STATISTICAL MODELING AND ANALYSIS FOR SURVIVAL DATA WITH A CURE FRACTION

by

JIANFENG XU

A thesis submitted to the
Department of Mathematics and Statistics
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
January 2012

Copyright © Jianfeng Xu, 2012
Abstract

The analysis of survival data with a possible cure fraction has attracted much interest in the last two decades. Various models and estimating methods have been proposed for such data and they have been applied in many fields, especially in cancer clinical trials.

One important issue for analyzing survival data with a cure fraction is the estimation of the cure rate. Most current statistical procedures for the cure rate estimation are either parametric or nonparametric. The parametric methods usually impose a strong structure and hence they may not provide a good fit for some data. The nonparametric methods and the semi-parametric methods, on the other side, are robust and have low risk of model misspecification. However, the current nonparametric methods for the cure rate estimation cannot handle continuous covariates. We propose a novel nonparametric estimation of the cure rate, which can handle both discrete and continuous covariates. The cure rate estimate is proved to be consistent and asymptotically normally distributed. Numerical studies showed that the proposed method generally outperforms existing methods.

Recently, transformation cure models have been proposed for survival data with a cure fraction. These models usually include two or more basic models as their special cases. In the thesis, we propose a new transformation cure model, which includes two basic cure models as its special cases. We show that the maximum likelihood estimates from the method are consistent and asymptotically normally distributed.
Simulation studies are conducted to illustrate the proposed model in small samples. The proposed model is applied to a colon cancer data set.

It is usually desirable to have a unified model which can handle both survival data with and without a cure fraction when there is little knowledge about the presence of cured subjects. In this thesis, we propose such a model based on the Box-Cox transformation. The transformation parameter governs whether the model is for survival data with or without a cure fraction. Two estimating methods, one parametric and the other semi-parametric, are presented. Numerical studies showed that the proposed model and the estimation method work well for survival data both with and without a cure fraction.

We also considered in this thesis the analysis of length-biased data with a cure fraction. We present weighted estimating equations and derive the weighted maximum likelihood estimates for both the population survival function and the cure rate. Simulation studies show that the proposed estimating method works better than the ordinary method without any adjustments.

Finally, we consider an extension of a proportional density model for two-sample data with a cure fraction. An empirical likelihood method is proposed to construct unbiased estimating equations. The unbiasedness of estimates are supported by simulation studies.
Acknowledgements

First of all, I would like to express my sincere gratitude to my Ph.D supervisor, Prof Yingwei Peng. With his guidance and his effort, he opened the door for me to research in survival analysis with a cure fraction. Throughout my thesis-writing period, he provided kind encouragement and sound advice. This thesis is impossible without his enthusiasm and energy.

I would also like to thank Prof Dongsheng Tu, member of my Ph.D thesis committee. When I was taking a seminar class organized by Prof Peng and Prof Tu in 2007, he provided good suggestion on my research work and shared his idea with me, which is discussed in Chapter 6.

I want to thank the Department of Mathematics and Statistics, who provided the financial support during my Ph.D study at Queen’s, and the Department of Cancer Care and Epidemiology of Cancer Research Institute, who provided me the office to work in. Special thanks to my PhD committee members, Prof Kai Wang from the University of Iowa, Prof Judy-anne Chapman from the Department of community health and epidemiology, Prof Bingshu Chen and Prof Ole Nielsen from the Department of Mathematics and Statistics, who provided insightful comments which led to a great improvement in my thesis.

I wish to thank my family for providing me a loving environment. My wife were particularly supportive, especially in the thesis-writing period. Thanks to my daughter for tossing around to relax me from work.
Statement of Originality

I certify that this thesis, and the research to which it refers, are the product of my own work and that any ideas or quotations from the work of other people, published or unpublished, are fully acknowledged in accordance with the standard referencing practices of the discipline.
# Contents

Abstract

Acknowledgements

Statement of Originality

Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Survival Models

1.2 Cure Models

1.3 Motivation

1.4 Organization of Thesis

Chapter 2: Nonparametric Estimation of the Cure Rate

2.1 Introduction

2.2 Nonparametric Cure Rate Estimator

2.3 Asymptotic Property

2.4 Simulation Study

2.5 Age Effects in Leukemia Patients with Bone Marrow Transplantation

2.6 Discussion

Chapter 3: Transformation Model for Survival Data with a Cure Fraction

3.1 Introduction

3.2 A New Transformation Cure Model

3.3 Estimation and Asymptotic Property

3.4 Simulation Study

3.5 Treatment Effect Analysis for Colon Cancer Patients

3.6 Discussion
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Biases and variances of cure rate estimates with discrete covariate $z$ from the proposed nonparametric model and the existing semiparametric models</td>
<td>30</td>
</tr>
<tr>
<td>2.2</td>
<td>Biases and variances of cure rate estimates with continuous covariate $z$ from the proposed nonparametric model and the existing semiparametric models</td>
<td>31</td>
</tr>
<tr>
<td>2.3</td>
<td>Age effects on cure rates for leukemia patients.</td>
<td>34</td>
</tr>
<tr>
<td>3.1</td>
<td>Simulation results of $\beta$ from 1000 replications</td>
<td>51</td>
</tr>
<tr>
<td>3.2</td>
<td>Estimated parameters and confidence intervals</td>
<td>54</td>
</tr>
<tr>
<td>4.1</td>
<td>Estimations with Weibull model from 1000 replications</td>
<td>73</td>
</tr>
<tr>
<td>4.2</td>
<td>Simulation result of hypothesis testing $H_0 : r \leq 0$</td>
<td>74</td>
</tr>
<tr>
<td>4.3</td>
<td>Simulation results with semi-parametric method</td>
<td>75</td>
</tr>
<tr>
<td>4.4</td>
<td>Parameter estimations in colon cancer data</td>
<td>76</td>
</tr>
<tr>
<td>4.5</td>
<td>Parameter estimations in lung cancer data</td>
<td>78</td>
</tr>
<tr>
<td>5.1</td>
<td>Simulation results of length-biased data based on 1000 replications</td>
<td>89</td>
</tr>
<tr>
<td>6.1</td>
<td>Parameter estimations based on 2000 replications.</td>
<td>98</td>
</tr>
<tr>
<td>6.2</td>
<td>Bias of $\hat{\beta}$ of three models when $\beta = 1.5$</td>
<td>98</td>
</tr>
</tbody>
</table>
# List of Figures

2.1 Kaplan-Meier survival estimate of the leukemia patients. ..................... 33

2.2 Estimated cure rates versus patient’s age and their 95% confidence bands from the three models. ................................. 35

2.3 Kaplan-Meier survival estimates of the leukemia patients in 6 groups. .... 36

3.1 Relationship between the transformation cure models and the basic cure models. ................................................. 42

3.2 The observed likelihood function versus the transformation parameter. .... 53

3.3 Kaplan-Meier curves and predicted survival curves of colon cancer data. .......................................................... 55

4.1 The hazard functions with different combination of $r$ and $\alpha$ ............ 65

4.2 Kaplan-Meier curves and predicted survival curves of the colon cancer data ....................................................... 77

4.3 Kaplan-Meier curves and predicted survival curves of the lung cancer trial data ....................................................... 79
Chapter 1

Introduction

1.1 Survival Models

The problem of analyzing survival data, which is also called failure time data or time-to-event data, arises in many applied fields, such as medicine, public health, epidemiology, engineering and economics. A common feature of such data is that they usually contain censored observations. Censored data arise when the event for an individual is only known to occur in a certain time period. The most popular type of censoring is right censoring when the observed time is shorter than the actual failure time. There are other types of censoring, such as left censoring and interval censoring. Our focus in this thesis is on the case of right censoring.

Before introducing models to analyze survival data, we first give the notation of the survival data. Let $Y_i$ be the failure time of the $i$th subject. Let $C_i$ be the $i$th censoring time and $t_i = \min(Y_i, C_i)$ be the observed failure time. Let $\delta_i$ be the censoring indicator with $\delta_i = 0$ when $t_i = C_i$ and $\delta_i = 1$ when $t_i = Y_i$; $z_i$ and $x_i$ are the corresponding covariates of $t_i$. The observed data is

\[
D = \{(t_i, \delta_i, z_i, x_i), i = 1, \ldots, n\}.
\]  

(1.1)
In ordinary survival analysis, we only have one set of covariate. However, in survival analysis with a cure fraction, we may have two sets of covariates: $x$ for uncured part and $z$ for cure part. For a general definition, we allow two sets of covariates in our survival data.

In survival analysis, the most frequently used quantities are survival function $S(t)$, hazard function $h(t)$, cumulative hazard function $\Lambda(t)$ and probability density function $f(t)$. The basic quantity employed to describe the time-to-event phenomena is the survival function, which is defined as

$$S(t) = Pr(T > t)$$

and the other three quantities can be derived from the survival function,

$$h(t) = -d \log(S(t))/dt, \Lambda(t) = - \log(S(t)), f(t) = -dS(t)/dt.$$ 

Survival models are usually modeled through one of these quantities.

Many statistical models have been proposed to analyze survival data in the literature, including parametric models, semi-parametric model and nonparametric models. Parametric models are popular for researchers as they offer insight into the nature of parameters based on various distributions. Popular distributions in parametric modeling include Weibull, Gamma, Log-Normal, Log-Logistic, Gompertz and generalized Gamma distributions. Nonparametric models are also extensively used in literature because of their robustness. The most important nonparametric model is proposed by Kaplan and Meier (1958), who gave the so called Product-Limit estimator of the
survival function,

$$\hat{S}(t) = \begin{cases} 
1 & \text{if } t < t_{(1)} \\
\prod_{t_{(i)} < t} [1 - \frac{d_i}{R_i}] & \text{if } t_{(1)} \leq t,
\end{cases} \quad (1.2)$$

where $t_{(1)}, t_{(2)}, \cdots$ are the uncensored failure times sorted in increasing order; $d_i$ is the number of failures at time $t_{(i)}$; $R_i$ is the number of subjects who are still alive and uncensored at $t_{(i)-}$, where $t_{(i)-} = \lim_{\Delta t \to 0} t - \Delta t$ for positive $\Delta t$. The product-limit estimator was proved consistent and asymptotically normal and it has been used as a standard tool to check the validation of other parametric models or semi-parametric models. The product-limit estimator is also named the Kaplan-Meier estimator in honor of the authors. Semi-parametric models are other approaches to analyze survival data by assuming only partial parametric forms in the model. The most widely used semi-parametric model for survival data is the Proportional Hazards model due to Cox (1972, 1975),

$$h(t|z) = h_0(t) \exp(\beta'z) \quad (1.3)$$

where $h(t|z)$ is the population hazard function given covariate $z$; $h_0(t)$ is the baseline hazard function (i.e. the hazard function condition on $z = 0$) and $\exp(\beta'z)$ is the multiplicative term with covariate $z$. In model (1.3), $h_0(t)$ is usually unspecified. The proportional hazards model has been extensively studied and applied in survival analysis since its origination and it is even treated as the corner-stone of modern survival analysis.

As we see from the proportional hazards model, it is simply assumed that the covariate effects act multiplicatively on the hazard function. When the covariate effects do not act in the multiplicative pattern, we may consider the Additive Hazards
Regression model (Lin and Ying, 1997), where the covariate effects act in additive fashion on the hazard function,

\[ h(t|z) = h_0(t) + \beta'z. \tag{1.4} \]

The Accelerated Failure Time model (Wei, 1992) is another regression model for survival data, where the covariate effects act additively on the logarithm of the failure time as a linear regression model,

\[ \log T = \mu + \beta'z + \sigma W, \tag{1.5} \]

where \( \mu \) is the mean of the logarithm of the failure time \( T \); \( \beta'z \) is the linear additive term where the coefficients \( \beta \) has an interpretation similar to those in standard linear regression; \( \sigma \) is the variance of \( \log T \) and \( W \) is the error term with specified or unspecified distribution function. Popular modeling for \( W \) is the extreme value distribution and logistic distribution. Through some algebra, the accelerated failure time model could be written as

\[ S(t|z) = S_0(\exp(-\beta'z)t), \tag{1.6} \]

where \( S(t|z) \) is the population survival function given covariate \( z \) and \( S_0(t) \) is the baseline survival function, the population survival function conditional on \( z = 0 \). From model (1.6), we see that the time is either accelerated by a constant factor or degraded by the constant factor \( \exp(-\beta'z) \) depending on the sign of \( \beta'z \) and therefore model (1.5) or model (1.6) is called the accelerated failure time model and factor \( \exp(-\beta'z) \) is called the acceleration factor. Another popular model in literature is
the Proportional Odds model (Bennett, 1983), where the covariate effects act multiplicatively on the odds ratio of the survival function,

\[
\frac{S(t|z)}{1 - S(t|z)} = \exp(\beta' z) \frac{S_0(t)}{1 - S_0(t)}.
\] (1.7)

Semi-parametric Transformation models have attracted much interest in the last two decades. In the transformation model, a family of transformation functions is imposed on the failure time, the hazard function or the survival function. The transformation function is usually parametrically specified. When the transformation function changes within the family, the transformation model generates a class of survival models, including some well-known survival models as its special cases. There are many examples of transformation survival models. For example, Ciampi et al. (1989) used the Box-Cox transformation for survival data generation. However, transformation models considered here are imposed on survival function, hazard function or cumulative hazard function in stead of failure time. To our knowledge, the first transformation model imposed on survival function was proposed by Cheng et al. (1995),

\[
g(S(t|z)) = h_0(t) + \beta' z,
\] (1.8)

where \(g(\cdot)\) is the transformation function; \(S(t|z)\) is the population survival function and \(h_0(t)\) is the baseline hazard function. When \(g(\cdot) = \log(- \log(\cdot))\), the model (1.8) becomes the Cox proportional hazards model; when \(g(\cdot) = -\logit(\cdot)\), it becomes the proportional odds model. Hence model (1.8) includes proportional hazards model and proportional odds model as its special cases. Since the origination of model (1.8), the transformation models have been widely studied by many researchers. A nice summary of the transformation models could be found in Zeng and Lin (2007),
where the authors presented several classes of semi-parametric transformation models as well as corresponding estimating methods.

1.2 Cure Models

In the analysis of survival data, an implicit assumption is that all subjects will eventually experience the event if the follow-up time is sufficiently long enough. However, it may not be true in many cases. In some cancer research studies, for example, there may be a certain percentage of patients who respond favorably to treatment and appear to be risk free after a sufficient follow-up time. We call such patients cured of the disease. In other words, only a proportion of subjects or patients from the population are susceptible to the event of interest and other subjects or patients are not susceptible to the event. The proportion of the cured subjects in the population is called the cure fraction or cure rate. A long and stable plateau at the tail of the Kaplan-Meier survival curves is an indication of the existence of cure fraction. Formal test of the existence of the cure fraction was given by Maller and Zhou (1994).

Survival data where cure is possible, or present, is called survival data with a cure fraction and survival data where cure is not present is called survival data without a cure fraction.

The model for the survival data with a cure fraction is called the cure rate model or simply cure model and the model to analyze the survival data without a cure fraction is referred to as non-cure model in sequel. Before introducing the cure models, we first give the notation of the survival data with a cure fraction. Survival data with a cure fraction is similar to the survival data without a cure fraction as defined in (1.1) except that the true “failure” time $Y$ of the subjects or patients can be $\infty$. However, the case where $Y = \infty$ is not observable because of the existence of right censoring and therefore the survival data with a cure fraction looks identical to the survival
data without a cure fraction (1.1).

In the cure model, the survival function \( S(t) \), the hazard function \( h(t) \), the cumulative hazard function \( \Lambda(t) \) or the probability density function \( f(t) \) are improper in the sense that \( \lim_{t \to \infty} S(t) > 0 \), \( \lim_{t \to \infty} \Lambda(t) = \lim_{t \to \infty} \int_0^t h(t) dt < \infty \) and \( \int_0^\infty f(t) dt < 1 \). Other properties of these quantities remain the same as the one in non-cure models. For example, \( S(t) \) of cure model is still monotone decreasing with \( S(0) = 1 \); \( h(t) \) is always non-negative and \( \Lambda(t) = \int_0^t h(t) dt \) is monotone non-decreasing; \( f(t) \) is always non-negative. By distinctions, counterparts of non-cure models are referred to as proper.

The most popular cure model is the \textit{Mixture Cure} model, which has the form

\[
S(t|z, x) = \pi(z) + (1 - \pi(z))S_0(t|x)
\]  

(1.9)

where \( S(t|z, x) \) is the improper population survival function; \( \pi(z) \) is the cure rate; \( S_0(t|x) \) is the proper survival function for the uncured patients; \( z \) is the covariate associated with \( \pi(z) \) and \( x \) is the covariate associated with \( S_0(t|x) \), where \( z \) and \( x \) may share some common elements.

In the mixture cure model, the cure rate \( \pi(z) \) is usually specified by a parametric logistic form, i.e. \( \log\left[\frac{\pi(z)}{1 - \pi(z)}\right] = \beta'z \), or a nonparametric form, i.e. a constant, when covariate \( z \) is suppressed. The survival function of the uncured subjects, \( S_0(t|x) \), can be parametrically specified or semi-parametrically specified. For parametric inference, Boag (1949) studied the case when \( S_0(t|x) \) is modeled by a log-normal distribution; Farewell (1982) studied the model when \( S_0(t|x) \) is modeled by Weibull distribution; Denham et al. (1996) studied another case with log-normal assumed \( S_0(t|x) \); Peng et al. (1998) proposed a generalized \( F \) distribution to model \( S_0(t|x) \). Many others could be found in literature. When \( S_0(t) \) is not fully specified, Kuk and Chen (1992); Peng and Dear (2000) and Sy and Taylor (2000) studied the mixture cure model by
assuming a proportional hazards model (1.3) for $S_0(t|x)$ with unspecified baseline hazard function; Li and Taylor (2002) and Zhang and Peng (2007) studied the mixture cure model when $S_0(t|x)$ is modeled by an accelerated failure time model with unspecified error distribution.

Another important cure model is the Proportional Hazards Cure model, which is given as

$$S(t|z, x) = \exp(-\theta(z)F_0(t|x))$$

where $S(t|z, x)$ is the improper population survival function; $F_0(t|x)$ is a proper baseline cumulative distribution function; $\theta(z)$ is a positive function of $z$ and is usually formulated as $\exp(\beta'z)$. The cure rate of this model is $S(\infty|z, x) = \exp(-\theta(z))$.

The proportional hazards cure model was first proposed by Yakovlev et al. (1993) and then studied by various researchers. Chen et al. (1999) studied the model in Bayesian frame. Tsodikov (1998b, 2001) studied the model by assuming that $F_0(t)$ is non-parametrically specified. Tsodikov (2003); Tsodikov and Garibotti (2007) studied the model again by assuming a Cox proportional hazards model for $F_0(t)$ with unspecified baseline hazard function; Tsodikov (1998a) studied this model when the covariate $z$ is time-dependent. Chen et al. (2002) studied the application of this model on multivariate survival data; many other examples can be found in the literature.

The advantage of the proportional hazards cure model (1.10) over the mixture cure model (1.9) is that model (1.10) has a proportional hazards structure, which is preferable for Bayesian inference, and a biological interpretation, which is explained as follows. For a given subject, let $N$ be the number of tumor cells that is capable of metastasizing after treatment and $(X_1, \ldots, X_N)$ be the failure times of each cell. Then the failure time of the subject is $\min(X_1, \ldots, X_N)$. Usually, $N$ is assumed to
follow a Poisson distribution with mean $\theta$ and $(X_1, \cdots, X_N)$ are assumed to be independent and identically distributed with a common cumulative distribution function $F(t)$ independent of $N$. Then the survival function is

$$S(t) = Pr(N = 0) + Pr(N > 1) \times Pr(X_1 > t, \cdots, X_N > t | N > 1)$$

$$= \exp(-\theta) + \sum_{j=1}^{\infty} \frac{(1 - F(t))^j \theta^j \exp(-\theta)}{j!} = \exp(-\theta F(t))$$

Because of its biologic derivation, model (1.10) is also called the *Promotion Time Cure* model in literature.

The third cure model, although less popular than the previous two in literature, is the *Proportional Odds Cure* model, which is defined analogous to the proportional odds model for survival data without a cure fraction. The survival function of the model is given as

$$S(t|z, x) = \frac{1}{1 + \exp(\beta^T z) F_0(t|x)} \quad (1.11)$$

where $S(t|z, x)$ is the improper population survival function and $F_0(t|x)$ is a proper cumulative distribution function. The cure rate of the proportional odds cure model is $S(\infty|z, x) = 1/(1 + \exp(\beta^T z))$. Tsodikov (2003) studied the model when $F_0(t)$ is non-parametrically specified; Zeng et al. (2006) studied the model as a special case of their transformation cure model; Gu et al. (2011) provided a biological derivation of model (1.11) and studied the model parametrically.

In addition to the three basic cure models, model (1.9), (1.10) and (1.11), transformation cure models have been proposed in the literature. Lu and Ying (2004)
proposed a transformation cure model in mixture cure pattern,

\[
\begin{aligned}
S(t|z) &= \pi(z) + (1 - \pi(z))S_0(t|z) \\
S_0(t|z) &= \exp[-\Lambda(H_0(t) + \beta'z)]
\end{aligned}
\]  \hspace{1cm} (1.12)

where \(S(t|z)\) is the improper population survival function; \(\pi(z)\) is the cure rate; \(S_0(t|z)\) is the improper survival function of the uncured subjects, which follows a similar transformation model as model (1.8); \(H_0(t)\) is an unspecified monotone increasing function; \(\Lambda(x)\) is a specified transformation function. When \(\Lambda(x) = \exp(x)\), \(S_0(t|z)\) follows the Cox proportional hazards model (1.3) with unspecified hazard function; when \(\Lambda(x) = \log(\exp(x)/(1 + \exp(x)))\), \(S_0(t|z)\) follows the proportional odds model (1.7).

Similar to model (1.12), Mao and Wang (2010) proposed another transformation cure model by imposing a parametric transformation model for \(S_0(t|z)\),

\[
\begin{aligned}
S(t|z) &= \pi(z) + (1 - \pi(z))S_0(t|z) \\
S_0(t|z) &= \frac{\exp(-\beta'z/\rho)}{[\exp(-\beta'z) + \rho H_0(t)]^{1/\rho}}
\end{aligned}
\]  \hspace{1cm} (1.13)

where \(H_0(t)\) is a proper, but unspecified, baseline cumulative hazard function. The proper survival function for uncured patients \(S_0(t|z)\) follows a transformation model, which is called the generalized proportional odds model proposed by Dabrowska and Doksum (1988). When the transformation parameter \(\rho = 1\) in model (1.13), \(S_0(t|z)\) follows the proportional odds model (1.7). When \(\rho \to 0\), \(S_0(t|z)\) follows the proportional hazards model (1.3).

It is seen that both model (1.12) and model (1.13) are based on the mixture cure model (1.9) and the transformation model is applied only on the proper baseline survival function. Transformation cure models which are not based on the mixture cure
model (1.9) were proposed in the literature too. Yin and Ibrahim (2005a) proposed a transformation cure model, by imposing the Box-Cox Transformation model on the survival function,

$$S(t|z) = \left\{ 1 - \frac{a \exp(\beta'z)}{1 + a \exp(\beta'z)} F_0(t) \right\}^{1/a} \quad (1.14)$$

where $S(t|z)$ is the improper population survival function and $F_0(t)$ is a proper cumulative distribution function; $a$ is the transformation parameter. The cure rate of this model is $S(\infty|z) = (1 + a \exp(\beta'z))^{-1/a}$. When $a = 1$, the model (1.14) becomes a mixture cure model

$$S(t|z) = \frac{1}{1 + \exp(\beta'z)} + \frac{\exp(\beta'z)}{1 + \exp(\beta'z)} S_0(t),$$

where $S_0(t) = 1 - F_0(t)$; when $a \to 0$, the model (1.14) becomes a proportional hazards cure model

$$S(t|z) = \exp(-\exp(\beta'z) F_0(t));$$

when $0 < a < 1$, the model (1.14) becomes an intermediate model between the mixture cure model and the proportional hazards cure model. Model (1.14) is still a valid cure model when $a > 1$.

