ANALYSIS OF FAILURE TIME DATA WITH MISSING AND INFORMATIVE AUXILIARY COVARIATES

by

Lipika Ghosh

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2011

Approved by:

Dr. Yanqing Sun

Dr. Jiancheng Jiang

Dr. Weihua Zhou

Dr. David Weggel

ABSTRACT

LIPIKA GHOSH. Analysis of failure time data with missing and informative auxiliary covariates.
(Under the direction of DR. YANQING SUN & DR. JIANCHENG JIANG)

In this dissertation we use Cox's regression model to fit failure time data with continuous informative auxiliary variables in the presence of a validation subsample. The work is motivated by a common problem of missing or mismeasured covariates in survival analysis as a result of which the relative risk function is not available for all the subjects in the sample. Here we introduce a two-stage procedure for estimating the parameters in the model. We first estimate the induced relative risk function with a kernel smoother based on the validation subsample, and then improve the estimation by utilizing the information from the non-validation subsample and the auxiliary observations from the primary sample. Asymptotic normality of the proposed estimator is obtained. The proposed method allows one to efficiently model the failure time data with informative multivariate auxiliary covariate. Comparison of the proposed approach with several existing methods is made via simulations. A real dataset is analyzed to illustrate the proposed method.

ACKNOWLEDGEMENTS

It is a pleasure to thank those who helped me in preparing and completing this thesis. First and foremost I owe my deepest gratitude to my advisors Dr Yanqing Sun and Dr Jiancheng Jiang for their valuable guidance, enormous patience, continuous encouragement and constant support. Throughout my research they instilled in me the interest and understanding of the subject and helped me to grow as an independent thinker. I also gratefully acknowledge the partial support I received from NSF grant DMS-0604576.

I am very thankful to the other members of my thesis committee Dr Weihua Zhou, Dr David Weggel and Dr Ming Dai. Besides them, my sincere thanks to all my professors at UNC Charlotte for their sincere effort in making me more knowledgable. Appreciation also extends to Dr Joel Avrin and Dr Mohammad Kazemi for their precious help and suggestions throughout my PhD. I am happy to acknowledge my debt to my high school teachers Anjana Bose and Bhaskar Bose for encouraging and motivating me to study mathematics and statistics.

Thanks are also due to my fellow graduate students. From time to time I received valuable advice and moral support from them. It has been a wonderful experience working in such a pleasant environment.

Finally I would like to thank my caring family, the greatest friends and loving husband Jaison. It would have been impossible to accomplish my goals without their immense support. I dedicate this thesis to them.

Last but not least, I am thankful to God for giving me the strength and ability to face all the difficult times in the past six years and complete this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Motivation & Background

In epidemiologic studies the researchers often wish to investigate the association between a particular risk factor or exposure variable with disease. The exposure variable may be hard or expensive to measure whereas some auxiliary variables vector are easy to measure for all subjects in the study cohort. Statistical methods that take advantage of existing auxiliary information about an expensive exposure variable are desirable in practice. For example, in a large scale nutritional study, it would be prohibitively expensive to obtain the exact dietary intake on each individual. Instead, a self administered quantitative food questionnaire is conducted on all subjects and a validation set consisting of a subset of the full study cohort is selected. The individuals in the validation set are asked to provide more detailed and accurate dietary information. Although the true covariates are missing, there exist some surrogates or auxiliary measurements which convey information about them and serve as common proxy measure. How to utilize the available auxiliary information is important for achieving higher statistical efficiency in the estimation of the effect of covariates. In this thesis, we study censored failure time regression with a continuous auxiliary covariate vector.

A variety of authors have contributed their work to this field. Related works include Prentice (1982), Pepe (1989), Lin and Ying (1993), Hughes (1993), Lipsitz and Ibrahim (1996), Zhou and Wang (2000), Fan and Wang (2009), Liu, Wu and Zhou (2010), etc. In particular, Prentice (1982) introduced a partial likelihood estimator based on the induced relative risk function. This method was further developed by Pepe (1989) using parametric modeling. Zhou and Pepe (1995) proposed an estimated partial likelihood method for discrete auxiliary covariates to relax the parametric assumptions on the frequency of events and the underlying distributions of covariates. This method was extended by Zhou and Wang (2000) to deal with continuous auxiliary variables, based on the Nadaraya-Watson kernel smoother method (Nadaraya, 1964; Watson, 1964). Fan and Wang, Liu (2009) and Wu, Zhou (2010) used the same approach for multivariate failure time data with auxiliary covariates. While Zhou and Wang's (2000) approach is useful in certain situations, there are some restrictions on it. First, the approach is effective only when the auxiliary variable W is of low dimension so that "curse of dimensionality" in nonparametric smoothing can be avoided. Secondly, it requires that, conditionally on X, W provides no additional information about the hazard of failure; that is, all of the effects of W on failure and censoring are mediated through X, which is somewhat restricted since W may not be a true surrogate and depends on the failure given X. In addition, the resulting estimators of the parameters are not efficient if the ratio of validation observations is small, which is mainly due to the fact that their smoothing method only used the data in the validation set to predict the induced relative risk function r_j for j in the non-validation set. Since the important information from the observations in the non-validation subsample is not fully utilized, this method cannot be efficient in certain situations. We here propose a new method to deal with the problems. The proposed method allows W to be highly dimensional and to be informative in the sense that, conditional X, it may provide additional information on the hazard of failure. We first estimate the induced relative risk function with a kernel smoother based on the validation sample, and then improve the estimation by utilizing the information on the incomplete observations from the non-validation subsample. In addition, the local linear smoother (see for example in Fan and Gijbels, 1996) is employed to enhance the performance of the kernel smoother in Zhou and Wang (2000) at the boundary regions. The newly proposed method will be expected to improve the efficiency of the estimators of parameters in various situations. Asymptotic normality of the proposed estimators is derived. The results in theory and practice show that the proposed method is efficient in certain situations even if auxiliary variable W is not very informative about X.

In the following sections of Chapter 1 we give a brief introduction of proportional hazards models and a brief overview of the remaining dissertation.

1.2 Proportional Hazard Models and Partial Likelihood

Proportional hazard models are popular models used in survival analysis that can be used to assess the importance of various covariates in the survival times of individuals or objects through the hazard function. In survival data, we need special techniques to explore the relationship between the survival times of an individual and the explanatory variables. The most frequently used model was proposed by Cox(1972) and is widely known as the Cox Proportional Hazards model. Prior to Cox Regression the leading approach to analyze mutivariate survival data was parametric which requires one to know the nature of the survival distribution. Also we need to be careful about violation of the model assumptions for some parametric models. Cox' regression model has the following advantages over those methods.

(1) Cox regression is a distribution free modeling approach.

(2) This model allows us to estimate the regression coefficients without specifying the baseline hazard function, and the estimates depend on the rank of the event times, not their numerical values.

(3) Since the model depends on ranks, the coefficients remain unchanged by any monotonic transformation of the hazard function.

(4) This model permits us to incorporate time varying covariates.

(5) With appropriate specification Cox's model can be employed to answer many challenging research questions.

1.2.1 Formulation of the Cox model

Let $\mathbf{z} = (z_1, z_2, ..., z_p)$ be a $p \times 1$ vector of covariates of risk factors and $\lambda(t|z)$ be the hazard function which depend on the covariates $\mathbf{z}$. The generalized form of the proportional hazards model is

$$\lambda(t|z) = \lambda_0(t) exp(\beta_1 z_1 + ... + \beta_p z_p)$$

where $\lambda_0(t)$ is the underlying baseline hazard function at time t and $\beta_1, \beta_2, ..., \beta_p$ are the regression coefficients.

This model is known as a semiparametric model. The nonparametric part is $\lambda_0(t)$ since it does not require any assumption about the shape of the underlying hazard function. The parametric part of the model reflects the effect of the predictors, $exp(\beta' z)$, which is called the risk function. Cox's model is also called the proportional hazards model since it assumes a constant ratio of hazards over time for any two individuals or units.

1.2.2 Partial Likelihood

The concept of Partial Likelihood was introduced by Cox (1972) for analysis of multiplicative hazard models. It was subsequently modified by many authors, such as Wong (1986) and Anderson & Gill(1982). Why partial likelihood is used instead of the full likelihood? First, we are interested in making inference about the regression parameters but not the form of the baseline hazard; second, the partial likelihood avoids misspecification of the baseline and hence assuages the modeling bias; third, under certain conditions the partial likelihood estimator is semiparametrically efficient. We will first give brief description of Partial Likelihood. Consider a sample of N individuals who are followed up in time prospectively. Suppose that k of these individuals die during the observed period. Also assume that, N-k individuals are right censored, that is they are still alive at the end of the observation period.

Let $t_1 \leq t_2 \leq ... \leq t_k$ be the ordered failure times for the k individuals who die during the observation period.

For the individual $j(i = 1, 2, ..N)$, let
t_j = observed follow up time, Z_j = vector of predictors, and $R(t_j)$ = the risk set at time t_j, that is the number of individuals who are alive and at risk at time t_j. The probability that the individual j with covariates z_j dies at time t_j given that individuals in $R(t_j)$ are at risk and only one individual dies at t_j is given by

$$L_j = \frac{exp(\beta' z_{(j)})}{\sum_{i \in R(t_j)} exp(\beta' z_{(i)})}$$

The partial likelihood (PL) is then obtained by taking the product of all these probabilities across all the individuals in the sample who failed. Therefore, the partial likelihood can be interpreted as the ratio of the risk for the individual who fails at a specific time with the risk of all other individuals at the same time. The estimates of the parameters can be obtained by maximizing the partial likelihood. We note that, the censored observations contribute information only in the denominator of the partial likelihood. Since each term in the partial likelihood contributes small information about the parameters β, the goodness of PL does not depend on the sample size but on the censoring rate. If the number of censored observations is large, partial likelihood is less informative. Cox's partial likelihood method is invalid when there are ties in the dataset. In case of tied dataset, that is multiple individuals having the same survival time, we can use Breslow's approximation to partial likelihood.

1.2.3 Time Dependent Covariates

A time-dependent covariate in a Cox model is a predictor whose values may vary with time. Fisher and Lin (1999) extended the cox model to include the time-dependent covariates. In this work $Xi(t)$, $Z_i(t)$ and $W_i(t)$ are time dependent, i.e, at time t, the measurements are $Xi(t)$, $Z_i(t)$ and $W_i(t)$ respectively. For simplicity sometimes X_i, Z_i and W_i are used instead of $Xi(t)$, $Z_i(t)$ and $W_i(t)$.

1.3 Overview

The rest of this dissertation is organized as follows. In Chapter 2 we introduce a new estimation approach to predict the induced relative risk for individuals in the non-validation subsample based on the local linear smoother. In Chapter 3 we establish asymptotic properties of the proposed estimators of the parameters. In Chapter 4 we conduct simulations to compare the performance of different estimating methods. In Chapter 5 we apply the proposed method to analyze a real dataset. In chapter 6 we summarize the dissertation and discuss future research work in this area.

CHAPTER 2: ESTIMATED PARTIAL LIKELIHOOD FOR THE COX MODEL

Motivated by the idea of the partial likelihood approach in Zhou & Pepe (1995) and Zhou Wang (2000) we introduce a new approach to estimate the induced relative risk function for an individual in the non-validation set.

2.1 Notations

To facilitate exposition, we here employ the notations in Zhou and Wang (2000). Suppose that there are n independent individuals in a study cohort. Let $\{X_i(t), Z_i(t)\}$ denote the covariate vectors for the i^{th} subject at time t ($i = 1, \cdots, n$). Assume that $X_i(\cdot)$ is observed only in the validation subsample which is chosen at the baseline under the ignorable missing mechanism condition (Rubin, 1976). Let $Z_i(\cdot)$ be the remaining covariate vector that is always observed and $W(\cdot)$ the informative auxiliary variables for $X(\cdot)$. Let η_i be an indicator variable with $\eta_i = 1$ if the i^{th} individual is in the validation set and 0 if in the non-validation set. Put $V = \{i : \ \eta_i = 1\}$ and $\bar{V} = \{i : \ \eta_i = 0\}$. We assume that individuals in the validation subsample are randomly selected and hence representative. Then observed data for the ith subject is $\{S_i, \delta_i, Z_i(\cdot), W_i(\cdot), X_i(\cdot)\}$, if $\eta_i = 1$ and $\{S_i, \delta_i, Z_i(\cdot), W_i(\cdot)\}$, if $\eta_i = 0$, where S_i is the observed event time for the ith subject which is the minimum of the potential failure time T_i and the censoring time C_i and δ_i is the indicator of failure. Now,we consider the following conditional hazard function of failure time

$$\begin{aligned} \lambda\{t; X_i(t), Z_i(t)\} &\equiv \lim_{\Delta t \downarrow 0} \left[\frac{1}{\Delta t} \Pr\{t \le T_i < t + \Delta t | T_i \ge t, X_i(t), Z_i(t)\} \right] \\ &= \lambda_0(t) \exp\{\beta_1' X_i(t) + \beta_2' Z_i(t)\}, \end{aligned} \tag{2.1}$$

where $\lambda_0(\cdot) \geq 0$ is unspecified which is called the base-line hazard rate and $\beta_0' = (\beta_1', \beta_2')'$ is the relative risk parameter vector to be estimated.

Model (2.1) can be fitted using the partial likelihood estimation based on the validation set V which leads to the complete-case partial likelihood estimator (see Cox, 1972). The resulting estimator is consistent, but it neglects the important information on the auxiliary W. For individuals in V, the relative risk functions are

$$\exp\{\beta_1' X_i(t) + \beta_2' Z_i(t)\}.$$

For subjects in $\bar{V}$, the true variate X is not observed, but the relative risk functions can be imputated by estimators of

$$\exp\{\beta_2' Z_i(t)\} E[\exp\{\beta_1' X_i(t)\} | T_i \geq t, Z_i(t)].$$

Then under the independent censoring assumption (Prentice, 1982), the induced relative risk for an individual i can be written as

$$\begin{aligned} r_i(\beta, t) &= \eta_i \exp\{\beta_1' X_i(t) + \beta_2' Z_i(t)\} \\ &\quad + (1 - \eta_i) \exp\{\beta_2' Z_i(t) E[\exp\{\beta_1' X_i(t)\} | S_i \geq t, Z_i(t)]. \end{aligned} \tag{2.2}$$

Then the partial likelihood function for the β is

$$PL(\beta) = \prod_{i=1}^{n} \left\{ \frac{r_i(\beta, S_i)}{\sum_{j \in \mathcal{R}(S_i)} r_j(\beta, S_i)} \right\}^{\delta_i}. \tag{2.3}$$

In order to estimate the parameters β based on the above partial likelihood, one needs an imputation value for the conditional expectation $E[\exp\{\beta_1' X_i(t)\} | S_i \geq t, Z_i(t)]$. Different imputation approaches generally yield different estimation of β. Zhou and Wang (2000) employed an imputation method for the relative risk functions for subjects in $\bar{V}$, where the relative risk functions are imputated by nonparametric estimators of

$$\exp\{\beta_2' Z_i(t)\} E[\exp\{\beta_1' X_i(t)\} | S_i \geq t, Z_i(t), W_i(t)]. \tag{2.4}$$

under the assumption that W is not informative, that is, all of the effects of W on failure and censoring are mediated through X, so that

$$\begin{aligned}
\lambda\{t; X_i(t), Z_i(t), W_i(t)\} &\equiv \lim_{\Delta t \downarrow 0} \big[\frac{1}{\Delta t} Pr\{t \leq T_i < t + \Delta t | T_i \geq t, X_i(t), Z_i(t), W_i(t)\}\big] \\
&= \lim_{\Delta t \downarrow 0} \big[\frac{1}{\Delta t} Pr\{t \leq T_i < t + \Delta t | T_i \geq t, X_i(t), Z_i(t)\}\big] \\
&= \lambda_0(t) \exp\{\beta_1' X_i(t) + \beta_2' Z_i(t)\} \\
&\equiv \lambda\{t; X_i(t), Z_i(t)\},
\end{aligned}$$

Zhou and Wang (2000) derived the consistency and asymptotic normality of the estimator. However, if W is informative, their method will generally be biased. In addition, this method directly used information in the auxiliary covariate W and estimated the conditional expectation in (2.4). So it may encounter the so-called "curse of dimensionality" if W is of high dimension. For the present study, the information in W will be used in a new way.

2.2 Local Linear Regression

We employ the kernel regression approach for estimating the relative risk function for the subjects with missing covariate measurements. Here, we give a brief description of the local linear regression. Local linear regression is a popular modeling procedure in nonparametric regression. Fan and Gijbels(1996) illustrated the techniques and theoretical properties in their literature. The local linear smoother possesses some advantages over the Nadaraya Watson (1964) method employed in Zhou & Wang (2000).

1. Local linear estimator has less bias while it does not increase the variance.

2. Local linear smoothing is very adaptable and can be applied for for different types of data design.

3. Local linear smoothing has the advantage that it adapts automatically to the boundary effects, and so no boundary modifications are needed.

Consider the bivariate data $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ which form an independent and identically distributed sample from the population (X,Y). We want to estimate the regression function $m(x_0) = E(Y|X = x_0)$ and its derivative $m'(x_0)$. The data is generated from the model

$$Y_i = m(X_i) + \epsilon_i \quad 1 \le i \le n,$$

where, $\{\epsilon_i\}_1^n$ denote zero mean random variables with variance σ^2.
Suppose that the second order derivative at x_0 exists. We then approximate the unknown regression function $m(x)$ locally by a linear equation. Using Taylor's expansion in the neighborhood of x_0 we have,

$$m(x) \approx m'(x_0) + (x - x_0)m'(x_0).$$

The above polynomial is locally fitted by a weighted least squares problem:
Minimize

$$\sum_{i=1}^{n} \{Y_i - \beta_0 - \beta_1(X_i - x_0)\}^2 K_h(X_i - x_0)$$

over β_j, j=0,1, where h is a smoothing parameter controlling the size of the local neighborhood and $K_h(.) = K(./h)/h$. Here K is a symmetric kernel function which assigns weight to each data point. We denote by $\hat{\beta}_j$, $j = 0, 1$, the solution of the above weighted least squares problem. From the Taylor's expansion we can see that $\hat{m}_\nu(x_0) = \nu!\hat{\beta}_\nu$ is an estimator of $m^{(\nu)}(x_0)$ $(\nu = 0, 1)$. The estimator $\hat{m}_0(x)$ is termed as a local linear regression smoother or a local linear fit. This estimator can be explicitly expressed as

$$\hat{m}_0(x) = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i},$$

$$w_i = K_h(X_i - x)S_{n,2} - (X_i - x)S_{n,1},$$

where, $S_{n,j} = \sum_{i=1}^n K_h(X_i - x)(X_i - x)^j$. For convenience we work with the matrix notation below.

Let X be the design matrix of the given least squares problem. Then,

$$X = \begin{pmatrix} 1 & (X_1 - x_0) \\ \vdots & \vdots \\ 1 & (X_n - x_0). \end{pmatrix}$$

Also, let $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n'; \end{pmatrix}$ and $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_p \end{pmatrix}$

Further, let $\mathbf{W}$ be the $n \times n$ diagonal matrix of weights .i.e.

$$W = diag\{K_h(X_i - x_0)\}.$$

The weighted least squares problem can then be written as

$$min_\beta(\mathbf{Y} - \mathbf{X}\beta)^T\mathbf{W}(\mathbf{Y} - \mathbf{X}\beta)$$

with $\beta = (\beta_0, \beta_1)^T$. The solution vector can be obtained by

$$\hat{\beta} = (X^TWX)^{-1}X^TWY.$$

2.3 Estimation Method

Throughout this dissertation, we assume that model (2.1) holds. In this section, we propose a new estimated partial likelihood approach to estimate the model parameters in (2.1).

2.3.1 Estimation of the Relative Risk Function

Denote $\gamma_i(\beta, t) = \exp\{\beta_1' X_i(t) + \beta_2' Z_i(t)\}$, and

$$\phi_i(\beta, t) = \exp\{\beta_2' Z_i(t) E[\exp\{\beta_1' X_i(t)\} | S_i \geq t, Z_i(t)].$$

Then,

$$r_i(\beta, t) = \eta_i \gamma_i(\beta, t) + (1 - \eta_i)\phi_i(\beta, t).$$

Put $\zeta_i(\beta_1, t) = \exp(\beta_1' X_i(t))$ and $\nu_j(\beta_1, t) = E[\zeta_j(\beta_1, t) | S_j \geq t, Z_j(t)]$. Since the validation subsample is representative, we can estimate based on the local linear regression which leads to the following estimators of $\nu_j(\beta_1, t)$ for $j \in \bar{V}$:

$$\hat{\nu}_j(\beta_1, t) = \sum_{i \in V} \omega_i(t, Z_j(t); h)\zeta_i(\beta_1, t), \tag{2.5}$$

where h is the bandwidth,

$$\omega_i(t, Z_j(t); h) = \frac{\{s_2 - (Z_i(t) - Z_j(t))s_1\} I_{[S_i \geq t]} K_h(Z_i(t) - Z_j(t))}{\sum_{i \in V}\{s_2 - (Z_i(t) - Z_j(t))s_1\} I_{[S_i \geq t]} K_h(Z_i(t) - Z_j(t))}$$

and $s_k = \sum_{i \in V}(Z_i(t) - Z_j(t))^k (I_{[S_i \geq t]} K_h(Z_i(t) - Z_j(t))$

with $K_h(\cdot) = h^{-d} K(\cdot / h)$ for a d-variate kernel function $K(\cdot)$ (d is the dimension of Z).

