THE UNIVERSITY OF CALGARY

Estimation for Censored Single-Index Models Using MAWVE and OPWG Methods

by

He, Zhulin

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

August, 2008

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Estimation for Censored Single-Index Models Using MAWVE and OPWG Methods" submitted by He, Zhulin in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE.

Xuewen Lu

Supervisor, Dr. Xuewen Lu
Department of Mathematics and Statistics

M. D. Burke

Dr. Murray Burke
Department of Mathematics and Statistics

Gemai Chen

Dr. Gemai Chen
Department of Mathematics and Statistics

A. David

Dr. Alexander David
Haskayne School of Business

Aug. 21, 2008

Date

Abstract

The accelerated failure time (AFT) model is one of the most popular regression models in survival analysis. The AFT model assumes a linear relationship between (log) event time and covariates. We propose a censored single-index model to extend the identity link function in the AFT model to an unknown link function, so that the new model is more flexible in capturing information between event time and covariates than the AFT model, and at the same time, it avoids the "curse of dimensionality". In this thesis, we provide two estimation methods for the new model. One is minimum average weighted conditional variance estimation (MAWVE); the other is estimation via outer product of weighted gradients (OPWG). These methods use the Kaplan-Meier weights in the least squares objective functions to account for censoring. The MAWVE method estimates parameters by minimizing the overall approximation errors which are calculated from the response variable and the estimated smooth link function. The OPWG method works on the eigenvector corresponding to the largest eigenvalue from the weighted outer product of gradients of the estimated link function. The weighed local linear estimate method is used to estimate the local link function, and the nonparametric 0.632 bootstrap is considered to estimate the variance and construct the bootstrap confidence intervals. Simulation studies and real data examples are used to evaluate and illustrate the proposed methods.

Acknowledgements

I would like to express my sincere thanks to many indeividuals who offered their invaluable assistance and generous help in the preparation of this thesis and also in my life.

My deepest gratitude goes first and foremost to my supervisor Dr. Xuewen Lu, for his tremendous support, invaluable advice, and great patience, not only on the thesis, but also on my graduate studies and my research life. His suggestions and encouragement helped me in all the time of research for and writing of this thesis. Without his constant encouragement and illuminating instruction, this thesis could not come true.

I want to thank the Department of Mathematics and Statistics for giving me all the chances and facilities to my graduate study. The faculties and staffs are very nice to the graduate students and always give me a lot of advices and help for my graduate study and life in Calgary. I really enjoy the peaceful and great research environment at the Department of Mathematics and Statistics.

I would also like to extend my sincere thanks to Dr. Karen Kopciuk, whose inspiring insights, generous encouragements, and enthusiastic instructions have facilitated me much through my research working and writing.

I would also wish to give thanks to my colleagues and friends, Piktan Wong, Chaoxiong(Michelle) Xia, Lei(Larry) Hua and Andrea Soo, for their assistance and advices in my graduate study.

Most importantly, I would express my thanks to my parents and my love, who always give me their endless love and boundless support. Thank you for giving me the great confidence and strong support in my life.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this Chapter, we introduce our models and briefly review the existing estimation methods for the related models. In Section 1.1, we describe the basic concepts and background of survival analysis. In Section 1.2, we outline the historical development and the basic set up of censored single-index models. Two methods, rMAVE and rOPG, on censored single-index models which were introduced by Xia (2006) are also reviewed in Section 1.3.

1.1 Introduction to Survival Analysis

Survival analysis is a branch of statistics which deals with survival time data. Survival analysis is an important research field in biostatistics, and it is widely used in actuarial science, business, biology, public health, medical science, engineering and sociology. This topic shares different terminologies in different areas, since the concepts may be a little bit different after several renovations, which make the topic much easier to be applied in the specific area. Survival analysis is also called reliability analysis or failure-time analysis in engineering, and duration analysis or event-history analysis in economics and sociology. In general, survival analysis involves with studying and modeling of time to event data. In this context, death or failure is considered the specific "event". This "event" may be death related to a disease, broken of electronic devices, criminal recidivism, divorce, unemployment, worker's compensation claims, subscription of a magazine, and graduation from school. Here in this thesis, we are concerned with survival analysis, in the biostatistical context, which is the application of statistics to questions about human health.

1.1.1 Censored Data

Survival time data set in survival analysis usually contains censored observations. The censored observations are defined as the specific event has not happened by the end of the study or the event-happening time is not observed or known; the specific event could also occur when people are lost to follow-up after the designated period of study. Otherwise, when there are no censored observations, we have complete observations.

There are mainly four types of censoring.

1. *Type I censoring*

   Type I censoring is defined as the specific event that is observed only if it occurs prior to some prespecified time. The censoring time may be different from individual to individual, but the censoring time is fixed to the specific individual. For example, a researcher may start with a fixed number of rats. The rats are given some treatment. Due to time or cost considerations, the researcher terminate the study or report the results before all the rats are observed for the specific event, say death. Type I censoring is the most common censoring type in censored data.

2. *Type II censoring*

   If a study continues until the death or failure of the first r individuals are observed, then we have Type II censoring. For instance, a study begins with n rats, which are given some treatment. After r rats are dead, the investigator will stop the study.

3. *Left censoring*

   Before an individual is observed for the event at time C_l, the specific event has already happened in the study. We call this left censoring. Here is an example, a question in a survey "When did you get your first car?". A response may like "I got a car before a certain time, but I can not remember when I got it". Then such a response is considered as left censored.

4. *Interval censoring*

   When the specific event is known to fall in a time interval (a, b), then we call it interval censoring. For example, the age at which a subject first develops a kind of disease is known to the before time b. But at time a, which is before time b, the physical examination to this disease shows no positive result. So the specific event or disease occurred in the time interval (a, b), and the response is subject to interval censoring.

Type I censoring and Type II censoring are also called right censoring. Another important type of right censorship is random censoring. It is very popular in biostatistics.

Suppose T is a random variable denoting the time-of-event, such as death and failure. Thus the survival function is the probability that an individual's time of event is beyond some specified time,

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

where t is a fixed time of interest. Here the survival function is the complement of the cumulative distribution function,

$$F(t) = Pr(T \leq t) = 1 - S(t).$$

The survival function $S(t)$ and the cumulative distribution function $F(t)$ always have the properties listed below,

1. At the beginning time t=0, we assume that $S(0) = 1$ and $F(0) = 0$, that is, no death or failure occur at the beginning of the study.

2. The survival function is a monotone and nonincreasing function. If $u > t$, then $S(u) \leq S(t)$ and $F(u) \geq F(t)$.

3. The survival function decreases to zero when time approaches to infinity, i.e. $S(t) \to 0$ and $F(t) \to 1$ as $t \to \infty$. This means for all individuals, the event of interest will occur if we have enough study time.

Kaplan-Meier estimator is the standard estimator of survival function for right censored data (see Kaplan & Meier, 1958). It is also called product limit estimator.

Suppose we have a study sample where D individuals are observed with the specific event at the observed time $t_1 \leq t_2 \leq \cdots \leq t_D$. At time t_i there are d_i events occuring right at that time. Let Y_i be the total number of individuals who are at risk (both survived and dead) at time t_i. The quantity d_i/Y_i shows the conditional probability that an individual who still survives up to time t_i may suffer the specific event at that time. The Kaplan-Meier estimator is defined as

$$\hat{S}(t) = \begin{cases} 1, & \text{if } t < t_1 \\ \prod_{t_i \leq t}[1 - \frac{d_i}{Y_i}], & \text{if } t_1 \leq t \end{cases} . \tag{1.1}$$

The advantage of this estimator is that there is no information loss, since it incorporates the information from all individuals, both censored and uncensored. We can see this point from its definition.

1.2 Introduction to Censored Single-Index Model

1.2.1 Background Information

One of the most important methods in statistics is regression analysis. It aims to estimate the regression function, which describes the relationship between a response variable $Y \in \mathbb{R}$ and an explanatory vector variable $X \in \mathbb{R}^p$. The regression analysis technique is initially based on a linear regression model, such as

$$Y = \alpha + \beta_0^T X + \varepsilon, \tag{1.2}$$

where X is a covariate vector which may depend on the error term ε, α is the intercept, and β_0 is the parameter vector. This model has been studied in many different ways, such as least squares estimation (LSE) method.

Accelerated failure time (AFT) models take logarithm of the survival time T with the ususal linear regression form

$$Y = \log T = \alpha + \beta_0^T X + \varepsilon, \tag{1.3}$$

where the variables are the same as those noted in (1.2).

Brillinger (1983) proposed a link function $g(\cdot)$ to the linear regression model, producing the generalized linear model (GLM)

$$Y = g(\beta_0^T X) + \varepsilon,$$

where the link function $g(\cdot)$ is known. The GLM consists of three elements:

1. A probability density function f for Y is from the exponential family.
2. A linear predictor $\eta = \beta_0^T X$.
3. A link function g such that $E(Y) = \mu, \eta = g^{-1}(\mu)$; g^{-1} is the inverse function of g.

The unknown parameter β_0 in the GLM is typically estimated with the maximum likelihood estimator (MLE), the maximum quasi-likelihood estimator (MQLE) or Bayesian techniques.

Stoker (1986) considered the same model form but with an unknown univariate smooth link function $g(\cdot)$, that is,

$$Y = g(\beta_0^T X) + \varepsilon, \tag{1.4}$$

where the norm of β_0 is 1 ($||\beta_0|| = 1$), $E(\varepsilon|X) = 0$, and the first component of β_0 is positive. If the first component of β_0 is negative, we can take $\tilde{\beta}_0 = -\beta_0$ and a new link function $\tilde{g}(\cdot)$ to satisfy that $g(\beta_0^T X) = g(-\tilde{\beta}_0^T X) = \tilde{g}(\tilde{\beta}_0^T X)$. This model is called the single-index model (SIM). Whether the link function $g(\cdot)$ is known or not is the primary difference between GLMs and SIMs. Recently, Xia (2006) considered SIMs and proposed two estimation methods, which are called rMAVE and rOPG methods, respectively (see section 1.3.1 and 1.3.2). He showed that the resultant estimators are asymptotically normal and his algorithms work more efficiently than others.

1.2.2 Censored Single-Index Model

Based on the previous work on the single-index model, we focus on the censored single-index model, where the response Y is randomly right censored. The censored single-index model (CSIM) is a semiparametric censored regression model, which generalizes the familiar accelerated failure time (AFT) models by assuming an unknown link function and an error term ε, where ε can be heteroscedastic. It can be used to model more flexible relationships between the survival response and the covariates than the AFT model. It also provides a technique for "dimension reduction" in nonparametric censored regression models. In the single-index model

$$Y_i = g(\beta_0^T X_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{1.5}$$

where Y_i is the ith object's survival time (on log scale), X_i is the covariate vector, $X_i \in \mathbb{R}^p$. β_0 is the coefficient parameter vector. Let C_i be the random censoring variable. Due to the censoring, we would not know the survival time Y_i, but the censoring indicator $\delta_i = I_{\{Y_i \le C_i\}}$ and the censored response $Z_i = \min\{Y_i, C_i\}$. So we have the data set in the form of $\{(X_i, Z_i, \delta_i)\}_{i=1}^n$.

Actually, the single-index model is a special case of the censored single-index model. When $\delta_i \equiv 1$, i.e. $Z_i \equiv Y_i$, $i = 1, \cdots, n$, the single-index model and the censored single-index model are the same. Lu and Burke (2005) considered a CSIM and proposed the average derivative estimator (ADE) for it.