Zeng et al. (2006) proposed another transformation cure model, which does not follow the mixture cure pattern,

$$S(t|z) = [1 + a \theta(z) F_0(t)]^{-1/a} \quad (1.15)$$

where $S(t|z)$ is the improper population survival function and $F_0(t)$ is a proper cumulative distribution function; $a$ is the transformation parameter and the cure rate
is $S(\infty|z) = [1 + a\theta(z)]^{-1/a}$. When $a = 1$, it results in the proportional odds cure model

$$
S(t|z) = \frac{1}{1 + \theta(z)F_0(t)}.
$$

When $a \to 0$, it results in the proportional hazards cure model

$$
S(t|z) = \exp(\theta(z)F_0(t)).
$$

When $0 < a < 1$, the model (1.15) produces a new intermediate cure model between the proportional odds cure model and the proportional hazards cure model; when $a > 1$, model (1.15) is still a valid cure model.

### 1.3 Motivation

In the analysis of survival data with a cure fraction, it is important to estimate the cure rate. However, the cure rate and covariate effects on the cure rate are primarily estimated parametrically in most existing methods. That is, the covariate effects on cure rate are modeled either by $\pi(z) = 1/[1 + \exp(\beta'z)]$ (the mixture cure model), by $\pi(z) = \exp(-\exp(\beta'z))$ (the proportional hazards cure model) or by $\pi(z) = 1/(1 + \exp(\beta'z))$ (the proportional odds cure model); these also apply for the transformation cure models. When the underlying covariate effects on the cure rate cannot be well approximated by one of the three parametric models (1.9), (1.10) and (1.11), using these cure models will lead to biased estimates of covariate effects on cure rate. Therefore a nonparametric method to estimate the cure rate and to assess the effects of covariates on cure rate is important in the analysis of survival data with a cure fraction.
Nonparametric estimating methods of cure rate without covariates have been pro-
posed in the literature. Maller and Zhou (1992) proposed a consistent nonpara-
metric estimator of the cure rate based on the value of Kaplan-Meier survival estimator at
the largest observed failure time. They proved that the estimator is consistent and
asymptotically normal. However, their method cannot handle covariates. Laska and
Meisner (1992) proposed another nonparametric estimator of cure rate as a nonpara-
metric maximum likelihood estimator. However, their method cannot handle general
covariates either. It motivates us to consider a nonparametric method to estimate
cure rate with both discrete and continuous covariate effects. The work on the non-
parametric estimator of cure rate with covariates will be discussed in Chapter 2.

Besides the estimation of the cure rate, we consider a new transformation cure
model to analyze survival data with a cure fraction. As introduced in the previous
section, both model (1.12) and model (1.13) restrained the models in mixture cure
pattern and the transformation technique is only applied on the uncured survival
function \( S_0(t|x) \), which are just straightforward extensions from the transformation
models of survival data without a cure fraction and will limit the application of their
models in the analysis of survival data with a cure fraction. For example, when it
is believed that the given dataset does not have a mixture cure pattern, both model
(1.12) and model (1.13) are inappropriate. In such cases, model (1.14) or model
(1.15) will be more suitable since both models are not confined to a fixed mixture
cure structure. However, Yin and Ibrahim’s model (1.14) only combines the mixture
cure model (1.9) and the proportional hazards cure model (1.10); Zeng et al.’s model
(1.15) combines the proportional hazards cure model (1.10) and the proportional odds
cure model (1.11). There are no transformation models which combine the mixture
cure model (1.9) and the proportional odds cure model (1.11). In this thesis, we
propose a new transformation cure model which includes the proportional odds cure
model and the mixture cure model as its special cases. This new model is presented in Chapter 3.

When we apply a model, an important issue is to check if the model is appropriate for the data. Nevertheless, such model checking is often overlooked in survival analysis. For instance, we may fit a transformation non-cure model if we believe that the data does not have a cure fraction or we fit a transformation cure model if we believe that the data does have a cure fraction. However, in reality, information required to determine the existence of the cure fraction is often limited or unavailable. It seems preferable to have a model that relies less on the subjective specification and more on the data itself. Particularly, it is desirable to have a transformation model which could handle survival data both with a cure fraction and without a cure fraction, which motivates us to consider a transformation model which includes both cure models and non-cure models as its special cases. This work is introduced in Chapter 4.

Other statistical techniques, which can be used for survival data without a cure fraction, can be modified to fit survival data with a cure fraction. In real applications, one may encounter biased samples. One important biased sample is the length-biased sample. We consider an unbiased estimating method of survival data with a cure fraction in length-biased sample. This work is given in Chapter 5. In clinical trials, it is very common to have two-sample data. A proportional density model is an important approach to analyze such data. We propose a transformation model based on the proportional density structure to analyze the two-sample survival data with a cure fraction. In our model, the empirical likelihood technique is applied. This work is given in Chapter 6.
1.4 Organization of Thesis

Nonparametric estimation of the cure rate with covariate effect is presented in Chapter 2. The transformation cure model which includes both mixture cure model and proportional odds cure model as its special cases is proposed in Chapter 3 and the transformation model which includes both cure and non-cure model as its special cases is given in Chapter 4. A length-biased data analysis is presented in Chapter 5. A new transformation model to analyze the two-sample data is given in Chapter 6. In Chapter 7, we conclude the thesis with a summary.
Chapter 2

Nonparametric Estimation of the Cure Rate

2.1 Introduction

Survival data with a possible cure fraction is very common in clinical and epidemiological studies. In these studies, a group of subjects with censored survival times may be immune to the event and can be considered cured. The main focus of analysing such data is to estimate the proportion of cured subjects or cure rate and to examine the effects of covariates on the cure rate. This is, however, not a trivial task because uncured subjects may have censored survival times as well as cured subjects, who are observed only at the censored time.

To estimate the cure rate and to model the effects of covariate $z$ on the cure rate, we may consider the application of three basic cure models, model (1.9), model (1.10) and model (1.11). It is worth noting that the existing semi-parametric estimation methods for these three models are mainly for the failure time distribution part of the model. The cure rate and covariate effects on the cure rate are primarily estimated parametrically. That is, the covariate effects on cure rate is modeled by $\pi(z) = 1/[1 + \exp(\beta^t z)]$ (the mixture cure model (1.9)), $\pi(z) = \exp(-\exp(\beta^t z))$ (the
proportional hazards cure model (1.10)) or \( \pi(z) = 1/[\exp(\beta' z) + 1] \) (the proportional odds cure model (1.11)). When the underlying covariate effects on the cure rate cannot be well approximated by one of the three parametric models, using these cure models will lead to biased estimates of covariate effects on cure rate. Therefore a nonparametric method to estimate the cure rate and the effects of covariates on cure rate will be desirable in modeling survival data with a cure fraction when the effects of covariates on cure rate may not be adequately modeled by the parametric models.

In this chapter, we propose a novel nonparametric estimation method and investigate the properties of the proposed method. The nonparametric estimator of cure rate is introduced in section 2.2. Large sample properties are proved in section 2.3. In section 2.4, we investigate the small sample properties through simulation studies. Analysis of leukemia data with bone marrow transplantation is given in section 2.5. In section 2.6, we give a short discussion.

2.2 Nonparametric Cure Rate Estimator

First, we introduce the survival data to be used in this chapter. For simplicity of presentation, we present the proposed method with single covariate. The multivariate case is an easy extension. The dataset we use in this chapter is given as

\[ D = \{(Y_i, \delta_i, z_i), i = 1, \cdots, n\}. \]  

Compared with the definition in (1.1), \( x \) is suppressed as it only affects the survival function of the uncured patients.

Nonparametric estimation methods of cure rate without covariates have been proposed in the literature. Maller and Zhou (1992) proposed a nonparametric estimator of the cure rate based on the value of Kaplan-Meier survival estimator (1.2) at the
largest uncensored failure time,

\[ \hat{\pi} = \hat{S}(T_{(n)}) \]  

(2.2)

where \( \hat{S}(\cdot) \) is the Kaplan-Meier estimate of the population survival function and \( T_{(n)} \) is the largest uncensored failure time. They proved that the estimator is consistent and asymptotically normal. We can see that there is no covariate included in the cure rate estimation. When a discrete covariate, such as treatment, is considered, the whole population must be divided into several groups depending on the discrete covariate and the cure rates are estimated in different groups respectively. However, this method cannot handle continuous covariates.

Laska and Meisner (1992) proposed another nonparametric estimator of the cure rate. Suppose that there are \( n + m \) subjects in the dataset and \( m \) of them are observed cured, where observed cure status is obtained by defining a cure threshold. Then Laska and Meisner (1992) proposed the generalized maximum likelihood estimation of the population survival function and the cure rate, which are given as follows,

\[ \hat{S}(t) = \prod_{t_{(i)} \leq t} \left( 1 - \frac{\delta_i}{n + m - i + \delta_i} \right) \]

and

\[ \hat{\pi} = \prod_{i=1}^{n} \left( 1 - \frac{\delta_i}{n + m - i + \delta_i} \right). \]

where \( t_{(1)} < t_{(2)} < \cdots < t_{(n)} \) are the ordered failure times. It is easy to see that when \( m = 0 \), the survival function estimation reduces to the Kaplan-Meier estimation and the cure rate estimation reduces to the estimate of Maller and Zhou (1992) of equation (2.2). To incorporate the discrete covariate, Laska and Meisner (1992) used the same
approach as one used in Maller and Zhou (1992): divide the whole population into groups depending on the discrete covariate and then estimate the cure rates in each group separately. This method cannot handle continuous covariates either.

Tsodikov (2001) proposed another estimation of the cure rate in the proportional hazards cure pattern (1.10) where \( F_0(t) \) is estimated by a step function with jumps only at uncensored failure times. After getting the estimations of \( \theta \), where covariate \( z \) is suppressed, the cure rate was estimated by,

\[
\hat{\pi} = \exp(-\hat{\theta})
\]

However, Tsodikov’s method can only handle discrete covariate in the same way used by Maller and Zhou (1992) and Laska and Meisner (1992); it cannot handle continuous covariates.

In this section, we propose a new nonparametric method to estimate the cure rate based on the generalized Kaplan-Meier estimator and both discrete and continuous covariate can be incorporated. As a generalization of the prominent Kaplan-Meier survival estimator, the generalized Kaplan-Meier estimator (Beran, 1981) has been widely used to estimate the survival function when a covariate effect on the survival time must be taken into account. For survival data \((Y_i, \delta_i, z_i), i = 1, \ldots, n\) without a cure fraction, the generalized Kaplan-Meier estimator is given by

\[
\hat{S}(t|z) = \begin{cases} 
\prod_{j=1}^{n} \left[ \frac{1 - \sum_{r=1}^{n} I(Y_r \leq Y_j) B_r(z)}{1 - \sum_{r=1}^{n} I(Y_r < Y_j) B_r(z)} \right] I(Y_j \leq t; \delta_j = 1) & \text{if } t \leq Y(n) \\
0 & \text{if } t > Y(n). \end{cases}
\]

(2.3)

where \( I(A) \) is the indicator function with \( I(A) = 1 \) if \( A \) is true and \( I(A) = 0 \) otherwise, \( B_j(z) \) is a proper weight function satisfying \( B_j(z) \geq 0 \) and \( \sum_{j=1}^{n} B_j(z) = 1 \), and \( Y(n) \) denotes the largest failure time. When the covariate \( z \) is compressed, model (2.3)
reduces to the Kaplan-Meier model (1.2).

For survival data (2.1) with a cure fraction, we propose to estimate $\pi(z)$ by

$$\hat{\pi}(z) = \hat{S}(Y_{(n)}^1|z)$$

(2.4)

where $\hat{S}(y|z)$ is the generalized Kaplan-Meier estimate with data (2.1) as if there is no cure fraction. $Y_{(n)}^1$ is the largest uncensored failure time. Two weight functions for $B(z)$ are considered. The first one is the Nadaraya-Watson weight function,

$$B_j(z) = \frac{K\left(\frac{z_j-z}{h}\right)}{\sum_{r=1}^{n} K\left(\frac{z_r-z}{h}\right)}, \quad j = 1, 2, \ldots, n$$

(2.5)

where $K(\cdot)$ is a proper kernel function. The second one is the Gasser-Müller weight function

$$B_j(z) = h^{-1} \int_{s_{j-1}}^{s_j} K\left(\frac{z-u}{h}\right)du, \quad j = 1, 2, \ldots, n.$$  

(2.6)

where $s_0 = 0, s_j = \frac{1}{2}(z_j + z_{j+1}), j = 1, \ldots, n - 1$ and $s_n = 1$. However, the Gasser-Müller weight function should only be used for a continuous covariate $z$.

Note that if there is no covariate, $B_r(z) = 1/n, r = 1, \ldots, n$ in (2.5), the proposed estimator reduces to the Kaplan-Meier estimator at $Y_{(n)}^1$. Hence the proposed estimator (2.4) reduces to the cure rate estimator of no-covariate case by Maller and Zhou (1992). Therefore the proposed estimator is a generalization of the one sample cure rate estimator to a sample with a covariate.
2.3 Asymptotic Property

The consistency and the asymptotic normality of the generalized Kaplan-Meier estimator (2.3) for survival data without a cure fraction was established by a few researchers. Beran (1981) proposed the generalized Kaplan-Meier estimator. Dabrowska (1987, 1989, 1992) studied the asymptotic property of the generalized Kaplan-Meier estimator with Nadaraya-Watson weight. Gonzalez-Manteiga and Cadarso-Suarez (1994) and Van Keilegom and Veraverbeke (1997) studied the asymptotic property of the Kaplan-Meier estimator with Gasser-Müller weight. In this section, we will show that the proposed estimator (2.4) is a consistent estimator of the true cure rate \( \pi(z) \) for any interior point \( z \) and it is asymptotically normal.

Let \( S_u(t|z) \) be the survival function conditional on \( T < \infty \), i.e. \( S_u(t|z) = Pr(T > t|T < \infty) \). Let \( H(t|z) \) be the survival function of \( Y = \min(T, C) \) and \( G(t|z) \) be the survival function of the censoring time \( C \). We assume that \( G(t|z) \) is a proper survival function, and \( T \) is independent with \( C \) given \( z \). Let \( \tau_H(z) = \inf_t \{t : H(t|z) = 0\} \), \( \tau_{S_u}(z) = \inf_t \{t : S_u(t|z) = 0\} \) and \( \tau_G(z) = \inf_t \{t : G(t|z) = 0\} \). Since \( S(t|z) \) is an improper survival function and \( H(t|z) = S(t|z) G(t|z) \), we have \( \tau_H(z) = \tau_G(z) \). Let \( \tau_u = \sup \{z \in D \} \tau_{S_u}(z) \), where \( D = [0,1] \) is the support of \( z \), and let \( \tau_u < \tau_G(z) \) for any \( z \), which guarantees that with probability 1, censored subjects beyond the largest uncensored failure time are cured. Similar assumption, \( \tau_u < \tau_G \), can be found in Maller and Zhou (1992, 1994). We further assume

(H1) The functions \( H(\cdot|z_i), S_u(\cdot|z_i) \) and \( G(\cdot|z_i), i = 1, 2, \ldots, n \), belong to the families \( \{H(\cdot|z) : z \in D\}, \{S_u(\cdot|z) : z \in D\} \) and \( \{G(\cdot|z) : z \in D\} \) respectively, where \( D \) is a compact set, and \( H, S_u \) and \( G \) have continuous first-order derivatives with respect to \( t \) for a given value of \( z \) and bounded second-order derivatives with respect to \( z \) for a given value of \( t \).
(H2) The weight function $K(\cdot) \geq 0$ has a compact support and is Lipschitz-continuous of order 1 with $\int K(u)du = 1$, $\int uK(u)du = 0$, $\int K^2(u)du < \infty$ and $\int u^2K(u)du < \infty$.

(H3) When $z$ is continuous, $z_1, \ldots, z_n$ satisfy $\max_i [z(i) - z(i-1)] = O(n^{-1})$, where $z(1) \leq \cdots \leq z(n)$ are sorted in increasing order. For the Gasser-Müller weight function, we additionally require $\max_i [s_i - s_{i-1}] = O(n^{-1})$.

Similar assumptions could be found in Gonzalez-Manteiga and Cadarso-Suarez (1994) and Van Keilegom and Veraverbeke (1997).

For simple notation, we use $p \to$ to denote convergence in probability and use $d \to$ to denote convergence in distribution.

To establish the consistency of $\hat{\pi}(z)$, we first demonstrate that the estimator $\hat{\pi}(z)$ can be written as a function of the estimators of $H(t|z)$ and $G(t|z)$. Since $(Y_i, z_i), i = 1, \ldots, n$ are random sample without censoring from $H(t|z)$, and $(c_i, 1 - \delta_i, z_i), i = 1, \ldots, n$ are random sample with censoring from $G(t|z)$, Beran (1981); Dabrowska (1987); Gonzalez-Manteiga and Cadarso-Suarez (1994) and Van Keilegom and Veraverbeke (1997) showed, with Nadaraya-Watson weight or Gasser-Müller weight, that $H(t|z)$ and $G(t|z)$ can be consistently estimated, respectively, by

$$\hat{H}(t|z) = 1 - \sum_{j=1}^{n} I(Y_j \leq t)B_j(z)$$

and

$$\hat{G}(t|z) = \begin{cases} \Pi_{j=1}^{n} \left[ \frac{1-\sum_{r=1}^{n} I(Y_r \leq Y_j)B_r(z)}{1-\sum_{r=1}^{n} I(Y_r < Y_j)B_r(z)} \right] I(Y_j \leq t; \delta_j = 0) \quad & \text{if } t \leq Y(n) \\ 0 \quad & \text{if } t > Y(n). \end{cases}$$

The following lemma shows that the nonparametric cure rate estimator (2.3) can be written as a function of $\hat{H}(t|z)$ and $\hat{G}(t|z)$:
Lemma 2.3.1. For any $t \in [0, \tau_u]$ and $z \in D$, we have $\hat{\pi}(z) = \hat{H}(Y_{(n)}^{1}|z)/\hat{G}(Y_{(n)}^{1}|z)$.

Proof. We first show that $\hat{S}(t|z) = \hat{H}(t|z)/\hat{G}(t|z)$ for any any $t \in [0, \tau_u]$ and $z \in D$. Without loss of generality, we assume that $0 \leq Y_1 \leq Y_2 \ldots \leq Y_n$. Since $\hat{S}$, $\hat{H}$ and $\hat{G}$ are all right continuous step functions, we only evaluate functions at jump points. For $Y_i$, if $\delta_i = 1$, then $\hat{S}(Y_i|z) = 1 - B_i(z)$, $\hat{H}(Y_i|z) = 1 - B_i(z)$ and $\hat{G}(Y_i|z) = 1$, we have $\hat{S}(Y_i|z) = \hat{H}(Y_i|z)/\hat{G}(Y_i|z)$. If $\delta_i = 0$, then $\hat{S}(Y_i|z) = \hat{H}(Y_i|z) = 1 - B_i(z)$ and $\hat{G}(Y_i|z) = 1 - B_i(z)$, we also have $\hat{S}(Y_i|z) = \hat{H}(Y_i|z)/\hat{G}(Y_i|z)$. Suppose it is true for $Y_i$, i.e. $\hat{S}(Y_i|z) = \hat{H}(Y_i|z)/\hat{G}(Y_i|z)$, We consider the functions at $Y_{i+1}$. If $\delta_{i+1} = 1$, then $\hat{S}(Y_{i+1}|z) = \hat{S}(Y_i|z) \frac{1 - \sum_{r=1}^{i+1} B_r(z)}{1 - \sum_{r=1}^{i} B_r(z)}$, $\hat{H}(Y_{i+1}|z) = 1 - \sum_{r=1}^{i+1} B_r(z)$ and $\hat{G}(Y_{i+1}|z) = \hat{G}(Y_i|z)$, hence

$$\frac{\hat{H}(Y_{i+1}|z)}{\hat{G}(Y_{i+1}|z)} = \frac{1 - \sum_{r=1}^{i} B_r(z)}{G(Y_i|z)} \times \frac{1 - \sum_{r=1}^{i+1} B_r(z)}{1 - \sum_{r=1}^{i} B_r(z)} = \hat{S}(y_{i+1}|z).$$

If $\delta_{i+1} = 0$, then $\hat{S}(Y_{i+1}|z) = \hat{S}(Y_i|z)$, $\hat{H}(Y_{i+1}|z) = 1 - \sum_{r=1}^{i+1} B_r(z)$ and $\hat{G}(Y_{i+1}|z) = \hat{G}(Y_i|z) \frac{1 - \sum_{r=1}^{i+1} B_r(z)}{1 - \sum_{r=1}^{i} B_r(z)}$, hence

$$\frac{\hat{H}(Y_{i+1}|z)}{\hat{G}(Y_{i+1}|z)} = \frac{1 - \sum_{r=1}^{i} B_r(z)}{G(Y_i|z)} = \hat{S}(Y_i|z) = \hat{S}(Y_{i+1}|z).$$

The lemma follows immediately from this result. \hfill \Box

The consistency of the proposed cure rate estimator (2.3) follows from the lemma,

Theorem 2.3.2. Suppose the assumptions (H1), (H2) and (H3) hold and $0 < \pi(z) < 1$, $nh^2 \to \infty$ and $nh/\log n \to \infty$ as $n \to \infty$, we have $\hat{\pi}(z) \stackrel{p}{\to} \pi(z)$ as $n \to \infty$.

Proof. We first prove that $Y_{(n)}^{1} \stackrel{p}{\to} \tau_u$. Suppose $S_u(\tau_u - \epsilon|z_m) = q > 0$ for some $z_m \in D$ and $\epsilon > 0$. By assumption (H1), there exists $\xi$ such that $|S_u(\tau_u - \epsilon|z^*) - S_u(\tau_u - \epsilon|z_m)| < \frac{\epsilon}{2}$ for all $z^* \in D$.
when \(|z^* - z_m| < \xi\). Define \(R_m(\xi) = \{(Y_i, \delta_i, z_i) : |z_m - z_i| < \xi, \delta_i = 1\}\) and denote the number of points in \(R_m(\xi)\) by \(n_m\). Obviously, \(n_m \to \infty\) when \(n \to \infty\). We have

\[
Pr(Y^1_{(n)} < \tau_u - \epsilon) \leq \prod_{R_m(\xi)} Pr(Y_i < \tau_u - \epsilon|z_i) < (1 - q/2)^{n_m} \to 0
\]

Therefore, \(Y^1_{(n)} \overset{p}{\to} \tau_u\). Hence we have \(S(Y^1_{(n)}|z) = \pi(z) + [1 - \pi(z)]S_u(Y^1_{(n)}|z) \overset{p}{\to} \pi(z) + [1 - \pi(z)]S_u(\tau_u|z) = \pi(z)\) as \(n \to \infty\). Due to the consistency of \(\hat{H}\) and \(\hat{G}\), we have

\[
\hat{\pi}(z) = \frac{\hat{H}(Y^1_{(n)}|z)}{G(Y^1_{(n)}|z)} \overset{p}{\to} \frac{H(Y^1_{(n)}|z)}{G(Y^1_{(n)}|z)} = S(Y^1_{(n)}|z) \overset{p}{\to} \pi(z)
\]

The proof is completed.

To establish the asymptotic normality of the estimator (2.3), we first consider an estimator of the cumulative hazard function \(\Lambda(t|z) = -\log(S(t|z))\). It is easy to show that \(\Lambda(t|z) = -\int_0^t \frac{dH^1(s|z)}{H(s - |z|)}\) where \(H^1(t|z) = Pr(Y > t, \delta = 1|z)\) is the conditional sub-survival function of \(Y\). Hence the nonparametric estimator of \(\Lambda(t|z)\) is given by

\[
\hat{\Lambda}(t|z) = -\int_0^t \frac{d\hat{H}^1(s|z)}{\hat{H}(s - |z|)}
\]

where \(\hat{H}^1(t|z) = \sum_{j=1}^n I(Y_j > t, \delta_j = 1)\hat{B}_j(z)\) is the nonparametric estimator of \(H^1(t|z)\). The following lemma establishes the consistency and asymptotic normality of \(\hat{\Lambda}(t|z)\).

