國 立 交 通 大 學

統計學研究所

碩 士 論 文

線性轉換模型之統計推論

Statistical Inference for Linear Transformation Models

研 究 生：林泰佐

指導教授：王維菁博士

中 華 民 國 九 十 七 年 六 月

線性轉換模型之統計推論

Statistical Inference for Linear Transformation Models

研 究 生：林泰佐　　Student：Tai-Tso Lin

指導教授：王維菁博士　　Advisor：Dr. Weijing Wang

國 立 交 通 大 學
統 計 學 研 究 所
碩 士 論 文

A Thesis
Submitted to Institute of Statistics
College of Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
in
Statistics

June 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年六月

線性轉換模型之統計推論

學生: 林泰佐　　　　　　　　　　　　指導教授: 王維菁 博士

國立交通大學統計學研究所

摘　　要

線性轉換模型是相當彈性的半母數迴歸模型。倖存分析常見的 Cox 模型與 Odds 模型，皆是線性轉換模型的特例。近年來許多研究針對線性轉換模型提出半母數推論方法，一套分析方法卻有廣泛的應用價值，是其吸引人的地方。我們以古典推論理論的兩個原則(動差法和概式法)為架構，檢視現有文獻的建構方式，希望此統整的角度有助辨識不同方法的特質。此外針對設限資料，我們除了討論現有文獻的做法外，並提出一個新的方法。所有的方法均透過模擬實驗檢驗其表現。

關鍵詞: Cox PH 模型; Odds 模型; 動差法

Statistical Inference for Linear Transformation Models

Student: Tai-Tso Lin
Advisor: Dr. Weijing Wang

Institute of Statistics
Nation Chiao Tung University
Hsinchu, Taiwan

ABSTRACT

The linear transformation model, which includes the proportional hazard model and the proportional odds model, has received considerable attentions in recent years due to its flexibility. In the thesis, we consider semi-parametric estimation for the regression parameter. We review existing literature under the framework of classical inference theory. Specifically we will see how these "old" principles, namely method of moment and likelihood estimation, are applied to the modern estimation problem which involves an infinite dimensional nuisance parameter in the model formulation. After examining common techniques of handling censored data, we also propose a new approach. All the methods are evaluated by Monte Carlo simulations.

Key words: Cox proportion hazard model; Proportional odds model; Counting process; Inverse probability weighting; Method of moment; Profile likelihood.

誌　　　謝

兩年前，懵懵懂懂的我因爲自己對數字的高敏感度以及對機率統計這門科學的熱愛，使我發奮努力成爲交大統研所的一員。這兩年寶貴的學習時光中，我時常砥礪自己要好好學習並且自動自發，在教授們辛勤的教導與同學們相互研究討論，以及所上提供良好的學習與研究環境(每週清交合作演講、電腦機房和所圖)，使我吸收到很多寶貴的知識與想法。統計是一門應用相當廣泛的科學，促使我更加努力學習其他方面的知識，自我學習能力是我這兩年最大的收穫。古人云:『學海無涯』，我將會繼續保持一個謙卑學習的態度面對接下來的任何挑戰。

這篇論文的完成，我最想要感謝我的指導老師王維菁教授這一年來諄諄教誨。老師十分強調正確的研究態度，任何事情都是從最簡單且最直覺的想法開始，想清楚其中的脈絡與思緒是最重要的。因此，在教授的指導協助下，從一開始概念的釐清、建構研究架構直到論文撰寫完成，在這從無到有的過程中，使我學習到面對問題時該有的正確態度。這兩年研究生活中，還要感謝我的好伙伴文廷，不管在學習、研究或生活上，他時常提供我意見與想法。讓我最欽佩的就是他如拼命三郎的學習態度，無形中也激勵我向上，文廷可以說是我的益友也是良師。此外還要感謝仲竹、翁賢、昱緯、重耕和瑜達，雖然我們研究的領域不同，不過大家一起努力一起學習，交換各種意見與心得，並且閒暇之餘聚在一起閒話家常，使我兩年研究生活增添許多歡樂。另外還有研究室的夥伴們－夙吟、政言、香菱、珮琦、佩芳、鈺婷、彥銘、郤嵐和健鈞，謝謝大家的陪伴，讓我在交大的研究生活更加多采多姿。感謝女友姮君在我遇到困難失落難過的時候陪伴我鼓勵我，讓我更加成長茁壯。最後要感謝我最親愛的家人，一路上栽培我的爸爸和媽嗎，以及陪伴我成長的哥哥和弟弟，你們是我最強而有力的後盾。

同時感謝口試委員陳鄰安教授、黃信誠教授和王秀瑛教授提供諸多建議，使得本論文更加完善。

最後，謹將此論文獻給敬愛的父母親和王維菁教授，感謝他們的無私奉獻，因爲有他們對我的栽培與指導，使我不怕面對未來的艱困與挑戰。希望所有關心我的人，能一同分享此篇論文完成之喜悅與榮耀，謝謝大家。

林泰佐 謹 誌于

國立交通大學統計研究所

中華民國九十七年六月

Contents

List of Tables

Chapter 1 Introduction

1.1 Motivation

In many biomedical applications, researchers are interested in studying how covariates affect a patient's survival time. Let T be the failure time of interest and Z be a $p \times 1$ vector of covariates. The most well-known regression model in survival analysis is perhaps the proportional hazard (PH) model proposed by Cox (1972) which can be written as

$$\lambda(t \mid Z) = \lambda_0(t) \times e^{Z^T \eta}, \tag{1.1}$$

where $\lambda(t \mid \mathrm{Z})$ is the hazard function given Z and $\lambda_0(t)$ is the hazard for the "baseline" group with $\mathrm{Z} = (0,0,...,0)^{\mathrm{T}}$. The Cox model can also be written as

$$\log(\text{-}\log\{S(t \mid Z)\}) = \log(\Lambda_0(t)) + Z^{\mathrm{T}}\eta, \tag{1.2}$$

where $S(t \mid Z)$ is the survival function of the failure time T given the covariate Z and $\Lambda_0(t) = \int_0^{\mathrm{t}} \lambda_0(\mathrm{u})\mathrm{du}$ is the baseline cumulative hazard function. Notice that the right-hand side in (1.2) shows a linear structure.

In recent years, there has been a trend to construct a general class of models which consist of existing models as special cases accordingly unified inference procedure applicable for all members in the class can be developed. Consider the following generalization of model (1.2):

$$\varphi(S(t \mid Z)) = h(t) + Z^{\mathrm{T}}\eta, \tag{1.3}$$

where $S(\mathrm{t} \mid \mathrm{Z}) = \Pr(T > t \mid Z)$ and $\varphi(.)$ is a decreasing link function. Here we consider the situation that $\varphi(.)$ is known but the form of $\mathrm{h}(.)$ is unspecified. Based on (1.2), we see that Cox proportional hazard model belongs to this class with link function $\varphi(\mathrm{x}) = \log(\text{-}\log(\mathrm{x}))$. Another example is the proportional odds model with $\varphi(\mathrm{x}) = \text{-}\log\left\{\frac{\mathrm{x}}{1-\mathrm{x}}\right\}$.

Model (1.3) has another more direct representation given by

$$h(T) = \text{-}Z^{\mathrm{T}}\eta + \varepsilon, \tag{1.4}$$

where $h(.)$ is a completely unknown strictly increasing function, η is a $p \times 1$ vector of unknown parameters and ε is the error term with distribution function $F_\varepsilon(x) = \Pr(\varepsilon \le x)$. Model (1.4) or its equivalent expression is called the linear transformation model. Note that $F_\varepsilon(x)$ is related to link function $\varphi(\mathrm{x})$. By simple calculations, we see that

$$\begin{aligned} F_\varepsilon(x) = \Pr(\varepsilon \le x) = \Pr(\varphi(S(T)) \le x) \\ = \Pr\left(S(T) \le \varphi^{-1}(x)\right) \\ = \Pr\left(F(T) \le 1\text{-}\varphi^{-1}(x)\right) \\ = 1\text{-}\varphi^{-1}(x), \end{aligned}$$

where the last two identities use the fact that $F(T) \sim \mathrm{U}(0,1)$. If ε has the extreme value distribution, model (1.4) is the Cox PH model. If ε has the standard logistic distribution, model (1.4) corresponds to the proportional odds model. Because of its generality and flexibility, the linear transformation model has attracted substantial attention.

1.2 Inference Methods

The parametric version of the linear transformation model (1.4), with $h(.)$ being specified up to a finite-dimensional parameter vector, was studied by Box & Cox (1964). In the thesis, we focus on semiparametric inference with $h(.)$ being unknown. For the proportional hazard model, Cox (1975) proposed the partial likelihood function for parameter estimation and its large sample was discussed by Tsiatis (1981). For the proportional odds model, Pettitt (1982) utilized properties of ranks (i.e. invariance under monotone increasing transformation) to construct a marginal likelihood based on ranks. Dabrowska & Doksum (1988) considered the proportional odds model under the two-sample setting.

In this thesis, we will review inference methods for estimating η for the class of linear transformation models when T is subject to censoring by another random variable C. These methods can handle all members in the model in (1.4) and hence are flexible compared with the aforementioned methods developed for a particular member ,say the Cox model in (1.1).

1.3 Outline

The thesis is organized as follows. In Chapter 2, we review fundamental inference ideas for analyzing survival data without covariates. We will see that these concepts are still useful for analyzing more complicated data structures. In Chapter 3, we discuss some ideas of inference based on complete data. In Chapter 4, we study how these methods are modified when data are subject to censoring. In chapter 5, we proposed a new simple method. Numerical analyses are summarized in Chapter 6. Concluding remarks are given in Chapter 7.

Chapter 2 An Overview of Survival Analysis

Recall that T denote the failure time with the distribution function $F(t) = \Pr(\mathrm{T} \le \mathrm{t})$ and survival function $S(t) = \Pr(T > t)$. The censoring variable is denoted by C with the survival function $G(t) = \Pr(C > t)$ and the density function $g(t)$. Here we temporarily ignore the information provided by the covariates. In presence of right censoring, the observed variables are (X, δ, Z), where $X = \min(T, C)$ and $\delta = I(T \le C)$. The sample contains independently and identically distributed observations of (X, δ, Z), denoted as $\{(X_i, \delta_i, Z_i)\,(i = 1,..., n)\}$.