Here, $\omega_i(t, Z_j(t); h)$ is known as the effective kernel (Fan & yao 2005). In Zhou and Wang (2000), the Nadaraya Watson (1964) estimator was used for the nonparametric smoothing in the estimation of $E[\gamma_i(\beta, t) | S_i \geq t, Z_i(t), W_i(t)]$, where "curse of dimensionality" can happen if W is of high dimension. The estimator is given by,

$$\hat{\nu}_j(\beta_1, t) = \sum_{i \in V} \tilde{\omega}_i(t, Z_j(t); h) \zeta_i(\beta_1, t), \tag{2.6}$$

where h is the bandwidth and

$$\tilde{\omega}_i(t, Z_j(t); h) = I_{[S_i \geq t]} K_h(Z_i(t) - Z_j(t)) / \sum_{i \in V} I_{[S_i \geq t]} K_h(Z_i(t) - Z_j(t))$$

with $K_h(\cdot) = h^{-d} K(\cdot/h)$ for a d-variate kernel function $K(\cdot)$ (d is the dimension of Z).

Note that the above estimation method uses only the complete observations in V and neglects the important information on incomplete observations in $\bar{V}$. It follows that this approach can not be expected to be efficient in certain situations. Also note that even for one dimensional Z and W, the method in Zhou and Wang (2000) requires a two-dimensional smoother while the new method needs only one-dimensional smoother. To have a performance comparable with that of one-dimensional nonparametric smoother using $M_1 = 50$ data points, for a 2-dimensional nonparameteric smoother, we need about $M = M_1^{1.2} = 109$ data points. Hence the loss of efficiency due to highly dimensional smoothing is large and increasing exponentially fast (see page 317 of Fan and Yao, 2003).

2.3.2 Improved Estimation of the Relative Risk Function

Recall that, W is an auxiliary variable for X and is hence correlated with X. Let $\xi_i(\alpha, t) = \exp(\alpha' W_i(t))$, where α is a parameter vector to be chosen. Considering the conditional expectation of $\psi_i(\alpha, t) = E[\xi_i(\alpha, t) | S_i \geq t, Z_i(t)]$, $\psi_i(\alpha, t)$ can also be estimated by local linear smoothing based on the data in V:

$$\hat{\psi}_j(\beta_1, t) = \sum_{i \in V} \omega_i(t, Z_j(t); h) \xi_i(\beta_1, t), \tag{2.7}$$

In the above, the weight function , $\omega_i(t, Z_j(t); h)$, kernel $K_h(.)$, bandwidth h and s_k have the same interpretation as in (2.5).

Proposition 2.1 *Suppose that the conditions in section (3.5) holds and n_v denotes the number of observations in the validation set. Given $(S_j \geq t, Z_j(t))$, $\sqrt{n_v h^d}[(\hat{\nu}_j(\beta_1, t) - \nu_j(\beta_1, t)), (\hat{\psi}_j(\alpha, t) - \psi_j(\alpha, t))]$ is jointly asymptotically normal with mean zero and covariance matrix*

$$\Sigma = v_0(K)p^{-1}(Z_j)\begin{bmatrix} \sigma_1^2(Z_j, t) & \rho^*_\alpha(Z_j, t)\sigma_1(Z_j, t)\sigma_2(Z_j, t) \\ \rho^*_\alpha(Z_j, t)\sigma_1(Z_j, t)\sigma_2(Z_j, t) & \sigma_2^2(Z_j, t) \end{bmatrix},$$

where $v_0(K) = \int K^2(u)du$, $\sigma_1^2(Z_j, t) = Var[\zeta_j | S_j \geq t, Z_j]$, $\sigma_2^2(Z_j, t) = Var[\xi_j | S_j \geq t, Z_j]$, $\rho^*_\alpha(Z_j, t)$ is the conditional correlation coefficient between ζ_j and ξ_j given $(S_j \geq t, Z_j)$, and $p(\cdot)$ is the density function of Z.

By the distribution theory for multivariate normal variates, the conditional distribution of $\sqrt{n_v h^d}[\hat{\nu}_j(\beta_1, t) - \nu_j(\beta_1, t)]$ given $\sqrt{n_v h^d}[\hat{\psi}_j(\alpha, t) - \psi_j(\alpha, t)]$ is asymptotically normal with mean

$$\rho^*_\alpha(Z_j, t)\frac{\sigma_1(Z_j, t)}{\sigma_2(Z_j, t)}\sqrt{nh^d}[\hat{\psi}_j(\alpha, t) - \psi_j(\alpha, t)].$$

The conditional mean can then be estimated by substituting consistent estimators based on the validation sample for $\rho^*_\alpha(Z_j, t)$, $\sigma_1(Z_j, t)$ and $\sigma_2(Z_j, t)$, and replacing $\psi_j(\alpha, t)$ with the primary sample based estimator

$$\bar{\psi}_j(\beta_1, t) = \sum_{i \in V \cup \bar{V}} \bar{\omega}_i(t, Z_j(t); h)\xi_i(\beta_1, t), \tag{2.8}$$

where h is the bandwidth and

$$\bar{\omega}_i(t, Z_j(t); h) = \frac{\{\bar{s_2} - (Z_i(t) - Z_j(t))\bar{s_1}\}I_{[S_i \geq t]}K_h(Z_i(t) - Z_j(t))}{\sum_{i \in V \cup \bar{V}}\{\bar{s_2} - (Z_i(t) - Z_j(t))\bar{s_1}\}I_{[S_i \geq t]}K_h(Z_i(t) - Z_j(t))}$$

and $\bar{s_k} = \sum_{i \in V \cup \bar{V}}(Z_i(t) - Z_j(t))^k(I_{[S_i \geq t]}K_h(Z_i(t) - Z_j(t))$

with $K_h(\cdot) = h^{-d}K(\cdot/h)$ for a d-variate kernel function $K(\cdot)$ (d is the dimension of Z).

Here, $\bar{\omega}_i(t, Z_j(t); h)$ is also known as the effective kernel. By equating $\sqrt{nh^d}[\hat{\nu}_j(\beta_1, t) - \nu_j(\beta_1, t)]$ with its estimated conditional mean and solving for $\nu_j(\beta_1, t)$, we obtain an improved estimate $\bar{\nu}_j(\beta_1, t)$:

$$\bar{\nu}_j(\beta_1, t) = \hat{\nu}_j(\beta_1, t) - \hat{\rho}^*_\alpha(Z_j, t)\frac{\hat{\sigma}_1(Z_j, t)}{\hat{\sigma}_2(Z_j, t)}[\hat{\psi}_j(\alpha, t) - \bar{\psi}_j(\alpha, t)]. \tag{2.9}$$

The updated estimator $\bar{\nu}_j$ depends on α which is related to the efficiency of the estimator.

Proposition 2.2 *Assume that the conditions in section (3.5) holds. Given* $(S_j \geq t, Z_j(t))$,

$$\sqrt{n_v h^d}[\bar{\nu}_j(\beta_1, t) - \nu_j(\beta_1, t)] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega),$$

where $\Omega(Z_j, t) = \sigma_1^2(Z_j, t)[1 - (1 - \rho)\rho^{*2}_\alpha(Z_j, t)]v_0(K)p^{-1}(Z_j)$.

When $\rho^*_\alpha = 0$, the estimator $\bar{\nu}_j$ is asymptotically equivalent to $\hat{\nu}_j$, which corresponds to the kernel regression estimator based on only the validation set V.

By Propositions 2.1 and 2.2, $\bar{\nu}_j$ is more efficient than $\hat{\nu}_j$. The proposed estimator is consistent for any α. However, its limiting covariance matrix depends on the choices of α. We chose the optimum value α_{opt} by minimizing the trace of the covariance matrix of the EPL estimator with $\hat{\beta}_{EPL}$ substituted by the initial estimator obtained from complete-cox regression which uses the data available only on the validation set. In particular, $\hat{\beta}_{EPL}(\alpha_{opt})$ is guaranteed to be more efficient than the complete-case estimator $\hat{\beta}_{EPL}(0)$. In this study α_{opt} is estimated by minimizing the trace of the covariance matrix of $\hat{\beta}_{EPL}$.

The proposed estimation method was similarly used in Chen and Chen (2000) for estimating parameters in a parametric regression model. Our estimation can be regarded as an extension of their estimation approach in nonparametric regression. In addition, we do not need a working model to specify the regression relation between the surrogate and the covariate, and hence there is no risk of mispecification of the working model.

We propose to estimate the reduced relative risk $r_i(\beta, t)$ by

$$\hat{r}_i(\beta, t) = \eta_i \gamma_i(\beta, t) + (1 - \eta_i)\bar{\phi}_i(\beta, t), \tag{2.10}$$

where $\bar{\phi}_i(\beta, t) = \bar{\nu}_i(\beta_1, t) \exp\{\beta_2' Z_i(t)\}$. Then the parameters β can be estimated by maximizing the following estimated partial likelihood function:

$$EPL(\beta) = \prod_{i=1}^{n} \left\{ \frac{\hat{r}_i(\beta, S_i)}{\sum_{j \in \mathcal{R}(S_i)} \hat{r}_j(\beta, S_i)} \right\}^{\delta_i}, \tag{2.11}$$

where $\mathcal{R}(S_i)$ is the risk set at time S_i. We denote $\hat{\beta}_{EPL} = \arg\max_\beta EPL(\beta)$.

For an extreme case with $W = Z$, the $\hat{\psi}_j$ equals $\bar{\psi}_j$, which leads to $\bar{\nu}_j = \hat{\nu}_j$ and that the resulting estimator $\hat{\beta}_{EPL}$ is the same as that in Zhou and Wang (2000). In above estimation of the reduced relative risk, we used an improved estimator $\phi_j(\beta, t)$ for $j \in \bar{V}$. The "curse of dimensionality " problem in Zhou & Wang (2000) can be avoided for a highly dimensional W. Our approach would be useful in cases where the number of variables in Z which are correlated with the missing covariate X is low, whereas the exposure variables of interest and their auxiliary variables may be of high dimension.

CHAPTER 3: ASYMPTOTIC RESULTS

3.1 Counting Process Formulation for the Cox Model

In this section we will develop the counting process formulation for Cox's type of model. We are going to use the framework developed in Anderson and Gill(1982) and the basic theory from Fleming and Harrington (1991). For simplicity, we assume the time interval to be finite. We take the time interval as [0,1] without loss of generality. To prove the asymptotic properties, we consider a sequence of models. A multivariate counting process with n components is a non-decreasing integer valued stochastic process which can be expressed as

$N^{(n)} = \{N_i^{(n)}(t) : 0 \leq t < \infty \, ; i = 1, 2, ..., n\}.$

Here, $N_i^{(n)}$ is the number of observed events in the life of the i^{th} subject ($i = 1, 2, ..., n$) in the n^{th} model (n=1,2,...) over the time interval [0,1]. For simplicity we shall drop the subscript n in the following sections.

It is assumed that $N_i(0) = 0$ for all i and the jump size is +1. This process may count the number of events in the n^{th} individual that happened upto time t. If it is the death of the individual then $N_i(t) \in \{0, 1\}$. $N_i(t)$ is right continuous and no two components of N jump at the same time. So there will be atmost one jump for each subject in the study. In our model we consider the nondecreasing family $\{\mathcal{F}_t : t \in [0, 1]\}$ of sub σ-algebra on the probability space $\{\Omega, \mathcal{F}, \mathcal{P}\}$. $\mathcal{F}_t$ is known as the filtration which is history of everything that happens upto time t. We shall use the results for counting processses and local martingales with respect to the filtration given above. Counting process is associated with a cumulative intensity process Λ

whose components are given by

$$
\begin{aligned}
\Lambda_i(t) &= \Lambda_i(t+dt) - \Lambda_i(t) \\
&= P(N_i(t+dt) - N_i(t) = 1|\mathcal{F}_{t-}),
\end{aligned}
$$

where $\mathcal{F}_{t-}$ represents everything that has happened upto just before t. This history includes paths of $N_i(.)$ and also other information about the predictor variables and censoring etc. A martingale with respect to a filtration $\mathcal{F}_t$ is a right-continuous stochastic process $M(t)$ with left-hand limits such that, in addition to some technical conditions:

(1) M(t) is adapted to history,

(2) $E|M(t)| < \infty$ for all t, and

(3) M(t) possesses the key martingale property $E(M(t)|\mathcal{F}_s) = M(s)$ for all $s \leq t$.

Following Anderson and Gill (1982), our model can be generalized as

$$
\begin{aligned}
\Lambda_i(t+dt) - \Lambda_i(t) &= \lambda_i(t)dt \\
&= Yi(t)\lambda_0(t)r_i^*(t)\}dt, \qquad (3.1)
\end{aligned}
$$

where

$$
\begin{aligned}
r_i^*(\beta, t) &= \eta_i \exp\{\beta_1' X_i(t) + \beta_2' Z_i(t)\} \\
&\quad + (1-\eta_i)\exp\{\beta_2' Z_i(t) E[\exp\{\beta_1' X_i(t)\}|S_i \geq t, Z_i(t), W_i(t)], \qquad (3.2)
\end{aligned}
$$

$Y_i(t) = 1$, if the i^{th} individual is under observation just before time t and 0 otherwise. $Y_i(.)$ is known as the"at-risk" indicator process and $\lambda_0(t)$ is the baseline hazard function. We assume that the covariate processes $X(t)$ and $Z(t)$ are predictable and locally bounded. Since $X(t)$ and $Z(t)$ are taken to be adapted and left continuous with right hand limits, these assumptions hold true as illustrated in Fleming and Harrington(1991). Therefore by considering the counting process N and associated intensity process λ, we can define the process $M_i(t)$ by

$$M_i(t) \quad = \quad N_i(t) - \int_0^t \lambda_i(u)du, i = 1, 2, ..., n, \;\; t \in [0, 1]. \tag{3.3}$$

Then $M_i(t)$ are local martingales on the time interval [0,1]. Then local martingales are local square integrable martingales since the intensity process $\lambda(.)$ is locally bounded. Following the theory and discussions in Fleming and Harrington (1991) , the predictable variation process of M(t) is given by

$$< M_i, M_i > \quad = \quad \int_0^t \lambda_i(u)du \tag{3.4}$$

and $< M_i, M_j >= 0$ when $i \neq j$.

The last equation implies M_i and M_j are orthogonal for $i \neq j$.

To prove the asymptotic properties of our estimator we use the the following theorem on local martingales.

Theorem 3.1 *If H_i is a locally bounded and $\mathcal{F}_{t-}-$predictable process, then $\sum_{i=1}^{n} \int H_i dM_i$ is a local square integrable Martingale, and the predictable covariance process is given by*

$$\sum_{i=1}^{n} < \int H_i dM_i, \int H_i dM_i >= \sum_{i=1}^{n} < \int H_i^2 d < M_i, M_i > .$$

For the proof of the above , see Theorem 2.4.3 (Page 70) in Fleming and Harrington (1991).

Using the new notation we write down the logarithm of the partial likelihood function using the information upto time t as

$$L(\beta, t) \quad = \quad \sum_{i=1}^{n} \int_0^t log\{r_i(u)\}dN_i(u) - \int_0^t log\{\sum_{i=1}^{n} Y_i(u) r_i(u)\} d\bar{N}(u), \tag{3.5}$$

where $\bar{N} = \sum_{i=1}^{n} N_i$. For the individuals in the non-validation set the induced relative risk function $r(\beta, t)$ is unknown. In section (2.2) we proposed an imputation method for the this function based on the kernel smoothing approach and then estimated the parameter vector $\beta = (\beta_1', \beta_2')'$ from the partial likelihood function given in (2.11). Therefore to obtain the proposed estimator $\hat{\beta}_{EPL}$ we need to find the solution of the estimating equation

$$\frac{\partial}{\partial \beta} L(\beta, t) = 0.$$

To obtain the above, we substitute $r(\beta, t)$ by $\hat{r}(\beta, t)$ given in (2.10). Then the vector of derivatives of the logarithm of partial likelihood function with respect to β can be expressed as

$$\begin{aligned} \hat{U}(\beta, t) &= \sum_{i=1}^{n} \int_0^t \frac{\hat{r}_i^{(1)}(u)}{\hat{r}_i(u)} dN_i(u) - \int_0^t \frac{\sum_{i=1}^{n} Y_i(u) \hat{r}_i^{(1)}(u)}{\sum_{i=1}^{n} Y_i(u) \hat{r}_i(u)} \bar{N}(u) \\ &= \sum_{i=1}^{n} \int_0^t \Delta(\hat{r}_i(u)) dN_i(u), \end{aligned} \tag{3.6}$$

where $\hat{r}_i^{(1)}(u) = \frac{\partial}{\partial \beta} \hat{r}_i(u)$ and

$$\Delta(\hat{r}_i(u)) = \frac{\hat{r}_i^{(1)}(u)}{\hat{r}_i(u)} - \frac{\sum_{i=1}^{n} Y_i(u) \hat{r}_i^{(1)}(u)}{\sum_{i=1}^{n} Y_i(u) \hat{r}_i(u)}.$$

Using the Doob Meyer decomposition, from (3.1) and (3.3) we rewrite the estimating equation as

$$\hat{U}(\beta, t) = \sum_{i=1}^{n} \int_0^t \Delta(\hat{r}_i(u)) dM_i(u) + \sum_{i=1}^{n} \int_0^t \Delta(\hat{r}_i(u)) \hat{r}_i^*(u) \lambda_0(u) du. \tag{3.7}$$

Also, with the estimator of β_0, $\hat{\beta}_{EPL}$, from the estimating score equation given above, the cumulative hazard $\Lambda_0(t) = \int_0^t \lambda_0(w)\, dw$ can be consistently estimated as

$$\hat{\Lambda}_0(t) \quad = \quad \int_0^t [\sum_{i=1}^n Y_i(u) r_i^*(\hat{\beta}_{EPL}, u)]^{-1} \sum_{i=1}^n dN_i(u). \tag{3.8}$$

3.2 Notations

In this section we will define some notations which will be used in the proofs. All the limits are taken as $n \to \infty$ unless otherwise stated. This implies numbers of subjects in the validation set and non-validation set, both, $n_v \to \infty$ and $(n - n_v) \to \infty$. Let d be the dimension of Z_i, n_v be the subsample size of the validation set, $\rho \in (0, 1]$ be the limit of ratio of validation observations, $\lim_{n\to\infty} n_v/n$. For a vector a, define $|a| = \sqrt{a'a} = \sqrt{a_i^2}$. Also, we write the matrix $aa' = a^{\otimes 2}$ and $(aa')(aa')' = a^{\otimes 4}$. For the relative risk function r (for $\hat{r}$, r^*, $\hat{r}^*$, ϕ and $\hat{\phi}$ as well) , let r^j denote the j^{th} derivative of r with respect to β, j=0,1,2, where $r^{(0)} = r$. Define

$$s^{(0)}(\beta, t) = E[Y_i(t) r_i(\beta, t)],$$

$$s^{(1)}(\beta, t) = (\partial/\partial\beta) s^{(0)}(\beta, t) = E[Y_i(t) r_i^{(1)}(\beta, t)],$$

$$s^{(2)}(\beta, t) = (\partial/\partial\beta^{\tau}) s^{(1)}(\beta, t) = E[Y_i(t) r_i^{(2)}(\beta, t)],$$

$$s^{(3)}(\beta, t) = E\big[Y(t)\big(\frac{r_i^{(1)}(\beta, t)}{r_i(\beta, t)}\big) r_i^*(\beta_0, t)\big],$$

$$s^{(4)}(\beta, t) = E\big[Y(t)\big(\frac{r_i^{(2)}(\beta, t)}{r_i(\beta, t)}\big) r_i^*(\beta_0, t)\big],$$

$$s^{(5)}(\beta, t) = E\big[Y(t)\big(\frac{r_i^{(1)}(\beta, t)}{r_i(\beta, t)}\big)^{\otimes 2} r_i^*(\beta_0, t)\big],$$

$$s^{(6)}(\beta, t) = E\big[Y(t)\big(\frac{r_i^{(2)}(\beta, t)}{r_i(\beta, t)}\big)^{\otimes 2} r_i^*(\beta_0, t)\big],$$

$$s^{(7)}(\beta, t) = E\big[Y(t)\big(\frac{r_i^{(1)}(\beta, t)}{r_i(\beta, t)}\big)^{\otimes 4} r_i^*(\beta_0, t)\big],$$

where $Y_i(t) = I_{[S_i \geq t]}$ is the at-risk indicator, $r_i^{(1)}(\beta, t) = (\partial/\partial\beta) r_i(\beta, t)$ and $r_i^{(2)}(\beta, t) = (\partial/\partial\beta) r_i^{(1)}(\beta, t)$.