1.3 Review of Estimation for Single-Index Model

In the past two decades, the estimation of the single-index parameter β_0 has been investigated in a series of papers. Typically, there are three main approaches. The first one is the average derivative estimation (ADE) method, proposed by Powell *et al.* (1989), Härdle *et al.* (1989 & 1993). The second one is the sliced inverse regression (SIR) suggested by Li (1989). The third method is the semiparametric least squares (SLS) estimation investigated by Ichimura (1993).

Suppose the mean response of Y given $X = x$ is denoted as

$$m(x) = E(Y|X = x).$$

Then the vector of "average derivative" is defined as

$$\eta = E[m'(X)],$$

where $m'(x) = \partial m(x)/\partial x$. It is the vector of partial derivatives and its expectation is taken with respect to the marginal distribution of X. Härdle et al. (1989 & 1993) proposed the average derivative estimation (ADE) on the SIM for studying the mean response $m(x)$ through the estimation of the d-dimensional vector η. ADE can be considered in two stages: first estimate β using $\hat{\beta}$, then approximate $m(x)$ by $\hat{m}(x) = \hat{g}(\hat{\beta}^T x)$.

Recently, Xia (2006) proposed two new methods: the refined minimum average conditional variance estimation (rMAVE) and the refined outer product of the gradients (rOPG). Xia (2006) summarized the ADE method proposed by Powell et al. (1989) and Härdle and Stoker (1989), and pointed out that this ADE method suffers two drawbacks. The first is the "curse of dimensionality", the second is that when $E\{g'(\beta_0^T X)\} = 0$, the method fails to estimate β_0. Then Xia introduced the rMAVE and rOPG to overcome these disadvantages.

In this thesis, we adopt Xia's methods and develop two new estimation methods for CSIMs. We first introduce rMAVE and rOPG for SIMs.

1.3.1 Introduction to the rMAVE Method

The basic idea of this method is to linearly approximate the smooth link function $g(\cdot)$ and to estimate β_0 by minimizing the overall approximation errors

$$E[Y - E(Y \mid X)]^2 = E[Y - g(\beta^T X)]^2,$$

where the unknown link function $g(\cdot)$ is locally approximated by a linear function (Taylor expansion) of $g(\beta_0^T X)$ at $\beta_0^T x$

$$g(\beta_0^T X) \approx a + d\beta_0^T (X - x),$$

where $a = g(\beta_0^T x)$ and $d = g'(\beta_0^T x)$. Hence, given a sample $\{(Y_i, X_i)\}_{i=1}^n$, rMAVE estimates β_0 by minimizing local conditional variance

$$\min_{\substack{\beta^T\beta=1\\ a_j, d_j, j=1,\cdots,n}} \sum_{i=1}^n \sum_{j=1}^n \left[Y_i - \{a_j + d_j \beta^T (X_i - X_j)\}\right]^2 w_{ij}, \tag{1.6}$$

where w_{ij} are the kernel weights to describe the local character of linear approximation,

$$w_{ij} = K_h(\beta^T(X_i - X_j)) \left[\sum_{i=1}^n K_h(\beta^T(X_i - X_j))\right]^{-1}, \tag{1.7}$$

$K_h(\cdot) = (1/h)K(\cdot/h)$, $K(\cdot)$ is a kernel function and h is the bandwidth.

The detailed algorithm of rMAVE could be expressed in three steps. Suppose β is an initial estimate of β_0. Let $X_{ij} = X_i - X_j$ and $\hat{f}_\beta(\beta^T X_j) = n^{-1}\sum_{i=1}^n K_h(\beta^T X_{ij})$.

Step 1: Calculate the a_j^β and d_j^β such that

$$\begin{pmatrix} a_j^\beta \\ d_j^\beta h \end{pmatrix} = \left[\sum_{i=1}^n K_h(\beta^T X_{ij}) \begin{pmatrix} 1 \\ \beta^T X_{ij}/h \end{pmatrix} \begin{pmatrix} 1 \\ \beta^T X_{ij}/h \end{pmatrix}^T\right]^{-1} \sum_{i=1}^n K_h(\beta^T X_{ij}) \begin{pmatrix} 1 \\ \beta^T X_{ij}/h \end{pmatrix} Y_i, \tag{1.8}$$

Step 2: Fix the weights $K_h(\beta^T X_{ij})$, a_j^β and d_j^β, calculate the solution to β in (1.6).

$$\begin{aligned}\beta =&\{\sum_{i,j=1}^{n} K_h(\beta^T X_{ij})\hat{\rho}_j^\beta\{d_\beta(X_j)\}^2 X_{ij}X_{ij}^T/\hat{f}_\beta(\beta^T X_j)\}^{-1}\\ &\sum_{i,j=1}^{n} K_h(\beta^T X_{ij})\hat{\rho}_j^\beta d_\beta(X_j)X_{ij}(Y_i - a_j^\beta)/\hat{f}_\beta(\beta^T X_j),\end{aligned} \tag{1.9}$$

where $d_\beta(X_j) = d_j^\beta$ in step 1, and $\hat{\rho}_j^\beta = \rho_n\{\hat{f}_\beta(X_j)\}$. Here $\rho_n(\cdot)$ is a trimming function used to handle the boundary points. In Xia's paper, for some fixed constant $\epsilon > 0$ and $c_0 > 0$,

$$\rho_n(v) = \begin{cases} 1, & \text{if } v \geq 2c_0 n^{-\epsilon} \\ \frac{\exp\{(2c_0 n^{-\epsilon} - v)^{-1}\}}{\exp\{(2c_0 n^{-\epsilon} - v)^{-1}\} + \exp\{(v - c_0 n^{-\epsilon})^{-1}\}}, & \text{if } 2c_0 n^{-\epsilon} > v > c_0 n^{-\epsilon} \\ 0, & \text{if } v \leq c_0 n^{-\epsilon} \end{cases}. \tag{1.10}$$

The choice of ϵ and c_0 will be given in the subsequent simulation study.

Step 3: iteratively repeat Step 1 and Step 2 until $\beta := \beta/||\beta||$ converges.

Normally the iteration could be stopped by some common convergence rules. We denote the converged rMAVE estimator by $\hat{\beta}_{rMAVE}$.

1.3.2 Introduction to the rOPG Method

Let $\lambda(x) = E(Y|X = x) = g(\beta_0^T x)$, where $\lambda(x)$ is the same to$m(x)$ in the rMAVE method, then it is easy to get

$$\nabla\lambda(x) = \frac{\partial}{\partial x}\lambda(x) = \frac{\partial}{\partial x}g(\beta_0^T x) = g'(\beta_0^T x)\beta_0, \tag{1.11}$$

where the derivative or gradient of the regression function at any point x shares the same direction as β_0. For rOPG method, we focus on $E\{\nabla\lambda(X)\nabla^T\lambda(X)\}$ instead of $E\{\nabla\lambda(X)\}$. Since $E\{\nabla\lambda(X)\nabla^T\lambda(X)\}$ has only one nonzero eigenvalue, which is 1, and

$$E\{\nabla\lambda(X)\nabla^T\lambda(X)\} = E\left[\{g'(\beta_0^T X)\}^2\right]\beta_0\beta_0^T, \tag{1.12}$$

the eigenvector of $E\{\nabla\lambda(X)\nabla^T\lambda(X)\}$, corresponding to its largest eigenvalue, is the index parameter β_0.

Thus, the rOPG method considers the local linear regression, it solves the minimization problem

$$\min_{a_j,b_j}\sum_{i=1}^{n}\{Y_i - a_j - b_j^T X_{ij}\}^2 w_{ij}, \tag{1.13}$$

where X_{ij} and w_{ij} are defined as in the rMAVE section.

Next, we introduce the algorithm of the rOPG method in three steps which are similar to those of the rMAVE method. Suppose β is an initial estimate of β_0, satisfying the condition (C5) stated in the next section.

Step 1: Calculate a_j^β and b_j^β such that

$$\begin{pmatrix} a_j^\beta \\ b_j^\beta \end{pmatrix} = \left[\sum_{i=1}^{n} K_h(\beta^T X_{ij}) \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}\begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^T\right]^{-1} \sum_{i=1}^{n} K_h(\beta^T X_{ij}) \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} Y_i. \tag{1.14}$$

Step 2: Calculate the first eigenvector corresponding to the largest eigenvalue of

$$\hat{\Sigma} = n^{-1}\sum_{j=1}^{n}\hat{\rho}_j^\beta b_j^\beta (b_j^\beta)^T, \tag{1.15}$$

where $\hat{\rho}_j^\beta = \rho_n(\hat{f}_\beta(\beta^T X_j))$, the same as $\hat{\rho}_j^\beta$ defined in the previous rMAVE section. Denote the first eigenvector to (1.15) again by β.

Step 3: Repeat Step 1 and Step 2 until β converges.

We denote the converged rOPG estimator by $\hat{\beta}_{rOPG}$. The iteration could be stopped by some common convergence rules, too.

1.3.3 Remarks on the rMAVE and rOPG Methods

Xia (2006) compared the two methods and got the conclusion that the rMAVE method is more efficient than the rOPG method. Under the following conditions, the two estimators

$\hat{\beta}_{rMAVE}$ and $\hat{\beta}_{rOPG}$ are asymptotically normal.

(C1) (Design). The density function $f_\beta(v)$ of $\beta^T X$ and its derivatives up to the third order are bounded on $\mathbb{R}$ for all $\beta : ||\beta - \beta_0|| < \delta$ where $\delta > 0$ is a constant, $E|X|^6 < \infty$, and $E|Y|^3 < \infty$.

(C2) (Link function). The conditional mean $g_\beta(v) = E(Y|\beta^T X = v)$, and its derivatives up to the third order are bounded for all $\beta : ||\beta - \beta_0|| < \delta$ where $\delta > 0$.

(C3) (Kernel function). $K(v)$ is a symmetric density function with finite moments of all orders and bounded derivative. Its Fourier transform is absolutely integrable.

(C4) (Bandwidth and trimming parameter). Bandwidth $h \propto n^{-1/5}$ and trimming parameter $\epsilon < \frac{1}{20}$.

(C5) (Small neighborhood). The initial value β is in a small neighbor of β_0: $\mathfrak{B} = \{\beta : ||\beta - \beta_0|| \le c_0 n^{1/2+c_0}\}$ with $c_0 < \frac{1}{20}$.

Under (C1)-(C5), Xia (2006) got the asymptotic results of the two estimators $\hat{\beta}_{rMAVE}$ and $\hat{\beta}_{rOPG}$. Use the Moore-Penrose method to calculate the inverse of symmetric matrix A, and denote it as A^+ (see Penrose, 1955). Let $\mu_\beta(x) = E(X|\beta^T X = \beta^T x), \nu_\beta(x) = \mu_\beta - x, w_\beta = E(XX^T|\beta^T X = \beta^T x), W_0(x) = \nu_{\beta_0}(x)\nu_{\beta_0}^T(x)$, and $W(x) = w_{\beta_0}(x) - \mu_{\beta_0}(X)\mu_{\beta_0}^T(X)$. Xia's results can be summarized in the following three theorems.

