_n(1/r_n) = r_n N^{1/2} + r_n^2 N/n^{1/2} < 2n^{1/2}$, where $r_n = N^{-1/2} n^{1/2} =$

$n^{(1-\nu)/2}$ with $0 < \nu < 0.5$. Hence $n^{(1-\nu)/2}d(\hat{\theta}, \theta_{n0}) = O_P(1)$ by Theorem 3.2.5 of van der Vaart and Wellner (1996). This, together with $d(\theta_{n0}, \theta_0) = O_p(n^{-r\nu})$ (Lemma A1 in Lu et al. 2007) yields that $d(\hat{\theta}, \theta_0) = O_p(n^{-(1-\nu)/2} + n^{-r\nu})$. The choice of $\nu = 1/(1 + 2r)$ yields the rate of convergence of $d(\hat{\theta}_n, \theta_0) = O_p(n^{-\frac{r}{1+2r}})$.

A.4.3 Proof of Theorem 5.3

Denote V as the linear span of $\Theta_0 - \theta_0$ where θ_0 denote the true value of $\theta = (\theta, \gamma, \psi)$ and Θ_0 denote the true parameter space. Let $l(\theta, W)$ be the log-likelihood for a sample of size one and $\delta_n = (n^{-(1-\nu)/2} + n^{-r\nu})$. For any $\theta \in \{\theta \in \Theta_0 : \|\theta - \theta_0\| = O(\delta_n)\}$, define the first order directional derivative of $l(\theta, X)$ at the direction $v \in V$ as

$$\dot{l}(\theta, X) = \left. \frac{dl(\theta + sv, X)}{ds} \right|_{s=0}, \quad (\text{A.19})$$

and the second order directional derivative as

$$\ddot{l}(\theta, X) = \left. \frac{d^2l(\theta + sv + \tilde{s}\tilde{v}, X)}{d\tilde{s}ds} \right|_{s=0} \Big|_{\tilde{s}=0} = \left. \frac{d\dot{l}(\theta + \tilde{s}\tilde{v}, X)}{d\tilde{s}} \right|_{\tilde{s}=0}. \quad (\text{A.20})$$

Define the Fisher inner product on the space V as

$$\langle v, \tilde{v} \rangle = P \left\{ \dot{l}(\theta, X)[v] \dot{l}(\theta, X)[\tilde{v}] \right\} \quad (\text{A.21})$$

and the Fisher norm for $v \in V$ as $\|v\|^{1/2} = \langle v, v \rangle$. Let \bar{V} be the closed linear span of V under the Fisher norm. Then $(\bar{V}, \|\cdot\|)$ is a Hilbert space.

Define the smooth functional of θ as

$$\gamma(\theta) = b'_1\beta + b'_2\gamma, \quad (\text{A.22})$$

where $b = (b'_1, b'_2)'$ is any vector of $2p$ dimension with $\|b\| \leq 1$. For any $v \in V$, we denote

$$\dot{\gamma}(\theta_0)[v] = \left. \frac{d\gamma(\theta_0 + sv)}{ds} \right|_{s=0} = r(v) \quad (\text{A.23})$$

whenever the right hand-side limit is well defined.

Note that $\gamma(\theta) - \gamma(\theta_0) = \dot{\gamma}(\theta_0)[\theta - \theta_0]$. It follows by the Riesz representation theorem that, there exists $v^* \in \bar{V}$ such that $\dot{\gamma}(\theta_0)[v] = \langle v^*, v \rangle$ for all $v \in \bar{V}$ and $\|v^*\|^2 = \|\dot{\gamma}(\theta_0)\|$.