Observe that,

$$s^{(0)}(\beta, t) = E[Y_i(t) r_i(\beta, t)] = E[Y_i(t) r_i^*(\beta, t)]$$

Next we define

$$S^{(0)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) r_i(\beta, t),$$

$$S^{(1)}(\beta, t) = (\partial/\partial\beta) S^{(0)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) r_i^{(1)}(\beta, t),$$

$$S^{(2)}(\beta, t) = (\partial/\partial\beta^{\tau}) S^{(1)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) r_i^{(2)}(\beta, t),$$

$$S^{(3)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Y(t) \big(\frac{r_i^{(1)}(\beta, t)}{r_i(\beta, t)} \big) r_i^*(\beta_0, t),$$

$$S^{(4)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Y(t) \big(\frac{r_i^{(2)}(\beta, t)}{r_i(\beta, t)} \big) r_i^*(\beta_0, t),$$

$$S^{(5)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Y(t) \big(\frac{r_i^{(1)}(\beta, t)}{r_i(\beta, t)} \big)^{\otimes 2} r_i^*(\beta_0, t),$$

$$S^{(6)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Y(t) \big(\frac{r_i^{(2)}(\beta, t)}{r_i(\beta, t)} \big)^{\otimes 2} r_i^*(\beta_0, t),$$

$$S^{(7)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Y(t) \big(\frac{r_i^{(1)}(\beta, t)}{r_i(\beta, t)} \big)^{\otimes 4} r^*(\beta_0, t),$$

For $k = 1, 2, ..., 7$, we similarly define $\hat{S}^{(k)}(\beta, t)$ with $r(\beta, t)$ replaced by $\hat{r}(\beta, t)$ and $r^*(\beta, t)$ by $\hat{r}^*(\beta, t)$, respectively.

Now, we define,

$$\phi_i^*(\beta, t) \quad = \quad \exp\{\beta_2' Z_i(t) E[\exp\{\beta_1' X_i(t)\} | S_i \geq t, Z_i(t), W_i(t)], \qquad (3.9)$$

and

$$r_i^*(\beta, t) \;\; = \;\; \eta_i \gamma_i(\beta, t) + (1 - \eta_i)\phi_i^*(\beta, t). \tag{3.10}$$

Let $N_i(t) = I_{[S_i < t, \delta_i = 1]}$ and

$$M_i(t) \;\; = \;\; N_i(t) - \int_0^t Y_i(u) r_i^*(\beta_0, u)\lambda_0(u)du, \tag{3.11}$$

which is a martingale (Kalbfleisch and Prentice (1980), Fleming and Harrington (1991)) as discussed in section 3.1.

Next, without loss of generality, we assume that $t \in [0, 1]$. Put

$$\Delta(\phi_i)(u) = \phi_i^{(1)}(u)/\phi_i(u) - s^{(1)}/s^{(0)},$$

$$\Delta(\gamma_i)(u) = \gamma_i^{(1)}(u)/\gamma_i(u) - s^{(1)}/s^{(0)},$$

$$Q_i = \int_0^1 \Delta(\phi_i)(u) Y_i(u)[\gamma_i(\beta_0, u) - \phi_i(\beta_0, u)]\lambda_0(u)du,$$

$$Q_i^* = \int_0^1 \Delta(\phi_i)(u) Y_i(u)\theta_i(u; \alpha)\lambda_0(u)du,$$

$$Q_i^{**} = \int_0^1 \Delta(\phi_i)(u) Y_i(u)[\phi_i^*(\beta_0, u) - \phi_i(\beta_0, u)]\lambda_0(u)du,$$

where $\phi_i^{(1)}(\beta, u) = (\partial/\partial\beta)\phi_i(\beta, u)$, and

$\theta_i(u; \alpha) = [\xi_i(\alpha, u) - \psi_i(\alpha, u)] \exp(\beta_2' Z_i(u))\rho_\alpha^*(Z_i, u)\sigma_1(Z_i, u)/\sigma_2(Z_i, u).$

By using counting process notations, the score function corresponding to the estimated partial likelihood function (2.11) at time point t can be written as

$$\hat{U}(\beta, t) = \sum_{i=1}^n \int_0^t \Delta(\hat{r}_i)(\beta, u) dM_i(u) + \sum_{i=1}^n \int_0^t \Delta(\hat{r}_i)(\beta, u) r_i^*(\beta_0, u) Y_i(u)\lambda_0(u)du, \tag{3.12}$$

where

$$\Delta(\hat{r}_i)(u) = \frac{\hat{r}_i^{(1)}(\beta,u)}{\hat{r}_i(\beta,u)} - \frac{\sum_{i=1}^n Y_i(u)\hat{r}_i^{1)}(\beta,u)}{\sum_{i=1}^n Y_i(u)\hat{r}_i(\beta,u)}.$$

Next we define $I(\beta)$, $\Sigma_1(\beta)$ and $\Sigma_2(\beta)$ and $\Sigma(\beta)$ respectively, which will be required in the proof of asymptotic normality of our estimator.

Let

$$I(\beta) = -E\left[\int_0^1 \left(\frac{r_i^{(2)}(\beta,u)}{r_i^{(0)}(\beta,u)} - \left\{\frac{r_i^{(1)}(\beta,u)}{r_i^{(0)}(\beta,u)}\right\}^{\otimes 2} - \frac{s^{(2)}(\beta,u)}{s^{(0)}(\beta,u)} + \left\{\frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)}\right\}^{\otimes 2}\right) dN_i(t)\right],$$

$$\Sigma_1(\beta) = E\left[\int_0^1 \Delta(\phi_i)(u)dM_i(u) - (1-\rho)Q_i^* + Q_i^{**}\right]^{\otimes 2},$$

$$\Sigma_2(\beta) = E\left[\int_0^1 \Delta(\gamma_i)(u)dM_i(u) - \frac{1-\rho}{\rho}\left\{Q_i - (1-\rho)Q_i^*\right\}\right]^{\otimes 2}, and$$

$$\Sigma(\beta) = \rho\Sigma_1(\beta) + (1-\rho)\Sigma_2(\beta).$$

From Theorem 3.6 proved in section (3.7) , the asymptotic covariance matrix of $\hat{\beta}_{EPL}$ is of sandwich form, which can be consistently be estimated by $\hat{\Omega} = \hat{I}^{-1}(\beta)\hat{\Sigma}(\beta)\hat{I}^{-1}(\beta)$, where $\hat{I}(\beta)$ and $\hat{\Sigma}(\beta)$ are the corresponding sample quantities, respectively. Specifically,

$$\begin{aligned}&\hat{I}(\beta)\\ = \;& -n^{-1}\sum_{i=1}^n \int_0^1 \left(\frac{\hat{r_i}^{(2)}(\beta,u)}{\hat{r_i}^{(0)}(\beta,u)} - \left\{\frac{\hat{r_i}^{(1)}(\beta,u)}{r_i^{(0)}(\beta,u)}\right\}^{\otimes 2} - \frac{\hat{S}^{(2)}(\beta,u)}{\hat{S}^{(0)}(\beta,u)} + \left\{\frac{\hat{S}^{(1)}(\beta,u)}{\hat{S}^{(0)}(\beta,u)}\right\}^{\otimes 2}\right) dN_i(t),\end{aligned}$$

$$\hat{\Sigma}_1(\beta) = n^{-1}\sum_{i=1}^n \left\{\int_0^1 \Delta(\hat{\phi}_i)(t)[dN_i(t) - Y_i(t)\hat{r}_i(\beta,t)\,d\hat{\Lambda}_0(t)] - (1-\hat{\rho})\hat{Q}_i^* + \hat{Q}_i^{**}\right\}^{\otimes 2},$$

$$\hat{\Sigma}_2(\beta) = n_v^{-1} \sum_{i=1}^{n_v} \Big\{ \int_0^1 \Delta(\hat{\gamma}_i)(t)[dN_i(t) - Y_i(t)\hat{r}_i(\beta,t)\, d\hat{\Lambda}_0(t)] - \frac{1-\hat{\rho}}{\hat{\rho}}[\hat{Q}_i - (1-\hat{\rho})\hat{Q}_i^*] \Big\}^{\otimes 2},$$

where

$$\hat{Q}_i = \int_0^1 \Delta(\hat{\phi}_i)(t) Y_i(t)[\hat{r}_i(\beta,t) - \hat{\phi}_i(\beta,t)]\, d\hat{\Lambda}_0(t),$$

$$\hat{Q}_i^* = \int_0^1 \Delta(\hat{\phi}_i)(t) Y_i(t)\hat{\theta}_i(t;\alpha)\, d\hat{\Lambda}_0(t),$$

$\hat{Q}_i^{**} = \int_0^1 \Delta(\hat{\phi}_i)(t) Y_i(t)[\hat{\phi}_i^{\,*}(\beta,t) - \hat{\phi}_i(\beta,t)]\, d\hat{\Lambda}_0(t)$, $\hat{\rho} = n_v/n$,

$$\Delta(\hat{\phi}_i)(t) = \hat{\phi}_i^{(1)}(\beta,t)/\hat{\phi}_i(\beta,t) - \hat{S}^{(1)}(\beta,t)/\hat{S}^{(0)}(\beta,t),$$

$$\Delta(\hat{\gamma}_i)(t) = \hat{\gamma}_i^{(1)}(\beta,t)/\hat{\gamma}_i(\beta,t) - \hat{S}^{(1)}(\beta,t)/\hat{S}^{(0)}(\beta,t),$$

and

$$\hat{\theta}_i(t;\alpha) = \big[\xi_i(\alpha,t) - \bar{\psi}_i(\alpha,t)\big] \exp(\beta_2^\tau Z_i(t))\hat{\rho}_\alpha^*(Z_i,t)\hat{\sigma}_1(Z_i,u)/\hat{\sigma}_2(Z_i,t).$$

3.3 Consistency of $\hat{\beta}_{EPL}$

To show the consistency of the estimator $\hat{\beta}_{EPL}$ we use the inverse function theorem from Walter and Rudin(1964) and Foutz's (1977) argument.

Inverse Function Theorem: Suppose f is a mapping from an open set Θ in Euclidean p space $\mathcal{R}_p$ into $\mathcal{R}_p$, the partial derivatives of f exist and are continuous on Θ, and the matrix derivatives $f'(\theta^*)$ has inverse $f'(\theta^*)^{-1}$ for some $\theta^* \in \Theta$. Write

$$\lambda = 1/4 \parallel f'(\theta^*)^{-1} \parallel).$$

Use the continuity of elements of $f'(\theta^*)$ to fix a neighborhood $\mathbf{U}_\delta$ of θ^* of sufficiently small radius $\delta > 0$ to insure $\parallel f'(\theta) - f'(\theta^*) \parallel) < 2\lambda$, whenever $\theta \in \mathbf{U}_\delta$. Then

(a) for every θ_1, θ_2 in $\mathbf{U}_\delta$,

$$|f(\theta_1) - f(\theta_2)| \geq 2\lambda|\theta_1 - \theta_2|,$$

and (b) the image set $f(\mathbf{U}_\delta)$ contains the open neighborhood with radius $\lambda\delta$ about $f(\theta^*)$.

(a) insures that f is one-to-one on $\mathbf{U}_\delta$ and that f^{-1} is well defined on the image set $f(\mathbf{U}_\delta)$. The proof of the theorem is given in p 194 (Walter & Rudin, 1964). Consider the inverse function $\frac{1}{n}\hat{U}^{-1}$ which is a mapping from p-dimensional Euclidean space to an open subset of $\mathcal{B}$. $\hat{\beta}_{EPL}$ is the value at 0 of this function. In the later section we show that, this inverse function is well defined in an open neighborhood about 0 with probability tending to 1. Then we can prove that $\hat{\beta}_{EPL} = \frac{1}{n}\hat{U}^{-1}(0)$ is a consistent estimate of β_0.

3.4 Asymptotic Normality of $\hat{\beta}_{EPL}$

To prove the asymptotic normality of the estimator $\hat{\beta}_{EPL}$ we use martingale approach under multivariate counting process framework. The main techniques we employed are Taylor's expansion of the score function corresponding to the estimated likelihood function (2.11), Lenglart inequality, the martingale central limit theorem (see e.g. Fleming and Harrington, 1991), and nonparametric regression techniques.

We use the first order Taylor's expansion of the score function $\hat{U}(\beta, 1)$ around β_0, which gives

$$\hat{U}(\beta, 1) - \hat{U}(\beta_0, 1) = \frac{\partial}{\partial \beta^*}\hat{U}(\beta^*, 1)(\hat{\beta} - \beta_0),$$

where, β^* is between $\hat{\beta}$ and β_0. Since $\hat{\beta}_{EPL}$ is the solution of the score equation $\hat{U}(\beta, 1) = 0$, we can rewrite the above equation as

$$n^{-1/2}\hat{U}(\beta_0, 1) = \{-n^{-1}\frac{\partial}{\partial \beta^*}\hat{U}(\beta^*, 1)\}n^{1/2}(\hat{\beta}_{EPL} - \beta_0).$$

We will show that

$$-n^{-1}\frac{\partial \hat{U}(\beta_*,1)}{\partial \beta} \xrightarrow{P} I(\beta_0).$$

Then the asymptotic normality of $n^{1/2}(\hat{\beta}_{EPL}-\beta_0)$ follows by showing that $n^{-1/2}\hat{U}(\beta_0,1)$ is asymptotically normal with mean 0 and variance $(1-\rho)\Sigma_1(\beta_0)+\rho\Sigma_2(\beta_0)$, where ρ, $I(\beta_0)$, $\Sigma_1(\beta_0)$ and $\Sigma_2(\beta_0)$ are defined in the section (3.2).

3.5 Definitions and Conditions

The following conditions are needed throughout the remaining part of the dissertation:

(1) $\int_0^1 \lambda_0(s)ds < \infty$.

(2) $Pr(Y(1)=1|V) > 0$ for any V.

(3)There exists an open subset $\mathcal{B}$, containing the true β, β_0, of the Euclidean space $\mathcal{R}_p$. In addition, $r_i^{(2)}(\beta,t)$ with elements $(\partial^2/\partial\beta_i\partial\beta_j)r(\beta,t)$ exists and is continuous on $\mathcal{B}$ for each $t\in[0,1]$, uniform in t, and $\phi(\beta,t)$ is bounded away from 0 on $\mathcal{B}\times[0,1]$. Furthermore, $I(\beta_0)$ defined in section (3.2) is positive definite.

(4)

$$E\{\sup_{\mathcal{B}\times[0,1]} |Y(t)r^{*(j)}(\beta,t)|\} < \infty, \quad j=0,1,2,$$

$$E\left\{\sup_{\mathcal{B}\times[0,1]} |Y(t)\left(\frac{r^{(1)}(\beta,t)}{r(\beta,t)}\right)^{\otimes 2j} r^*(\beta_0,t)|\right\} < \infty, \quad j=1,2,$$

$$E\left\{\sup_{\mathcal{B}\times[0,1]} |Y(t)\left(\frac{r^{(2)}(\beta,t)}{r(\beta,t)}\right)^{\otimes j} r^*(\beta_0,t)|\right\} < \infty, \quad j=1,2.$$

(5) Let $F_{Y(t),Z}$ be the joint distribution of $(Y(t),Z)$, and $f(t,z)=(\partial/\partial z)F_{Y(t),z}(1,z)$. For each $t\in[0,1]$, both $f(t,z)$ and $\phi(\beta,t)$ have the 2nd continuous derivative almost everywhere.

(6) $h\to 0$, $nh^{2d+3}\to 0$ and $nh^d(\log n)^2\to\infty$, as $n\to\infty$.

3.6 Properties of Local Polynomial estimators

Our proposed estimator is based on local linear estimation which we introduced in section 2.2. In particular, we employed the local linear kernel smoother ($p = 1$) in section (2.2) of chapter 2. In this section we will mention some properties of local linear estimators. Most of the proofs are given in Fan and Gijbels (1996), Fan and Yao (2005). First we show, how a local polynomial kernel smoother can be expressed as usual kernel estimator introduced by Nadaraya and Watson. Define S_k, for $k = 0, 1, ..p$ given by,

$$S_k = \sum_{i=1}^{n} (X_i - x_0)^k K_h(X_i - x_0).$$

Now, let, $S = X^T W X$, the $(p+1) \times (p+1)$ matrix $S_{k+l},\ 0 \leq k, l \leq p$.
Then the estimator $\hat{\beta}_\nu$ from section (1.4) can be written as

$$\begin{aligned} \hat{\beta}_\nu &= e^T_{\nu+1} \hat{\beta} \\ &= e^T_{\nu+1} S^{-1} X^T W y \\ &= \sum_{i \in V} W_\nu \Big(\frac{X_i - x_0}{h}) Y_i(t), \end{aligned} \tag{3.13}$$

where, W_ν is called the effective kernel and can be expresses as the following

$$W_\nu(x) \quad = \quad e^T_{\nu+1} S^{-1} \{1, xh, ..., (xh)^p\} K(x)/h \tag{3.14}$$

In the expression above W_ν depends on the design points and locations. That is why it can adapt automatically to various designs and to boundary estimation. The weights W_ν satisfies the following discrete moment conditions

$$\sum_{i=1}^{n} (X_i - x_0)^q W_\nu \Big(\frac{X_i - x_0}{h}) \quad = \quad \delta_{\nu,q} \ \ 0 \leq \nu, q \leq p, \tag{3.15}$$

where $\delta_{\nu,q} = 0$ if $\nu \neq q$ and 1, otherwise. It follows from the above result, that the local polynomial estimator is unbiased for estimating β_ν when the true regression function $m(x)$ is polynomial of order p. To prove the asymptotic properties we need the asymptotic form of the estimator in (3.13) .

Let $\mathbf{S}$ be the $(p+1) \times (p+1)$ matrix whose$(i,j)^{th}$ element is μ_{i+j-2}, where, $\mu_j = \int_{-\infty}^{\infty} u^j K(u)du$. With these notations we can define the equivalent kernel by,

$$K_\nu^*(x) = e_{\nu+1}^T S^{(-1)}\{1, x, ..., (x)^p\}K(x) = \Big(\sum_{l=0}^{p} \mathbf{S}^{\nu l} x^l\Big)K(x), \tag{3.16}$$

where $S^{\nu l}$ is the $(\nu+1, l+1)$-element of $\mathbf{S}^{-1}$..

Note that,

$$S_k \;\;=\;\; nh^k f(x_0)\mu_k\{1 + o_p(1)\} \tag{3.17}$$

From this, it follows that,

$$S \;\;=\;\; nh^k f(x_0)\mathbf{HSH}\{1 + o_p(1)\} \tag{3.18}$$

where, $\mathbf{H} = diag(1, h, .., h^p)$. substituting the above in the definition of W^ν, we have

$$W^\nu(x) = \frac{1}{nh\nu + 1 f(x_0)} e_{\nu+1}^T S^{-1}\{1, x, ..., (x)^p\}K(x)\{1 + o_p(1)\} \tag{3.19}$$

Therefore,

$$\hat{\beta}_\nu \;\;=\;\; \frac{1}{nh^{\nu+1} f(x_0)} \sum_{i=1}^{n} K_\nu^*\Big(\frac{X_i - x_0)}{h} Y_i\{1 + o_p(1)\} \tag{3.20}$$

where $K_\nu^*(x)$ is already defined in (3.16).

The kernel K_ν^* satisfies the following moment conditions:

$$\int u^q K_\nu^*(u)du = \delta_{\nu q} \; 0 \leq \nu, \; q \leq p,$$

This is an asymptotic version of the discrete moment conditions in (3.15). Next we give the expressions for bias and variance of the estimator $\hat{m}^\nu(x_0)$ with respect to the equivalent kernels K_ν^*

$$bias(\hat{m}_\nu(x_0)) = \Big(\int u^{p+1} K_\nu^*(u)du\Big)\frac{\nu!}{(p+1)!} m^{p+1}(x_0)h^{p+1-\nu} + o_p(h^{p+1-\nu}) \tag{3.21}$$

and

$$Var(\hat{m}_\nu(x_0)) = \int K_\nu^{*2}(u)du\frac{\nu!^2\sigma^2(x_0)}{f(x_0)nh^{1+2\nu}} + o_p\Big(\frac{1}{nh^{1+2\nu}}\Big). \tag{3.22}$$

Finally, we state two important results. The proofs are given in fan and Gijbels (1996).