**Lemma 2.3.3.** Suppose the assumptions (H1), (H2) and (H3) hold, the nonparametric estimator \(\hat{\Lambda}(t|z)\) satisfies

\[
(nh)^{1/2}(\hat{\Lambda}(t|z) - \Lambda(t|z)) \overset{d}{\to} N(0, \sigma^2(t, z))
\]

(2.9)
in distribution for any \( t \in [0, \tau_u] \) with

\[
\sigma^2(t, z) = \int_0^t H^{-1}(s|z)d\Lambda(s|z).
\]

Proofs of Lemma 2.3.3 can be found in Dabrowska (1987) for Nadaraya-Watson weight and Gonzalez-Manteiga and Cadarso-Suarez (1994) for Gasser-Müller weight. We omit the proof here.

The following theorem establishes the asymptotic property of the proposed non-parametric cure rate estimator (2.3):

**Theorem 2.3.4.** Under the assumptions of Theorem 2.3.2, we have

\[
(nh)^{1/2}(\hat{\pi}(z) - \pi(z)) \xrightarrow{d} N(0, \pi^2(z)\sigma^2(z))
\]

where

\[
\sigma^2(z) = \int_0^{T_u} H^{-1}(s|z)d\Lambda(s|z)ds \int K^2(t)dt
\]

Proof. We first decompose \( \hat{\pi}(z) - \pi(z) \) as follows

\[
\hat{\pi}(z) - \pi(z) = \left[ \hat{S}(Y_{(n)}^1|z) - \exp(-\hat{\Lambda}(Y_{(n)}^1|z)) \right] \\
+ \left[ \exp(-\hat{\Lambda}(Y_{(n)}^1|z)) - S(Y_{(n)}^1|z) \right] \\
+ \left[ S(Y_{(n)}^1|z) - \pi(z) \right].
\]

Since \( S(Y_{(n)}^1|z) = \exp(-\Lambda(Y_{(n)}^1|z)) \), we take Taylor expansion of the second term and have

\[
(nh)^{1/2}(\hat{\pi}(z) - \pi(z)) = R^* + R_1 + R_2 + R_3.
\]
where

\[ R^* = S(Y_{(n)}^1|z) (nh)^{1/2} (\hat{\Lambda}(Y_{(n)}^1|z) - \Lambda(Y_{(n)}^1|z)) \]

\[ R_1 = (nh)^{1/2} \exp(-\Lambda(Y_{(n)}^1|z)) (\Lambda^*(Y_{(n)}^1|z) - \Lambda(Y_{(n)}^1|z))^2 / 2, \]

\[ R_2 = (nh)^{1/2} \left[ \hat{S}(Y_{(n)}^1|z) - \exp(-\hat{\Lambda}(Y_{(n)}^1|z)) \right], \]

\[ R_3 = (nh)^{1/2} (S(Y_{(n)}^1|z) - \pi(z)), \]

where \( \Lambda^*(t|z) \) is a value between \( \hat{\Lambda}(t|z) \) and \( \Lambda(t|z) \). It is easy to see that \( R^* \) converges to a normal random variable with mean zero and variance \( \pi^2(z)\sigma^2(z) \) by Lemma 2.3.3. Hence proving the main result in this theorem is equivalent to showing that the three remainders \( R_1, R_2 \) and \( R_3 \) all converge to 0 in probability. We prove it one by one.

For \( R_1 \), we have \( \exp(-\Lambda(Y_{(n)}^1|z)) \xrightarrow{p} \pi(z) \),

\[ \left[ (nh)^{1/2} (\Lambda^*(Y_{(n)}^1|z) - \Lambda(Y_{(n)}^1|z)) \right]^2 \leq \left[ (nh)^{1/2} (\hat{\Lambda}(Y_{(n)}^1) - \Lambda(Y_{(n)}^1|z)) \right]^2 \xrightarrow{p} c_0 \chi_1^2 \]

and \( (nh)^{-1/2} \to 0 \), we have

\[ R_1 = (nh)^{-1/2} / 2 \times \exp(-\Lambda^*(Y_{(n)}^1|z)) \times \left[ (nh)^{1/2} (\Lambda^*(Y_{(n)}^1|z) - \Lambda(Y_{(n)}^1|z)) \right]^2 \xrightarrow{p} 0. \]

For \( R_2 \), we have \( \hat{S}(Y_{(n)}|z) = \exp\{\sum_{i:Y_i \leq Y_{(n)}^1} \log(1 - \Delta \hat{\Lambda}(Y_i|z))\} \) with \( \Delta \hat{\Lambda}(t|z) = \hat{\Lambda}(t|z) - \hat{\Lambda}(t - |z). \) By Taylor expansion, we have

\[ |R_2| \leq (nh)^{1/2} \sum |\log(1 - \Delta \hat{\Lambda}(Y_i|z)) + \Delta \hat{\Lambda}(Y_i|z)| \]

\[ \leq (nh)^{1/2} \sum (\Delta \hat{\Lambda}(Y_i|z))^2 / 2 \]

\[ = (nh)^{1/2} \sum \hat{H}^{-2}(Y_i - |z)(\Delta \hat{H}^1(Y_i|z))^2 \]

\[ \leq (nh)^{1/2} \sup_{i:Y_i < Y_{(n)}^1} \Delta \hat{H}^1(Y_i|z) \sup_{i:Y_i < Y_{(n)}^1} \hat{H}^{-2}(Y_i - |z) \]

26
\[
\times \int_0^{\tau_u(z)} d(-\hat{H}^1(s|z))
\]

where the sum is over \( \{ i : Y_i \leq Y_{(n)}^1 \} \). As we have \( \int_0^{\tau_u(z)} d(-\hat{H}^1(s|z)) \overset{p}{\to} 1 \), \((nh)^{-1/2} \sup |\Delta \hat{H}^1(t|z)| = \sup((nh)^{1/2} B_r(t|z)) \overset{p}{\to} 0 \) for any \( z \) and \( \sup \hat{H}^{-2}(Y_i - |z|) \to H^{-2}(\tau_u|z) = \pi^{-2}(z)G^{-2}(\tau_u|z) < \infty \), we have \( R_2 \overset{p}{\to} 0 \). For \( R_3 \), we need to show that for any \( \epsilon > 0 \), \( Pr((nh)^{-1/2}(S(Y_{(n)}^1|z) - \pi(z)) > \epsilon) \overset{p}{\to} 0 \) or equivalently, \( Pr((nh)^{-1/2}(1 - \pi(z)))S_u(Y_{(n)}^1|z) > \epsilon) \overset{p}{\to} 0 \). Let \( a_n(\epsilon, z) \) be the value satisfying

\[
S_u(a_n(\epsilon, z)|z) \leq \frac{\epsilon}{(nh)^{1/2}(1 - \pi(z))} \leq S_u(a_n(\epsilon, z) - |z|)
\]

For any \( z^* \in N_z = (z \pm k(\epsilon, z)(nh)^{-1/2}) \), where \( k(\epsilon, z) \) is a constant independent of \( n \) and \( h \), we have

\[
\left| S_u(a_n(\epsilon, z) - |z|) - S_u(a_n(\epsilon, z) - |z^*|) \right| \leq \frac{\epsilon}{2(1 - \pi(z))}(nh)^{-1/2}
\]

\[
\leq \frac{1}{2} S_u(a_n(\epsilon, z) - |z|)
\]

By (H1), such \( k(\epsilon, z) \) exists. Hence we have \( S_u(a_n(\epsilon, z) - |z^*|) \geq \frac{1}{2} S_u(a_n(\epsilon, z) - |z|) \) for any \( z^* \in N_z \). Define \( R_z = \{(Y_i, \delta_i, z_i) : z_i \in N_z, \delta_i = 1 \} \) and denote the number of subjects in \( R_z \) by \( n_z \). By (H3), we have \( n_z = O((nh)^{-1/2} \times n) = O(n^{1/2}h^{-1/2}) \). Let \( p_r(z_i) = G(\tau_u|z_i) \), we have

\[
Pr((nh)^{-1/2}(1 - \pi(z)))S_u(Y_{(n)}^1|z) > \epsilon) = Pr(S_u(Y_{(n)}^1|z) > \frac{\epsilon}{(nh)^{1/2}(1 - \pi(z))})
\]

\[
= Pr(Y_{(n)}^1 < a_n(\epsilon, z) - |z|) \leq \prod_{R_z} \{ 1 - Pr[a_n(\epsilon, z) - \leq Y_i < C_i|z_i] \}
\]

\[
= \prod_{R_z} \left[ 1 - \int_{a_n(\epsilon, z) -}^{\tau_u} G(s|z_i)d(1 - S(s|z_i)) \right]
\]

\[
\leq \prod_{R_z} \left[ 1 + (1 - \pi(z_i)) \int_{a_n(\epsilon, z) -}^{\tau_u} p_r(z_i)dS_u(s|z_i) \right]
\]

27
\[
\leq \prod_{R_z} [1 - c_0 S_u(a_n(\epsilon, z) - |z_i|)] \leq \left[ 1 - c_0 \times \frac{1}{2} S_u(a_n(\epsilon, z)|z) \right]^{n_z}
\rightarrow \left[ 1 - c_0 \frac{\epsilon/2}{(nh)^{1/2}(1 - \pi(z))} \right]^{O(n^{1/2} h^{-1/2})} \rightarrow 0
\]

where \( C_i \) is the censoring variable associated with \( Y_i \) and \( c_0 = \min_{z_i \in N_z} \{(1 - \pi(z_i))p_r(z_i)\} \), a positive constant. The proof is completed. \( \square \)

For practical applications, we may use consistent estimates

\[
\hat{\sigma}^2(z) = \int_0^{Y_n} \frac{-d\hat{H}^1(s|z)}{\hat{H}(s|z)\hat{H}(s - |z|)} ds \int K^2(t)dt = \int K^2(t)dt \sum_{j=1}^{n} \frac{B_j(z)I(\delta_j = 1)}{H(Y_j|z)\hat{H}(Y_j - |z|)}.
\]

for \( \sigma^2(z) \) and \( \hat{\pi}(z) \) for \( \pi(z) \) respectively to obtain the variance of the estimator (2.3).

### 2.4 Simulation Study

We conducted a simulation study to investigate the performance of the proposed model. In the simulation study, we first generated survival data with a cure fraction from the mixture cure model (1.9) and the proportional hazard cure model (1.10). A single covariate \( z \) is considered and it is either a binary variable with equal probability to take 0 and 1 or a continuous variable from the standard uniform distribution. The baseline distribution in the two models is assumed to be the truncated exponential distribution with \( S_u(t) = \frac{\exp(-\lambda t) - \exp(-\lambda \tau_0)}{1 - \exp(-\lambda \tau_0)} \) when \( t \in (0, \tau_0) \) and \( S_u(t|z) = 0 \) when \( t > \tau_0 \), where \( \lambda = \exp(z) \). Therefore, we have \( S_0(t) = S_u(t) \) in the mixture cure model (1.9) and \( F_0(t) = 1 - S_u(t) \) in the proportional hazards model (1.10). We choose \( \tau_0 = 4.605 \), the 99th percentile of the standard exponential distribution. For a discrete covariate, \( z \) is generated from a Bernoulli distribution and the coefficients \( \beta_0 \) and \( \beta_1 \) are set to the values so that the resulting cure rates are from 0.1 to 0.7. For a continuous covariate, \( z \) is generated from the uniform distribution on \((0, 1)\) and
the coefficients are $\beta_0 = 0.476$ and $\beta_1 = 0.358$. We consider cure rate estimation at $z = 0.1, 0.3, 0.5, 0.7$ and 0.9. The censoring variable is generated from exponential distribution with hazard rate 0.3, and thus $\tau_G = \infty > \tau_{S_u}(z)$.

We generated 2000 independent simulation data sets each with a sample size 100. When fitting the proposed nonparametric cure model, we used both Nadaraya-Watson weight (denoted as N-W) and Gasser-Müller weight (denoted as G-M) for continuous $z$ and used Nadaraya-Watson weight only for discrete $z$. The bandwidth in the weight functions is $h = \alpha n^{-1/3}$ according to Theorem 2.3.4. We tried $\alpha$ from 0.1 to 5 and found that $h = 0.4$ works very well under all scenarios and hence it is used in the following model fitting. The kernel function in the weight is set to the cosine kernel function with $K(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2} x\right)$ if $-1 \leq x \leq 1$ and $K(x) = 0$ otherwise. Other kernel functions with a compact support may be considered, such as tri-weight and triangular functions. We found that different kernel functions produced very similar results so we report only the results based on the cosine kernel function. The simulation results, including the biases of the cure rate estimates at given $z$ values, their corresponding empirical variances ($\text{Var}^*$), and the average of the estimated variances based on the asymptotic property of the estimator in Theorem 2.3.4 ($\text{Var}^{**}$), are summarized in Table 2.1 for discrete $z$ and in Table 2.2 for continuous $z$.

We further generated data from the mixture cure model (1.9) with the cure rate $\pi(z) = (1 + \exp(\beta_0(z - 0.5)^2))^{-1}$ and $\beta_0 = 1.85$ for continuous covariate $z$ to compare the performance of the proposed nonparametric method with the existing semiparametric methods. The main difference between the data in the current setting and the data we considered above is that the effect $z$ on the cure rate is not in the form of model (1.9) or (1.10). We fit the data using the proposed method and examine the biases and variances of the cure rate estimates at $z = 0.1, 0.3, 0.5, 0.7$, and 0.9. This result is reported in Table 2.2.
For comparison, we also fit the data in all scenarios with the semiparametric mixture cure model (1.9) using the method proposed by Peng and Dear (2000) and with the semiparametric proportional hazards cure model (1.10) using the method proposed by Tsodikov (2003). The biases of the cure rate estimates from these two models are also reported in Table 2.1 and Table 2.2.

Table 2.1: Biases and variances of cure rate estimates with discrete covariate \( z \) from the proposed nonparametric model and the existing semiparametric models

<table>
<thead>
<tr>
<th>True model</th>
<th>True cure rate ( \pi(z) )</th>
<th>Proposed model(^c)</th>
<th>Peng-Dear</th>
<th>Tsodikov</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC(^a)</td>
<td>.1(0)</td>
<td>.006 .0040 .0036 .902</td>
<td>.002(.0044)</td>
<td>.025(.0031)</td>
</tr>
<tr>
<td></td>
<td>.2(1)</td>
<td>.005 .0062 .0058 .929</td>
<td>.004(.0061)</td>
<td>−.023(.0051)</td>
</tr>
<tr>
<td></td>
<td>.2(0)</td>
<td>.006 .0068 .0062 .932</td>
<td>.004(.0066)</td>
<td>.015(.0048)</td>
</tr>
<tr>
<td></td>
<td>.3(1)</td>
<td>.003 .0079 .0072 .940</td>
<td>.003(.0071)</td>
<td>−.014(.0066)</td>
</tr>
<tr>
<td></td>
<td>.5(1)</td>
<td>.005 .0086 .0079 .934</td>
<td>.005(.0083)</td>
<td>.009(.0079)</td>
</tr>
<tr>
<td></td>
<td>.7(0)</td>
<td>.002 .0062 .0062 .933</td>
<td>.002(.0056)</td>
<td>−.008(.0057)</td>
</tr>
<tr>
<td>PHC(^b)</td>
<td>.1(0)</td>
<td>.003 .0033 .0028 .910</td>
<td>.014(.0025)</td>
<td>.004(.0032)</td>
</tr>
<tr>
<td></td>
<td>.2(1)</td>
<td>−.003 .0057 .0049 .941</td>
<td>−.017(.0046)</td>
<td>−.003(.0054)</td>
</tr>
<tr>
<td></td>
<td>.2(0)</td>
<td>.005 .0059 .0049 .935</td>
<td>.007(.0043)</td>
<td>−.005(.0057)</td>
</tr>
<tr>
<td></td>
<td>.3(1)</td>
<td>−.002 .0072 .0064 .945</td>
<td>−.009(.0065)</td>
<td>.004(.0068)</td>
</tr>
<tr>
<td></td>
<td>.5(1)</td>
<td>.003 .0080 .0074 .949</td>
<td>−.001(.0074)</td>
<td>.000(.0077)</td>
</tr>
<tr>
<td></td>
<td>.7(0)</td>
<td>.000 .0062 .0061 .945</td>
<td>−.003(.0059)</td>
<td>.003(.0059)</td>
</tr>
</tbody>
</table>

\(^a\) mixture cure model.

\(^b\) proportional hazards cure model.

\(^c\) use the Nadaraya-Watson weight.

\(^d\) coverage probability with nominal level 95%.

The results in the tables show that the proposed model estimates the covariate effect with smaller biases no matter what the true model is. The variance estimates based on Theorem 2.3.4 with adjustments discussed following the proof of the theorem are comparable to the empirical variances. However, the coverage probability is a bit smaller than the nominal level 95% (Table 2.1), which means that the variance based on Theorem 2.3.4 may be underestimated in some cases. For a continuous covariate, the biases and variances of the estimates become slightly larger when \( z \) is close to 0.1 or 0.9, which is similar to other kernel-based estimation methods at boundary.
Table 2.2: Biases and variances of cure rate estimates with continuous covariate $z$
from the proposed nonparametric model and the existing semiparametric models

<table>
<thead>
<tr>
<th>True model</th>
<th>$\pi(z)$</th>
<th>True cure rate</th>
<th>Proposed model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N-W weight $^a$</td>
<td>G-M weight $^b$</td>
</tr>
<tr>
<td>MC$^c$</td>
<td>.375(.1)</td>
<td>-.003 .0120 .0077</td>
<td>-.003 .0121 .0101</td>
</tr>
<tr>
<td>.358(.3)</td>
<td>.000 .0081 .0078</td>
<td>.000 .0081 .0078</td>
<td>.002(.0059)</td>
</tr>
<tr>
<td>.342(.5)</td>
<td>.001 .0082 .0077</td>
<td>.001 .0082 .0077</td>
<td>.000(.0045)</td>
</tr>
<tr>
<td>.326(.7)</td>
<td>.004 .0089 .0076</td>
<td>.004 .0088 .0076</td>
<td>.000(.0063)</td>
</tr>
<tr>
<td>.310(.9)</td>
<td>.010 .0123 .0072</td>
<td>.009 .0124 .0095</td>
<td>.005(.0106)</td>
</tr>
<tr>
<td>PHC$^d$</td>
<td>.189(.1)</td>
<td>-.006 .0073 .0044</td>
<td>-.006 .0074 .0059</td>
</tr>
<tr>
<td>.167(.3)</td>
<td>-.002 .0045 .0044</td>
<td>-.002 .0045 .0044</td>
<td>.003(.0050)</td>
</tr>
<tr>
<td>.146(.5)</td>
<td>-.003 .0047 .0051</td>
<td>-.004 .0047 .0052</td>
<td>.005(.0051)</td>
</tr>
<tr>
<td>.126(.7)</td>
<td>-.004 .0042 .0036</td>
<td>-.005 .0042 .0036</td>
<td>.002(.0047)</td>
</tr>
<tr>
<td>.108(.9)</td>
<td>.011 .0057 .0031</td>
<td>.011 .0057 .0040</td>
<td>.008(.0062)</td>
</tr>
<tr>
<td>MC-NL$^e$</td>
<td>.427(.1)</td>
<td>.000 .0078 .0062</td>
<td>.000 .0078 .0068</td>
</tr>
<tr>
<td>.482(.3)</td>
<td>.002 .0060 .0062</td>
<td>.002 .0060 .0062</td>
<td>-.020(.0052)</td>
</tr>
<tr>
<td>.500(.5)</td>
<td>-.001 .0060 .0062</td>
<td>-.001 .0060 .0062</td>
<td>-.037(.0033)</td>
</tr>
<tr>
<td>.482(.7)</td>
<td>.001 .0065 .0062</td>
<td>.001 .0065 .0062</td>
<td>-.019(.0044)</td>
</tr>
<tr>
<td>.427(.9)</td>
<td>.006 .0078 .0061</td>
<td>.005 .0078 .0067</td>
<td>.037(.0082)</td>
</tr>
</tbody>
</table>

$^a$ Nadaraya-Watson weight  
$^b$ Gasser-Müller weight  
$^c$ mixture cure model  
$^d$ proportional hazards cure model  
$^e$ mixture cure model with nonlinear effects of $z$

points. The results from the Nadaraya-Watson weight and the Gasser-Müller weight are comparable too.

When the true effect of the covariate is non-linear, both semiparametric models do not perform well because of the misspecification of the covariate effect on the cure rate. The proposed nonparametric method, on the other side, produces estimates with smaller biases and captures the effect of $z$ on the cure rate.

In summary, the proposed nonparametric cure rate estimation performs well and tends to have smaller biases compared to the existing semiparametric methods when the semiparametric models correctly specify the effect of a covariate on cure rate.
When the semiparametric models misspecify the effect of the covariate on cure rate, the proposed nonparametric method outperforms both methods of Peng and Dear (2000) and Tsodikov (2003) and produces estimates with smaller biases.

2.5 Age Effects in Leukemia Patients with Bone Marrow Transplantation

We applied the proposed model to the data from a leukemia study on patients with bone marrow transplantation. The data set of the study was first reported by Copelan et al. (1991) and was re-studied by Klein and Moeschberger (2003). There are 137 leukemia patients registered in the study and they were followed up to 2640 days. The event of interest is relapse or death due to the leukemia following bone marrow transplantation, and we are interested in the failure time to the event. Among the 137 patients, 42 are censored at the end of the study. The time to event may depend on several factors, such as patient and/or donor age and sex, the stage of initial disease, the time from diagnosis to transplantation. The study also collected patients’ ages, ranging from 7 to 52 years. We will investigate via the proposed model how the patient’s age affects the chance of being cured after the bone marrow transplantation.

To demonstrate the possible presence of cured patients in this study, we plot the Kaplan-Meier survival estimate in Figure 2.1. The curve levels off above 0.6, which clearly indicates that the follow-up time of the study is sufficient and there exists a potential cure fraction.

When applying the proposed cure rate estimation to the data, we consider the bandwidth \( h = 0.4 \), the same value used in the simulation study, and a standardized age variable with values in \((0, 1)\). The fit model provides a nonparametric assessment of the effect of age on the cure rate. For simplicity, we only examine the estimated cure rates at age 7, 18, 30, 41 and 52, which correspond to the minimum, first quartile,
median, third quartile and maximum of the age values in the sample, respectively. The estimated cure rates and their standard errors (s.e.) are reported in Table 2.3. As a comparison, we also reported the estimated cure rates and their standard errors from the two semiparametric cure models considered by Peng and Dear (2000) and Tsodikov (2003). The standard error of nonparametric estimates is calculated according to Theorem 2.3.4 while the standard error of the two semiparametric estimates is calculated by the bootstrap method with 1000 resampling samples.

Both nonparametric and semiparametric models show that the cure rates are substantial for ages under consideration, and the cure rates range from 57% to 67% in
the nonparametric model and from 59% to about 65% in the semiparametric models.

The cure rates generally decrease as age increases, particularly in the semiparametric models. However, when age is very small or very large, the cure rates tend to increase as age increases, demonstrating an S-shape curve. For a full picture of how age affects the cure rate, we estimated the cure rate at $z = 0, 0.02, \ldots, 1$, which corresponds to the actual age from 7 to 52, and plotted the relation between the cure rate and age in Figure 2.2.

