國立交通大學

統計學研究所

博 士 論 文

治癒模式之半母數迴歸分析

Semiparametric Regression Analysis in Presence of Non-susceptibility

研究生： 許秋婷

指導教授： 王維菁 教授

中 華 民 國 九十九 年 六 月

治癒模式之半母數迴歸分析

Semi-parametric Regression Analysis in Presence of Non-susceptibility

研　究　生：許秋婷　　Student：Chew-Teng Kor

指導教授：王維菁　教授　　Advisor：Weijing Wang

國 立 交 通 大 學

統 計 學 研 究 所

博 士 論 文

A Dissertation
Submitted to Institute of Statistics
College of Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in
Statistics

Institute of Statistics, National Chiao Tung University
Hsinchu, Taiwan, R.O.C.

中華民國九十九年六月

治癒模式之半母數迴歸分析

學生：許秋婷　　　　　　　　　　　　　　　　指導教授：王維菁

國立交通大學統計學研究所 博士班

摘　　　　　　　　　要

本論文針對存活資料，考慮“不受感染體質”(nonsusceptibility)者之存在，在混合模式架構下提出半母數迴歸分析方法。我們採用邏輯斯模式分析解釋變數與“發病與否”的關係。針對受感染體質者之“潛在發病時間”，我們探討兩類迴歸模式之推論問題。第一類模式包含常見的加速失敗模式和位移模式，我們利用計數程序之機率性質以建構估計函數，並進一步提出模式選取方法。第二類為線性轉換模式，包含等比風險模式與等比勝負比模式。我們採用概似函數法做為參數估計的原則，除了分析獨立設限的情況外，並進一步提出當存在競爭風險時，如何修正模式假設與推論方法。兩個研究方向都利用 EM 的技巧，以補插法處理感染體質不確定之觀測值。我們透過模擬實驗評估所提出方法在有限樣本下之表現。

關鍵字: 混合模式，不受感染體質，補插法，半母數線性迴歸，線性轉換模式，鞅估計函數，對數秩統計量，競爭風險。

Semi-parametric Regression Analysis in Presence of Non-susceptibility

Student: ChewTeng Kor Advisor:Weijing Wang

Institute of Statistics
National Chiao Tung University

Abstract

In this thesis, we consider semiparametric regression analysis for survival data in presence of non-susceptibility or cure. The mixture framework is adopted in analysis of such data. The incidence rate is assumed to follow the logistic regression model and the latency distribution is studied under two types of semiparametric regression models. One class refers to the semi-parametric linear regression model which includes the AFT and location-shift models as special cases. We propose estimating functions and also a model checking procedure based on properties of counting processes. The other class is known as transformation models which contain the proportional hazards model and proportional odds model. The likelihood principle is adopted for parameter estimation. We examine two situations of independent and dependent censoring respectively. In both research directions, the principle of EM is applied to handle uncertain susceptibility status. Simulation results are provided to examine the finite-sample properties of the proposed methods.

Keywords: Competing risk; EM, Logistic regression; Linear regression model; Latency distribution; Log-rank statistic; Transformation model; Martingale; Mixture model; Non-susceptibiblity.

誌　　　　　謝

人生的機遇總是很奇妙，如同蝴蝶效應般的影響著往後的人生規劃。或許很多決定是不假思索順勢而為，感謝上帝的眷顧，讓我擁有這段美麗的研究生生活。博士畢業也代表我在台灣 12 年學生旅程的 ending，接下來我將轉換身份繼續出發，期許自己能有更好的表現答謝支持鼓勵我的師長，家人和朋友們。

首先，我要感謝我的指導教授王維菁教授。我自覺自己是個迷糊，缺乏組織能力的學生。在這方面，王老師時而充滿著母愛般耐心的提醒我，時而如嚴父般地教誨改進我缺點，我內心充滿感激老師的用心良苦。回想當初在我研究遇到瓶頸，心情沮喪挫折之時，王老師比任何人更擔憂我的前途，也與我討論了許多改進的方向。我由衷的感謝老師的教育。

我還要感謝洪慧念教授的輔助指導，當我在理論證明或是模擬程式有困難的時候，洪老師總是不厭其煩的指導我，並指示出可能出錯的問題。我也很感謝洪老師分享生活上經驗予我參考。

另外，我也要感謝我的口試委員 鄭光甫教授，陳珍信教授和程毅豪教授特地在來新竹幫我口試。老師們撥空對我論文的架構提出寶貴的意見，使我的博士論文更為完善。當中老師們也提供了許多論文題目予我分享，對後續研究方向的確立幫助甚大。我將繼續保持謙卑，不倦的學習態度面對往後的研究工作，再次感謝老師們的指導與教育。

在交大八年的研究生生涯中，我也要感謝統計所所有的老師的教育。感謝所辦郭姐身兼諮商老師輔導我在家庭相處，婚姻感情的處理態度，我懷念跟郭姐一起午休散步的日子。感謝研究室的伙伴們和一起在課業打拼的同學們，我很高興有你們的陪伴。因為有你們的陪伴學習，使我在交大有美好的學習回憶。

在此，我要特別的感謝我先生新淵永遠支持我任何的決定。當我學習上遇到挫折沮喪時常常鼓勵我，並安慰我說“蹲下是為了跳更高”。當我在煩惱畢業問題時，總是安慰我說“不急慢慢來，讓我照顧你”。因為有你的陪伴，給了我重新出發勇氣。最後也要感謝我親愛的家人們，因為有你們的栽培，才能成就今天的我。

最後，僅將此論文獻給敬愛的家人，老師，同學和關心我的人，與我一同分享此論文的喜悅與榮耀。

許秋婷 謹誌於

國立交通大學統計所

中華民國九十九年六月

Table of Contents

1 Introduction 1

1.1 Literature background 1

1.2 Outline of the thesis 2

2 Literature Review for Semi-parametric Linear Models with Cure 4

2.1 Overview 4.

2.2 M-Estimation by Li & Taylor (2002) 7

2.3 Log-rank type Estimation by Zhang & Peng (2007) 8

2.4 Sketch of Numerical Algorithm for EM-type Estimation 9

3 Proposed Approach for Semiparametric Linear Models 11

3.1 Martingale estimating function based on complete data 11

3.2 The proposed estimating functions 12

3.3 Large sample analysis 13

3.4 Numerical algorithm and variance estimation 17

3.4.1 Re-sampling based on bootstrap approach 18

3.4.2 Re-sampling based on pivotal estimating function 18

3.5 Model Checking 20

3.6 Simulation analysis 22

3.6.1 Data generation 22

3.6.2 Simulation results 22

4 Literature Review for Transformation Models with Cure 25

4.1 Background 26

4.2 Different model expressions 26

4.3 Likelihood approach under the PH model 26
4.4 Moment approach for the class of transformation models 29
5 Proposed Approach for Transformation Cure Model under Independent Censoring 32
5.1 Model assumptions 32
5.2 Likelihood analysis for complete data 33
5.3 Imputation for handling missing data 36
5.4 Numerical algorithms 37
5.5 Simulation analysis 37
5.5.1 Data generation 37
5.5.2 Simulation results 38
6 Proposed Approach for Transformation Cure Model Under Dependent Censoring 40
6.1 Model assumptions 40
6.2 Imputation under the new models 41
6.3 Likelihood analysis under dependent censorship 42
6.4 Score equations under dependent censorship 45
6.5 Numerical algorithm 49
6.6 Simulation analysis 50
6.6.1 Data generation 50
6.6.2 Simulation results 52
Conclusion 53
References 54
Appendix 56
Tables and figures 69

Chapter 1

Introduction

1.1 Literature background

Traditional survival models assume that every subject in the study will eventually experience the event of interest. However, Kaplan-Meier curves based on empirical data often level off at the right tail and exhibit a stable plateau. Survival analysis which accounts for the possibility of cure or non-susceptibility has received increasing attentions in the literature since it provides reasonable explanations for some scientific phenomenon. The most popular approach to analyzing survival data in presence of cure is to represent the population as a mixture of susceptible and cured subjects. Define ς as the indicator of susceptibility. The population is divided into two groups: the susceptible with $\varsigma = 1$ and the cure with $\varsigma = 0$. For $\varsigma = 1$, define T as the time to the failure event. When $\varsigma = 0$, T is undefined or conventionally set to be infinity. Accordingly one can write

$$\Pr(T > t) = \Pr(T > t \mid \varsigma = 1)\Pr(\varsigma = 1) + \Pr(\varsigma = 0).$$

In presence of covariates denoted by Z, the mixture model can be written as

$$\Pr(T > t \mid Z) = \Pr(T > t \mid \varsigma = 1, Z)\Pr(\varsigma = 1 \mid Z) + \Pr(\varsigma = 0 \mid Z). \quad (1.1)$$

Under the above mixture framework, most literature assumes that the incidence function follows the logistic regression model which can be written as

$$\Pr(\varsigma = 1 \mid Z) = \pi(\theta_0 \mid Z) = \frac{\exp(Z^T\theta_0)}{1 + \exp(Z^T\theta_0)}. \quad (1.2)$$

Different proposals for modeling the latency variable $T \mid \varsigma = 1$ have appeared in the literature. Parametric models including the Weibull, generalized Gamma and generzlized F have been proposed by Farewell (1982), Yamaguchi (1992) and Peng et al. (1998) respectively. Semi-parametric models are more popular choices due to their flexibility and robustness. Most popular semi-parametric models, after some transformation, can be written as a linear regression form. For modeling $T \mid \varsigma = 1$, one can write

$$h(T) = Z^T \beta_0 + \varepsilon, \qquad (1.3)$$

where $\beta_0 : p \times 1$ is the unknown regression parameter of interest, $h(\cdot)$ is a monotone functions and ε is the error term whose distribution does not depend on Z.

Now we discuss two general classes of model (1.3). One type of models, which refers to semi-parametric linear models, assumes that $h(\cdot)$ is given but the error distribution is unknown. For example if $h(t) = \log(t)$, the model becomes an accelerated failure time model. If $h(t) = t$, it comes a location-shift model. Hence unknown parameters become β_0 and $F_\varepsilon^0(t) = \Pr(\varepsilon \le t)$. The other class, known as transformation models, assumes that $h(\cdot)$ is unknown but the distribution of ε is specified. For the Cox proportional hazards (PH) model, ε follows the extreme value distribution with $S_\varepsilon^0(t) = \exp\{-\exp(t)\}$ and $\Lambda_\varepsilon^0(t) = \exp(t)$. For the proportional odds model, ε follows the logistic distribution with $S_\varepsilon^0(t) = \exp(t)/\{1+\exp(t)\}$. Unknown parameters contain β_0 and $h(\cdot)$.

Nonparametric analysis for cure models with right censored observations may suffer from the inherent non-identifiability problem. A censored observation indicates two possible situations: the subject may be susceptible but the event has not occurred by the end of study; or he/she is cured. To distinguish the two different cases, the follow-up period has to be long enough to observe the susceptible ones as much as possible. The book of Maller and Zhou (1996) discusses the issue of identifiability and presents nonparametric tests to verify the condition of sufficient follow-up. Despite the theoretical contribution, these tests are not practical due to their low power. Therefore for practical applications, expert opinions about whether cure exists or not are important for choosing an appropriate model (Farewell, 1986).

1.2 Outline of the thesis

This thesis considers semi-parametric inference based on models in (1.2) and (1.3). The problem of non-identifiability is not as serious as in the nonparametric setting since additional model assumptions will be imposed. For the latency distribution, we consider both classes of models. We review the literature for semi-parametric linear models and present our proposal in Chapter 2 and 3 respectively. Then we consider transformation models in Chapters 4, 5 and 6. Chapter 4 reviews existing literature and Chapters 5 and 6 present our proposals under independent censoring and dependent censoring respectively. Chapter 7 contains concluding remarks.

Chapter 2 Literature Review for Semi-parametric Linear Models with Cure

2.1 Overview

Under the mixture framework, assume that

$$\Pr(\varsigma_i = 1 | Z_i) = \pi_i(\theta_0 | Z) = \frac{\exp(Z_i^T\theta_0)}{1+\exp(Z_i^T\theta_0)}$$

and for $\varsigma_i = 1$, we have

$$h(T_i) = Z_i^T\beta_0 + \varepsilon_i ,$$

where $h(\cdot)$ is specified and ε_i $(i=1,\ldots,n)$ form an iid sample with an unknown marginal distribution independent of Z_i. Define $f_\varepsilon^0(t)$, $F_\varepsilon^0(t)$, $S_\varepsilon^0(t)$ and $\Lambda_\varepsilon^0(t)$ as the density, distribution, survival and cumulative hazard functions of ε respectively, all of which are unspecified. In this chapter, we review existing literature for estimating (θ_0, β_0) in presence of the nuisance function $S_\varepsilon^0(t)$. Note that when $h(t) = \log(t)$, the model becomes the AFT model; and if $h(t) = t$, the model becomes the location-shift model.

Let C_i be the censoring variable for the *i*th subject. We will assume that C_i and T_i are independent. Denote observed data as $\{(X_i, \delta_i, Z_i),\ i=1,\ldots,n\}$, where $X_i = T_i \wedge C_i$ and $\delta_i = I(T_i \le C_i)$. Before we discuss specific methods, it is useful to examine the inference problem using the classical likelihood approach. One can express the data under the scale of the error variable. Let $\varepsilon_i(\beta) = h(T_i) - Z_i^T\beta$, $\varepsilon_i^C(\beta) = h(C_i) - Z_i^T\beta$ and $\tilde{\varepsilon}_i(\beta) = h(X_i) - Z_i^T\beta$. Note that $\varepsilon_i(\beta_0)$ has the same distribution as ε when β_0 is the true value of β. The likelihood function can be written as

$$L(\beta,\theta,f_\varepsilon^0) = \prod_{i=1}^{n} \left[\pi_i(\theta) f_\varepsilon^0\left(\tilde{\varepsilon}_i(\beta)\right)\right]^{I(\delta_i=1)} \times \left[\pi_i(\theta) S_\varepsilon^0\left(\tilde{\varepsilon}_i(\beta)\right) + (1-\pi_i(\theta))\right]^{I(\delta_i=0)}, \qquad (2.1a)$$

where $\pi_i(\theta) = \exp(Z_i^T\theta)/\{1+\exp(Z_i^T\theta)\}$.

The second component in the right-hand side of (2.1a) becomes complicated after taking logarithm.

The idea of EM algorithm is often adopted in statistical inference of cure models. If "complete" data denoted as $\{(X_i, Z_i, \varsigma_i, \delta_i),\ i=1,...,n\}$ are available, the above likelihood in (2.1a) can be simplified as:

$$\prod_{i=1}^{n} \left\{ \left[\pi_i(\theta) f_\varepsilon^0\left(\tilde{\varepsilon}_i(\beta)\right)\right]^{I(\delta_i=1)} \left[S_\varepsilon^0\left(\tilde{\varepsilon}_i(\beta)\right)\pi_i(\theta)\right]^{I(\delta_i=0,\varsigma_i=1)} \left[1-\pi_i(\theta)\right]^{I(\delta_i=0,\varsigma_i=0)} \right\}. \quad (2.1b)$$

The resulting log-likelihood function can be written as:

$$\log L(\beta,\theta,f_\varepsilon^0) = \ell(\theta,\beta,f_\varepsilon^0) = \ell_1(\theta) + \ell_2(\beta,f_\varepsilon^0)$$

where

$$\ell_1(\theta) = \sum_{i=1}^{n} \varsigma_i \log \pi_i(\theta) + \sum_{i=1}^{n} (1-\varsigma_i)\cdot \log\{1-\pi_i(\theta)\}; \quad (2.3a)$$

$$\ell_2(\beta,f_\varepsilon^0) = \sum_{i=1}^{n} \delta_i \log f_\varepsilon^0\left(\tilde{\varepsilon}_i(\beta)\right) + \sum_{i=1}^{n} (1-\delta_i)\cdot\varsigma_i \log S_\varepsilon^0\left(\tilde{\varepsilon}_i(\beta)\right). \quad (2.3b)$$

Notice that the parameters θ and β become separated in (2.3a) and (2.3b) respectively. Accordingly the score functions for θ and β become

$$\partial\ell_1(\theta)/\partial\theta = \sum_{i=1}^{n} \varsigma_i \frac{\dot{\pi}_i(\theta)}{\pi_i(\theta)} + \sum_{i=1}^{n} (1-\varsigma_i)\cdot \frac{\{-\dot{\pi}_i(\theta)\}}{1-\pi_i(\theta)}, \quad (2.4a)$$

$$\partial\ell_2(\beta,f_\varepsilon^0)/\partial\beta = \sum_{i=1}^{n} Z_i \left[-\delta_i \frac{\dot{f}_\varepsilon^0(\tilde{\varepsilon}_i(\beta))}{f_\varepsilon^0(\tilde{\varepsilon}_i(\beta))} + (1-\delta_i)\varsigma_i \frac{f_\varepsilon^0(\tilde{\varepsilon}_i(\beta))}{S_\varepsilon^0(\tilde{\varepsilon}_i(\beta))}\right], \quad (2.4b)$$

where $\dot{f}(t) = \partial f(t)/\partial t$ and $\dot{\pi}(\theta) = \partial\pi(\theta)/\partial\theta$. The above derivations imply that θ and β can be estimated separately if the value of ς_i could be observed for all $i=1,...,n$ and $f_\varepsilon^0(t)$, or at least its parametric form, is known. However these two conditions often do not hold in practical applications. Now we discuss how to handle these problems.

To deal with possibly unknown value of ς_i, a common approach is to replace it by an imputed value, often an estimate of its conditional mean given observed data. Notice that when $\delta_i = 1$, $\varsigma_i = 1$; but when $\delta_i = 0$, ς_i is unknown. It follows that

$$E(\varsigma_i \mid X_i, \delta_i, Z_i) = \delta_i + E(\varsigma_i \mid T_i > C_i = X_i, Z_i).$$

Under the imposed models, we write

$$E(\varsigma_i \mid X_i, \delta_i, Z_i) = \delta_i + (1-\delta_i) w_i(\theta_0, \beta_0, S_\varepsilon^0), \tag{2.5}$$

where

$$w_i(\theta, \beta, S_\varepsilon) \quad = \frac{\pi_i(\theta) \times S_\varepsilon(\tilde{\varepsilon}_i(\beta))}{\pi_i(\theta) \times S_\varepsilon(\tilde{\varepsilon}_i(\beta)) + \{1-\pi_i(\theta)\}}. \tag{2.6}$$

In estimation, the weight $w_i(\theta, \beta, S_\varepsilon)$ is often treated as a fixed value by plugging in previous estimates of $(\theta_0, \beta_0, S_\varepsilon^0)$. This technique is commonly seen for analyzing missing data.

Under the semi-parametric setting, the major challenge is the log-likelihood function in (2.3b) or the score equation in (2.4b) which involves the nuisance functions $\dot{f}_\varepsilon^0(.)$, $f_\varepsilon^0(.)$ and $S_\varepsilon^0(.)$, the first two of which are complicated. Existing methods try to get rid of the density function $f_\varepsilon^0(.)$ in the estimation but still keep the survival function $S_\varepsilon^0(.)$ since it is easier to handle. To see this, there exist two estimators of $S_\varepsilon^0(t)$ based on complete data given by

$$\hat{S}_\varepsilon^0(t \mid \beta) = \prod_{u \le t} \{1 - \frac{\sum_{i=1}^n I(\tilde{\varepsilon}_i(\beta) = u, \delta_i = 1)}{\sum_{i=1}^n \varsigma_i I(\tilde{\varepsilon}_i(\beta) \ge u)}\}, \tag{2.7a}$$

$$\hat{S}_\varepsilon^0(t \mid \beta) = \exp(-\sum_{u \le t} \frac{\sum_{i=1}^n I(\tilde{\varepsilon}_i(\beta) = u, \delta_i = 1)}{\sum_{i=1}^n \varsigma_i I(\tilde{\varepsilon}_i(\beta) \ge u)}). \tag{2.7b}$$

Note we use the same notations in (2.7a) and (2.7b) to simplify the presentation since these two functions are asymptotically equivalent. Replacing ς_i by $\delta_i + (1-\delta_i) w_i(\theta, \beta, S_\varepsilon)$, we have

$$\hat{S}_\varepsilon(t \mid \beta, w) = \prod_{u \le t} \{1 - \frac{\sum_{i=1}^n I(\tilde{\varepsilon}_i(\beta) = u, \delta_i = 1)}{\sum_{i=1}^n \{\delta_i + (1-\delta_i) w_i\} I(\tilde{\varepsilon}_i(\beta) \ge u)}\}; \tag{2.8a}$$

or

$$\hat{S}_\varepsilon(t \mid \beta, w) = \exp\{-\sum_{u \le t} \frac{\sum_{i=1}^{n} I(\tilde{\varepsilon}_i(\beta) = u, \delta_i = 1)}{\sum_{i=1}^{n} \{\delta_i + (1-\delta_i) w_i\} I(\tilde{\varepsilon}_i(\beta) \ge u)}\}, \tag{2.8b}$$

where $w_i = w_i(\theta, \beta, S_\varepsilon)$ and $w = \{w_j, j = 1, \ldots, n\}$. Since $\hat{S}_\varepsilon(t \mid \beta, w)$ depends on $(\theta, \beta, S_\varepsilon)$, the expressions in (2.8a) and (2.8b) are not explicit estimators of $S_\varepsilon(t)$ but can be used as an estimating equation along with the score equations in (2.4a) and (2.4b) or its modified version.

We now introduce two papers which provide different ways of modifying the second score equation. Note that, since the transformation $h(\cdot)$ is known, the papers usually assume the accelerated failure time model with $h(t) = \log(t)$.

2.2 M-Estimation by Li & Taylor (2002)

Li and Taylor (2002) extended the idea of M-estimators by Ritov (1990) to cure models. First, the covariate Z is centered to exclude the unknown intercept term:

$$\sum_{i=1}^{n} (Z_i - \bar{Z}) \left[-\delta_i \varsigma_i \frac{\dot{f}_\varepsilon^0(\tilde{\varepsilon}_i(\beta))}{f_\varepsilon^0(\tilde{\varepsilon}_i(\beta))} + (1-\delta_i)\varsigma_i \frac{f_\varepsilon^0(\tilde{\varepsilon}_i(\beta))}{S_\varepsilon^0(\tilde{\varepsilon}_i(\beta))} \right], \tag{2.9}$$

where $\bar{Z} = \sum_{i=1}^{n} Z_i / n$. Following Ritov (1990), Li and Taylor (2002) suggested to replace $-\frac{\dot{f}_\varepsilon^0(.)}{f_\varepsilon^0(.)}$ by a reasonable score function $g(.)$. Notice that

$$f_\varepsilon^0(t) = \int_t^\infty -\frac{\dot{f}_\varepsilon^0(x)}{f_\varepsilon^0(x)} f_\varepsilon^0(x) dx = \int_t^\infty -\frac{\dot{f}_\varepsilon^0(x)}{f_\varepsilon^0(x)} dF_\varepsilon^0(x)$$

given that $f_\varepsilon^0(\infty) = 0$. Accordingly $f_\varepsilon(\tilde{\varepsilon}_i(\beta))$ can be replaced by $\int_{\tilde{\varepsilon}_i(\beta)}^{\infty} g(x) dF_\varepsilon(x)$. Here are the examples of $g(.)$ given in the paper:

(i) $g(u) = u$;

(ii) $g(u) = \begin{cases} -3 & \text{if } u < -3 \\ u & \text{if } |u| \le 3 \\ 3 & \text{if } u > 3 \end{cases}$.

Finally Li and Taylor (2002) proposed to modify (2.4b) by

$$U^{LT}(\beta \mid w, S_\varepsilon) = \sum_{i=1}^{n} (\mathrm{Z_i} - \bar{Z}) \left[\delta_i g(\tilde{\varepsilon}_i(\beta)) + (1-\delta_i) w_i \frac{\int_{u=\tilde{\varepsilon}_i(\beta)}^{\infty} g(u) dF_\varepsilon(u)}{S_\varepsilon(\tilde{\varepsilon}_i(\beta))} \right]. \tag{2.10}$$

2.3 Log-rank type Estimation by Zhang & Peng (2007)

The log-likelihood function in (2.1b) is expressed in terms of density and survival functions. Zhang and Peng (2007) re-wrote the function in terms of hazard and survival function such that

$$\ell_2(\beta, f_\varepsilon^0) = \sum_{i=1}^{n} \varsigma_i \delta_i \log[\lambda_\varepsilon^0(\tilde{\varepsilon}_i(\beta))] + \varsigma_i \log[S_\varepsilon^0(\tilde{\varepsilon}_i(\beta))]. \tag{2.11}$$

Zhang and Peng (2007) found new insights from (2.11). Specifically, replacing ς_i by $\tilde{\varsigma}_i = \delta_i + (1-\delta_i) w_i$, the above function can be written as

$$\tilde{l}_2(\beta, f_\varepsilon^0) = \sum_{i=1}^{n} \delta_i \log[\lambda_\varepsilon^0(\tilde{\varepsilon}_i(\beta))] + \tilde{\varsigma}_i \log[S_\varepsilon^0(\tilde{\varepsilon}_i(\beta))] \tag{2.12a}$$

which equals

$$\sum_{i=1}^{n} \delta_i \log[\tilde{\varsigma}_i \lambda_\varepsilon^0\{\tilde{\varepsilon}_i(\beta)\}] + \tilde{\varsigma}_i \log[S_\varepsilon^0\{\tilde{\varepsilon}_i(\beta)\}]. \tag{2.12b}$$

In particular, the expression in (2.12b) can be viewed as the likelihood from the model such that

$$h(T_i) = Z_i^T \beta + \varepsilon_i^*, \tag{2.13a}$$

where ε_i^* has the hazard function $\tilde{\varsigma}_i \lambda_\varepsilon^0(\cdot)$. Notice that the problem in (2.13a) becomes a semi-parametric model without cure. It has the form of a semi-parametric linear model since $h(.)$ is specified while the distribution of ε_i^* is unknown.

The proposal of Zhang and Peng (2007) was motivated by the work of Wei (1992) who incorporated the rank estimation method under the framework of PH models. Specifically (2.13a) can be written as

$$\varepsilon_i^*(\beta) = h(T_i) - Z_i^T \beta .$$

As mentioned earlier ε_i^* has the hazard function $\tilde{\varsigma}_i \lambda_\varepsilon^0(\cdot)$. Consider a more general type of proportional hazards model for ε^*

$$\lambda_{\text{PH}}(t) = \tilde{\varsigma}\lambda_\varepsilon^0(t)\exp(Z^T\gamma)\,. \tag{2.14a}$$

The score equation for γ deriving from the partial likelihood function based on model (2.14a) is given by

$$\psi(\gamma) = \sum_{i=1}^n \delta_i \left(Z_i - \frac{\sum_{j=1}^n Z_j \tilde{\varsigma}_j \exp(Z_j^T\gamma) I\{\varepsilon_j^*(\beta) \ge \varepsilon_i^*(\beta)\}}{\sum_{j=1}^n \tilde{\varsigma}_j \exp(Z_j^T\gamma) I\{\varepsilon_j^*(\beta) \ge \varepsilon_i^*(\beta)\}} \right). \tag{2.14b}$$

Notice that (2.14b) has the form of log-rank statistics. When $\gamma = 0$, which reduces to the true model (2.13a), the above score function becomes

$$\psi(0) = \sum_{i=1}^n \delta_i \left(Z_i - \frac{\sum_{j=1}^n Z_j \tilde{\varsigma}_j I(\tilde{\varepsilon}_j(\beta) \ge \tilde{\varepsilon}_i(\beta))}{\sum_{j=1}^n \tilde{\varsigma}_j I(\tilde{\varepsilon}_j(\beta) \ge \tilde{\varepsilon}_i(\beta))} \right). \tag{2.14c}$$

Zhang and Peng (2007) suggested to add a weight function $\psi(0)$ to (2.14c) and proposed the following estimating function:

$$U^{ZP}(\beta \mid w) = \sum_{i=1}^n \delta_i W\{\tilde{\varepsilon}_i(\beta)\} \left(Z_i - \frac{\sum_{j=1}^n Z_j \tilde{\varsigma}_j I\{\tilde{\varepsilon}_j(\beta) \ge \tilde{\varepsilon}_i(\beta)\}}{\sum_{j=1}^n \tilde{\varsigma}_j I\{\tilde{\varepsilon}_j(\beta) \ge \tilde{\varepsilon}_i(\beta)\}} \right), \tag{2.15}$$

where $W(.)$ is a weight function and $\tilde{\varsigma}_i = \delta_i + (1-\delta_i)w_i$ with w_i depending on $(\theta, \beta, S_\varepsilon)$.