2.1 Likelihood Inference

If a parametric distribution is imposed on T, we can write $F(t)$ and $S(t)$ as $F_\theta(t)$ and $S_\theta(t)$ respectively. Denote $f_\theta(t)$ as the corresponding density. The parametric likelihood of θ can be denoted as

$$L(\theta) = \prod_{i=1}^{n} \left[f_\theta(x_i)^{\delta_i} S_\theta(x_i)^{1-\delta_i} \right] \times \left[G(x_i)^{\delta_i} g(x_i)^{1-\delta_i} \right]$$

$$\propto \prod_{i=1}^{n} \left[f_\theta(x_i)^{\delta_i} S_\theta(x_i)^{1-\delta_i} \right]. \tag{2.1}$$

The maximum likelihood estimator of θ can be obtained by solving the equation $\partial L(\theta) / \partial\theta = 0$ given that $\mathrm{L}(\theta)$ is differentiable with respect to θ.

If the distribution of T is not completely specified, θ often involves high-dimensional nuisance parameters and direct direction by solving $\partial L(\theta) / \partial\theta = 0$ is not easy. Consequently modified versions of the likelihood function, such as marginal, conditional or profile likelihoods, have appeared in the literature. For example, for the profile log-likelihood, θ is decomposed as $\theta = (\rho, \nu(\rho))$. For fixed ρ, $\nu(\rho)$ can be estimated by $\hat{\nu}(\rho)$. Then the profile likelihood replaces the original likelihood $L(\theta) = L(\rho, \nu(\rho))$ by $L(\rho, \hat{\nu}(\rho))$.

2.2 Nonparametric Estimation

Suppose that the distribution of T is completely not specified. If complete data are available, $S(t) = \Pr(T > t)$ can be estimated by the empirical estimators:

$$\bar{S}(t) = \sum_{i=1}^{n} I(T_i > t) / n .$$

In presence of right censoring, Kaplan and Meier (1958) proposed the following product-limit estimator:

$$\hat{S}(t) = \prod_{u \le t} \{1 - \frac{\sum_{i=1}^{n} I(X_i = u, \delta_i = 1)}{\sum_{i=1}^{n} I(X_i \ge u)}\} . \quad (2.2)$$

It can be shown that $\hat{S}(t)$ reduces to $\bar{S}(t)$ when $\delta_i = 1$ for all $i = 1,...,n$. Similarly the cumulative hazard function $\Lambda(t) = \int_0^t \lambda(u) du$ can be estimated by

$$\hat{\Lambda}(t) = \sum_{u \le t} \frac{\sum_{i=1}^{n} I(X_i = u, \delta_i = 1)}{\sum_{i=1}^{n} I(X_i \ge u)} .$$

Now we discuss some nice properties of the Kaplan-Meier estimator since these ideas can be further utilized to solve more complicated problems. We can view

$$S(t) = \Pr(T \ge t) = E[I(T \ge t)]. \quad (2.3)$$

When the data are complete, the empirical estimator of $S(t) = E[I(T \ge t)]$ is given by $\bar{S}(t) = \frac{1}{n} \sum_{i=1}^{n} [I(T_i \ge t)]$, which utilizes the method of moment. In presence of censoring, $I(T_i \ge t)$ may not be completely observable. Two useful techniques for handling missing data can be applied. One is imputation and the other is weighting.

The idea of imputation is to replace $I(T_i \ge t)$ by an estimation of its conditional expectation given the data. Specifically we have

$$\hat{E}[I(T_i \ge t) \mid X_i, \delta_i] = \delta_i I(X_i \ge t) + I(X_i < t, \delta_i = 0) \hat{E}[I(T_i \ge t) \mid X_i, T_i > X_i]$$

$$= \delta_i I(X_i \geq t) + I(X_i < t, \delta_i = 0)\frac{\hat{S}(t)}{\hat{S}(X_i)}.$$

The above idea has been utilized in constructing the following self-consistency equation:

$$\hat{S}(t) = \frac{1}{n}\sum_{i=1}^{n}\hat{E}[I(T_i \geq t) \mid X_i, \delta_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[\delta_i I(X_i \geq t) + I(X_i < t, \delta_i = 0)\frac{\hat{S}(t)}{\hat{S}(X_i)}\right]. \tag{2.4}$$

The above equation can be solved successively from the smallest observed value of T. It is well-known that the above equation has a unique solution which is the Kaplan & Meier estimator. Equation (2.4) can be modified for more complicated data structures such as interval censoring (Turnbull, 1976).

Weighting is another way of handling missing data. We can view $I(X \geq t)$ as a proxy of $I(T \geq t)$. To correct the bias of $I(X \geq t)$, we find that $E[\frac{I(X \geq t)}{G(t)}] = E[I(T > t)]$. The Kaplan estimator $\hat{S}(t)$ can be written as $\hat{S}(t) = \sum_{i=1}^{n}\frac{I(X_i > t)}{\hat{G}(t)}$, where $\hat{G}(t)$ is the Kaplan-Meier estimator of $G(t)$ such that

$$\hat{G}(t) = \prod_{u \leq t}\{1 - \frac{\sum_{i=1}^{n} I(X_i = u, \delta_i = 0)}{\sum_{i=1}^{n} I(X_i \geq u)}\}.$$

2.3 Counting Process

Aalen (1975) analyzed survival data under the framework of counting processes and martingales. This approach provides an elegant and relatively simple structure for theoretical analysis of many well-known inference methods for analyzing survival data.

Consider the counting process defined as $N_i(t) = I(X_i \leq t,\ \delta_i = 1)$. The corresponding filtration can be written as

$$F_t = \sigma\{I(X_i \leq t, \delta_i = 1), I(X_i \leq t, \delta_i = 0)(i = 1, \ldots, n)\}.$$

Note that F_t records the history of the counting process at or prior to time t and satisfies the nested property such that $F_{\mathrm{s}} \subset F_{\mathrm{t}}$ for $s \le t$. It follows that

$$E[dN(t)|F_{t-}] = I(X \ge t)\Pr(X = t, \delta = 1|X \ge t)$$

$$= Y(t) \times \lambda(t)dt,$$

where $Y(t) = I(X \ge t)$ is the "at-risk" process for the failure event under censoring. Define

$$\Lambda(t) = \int_0^t Y(s)\lambda(s)ds,$$

which is the compensator for $N(t)$. Based on the Doob-Meyer decomposition, we have

$$N(t) = \Lambda(t) + M(t),$$

where $M(t)$ is a mean-zero martingale, satisfying $\mathrm{E}(M(t)|F_s) = M(s)$ for $s \le t$. The counting process representation is not only useful for theoretical analysis but also provides a nice structure for parameter estimation. We will see how the idea is applied under the regression setting.

Chapter 3 Regression Analysis without Censoring

The linear transformation regression model can be written as

$$h(T_i) = \text{-}Z_i^T \eta + \varepsilon_i \ , \ i = 1,2,\ldots,n$$

where Z_i is a $p \times 1$ vector of covariates for subject i. The error terms ε_i $(i = 1,2,\ldots,n)$ are identically and independently distributed with a known distribution $F_\varepsilon(x) = \Pr(\varepsilon \le x)$. Semiparametric inference of η without knowing $h(.)$ has received substantial interest in the literature due to the model flexibility. In this chapter, we review existing modern methods based classical inference principles, namely the method of moment and likelihood estimation. To simplify the presentation, we ignore censoring temporarily. Hence the data can be denoted as $\{(T_i, Z_i), (i = 1,2,\ldots,n)\}$.

3.1 Moment-based Inference

The method of moment is attractive since it does not require strong distributional assumption and usually provides more robust results. The method of moment or its modified version is constructed based on some moment properties of a chosen response variable. Suppose O_i is a chosen response variable. Denote E_i as $E(O_i|Z_i)$ or $E(O_i|Z_i, A)$, where A is a statistic, which is a function of η or other nuisance parameter denoted as γ. In our model, γ is related to $h(.)$. The following unbiased estimating equation can be constructed

$$U(\eta,\gamma) = \sum_{i=1}^{n} W_i \times (O_i \text{-} E_i(\eta,\gamma)) = 0 \ ,$$

where W_i is a weight function for subject i. If additional parameter γ is involved, other estimating equation is needed. The property of the resulting estimator is related to whether $E_i(\eta,\gamma)$ is a nice function of η and whether the weight W_i is properly chosen. We will see that there exist several candidates for O_i.

3.1.A The pairwise order indicator as the chosen response

In the paper of Cheng, Wei and Ying (1995), they used $I(T_i \geq T_j)$ as the response variable, where $I(.)$ denotes the indicator function. Based on the assumption that $h(.)$ is a strictly increasing function, it follows that

$$I(T_i \geq T_j) = I(\mathrm{h}(T_i) \geq \mathrm{h}(T_j)) = I(\varepsilon_\mathrm{i} - \varepsilon_\mathrm{j} \geq (\mathrm{Z_{ij}}^\mathrm{T}\eta)).$$

This implies that

$$\begin{aligned} \mathrm{E}[I(T_i \geq T_j) \mid \mathrm{Z}] &= \mathrm{E}[I(\varepsilon_\mathrm{i} - \varepsilon_\mathrm{j} \geq \{\mathrm{Z_i}^\mathrm{T} - \mathrm{Z_j}^\mathrm{T}\}\eta))] \\ &= \Pr(\varepsilon_\mathrm{i} - \varepsilon_\mathrm{j} \geq (\mathrm{Z_{ij}}^\mathrm{T}\eta)) \\ &= \xi(\mathrm{Z_{ij}}^\mathrm{T}\eta), \end{aligned}$$

where $\xi(t) = \int_{-\infty}^{\infty} \{1\text{-}F_\varepsilon(t+s)\} dF_\varepsilon(s)$ and $F_\varepsilon(t) = Pr(\varepsilon \leq t)$. Appendix 1 contains more detailed derivations. A nice feature of using $I(T_i \geq T_j)$ as the response is that the corresponding expectation does not involve the unknown nuisance function $\mathrm{h}(.)$. It implies no additional estimating functions are needed. The resulting estimating equation becomes:

$$\sum_{i \neq j}\sum W(Z_{ij}{}^T\eta) Z_{ij} \times \left(I(T_i \geq T_j) - \xi(Z_{ij}{}^T\eta)\right) = 0. \qquad (3.1)$$

The solution is an unbiased and consistent estimator of η.