If the design density f is uniformly continuous on $[a, b]$ with $inf_{x\in[a,b]}f(x) > 0$, then the local polynomial estimator has the following uniform convergence under the condition

$$sup_{x\in[a,b]}|\hat{m}(x) - m(x)| = O_p\Big(h^{p+1} + \{\frac{nh}{log(1/h)}\}^{-1/2}\Big). \tag{3.23}$$

Under condition (1) in §6.6.2 of Fan and Yao(2005) and if $h = O(n^{1/(2p+3)})$ and $m^{(p+1)}(.)$ is continuous at the point x, then as $n \to \infty$,

$$\begin{aligned} &\sqrt{nh}\Big[diag(1, h, ..., h^p)\{\hat{\beta}^p(x) - \beta_0(x)\} - \frac{h^{p+1}m^{(p+1)}(x)}{(p+1)!}\mathbf{S}^{-1}\mathbf{c}_p\Big] \\ \xrightarrow{\mathcal{D}}\; & N\{0, \sigma^2(x)\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}/f(x)\}, \end{aligned} \tag{3.24}$$

where $S = \{\mu_{i+j-2}\}_{i,j=1}^{p+1,p+1}$, $S^* = \{\nu_{i+j-2}\}_{i,j=1}^{p+1,p+1}$ with $\mu_j = \int_{-\infty}^{\infty} u^j K(u)du$ and $\nu_j =$

$\int_{-\infty}^{\infty} u^j K^2(u)du$.

3.7 Proofs

First we prove the two propositions mentioned in Chapter 2.

Proof of Proposition 2.1. Note that $\hat{\nu}_j - \nu_j = \sum_{i \in V} \omega_i(\nu_i - \nu_j) + \sum_{i \in V} \omega_i(\zeta_i - \nu_i)$. In the above ω_i is the effective kernel weight and can be expressed as

$$\omega_i(t, Z_j(t); h) = \frac{\{s_2 - (Z_i(t) - Z_j(t))s_1\} I_{[S_i \geq t]} K_h(Z_i(t) - Z_j(t))}{\sum_{i \in V} \{s_2 - (Z_i(t) - Z_j(t))s_1\} I_{[S_i \geq t]} K_h(Z_i(t) - Z_j(t))} = W_{\nu}^{n_v}$$

By standard nonparametric regression techniques (see for example Härdle, 1990; Fan and Gijbels, 1996), it can be shown that the first term above is $O_p(h^{p+1})$ as in (3.21) ($\nu = 0$), which is of order $o_p(1/\sqrt{n_v h^d})$ if one uses an undersmoothing bandwidth such that $nh^{2p+3} \to 0$, so that $\hat{\nu}_j - \nu_j = \sum_{i \in V} \omega_i(\zeta_i - \nu_i) + o_p(1/\sqrt{n_v h^d})$.

Similarly, $\hat{\psi}_j - \psi_j = \sum_{i \in V} \omega_i(\xi_i - \psi_i) + o_p(1/\sqrt{n_v h^d})$. Then the asymptotic normality can be obtained by using the Cramé-Wald device and directly computing the asymptotic mean and variance (see, for example the Lemma 6.3 in Jiang and Mack, 2001).

Let,

$$V_{n_1} = \sqrt{n_v h^d}[\hat{\nu}_j - \nu_j] = \sum w_i(\zeta_i - \nu_i) + o_p(1)$$

and

$$W_{n_2} = \sqrt{n_v h^d}[\hat{\xi}_j - \psi_j] = \sum w_i(\zeta_i - \nu_i) + o_p(1)$$

Let $W_n = aV_{n_1} + bV_{n_2} + o_p(1)$, where a, b are scalars.

Now, $E\{W_n | S_j \geq t, Z_j(t)\} = E\{aV_{n_1} + bV_{n_2} | S_j \geq t, Z_j(t)\}$

$$\begin{aligned} E\{aV_{n_1} | S_j \geq t, Z_j(t)\} &= E\Big[a \sum \{\omega_i(\zeta_i - E(\zeta_i | S_j \geq t, Z_j(t)\} | S_i \geq t, Z_i(t)\Big] \\ &\xrightarrow{P} 0 \end{aligned}$$

Similarly, $E\{bV_{n_2} | S_j \geq t, Z_j(t)\} \xrightarrow{P} 0$.

Now,

$$
\begin{aligned}
& Var\Big[W_n|S_j \geq t, Z_j(t)\Big] \\
= \; & Var\{aV_{n_1} + bV_{n_2}|S_j \geq t, Z_j(t)\} \\
= \; & a^2 Var\Big[\sum \omega_i(\zeta_i - \nu_i)|S_i \geq t, Z_i(t)\Big] + b^2 Var\Big[\sum \omega_i(\xi_i - \psi_i)|S_i \geq t, Z_i(t)\Big] \\
& +2abCov\Big[\Big(\sum \omega_i(\zeta_i - \nu_i), \sum \omega_i(\xi_i - \psi_i)\Big)|S_i \geq t, Z_i(t)\Big] \\
& \xrightarrow{P} (a \;\; b)\Sigma(a \;\; b)',
\end{aligned}
$$

where,

$$
\Sigma = v_0(K)p^{-1}(Z_j)\begin{bmatrix} \sigma_1^2(Z_j, t) & \rho_\alpha^*(Z_j, t)\sigma_1(Z_j, t)\sigma_2(Z_j, t) \\ \rho_\alpha^*(Z_j, t)\sigma_1(Z_j, t)\sigma_2(Z_j, t) & \sigma_2^2(Z_j, t) \end{bmatrix},
$$

where $v_0(K) = \int K^2(u)du$, $\sigma_1^2(Z_j, t) = Var[\zeta_j|S_j \geq t, Z_j]$, $\sigma_2^2(Z_j, t) = Var[\xi_j|S_j \geq t, Z_j]$, $\rho_\alpha^*(Z_j, t)$ is the conditional correlation coefficient between ζ_j and ξ_j given $(S_j \geq t, Z_j)$, and $p(\cdot)$ is the density function of Z.

Now, by properties of normal distribution, the result in (3.24) and Cramer-Wold device

$\sqrt{n_v h}[\{\hat{\nu}_j(\beta_1, t) - \nu_j(\beta_1, t)\}, \{\hat{\psi}_j(\alpha_1, t) - \psi_j(\alpha, t)\}]$ is jointly asymptotically normal with covariance matrix Σ defined above. Hence the proof is completed.

Proof of Proposition 2.2. Note that from (2.9)

$$
\begin{aligned}
[\bar{\nu}_j - \nu_j] \;\; = \;\; & [\hat{\nu}_j - \nu_j] \\
& -\rho^*(Z_j, t)\frac{\sigma_1(Z_j, t)}{\sigma_2(Z_j, t)}[(\hat{\psi}_j - \psi_j) - (\bar{\psi}_j - \psi_j)](1 + o_p(1)).
\end{aligned}
$$

The asymptotic normality of $\sqrt{n_v h^d}(\bar{\nu}_j - \nu_j)$ is obtained by the asymptotic normality of $\sqrt{n_v h^d}(\hat{\nu}_j - \nu_j)$, $\sqrt{n_v h^d}(\hat{\psi}_j - \psi_j)$ and $\sqrt{n h^d}(\bar{\psi}_j - \psi_j)$.

Note that $\frac{n_v}{n} \to \rho \; as \; n \to \infty$.

Using the property of multivariate normal variables and Slutsky's theorem

$$\sqrt{n_v h^d}(\bar{\nu}_j - \nu_j) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega),$$

where

$$\begin{aligned}
&\Omega = v_0(K)p^{-1}(Z_j)\Big[\sigma_1^2(Z_j,t) + \rho_\alpha^{*2}(Z_j,t)\frac{\sigma_1^2(Z_j,t)}{\sigma_2^2(Z_j,t)}\sigma_2^2(Z_j,t) \\
&+\rho_\alpha^{*2}(Z_j,t)\frac{\sigma_1^2(Z_j,t)}{\sigma_2^2(Z_j,t)}\rho\sigma_2^2(Z_j,t) - 2\rho_\alpha^{*2}(Z_j,t)\frac{\sigma_1(Z_j,t)}{\sigma_2(Z_j,t)}\sigma_1(Z_j,t)\sigma_2(Z_j,t) \\
&-2\rho\rho_\alpha^{*2}(Z_j,t)\frac{\sigma_1^2(Z_j,t)}{\sigma_2^2(Z_j,t)}\sigma_2^2(Z_j,t) + 2\rho\rho_\alpha^{*2}(Z_j,t)\frac{\sigma_1^2(Z_j,t)}{\sigma_2^2(Z_j,t)}\sigma_2^2(Z_j,t)\Big] \\
&= \sigma_1^2(Z_j,t)[1-(1-\rho)\rho_\alpha^{*2}(Z_j,t)]v_0(K)p^{-1}(Z_j)
\end{aligned}$$

Hence proved.

Next we need the following theorems and lemmas to show the consistency and asymptotic normality of our proposed estimator. Some of the proofs are given in Anderson and Gill(1982) and Zhou PhD Dissertation(1992). We follow their idea and the proofs relevant to our model.

Theorem 3.2 *Under the conditions in section (3.5)*

$\sup_{\mathcal{B}\times[0,1]} \| \bar{\phi}(\beta,t) - \phi(\beta,t) \| \to 0$ *a.s* *and* $\sup_{\mathcal{B}\times[0,1]} \| \hat{r}(\beta,t) - r(\beta,t) \| \to 0$ *a.s*

Proof. Consider the notations defined in section 2.2. Note that from (2.9)

$$\begin{aligned}
[\bar{\nu}_j - \nu_j] &= [\hat{\nu}_j - \nu_j] \\
&\quad -\hat{\rho}^*(Z_j,t)\frac{\hat{\sigma}_1(Z_j,t)}{\hat{\sigma}_2(Z_j,t)}[(\hat{\psi}_j - \psi_j) + (\bar{\psi}_j - \psi_j)]
\end{aligned}$$

Now applying the theorem 6.5 in Fan and Yao (2005) given in (3.23) and the condition (6) in section 3.5, we have $\hat{\nu}_j - \nu_j \xrightarrow{a.s} 0$.

Similarly from the definition of $\hat{\psi}_j$ and $\bar{\psi}_j$, and the same argument for local polynomial estimators

$(\hat{\psi}_j - \psi_j) \xrightarrow{a.s} 0$

$(\bar{\psi}_j - \psi_j) \xrightarrow{a.s} 0.$

Also the local polynomial kernel estimates $\hat{\rho}^*_\alpha(Z_j, t)$, $\hat{\sigma}_1(Z_j, t)$ and $\hat{\sigma}_1(Z_j, t)$ converges to $\rho^*_\alpha(Z_j, t)$, $\sigma_1(Z_j, t)$ and $\sigma_2(Z_j, t)$ respectively.

Therefore,

$\sup_{\mathcal{B}\times[0,1]} \parallel \bar{\phi}(\beta, t) - \phi(\beta, t) \parallel \to 0 \ a.s.$

Now, $(\hat{r_i}(\beta, t) - r_i(\beta, t))$=0 when $\eta_i = 1$.

So $\sup_{\mathcal{B}\times[0,1]} \parallel \hat{r}(\beta, t) - r(\beta, t) \parallel \to 0 \ a.s.$

Theorem 3.3 *Under the conditions in section , for* $k = 0, 1, \ldots\ldots, 4$

$$\sup_{\mathcal{B}\times[0,1]} \parallel \hat{S}^{(k)}(\beta, t) - S^{(k)}(\beta, t) \parallel \to 0 \ a.s$$

and

$$\sup_{\mathcal{B}\times[0,1]} \parallel \hat{S}^{(k)}(\beta, t) - s^{(k)}(\beta, t) \parallel \to 0 \ a.s$$

Proof. We shall prove the above result for for $k = 0$. The remaining results can be proved in a similar way. By the definition of $\hat{S}^{(0)}(\beta, t)$ and $S^{(0)}(\beta, t)$ and by the theorem (3.2)

$\sup_{\mathcal{B}x[0,1]} \parallel \hat{S}^{(0)}(\beta, t) - S^{(0)}(\beta, t) \parallel \to 0 \ a.s$

Next, by the definition of $s^{(0)}$ and applying the uniform stron law of large numbers

$$\sup_{\mathcal{B}\times[0,1]} \parallel S^{(0)}(\beta, t) - s^{(0)}(\beta, t) \parallel \to 0 \ a.s$$

Now,

$$\begin{aligned}\sup_{\mathcal{B}\times[0,1]} \parallel \hat{S}^{(0)}(\beta, t) - s^{(0)}(\beta, t) \parallel \quad & \leq \sup_{\mathcal{B}\times[0,1]} \parallel S^{(0)}(\beta, t) - S^{(0)}(\beta, t) \parallel \\ & + \sup_{\mathcal{B}\times[0,1]} \parallel S^{(0)}(\beta, t) - s^{(0)}(\beta, t) \parallel\end{aligned}$$

Hence, it follows directly,

$$\sup_{\mathcal{B}\times[0,1]} \parallel \hat{S}^{(0)}(\beta,t) - s^{(0)}(\beta,t) \parallel \to 0 \ a.s$$

Lemma 3.1

$$n^{-1/2}\sum_{i=1}^{n}\int_0^1 (\hat{r}_i^{(k)}(\beta_0,w) - r_i^{(k)}(\beta_0,w))^2 Y_i(w) r_i^*(\beta_0,w)\lambda_0(w)\,dw \xrightarrow{p} 0, \quad k=0,1$$

$$n^{-1/2}\sum_{i=1}^{n}\int_0^1 (\hat{S}^{(k)}(\beta_0,w) - S^{(k)}(\beta_0,w))^2 Y_i(w) r_i^*(\beta_0,w)\lambda_0(w)\,dw \xrightarrow{p} 0, \quad k=0,1.$$

Proof. The proof of the above theorem is similar to the lemma 2.4 of Zhou (1992, PhD dissertation)

Lemma 3.2

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(\hat{r}_i)(\beta_0,w) Y_i(w) r_i^*(\beta_0,w)\lambda_0(w)\,dw \\ = \; & -n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(r_i)(\beta_0,w) Y_i(w)[\hat{r}_i(\beta_0,w) - r_i(\beta_0,w)]\lambda_0(w)\,dw \\ & -n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(r_i)(\beta_0,w) Y_i(w)[r_i(\beta_0,w) - r_i^*(\beta_0,w)]\lambda_0(w)\,dw + o_p(1) \end{aligned}$$

Proof. By the Taylor expansion, the second term of $\hat{U}(\beta,t)$ in (3.7) admits the following decomposition

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(\hat{r}_i)(\beta_0,w) Y_i(w) r_i^*(\beta_0,w)\lambda_0(w)\,dw \\ = \; & -n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(r_i)(\beta_0,w) Y_i(w)[\hat{r}_i(\beta_0,w) - r_i(\beta,w)]\lambda_0(w)dw \\ & -n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(r_i)(\beta_0,w) Y_i(w)[r_i(\beta_0,w) - r_i^*(\beta_0,w)]\lambda_0(w)dw + o_p(1). \end{aligned}$$

$$\begin{aligned} f(x,y) &= f(x_0,y_0) + \frac{\partial f(x,y)}{\partial x}\Big|_{x_0,y_0}(x-x_0) \\ &+ \frac{\partial f(x,y)}{\partial y}\Big|_{x_0,y_0}(y-y_0) + O((x-x_0)^2+(y-y_0)^2), \end{aligned}$$

if $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$, and $\frac{\partial^2 f}{\partial x \partial y}$ are finite. Then $\frac{\hat{r}^{(1)}}{\hat{r}} = \frac{\hat{r}^{(1)}}{r} - \frac{r^{(1)}(\hat{r}-r)}{r^2} + O[(\hat{r}-r)^2 + (\hat{r}^{(1)} - r^{(1)})^2]$ $\frac{\hat{S}^{(1)}}{\hat{S}^{(0)}} = \frac{\hat{S}^{(1)}}{S^{(0)}} - \frac{S^{(1)}(\hat{S}-S^{(0)})}{S^{(0)2}} + O[(\hat{S}-S^{(0)})^2 + (\hat{S}^{(1)} - S^{(1)})^2]$.

Note that $\sum_i \Delta \hat{r}_i(u)\hat{r}_i(u)Y_i(u) = 0$.

It follows that the left side of the result in the lemma can be expressed as

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(\hat{r}_i)(\beta_0,w)Y_i(w)r_i^*(\beta_0,w)\lambda_0(w)\,dw \\ = & -n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(\hat{r}_i)(\beta_0,w)Y_i(w)[\hat{r}_i(\beta_0,w) - r_i^*(\beta_0,w)]\lambda_0(w)\,dw \\ = & -n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(r_i)(\beta_0,w)Y_i(w)[\hat{r}_i(\beta_0,w) - r_i(\beta_0,w) + r_i(\beta_0,w) \\ & -r_i^*(\beta_0,w)]\lambda_0(w)\,dw + o_p(1) \\ = & -n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(r_i)(\beta_0,w)Y_i(w)[\hat{r}_i(\beta_0,w) - r_i(\beta_0,w)]\lambda_0(w)\,dw \\ & -n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(r_i)(\beta_0,w)Y_i(w)[r_i(\beta_0,w) - r_i^*(\beta_0,w)]\lambda_0(w)\,dw + o_p(1) \end{aligned}$$

where the last equality is from Lemma 2.4 of Zhou (1992). Therefore the result holds.

Lemma 3.3

$$sup_{\beta\in\mathcal{B}}\| - \frac{1}{n}\frac{\partial \hat{U}(\beta,1)}{\partial \beta} - I(\beta)\| \xrightarrow{P} 0.$$

Furthermore, $-\frac{1}{n}\frac{\partial \hat{U}(\beta_0)}{\partial \beta}$ *is positive definite with probability going to 1.*

Proof. From equation (3.6) we have

$$\hat{U}(\beta,t) = \sum_{i=1}^{n}\int_0^t \Big[\frac{\hat{r}_i^{(1)}(\beta,u)}{\hat{r}_i(\beta,u)} - \frac{\sum_{i=1}^n Y_i(u)\hat{r}_i^{(1)}(\beta,u)}{\sum_{i=1}^n Y_i(u)\hat{r}_i(\beta,u)}\Big]dNi(u) \qquad (3.25)$$

Differentiating with respect to β, we have,

$$\begin{aligned} \partial \hat{U}(\beta,1)/\partial\beta &= \int_0^1 \sum_{i=1}^n \Big[\frac{\hat{r}_i^{(2)}(\beta,t)}{\hat{r}_i^{(0)}(\beta,t)} - \Big(\frac{\hat{r}_i^{(1)}(\beta,t)}{\hat{r}_i^{(0)}(\beta,t)}\Big)^{\otimes 2} \\ &\quad -\frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} + \Big(\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\Big)^{\otimes 2}\Big] dN_i(t). \end{aligned}$$

Now, we define the process,

$$\begin{aligned} & C(\beta,t) \\ = & \int_0^1 \sum_{i=1}^n \Big[\frac{\hat{r}_i^{(2)}(\beta,t)}{\hat{r}_i^{(0)}(\beta,t)} - \Big(\frac{\hat{r}_i^{(1)}(\beta,t)}{\hat{r}_i^{(0)}(\beta,t)}\Big)^{\otimes 2} \\ & -\frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} + \Big(\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\Big)^{\otimes 2}\Big] Yi(t) r_i^*(\beta_0,t)\lambda_0(t)dt. \end{aligned}$$

Then,

$$\begin{aligned} n^{-1}\partial \hat{U}(\beta,1)/\partial\beta - n^{-1}C(\beta,1) &= \int_0^1 n^{-1} \sum_{i=1}^n \Big[\frac{\hat{r}_i^{(2)}(\beta,t)}{\hat{r}_i^{(0)}(\beta,t)} - \Big(\frac{\hat{r}_i^{(1)}(\beta,t)}{\hat{r}_i^{(0)}(\beta,t)}\Big)^{\otimes 2} \\ &\quad -\frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} + \Big(\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\Big)^{\otimes 2}\Big] dM_i(t). \end{aligned}$$

which is a local square integrable martingale by condition (3) and the covariance process is given by

$$\langle n^{-1}\partial \hat{U}(\beta,.)/\partial\beta - n^{-1}C(\beta,.), n^{-1}\partial \hat{U}(\beta,1)/\partial\beta - n^{-1}C(\beta,t)\rangle = B(\beta,.),$$

where

$$
\begin{aligned}
B(\beta,1) \quad = \quad & \int_0^1 \Big[\frac{1}{n^2}\sum_{i=1}^{n}\Big(\frac{\hat{r}_i^{(2)}(\beta,t)}{\hat{r}_i^{(0)}(\beta,t)}\Big)^{\otimes 2} r_i^*(\beta_0,t)\lambda_0(t) \\
& + \frac{1}{n^2}\sum_{i=1}^{n}\Big((\frac{\hat{r}_i^{(1)}(\beta,t)}{\hat{r}_i^{(0)}(\beta,t)}\Big)^{\otimes 4} r_i^*(\beta_0,t)\lambda_0(t) \\
& + \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\Big)^{\otimes 2}\frac{1}{n}\sum_{i=1}^{n} Y_i(t) r_i^*(\beta_0,t)\lambda_0(t) \\
& + \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\Big)^{\otimes 4}\frac{1}{n}\sum_{i=1}^{n} Y_i(t) r_i^*(\beta_0,t)\lambda_0(t)\Big] dt \\
& + \, remaining \;\; terms
\end{aligned}
$$

By the definitions in section (3.2) and conditions in (3.5) all the terms converge to zero. Therefore $\|B(\beta,1)\|_{\mathcal{B}} \xrightarrow{P} 0$. Now, by Lenglart's inequality(Apendix I Anderson and Gill 1982) it follows that $\frac{1}{n}\hat{U}(\beta,t)$ and $\frac{1}{n}C(\beta,t)$ converges in probability to the same limit uniformly in $\beta \in \mathcal{B}$. By theorem (3.3) and conditions in section (3.5)

$$
\begin{aligned}
& \frac{1}{n}C(\beta,t) \\
& \xrightarrow{P} \int_0^1 \sum_{i=1}^{n}\Big[s^{(4)}(\beta,t) - s^{(5)}(\beta,t) - \frac{s^{(2)}(\beta,t)}{s^{(0)}(\beta,t)} s^{(0)}(\beta_0,t) + \Big(\frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}\Big)^{\otimes 2} s^{(0)}(\beta_0,t)\Big]\lambda_0(t)dt
\end{aligned}
$$

$\equiv -I(\beta)$ uniformly in β in the neighborhood of β_0.