2.4 Sketch of Numerical Algorithm for EM-type Estimation

Now we discuss how to implement the estimation procedures which will also be adopted by the proposed approach discussed in the next chapter. We need to solve two estimating equations: $\kappa(\theta \mid w) = 0$ and $U^*(\beta \mid w) = 0$ where

$$\kappa(\theta \mid w) = \sum_{i=1}^n \tilde{\varsigma}_i \frac{\dot{\pi}_i(\theta)}{\pi_i(\theta)} + \sum_{i=1}^n (1-\tilde{\varsigma}_i) \cdot \frac{\{-\dot{\pi}_i(\theta)\}}{1-\pi_i(\theta)}$$

and * = "LT" based on (2.10) or "ZP" based on (2.15). Define $U^{LT}(\beta \mid w) = U^{LT}(\beta \mid w, \hat{S}_\varepsilon)$ since $\hat{S}_\varepsilon$ depends on (β, w) and the data.

For numerical implementation, let $w^{(m)}$ be the mth step estimate of w based on $(\hat{\theta}^{(m)}, \hat{\beta}^{(m)}, \hat{S}_{\varepsilon}^{(m)})$. It is used to solve $\kappa(\theta \mid w^{(m)}) = 0$ and $U^{*}(\beta \mid w^{(m)}) = 0$ to obtain $(\hat{\theta}^{(m+1)}, \hat{\beta}^{(m+1)})$ and

$$\hat{S}_{\varepsilon}^{(m+1)}(t) = \exp\{-\sum_{u \le t} \frac{\sum_{i=1}^{n} I(\tilde{\varepsilon}_i(\beta) = u, \delta_i = 1)}{\sum_{i=1}^{n} \delta_i I(\tilde{\varepsilon}_i(\beta) \ge u) + \sum_{i=1}^{n} w^{(m)} I(\tilde{\varepsilon}_i(\beta) \ge u)}\}.$$

The procedure is repeated for $m = 0, 1, 2, \ldots$ until convergence.

It is important to note that solving $U^{LT}(\beta \mid w^{(m)}, \hat{S}_{\varepsilon}^{(m)}) = 0$ is more difficult than $U^{ZP}(\beta \mid w^{(m)}) = 0$ since $\hat{S}_{\varepsilon}^{(m)}$ plays a more important role in the former equation. As a result, a grid search with a large number of finely spaced points is suggested by Li and Taylor (2002). In the simulation studies conducted by Zhang and Peng (2007), the estimator proposed by Li and Taylor (2002) may fail to produce a consistent estimator.

The dependency of w on $(\theta, \beta, S_{\varepsilon})$ complicates theoretical analysis. Both papers did not derive asymptotic properties of their proposed estimators. The bootstrap approach was suggested by Zhang and Peng (2007) for variance estimation. We will briefly discuss this approach in the next chapter.

Chapter 3 Proposed Approach for Semiparametric Linear Models

In this chapter we present our proposal to replace the second score function in (2.4b):

$$\partial \ell_2(\beta, f_\varepsilon^0)/\partial\beta = \sum_{i=1}^{n} Z_i^T \left[-\delta_i \frac{\dot{f}_\varepsilon^0(\tilde{\varepsilon}_i(\beta))}{f_\varepsilon^0(\tilde{\varepsilon}_i(\beta))} + (1-\delta_i)\varsigma_i \frac{f_\varepsilon^0(\tilde{\varepsilon}_i(\beta))}{S_\varepsilon^0(\tilde{\varepsilon}_i(\beta))} \right].$$

3.1 Martingale estimating function based on complete data

Temporarily we assume that the information of ς_i is available. Recall that $\varepsilon_i(\beta) = h(T_i) - Z_i^T\beta$ and $\tilde{\varepsilon}_i(\beta) = h(X_i) - Z_i^T\beta$. Define the observable counting process for $\varepsilon_i(\beta)$ as

$$N_i(t;\beta) = I(\tilde{\varepsilon}_i(\beta) \le t, \delta_i = 1, \varsigma_i = 1) = I(\tilde{\varepsilon}_i(\beta) \le t, \delta_i = 1), \tag{3.1a}$$

the at-risk process for a susceptible subject:

$$Y_i(t;\beta) = I(\tilde{\varepsilon}_i(\beta) \ge t, \varsigma_i = 1) \tag{3.1b}$$

and the corresponding filtration for the susceptible group:

$$F_t(t,\beta) = \sigma\left\{ I(\tilde{\varepsilon}_i(\beta) \le u, \delta_i = 1), I(\tilde{\varepsilon}_i(\beta) \le u, \delta_i = 0, \varsigma_i = 1), Z_i^T \mid 0 < u \le t, i = 1,\ldots,n \right\}.$$

Define

$$M_i(t;\beta) = N_i(t;\beta) - \int_0^t Y_i(u;\beta)\, d\Lambda_\varepsilon^0(u). \tag{3.1c}$$

When β equals its true value β_0, the Doob-Meyer decomposition says that $M_i(t;\beta_0)$ is a mean-zero martingale with respect to $F_t(t,\beta_0)$.

The martingale property of $M_i(t;\beta_0)$ can be used to construct an estimating function for β when ς_i $(i = 1,\ldots,n)$ are available. Consider

$$\tilde{U}(\beta) = \int_0^\infty \sum_{i=1}^{n} Z_i \left\{ dN_i(t;\beta) - Y_i(t;\beta)\, d\tilde{\Lambda}_\varepsilon(t;\beta) \right\}, \tag{3.2a}$$

where

$$\tilde{\Lambda}_\varepsilon(t \mid \beta) = \int_0^t \sum_{j=1}^{n} dN_j(u;\beta) \Big/ \sum_{j=1}^{n} Y_j(u;\beta). \tag{3.2b}$$

We can express $\tilde{U}(\beta)$ in terms of the log-rank statistics. It follows that

$$\sum_{i=1}^{n} Z_i Y_i(t;\beta) d\tilde{\Lambda}_{\varepsilon}(t;\beta) = \sum_{i=1}^{n} Z_i Y_i(t;\beta) \frac{\sum_{j=1}^{n} dN_j(t;\beta)}{\sum_{j=1}^{n} Y_j(t;\beta)} = \sum_{j=1}^{n} \tilde{Z}(t;\beta) dN_j(t;\beta),$$

where $\tilde{Z}(t;\beta) = \sum_{i=1}^{n} Z_i Y_i(t;\beta) / \sum_{j=1}^{n} Y_j(t;\beta)$. Accordingly we can write

$$\tilde{U}(\beta) = \int_0^{\infty} \sum_{i=1}^{n} Z_i \left\{ dN_i(t;\beta) - Y_i(t;\beta) d\tilde{\Lambda}_{\varepsilon}(t;\beta) \right\}. \tag{3.3a}$$

$$= \int_0^{\infty} \sum_{i=1}^{n} \{ Z_i - \tilde{Z}(t;\beta) \} dN_i(t;\beta). \tag{3.3b}$$

3.2 The proposed estimating functions

A possibly unknown ς_i can be replaced by its imputed value: $\tilde{\varsigma}_i = \delta_i + (1-\delta_i) w_i$, where

$$w_i = w_i(\theta, \beta, S_{\varepsilon}) = \frac{\pi_i(\theta) \times S_{\varepsilon}(\tilde{\varepsilon}_i(\beta))}{\pi_i(\theta) \times S_{\varepsilon}(\tilde{\varepsilon}_i(\beta)) + \{1 - \pi_i(\theta)\}}.$$

The at-risk process $Y_i(t;\beta)$ can be replaced by

$$\tilde{Y}_i(t \mid \beta, w_i) = I(\tilde{\varepsilon}_i(\beta) \geq t, \delta_i = 1) + w_i I(\tilde{\varepsilon}_i(\beta) \geq t, \delta_i = 0)$$

and define

$$\bar{Z}(t;\beta,w) = \sum_{i=1}^{n} Z_i \tilde{Y}_i(t;\beta,w_i) / \sum_{i=1}^{n} \tilde{Y}_i(t;\beta,w_i)$$

and

$$\hat{\Lambda}_{\varepsilon}(t \mid \beta, w) = \int_0^t \sum_{i=1}^{n} dN_i(u;\beta) / \sum_{j=1}^{n} \tilde{Y}_i(u;\beta,w_i).$$

We propose the following estimating function for β

$$U(\beta \mid w) = \int_0^{\infty} \sum_{i=1}^{n} Z_i \left\{ dN_i(t;\beta) - \tilde{Y}_i(t;\beta,w_i) d\hat{\Lambda}_{\varepsilon}(t;\beta,w) \right\} \tag{3.4a}$$

$$= \int_0^{\infty} \sum_{i=1}^{n} \{ Z_i - \bar{Z}(t;\beta,w) \} dN_i(t;\beta), \tag{3.4b}$$

where $\bar{Z}(t;\beta,w) = \sum_{i=1}^{n} Z_i \tilde{Y}_i(t;\beta,w_i) / \sum_{i=1}^{n} \tilde{Y}_i(t;\beta,w_i)$. Recall that the other estimating function of θ is

$$\kappa(\theta \mid w) = \sum_{i=1}^{n} \tilde{\varsigma}_i \frac{\dot{\pi}_i(\theta)}{\pi_i(\theta)} + \sum_{i=1}^{n} (1 - \tilde{\varsigma}_i) \frac{\{-\dot{\pi}_i(\theta)\}}{1 - \pi_i(\theta)},$$

where $\dot{\pi}(\theta) = \frac{\partial}{\partial \theta} \pi(\theta)$.

The expression of $U(\beta \mid w)$ in (3.4b) is equivalent to $U^{ZP}(\beta \mid w)$ proposed by Zhang and Peng (2007) despite that the two proposals are developed based on different ideas. Nevertheless our approach starts from the concept of martingales which provides a useful framework for further analysis including large-sample analysis, variance estimation and model checking.

3.3 Large sample analysis

Recall that the proposed estimators of (θ_0, β_0), denoted as $(\hat{\theta}, \hat{\beta})$ solve

$$\begin{bmatrix} \kappa(\theta \mid w\{\theta, \beta, \hat{S}_\varepsilon\}) \\ U(\beta \mid w\{\theta, \beta, \hat{S}_\varepsilon\}) \end{bmatrix} = \begin{bmatrix} \kappa(\theta \mid \hat{w}) \\ U(\beta \mid \hat{w}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

where $\hat{w} = \left\{ \hat{w}_j, j = 1, ..., n \right\}$ and $\hat{w}_i = w_i\{\theta, \beta, \hat{S}_\varepsilon(. \mid \beta, \hat{w})\}$,

$$\hat{S}_\varepsilon(t \mid \beta, w) = \exp(-\sum_{u \le t} \frac{\sum_{i=1}^{n} I(\tilde{\varepsilon}_i(\beta) = u, \delta_i = 1)}{\sum_{i=1}^{n} \{\delta_i + (1-\delta_i) w_i\} I(\tilde{\varepsilon}_i(\beta) \ge u)}).$$

We also define

$$w_i^* = w_i(\theta, \beta, S_\varepsilon^0) = \frac{\pi_i(\theta) \times S_\varepsilon^0(\tilde{\varepsilon}_i(\beta))}{\pi_i(\theta) \times S_\varepsilon^0(\tilde{\varepsilon}_i(\beta)) + \{1 - \pi_i(\theta)\}},$$

where $S_\varepsilon^0(\cdot)$ is true survival function. It is not easy to establish asymptotic properties of $(\hat{\theta}, \hat{\beta})$ jointly since $\hat{w}$ still depends on (θ, β) in a complicated way. To precede the theoretical development, we need to assume

$$\textit{Assumption:} \ \sup_i \left\| \hat{w}_i - w_i^* \right\| \le o(n^{-1/3}) \ \textit{a.s. for all} \ \theta \ \textit{and} \ \beta.$$

Note that the quality of weights still plays an important role. We ran simulations to evaluate the effect of using arbitrary weights but the results lead to a biased estimator of β. The imposed assumption is a condition to assure that $\hat{w}$ is a good weight. Due to this assumption, $\kappa(\theta \mid \hat{w})$ can be ignored in the evaluation of $\hat{\beta}$ since $\hat{\theta}$ affects $\hat{\beta}$ only through $\hat{w}$.

Now we can focus on

$$U(\beta \mid \hat{w}) = \int_0^\infty \sum_{i=1}^n \{Z_i - \bar{Z}(t;\beta,\hat{w})\} dN_i(t;\beta). \tag{3.5}$$

Temporarily ignoring the estimated weight, first we examine the property of

$$U(\beta \mid w^*) = \int_0^\infty \sum_{i=1}^n \{Z_i - \bar{Z}(t;\beta,w^*)\} dN_i(t;\beta), \tag{3.6}$$

where w^* is the true weight. Following Lin et al. (1998), we can write $U(\beta \mid w^*) - U(\beta_0 \mid w^*)$ as the sum of the following three terms:

$$B_{1n} = \int_0^\infty \sum_{i=1}^n \left\{ Z_i - \bar{Z}(t;\beta,w^*) \right\} \left(dN_i(t;\beta) - Y_i^*(t;\beta) d\Lambda_i^*(t;\beta) \right),$$

$$B_{2n} = -\int_0^\infty \sum_{i=1}^n \left\{ Z_i - \bar{Z}(t;\beta_0,w^*) \right\} \left(dN_i(t;\beta_0) - Y_i^*(t;\beta_0) d\Lambda_\varepsilon^*(t;\beta_0) \right),$$

$$B_{3n} = \int_0^\infty \sum_{i=1}^n \left\{ Z_i - \bar{Z}(t;\beta,w^*) \right\} Y_i^*(t;\beta) \left(d\Lambda_i^*(t;\beta) - d\Lambda_\varepsilon^*(t;\beta_0) \right)$$

where $Y_i^*(t;\beta) = \tilde{Y}_i(t \mid \beta, w_i^*)$ and $\Lambda_\varepsilon^*(t;\beta_0)$ is the limit of

$$\int_0^t \sum_{j=1}^n dN_j(u;\beta_0) / \sum_{j=1}^n \tilde{Y}_j(u;\beta_0,w_j^*)$$

and $\Lambda_i^*(t;\beta) = \Lambda_\varepsilon^*\left((t - (\beta_0 - \beta) Z_i); \beta_0\right)$. Note that

$$\int_0^\infty \sum_{i=1}^n \left\{ Z_i - \bar{Z}(t;\beta_0,w^*) \right\} Y_i^*(t;\beta_0) d\Lambda_\varepsilon^*(t;\beta_0) = 0,$$

And

$$\int_0^\infty \sum_{i=1}^n \left\{ Z_i - \bar{Z}(t;\beta,w^*) \right\} Y_i^*(t;\beta) d\Lambda_\varepsilon^*(t;\beta_0) = 0,$$

where $\dot{\lambda}_\varepsilon^*(t;\beta_0) = \partial \lambda_\varepsilon^*(t;\beta_0)/\partial t$, $\lambda_\varepsilon^*(t;\beta_0) = \partial \Lambda_\varepsilon^*(t;\beta_0)/\partial t$ and $\Lambda_\varepsilon^*(t;\beta_0)$ is the limit of $\int_0^t \sum_{j=1}^n dN_j(u;\beta_0) / \sum_{j=1}^n \tilde{Y}_j(u;\beta_0,w_j^*)$. We apply similar techniques of Ying (1993) to prove that $B_{1n} + B_{2n}$ has the order $o(n^{1/2})$ in a $o(n^{1/3})$ neighborhood of β_0. See Appendix 1 for the proof.

By the Taylor's expansion,

$$\Lambda_\varepsilon^*(t;\beta) = \Lambda_\varepsilon^*(t;\beta_0) + \left\{ \lambda_\varepsilon^*(t;\beta_0) + o(1) \right\} Z^T (\beta - \beta_0),$$

where $\lambda_\varepsilon^*(t;\beta) = \partial \Lambda_\varepsilon^*(t;\beta)/\partial t$ or equivalently

$$d\Lambda_\varepsilon^*(t;\beta) = d\Lambda_\varepsilon^*(t;\beta_0) + \left\{\dot{\lambda}_\varepsilon^*(t;\beta_0)dt + o(1)\right\} Z^T(\beta-\beta_0)$$

where $\dot{\lambda}_\varepsilon^*(t;\beta_0)dt = d\lambda_\varepsilon^*(t;\beta_0)$. Thus B_{3n} can be written as

$$\int_0^\infty \sum_{i=1}^n \left\{ Z_i - \bar{Z}(t;\beta,w^*)\right\} Z_i^T Y_i^*(t;\beta)\dot{\lambda}_\varepsilon^*(t;\beta_0)dt(\beta-\beta_0) + o\left(n\cdot\|\beta-\beta_0\|\right)$$

$$= \int_0^\infty \sum_{i=1}^n \left\{ Z_i - \bar{Z}(t;\beta,w^*)\right\}^{\otimes 2} \frac{\dot{\lambda}_\varepsilon^*(t;\beta_0)}{\lambda_\varepsilon^*(t;\beta)} dN(t;\beta)(\beta-\beta_0) + o\left(n\cdot\|\beta-\beta_0\|\right),$$

where $N(t;\beta) = \sum_{j=1}^n N_i(t;\beta)$ and $M^{\otimes 2} = MM^T$. In summary we have proved

$$U(\beta\,|\,w^*) - U(\beta_0\,|\,w^*)$$

$$= \int_0^\infty \sum_{i=1}^n \left\{ Z - \bar{Z}(t;\beta,w^*)\right\}^{\otimes 2} \frac{\dot{\lambda}_\varepsilon^*(t;\beta_0)}{\lambda_\varepsilon^*(t;\beta)} dN(t;\beta)(\beta-\beta_0) + o(n\|\beta-\beta_0\| + n^{1/2})$$

which is equivalent to

$$U(\beta\,|\,w^*) - U(\beta_0\,|\,w^*) = A_n n(\beta-\beta_0) + o(n^{1/2} + n\|\beta-\beta_0\|), \tag{3.7a}$$

where $A_n = \frac{1}{n}\sum_{i=1}^n \left\{ Z_i - \bar{Z}(t;\beta,w^*)\right\}^{\otimes 2} \frac{\dot{\lambda}_\varepsilon^*(t;\beta_0)}{\lambda_\varepsilon^*(t;\beta)} dN(t;\beta)$ .Now let's incorporate the influence of the estimated weight. Our original goal is to show that

$$n^{-1/2}U(\beta\,|\,\hat{w}) = n^{-1/2}U(\beta\,|\,w^*) + o_p(1).$$

However we only obtained the result: $U(\beta\,|\,\hat{w}) - U(\beta\,|\,w^*) = o(n^{2/3})$ which implies that $\frac{1}{\sqrt{n}}\left(U(\beta\,|\,\hat{w}) - U(\beta\,|\,w^*)\right) = o(n^{1/6})$ which des not converge to $o_p(1)$ when n is large. Note that if we impose a more strict condition on $\hat{w}$ (say $\sup_i \|\hat{w}_i - w_i^*\| \le o(n^{-1/2})$ for all (θ,β)), we will get the desirable property. However since this is not a realistic assumption, we have to try other approaches.

We obtain some intermediate results. Applying similar techniques of expansion, we can write

$$U(\beta\,|\,\hat{w}) - U(\beta_0\,|\,\hat{w}) = \hat{A}_n n(\beta-\beta_0) + r_n, \tag{3.7b}$$

where the components of $\hat{A}_n$ are similar to A_n with w^* being replaced by $\hat{w}$. The difference of (3.7a) and (3.7b) directly follows that

$$[U(\beta \mid w^*) - U(\beta_0 \mid w^*)] - [U(\beta \mid \hat{w}) - U(\beta_0 \mid \hat{w})]$$

$$= \int_0^\infty \sum_{i=1}^n \left\{ \bar{Z}(t;\beta,w^*) - \bar{Z}(t;\beta,\hat{w}) \right\} dN_i(t;\beta) - \int_0^\infty \sum_{i=1}^n \left\{ \bar{Z}(t;\beta_0,w^*) - \bar{Z}(t;\beta_0,\hat{w}) \right\} dN_i(t;\beta_0) = d_n .$$

In Appendix 2, we show that $d_n = o(n^{1/3})$. Notice that, based on the right-hand sides of (3.7a) and (3.7b), one can also write

$$d_n = A_n n(\beta - \beta_0) + o(n^{1/2} + n\|\beta - \beta_0\|) - \hat{A}_n n(\beta - \beta_0) + r_n .$$

In Appendix 3, we show that $r_n = o(n^{1/2})$ and hence $A_n \approx \hat{A}_n$. We aim to establish the result:

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \Rightarrow Normal(0, (A)^{-1}\Sigma(A)^{-1}), \tag{3.8}$$

where A is the limit of A_n and Σ is the limit of

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^n \left\{ Z_i - \bar{Z}(t;\beta_0,w^*) \right\}^{\otimes 2} dN(t;\beta_0) .$$

However the above proofs are not enough to make this conclusion. Let's summarize the results that we have obtained:

$$U(\beta \mid \hat{w}) = U(\beta_0 \mid w^*) + \{U(\beta_0 \mid \hat{w}) - U(\beta_0 \mid w^*)\} + A_n n(\beta - \beta_0) + o(n^{1/2}) .$$

Note that $U(\beta_0 \mid w^*) \Rightarrow Normal(0,\Sigma)$. If $U(\beta_0 \mid \hat{w}) = U(\beta_0 \mid w^*) + o(n^{1/2})$, it follows that asymptotically, $0 = n^{-1/2} U(\hat{\beta} \mid \hat{w}) + A\sqrt{n}(\hat{\beta} - \beta_0)$ which implies the normality of $\hat{\beta}$. In developing the variance estimator of $\hat{\beta}$, we still rely on the result in (3.8).

In Appendix 4, we show that for each t, $\bar{Z}(t;\beta_0,\hat{w}) = \bar{Z}(t;\beta_0,w^*) + o_p(1)$, but the order after taking the sum is still not derived yet. The final goal is to prove

$$\frac{1}{\sqrt{n}} U(\beta_0 \mid \hat{w}) = \frac{1}{\sqrt{n}} U(\beta_0 \mid w^*) + o_p(1) .$$

Note that

$$U(\beta_0 \mid \hat{w}) - U(\beta_0 \mid w^*) = \int_0^\infty \sum_{i=1}^n \left\{ \bar{Z}(t;\beta_0,w^*) - \bar{Z}(t;\beta_0,\hat{w}) \right\} dN_i(t;\beta_0) .$$

In Appendix 4, we show that for each t, $\bar{Z}(t;\beta_0,\hat{w}) = \bar{Z}(t;\beta_0,w^*) + o_p(1)$, but the order after taking the sum is still not derived yet. The difficulty comes from the dynamic weight which is a complicated function of (θ, β). We have conducted simulations to check whether $r_n = \frac{1}{\sqrt{n}}\left(U(\beta_0 \mid \hat{w}) - U(\beta_0 \mid w^*)\right)$ gets close to zero as the sample size increases. In Table 2D, we can see that the sample size changes from $n = 100$ to $n = 2000$, the value of r_n (or $|r_n|$) decreases and is close to zero.

3.4 Numerical algorithm and variance estimation

The proposed estimators solve

$$\kappa(\theta \mid w) = \sum_{i=1}^{n} \tilde{\varsigma}_i \frac{\dot{\pi}_i(\theta)}{\pi_i(\theta)} + \sum_{i=1}^{n} (1-\tilde{\varsigma}_i) \cdot \frac{\{-\dot{\pi}_i(\theta)\}}{1-\pi_i(\theta)}; \tag{3.9a}$$

$$U(\beta \mid w) = \int_0^\infty \sum_{i=1}^{n} \{Z_i - \bar{Z}(t;\beta,w)\} dN_i(t;\beta), \tag{3.9b}$$

where $\tilde{\varsigma}_i = \delta_i + (1-\delta_i) w_i$ and

$$w_i = w_i(\theta, \beta, S_\varepsilon) = \frac{\pi_i(\theta) \times S_\varepsilon(\tilde{\varepsilon}_i(\beta))}{\pi_i(\theta) \times S_\varepsilon(\tilde{\varepsilon}_i(\beta)) + \{1-\pi_i(\theta)\}}$$

with S_ε being replaced by the following explicit formula:

$$\hat{S}_\varepsilon(t) = \exp\{-\sum_{u \le t} \frac{\sum_{i=1}^{n} I(\tilde{\varepsilon}_i(\beta) = u, \delta_i = 1)}{\sum_{i=1}^{n} \{\delta_i + (1-\delta_i) w_i\} I(\tilde{\varepsilon}_i(\beta) \ge u)}\}.$$

The estimation procedure requires many iterations by updating the weights using previous estimates. Specifically let $w^{(m)}$ be the m-th step estimate of w based on $(\hat{\theta}^{(m)}, \hat{\beta}^{(m)}, \hat{S}_\varepsilon^{(m)})$. Treating $w^{(m)}$ as fixed value, one can solve $\kappa(\theta \mid w^{(m)}) = 0$ and $U(\beta \mid w^{(m)}) = 0$ to obtain $\hat{\theta}^{(m+1)}$ $\hat{\beta}^{(m+1)}$ respectively and then update

$$\hat{S}_\varepsilon^{(m+1)}(t) = \exp\{-\sum_{u \le t} \frac{\sum_{i=1}^{n} I(\tilde{\varepsilon}_i(\beta) = u, \delta_i = 1)}{\sum_{i=1}^{n} \{\delta_i + (1-\delta_i) w_i^{(m)}\} I(\tilde{\varepsilon}_i(\beta) \ge u)}\}.$$

The procedure is repeated for $m = 0, 1, 2, \ldots$ until convergence.

3.4.1 Re-sampling based on bootstrap approach

The non-differentiability and the complicated and dynamic weight components make it difficult to derive an analytic formula for variance estimation. The bootstrap approach provides a simulation scheme without extra analytic work. Say R bootstrap samples are drawn from the original data $\left\{(X_i,\delta_i,Z_i^T),\ i=1,...,n\right\}$. For each bootstrap sample, we perform the estimation procedure which involves $\kappa(\theta\,|\,w^{(m)})=0$ and $U(\beta\,|\,w^{(m)})$ for say $m=1,...,M$. The sampling distributions of $\hat{\theta}$ and $\hat{\beta}$ can be approximated based on the K bootstrap estimates. The bootstrap method is time-consuming which involves solving the roots $R\times M$ times. Note that solving $U(\beta\,|\,w^{(m)})=0$ even once is not an easy task.

3.4.2 Re-sampling based on pivotal estimating functions

Parzen, Wei and Ying (1994) proposed a re-sampling method which has become a popular tool for variance estimation for many semi-parametric inference problems. This approach is useful when the estimating function is not smooth. In other situations, the derivative of the score function can be derived under some regularity conditions but still contains unknown density functions which cannot be estimated based on the simple plug-in approach.

Now we apply and modify the idea of Parzen et al. (1994). For our problem, the pivotal estimating function are the asymptotic distributions of

$$\begin{bmatrix}\kappa(\theta_0\,|\,w^*)\\ U(\beta_0\,|\,w^*)\end{bmatrix}=\begin{bmatrix}\mathrm{K}_1\\ U_1\end{bmatrix}.$$

Directly applying this approach, we first need to generate many replicates from the pivotal distribution denoted as (K_j,U_j) for $j=1,...,R$. Then solve

$$\begin{bmatrix}\kappa(\theta\,|\,w\{\theta,\beta,\hat{S}_\varepsilon\})\\ U(\beta\,|\,w\{\theta,\beta,\hat{S}_\varepsilon\})\end{bmatrix}=\begin{bmatrix}\mathrm{K}_j\\ U_j\end{bmatrix}. \tag{3.9c}$$

Let $(\tilde{\theta}_j,\tilde{\beta}_j)$ be the corresponding solution for $j=1,...,R$. Then the conditional distribution of

$\begin{bmatrix} \hat{\theta} \\ \hat{\beta} \end{bmatrix} - \begin{bmatrix} \tilde{\theta}_j \\ \tilde{\beta}_j \end{bmatrix}$, given the observed sample, is asymptotic equivalent to the unconditional distribution of $\begin{bmatrix} \hat{\theta} \\ \hat{\beta} \end{bmatrix} - \begin{bmatrix} \theta_0 \\ \beta_0 \end{bmatrix}$ for each j. It implies that the empirical distributions of $\{(\tilde{\theta}_j, \tilde{\beta}_j)\ j = 1, ..., R\}$, conditional on the observed sample, can be used to approximate the unconditional distribution of $(\hat{\theta}, \hat{\beta})$.