Theorem 1.3-1. *(see Xia, 2006)*

$$\sqrt{n}(\hat{\beta}_{rMAVE} - \beta_0) \rightarrow N(0, \Sigma_{rMAVE}), \tag{1.16}$$

where

$$\begin{aligned}\Sigma_{rMAVE} =& [E\{g'(\beta_0^T X)^2 W(X)\}]^+ E\{g'(\beta_0^T X)^2 W_0(X)\varepsilon^2\} \\ & \times [E\{g'(\beta_0^T X)^2 W(X)\}]^+.\end{aligned} \tag{1.17}$$

Theorem 1.3-2. *(see Xia, 2006)*

$$\sqrt{n}(\hat{\beta}_{rOPG} - \beta_0) \rightarrow N(0, \Sigma_{rOPG}), \tag{1.18}$$

where

$$\Sigma_{rOPG} = \quad E\{g'(\beta_0^T X)^2 W(X)^+ W_0(X) W(X)^+ \varepsilon^2\}/\{Eg'(\beta_0^T X)^2\}^2. \tag{1.19}$$

Theorem 1.3-3. *(see Xia, 2006) If ε is independent of X, then*

$$\Sigma_{rOPG} \geq \Sigma_{rMAVE}. \tag{1.20}$$

When X is normal, these two methods perform equivalently in terms of asymptotic efficiency, i.e., $\Sigma_{rOPG} = \Sigma_{rMAVE}$. Theorem $1.3-3$ shows that rMAVE is more efficient than rOPG. But in computing the estimates, rOPG is easier to implement and faster to compute than rMAVE.

Chapter 2

Estimation of CSIMs

In last Chapter, we introduced CSIMs and SIMs, and also introduced the two estimators of SIMs: β_{rMAVE} and β_{rOPG}. Now we present the core work of this thesis: two estimators of CSIMs. In Section 2.1, we describe the model assumptions. In Section 2.2, we propose two estimation procedures for CSIMs, which are modified from rMAVE and rOPG for SIMs. We discuss the properties of the estimators in Section 2.3.

2.1 Model Assumptions

Now, we study the CSIMs. Consider the following censored single-index model,

$$Y_i = g(\beta_0^T X_i) + \varepsilon_i, \tag{2.1}$$

where Y_i is the ith subject's survival time (on the log scale), X_i is the covariate vector, $X_i \in \mathbb{R}^p$, β_0 is the parameter p-vector, $\beta_0 \in \mathbb{R}^p$, satisfying $||\beta_0|| = 1$, the first component of β_0 is positive, ε_i is the random error, satisfying $E(\varepsilon_i|X_i) = 0$. Let C_i be the random censoring variable, independent of X_i and Y_i. Then we have the censor indicator $\delta_i = I_{\{Y_i \leq C_i\}}$, and at the same time, we have the censored response $Z_i = min\{Y_i, C_i\}$. Hence we have an i.i.d random sample $\{(X_i, Z_i, \delta_i)\}_{i=1}^n$ from the population $\{X, Z, \delta\}$. Assume the distribution of Y, C and Z are $F(t) = P(Y \leq t)$, $G(t) = P(C \leq t)$ and $H(t) = P(Z \leq t)$, respectively.

We assume the same assumptions (C1)-(C5) in Chapter 1. In addition, we assume that the following conditions for the Kaplan-Meier integrals hold,

(C6). Let $u = \beta_0^T x$, $H(t) = P(Z \leq t)$, denote by $F(y|u)$ the conditional distribution of Y given $U = \beta_0^T X = u$, assume

$$\int_0^{\tau_H} y^2 \{1 - G(y)\}^{-1} dF(y|u) < \infty,$$

where $\tau_H = \inf\{t : H(t) = 1\}$.

(C7). Denote $\tau_F = \inf\{t : F(t) = 1\}$ and $\tau_G = \inf\{t : G(t) = 1\}$. Suppose $\tau_F \leq \tau_G$ which implies $1 - H(t) = (1 - F(t))(1 - G(t))$, $\tau_H = \tau_F$; and assume

$$\int_0^{\tau_H} |y| c^{1/2}(y) dF(y|u) < \infty,$$

where $c(y) = \int_{-\infty}^{y} \{1 - H(s)\}^{-1} \{1 - G(s)\}^{-1} dG(s)$. See Stute (1996) for more details about these conditions.

2.2 Estimation for CSIMs

2.2.1 MAWVE Method

This approach is to linearly approximate the smooth link function $g(\cdot)$ of the CSIM and then estimate the parameter β_0 by minimizing the overall weighted approximation errors. The method is called minimum average weighted conditional variance estimation (MAWVE).

Suppose β is an initial estimate of β_0, G is known. Take $X_{ij} = X_i - X_j$, $\hat{f}_\beta(X_j) = n^{-1} \sum_{i=1}^{n} K_h(\beta^T X_{ij})$ and $w_{iG} = \delta_i/(1 - G(Z_i-))$, where $G(\cdot-)$ is the left-continuous version of G. When G is unknown, we replace G by its Kaplan-Meier estimator $\hat{G}$, i.e. $w_{i\hat{G}} = \delta_i/(1 - \hat{G}(Z_i-))$. In fact, $w_{i\hat{G}}$ is equivalent to the Kaplan-Meier weight defined in Stute (1993, 1996), see Satten and Datta (2001). Then we have conditional expectation equation $E(Y_i|X_i) = E(w_{iG} Z_i|X_i)$. This indicates that we can use the Kaplan-Meier weight to construct new estimators of β_0 under censorship.

In order to estimate β, we need to estimate $g(\cdot)$ first. To do this, assume β is fixed and similar to the rMAVE method, we have the Taylor expansion of the link function

$g(\beta^T X_i)$ at $g(\beta^T x)$ given by

$$g(\beta^T X_i) \approx a + d\beta^T (X_i - x), \tag{2.2}$$

where $a = g(\beta^T x)$, $d = g'(\beta^T x)$. we would like to calculate the minimum value of the estimated conditional variance with respect to a and d using the Kaplan-Meier weights,

$$\sigma_n^2(x|\beta) = \min_{a,d} \{n\hat{f}_\beta(x)\}^{-1} \sum_{i=1}^{n} [Z_i - \{a + d\beta^T (X_i - x)\}]^2 w_{i\hat{G}} K_h\{\beta^T (X_i - x)\}, \tag{2.3}$$

where $\hat{f}_\beta(x) = n^{-1} \sum_{i=1}^{n} K_h\{\beta^T (X_i - x)\}$ and $K_h(u) = K(u/h)/h$, where K is Gaussian kernel function, $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$, and h is the bandwidth.

Write the right-hand-side of (2.3) as

$$l_n^2(x, a, d, \beta, Z) = \{n\hat{f}_\beta(x)\}^{-1} \sum_{i=1}^{n} \left[Z_i - \{a + d\beta^T (X_i - x)\}\right]^2 w_{iG} K_h\{\beta^T (X_i - x)\}. \tag{2.4}$$

Let

$$l_n^2(x, a, d, \beta, Y) = \{n\hat{f}_\beta(x)\}^{-1} \sum_{i=1}^{n} \left[Y_i - \{a + d\beta^T (X_i - x)\}\right]^2 K_h\{\beta^T (X_i - x)\}, \tag{2.5}$$

which is the estimated conditional variance when data are fully observed. We can see that

$$
\begin{aligned}
&E\{l_n^2(x,a,d,\beta,Z)|X_1,\cdots,X_n\}\\
=&E\{\{n\hat{f}_\beta(x)\}^{-1}\sum_{i=1}^{n}\left[Z_i-\{a+d\beta^T(X_i-x)\}\right]^2\frac{I_{\{Y_i\le C_i\}}}{1-G(Z_i-)}K_h\{\beta^T(X_i-x)\}|X_1,\cdots,X_n\}\\
=&E\{\{n\hat{f}_\beta(x)\}^{-1}\\
&\sum_{i=1}^{n}\left[\min\{Y_i,C_i\}-\{a+d\beta^T(X_i-x)\}\right]^2\frac{I_{\{Y_i\le C_i\}}}{1-G(\min\{Y_i,C_i\})}K_h\{\beta^T(X_i-x)\}|X_1,\cdots,X_n\}\\
=&E\{\{n\hat{f}_\beta(x)\}^{-1}\sum_{i=1}^{n}E\{\frac{I_{\{Y_i\le C_i\}}}{1-G(Y_i)}\\
&\left[Y_i-\{a+d\beta^T(X_i-x)\}\right]^2K_h\{\beta^T(X_i-x)\}|X_1,\cdots,X_n,Y_1,\cdots,Y_n\}|X_1,\cdots,X_n\}\\
=&E\{\{n\hat{f}_\beta(x)\}^{-1}\{1-G(Y_i)\}\frac{1}{1-G(Y_i)}\\
&\sum_{i=1}^{n}\left[Y_i-\{a+d\beta^T(X_i-x)\}\right]^2K_h\{\beta^T(X_i-x)\}|X_1,\cdots,X_n\}\\
=&E\{l_n^2(x,a,d,\beta,Y)|X_1,\cdots,X_n\}.
\end{aligned}
$$

Hence, we propose the weighted estimator of the conditional variance when data are randomly right censored.

An important class of estimators is one that minimizes the overall sum of $\sigma_n^2(x|\beta)$ at all $x=X_j, j=1,\cdots,n$. Hence, in the MAWVE method, the estimator of β_0 is to minimize the overall local linear approximation of the conditional variance,

$$
Q_n(\beta)=\sum_{j=1}^{n}\sigma_n^2(X_j|\beta), \tag{2.6}
$$

where β satisfies $||\beta||=1$.

The algorithm of the MAWVE method can be expressed in three steps. Suppose β is an initial estimate of β_0.

Step 1: Fix β, calculate the a_j^β and d_j^β such that

$$\begin{pmatrix} a_j^\beta \\ d_j^\beta h \end{pmatrix} = \left[\sum_{i=1}^{n} w_{i\hat{G}} K_h(\beta^T X_{ij}) \begin{pmatrix} 1 \\ \beta^T X_{ij}/h \end{pmatrix} \begin{pmatrix} 1 \\ \beta^T X_{ij}/h \end{pmatrix}^T \right]^{-1}$$
$$\sum_{i=1}^{n} w_{i\hat{G}} K_h(\beta^T X_{ij}) \begin{pmatrix} 1 \\ \beta^T X_{ij}/h \end{pmatrix} Z_i. \tag{2.7}$$

Step 2: Fix $K_h(\beta^T X_{ij})$, a_j^β and d_j^β, calculate the solution of β to a minimization formula,

$$\begin{aligned}\beta =& \{\sum_{i,j=1}^{n} w_{i\hat{G}} K_h(\beta^T X_{ij}) \hat{\rho}_j^\beta \{d_\beta(X_j)\}^2 X_{ij} X_{ij}^T / \hat{f}_\beta(\beta^T X_j)\}^{-1} \\ & \sum_{i,j=1}^{n} w_{i\hat{G}} K_h(\beta^T X_{ij}) \hat{\rho}_j^\beta d_\beta(X_j) X_{ij} (Z_i - a_j^\beta) / \hat{f}_\beta(\beta^T X_j),\end{aligned} \tag{2.8}$$

where $d_\beta(X_j) = d_j^\beta$ in Step 1 and $\hat{\rho}_j^\beta = \rho_n\{\hat{f}_\beta(X_j)\}$.

Step 3: iteratively repeat Step 1 and Step 2 until $\beta := \beta/||\beta||$ converges.

The iteration can be stopped by some common convergence rules (see Section 2.3.1). We denote the converged MAWVE estimator by $\hat{\beta}_{MAWVE}$.

2.2.2 OPWG Method

The basic idea of this method is to work on the eigenvector corresponding to the largest eigenvalue from the outer product of weighted gradients of the estimated link function. The method is called the outer product of weighted gradients (OPWG) method. Let $\lambda(x) = E(Y|X=x)$, which is the same in the rOPG section. Define

$$\nabla\lambda(x) = \frac{\partial}{\partial x}\lambda(x) = \frac{\partial}{\partial x} g'(\beta_0^T x) = g'(\beta_0^T x)\beta_0. \tag{2.9}$$

Then

$$E\{\nabla\lambda(x)\nabla^T\lambda(x)\} = E\left[\{g'(\beta_0^T x)\}^2\right]\beta_0\beta_0^T. \tag{2.10}$$

So the index parameter β_0 is the corresponding eigenvector when the mean outer product of the gradients $E\{\nabla\lambda(X)\nabla^T\lambda(X)\}$ reaches its largest eigenvalue.