3.1.B The at-risk process as the chosen response

Recall that the at-risk process is defined as $Y(t) = I(T \geq t)$. Cai, Wei and Wilcox (2000) suggested to use $Y(t)$ as the response. Its expected value under the model is

$$S(t \mid Z) = \varphi^{-1}(h(t) + Z^T\eta).$$

The resulting estimating equation is given by

$$\sum_{i=1}^{n}[I(T_i \geq t) - \varphi^{-1}\{h(\mathrm{t}) + Z_\mathrm{i}{}^T\eta\}] = 0, \quad t \in (\tau_a, \tau_b).. \qquad (3.2a)$$

Note that (3.2) provides a set of equations for t being the observed values of T_i $(i = 1,\dots,n)$. If η is one-dimensional, there are $n+1$ unknown parameters in (3.2). Therefore we need one more equation. Cai et al. (2000) suggested the following equation

$$\sum_{i=1}^{n}\int_{\tau_a}^{\tau_b} Z_i [I(T_i \geq t) - \varphi^{-1}\{h(t) + Z_i^T \eta\}]dt = 0\,, \tag{3.2b}$$

where (τ_a, τ_b) is a re-specified range that contains enough data information. The authors suggested to solve equations (3.2a) and (3.2b) iteratively. That is with an initial value of η, $h(t_i)$ for $i = 1,...,n$ can be estimated based on equation (3.2). These estimators are plugged into equation (3.3) and then η can be estimated. The procedure is implemented iteratively between (3.2a) and (3.2b) until the convergence criteria is reached. This method seems to be more complicated than the previous one since the nuisance function $h(.)$ has to be estimated as well and the region (τ_a, τ_b) has to be determined.

3.1.C The counting process as the chosen response

Recall that the counting process $N(t) = I(T \leq t)$. The corresponding expectation conditional on the filtration F_{t-} under the transformation model can be written as

$$\int_0^t Y(s) d\Lambda(\eta^T Z_i + h(s)).$$

With the true parameter values, $N(t) - \int_0^t Y(s) d\Lambda(\eta^T Z_i + h(s))$ is a mean-zero martingale. This property can be used for constructing estimating equations.

Chen, Jin and Ying (2002) suggested two estimating equations. One for estimating $h(t)$ is given by

$$\sum_{i=1}^{n}[dN_i(t) - Y_i(t) d\Lambda\{\eta^T Z_i + h(t)\}] = 0\,. \tag{3.3a}$$

The second estimating equation for η is given by

$$\sum_{i=1}^{n}\int_0^\infty Z_i [dN_i(t) - Y_i(t) d\Lambda\{\eta^T Z_i + h(t)\}] = 0\,. \tag{3.3b}$$

The same idea of iteration mentioned in the previous sub-section is also applied to solve (3.3a) and (3.3b).

3.2 Likelihood Inference

Under the linear transformation model (1.4), $h(T) = -Z^{\mathrm{T}}\eta + \varepsilon$, the survival function can be written as $S(t \mid Z) = \varphi^{-1}\left(h(t) + Z^{T}\eta\right)$. The likelihood function can be written as

$$-\prod_{i=1}^{n} \frac{\partial S(t \mid Z)}{\partial t}\Big|_{t=t_i} = -\prod_{i=1}^{n} \frac{\partial \varphi^{-1}\left(h(t) + Z^{T}\eta\right)}{\partial t}\Big|_{t=t_i}. \tag{3.4}$$

The above function is very complicated and straightforward maximization is impossible. We will present the related work in the next chapter which accounts for the presence of censoring.

Chapter 4 Regression Analysis with Right Censoring

In practice, patients may drop out from the study or do not develop the event of interest during the study period. Therefore, T is often subject to right censoring. In this chapter, we discuss how the aforementioned methods adjust for the presence of censoring. In Section 4.1, we review three ways of modification for the moment-based estimators. In Section 4.2, we review the likelihood method in presence of censoring. Suppose that under model (1.4), T is subject to censoring by C with the survival function $G(t)=\Pr(C\geq t)$. Observable data become $\{(X_i,\delta_i,Z_i),(i=1,2,...,n)\}$ which are random replications of (X,δ,Z), where $X=\min(T,C)$ and $\delta=I(T\leq C)$.

4.1 Moment-based Inference

The chosen response variables discussed in Chapter 3 are not completely observed. To handle this problem, two useful techniques for analyzing missing data, namely the weighting and imputation approaches, are frequently used.

Now we illustrate the technique of weighting. For the response variable $I(T\leq t)$, a natural proxy under censoring is $I(X\leq t,\delta=1)$, which however is biased.

$$\begin{aligned} E(I(X\leq t,\delta=1)) &= E[E[I(T\leq t,C\geq T)\mid T]] \\ &= E[I(T\leq t)G(T)]. \end{aligned}$$

Therefore,

$$E\left(\frac{I(X\leq t,\delta=1)}{G(X)}\right)=E(I(T\leq t)),\text{ if } G(X)>0 \tag{4.1}$$

Since $G(\mathrm{t})$ is often unknown, the Kaplan-Meier estimator $\hat{G}(X)$ is a suitable candidate to replace $G(X)$. Specifically,

$$\hat{\mathrm{G}}(t)=\prod_{u\leq t}\{1-\frac{\sum_{i=1}^{n}I(X_i=u,\delta_i=0)}{\sum_{i=1}^{n}I(X_i\geq u)}\}.$$

4.1.A The pairwise order indicator as the chosen response

Cheng et al. (1995), suggested to use $\dfrac{I(X_i \ge X_j, \delta_j = 1)}{G^2(X_j)}$ as the proxy of $I(T_i \ge T_j)$.

Notice that

$$E[I(X_i \ge X_j, \delta_j = 1)] = E[E[T_i \ge T_j, C_i \ge T_j, C_j \ge T_j \mid T_i, T_j]]$$
$$= E[I(T_i \ge T_j)G^2(T)].$$

This implies that

$$E[\frac{I(X_i \ge X_j, \delta_j = 1)}{G^2(X_j)}] = E[I(T_i \ge T_j)],$$

if the denominator is not zero. See Appendix 2 for the details.

The estimating equation in (3.1) can be modified as

$$\sum_{i=1}^{n}\sum_{i \ne j} W(Z_{ij}^{T}\eta)Z_{ij} \times \left(\frac{\delta_j I(X_i \ge X_j)}{\hat{G}^2(X_j)} - \xi(Z_{ij}^{T}\eta) \right) = 0, \tag{4.2}$$

where $\hat{G}(.)$ is the Kaplan-Meier estimator $\Pr(C \ge t)$.

Despite its simplicity and convenience, the weighting approach has a series drawback. First of all, equation (4.2) produces (asymptotically) unbiased result only if the censoring support lies within the support of T. Specifically define $\tau_c = \sup_t\{t : G(t) > 0\}$ and $\tau_T = \sup_t\{t : \Pr(T > t) > 0\}$. The validity of (4.2) requires $\tau_T \le \tau_c$ which eliminate the situation $G(T_i) = 0$. However, this assumption rarely holds in practice since the study period is often limited which makes $\tau_T > \tau_c$. This situation is developed in Figure (4.1).

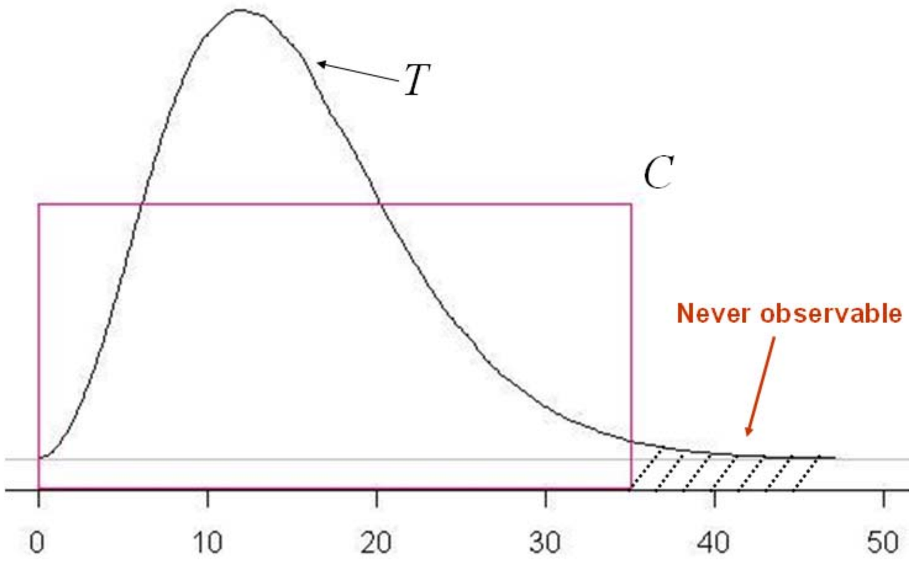


Figure 4.1 *Problem for the weighting technique*

To overcome this problem, Fine, Ying and Wei (1998) suggested to impose a truncation point t_0 (Figure 4.2) such that

$$\frac{I(\min(X_i, t_0) \geq X_j, \delta_j = 1)}{G^2(X_j)},$$

where $G(t_0) > 0$ and t_0 is a prespecified constant satisfying.