The above procedure, however, is very time-consuming since it still involves many iterations to obtain the solution of (3.9c). We propose to modify the procedure by solving

$$\begin{bmatrix} \kappa(\theta \mid \hat{w}^*) \\ U(\beta \mid \hat{w}^*) \end{bmatrix} = \begin{bmatrix} \mathrm{K}_j \\ U_j \end{bmatrix}, \tag{3.10}$$

where $\hat{w}^*$ denotes the final estimated weight. This modification can avoid the time-consuming iterations within each re-sampling run. In Appendix 3, we will see that this modification still produces valid results for variance estimation.

Now we derive the algorithm to simulate random samples from the pivotal distributions. Since our interest is in β, we only need to focus on $U(\beta \mid w^*)$ since it does not involve other parameters when w^* is a fixed value. We have

$$\begin{aligned} U(\beta \mid w^*) &= \int_0^\infty \sum_{i=1}^n \{Z_i - \bar{Z}(t;\beta,w^*)\} dN_i(t;\beta) \\ &= \sum_{i=1}^n \int_0^\infty \{Z_i - \bar{Z}(t;\beta,w^*)\} dM_i(t;\beta) \\ &= \sum_{i=1}^n \phi_i(\beta, w^*)\,, \end{aligned} \tag{3.10a}$$

where $\phi_i(\beta, w) = \int_0^\infty \{Z_i - \bar{Z}(t;\beta,w)\} dM_i(t;\beta)$. It can be shown that

$$n^{-1/2} U(\beta_0; w^*) \sim N_p(0, \Sigma)\,, \tag{3.10b}$$

where $\Sigma = E[\phi_i(\beta_0, w^*)\phi_i(\beta_0, w^*)^T]$. Plugging in the final estimated weight, asymptotically we have

$$n^{-1/2}U(\beta_0;\hat{w}^*) \sim N_p(0,\Sigma), \tag{3.10c}$$

where the covariance matrix can be estimated by

$$\hat{\Sigma} = \sum_{i=1}^{n} \phi_i(\hat{\beta},\hat{w})\phi_i(\hat{\beta},\hat{w})^T / n. \tag{3.10d}$$

In Section 3.3, we have shown that for β in a small neighborhood of β_0,

$$n^{-1/2}U(\beta \mid \hat{w}^*) = n^{-1/2}U(\beta_0;\hat{w}^*) + An^{1/2}(\beta-\beta_0) + o_p(1),$$

where A is the asymptotic slope matrix of $n^{-1/2}U(\beta_0;w^*)$. We simulate $G_i \sim N(0,1)$ independently for $i=1,\ldots,n$. Let $\hat{\beta}^*$ be the solution to

$$U(\beta \mid \hat{w}^*) = \sum_{i=1}^{n} \phi_i(\hat{\beta},\hat{w}^*)G_i\ . \tag{3.10e}$$

We can show that the conditional distribution of $n^{-1/2}\sum_{i=1}^{n}\phi_i(\hat{\beta},\hat{w}^*)G_i$, given the observed data, is also $N_p(0,\Sigma)$. Accordingly the conditional distribution of $n^{1/2}(\hat{\beta}^*-\hat{\beta})$ follows $N_p(0,A^{-1}\Sigma A^{-1})$, which is equivalent to the unconditional distribution of $n^{1/2}(\hat{\beta}-\beta_0)$. To implement the re-sampling algorithm, we repeat (3.10e) for R times and then obtain $\hat{\beta}_j^*$ for $j=1,\ldots,R$. The sample variance can be used to estimate $Var(\hat{\beta})$. The proposed re-sampling procedure is much faster than the bootstrap approach since no iteration is needed in solving (3.10e) and also there is no need to deal with the estimating function of θ.

3.5 Model Checking

We utilize the martingale framework to construct a model checking procedure for the latency distribution. Here the model assumption refers to the chosen form of $h(.)$. Define the residual process:

$$V(t;\beta) = n^{-1/2}\sum_{i=1}^{n} Z_i\,\hat{M}_i(t;\beta), \tag{3.11a}$$

where

$$\hat{M}_i(t;\beta) = N_i(t;\beta) - \int_0^t \tilde{Y}_i(u;\beta,\hat{w}_i^*)d\hat{\Lambda}_\varepsilon(u) \tag{3.11b}$$

and

$$\hat{\Lambda}_\varepsilon(t) = \int_0^t \sum_{j=1}^n dN_j(u;\beta) / \sum_{j=1}^n \tilde{Y}_i(u;\beta,\hat{w}_i^*) . \tag{3.11c}$$

The Kolmogorov-type test based on $\sup_t \left\| n^{-1/2} V(t;\hat{\beta}) \right\|$ can be used to measure the degree of departure from the imposed model.

First of all we need to show that, under the assumed model, $V(t;\hat{\beta})$ converges weakly to a mean-zero Gaussian process. The argument is similarly to Ghosh (2003). Here we summarize the sketch of proof. One can write

$$\begin{aligned} V(t;\hat{\beta}) &= n^{-1/2} \sum_{i=1}^n Z_i \hat{M}_i(t;\hat{\beta}) \\ &= n^{-1/2} \sum_{i=1}^n \int_0^t \left(Z_i - \bar{Z}(u;\hat{\beta},\hat{w}^*) \right) d\hat{M}_i(u;\hat{\beta},\hat{w}^*) . \end{aligned}$$

Notice that

$$\begin{aligned} \sum_{i=1}^n \hat{M}_i(u;\hat{\beta},\hat{w}^*) &= \sum_{i=1}^n \left(I(\tilde{\varepsilon}_i(\hat{\beta}) = u, \delta_i = 1) - \hat{w}_i^* I(\tilde{\varepsilon}_i(\hat{\beta}) \ge u) \frac{\sum_{j=1}^n I(\tilde{\varepsilon}_j(\hat{\beta}) = u, \delta_j = 1)}{\sum_{j=1}^n \hat{w}_j^* I(\tilde{\varepsilon}_j(\hat{\beta}) \ge u)} \right) \\ &= \sum_{i=1}^n I(\tilde{\varepsilon}_i(\hat{\beta}) = u, \delta_i = 1) - \sum_{i=1}^n \hat{w}_i^* I(\tilde{\varepsilon}_i(\hat{\beta}) \ge u) \frac{\sum_{j=1}^n I(\tilde{\varepsilon}_j(\hat{\beta}) = u, \delta_j = 1)}{\sum_{j=1}^n \hat{w}_j^* I(\tilde{\varepsilon}_j(\hat{\beta}) \ge u)} \\ &= 0 \text{, for } 0 < u < t . \end{aligned}$$

Using (3.8), $V(t;\hat{\beta})$ can be rewritten as

$$n^{-1/2} \sum_{i=1}^n \int_0^t (Z_i - \bar{Z}(t;\beta_0,\hat{w}^*)) d\hat{M}_{1i}(t;\beta_0,\hat{w}^*) + A_n(t) \cdot n^{1/2}(\hat{\beta} - \beta_0) + o_p(1) .$$

By the martingale central limit theorem and the consistency of $\hat{\beta}$,

$$V(t;\hat{\beta}) \xrightarrow{d} N_p(0,\Sigma) ,$$

where the covariance matrix (Σ) can be estimated by (3.10d). Furthermore its asymptotic distribution can be approximated by

$$\hat{V}(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t \{Z_i - \bar{Z}(u;\hat{\beta},\hat{w}^*)\}d\hat{M}_i(u;\hat{\beta})G_i + V(t;\tilde{\beta}) - V(t;\hat{\beta}). \quad (3.12)$$

For informal model diagnostics, we can plot the sample curve of $V(t;\hat{\beta})$ along with several simulated curves of $\hat{V}(t)$. If the sample curve is located within the range of simulated curves, the model assumption is reasonable. Formally, we can generate many replicates of $\hat{V}(t)$ and compute the value of $\sup_t \left\| n^{-1/2}\hat{V}(t) \right\|$ for the model candidates under consideration. The p-value refers to the empirical frequency that the observed value of $\sup_t \left\| n^{-1/2}V(t;\hat{\beta}) \right\|$ exceeds the simulated values of $\sup_t \left\| n^{-1/2}\hat{V}(t) \right\|$.

3.6 Simulation analysis

3.6.1 Data generation

We first generate Z_1 from Bernoulli (0.5) and compute

$$Z^T\theta_0 = (1, Z_1)^T(\theta_0^*, \theta_0^{(1)}) = \theta_0^* + \theta_0^{(1)}Z_1,$$

where the values of θ_0^* are $\theta_0^{(1)}$ are specified. Then generate ς ~ Bernoulli(p_Z) with

$$p_Z = \pi(\theta \mid Z) = \frac{\exp(Z^T\theta_0)}{1+\exp(Z^T\theta_0)}.$$

If $\varsigma = 1$, we generate the latency variable T which follows

$$\log T = \beta_0^{(1)}Z_1 + \beta_0^{(2)}Z_2 + \varepsilon$$

where ε follows the log-exponential distribution. If $\varsigma = 0$, we set T to be a very large number exceeding the support of C which follows a uniform distribution. Observed variables include replications of (X, δ, Z), where $X = T \wedge C$ and $\delta = I(T \le C)$. We consider two settings with A: $Z_1 \sim Ber(0.5)$ and $Z_2 \sim Ber(0.5)$; and B: $Z_1 \sim Ber(0.5)$ and $Z_2 \sim Unif(0,1)$.

3.6.2 Simulation results

Tables 1A and 1B show the results for estimating $\theta_0^{(0)}$, $\theta_0^{(1)}$,

$$p_0 = \pi(\theta \mid Z=0) = \frac{\exp(\theta_0^{(0)})}{1+\exp(\theta_0^{(0)})},$$

and

$$p_1 = \pi(\theta \mid Z=1) = \frac{\exp(\theta_0^{(0)}+\theta_0^{(1)})}{1+\exp(\theta_0^{(0)}+\theta_0^{(1)})}.$$

We calculate the average bias and standard deviation based on 1000 replications. In the two tables, the estimators have reasonable performances which improve as the sample size increases. Transforming $(\theta_0^{(0)},\theta_0^{(1)})$ into the probability scale based on (p_0, p_1), the performances of $(\hat{p}_0, \hat{p}_1)$ look satisfactory. Comparing the two tables which differ in the values of p_1, we see that the corresponding estimator becomes more variable when p_1 is closer to 0.5.

Our main proposal is developed for estimating β_0 in the latency model. Table 2A and Table 2B correspond to the incidence models in Table 1A and Table 1B respectively. The proposed estimators of $\beta_0^{(1)}$ and $\beta_0^{(2)}$ have reasonable performances but sometimes produce larger bias when the sample size is small or the censoring rate is high. Our another important proposal is the re-sampling scheme for variance estimation. To examine the performance, we first check whether the sample average of $\hat{\beta}_j^*$ which solves (3.10e) is close to the true parameter value. Then we examine whether the proposed estimator $\hat{\sigma}(\hat{\beta}_j)$, which is sample standard deviation of $\hat{\beta}_j^*$ $(j=1,...,R)$, is close to the simulated estimate denoted as $se(\hat{\beta}_j)$. The results are satisfactory. As a consequence, the coverage probability is close to the 95% nominal level in most cases. Notice that the results in Table 2B appear to be better than those in Table 2A since the former corresponds to higher incidence rate which provides more data to estimate the latency distribution.

Finally we examine the proposed model checking procedure. We first simulate data from an AFT model and then analyze it by an AFT model. Figures 4.1A and 4.1B show the two components of $V(t;\hat{\beta})$ based on Z_1 and Z_2 respectively. The observed curves are mostly located within 20 simulated curves which show that the fitted model is acceptable. Then we generate an AFT model and fit a location shift model. Figures 4.2A and 4.2B, the observed curves are located outside the simulated curves which show that the fitted model is not satisfactory.

Chapter 4

Literature Review for

Transformation Models with Cure

4.1 Background

In this chapter we consider the second class of models with the incidence model given by

$$\Pr(\varsigma = 1 | Z) = \pi(\theta_0 | Z) = \frac{\exp(Z^T\theta_0)}{1+\exp(Z^T\theta_0)}.$$

And for $\varsigma = 1$, T_i follows a transformation model of the form

$$h(T) = Z^T\beta_0 + \varepsilon,$$

where $h(\cdot)$ is a unknown monotone function but the distribution of ε is completely specified. Note that we denote the distribution, survival and cumulative hazard functions of ε as F_ε, S_ε and Λ_ε which are fully specified. The most well-known example is the proportional hazards (PH) model in which ε follows the extreme value distribution with $F_\varepsilon(s) = 1 - \exp\{-\exp(s)\}$. When ε follows the standard logistic distribution with $F_\varepsilon(s) = \exp(s)/\{1+\exp(s)\}$, the model becomes the proportional odds (PO) model.

For the discussions in this chapter, observed data are denoted as $\{(X_i, \delta_i, Z_i),\ i = 1,...,n\}$, where $X_i = T_i \wedge C_i$ and $\delta_i = I(T_i \le C_i)$. The parameters of interest are (θ, β) while $h(t)$ is an infinite-dimensional nuisance function. In the early stage of methodology development, statisticians including Kuk and Chen (1992), Sy and Taylor (2000) and Peng and Dear (2000) focused on the special case that the latency distribution follows the PH model. Then a new trend starting from Lu and Ying (2004) considers statistical inference for the whole of class of transformation models. In this chapter, we review existing literature for transformation cure models. Roughly speaking, existing inference approaches can be classified into two types. One is based on the likelihood principle and the other is based on moment properties.

4.2 Different model expressions

We first review different formulations of a transformation model since the form of model expression affects subsequent inference development. The most well-known representation is given by

$$h(T) = Z^T \beta_0 + \varepsilon\,, \tag{4.1a}$$

which states that the failure time T for a susceptible subject can be written as a parametric linear model after an unknown monotone transformation. Alternatively one can also write

$$\varphi\{\tilde{S}_Z(t)\} = h(t) + Z^T \beta_0\,, \tag{4.1b}$$

where $\tilde{S}_Z(t) = \Pr(T > t \mid \varsigma = 1, Z)$ is the survival function of $T \mid \varsigma = 1, Z$ and $F_\varepsilon(t) = 1 - \varphi^{-1}(t)$ which is a known function. The representation of (4.1b) says that a known transformation of the survival function leads to a linear structure in the parameters which contains an un-specified intercept function. One can also write (4.1b) in terms of the cumulative hazard function, defined as $\tilde{\Lambda}_Z(t) = -\log\{\tilde{S}_Z(t)\}$, such that

$$\tilde{\Lambda}_Z(t) = -\log[\varphi^{-1}\{h(t) + Z^T \beta_0\}] = H\{h(t) + Z^T \beta_0\}\,, \tag{4.1c}$$

where $H(t) = -\log\{\varphi^{-1}(t)\}$ is also completely specified. Notice that the above three equivalent expressions only allow for time-independent covariates. Later in Chapter 5, we will discuss the extension of including time-dependent covariates.

4.3 Likelihood approach under the PH model

The likelihood function under the transformation model can be written as

$$\prod_{i=1}^{n} [\pi_i(\theta) \cdot \tilde{f}_{Z_i}(x_i)]^{\delta_i} [\tilde{S}_{Z_i}(x_i) \cdot \pi_i(\theta) + 1 - \pi_i(\theta)]^{1-\delta_i}\,, \tag{4.2a}$$

where $\tilde{S}_Z(t) = \varphi^{-1}\{h(t) + Z^T \beta\}$ and $\tilde{f}_Z(t) = -\partial \tilde{S}_Z(t) / \partial t$. Expressing the function in terms of hazard and survival functions, we obtain

$$\prod_{i=1}^{n} \left\{ \left[\pi_i(\theta) \tilde{\lambda}_{Z_i}(x_i) \tilde{S}_{Z_i}(x_i)\right]^{\delta_i} \left[\pi_i(\theta) \tilde{S}_{Z_i}(x_i) + 1 - \pi_i(\theta)\right]^{1-\delta_i} \right\}. \tag{4.2b}$$

When the latency follows the PH model, (4.2b) can be written as

$$\prod_{i=1}^{n}\left\{ \left[\pi_i(\theta)\cdot\tilde{\lambda}_0(x_i)e^{\beta^T Z_i}\cdot\tilde{S}_0(x_i)^{\exp(\beta^T Z_i)}\right]^{\delta_i}\left[\pi_i(\theta)\cdot\tilde{S}_0(x_i)^{\exp(\beta^T Z_i)}+1-\pi_i(\theta)\right]^{1-\delta_i} \right\} \tag{4.3}$$

where $\tilde{S}_0(.)$ and $\tilde{\lambda}_0(.)$ are the baseline survival and hazard functions for the susceptible group.

Equation (4.3) can be simplified by first considering complete data with ς_i being observed. The completed likelihood function for (θ,β) can be written as follows: $\prod_{i=1}^{n}\left\{ \left[\pi_i(\theta)\cdot\tilde{\lambda}_0(x_i)e^{\beta^T Z_i}\cdot\tilde{S}_0(x_i)^{\exp(\beta^T Z_i)}\right]^{\delta_i}\left[\pi_i(\theta)\cdot\tilde{S}_0(x_i)^{\exp(\beta^T Z_i)}\right]^{(1-\delta_i)\varsigma_i}\left[1-\pi_i(\theta)\right]^{(1-\delta_i)(1-\varsigma_i)}\right\}$ which can be written as the product of the following two terms:

$$L_1(\theta)=\prod_{i=1}^{n}\left\{ \left[\pi_i(\theta)\right]^{\varsigma_i}\left[1-\pi_i(\theta)\right]^{(1-\varsigma_i)} \right\} \tag{4.5a}$$

and

$$L_2(\beta,\tilde{\Lambda}_0)=\prod_{i=1}^{n}\left\{ \left[\tilde{\lambda}_0(x_i)e^{\beta^T Z_i}\cdot\tilde{S}_0(x_i)^{\exp(\beta^T Z_i)}\right]^{\delta_i}\left[\tilde{S}_0(x_i)^{\exp(\beta^T Z_i)}\right]^{(1-\delta_i)\varsigma_i} \right\}. \tag{4.5b}$$

Both terms involve possibly missing ς_i and the second equation involves the nuisance function $\tilde{S}_0(.)$ or $\tilde{\Lambda}_0(.)$.

Kuk and Chen (1992) considered the marginal likelihood function $\sum_{\varsigma\in\Omega} L_1(\theta)L_2(\beta,\tilde{\Lambda}_0)$, where Ω is the collection of all n_c-tuples of 0's and 1's, $n_c=\sum_{i=1}^{n} I(\delta_i=0)$ and $\varsigma\in\Omega$ is a realization of ς_i for those observations with $\delta_i=0$ $(i=1,\ldots,n)$. However since this marginal likelihood, which involves the complicated summation, is not easy to handle numerically, a Monte Carlo approach was suggested to implement the procedure.

Later researchers proposed to analyze the two terms in (4.5a) and (4.5b) separately. The log-likelihood can be rewritten as the sum of

$$l_1(\theta)=\sum_{i=1}^{n}\left\{\varsigma_i\log\left(\pi_i(\theta)\right)+(1-\varsigma_i)\log\left(1-\pi_i(\theta)\right)\right\} \tag{4.6a}$$

and

$$l_2(\beta,\tilde{S}_0)=\sum_{i=1}^{n}\left\{\delta_i\log\left(\beta^T Z\right)+\varsigma_i\log\left(\tilde{S}_0(x_i)^{\exp(\beta^T Z_i)}\right)\right\}. \tag{4.6b}$$

There are two ways of handling the nuisance function in (4.5b) – ignore it as in the classical analysis without cure or estimate it using explicit formula. Now we discuss both approaches.

Motivated by the marginal distribution of ranks discussed in Kalbfleisch and Prentice (1973), Peng and Dear (2000) suggested to ignore the nuisance function $\tilde{\Lambda}_0(.)$ in (4.5b). To simplify the discussion, assume there are no ties. Let $x_{(1)} < \cdots < x_{(k)}$ be ordered uncensored failure times with corresponding covariates $Z_{(1)},\cdots,Z_{(k)}$. The partial likelihood for β, assuming that ς_i $(i=1,...,n)$ are available, is given by

$$L_2^*(\beta) = \left(\prod_{i=1}^{k} \frac{\exp\left(\beta^T Z_{(i)}\right)}{\sum_{j \in R(x_{(i)})} \varsigma_j \exp\left(\beta^T Z_j\right)} \right), \tag{4.7}$$

where $R(t) = \{i : X_i \geq t\}$ is the risk set at time t. Peng and Dear (2000) further proposed to impute ς_i by an estimator of its conditional mean:

$$E\left(\varsigma_i \mid x_i, \delta_i, z_i\right) = \delta_i + (1-\delta_i)\Pr(\varsigma_i \mid T_i > x_i, \delta_i = 0, z_i),$$

where

$$w_i = \Pr(\varsigma_i = 1 \mid T_i > x_i, \delta_i = 0, z_i) = \frac{\pi_i(\theta)\tilde{S}_{Z_i}(x_i)}{\pi_i(\theta)\tilde{S}_{Z_i}(x_i) + 1 - \pi_i(\theta)}.$$

The baseline function $\tilde{S}_0(t)$ can be estimated by

$$\breve{S}_0(t) = \exp\left(-\sum_{j: x_{(j)} < t} \frac{1}{\sum_{l \in R_j} \{\delta_l + (1-\delta_l)w_l\} \exp\left(\beta^T z_l\right)} \right).$$

Since $\tilde{S}_{Z_i}(t) = \tilde{S}_0(t)^{\exp(\beta^T z_i)}$, $\tilde{S}_{Z_i}(x_i)$ can be estimated by $\breve{S}_0(x_i)^{\exp(\beta^T z_i)}$. Finally an estimator of β can be obtained by maximizing

$$\tilde{L}_2^*(\beta) = \left(\prod_{i=1}^{k} \frac{\exp\left(\beta^T Z_{(i)}\right)}{\sum_{j \in R(x_{(i)})} \{\delta_i + (1-\delta_i) w_i\} \exp\left(\beta^T Z_j\right)} \right).$$

In each iteration of the maximization, w_i is treated as a fixed value by plugging in previous estimates of $(\theta, \beta, \tilde{S}_0)$. The final estimator is obtained when the convergence criteria is satisfied.

Sy and Taylor (2000) proposed two methods for handling (4.5b). Their first proposal suggested to substitute $\tilde{\Lambda}_0(t)$ in $L(\beta, \tilde{\Lambda}_0)$ by the Breslow estimator given by $\sum_{j: x_{(j)} < t} \{1 / \sum_{l \in R_j} \varsigma_l \exp\left(\beta^T z_l\right)\}$. Their first approach becomes very similar to that of Peng and Dear (2000). Their second proposal was motivated by Kalbfleisch and Prentice (1980) such that the nuisance function $\tilde{S}_0(t)$ is decomposed as the product-limit form of hazard rates at observed failure times. Then $L_2(\beta, \tilde{\Lambda}_0)$ can be expressed as a function of β and the (baseline) hazard rates. The idea of profile likelihood estimation is adopted such that the hazard rates are estimated first, assuming that β is known, and then β is maximized based on the profile likelihood.

4.4 Moment approach for the class of transformation models

Lu and Ying (2004) extended the analysis to the whole class of transformation models. Unknown parameters become $\{\theta, \beta, h(\cdot)\}$ with true values denoted by $\{\theta_0, \beta_0, h_0(\cdot)\}$ respectively. Properties of counting processes and martingales are applied to construct estimating functions. Specifically define $N_i(t) = I(X_i \le t, \delta_i = 1)$ and F_t as the corresponding filtration up to time t. It follows that

$$E[dN_i(t) \mid F_{t-}] = I(X_i \ge t) d\Lambda_Z(t) = I(X_i \ge t) \frac{\partial - \log\{S_Z(t)\}}{\partial t},$$

where

$$S_Z(t) = \frac{1}{1 + \exp(Z^T \theta_0)} + \frac{\exp(Z^T \theta_0)}{1 + \exp(Z^T \theta_0)} \tilde{S}_Z(t).$$

Additional re-parameterization is needed to express $E[dN_i(t) \mid F_{t-}]$ as a simpler function of the unknown parameters. It follows that

$$S_Z(t) = \frac{1+\exp(Z^T\theta_0 - \tilde{\Lambda}_Z(t))}{1+\exp(Z^T\theta_0)} = \frac{1+\exp(Z^T\theta_0 - H\{h_0(t) + Z^T\beta_0\})}{1+\exp(Z^T\theta_0)},$$

where $H(x) = -\log\{\varphi^{-1}(x)\}$ is a known function. To simplify the expression, define $\psi(x) = \exp(x)/\{1+\exp(x)\}$ as the logistic function and $\bar{\psi}(x) = 1-\psi(x)$. Then

$$S_Z(t) = \frac{\bar{\psi}(Z^T\theta_0)}{\bar{\psi}(Z^T\theta_0 - \tilde{\Lambda}_Z(t))} = \frac{\bar{\psi}(Z^T\theta_0)}{\bar{\psi}[Z^T\theta_0 - H\{h_0(t) + Z^T\beta_0\}]}$$

and

$$\Lambda_Z(t) = -\log\{\bar{\psi}(Z^T\theta_0)\} + \log\{\bar{\psi}[Z^T\theta_0 - H\{h_0(t) + Z^T\beta_0\}]\}.$$

Note that the first term in $\Lambda_Z(t)$ does not depend on t. Accordingly one can define

$$M_i(t;\theta,\beta,h) = N_i(t) - I(X_i \ge t)\ d\log\{\bar{\psi}[Z_i^T\theta - H\{h(t) + Z_i^T\beta\}]\}$$

$$= N_i(t) - I(X_i \ge t) d\log\{\psi[-Z_i^T\theta + H\{h(t) + Z_i^T\beta\}]\},$$

where the second term uses the fact that $\psi(-x) = \bar{\psi}(x)$. An important property is that $M_i(t;\theta_0,\beta_0,h_0)$ is a mean-zero martingale. This expression is nice since it is a tractable function of linear terms in the unknown parameters. Two sets of estimating functions can be constructed:

$$\sum_{i=1}^{n} dN_i(t) - Y_i(t) d\log\{\bar{\psi}[Z_i^T\theta - H\{h(t) + Z_i^T\beta\}]\} = 0;$$

$$\sum_{i=1}^{n} \int_0^\infty Z_i\{dN_i(t) - Y_i(t) d\log\{\bar{\psi}[Z_i^T\theta - H\{h(t) + Z_i^T\beta\}]\} = 0,$$

where the first one can be viewed as for estimating $h(t)$ and the second one is for β. Notice that for the transformation cure model, estimation for the latency distribution does not use the idea of E-step since the compensator $I(X_i \ge t) d\Lambda_Z(t)$ does not need the information of ς_i. Nevertheless since θ is also unknown, additional set of estimating equation is needed. Lu and Ying (2004) suggested to modify the estimating function:

$$\sum_{i=1}^{n} \int_0^\infty Z_i\{\tilde{\varsigma}_i - \frac{\exp(Z_i^T\theta)}{1+\exp(Z_i^T\theta)}\} = 0,$$

where $\tilde{\varsigma}_i = \delta_i + (1-\delta_i)w_i$ and $w_i = \Pr(\varsigma_i = 1 \mid T_i > x_i, \delta_i = 0, z_i)$ is expressed as a function of

$\{\theta, \beta, h(\cdot)\}$.

The approach proposed by Lu and Ying is attractive since it can be applied to the whole class of transformation models. Their idea is based on martingale properties which can be viewed as a moment approach. In the next chapter, we also consider general transformation models but adopt the likelihood principle for inference.