Hence,

$$
sup_{\beta\in\mathcal{B}}\|\frac{1}{n}\frac{\partial \hat{U}(\beta,1)}{\partial\beta} - I(\beta)\| \xrightarrow{P} 0.
$$

At $\beta = \beta_0$, $I(\beta) = I(\beta_0)$ which is positive definite by condition (3). Hence the proof is completed.

Lemma 3.4 *Under the conditions in section (3.5) , we have:*

$$
\begin{aligned}
& n^{-1/2}\hat{U}(\beta_0, 1) \\
= \; & n^{-1/2}\sum_{i\in\bar{V}}\left[\int_0^1 \left\{\frac{\phi_i^{(1)}(\beta_0,u)}{\phi_i(\beta_0,u)} - \frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)}\right\} dM_i(s) - \frac{n-n_v}{n}Q_i^* + Q_i^{**}\right] + o_p(1) \\
+ \; & n^{-1/2}\sum_{i\in V}\left[\int_0^1 \left\{\frac{r_i^{(1)}(\beta_0,u)}{r_i(\beta_0,u)} - \frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)}\right\} dM_i(s) - \frac{n-n_v}{n_v}(Q_i - Q_i^*\frac{n-n_v}{n})\right].
\end{aligned}
$$

where Q, Q^*, Q^{**} are defined in section (3.2).

Proof. Note that $\hat{r}_i - r_i = (1-\eta_i)(\bar{\phi}_i - \phi_i)$ and $r_i - r_i^* = (1-\eta_i)(\phi_i - \phi_i^*)$. Applying the first order expansion $x/y = x_0/y_0 + (x-x_0)/y_0 - (y-y_0)x_0/y_0^2 + O((x-x_0)^2 + (y-y_0)^2)$ to $\hat{r}^{(1)}/\hat{r}$ and $\hat{S}^{(1)}/\hat{S}^{(0)}$ around $(r^{(1)}, r)$ and $(s^{(1)}, s^{(0)})$, respectively, and by lemma 3.2 we can rewrite the second summation of $n^{-1/2}\hat{U}(\beta_0, 1)$ in (3.12) as

$$
\begin{aligned}
& \Delta(r_i)(\beta_0,w)Y_i(w)[\hat{r}_i(\beta_0,w) - r_i(\beta_0,w) + r_i(\beta_0,w) - r_i^*(\beta_0,w)]\lambda_0(w)\,dw + o_p(1) \\
= \; & -n^{-1/2}\sum_{i=1}^n \int_0^1 \Delta(r_i)(\beta_0,w)Y_i(w)[\hat{r}_i(\beta_0,w) - r_i(\beta_0,w)]\lambda_0(w)\,dw \\
& -n^{-1/2}\sum_{i=1}^n \int_0^1 \Delta(r_i)(\beta_0,w)Y_i(w)[r_i(\beta_0,w) - r_i^*(\beta,w)]\lambda_0(w)\,dw + o_p(1) \\
= \; & -n^{-1/2}\sum_{j\in\bar{V}} \int_0^1 (\bar{\phi}_j - \phi_j)\Delta(\phi_j)(u)Y_j(u)\lambda_0(u)du \\
& +n^{-1/2}\sum_{j\in\bar{V}} \int_0^1 (\phi_j^* - \phi_j)\Delta(\phi_j)(u)Y_j(u)\lambda_0(u)du + o_p(1) \\
= \; & I_{n_1} + I_{n_2} + o_p(1). \qquad (3.26)
\end{aligned}
$$

Note that $\hat{\phi}_j(\beta, t) = \hat{\nu}_j(\beta_1, t)\exp\{\beta_2' Z_j(t)\}$.

Since

$$\begin{aligned}
\bar{\phi}_j - \phi_j &= (\hat{\phi}_j - \phi_j) - \exp\{\beta_2' Z_j(u)\}\rho_\alpha^*(Z_j, u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}(\hat{\psi}_j - \bar{\psi}_j)(1+o_p(1)) \\
&= \sum_{i\in V}\omega_i(\gamma_i - \phi_j) - \exp\{\beta_2' Z_j(u)\}[\sum_{i\in V}\omega_i(\xi_i - \psi_j)\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)} \\
&\quad - \sum_{i\in V\cup\bar{V}}\bar{\omega}_i(\xi_i - \psi_j)\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}](1+o_p(1)) + o_p(\frac{1}{\sqrt{n}}) \\
&= \sum_{i\in V}\omega_i[(\gamma_i - \phi_j) - \exp\{\beta_2' Z_j(u)\}\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}(\xi_i - \psi_j)](1+o_p(1)) \\
&\quad + \sum_{i\in V\cup\bar{V}}\bar{\omega}_i(\xi_i - \psi_j)\exp\{\beta_2' Z_j(u)\}\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}(1+o_p(1)) + o_p(\frac{1}{\sqrt{n}}),
\end{aligned}$$

the first term in (3.26) can be rewritten as

$$\begin{aligned}
& I_{n_1} = -n^{-1/2}\sum_{j\in\bar{V}}\int_0^1 \Delta(\phi_j)(u)Y_j(u)\lambda_0(u) \\
& \times\Big\{\sum_{i\in V}\omega_i[(\gamma_i - \phi_j) - \exp\{\beta_2' Z_j(u)\}\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}(\xi_i - \psi_j)] \\
& + \sum_{i\in V\cup\bar{V}}\bar{\omega}_i(\xi_i - \psi_j)\exp\{\beta_2' Z_j(u)\}\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}\Big\}du(1+o_p(1)) + o_p(1) \\
\equiv\ & J_{\tilde{n1}} + J_{\tilde{n2}} + o_p(1).
\end{aligned}$$

Note that

$$n_v^{-1}\sum_{i\in V}Y_i(t)K_h(Z_i - Z_j) = f(t,Z_j)(1+o_p(1)),$$

$$n^{-1}\sum_{i\in V\cup\bar{V}}Y_i(t)K_h(Z_i - Z_j) = f(t,Z_j)(1+o_p(1)),$$

$$\omega_i(t,Z_j;h) = f^{-1}(t,Z_j)(1+o_p(1))\ n_v^{-1}Y_i(t)K_h(Z_i - Z_j),$$

$$\bar{\omega}_i(t,Z_j;h) = f^{-1}(t,Z_j)(1+o_p(1))\ n^{-1}Y_i(t)K_h(Z_i - Z_j),$$

uniformly for $j = 1, \cdots, n$. Then

$$\begin{aligned}
J_{\tilde{n}_1} &= -\frac{1}{\sqrt{n}} \sum_{j \in \bar{V}} \int_0^1 \Delta(\phi_j)(u) Y_j(u) \lambda_0(u) f^{-1}(u, Z_j) \cdot \\
& \frac{1}{n_v} \sum_{i \in V} Y_i(u) K_h(Z_i - Z_j)[(\gamma_i - \phi_j) - \exp\{\beta_2' Z_j(u)\} \rho_\alpha^*(Z_j, u) \frac{\sigma_1(Z_j, u)}{\sigma_2(Z_j, u)}(\xi_i - \psi_j)] du \\
&+ o_p(1) \\
&= -\frac{1}{\sqrt{n}} \frac{n - n_v}{n_v} \sum_{i \in V} [Q_i - Q_i^*] + o_p(1),
\end{aligned}$$

$$\begin{aligned}
J_{\tilde{n}_2} &= -\frac{1}{\sqrt{n}} \sum_{j \in \bar{V}} \int_0^1 \Delta(\phi_j)(u) Y_j(u) \lambda_0(u) \exp\{\beta_2' Z_j(u)\} \rho_\alpha^*(Z_j, u) \frac{\sigma_1(Z_j, u)}{\sigma_2(Z_j, u)} \\
& \times \frac{1}{n} \sum_{i \in V \cup \bar{V}} Y_i(u) K_h(Z_i - Z_j)(\xi_i - \psi_j) f^{-1}(u, Z_j) du + o_p(1) \\
&= -\frac{1}{\sqrt{n}} \frac{n - n_v}{n} \sum_{i \in V \cup \bar{V}} Q_i^* + o_p(1).
\end{aligned}$$

Again, since $\phi_j^* - \phi_j = \phi_j^* - E\{\phi_j^* | Y_i(t) = 1, Z_i(t)\}$

The second term in (3.26) can be rewritten as

$$\begin{aligned}
In_2 &= n^{-1/2} \sum_{j \in \bar{V}} \int_0^1 [\phi_j^* - E\{\phi_j^* | Y_i(u) = 1, Z_i(u)] \Delta(\phi_j)(u) Y_j(u) \lambda_0(u) \\
&\equiv n^{-1/2} \sum_{j \in \bar{V}} Q_i^{**}.
\end{aligned}$$

Therefore, the second summation of $n^{-1/2}\hat{U}(\beta, 1)$ in (3.12) equals

$$-\frac{1}{\sqrt{n}} \frac{n - n_v}{n_v} \sum_{i \in V} [Q_i - Q_i^*] - \frac{1}{\sqrt{n}} \frac{n - n_v}{n} \sum_{i \in V \cup \bar{V}} Q_i^* + n^{-1/2} \sum_{j \in \bar{V}} Q_i^{**} + o_p(1).$$

Hence, $n^{-1/2}\hat{U}(\beta, 1)$ can be expressed as

$$
\begin{aligned}
& n^{-1/2}\sum_{i\in\bar{V}}\left[\int_0^1\left\{\frac{\phi_i^{(1)}(\beta_0,u)}{\phi_i(\beta_0,u)}-\frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)}\right\}dM_i(s)-\frac{n-n_v}{n}Q_i^*+Q_i^{**}\right]+o_p(1)\\
+\ & n^{-1/2}\sum_{i\in V}\left[\int_0^1\left\{\frac{r_i^{(1)}(\beta_0,u)}{r_i(\beta_0,u)}-\frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)}\right\}dM_i(s)-\frac{n-n_v}{n_v}(Q_i-Q_i^*\frac{n-n_v}{n})\right].
\end{aligned}
$$

Lemma 3.5 *Under the conditions in section (3.5)*

$$n^{-1}\hat{U}(\beta_0,1)\xrightarrow{a.s}0.$$

Proof. From lemma (3.4) we can write,

$$
\begin{aligned}
& n^{-1/2}\hat{U}(\beta_0,1)\\
=\ & n^{-1/2}\sum_{i\in\bar{V}}\left[\int_0^1\left\{\frac{\phi_i^{(1)}(\beta_0,u)}{\phi_i(\beta_0,u)}-\frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)}\right\}dM_i(s)-\frac{n-n_v}{n}Q_i^*+Q_i^{**}\right]+o_p(1)\\
+\ & n^{-1/2}\sum_{i\in V}\left[\int_0^1\left\{\frac{r_i^{(1)}(\beta_0,u)}{r_i(\beta_0,u)}-\frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)}\right\}dM_i(s)-\frac{n-n_v}{n_v}(Q_i-Q_i^*\frac{n-n_v}{n})\right].
\end{aligned}
$$

Note that $M_i(t)$ is a martingale with mean zero. Also $E[Q_i^*]=0$, $E[Q_i^{**}]=0$ and $E[Q_i]=0$. Then, by strong law of large numbers, we have

$$n^{-1}\hat{U}(\beta_0,1)\xrightarrow{a.s}0$$

3.7.1 Proof of consistency of $\hat{\beta}_{EPL}$

Theorem 3.4 *$\hat{\beta}_{EPL}$ is a consistent estimator for β_0.*

Proof. We have shown that $n^{-1}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)$ exists and is continuous in an open neighborhood $\mathcal{B}$ of β_0. Now, by the lemma (3.3) $-n^{-1}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)$ converges in probability to a fixed function $I(\beta)$, uniformly in an open neighborhood of β_0. Also every element of $I(\beta)$ is a continuous function of β in the neighborhood of β_0 and $I^{-1}(\beta_0)$

exists. Next, by condition 3 in section (3.5) $-n^{-1}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)$ is positive-definite with probability going to 1. Finally from lemma (3.5) we have, $n^{-1}\frac{\partial}{\partial\beta}\hat{U}(\beta_0,1) \xrightarrow{a.s} 0$.

Using the above results, Inverse Function Theorem given in section 3.3 and following closely the arguments of Foutz (1977), $\hat{\beta}_{EPL}$ is a consistent estimator for β_0.

3.7.2 Proof of Asymptotic Normality of $\hat{\beta}_{EPL}$

Theorem 3.5

$$-\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)|_{\beta=\beta^*} \xrightarrow{a.s} I(\beta_0)$$

Proof. In lemma (3.3) we have shown that,

$-\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1) \xrightarrow{a.s} I(\beta)$ for any $\beta \in \mathcal{B}$ and that $I(\beta_0)$ is positive definite, where

$$\begin{aligned}
&I(\beta_0)\\
&= \int_0^1 \Big[s^{(3)}(\beta_0,t) - s^{(4)}(\beta_0,t) - \frac{s^{(2)}(\beta_0,t)}{s^{(0)}(\beta_0,t)}s^{(0)}(\beta_0,t) + \Big(\frac{s^{(1)}(\beta_0,t)}{s^{(0)}(\beta_0,t)}\Big)^{\otimes 2} s^{(0)}(\beta_0,t)\Big]\lambda_0(t)dt\\
&= \int_0^1 \Big[\frac{s^{(2)}(\beta_0,t)}{s^{(0)}(\beta_0,t)} - \Big(\frac{s^{(1)}(\beta_0,t)}{s^{(0)}(\beta_0,t)}\Big)^{\otimes 2}\Big]s^{(0)}(\beta_0,t)\lambda_0(t)dt
\end{aligned}$$

$I(\beta)$ is continuous in β. Now, for β_* lying between $\hat{\beta}_{EPL}$ and β_0

$$\begin{aligned}
&\mid -\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1) - I(\beta_0) \mid\\
&\qquad = \mid -\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1) - I(\beta^*) + I(\beta^*) - I(\beta_0) \mid\\
&\qquad \leq \mid -\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1) - I(\beta) \mid \ + |I(\beta^*) - I(\beta_0)|
\end{aligned}$$

The first term on the right hand side goes to zero in probability by lemmma 2. Since $\hat{\beta}_{EPL}$ is consistent estimator for β_0 by theorem 3.4 and I is continuous, the second term converges to zero in probability. Hence,

$$-\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)|_{\beta=\beta^*} \xrightarrow{a.s} I(\beta_0) \ as \ n \ \to \ \infty$$

.

Theorem 3.6 *Suppose that Conditions in section 3.5 holds. Then* $\hat{\beta}_{EPL}$ *satisfies*

$$\sqrt{n}(\hat{\beta}_{EPL} - \beta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega(\beta_0)),$$

where $\Omega(\beta_0) = I^{-1}(\beta_0)\Sigma(\beta_0)I^{-1}(\beta_0)$ *with* $\Sigma(\beta_0) = (1-\rho)\Sigma_1(\beta_0) + \rho\Sigma_2(\beta_0)$,

Proof: By (3.12), $\hat{\beta}_{EPL}$ solves the equation $\hat{U}(\beta, 1) = 0$. By Taylor's expansion,

$$n^{-1/2}\hat{U}(\beta_0, 1) = -n^{-1}\frac{\partial \hat{U}(\beta_*, 1)}{\partial \beta}\sqrt{n}(\hat{\beta}_{EPL} - \beta_0), \tag{3.27}$$

where β_* is between $\hat{\beta}_{EPL}$ and β_0. By Lemma (3.3) and consistency of $\hat{\beta}_{EPL}$,

$$-n^{-1}\frac{\partial \hat{U}(\beta_*, 1)}{\partial \beta} \xrightarrow{P} I(\beta_0).$$

Therefore, to prove the asymptotic normality it suffices to show that $n^{-1/2}\hat{U}(\beta_0, 1)$ is asymptotically normal with mean 0 and variance $(1-\rho)\Sigma_1(\beta_0) + \rho\Sigma_2(\beta_0)$.

From lemma (3.4) we have,

$$\begin{aligned} & n^{-1/2}\sum_{i \in \bar{V}}\left[\int_0^1 \left\{\frac{\phi_i^{(1)}(\beta_0, u)}{\phi_i(\beta_0, u)} - \frac{s^{(1)}(\beta_0, u)}{s^{(0)}(\beta_0, u)}\right\} dM_i(s) - \frac{n - n_v}{n}Q_i^* + Q_i^{**}\right] + o_p(1) \\ + \quad & n^{-1/2}\sum_{i \in V}\left[\int_0^1 \left\{\frac{r_i^{(1)}(\beta_0, u)}{r_i(\beta_0, u)} - \frac{s^{(1)}(\beta_0, u)}{s^{(0)}(\beta_0, u)}\right\} dM_i(s) - \frac{n - n_v}{n_v}(Q_i - Q_i^*\frac{n - n_v}{n})\right]. \end{aligned}$$

Now, $\Delta\phi(\beta_0, u)$ is locally bounded by the given conditions. By the martingale central limit theorem the first term above converges weakly to a gaussian process with covariance $(1-\rho)\Sigma_1(\beta_0)$. The third term above is a sum of independently distributed terms with mean zero from the validation subsample. Then this term is asymptotically normal with mean zero and variance $\rho\Sigma_2(\beta_0)$. By independence of the two terms, $n^{-1/2}\hat{U}(\beta_0, 1) \xrightarrow{P} N(0, \Sigma(\beta_0))$ with $\Sigma(\beta_0) = (1-\rho)\Sigma_1(\beta_0) + \rho\Sigma_2(\beta_0)$.

CHAPTER 4: SIMULATIONS

In this section, we conduct finite-sample simulations. The aims of the simulations are three-fold: one is to examine the small sample behavior of $\hat{\beta}_{EPL}$, another is to compare the performance of our estimator with some existing estimators under various situations, and the third and the most important is to illustrate that the proposed estimation allows for an informative auxiliary vector W.

4.1 Generation of Data

The covariates (X, Z) are generated from the following transformation to create correlation:

$$\begin{pmatrix} X \\ Z \end{pmatrix} = \begin{pmatrix} 1 & 0.0 \\ 0.5 & 1 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}, \tag{4.1}$$

where U_i's are independent and identically distributed as $U(0, 2)$. The failure time T conditional on covariate X is from an exponential distribution with hazard function

$$\lambda(t; X) = \lambda exp(\beta_1 X + \beta_2 Z),$$

where, λ is the baseline constant hazard. We only consider the case $\lambda = 1$.

Then

$$f(t; X, Z) = exp(\beta_1 X_1 + \beta_2 X_2) exp(-te^{(\beta_1 X_1 + \beta_2 X_2)}).$$

The auxiliary variable W is generated from

$$W = X + \gamma \log(T) + e, \tag{4.2}$$

where $e \sim \mathcal{N}(0, \sigma^2)$ and σ^2 is the parameter controlling the strength of the association between X and W. We consider the settings with $\gamma = 0$, 2 and 4. The model (4.2) with $\gamma = 2$ allows one to explore the effectiveness of the proposed method with an informative surrogate W. For $\gamma = 0$, it also allows us to compare the performance of the newly proposed method and that in Zhou and Wang (2000). We do simulations for $\sigma = 0.2$, 0.8 and 1.6. The censoring variable C is uniformly distributed and is independent of the failure time variable. It is generated from the uniform $(0, c)$ distribution where c is a parameter which determines the percentage of censoring in the sample.