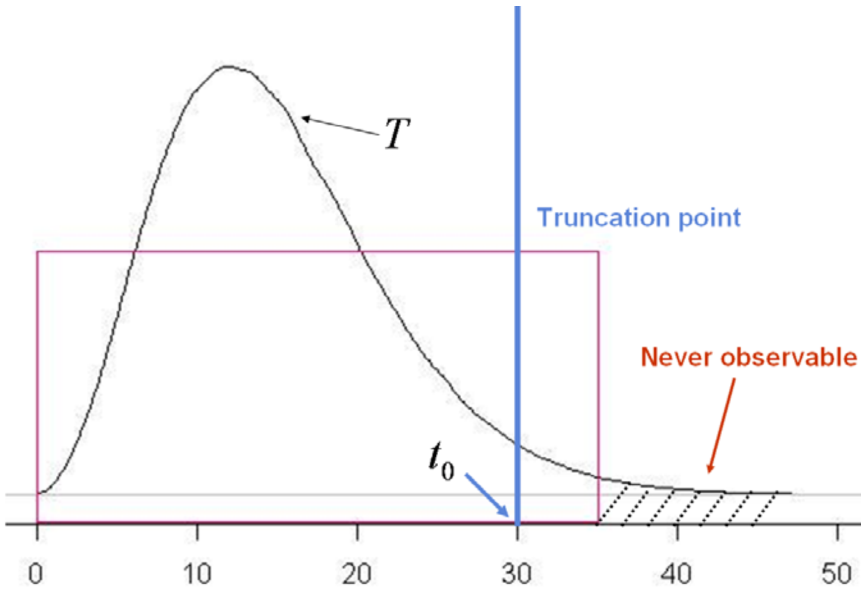


Figure 4.2 *Imposing a truncation point to overcome the problem*

The corresponding expected value for the adjusted response is given by

$$\xi_{ij}^{\ *} = E[\frac{\delta_j I\left(\min(X_i, t_0) \geq X_j\right)}{G^2(X_j)}] = \xi_{ij}(\eta) - \Pr\left(X_i \geq X_j \geq t_0\right)$$

$$= E[I(T_j \leq t_0) I(T_i \geq T_j)]$$

$$= \int_{-\infty}^{h(t_0)} \{1 - F_\varepsilon(t + Z_{ij}^{\ T}\eta)\} dF_\varepsilon(t), \tag{4.3}$$

which is a function of η and $h(t_0)$. See Appendix 3 for the details. Furthermore instead of using the first moment condition, Fine et al. (1998) proposed to use the least square principle by minimizing the objective function:

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{i \neq j} W(Z_{ij}^{\ T}\eta) \times \left(\frac{\delta_j I\left(\min(X_i, t_0) \geq X_j\right)}{\hat{G}^2(X_j)} - \xi_{ij}^{\ *}(\eta) \right)^2,$$

which leads to the estimating equation

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{i \neq j} W(Z_{ij}^{\ T}\eta) \xi_{ij}^{\ '}(\eta) \times \left(\frac{\delta_j I\left(\min(X_i, t_0) \geq X_j\right)}{\hat{G}^2(X_j)} - \xi_{ij}^{\ *}(\eta) \right) = 0.$$

For estimating $h(t_0)$, they also proposed another estimating function.

Subramanian (2004) proposed a different way of modifying equation (3.1). The idea of Subramanian is to replace the original response $I\left(T_i \geq T_j\right)$ by an estimator of its nonparametric estimation. However, this method assumes that the covariate Z takes discrete values.

$$E\left(I\left(T_i \geq T_j\right) Z_i, Z_j\right) = \Pr\left(T_i \geq T_j | Z_i, Z_j\right) = \int_0^\infty S\left(t | Z_i\right) dF\left(t | Z_i\right)..$$

The Kaplan-Meier estimator can be applied to estimate $S\left(t | Z\right)$ such

$$\hat{\mathrm{S}}\left(\mathrm{t} \mid \mathrm{Z}\right) = \prod_{u \leq t} \{1 - \frac{\sum_{i=1}^{n} I(X_i = u, \delta_i = 1, Z_i = \mathrm{z})}{\sum_{i=1}^{n} I(X_i \geq u, Z_i = \mathrm{z})}\}.$$

Therefore, Subramanian develop its estimating equation,

$$\sum\sum_{i\neq j} W(Z_{ij}{}^T\eta)Z_{ij}\times\left(\int_0^\infty \hat{S}(t|Z_i)d\hat{F}(t|Z_i)-\xi\left(Z_{ij}{}^T\eta\right)\right)=0. \tag{4.4}$$

where $\hat{S}(t|Z_i)$ is the K-M estimation and $\xi(t)=\int_{-\infty}^{\infty}\{1\text{-}F_\varepsilon(t+s)\}dF_\varepsilon(s)$. But, t This method is also vulnerable to the tail problem since the Kaplan-Meier estimator can not catch the tail information either if $\tau_T>\tau_c$. Therefore, Subramanian used the same technique, imposing the truncation point t_0, to develop the modified estimating equation,

$$U_s(\eta)=\sum\sum_{i\neq j} W(Z_{ij}{}^T\eta)Z_{ij}\times\left(\int_0^{h(t_0)} \hat{S}(t|Z_i)d\hat{F}(t|Z_i)-\xi^*\left(Z_{ij}{}^T\eta\right)\right)=0.$$

4.1.B The at-risk process as the chosen response

Recall 3.1.B, Cai, Wei and Wilcox (2000) suggested to use $Y(t)$ as the response. Its expected value under the model is

$$S(t\,|\,Z)=\varphi^{-1}\left(h(t)+Z^T\eta\right).$$

Under right censoring data, the corresponding response variable is $I(X_i\geq t)$. Thus, we can derived the expectation:

$$E[I(X_i\geq t)]=\Pr(T_i\geq t, C_i>T_i)=\varphi^{-1}\{h(t)+Z_i{}^T\eta\}G(t).$$

Cai et al. modify the equation (3.2a) as

$$\sum_{i=1}^{n}[I(X_i\geq t)-\varphi^{-1}\{h(\mathrm{t})+Z_{\mathrm{i}}{}^T\eta\}\hat{G}(\mathrm{t})]=0,\quad t\in(\tau_a,\tau_b)\quad. \tag{4.5a}$$

Note that (4.5) provides a set of equations for t being the observed values of T_i $(i=1,\ldots,n)$. If η is one-dimensional, there are $n+1$ unknown parameters in (4.5). Therefore we need one more equation. Cai et al. (2000) suggested the following equation

$$\sum_{i=1}^{n}\int_{\tau_a}^{\tau_b}\mathrm{Z_i}[I(T_i\geq t)-\varphi^{-1}\{h(t)+Z_{\mathrm{i}}{}^T\eta\}\hat{G}(\mathrm{t})]\mathrm{dt}=0, \tag{4.5b}$$

where (τ_a,τ_b) is a re-specified range that contains enough data information. Solve equations (4.5a) and (4.5b) iteratively. The following numerical operation is the same as 3.1.B we mentioned.

4.1.C The counting process as the chosen response

With censoring data structure, using the estimating equation based on counting process is easily modifying. We would not change the formation we mentioned in 3.1.B. That is to say, it is very generalized method in constructing the estimating equation in linear transformation model.

4.2 Likelihood Inference

The likelihood function in (3.6) can be extended to the censoring situation as follows:

$$-\prod_{i=1}^{n}\left[\frac{\partial\varphi^{-1}(\eta,\mathrm{z_i},h(t))}{\partial t}\Big|_{t=x_i}\right]^{\delta_i}\varphi^{-1}(\eta,\mathrm{z_i},h(x_i))^{1-\delta_i}\ .$$

Since direct maximization is impossible, how to handle the nuisance function $h(.)$ is the key.

4.2.1 Partial Likelihood – Cox model

Here we illustrate the way Cox (1972, 1975) used to handle the nuisance baseline hazard function $\lambda_0(t)$ under the model $\lambda(t|Z)=\lambda_0(t)\times\exp(Z^T\eta)$. At time t, the probability that the failure event is for patient i given the risk set information is

$$\frac{I(X_i=t,\delta_i=1)\lambda_0(t)\times\exp(Z_i^T\eta)}{\sum_{j\in R(t)}\lambda_0(t)\times\exp(Z_j^T\eta)}=\frac{I(X_i=t,\delta_i=1)\exp(Z_i^T\eta)}{\sum_{j\in R(t)}\exp(Z_j^T\eta)},$$

where $R(t)=\{j:X_j\geq t,\delta_{j=1}\}$ is the risk set at time at time t. The important point is that the same $\lambda_0(t)$ appears in both the numerator and denominator and hence gets cancelled out. Thus the above conditional probability is only the function of η. The so-called partial likelihood can be written as

$$\prod_{\text{all failture points } u}\left\{\frac{\sum_{i=1}^{n}I(X_i=u,\delta_i=1)\times\exp(Z_i^T\eta)}{\sum_{j=1}^{n}I(X_j\geq u)\times\exp(Z_j^T\eta)}\right\}, \qquad (4.6)$$

Since $\lambda_0(t)$ disappears, maximization of (4.6) becomes easy. The corresponding score function can be written as

$$\mathrm{U}(\eta)=\sum_{\mathrm{i}=1}^{\mathrm{D}}[Z_{(i)}-\frac{\sum_{j=1}^{n} I\left(X_j \geq t_{(i)}\right) Z_j \times \exp\left(\mathrm{Z}_j{}^T \eta\right)}{\sum_{j=1}^{n} I\left(X_j \geq t_{(i)}\right) \times \exp\left(\mathrm{Z}_j{}^T \eta\right)}],$$

where $t_{(i)}$ is the order values of X_j with $\delta_j = 1$ for $j = 1,...,n$ and $D = \sum_{j=1}^{n} \delta_j$ is the total number of observed failure events.

4.2.2 Conditional Profile Likelihood

The amazing cancellation for the Cox partial likelihood does not happen to the more general class of transformation models. Therefore if the likelihood approach is pursued, the nuisance function has to be dealt with directly.