Chapter 5

Proposed Approach for Transformation Cure Model under Independent Censoring

Recently Zeng and Lin (2006) extended the transformation model without cure to allow for time-dependent covariates. The model assumption is imposed on the cumulative intensity function which can be expressed as

$$A_Z(t) = G\left\{\int_0^t I(T > s)\exp\{\beta_0^T Z(s)\}dR(s)\right\}, \tag{5.1}$$

where $Z(t)$ denotes the vector of time-dependent covariates, $G(.)$ is a known transformation and $R(.)$ is an unknown increasing function. We derive the connection of model (5.1) with the original model expressions discussed in Chapter 4 under time-independent covariates with $Z(t) = Z$. Notice that the cumulative hazards function can be written as:

$$\begin{aligned}\Lambda_Z(t) &= G\left\{\int_0^t \exp\{\beta_0^T Z(s)\}dR(s)\right\} \\ &= G\left\{\exp\left(\beta_0^T Z + \log[R(t)]\right)\right\} \\ &= H\{\beta_0^T Z + h(t)\},\end{aligned} \tag{5.2}$$

where $H(t) = G\{\exp(t)\}$ and $h(t) = \log R(t)$ which correspond to (4.1c). The advantage of (5.1) is the inclusion of time-dependent covariates in the model. Besides the model extension, Zeng and Lin (2006) also proposed likelihood-based inference methods which may yield more efficient results. We consider two extensions. In this chapter we apply the approach to the mixture model when the latency distribution follows the transformation models. In Chapter 6, we discuss a complicated situation of the cure model when dependent censoring exists.

5.1 Model assumptions

Let ς be the susceptible indicator which, given the covariate Z, follows the logistic model:

$$\pi(\theta \mid Z) = \frac{\exp(Z^T\theta)}{1+\exp(Z^T\theta)}. \tag{5.3a}$$

For $\varsigma = 1$, the cumulative intensity function can be written as

$$\tilde{A}_Z(t \mid Z) = G\left(\int_0^t I(T > s)e^{\beta^T Z(s)}dR(s)\right), \tag{5.3b}$$

where $G(.)$ is a known function and $R(.)$ is an unknown increasing function. Define $Y(s) = I(T > s)$ and

$$\xi(t;\beta,R) = \int_0^t Y(s)e^{\beta^T Z(s)}dR(s).$$

Accordingly for $\varsigma = 1$, the intensity function can be written as

$$\begin{aligned}
\tilde{a}(t) &= \frac{\partial}{\partial t}G\left(\xi(t;\beta,R)\right) \\
&= \left(\frac{\partial}{\partial\xi(t;\beta,R)}G\left(\xi(t;\beta,R)\right)\right)\cdot\left(\frac{\partial\xi(t;\beta,R)}{\partial t}\right) \\
&= g\left\{\xi(t;\beta,R)\ \right\}I(T > t)e^{\beta^T Z(t)}dR(t),
\end{aligned} \tag{5.3c}$$

where $g(t) = \frac{\partial}{\partial t}G(t)$. For time-independent covariates, the cumulative hazard function and hazard function are defined as $\tilde{\Lambda}_Z(t)$ and $\tilde{\lambda}_Z(t)$ which have similar expressions as $\tilde{A}_Z(t)$ and $\tilde{a}_Z(t)$ respectively but without the stochastic component $I(T > t)$. Accordingly we have

$$S_Z(t) = \Pr(T > t \mid Z) = \frac{1}{1+\exp(Z^T\theta)} + \frac{\exp(Z^T\theta)}{1+\exp(Z^T\theta)}\exp\{-\tilde{\Lambda}_Z(t)\}$$

and

$$\Lambda_Z(t) = -\log\{S_Z(t)\} = -\log\left[\frac{1}{1+\exp(Z^T\theta)} + \frac{\exp(Z^T\theta)}{1+\exp(Z^T\theta)}\exp\{-\tilde{\Lambda}_Z(t)\}\right].$$

5.2 Likelihood analysis for complete data

Suppose we have complete data $\left\{\left(X_i, \delta_i, \varsigma_i\right), i = 1,\dots,n\right\}$. Let $N_i(t) = I(T_i \le t, \delta_i = 1)$, $Y_i(s) = I(T_i > s)$ and $\xi_i(t;\beta,R) = \int_0^t Y_i(s)e^{\beta^T Z_i(s)}dR(s)$. The log-likelihood function for $\left(dR, \beta\right)$ is proportion to

$$\sum_{i=1}^{n}\left\{ \int_0^{\tau}\delta_i\left[\log g\left(\xi_i(t;\beta,R)\right)+\beta^T Z_i(t)+\log dR(t)\right]-\varsigma_i G\left(\xi_i(t;\beta,R)\right) \right\}$$

$$=\sum_{i=1}^{n}\left\{ \int_0^{\tau}\left[\log g\left(\xi_i(t;\beta,R)\right)+\beta^T Z_i(t)+\log dR(t)\right]dN_i(t)-\varsigma_i\int_0^{\tau} g\left(\xi_i(t-;\beta,R)\right)Y_i(t)e^{\beta^T Z_i(t)}dR(t)\right\}.$$

Directly maximizing the above likelihood is difficult. Nevertheless Chen (2009) proposed closed-form score equations which yield nice analytical results. Here we modify his results for the extended cure model. We need to assume that $R(t)$ is a step function which takes jumps only at observed failure points. Let $t_* \in (0,\tau]$ be an observed time and $R_* = R(t_*)$ is a step function which jumps at time t_*. The score function for dR involves the following two derivative equations:

$$\frac{\partial}{\partial dR_*}\int_0^{\tau}\log g\left(\xi_i(t-;\beta,R)\right)dN_i(t)$$

$$=\int_{t_*+}^{\tau}\frac{g'\left(\xi_i(t-;\beta,R)\right)}{g\left(\xi_i(t-;\beta,R)\right)}\left[\frac{\partial}{\partial dR_*}\int_0^{\tau}Y_i(t)e^{\beta^T Z_i(t)}dR_i(t)\right]dN_i(t)$$

$$=\int_{t_*+}^{\tau}\frac{g'\left(\xi_i(t-;\beta,R)\right)}{g\left(\xi_i(t-;\beta,R)\right)}Y_i(t_*)e^{\beta^T Z_i(t)}dN_i(t)\,;$$

and

$$\frac{\partial}{\partial dR_*}G\left(\xi_i(\tau;\beta,R)\right)$$

$$=\frac{\partial}{\partial dR_*}\int_0^{\tau}Y_i(t)g\left(\xi_i(t-;\beta,R)\right)\cdot e^{\beta^T Z_i(t)}dR_i(t)$$

$$=\left[\int_{t_*+}^{\tau}g'\left(\xi_i(t-;\beta,R)\right)Y_i(t)e^{\beta^T Z_i(t)}Y_i(t_*)e^{\beta^T Z_i(t)}dR_i(t)+Y_i(t_*)e^{\beta^T Z_i(t)}g\left(\xi_i(t_*-;\beta,R)\right)\right].$$ It

follows that

$$\frac{\partial l(\beta,dR)}{\partial dR_*}$$

$$=\frac{\partial}{\partial dR_*}\sum_{i=1}^{n}\left\{ \int_0^{\tau}\left[\log dR(t)+\log g\left(\xi_i(t-;\beta,R)\right)+\beta^T Z_i(t)\right]dN_i(t)-\varsigma_i G\left(\xi_i(\tau;\beta,R)\right)\right\}$$

$$=\sum_{i=1}^{n}\left\{\frac{dN_i(t_*)}{dR(t_*)}+\frac{\partial}{\partial dR_*}\int_0^{\tau}\log g\left(\xi_i(t-;\beta,R)\right)dN_i(t)-\frac{\partial}{\partial dR_*}\varsigma_i G\left(\xi_i(\tau;\beta,R)\right) \right\}$$

$$= \sum_{i=1}^{n} \left\{ \frac{dN_i(t_*)}{dR(t_*)} - w_i(t_*;\beta,R) Y_i(t_*) \varsigma_i e^{\beta^T Z_i(t_*)} g\left(\xi_i(t_*-;\beta,R)\right) \right\} \tag{5.4}$$

where

$$w_i(t_*;\beta,R) = 1 - \frac{\kappa_i(t_*;\beta,R)}{g\left(\xi_i(t_*-;\beta,R)\right)}, \tag{5.5a}$$

$$\kappa_i(t_*;\beta,R) = \int_{t_*+}^{\tau} \frac{g'\left(\xi_i(u-;\beta,R)\right)}{g\left(\xi_i(u-;\beta,R)\right)} dM_i(u), \tag{5.5b}$$

$$dM_i(t) = dN_i(t) - g\left(\xi_i(t;\beta,R)\right) Y_i(t) e^{\beta^T Z_i(t)} dR(t). \tag{5.5c}$$

The score equation for β is given by

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \left\{ \frac{\partial}{\partial \beta} \int_0^{\tau} \left[\log g\left(\xi_i(t;\beta,R)\right) + \beta^T Z_i(t) + \log dR(t) \right] dN_i(t) \right.$$
$$\left. - \frac{\partial}{\partial \beta} \int_0^{\tau} g\left(\xi_i(t-;\beta,R)\right) \varsigma_i Y_i(t) e^{\beta^T Z_i(t)} dR(t) \right\}$$

$$= \sum_{i=1}^{n} \left\{ \int_0^{\tau} Z_i(t) dN_i(t) + \int_0^{\tau} \frac{g'\left(\xi_i(t;\beta,R)\right)}{g\left(\xi_i(t;\beta,R)\right)} \left[\int_0^{t} Z_i(u) Y_i(u) e^{\beta^T Z_i(u)} dR(u) \right] dN_i(t) \right.$$

$$- \int_0^{\tau} g'\left(\xi_i(t-;\beta,R)\right) \left[\int_0^{t} Z_i(u) Y_i(u) e^{\beta^T Z_i(u)} dR(u) \right] \varsigma_i Y_i(t) e^{\beta^T Z_i(t)} dR(t)$$

$$\left. - \int_0^{\tau} g\left(\xi_i(t-;\beta,R)\right) \varsigma_i Y_i(t) e^{\beta^T Z_i(t)} Z_i(t) dR(t) \right\}.$$

Re-arranging the terms in the above equation and applying the formula in (5.5a)-(5.5c), $\frac{\partial l}{\partial \beta}$ can be written as

$$\sum_{i=1}^{n} \left\{ \int_0^{\tau} Z_i(t) dN_i(t) + \int_0^{\tau} \frac{g'\left(\xi_i(t;\beta,R)\right)}{g\left(\xi_i(t;\beta,R)\right)} \left[\int_0^{t} Z_i(u) \varsigma_i Y_i(u) e^{\beta^T Z_i(u)} dR(u) \right] dM_i(t) \right.$$

$$\left. - \int_0^{\tau} g\left(\xi_i(t-;\beta,R)\right) \varsigma_i Y_i(t) e^{\beta^T Z_i(t)} Z_i(t) dR(t) \right\}$$

$$= \sum_{i=1}^{n} \left\{ \int_0^{\tau} Z_i(t) dN_i(t) + \int_0^{\tau} \frac{\left[\int_t^{\tau} \frac{g'\left(\xi_i(u;\beta,R)\right)}{g\left(\xi_i(u;\beta,R)\right)} dM_i(u) \right]}{g\left(\xi_i(t;\beta,R)\right)} g\left(\xi_i(t;\beta,R)\right) Z_i \varsigma_i Y_i(t) e^{\beta^T Z_i(t)} dR(t) \right.$$

$$\left. - \int_0^{\tau} g\left(\xi_i(t-;\beta,R)\right) \varsigma_i Y_i(t) e^{\beta^T Z_i(t)} Z_i dR(t) \right\}$$

$$= \sum_{i=1}^{n} \left\{ \int_0^\tau Z_i(t) dN_i(t) - \int_0^\tau w_i(t;\beta,R) g\left(\xi_i(t;\beta,R)\right) Z_i(t) \varsigma_i Y_i(t) e^{\beta^T Z_i(t)} dR(t) \right\}. \tag{5.6}$$

Finally we can obtain the score equations for $dR(t_*)$ and β as follows:

$$dR(t) = \frac{\sum_{i=1}^{n} dN_i(t)}{\sum_{i=1}^{n} w_i(t;\beta,R) \varsigma_i Y_i(t) e^{\beta^T Z_i(t)} g\left(\xi_i(t-;\beta,R)\right)}; \tag{5.7}$$

$$\sum_{i=1}^{n} \int_0^\tau \left\{ Z_i(t) - \bar{Z}(t) \right\} dN_i(u) = 0, \tag{5.8}$$

where

$$\bar{Z}(t) = \frac{\sum_{i=1}^{n} w_i(t;\beta,R) g\left(\xi_i(t;\beta,R)\right) \varsigma_i Y_i(t) e^{\beta^T Z_i(t)} Z_i(t)}{\sum_{i=1}^{n} w_i(t;\beta,R) g\left(\xi_i(t;\beta,R)\right) \varsigma_i Y_i(t) e^{\beta^T Z_i(t)}}.$$

5.3 Imputation for handling missing data

The next job is to deal with the unknown value of ς_i. We can replace ς_i by $\hat{\varsigma}_i = \delta_i + (1-\delta_i) w_i^{EM}$, where

$$w_i^{EM} = w_i^{EM}(\theta,\beta,S,R) = \frac{\pi_i(\theta) \times S(t_i)}{\pi_i(\theta) \times S(t_i) + 1 - \pi_i(\theta)}.$$

Notice that $\varsigma_i Y_i(u) = I(T_i > u, \delta_i = 1) + \varsigma_i I(T_i > u, \delta_i = 0)$ and hence

$$\hat{\varsigma}_i Y_i(u) = I(T_i > u, \delta_i = 1) + w_i^{EM} I(T_i > u, \delta_i = 0).$$

The final estimating equations can be written as

$$\sum_{i=1}^{n} \hat{\varsigma}_i \frac{\dot{\pi}_i(\theta)}{\pi_i(\theta)} + \sum_{i=1}^{n} (1-\hat{\varsigma}_i) \frac{\{-\dot{\pi}_i(\theta)\}}{1-\pi_i(\theta)} = 0; \tag{5.9a}$$

$$dR(t) = \frac{\sum_{i=1}^{n} dN_i(t)}{\sum_{i=1}^{n} w_i(t;\beta,R) \hat{\varsigma}_i Y_i(t) e^{\beta^T Z_i(t)} g\left(\xi_i(t-;\beta,R)\right)}; \tag{5.9b}$$

$$\sum_{i=1}^{n} \int_0^\tau \left\{ Z_i - \frac{\sum_{i=1}^{n} w_i(t;\beta,R) g\left(\xi_i(t;\beta,R)\right) \hat{\varsigma}_i Y_i(t) e^{\beta^T Z_i(t)} Z_i(t)}{\sum_{i=1}^{n} w_i(t;\beta,R) g\left(\xi_i(t;\beta,R)\right) \hat{\varsigma}_i Y_i(t) e^{\beta^T Z_i(t)}} \right\} dN_i(t) = 0. \tag{5.9c}$$

5.4 Numerical algorithm

Implementation of the proposed estimation procedure is stated as follows.

i. Starting with initial the Breslow estimator $dR^{(0)}(t_*) = 1/n$ and $\left(\theta^{(0)}, \beta^{(0)}\right) = (0,0)$, we obtain

$$S^{(0)}(t \mid Z) = \exp\left\{-G\left(\frac{t}{n}\right)\right\},$$

and

$$w_i^{EM(0)} = \frac{S^{(0)}(t_i)}{S^{(0)}(t_i)+1}.$$

ii. Denote k as the indicator of iterations. Given $w_i^{(k)}$ and $w_i^{EM(k)}$, first obtain $dR^{(k+1)}$ from (5.9b) and then $\left(\theta^{(k+1)}, \beta^{(k+1)}\right)$ from (5.9a) and (5.9c).

iii. The estimate of the survival function is updated as

$$S^{(k+1)}(t \mid Z) = \exp\left\{-G\left(\int_0^t e^{\beta^{(k)}Z} dR^{(k)}(s)\right)\right\},$$

which is applied to obtain

$$w_i^{EM(k+1)} = \frac{\pi_i(\theta^{(k)}) \times S^{(k)}(t_i)}{\pi_i(\theta^{(k)}) \times S^{(k)}(t_i) + 1 - \pi_i(\theta^{(k)})}$$

and $\hat{\varsigma}_i^{(k+1)} = \delta_i + (1-\delta_i) w_i^{EM(k+1)}$. Then the weights $w_i^{(k+1)}$ are obtained from (5.5a),(5.5b) and (5.5c).

iv. Repeat the steps (ii) and (iii) for $k = 0,1,2,...$ until convergence.

5.5 Simulation analysis

5.5.1 Data generation

We first generate covariate $Z = (Z_1, Z_2)^T$ where $Z_1 \sim Ber(0.5)$ and $Z_2 \sim N(0,1)$ truncated at ± 2 and. Then we generate ς which follows the Bernoulli distribution with probability

$$p_Z = \pi(\theta_0 \mid Z) = \frac{\exp(\theta_0^{(0)} + \theta_0^{(1)} Z_1)}{1+\exp(\theta_0^{(0)} + \theta_0^{(1)} Z_1)},$$

where $\theta_0 = (\theta_0^{(0)}, \theta_0^{(1)})^T$.

If $\varsigma = 1$, we generate the latency variable T with

$$\tilde{S}_Z(t) = \exp\left(-G\left\{\int_0^t e^{Z^T\beta_0} dR(u)\right\}\right),$$

where $Z^T\beta_0 = (Z_1, Z_2)^T(\beta_0^{(1)}, \beta_0^{(2)}) = \beta_0^{(1)}Z_1 + \beta_0^{(2)}Z_2$. To obtain T, we can generate $U \sim Unif(0,1)$ and then solve

$$U = \tilde{S}_Z(T) = \exp\left(-G\left\{\int_0^T \mathrm{e}^{Z^T\beta_0} dR(u)\right\}\right).$$

Accordingly

$$G^{-1}(-\log U) = e^{Z^T\beta_0}\int_0^T dR(u).$$

In the simulations, we choose the proportional odds model with $G(t) = \log(1+t)$ and $R(t) = t^2$. Hence $G^{-1}(t) = \exp(t) - 1$ and $e^{-\log U} - 1 = e^{Z^T\beta_0} \cdot T^2$. Finally for $\varsigma = 1$, we let

$$T = \sqrt{e^{-Z^T\beta_0}\left(e^{-\log U} - 1\right)} = \sqrt{\frac{1-U}{U}e^{-Z^T\beta_0}}.$$

If $\varsigma = 0$, we set T to be a very large number exceeding the support of the censoring variable C which follows the uniform distribution $Uniform(0, \tau_C)$. The values of τ_C are set to yield the censoring rates $\Pr(\delta = 0)$ with 25% and 40% respectively. Note that when $\Pr(\varsigma = 1) = 1$, this setting is the same as that in Chen (2009).

5.5.2 Simulation results

Tables 3A and 3B summarize the results for $n = 200$ and $n = 500$ based on 1000 replications. In Table 3A, the bias and standard error of $\hat{\theta}_0$ and $\hat{\theta}_1$ as well as those of $\hat{p}_0$ ($p_0 = 0.88$) and $\hat{p}_1$ ($p_1 = 0.73$) look reasonable. From Table 3B, the proposed parameter estimators of β_j $(j = 1, 2)$ are virtually unbiased and have satisfactory standard deviations which improve as the sample size increases. The performances of $R(t)$ are evaluated at selected value of $t = t_p$ which is the *p*th percentile of $S(t \mid Z = 0)$. The estimator of R_p is roughly unbiased but becomes more variable as $t = t_p$ increases. Recall that the setting is an extension of the one in

Chen (2009) in which $\Pr(\varsigma = 1)$ and hence we can make comparison to examine the effect of cure on estimation of β. In present of cure, the proposed estimators are still roughly unbiased, but the standard deviations slightly increase.

Chapter 6

Proposed Approach for Transformation Cure Model Under Dependent Censoring

6.1 Model assumptions

Now we consider the more complicated but commonly-seen situation that, under the mixture framework, the event of interest is subject to competing risks such as death. Let T_1 be the time to the event of interest, say disease onset and T_2 be the time to dependent censoring. We denote Z as the vector of covariates. Also assume the mixture framework such that only a proportion of subjects with $\varsigma = 1$ develop the disease. The incidence rate is also described by the logistic model:

$$\pi(\theta \mid Z) = \frac{\exp(\theta^T Z)}{1 + \exp(\theta^T Z)} . \tag{6.1a}$$

For $\varsigma = 1$, the marginal cumulative hazard function of $T_j \mid \varsigma = 1$ can be written as

$$\tilde{\Lambda}_j(t \mid Z) = G_j\left(\int_0^t e^{\beta_j^T Z(s)} dR_j(s)\right), \tag{6.1b}$$

where $G_j(.)$ is a known functions and $R_j(.)$ is an unknown increasing function for both $j = 1,2$.

For those with $\varsigma = 0$, we assume that $T_1 = \infty$ and

$$\Pr(T_2 > t \mid \varsigma = 0, Z) = \Pr(T_2 > t \mid \varsigma = 1, Z) = \Pr(T_2 > t \mid Z) . \tag{6.1c}$$

This assumption implies that the competing risk events follow the same distribution for the susceptible and cured populations. The joint survival function of $(T_1, T_2) \mid \varsigma = 1$ is assumed to follow a copula model of the form:

$$\tilde{S}(s,t) = \Pr(T_1 > s, T_2 > t \mid Z, \varsigma = 1) = C_\alpha\{\tilde{S}_1(s \mid Z), \tilde{S}_2(t \mid Z)\} , \tag{6.1d}$$

where $\tilde{S}_j(t \mid Z) = \exp\{-\tilde{\Lambda}_j(t \mid Z)\}$ and α measures the degree of association between T_1 and T_2.

We introduce some popular copula models.

a. *Clayton copula* (Clayton, 1978): $C_\alpha(u,v) = (u^{1-\alpha} + v^{1-\alpha} - 1)^{1/(1-\alpha)}, \alpha > 1$;

b. *Frank copula* (Frank,1979): $C_\alpha(u,v) = \log_\alpha\left\{1 - \frac{(1-u^\alpha)(1-v^\alpha)}{1-\alpha}\right\}, 0 < \alpha < 1$;

c. *Positive stable copula* (Hougaard, 1986):

$$C_\alpha(u,v) = \exp\left[-\left\{(-\log u)^{1/\alpha} + (-\log v)^{1/\alpha}\right\}^\alpha\right], 0 < \alpha < 1.$$

Note that the independent copula is degenerated case with $C_\alpha(u,v) = uv$.

It is important to mention that when $\Pr(\varsigma = 1) = 1$, the model assumptions reduce to the framework discussed in Chen (2010) who adopted the likelihood approach for parameter estimation. For inference, we will take the same likelihood principle combined with the EM technique. Before presenting the detailed likelihood derivations, it is useful to discuss how to impute the unknown ς_i under the new assumptions.

6.2 Imputation under the new models

In absence of external censoring, data consists of $\{(T_i, \Delta_i, Z_i)(i = 1,...,n)\}$, where $T_i = T_{1i} \wedge T_{2i}$ and $\Delta_i = I(T_{1i} \le T_{2i})$. Notice that

$$E(\varsigma_i = 1 \mid T_i, \Delta_i) = \Delta_i + (1 - \Delta_i)E(\varsigma_i = 1 \mid T_i, \Delta_i = 0).$$

and

$$\Pr(\varsigma_i = 1 \mid T_i, \Delta_i = 0) = \frac{\Pr(\varsigma_i = 1, T_{1i} > T_i, T_{2i} = T_i)}{\Pr(\varsigma_i = 1, T_{1i} > T_i, T_{2i} = T_i) + \Pr(\varsigma_i = 0, T_{1i} > T_i, T_{2i} = T_i)}$$
$$= \frac{\Pr(T_{1i} > T_i, T_{2i} = T_i \mid \varsigma_i = 1)\Pr(\varsigma_i = 1)}{\Pr(T_{1i} > T_i, T_{2i} = T_i \mid \varsigma_i = 1)\Pr(\varsigma_i = 1) + \Pr(\varsigma_i = 0)\Pr(T_{2i} = T_i \mid \varsigma_i = 0)},$$

where

$$\Pr(T_{1i} > t, T_{2i} = t \mid \varsigma_i = 1) = C_\alpha^{01}\{\tilde{S}_1(s \mid Z), \tilde{S}_2(t \mid Z)\}\tilde{S}_2(\Delta t \mid Z),$$

$$\Pr(T_{2i} = t \mid \varsigma_i = 0) = \tilde{S}_2(\Delta t)$$

and $\tilde{S}_j(\Delta t \mid Z) = \exp\{-\tilde{\Lambda}_j(t \mid Z)\}\tilde{\Lambda}_j(\Delta t \mid Z)$. Note that if $g(t)$ is increasing, $g(\Delta t) = g(t) - g(t-)$; while if $g(t)$ is decreasing, $g(\Delta t) = g(t-) - g(t)$.

In presence of censoring, data consists of $\{(T_i, \delta_{1i}, \delta_{2i}, Z_i)(i=1,...,n)\}$, where $T_i = T_{1i} \wedge T_{2i} \wedge C_i$, $\delta_{1i} = I(T_{1i} = T_i)$, $\delta_{2i} = I(T_{2i} = T_i)$ and C_i is the censoring variable independent of ς_i and (T_{1i}, T_{2i}). Notice that $E[I(\varsigma_i = 1 | T_i, \delta_{1i}, \delta_{2i})]$ can be written as

$$\delta_{1i} + \delta_{2i} \Pr(\varsigma_i = 1 | T_i, \delta_{2i} = 1)] + (1-\delta_{1i})(1-\delta_{2i}) \Pr(\varsigma_i = 1 | T_i, \delta_{1i} = \delta_{2i} = 0).$$

We have

$$\begin{aligned}
&\Pr(\varsigma_i = 1 | T_i, \delta_{2i} = 1) \\
&= \frac{\Pr(\varsigma_i = 1, T_{1i} > T_i, T_{2i} = T_i, C_i > T_i)]}{\Pr(\varsigma_i = 1, T_{1i} > T_i, T_{2i} = T_i, C_i > T_i) + \Pr(\varsigma_i = 0, T_{2i} = T_i, C_i > T_i)} \\
&= \frac{\Pr(T_{1i} > T_i, T_{2i} = T_i, C_i > T_i | \varsigma_i = 1) \Pr(\varsigma_i = 1)}{\Pr(T_{1i} > T_i, T_{2i} = T_i, C_i > T_i | \varsigma_i = 1) \Pr(\varsigma_i = 1) + \Pr(\varsigma_i = 0) \Pr(T_{2i} = T_i, C_i > T_i | \varsigma_i = 0)} \\
&= \frac{\Pr(T_{1i} > T_i, T_{2i} = T_i | \varsigma_i = 1) \Pr(C_i > T_i) \Pr(\varsigma_i = 1)}{\Pr(T_{1i} > T_i, T_{2i} = T_i | \varsigma_i = 1) \Pr(C_i > T_i) \Pr(\varsigma_i = 1) + \Pr(\varsigma_i = 0) \tilde{S}_2(\Delta T_i) \Pr(C_i > T_i)} \\
&= \frac{\Pr(T_{1i} > T_i, T_{2i} = T_i | \varsigma_i = 1) \Pr(\varsigma_i = 1)}{\Pr(T_{1i} > T_i, T_{2i} = T_i | \varsigma_i = 1) \Pr(\varsigma_i = 1) + \Pr(\varsigma_i = 0) \tilde{S}_2(\Delta T_i)}; \qquad (6.2a)
\end{aligned}$$

and

$$\begin{aligned}
&\Pr(\varsigma_i = 1 | T_i, \delta_{1i} = \delta_{2i} = 0) \\
&= \frac{\Pr(\varsigma_i = 1, T_{1i} > T_i, T_{2i} > T_i, C_i = T_i)}{\Pr(\varsigma_i = 1, T_{1i} > T_i, T_{2i} > T_i, C_i = T_i) + \Pr(\varsigma_i = 0, T_{2i} > T_i, C_i = T_i)} \\
&= \frac{\Pr(T_{1i} > T_i, T_{2i} > T_i, C_i = T_i | \varsigma_i = 1) \Pr(\varsigma_i = 1)}{\Pr(T_{1i} > T_i, T_{2i} > T_i, C_i = T_i | \varsigma_i = 1) \Pr(\varsigma_i = 1) + \Pr(\varsigma_i = 0) \Pr(T_{2i} > T_i, C_i = T_i | \varsigma_i = 0)} \\
&= \frac{\Pr(T_{1i} > T_i, T_{2i} > T_i | \varsigma_i = 1) S_C(\Delta T_i) \Pr(\varsigma_i = 1)}{\Pr(T_{1i} > T_i, T_{2i} > T_i | \varsigma_i = 1) S_C(\Delta T_i) \Pr(\varsigma_i = 1) + \Pr(\varsigma_i = 0) \tilde{S}_2(T_i) S_C(\Delta T_i)} \\
&= \frac{\Pr(T_{1i} > T_i, T_{2i} > T_i | \varsigma_i = 1) \Pr(\varsigma_i = 1)}{\Pr(T_{1i} > T_i, T_{2i} > T_i | \varsigma_i = 1) \Pr(\varsigma_i = 1) + \Pr(\varsigma_i = 0) \tilde{S}_2(T_i)}. \qquad (6.2b)
\end{aligned}$$

Notice that the effect of censoring is cancelled out in the above derivations which implies that the conditional means only contain model parameters. This simplification avoids estimation of nuisance parameter in subsequent inference.