For $\beta = (ln(2), 0.5))'$, the values of c for 20%, 50% and 80% censoring obtained are 0.1353, 0.372 and 0.0965 respectively. The observation time is then obtained from $S = T \wedge C$.

The validation set is randomly selected by using $P(\eta_i = 1) = 0.5$. We choose the Gaussian kernel function with the bandwidths $h = (\hat{\sigma}_Z n^{-1/3}$ which satisfy the bandwidth conditions in section 3.5, where $\hat{\sigma}_Z$ is the sample standard deviation of Z.

4.2 Implementation method in finite samples

In the proposed estimation method we obtained the pseudo-partial likelihood by considering all the subjects both in the validation and non-validation sets. For the subjects in the non-validation set we estimated the relative risk function by the kernel smoothing approach. Recall that, the estimated relative risk defined in (2.9) is given by

$$\hat{r}_i(\beta, t) = \eta_i \gamma_i(\beta, t) + (1 - \eta_i)\bar{\phi}_i(\beta, t),$$

where $\bar{\phi}_i(\beta, t) = \bar{\nu}_i(\beta_1, t) \exp\{\beta_2' Z_i(t)\}$ and for $j \in \bar{V}$,

$$\bar{\nu}_j(\beta_1, t) = \hat{\nu}_j(\beta_1, t) - \hat{\rho}^*_\alpha(Z_j, t)\frac{\hat{\sigma}_1(Z_j, t)}{\hat{\sigma}_2(Z_j, t)}[\hat{\psi}_j(\alpha, t) - \bar{\psi}_j(\alpha, t)].$$

Now consider the weight function in the definitions of $\hat{\nu}_j(\beta_1, t)$, $\hat{\psi}_i(\beta, t)$ and $\bar{\psi}_i(\beta, t)$ in Section (4.1). $\hat{\nu}_j(\beta_1, t)$ is undefined when the denominator is zero. It happens when the risk set in the validation set is a null set. Similar situations occur in estimation of $\hat{\psi}_i(\beta, t)$ and $\bar{\psi}_i(\beta, t)$. In practice, when we have a finite sample it is indeed possible that there will be no subject in the validation set at time t which usually happen in the latter part of the time interval being studied. Consequently $\hat{r}_i(\beta_1, t)$ $(i = 1, 2, ..., n)$ becomes impossible to calculate. In this case we could use either of the following two approaches

(a) perform estimation without using those points where the risk set in the validation set becomes empty.

(b) perform estimation after imputation of the relative risk function at those points by interpolation based on neighboring points.

Since Z is assumed to be a continuous variable, we employed the latter approach in our study to deal with the problem. For those observations, for which the risk set is empty, the relative risk functions can be estimated by the relative risk function of the subject with maximum observation time at risk in the validation set. Then the parameters β can be estimated by maximizing the following estimated partial likelihood function:

$$EPL(\beta) = \prod_{i=1}^{n} \left\{ \frac{\hat{r}_i(\beta, S_i)}{\sum_{j \in \mathcal{R}(S_i)} \hat{r}_j(\beta, S_i)} \right\}^{\delta_i},$$

where $\mathcal{R}(S_i)$ is the risk set at time S_i. We denote $\hat{\beta}_{EPL} = \arg\max_\beta EPL(\beta)$. The performance of the proposed estimator in finite sample is illustrated in the following section.

4.3 Simulation Results

Tables 4.1-4.9 provide the results for the following settings:

1. β_0: $[ln(2), 0.5]'$.

2. n: 100, 200 and 300.

3. Censoring percentage: 20%, 50% and 80%.

4. σ : 0.2, 0.8, 1.6.

5. γ : 0, 2, 4.

6. validation fraction ρ: 30%, 50% and 70%.

In this section we will discuss the results in regards to bias and asymptotic normality of $\hat{\beta}_{EPL}$. We will also observe the performance of the variance estimator proposed in our study.

4.3.1 Bias of $\hat{\beta}_{EPL}$

Examining the first column in the Tables 4.1-4.9, we find that there exists a bias in different situations which tends to zero. In all the situations $\hat{\beta}_{EPL}$ is observed to be a consistent estimator of true β_0.

We observed the effect of four different factors on the bias of the estimator $\hat{\beta}_{EPL}$ which is illustrated below.

1. n : As the sample size n increases, the bias decreases.

2. Censoring Percentage: We did not observe any significant effect of censoring percentage on the bias of $\hat{\beta}_{EPL}$.

3. σ: σ represents the strength of association between X and W. Since we include the information contained in W both from the validation and non-validation sets the effect of σ^2 on the bias of the estimator is not dramatic. When n increases, the bias of $\hat{\beta}_{EPL}$ goes to zero.

4. Validation fraction: As the validation fraction increases the bias decreases. In Tables 4.10 and 4.11, we have shown the bias of estimator for different validation fractions and $n = 300$.

4.3.2 Normality of $\hat{\beta}_{EPL}$

In chapter 3 we proved the asymptotic normality of the proposed estimator $\hat{\beta}_{EPL}$. In Figures 4.1-4.6 we draw the QQplot of the estimates for different values of σ and γ when n equals 300. We observe that the plot is close to a straight line. As n increases, the points lie closer to the line. Therefore, we can conclude that $\hat{\beta}_{EPL} = (\hat{\beta}_1, \hat{\beta}_2)'$ are approximately normally distributed.

4.3.3 Performance of estimator of standard error of $\hat{\beta}_{EPL}$

The sample standard error of $\hat{\beta}_{EPL}$ is calculated from 500 simulations and shown in the Tables 4.1-4.9 for different settings. This simulation standard error can be used as an estimate of the true standard error of the estimator. We also calculated the mean of the 500 estimates of standard error using the variance estimator of $\hat{\beta}_{EPL}$ suggested in chapter 3. By examining the corresponding columns in the tables, we observe that the estimated standard errors provide very good estimates of the true standard errors of the EPL estimator. The mean of the estimated standard error is very close to the simulated standard errors of $\hat{\beta}_{EPL}$.

We also calculated the nominal 95% confidence intervals using the following formula

$$\hat{\beta}_{EPL} \pm 1.96\widehat{s.e}(\beta_{EPL}).$$

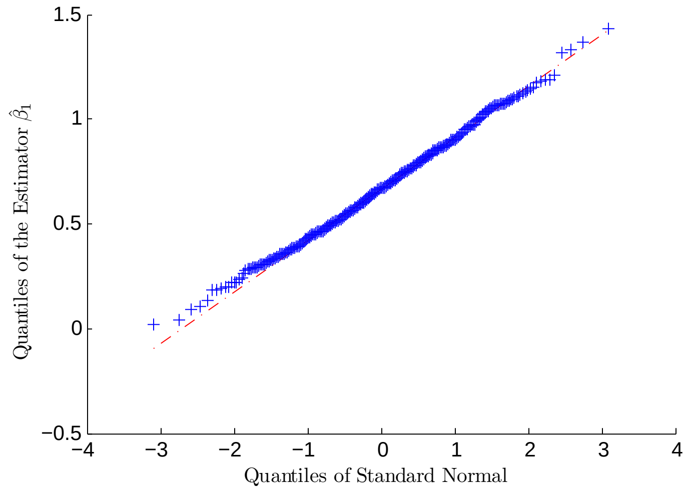
The coverage probabilities are listed in the table which ranges from .91-.96 in most of the cases. This implies that standard error estimates of $\hat{\beta}_{EPL}$ are quite reasonable.

4.3.4 Results

Table 4.1: Simulation Results with $\beta = [0.693\ 0.5]'$, $\gamma = 0$ and 50% censoring

n	σ	$\hat{\beta}_{EPL}$	*mean*	*median*	*se*	*mean*($\hat{se}$)	cp
100	0.2	$\hat{\beta}_1$	0.701	0.679	0.436	0.405	0.948
		$\hat{\beta}_2$	0.509	0.513	0.310	0.277	0.926
	0.8	$\hat{\beta}_1$	0.744	0.726	0.443	0.399	0.940
		$\hat{\beta}_2$	0.505	0.499	0.307	0.279	0.926
	1.6	$\hat{\beta}_1$	0.761	0.745	0.438	0.404	0.950
		$\hat{\beta}_2$	0.503	0.495	0.304	0.277	0.936
200	0.2	$\hat{\beta}_1$	0.727	0.700	0.304	0.283	0.936
		$\hat{\beta}_2$	0.503	0.516	0.201	0.194	0.944
	0.8	$\hat{\beta}_1$	0.761	0.720	0.312	0.287	0.930
		$\hat{\beta}_2$	0.492	0.510	0.205	0.197	0.944
	1.6	$\hat{\beta}_1$	0.763	0.727	0.311	0.290	0.922
		$\hat{\beta}_2$	0.494	0.514	0.204	0.199	0.944
300	0.2	$\hat{\beta}_1$	0.671	0.671	0.238	0.249	0.940
		$\hat{\beta}_2$	0.505	0.504	0.163	0.161	0.946
	0.8	$\hat{\beta}_1$	0.693	0.697	0.243	0.227	0.930
		$\hat{\beta}_2$	0.498	0.503	0.166	0.159	0.942
	1.6	$\hat{\beta}_1$	0.697	0.697	0.242	0.228	0.938
		$\hat{\beta}_2$	0.497	0.504	0.166	0.160	0.952

[1] $\hat{\beta}_{EPL}$ denotes the proposed estimator, *se* is the standard error of $\hat{\beta}_{EPL}$ from simulation, *mean*($\hat{se}$) denotes the mean of the estimated standard errors and *cp* denotes the 95% coverage probability.



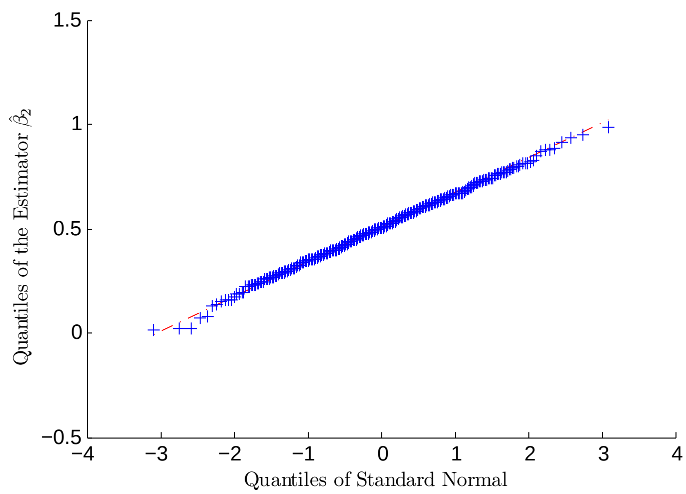


Figure 4.1: QQplot of $\hat{\beta}_{EPL} = (\hat{\beta}_1 \;\; \hat{\beta}_2)'$ for $n = 300$, $\gamma = 0$ and $\sigma = 0.2$

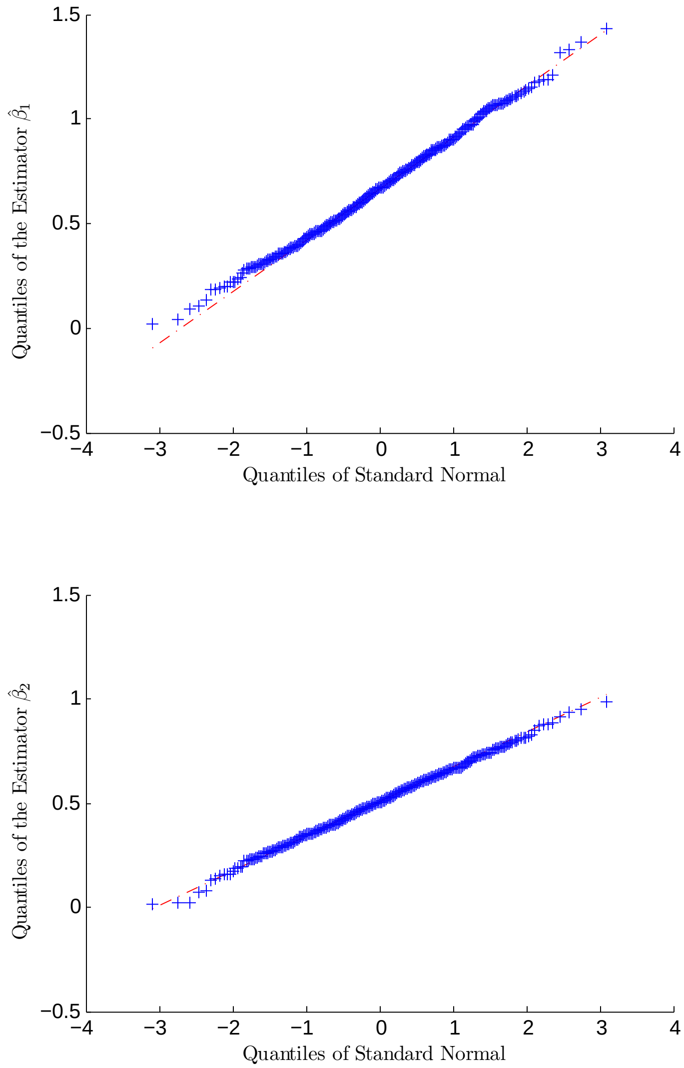
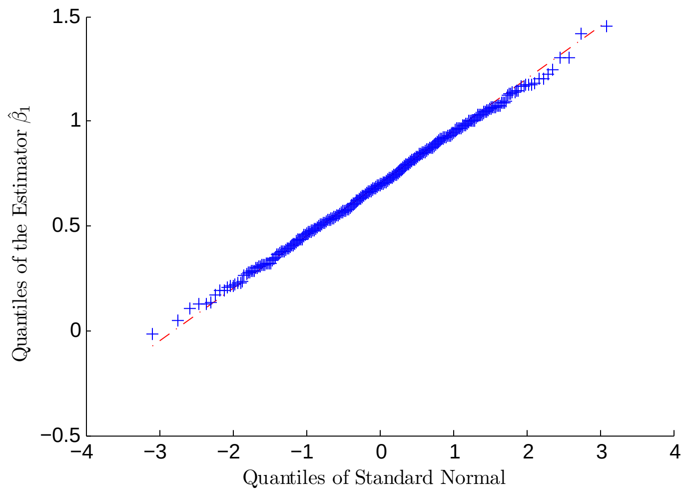


Figure 4.2: QQplot of $\hat{\beta}_{EPL} = (\hat{\beta}_1 \;\; \hat{\beta}_2)'$ for $n = 300$, $\gamma = 0$ and $\sigma = 0.8$



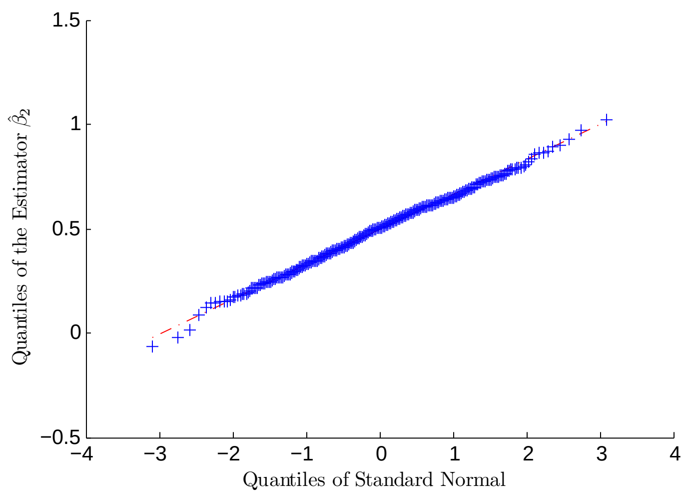


Figure 4.3: QQplot of $\hat{\beta}_{EPL} = (\hat{\beta}_1 \ \hat{\beta}_2)'$ for $n = 300$, $\gamma = 0$ and $\sigma = 1.6$

Table 4.2: Simulation Results with $\beta = [0.693\ 0.5]'$, $\gamma = 2$ and 50% censoring

n	σ	$\hat{\beta}_{EPL}$	*mean*	*median*	*se*	$mean(\hat{se})$	cp
100	0.2	$\hat{\beta}_1$	0.745	0.727	0.432	0.403	0.948
		$\hat{\beta}_2$	0.516	0.514	0.312	0.280	0.928
	0.8	$\hat{\beta}_1$	0.747	0.740	0.436	0.406	0.942
		$\hat{\beta}_2$	0.515	0.508	0.310	0.281	0.920
	1.6	$\hat{\beta}_1$	0.741	0.717	0.425	0.412	0.950
		$\hat{\beta}_2$	0.516	0.507	0.311	0.278	0.922
200	0.2	$\hat{\beta}_1$	0.747	0.713	0.296	0.287	0.944
		$\hat{\beta}_2$	0.502	0.513	0.202	0.198	0.952
	0.8	$\hat{\beta}_1$	0.748	0.719	0.291	0.287	0.948
		$\hat{\beta}_2$	0.502	0.512	0.203	0.198	0.952
	1.6	$\hat{\beta}_1$	0.758	0.740	0.296	0.289	0.942
		$\hat{\beta}_2$	0.498	0.503	0.202	0.199	0.950
300	0.2	$\hat{\beta}_1$	0.691	0.683	0.242	0.233	0.938
		$\hat{\beta}_2$	0.500	0.509	0.169	0.161	0.938
	0.8	$\hat{\beta}_1$	0.689	0.687	0.234	0.226	0.942
		$\hat{\beta}_2$	0.500	0.512	0.163	0.159	0.948
	1.6	$\hat{\beta}_1$	0.688	0.677	0.242	0.227	0.942
		$\hat{\beta}_2$	0.501	0.513	0.167	0.159	0.944

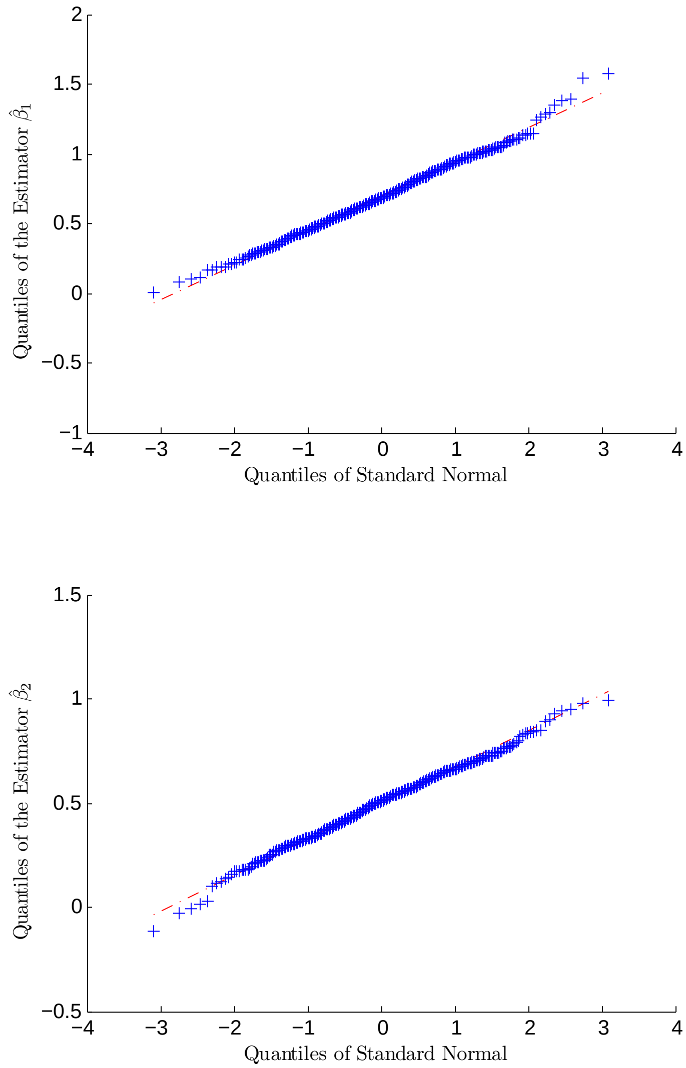


Figure 4.4: QQplot of $\hat{\beta}_{EPL} = (\hat{\beta}_1 \;\; \hat{\beta}_2)'$ for $n = 300$, $\gamma = 2$ and $\sigma = 0.2$