Let ε_n be any positive sequence satisfying $\varepsilon_n = o(n^{-1/2})$. For any $v^* \in \Theta_0$, by (A4), Corollary 6.21 of Schumaker (1981) (p.227), there exists $\Pi_n v^* \in \Theta_n$ such that $\|\Pi_n v^* - v^*\| = o(1)$ and $\delta_n \|\Pi_n v^* - v^*\| = o(n^{-1/2})$. Also define $r[\theta - \theta_0, X] = l(\theta, X) - l(\theta_0, X) - \dot{l}(\theta, X)[\theta - \theta_0]$. Then by definition of $\hat{\theta}$, we have

$$\begin{aligned} 0 &\leq P_n[l(\hat{\theta}, X) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, X)] \\ &= (P_n - P)[l(\hat{\theta}, X) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, X)] + P[l(\hat{\theta}, X) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, X)] \\ &= \pm \varepsilon_n P_n \dot{l}(\theta, X)[\Pi_n v^*] + (P_n - P) \left\{ r[\theta - \theta_0, X] - r[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, X] \right\} \\ &\quad + P \left\{ r[\theta - \theta_0, X] - r[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, X] \right\} \\ &= \mp \varepsilon_n P_n \dot{l}(\theta, X)[v^*] \pm \varepsilon_n P_n \dot{l}(\theta, X)[\Pi_n v^* - v^*] + (P_n - P) \left\{ r[\theta - \theta_0, X] \right. \\ &\quad \left. - r[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, X] \right\} + P \left\{ r[\theta - \theta_0, X] - r[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, X] \right\} \\ &:= \mp \varepsilon_n P_n \dot{l}(\theta, X)[v^*] + I_1 + I_2 + I_3. \end{aligned}$$

By (A1)-(A2) and Chebyshev inequality, independent and identical distribution data, and $\|\Pi_n v^* - v^*\| = o_p(1)$, we have $I_1 = o_p(n^{-1/2})$.

For I_2 , we have

$$\begin{aligned} I_2 &= (P_n - P) \left\{ l(\hat{\theta}, X) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, X) \pm \varepsilon_n \dot{l}(\theta_0, X) [\Pi_n v^*] \right\} \\ &= \mp \varepsilon_n (P_n - P) \left\{ \dot{l}(\tilde{\theta}, X) - \dot{l}(\theta_0, X) [\Pi_n v^*] \right\}, \end{aligned}$$

where $\tilde{\theta}$ lies between $\hat{\theta}$ and $\hat{\theta} \pm \varepsilon_n \Pi_n v^*$. By Theorem 2.8.3 in of van der Vaart and Wellner (1996), we know that $\{\dot{l}(\theta, X) [\Pi_n v^*] : \|\theta - \theta_0\| = O(\delta_n)\}$ is Donsker class. Therefore, by Theorem 2.11.23 of van der Vaart and Wellner (1996), we have $I_2 = \varepsilon_n \times o_p(n^{-1/2})$.

Note that

$$\begin{aligned} P(r[\theta - \theta_0, X]) &= P\{l(\theta, X) - l(\theta_0, X) - \dot{l}(\theta_0, X)[\theta - \theta_0]\} \\ &= 2^{-1} P\{\ddot{l}(\tilde{\theta}, X)[\theta - \theta_0, \theta - \theta_0] - \ddot{l}(\theta_0, X)[\theta - \theta_0, \theta - \theta_0]\} \\ &\quad + 2^{-1} P\{\ddot{l}(\theta_0, X)[\theta - \theta_0, \theta - \theta_0]\} \\ &= 2^{-1} P\{\ddot{l}(\theta_0, X)[\theta - \theta_0, \theta - \theta_0]\} + \varepsilon_n \times o_p(n^{-1/2}) \end{aligned}$$

where $\tilde{\theta}$ lies between θ_0 and θ and the last equation is due to Taylor expansion,