6.3 Likelihood analysis under dependent censorship

We extend the results of Chen (2010) by temporarily assuming that ς_i is observable. Additional step of imputation is needed for further implementation. Based on completed data $\{(T_i, \delta_{1i}, \delta_{2i}, \varsigma_i), i = 1, ..., n\}$, the log-likelihood function can be written as

$$\sum_{i=1}^{n}\left\{\delta_{1i}\log\frac{\Pr(T_1\in[t_i,t_i+\Delta),T_2>t_i\mid\varsigma_i=1)}{\Pr(T_1\geq t_i,T_2\geq t_i\mid\varsigma_i=1)}+\delta_{2i}\varsigma_i\log\frac{\Pr(T_1>t_i,T_2\in[t_i,t_i+\Delta)\mid\varsigma_i=1)}{\Pr(T_1\geq t_i,T_2\geq t_i\mid\varsigma_i=1)}\right.$$

$$+\delta_{2i}(1-\varsigma_i)\log\frac{\Pr(T_2\in[t_i,t_i+\Delta)\mid\varsigma_i=0)}{\Pr(T_2\geq t_i\mid\varsigma_i=0)}+\varsigma_i\log\Pr(T_1\geq t_i,T_2\geq t_i\mid\varsigma_i=1)$$

$$+(1-\varsigma_i)\log\Pr(T_2\geq t_i\mid\varsigma_i=0)+\left(\delta_{1i}+\delta_{2i}\varsigma_i\right)\log\Pr(\varsigma_i=1)$$

$$\left.+(1-\delta_{1i})(1-\varsigma_i)\log\Pr(\varsigma_i=0)\right\}.$$

Notice that the log-likelihood function contains the cause-specified hazard probabilities:

$$\tilde{\Lambda}_1^{CS}(\Delta t)=\Pr(T_1\in[t,t+\Delta),T_2>t\mid T_1\geq t,T_2\geq t,\varsigma=1)=\frac{C_\alpha^{10}\left(\tilde{S}_1(t),\tilde{S}_2(t)\right)\cdot\left(\tilde{S}_1(\Delta t)\right)}{C_\alpha\left(\tilde{S}_1(t),\tilde{S}_2(t)\right)};$$

$$\tilde{\Lambda}_2^{CS}(\Delta t)=\Pr(T_1>t,T_2\in[t,t+\Delta)\mid T_1\geq t,T_2\geq t,\varsigma=1)$$

$$=\frac{C_\alpha^{01}\left(\tilde{S}_1(t),\tilde{S}_2(t)\right)\tilde{S}_2(\Delta t)}{C_\alpha\left(\tilde{S}_1(t),\tilde{S}_2(t)\right)},$$

where $C_\alpha^{10}(u_1,u_2)=\dfrac{C_\alpha(u_1,u_2)}{\partial u_1}$ and $C_\alpha^{01}(u_1,u_2)=\dfrac{C_\alpha(u_1,u_2)}{\partial u_2}$.

$$=\sum_{i=1}^{n}\left\{\delta_{1i}\log\tilde{\Lambda}_1^{CS}(\Delta t_i)+\delta_{2i}\varsigma_i\log\tilde{\Lambda}_2^{CS}(\Delta t_i)+\delta_{2i}(1-\varsigma_i)\log\tilde{\Lambda}_2(\Delta t_i)+\right.$$

$$\varsigma_i\log C_\alpha\left(\tilde{S}_1(t_i),\tilde{S}_2(t_i)\right)+(1-\varsigma_i)\log\tilde{S}_2(t_i)+\left(\delta_{1i}+\delta_{2i}\varsigma_i\right)\log\Pr(\varsigma_i=1)+$$

$$\left.(1-\delta_{1i})(1-\varsigma_i)\log\Pr(\varsigma_i=0)\right\}.$$

Note that $\delta_{1i}+\delta_{2i}\varsigma_i=\varsigma_i$ and $\delta_{1i}=\delta_{1i}\varsigma_i$. Accordingly, the log-likelihood function can be written as:

$$\sum_{i=1}^{n}\left\{\delta_{1i}\log\tilde{\Lambda}_1^{CS}(\Delta t_i)+\delta_{2i}\varsigma_i\log\tilde{\Lambda}_2^{CS}(\Delta t_i)+\delta_{2i}(1-\varsigma_i)\log\tilde{\Lambda}_2(\Delta t_i)\right.$$

$$+\varsigma_i\log C_\alpha\left(\tilde{S}_1(t_i),\tilde{S}_2(t_i)\right)+(1-\varsigma_i)\log\tilde{S}_2(t_i)+\left(\delta_{1i}+\delta_{2i}\varsigma_i\right)\log\Pr(\varsigma_i=1)$$

$$\left.+(1-\delta_{1i})(1-\varsigma_i)\log\Pr(\varsigma_i=0)\right\}.$$

Including the effect of covariates, consider data of $\left\{\left(T_i,\delta_{1i},\delta_{2i},Z_i,\varsigma_i\right),i=1,...,n\right\}$. The next objective is to re-parameterize the log-likelihood function in terms of the model parameters in

(6.1a) ~ (6.1d). To simplify the presentation, we illustrate the analysis in terms of time-independent covariates since adapting to time-varying covariates is straightforward. For $j=1,2$, we derive the following quantities. The probability density function is given by

$$\tilde{S}_j(\Delta t\,|\,Z)=\exp\{-\tilde{\Lambda}_j(t\,|\,Z)\}\tilde{\Lambda}_j(\Delta t\,|\,Z)$$

$$=e^{-G_j\left\{\gamma_j(t;\beta,R)\right\}}\cdot g_j\left\{\gamma_j(t;\beta,R)\right\}\cdot e^{\beta_j^T Z_j(t)}dR_j(t)\,, \tag{6.3a}$$

where $g_j(t)=\frac{\partial}{\partial t}G_j(t)$ and $\gamma_j(t;\beta,R)=\int_0^t e^{\beta_j^T Z(s)}dR_j(s)$. The corresponding cause-specified hazard function is given by

$$\tilde{\Lambda}_j^{CS}(\Delta t)\ =\frac{\partial}{\partial u_j}\log C_\alpha\left(u_1,u_2\right)|_{u_j=\tilde{S}_j(t)}\ \tilde{S}_j(\Delta t)$$

$$=D\left(u_1,u_2\right)|_{u_j=\tilde{S}_j(t)}\ \tilde{S}_j(\Delta t)\,, \tag{6.3b}$$

where $\Phi(u_1,u_2)=-\log C_\alpha(u_1,u_2)$, $D_j(u_1,u_2)=-\frac{\partial}{\partial u_j}\Phi(u_1,u_2)$, $j=1,2$. Hence, $\tilde{\Lambda}_j^{CS}(\Delta t)$ can be re-expressed as follows:

$$\left(-\frac{\partial}{\partial u_j}\Phi\left(u_1,u_2\right)|_{u_j=\exp\left\{-G_j\left(\gamma_j(t;\beta,R)\right)\right\}}\right)e^{-G_j\left\{\gamma_j(t;\beta,R)\right\}}g_j\left\{\gamma_j(t;\beta,R)\right\}I(T>t)e^{\beta_j^T Z(t)}dR_j(t)$$

$$=I(T>t)D_j\left(e^{-G_1(\gamma_1(t;\beta,R))},e^{-G_2(\gamma_2(t;\beta,R))}\right)e^{-G_j\left\{\gamma_j(t;\beta,R)\right\}}g_j\left\{\gamma_j(t;\beta,R)\right\}e^{\beta_j^T Z(t)}dR_j(t)\,.$$

To simplify the notations, define

$$\eta_j(t;\beta,R)=D_j\left(e^{-G_1(\gamma_1(t;\beta_1,R_1))},e^{-G_2(\gamma_2(t;\beta_2,R_2))}\right)e^{-G_j\left\{\gamma_j(t;\beta_j,R_j)\right\}}g_j\left\{\gamma_j(t;\beta_j,R_j)\right\}. \tag{6.3c}$$

Thus,

$$\tilde{\Lambda}_1^{CS}(\Delta t;\beta_1,R_1)=\eta_1(t-;\beta,R)\cdot I(T\ge t)e^{\beta_1^T Z(t)}dR_1(t) \tag{6.4a}$$

and

$$\tilde{\Lambda}_2^{CS}(\Delta t;\beta_2,R_2)=\eta_2(t-;\beta,R)I(T\ge t)e^{\beta_2^T Z(t)}dR_2(t)\,. \tag{6.4b}$$

Notice that

$$\frac{\partial}{\partial t}\Phi\left(e^{-G_1(\gamma_1(t;\beta,R))},e^{-G_2(\gamma_2(t;\beta,R))}\right)$$

$$= \sum_{j=1}^{2} \left(-\frac{\partial}{\partial u_j} \Phi(u_1, u_2)|_{u_j = e^{-G(\gamma_j(t_j;\beta,R))}} \right) e^{-G(\gamma_j(t;\beta,R))} g\left(\gamma_j(t;\beta,R)\right) I(T \ge t) e^{\beta_j^T Z(t)} dR_j(t)$$

$$= \sum_{j=1}^{2} \eta_j(t-;\beta,R) I(T \ge t) e^{\beta_j^T Z(t)} dR_j(t),$$

where $\eta_j(t;\beta,R) = D_j\left(e^{-G_1(\gamma_1(t;\beta_1,R_1))}, e^{-G_2(\gamma_2(t;\beta_2,R_2))}\right) e^{-G_j\{\gamma_j(t;\beta_j,R_j)\}} g_j\{\ \gamma_j(t;\beta_j,R_j)\ \}$.

Thus, we obtain the re-expression for $\Phi\left(e^{-G(\gamma_1(\tau;\beta,R))}, e^{-G(\gamma_2(\tau;\beta,R))}\right)$ as

$$\sum_{j=1}^{2} \int_0^{\tau} \eta_j(t-;\beta,R) I(T \ge t) e^{\beta_j^T Z(t)} dR_j(t).$$

Accordingly we obtain

$$E\left\{dN_j(t) \mid F_{t-}\right\} = g_j\left(\gamma_j(t-;\beta,R)\right) \cdot Y(t) \cdot e^{\beta_j^T Z_j(t)} \cdot dR_j(t).$$

The cumulative intensity function can be re-expressed as follows:

$$G_j(\beta,R) = \int_0^{\tau} g_j\left(\gamma_j(t-;\beta_j,R_j)\right) Y(t) e^{\beta_j^T Z(t)} dR_j(t). \qquad (6.4c)$$

Therefore, the log-likelihood function for parameters $\left(dR_1, dR_2, \beta_1, \beta_2\right)$ is given by

$$\sum_{i=1}^{n} \Big\{ \varsigma_i \int_0^{\tau} \left(\log \eta_{1i}(t-;\beta,R) + \beta_1^T Z_i(t) + \log dR_1(t)\right) dN_{1i}(t)$$

$$+ \varsigma_i \int_0^{\tau} \left(\log \eta_{2i}(t-;\beta,R) + \beta_2^T Z_i(t) + \log dR_2(t)\right) dN_{2i}(t)$$

$$+ (1-\varsigma_i) \int_0^{\tau} \left[\log g_2\left(\gamma_{2i}(t;\beta_2,R_2)\right) + \beta_2^T Z_i(t) + \log dR_2(t)\right] dN_{2i}(t)$$

$$-(1-\varsigma_i) \cdot \int_0^{\tau} g_2\left(\gamma_{2i}(t-;\beta_2,R_2)\right) Y_i(t) e^{\beta_2^T Z_i(t)} dR_2(t)$$

$$-\varsigma_i \cdot \sum_{j=1}^{2} \int_0^{\tau} \eta_j(t;\beta,R) \cdot I(T > t) e^{\beta_j^T Z_i(t)} dR_j(t) \Big\}. \qquad (6.5)$$

6.4 Score equations under dependent censorship

Let t_* be the observed event-*j* time and assume $R_j(t_*)$ is step function at jump time t_*, for $j = 1,2$. Differentiating the log-likelihood function with respect to $dR_j(t_*)$ involves the following two derivative equations. By (6.3c), we can define the derivative $\eta'_{jk}(t;\beta,R)$ as the differentiation of $\eta_j(t;\beta,R)$ with respect to $\gamma_k(t;\beta_j,R_j)$ $(k,j=1,2)$ as follows.

If $k=j$,

$$\eta'_j(t;\beta,R)=\frac{\partial}{\partial u_j}D_j\left(u_1,u_2\right)|_{u_j=e^{-G_j(\gamma_j(t;\beta,R))}}\cdot e^{-2G_j\{\gamma_j(t;\beta_j,R_j)\}}\cdot g_j^2\left\{\gamma_j(t;\beta_j,R_j)\right\}$$
$$+D_j\left(e^{-G_1(\gamma_1(t;\beta_1,R_1))},e^{-G_2(\gamma_2(t;\beta_2,R_2))}\right)\cdot e^{-G_j\{\gamma_j(t;\beta_j,R_j)\}}\cdot g_j^2\left\{\gamma_j(t;\beta_j,R_j)\right\}$$
$$+D_j\left(e^{-G_1(\gamma_1(t;\beta_1,R_1))},e^{-G_2(\gamma_2(t;\beta_2,R_2))}\right)\cdot e^{-G_j\{\gamma_j(t;\beta_j,R_j)\}}\cdot g'_j\left\{\gamma_j(t;\beta_j,R_j)\right\}; \quad (6.6a)$$

if $k\neq j$,

$$\eta'_{jk}(t;\beta,R)=-\frac{\partial}{\partial u_k}D_j\left(u_1,u_2\right)|_{u_j=e^{-G_j(\gamma_j(t;\beta,R))}}\cdot\prod_{j=1}^{2}e^{-G_j\{\gamma_j(t;\beta_j,R_j)\}}\cdot g_j\left\{\gamma_j(t;\beta_j,R_j)\right\}. \quad (6.6b)$$

Hence, the score function for $dR_1(t_*)$ can be written as

$$\frac{\partial l}{\partial dR_1(t_*)}=\sum_{i=1}^{n}\left\{\frac{dN_{1i}(t_*)}{dR_1(t_*)}+\frac{\partial}{\partial\, dR_1(t_*)}\int_0^\tau\varsigma_i\cdot\log\eta_{1i}(t-;\beta,R)\,dN_{1i}(t)\right.$$
$$+\frac{\partial}{\partial\, dR_1(t_*)}\int_0^\tau\varsigma_i\cdot\log\eta_{2i}(t-;\beta,R)\,dN_{2i}(t)$$
$$\left.-\varsigma_i\frac{\partial}{\partial\, dR_1(t_*)}\sum_{j=1}^{2}\int_0^\tau\eta_j(t;\beta,R)I(T_i>t)e^{\beta_j^TZ_i(t)}dR_j(t)\right\}$$

$$=\sum_{i=1}^{n}\left\{\frac{dN_{1i}(t_*)}{dR_1(t_*)}-w_{1i}(t_*-;\beta,R)\cdot\varsigma_i\cdot\eta_{1i}(t_*-;\beta,R)\cdot I(T_i>t_*)\cdot e^{\beta_1^TZ_i(t)}\right\},$$

where $$w_{1i}(t_{1*};\beta,R)=1-\frac{\int_{t_*+}^{\tau}\psi_{1i}(t-;\beta,R)\cdot dM_{1i}(t)}{\eta_{1i}(t_*-;\beta,R)}-\frac{\int_{t_*+}^{\tau}\psi_{21i}(t-;\beta,R)\cdot dM_{2i}(t)}{\eta_{1i}(t_*-;\beta,R)}, \quad (6.7a)$$

$$\psi_{1i}(t-;\beta,R)=\frac{\eta'_{1i}(t-;\beta,R)}{\eta_{1i}(t-;\beta,R)} \quad \text{and} \quad \psi_{21i}(t-;\beta,R)=\frac{\eta'_{21i}(t-;\beta,R)}{\eta_{2i}(t-;\beta,R)}, \quad (6.7b)(6.7c)$$

$$dM_{1i}(t)=dN_{1i}(t)-\eta_{1i}(t-;\beta,R)\cdot I(T_i>t)\cdot e^{\beta_1^TZ_i(t)}\cdot dR_1(t) \quad \text{and} \quad (6.7d)$$

$$dM_{2i}(t)=dN_{2i}(t)-\eta_{2i}(t-;\beta,R)\cdot I(T_i>t)\cdot e^{\beta_2^TZ_i(t)}\cdot dR_2(t). \quad (6.7e)$$

The resulting maximum likelihood estimator for $dR_1(t_*)$ is given by

$$dR_1(t_*) = \frac{\sum_{i=1}^{n} dN_{1i}(t_*)}{\sum_{i=1}^{n} w_{1i}(t_*-;\beta,R)\cdot\varsigma_i\cdot\eta_{1i}(t_*-;\beta,R)\cdot I(T_i>t_*)\cdot e^{\beta_1^T Z_i(t)}}. \tag{6.8}$$

The score function for $dR_2(t_*)$ is given by

$$\begin{aligned}
\frac{\partial l}{\partial dR_2(t_*)} &= \sum_{i=1}^{n}\left\{ \frac{\partial}{\partial dR_2(t_*)}\int_0^\tau \varsigma_i\left(\log\eta_{1i}(t-;\beta,R)+\beta_1^T Z_i(t)+\log dR_1(t)\right)dN_{1i}(t)\right.\\
&\quad + \frac{\partial}{\partial dR_2(t_*)}\int_0^\tau \varsigma_i\left(\log\eta_{2i}(t-;\beta,R)+\beta_2^T Z_i(t)+\log dR_2(t)\right)dN_{2i}(t)\\
&\quad + \frac{\partial}{\partial dR_2(t_*)}\int_0^\tau (1-\varsigma_i)\cdot\left[\log g_2\left(\gamma_{2i}(t;\beta_2,R_2)\right)+\beta_2^T Z_i(t)+\log dR_2(t)\right]dN_{2i}(t)\\
&\quad -(1-\varsigma_i)\cdot\frac{\partial}{\partial dR_2(t_*)}\int_0^\tau g_2\left(\gamma_{2i}(t-;\beta_2,R_2)\right)\cdot I(T_i>t)\cdot e^{\beta_2^T Z_i(t)}\cdot dR_2(t)\\
&\quad \left. -\varsigma_i\cdot\frac{\partial}{\partial dR_2(t_*)}\sum_{j=1}^{2}\int_0^\tau \eta_j(t;\beta,R)\cdot I(T_i>t)\, e^{\beta_j^T Z_i(t)} dR_j(t)\right\}\\
&= \sum_{i=1}^{n}\left\{ \frac{dN_{2i}(t)}{dR_2(t_*)} - \varsigma_i w_{2i}(t_*;\beta,R)\eta_{2i}(t_*-;\beta,R) I(T_i>t_*)\, e^{\beta_2^T Z_i(t)}\right.\\
&\quad \left. -(1-\varsigma_i)\tilde{w}_{2i}(t_*;\beta_2,R_2) g_2\left(\gamma_{2i}(t_*-;\beta_2,R_2)\right) I(T_i>t_*) e^{\beta_2^T Z_i(t)}\right\},
\end{aligned}$$

where

$$\psi_{2i}(t-;\beta,R) = \frac{\eta'_{2i}(t-;\beta,R)}{\eta_{2i}(t-;\beta,R)}, \tag{6.8a}$$

$$\psi_{12i}(t-;\beta,R) = \frac{\eta'_{12i}(t-;\beta,R)}{\eta_{1i}(t-;\beta,R)}, \tag{6.8b}$$

$$\varphi_{2i}(t;\beta_2,R_2) = \frac{g'_2\left(\gamma_{2i}(t;\beta_2,R_2)\right)}{g_2\left(\gamma_{2i}(t;\beta_2,R_2)\right)}, \tag{6.8c}$$

$$w_{2i}(t_*;\beta,R) = 1 - \frac{\int_{t_*+}^{\tau}\psi_{2i}(t-;\beta,R)\cdot dM_{2i}(t)}{\eta_{2i}(t_*-;\beta,R)} - \frac{\int_{t_*+}^{\tau}\psi_{12i}(t-;\beta,R)\cdot dM_{1i}(t)}{\eta_{2i}(t_*-;\beta,R)}, \tag{6.8d}$$

$$\tilde{w}_i(t_*;\beta,R) = 1 - \frac{\int_{t_*+}^{\tau}\phi_{2i}(t-;\beta,R)\cdot d\tilde{M}_{2i}(t)}{g_{2i}(t_*-;\beta,R)} . \tag{6.8e}$$

and

$$d\tilde{M}_{2i}(t) = dN_{2i}(t) - g_2\left(\gamma_{2i}(t;\beta_2,R_2)\right)Y_i(t)e^{\beta_2^T Z_i(t)}dR_2(t)\,. \qquad (6.8f)$$

Accordingly the maximum likelihood estimator of $dR_2(t_*)$ is given by

$$dR_2(t_*) = \frac{\sum_{i=1}^{n} dN_{2i}(t_*)}{\left(\begin{array}{l}\sum_{i=1}^{n} w_{2i}(t_*-;\beta,R)\varsigma_i\eta_{2i}(t_*-;\beta,R)I(T_i>t_*)e^{\beta_2^T Z_i(t)} \\ +\sum_{i=1}^{n}\tilde{w}_{2i}(t_*-;\beta,R)(1-\varsigma_i)g_{2i}(t_*-;\beta_2,R_2)I(T_i>t_*)\,e^{\beta_2^T Z_i(t)}\end{array}\right)}\,. \qquad (6.9)$$

The score equation for β_1 is given by:

$$\begin{aligned}\frac{\partial l}{\partial \beta_1} &= \sum_{i=1}^{n}\Bigg\{ \frac{\partial}{\partial \beta_1}\int_0^{\tau}\varsigma_i\left(\log\eta_{1i}(t-;\beta,R)+\beta_1^T Z_i(t)+\log dR_1(t)\right)dN_{1i}(t) \\ &\quad +\frac{\partial}{\partial \beta_1}\int_0^{\tau}\varsigma_i\left(\log\eta_{2i}(t-;\beta,R)+\beta_2^T Z_i(t)+\log dR_2(t)\right)dN_{2i}(t) \\ &\quad -\varsigma_i\frac{\partial}{\partial \beta_1}\sum_{j=1}^{2}\int_0^{\tau}\eta_j(t;\beta,R)I(T_i>t)e^{\beta_j^T Z_i(t)}dR_j(t)\Bigg\} \\ &= \sum_{i=1}^{n}\Bigg\{ \int_0^{\tau}\varsigma_i Z_i(t)dN_{1i}(t) \\ &\quad -\varsigma_i\int_0^{\tau}\left(1-\frac{\int_s^{\tau}\psi_{1i}(t-;\beta,R)\cdot dM_{1i}(t)}{\eta_{1i}(t-;\beta,R)}-\frac{\int_s^{\tau}\psi_{21i}(t-;\beta,R)\cdot dM_{2i}(t)}{\eta_{1i}(t-;\beta,R)}\right)\eta_{1i}(s;\beta,R)I(T_i>s)e^{\beta_1^T Z_i(s)}Z_i(s)dR_1(s)\Bigg\} \\ &= \sum_{i=1}^{n}\left\{ \int_0^{\tau}\varsigma_i Z_i(t)dN_{1i}(t)-\varsigma_i\int_0^{\tau}w_{1i}(t;\beta,R)\eta_{1i}(t;\beta,R)I(T_i>t)e^{\beta_1^T Z_i(t)}Z_i(t)dR_1(t)\right\}.\end{aligned}$$

Replacing $dR_1(t)$ by the maximum likelihood estimator for $dR_1(t)$, it implies that

$$\frac{\partial l}{\partial \beta_1} = \sum_{i=1}^{n}\int_0^{\tau}\left\{ Z_i(t)-\frac{\sum_{i=1}^{n} w_{1i}(t;\beta,R)\cdot\varsigma_i\cdot\eta_{1i}(t;\beta,R)\cdot I(T_i>t)\cdot e^{\beta_1^T Z_i(t)}\cdot Z_i(t)}{\sum_{i=1}^{n} w_{1i}(t;\beta,R)\cdot\varsigma_i\cdot\eta_{1i}(t;\beta,R)\cdot I(T_i>t)\cdot e^{\beta_1^T Z_i(t)}}\right\}dN_{1i}(t)\,. \qquad (6.10)$$

The score equation for β_2 is given by:

$$\begin{aligned}\frac{\partial l}{\partial \beta_2} &= \sum_{i=1}^{n}\Bigg\{ \frac{\partial}{\partial \beta_2}\int_0^{\tau}\varsigma_i\left(\log\eta_{1i}(t-;\beta,R)+\beta_1^T Z_i(t)+\log dR_1(t)\right)dN_{1i}(t) \\ &\quad +\frac{\partial}{\partial \beta_2}\int_0^{\tau}\varsigma_i\left(\log\eta_{2i}(t-;\beta,R)+\beta_2^T Z_i(t)+\log dR_2(t)\right)dN_{2i}(t)\end{aligned}$$

$$+\frac{\partial}{\partial\beta_2}\int_0^\tau (1-\varsigma_i)\Big[\ \log g_2\left(\gamma_{2i}(t;\beta_2,R_2)\right)+\beta_2^T Z_i(t)+\log dR_2(t)\Big]dN_{2i}(t)$$

$$-(1-\varsigma_i)\cdot\frac{\partial}{\partial\beta_2}\int_0^\tau g_2\left(\gamma_{2i}(t-;\beta_2,R_2)\right)Y_i(t)e^{\beta_2^T Z_i(t)}dR_2(t)$$

$$\left.-\varsigma_i\frac{\partial}{\partial\beta_2}\sum_{j=1}^{2}\int_0^\tau \eta_j(t;\beta,R)I(T_i>t)e^{\beta_j^T Z_i(t)}dR_j(t)\right\},$$

$$=\sum_{i=1}^{n}\Big\{\int_0^\tau Z_i\Big(dN_{2i}(t)-\varsigma_i w_{2i}(t-;\beta,R)\,I(T_i>t)\eta_{2i}(t-;\beta,R)\,e^{\beta_2^T Z_i(t)}dR_2(t)$$

$$-(1-\varsigma_i)\tilde{w}_{2i}(t-;\beta,R)\,I(T_i>t)g_{2i}(t-;\beta,R)\,e^{\beta_2^T Z_i(t)}dR_2(t)\Big)\Big\}.$$

Replacing the m.l.e of $dR_2(t)$, the score function for β_2 becomes

$$\frac{\partial l}{\partial\beta_2}=\sum_{i=1}^{n}\int_0^\tau\left\{Z_i(t)-\frac{\left(\begin{array}{l}\sum_{i=1}^{n}w_{2i}(t-;\beta,R)Z_i(t)\varsigma_i\eta_{2i}(t-;\beta,R)I(T_i>t)\,e^{\beta_2^T Z_i(t)}\\+\sum_{i=1}^{n}\tilde{w}_{2i}(t-;\beta,R)Z_i(t)(1-\varsigma_i)g_{2i}(t-;\beta_2,R_2)I(T_i>t)\,e^{\beta_2^T Z_i(t)}\end{array}\right)}{\left(\begin{array}{l}\sum_{i=1}^{n}w_{2i}(t-;\beta,R)\varsigma_i\eta_{2i}(t-;\beta,R)I(T_i>t)\,e^{\beta_2^T Z_i(t)}\\+\sum_{i=1}^{n}\tilde{w}_{2i}(t-;\beta,R)(1-\varsigma_i)g_{2i}(t-;\beta_2,R_2)I(T_i>t)\,e^{\beta_2^T Z_i(t)}\end{array}\right)}\right\}dN_{2i}(t).\quad(6.11)$$

6.5 Numerical algorithm

Implementation of the proposed estimation procedure is stated as follows.

i. Starting with initial the Breslow estimator $dR_j^{(0)}(t_*)=1/n$, for $j=1,2$, and $\left(\theta^{(0)},\beta^{(0)}\right)=(0,0)$,

we obtain

$$S_j^{(0)}(t\,|\,Z)=\exp\left\{-G_j\left(\frac{t}{n}\right)\right\},\text{ for } j=1,2,$$

$$C_{00}^{(0)}(t\,|\,Z)=C_\alpha\left(S_1^{(0)}(t\,|\,Z),S_2^{(0)}(t\,|\,Z)\right),$$

$$C_{01}^{(0)}\left(t\,|\,Z\right)=\frac{\partial}{\partial u_2}C_\alpha\left(u_1,u_2\right)|_{u_j=S_j^{(0)}(t|Z)},$$

$$w_i^{EM(0)}=\delta_{1i}+\delta_{2i}\frac{C_{01}^{(0)}(t_i)S_j^{(0)}(\Delta t\,|\,Z)}{C_{01}^{(0)}(t_i)S_j^{(0)}(\Delta t\,|\,Z)+S_j^{(0)}(\Delta t\,|\,Z)}+(1-\delta_{1i})(1-\delta_{2i})\frac{C_{00}^{(0)}(t_i)}{C_{00}^{(0)}(t_i)+S_j^{(0)}(t\,|\,Z)}$$ and the

weights $w_{1i}^{(0)},w_{2i}^{(0)},\tilde{w}_i^{(0)}$ are obtained from (6.7a)~(6.7e),(6.8a)~(6.8e).

ii. Denote k as the indicator of iterations. Given $w_{1i}^{(k)}, w_{2i}^{(k)}, \tilde{w}_i^{(k)}$ and $w_i^{EM(k)}$, first obtain $dR_1^{(k+1)}$ from (6.8), $dR_2^{(k+1)}$ from (6.9) and then $\left(\theta^{(k+1)}, \beta_1^{(k+1)}, \beta_2^{(k+1)}\right)$ from (6.10) and (6.11).

iii. The estimate of the survival function is updated as

$$S_j^{(k+1)}(t\,|\,Z) = \exp\left\{-G_j\left(\int_0^t e^{\beta_j^{(k)}Z} dR_j^{(k)}(s)\right)\right\}, \text{ for } j = 1,2,$$

$$C_{00}^{(k+1)}(t\,|\,Z) = C_\alpha\left(S_1^{(k+1)}(t\,|\,Z), S_2^{(k+1)}(t\,|\,Z)\right)$$

which is applied to obtain

$$w_i^{EM(k+1)} = \delta_{1i} + \delta_{2i}\frac{C_{01}^{(k)}(t_i)S_2^{(k)}(\Delta t\,|\,Z)}{C_{01}^{(k)}(t_i)S_2^{(k)}(\Delta t\,|\,Z) + S_2^{(k)}(\Delta t\,|\,Z)} + (1-\delta_{1i})(1-\delta_{2i})\frac{C_{00}^{(k)}(t_i)}{C_{00}^{(k)}(t_i) + S_2^{(k)}(t\,|\,Z)}$$

and $\hat{\varsigma}_i^{(k+1)} = \delta_i + (1-\delta_i)w_i^{EM(k+1)}$. Then the weights $w_{1i}^{(0)}$, $w_{2i}^{(0)}$, $\tilde{w}_i^{(0)}$ are obtained from (6.7a)~(6.7e),(6.8a)~(6.8e).

iv. Repeat the steps (ii) and (iii) for $k = 0,1,2,\ldots$ until convergence.