For the general transformation model in (1.4),

$$\Pr(T > t | Z = z) = \Pr\{\varepsilon > h(t) + z^T \eta\} = \varphi^{-1}\left(h(t) + Z^T \eta\right),$$

Chen, H. Y. (2001) proposed a likelihood approach for the case-cohort study, the covariate Z has an unknown distribution $\pi(\mathrm{z}) = \Pr(\mathrm{Z} \leq \mathrm{z})$, which is a more complex data structure than that considered in the thesis. Now we organize his method based on our data structure. By writing the full likelihood as

$$\prod_{i=1}^{n}\left[f_\eta(x_i)^{\delta_i} S_\eta(x_i)^{1-\delta_i}\right]$$
$$=\prod_{i=1}^{n}\left[\left(-\frac{\partial}{\partial t}\varphi^{-1}\{\eta, Z_i, h(x_i)\}\right) d\pi(z)\right]^{\delta_i}\left[\varphi^{-1}\{\eta, Z_i, h(x_i)\} d\pi(z)\right]^{1-\delta_i},$$

Chen suggested to express the function in terms of η and the marginal survival distribution of T, such that

$$R(t) = \Pr(T \geq t) = \int \varphi^{-1}\left(h(t) + Z^T \eta\right) d\pi(\mathrm{z}), \tag{4.7}$$

where $\pi(\mathrm{z}) = \Pr(\mathrm{Z} \leq \mathrm{z})$ is the distribution function of Z. The motivation of the above transformation is that $R(t)$ can be estimated by the Kaplan-Meier estimator

$$\hat{R}(t)=\prod_{u\le t}\{1-\frac{\sum_{i=1}^{n}I(X_i=t,\delta_i=1)}{\sum_{i=1}^{n}I(X_i\ge t)}\}.$$

This implies that $R(t)$, as a complicated function of η, $h(.)$ and $H(.)$, can be estimated. The distribution function $\pi(.)$ can also be estimated explicitly by $\hat{\pi}(z)=\sum_{i=1}^{n}I(Z_i\le z)/n$. Based on (4.7), one can derive the relationship between $h(t)$ and $\{R(t),\eta,H(.)\}$ by inverse transformation. Denote $v=v\{\eta,\pi,R\}$ as the transformation. The technical issue is not the focus of the thesis so that we do not state the details. Finally, after the transformation, η can be estimated by maximizing the following profile likelihood function:

$$\prod_{i=1}^{n}\left[\frac{\frac{\partial}{\partial}\varphi^{-1}\{\eta,z_i,v(\eta,\hat{\pi}(z_i),\hat{R}(x_i))\}}{\int\frac{\partial}{\partial}\varphi^{-1}\{\eta,z_i,v(\eta,\hat{\pi}(z_i),\hat{R}(x_i))\}d\hat{\pi}(z_i)}\right]^{\delta_i}\left[\frac{\varphi^{-1}\{\eta,z_i,v(\eta,\hat{\pi}(z_i),\hat{R}(x_i))\}}{\hat{R}(x_i)}\right]^{1-\delta_i}.$$

This approach is very complicated and difficult to implement. The validity of the resulting estimator depends on whether the suggested transformation has to be a one-to-one mapping.

Chapter 5 New Proposed Method

Under pairwise method, we propose to directly modify the whole equation:

$$\sum\sum_{i \neq j} W(Z_{ij}{}^T \eta) Z_{ij} \times \left(I(T_i \geq T_j) - \xi\left(Z_{ij}{}^T \eta\right)\right) = 0 .$$

Specifically we only select pairs with the value of $I(T_i \geq T_j)$ being exactly known. To illustrate the idea, we can examine the two cases.

Case 1: If $I(X_i \geq X_j, \delta_j = 1)$, we know that $X_j = T_j$ and $I(T_i \geq T_j) = 1$.

Case 2: If $I(X_i \geq X_j, \delta_j = 0)$, we have $T_j > X_j$ and $T_i > X_j$ but the order of (T_i, T_j) is uncertain.

The following figure depicts the possible order relationships for a pair subject to censoring.

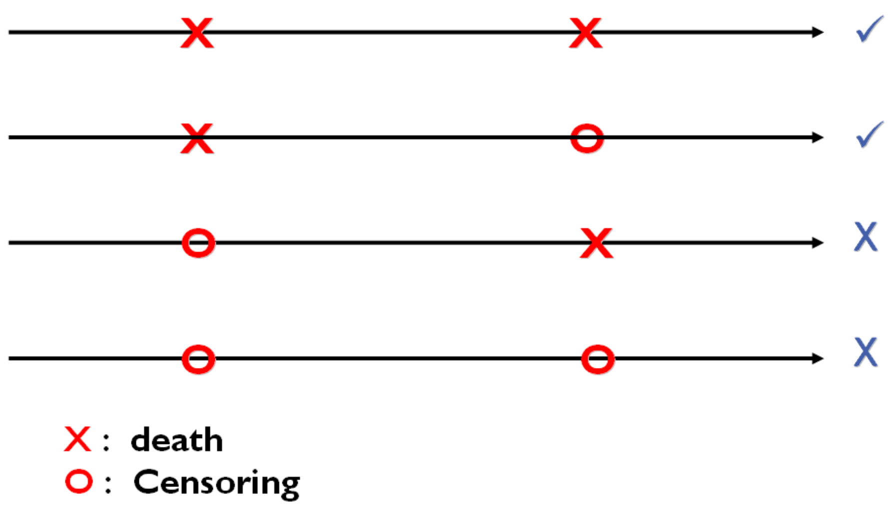


Figure 5.1: *The order relationship for a pair subject to censoring*

Define $\Delta_{ij} = \mathrm{I}\left((\delta_i = 1, \delta_j = 1) \cup (X_i > X_j, \delta_i = 0, \delta_j = 1) \cup (X_i < X_j, \delta_i = 1, \delta_j = 0)\right)$ as the orderable indicator which corresponds to $I(X_i \geq X_j, \delta_j = 1)$ or $I(X_j \geq X_i, \delta_i = 1)$. The corresponding estimating function is given by

$$U_{\text{new}}(\eta) = \sum\sum_{i \neq j} W\left(Z_{ij}{}^T \eta\right) \times Z_{ij} \times \Delta_{ij} \times \left(I(X_i \geq X_j) - \xi(Z_{ij}\eta)\right). \tag{5.1}$$

The proposed estimator is obtained by solving $U_{\text{new}}(\eta) = 0$. In Appendix 4, unbiasness of the estimator is proved.

We may study the proportion of the data that has been used in estimation. Define $p_1 = \sum_{i=1}^{n} I(\delta_i = 1)/n$ and $p_2 = \sum_{i=1}^{n}\sum_{j \neq i} I(\Delta_{ij} = 1)/\binom{n}{2}$. By changing the censoring proportions, we see that $1 > p_2 > p_1$. The new propose method uses p_2 of the data. However all the rest estimators use almost 100% of the data. The major advantage of our method is that there is no need to estimator other nuisance parameters and remains unbiasness even under censoring. The disadvantage is that $1 - p_2$ of the data is deleted. The loss of efficiency under heady censoring is expected.

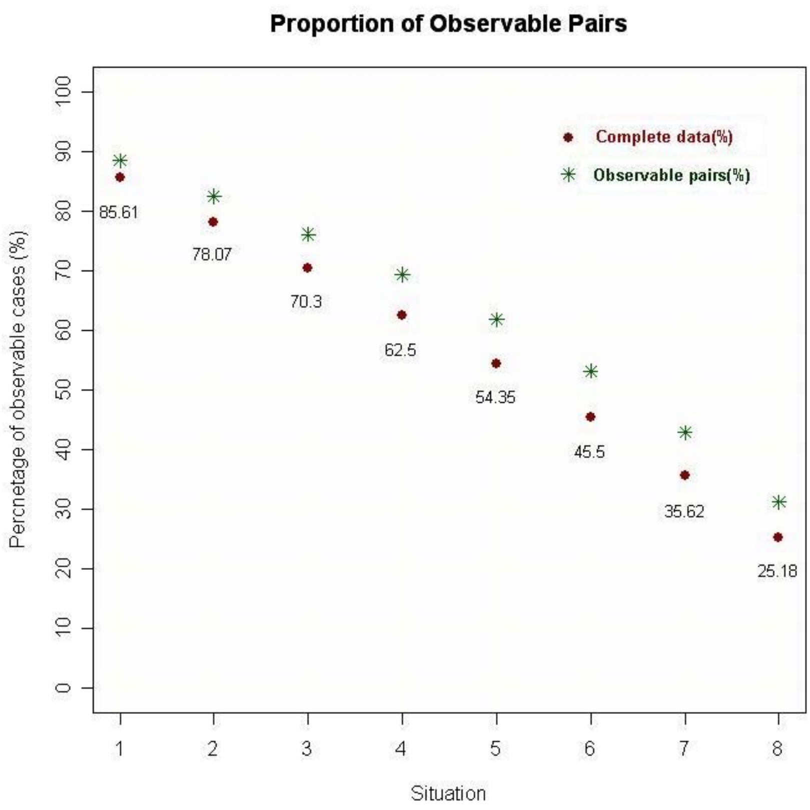


Figure 5.2: *The observable proportion: Original data vs. Paired data*

In addition, the new method losses too many data in using the comparable indicator Δ_{ij} and causes the inefficient outcome. Thus, we want to improve this estimating equation to avoid missing too many information. We consider the following situation in Figure 5.3. Using comparable indicator Δ_{ij}, we drop all the data in this situation. We use imputation technique and Kaplan-Meier estimator to modify the original equation.

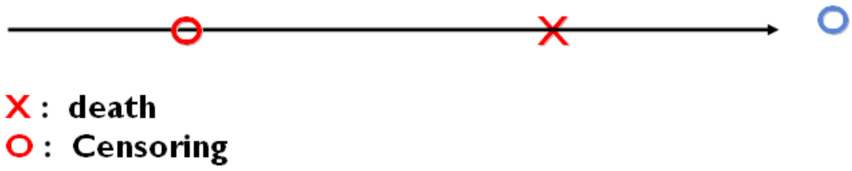


Figure 5.3: $(X_i > X_j, \delta_i = 1, \delta_j = 0) \cup (X_j > X_i, \delta_j = 1, \delta_i = 0)$

We imposed new comparable indicator $\Delta_{ij}^{*} = I(X_i > X_j, \delta_i = 1, \delta_j = 0)$ and imputation technique to renew our estimating equation:

$$\sum\sum_{i \neq j} W(Z_{ij}^{T}\eta) \times Z_{ij} \times \left\{ \Delta_{ij} \times \left[I(X_i \geq X_j) - \xi(Z_{ij}\eta) \right] + \Delta_{ij}^{*} \times \left[\left(1 - \hat{S}_{Z_i}(X_i) / \hat{S}_{Z_j}(X_j) \right) - \xi(Z_{ij}\eta) \right] \right\},$$

where $\hat{S}(X)$ is the Kaplan-Meier estimator.

Chapter 6 Simulation

We conduct Monte Carlo simulations to examine the finite-sample performances of the methods in the thesis. Based on the model, $h(T) = -\eta^T Z + \varepsilon$, we consider its special case, namely the Cox proportional hazard models. These Cox model is the most popular in practical applications. The covariate Z follows $\text{Bernoulli}(0.5)$ or $U(0,1)$. The censoring variable C is generated from uniform distributions. Two sample sizes are considered, namely n =100 or 250. For each setting, 2000 replications are run. The average bias and standard deviation of each method are reported.