To show that the S-shape relationship between the cure rate and age is supported by the data, we divided the original data into 6 groups based on age: $\leq 14$, (14, 21], (21, 28], (28, 35], (36, 45), and $\geq 45$. The method of Maller and Zhou (1992) was applied in each of the 6 groups to estimate the cure rates, which are shown in Figure 2.3. The Figure shows that the cure rates are decreasing when age increases from group 2 to group 5. However, the cure rates of group 1 and group 6 do not follow this pattern and the former is smaller and the latter is greater than the cure rates of their respective adjacent groups. This is in agreement with the finding by the nonparametric model.

The standard errors from the nonparametric model are consistently smaller than those from the semiparametric models, indicating a higher efficiency of the nonparametric model over the two semi-parametric methods in this example. However, the
standard errors are not small enough to reject the results from the semiparametric models because the fit lines from the semiparametric models are within the 95% confidence band of the nonparametric estimates.

In summary, the nonparametric model reveals an interesting age effect for the leukemia patients following bone marrow transplantation. The finding is supported by the data and could not be obtained with the existing semiparametric models. Unfortunately the data evidence is not strong enough to reject the results from the semiparametric models and a larger sample would be required to see if there is a more convincing conclusion.
2.6 Discussion

We proposed a new nonparametric method to explore the effect of a covariate on cure rate for survival data with a cure fraction. Two weight functions, Nadaraya-Watson weight function and Gasser-Müller weight function, were considered in the proposed method. The proposed method extends the work of Maller and Zhou (1992) and it allows either a discrete or a continuous covariate. When there is no covariate, the proposed method reduces to the method of Maller and Zhou (1992). Our numerical study shows that the proposed nonparametric method is flexible and can accommodate a complex effect of a covariate on cure rate that may cause a large bias if using
the existing semiparametric methods. The estimates from the nonparametric method tend to have small biases and variances, which makes the proposed method a viable alternative to assess the effect of a covariate on cure rate.

In developing the asymptotic results, \( \tau_{S_u}(z) \) is supposed to be finite. However, it may not be necessary in practice. Our numerical experience shows if the censoring distribution \( G(t|z) \) has heavier tail than \( S_u(t|z) \), regardless whether the right extreme of \( S_u(t|z) \) is or is not finite, the estimates from the proposed method will tend to have small biases. This is useful in practice because it may be difficult to determine whether a failure time distribution of uncured patients has a finite right extreme or not, or how large it is if it does exist. Similar observations can be found in Maller and Zhou (1992, 1994) for no covariate cases.

The proposed nonparametric method for a single covariate may be extended to a multiple-covariate case. However, like the other kernel-based nonparametric methods, the performance of the method can be compromised in the multiple-covariate case, especially when the number of covariates is more than two and the sample size is not large.
Chapter 3

Transformation Model for Survival Data with a Cure Fraction

3.1 Introduction

As introduced in Chapter 1, there exist three basic models for analyzing survival data with a cure fraction. The mixture cure model (1.9) assumes that the population is divided into two subpopulations so that an individual is either in the cured subpopulation with probability $\pi$; or the individual is in the un-cured subpopulation with probability $1 - \pi$ and the individual’s failure time follows a proper survival function $S_0(t)$. Therefore the mixture cure model is easy to interpret in regard to the cure rate and the survival function of uncured patients. The proportional hazards cure model (1.10), on the other hand, assumes proportional hazard structure with a bounded cumulative hazard function $H_0(t)$. Hence the proportional hazards cure model is capable of making inference in parallel with the standard Cox proportional hazards model (1.3) and it also has a biological explanation. Another advantage of the proportional hazards cure model (1.10) is that the proportional hazards cure model is suitable for Bayesian inference. However, the proportional hazards structure must be checked when we use this model. The proportional odds model (1.11) keeps
the proportional odds structure and the parameters are easy to interpret in terms of
odds ratio of survival functions. The proportional odds cure model also has a bio-
logical explanation. However it lacks the properties of the mixture cure model and the
proportional hazard cure model.

Since each of the three basic cure models has their own advantages, it is of inter-
est to propose some transformation models which can combine the basic cure models
together so that we have an interface between the basic cure models. As discussed in
Chapter 1, Lu and Ying (2004) and Mao and Wang (2010) proposed the transforma-
tion cure model in mixture cure model pattern. However, their models do not combine
any other basic cure models except the mixture cure model. Yin and Ibrahim (2005a)
proposed a transformation cure model (1.14), which includes the mixture cure model
(1.9) and the proportional hazards cure model (1.10) as its special cases. Zeng et al.
(2006) proposed the transformation cure model (1.15), which includes the propor-
tional hazards cure model (1.10) and the proportional odds cure model (1.11) as its
special cases.

By comparing the model of Yin and Ibrahim (2005a) and the model of Zeng et al.
(2006), one may have interest in a new transformation model which has the mixture
cure model (1.9) and the proportional odds cure model (1.11) as its special cases. In
the current chapter, we propose a new transformation cure model which includes the
mixture cure model (1.9) and proportional odds cure model (1.11) as its special cases.
The proposed model is constructed through the Box-Cox transformation.

The remainder of this chapter is organized as follows. In section 3.2, we introduce
the new transformation cure model and its properties. In section 3.3, we present an
algorithm of nonparametric maximum likelihood estimation and derive the asymptotic
property of the estimators, including consistency and asymptotic normality. In section
3.4, we conduct simulation studies to evaluate the finite-sample properties of the
estimations. In Section 3.5, we applied the proposed model to a colon cancer therapy data. A brief discussion is given in section 3.6.

3.2 A New Transformation Cure Model

In Zeng et al. (2006)’s transformation model (1.15), a Box-Cox transformation is imposed on a baseline function on the right,

\[ S(t|z) = [1 + a\theta(z)F_0(t)]^{-1/a}. \]

While in Yin and Ibrahim (2005a)’s transformation model (1.14), a Box-Cox transformation is applied on the population survival function on the left,

\[ \frac{S(t|z, x)^a - 1}{a} = -\theta(z)F_0(t|x), \]

which results in model (1.14) following some algebra. These two transformation cure models inspire us to consider applying the Box-Cox transformation on both the population survival function and the baseline function, which leads to the following model

\[ -B_a(S(t|z)) = B_a(1 + \theta(z)F_0(t)), \quad (3.1) \]

where \( S(t|z) \) is the improper population survival function; \( \theta(z) \) should be in the interval \((0, 1)\). In sequel, \( \theta(z) \) is modeled as \( \frac{\exp(\beta'z)}{1 + a \exp(\beta'z)} \), which is similar to model (1.14); \( F_0(t) \) is a proper cumulative distribution function and \( a \) is the transformation
parameter; \( B_a(x) \) is the Box-Cox transformation function,

\[
B_a(x) = \begin{cases} 
\frac{x^a - 1}{a}, & \text{if } a \neq 0; \\
\log(x), & \text{if } a = 0.
\end{cases}
\]

Model (3.1) defines a class of transformation cure models. When \( a = 0 \), (3.1) becomes

\[
-\log S(t|z) = \log(1 + \exp(\beta'z)F_0(t)),
\]

which can be rewritten as

\[
S(t|z) = \left[1 + \exp(\beta'z)F_0(t)\right]^{-1},
\]

a proportional odds cure model (1.11). When \( a = 1 \), model (3.1) becomes

\[
S(t|z) = 1 - \frac{\exp(\beta'z)}{1 + \exp(\beta'z)}F_0(t) - \frac{\exp(\beta'z)}{1 + \exp(\beta'z)}S_0(t)
\]

which is a mixture cure model (1.9) with \( S_0(t) = 1 - F_0(t) \) being a proper survival function. When \( 0 < a < 1 \), model (3.1) defines some intermediate cure model between the proportional odds cure model (1.11) and the mixture cure model (1.9).

The relationship between the basic cure models, model (1.9), model (1.10) and model (1.11), and the three transformation cure models, model (1.14), model (1.15) and model (3.2), is depicted in Figure 3.1, where MC model denotes the mixture cure model (1.9), PHC model denotes the proportional hazards cure model (1.10) and POC denotes the proportional odds cure model (1.11).

Through some algebra, we rewrite equation (3.1) as

\[
S(t|z) = \left\{2 - \left[1 + \frac{\exp(\beta'z)}{1 + a \exp(\beta'z)}F_0(t)\right]^a\right\}^{1/a}
\]

In the transformation model, the parameter \( a \) is restricted in the interval \([0, 1]\). It is not only because model (3.2) yields an intermediate model between the proportional odds cure model (1.11) and the mixture cure model (1.9) when \( a \in (0, 1) \) but also...
Figure 3.1: Relationship between the transformation cure models and the basic cure models.

because model (3.2) fails to be a valid cure model for some specific $\exp(\beta'z)$ when $a \notin [0, 1]$. Restrictions of the transformation parameter exist in model (1.15) and model (1.14) too, though the restriction of the transformation parameter of model (1.14) was not addressed by Yin and Ibrahim, where the parameter $a$ should be non-negative only while the authors claimed that $a$ can be any value in the real line. For our proposed model, we can prove that model (3.2) is a proper cure model if and only if $a \in [0, 1]$, which is summarized in the following proposition.

**Proposition 1.** In model (3.2), we have (1) $S(t|z)$ is monotone decreasing in $t$ and (2) $S(t|z) \in (0, 1]$ for any $z$ if and only if $a \in [0, 1]$.

*Proof.* The proof of the monotonicity in the first part is trivial. For the second part of the proposition, we divide the whole real line into three intervals, $0 \leq a \leq 1$, $a < 0$ and $a > 1$, and discuss them respectively. First, we show that $\left[1 + \frac{\exp(\beta'z)}{1 + a \exp(\beta'z)} F_0(t)\right]^a \in [1, 2]$, or simply

$$\left[1 + \frac{\exp(\beta'z)}{1 + a \exp(\beta'z)}\right]^a < 2 \text{ for any } a \in [0, 1].$$

(3.3)
The left hand side can be proved to be monotone increasing in $a$ and hence we have
\[
\left[1 + \frac{\exp(\beta'z)}{1 + a \exp(\beta'z)}\right]^a < \left[1 + \frac{\exp(\beta'z)}{1 + \exp(\beta'z)}\right] < 2.
\]

For the “only if” part, we show that $S(t|z) \not\in (0,1]$ for some $t$ and $z$ when $a \not\in [0,1]$.

When $a < 0$, it is easy to see that $1 + \frac{\exp(\beta'z)}{1 + a \exp(\beta'z)}F_0(t)$ could be negative when \(\exp(\beta'z)\) is large enough and hence model (3.2) fails to be a valid model; when $a > 1$, let $\exp(\beta'z) \to \infty$ and $t \to \infty$, we have
\[
\left[1 + \frac{\exp(\beta'z)}{1 + a \exp(\beta'z)}F_0(t)\right]^a \to (1 + 1/a)^a > (1 + 1/1)^1 = 2
\]
and the model (3.2) is no longer a valid model. The proof is completed.

An important issue in a transformation cure model is identifiability. For example, the identifiability problem of model (1.15) was addressed in Zeng et al. (2006). Model (1.15) is proved to be identifiable when the transformation parameter $a$ is given. To introduce the identifiability, we give a regularity condition,

(C1) The covariate $z$ is bounded with probability 1, and if there exists a vector $\beta$ such that $\beta'z = 0$ with probability 1, then $\beta = 0$.

The boundedness in probability 1 is defined as: $Pr(\sup_{z \in \mathcal{Z}} ||z|| < \infty) = 1$, where $\mathcal{Z}$ is the space of $z$ and $||z||$ denotes the Euclidean distance.

The following proposition shows that our proposed model (3.2) is identifiable when the transformation parameter $a$ is known.

**Proposition 2.** Model (3.2) is identifiable when $a$ is given under the regularity condition (C1).

**Proof.** Suppose that two parameter sets, $(\beta, F_0(t))$ and $(\beta^*, F_0^*(t))$, produce the same
survival function \((3.2)\), we have

\[
1 + \frac{\exp(\beta'z)}{1 + a \exp(\beta'z)} F_0(t) = 1 + \frac{\exp(\beta'^*z)}{1 + a \exp(\beta'^*z)} F^*_0(t) \tag{3.4}
\]

let \(t \to \infty\), we have

\[
\frac{\exp(\beta'z)}{1 + a \exp(\beta'z)} = \frac{\exp(\beta'^*z)}{1 + a \exp(\beta'^*z)}.
\]

Because of the monotonicity of the function \(f(x) = x/(1 + ax)\) in \(x\), we have \(\exp(\beta'z) = \exp(\beta'^*z)\) and hence \(\beta = \beta^*\) by regularity condition \((C1)\). Plugging \(\beta = \beta^*\) into equation \((3.4)\), we have \(F_0(t) = F^*_0(t)\). The proof is completed. \(\square\)

However, model \((3.2)\) is not identifiable when \(a\) is unknown. Suppose there are two sets of parameters, one set is \(\gamma_1: a = 0.5, \beta = (\log(10/3), \beta_1), F_0(t) = F_1(t)\) and the other set is \(\gamma_2: a = 1, \beta = (\log(3), \beta_1), F_0(t) = F_2(t)\) where

\[
F_2(t) = \frac{4}{3} \left\{ 1 - \left\{ 2 - \left[ 1 + 5F_1(t)/4 \right]^{0.5} \right\}^2 \right\}.
\]

We can show that \(F_2(t)\) is monotone increasing with \(F_2(0) = 0\) and \(F_2(\infty) = 1\), which means that \(F_2(t)\) is a valid cumulative distribution function. It is easy to show that the population survival function \(S(t|z = 0)\) under parameter set \(\gamma_1\) is identical to \(S(t|z = 0)\) under parameter set \(\gamma_2\). Therefore, model \((3.2)\) is not identifiable when the transformation parameter is not fixed.

The identifiability problem also arises in other transformation cure models when the transformation parameter is not fixed. Model \((1.15)\) was proved to be identifiable when the transformation parameter is given and unidentifiable when the transformation parameter is not given. The identifiability problem when the transformation parameter is not given was not addressed in Yin and Ibrahim \((2005a)\) for model.
However, the authors claimed that the profile likelihood function of the transformation parameter $a$ is flat and it is difficult to estimate $a$ by maximizing the profile likelihood, which implies that model (1.14) is not identifiable when the transformation parameter $a$ is not given.

From model (3.2), we can derive the population density function and population hazard function

$$f(t|z) = \left\{ 2 - [1 + \theta(z)F_0(t)]^a \right\}^{1/a-1}[1 + \theta(z)F_0(t)]^{a-1}\theta(z)f_0(t)$$

and

$$h(t|z) = \left\{ 2 - [1 + \theta(z)F_0(t)]^a \right\}^{-1}[1 + \theta(z)F_0(t)]^{a-1}\theta(z)f_0(t)$$

where $f_0(t) = \frac{\partial F_0(t)}{\partial t}$. The cure rate of model (3.2) is

$$\pi(z) = \lim_{t \to \infty} S(t|z) = \left[ 2 - (1 + \theta(z))^a \right]^{1/a}$$

and the survival function of the uncured patients is derived as

$$S_{\text{uncured}}(t|z) = \frac{S(t|z) - \pi(z)}{1 - \pi(z)}$$

which is a proper survival function.

### 3.3 Estimation and Asymptotic Property

Throughout the chapter, $a$ is assumed known and we focus on the estimation of $\beta$ and $F_0(t)$ only.

For a concise presentation, model (3.2) is rewritten through a transformation
function,

\[ S(t|z) = G(\theta(z)F_0(t)). \tag{3.5} \]

where \( G(x) = (2 - (1 + x)^a)^{1/a} \) is the transformation function. Given the observed data (2.1), we have the likelihood function

\[ L = \prod_{i=1}^{n} \left\{ -G'[\theta(z_i)F_0(t_i)]\theta(z_i)f_0(t_i) \right\}^{\delta_i} \left[ G(\theta(z_i)F_0(t_i)) \right]^{1-\delta_i}. \tag{3.6} \]

The estimates of \( \beta \) and \( F_0(t) \) are obtained by maximizing the likelihood function \( L \).

To maximize \( L \), \( F_0(t) \) may be modeled by a parametric distribution. The Weibull or Gamma distribution can be used to model \( F_0(t) \). For a robust inference, we do not assume any parametric form for \( F_0(t) \) in sequel. However, the maximum of the likelihood function (3.6) does not exist when \( F_0(t) \) is continuous because we can choose some \( f(t_i) = \infty \) for some \( t_i \) with \( \delta_i = 1 \) and then the likelihood function (3.6) is infinity. As a natural modification, we assume that \( F_0(t) \) is a discrete function with jumps only at uncensored survival times. A similar argument can be found in Section 4.3 of Kalbfleisch and Prentice (2002). To maximize a likelihood function in regard to a cumulative hazard function \( \Lambda(t) \), Kalbfleisch and Prentice assumed that \( \Lambda(t) \) is a right-continuous step function and takes jumps only on uncensored failure times. Zeng et al. (2006) provided another example of maximizing a likelihood function in regard to a cumulative distribution function \( F(t) \) in model (1.15). They assumed that \( F(t) \) is discrete and takes value only at uncensored failure times with certain restrictions. The similar argument can also be found in Tsodikov (1998b, 2001).

We apply the same technique used in Zeng et al. (2006) to maximize the likelihood function (3.6). Specifically, we suppose that \( t_{(1)}, t_{(2)}, \cdots, t_{(m)} \) are the increasingly ordered uncensored failure times and \( p_{(1)}, p_{(2)}, \cdots, p_{(m)} \) are the corresponding jumps
with constraint \( \sum_{i=1}^{m} p(i) = 1 \). The estimators are obtained by maximizing the following log-likelihood function,

\[
l = \sum_{i=1}^{m} \{ \log(-G'[\theta(z(i))F_0(t(i))]) + \log(\theta(z(i))) + \log p(i) \} \\
+ \sum_{i=1}^{n} (1 - \delta_i) \log(G(\theta(z_i)F_0(t_i))) + \lambda(1 - \sum_{i=1}^{m} p(i)) \tag{3.7}
\]

where \( F_0(t_j) = \sum_{t(i) \leq t_j} p(i) \) and \( \lambda \) is the Lagrange multiplier. Taking the first derivative of \( l \) with respect to \( p(k) \), for \( k = 1, \ldots, m \), we have

\[
\frac{1}{p(k)} + \sum_{i=1}^{m} \frac{G''[\theta(z(i))F_0(t(i))]\theta(z(i))I(t(i) \geq t(k))}{G'[\theta(z(i))F_0(t(i))]} \\
+ \sum_{j=1}^{n} (1 - \delta_j) \frac{G''[\theta(z_j)F_0(t_j)]\theta(z_j)I(t_j \geq t(k))}{G[\theta(z_j)F_0(t_j)]} - \lambda = 0,
\]

where \( G''(x) = \frac{\partial^2 G(x)}{\partial x^2} \), the second derivative of \( G(x) \) with respect to \( x \). Taking the difference between the \((k+1)\)-th equation and \( k \)-th equation for \( k = m-1, \ldots, 1 \), we have

\[
\frac{1}{p(k)} - \frac{1}{p(k+1)} = \frac{G''(\theta(z(k))F_0(t(k)))\theta(z(k))}{G'(\theta(z(k))F_0(t(k)))} \\
- \sum_{t(k) < t_j < t(k+1)} \frac{G'(\theta(z_j)F_0(t_j))\theta(z_j)}{G(\theta(z_j)F_0(t_j))} \tag{3.8}
\]

Through these equations, \( p_{(m-1)}, \ldots, p_{(1)} \) could be numerically solved when \( p_{(m)}, \beta \) and \( F_0(t) \) are given. Therefore, we can treat the log-likelihood function \( l \) as the functions of \( \beta, \lambda \) and \( p_{(m)} \) while \( p_{(m-1)}, \ldots, p_{(1)} \) are referred to as the functions of \( \beta \) and \( p_{(m)} \). Plugging in the estimators of \( p_{(1)}, \ldots, p_{(m-1)} \) by equation (3.8) into equation (3.7), we obtain the profile likelihood \( l_{pr}(\beta, \lambda, p_{(m)}) \). The score equations of \( \beta, \lambda \) and
To prove the consistency and asymptotic normality of the estimates, the following algorithm to solve the equations. The first and second derivative of \( p \)

Setting these score equations to zero, we get the estimating equations for \( \beta, \lambda \)

To obtain the estimations from the estimating equations, we first use the first two equations of equations (3.9) to eliminate the \( \lambda \) and then apply the Newton-Raphson algorithm to solve the equations. The first and second derivative of \( p(i) \) with respect to \( \beta \) and \( p(m) \) could be computed by equation (3.8).

After getting the maximum likelihood estimations \( \hat{\beta} \) of \( \beta \) and \( \hat{p}(m) \) of \( p(m) \), we can estimate the asymptotic variances of \( \hat{\beta} \) and \( \hat{p}(m) \) based on \( l_{pr}(\beta, p(m)) \), which could be written as \( l_{pr}(\beta, p(m)) \) since \( \lambda \) disappears in equation (3.7) when \( 1 - \sum_{i=1}^{m} p(i) = 0 \). The asymptotic variance of \( \hat{\beta} \) and \( \hat{p}(m) \) is estimated by the negative inverse matrix of the Hessian matrix of \( l_{pr}(\beta, p(m)) \) evaluated at \( (\hat{\beta}, \hat{p}(m)) \), which is

\[
V = - \left( \begin{array}{cc} \frac{\partial^2 l_{pr}(\beta, p(m))}{\partial \beta^2} & \frac{\partial^2 l_{pr}(\beta, p(m))}{\partial \beta \partial p(m)} \\ \frac{\partial^2 l_{pr}(\beta, p(m))}{\partial p(m) \partial \beta} & \frac{\partial^2 l_{pr}(\beta, p(m))}{\partial p(m)^2} \end{array} \right)^{-1} \bigg|_{\beta=\hat{\beta}, p(m)=\hat{p}(m)} \tag{3.10}
\]

To prove the consistency and asymptotic normality of the estimates, the following
regularity conditions, as well as condition \((C1)\), are required:

(C2) Conditional on \(z\), the right-censoring time \(C\) is independent of \(T\).

(C3) The true value of \(\beta\), denoted by \(\beta_0\), belongs to the interior of a known compact set \(\beta\) and the true baseline function of \(F_0(t)\) is differentiable with \(F_0'(x) > 0\) for all \(x \in R^+\).

(C4) The link function \(\theta(z)\) is strictly increasing and twice-continuously differentiable with respect to \(\beta'z\) and \(\theta(z) > 0\). The transformation \(G\) satisfies

\[
G(0) = 1, \ G(x) > 0, \ G'(x) < 0, \ G^{(3)}(x) \text{ exists and is continuous.}
\]

where \(G^{(3)}\) is the third derivative of \(G(x)\).

The asymptotic properties of \(\hat{\beta}\) and \(\hat{F}_0(t)\) are summarized in the following theorem,

**Theorem 3.3.1.** Under the regularity conditions \((C1)-(C4)\), we have \(|\hat{\beta} - \beta_0| \to 0\) and \(\sup_{t \in R^+} |\hat{F}_0(t) - F_0(t)| \to 0\) almost surely and \(\sqrt{n}(\hat{\beta} - \beta_0, \hat{F}_0(t) - F_0(t))\) converges to a mean zero Gaussian process in the metric space \(l^\infty(\mathcal{H})\), where

\[
\mathcal{H} = \{(h_1, h_2) : h_1 \in R^d, h_2 \text{ is a function on } [0, \infty); \|h_1\| \leq 1, \|h_2\|_V \leq 1\}
\]

here \(d\) is the dimension of \(\beta\) and \(\|h_1\|\) is the Euclidean metric of \(h_1\); \(\|h_2\|_V = \sup \sum_{i=1}^k |h_2(t_{i+1}) - h_2(t_i)|\) over all finite partition \(0 = t_1 < t_2 < \cdots < t_{k+1} = \infty\), called the total variance of \(h_2\) in \([0, \infty]\); \(l^\infty(\mathcal{H})\) is the space of all bounded linear functions on \(\mathcal{H}\).