(A1)-(A2) and $r > 2$. Therefore,

$$\begin{aligned}
I_3 &= -2^{-1}\{\|\hat{\theta} - \theta_0\|^2 - \|\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0\|^2\} + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm \varepsilon_n \langle \hat{\theta} - \theta_0, \Pi_n v^* \rangle + 2^{-1}\|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + 2^{-1}\|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + \varepsilon_n \times o_p(n^{-1/2})
\end{aligned}$$

where the last equality holds since $\delta_n \|\Pi_n v^* - v^*\| = o(n^{-1/2})$, Cauchy-Schwartz inequality, and $\|\Pi_n v^*\|^2 \rightarrow \|v^*\|^2$. Combing the above facts, together with $P\dot{l}(\theta_0, X)[v^*] = 0$, we can establish that

$$\begin{aligned}
0 &\leq P_n\{l(\hat{\theta}, X) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, X)\} \\
&= \mp \varepsilon_n P_n \dot{l}(\theta_0, X)[v^*] \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \mp \varepsilon_n (P_n - P)\{\dot{l}(\theta_0, X)[v^*]\} \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + \varepsilon_n \times o_p(n^{-1/2}).
\end{aligned}$$

Therefore, we obtain $\sqrt{n} \langle \hat{\theta} - \theta_0, v^* \rangle = \sqrt{n}(P_n - P)\{\dot{l}(\theta_0, X)[v^*]\} + o_p(1) \rightarrow N(0, \|v^*\|^2)$, where the asymptotic normality is guaranteed by Central limits Theorem (Theorem 4.1 in Chapter 14 of Shorack, 2000) and the asymptotic variance being equal to $\|v^*\|^2 = \|\dot{l}(\theta_0, X)\|^2$. This, together with A5 imply $n^{1/2}(\gamma(\hat{\theta}) - \gamma(\theta_0)) = n^{1/2} \langle \hat{\theta} - \theta_0, v^* \rangle + o_p(1) \rightarrow N(0, \|v^*\|^2)$ in distribution. The semiparametric efficiency can be established by applying the result of Bickel and Kwon (2001) or Theorem 4 in Shen (1997).

For each component ϑ_q , $q = 1, 2, \dots, 2p$, we denote by $\psi_q^* = (b_{1q}^*, b_{2q}^*)$ the solution

to

$$\inf_{\psi_q^*} E \left\{ l_\vartheta \cdot e_q - l_{b_1^*} \cdot b_{1q}^* - l_{b_2^*} \cdot b_{2q}^* \right\}^2. \quad (\text{A.24})$$

where $l_\vartheta = (l'_\beta, l'_\gamma)'$, $l_{b_1^*}$ and $l_{b_2^*}$ can be calculated as $l_{\Lambda_1}(\theta, X)$ and $l_{\Lambda_2}(\theta, X)$ in the proof of Theorem 5.1. Now let $\psi^* = (\psi_1^*, \dots, \psi_q^*)$. By the calculations of Chen et al. (2006), we have $\|v^*\|^2 = \|\dot{\gamma}(\theta_0)\| = \sup_{v \in \bar{V}: \|v\| > 0} \frac{|\dot{\gamma}(\theta_0)[v]|}{\|v\|} = b' \Sigma b$, where $\Sigma = E(S_\vartheta S'_\vartheta)$, $S_\vartheta = \{l_\vartheta - l_{b_1^*} b_1^* - l_{b_2^*} b_2^*\}$. Now, since $b'((\hat{\beta} - \beta_0)', (\hat{\gamma} - \gamma_0)') = \langle \hat{\theta} - \theta_0, v^* \rangle$, the conclusion of the theorem follows by the Cramér-Wold device.

Bibliography

- [1] Andersen, P., Gill, R. (1982). Cox's regression model for counting process: a large sample study. *Ann. Statist.* **10**,1100-1120.
- [2] Andersen, P. and Ronn, B. (1995). A nonparametric test for comparing two samples where all observations are either left- or right-censored. *Biometrics* **51**, 323-329.
- [3] Betensky, R. and Finkelstein, D. (2002). Testing for dependence between failure time and visit compliance with interval-censored data. *Biometrics* **58**, 58-63.
- [4] Betensky, R., Rabinowitz, D. and Tsiatis, A. (2001). Computationally simple accelerated failure time regression for interval censored data. *Biometrika*, **88**, 703-711.
- [5] Bickel, P. and J. Kwon (2001). Inference for semiparametric models: some questions and an answer. *Statistica Sinica*, **11**, 863-960.
- [6] Chen, C., Lu, T., Chen, M., Hsu, C. (2012). Semiparametric transformation models for current status data with informative censoring. *Biometrical Journal* **54**, 641-656.
- [7] Chen, D., Sun, J., Peace K.E. (2012). *Interval-censored Time-to Event Data: Methods and Applications*. CRC Press LLC.
- [8] Chen, K., Jin, Z., Ying, Z. 2002. Semiparametric analysis of transformation models with censored data. *Biometrika* **89**: 659-668.