First we estimate the gradients by local polynomial smoothing. It is implemented by solving the following minimization problem of the weighted local linear regression function

$$\min_{a_j,b_j}\sum_{i=1}^{n}\{Z_i - a_j - b_j^T X_{ij}\}^2 w_{i\hat{G}} w_{ij}, \tag{2.11}$$

where $X_{ij} = X_i - X_j$, the same as that presented in the previous section, w_{ij} are the kernel weights depending on the distance between X_i and X_j in equation (1.7), the same as those defined in (1.7) in the rMAVE section.

As we did for the MAWVE estimator, we can show that

$$E\left[\sum_{i=1}^{n}\{Z_i - a_j - b_j^T X_{ij}\}^2 w_{iG} w_{ij}|X_1,\cdots,X_n\right] = E\left[\sum_{i=1}^{n}\{Y_i - a_j - b_j^T X_{ij}\}^2 w_{ij}|X_1,\cdots,X_n\right]. \tag{2.12}$$

This indicates that the weighted objective function under censoring is equal to the unweighted objective function under non-censoring, conditional on the observed covariates $X_1,\cdots,X_n$. Therefore, the new estimator is called the OPWG estimator.

Specifically, we would calculate the matrix

$$\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}\hat{b}_j\hat{b}_j^T, \tag{2.13}$$

where $\hat{b}_j$ is b_j corresponding to the minimum of the weighted local linear regression in (2.11). Hence the first eigenvector of $\hat{\Sigma}$, $\hat{\beta}$, is an estimator of the index parameter β_0, where $\hat{\beta}$ satisfies $||\hat{\beta}|| = 1$.

Similar to MAWVE, OPWG can also be expressed in three steps. The definitions of symbols X_{ij}, $\hat{f}_\beta(X_j)$, $w_{i\hat{G}}$ and $K_h(u) = K(u/h)/h$ are given in the previous sections.

Step 1: Calculate the a_j^β and b_j^β such that

$$\begin{pmatrix} a_j^\beta \\ b_j^\beta \end{pmatrix} = \left\{ \sum_{i=1}^n w_{i\hat{G}} K_h(\beta^T X_{ij}) \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^T \right\}^{-1} \sum_{i=1}^n w_{i\hat{G}} K_h(\beta^T X_{ij}) \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} Z_i. \tag{2.14}$$

Step 2: Calculate the first eigenvector of

$$\hat{\Sigma} = n^{-1} \sum_{j=1}^n \hat{\rho}_j^\beta b_j^\beta (b_j^\beta)^T, \tag{2.15}$$

where $\hat{\rho}_j^\beta = \rho_n(\hat{f}_\beta(\beta^T X_j))$, which is defined in the MAWVE section, and $\hat{f}_\beta(\beta^T X_j) = n^{-1} \sum_{i=1}^n K_h(\beta^T X_{ij})$. Denote the eigenvector as β.

Step 3: iteratively repeat Step 1 and Step 2 until $\beta := \beta/||\beta||$ converges.

We denote the final converged OPWG estimator by $\hat{\beta}_{OPWG}$, which satisfies $||\hat{\beta}_{OPWG}|| = 1$.

2.3 Inference for CSIMs

2.3.1 Consistency

Huber (1964) introduced the concept of M-estimators. Generalized maximum likelihood estimation is proposed to the minimization of

$$\sum_{i=1}^n \rho(x_i, \theta), \tag{2.16}$$

where ρ is an arbitrary function that is differentiable with respect to θ and could be solved for the root of the derivative. Then the solution $\hat{\theta}$ to the minimization of (2.16) is called an M-estimator. We can easily find that $\hat{\beta}_{MAWVE}$ and $\hat{\beta}_{OPWG}$ are both M-estimators since the functions we used in (2.6) and (2.10) satisfy the conditions for $\rho(\cdot)$.

Next, we want to check if the M-estimators $\hat{\beta}_{MAWVE}$ and $\hat{\beta}_{OPWG}$ are consistent. Before proving the consistency, suppose $M_n(\beta)$ and $M(\beta)$ are criterion functions. M-estimator $\hat{\beta}_n$ maximizes a random criterion function

$$\beta \mapsto M_n(\beta).$$

We have the theorem below for the M-estimator $\hat{\beta}_n$ to be consistent.

Theorem 2.3-1. *(see van der Vaart, 1998) Let M_n be a random function and let M be a fixed function of β such that for every $\varepsilon > 0$*

$$\sup_{\beta} |M_n(\beta) - M(\beta)| \xrightarrow{P} 0,$$

$$\sup_{\beta:\ d(\beta,\beta_0)\geq\varepsilon} M(\beta) < M(\beta_0).$$

Then any sequence of estimators $\hat{\beta}_n$ with $M_n(\hat{\beta}_n) \geq M_n(\beta_0) - o_p(1)$ converges in probability to β_0.

For MAWVE, we have

$$M_n(\beta) = n/Q_n(\beta) = \frac{n}{\sum_{j=1}^{n} \sigma_n^2(X_j|\beta)},$$

where the M-estimator $\hat{\beta}_{MAWVE}$ maximizes the random function M_n.

Also from the idea of MAWVE, we may find a determinant function $M(\beta)$ such that it satisfies the conditions of Theorem 2.3-1 and β_0 maximizes it. Hence, through Theorem 2.3-1 we can get the conclusion that

$$\hat{\beta}_{MAWVE} \xrightarrow{P} \beta_0.$$

For the OPWG method, we have the same convergence conclusion,

$$\hat{\beta}_{OPWG} \xrightarrow{P} \beta_0.$$

2.3.2 Estimation for the Link Function

Fan and Gijbels (1994) proposed methods of data transformation and local linear fit to estimate the nonparametric regression function when data are censored. Lu and Burke (2005) applied their techniques to the censored single-index models to estimate the link function. Cai (2003) considered a weighted local linear method to estimate the nonparametric regression function with censored data.

For the CSIMs, after we obtain the estimator $\hat{\beta}$ of the single-index parameter, we use Cai's method to estimate the link function. The estimation procedure is outlined as follows. In fact, such an estimator, which coincides with a_j^{β} in the final step of the iterative algorithm of MAWVE or OPWG at each $\{X_j\}_{j=1}^n$, is the estimate of the link function at that point.

To estimate $g(\cdot)$, with $\hat{\beta}$ from the final step of the iterative algorithm, we calculate

$$U_i = \hat{\beta}^T X_i.$$

Thus, we have the data $\{(U_i, Z_i, \delta_i)\}_{i=1}^n$ with the weights $\{w_i = w_{i\hat{G}}\}_{i=1}^n$. Following Cai (2003), the weighted local linear least squares equation based on the censored data $\{(U_i, Z_i, \delta_i)\}_{i=1}^n$ is expressed as

$$\sum_{i=1}^{n} w_i \{Z_i - a_1 - a_2(U_i - u)\}^2 K_h(U_i - u). \tag{2.17}$$

By minimizing (2.17) with respect to a_1 and a_2, we can obtain the weighted local linear estimator of the link function $g(\cdot)$ at point u, $\hat{g}(u;\hat{\beta}) = \hat{a}_1$, which can be expressed as below,

$$\hat{g}(u;\hat{\beta}) = \sum_{i=1}^{n} m_i(u;\hat{\beta}) Z_i \Big/ \sum_{i=1}^{n} m_i(u;\hat{\beta}), \tag{2.18}$$

with

$$m_i(u;\hat{\beta}) = w_i K_h\left(U_i - u\right)\left\{S_{n,2} - (U_i - u) S_{n,1}\right\},$$

where

$$S_{n,l} = \sum_{i=1}^{n} w_i K_h (U_i - u) (U_i - u)^l, \quad l = 1, 2,$$

and h is the bandwidth used in the MAWVE and OPWG methods.

2.3.3 The Nonparametric 0.632 Bootstrap

We use the nonparametric 0.632 bootstrap, which is introduced by Efron and Tibshirani (1993 and 1997), to estimate the variance of $\hat{\beta}_{MAWVE}$ and $\hat{\beta}_{OPWG}$. For one sample set with sample size n, the probability for the subject in the sample not chosen after n samples is about

$$(1 - 1/n)^n \approx e^{-1} \approx 0.368.$$

So the expected bootstrap sample size from the original sample set is appearing to be $0.632n$.

The 0.632 bootstrap estimator has the following error form:

$$\widehat{Err}^{0.632} = 0.368 \cdot \overline{err} + 0.632 \cdot \widehat{Err}^{(1)},$$

where $\widehat{Err}^{0.632}$ denotes the 0.632 bootstrap estimation error, $\overline{err}$ presents apparent error rate (or resubstitution estimation error) and $\widehat{Err}^{(1)}$ is for leave-one-out bootstrap estimation error. Here leave-one-out bootstrap is the bootstrap procedure that leaves out one subject at one bootstrap time.

Thus for the 0.632 bootstrap, we resample the data with sample size $\tilde{n} = 0.632n$ for m times without replacement. Then we would have the bootstrap estimators $\hat{\beta}_1, \cdots, \hat{\beta}_m$. The variance of $\{\hat{\beta}_i\}$ approxiimates as the variance of $\hat{\beta}$.

Chapter 3

Simulation Studies

In this Chapter, two Monte Carlo simulation studies are performed to check the consistency and efficiency of the MAWVE and OPWG methods. In Section 3.1, we introduce a model which is developed from the "sine-bump" model and present the results from MAWVE and OPWG. Another model is called the exponential model, which is modified from Xia (2006). This exponential model is used in the second simulation study in Section 3.2. In Section 3.3, we conduct a small simulation study on the "sine-bump" model to estimate the variances and to construct the bootstrap confidence intervals of the index parameters with MAWVE and OPWG.

3.1 Simulation Study I: Sine-Bump Model

3.1.1 Model Introduction

Consider a censored single-index model which is developed from the "sine-bump" model (see Carroll, Fan, Gijbels & Wand, 1997; Yu & Ruppert, 2002):

$$Y = \sin\left\{\frac{\pi(\beta_0^T X - A)}{B - A}\right\} + \varepsilon, \tag{3.1}$$

where X is a trivariate generated independently from the standard normal distribution $N(0, 1)$, ε is independent of X and follows the normal distribution $N(0, \sigma^2)$, where σ is the standard deviation of the error. Set $A = \sqrt{3}/2 - 1.645/\sqrt{12}$ and $B = \sqrt{3}/2 + 1.645/\sqrt{12}$. Assume that C follows $N(\mu, 1)$, where μ determines the censoring rate. The true value of the parameter is set to be $\beta_0 = (1, 1, 1)^T/\sqrt{3}$.

In this simulation study, we consider the CSIM with respect to different σ and μ. We choose $\sigma = 0.1$ and 0.5, each with three different levels of censoring: $\mu = 0.5$, 1.0, and 2.0. When $\mu = 0.5$, the censoring rate is approximately 55%; when $\mu = 1.0$, the censoring rate is about 37%; when $\mu = 2.0$, the censoring rate is around 10%. The higher censoring rate, i.e. smaller μ, the more censored data we will get.

3.1.2 Simulation Results

The simulation is based on 1000 replications for different σ and μ using MAWVE and OPWG methods. In each replication, the sample size is 100. Thus we have 1000 estimates $\hat{\beta}_i$, $i = 1, \cdots, 1000$, from each replication. We calculate the mean of the estimates $\bar{\beta}$ and the Monte Carlo variance with 1000 replications. With the mean $\bar{\beta}$ and the Monte Carlo variance, we obtain the bias and the mean squared errors (MSE) of the estimators, which are used to check the performance of MAWVE and OPWG.