This theorem establishes the strong consistency and the asymptotic normality of the nonparametric maximum likelihood estimates based on the profile likelihood theory of **Murphy and van der Vaart (2000)**. The regularity conditions \((C1)-(C4)\) were
used in Zeng et al. (2006). Condition (C1) is a usual condition for a design matrix in regression; Condition (C2) is slightly different from the one used in Zeng et al. (2006), where they added $Pr(C = \infty|x) > 0$ to ensure that some cured patients are observed by defining a cure threshold. However, their condition can be simplified to condition (C2) because all the censored subjects beyond the largest observed failure time can be treated as cured since $\hat{F}(t) = 1$ when $t \geq t_{(m)}$. Therefore the largest uncensored failure time $t_{(m)}$ can be treated as a cure threshold in our proposed model (3.2) and then our condition (C2) is equivalent to the counterpart in Zeng et al. (2006). The boundedness and smoothness of condition (C3) are often imposed in semi-parametric inference. For condition (C4), it is easy to show that $\theta(z) = \frac{\exp(\beta'z)}{1 + a \exp(\beta'z)}$ and $G(x) = [2 - (1 + x)^a]^{1/a}$ satisfy the requirements. The proof of the consistency and asymptotic normality is similar to the proof in Zeng et al. (2006) and we omit the detailed proof here.

### 3.4 Simulation Study

In this section, we conduct simulation studies to examine the small sample performance of the proposed estimating method. The survival function of the transformation cure model is

$$S(t|z) = \left\{ 2 - \left[ \frac{\exp(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2)}{1 + a \exp(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2)}[1 - \exp(-t)]^a \right]^{1/a} \right\}$$

where $Z_1$ is a Bernoulli random variable, $Z_2$ is a uniformly distributed random variable in $(0, 1)$, $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = -0.5$. We select $a = 0.00, 0.25, 0.5, 0.75$ and $1.00$. The censoring time is generated from $1 - \exp(-\gamma x)$, where $\gamma$ varies to generate a 40% censoring rate for different values of $a$. For each simulated dataset, the proposed method was implemented to get the estimates of $\beta_0$ and $\beta_1$ and their standard errors.
Table 3.1: Simulation results of $\beta$ from 1000 replications

<table>
<thead>
<tr>
<th>$a$</th>
<th>$n$</th>
<th>Parameter</th>
<th>True Value</th>
<th>Bias</th>
<th>ESE$^a$</th>
<th>SE$^b$</th>
<th>CP(%)$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>100</td>
<td>$\beta_0$</td>
<td>0.5</td>
<td>-.037</td>
<td>0.424</td>
<td>0.444</td>
<td>96.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>1.0</td>
<td>.042</td>
<td>0.659</td>
<td>0.661</td>
<td>95.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_2$</td>
<td>-.5</td>
<td>-.008</td>
<td>0.379</td>
<td>0.377</td>
<td>95.7</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>$\beta_0$</td>
<td>0.5</td>
<td>-.004</td>
<td>0.314</td>
<td>0.311</td>
<td>94.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>1.0</td>
<td>.006</td>
<td>0.456</td>
<td>0.462</td>
<td>95.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_2$</td>
<td>-.5</td>
<td>-.007</td>
<td>0.263</td>
<td>0.264</td>
<td>95.9</td>
</tr>
<tr>
<td>0.25</td>
<td>100</td>
<td>$\beta_0$</td>
<td>0.5</td>
<td>.000</td>
<td>0.574</td>
<td>0.546</td>
<td>95.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>1.0</td>
<td>.088</td>
<td>0.903</td>
<td>0.876</td>
<td>95.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_2$</td>
<td>-.5</td>
<td>-.041</td>
<td>0.488</td>
<td>0.493</td>
<td>97.6</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>$\beta_0$</td>
<td>0.5</td>
<td>.015</td>
<td>0.385</td>
<td>0.372</td>
<td>95.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>1.0</td>
<td>.019</td>
<td>0.582</td>
<td>0.591</td>
<td>96.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_2$</td>
<td>-.5</td>
<td>-.015</td>
<td>0.360</td>
<td>0.336</td>
<td>94.2</td>
</tr>
<tr>
<td>0.50</td>
<td>100</td>
<td>$\beta_0$</td>
<td>0.5</td>
<td>.032</td>
<td>0.595</td>
<td>0.606</td>
<td>97.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>1.0</td>
<td>.059</td>
<td>0.992</td>
<td>0.976</td>
<td>97.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_2$</td>
<td>-.5</td>
<td>-.036</td>
<td>0.571</td>
<td>0.562</td>
<td>97.2</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>$\beta_0$</td>
<td>0.5</td>
<td>-.008</td>
<td>0.403</td>
<td>0.405</td>
<td>96.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>1.0</td>
<td>.070</td>
<td>0.689</td>
<td>0.658</td>
<td>95.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_2$</td>
<td>-.5</td>
<td>-.017</td>
<td>0.366</td>
<td>0.373</td>
<td>96.0</td>
</tr>
<tr>
<td>0.75</td>
<td>100</td>
<td>$\beta_0$</td>
<td>0.5</td>
<td>-.010</td>
<td>0.695</td>
<td>0.617</td>
<td>96.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>1.0</td>
<td>.125</td>
<td>1.124</td>
<td>1.007</td>
<td>97.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_2$</td>
<td>-.5</td>
<td>-.025</td>
<td>0.600</td>
<td>0.570</td>
<td>95.6</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>$\beta_0$</td>
<td>0.5</td>
<td>-.000</td>
<td>0.426</td>
<td>0.416</td>
<td>96.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>1.0</td>
<td>.044</td>
<td>0.674</td>
<td>0.674</td>
<td>96.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_2$</td>
<td>-.5</td>
<td>-.012</td>
<td>0.399</td>
<td>0.383</td>
<td>94.7</td>
</tr>
<tr>
<td>1.00</td>
<td>100</td>
<td>$\beta_0$</td>
<td>0.5</td>
<td>.015</td>
<td>0.615</td>
<td>0.592</td>
<td>95.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>1.0</td>
<td>.089</td>
<td>0.979</td>
<td>0.953</td>
<td>95.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_2$</td>
<td>-.5</td>
<td>-.034</td>
<td>0.548</td>
<td>0.539</td>
<td>96.0</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>$\beta_0$</td>
<td>0.5</td>
<td>.004</td>
<td>0.415</td>
<td>0.401</td>
<td>95.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>1.0</td>
<td>.035</td>
<td>0.685</td>
<td>0.646</td>
<td>94.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_2$</td>
<td>-.5</td>
<td>-.016</td>
<td>0.366</td>
<td>0.368</td>
<td>95.2</td>
</tr>
</tbody>
</table>

$^a$ ESE is the empirical standard error
$^b$ SE is the average of estimated standard errors by Theorem 3.3.1
$^c$ CP is the coverage probability with nominal level 95%.

The simulation result is summarized in Table 3.1 based on 1000 replications. The simulation result shows that the proposed method performs well with sample size 100 and 200. The biases are small and the estimated standard errors are comparable with the empirical standard errors. The coverage probabilities are close to the
nominal level 95%, especially when the sample size is large \((n = 200)\). The standard errors are improved significantly when the sample size is increased from 100 to 200;

### 3.5 Treatment Effect Analysis for Colon Cancer Patients

We applied the proposed transformation cure model on a data set about colon cancer therapy (Moertel et al., 1990, 1995). The trial started in March 1984 and completed in October 1987. In the trial, there are three groups of patients who received observation, levamisole alone treatment and levamisole combined with fluorouracil treatment, respectively, at different stages. We are interested in the effect of levamisole combined with fluorouracil treatment at stage C, which is compared with the null treatment (observation). The two groups are referred to as control group and treatment group, respectively. There are 315 patients in the control group and 304 patients in the treatment group. By the end of study, 177 patients in the control group and 103 patients in the treatment group had recurrence. The failure time in our study is the time to the cancer recurrence.

We first drew the Kaplan-Meier curves of the two groups in Figure 3.3. From the figure, we can see that there are positive plateau at the tails of both survival curves. Hence a cure model is a suitable approach for this data set. Before we fit the proposed model, we selected the transformation parameter \(a\) as the one that maximizes AIC criterion. It is equivalent to the one that maximizes the observed likelihood function (3.7) because the number of parameters is constant. The observed likelihood function is maximized around \(a = 0.9\) (Figure 3.2) and therefore we choose \(a = 0.9\) in the model fitting.

By fitting the proposed model, we have estimates summarized in Table 3.2.

From the table, we see that \(\beta_1\) is significantly negative, which means that the treatment has negative effect on the failure time. The resulted cure rates of the two
groups are 0.416 and 0.601. We also plot the predicted survival curves in Figure 3.3. We can see that the predicted population survival function of the treatment group is always greater than the predicted survival function of the control group. We conclude that the treatment, levamisole combined with fluorouracil, not only prolongs the patients' survival times but also increases the cure rate of the patients. Our result agrees with the common conclusion from the cancer study that the therapy
Table 3.2: Estimated parameters and confidence intervals

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>s.e.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_0$</td>
<td>0.458</td>
<td>0.198</td>
<td>(0.071, 0.845)</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>-0.724</td>
<td>0.171</td>
<td>(-1.059, -0.389)</td>
</tr>
</tbody>
</table>

with levamisole plus fluorouracil is effective with surgically treated Stage C colon carcinoma (Moertel et al., 1990, 1995).

3.6 Discussion

We proposed a class of cure models by imposing the Box-Cox transformation on both the survival function and the baseline function. The proposed model includes two popular cure models, the mixture cure model (1.9) and the proportional odds cure model (1.11), as its special cases.

An algorithm to calculate the nonparametric maximum likelihood estimations is proposed. Our simulation experience showed that for small sample cases, the nonparametric maximum likelihood estimates of both unknown parameters $\beta$ and baseline function $F_0(t)$ are both unbiased.

By comparing the three transformation cure models, model (1.14), model (1.15) and the proposed model in this chapter, one may be interested in a more general transformation cure model, which includes the mixture cure model (1.9), the proportional hazards cure model (1.10) and the proportional odds cure model (1.11) as its special cases. If such general transformation cure model exists, we will have a unified approach for transformation cure models. We discuss one such general model later in Chapter 7.
Figure 3.3: Kaplan-Meier curves and predicted survival curves of colon cancer data.
Chapter 4

Transformation Model for Survival Data with or without a Cure Fraction

4.1 Introduction

In survival analysis, the survival models, including cure models and non-cure models, only fit specific types of survival data. For example, the Cox proportional hazards model (1.3) and the accelerated failure time model (1.5) are only suitable for survival data without a cure fraction while the mixture cure model (1.9) only handles survival data with a cure fraction. It is true for the transformation models as well. Transformation non-cure models are specified for survival data without a cure fraction while the transformation cure models, such as model (1.14), model (1.15) and model (3.2), are only suitable for survival data with a cure fraction. Transformation cure models are not appropriate for the survival data without a cure fraction and vice versa. Therefore, the application of the transformation models requires prior information about the existence of cure fraction of survival data. In reality, such information is often limited or unavailable. It is preferable to have a model that relies less on prior information and is suitable for both types of data.

To our knowledge, there is very limited work on such modeling. Yin and Ibrahim
(2005b) proposed a transformation model, which includes both cure models and non-cure models as its special cases,

\[
h_{\text{pop}}(t|z) = \begin{cases} 
\frac{f_0(t) + r \beta' z}{r} & \text{if } 0 < r \leq 1 \\
\exp(\beta' z) f_0(t) & \text{if } r = 0,
\end{cases}
\]

where \( h_{\text{pop}}(t|z) \) is the population hazard function; \( f_0(t) \) is a probability density function; \( \beta \) is an unknown coefficient associated with covariate \( z \); \( r \) is the transformation parameter. It is easy to see that model (4.1) becomes the proportional hazards cure model (1.10) when \( r = 0 \) and non-cure model when \( 0 < r \leq 1 \). However, it is easy to see that model (4.1) has a parameter constraint: \( f_0(t)^r + r \beta' z \) should be positive when \( 0 < r \leq 1 \), which could be violated in real applications. Another disadvantage of model (4.1) is that the transformation parameter \( r \) must be given and hence the type of model, cure model or non-cure model, is determined before model fitting. These properties restrict the application of model (4.1) and motivate us to consider a new transformation model, which could handle both survival data with and without a cure fraction. In this chapter, we propose a new such transformation model, which includes both cure models and non-cure models as its special cases. Unlike the model (4.1), our proposed model has no parameter constraints and the estimation procedure is straightforward and easy to implement. The transformation parameter, which determines the cure model or non-cure model, is estimated from data. Two estimation procedures, one parametric procedure and one semi-parametric procedure, are presented.

The rest of this chapter is organized as follows. The new transformation model is introduced in Section 4.2. In Section 4.3, we present a parametric estimation method by a parametric maximum likelihood method. A semi-parametric inference is introduced in section 4.4. We present extensive simulation studies in Section 4.5.
and two real data applications, one with a cure fraction and the other without a cure fraction, are presented in Section 4.6. A brief discussion is given in Section 4.7.

### 4.2 A New Transformation Model

The proportional hazards model for survival data, with or without a cure fraction, is given as

$$ \Lambda(t|z) = \exp(\beta'z)\Lambda_0(t) $$

(4.2)

where $\Lambda(t|z)$ is the population cumulative hazard function; $\beta$ is an unknown covariate coefficient; $\Lambda_0(t)$ is a baseline cumulative hazard function and it could be parametrically specified or non-parametrically specified. If $\Lambda_0(t)$ is proper, $\Lambda(t|z)$ is proper too and hence model (4.2) becomes the Cox proportional hazards non-cure model (1.3) for survival data without a cure fraction. If $\Lambda_0(t)$ is improper, $\Lambda(t|z)$ is also improper and hence model (4.2) becomes the proportional hazards cure model (1.10) for the survival data with a cure fraction.

Since the improperness of $\Lambda_0(t)$ determines the type of the model, the cure model or the non-cure model, it motivates us to consider a transformation model as follows,

$$ \Lambda(t|z) = \exp(\beta'z)G_r(\Lambda_0(t)) $$

(4.3)

where the baseline cumulative hazard function $\Lambda_0(t)$ is proper; $G_r(s)$ is a transformation function defined as follows

$$ G_r(s) = \begin{cases} 
[1 - \exp(-rs)]/r & \text{if } r \neq 0, \\
 s & \text{if } r = 0.
\end{cases} $$

(4.4)
It is easy to see that model (4.3) leads to a cure model if the population cumulative hazard function \( \Lambda(t|z, x) \) is improper or a non-cure model if \( \Lambda(t|z, x) \) is proper.

For model flexibility, we may incorporate another covariate \( x \) into \( \Lambda_0(t) \) and we have \( \Lambda_0(t|x) \) and \( \Lambda(t|z, x) \), where \( x \) may not share common elements with \( z \) for the identifiability purpose. For example, when \( \Lambda_0(t|x) \) follows Weibull distribution (4.6) and \( r = 1 \), we have

\[
\Lambda(t|z, x) = \exp(\beta'z + \gamma'x)t^\alpha.
\]

It is seen that \( z \) and \( x \) cannot have common elements, including the intercept terms. In sequel, we assume that \( x \) contains an intercept term while \( z \) does not.

Through (4.3) and (4.4), we have a general class of transformation models. Specifically, when \( r = 0 \), we have

\[
\Lambda(t|z, x) = \exp(\beta'z)\Lambda_0(t|x),
\]

which is a parametric proportional hazards non-cure model (1.3); when \( r = 1 \), we have

\[
\Lambda(t|z, x) = \exp(\beta'z)[1 - \exp(-\Lambda_0(t|x))]
\]

which is a proportional hazards cure model (1.10). Generally for any \( r > 0 \), we have,

\[
\lim_{t \to \infty} \Lambda(t|z, x) = \lim_{t \to \infty} \exp(\beta'z)\frac{1 - \exp(-r\Lambda_0(t|x))}{r} = \exp(\beta'z)/r < \infty,
\]

which means that the cumulative hazard function is improper and hence model (4.3)
results in a cure model; when $r \leq 0$, we have $\Lambda(t|z, x) = \exp(\beta' z)\Lambda_0(t|x)$ for $r = 0$ or

$$
\lim_{t \to \infty} \Lambda(t|z, x) = \lim_{t \to \infty} \exp(\beta' z) \frac{\exp[(-r)\Lambda_0(t|x)] - 1}{(-r)} = \infty, \text{ for } r < 0
$$

and model (4.3) results in a non-cure model. Therefore the proposed model (4.3) is either a cure model or a non-cure model depending on whether $r > 0$ or not. When $r > 0$, the cure rate is,

$$
\pi(z) = \lim_{t \to \infty} \exp(-\Lambda(t|x)) = \exp(-\exp(\beta' z)/r).
$$

which tends to 0 when $r \to 0^+$. Therefore $r$ acts as a bridge smoothly connecting the cure models and non-cure models.

It is easy to derive the population survival function, probability density function and hazard function,

$$
S(t|z, x) = \exp(-\exp(\beta' z)G_r(\Lambda_0(t|x))),
$$

$$
f(t|z, x) = S(t|x, z) \exp(\beta' z)G'_r(\Lambda_0(t|x))h_0(t|x), \tag{4.5}
$$

$$
h(t|z, x) = \exp(\beta' z)G'_r(\Lambda_0(t|x))h_0(t|x).
$$

where $G'_r(t) = \exp(-rt)$ denotes the first derivative of $G_r(t)$ with respect to $t$ and $h_0(t|x) = \Lambda'_0(t|x)$.

### 4.3 Parametric Estimation

We assume a Weibull distribution for $\Lambda_0(t|x)$ in this section. Other distributions, such as Gamma distribution, could be specified as well. The hazard function of Weibull
distribution is

\[ \Lambda_0(t|x) = \exp(\gamma' x)t^\alpha \]  

(4.6)

With this parametric assumption, model (4.3) is identifiable under the following regularity condition,

(A1) The covariates \( z \) and \( x \) are bounded with probability 1, and if there exist vector \( \beta \) and \( \gamma \) such that \( \beta'z = \gamma'x = 0 \) with probability 1, then \( \beta = \gamma = 0 \).

Condition (A1) is a common condition for a design matrix in linear regression setting and it is similar to the condition (C1) in Chapter 3.

**Proposition 3.** Under the parametric assumption (4.6) and regularity condition (A1), the transformation model (4.3) is identifiable.

**Proof.** Suppose that there exist \( \theta_1 = (\beta'_1, \gamma'_1, \alpha_1, r_1) \) and \( \theta_0 = (\beta'_0, \gamma'_0, \alpha_0, r_0) \) giving the same survival function (4.5), which is

\[
\exp \left\{ \exp(\beta'_1 z) \frac{\exp(-r_1 \exp(\gamma'_1 x)t^{\alpha_1}) - 1}{r_1} \right\} = \exp \left\{ \exp(\beta'_0 z) \frac{\exp(-r_0 \exp(\gamma'_0 x)t^{\alpha_0}) - 1}{r_0} \right\}
\]

or equivalently

\[
\exp(\beta'_1 z) \frac{\exp(-r_1 \exp(\gamma'_1 x)t^{\alpha_1}) - 1}{r_1} = \exp(\beta'_0 z) \frac{\exp(-r_0 \exp(\gamma'_0 x)t^{\alpha_0}) - 1}{r_0}.
\]

(4.7)

It is obvious that \( r_1 \) and \( r_0 \) share the same sign since positive \( r \) leads to a cure model while non-positive \( r \) leads to a non-cure model. When both \( r_1 \) and \( r_0 \) are positive, we have \( \exp(\beta'_1 z)/r_1 = \exp(\beta'_0 z)/r_0 \) by letting \( t \to \infty \), which could be written as,

\[
\exp((\beta_1 - \beta_0)'z) = r_1/r_0
\]
where the left hand side is a function of $z$ without intercept terms and the right hand side is a constant, therefore we have $(\beta_1 - \beta_0)'z = 0$ and $r_1 = r_0$. By condition (A1), we have $\beta_1 = \beta_0$. Plugging this result back into (4.7), we have $\exp(-r_1 \exp(\gamma_1'x)t^{\alpha_1}) = \exp(-r_0 \exp(\gamma_0'x)t^{\alpha_0})$. Hence we have

$$\exp(\gamma_1'x)t^{\alpha_1} = \exp(\gamma_0'x)t^{\alpha_0}$$

or

$$\exp((\gamma_1 - \gamma_0)'x) = t^{\alpha_0 - \alpha_1}. \quad (4.8)$$

Because the left hand side of (4.8) is a function of $x$ and the right hand side is a function of $t$, we have $\alpha_1 = \alpha_0$ and $\exp(\gamma_1'x) = \exp(\gamma_0'x)$, which gives $\gamma_1 - \gamma_0 = 0$ by condition (A1). The proof for the case when $r_1 > 0$ and $r_0 > 0$ is completed.

When one of $r_0$ and $r_1$ is zero, it is easy to see from (4.7) that the other one is zero too and then the identifiability problem reduces to the identifiability problem of the Cox proportional hazards model (1.3) with a specified baseline hazard function $\Lambda_0(t|x)$. It is well known that model (1.3) is identifiable and the proof is omitted here. When both $r_1$ and $r_0$ are negative, we have,

$$\exp((\beta_1 - \beta_0)'z) = \frac{r_1[\exp(-r_0 \exp(\gamma_0'x)t^{\alpha_0}) - 1]}{r_0[\exp(-r_1 \exp(\gamma_1'x)t^{\alpha_1}) - 1]}$$

Because the left hand side is a function of $z$ without intercept term and right hand side is independent of $z$, we have $(\beta_1 - \beta_0)'z = 0$ and hence $\beta_1 = \beta_0$ by (A1). Plugging this result into (4.7), we have, through some algebra,

$$r_0 \exp[r_0 \exp(\gamma_0'x)t^{\alpha_0} - r_1 \exp(\gamma_1'x)t^{\alpha_1}] = r_1 + (1/r_0 - 1/r_1) \exp[r_0 \exp(\gamma_0'x)t^{\alpha_0}]$$
Let \( t \to \infty \), the right hand side goes to \( r_1 \), a negative constant. Meanwhile, the left hand side goes to 0 when \( \alpha_0 > \alpha_1 \) or goes to \(-\infty\) when \( \alpha_0 < \alpha_1 \). Hence we have \( \alpha_0 = \alpha_1 \). Following the same argument, we have \( r_0 \exp(\gamma_0' x) = r_1 \exp(\gamma_1' x) \) in the left hand side of the above equation. Then the left hand side of the above equation is \( r_0 \) while the right hand side is \( r_1 \), which gives \( r_0 = r_1 \) and hence \( \exp(\gamma_0' x) = \exp(\gamma_1' x) \). Combined with condition \((A1)\), it is easy to see that \( \gamma_0 = \gamma_1 \). The proof is completed.

With the Weibull assumption \((4.6)\), model \((4.3)\) encompasses a very rich family of parametric models, including both cure models and non-cure models. This property is summarized in the following proposition.

**Proposition 4.** The transformation model \((4.3)\) with the assumption \((4.6)\) has the following models as its special cases,

I when \( r > 0 \), it is a cure model.

- when \( \alpha = 1 \), it is the generalized Gompertz cure model.

II when \( r \leq 0 \), it is a non-cure model.

- when \( r = 0 \), it is a proportional hazards model with Weibull baseline;
- when \( r < 0 \) and \( \alpha = 1 \), it is a proportional hazards model with Gompertz baseline;
- when \( r = -1 \), it is an exponential power distribution.

The generalized Gompertz cure model was proposed and studied by Gieser et al. (1998). Since model \((4.3)\) has Weibull model as its special case, it can be treated as a generalized Weibull distribution. Another type of generalized Weibull distribution has been studied by Mudholkar et al. (1996). However their model can only handle the
survival data without a cure fraction and it requires bounded support. Our proposed model (4.3), on the other hand, can handle both survival data with and without a cure fraction and has no bounded support problems.