- [9] Chen, L. & Sun, J. (2009). A multiple imputation approach to the analysis of current status data with the additive hazards model. *Commun. Statist. Theor. Meth.* **38**, 1009-1018.
- [10] Chen, L., Tong, X. & Sun, J. (2009). A frailty model approach for regression analysis of multivariate current status data. *Statist. Med.* **28**, 3424-3436.
- [11] Chen, M., Tong, X., Sun, J. (2007). The proportional odds model for multivariate interval-censored failure time data. *Statist. Med.* **26**: 5147-5161.
- [12] Chen, S. (1994). Empirical likelihood confidence intervals for linear regression coefficients. *J. Multivariate Anal.* **49**:24-40.
- [13] Chen, X., Fan, Y. and Tsyrennikov, V. (2006). Efficient Estimation of Semiparametric Multivariate Copula Models. *Journal of the American Statistical Association*, **101**, 1228–1240.
- [14] Chen, Y.-H. (2010). Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *Journal of the Royal Statistical Society, Ser B*, **72**, 235–251.
- [15] Cox, D. (1972). Regression models and life tables (with discussion). *J. R. Stat. Soc. B* **34**: 187-220.
- [16] De Gruttola, V., Lagakos S. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* **45**, 1-11.
- [17] Dewanji, A. and kalbfleisch, J. (1986). Nonparametric methods for survival/sacrifice experiments. *Biometrics* **42**, 325-341.
- [18] Dinse, G. (1991). Constant risk differences in the analysis of animal tumorigenicity data. *Biometrics* **47**, 681-700.
- [19] Dunson, D. and Dinse, G. (2002). Bayesian models for multivariate current status data with informative censoring. *Biometrics* **58**, 79-88.

- [20] Fan, J., Lin, H. & Zhou, Y. (2006). Local partial-likelihood estimation for lifetime data. *Ann. Statist.* **34**, 290-325.
- [21] Finkelstein, D. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics.* 42:845-854.
- [22] Finkelstein, D., Goggins, W. and Schoenfeld, D. (2002). Analysis of failure time data with dependent interval censoring. *Biometrics* **58**, 298-304.
- [23] Fleming, T. & Harrington, D. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.
- [24] Foutz, R. (1977). On the Unique consistent solution to the likelihood equations. *J. Amer. Statist. Assoc.* **72**, 147-148.
- [25] Goggins, W. and Finkelstein, D. (2000). A proportional hazards model for multivariate interval-censored failure time data. *Biometrics*, **56**, 940-943.
- [26] Goggins, W., Finkelstein, D., Schoenfeld, D., Zaslavsky, A. (1998). A Markov chain Monte Carlo EM algorithm for analyzing interval censored data under the Cox proportional hazards model. *Biometrics***54**, 1498-1507.
- [27] Gómez, G., Espinal, A. and Lagakos, S. (2003). Inference for a linear regression model with an interval-censored covariate. *Stat. Med.* **22**, 409-425.
- [28] Groeneboom, P., Wellner, J.A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser Verlag, Basel, Switzerland.
- [29] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* 19: 293-325.
- [30] Hu, C. & Lin, D. (2002) Cox regression with covariate measurement error. *Scand. J. Statist.* **29**, 637-655.