In the simulation, we choose $(1,2,3)/\sqrt{14}$ as the initial value of β in the algorithm. We use Gaussian kernel function $K(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$ with bandwidth $h = 0.14$. We set $\epsilon = 1/10$ and $c_0 = 0.01$ in the trimming function (1.10).

The results for the bias and MSE with respect to different σ and censoring rates are showed in Table 3.1. Smaller bias and MSE indicate good fitness of the estimates. We can easily see that the estimates have small bias and MSE, even when the censoring rate is 55%. For smaller σ, the bias and MSE of the estimates are smaller; when the censoring rate is smaller, i.e. μ is larger, the bias and MSE of the estimates are smaller too. From Table 3.1, we conclude that the MAWVE method performs slightly better than the OPWG method, since the MSE values of the MAWVE method are generally smaller than those of the OPWG method.

σ	$cr(\%)$	μ	$(\beta_1, \beta_2, \beta_3)$ Bias	$(\beta_1, \beta_2, \beta_3)$ MSE ($\times 100$)
			MAWVE	
0.1	10	2.0	(-0.001, 0.000, 0.001)	(0.031, 0.031, 0.028)
0.1	37	1.0	(-0.001, 0.000, 0.001)	(0.043, 0.047, 0.042)
0.1	55	0.5	(-0.002, 0.000, 0.000)	(0.074, 0.077, 0.071)
0.5	10	2.0	(-0.009, -0.010, -0.002)	(0.764, 0.833, 0.775)
0.5	36	1.0	(-0.019, -0.015, -0.006)	(1.404, 1.636, 1.588)
0.5	55	0.5	(-0.031, -0.026, -0.026)	(2.425, 3.559, 3.605)
			OPWG	
0.1	10	2.0	$(-0.001, -0.001, 0.001)$	(0.031, 0.031, 0.029)
0.1	37	1.0	$(-0.001, -0.001, 0.001)$	(0.042, 0.047, 0.041)
0.1	55	0.5	$(-0.002, -0.001, 0.001)$	(0.075, 0.079, 0.068)
0.5	10	2.0	$(-0.009, -0.011, -0.002)$	(0.855, 0.886, 0.831)
0.5	36	1.0	$(-0.018, -0.018, -0.007)$	(1.495, 1.844, 1.604)
0.5	55	0.5	$(-0.029, -0.036, -0.033)$	(2.664, 4.104, 4.542)

Table 3.1: Sine-Bump Model: Results of Monte Carlo simulations with the MAWVE method and the OPWG method (σ: standard deviation of the error; μ: censoring parameter; cr: censoring rate in percentage; Bias: bias of the estimates; MSE: mean squared error).

After obtaining the index parameter estimator $\hat{\beta}$, we can estimate the link function using weighted local linear method (see (2.18)) and compare it with the true link function. For example, we show the estimation from a typical simulated sample in Figure 3.1 with the MAWVE method, under the condition that standard deviation for the error $\sigma = 0.1$ and the censoring parameter $\mu = 1.0$ (the corresponding censoring rate is about 37%). Figure 3.2 is the results using the OPWG method under the same situation. In both figures, the link function estimations fit the true link function very well, except for some small bias in the peak.

3.2 Simulation Study II: Exponential Model

3.2.1 Model Introduction

Consider the exponential single-index model modified from Xia (2006).

$$Y = 1 + (\beta_0^T X)^2 \exp(\beta_0^T X) + \varepsilon, \tag{3.2}$$

where $X = (\mathrm{x}_1, \cdots, \mathrm{x}_{10})^T$, $\beta_0 = (2, 1, 0, \cdots, 0)^T/\sqrt{5}$, $\varepsilon \sim N(0, \sigma^2)$. Let $(\mathrm{x}_1 + 1)/2 \sim Beta(\tau, 1)$ and $P(\mathrm{x}_k = \pm 0.5) = 0.5$, $k = 2, 3, 4, \cdots, 10$. The censoring variable C follows $N(\mu, 1.5^2)$, where μ determines the censoring rate. Assume that the random variables $\mathrm{x}_1, \cdots, \mathrm{x}_{10}$, ε and C are independent with each other.

We conduct this simulation study with respect to different σ and μ values. We choose $\sigma = 0.1$ and 0.5, $\mu = 1.2$, 1.8 and 3.6. When $\mu = 3.6$, the censoring rate is about 10%; when $\mu = 1.8$, the censoring rate is around 39%; when $\mu = 1.2$, the censoring rate is approximately 54%.

3.2.2 Simulation Results

The simulation runs 250 replications for different σ and μ with the sample size $n = 200$ in each replication. Hence, similar to the "sine-bump" simulation, we have 250 estimates $\hat{\beta}_i$, $i = 1, \cdots, 250$ from each of the MAWVE and OPWG methods. We calculate the mean of the estimates $\bar{\beta}$ and Monte Carlo variance in 250 replications. The bias and MSE of the estimators are also calculated for comparison.

In the simulation, we use Gaussian kernel function $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$ with bandwidth $h = 0.22$. We also set $\epsilon = 1/10$ and $c_0 = 0.01$ in the trimming function (1.10). Use $(10, 9, 8, \cdots, 1)/\sqrt{385}$ as the initial value of β in the algorithm.

Table 3.2 shows the results from both the MAWVE and OPWG methods. Similar to the "sine-bump" model simulation study, we can find that when σ is smaller, the bias and MSE are smaller; when the censoring rate is smaller, the bias and MSE are smaller as well. Smaller bias and MSE indicate better estimates. From Table 3.2, we can notice that both the MAWVE and OPWG methods perform well on the simulation data. MAWVE works better than OPWG, since the results of the MAWVE method always show smaller bias and MSE. Comparing the results in Table 3.1, we can also find that the bias and MSE of larger dimension of covariates ($p = 10$) appear larger than those of smaller dimension of covariates ($p = 3$).

We can also estimate the link function using the weighted local linear method (see (2.18)), and compare it with the true link function. For illustration, we show one example in Figure 3.3 with the MAWVE method, in which the data are from a typical simulated sample with standard deviation for the error $\sigma = 0.1$ and the censoring parameter $\mu = 1.8$ (the corresponding censoring rate is about 39%). Figure 3.4 shows the estimated link function with the OPWG method, along with the true link function from a same simulated sample.

Figure 3.3 and Figure 3.4 show the performance of the estimation of link function. In both figures, the link function estimations fit the true link function well, except for the right boundary of the x-axis. This results from lack of information, such as number of data, in the boundary.

3.3 A Small Simulation Study On Variance Estimation and Bootstrap Confidence Interval

As illustrated in Section 2.3, we can make inference on β if we know the variance of the estimator. To do this, we consider the bootstrap estimation of the variance. We use the "sine-bump" model as an example to explain the method. In a simulation of this model, we set the sample size $n = 100$ and the replication number $r = 100$. In the ith replication, we estimate the variance of the estimator $\hat{\beta}_i$ using the 0.632 bootstrap without replacement (see Efron & Tibshirani, 1993 or explanation in Chapter 2). The bandwidth used is $h = 0.14$ and the number of bootstrap times is $b = 100$.

For the accuracy of the bootstrap variance, the results are shown in Table 3.3. We compare it with the Monte Carlo variance of the $r = 100$ estimates. We also calculate the coverage probability of the bootstrap confidence intervals.

We find that when σ is small, the bootstrap variance agrees with the Monte Carlo variance. Otherwise, the bootstrap variance tends to overestimate the variances of the estimators.

We notice that the average coverage probability deviates from the nominal level 95%, which reflects the inaccuracy of the bootstrap variance. OPWG has higher coverage probability than MAWVE, but it appears to have larger Monte Carlo variance and bootstrap variance. Hence, we may conclude that OPWG is less efficient but has better coverage probability than MAWVE.

This is a small scale simulation study. We believe that we can improve the results by increasing the sample size and the number of bootstrap times. However, a faster computing method is needed to achieve this goal.

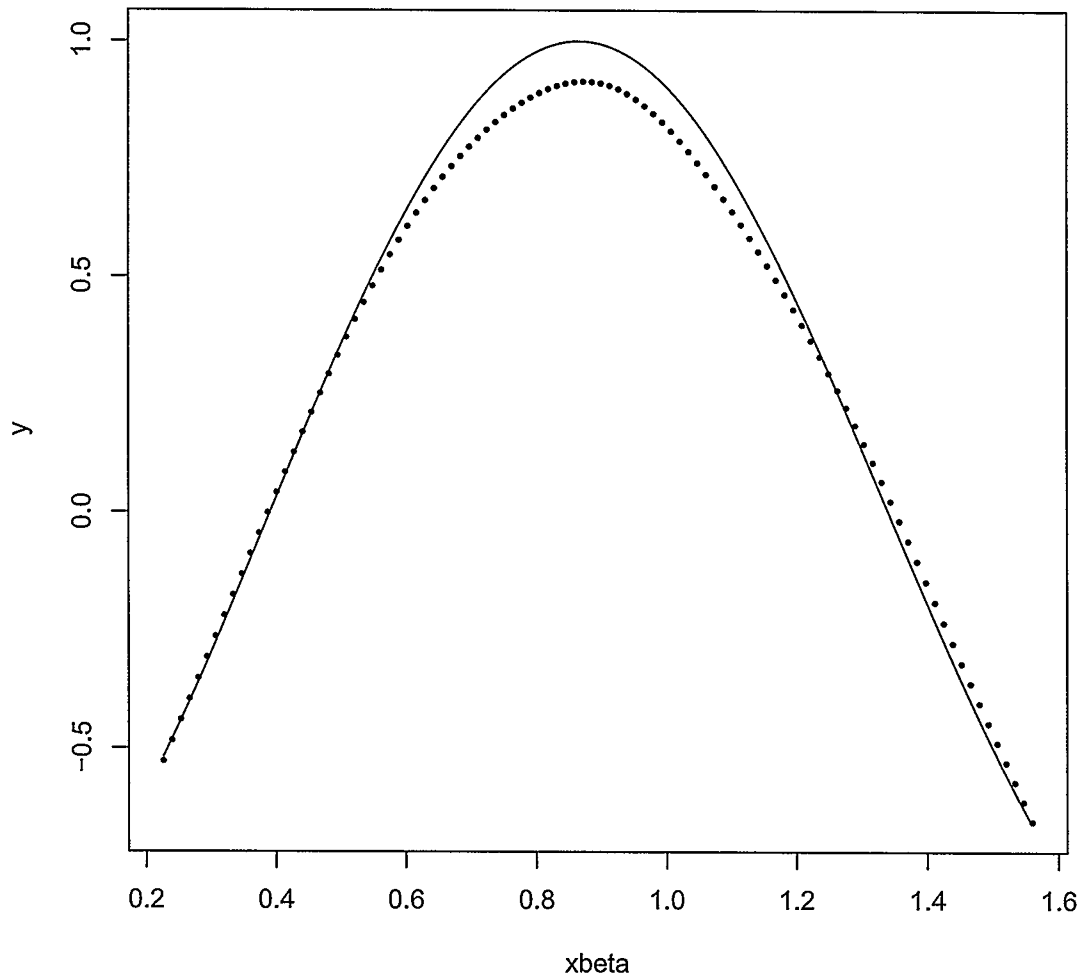

Figure 3.1: Sine-Bump Model: Curve estimation for the link function (standard deviation for the error $\sigma = 0.1$, censoring rate $cr = 37\%$) with the MAWVE method and a typical simulated sample. The dotted curve is the estimated link function. The solid curve is the true link function.