In the survival analysis, hazard function presents rich information. In the following proposition, we summarize the hazard functions generated from model (4.3) under the Weibull assumption (4.6),

**Proposition 5.** Under the Weibull assumption (4.6), model (4.3) has the population hazard function

\[
h(t|z) = \exp(\beta'z + \gamma'x - r \exp(\gamma'x)t^\alpha)t^{\alpha-1}
\]

which is

a. monotone decreasing with bounded cumulative hazard function when \( r > 0 \) and \( \alpha \leq 1 \);

b. hump-shaped with bounded cumulative hazard function when \( r > 0 \) and \( \alpha > 1 \);

c. constant with unbounded cumulative hazard function when \( r = 0 \) and \( \alpha = 1 \);

d. monotone decreasing with unbounded cumulative hazard function when \( r = 0 \) and \( \alpha < 1 \);

e. monotone increasing with unbounded cumulative hazard function when \( r \leq 0 \) and \( \alpha \geq 1 \) (excluding case c.);

f. bathtub-shaped with unbounded cumulative hazard function when \( r < 0 \) and \( \alpha < 1 \).

For a graphical illustration, we plot the hazard functions in different combinations of parameters \( r \) and \( \alpha \) in Figure 4.1. In the figure, there are three curves representing
different values of $r$ and $\alpha$ within each case except case c. From Figure 4.1, we can see that the hazard function of model (4.3) with Weibull assumption (4.6) contains many common types of hazard functions.

Figure 4.1: The hazard functions with different combination of $r$ and $\alpha$
4.3.1 Likelihood Estimation

Since $\Lambda_0(t)$ in model (4.3) is specified by the Weibull distribution, we use the parametric maximum likelihood estimating method to get estimates of unknown parameters. Suppose $D = \{(t_i, \delta_i, z_i, x_i), i = 1, \cdots, n\}$ be the observed data (1.1), then the likelihood function is

\[
L(\theta) = \prod_{i=1}^{n} S(t_i|z_i, x_i)[h(t_i|z_i, x_i)]^{\delta_i}
\]

where $\theta = (\beta', \gamma', \alpha, r)$, denoting all unknown parameters. $S(t|z, x)$ is the population survival function and $h(t|z, x)$ is the population hazard function. The maximum likelihood estimate of $\theta$, denoted by $\hat{\theta}$, is obtained by maximizing the likelihood function $L(\theta)$. The Newton-Raphson algorithm can be used to maximize $L(\theta)$. Denote the true parameter of $\theta$ by $\theta_0 = (\beta'_0, \gamma'_0, \alpha_0, r_0)$, we have, under the mild regularity conditions (Casella and Berger, 2001),

\[
\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0, I^{-1}(\theta_0)), \text{ for } n \to \infty. \tag{4.9}
\]

where $I(\theta_0)$ is the Fisher information matrix and $\theta_0$ is usually replaced by $\hat{\theta}$ in $I(\theta)$ in real applications.

The asymptotic distribution given in equation (4.9) can be used to construct the approximate confidence interval or confidence region for individual parameters. For example, the sign of $r$ is of interest because it determines whether it is cure model or non-cure model and we need to test whether $r$ is positive or non-positive, which leads to the following hypothesis testing,

\[
H_0 : r \leq 0 \text{ versus } H_1 : r > 0. \tag{4.10}
\]
Suppose the standard error of $\hat{r}$ is $se_r$ and hence the one-sided accepting interval with 95% confidence level is $(-\infty, se_r \times z_{0.95})$, where $z_{0.95}$ is the 95% quantile of standard normal distribution. If $se_r \times z_{0.95} < \hat{r}$, $H_0$ is rejected and it is believed that there is a cure fraction and cure models will be more appropriate; If $se_r \times z_{0.95} > \hat{r}$, $H_0$ is not rejected and non-cure models are more suitable.

4.4 Semiparametric Estimation

Although parametric methods are efficient and easy to apply, they may result in misleading inference if the baseline model is misspecified. A nonparametric assumption of the baseline cumulative hazard function, on the other hand, will avoid the risk of misspecification. In this section, we consider a semi-parametric estimating approach based on a nonparametric assumption of $\Lambda_0(t)$ in model (4.3). However, it is computational expensive to incorporate $x$ into a nonparametrically specified $\Lambda_0(t)$. We will not consider covariates in $\Lambda_0(t)$ in this work.

By a similar argument used by Kalbfleisch and Prentice (2002), which has also been discussed in Chapter 3, we can show that a nonparametrically specified function $\Lambda_0(t)$ should be a right-continuous step function with jumps only at uncensored failure times. Specifically, let $t_{(1)} < t_{(2)} < \cdots < t_{(m)}$ be the observed failure times and $h_{0i}$ be the jump of $\Lambda_0(t)$ at $t_{(i)}$, we have

$$\Lambda_0(t) = \begin{cases} \sum_{i=1}^{n} h_{0i} I(t \geq t_{(i)}) & \text{if } t \leq t_{(m)} \\ \infty & \text{if } t > t_{(m)}, \end{cases} \quad (4.11)$$

where $I(A)$ is the indicator function: $I(A) = 1$ when event $A$ is true and $I(A) = 0$ otherwise. The setting that $\Lambda_0(t) = \infty$ when $t > t_{(m)}$ implies that subjects which are still alive beyond the largest uncensored failure time $t_{(m)}$ are treated as cured. Such
setting is commonly used in cure models. One may check Maller and Zhou (1992); Peng and Dear (2000); Sy and Taylor (2000) for more details. An alternative setting is defining a cure threshold: subjects still alive beyond the cure threshold are treated as cured. This setting is used in the transformation cure model in Zeng et al. (2006). It is obvious that when the cure threshold is defined at the largest uncensored failure time $t_{(m)}$, the resulted likelihood function is equivalent to the likelihood function derived in this section.

Since $\Lambda_0(t)$ is non-parametrically specified, we must consider the identifiability problem again. Suppose there are two parameter sets $(\beta_0, r_0, \Lambda_0(t))$ and $(\beta_0, 0, \Lambda_0^*(t))$ with $r_0 < 0$ and $\Lambda_0^*(t) = \left[ \exp(-r_0\Lambda_0(t)) - 1 \right]/(-r_0)$. Because $\Lambda_0(t)$ is a nonparametrically specified, right-continuous step function, $\Lambda_0^*(t)$ is also a nonparametrically specified, right-continuous step function with the same jump points as $\Lambda_0(t)$. It is also easy to show that $\Lambda_0^*(t)$ is a proper cumulative hazards function, i.e. $\Lambda_0^*(0) = 0$ and $\Lambda_0^*(\infty) = \infty$. Therefore $(\beta_0, 0, \Lambda_0^*(t))$ is a valid parameter set for model (4.3). Through some algebra, we can show that the survival functions under $(\beta_0, r_0, \Lambda_0(t))$ and $(\beta_0, 0, \Lambda_0^*(t))$ are identical. Hence model (4.3) is not identifiable when $r < 0$. Due to this reason, we assume that $r \geq 0$ in this section. We can show that the model (4.3) is identifiable when $r \geq 0$.

**Proposition 6.** When $r \geq 0$, model (4.3) is identifiable under condition (A1).

**Proof.** Suppose that there are two parameter sets, $(\beta, r, \Lambda_0(t))$ and $(\beta^*, r^*, \Lambda_0^*(t))$, which produce common survival functions (4.5), we have

$$
\exp \left( - \exp(\beta'z) \frac{1 - \exp(-r\Lambda_0(t))}{r} \right) = \exp \left( - \exp(\beta'^*z) \frac{1 - \exp(-r^*\Lambda_0^*(t))}{r^*} \right)
$$
or equivalently

\[ \exp(\beta'z) \frac{1 - \exp(-r\Lambda_0(t))}{r} = \exp(\beta^* z) \frac{1 - \exp(-r^*\Lambda_0^*(t))}{r^*} \]  \hspace{1cm} (4.12)

Suppose that both \( r \) and \( r^* \) are positive and let \( t \to \infty \), we have \( \exp(\beta'z)/r = \exp(\beta^* z)/r^* \), which gives

\[ \exp[(\beta - \beta^*)'z] = r/r^* \]

where the left hand side is a function of covariate \( z \) and the right hand side is a constant. Because there is no intercept term in \( z \), we have \( (\beta - \beta^*)'z = 0 \), which leads to \( \beta = \beta^* \) by condition (A1) and hence \( r/r^* = 1 \). Plugging these results into equation (4.12), we have \( \Lambda_0(t) = \Lambda_0^*(t) \). When \( r = 0 \), it is easy to see that \( r^* = 0 \) and hence model (4.3) becomes the Cox proportional hazards model (1.3) and it is well known that the model (1.3) is identifiable. The proof is completed. \( \square \)

The estimates are obtained by maximizing the log-likelihood with respect to \( \beta \) and \( r \) as well as to \( \Lambda_0(t) \). Because \( r > 0 \) leads to a cure model and \( r = 0 \) leads to a non-cure model, we present two estimating approaches depending on \( r \). When \( r = 0 \), the log-likelihood function is

\[ l(\beta, \Lambda_0(t)) = \sum_{j=1}^{m} \left[ \log(h_{0j}) + \beta'z_{(j)} \right] - \sum_{i=1}^{n} \left\{ \sum_{j=1}^{m} h_{0j}I(t_i \geq t_{(j)}) \exp(\beta'z_i) \right\} \]  \hspace{1cm} (4.13)

Taking the first derivative of \( l(\beta, \Lambda_0(t)) \) with respect to \( h_{0j} \) and setting it to zero, we have

\[ \hat{h}_{0j} = \left\{ \sum_{t_{i \geq t_{(j)}}} \exp(\beta'z_i) \right\}^{-1}. \]  \hspace{1cm} (4.14)
Substituting \( h_{0j} \) into (4.13), we have the profile likelihood

\[
l_{pr}(\beta, 0) = \sum_{j=1}^{n} \left\{ \beta' z_{(j)} - \log \left( \sum_{t_i \geq t_{(j)}} \exp(\beta' z_i) \right) \right\} - m \tag{4.15}
\]

which is the Cox's partial likelihood \( \text{(Cox, 1972, 1975)} \) by dropping the constant \( m \). Then the estimation of \( \beta \) could be obtained by maximizing the profile likelihood as discussed in \( \text{Cox (1972, 1975)} \).

When \( r > 0 \), the log-likelihood is

\[
l(\beta, r, \Lambda_0(t)) = \sum_{i=1}^{m} \left[ \log(h_{0i}) + \beta' z_{(i)} - r \sum_{j=1}^{m} h_{0j} I(t_{(i)} \geq t_{(j)}) \right] - \sum_{i=1}^{n} \exp(\beta' z_i) \frac{1 - \exp(-r \sum_{j=1}^{m} h_{0j} I(t_i \geq t_{(j)}))}{r} \tag{4.16}
\]

Taking the first derivative of \( l(\beta, r, \Lambda_0(t)) \) with respect to \( h_{0i} \) and setting it to zero, we have

\[
\hat{h}_{0i} = \frac{1}{r(m - i + 1) + \phi(h_{0i})}.
\]

where

\[
\phi(h_{0i}) = \sum_{i=1}^{n} \exp(\beta' z_i) \frac{\partial \left[ 1 - \exp(-r \sum_{j=1}^{m} h_{0j} I(t_i \geq t_{(j)})) \right]}{\partial h_{0i}} / r,
\]

which is the first derivative of the second term in the right hand side of equation (4.16) with respect to \( h_{0i} \). Because \( h_{0i} \) is unknown in \( \phi(h_{0i}) \), we use the following iterative equation in the estimation,

\[
\hat{h}_{0i}^{(k+1)} = \frac{1}{r(m - i + 1) + \phi(h_{0i}^{(k)})}, \quad i = 1, 2, \cdots, m. \tag{4.17}
\]
where \( k \) is the iteration number; \( \hat{h}_{0i}^{(k)} \) is the estimate of \( h_{0i} \) in the \( k \)th iteration. A similar technique can be found in Tsodikov (2003). Substituting \( \hat{h}_{0i} \) into (4.16), we have the profile likelihood \( l_{pr}(\beta, r) \) with unknown parameters \( \beta \) and \( r \). The estimates of \( \beta \) and \( r \) is then obtained by maximizing \( l_{pr}(\beta, r) \), which could be completed by some nonlinear maximizing method (Press et al., 2007). The estimating procedure in this section is summarized as follows:

1. Select the initial value for \( h_{0i}, i = 1, \cdots, m, \beta \) and \( r \). We use \( h_{0i} = 0, i = 1, \cdots, m \) and \( \beta = 0, r = 1 \) in the numerical studies.

2. In the \((k + 1)\)st iteration, we get \( \hat{h}_{0i}^{(k+1)} \) by equation (4.17) if \( \hat{r}^{(k)} > \delta \) and by equation (4.14) if \( \hat{r}^{(k)} < \delta \).

3. Given \( \hat{h}_{0i}^{(k+1)} \), we get \( \beta^{(k+1)} \) and \( \hat{r}^{(k+1)} \) by maximizing \( l_{pr}(\beta, r) \) if \( \hat{r}^{(k)} > \delta \) or \( l_{pr}(\beta, 0) \) if \( \hat{r}^{(k)} < \delta \).

4. Continue step 2 and step 3 until a predetermined convergence criterion is met.

Here \( \delta \) is the predetermined threshold. By maximizing \( l_{pr}(\beta, r) \), the resulted \( \hat{r} \) is always positive. Therefore we can simply solve equation (4.17) for \( \hat{h}_{0i} \) in step 2 and maximize \( l_{pr}(\beta, r) \) for \( \hat{\beta} \) and \( \hat{r} \) in step 3. From our numerical experiences, the estimates by the simplified estimation procedure are in agreement with those by the original estimation procedure (\( \delta = 0.005 \)), even though the true \( r = 0 \). This algorithm is stable and it usually takes fewer than 10 iteration steps to reach convergence according to our simulation studies.

We denote the nonparametric maximum likelihood estimations of \( \beta \) and \( r \) by \( \hat{\beta} \) and \( \hat{r} \). The asymptotic variance of \((\hat{\beta}, \hat{r})\) could be derived as the negative inverse of
the second derivation of \( l_{pr}(\beta, r) \) with respect to \( \beta \) and \( r \),

\[
- \left( \begin{array}{c}
\frac{\partial^2}{\partial \beta^2} l_{pr}(\beta, r) & \frac{\partial^2}{\partial \beta \partial r} l_{pr}(\beta, r) \\
\frac{\partial^2}{\partial r \partial \beta} l_{pr}(\beta, r) & \frac{\partial^2}{\partial r^2} l_{pr}(\beta, r)
\end{array} \right)^{-1} \bigg|_{\beta=\hat{\beta}, r=\hat{r}}
\] (4.18)

The justification of the variance estimation is based on the profile likelihood theory of Murphy and van der Vaart (2000).

4.5 Simulation Study

4.5.1 Parametric Estimation Method

In the first scenario, the survival function of the transformation model is

\[
S(t|z,x) = \exp \left[ -\exp(\beta_1 z) \frac{1 - \exp(-r \Lambda_0(t|x))}{r} \right]
\]

where \( \Lambda_0(t|x) = \exp(\gamma_0 + \gamma_1 x) t^\alpha \); \( z \) is generated from Bernoulli distribution with success probability 0.5; \( x \) is generated from uniform distribution on \((0,1)\); \( \beta_1 = 1.0, \gamma_0 = 0.5 \) and \( \gamma_1 = 1.0 \); \( \alpha \) varies from 0.5 to 1.5; \( r \) is selected from 1.0 to -0.5. The censoring time is generated from an exponential distribution \( F(t) = 1 - \exp(-0.2t) \), producing a censoring rate from 15\% to 40\%.

Table 4.1 summarizes the result from 1000 replications with sample size 200. From the table, we see that the biases of the estimates are small. The estimated standard error is comparable with the empirical error.

Because the sign of \( r \) determines the cure model or non-cure model, we report both the probability of type I error and the power of the hypothesis testing (4.10). Using the Wald test statistic and a rejection region from the normal distribution, the probability of type I error is calculated as the total number of rejecting \( H_0 \) divided by 1000 when \( H_0 \) is true while the power is calculated as the total number of accepting
Table 4.1: Estimations with Weibull model from 1000 replications

<table>
<thead>
<tr>
<th>True Parameter</th>
<th>$\beta_1$</th>
<th>$\gamma_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\gamma}_1$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\gamma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$r$</td>
<td>0.5</td>
<td>1.0</td>
<td>1.008</td>
<td>0.514</td>
<td>1.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.506</td>
<td>0.170</td>
<td>0.256</td>
<td>0.361</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>1.010</td>
<td>0.515</td>
<td>1.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.137</td>
<td>0.171</td>
<td>0.245</td>
<td>0.356</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5</td>
<td>1.0</td>
<td>1.011</td>
<td>0.514</td>
<td>1.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.137</td>
<td>0.173</td>
<td>0.244</td>
<td>0.356</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>1.013</td>
<td>0.510</td>
<td>1.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.080</td>
<td>0.160</td>
<td>0.232</td>
<td>0.334</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.5</td>
<td>1.014</td>
<td>0.510</td>
<td>1.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.068</td>
<td>0.161</td>
<td>0.221</td>
<td>0.329</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5</td>
<td>0.5</td>
<td>1.013</td>
<td>0.511</td>
<td>1.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.067</td>
<td>0.163</td>
<td>0.219</td>
<td>0.329</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.0</td>
<td>1.021</td>
<td>0.483</td>
<td>0.990</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.201</td>
<td>0.158</td>
<td>0.226</td>
<td>0.276</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.0</td>
<td>1.023</td>
<td>0.479</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.194</td>
<td>0.160</td>
<td>0.223</td>
<td>0.275</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5</td>
<td>0.0</td>
<td>1.024</td>
<td>0.485</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.190</td>
<td>0.161</td>
<td>0.226</td>
<td>0.277</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>−0.5</td>
<td>1.020</td>
<td>0.486</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.566</td>
<td>0.158</td>
<td>0.298</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>−0.5</td>
<td>1.012</td>
<td>0.489</td>
<td>1.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.574</td>
<td>0.160</td>
<td>0.301</td>
<td>0.228</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5</td>
<td>−0.5</td>
<td>1.019</td>
<td>0.486</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.579</td>
<td>0.161</td>
<td>0.303</td>
<td>0.230</td>
</tr>
</tbody>
</table>

$^a$ The first line denotes the empirical standard error based on 1000 replications

$^b$ The second line denotes the averages of the standard error by formula (4.9).
$H_0$ divided by 1000 when $H_0$ is false. For comparison, we also reported the coverage probability of the confidence interval of $\hat{r}$ containing true parameter $r$. The nominal confidence level is 95%. The results are summarized in Table 4.2

<table>
<thead>
<tr>
<th>$r$</th>
<th>$\alpha$</th>
<th>Probability of Type I error</th>
<th>Power</th>
<th>Coverage Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>0.5</td>
<td>-</td>
<td>1.000</td>
<td>0.947</td>
</tr>
<tr>
<td>1.0</td>
<td>-</td>
<td>1.000</td>
<td>0.944</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>-</td>
<td>1.000</td>
<td>0.949</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>-</td>
<td>0.997</td>
<td>0.963</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>-</td>
<td>0.999</td>
<td>0.954</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>-</td>
<td>0.999</td>
<td>0.961</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.5</td>
<td>0.058</td>
<td>-</td>
<td>0.960</td>
</tr>
<tr>
<td>1.0</td>
<td>-</td>
<td>0.042</td>
<td>-</td>
<td>0.967</td>
</tr>
<tr>
<td>1.5</td>
<td>-</td>
<td>0.043</td>
<td>-</td>
<td>0.971</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.5</td>
<td>0.000</td>
<td>-</td>
<td>0.944</td>
</tr>
<tr>
<td>1.0</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>0.953</td>
</tr>
<tr>
<td>1.5</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>0.957</td>
</tr>
</tbody>
</table>

From the table, we see that the coverage probability is close to the nominal level 95%. The power is close to 1 when $r = 1.0$ and $r = 0.5$. When $r = 0.0$, the type I error is close to the nominal level 5%. It shows that the proposed model has a reasonable power to differentiate cure models from non-cure models.

### 4.5.2 Semi-parametric Estimation Method

We now illustrate the performance of the proposed semi-parametric estimating method of Section 4.4. The survival function of the data is

$$S(t|z) = \exp \left[ - \exp(\beta'z) \frac{1 - \exp(-r\Lambda_0(t))}{r} \right]$$

where $\Lambda_0(t) = t$; $z$ is generated from Bernoulli distribution with equal probabilities. We select $\beta = -1$ or $\beta = 0$ and $r$ varies from 0 to 1. Table 4.3 summarizes the results from 1000 repetitions with sample size 200.
Table 4.3: Simulation results with semi-parametric method

<table>
<thead>
<tr>
<th>( \hat{\beta} )</th>
<th>( r )</th>
<th>( \hat{\beta} )</th>
<th>SE(^a)</th>
<th>ESE(^b)</th>
<th>( \hat{r} )</th>
<th>SE</th>
<th>ESE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.0</td>
<td>1.50</td>
<td>-1.022</td>
<td>0.385</td>
<td>0.388</td>
<td>1.565</td>
<td>0.341</td>
<td>0.344</td>
</tr>
<tr>
<td>-1.0</td>
<td>1.25</td>
<td>-1.023</td>
<td>0.359</td>
<td>0.364</td>
<td>1.300</td>
<td>0.268</td>
<td>0.272</td>
</tr>
<tr>
<td>-1.0</td>
<td>1.00</td>
<td>-1.021</td>
<td>0.329</td>
<td>0.326</td>
<td>1.029</td>
<td>0.200</td>
<td>0.200</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.75</td>
<td>-1.022</td>
<td>0.298</td>
<td>0.288</td>
<td>0.764</td>
<td>0.142</td>
<td>0.144</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.50</td>
<td>-1.012</td>
<td>0.268</td>
<td>0.259</td>
<td>0.501</td>
<td>0.094</td>
<td>0.096</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.25</td>
<td>-1.000</td>
<td>0.243</td>
<td>0.242</td>
<td>0.248</td>
<td>0.058</td>
<td>0.061</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.00</td>
<td>-0.959</td>
<td>0.236</td>
<td>0.234</td>
<td>0.028</td>
<td>0.040</td>
<td>0.045</td>
</tr>
<tr>
<td>0.0</td>
<td>1.50</td>
<td>0.004</td>
<td>0.296</td>
<td>0.288</td>
<td>1.562</td>
<td>0.338</td>
<td>0.344</td>
</tr>
<tr>
<td>0.0</td>
<td>1.25</td>
<td>0.009</td>
<td>0.279</td>
<td>0.273</td>
<td>1.295</td>
<td>0.265</td>
<td>0.268</td>
</tr>
<tr>
<td>0.0</td>
<td>1.00</td>
<td>0.011</td>
<td>0.260</td>
<td>0.252</td>
<td>1.025</td>
<td>0.197</td>
<td>0.198</td>
</tr>
<tr>
<td>0.0</td>
<td>0.75</td>
<td>0.004</td>
<td>0.242</td>
<td>0.238</td>
<td>0.760</td>
<td>0.140</td>
<td>0.140</td>
</tr>
<tr>
<td>0.0</td>
<td>0.50</td>
<td>-0.007</td>
<td>0.225</td>
<td>0.218</td>
<td>0.496</td>
<td>0.092</td>
<td>0.094</td>
</tr>
<tr>
<td>0.0</td>
<td>0.25</td>
<td>0.002</td>
<td>0.215</td>
<td>0.216</td>
<td>0.238</td>
<td>0.064</td>
<td>0.071</td>
</tr>
<tr>
<td>0.0</td>
<td>0.00</td>
<td>0.001</td>
<td>0.212</td>
<td>0.214</td>
<td>0.027</td>
<td>0.057</td>
<td>0.065</td>
</tr>
</tbody>
</table>

\(^a\) SE denotes the average of the standard errors by the variance estimation (4.18)

\(^b\) ESE denotes the empirical standard error based on 1000 replications.