- [31] Huang, J. (1995). Maximum likelihood estimation for proportional odds regression model with current status data. *Analysis of censored data, IMS Lecture Notes - Monograph Series 27*: 129-146.
- [32] Huang, J. (1996) Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* 24:540-568.
- [33] Huang, J. and Rossini, A. (1997). Sieve estimation for the proportional odds failure-time regression model with interval censoring. *Journal of the American Statistical Association***92**, 960-967.
- [34] Huang, J. and Wellner, J. A. (1997). Interval censored survival data: A review of recent progress. In *Proceedings of the first Seattle Symposium in Biostatistics: Survival Analysis*. New York: Springer-Verlag, 123-169.
- [35] Huang, X. and Zhang, N. (2008). Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach. *Biometrics* **64**, 1090–1099.
- [36] Huang, Y. & Wang, C.Y. (2000). Cox regression with accurate covariates uncertain-able??a nonparametric approach. *J. Amer. Statist. Assoc.* **95**, 1209-1219.
- [37] Jewell, N. & van der Laan, M. (1995). Generalizations of current status data with applications. *Lifetime Data Anal.* **1**, 101-110.
- [38] Jing, B.Y., Yuan, J., Zhou W. (2008). Empirical likelihood for non-degenerate U-statistics. *Statist. Probab. Lett.* 78: 599-607.
- [39] Kalbfleisch, J., Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York.
- [40] Lagakos, S. and Louis, T. (1988). Use of tumour lethality to interpret tumorigenicity experiments lacking cause-of-death data. *Applied Statistics* **37**, 169-179.
- [41] Lawless, J., Babineau, D. (2006) Models of interval censoring and simulation-based inference for lifetime distributions. *Biometrika***93**, 671-686

- [42] Lin, D., Oakes, D. & Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika* **85**, 289-298.
- [43] Lin, D. & Ying, Z. (1993). Cox regression with incomplete covariate measurements. *J. Amer. Statist. Assoc.* **88**, 1341-1349.
- [44] Lin, D. & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.
- [45] Lindsey, J. and Ryan, L. (1993). A three-state multiplicative model for rodent tumorigenicity experiments. *Applied Statistics* **42**, 283-300.
- [46] Liu, Y., Zhou, H. & Cai, J. (2009). Estimated pseudo-partial-likelihood method for correlated failure time data with auxiliary covariates. *Biometrics* **65**, 1184-1193.
- [47] Lu, M., Zhang, Y., and Huang, J. (2007), Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94**(3), 705–718.
- [48] Martinussen, T. & Scheike, T. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika* **89**, 649-658.
- [49] Nadaraya, E. (1964). On estimating regression, *Theor. Prob. Appl.* **10**, 186-190.
- [50] Nelsen, R. (2006). *An introduction to copulas (Second Edition)*. Springer, New York.
- [51] Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*. 75:237-249.
- [52] Owen, A. (1990). Empirical likelihood and confidence regions. *Ann. Statist.* 18:90-120.
- [53] Pettitt, A. (1982). Inference for the linear model using a likelihood based on ranks. *J. R. Statist. Soc. B* 44:234-243.
- [54] Pettitt, A. (1984). Proportional odds model for survival data and estimates using ranks. *Appl. Statist.* 33:169-175.
- [55] Pollard, D. (1984), *Convergence of Stochastic Processes*. Springer, New York.

- [56] Prentice, R. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69**, 331-342.
- [57] Qin, J., Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* 22:300-325.
- [58] Ramsay, J. (1988). Monotone regression splines in action. *Statist. Sci.* **3**, 425–41.
- [59] Rao, J., Scott, A. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *J. Amer. Statist. Assoc.* 76:221-230.
- [60] Rossini, A. & Tsiatis, A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *J. Amer. Statist. Assoc.* **91**, 713-721.
- [61] Satten, G. (1996). Rank-based inference in the proportional hazards model for interval censored data. *Biometrika***83**, 355-370
- [62] Schumaker, L. (1981). *Spline Function: Basic Theory*. New York: Wiley.
- [63] Seaman, S. and Bird, S. (2001). Proportional hazards model for interval-censored failure times and time dependent covariates: application to hazard of HIV infection of injecting drug users in prison. *Stat. Med.* **20**, 1855-1870.
- [64] Shen, X. and Wong, W. (1994). Convergence rate of sieve estimates. *Annals of Statistics*, **22**, 580–615.
- [65] Shen, X. (1997). On methods of sieves and penalization. *Annals of Statistics*, **25**, 2555–2591.
- [66] Stone, C. (1980). Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, **8**, 1348–1360.
- [67] Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer: New York.