6.1 Cox proportional hazard model

The error term follows the extreme value distribution with $F_\varepsilon(s) = 1 - \exp(-\exp(s))$. It follows that

$$\xi(s) = \int_{-\infty}^{\infty} \{1 - F(t+s)\} dF(t) = \frac{1}{1+\exp(s)}$$

and

$$\xi^*(p, s \mid t_0) = \frac{e^p \left(1 - e^{-e^{h(t_0)-p-s}\left(e^p + e^s\right)}\right)}{e^p + e^s}.$$

The latter is used in the work of Fine et al. (1998) which requires specifying the truncation point related to the range of integration. Here we choose $t_0 = 0.5$. For the estimating function proposed by Cai et al. (2000), we set (τ_a, τ_b) to be $(T_{0.05}, T_{0.95})$ and $(T_{0.02}, T_{0.98})$.

6.2 Discussion

Besides the results of the moment-based methods, we also report the result of the Cox partial likelihood estimator. Recall that the methods discussed in thesis are suitable of all members in the model class but the partial likelihood estimator is developed only for the Cox model.

In absence of censoring, the results produced by the pairwise comparison approach are similar to those of the partial likelihood method. However the weighting adjustment by Cheng et al. (1995) is problematic when censoring becomes heavier due to the tail problem mentioned earlier. The modification by Fine et al. (1998) by adding a truncation point successfully fixes the tail problem. Subramanian (2004) suggest using the nonparametric estimation to replace the pairwise indicator when data is censoring. However, the method of Subramanian is suitable under one important assumptation, the covariates Z follows discrete distribution. Our method uses the idea of choosing only comparable pairs to develop the unbiased estimating equation. However, we find that method losses too many data when censoring is heavy, the performance of estimation is inefficient (variance is relatively large). Thus, we use imputation technique to add a new comparable indicator Δ_{ij}^{*} to avoid losing too many data. In numerical outcome, we find that the equation imposing new comparable indicator Δ_{ij}^{*} is more efficient.

The approach by Cai et al. (2000) used at-risk process approach which requires specification of (τ_a, τ_b) and we see that this choice affects the result. If the range is set too wide including the extreme value, the variation of the estimator gets larger. How to choose the suitable (τ_a, τ_b) becomes the biggest problem in this method.

The counting process approach is the ideal method in the linear transformation models. In addition, we prove that counting process approach is equal to the specific method, Cox partial likelihood (1975), under Cox PH model in appendix 5. Because counting process approach uses the advantage of martingale, we find the numerical results of Chen et al.'s (2002) in below tables is more efficient than other unified methods. In theory, Chen's method also has good properties and generalization.

Table 1 *Comparison for different estimators of* $\eta_0 = 2$ *when the covariate* Z *is i.i.d.* $Unif(0,1)$ *and the sample size equal to* 100

N=100 (Re-sampling Size=2000)										
True Parameter	Censoring Proportion		Partial Likelihood	$I(X_i \ge X_j),\ i \ne j$				$I(X_i \ge t)$		$I(X_i \le t)$
				Cheng et al.	Fine et al.	New				
						Δ_{ij}	Δ_{ij}^*	$(T_{0.05}, T_{0.95})$	$(T_{0.02}, T_{0.98})$	
$\eta_0 = 2$	0%	Bias	0.06	-0.03				0.02	0.06	0.06
		(SD)	(0.41)	(0.43)				(0.43)	(0.45)	(0.41)
	14%	Bias	-0.02	-0.03	-0.02	-0.02	-0.01	-0.03	-0.07	-0.02
		(SD)	(0.43)	(0.47)	(0.43)	(0.50)	(0.48)	(0.44)	(0.49)	(0.43)
	22%	Bias	0.02	0.03	0.04	0.04	-0.05	0.05	0.03	0.02
		(SD)	(0.44)	0.49	0.47	(0.51)	(0.49)	(0.45)	(0.50)	(0.44)
	30%	Bias	0.06	0.06	0.07	0.08	-0.08	0.06	0.06	0.06
		(SD)	(0.48)	(0.51)	(0.49)	(0.55)	(0.52)	(0.48)	(0.52)	(0.48)
	38%	Bias	0.02	-0.01	0.02	0.05	-0.09	0.08	0.09	0.02
		(SD)	(0.49)	(0.51)	(0.49)	(0.57)	(0.50)	(0.49)	(0.55)	(0.49)
	46%	Bias	0.01	-0.06	-0.03	0.02	-0.08	0.02	0.07	0.01
		(SD)	(0.53)	(0.55)	(0.54)	(0.61)	(0.55)	(0.54)	(0.59)	(0.53)
	54%	Bias	0.03	-0.14	0.02	0.07	-0.06	0.06	0.06	0.03
		(SD)	(0.56)	(0.61)	(0.57)	(0.68)	(0.62)	(0.57)	(0.62)	(0.56)
	65%	Bias	0.02	-0.29	-0.08	0.06	-0.04	0.01	0.05	0.02
		(SD)	(0.63)	(0.65)	(0.65)	(0.76)	(0.72)	(0.65)	(0.70)	(0.63)

Table 2 *Comparison for different estimators of* $\eta_0 = 2$ *when the covariate* Z *is i.i.d.* $Unif(0,1)$ *and the sample size equal to* 250

N=250 (Re-sampling Size=2000)										
True Parameter	Censoring Proportion		Partial Likelihood	$I(X_i \ge X_j), i \ne j$				$I(X_i \ge t)$		$I(X_i \le t)$
				Cheng et al.	Fine et al.	New				
						Δ_{ij}	Δ_{ij}^*	$(T_{0.05}, T_{0.95})$	$(T_{0.02}, T_{0.98})$	
$\eta_0 = 2$	0%	Bias	0.01	0.01				0.02	0.04	0.01
		(SD)	(0.25)	(0.25)				(0.26)	(0.30)	(0.25)
	14%	Bias	0.00	0.00	0.00	0.00	-0.05	0.01	0.04	0.00
		(SD)	(0.27)	(0.31)	(0.27)	(0.32)	(0.30)	(0.29)	(0.31)	(0.27)
	22%	Bias	0.01	0.01	0.01	0.01	-0.07	0.02	0.05	0.01
		(SD)	(0.25)	(0.29)	(0.26)	(0.31)	(0.29)	(0.29)	(0.33)	(0.25)
	30%	Bias	0.02	0.01	0.02	0.03	-0.08	0.02	0.06	0.02
		(SD)	(0.28)	(0.32)	(0.29)	(0.34)	(0.29)	(0.30)	(0.35)	(0.28)
	38%	Bias	0.02	-0.01	0.01	0.03	-0.09	0.01	0.04	0.02
		(SD)	(0.28)	(0.32)	(0.31)	(0.35)	(0.31)	(0.33)	(0.36)	(0.28)
	46%	Bias	0.02	-0.04	0.02	0.03	-0.08	-0.02	0.05	0.02
		(SD)	(0.30)	(0.34)	(0.33)	(0.38)	(0.33)	(0.34)	(0.38)	(0.30)
	54%	Bias	0.01	-0.13	0.00	0.02	-0.06	-0.01	0.08	0.01
		(SD)	(0.35)	(0.38)	(0.37)	(0.42)	(0.36)	(0.38)	(0.40)	(0.35)
	65%	Bias	0.00	-0.33	-0.01	0.02	-0.04	0.03	0.06	0.00
		(SD)	(0.37)	(0.40)	(0.40)	(0.46)	(0.43)	(0.40)	(0.42)	(0.37)

Table 3 *Comparison for different estimators of* $\eta_0 = 2$ *when the covariate* Z *is i.i.d.* $Ber(0.5)$ *and the sample size equal to* 100

N=100 (Re-sampling Size=2000)

True Parameter	Censoring Proportion		Partial Likelihood	$I(X_i \geq X_j),\ i \neq j$					$I(X_i \geq t)$		$I(X_i \leq t)$
				Cheng et al.	Fine et al.	Subramanian	New		$(T_{0.05}, T_{0.95})$	$(T_{0.02}, T_{0.98})$	
							Δ_{ij}	Δ_{ij}^{*}			
$\eta_0 = 2$	0%	Bias	0.04	0.04					0.06	0.04	0.04
		(SD)	(0.30)	(0.34)					(0.30)	(0.32)	(0.30)
	14%	Bias	0.03	0.05	0.05	0.05	0.05	0.00	0.05	-0.05	0.03
		(SD)	(0.30)	(0.36)	(0.32)	(0.35)	(0.36)	(0.35)	(0.32)	(0.35)	(0.30)
	22%	Bias	0.03	0.04	0.04	0.04	0.05	-0.03	0.02	0.06	0.03
		(SD)	(0.31)	(0.36)	(0.33)	(0.37)	(0.37)	(0.34)	(0.34)	(0.36)	(0.31)
	30%	Bias	0.04	0.06	0.06	0.02	0.06	-0.04	0.00	0.08	0.04
		(SD)	(0.33)	(0.40)	(0.35)	(0.39)	(0.41)	(0.35)	(0.36)	(0.39)	(0.33)
	38%	Bias	0.05	0.06	0.06	0.03	0.09	-0.06	0.01	-0.05	0.05
		(SD)	(0.34)	(0.39)	(0.35)	(0.40)	(0.42)	(0.36)	(0.37)	(0.41)	(0.34)
	46%	Bias	0.05	0.03	0.06	0.04	0.08	-0.06	0.02	0.04	0.05
		(SD)	(0.40)	(0.46)	(0.41)	(0.46)	(0.49)	(0.40)	(0.42)	(0.48)	(0.40)
	54%	Bias	0.06	0.12	0.08	0.03	0.07	-0.07	0.08	-0.06	0.06
		(SD)	(0.42)	(0.46)	(0.43)	(0.47)	(0.48)	(0.41)	(0.43)	(0.51)	(0.42)
	65%	Bias	0.08	-0.35	-0.09	-0.08	0.07	0.07	0.04	0.08	0.08
		(SD)	(0.46)	(0.51)	(0.46)	(0.52)	(0.55)	(0.45)	(0.49)	(0.53)	(0.46)

Table 4 *Comparison for different estimators of* $\eta_0 = 2$ *when the covariate* Z *is i.i.d.* $Ber(0.5)$ *and the sample size equal to* 250