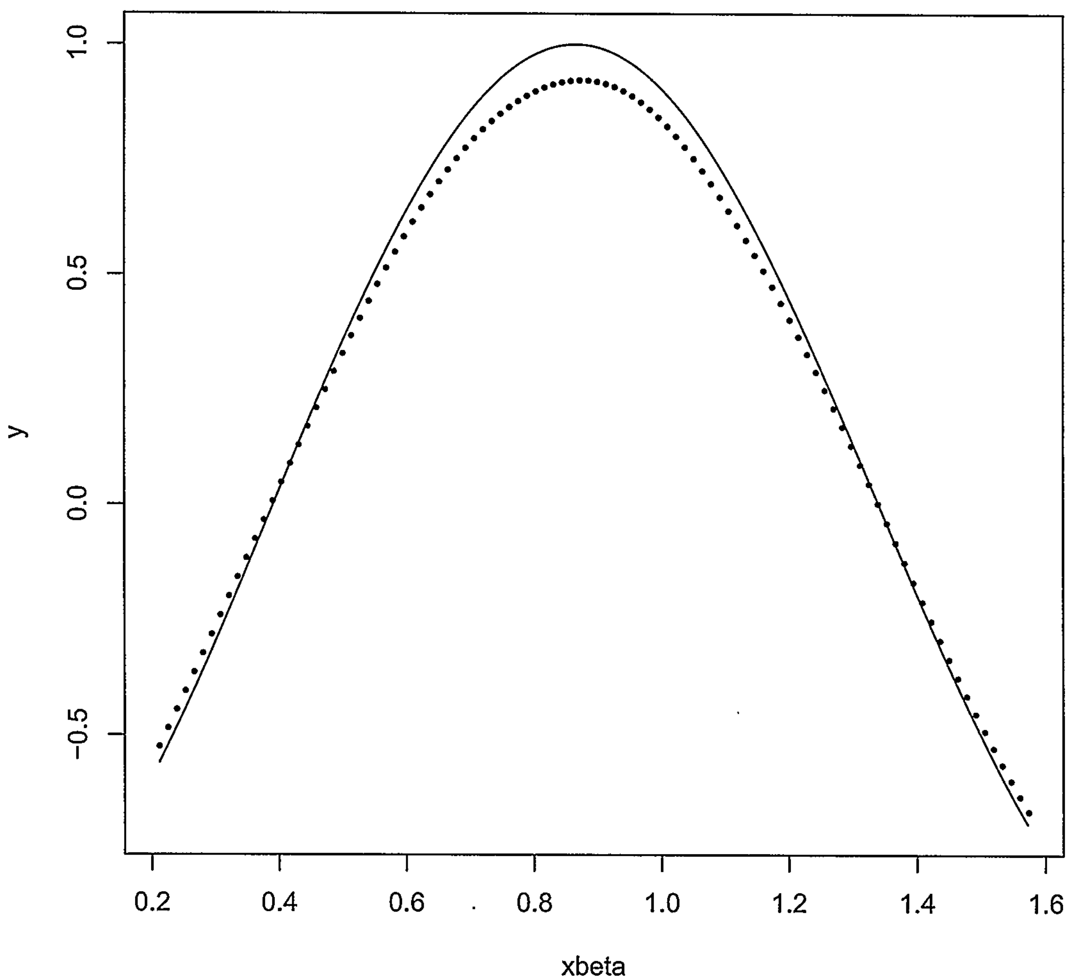

Figure 3.2: Sine-Bump Model: Curve estimation for the link function (standard deviation for the error $\sigma = 0.1$, censoring rate $cr = 37\%$) with the OPWG method and a typical simulated sample. The dotted curve is the estimated link function. The solid curve is the true link function.

σ	$cr(\%)$	μ		MAWVE									
				β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
0.1	10	3.6	Bias:	-0.008	0.012	0.000	0.001	0.000	0.000	0.001	-0.001	0.000	-0.001
			MSE($\times$100):	0.012	0.036	0.036	0.036	0.036	0.033	0.033	0.036	0.036	0.032
0.1	39	1.8	Bias:	-0.014	0.020	0.000	0.003	-0.004	0.001	0.001	-0.001	-0.003	-0.002
			MSE($\times$100):	0.031	0.073	0.086	0.120	0.090	0.078	0.086	0.086	0.089	0.079
0.1	54	1.2	Bias:	-0.025	0.014	-0.004	0.003	-0.005	0.005	-0.001	0.000	0.001	0.002
			MSE($\times$100):	0.654	0.600	0.369	0.263	0.255	0.333	0.139	0.233	0.287	0.121
0.5	10	3.6	Bias:	-0.024	0.015	-0.001	0.004	-0.003	0.002	0.006	-0.005	0.000	-0.004
			MSE($\times$100):	0.154	0.288	0.281	0.383	0.260	0.320	0.324	0.270	0.319	0.314
0.5	39	1.8	Bias:	-0.077	0.005	0.007	0.015	0.005	0.009	0.016	0.001	-0.002	0.002
			MSE($\times$100):	1.960	1.306	1.142	1.492	1.252	1.196	1.299	1.256	1.351	1.043
0.5	54	1.2	Bias:	-0.151	0.006	0.055	0.047	0.041	0.032	0.002	0.024	0.002	0.000
			MSE($\times$100):	5.079	1.745	2.795	3.234	3.020	2.697	2.314	2.081	1.805	1.667
				OPWG									
0.1	10	3.6	Bias:	-0.001	-0.001	0.000	0.001	0.000	0.001	0.001	0.000	0.000	-0.001
			MSE($\times$100):	0.009	0.034	0.040	0.037	0.037	0.035	0.034	0.040	0.034	0.034
0.1	39	1.8	Bias:	-0.013	0.014	-0.001	0.004	-0.003	0.001	0.000	-0.003	-0.005	-0.004
			MSE($\times$100):	0.034	0.080	0.108	0.146	0.124	0.105	0.095	0.117	0.125	0.104
0.1	54	1.2	Bias:	-0.023	0.020	-0.002	0.000	-0.006	0.002	0.002	0.000	-0.003	-0.001
			MSE($\times$100):	0.162	0.173	0.344	0.304	0.253	0.265	0.241	0.248	0.227	0.187
0.5	10	3.6	Bias:	-0.018	0.004	0.000	0.004	-0.002	0.002	0.006	-0.005	0.001	-0.003
			MSE($\times$100):	0.139	0.255	0.281	0.364	0.296	0.348	0.317	0.281	0.302	0.319
0.5	39	1.8	Bias:	-0.089	0.001	0.001	0.013	-0.002	0.001	0.006	-0.007	-0.011	0.002
			MSE($\times$100):	2.435	1.603	1.265	1.684	1.498	1.556	1.814	1.218	1.345	1.381
0.5	54	1.2	Bias:	-0.213	-0.04	0.033	0.029	0.015	0.007	0.018	0.011	0.006	-0.001
			MSE($\times$100):	9.164	3.802	4.060	4.278	3.395	3.555	3.126	3.563	3.455	3.219

Table 3.2: Exponential Model: Results of Monte Carlo simulations with MAWVE and OPWG methods (σ: standard deviation of the error; μ: censoring parameter; cr: censoring rate in percentage; Bias: bias of the estimates; MSE: mean squared error).

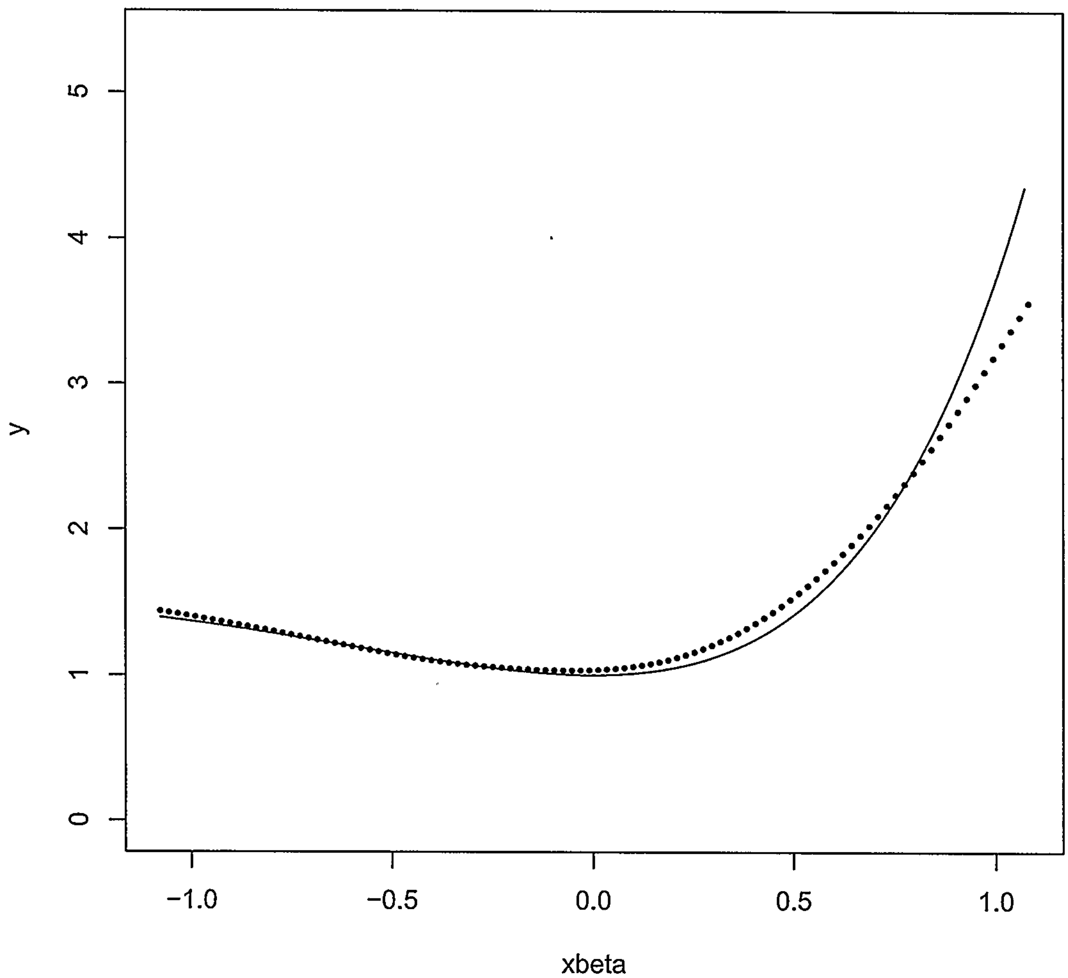

Figure 3.3: Exponential Model: Curve estimation for the link function (standard deviation for the error $\sigma = 0.1$, censoring rate $cr = 39\%$) with the MAWVE method and a typical simulated sample. The dotted curve is the estimated link function. The solid curve is the true link function.