6.6 Simulation analysis

6.6.1 Data generation

We also generate covariate $Z = (Z_1, Z_2)^T$ where $Z_1 \sim Ber(0.5)$ and $Z_2 \sim N(0,1)$ truncated at ± 2 and set

$$p_Z = \pi(\theta_0\,|\,Z) = \frac{\exp(\theta_0^{(0)} + \theta_0^{(1)}Z_1)}{1+\exp(\theta_0^{(0)} + \theta_0^{(1)}Z_1)},$$

where $\theta_0 = (\theta_0^{(0)}, \theta_0^{(1)})^T$.

We assume that (Z_1, Z_2) affects the (T_1, T_2). Marginally for $T_1\,|\,\varsigma = 1$, we set $G_1(t) = \log(1+t)$ and $R_1(t) = t^2$ which corresponds to a proportional odds model with the survival function:

$$\tilde{S}_1(t\,|\,Z) = \exp\left\{-G_1\left(\int_0^t e^{\beta_1 Z_1 + \beta_2 Z_2} dR_1(u)\right)\right\}.$$

For $T_2\,|\,\varsigma = 1$, we set $G_2(t) = t$ and $R_2(t) = ct$ which corresponds to the survival function:

$$\tilde{S}_2(t \mid Z) = \exp\left\{-G_2\left(\int_0^t e^{\beta_3 Z_1 + \beta_4 Z_2} dR_2(u)\right)\right\}.$$

The value of c controls the proportion of experiencing the two events (i.e. $\Pr(T_1 \le T_2 \mid \varsigma = 1)$) and in the simulations we set $c = 0.3$ and $c = 0.15$.

Now we describe how to simulate $(T_1, T_2) \mid \varsigma = 1, Z_1, Z_2$ which jointly follows a Clayton model. At first, we generate a pair of correlated failure times (Y_1, Y_2) following the Clayton distribution with exponential marginals and the association parameter α related to Kendall's tau τ such that $\alpha = \dfrac{1+\tau}{1-\tau}$. To attain this, we perform the following steps:

i. Generate independent U_1 and U_2, both of which follow $Uniform(0,1)$;

ii. Let $aa = (1-U_1)^{1-\alpha}$. Then set

$$Y_1 = -\log(1-U_1) \text{ and } Y_2 = \frac{1}{\alpha - 1}\log(1 - aa + aa(1-U_2)^{(1-\alpha)/\alpha}).$$

Secondly we obtain $(T_1, T_2) \mid \varsigma = 1, Z_1, Z_2$ from $(\varsigma, Y_1, Y_2, Z_1, Z_2)$. Recall that $Y_j \sim \exp(1)$ and hence $S_j = \exp(-Y_j)$ follows a uniform distribution for both $j = 1, 2$. If $\varsigma = 1$, set

$$T_1 = \sqrt{\frac{1-S_1}{S_1} \mathrm{e}^{-(\beta_1 Z_1 + \beta_2 Z_2)}} \text{ and } T_2 = -\frac{1}{c}\log(S_2)\mathrm{e}^{-(\beta_3 Z_1 + \beta_4 Z_2)};$$

while if $\varsigma = 0$, set $T_1 = \infty$ (a very large number) and

$$T_2 = -\frac{1}{c}\log(S_2)\mathrm{e}^{-(\beta_3 Z_1 + \beta_4 Z_2)}.$$

Repeating the procedure n times, we have $\left\{\left(T_{1i}, T_{2i}, \varsigma_i, Z_{1i}, Z_{2i}\right), i = 1, \ldots, n\right\}$. Then we simulate the external censoring variables $C_i \sim uniform(0, \tau_c)$ $(i = 1, \ldots, n)$. By setting $T_i = \min(T_{1i}, T_{2i}, C_i)$, $\delta_{1i} = I(T_i = T_{1i})$ and $\delta_{2i} = I(T_i = T_{2i})$, observed data can be written as $\left\{\left(T_i, \delta_{1i}, \delta_{2i}, Z_{1i}, Z_{2i}\right), i = 1, \ldots, n\right\}$. The censoring support τ_c affects the censoring rate which is set to be 10 and 8 yielding about 1% and 5% rates of external censoring respectively. We set the value of τ to be 0.3 and 0.5 which controls the association of $\left(T_1, T_2\right) \mid \varsigma = 1$. The sample size is set to be $n = 300$. Two settings are evaluated:

Setting A: $(c,\tau_c)=(0.3,10)$ which corresponds to

$$\{\Pr(\delta_1=1),\Pr(\delta_2=1),\Pr(\delta_1=0,\delta_2=0)\}=(0.5,0.49,0.01);$$

Setting B: $(c,\tau_c)=(0.15,8)$ which corresponds to

$$\{\Pr(\delta_1=1),\Pr(\delta_2=1),\Pr(\delta_1=0,\delta_2=0)\}=(0.63,0.32,0.05).$$

Simulation results based on 1000 replications are provided.

6.6.2 Simulation results

Table 4A presents the results for the bias and standard error of $\hat{\theta}_0$ and $\hat{\theta}_1$ as well as those of $\hat{p}_0$ ($p_0=0.88$) and $\hat{p}_1$ ($p_1=0.73$). Most results look reasonable. However under setting A with lower $\Pr(\delta_1=1)$, $\hat{p}_1$ has larger variation. From Tables 4B & 4C, we see that the proposed parameter estimators of β are virtually unbiased. For estimating $R(t)$, the performances are better for small t. The performances seem not much affected by the chosen values of τ. Based on Tables 4B which evaluates the latency estimation, we see that Setting B which gives higher $\Pr(\delta_1=1)$ yields better results. On the other hand based on Tables 4C which evaluates the survival estimation, we see that Setting A which gives higher $\Pr(\delta_2=1)$ yields better results. It is worthy to mention that Chen (2010) considered dependent censoring without cure. The settings in Tables 4B and 4C mimic Table 1 of Chen (p243, JRSSB 2010) so that we can assess the effect of cure on the results. In present of cure, the proposed estimators are still roughly unbiased, but the standard deviations slightly increase.

To evaluate the effect of dependent censoring on estimation, we design two settings under independent censoring:

Setting A*: $\{\Pr(\delta=1),\Pr(\delta=0)\}=(0.5,0.5)$;

Setting B*: $\{\Pr(\delta=1),\Pr(\delta=0)\}=(0.63,0.37)$.

The results using the proposed methods in Chapter 5 are given in Tables 4D and 4E. We see that the presence of dependent censoring increases the variation of the proposed estimators for both the incidence and latency models.

Conclusion

Cure model provides a useful approach to describing failure time data when some subjects will never experience of the event. In the thesis, we adopt the mixture framework to analyze such data. The latency distribution is modeled by two general types of semi-parametric models.

For the first class of semi-parametric linear models, the proposed estimating functions are originally constructed based on martingale properties for complete data under the error scale. Then the information of uncertain susceptibility status is imputed by its conditional mean. Our proposal turns out to coincide with the log-rank estimating function proposed by Zhang and Peng (2007). However, the proposed approach can utilize the nice martingale structure in further inference problems. For example based on the large sample analysis, we propose a fast algorithm for variance estimation which does not require doing iterations in each re-sampling step. We also propose a model diagnostic approach and a test for model checking. For the second class of transformation models which permit r time-dependent covariates, we extended the results of Zeng and Lin (2006) and Chen (2009) to cure models. Besides independent censorship, we also consider the situation of competing risks which is an extension of the work by Chen (2010).

For practical applications, whether a cure model is appropriate at hand should consult with experts in the field. Although there exist nonparametric tests as described in the book of Maller and Zhau (1996), the condition of sufficient follow-up may not satisfied and the tests may have low power. We have performed simulations to examine the effect of fitting survival data without cure by the proposed approach assuming cure. We obtained high incidence probability close to one and unbiased (but with larger variance) estimates for the parameters in the latency model. Thus our approach is in some sense robust. However under this situation, the parameter is located on the boundary of the parameter space, the distribution of the proposed estimator may no longer be normally distributed. Model checking for the second class of models will be one future work. Extending the likelihood approach to the first class of models may deserve some investigation.

References

Chen, Y.-H. (2009). Weighted Breslow-type estimator and maximum likelihood estimation in semiparametric transformation models. *Biometrika*, **96**, 591–600.

Chen, Y.-H. (2010). Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *J. R. Statist. Soc. B*, **72**, 235 – 251**.**

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

Ghosh, D. and Lin, D. Y. (2003). Semiparametric Analysis of Recurrent Events Data in the Presence of Dependent Censoring. *Biometrics*, **59**, 877-885.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*; **38**, 1041–1046.

Farewell, V. T. (1986). Mixture Models in Survival Analysis: Are They Worth the Risk? *The Canadian Journal of Statistics*, **14**, 257-262

Frank. M. J. (1979). On the simulataneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematicae*, **19**, 194-226.

Haugaard, P. (1986). A class of multivariate failure time distributions. *Biometrika*, **73**, 671-678.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Kuk, A. Y. C. and Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79**, 531-541.

Li, C.-S. and Taylor, J. M. G. (2002). A semi-parametric accelerated failure time cure model. *Statistic in Medicine*, **21**, 3235-3247.

Lin, D. Y., Wei, L. J., and Ying, Z. (1998). Accelerated failure time models for counting processes.

Biometrika **85**, 605- 618.

Lu, W. and Ying, Z. (2004). On semiparametric transformation cure models. *Biometrika,* **91**, 331-343.

Maller, R. A., Zhou, X. (1996). *Survival Analysis with Long-term Survivors.* Wiley.

Parzen, M. I., Wei, L. J., and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341-350.

Peng Y, Dear, K. B. G., Denham, J. W. (1998). A generalized F mixture model for cure rate estimation. *Statistics in Medicine*; 17:813–830.

Peng Y. and Dear, K. B. G. (2000). A Nonparametric Mixture Model for Cure Rate Estimation. *Biometrics*, **56**, 237-243.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Annals of Statistics*; **18**, 303–328.

Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox Proportional Hazards Cure Model. *Biometrics*, **56**, 227-236.

Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis (Disc: P1881–1885). *Statistics in Medicine*, **11**, 1871–1879.

Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of 'permanent employment' in Japan. *Journal of the American Statistical Association*; **87**, 284–292.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Annals of Statistics,* **21**, 76-99.

Zhang, J. and Peng, Y. (2007). A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statistic in Medicine,* **26**, 3157-3171.

Zeng, D. & Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika,* **93**, 627–40.

Appendix 1: Proof for the order of $B_{1n}+B_{2n}$

Define $N_i(t;\beta)=I(\tilde{\varepsilon}_i(\beta)\le t,\delta_i=1)$ and $\tilde{N}_i(t;\beta)=I(\tilde{\varepsilon}_i(\beta)\ge t,\delta_i=1)$ .Notice that $dN_i(t;\beta)=N_i(t;\beta)-N_i(t-;\beta)=-d\tilde{N}_i(t;\beta)$. Recall that $Y_i^*(t;\beta)=\tilde{Y}_i(t;\beta,w_i^*)$ and $\Lambda_\varepsilon^*(t\,|\,\beta)$ is the limit of $\int_0^t\sum_{j=1}^n dN_j(u;\beta)/\sum_{j=1}^n\tilde{Y}_j(u;\beta,w_j^*)$. Define $dEN_i(t;\beta)=Y_i^*(t;\beta)d\Lambda_i^*(t;\beta)$.

Now we show $B_{1n}+B_{2n}=o(n^{1/2+\varepsilon})$ for a small value $\varepsilon>0$. We can write $B_{1n}+B_{2n}$ as follows:

$$\begin{aligned}
&\int_0^\infty\sum_{i=1}^n\left\{Z_i-\bar{Z}(t;\beta,w^*)\right\}\left(dN_i(t;\beta)-dEN_i(t;\beta))\right)-\int_0^\infty\sum_{i=1}^n\left\{Z_i-\bar{Z}(t;\beta_0,w^*)\right\}\left(dN_i(t;\beta_0)-dEN_i(t;\beta_0)\right)\\
&=\int_0^\infty\sum_{i=1}^n Z_i\left(dN_i(t;\beta)-dEN_i(t;\beta)-dN_i(t;\beta_0)-dEN_i(t;\beta_0)\right)\\
&\quad-\int_0^\infty\sum_{i=1}^n\frac{Y_Z^*(t;\beta)}{Y^*(t;\beta)}\left(dN_i(t;\beta)-dEN_i(t;\beta))\right)+\int_0^\infty\sum_{i=1}^n\frac{Y_Z^*(t;\beta_0)}{Y^*(t;\beta_0)}\left(dN_i(t;\beta_0)-dEN_i(t;\beta_0)\right).\\
&=Q_{1n}+Q_{2n}+Q_{3n},
\end{aligned}$$

where $Y^*(t;\beta)=\sum_{i=1}^n\tilde{Y}_i(t;\beta,w_i^*)$ and $Y_Z^*(t;\beta)=\sum_{i=1}^n Z_i\tilde{Y}_i(t;\beta,w_i^*)$. For $\alpha\in[0,1)$, define

$$t_\beta(\alpha)=\inf\left\{t:EY^*(t;\beta)\le n^{1-\alpha}\right\}.$$

We will use the facts that

i. $EN(0;\beta)=0$.

ii. $E\tilde{N}(t;\beta)\le EY^*(t;\beta)\le n^{1-\alpha}$ for $t>t_\beta(\alpha)$.

Then for Q_{1n}, we have

$$\begin{aligned}
&\sup_{\|\beta-\beta_0\|\le n^{-1/3}}\left\|\int_0^\infty\sum_{i=1}^n Z_i\left(dN_i(t;\beta)-dEN_i(t;\beta)-dN_i(t;\beta_0)-dEN_i(t;\beta_0)\right)\right\|\\
&\le\sup_{\|\beta-\beta_0\|\le n^{-1/3}}\left\|\int_0^{t_{\beta_0}(\alpha)}\sum_{i=1}^n Z_i\left(dN_i(t;\beta)-dEN_i(t;\beta)-dN_i(t;\beta_0)-dEN_i(t;\beta_0)\right)\right\|\\
&\quad+\sup_{\|\beta-\beta_0\|\le n^{-1/3}}\left\|\int_{t_{\beta_0}(\alpha)}^\infty\sum_{i=1}^n Z_i\left(dN_i(t;\beta)-dEN_i(t;\beta)-dN_i(t;\beta_0)-dEN_i(t;\beta_0)\right)\right\|
\end{aligned}$$

$$= \sup_{\|\beta-\beta_0\|\le n^{-1/3}} \left\| \int_0^{t_{\beta_0}(\alpha)} \sum_{i=1}^n Z_i \left(dN_i(t;\beta) - dEN_i(t;\beta) - dN_i(t;\beta_0) - dEN_i(t;\beta_0) \right) \right\|$$

$$+ \sup_{\|\beta-\beta_0\|\le n^{-1/3}} \left\| \int_{t_{\beta_0}(\alpha)}^{\infty} \sum_{i=1}^n Z_i \left(d\tilde{N}_i(t;\beta) - dE\tilde{N}_i(t;\beta) - d\tilde{N}_i(t;\beta_0) - dE\tilde{N}_i(t;\beta_0) \right) \right\| .$$

If we write

$$\int_0^\infty \sum_{i=1}^n Z_i \, dN_i(t;\beta) = N^Z(\infty;\beta) ; \quad \int_0^\infty \sum_{i=1}^n Z_i \, dEN_i(t;\beta) = EN^Z(\infty;\beta) ;$$

$$\int_0^{t_{\beta_0}(\alpha)} \sum_{i=1}^n Z_i \, dN_i(t;\beta) = N^Z(t_{\beta_0}(\alpha);\beta) ; \quad \int_0^{t_{\beta_0}(\alpha)} \sum_{i=1}^n Z_i \, dEN_i(t;\beta) = EN^Z(t_{\beta_0}(\alpha);\beta) ;$$

$$\int_{t_{\beta_0}(\alpha)}^\infty \sum_{i=1}^n Z_i \, d\tilde{N}_i(t;\beta) = \tilde{N}^Z(t_{\beta_0}(\alpha);\beta) \quad \text{and} \quad \int_{t_{\beta_0}(\alpha)}^\infty \sum_{i=1}^n Z_i \, dE\tilde{N}_i(t;\beta) = E\tilde{N}^Z(t_{\beta_0}(\alpha);\beta) .$$

It follows that

$$\sup_{\|\beta-\beta_0\|\le n^{-1/3}} \left\| N^Z(\infty;\beta) - EN^Z(\infty;\beta) - N^Z(\infty;\beta_0) - EN^Z(\infty;\beta_0) \right\|$$

$$\le \sup_{\|\beta-\beta_0\|\le n^{-1/3}} \left\| N^Z(t_{\beta_0}(\alpha);\beta) - EN^Z(t_{\beta_0}(\alpha);\beta) - N^Z(t_{\beta_0}(\alpha);\beta_0) - EN^Z(t_{\beta_0}(\alpha);\beta_0) \right\|$$

$$+ \sup_{\|\beta-\beta_0\|\le n^{-1/3}} \left\| \tilde{N}^Z(t_{\beta_0}(\alpha);\beta) - E\tilde{N}^Z(t_{\beta_0}(\alpha);\beta) - \tilde{N}^Z(t_{\beta_0}(\alpha);\beta_0) - E\tilde{N}^Z(t_{\beta_0}(\alpha);\beta_0 \right\|$$

which is the case with $v_i = N_i(t;\beta)$ or $\tilde{N}_i(t;\beta)$ in Lemma 2 and Lemma 3 of Ying (1993).

Hence

$$Q_{1n} = o(n^{(1-\alpha)/2}) , \quad \text{for} \quad \alpha \in [0,1) .$$

For Q_{2n}, without loss of generality, we assume $\sup_i \|Z_i^T\| < 1$. Using the fact of the total variation, we have $B > 0$,

$$\sup_{\|\beta\|\le B} \int_0^\infty \left| \frac{dY_Z^*(t;\beta)}{Y^*(t;\beta)} \right| \le \sup_{\|\beta\|\le B} \int_0^\infty \left| \frac{dY^*(t;\beta)}{Y^*(t;\beta)} \right| = O(\log n) .$$

By lemma 1(a) of Ying (1993) let $v_i = N_i(t;\beta)$, we have

$$\sup_{\|\beta\|\le B} \left\| N(\infty;\beta) - EN(\infty;\beta) \right\| = o(n^{(1-\alpha)/2}) ,$$

where $\int_0^\infty \sum_{i=1}^n dN_i(t;\beta) = N(\infty;\beta)$ and $\int_0^\infty \sum_{i=1}^n dEN_i(t;\beta) = EN(\infty;\beta)$.

By integration by part, it follows that

$$\sup_{\|\beta\|\le B}\left\|\int_0^\infty \frac{Y_Z^*(t;\beta)}{Y^*(t;\beta)}\left(dN(t;\beta)-dEN(t;\beta)\right)\right\|$$

$$\le \sup_{\|\beta\|\le B}\left\|dN(\infty;\beta)-dEN(\infty;\beta)\right\| + \sup_{\|\beta\|\le B}\left\|\int_0^\infty \left(dN(t;\beta)-dEN(t;\beta)\right)\frac{dY_Z^*(t;\beta)}{Y^*(t;\beta)}\right\|$$

$$= o(n^{(1-\alpha)/2}) + o(n^{(1-\alpha)/2})\cdot O(\log n)$$

$$= o(n^{(1-\alpha)/2+\varepsilon}).$$

For Q_{3n}, we can follow the idea for proving Q_{2n} and get

$$\sup_{\|\beta\|\le B}\left\|\int_0^\infty \sum_{i=1}^{n}\frac{Y_Z^*(t;\beta_0)}{Y^*(t;\beta_0)}\left(dN_i(t;\beta_0)-dEN_i(t;\beta_0)\right)\right\| = o(n^{1/2}) \quad \text{a.s.}$$

Appendix 2: Proof for $\sup_{\|\beta-\beta_0\|<n^{-1/3}} \|d_n\| = o\left(n^{1/3}\right)$

The proofs are derived under the assumption that

$$\sup_i \| w_i^* - \hat{w}_i \| \le o(n^{-1/3}) \quad \text{a.s. for all i,}$$

where

$$w_i^* = w_i(\theta, \beta, S_\varepsilon^0) \quad = \frac{\pi_i(\theta) \times S_\varepsilon^0(\tilde{\varepsilon}_i(\beta))}{\pi_i(\theta) \times S_\varepsilon^0(\tilde{\varepsilon}_i(\beta)) + \{1 - \pi_i(\theta)\}}.$$

We can write

$$\int_0^\infty \sum_{i=1}^n \left\{\overline{Z}(t;\beta,w^*) - \overline{Z}(t;\beta,\hat{w})\right\} dN_i(t;\beta) - \int_0^\infty \sum_{i=1}^n \left\{\overline{Z}(t;\beta_0,w^*) - \overline{Z}(t;\beta_0,\hat{w})\right\} dN_i(t;\beta_0)$$

$$= \int_0^\infty \sum_{i=1}^n \left\{\overline{Z}(t;\beta,w^*) - \overline{Z}(t;\beta,\hat{w}) - \overline{Z}(t;\beta_0,w^*) + \overline{Z}(t;\beta_0,\hat{w})\right\} dN_i(t;\beta)$$

$$-\int_0^\infty \sum_{i=1}^n \left\{\overline{Z}(t;\beta_0,w^*) - \overline{Z}(t;\beta_0,\hat{w})\right\} \left(dN_i(t;\beta) - dN_i(t;\beta_0)\right)$$

$$= Ⓐ + Ⓑ,$$

where

$$Ⓐ = \int_0^\infty \sum_{i=1}^n \left\{\overline{Z}(t;\beta,w^*) - \overline{Z}(t;\beta,\hat{w}) - \overline{Z}(t;\beta_0,w^*) + \overline{Z}(t;\beta_0,\hat{w})\right\} dN_i(t;\beta)$$

$$= \int_0^\infty \left\{ \frac{Y_Z^*(t;\beta)}{Y^*(t;\beta)} - \frac{\hat{Y}^Z(t;\beta)}{\hat{Y}(t;\beta)} - \frac{Y_Z^*(t;\beta_0)}{Y^*(t;\beta_0)} + \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \right\} dN(t;\beta),$$

$$Ⓑ = \int_0^\infty \sum_{i=1}^n \left\{\overline{Z}(t;\beta_0,w^*) - \overline{Z}(t;\beta_0,\hat{w})\right\} \left(dN_i(t;\beta) - dN_i(t;\beta_0)\right)$$

$$= \int_0^\infty \left\{ \frac{Y_Z^*(t;\beta_0)}{Y^*(t;\beta_0)} - \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \right\} \left(dN(t;\beta) - dN(t;\beta_0)\right).$$

We derive that

$$\frac{Y_Z^*(t;\beta)}{Y^*(t;\beta)} - \frac{\hat{Y}^Z(t;\beta)}{\hat{Y}(t;\beta)} = \frac{Y_Z^*(t;\beta)\hat{Y}(t;\beta) - Y^*(t;\beta)\hat{Y}^Z(t;\beta)}{Y^*(t;\beta)\hat{Y}(t;\beta)}$$

$$= \frac{Y_Z^*(t;\beta) - \hat{Y}^Z(t;\beta)}{Y^*(t;\beta)} - \frac{\hat{Y}^Z(t;\beta)}{\hat{Y}(t;\beta)} \cdot \frac{Y^*(t;\beta) - \hat{Y}(t;\beta)}{Y^*(t;\beta)}.$$

Similarly,

$$\frac{Y_Z^*(t;\beta_0)}{Y^*(t;\beta_0)}-\frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)}=\frac{Y_Z^*(t;\beta_0)-\hat{Y}^Z(t;\beta_0)}{Y^*(t;\beta_0)}-\frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)}\cdot\frac{Y^*(t;\beta_0)-\hat{Y}(t;\beta_0)}{Y^*(t;\beta_0)}.$$

Therefore, we obtain

$$\left\|\frac{Y_Z^*(t;\beta)}{Y^*(t;\beta)}-\frac{\hat{Y}^Z(t;\beta)}{\hat{Y}(t;\beta)}-\frac{Y_Z^*(t;\beta_0)}{Y^*(t;\beta_0)}+\frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)}\right\|$$

$$\leq\left\|\frac{Y_Z^*(t;\beta)-\hat{Y}^Z(t;\beta)}{Y^*(t;\beta)}-\frac{Y_Z^*(t;\beta_0)-\hat{Y}^Z(t;\beta_0)}{Y^*(t;\beta_0)}\right\| \tag{A1}$$

$$+\left\|\frac{\hat{Y}^Z(t;\beta)}{\hat{Y}(t;\beta)}\cdot\frac{Y^*(t;\beta)-\hat{Y}(t;\beta)}{Y^*(t;\beta)}-\frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)}\cdot\frac{Y^*(t;\beta_0)-\hat{Y}(t;\beta_0)}{Y^*(t;\beta_0)}\right\|. \tag{A2}$$