N=250 (Re-sampling Size=2000)											
True Parameter	Censoring Proportion		Partial Likelihood	$I(X_i \geq X_j), i \neq j$					$I(X_i \geq t)$		$I(X_i \leq t)$
				Cheng et al.	Fine et al.	Subramanian	New		$(T_{0.05}, T_{0.95})$	$(T_{0.02}, T_{0.98})$	
							Δ_{ij}	Δ_{ij}^*			
$\eta_0 = 2$	0%	Bias	0.00	0.02					0.02	0.04	0.00
		(SD)	(0.17)	(0.20)					(0.19)	(0.21)	(0.17)
	14%	Bias	0.02	0.03	0.03	0.05	0.03	-0.03	0.03	0.05	0.02
		(SD)	(0.19)	(0.22)	(0.19)	(0.20)	(0.23)	(0.19)	(0.22)	(0.23)	(0.19)
	22%	Bias	0.03	0.02	0.03	-0.03	0.03	-0.06	-0.05	0.07	0.03
		(SD)	(0.18)	(0.21)	(0.19)	(0.21)	(0.22)	(0.21)	(0.22)	(0.24)	(0.18)
	30%	Bias	0.01	0.02	0.02	-0.02	0.02	-0.06	0.02	0.06	0.01
		(SD)	(0.20)	(0.24)	(0.21)	(0.24)	(0.25)	(0.21)	(0.23)	(0.25)	(0.20)
	38%	Bias	0.01	0.01	0.02	0.05	0.02	-0.08	0.01	-0.08	0.01
		(SD)	(0.21)	(0.23)	(0.20)	(0.22)	(0.24)	(0.22)	(0.24)	(0.28)	(0.21)
	46%	Bias	0.01	-0.02	0.02	0.02	0.03	-0.09	-0.01	0.06	0.01
		(SD)	(0.23)	(0.27)	(0.23)	(0.28)	(0.28)	(0.26)	(0.26)	(0.30)	(0.23)
	54%	Bias	0.03	-0.08	-0.02	0.02	0.05	-0.08	0.05	-0.05	0.03
		(SD)	(0.26)	(0.34)	(0.31)	(0.32)	(0.31)	(0.25)	(0.29)	(0.32)	(0.26)
	65%	Bias	0.03	-0.33	-0.07	0.08	0.04	-0.08	0.04	0.06	0.03
		(SD)	(0.30)	(0.33)	(0.29)	(0.33)	(0.35)	(0.30)	(0.33)	(0.35)	(0.30)

Chapter 7 Concluding Remarks

In this thesis, we consider semiparametric inference for linear transformation models which form a general class of regression models. We review existing literature under the classical framework of estimation theory. Specifically the method of moment and likelihood method are the two most important principles for parameter estimation. Here we see that how these approaches are adapted to the semi-parametric structure. The method of moment usually yields simpler solutions than the likelihood method in presence of high-dimensional nuisance parameter, namely $h(.)$, in our problem.

Another focus of thesis is to review how censoring is handled in an estimation procedure. The weight approach is appealing due to its simplicity. However we have seen that the suggested weight, in terms of the reciprocal of a Kaplan-Meier estimator, is sensitive to the tail estimation. When the censoring support lies within the support of the true lifetime, the weighting method can lead to bias solution. Nevertheless, Fine et al. (1998) proposed to further truncate the tail area to fix this problem. The method utilizing the martingale theory proposed by Chen, Jin and Ying (2002), is appealing in presence of censoring. First of all, it can be easily modified for censored data without estimation of $G(.)$. It does not need to select the range of integration as Cai et al. (2000) did for (τ_a, τ_b) since the expectation conditional on F_{t-} can update the most recent information and hence flexible. The likelihood approach is useful for the Cox model since the nuisance function gets cancelled out in estimating the conditional hazard. However likelihood inference becomes very complicated under the more general model setting.

Appendix

Appendix 1: *Prove* $\xi(\mathrm{s}) = \Pr(\varepsilon_i - \varepsilon_j \geq \mathrm{s}) = \int_{-\infty}^{\infty} \{1 - F_\varepsilon(t+s)\} dF_\varepsilon(\mathrm{t})$

Let $V = \varepsilon_i - \varepsilon_j$ and $\varepsilon \overset{\text{iid}}{\sim} \mathrm{f}_\varepsilon(\varepsilon)$

By convolution, we can find the following result:

$$f_T(t) = \int_{-\infty}^{\infty} f_{\varepsilon_i,\varepsilon_j}(v+\varepsilon_j, \varepsilon_j) d\varepsilon_j = \int_{-\infty}^{\infty} f_\varepsilon(v+\varepsilon_j) \times f_\varepsilon(\varepsilon_j) d\varepsilon = \int_{-\infty}^{\infty} f_\varepsilon(v+\varepsilon_j) dF_\varepsilon(\varepsilon_j).$$

Thus,

$$\begin{aligned}
\xi(\mathrm{s}) = \Pr(\varepsilon_i - \varepsilon_j \geq \mathrm{s}) &= \int_s^{\infty} \int_{-\infty}^{\infty} f_\varepsilon(v+\varepsilon_j) dF_\varepsilon(\varepsilon_j) dv \\
&= \int_{-\infty}^{\infty} \int_s^{\infty} f_\varepsilon(v+\varepsilon_j) dv \, dF_\varepsilon(\varepsilon_j) \\
&= \int_{-\infty}^{\infty} 1 - \left(\int_{-\infty}^{s} f_\varepsilon(s+\varepsilon_j) dv \right) dF_\varepsilon(\varepsilon_j) \\
&= \int_{-\infty}^{\infty} 1 - F_\varepsilon(\varepsilon_j + s) dF_\varepsilon(\varepsilon_j) \\
&= \int_{-\infty}^{\infty} \{1 - F(t+s)\} \, dF(t),
\end{aligned}$$

where $F(\mathrm{s}) = \Pr(\varepsilon \leq \mathrm{s})$.

Appendix 2: *Unbiased property for Cheng's*

Recall that Cheng's estimating equation under right censoring:

$$U(\eta) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{i \neq j} W(Z_{ij}{}^T \eta) Z_{ij} \times \left(\frac{\delta_j I(X_i \geq X_j)}{G^2(X_j)} - \xi(Z_{ij}{}^T \eta) \right), \qquad \text{(A-1)}$$

where $\xi(s) = \int_{-\infty}^{\infty} \{1 - F_\varepsilon(t+s)\} dF_\varepsilon(s)$.

Now, we want to show that equation (A-1) is unbiased. That is to say, we will prove

$$\mathrm{E}[\frac{\delta_j I(X_i \geq X_j)}{\mathrm{G}^2(X_j)} \mid Z_i, Z_j] = \xi(Z_{ij}{}^T \eta).$$

We use the conditional double expectation technique in the following:

$$
\begin{aligned}
\mathrm{E}[\frac{\delta_j I(X_i \geq X_j)}{\mathrm{G}^2(X_j)} \mid Z_i, Z_j] &= E[\frac{I(C_j \geq T_j) I(X_i \geq X_j)}{G^2(X_j)} | Z_i, Z_j] \\
&= E[E[\frac{I(min(C_j, X_i) \geq T_j)}{G^2(T_j)} | Z_i, Z_j, T_j]] \\
&= E[E[\frac{I(T_i \geq T_j) I(min(C_i, C_j) \geq T_j)}{G^2(T_j)} | Z_i, Z_j, T_j]] \\
&= E[\frac{\Pr(min(C_i, C_j) \geq T_j)}{G^2(T_j)} E[I(T_i \geq T_j) | Z_i, Z_j, T_j]] \\
&= E[E[I(T_i \geq T_j) | Z_i, Z_j, T_j]] \\
&= E[I(T_i \geq T_j) | Z_i, Z_j] \\
&= \Pr(h(T_i) \geq h(T_j) \mid Z_i, Z_j) \\
&= \Pr(\varepsilon_i - \varepsilon_j \geq Z_{ij}{}^T \eta \mid Z_i, Z_j) \\
&= \xi(Z_{ij}{}^T \eta).
\end{aligned}
$$

Appendix 3: *Unbiased property for Fine's*

Fine et al. added a truncation point t_0 to overcome the biased problem of Cheng et al.'s estimating equation. We recall Fine's estimating equation:

$$U_{\mathrm{f}}(\eta) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{i \neq j} W(Z_{ij}{}^T \eta) \xi_{ij}{}'(\eta) \times \left(\frac{\delta_j I(\min(X_i, t_0) \geq X_j)}{G^2(X_j)} - \xi^*{}_{ij}(\eta) \right), \qquad \text{(A-2)}$$

where $\xi^*{}_{ij}(\eta_0) = \int_{-\infty}^{h(t_0)} \{1 - F_\varepsilon(t + Z_{ij}{}^T \eta_0)\} dF_\varepsilon(t)$.

Now, we want to prove equation (A-2) is unbiased. Thus, we will show that

$$E\left[\frac{\delta_j I(\min(X_i,t_0)\ge X_j)}{G^2(X_j)}|Z_i,Z_j,t_o\right]=\xi^*_{ij}(\eta).$$

Again, we take advantage of the conditional double expectation technique in our procedure:

$$
\begin{aligned}
E[\frac{\delta_j I(min(X_i,t_0)\ge X_j)}{G^2(X_j)}|Z_i,Z_j,t_o] &= E[\frac{I(C_i\ge T_j)I(X_i\ge X_j)I(X_j\le t_0)}{G^2(X_j)}|Z_i,Z_j,t_o]\\
&= E[E[\frac{I(min(X_i,C_i)\ge T_j)I(T_j\le t_0)}{G^2(X_j)}|Z_i,Z_j,t_o,T_j]]\\
&= E[E[\frac{I(T_i\ge T_j)I(min(C_i,C_j)\ge T_j)I(T_j\le t_0)}{G^2(T_j)}|Z_i,Z_j,t_o,T_j]]\\
&= E[\frac{Pr(min(C_i,C_i)\ge T_j)}{G^2(T_j)}I(T_j\le t_0)E[I(T_i\ge T_j)|Z_i,Z_j,t_o,T_j]\\
&= E[I(T_j\le t_0)E[I(T_i\ge T_j)|Z_i,Z_j,t_o,T_j]]\\
&= E[I(T_j\le t_0)I(T_i\ge T_j)|Z_i,Z_j,t_o]\\
&= \Pr(T_j\le t_0,\ T_i\ge T_j|Z_i,Z_j t_o)\\
&= \Pr(T_i\ge T_j|Z_i,Z_j)\text{-}\Pr(T_j\ge t_0,\ T_i\ge T_j|Z_i,Z_j t_o)\\
&= \Pr(h(T_i)\ge h(T_j)|Z_i,Z_j)\ \text{-}\ \Pr(T_i\ge T_j\ge t_o|Z_i,Z_j,t_o)\\
&= \Pr(h(T_i)\ge h(T_j)|Z_i,Z_j)\text{-}\ \Pr(h(T_i)\ge h(T_j)\ge h(t_o)|Z_i,Z_j,t_o)\\
&= \int_{-\infty}^{h(t_0)}\{1-F_\varepsilon(t+Z_{ij}{}^T\eta)\}dF_\varepsilon(t)\\
&= \xi^*_{ij}(\eta).
\end{aligned}
$$