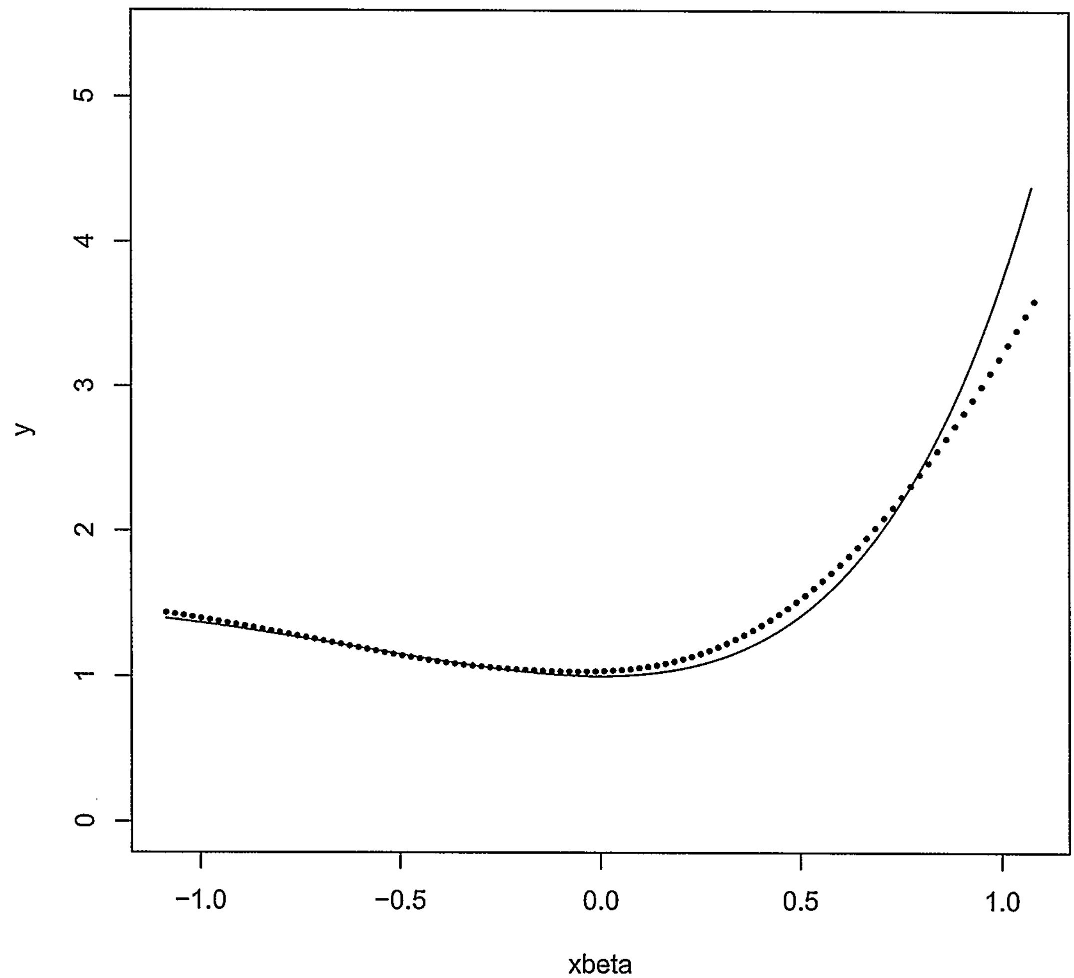

Figure 3.4: Exponential Model: Curve estimation for the link function (standard deviation for the error $\sigma = 0.1$, censoring rate $cr = 39\%$) with the OPWG method and a typical simulated sample. The dotted curve is the estimated link function. The solid curve is the true link function.

σ	cr(%)	μ	Monte Carlo Variance (×100)	Bootstrap Variance (×100)	Bootstrap CI Coverage (%)
				MAWVE	
0.1	10	2.0	(0.032, 0.030, 0.026)	(0.020, 0.021, 0.020)	(85, 83, 87)
0.1	37	1.0	(0.041, 0.038, 0.041)	(0.034, 0.034, 0.034)	(88, 93, 89)
0.1	55	0.5	(0.070, 0.069, 0.072)	(0.067, 0.068, 0.058)	(90, 88, 88)
0.5	10	2.0	(0.656, 0.730, 0.615)	(0.668, 0.882, 0.910)	(90, 93, 85)
0.5	36	1.0	(1.605, 1.341, 1.458)	(1.342, 2.603, 2.386)	(87, 92, 91)
0.5	55	0.5	(2.373, 2.981, 2.053)	(2.142, 5.043, 4.462)	(91, 95, 93)
				OPWG	
0.1	10	2.0	(0.086, 0.105, 0.088)	(0.080, 0.077, 0.077)	(86, 83, 90)
0.1	37	1.0	(0.094, 0.109, 0.091)	(0.096, 0.095, 0.091)	(91, 92, 89)
0.1	55	0.5	(0.120, 0.133, 0.119)	(0.157, 0.238, 0.212)	(92, 93, 93)
0.5	10	2.0	(0.815, 0.871, 0.622)	(0.993, 1.958, 2.412)	(91, 93, 92)
0.5	36	1.0	(1.873, 3.871, 3.001)	(1.811, 5.012, 5.027)	(95, 97, 92)
0.5	55	0.5	(2.562, 5.883, 8.938)	(2.95, 10.136, 10.001)	(98, 96, 94)

Table 3.3: Sine-Bump Model: Results of Monte Carlo variances, Bootstrap variances and Bootstrap confidence intervals with MAWVE and OPWG methods (σ: standard deviation of the error; μ: censoring parameter; cr: censoring rate in percentage; Bootstrap CI Coverage: the probability that the true parameter value falls in the 95% bootstrap confidence interval).

Chapter 4

Case Study: Stanford Heart Transplantation Data

In this chapter, we consider Stanford Heart Transplantation data (see Miller & Halpern, 1982). We use the MAWVE and OPWG methods to estimate the index parameter under the censored single-index model. We first introduce the data and the necessary background in Section 4.1. We perform analysis on the data and compare the results with other methods, such as the linear least squares estimates using the Buckley-James method in Section 4.2.

4.1 Data Description

The Stanford heart transplantation program was conducted from October 1967 to February 1980. In this program, 184 patients had received heart transplants. Their survival times, uncensored or censored, were recorded by the end of the program, in February 1980. Their ages at the time of the first transplant, noted by "age", were recorded too. The patients' T5 mismatch scores measure the degree of tissue incompatability between the initial donor and recipient hearts with respect to HLA antigens.

We do not use the whole data of the Stanford heart transplantation program. We refer readers to Miller & Halpern (1982) for the whole data set. We use base-10 logarithm of the survival time as the response. As Miller & Halpern (1982) and Zhou (1992) did, we only consider 157 patients who survived for at least 10 days after transplantation. The censoring rate in this data set is 36%.

4.2 Study Results

There are several regression techniques used on the Stanford heart transplantation data, we refer to Buckley & James (1979) and Huang & Jin (2007) for the Buckley-James method, Miller (1976) for the least squares regression method, and Cox (1972) for the Cox regression model. These authors considered linear models but suggested nonlinear effects of age by introducing a quadratic term age^2.

In this analysis of Stanford heart transplantation data, we consider it with a CSIM. We use the 157 patients' base-10 logarithm survival time (Z_i) and two covariates: T5 (X_{1i}) and age (X_{2i}). We define $X_i = (X_{1i}, X_{2i})^T$. We also denote the censoring status (dead=1, censored=0) as δ_i, the original log10 survival time as Y_i, and the censoring variable as C_i, then we have the data set $\{(Z_i, X_i, \delta_i)\}_{i=1}^{157}$ with a CSIM

$$Y_i = g(\beta_1 X_{1i} + \beta_2 X_{2i}) + \varepsilon_i, \tag{4.1}$$

where $Z_i = \min\{Y_i, C_i\}$ and $\delta_i = I_{\{Y_i \leq C_i\}}$, for $i = 1, \cdots, 157$.

Our model is different from the linear models considered by other authors. We don't introduce any higher order polynomial terms of the variables such as age^2 in the model. We assume an arbitrary link function $g(\cdot)$, and let the model automatically adapt to the data for possible nonlinear effects.

The results from the MAWVE and OPWG methods compared with the Buckley-James method are presented in Table 4.1. We use Gaussian kerneal function $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$ and bandwidth $h = 0.4$, $\epsilon = 1/10$ and $c_0 = 0.01$ in the trimming function (1.10), and $\tilde{\beta} = (1, 2)/\sqrt{5}$ as the initial value in the algorithm.

To calculate the variance (or standard deviation) of the estimators, we use the non-parametric 0.632 bootstrap for 100 resampling times without replacement (see Efron & Tibshirani, 1993 and the explanation in Chapter 2). That is, the sample size for each bootstrap is 96. Table 4.1 shows the estimation results and bootstrap standard deviations, including the results from the linear model with three covariates (T5, age, age^2) fitted by the Buckley-James method for comparison.

Model	Method	Estimates (T5, age)	Bootstrap Standard Deviation
CSIM	MAWVE	(0.3369, 0.9003)	(0.2458, 0.1279)
CSIM	OPWG	(0.4332, 0.8541)	(0.2313, 0.1735)
		Estimates (T5, age, age^2)	Standard Error
Linear	Buckley-James	(0.0289, 0.1140, -0.0017)	(0.1432, 0.0537, 0.6624 $\times 10^{-3}$)

Table 4.1: Stanford Heart Transplantation Data: Results of the single-index model and the linear model with MAWVE, OPWG and Buckley-James methods.

In the linear model, both age and age^2 are significant, T5 is not. In the CSIM, we have similar conclusion, age is significant and T5 is not. However, we don't use high order terms such as age^2 to model the nonlinear effect of age, instead, we use the nonparametric link function to model possible nonlinear effects and interactions. Figure 4.1 and 4.2 are plots of the estimated link function with the MAWVE and OPWG methods when the bandwidth $h = 0.4$. It is clear that the identity link is not appropriate for this data set if we fit a linear model with only two covariates T5 and age in the model. These figures also suggest for practitioners to include high order terms such as T5 $\times$ age and age^2 if they want to use a linear model.

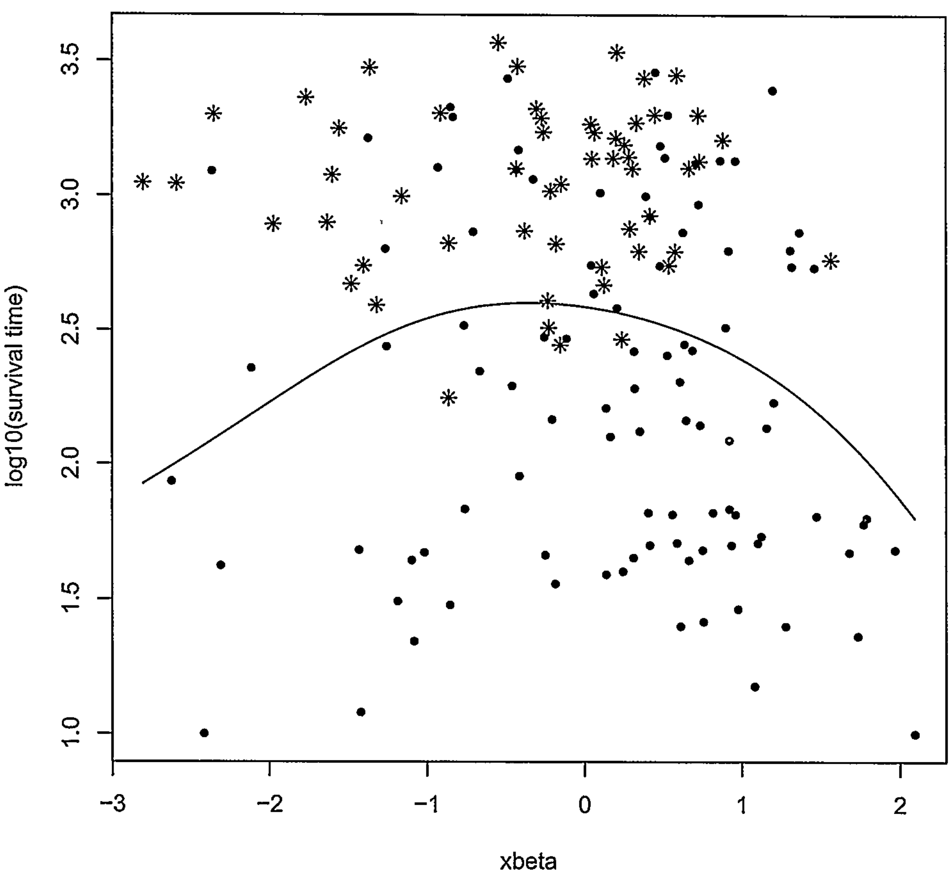

Figure 4.1: Single-index model for Stanford Heart Transplantation Data. The data are presented by stars for censored data ($\delta = 0$) and by solid points for observed data ($\delta = 1$). The solid curve is the estimated link function with MAWVE method.