We derive that

$$Y_Z^*(t;\beta)-\hat{Y}^Z(t;\beta)=\sum_{i=1}^n(1-\delta_i)Z_i\left(w_i^*-\hat{w}_i\right)I(\tilde{\varepsilon}_i(\beta)\geq t)\leq o(n^{-1/3})R^Z(t;\beta);$$

$$Y^*(t;\beta_0)-\hat{Y}(t;\beta_0)=\sum_{i=1}^n(1-\delta_i)Z_i\left(w_i^*-\hat{w}_i\right)I(\tilde{\varepsilon}_i(\beta_0)\geq t)\leq o(n^{-1/3})R^Z(t;\beta_0),$$

where

$$R^Z(t;\beta)=\sum_{i=1}^n(1-\delta_i)Z_i\,I(\tilde{\varepsilon}_i(\beta)\geq t),\,R^Z(t;\beta_0)=\sum_{i=1}^n(1-\delta_i)Z_i\,I(\tilde{\varepsilon}_i(\beta\ )\geq t).$$

Accordingly

$$\left\|\frac{Y_Z^*(t;\beta)-\hat{Y}^Z(t;\beta)}{Y^*(t;\beta)}-\frac{Y_Z^*(t;\beta_0)-\hat{Y}^Z(t;\beta_0)}{Y^*(t;\beta_0)}\right\|$$

$$\leq o(n^{-1/3})\cdot\left\|\frac{R^Z(t;\beta)}{Y^*(t;\beta)}-\frac{R^Z(t;\beta_0)}{Y^*(t;\beta_0)}\right\|$$

$$=o(n^{-1/3})\cdot\left\|\frac{R^Z(t;\beta)-R^Z(t;\beta_0)}{Y^*(t;\beta)}-\frac{R^Z(t;\beta_0)}{Y^*(t;\beta_0)}\frac{Y^*(t;\beta)-Y^*(t;\beta_0)}{Y^*(t;\beta)}\right\|,$$

where

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}}\left\|R^Z(t;\beta)-R^Z(t;\beta_0)\right\|$$

$$=\sup_{\|\beta-\beta_0\|<n^{-1/3}}\left\|\sum_{i=1}^n(1-\delta_i)Z_i\left(I(\tilde{\varepsilon}_i(\beta_0)\geq t+(\beta-\beta_0)Z_i)-I(\tilde{\varepsilon}_i(\beta_0)\geq t)\right)\right\|$$

$$\leq n \cdot \|\beta - \beta_0\| \cdot f_{\varepsilon_0}(t;\beta_0) \leq n \cdot n^{-\frac{1}{3}} = n^{\frac{2}{3}}.$$

Hence

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{R^Z(t;\beta_0)}{Y^*(t;\beta_0)} \left(Y^*(t;\beta) - Y^*(t;\beta_0) \right) \right\|$$

$$= \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{\sum_{i=1}^{n}(1-\delta_i)Z_i I(\tilde{\varepsilon}_i(\beta_0) \geq t)}{\sum_{i=1}^{n}\left(\delta_i + (1-\delta_i)w_i^*\right) I(\tilde{\varepsilon}_i(\beta_0) \geq t)} \sum_{i=1}^{n}(1-\delta_i)w_i^* \left(I(\tilde{\varepsilon}_i(\beta) \geq t) - I(\tilde{\varepsilon}_i(\beta_0) \geq t) \right) \right\|$$

$$= \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{\sum_{i=1}^{n}(1-\delta_i)Z_i I(\tilde{\varepsilon}_i(\beta_0) \geq t)}{\sum_{i=1}^{n}\left(\delta_i + (1-\delta_i)w_i^*\right) I(\tilde{\varepsilon}_i(\beta_0) \geq t)} \sum_{i=1}^{n}(1-\delta_i)w_i^* \left[I(\tilde{\varepsilon}_i(\beta_0) \geq t + (\beta-\beta_0)Z_i) - I(\tilde{\varepsilon}_i(\beta_0) \geq t) \right] \right\|$$

$$= \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{O(n)}{O(n)} \cdot \sum_{i=1}^{n}(1-\delta_i) \cdot w_i^* \cdot \left[I(\tilde{\varepsilon}_i(\beta_0) \geq t + (\beta-\beta_0)Z_i) - I(\tilde{\varepsilon}_i(\beta_0) \geq t) \right] \right\|$$

$$= \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{O(n)}{O(n)} \cdot \sum_{i=1}^{n}(1-\delta_i) \cdot w_i^* \cdot \|\beta-\beta_0\| \cdot f_{\varepsilon_0}(t;\beta_0) \right\|$$

$$\leq O(1) \cdot O(n) \cdot n^{-\frac{1}{3}} = o(n^{\frac{2}{3}}).$$

Therefore for (A1), we have

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{Y_Z^*(t;\beta) - \hat{Y}^Z(t;\beta)}{Y^*(t;\beta)} - \frac{Y_Z^*(t;\beta_0) - \hat{Y}^Z(t;\beta_0)}{Y^*(t;\beta_0)} \right\| \leq o(n^{-1/3}) \cdot \left(\frac{n^{\frac{2}{3}} + o(n^{\frac{2}{3}})}{Y^*(t;\beta)} \right) = \frac{o(n^{\frac{1}{3}})}{Y^*(t;\beta)}.$$

For the second term (A2), we obtain

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{\hat{Y}^Z(t;\beta)}{\hat{Y}(t;\beta)} \cdot \frac{Y^*(t;\beta) - \hat{Y}(t;\beta)}{Y^*(t;\beta)} - \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \cdot \frac{Y^*(t;\beta_0) - \hat{Y}(t;\beta_0)}{Y^*(t;\beta_0)} \right\|$$

$$= \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \left(\frac{\hat{Y}^Z(t;\beta)}{\hat{Y}(t;\beta)} - \frac{\hat{Y}^Z(t;\beta_0)}{Y(t;\beta_0)} \right) \cdot \frac{Y^*(t;\beta) - Y(t;\beta)}{Y^*(t;\beta)} \right\| \tag{A3}$$

$$+ \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \cdot \left(\frac{Y^*(t;\beta_0) - Y(t;\beta_0)}{Y^*(t;\beta_0)} - \frac{Y^*(t;\beta) - \hat{Y}(t;\beta)}{Y^*(t;\beta)} \right) \right\|, \tag{A4}$$

where

$$\left\|Y^*(t;\beta)-\hat{Y}(t;\beta)\right\| = \left\|\sum_{i=1}^{n}(1-\delta_i)\left(w_i^*-\hat{w}_i\right)I(\tilde{\varepsilon}_i(\beta)\geq t)\right\| \leq o(n^{-1/3})\cdot O(n)=o(n^{2/3}).$$

We have

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}}\left\|\frac{\hat{Y}^Z(t;\beta)}{\hat{Y}^Z(t;\beta_0)}-1\right\|=o(1) \quad \text{a.s and} \quad \sup_{\|\beta-\beta_0\|<n^{-1/3}}\left\|\frac{\hat{Y}(t;\beta)}{\hat{Y}(t;\beta_0)}-1\right\|=o(1) \quad \text{a.s.}$$

It can be shown that

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}}\left\|\frac{\hat{Y}^Z(t;\beta)}{\hat{Y}(t;\beta)}-\frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)}\right\| \leq K\left(\|\beta-\beta_0\|\right)=o(n^{-\frac{1}{3}}).$$

Thus for (A3), we have

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}}\left\|\left(\frac{\hat{Y}^Z(t;\beta)}{\hat{Y}(t;\beta)}-\frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)}\right)\cdot\frac{\hat{Y}(t;\beta)-Y^*(t;\beta)}{Y^*(t;\beta)}\right\| \leq \frac{o(n^{-\frac{1}{3}})\cdot o(n^{\frac{2}{3}})}{Y^*(t;\beta)}=\frac{o(n^{\frac{1}{3}})}{Y^*(t;\beta)}.$$

By the assumption (condition) with $\sup_i \| Z_i \| \leq 1$, then

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}}\left\|\frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)}\right\| \leq \sup_{\|\beta-\beta_0\|<n^{-1/3}}\left\|\frac{\hat{Y}(t;\beta_0)}{\hat{Y}(t;\beta_0)}\right\|=O(1).$$

It follows that

$$\begin{aligned}
&\left\|Y^*(t;\beta)-Y^*(t;\beta_0)\right\| \\
&=\left\|\sum_{i=1}^{n}\left[\delta_i+(1-\delta_i)w_i^*\right]\cdot I(\tilde{\varepsilon}_i(\beta)\geq t)-\sum_{i=1}^{n}\left[\delta_i+(1-\delta_i)w_i^*\right]\cdot I(\tilde{\varepsilon}_i(\beta_0)\geq t)\right\| \\
&=\left\|\sum_{i=1}^{n}\left[\delta_i+(1-\delta_i)w_i^*\right]\cdot\left[I(\tilde{\varepsilon}_i(\beta_0)\geq t+(\beta-\beta_0)Z_i^T)-I(\tilde{\varepsilon}_i(\beta_0)\geq t)\right]\right\| \\
&\leq O(n)\cdot\|\beta-\beta_0\|\cdot f_{\varepsilon_0}(t;\beta_0)\leq O(n)\cdot o(n^{-1/3})=o(n^{2/3}).
\end{aligned}$$

Since $\left\|\dfrac{1}{Y^*(t;\beta_0)}\right\|=\dfrac{1}{O(n)}$, then

$$\left\|\frac{1}{Y^*(t;\beta_0)}-\frac{1}{Y^*(t;\beta)}\right\|=\left\|\frac{Y^*(t;\beta)-Y^*(t;\beta_0)}{Y^*(t;\beta_0)Y^*(t;\beta)}\right\|=\frac{o(n^{2/3})/O(n)}{Y^*(t;\beta)}=\frac{o(n^{-1/3})}{Y^*(t;\beta)}.$$

We have $\left\|Y^*(t;\beta)-\hat{Y}(t;\beta)\right\|=o(n^{2/3})$ for all β, so that

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{\hat{Y}(t;\beta_0)-Y^*(t;\beta_0)}{Y^*(t;\beta_0)} - \frac{\hat{Y}(t;\beta)-Y^*(t;\beta)}{Y^*(t;\beta)} \right\| = o(n^{\frac{2}{3}}) \left\| \frac{1}{Y^*(t;\beta_0)} - \frac{1}{Y^*(t;\beta)} \right\|$$

$$= o(n^{\frac{2}{3}}) \cdot \frac{o(n^{-1/3})}{Y^*(t;\beta)} = \frac{o(n^{\frac{1}{3}})}{Y^*(t;\beta)}.$$

Thus for (A4),

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \cdot \left(\frac{\hat{Y}(t;\beta_0)-Y^*(t;\beta_0)}{Y^*(t;\beta_0)} - \frac{\hat{Y}(t;\beta)-Y^*(t;\beta)}{Y^*(t;\beta)} \right) \right\| = \frac{o(n^{\frac{1}{3}})}{Y^*(t;\beta)}.$$

Therefore, we obtain

$$\text{Ⓐ} = \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \int_0^\infty \left\{ \frac{Y^Z(t;\beta)}{Y^*(t;\beta)} - \frac{\hat{Y}^Z(t;\beta)}{\hat{Y}(t;\beta)} - \frac{Y^Z(t;\beta_0)}{Y^*(t;\beta_0)} + \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \right\} dN(t;\beta) \right\|$$

$$\le \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \int_0^\infty \left(\frac{\hat{Y}^Z(t;\beta)}{\hat{Y}(t;\beta)} - \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \right) \cdot \frac{\hat{Y}(t;\beta)-Y^*(t;\beta)}{Y^*(t;\beta)} dN(t;\beta) \right\|$$

$$+ \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \cdot \left(\frac{\hat{Y}(t;\beta_0)-Y^*(t;\beta_0)}{Y^*(t;\beta_0)} - \frac{\hat{Y}(t;\beta)-Y^*(t;\beta)}{Y^*(t;\beta)} \right) \right\|$$

$$\le o(n^{\frac{1}{3}}) \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \int_0^\infty \frac{1}{Y^*(t;\beta)} dN(t;\beta) \right\| = o(n^{\frac{1}{3}}) O(\log n) = o(n^{\frac{1}{3}})$$

and also

$$\text{Ⓑ} = \int_0^\infty \left\{ \frac{Y_Z^*(t;\beta_0)}{Y^*(t;\beta_0)} - \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \right\} (dN(t;\beta) - dN(t;\beta_0))$$

$$= \int_0^\infty \left\{ \frac{Y_Z^*(t;\beta_0)}{Y^*(t;\beta_0)} - \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \right\} (dN(t;\beta) - dN(t;\beta_0)),$$

where

$$\|dN(t;\beta) - dN(t;\beta_0)\| = \left\| \sum_{i=1}^n \delta_i \left(I\left(\tilde{\varepsilon}_i(\beta_0) = t + (\beta-\beta_0) Z_i^T\right) - I\left(\tilde{\varepsilon}_i(\beta_0) = t\right) \right) \right\|$$

$$\le n \|\beta - \beta_0\| f'_{\varepsilon_0}(t;\beta_0) dt$$

and

$$\left\| Y_Z^*(t;\beta_0) - \hat{Y}^Z(t;\beta_0) \right\| = \left\| \sum_{i=1}^{n} (1-\delta_i) Z_i^T \left(w_i^* - \hat{w}_i \right) I(\tilde{\varepsilon}_i(\beta_0) \ge t) \right\|$$

$$\le o(n^{-1/3}) \cdot O(n) = o(n^{2/3}) .$$

Similarly, $\left\| Y^*(t;\beta_0) - \hat{Y}(t;\beta_0) \right\| = o(n^{2/3})$, and we have

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \right\| \le \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \frac{\hat{Y}(t;\beta_0)}{\hat{Y}(t;\beta_0)} \right\| = O(1) ,$$

$$\sup \left\| \frac{Y_Z^*(t;\beta_0)}{Y^*(t;\beta_0)} - \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \right\| \le \sup \left\| \frac{Y_Z^*(t;\beta_0) - \hat{Y}^Z(t;\beta_0)}{Y^*(t;\beta_0)} \right\| + \sup \left\| \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \cdot \frac{Y^*(t;\beta_0) - \hat{Y}(t;\beta_0)}{Y^*(t;\beta_0)} \right\|$$

$$\le \frac{o(n^{2/3})}{O(n)} = o(n^{-1/3}) .$$

Hence,

$$\text{Ⓑ} = \sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \int_0^\infty \left\{ \frac{Y_Z^*(t;\beta_0)}{Y^*(t;\beta_0)} - \frac{\hat{Y}^Z(t;\beta_0)}{\hat{Y}(t;\beta_0)} \right\} \left(dN(t;\beta) - dN(t;\beta_0) \right) \right\|$$

$$\le \int_0^\infty o(n^{-1/3}) \cdot n \cdot \| \beta - \beta_0 \| f'_{\varepsilon_0}(t;\beta_0)\, dt$$

$$\le o(n^{-1/3}) \cdot n \cdot n^{-1/3} \int_0^\infty f'_{\varepsilon_0}(t;\beta_0)\, dt$$

$$= o(n^{1/3}) .$$

By Ⓐ and Ⓑ, we complete the proof for

$$\sup_{\|\beta-\beta_0\|<n^{-1/3}} \left\| \int_0^\infty \sum_{i=1}^{n} \left\{ \overline{Z}(t;\beta,w^*) - \overline{Z}(t;\beta,\hat{w}) \right\} dN_i(t;\beta) - \int_0^\infty \sum_{i=1}^{n} \left\{ \overline{Z}(t;\beta_0,w^*) - \overline{Z}(t;\beta_0,\hat{w}) \right\} dN_i(t;\beta_0) \right\| \le o\left(n^{1/3} \right)$$

Appendix 3: Proof for $r_n = o(n^{1/2})$

The technique of verifying $r_n = o(n^{1/2})$ is similar as proving $B_{1n} + B_{2n} = o(n^{1/2})$ which involves applying the property of $dN_i(t;\beta) - dEN_i(t;\beta))$ which is $o(n^{(1-\alpha)/2})$. The difference is in the at-risk set. We replace w_i^* in $\tilde{Y}_i(t;\beta, w_i^*)$ by $\hat{w}_i$ and obtain $\hat{Y}_i(t;\beta) = Y_i(t;\beta,\hat{w}_i)$. Note that $\hat{Y}(t;\beta) = \sum_{i=1}^{n} Y_i(t;\beta,\hat{w}_i)$ and $\hat{Y}_Z(t;\beta) = \sum_{i=1}^{n} Z_i Y_i(t;\beta,\hat{w}_i)$.

Then we can also obtain that

$$\sup_{\|\beta\|\le B} \int_0^\infty \left| \frac{d\hat{Y}_Z(t;\beta)}{\hat{Y}(t;\beta)} \right| \le \sup_{\|\beta\|\le B} \int_0^\infty \left| \frac{d\hat{Y}(t;\beta)}{\hat{Y}(t;\beta)} \right| = O(\log n)$$

Apply the integration by part, we can get $r_n = o(n^{1/2})$.

Appendix 4: Derivation of the difference between

$\bar{Z}(t;\beta_0,\hat{w})$ and $\bar{Z}(t;\beta_0,w^*)$.

We derive that

$$\bar{Z}(t;\beta_0,\hat{w})=\frac{\sum_{i=1}^{n} Z_i\left(\delta_i+(1-\delta_i)\hat{w}_i\right)I(\tilde{\varepsilon}(\beta_0)\geq t)}{\sum_{i=1}^{n}\left(\delta_i+(1-\delta_i)\hat{w}_i\right)I(\tilde{\varepsilon}(\beta_0)\geq t)}$$

$$=\frac{\left(\begin{array}{l}\sum_{i=1}^{n} Z_i\left(\delta_i+(1-\delta_i)\hat{w}_i\right)I(\tilde{\varepsilon}(\beta_0)\geq t)-\sum_{i=1}^{n} Z_i\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)\\ +\sum_{i=1}^{n} Z_i\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)\end{array}\right)}{\left(\begin{array}{l}\sum_{i=1}^{n}\left(\delta_i+(1-\delta_i)\hat{w}_i\right)I(\tilde{\varepsilon}(\beta_0)\geq t)-\sum_{i=1}^{n}\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)\\ +\sum_{i=1}^{n}\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)\end{array}\right)}$$

$$=\frac{\sum_{i=1}^{n} Z_i(1-\delta_i)(\hat{w}_i-w_i^*)I(\tilde{\varepsilon}(\beta_0)\geq t)+\sum_{i=1}^{n} Z_i\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)}{\sum_{i=1}^{n}(1-\delta_i)(\hat{w}_i-w_i^*)I(\tilde{\varepsilon}(\beta_0)\geq t)+\sum_{i=1}^{n}\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)}$$

$$=\frac{\left(\dfrac{\sum_{i=1}^{n} Z_i(1-\delta_i)(\hat{w}_i-w_i^*)I(\tilde{\varepsilon}(\beta_0)\geq t)}{\sum_{i=1}^{n} Z_i\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)}+1\right)}{\left(\dfrac{\sum_{i=1}^{n}(1-\delta_i)(\hat{w}_i-w_i^*)I(\tilde{\varepsilon}(\beta_0)\geq t)}{\sum_{i=1}^{n}\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)}+1\right)}\times\frac{\sum_{i=1}^{n} Z_i\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)}{\sum_{i=1}^{n}\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)},$$

which, based on the assumption that $\sup_i\left\|\hat{w}-w^*\right\|<o(n^{-1/3})$, can be written as

$$\frac{\left(\dfrac{o(n^{2/3})}{O(n)}+1\right)}{\left(\dfrac{o(n^{2/3})}{O(n)}+1\right)}\times\frac{\sum_{i=1}^{n} Z_i\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)}{\sum_{i=1}^{n}\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)}$$

which converges in probability to

$$\bar{Z}(t;\beta_0,w^*)=\frac{\sum_{i=1}^{n}Z_i\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)}{\sum_{i=1}^{n}\left(\delta_i+(1-\delta_i)w_i^*\right)I(\tilde{\varepsilon}(\beta_0)\geq t)}.$$

Appendix 5: Discussion on the validity of the modified re-sampling algorithm

Consider three weights, denoted as $\hat{w}$, w^*, $\hat{w}^*$, representing the proposed weight formula as a function of (θ,β), the true weight and the final estimated weight. Consider three estimators of β_0, denoted $\hat{\beta}$, $\hat{\beta}^*$ and β^*, which solve $U(\beta\,|\,\hat{w})=0$, $U(\beta\,|\,\hat{w}^*)=0$ and $U(\beta\,|\,w^*)=0$ respectively. We aim to claim that the asymptotic variances of $\hat{\beta}$, $\hat{\beta}^*$ and β^* are the same. The results depend on whether $U(\beta_0\,|\,\hat{w})=U(\beta_0\,|\,w^*)+o(n^{1/2})$ and $U(\beta_0\,|\,\hat{w}^*)=U(\beta_0\,|\,w^*)+o(n^{1/2})$. When the two statements are true, the asymptotic normality of $U(\beta_0\,|\,w^*)$ yield the following results:

$$\sqrt{n}\left(\beta^*-\beta_0\right)\Rightarrow Normal(0,(A)^{-1}\Sigma(A)^{-1})$$

and

$$\sqrt{n}\left(\hat{\beta}^*-\beta_0\right)\Rightarrow Normal(0,(A)^{-1}\Sigma(A)^{-1}).$$

For variance estimation, we will use $U(\beta\,|\,\hat{w}^*)=0$ as the basis in the re-sampling algorithm since, unlike w^*, $\hat{w}^*$ is available and does not involve performing iterations.

			$\theta_0^{(0)}=0$		$\theta_0^{(1)}=0.5$		$p_0=0.5$		$p_1=0.6225$	
Setting	Censoring Rate	n	Bias($\hat{\theta}_0^{(0)}$)	$\sigma(\hat{\theta}_0^{(0)})$	Bias($\hat{\theta}_0^{(1)}$)	$\sigma(\hat{\theta}_0^{(1)})$	Bias($\hat{p}_0$)	$\sigma(\hat{p}_0)$	Bias($\hat{p}_1$)	$\sigma(\hat{p}_1)$
A	0.4630	100	0.0128	0.2919	0.0270	0.4412	0.0031	0.0714	0.0061	0.0737
		200	0.0014	0.2042	0.0130	0.3043	0.0004	0.0505	0.0020	0.0507
		500	0.0088	0.1386	0.0012	0.2063	0.0022	0.0345	0.0018	0.0331
	0.5741	100	0.0617	0.3518	0.0168	0.5559	0.0148	0.0851	0.0125	0.0939
		200	0.0493	0.2465	0.0190	0.3926	0.0121	0.0605	0.0132	0.0660
		500	0.0234	0.1546	0.0031	0.2449	0.0058	0.0384	0.0052	0.0429
B	0.4611	100	0.0184	0.3046	0.0069	0.4517	0.0045	0.0746	0.0029	0.0733
		200	0.0105	0.2162	-0.0036	0.3165	0.0026	0.0534	0.0003	0.0506
		500	0.0011	0.1323	-0.0016	0.1935	0.0003	0.0329	-0.0006	0.0319
	0.5340	100	0.0638	0.3389	0.0154	0.5592	0.0154	0.0824	0.0125	0.0913
		200	0.0388	0.2343	0.0123	0.3744	0.0096	0.0578	0.0096	0.0630
		500	0.0315	0.1440	0.0101	0.2256	0.0078	0.0358	0.0089	0.0388

Table 1A: Finite-sample performances for estimating $\theta_0^{(0)}, \theta_0^{(1)}$, p_0 and p_1

under AFT model based on 1000 replications

			$\theta_0^{(0)} = 0$		$\theta_0^{(1)} = 3.0$		$p_0 = 0.5$		$p_1 = 0.9526$	
Setting	Censoring Rate	n	Bias($\hat{\theta}_0^{(0)}$)	$\sigma(\hat{\theta}_0^{(0)})$	Bias($\hat{\theta}_0^{(1)}$)	$\sigma(\hat{\theta}_0^{(1)})$	Bias($\hat{p}_0$)	$\sigma(\hat{p}_0)$	Bias($\hat{p}_1$)	$\sigma(\hat{p}_1)$
A	0.3083	100	0.0181	0.3047	-0.0686	0.6554	0.0044	0.0744	-0.0098	0.0312
		200	-0.0012	0.2062	0.0986	0.5897	-0.0003	0.0510	-0.0012	0.0227
		500	0.0081	0.1386	0.0380	0.4096	0.0020	0.0345	-0.0005	0.0152
	0.4584	100	0.0159	0.3568	-0.3540	0.6263	0.0037	0.0864	-0.0256	0.0383
		200	0.0118	0.2490	-0.0169	0.5999	0.0029	0.0613	-0.0067	0.0278
		500	0.0147	0.1547	0.0103	0.4333	0.0036	0.0384	-0.0018	0.0191
B	0.3054	100	0.0192	0.3020	-0.0342	0.6573	0.0046	0.0740	-0.0080	0.0305
		200	0.0085	0.2163	0.0837	0.5832	0.0021	0.0535	-0.0012	0.0221
		500	0.0000	0.1324	0.0572	0.3818	0.0000	0.0329	0.0002	0.0146
	0.4052	100	0.0163	0.3435	-0.3611	0.6013	0.0033	0.0842	-0.0272	0.0505
		200	0.0097	0.2279	0.0206	0.5837	0.0024	0.0562	-0.0095	0.0280
		500	0.0016	0.1426	0.0096	0.4293	0.0004	0.0355	0.0005	0.0189

Table 1B: Finite-sample performances for estimating $\theta_0^{(0)}, \theta_0^{(1)}$, p_0 and p_1

under AFT model based on 1000 replications

			Estimation of $\beta_0^{(1)}$					Estimation of $\beta_0^{(2)}$				
Setting	**Censoring rate**	**Sample size**	$Bias(\hat{\beta}_1)$	$Bias(\hat{\beta}_1^*)$	$se(\hat{\beta}_1)$	Avg $\hat{\sigma}(\hat{\beta}_1)$	CP	$Bias(\hat{\beta}_2)$	$Bias(\hat{\beta}_2^*)$	$se(\hat{\beta}_2)$	Avg $\hat{\sigma}(\hat{\beta}_2)$	CP
A	0.4630	100	-0.0058	-0.0199	0.3040	0.3378	0.96	0.0166	-0.0082	0.2965	0.3367	0.94
		200	0.0090	-0.0387	0.2083	0.2228	0.95	-0.0072	-0.0046	0.2036	0.2196	0.92
		500	0.0019	-0.0049	0.1282	0.1272	0.93	-0.0051	-0.0034	0.1309	0.1245	0.92
	0.5741	100	-0.0182	-0.0050	0.3776	0.4494	0.975	-0.0209	0.0229	0.4711	0.4452	0.915
		200	-0.0070	-0.0405	0.2534	0.2823	0.935	-0.0080	-0.0136	0.3355	0.2667	0.84
		500	0.0009	0.0029	0.1558	0.1648	0.955	-0.0020	-0.0047	0.2025	0.1592	0.895
B	0.4611	100	0.0164	-0.0166	0.3044	0.3273	0.97	0.0047	0.0083	0.5287	0.5599	0.905
		200	0.0023	-0.0442	0.2080	0.2165	0.945	0.0092	0.0082	0.3693	0.3629	0.945
		500	-0.0032	-0.0033	0.1234	0.1266	0.95	0.0052	-0.0022	0.2216	0.2153	0.94
	0.5340	100	-0.0094	-0.013	0.3659	0.3977	0.94	-0.0185	-0.0274	0.7294	0.6700	0.9
		200	-0.0080	-0.0584	0.2566	0.2487	0.92	0.0106	-0.0435	0.4926	0.4168	0.89
		500	-0.0028	-0.0033	0.1527	0.1470	0.935	0.0037	-0.0221	0.2998	0.2542	0.905

Table 2A: Performances of proposed estimators of $\beta_0^{(j)}$ and $\sigma(\hat{\beta}_j)$ under SP model based on 1000 replications with $p_1 = 0.6225$.