- [68] Sun, J. & Shen, J. (2009). Efficient estimation for the proportional hazards model with competing risks and current status data. *Canad. J. Statist.* **37**, 592-606.
- [69] Sun, J. & Sun, L. (2005). Semiparametric linear transformation models for current status data. *The Canad. J. Statist.* **33**, 85-96.
- [70] Titman A.C. (2013). A pool-adjacent-violators type algorithm for non-parametric estimation of current status data with dependent censoring. *Lifetime Data Analysis*, to appear.
- [71] van der Lann, M. and Robins, J. (1998). Locally efficient estimation with current status data and time-dependent covariates. *Journal of the American Statistical Association* **93**, 693-701.
- [72] van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer, Berlin.
- [73] Wang, C., Sun, J., Sun, L., Zhou, J. and Wang, D. (2012). Nonparametric Estimation of Current Status Data with Dependent Censoring. *Lifetime Data Analysis*, **18**, 434–445.
- [74] Wang, L., Sun, J., Tong, X. (2010). Regression analysis of case II interval-censored failure time data with additive hazards model. *Statistica Sinica*. 20:1709-1723.
- [75] Wang, N., Lin, X., Gutierrez, R. & Carroll, R. (1998). Bias analysis and SIMEX approach in generalized linear mixed measurement error models. *J. Amer. Statist. Assoc.* **93**, 249-261.
- [76] Wang, Q., Rao, J. (2001). Empirical likelihood for linear regression models under imputation for missing responses. *Canad. J. Statist.* 29:597-608.
- [77] Watson, G. (1964). Smooth regression analysis, *Sankhya, A* **26**, 359-372.
- [78] Yu, W., Sun, Y., Zheng, M. (2011). Empirical likelihood method for linear transformation modes. *Ann. Inst. Statist. math.* 63:331-346.

- [79] Zeng, D., Cai, J. and Shen, Y. (2006). Semiparametric additive risks model for interval-censored data. *Statistica Sinica***16**, 287-302.
- [80] Zhang, Y., Hua, L., and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scandinavian Journal of Statistics*, **37**, 338–354.
- [81] Zhang, Z., Sun, J. and Sun, L. (2005). Statistical Analysis of current status data with informative observation times. *Statistics in Medicine* **24**, 1399-1407.
- [82] Zhang, Z., Sun, L., Sun, J. and Finkelstein D. (2007). Regression analysis of failure time data with informative interval-censoring. *Statistics in Medicine***26**: 2533-2546.
- [83] Zhang, Z., Sun, L., Zhao, X., Sun, J. (2005). Regression analysis of interval-censored failure time data with linear transformation models. *Canad. J. Statist.* 33:61-70.
- [84] Zhao, Y. (2010). Semiparametric inference for transformation models via empirical likelihood. *J. Multivariate Anal.* 101:1846-1858.
- [85] Zheng, M., Klein, J.P. (1995). Estimates of marginal survival for dependent competing risk based on an assumed copula. *Biometrika*, **82**, 127–138.
- [86] Zhou, H., & Pepe, M.S. (1995). Auxiliary covariate data in failure time regression. *Biometrika* **82**, 139-149.
- [87] Zhou, H. & Wang C.Y. (2000). Failure time regression with continuous covariates measured with error. *J. R. Statist. Soc. B* **62**, 657-665.
- [88] Zhu, L., Tong, X. and Sun, J. (2008). A transformation approach to the analysis of interval-censored failure time data. *Lifetime Data Analysis***14**, 167-178.

VITA

Ling Ma was born in 1988 in Taixing, Jiangsu Province, China. She received her B.S. in statistics from University of Science and Technology of China in 2010. Then she joined the Ph.D. program in the Department of Statistics at the University of Missouri Columbia in August 2010. She will receive her Ph.D. in Statistics in July of 2014.