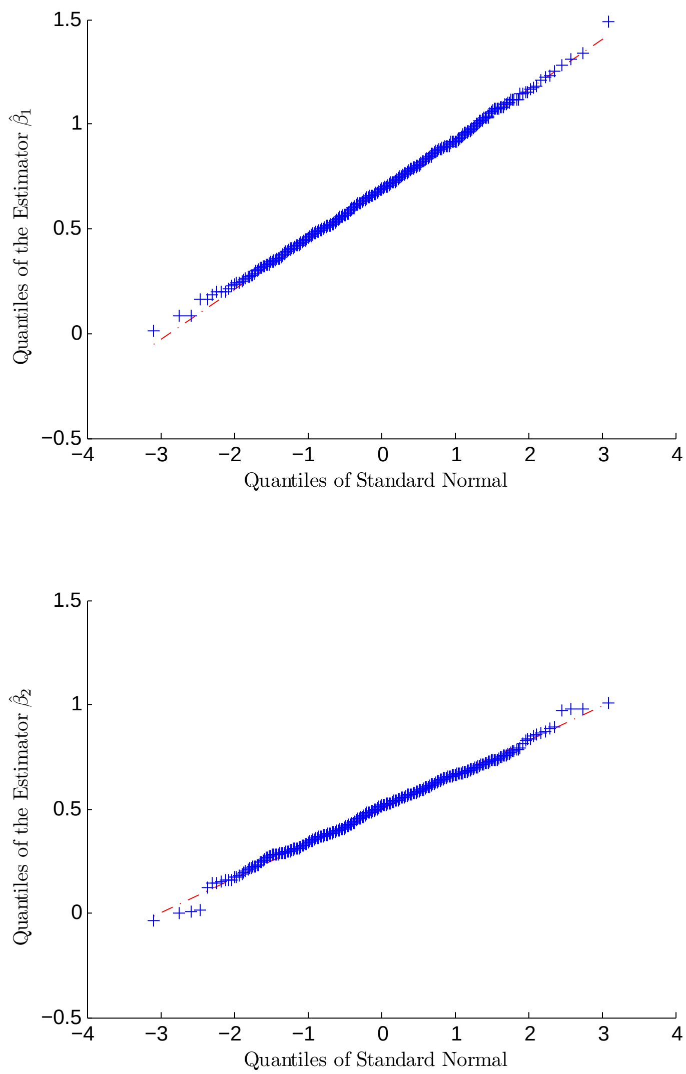
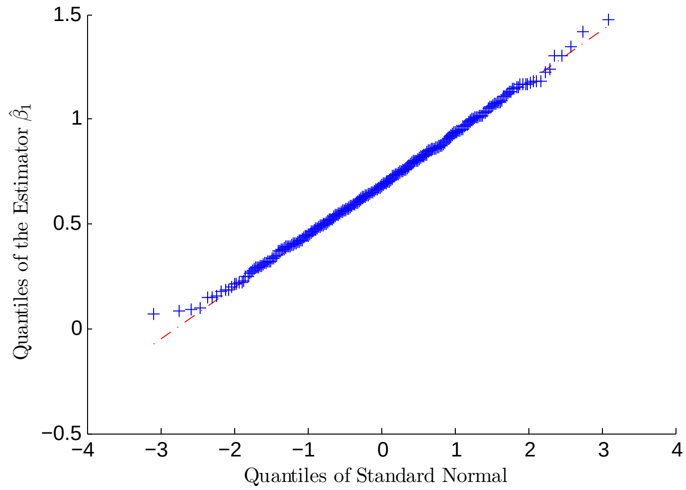


Figure 4.5: QQplot of $\hat{\beta}_{EPL} = (\hat{\beta}_1\ \ \hat{\beta}_2)'$ for $n = 300$, $\gamma = 2$ and $\sigma = 0.8$



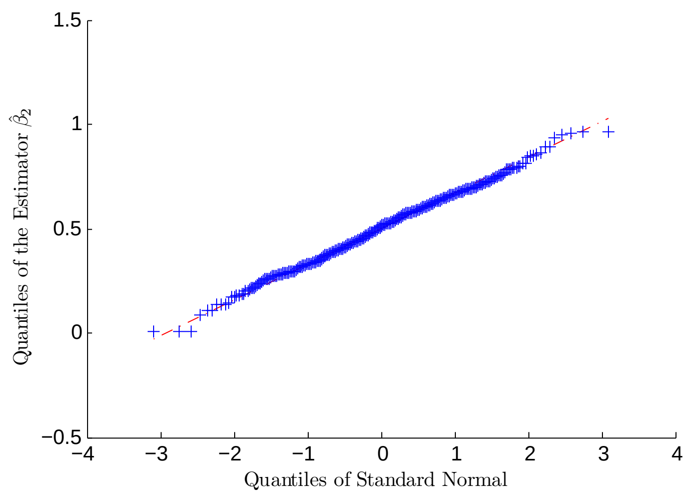
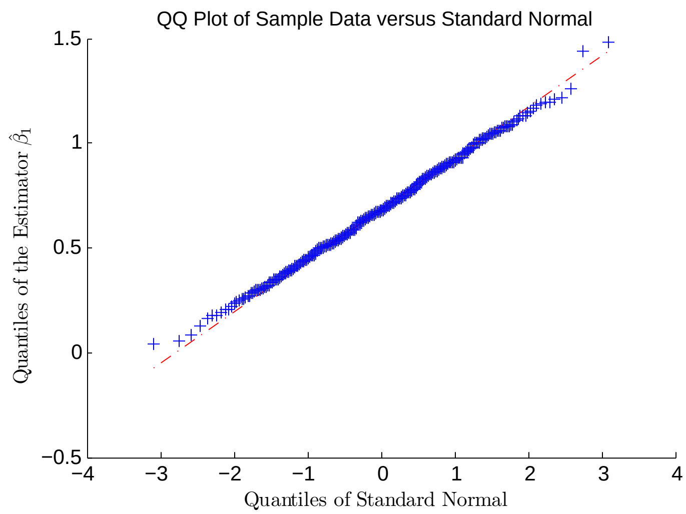


Figure 4.6: QQplot of $\hat{\beta}_{EPL} = (\hat{\beta}_1 \ \ \hat{\beta}_2)'$ for $n = 300$, $\gamma = 2$ and $\sigma = 1.6$

Table 4.3: Simulation Results with $\beta = [0.693\ \ 0.5]'$, $\gamma = 4$ and 50% censoring

n	σ	$\hat{\beta}_{EPL}$	$mean$	$median$	se	$mean(\hat{se})$	cp
100	0.2	$\hat{\beta}_1$	0.733	0.714	0.436	0.415	0.954
		$\hat{\beta}_2$	0.521	0.511	0.314	0.285	0.922
	0.8	$\hat{\beta}_1$	0.735	0.727	0.435	0.432	0.946
		$\hat{\beta}_2$	0.517	0.509	0.314	0.295	0.922
	1.6	$\hat{\beta}_1$	0.732	0.716	0.443	0.478	0.952
		$\hat{\beta}_2$	0.513	0.510	0.326	0.299	0.926
200	0.2	$\hat{\beta}_1$	0.738	0.713	0.297	0.287	0.946
		$\hat{\beta}_2$	0.504	0.515	0.205	0.198	0.944
	0.8	$\hat{\beta}_1$	0.738	0.717	0.287	0.285	0.944
		$\hat{\beta}_2$	0.504	0.509	0.202	0.197	0.944
	1.6	$\hat{\beta}_1$	0.743	0.725	0.291	0.288	0.944
		$\hat{\beta}_2$	0.502	0.518	0.201	0.198	0.944
300	0.2	$\hat{\beta}_1$	0.681	0.677	0.237	0.226	0.944
		$\hat{\beta}_2$	0.503	0.509	0.167	0.159	0.950
	0.8	$\hat{\beta}_1$	0.681	0.679	0.234	0.226	0.942
		$\hat{\beta}_2$	0.504	-0.513	0.166	0.159	0.944
	1.6	$\hat{\beta}_1$	0.685	0.684	0.236	0.226	0.946
		$\hat{\beta}_2$	0.504	0.512	0.165	0.159	0.950



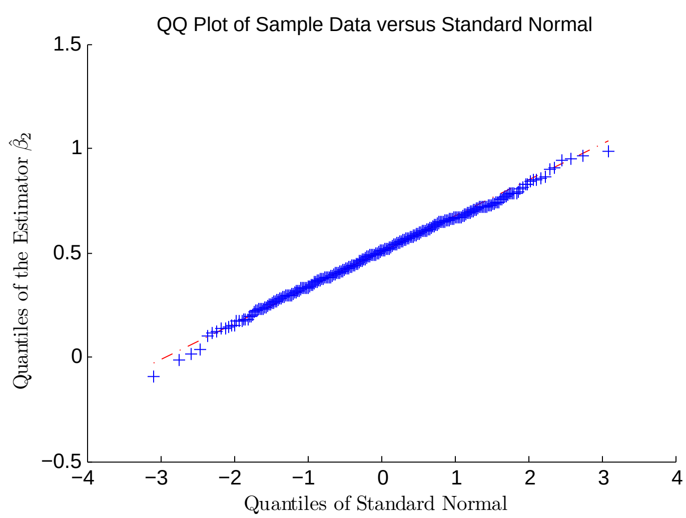


Figure 4.7: QQplot of $\hat{\beta}_{EPL} = (\hat{\beta}_1 \ \ \hat{\beta}_2)'$ for $n = 300$, $\gamma = 4$ and $\sigma = 0.2$

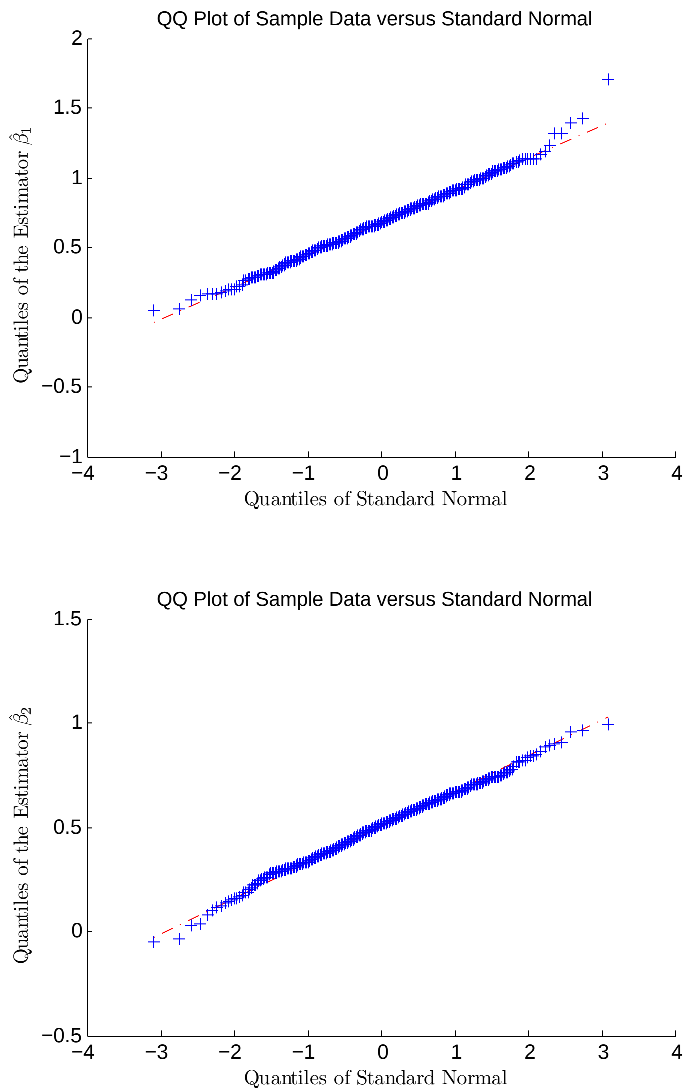


Figure 4.8: QQplot of $\hat{\beta}_{EPL} = (\hat{\beta}_1 \;\; \hat{\beta}_2)'$ for $n = 300$, $\gamma = 4$ and $\sigma = 0.8$

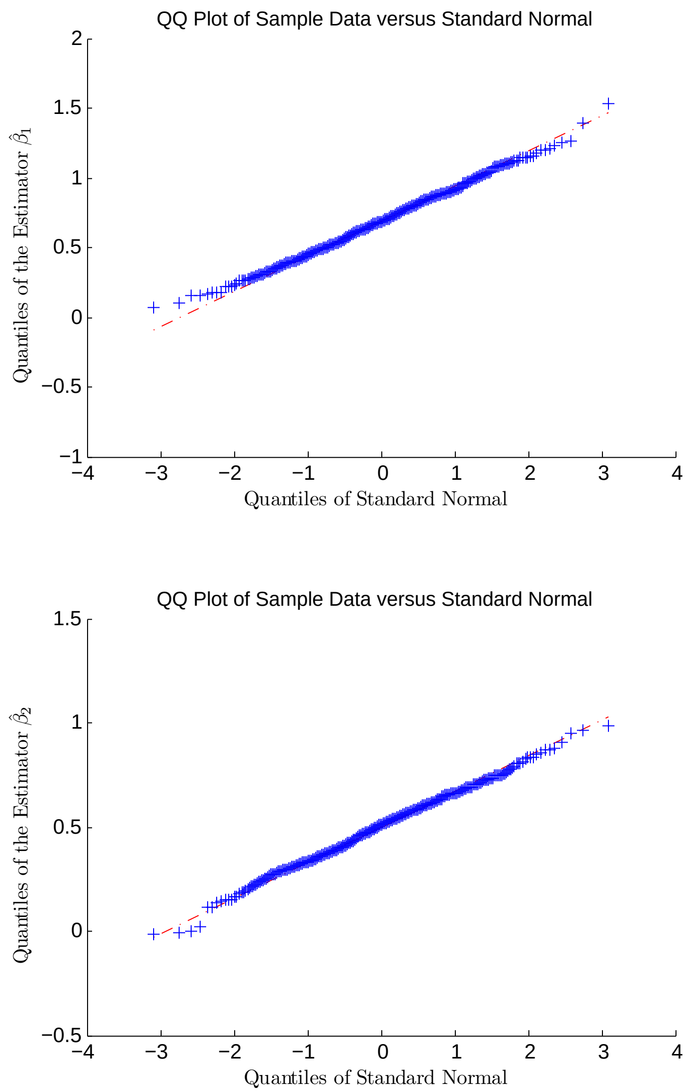


Figure 4.9: QQplot of $\hat{\beta}_{EPL} = (\hat{\beta}_1 \ \hat{\beta}_2)'$ for $n = 300$, $\gamma = 4$ and $\sigma = 1.6$

Table 4.4: Simulation Results with $\beta = [0.693\ 0.5]'$, $\gamma = 0$ and 20% censoring

n	σ	$\hat{\beta}_{EPL}$	*mean*	*median*	*se*	$mean(\hat{se})$	cp
100	0.2	$\hat{\beta}_1$	0.705	0.692	0.346	0.308	0.914
		$\hat{\beta}_2$	0.506	0.503	0.224	0.213	0.934
	0.8	$\hat{\beta}_1$	0.740	0.718	0.350	0.314	0.922
		$\hat{\beta}_2$	0.517	0.504	0.229	0.216	0.940
	1.6	$\hat{\beta}_1$	0.751	0.742	0.347	0.318	0.914
		$\hat{\beta}_2$	0.504	0.503	0.231	0.220	0.940
200	0.2	$\hat{\beta}_1$	0.722	0.702	0.250	0.220	0.920
		$\hat{\beta}_2$	0.498	0.491	0.167	0.151	0.930
	0.8	$\hat{\beta}_1$	0.754	0.736	0.260	0.229	0.904
		$\hat{\beta}_2$	0.492	0.491	0.168	0.155	0.936
	1.6	$\hat{\beta}_1$	0.761	0.746	0.255	0.225	0.928
		$\hat{\beta}_2$	0.489	0.489	0.171	0.156	0.936
300	0.2	$\hat{\beta}_1$	0.682	0.670	0.192	0.178	0.942
		$\hat{\beta}_2$	0.497	0.495	0.126	0.123	0.940
	0.8	$\hat{\beta}_1$	0.700	0.695	0.197	0.205	0.932
		$\hat{\beta}_2$	0.494	0.494	0.131	0.136	0.950
	1.6	$\hat{\beta}_1$	0.706	0.690	0.194	0.288	0.936
		$\hat{\beta}_2$	0.493	0.491	0.129	0.173	0.952

Table 4.5: Simulation Results with $\beta = [0.693 \ 0.5]'$, $\gamma = 2$ and 20% censoring

n	σ	$\hat{\beta}_{EPL}$	*mean*	*median*	*se*	*mean*($\hat{se}$)	cp
100	0.2	$\hat{\beta}_1$	0.743	0.737	0.337	0.318	0.926
		$\hat{\beta}_2$	0.513	0.506	0.233	0.222	0.944
	0.8	$\hat{\beta}_1$	0.737	0.742	0.332	0.315	0.930
		$\hat{\beta}_2$	0.515	0.513	0.231	0.219	0.952
	1.6	$\hat{\beta}_1$	0.737	0.738	0.345	0.326	0.920
		$\hat{\beta}_2$	0.513	0.507	0.231	0.217	0.932
200	0.2	$\hat{\beta}_1$	0.754	0.742	0.243	0.222	0.922
		$\hat{\beta}_2$	0.492	0.485	0.167	0.154	0.944
	0.8	$\hat{\beta}_1$	0.748	0.735	0.246	0.222	0.934
		$\hat{\beta}_2$	0.500	0.488	0.170	0.154	0.942
	1.6	$\hat{\beta}_1$	0.754	0.732	0.253	0.223	0.928
		$\hat{\beta}_2$	0.493	0.490	0.170	0.155	0.946
300	0.2	$\hat{\beta}_1$	0.701	0.689	0.190	0.179	0.940
		$\hat{\beta}_2$	0.494	0.489	0.128	0.125	0.956
	0.8	$\hat{\beta}_1$	0.706	0.691	0.192	0.180	0.944
		$\hat{\beta}_2$	0.493	0.486	0.128	0.126	0.952
	1.6	$\hat{\beta}_1$	0.700	0.683	0.193	0.184	0.938
		$\hat{\beta}_2$	0.495	0.491	0.132	0.127	0.946

Table 4.6: Simulation Results with $\beta = [0.693\ 0.5]'$, $\gamma = 4$ and 20% censoring

n	σ	$\hat{\beta}_{EPL}$	*mean*	*median*	*se*	$mean(\hat{se})$	cp
100	0.2	$\hat{\beta}_1$	0.716	0.701	0.334	0.334	0.928
		$\hat{\beta}_2$	0.523	0.510	0.232	0.231	0.942
	0.8	$\hat{\beta}_1$	0.731	0.715	0.347	0.318	0.922
		$\hat{\beta}_2$	0.522	0.515	0.235	0.224	0.944
	1.6	$\hat{\beta}_1$	0.736	0.712	0.344	0.319	0.940
		$\hat{\beta}_2$	0.520	0.517	0.233	0.222	0.940
200	0.2	$\hat{\beta}_1$	0.738	0.718	0.242	0.222	0.936
		$\hat{\beta}_2$	0.499	0.498	0.169	0.155	0.942
	0.8	$\hat{\beta}_1$	0.741	0.725	0.246	0.222	0.920
		$\hat{\beta}_2$	0.496	0.496	0.170	0.154	0.942
	1.6	$\hat{\beta}_1$	0.740	0.729	0.253	0.225	0.926
		$\hat{\beta}_2$	0.497	0.492	0.171	0.162	0.938
300	0.2	$\hat{\beta}_1$	0.690	0.681	0.187	0.178	0.940
		$\hat{\beta}_2$	0.499	0.495	0.129	0.125	0.946
	0.8	$\hat{\beta}_1$	0.692	0.687	0.190	0.179	0.934
		$\hat{\beta}_2$	0.500	0.501	0.127	0.126	0.950
	1.6	$\hat{\beta}_1$	0.690	0.676	0.189	0.182	0.942
		$\hat{\beta}_2$	0.500	0.503	0.129	0.127	0.952

Table 4.7: Simulation Results with $\beta = [0.693\ 0.5]'$, $\gamma = 0$ and 80% censoring

n	σ	$\hat{\beta}_{EPL}$	$mean$	$median$	se	$mean(\hat{se})$	cp
100	0.2	$\hat{\beta}_1$	0.717	0.699	0.756	0.784	0.946
		$\hat{\beta}_2$	0.531	0.489	0.498	0.490	0.944
	0.8	$\hat{\beta}_1$	0.743	0.710	0.803	0.774	0.952
		$\hat{\beta}_2$	0.532	0.507	0.517	0.468	0.944
	1.6	$\hat{\beta}_1$	0.765	0.752	0.752	0.713	0.960
		$\hat{\beta}_2$	0.529	0.504	0.494	0.463	0.946
200	0.2	$\hat{\beta}_1$	0.742	0.738	0.492	0.470	0.942
		$\hat{\beta}_2$	0.501	0.489	0.319	0.317	0.956
	0.8	$\hat{\beta}_1$	0.776	0.764	0.491	0.472	0.942
		$\hat{\beta}_2$	0.492	0.491	0.318	0.317	0.964
	1.6	$\hat{\beta}_1$	0.772	0.763	0.497	0.475	0.952
		$\hat{\beta}_2$	0.496	0.502	0.318	0.320	0.960
300	0.2	$\hat{\beta}_1$	0.690	0.681	0.187	0.178	0.940
		$\hat{\beta}_2$	0.499	0.495	0.129	0.125	0.946
	0.8	$\hat{\beta}_1$	0.692	0.687	0.190	0.179	0.934
		$\hat{\beta}_2$	0.500	0.501	0.127	0.126	0.950
	1.6	$\hat{\beta}_1$	0.700	0.692	0.397	0.376	0.934
		$\hat{\beta}_2$	0.509	0.511	0.259	0.259	0.960