Notes: $\hat{\beta}$ is the average of proposed estimator β solving (3.4b) and $bias(\hat{\beta})$ is the average bias in 1000 replications. $\hat{\beta}^*$ is the average of the solution to (3.10e) based on R= 200 re-sampling runs and $bias(\hat{\beta}^*)$ is the average bias in 1000 replications. $se(\hat{\beta}_1)$ is the sample standard error of $\hat{\beta}$ based on 1000 replications. $\hat{\sigma}(\hat{\beta}_1)$ is the sample standard deviation of $\hat{\beta}^*$ based on R= 200 re-sampling runs and Avg $\hat{\sigma}(\hat{\beta}_1)$ is the average in 1000 replications. CP is the coverage probability of the Wald 95% confidence interval using $\hat{\sigma}(\hat{\beta}_1)$ in the formula.

			Estimation of $\beta_0^{(1)}$					Estimation of $\beta_0^{(2)}$				
Setting	**Censored rate**	**Sample size**	$Bias(\hat{\beta}_1)$	$Bias(\hat{\beta}_1^*)$	$se(\hat{\beta}_1)$	Avg $\hat{\sigma}(\hat{\beta}_1)$	CP	$Bias(\hat{\beta}_2)$	$Bias(\hat{\beta}_2^*)$	$se(\hat{\beta}_2)$	Avg $\hat{\sigma}(\hat{\beta}_2)$	CP
A	0.3083	100	-0.0137	-0.0397	0.2717	0.2742	0.945	-0.0240	-0.0233	0.2585	0.2602	0.93
		200	0.0061	-0.0356	0.1877	0.1846	0.96	-0.0117	0.0058	0.1798	0.1789	0.905
		500	0.0007	0.0011	0.1166	0.1164	0.94	-0.0062	-0.0083	0.1140	0.1105	0.96
	0.4584	100	-0.0229	-0.0206	0.3240	0.3641	0.97	-0.0373	0.029	0.3683	0.3823	0.94
		200	-0.0012	-0.0131	0.2221	0.2436	0.98	0.0045	0.0289	0.2512	0.2482	0.905
		500	0.0037	0.0006	0.1371	0.1493	0.98	-0.0016	0.0035	0.1595	0.1483	0.94
B	0.3054	100	0.0113	-0.0483	0.2713	0.2722	0.935	-0.0178	-0.0667	0.4646	0.4435	0.915
		200	0.0028	-0.0377	0.1878	0.1855	0.955	0.0105	0.0137	0.3227	0.3004	0.935
		500	-0.0031	-0.0023	0.1103	0.1159	0.94	0.0075	-0.0149	0.1920	0.1908	0.97
	0.4052	100	-0.0542	-0.0604	0.3096	0.3219	0.95	-0.0793	-0.0019	0.5766	0.5384	0.905
		200	-0.0208	-0.0551	0.2164	0.2197	0.945	-0.0172	0.0094	0.3890	0.3695	0.94
		500	-0.0082	-0.0021	0.1390	0.1346	0.95	0.0050	-0.0095	0.2417	0.227	0.95

Table 2B: Performances of proposed estimators of $\beta_0^{(j)}$ and $\sigma(\hat{\beta}_j)$ under SP model based on 1000 replications with $p_1 = 0.9526$.

Notes: $\hat{\beta}$ is the average of proposed estimator β solving (3.4b) and $bias(\hat{\beta})$ is the average bias in 1000 replications. $\hat{\beta}^*$ is the average of the solution to (3.10e) based on R= 200 re-sampling runs and $bias(\hat{\beta}^*)$ is the average bias in 1000 replications. $se(\hat{\beta}_1)$ is the sample standard error of $\hat{\beta}$ based on 1000 replications. $\hat{\sigma}(\hat{\beta}_1)$ is the sample standard deviation of $\hat{\beta}^*$ based on R= 200 re-sampling runs and Avg $\hat{\sigma}(\hat{\beta}_1)$ is the average in 1000 replications. CP is the coverage probability of the Wald 95% confidence interval using $\hat{\sigma}(\hat{\beta}_1)$ in the formula.

Sample size = 200	$Z_1 \sim Ber(0.5)$			$Z_2 \sim U(0,1)$		
Censoring rate	P-value AFT	P-value LS	Proportion selection for AFT	P-value AFT	P-value LS	Proportion selection for AFT
0.3075	0.8390	0.0330	0.99	0.8313	0.0550	0.99
0.49	0.7347	0.0516	1.00	0.5705	0.0771	1.00
0.4628	0.7943	0.0514	1.00	0.7722	0.0774	0.98
0.5727	0.7833	0.0525	1.00	0.6036	0.0807	1.00

Table 2C: Results of model diagnostics based on p-value and selected proportion with fitting AFT and LS model based on 100 datasets simulated from AFT model.

			Average mean for r_n				Average mean for $\lvert r_n \rvert$			
Setting			$Z_1 \sim Ber(0.5)$		$Z_1 \sim U(0,1)$		$Z_1 \sim Ber(0.5)$		$Z_1 \sim U(0,1)$	
θ_0	(β_1,β_2)	Censored rate	$n=100$	$n=2000$	$n=100$	$n=2000$	$n=100$	$n=2000$	$n=100$	$n=2000$
3.0	$(0.5,1)$	0.3054	-0.0011	-0.0001	-0.0009	-0.0002	0.0046	0.0018	0.0036	0.0018
0.5	$(0.5,1)$	0.4611	-0.0010	0.0001	-0.0008	0.0002	0.0043	0.0024	0.0028	0.0018
3.0	$(0.8,3)$	0.4052	-0.0020	-0.0019	-0.0023	-0.0011	0.0146	0.0063	0.0090	0.0078
0.5	$(0.8,3)$	0.5340	-0.0009	-0.0007	-0.0031	-0.0006	0.0150	0.0064	0.0128	0.0099

Table 2D: Performances of $r_n = \frac{1}{\sqrt{n}}\left(U(\beta_0 \mid \hat{w}) - U(\beta_0 \mid w^*)\right)$ **with n = 100 and n = 2000**

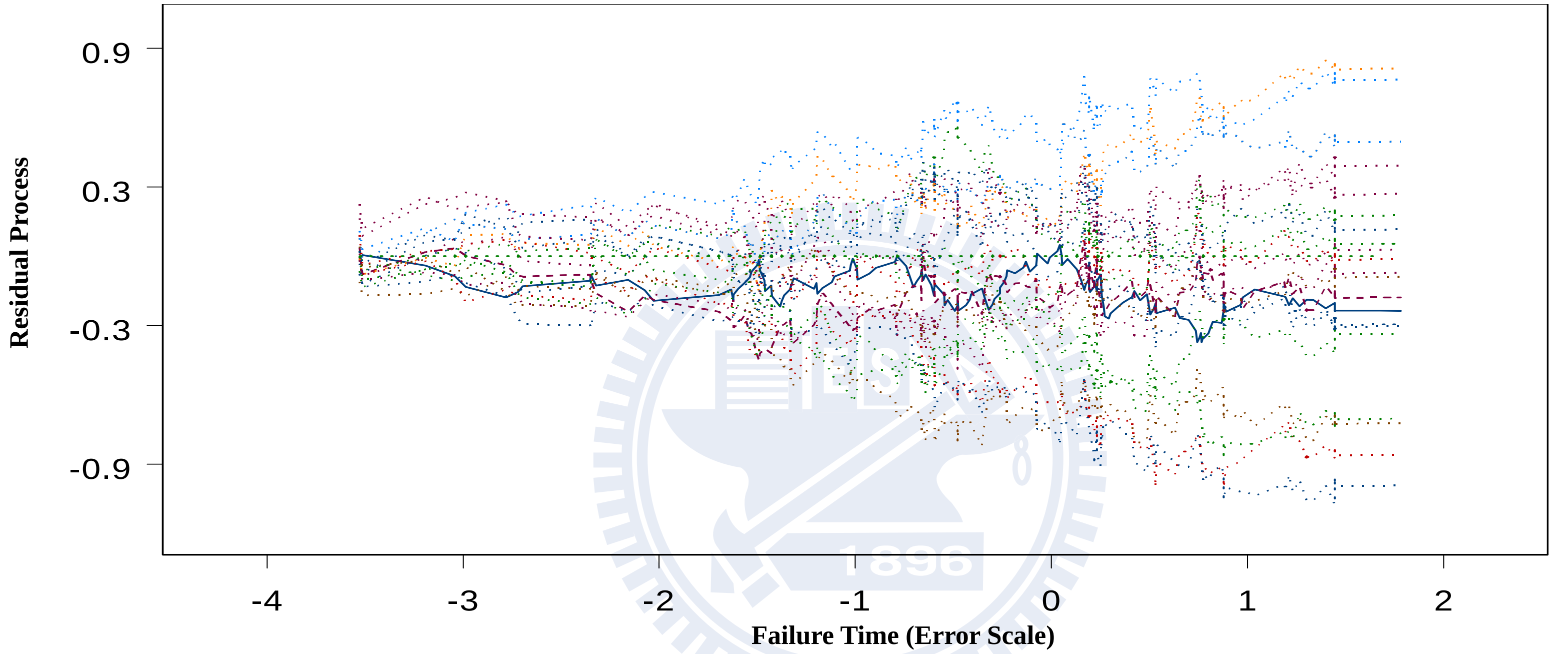

Figure 4.1a: Diagnostic plot of $\hat{V}(t)$ in (3.12) based on Z_1 when the true and imposed model are both AFT model.

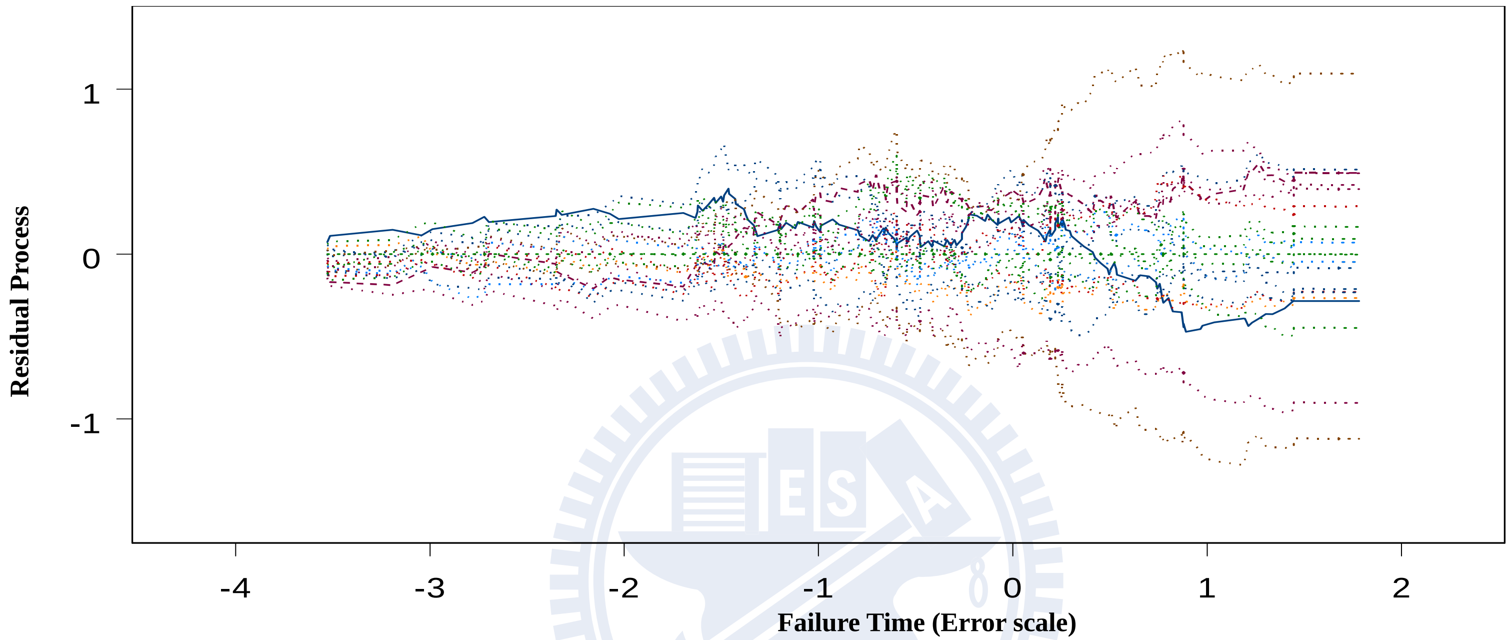

Figure 4.1b: Diagnostic plot of $\hat{V}(t)$ in (3.12) based on Z_2 when the true and imposed model are both AFT model.

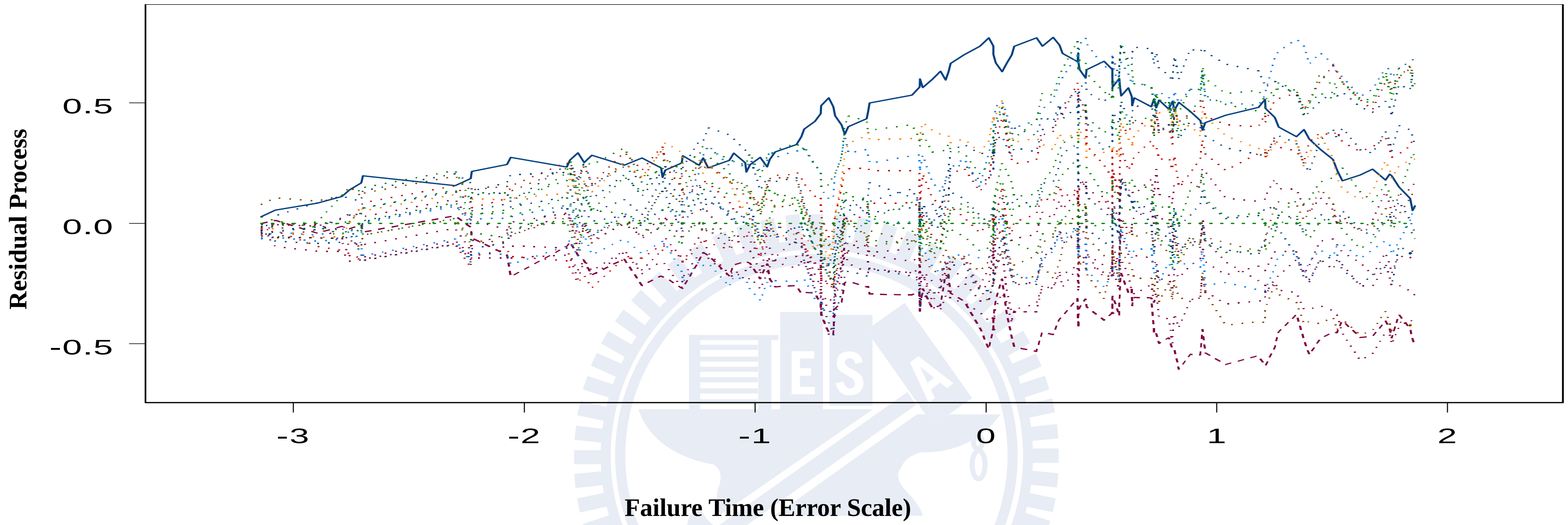

Figure 4.2a: Diagnostic plot of $\hat{V}(t)$ in (3.12) based on Z_1 when the true model is AFT model and imposed model is LS model

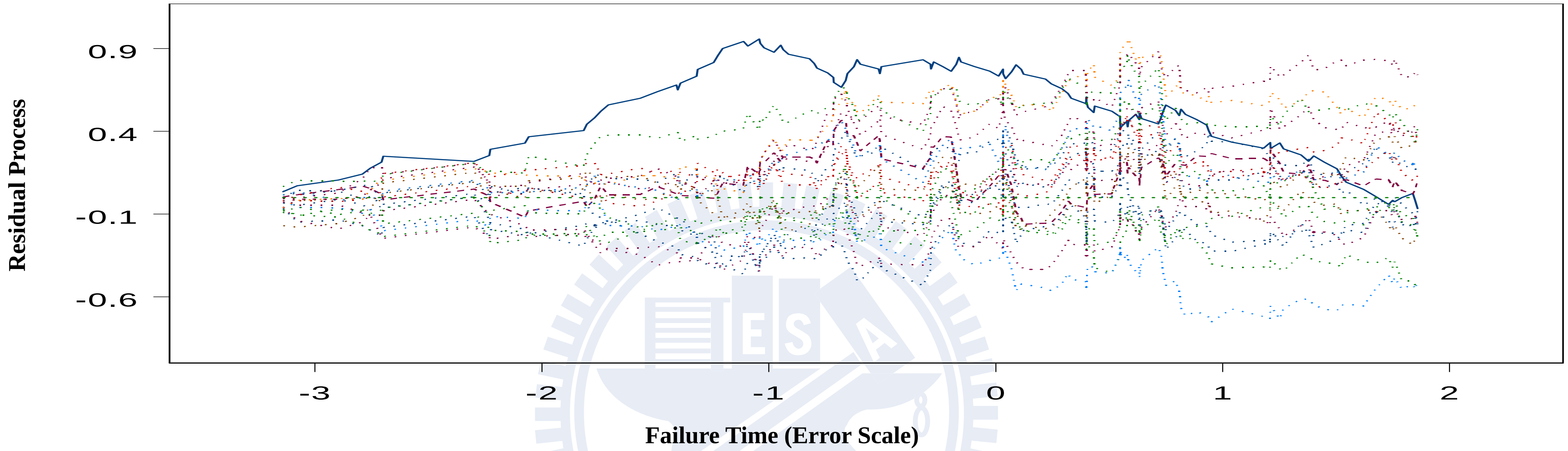

Figure 4.2b: Diagnostic plot of $\hat{V}(t)$ in (3.12) based on Z_2 when the true model is AFT model and imposed model is LS model

parameter		$\theta_0 = 2$		$\theta_1 = -1$		$p_0 = 0.88$		$p_1 = 0.73$	
censoring rate	Sample Size	BS($\hat{\theta}_0$)	σ($\hat{\theta}_0$)	BS($\hat{\theta}_1$)	σ($\hat{\theta}_1$)	BS($\hat{p}_0$)	σ($\hat{p}_0$)	BS($\hat{p}_1$)	σ($\hat{p}_1$)
0.25	200	0.3302	0.5248	-0.0033	0.6338	0.0228	0.0360	0.0521	0.0593
	500	0.1603	0.2393	0.0221	0.2656	0.0138	0.0216	0.0325	0.0353
0.40	200	0.3503	0.6282	0.0085	0.7573	0.0209	0.0457	0.0523	0.0770
	500	0.1861	0.4087	0.0267	0.4524	0.0126	0.0344	0.0345	0.0544

Table 3A: Finite-sample performances for estimating $\theta_0^{(0)}, \theta_0^{(1)}$, p_0 and p_1 in the incidence model based on 1000 replications

parameter		$\beta_1 = -1$		$\beta_2 = 1$		$R_{0.75} \equiv R(t_{0.75}) = 0.3333$		$R_{0.5} \equiv R(t_{0.5}) = 1$		$R_{0.25} \equiv R(t_{0.25}) = 3$	
censoring rate	Sample Size	BS($\hat{\beta}_1$)	σ($\hat{\beta}_1$)	BS($\hat{\beta}_2$)	σ($\hat{\beta}_2$)	BS($\hat{R}_{0.75}$)	σ($\hat{R}_{0.75}$)	BS($\hat{R}_{0.5}$)	σ($\hat{R}_{0.5}$)	BS($\hat{R}_{0.25}$)	σ($\hat{R}_{0.25}$)
0.25	200	-0.0612	0.3064	-0.0133	0.1725	-0.0054	0.0762	-0.0260	0.2163	-0.1572	0.6968
	500	-0.0756	0.1972	-0.0052	0.1209	0.0083	0.0489	0.0222	0.1329	-0.0510	0.4092
0.40	200	-0.0548	0.3369	-0.0179	0.1802	-0.0059	0.0814	-0.0276	0.2256	-0.1460	0.7714
	500	-0.0488	0.2149	-0.0223	0.1189	0.0014	0.0530	-0.0062	0.1472	-0.1029	0.5174

Table 3B: Finite-sample performances for estimating β_j in the latency model based on 1000 replications.

Note: $t_{0.75} = 0.5773$ such that $S(t_{0.75} \mid Z=0) = 0.75$ and $R(t_{0.75}) = 0.3333$; $t_{0.5} = 1$ such that $S(t_{0.5} \mid Z=0) = 0.50$ and $R(t_{0.5}) = 1$; $t_{0.25} = 1.732$ such that $S(t_{0.25} \mid Z=0) = 0.25$ and $R(t_{0.25}) = 3$

Parameter		$\theta_0 = 2$		$\theta_1 = -1$		$p_0 = 0.88$		$p_1 = 0.73$	
Kendall's tau	Setting	BS($\hat{\theta}_0$)	$\sigma(\hat{\theta}_0)$	BS($\hat{\theta}_1$)	$\sigma(\hat{\theta}_1)$	BS($\hat{p}_0$)	$\sigma(\hat{p}_0)$	BS($\hat{p}_1$)	$\sigma(\hat{p}_1)$
0.30	A	-0.1336	0.3747	-0.1465	0.6260	-0.0204	0.0409	-0.0692	0.1008
	B	-0.1369	0.2787	-0.0657	0.3646	-0.0184	0.0317	-0.0441	0.0519
0.50	A	-0.1017	0.3272	-0.1825	0.7167	-0.0154	0.0354	-0.0699	0.1138
	B	-0.1213	0.2682	-0.0650	0.3472	-0.0163	0.0305	-0.0404	0.0486

Table 4A: Finite-sample performances for estimating the incidence model based on n=300 and 1000 replications

Parameter		$\beta_1 = -0.7$		$\beta_2 = 0.3$		$R_{0.75} \equiv R_1(t_{0.75}) = 0.3333$		$R_{0.5} \equiv R_1(t_{0.5}) = 1$		$R_{0.25} \equiv R(t_{0.25}) = 3$	
Kendall's tau	Setting	BS($\hat{\beta}_1$)	$\sigma(\hat{\beta}_1)$	BS($\hat{\beta}_2$)	$\sigma(\hat{\beta}_2)$	BS($\hat{R}_{.75}$)	$\sigma(\hat{R}_{.75})$	BS($\hat{R}_{.5}$)	$\sigma(\hat{R}_{.5})$	BS($\hat{R}_{.25}$)	$\sigma(\hat{R}_{.25})$
0.30	A	0.0158	0.3761	0.0007	0.1615	0.0106	0.0769	0.0648	0.2197	0.3762	0.8367
	B	-0.0136	0.2824	-0.0080	0.1397	0.0111	0.0701	0.0596	0.2008	0.3435	0.7078
0.50	A	0.0217	0.4314	-0.0076	0.1691	0.0096	0.0754	0.0485	0.2040	0.2399	0.7073
	B	-0.0099	0.2690	-0.0070	0.1370	0.0110	0.0703	0.0517	0.1935	0.2752	0.6481

Table 4B: Finite-sample performances for estimating the latency model (T_1) based on n=300 and 1000 replications.

Note: $t_{0.75} = 0.5773$ such that $\tilde{S}_1(t_{0.75} \mid Z=0) = 0.75$ and $R_1(t_{0.75}) = 0.3333$; $t_{0.5} = 1$ such that $\tilde{S}_1(t_{0.5} \mid Z=0) = 0.50$ and $R_1(t_{0.5}) = 1$; $t_{0.25} = 1.732$ such that $\tilde{S}_1(t_{0.25} \mid Z=0) = 0.25$ and $R_1(t_{0.25}) = 3$

Parameter		$\beta_3 = 0.5$		$\beta_4 = 0.00$		$R_{0.75} \equiv R(t_{0.75}) = ct_{0.75}$		$R_{0.5} \equiv R(t_{0.5}) = ct_{0.5}$		$R_{0.25} \equiv R(t_{0.25}) = ct_{0.25}$	
Kendall's tau	Setting	Bias($\hat{\beta}_3$)	σ($\hat{\beta}_3$)	Bias($\hat{\beta}_4$)	σ($\hat{\beta}_4$)	Bias($\hat{R}_{0.75}$)	$\sigma(\hat{R}_{.75})$	Bias($\hat{R}_{0.5}$)	σ($\hat{R}_{0.5}$)	Bias($\hat{R}_{0.25}$)	σ($\hat{R}_{0.25}$)
0.30	A	0.0099	0.1854	-0.0062	0.0924	-0.0033	0.0468	-0.0230	0.1106	-0.0941	0.2267
	B	-0.0030	0.2298	0.0015	0.1095	0.0004	0.0309	-0.0041	0.0722	-0.0300	0.1394
0.50	A	-0.0179	0.1898	0.0048	0.0960	-0.0017	0.0504	-0.0336	0.1135	-0.1645	0.2237
	B	-0.0061	0.2483	0.0053	0.1164	-0.0002	0.0331	-0.0045	0.0825	-0.0420	0.1495

Table 4C: Finite-sample performances for estimating β_j and $\sigma(\hat{\beta}_j)$ in the survival model (T_2) based on n=300 and 1000 replications.

Note that:

Setting A: $t_{0.75} = 0.9589$ such that $S(t_{0.75} \mid Z = 0) = 0.75$ and $R(t_{0.75}) = 0.28767$; $t_{0.5} = 2.3105$ such that $S(t_{0.5} \mid Z = 0) = 0.5$ and $R(t_{0.5}) = 0.69315$; $t_{0.25} = 4.621$ such that $S(t_{0.25} \mid Z = 0) = 0.25$ and $R(t_{0.25}) = 1.3863$

Setting B: $t_{0.75} = 0.9589$ such that $S(t_{0.75} \mid Z = 0) = 0.87$ and $R(t_{0.75}) = 0.1438$; $t_{0.5} = 2.3105$ such that $S(t_{0.5} \mid Z = 0) = 0.7071$ and $R(t_{0.5}) = 0.3466$; $t_{0.25} = 4.621$ such that $S(t_{0.25} \mid Z = 0) = 0.50$ and $R(t_{0.25}) = 0.69315$

Setting		$\beta_1 = -0.7$		$\beta_2 = 0.3$		$R_{0.75} \equiv R(t_{0.75}) = 0.3333$		$R_{0.5} \equiv R(t_{0.5}) = 1$		$R_{0.25} \equiv R(t_{0.25}) = 3$	
$\Pr(\delta = 1)$	$\Pr(\delta = 0)$	BS($\hat{\beta}_1$)	$\sigma(\hat{\beta}_1)$	BS($\hat{\beta}_2$)	$\sigma(\hat{\beta}_2)$	BS($\hat{R}_{0.75}$)	$\sigma(\hat{R}_{0.75})$	BS($\hat{R}_{0.5}$)	$\sigma(\hat{R}_{0.5})$	BS($\hat{R}_{0.25}$)	$\sigma(\hat{R}_{0.25})$
0.50	0.50	0.0085	0.2227	0.0042	0.1335	0.0002	0.0581	0.0272	0.1544	-0.2781	0.3567
0.63	0.37	-0.0028	0.2051	-0.0015	0.1203	-0.0019	0.0558	-0.0107	0.1396	-0.0553	0.3341

Table 4E: Finite-sample performances for estimating β_j and $\sigma(\hat{\beta}_j)$ in the latency model based on n= 300 and 1000 replications.

Note: $t_{0.75} = 0.5773$ such that $S(t_{0.75} \mid Z = 0) = 0.75$ and $R(t_{0.75}) = 0.3333$; $t_{0.5} = 1$ such that $S(t_{0.5} \mid Z = 0) = 0.50$ and $R(t_{0.5}) = 1$; $t_{0.25} = 1.732$ such that $S(t_{0.25} \mid Z = 0) = 0.25$ and $R(t_{0.25}) = 3$

Setting		$\theta_0 = 2$		$\theta_1 = -1$		$p_0 = 0.88$		$p_1 = 0.73$	
$\Pr(\delta = 1)$	$\Pr(\delta = 0)$	BS($\hat{\theta}_0$)	$\sigma(\hat{\theta}_0)$	BS($\hat{\theta}_1$)	$\sigma(\hat{\theta}_1)$	BS($\hat{p}_0$)	$\sigma(\hat{p}_0)$	BS($\hat{p}_1$)	$\sigma(\hat{p}_1)$
0.5	0.5	0.0077	0.2889	-0.1275	0.3738	-0.0025	0.0302	-0.0256	0.0380
0.63	0.37	0.0462	0.2428	-0.0322	0.3014	0.0025	0.0244	0.0023	0.0299

Table 4D: Finite-sample performances for estimating $\theta_0^{(0)}, \theta_0^{(1)}, p_0$ and p_1 in the incidence model based on 300 sample size and 1000 replications