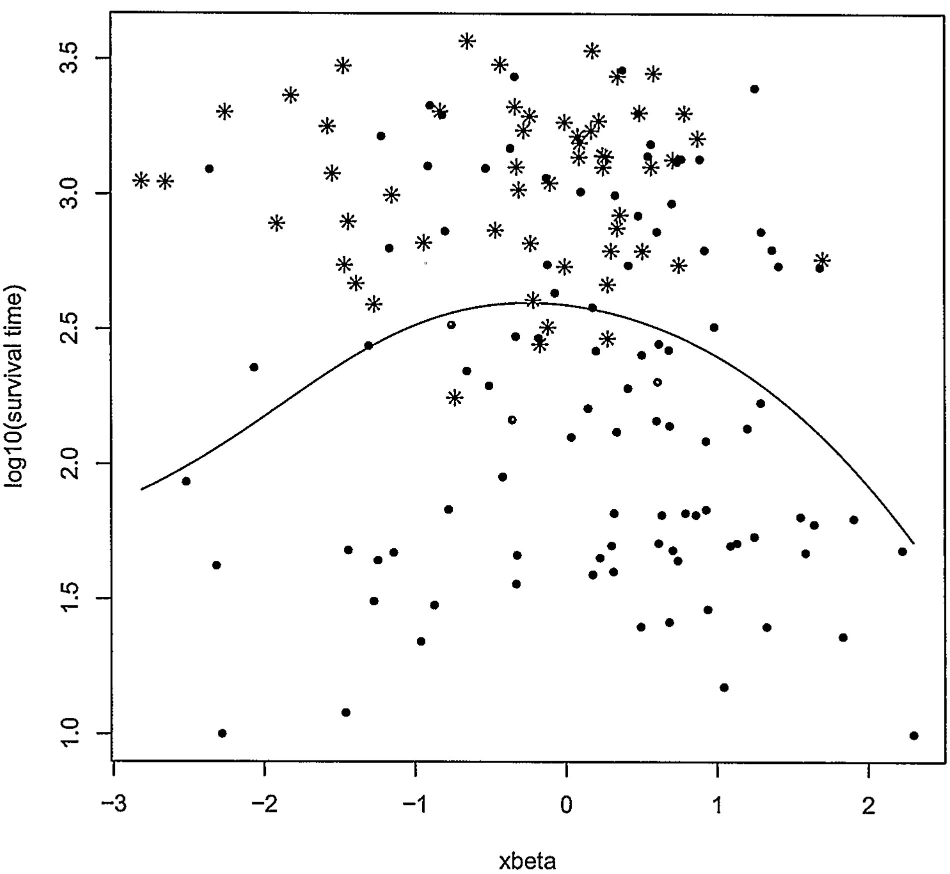

Figure 4.2: Single-index model for Stanford Heart Transplantation Data. The data are presented by stars for censored data ($\delta = 0$) and by solid points for observed data ($\delta = 1$). The solid curve is the estimated link function with OPWG method.

Chapter 5

Conclusion and Discussion

The MAWVE and OPWG methods are proposed as tools for studying the censored single-index models. These two methods are extensious from rMAVE and rOPG of Xia (2006) for censored data analysis. When there is no censored response, MAWVE and OPWG are equivalent to rMAVE and rOPG respectively.

The MAWVE and OPWG methods are both based on the weighted least squares principle. The weights are Kaplan-Meier weights, the accuracy of the parameter estimation depends on how accurate these weights are in estimating the theoretical weights and other unknown quantities such as the link function. The factors affecting the performance of the estimation include the standard deviation of the error, the censoring rate, the sample size and the dimension of covariates. From the simulation studies, we find that when the sample size is moderate ($n = 100$), the dimension of covariates is low ($p < 10$), the two methods work quite well. When the dimension of covariates is large ($p \geq 10$), the performance becomes worse. It seems that large sample size is needed when p becomes large. We hope that we can investigate the issue of small n and large p when data are censored in our future work.

Since we are unable to establish the asymptotic distribution theory of the estimators, we prove the consistency of the estimates and use the bootstrap method to make inference about the parameters. Our simulation results show that it works reasonably well. But the bootstrap method requires a lot of computing time, it is desirable to develop a faster inference procedure for the CSIMs.

We have successfully applied our methods to the Stanford heart transplantation data. We are able to discover the nonlinear effects of the covariates. To get similar results, the users of linear models must include some high order terms chosen subjectively. While our methods are data-driven methods — they do it automatically.

From all above, we conclude that the CSIMs are useful models in analyzing censored data, they are more flexible in modelling effects of covariates than the traditional linear models for censored data. The proposed MAWVE and OPWG methods have some good features in estimating the CSIMs. They are promising techniques in dimension reduction. It is worthwhile to further explore the large sample properties of the new estimators and the computational issues regarding the CSIMs in future work.

Bibliography

[1] Brillinger, D.R. (1983). A generalized linear model with "Gaussian" regressor variables. *A Festschrift for Erich L. Lehmann*, 97-114.

[2] Buckley, J., & I.R. James (1979). Linear regression with censored data. *Biometrika* 66, 429-436.

[3] Cai, Z. (2003). Weighted Local Linear Approach to Censored Nonparametric Regression. *Recent Advances and Trends in Nonparametric Statistics*, M.G. Akritas & D.N. Politis, 217-231.

[4] Carroll, R.J., J. Fan, I. Gijbels, & M.P. Wand (1997). Generalized parrtially linear single-index models. *Journal of the American Statistical Associaltion* 92, 477-489.

[5] Chernick, M.R. (1999). *Bootstrap methods: a practitioner's guide.* New York: John Wiley & Sons, Inc.

[6] Čížek, P., & W. Härdle (2006). Robust estimation of dimension reduction space. *Computational Statistics & Data Analysis* 51, 545-555.

[7] Cox, D.R. (1972). Regression models and life-tables (with Discussion). *Journal of the Royal Statistical Society* B 34, 187-220.

[8] Davison, A.C., & D.V. Hinkley (1997). *Bootstrap method and their application.* Cambridge University Press.

[9] Delecroix, M., W. Härdle, & M. Hristache (2003). Efficient estimation in conditional single-index regression. *Journal of Multivariate Analysis* 86, 213-226.

[10] Efron, B. (1983). Estimating the error rate of a prediction rule: improvement on crossvalidation. *Journal of the American Statistical Association* 78, 316-331.

[11] Efron, B., & R. Tibshirani (1993). *An introduction to the bootstrap.* New York: Champman and Hall.

[12] Efron, B., & R. Tibshirani (1997). Improvement on cross-validation: The .632 bootstrap methods. *Journal of the American Statistical Associaltion* 92, 548-560.

[13] Escobar, L.A., & W.Q. Meeker (1992). Assessing influence in regression analysis with censored data. *Biometrics* 48, 507-528.

[14] Györfi, L., M. Kohler, A. Krzyżak, & H. Walk (2002). *A distribution-free theory of nonparametric regression.* New York: Springer-Verlag.

[15] Härdle, W., & T.M. Stoker (1989). Investigating smooth multiple regression by method of average derivatives. *Journal of the American Statistical Associaltion* 84, 986-995.

[16] Härdle, W., P. Hall, & H. Ichimura (1993). Optimal smoothing in single-index models. *Annals of Statistics* 21, 157-178.

[17] Härdle, W., M. Müller, S. Sperlich, & A. Werwatz (2004). *Nonparametric and semiparametric models.* Berlin, Heidelberg: Springer-Verlag.

[18] Hosmer, D.W., & S. Lemeshow (1999). *Applied survival analysis: regression modeling of time to event.* Canada: John Wiley & Sons.

[19] Hristache, M., A. Juditsky, & V. Spokoiny (2001). Direct estimation of the single-index coefficients in single-index models. *Annals of Statistics* 29, 595-623.

[20] Huang, J., S. Ma, & H. Xie (2007). Least absolute deviations estimation for the accelerated failure time model. *Statistica Sinica* 17, 1533-1548.

[21] Huang, L., & Jin, Z (2007). LSS: Splus/R program for the accelerated failure time model based on least-squares principle. *Computer Methods and Programs in Biomedicine* 86, 45-50.

[22] Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics.* 35, 73-101.

[23] Ichimura, H. (1993). Smiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* 58, 71-120.

[24] James, I.R. (1986). On estimating equations with censored data. *Biometrika* 73, 35-42.

[25] Jin, Z., D.Y. Lin, & Z. Ying (2006). On least-squares regression with censored data. *Biometrika* 93, 147-161.

[26] Kaplan, E.L., & P. Meier (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association.* 53, 457-481.

[27] Klein, J.P., & M.L. Moeschberger (2003). *Survival anlysis: techniques for censored nd truncated data.*New York: Springer-Verlag.

[28] Krause, A., & M. Olson (2005). *The basics of S-PLUS.* New York : Springer-Verlag.

[29] Lee, E.L., & J.W. Wang (2003). *Statistical methods for survival data analysis.* Hobken, New Jersey: John Wiley & Sons.

[30] Li, K.-C., & N. Duan (1989). Regression under link violation. *Annals of Statistics* 17, 1009-1052.

[31] Lu, X., & M.D. Burke (2005). Censored multiple regression by the method of average derivatives. *Journal of Multivariate Analysis* 95, 182-205.

[32] Lu, X., & T.L. Cheng (2006). Randomly censored partially linear single-index models. *Journal of Multivariate Analysis* 98, 1895-1922.

[33] Miller, R.G. (1976). Least squares regression with censored data. *Biometrika* 63, 449-464.

[34] Miller, R.G., & J. Halpern (1982). Regression with censored data. *Biometrika* 69, 521-531.

[35] Nichiyama, Y., & P.M. Robinson (2005). The bootstrap and the edgeworth correction for semiparametric averaged derivatives. *Econometrica* 73, 903-948.

[36] Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society* 51, 406-413.

[37] Powell, J.L., J.H. Stock, & T.M. stoker (1989). Semiparametric estimation of index coefficients. *Econometrica* 57, 1403-1430.

[38] Satten, G. A., & Datta, S. (2001). The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician* 55, no.3, 207-210.

[39] Stoker, T.M. (1986). Consistent estimation of scaled coefficients. *Econometrica* 54, 1461-1481.

[40] Stute, W. (1993). Consistent estimation under random censorship when covariables are available. *Journal of Multivariate Analysis* 45, 89-103.

[41] Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinivian Journal of Statistics* 23, 461-471.

[42] Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica* 9, 1089-1102.

[43] van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge: Cambridge University Press.

[44] Xia, Y., H. Tong, W.K. Li, & L. Zhu (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society, Series B* , 64, 363-410.

[45] Xia, Y. (2006). Asymptotic distributions for two estimators of the single-index model. *Econometric Theory* 22, 1112-1137.

[46] Yu, Y., & D. Ruppert (2002). Penalized Spline Estimation for Partially Linear Single-Index Models. *Journal of the American Statistical Association* 97, 1042-1054.

[47] Zhou, M. (1992). M-estimation in censored linear models. *Biometrika* 20, 1002-1021.

[48] Zhou, M. (1994). Linear regression with censoring. *Journal of Multivariate Analysis* 49, 179-201.