Appendix 4: *Unbiased property for New proposed method*

We take the observing pairs to be our main idea to develop our simple method. Therefore, we use an indicated function $\mathrm{I}((T_i\wedge T_j)<(C_i\wedge C_j))$ to construct our estimating equation. Now, we would show that new estimating equation we proposed is unbiased. For convenience, we let $\mathrm{I}((T_i\wedge T_j)<(C_i\wedge C_j))=\Delta_{\mathrm{ij}}$, and recall our simple estimating equation:

$$U_{\text{new}}(\eta)=\sum\sum_{i\neq j} W\left(Z_{ij}{}^{T}\eta\right)\times\xi'\left(Z_{ij}\eta\right)\times\Delta_{ij}\times\left(I\left(X_i\geq X_j\right)-\xi\left(Z_{ij}\eta\right)\right).$$

In the beginning, we want to find $\mathrm{E}\left[\mathrm{I}\left(X_i\geq X_j\right)-\Pr\left(T_i\geq T_j\right)\mid\left(T_i\wedge T_j\right)<\left(C_i\wedge C_j\right)\right]$. Thus, we use two cases to help us analyze the problem:

1. Given $\left(T_i\wedge T_j\right)=T_i$ and $\left(T_i\wedge T_j\right)<\left(C_i\wedge C_j\right)$

   Then T_i is the smallest of T_i, T_j, C_i and C_j

   Therefore, we can get $T_j \wedge C_j\geq T_i$ => $X_j\geq T_i$

2. Given $\left(T_i\wedge T_j\right)=T_j$ and $\left(T_i\wedge T_j\right)<\left(C_i\wedge C_j\right)$

   Then T_j is the smallest of T_i, T_j, C_i and C_j

   Therefore, we can get $T_i \wedge C_i\geq T_j$ => $X_i\geq T_j$

Then, we can make some inference based on above information:

$$\mathrm{E}\left[\mathrm{I}\left(X_i\geq X_j\right)-\Pr\left(T_i\geq T_j\right)\mid\left(T_i\wedge T_j\right)<\left(C_i\wedge C_j\right)\right]$$

$$=\mathrm{E}\left[\mathrm{I}\left(X_i\geq X_j\right)-\Pr\left(T_i\geq T_j\right)\mid\left(T_i\wedge T_j\right)<\left(C_i\wedge C_j\right),\left(T_i\wedge T_j\right)=T_i\right]\times\Pr\left(\left(T_i\wedge T_j\right)=T_i\right)+$$

$$\mathrm{E}\left[\mathrm{I}\left(X_i\geq X_j\right)-\Pr\left(T_i\geq T_j\right)\mid\left(T_i\wedge T_j\right)<\left(C_i\wedge C_j\right),\left(T_i\wedge T_j\right)=T_j\right]\times\Pr\left(\left(T_i\wedge T_j\right)=T_j\right)$$

$$=\Pr\left(\mathrm{T_i}\geq \mathrm{X_j}\mid\left(T_i\wedge T_j\right)<\left(C_i\wedge C_j\right),\left(T_i\wedge T_j\right)=T_i\right)\times\Pr\left(\left(T_i\wedge T_j\right)=T_i\right)+$$

$$\Pr\left(X_i\geq T_j\mid\left(T_i\wedge T_j\right)<\left(C_i\wedge C_j\right),\left(T_i\wedge T_j\right)=T_j\right)\times\Pr\left(\left(T_i\wedge T_j\right)=T_j\right)$$

$$=(0\text{-}0)\times\Pr\left(T_i\wedge T_j=T_i\right)+(1\text{-}1)\times\Pr\left(T_i\wedge T_j=T_j\right)$$

$$=0$$

Appendix 5: *Chen et al. (2002) is equal to partial likelihood method under Cox model*

Consider the special case of the Cox model, in which $\lambda(\mathrm{t})=\exp(\mathrm{t})$. We use this result $\lambda(\mathrm{t})=\exp(\mathrm{t})$, plug into equation (3.3). We can find the following result,

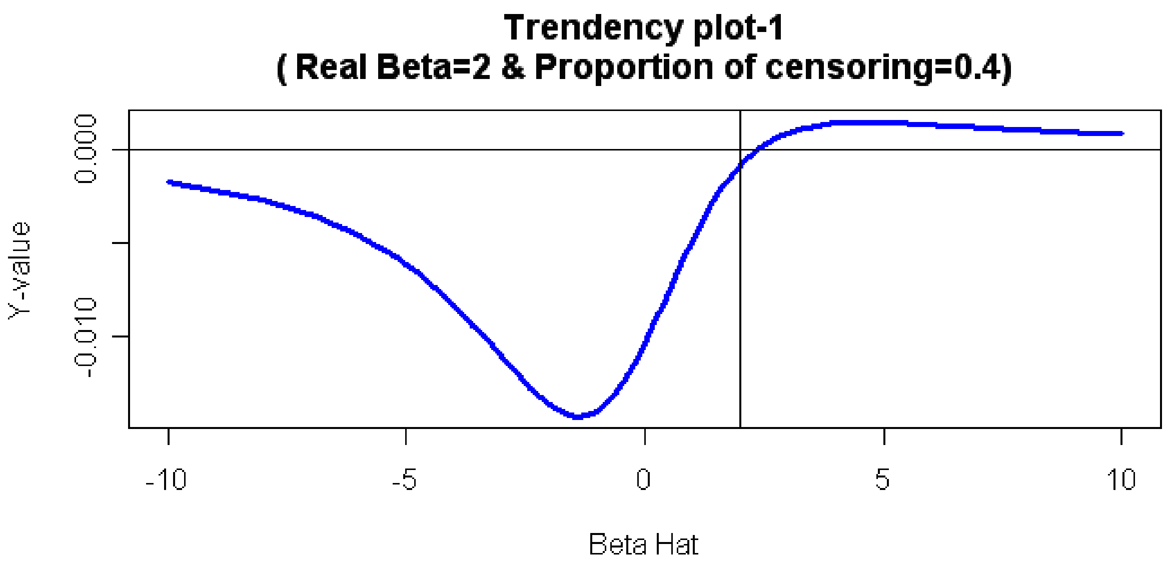
$$d\left[e^{h(t)}\right]=\frac{\sum_{i=1}^{n} dN_i(t)}{\sum_{i=1}^{n} Y_i(t)\times exp\left(Z^T\eta\right)}\ .$$

If we plug the result (3.5) in the martingale integral equation (3.4), we obtain

$$\mathrm{U}(\eta)=\sum_{\mathrm{i=1}}^{\mathrm{D}}[Z_{(i)}-\frac{\sum_{j=1}^{n} I\left(T_j \geq T_{(i)}\right) z_j \times exp\left(z_j^{\ T}\eta\right)}{\sum_{j=1}^{n} I\left(T_j \geq T_{(i)}\right)\times exp\left(z_j^{\ T}\eta\right)}]=0,$$

which is precisely the Cox partial likelihood score equation.

Appendix 6: *trend plot of new proposed method*



References

[1] Andersen, Per Kragh, Borgan, Ornulf, Gill, Richard D.& Keiding, Niels (1997) **Statistical Models Based on Counting Process.** Springer Series in Statistics

[2] Buckley, Jonathan & James, Ian (1979). **Linear Regression with Censored Data.** *Biometrika*, Vol. 66, No. 3., pp. 429-436.

[3] Cai, T., Wei L. J., Wilcox, M. (2000). **Semiparametric regression analysis for clustered failure time data.** *Biometrika*, 87, 4, *pp.* 867-878

[4] Chen, H. Y.(2001). **Fitting Semiparameteric Transformation Regression Models to Data form a Modified Case-Cohort Design.** *Biometrika*, 88, 1, *pp.* 255-268

[5] Chen, K., Jin, Z. & Ying, Z. (2002). **Semiparametric analysis of transformation models with censored data.** *Biometrika*, 89, 3, *pp.* 659-668

[6] Cheng, S. C., Wei, L. J. & Ying, Z. (1995). **Analysis of transformation models with censored data.** *Biometrika*, 82, 4, *pp.* 835-45

[7] Cheng, S. C., Wei, L. J. & Ying, Z. (1997). **Prediction of survival probabilities with semi-parametric transformation models.** *JASA*, 92, *pp.* 227-76

[8] Cox, D. R. (1972). **Regression models and life tables (with Discussion).** *J. R. Statist. Soc. B* 34, 187-220.

[9] Cox, D. R. (1975). **Partial likelihood.** *Biometrika* 62, 269-76.

[10] Fine, J. P., Ying, Z. & Wei, L. J. (1998). **On the linear transformation model with censored data.** *Biometrika*, 85, *pp.* 980-6

[11] Green, P.J. and Silverman, B.W. (1994) **Nonparametic Regression and Generalized Linear Models. A roughness penalty approach.** Monographs on Statistics and Applied Probability 58

[12] Klein, J. P. and Moeschberger, M. L. (1997). **Survival Analysis: Techniques for Censored and Truncated Data**, Springer-Verlag, New York.

[13] Miller, R. & Halpern, J. (1982). **Regression with censored data.** *Biometrika* 69, 521-31.

[14] Pettitt, A. N. (1984). **Proportional odds model for survival data and estimates using ranks.** *Appl. Statist.* 33, 169-75.

[15] Subramanian, S. (2003). **Semiparametric transformation models and the missing information principle.** *Journal of Statistical Planning and Inference* 115 327 V 348

[16] Subramanian, S. (2004). **A correction note on "Semiparametric transformation models and the missing information principle".** *Journal of Statistical Planning and Inference* 126 397 - 399

[17] Tsiatis, Anastasios A.(2006). **Semiparametric Theory and Missing Data.** Springer Series in Statistics