From the table, the biases of the estimates are reasonable. The estimated standard errors by profile variance estimation are comparable with the empirical standard errors. From the simulation results, we conclude that the proposed nonparametric maximum likelihood estimating method yields estimates with small biases in small samples.

### 4.6 Real Examples

#### 4.6.1 Treatment Effect Analysis for Colon Cancer Patients — Revisit

As an illustration, we re-analyzed the colon cancer data by the proposed estimating methods. The colon cancer data has been studied in Section 3.5 and detailed information about the data can be found there. We still focus on the effect of the levamisole combined with fluorouracil treatment on the survival time, the time to cancer recurrence, and the cure rate.
Firstly, we fit the data by the parametric likelihood method of Section 4.3. The covariate associated with $\beta_1$ is treatment status, 1 for levamisole combined with fluorouracil treatment and 0 otherwise. The estimates are summarized in Table 4.4. The estimated cure rates are 0.417 and 0.600. It is seen that $\beta_1 < 0$ and $r > 0$ and hence there exist cure fractions in both groups and the treatment has negative effect on the survival time and the cure rate.

Secondly, we fit the data by the semi-parametric method of Section 4.4. The result is also summarized in Table 4.4. The estimated cure rates are 0.417 in control group and 0.601 in treatment group, respectively. It is seen that the estimates by the semi-parametric method are comparable with the estimates by the parametric method. We conclude that the levamisole combined with fluorouracil treatment has effect of prolonging the population survival time and increasing the cure rate. The result agrees with the result in Section 3.5 and Moertel et al. (1990, 1995), etc.

Table 4.4: Parameter estimations in colon cancer data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parametric Method</th>
<th>Semi-parametric Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimation</td>
<td>SE a</td>
</tr>
<tr>
<td>$r$</td>
<td>1.143</td>
<td>0.091</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$-0.524$</td>
<td>0.123</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>$-1.049$</td>
<td>0.110</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1.240</td>
<td>0.060</td>
</tr>
</tbody>
</table>

*a standard error by respective method
b 95% confidence interval using SE.

For a graphical illustration of the analysis on the colon cancer data, we plot the estimated survival curves by the proposed methods along with the Kaplan-Meier survival curves in Figure 4.2. It is seen that the estimated survival curves by the proposed methods are very close to the Kaplan-Meier curves.
4.6.2 Effect of Performance Status in the Veterans’ Administration Lung Cancer Trial

As a second sample, we consider the survival data from the veterans’ administration lung cancer trial (Prentice, 1973). Only 97 patients without prior therapy and one covariate, the performance status measured on a scale from 0 to 100, are considered.
here. In the dataset, 6 survival times are censored. The subset of data for the 97 patients without prior therapy has been analyzed as a survival dataset without a cure fraction by many authors, including Bennett (1983); Cheng et al. (1995); Pettitt (1984); Murphy et al. (1997); Chen et al. (2002) and Zeng and Lin (2006). We divided the data set into two groups, the high performance group contains the patients with performance status \( \geq 50 \) and the lower performance group contains the remaining patients. The Kaplan-Meier survival curves of two groups were plotted in Figure 4.3. From the Kaplan-Meier curves, we can see that there does not exist a long and stable plateau in the tails with heavy censoring as the Kaplan-Meier curves of colon cancer data showed in the previous example. Therefore, it is most likely that there is no cure fractions in both groups.

We fit the data by both the parametric method and the semi-parametric method. The results are summarized in Table 4.5. From the table, it is seen that \( \beta_1 \) is significantly negative by both methods and we conclude that the high performance status prolongs the survival time. The estimated cure rates of the two groups by both methods are less than 0.00001. By the result, we conclude that there is no potential cure fractions in the veterans' administration lung cancer trial data, which agrees with other studies (Prentice, 1973; Zeng and Lin, 2006).

Table 4.5: Parameter estimations in lung cancer data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parametric Method</th>
<th>Semi-parametric Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimation</td>
<td>SE</td>
</tr>
<tr>
<td>( r )</td>
<td>0.023</td>
<td>0.095</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>(-0.763)</td>
<td>0.123</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>(-4.200)</td>
<td>0.110</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.999</td>
<td>0.060</td>
</tr>
</tbody>
</table>

\(^a\) standard error by respective method  
\(^b\) 95% confidence interval using SE.  
\(^c\) confidence interval left truncated at 0.0.

We plot the estimated survival curves by the proposed methods along with the
Kaplan-Meier survival curves in Figure 4.3 for a graphical illustration. It is seen that the estimated survival curves by the proposed methods are close to the Kaplan-Meier curves.

Figure 4.3: Kaplan-Meier curves and predicted survival curves of the lung cancer trial data
4.7 Discussion

Non-cure models, for the survival data without a cure fraction, and cure models, for the survival data with a cure fraction, are extensively studied in the last two decades. However, very little work has been devoted to a unified model for both types of data. We proposed a new transformation model for both the survival data without a cure fraction and the survival data with a cure fraction. The transformation parameter, which determines the type of the model, cure model or non-cure model, is estimated from data. Two estimating methods, one parametric method and one semi-parametric method, were presented. Our numerical studies show that both methods provide good estimations for both types of datasets.

It is worth noting that the parametric model with a Weibull assumption (4.6) does not have proportional hazards for different covariate values of $z$ and $x$, while the semi-parametric model with (4.11) still has the proportional hazards structure because $z$ does not appear in $\Lambda(t)$. 
Chapter 5

Length-biased Data Analysis

5.1 Introduction

In observational studies, one may encounter samples where the subjects are not randomly sampled but subject to length-biased sampling scheme. In one special case, the probability of observing a length-biased failure time $t$ is proportional to $t$ itself. Many methodologies have been developed to remove the bias caused by length-biased sampling. Specifically, let $f(t)$ be the density of the target population and let $g(t)$ be the density of the length-biased data. Cox (1969) proved that $g(t) = tf(t)/\mu$, where $\mu$ is the mean time of the unbiased data. It is seen that it causes biased inference if we use the density of the biased sample as the estimate of the density of the original sample. Cox’s result is widely used in the literature. However, when the survival time from the target population has a cure fraction, Cox’s result is no longer valid because we have $\mu = \infty$ as the density function $f(t)$ is improper, which leads to $g(t) = 0$.

In this chapter, we consider a weighted method to adjust the bias in the length-biased data and provide an unbiased estimation of both population survival function and cure rate for the survival data with a cure fraction. A weighted method has been proposed for length-biased data analysis in literature. Horvitz and Thompson (1952) proposed a weighted method to produce an unbiased estimate of population mean.
Binder (1992) and Lin (2000) proposed weighted methods combined with the Cox proportional hazards model for length-biased survival data without a cure fraction. In their methods, the so-called inverse-probability-of-selection weight is supposed to be known. Recently, Pan and Schaubel (2008) proposed another weighted method where the weights were empirically estimated using the original sample, the sample before the length-biased sampling. However the original sample is not always available in real application. In this chapter, we propose a weighted method for length-biased data with a cure fraction, where the weight is empirically estimated using the given length-biased sample only.

5.2 Length-Biased Data

Suppose $T^*$ be the failure time in the original sample and $T^*$ is assumed to follow the mixture cure model (1.9) with known cure status $g$: $g = 1$ if the subject is cured and 0 otherwise. We also define a cure threshold $u$: $Pr(g = 1|T^* = u) = 1$ and $Pr(T^* < u|g = 0) = 1$ and denote $D = \{(T^*_i, g_i) : i = 1, \cdots, N\}$ as the original sample. The procedure of length-biased sampling is presented as follows: suppose $A_i$ and $C_i$ are two random variables independent with $T^*_i$,

- Select $T^*_i$ if $T^*_i > A_i$ and drop $T^*_i$ if $T^*_i < A_i$;
- $T^*$ is observed when $T^*_i < A_i + C_i$ with $\delta_i = 1$ and $T_i = T^*_i$;
- $T^*$ is censored when $T^*_i > A_i + C_i$ with $\delta_i = 0$ and $T_i = A_i + C_i$;

The length-biased sample is denoted by $D_{lb} = \{(T_i, \delta_i, g_i), i = 1, \cdots, n.\}$, where $n < N$ is the number of $T^*_i$'s being selected into the sample. Our goal is to estimate the survival function and the cure rate of the original sample $D$ through $D_{lb}$.
5.3 Weighted Model

5.3.1 Uncensored Case

We start with the case when no censoring is presented, i.e. \( C_i = \infty \) for \( i = 1, \cdots, N \). In this case, we can rewrite the length-biased sample \( D_{lb} \) as \( D_{lb}^* = \{(T_{i}^*, g_i, s_i) : i = 1, \cdots, N\} \), where \( s_i \) is the selection indicator, \( s_i = 1 \) if the \( i \)th subject is selected into the length-biased sample and 0 otherwise. It is obvious that \( n = \sum_{i=1}^{N} s_i \). Under the length-biased sampling scheme and given \( T_{i}^* \), we have \( \Pr(s_i = 1) \propto T_{i}^* \). Suppose \( \Pr(s_i = 1) = T_{i}^*/\alpha \), where \( \alpha > u \) is a constant. By the assumption that \( T_{i}^* \) follows the mixture cure model (1.9), the observed likelihood can be constructed as (Laska and Meisner, 1992),

\[
L = \prod_{i=1}^{N} \pi^{g_i s_i} \left[(1-\pi)\Delta S_0(T_{i}^*)^{(1-g_i)s_i}\right] \quad (5.1)
\]

where \( \Delta S_0(t) \) is the density function or the probability mass function of \( T_{i}^* \) depending on whether \( T_{i}^* \) is discrete or continuous. In this chapter, we assume that \( S_0(t) \) is a right-continuous step function. It is easy to show that the expectation of the logarithm of the likelihood function \( L \) with respect to \( s_i \) is

\[
E(\log L) = E\left(\sum_{i=1}^{N} \log \left\{ \pi^{c_i s_i} \left[(1-\pi)\Delta S_0(T_{i}^*)^{(1-c_i)s_i}\right]\right\}\right) \\
= \sum_{i=1}^{N} \log \left\{ \pi^{c_i} \left[(1-\pi)\Delta S_0(T_{i}^*)\right]^{1-c_i}\right\} \frac{T_{i}^*}{\alpha} \quad (5.2)
\]

Compared to the log-likelihood function of the original sample \( D \), which is

\[
\sum_{i=1}^{N} \log \left\{ \pi^{c_i} \left[(1-\pi)\Delta S_0(T_{i}^*)\right]^{1-c_i}\right\} \quad (5.3)
\]
the log-likelihood function of (5.1) is biased. Following Little and Rubin (2002), we weight each subject by the inverse of the inclusion probability. Because the inclusion probability is \( Pr(s_i = 1) = T_i^*/\alpha \) and \( \alpha \) is a constant, we use the weight \( w_i = 1/T_i^* \). The resulted weighted log-likelihood function is

\[
l_w = \sum_{i=1}^{N} \log \left\{ \pi^{c_i s_i} ((1 - \pi) \Delta S_0(T_i))^{(1-c_i)s_i} \right\} / T_i^*
\]

(5.4)

It is easy to show that \( E(l_w) \) is identical to the log-likelihood (5.3) by suppressing the constant \( \alpha \) and hence the weighted log-likelihood (5.4) is unbiased.

5.3.2 Censored Case

When censoring is present, the true failure times of censored subjects are unknown except for the subjects which are censored at the cure threshold \( u \). We need to estimate the true failure time of the censored subjects. Suppose there are \( m \) observed cured subjects in \( D_{lb} \) and the weighted log-likelihood function is

\[
l_w = w_u m \log(\pi) + \sum_{i=1}^{n-m} w_i [\delta_i \log((1 - \pi) \Delta S_0(t_i)) + (1 - \delta_i) \log(\pi + (1 - \pi) S_0(t_i))] \]

(5.5)

where \( w_u = 1/u \) since the “failure” times of the cured subjects are \( u \); \( w_i = 1/T_i \) when \( \delta_i = 1 \) and \( w_i = 1/\hat{T}_i \) when \( \delta_i = 0 \), where \( \hat{T}_i \) is an unbiased estimate of the true failure time of the \( i \)-th censored subject. Since we only have the length-biased data \( D_{lb} \), the estimate of \( T_i \) must be estimated from \( D_{lb} \). A natural estimate of \( T_i \) is

\[
\hat{T}_i = \int_{T_i}^{\infty} td(-S_{lb}(t|t > T_i)) = \int_{T_i}^{\infty} td(-S_{lb}(t))/S_{lb}(T_i),
\]
where \( S_{lb}(t) \) is the survival function of \( D_{lb} \); Following Maller and Zhou (1992), subjects who are still alive beyond \( T_{(n)} \) can be treated as cured and hence we rewrite the estimate \( \hat{T}_i \) as

\[
\hat{T}_i = \left[ \int_{T_i}^{T_{(n)}} td(-S_{lb}(t)) + \pi_{lb} \times u \right] S_{lb}^{-1}(T_i),
\]

(5.6)

where \( T_{(n)} = \max_i \{ T_i : \delta_i = 1, i = 1, \cdots, n \} \) is the largest uncensored failure time in \( D_{lb} \); \( \pi_{lb} \) is the cure rate in \( D_{lb} \); \( u \) is the cure threshold. A natural estimation of \( S_{lb}(t) \) is the Kaplan-Meier estimation \( \hat{S}_{KM}^{lb}(t) \) in \( D_{lb} \). Following Wang (1987), \( \hat{S}_{KM}^{lb}(t) \to S_{lb}(t) \) uniformly on \((0, T_{(n)})\); Following Maller and Zhou (1992), the cure rate \( \pi_{lb} \) could be consistently estimated by \( \hat{S}_{KM}^{lb}(T_{(n)}) \). Therefore, we have

\[
\hat{T}_i = \left[ \int_{T_i}^{T_{(n)}} td(-\hat{S}_{KM}^{lb}(t)) + \hat{\pi} \times u \right] [\hat{S}_{KM}^{lb}]^{-1}(T_i),
\]

When \( u \) is unknown, we use the maximum failure time \( T_n = \max(T_1, \cdots, T_n) \) of \( D_{lb} \) as the estimate of \( u \).

5.3.3 Maximum Likelihood Estimations

After we get \( \hat{T}_i \)'s, the estimates of the survival function \( S_0(t) \) and the cure rate \( \pi \), as well as the population survival function \( S_p(t) = \pi + (1-\pi)S_0(t) \), of the original sample \( D \) is obtained by maximizing the likelihood function (5.5), which is summarized in the following theorem.

**Theorem 5.3.1.** Suppose that in the length-biased sample \( D_{lb} = \{(T_i, \delta_i, g_i), i = 1, \cdots, n\} \), \( m \) subjects are cured with known cure status. Then the maximum likelihood estimates are given as

\[
\hat{S}_{wp}(t) = \prod_{t_i \leq t} \left( 1 - \frac{\delta_i w_i}{\sum_{j: t_j \geq t} w_j + w_u m} \right)
\]
\[ \hat{\pi}_w = \prod_{i=1}^{n-m} \left( 1 - \frac{\delta_i w_i}{\sum_{j:t_j \geq t_i} w_j + u w_m} \right) \]  
\[ \hat{S}_{w0}(t) = (\hat{S}_{wp}(t) - \hat{\pi})/(1 - \hat{\pi}) \]

where \( \hat{S}_{wp}(t) \) is the estimate of the population survival function, \( \hat{\pi} \) is the estimate of cure rate and \( \hat{S}_{w0}(t) \) is the estimate of \( S_0(t) \) in the original sample \( D \).

**Proof.** Following a similar argument in Kalbfleisch and Prentice (2002), we assume that \( S_0(t) \) of the original sample is a right-continuous step function with jumps only at uncensored failure times. Suppose \( t_1 \leq \cdots \leq t_{n-m} < u \) be the sorted uncensored failure times in increasing order and \( \Delta S_0(t_i) \)'s be the corresponding jumps. To maximize the log-likelihood function (5.5), we write

\[ p_i = (1 - \pi) \Delta S_0(t_i) \quad \text{for} \quad i = 1, \cdots, n-m, \]

then \( \pi = 1 - \sum_{j=1}^{n-m} p_j \) and \( \pi + (1 - \pi) S_0(t_i) = 1 - \sum_{j=1}^{i} p_j \) and we have

\[ l_w = w_u m \log(1 - \sum_{j=1}^{n-m} p_j) + \sum_{i=1}^{n-m} \left\{ w_i \delta_i \log(p_i) + w_i (1 - \delta_i) \log(1 - \sum_{j=1}^{i} p_j) \right\} \]  
\[ (5.8) \]

by solving \( \frac{\partial l_w}{\partial p_{n-m}} = 0 \), we have

\[ p_{n-m} = \frac{\delta_{n-m} w_{n-m}}{m w_u + w_{n-m}} (1 - \sum_{j=1}^{n-m-1} p_j). \]

Plugging the estimation of \( p_{n-m} \) into equation (5.8), we have, through some algebra,

\[ l_w = (w_u m + w_{n-m}) \log(1 - \sum_{j=1}^{n-m-1} p_j) + \sum_{i=1}^{n-m-1} \left\{ w_i \delta_i \log(p_i) + w_i (1 - \delta_i) \log(1 - \sum_{j=1}^{i} p_j) \right\} + a \]

where \( a \) is a constant independent of \( p_i \)'s, \( i = 1, \cdots, n-m-1 \). Recursively solving
\[ \partial l_w / \partial p_j = 0, \text{ we have} \]

\[ p_j = \frac{\delta_j w_j}{m w_u + \sum_{k=j}^{n-m} w_k} \left(1 - \sum_{k=1}^{j-1} p_k\right), \ j = n - m - 1, \cdots, 1 \]

where \( \sum_{k=1}^{0} p_k = 0 \). By \( \hat{S}_{wp}(t) = 1 - \sum_{i \leq t} p_i \) and \( \hat{\pi}_w = 1 - \sum_{i=1}^{n-m} p_i \), we have equation (5.7). The proof is completed.

Similar results of estimations of survival function and cure rate, without weight function, can be found in Laska and Meisner (1992). Our derivation is slightly different from theirs. In Laska and Meisner (1992), the authors assumed that the survival function is left continuous, i.e. \( S(t) = Pr(T \geq t) \). Our derivation assumes that the survival function is right continuous, i.e. \( S(t) = Pr(T > t) \), which is more popular in literature (Kalbfleisch and Prentice, 2002; Tsodikov, 2003; Zeng et al., 2006). When the weights are compressed, our estimates reduce to Laska and Meisner’s result with a slight difference in expression because of the different assumptions of \( S(t) \); when there is no cure fractions, our estimate reduces to the weighted Kaplan-Meier estimate (Satten and Datta, 2001); when both weights and cure fraction are absent, our estimate reduces to the standard Kaplan-Meier estimate (Kaplan and Meier, 1958).

The estimation of the variance of \( \hat{\pi} \) can be derived by Greenwood’s formula (Greenwood, 1926),

\[ \hat{V}(\hat{\pi}_w) = \hat{\pi}_w^2 \sum_{i=1}^{n-m} \frac{\delta_i w_i}{\sum_{k \geq i} w_k \sum_{k > i} w_k} \]

(5.9)

### 5.4 Simulation Study

We conducted a simulation study to illustrate the performance of the proposed method. In the simulation study, the original data was generated from the mixture cure model (1.9) with cure rates equaling 0.1, 0.2, 0.3 or 0.4. The “failure time”s
of cured subjects are set at the cure threshold \( u = 18 \). The failure times of uncured subjects are generated from the exponential distribution \( \exp(0.5t) \). After \( N \) failure times are generated, we generate the truncation time \( A_i, i = 1, \cdots, N \), from uniform \( (0, u + 1) \) and compare the failure time \( \tilde{T}_i \) with \( A_i \). We only kept the failure times which were greater than the truncation times. The sample size after truncation is denoted by \( n \). Next, we generated censoring times \( C_i \) from \( \text{unif}(5, 12) \) for \( i = 1, \cdots, n \). If \( \tilde{T}_i < A_i + C_i \), the \( i \)-th subject is observed with \( \delta_i = 1 \) and the observed failure time is \( T_i = \tilde{T}_i \); If \( \tilde{T}_i > A_i + C_i \), the \( i \)-th subject is censored with \( \delta_i = 0 \) and the observed censored time is \( T_i = A_i + C_i \). Here \( i \) runs from 1 to \( n \). The generated length-biased data is \( D_{lb} = \{(T_i, \delta_i, c_i), i = 1, \cdots, n\} \). The cure status \( c_i \) is unknown when the \( i \)-th subject is censored and \( T_i < u \); when \( \delta_i = 1 \) and \( T_i < u, c_i = 0 \); when \( T_i = u, c_i = 1 \).

For each scenario, we consider different sample sizes, \( N = 500, 1000 \) and \( 1500 \) for the original sample. For each case, we generated 1000 independent simulation data sets. For illustration purpose, we recorded the size \( n \) of the length-biased sample as well as the censoring rate in each case. For comparison purposes, we also reported the unweighted estimation of the cure rate, \( \hat{\pi}_u \), by Maller and Zhou’s method using the length-biased sample. The result is summarized in Table 5.1.

From the table, it is seen that the weighted estimations are pretty close to the true cure rates in all scenarios. By comparison, the estimate of the cure rate without any adjustments, \( \hat{\pi}_u \), is far from the true cure rate.
Table 5.1: Simulation results of length-biased data based on 1000 replications

<table>
<thead>
<tr>
<th>π</th>
<th>N&lt;sup&gt;a&lt;/sup&gt;</th>
<th>n&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Censoring %</th>
<th>( \hat{\pi}_w )</th>
<th>( \hat{\pi}_u )&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>500</td>
<td>95.3</td>
<td>0.511</td>
<td>0.107</td>
<td>0.504</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>189.8</td>
<td>0.510</td>
<td>0.104</td>
<td>0.504</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>284.2</td>
<td>0.511</td>
<td>0.103</td>
<td>0.504</td>
</tr>
<tr>
<td>0.2</td>
<td>500</td>
<td>137.4</td>
<td>0.701</td>
<td>0.215</td>
<td>0.697</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>273.7</td>
<td>0.699</td>
<td>0.208</td>
<td>0.696</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>409.6</td>
<td>0.699</td>
<td>0.204</td>
<td>0.695</td>
</tr>
<tr>
<td>0.3</td>
<td>500</td>
<td>179.0</td>
<td>0.799</td>
<td>0.317</td>
<td>0.796</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>358.9</td>
<td>0.798</td>
<td>0.309</td>
<td>0.796</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>537.0</td>
<td>0.799</td>
<td>0.307</td>
<td>0.796</td>
</tr>
<tr>
<td>0.4</td>
<td>500</td>
<td>221.0</td>
<td>0.859</td>
<td>0.416</td>
<td>0.858</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>443.2</td>
<td>0.861</td>
<td>0.412</td>
<td>0.859</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>663.4</td>
<td>0.860</td>
<td>0.408</td>
<td>0.858</td>
</tr>
</tbody>
</table>

<sup>a</sup> the original sample size (before truncation)

<sup>b</sup> the length-biased sample size (after truncation)

<sup>c</sup> the estimates of cure rate without adjustments
Chapter 6

Proportional Density Model

6.1 Introduction

In survival analysis, two-sample data is very common, especially in clinical trials. For example, when we compare a new treatment with an existing treatment, or when we want to test the effect of a new drug, two sample data arises. A so-called proportional density model has been proposed in the literature for two sample data,

\[-dS_g(t) = \exp(\alpha + \beta h(t))[-dS_f(t)]\]  

(6.1)

where \(-dS_f(t)\) and \(-dS_g(t)\) are the probability density functions or probability mass functions of two groups, respectively. One group is referred to as control group, corresponding to \(-dS_f(t)\), and the other group as treatment group, corresponding to \(-dS_g(t)\). \(h(t)\) is a known function. When no cure fraction is presented, \(dS_f(t)\) and \(dS_g(t)\) are both proper, i.e. \(\int d[-S_f(t)] = \int d[-S_g(t)] = 1\). Recently, Shen et al. (2007) studied model (6.1) for the case when a cure fraction is present. They assumed that the failure times in both groups follow a mixture cure model (1.9), which leads
to the following model

\[
\begin{align*}
S_f(t) &= 1 - \pi_0 + \pi_0 S_{f0}(t) \\
S_g(t) &= 1 - \pi_1 + \pi_1 S_{g0}(t) \\
dS_{g0}(t) &= \exp(\alpha + \beta h(t))dS_{f0}(t)
\end{align*}
\] (6.2)

where \( S_f(t) \) and \( S_g(t) \) are the population survival functions of respective groups; \( S_{f0} \) and \( S_{g0} \) are the survival functions for the uncured patients in respective groups. To estimate the parameters \( \alpha \) and \( \beta \), the authors first used the method proposed by Maller and Zhou (1992) to estimate the cure rate \( \pi_0 \) and \( \pi_1 \) and then used a nonparametric right-continuous step function to specify \( S_{f0}(t) \), which leads to the involvement of the empirical likelihood method.

The empirical likelihood method was proposed by Owen (1988, 1990) as a non-parametric method. Compared with parametric likelihood method, the empirical likelihood method does not require distributional assumptions and still has similar asymptotic properties as the parametric likelihood. It has attracted much interest in various fields of statistics in the literature since its origin. A good introduction of empirical likelihood can be found in Owen (2001).

An important issue of model (6.1) or model (6.2) is the selection of \( h(t) \). Various models have been proposed to model \( h(t) \) in the literature. Anderson (1979) studied model (6.1) by assuming \( h(t) = t \); Cox (1969) and Wang (1996) studied model (6.1) with the assumption of \( h(t) = \log(t) \) and \( \beta = 1 \). Kay and Little (1987) discussed several choices of \( h(t) \). For survival data with a cure fraction, Shen et al. (2007) simply assumed that \( h(t) \) is known in model (6.2) and used a polynomial in their numerical studies. Recent applications have shown that model (6.1) or model (6.2) with a proper choice of \( h(t) \) provides a good fit to the data (Qin and Zhang, 1997; Qin and Leung. 91
However, selection of a proper $h(t)$ is still a challenge. In this chapter, we propose a new transformation model for $h(t)$. The transformation parameter of the model will be determined by data and it includes two popular settings of $h(t)$ as its special cases.

### 6.2 Model

Based on model (6.2), we impose the Box-Cox transformation model on $h(t)$ with unknown parameter $\lambda$, which leads to

\[
\begin{align*}
S_f(t) &= 1 - \pi_0 + \pi_0 S_{f_0}(t) \\
S_g(t) &= 1 - \pi_1 + \pi_1 S_{g_0}(t) \\
dS_{g_0}(t) &= \exp \left[ \alpha + \beta h(t, \lambda) \right] dS_{f_0}(t)
\end{align*}
\]

(6.3)

where

\[
h(t, \lambda) = \begin{cases} 
(t^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\
\log(t) & \text{if } \lambda = 0.
\end{cases}
\]

It is easy to see that $h(t, 1) = t$ and $h(t, 0) = \log t$, hence the proposed model includes two popular settings of $h(t)$ as it special cases. Unlike the other modeling of $h(t)$, the transformation parameter $\lambda$ is estimated from the given data in our model (6.3) and hence model misspecification could be avoided.

To estimate $\alpha$, $\beta$ and $\lambda$ in model (6.3), we use the empirical likelihood technique. Let $dF$ and $dG$ denote the probability density functions or probability mass functions of the uncensored failure time in the control group and the treatment group,
respectively. That is
\[ dF(t) = \Pr(T = t|\delta = 1) = \frac{\Pr(T = t, U > t)}{\Pr(T < U)} \]

where \( U \) is the censoring variable. Under the independent censoring scheme, we have
\[ \Pr(T = t, U > t) = \Pr(T = t) \times \Pr(U > t) = \pi_0 dS_{f0}(t) \times \bar{H}_0(t) \]

and
\[ \Pr(T < U) = \int_{0}^{\infty} \pi_0 dS_{f0}(t) \int_{t}^{\infty} (-d\bar{H}_0(v)) = \int_{0}^{\infty} \pi_0 \bar{H}_0(t)dS_{f0}(t) \]

where \( \bar{H}_0(t) \) is the survival function of the censoring variable in the control group. Hence we have
\[ dF(t) = \frac{\bar{H}_0(t)dS_{f0}(t)}{\int_{0}^{\infty} \pi_0 \bar{H}_0(v)dS_{f0}(v)} = \frac{\bar{H}_0(t)dS_{f0}(t)}{\mu(dS_{f0}, \bar{H}_0)} \]

where \( \mu(dS_{f0}, \bar{H}_0) = \int_{0}^{\infty} \bar{H}_0(v)dS_{f0}(v) \) is a constant with respect to time \( t \). Similarly in the treatment group, we have
\[ dG(t) = \frac{\bar{H}_1(t)dS_{g0}(t)}{\int_{0}^{\infty} \pi_1 \bar{H}_1(v)dS_{g0}(v)} = \frac{\bar{H}_1(t)dS_{g0}(t)}{\mu(dS_{g0}, \bar{H}_1)} \]

where \( \mu(dS_{g0}, \bar{H}_1) = \int_{0}^{\infty} \bar{H}_1(v)dS_{g0}(v) \) and \( \bar{H}_1 \) is the survival function of the censoring variable in the treatment group. By model (6.3), we have
\[ \frac{dG(y)}{dF(y)} = \frac{\bar{H}_1(y)\mu(dS_{f0}, \bar{H}_0)dS_{g0}(y)}{\bar{H}_0(y)\mu(dS_{g0}, \bar{H}_1)dS_{f0}(y)} \]
\[ = \exp(\alpha^* + \beta h(y, \lambda) + \psi(y)) \]
where \( \alpha^* = \alpha + \log(\mu(dS_{f0}, \bar{H}_0)) - \log(\mu(dS_{g0}, \bar{H}_1)) \) and \( \psi(y) = \log(\bar{H}_1(y)) - \log(\bar{H}_0(y)) \). It is worth noting that \( \psi(y) = 0 \) when the censoring distributions of the two groups are identical.

Suppose \( x_1, \cdots, x_{n_0} \) are the sample of control group and \( x_{n_0+1}, \cdots, x_{n_0+n_1} \) are the sample of treatment group; \( y_1, \cdots, y_{m_0} \) and \( y_{m_0+1}, \cdots, y_{m_0+m_1} \) are the uncensored samples from each group. Let \( m = m_0 + m_1 \) and \( n = n_0 + n_1 \). By applying the empirical likelihood technique, we have the log-likelihood function conditional on the uncensored sample as

\[
l_0(\alpha, \beta, \lambda, F) = \sum_{i=1}^{m_0} \log dF(y_i) + \sum_{j=m_0+1}^{m} \left( \log dF(y_j) + \alpha + \beta h(y_j, \lambda) + \psi(y_j) \right)
\]

subject to the constraint that \( \sum_{i=1}^{m} p_i = 1 \), \( p_i \geq 0 \) and \( \sum_{i=1}^{m} q_i = 1 \) where \( p_i = dF_1(y_i) \) and \( q_i = p_i \exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i)) \). Using Lagrange multipliers, the likelihood function is written as

\[
l_0(\alpha, \beta, \lambda, F_1) = \sum_{i=1}^{m} \log p_i + \sum_{j=m_0+1}^{m} (\alpha + \beta h(y_j, \lambda) + \psi(y_j)) + \gamma_1(\sum_{i=1}^{m} p_i - 1) + \gamma_2 \left\{ \sum_{i=1}^{m} p_i[\exp(\alpha + \beta h(t, \lambda) + \psi(y_i)) - 1] \right\}
\]

Taking the first derivative with respect to \( p_i \), we have

\[
\frac{\partial l_0(\alpha, \beta, \lambda, F_1)}{\partial p_i} = 1/p_i + \gamma_1 + \gamma_2[\exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i)) - 1]
\]
By setting $\frac{\partial l_0}{\partial p_i} = 0$, we have $\sum_{i=1}^m p_i \frac{\partial l_0}{\partial p_i} = m + \gamma_1$, which leads to $\gamma_1 = -m$ and

$$p_i = \frac{1}{\{m - \gamma_2[\exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i)) - 1]\}} \quad (6.5)$$

where $\gamma_2$ is determined by equation

$$f(\gamma_2) = \sum_{i=1}^m \frac{\exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i)) - 1}{m - \gamma_2[\exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i)) - 1]} = 0.$$ 

It is easy to verify that $\gamma_2 = -m_1$ is a solution of $f(\gamma_2) = 0$. Because

$$\frac{\partial f(\gamma_2)}{\partial \gamma_2} = \sum_{i=1}^m \frac{\{\exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i)) - 1\}^2}{\{m - \gamma_2[\exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i)) - 1]\}^2} > 0,$$

$\gamma_2$ is uniquely determined. Therefore $\gamma_2 = -m_1$ is the unique solution and the maximum likelihood is attained at $p_i = m_0^{-1}\{1 + \rho \exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i))\}^{-1}$, where $\rho = m_1/m_0$. Plugging $\hat{p}_i$ into the log-likelihood (6.4), we have the profile likelihood

$$l(\alpha, \beta, \lambda) = \sum_{i=m_0+1}^m (\alpha + \beta h(y_i, \lambda)) - \sum_{i=1}^m \log(1 + \rho \exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i))). \quad (6.6)$$

For given $\psi(y_i)$, the score equations for $(\alpha, \beta, \lambda)$ could be derived as

$$\frac{\partial l}{\partial \alpha} = \sum_{i=m_0+1}^m 1 - \sum_{i=1}^m \frac{\rho \exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i))}{1 + \rho \exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i))}$$

$$\frac{\partial l}{\partial \beta} = \sum_{i=m_0+1}^m h(y_i, \lambda) - \sum_{i=1}^m \frac{\rho \exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i))}{1 + \rho \exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i))} h(y_i, \lambda) \quad (6.7)$$

$$\frac{\partial l}{\partial \lambda} = \sum_{i=m_0+1}^m \beta h'_\lambda(y_i, \lambda) - \sum_{i=1}^m \frac{\rho \exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i))}{1 + \rho \exp(\alpha + \beta h(y_i, \lambda) + \psi(y_i))} \beta h'_\lambda(y_i, \lambda)$$

where $h'_\lambda(y, \lambda)$ is the partial derivative of $h(y, \lambda)$ with respect to $\lambda$. Under the mild
regularity condition (Casella and Berger, 2001), the system of equations (6.7) produce consistent and unique estimators of \((\alpha, \beta, \lambda)\).

In the estimation, \(\psi\) is in general unknown. Applying a method similar to that of Cheng et al. (1995), we replace \(\psi\) in the estimating equation by \(\hat{\psi} = \log(\hat{H}_1) - \log(\hat{H}_0)\), where \(\hat{H}_1\) and \(\hat{H}_0\) are the Kaplan-Meier estimators for the survival functions of the censoring variables in separate groups. With the estimations \((\hat{\alpha}, \hat{\beta}, \hat{\lambda})\) and \(\hat{\psi}\), we have the estimate of \(F(t)\): \(\hat{F}(t) = \sum_{i:y_i<t} \hat{p}_i\), and the estimate of \(G(t)\): \(\hat{G}(t) = \sum_{i:y_i<t} \hat{q}_i\), respectively, where

\[
\hat{p}_i = m_0^{-1} \left\{ 1 + \rho \exp(\hat{\alpha} + \hat{\beta} h(y_i, \hat{\lambda}) + \hat{\psi}(y_i)) \right\}^{-1}, \quad \hat{q}_i = \exp(\hat{\alpha} + \hat{\beta} h(y_i, \hat{\lambda}) + \hat{\psi}(y_i)) \hat{p}_i.
\]

Since both \(F(t)\) and \(G(t)\) are the conditional distribution functions, we must find the unconditional distribution functions \(S_f(t)\) and \(S_g(t)\). Following the same procedure used in Shen et al. (2007), we have

\[
\hat{S}_f(t) = 1 - \hat{\mu}(dS_{f0}, \hat{H}_0) \sum_{i=1}^{m} \hat{H}_0^{-1}(y_i) \hat{p}_i I(y_i \leq t)
\]

where \(\hat{\mu}(dS_{f0}, \hat{H}_0) = \left[ \sum_{i=1}^{m} \frac{\hat{p}_i}{\hat{H}_0(y_i)} \right]^{-1}\), the estimate of \(\mu(dS_{f0}, \hat{H}_0)\); \(\hat{H}_0(t)\) is the Kaplan-Meier estimator of the censoring survival function in the control group. Similarly, we have

\[
\hat{S}_g(t) = 1 - \hat{\mu}(dS_{g0}, \hat{H}_1) \sum_{i=1}^{m} \hat{H}_1^{-1}(y_i) \hat{q}_i I(y_i \leq t)
\]

with \(\hat{\mu}(dG_c, \hat{H}_1) = \left[ \sum_{i=1}^{m} \frac{\hat{q}_i}{\hat{H}_1(y_i)} \right]^{-1}\) and \(\hat{H}_1(t)\) is the Kaplan-Meier estimator of the survival function of the censoring variable in the treatment group. Cure fractions can

96
be estimated non-parametrically by Maller and Zhou’s method,

\[ \hat{\pi}_0 = \tilde{F}_{n_0}(x(n_0)), \quad \hat{\pi}_1 = \tilde{G}_{n_1}(x(n_1)). \]

where \( \tilde{F}_{n_0} \) and \( \tilde{G}_{n_1} \) are the conventional Kaplan-Meier estimators of survival functions in the two groups, and \( x(n_0) \) and \( x(n_1) \) are the largest uncensored failure times of the two groups.

The variance estimates of \( \hat{\beta} \) and \( \hat{\lambda} \) could be obtained through the negative inverse of the second differentiation matrix of (6.7). However, we have not got explicit forms of the variances and are still working on that. As an alternative, the Bootstrap method may be used.

### 6.3 Simulation Study

In the simulation study, we assess the point estimators of \((\alpha, \beta, \lambda)\) by estimating equations (6.7), along with the nonparametric estimators for \(\pi_0\) and \(\pi_1\) by Maller and Zhou (1992). In the simulation, the probability density function in the control group is

\[ -dS_{f_0}(t) = 3 t^2 \exp(-t^3), \]

a Weibull distribution. The density in the treatment group is then

\[ -dS_{g_0}(t) = \exp(\alpha + \beta(t^\lambda - 1)/\lambda)[-dS_{f_0}(t)] \]

We choose \( \beta = 1.5 \). Because \( \alpha \) is determined by \( \lambda \) and \( \beta \) through \(- \int dS_{g_0}(t) = 1\), the estimates of \( \alpha \) are not reported. The cure rates are 0.15 in the control group and 0.20 in treatment group, respectively. For simplicity, \( 1 - \exp(x/10) \) is the censoring
distribution in both groups. Therefore, we have $\psi = 0$. The censoring rates vary from 25% to 45% under different settings. The sample size is 200 in all scenarios. Table 6.1 summarizes the estimates of the unknown parameters with corresponding variances by 2000 replications.

Table 6.1: Parameter estimations based on 2000 replications.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\hat{\beta}$</th>
<th>Var</th>
<th>$\hat{\lambda}$</th>
<th>Var</th>
<th>$\hat{\pi}_0$</th>
<th>Var</th>
<th>$\hat{\pi}_1$</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.5</td>
<td>1.4706</td>
<td>0.0924</td>
<td>-0.6118</td>
<td>0.3507</td>
<td>0.1505</td>
<td>0.0003</td>
<td>0.2007</td>
<td>0.0004</td>
</tr>
<tr>
<td>0.0</td>
<td>1.4818</td>
<td>0.0851</td>
<td>-0.0230</td>
<td>0.4748</td>
<td>0.1499</td>
<td>0.0003</td>
<td>0.2001</td>
<td>0.0004</td>
</tr>
<tr>
<td>0.5</td>
<td>1.4851</td>
<td>0.0698</td>
<td>0.5057</td>
<td>0.3995</td>
<td>0.1498</td>
<td>0.0003</td>
<td>0.1993</td>
<td>0.0004</td>
</tr>
<tr>
<td>1.0</td>
<td>1.4747</td>
<td>0.0624</td>
<td>1.0307</td>
<td>0.5219</td>
<td>0.1495</td>
<td>0.0003</td>
<td>0.1996</td>
<td>0.0004</td>
</tr>
<tr>
<td>1.5</td>
<td>1.4765</td>
<td>0.0613</td>
<td>1.5520</td>
<td>0.6387</td>
<td>0.1500</td>
<td>0.0003</td>
<td>0.1993</td>
<td>0.0004</td>
</tr>
<tr>
<td>2.0</td>
<td>1.4741</td>
<td>0.0657</td>
<td>2.0441</td>
<td>0.6990</td>
<td>0.1507</td>
<td>0.0003</td>
<td>0.2006</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

From the Table, we see that model (6.3) produces estimates of parameters $\beta, \lambda$ and estimates of cure rates with small biases.

For the comparison of the proposed model (6.3) with model (6.2) with specified $h(t)$, we reported the estimates of $\beta$ when $\lambda$ varies with three models: the proposed model (6.3), model (6.2) with $h(t) = t$ and model (6.2) with $h(t) = \log t$. The biases of estimates $\hat{\beta}$ are summarized in Table 6.2.

Table 6.2: Bias of $\hat{\beta}$ of three models when $\beta = 1.5$

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>Bias of the estimation $\hat{\beta}$</th>
<th>$h(t, \lambda)$</th>
<th>$h(t) = t$</th>
<th>$h(t) = \log t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.5</td>
<td></td>
<td>-0.0708</td>
<td>0.3150</td>
<td>0.1769</td>
</tr>
<tr>
<td>0.0</td>
<td></td>
<td>-0.0439</td>
<td>0.1860</td>
<td>0.0005</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td>-0.0086</td>
<td>0.1145</td>
<td>-0.1189</td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td>-0.0375</td>
<td>0.0084</td>
<td>-0.2643</td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td>-0.0235</td>
<td>-0.0350</td>
<td>0.3325</td>
</tr>
<tr>
<td>2.0</td>
<td></td>
<td>-0.0097</td>
<td>-0.0141</td>
<td>0.3514</td>
</tr>
</tbody>
</table>

From the table, the bias of $\beta$ is small for different values of $\lambda$ in model (6.3). The bias of model (6.2) with $h(t) = t$ is small only when $\lambda = 1$ and it is a bit large when
λ \neq 1. It is also true for model (6.2) with \( h(t) = \log t \): the bias is small when \( \lambda = 0 \) and the bias is large when \( \lambda \neq 0 \). In general, model (6.3) outperforms the other two models.
Chapter 7

Conclusions and Discussions

In this thesis, we first reviewed the notations and models for the survival data with and without a cure fraction. We then developed a new nonparametric estimation method for the cure rate. Two new transformation models, one for survival data with a cure fraction only, and one for survival data both with and without a cure fraction, were proposed with corresponding estimating methods. With certain modifications, a weighted log-likelihood function was proposed for survival data with a cure fraction in a length-biased sample. In the two-sample proportional density model, a transformation model was proposed for a function which was assumed known in literature.

In Chapter 2, we developed a new nonparametric estimation of the cure rate for survival data with a cure fraction. The proposed method could adjust the nonparametric effect of the continuous covariate, which is not possible by applying existing methods, and has Maller and Zhou’s method as its special case. The consistency and asymptotic normality of the estimations were proved by theoretical derivation. In the thesis, the theoretical derivation and numerical studies are based on the univariate case. It can be generalized to the multivariate case. However, from our simulation experience, the performance of the proposed method even in bivariate case is not as
good as in univariate case. We suggest the method of Peng and Dear (2000) or the method of Tsodikov (2003) for multivariate case.

In Chapter 3, we proposed a new transformation model for survival data with a cure fraction. The proposed transformation cure model, which was stimulated by the work of Yin and Ibrahim (2005a) and of Zeng et al. (2006), includes the mixture cure model and the proportional odds cure model as its special cases. By comparing the three transformation cure models introduced in Chapter 3, a general transformation model which includes the three basic cure models, the mixture cure model (1.9), the proportional hazards cure model (1.10) and the proportional odds cure model (1.11), is of interest. We propose one such model as follows,

$$
\frac{S_{\text{pop}}(t|z)^{-a} - 1}{a} = \frac{\exp(\beta' z)}{1 - aI(a \leq 0)\exp(\beta' z)} F_0(t)
$$

where $S_{\text{pop}}(t|z)$ is the improper population survival function; $a$ is the transformation parameter; $F_0(t)$ is a proper cumulative distribution function. When $a = 0$, we have

$$
-\log[S_{\text{pop}}(t|z)] = \exp(\beta' z) F_0(t) \text{ or }
$$

$$
S_{\text{pop}}(t|z) = \exp[- \exp(\beta' z) F_0(t)]
$$

which results in the proportional hazards cure model (1.10); When $a = -1$, we have

$$
1 - S_{\text{pop}}(t|z) = \frac{\exp(\beta' z)}{1 + \exp(\beta' z)} F_0(t)
$$

or equivalently

$$
S_{\text{pop}}(t|z) = 1 - \frac{\exp(\beta' z)}{1 + \exp(\beta' z)} F_0(t) = \frac{1}{1 + \exp(\beta' z)} + \frac{\exp(\beta' z)}{1 + \exp(\beta' z)} S_0(t),
$$

which is the mixture cure model (1.9), where $S_0(t) = 1 - F_0(t)$ is the proper survival function for the uncured subjects. When $a = 1$, we have $S_{\text{pop}}(t|z)^{-1} - 1 =$
\[ S_{\text{pop}}(t|z) = \frac{1}{1 + \exp(\beta'z)F_0(t)} \]

resulting in the proportional odds cure model (1.11). Hence model (7.1) includes all three basic cure models as its special cases.

Model (7.1) could be rewritten as

\[ S_{\text{pop}}(t|z) = \left\{ 1 + \frac{a \exp(\beta'z)}{1 - a \mathbb{I}(a \leq 0) \exp(\beta'z)} F_0(t) \right\}^{-1/a} \]

In general, it is a valid cure model for any \( a \) in the real line and the cure rate

\[ \pi(z) = \left\{ 1 + \frac{a \exp(\beta'z)}{1 - a \mathbb{I}(a \leq 0) \exp(\beta'z)} \right\}^{-1/a}. \]

A nonparametric distribution or a parametric distribution may be specified for \( F_0(t) \) in the model. The asymptotic property may be adjusted using the profile likelihood theory of Murphy and van der Vaart (2000).

Transformation models proposed in Chapter 3, Chapter 4 and Chapter 5 are constructed through the Box-Cox transformation. In linear models, it is assumed that the population being sampled or investigated is normally distributed. When this assumption is violated, Box and Cox suggested the Box-Cox transformation. The transformation is imposed on the response variables and the normality assumption is then nearly satisfied (Box and Cox, 1964). In this thesis, however, the Box-Cox transformation was imposed not on the response variables but on the functions, such as survival functions or hazard functions. Our goal of using the Box-Cox transformation is not for satisfying the normality assumption but for getting a unified approach.

In Chapter 3, a transformation model was proposed for cure models selection. In
Chapter 4, a new transformation model was introduced for analyzing the survival both with and without a cure fraction. In Chapter 6, a transformation model was proposed to select $h(t)$. All three transformation models are constructed through the Box-Cox transformation model and provide unified solutions in different cases. Both the transformation model in Chapter 3 and the model in Chapter 4 are proposed for model selection. However, the model in Chapter 4 is proposed mainly for detecting cure models from non-cure models.

In Chapter 5 and Chapter 6, the performances of the proposed methods were illustrated through limited simulation works. Extensive simulation studies and real data applications are desired in the future. Theoretical works are also needed to adjust asymptotic properties of the proposed estimates.
Bibliography