Table 4.8: Simulation Results with $\beta = [0.693\ 0.5]'$, $\gamma = 2$ and 80% censoring

n	σ	$\hat{\beta}_{EPL}$	*mean*	*median*	*se*	$mean(\hat{se})$	cp
100	0.2	$\hat{\beta}_1$	0.752	0.690	0.740	0.723	0.962
		$\hat{\beta}_2$	0.536	0.503	0.506	0.474	0.952
	0.8	$\hat{\beta}_1$	0.762	0.735	0.763	0.819	0.958
		$\hat{\beta}_2$	0.532	0.511	0.503	0.505	0.956
	1.6	$\hat{\beta}_1$	0.772	0.731	0.751	0.742	0.956
		$\hat{\beta}_2$	0.536	0.504	0.501	0.485	0.948
200	0.2	$\hat{\beta}_1$	0.757	0.768	0.489	0.469	0.950
		$\hat{\beta}_2$	0.501	0.496	0.323	0.318	0.952
	0.8	$\hat{\beta}_1$	0.752	0.760	0.481	0.469	0.952
		$\hat{\beta}_2$	0.501	0.500	0.319	0.317	0.960
	1.6	$\hat{\beta}_1$	0.771	0.785	0.484	0.473	0.950
		$\hat{\beta}_2$	0.496	0.502	0.323	0.319	0.956
300	0.2	$\hat{\beta}_1$	0.688	0.677	0.391	0.374	0.942
		$\hat{\beta}_2$	0.513	0.518	0.259	0.257	0.960
	0.8	$\hat{\beta}_1$	0.689	0.677	0.374	0.368	0.938
		$\hat{\beta}_2$	0.512	0.512	0.260	0.257	0.960
	1.6	$\hat{\beta}_1$	0.692	0.674	0.391	0.375	0.938
		$\hat{\beta}_2$	0.513	0.517	0.262	0.258	0.956

Table 4.9: Simulation Results with $\beta = [0.693\ 0.5]'$, $\gamma = 4$ and 80% censoring

n	σ	$\hat{\beta}_{EPL}$	*mean*	*median*	*se*	*mean*($\hat{se}$)	cp
100	0.2	$\hat{\beta}_1$	0.752	0.690	0.740	0.723	0.962
		$\hat{\beta}_2$	0.536	0.503	0.506	0.474	0.952
	0.8	$\hat{\beta}_1$	0.761	0.719	0.762	1.077	0.950
		$\hat{\beta}_2$	0.533	0.481	0.503	0.732	0.946
	1.6	$\hat{\beta}_1$	0.772	0.718	0.745	0.807	0.956
		$\hat{\beta}_2$	0.530	0.505	0.493	0.509	0.946
200	0.2	$\hat{\beta}_1$	0.745	0.761	0.487	0.469	0.940
		$\hat{\beta}_2$	0.505	0.505	0.323	0.318	0.962
	0.8	$\hat{\beta}_1$	0.747	0.777	0.487	0.472	0.950
		$\hat{\beta}_2$	0.505	0.505	0.323	0.318	0.964
	1.6	$\hat{\beta}_1$	0.762	0.781	0.481	0.471	0.956
		$\hat{\beta}_2$	0.500	0.500	0.323	0.319	0.958
300	0.2	$\hat{\beta}_1$	0.685	0.670	0.392	0.374	0.934
		$\hat{\beta}_2$	0.515	0.518	0.259	0.258	0.956
	0.8	$\hat{\beta}_1$	0.682	0.674	0.391	0.440	0.938
		$\hat{\beta}_2$	0.517	0.524	0.259	0.285	0.960
	1.6	$\hat{\beta}_1$	0.686	0.675	0.395	0.374	0.936
		$\hat{\beta}_2$	0.517	0.517	0.257	0.258	0.956

In the following two tables we illustrate the effect of validation fraction on the estimator.

Table 4.10: Simulation Results with $\sigma = 0.2$

n	$Validation\,Prop$	$\hat{\beta}_{EPL}$	$mean$	$median$	se	$mean(\hat{se})$	cp
100	0.3	$\hat{\beta}_1$	0.745	0.756	0.537	0.528	0.924
		$\hat{\beta}_2$	0.514	0.500	0.314	0.308	0.928
	0.5	$\hat{\beta}_1$	0.701	0.679	0.436	0.405	0.948
		$\hat{\beta}_2$	0.509	0.513	0.310	0.277	0.926
	0.7	$\hat{\beta}_1$	0.700	0.674	0.370	0.338	0.918
		$\hat{\beta}_2$	0.507	0.497	0.292	0.261	0.912
200	0.3	$\hat{\beta}_1$	0.731	0.704	0.371	0.352	0.936
		$\hat{\beta}_2$	0.503	0.511	0.215	0.212	0.944
	0.5	$\hat{\beta}_1$	0.727	0.700	0.304	0.283	0.936
		$\hat{\beta}_2$	0.503	0.516	0.201	0.194	0.944
	0.7	$\hat{\beta}_1$	0.713	0.717	0.224	0.269	0.948
		$\hat{\beta}_2$	0.509	0.516	0.185	0.204	0.958
300	0.3	$\hat{\beta}_1$	0.709	0.703	0.320	0.279	0.920
		$\hat{\beta}_2$	0.495	0.495	0.180	0.170	0.918
	0.5	$\hat{\beta}_1$	0.670	0.671	0.238	0.249	0.940
		$\hat{\beta}_2$	0.505	0.504	0.163	0.161	0.946
	0.7	$\hat{\beta}_1$	0.670	0.661	0.191	0.193	0.946
		$\hat{\beta}_2$	0.505	0.513	0.152	0.151	0.956

Table 4.11: Simulation Results with $\sigma = 0.8$

n	$Validation\,Prop$	$\hat{\beta}_{EPL}$	$mean$	$median$	se	$mean(\hat{se})$	cp
100	0.3	$\hat{\beta}_1$	0.805	0.788	0.540	0.633	0.938
		$\hat{\beta}_2$	0.511	0.496	0.330	0.357	0.936
	0.5	$\hat{\beta}_1$	0.744	0.726	0.443	0.399	0.940
		$\hat{\beta}_2$	0.505	0.499	0.307	0.279	0.926
	0.7	$\hat{\beta}_1$	0.720	0.701	0.376	0.340	0.920
		$\hat{\beta}_2$	0.507	0.502	0.293	0.261	0.916
200	0.3	$\hat{\beta}_1$	0.782	0.762	0.385	0.372	0.950
		$\hat{\beta}_2$	0.496	0.503	0.221	0.229	0.944
	0.5	$\hat{\beta}_1$	0.761	0.720	0.312	0.287	0.930
		$\hat{\beta}_2$	0.492	0.510	0.205	0.197	0.944
	0.7	$\hat{\beta}_1$	0.723	0.725	0.226	0.228	0.948
		$\hat{\beta}_2$	0.506	0.514	0.186	0.188	0.960
300	0.3	$\hat{\beta}_1$	0.746	0.727	0.335	0.287	0.904
		$\hat{\beta}_2$	0.485	0.493	0.187	0.178	0.932
	0.5	$\hat{\beta}_1$	0.693	0.697	0.243	0.227	0.930
		$\hat{\beta}_2$	0.498	0.503	0.166	0.159	0.942
	0.7	$\hat{\beta}_1$	0.676	0.668	0.193	0.193	0.952
		$\hat{\beta}_2$	0.503	0.513	0.152	0.152	0.950

4.4 Comparison of different methods

In this section, we have explored five different methods to estimate the unknown parameter in our simulation study.

1. $\hat{\beta}_F$-The full-data Cox regression estimator which uses the the full data without any missing covariate.

2. $\hat{\beta}_{CC}$-The complete-case Cox regression estimator which uses the data available only on the validation set.

3. $\hat{\beta}_N$-The Cox regression estimator with W substitued for the missing X for the subjects in the non-validation set.

4. $\hat{\beta}_{ZW}$-The estimated partial likelihood estimator suggested by Zhou and Wang (2000), and

5. $\hat{\beta}_{EPL}$-The newly proposed estimator.

4.4.1 Performance of different methods

Tables 4.11-4.21 provide the results for the following settings for different methods discussed above.

1. $\beta_0 = (\beta_{01}, \beta_{02})'$: $[ln(2), 0.5]'$.

2. n: 100 and 300.

3. Censoring percentage: 20% and 50%.

4. σ : 0.2, 0.8, 1.6.

5. γ : 0,2, 4.

6. Validation fraction ρ: 50%.

4.4.2 Results

In Tables 4.12-4.29 we present our simulation results obtained using the estimation procedures described above.

For a given sample size, mean($\hat{\beta}_{EPL}) - \beta_0$)(bias), median($\hat{\beta}_{EPL}$)-$\beta_0$,(robust bias), standard errors, mean of the estimated standard errors and 95% confidence intervals for the estimators are obtained using 500 independent runs. In this section we discuss the results with regards to the bias and variance of the estimators. We also discuss the performance of the estimated variances obtained using these methods.

Bias: From these results we observe that, the full-data Cox Regression estimator ($\hat{\beta}_F$) the complete-case Cox Regression estimator($\hat{\beta}_{CC}$) and our proposed estimator($\hat{\beta}_{EPL}$) have acceptable small bias for all values of σ and γ considered. The estimated partial likelihood method (($\hat{\beta}_{ZW}$)) proposed by Zhou and Wang (2000) works for $\gamma = 0$ but biased for $\gamma \neq 0$. The naive estimator β_N is biased.

Variance: The standard errors of all the estimators are obtained. We note that, ($\hat{\beta}_F$) is the most efficient estimator. The proposed estimator ($\hat{\beta}_{EPL}$) is more efficient than ($\hat{\beta}_{cc}$) in all the situations considered. We also compared our estimator with Zhou and Wang's estimator ($\hat{\beta}_{ZW}$). When $\gamma = 0$, Zhou's method is more efficient for very small value of σ but for higher values of σ both methods are almost equally efficient and their estimator has more bias than our estimator. For situations where $\gamma \neq 0$, the consistency property of their estimator does not hold. So, we cannot compare the efficiency of both the estimators in that case.

Estimated Standard Error: We observe that the estimator of the standard error performs very well for all the estimators when $\gamma = 0$. When $\gamma \neq 0$, the standard error estimates of both $\hat{\beta}_N$ and $\hat{\beta}_{ZW}$ are biased whereas the standard error estimates of $\hat{\beta}_{EPL}$ remains to be consistent for all values of γ and σ. The 95% coverage probabilities shown in the Tables also demonstrate this fact.

Table 4.12: $n = 100, \ \ \sigma = 0.2$ and $\ \gamma = 0$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.018	0.013	0.323	0.292	0.916
		$\hat{\beta}_2$	0.004	-0.015	0.281	0.258	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.930
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	0.068	0.076	0.158	0.158	0.926
		$\hat{\beta}_2$	0.024	0.028	0.146	0.145	0.954
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	0.045	0.033	0.346	0.329	0.940
		$\hat{\beta}_2$	0.013	-0.001	0.283	0.257	0.914
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.008	-0.014	0.436	0.405	0.948
		$\hat{\beta}_2$	0.009	0.013	0.310	0.277	0.926
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.021	0.016	0.248	0.232	0.936
		$\hat{\beta}_2$	0.001	-0.008	0.211	0.205	0.938
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.020	0.014	0.339	0.340	0.956
		$\hat{\beta}_2$	0.014	-0.001	0.305	0.302	0.956
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.031	-0.029	0.234	0.223	0.934
		$\hat{\beta}_2$	0.018	0.011	0.210	0.204	0.938
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	0.044	0.038	0.272	0.263	0.966
		$\hat{\beta}_2$	0.013	0.005	0.212	0.204	0.952
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.012	-0.001	0.346	0.308	0.914
		$\hat{\beta}_2$	0.006	0.003	0.224	0.213	0.934

[1] $\hat{\beta}_{EPL}$ denotes the full data Cox regression estimator, $\hat{\beta}_{CC}$ denotes the complete case Cox regression estimator, $\hat{\beta}_N$ denotes the naive Cox regression estimator replacing missing X by W, $\hat{\beta}_{ZW}$ denotes the partial likelihood estimator proposed by Zhou and Wang(2000) and $\hat{\beta}_{EPL}$ denotes the proposed estimator; C represents percentage of censoring, se is the standard error of $\hat{\beta}_{EPL}$ from simulation, $mean(\hat{se})$ denotes the mean of the estimated standard errors and cp denotes the 95% coverage probability.

Table 4.13: $n = 100$, $\sigma = 0.8$ and $\gamma = 0$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.018	0.013	0.323	0.292	0.916
		$\hat{\beta}_2$	0.004	-0.015	0.281	0.258	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.930
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.369	-0.381	0.206	0.196	0.504
		$\hat{\beta}_2$	0.139	0.028	0.260	0.249	0.914
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.045	-0.052	0.374	0.356	0.940
		$\hat{\beta}_2$	0.055	0.051	0.287	0.262	0.932
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.051	0.033	0.443	0.399	0.940
		$\hat{\beta}_2$	0.005	-0.001	0.307	0.279	0.926
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.021	0.016	0.248	0.232	0.936
		$\hat{\beta}_2$	0.001	-0.008	0.211	0.205	0.938
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.020	0.014	0.339	0.340	0.956
		$\hat{\beta}_2$	0.014	-0.001	0.305	0.302	0.956
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.368	-0.372	0.165	0.156	0.352
		$\hat{\beta}_2$	0.126	0.122	0.202	0.198	0.918
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.046	-0.052	0.272	0.263	0.966
		$\hat{\beta}_2$	0.052	0.053	0.213	0.207	0.942
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.046	0.025	0.350	0.314	0.922
		$\hat{\beta}_2$	0.007	0.004	0.229	0.217	0.940

Table 4.14: $n = 100$, $\sigma = 1.6$ and $\gamma = 0$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.018	0.013	0.323	0.292	0.916
		$\hat{\beta}_2$	0.004	-0.015	0.281	0.258	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.930
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.566	-0.572	0.130	0.112	0.026
		$\hat{\beta}_2$	0.208	0.191	0.252	0.244	0.878
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.102	-0.098	0.391	0.364	0.928
		$\hat{\beta}_2$	0.077	0.074	0.291	0.263	0.920
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.004	0.004	0.242	0.228	0.938
		$\hat{\beta}_2$	-0.003	0.004	0.166	0.160	0.952
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.021	0.016	0.248	0.232	0.936
		$\hat{\beta}_2$	0.001	-0.008	0.211	0.205	0.938
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.020	0.014	0.339	0.340	0.956
		$\hat{\beta}_2$	0.014	-0.001	0.305	0.302	0.956
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.567	-0.574	0.105	0.096	0.002
		$\hat{\beta}_2$	0.188	0.182	0.196	0.195	0.864
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.046	-0.052	0.272	0.263	0.966
		$\hat{\beta}_2$	0.052	0.053	0.213	0.207	0.942
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.058	0.049	0.347	0.318	0.914
		$\hat{\beta}_2$	0.004	0.003	0.231	0.220	0.940

Table 4.15: $n = 300,\ \ \sigma = 0.2$ and $\ \gamma = 0$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.019	-0.028	0.161	0.164	0.944
		$\hat{\beta}_2$	0.006	0.007	0.146	0.146	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.942
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	0.068	0.076	0.158	0.158	0.926
		$\hat{\beta}_2$	0.024	0.028	0.146	0.145	0.954
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.007	-0.012	0.176	0.177	0.950
		$\hat{\beta}_2$	0.010	0.018	0.150	0.147	0.944
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.022	0.022	0.238	0.249	0.940
		$\hat{\beta}_2$	0.005	0.004	0.163	0.161	0.946
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.007	-0.019	0.131	0.131	0.948
		$\hat{\beta}_2$	-0.001	-0.002	0.116	0.116	0.952
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	-0.009	-0.016	0.190	0.187	0.960
		$\hat{\beta}_2$	0.008	0.002	0.164	0.166	0.954
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.056	-0.064	0.127	0.126	0.928
		$\hat{\beta}_2$	0.016	0.011	0.116	0.115	0.952
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.006	-0.006	0.141	0.142	0.966
		$\hat{\beta}_2$	0.004	0.003	0.119	0.117	0.952
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	-0.011	-0.023	0.192	0.178	0.942
		$\hat{\beta}_2$	-0.003	-0.005	0.126	0.123	0.940

Table 4.16: $n = 300$, $\sigma = 0.8$ and $\gamma = 0$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.019	-0.028	0.161	0.164	0.944
		$\hat{\beta}_2$	0.006	0.007	0.146	0.146	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.942
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.392	-0.392	0.114	0.108	0.068
		$\hat{\beta}_2$	0.139	0.139	0.142	0.140	0.830
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.056	-0.055	0.198	0.223	0.932
		$\hat{\beta}_2$	0.033	0.044	0.156	0.157	0.946
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.000	0-.004	0.243	0.227	0.930
		$\hat{\beta}_2$	-0.002	0.003	0.166	0.159	0.942
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.007	-0.019	0.131	0.131	0.948
		$\hat{\beta}_2$	-0.001	-0.002	0.116	0.116	0.952
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	-0.009	-0.016	0.190	0.187	0.960
		$\hat{\beta}_2$	0.008	0.002	0.164	0.166	0.954
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.390	-0.393	0.092	0.086	0.018
		$\hat{\beta}_2$	0.123	0.123	0.115	0.112	0.792
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.048	-0.058	0.159	0.172	0.954
		$\hat{\beta}_2$	0.026	0.028	0.123	0.123	0.944
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.007	0.004	0.243	0.227	0.930
		$\hat{\beta}_2$	-0.007	0.031	0.166	0.159	0.942

Table 4.17: $n = 300$, $\sigma = 1.6$ and $\gamma = 0$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.019	-0.028	0.161	0.164	0.944
		$\hat{\beta}_2$	0.006	0.007	0.146	0.146	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.942
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.579	-0.580	0.071	0.067	0.0
		$\hat{\beta}_2$	0.207	0.207	0.139	0.138	0.684
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.096	-0.045	0.209	0.229	0.908
		$\hat{\beta}_2$	0.097	0.053	0.157	0.158	0.936
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.000	0-.004	0.243	0.227	0.930
		$\hat{\beta}_2$	-0.002	0.003	0.166	0.159	0.942
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.007	-0.019	0.131	0.131	0.948
		$\hat{\beta}_2$	-0.001	-0.002	0.116	0.116	0.952
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	-0.009	-0.016	0.190	0.187	0.960
		$\hat{\beta}_2$	0.008	0.002	0.164	0.166	0.954
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.580	-0.582	0.057	0.053	0.0
		$\hat{\beta}_2$	0.184	0.187	0.112	0.110	0.614
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.084	-0.089	0.170	0.192	0.904
		$\hat{\beta}_2$	0.034	0.042	0.192	0.125	0.942
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.007	0.004	0.243	0.227	0.930
		$\hat{\beta}_2$	-0.007	0.031	0.166	0.159	0.942

Table 4.18: $n = 100$, $\sigma = 0.2$ and $\gamma = 2$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.018	0.013	0.323	0.292	0.916
		$\hat{\beta}_2$	0.004	-0.015	0.281	0.258	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.930
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.988	-0.983	0.084	0.068	0.0
		$\hat{\beta}_2$	0.169	0.160	0.237	0.244	0.920
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.434	-0.484	0.574	0.373	0.670
		$\hat{\beta}_2$	0.174	0.172	0.324	0.274	0.892
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.052	0.034	0.432	0.403	0.948
		$\hat{\beta}_2$	0.016	0.014	0.312	0.280	0.928
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.021	0.016	0.248	0.232	0.936
		$\hat{\beta}_2$	0.001	-0.008	0.211	0.205	0.938
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.020	0.014	0.339	0.340	0.956
		$\hat{\beta}_2$	0.014	-0.001	0.305	0.302	0.956
	$\hat{\beta}_N$	$\hat{\beta}_1$	-1.003	-1.000	0.071	0.062	0.0
		$\hat{\beta}_2$	0.155	0.151	0.180	0.195	0.924
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.455	-0.466	0.467	0.292	0.550
		$\hat{\beta}_2$	0.163	0.159	0.241	0.214	0.862
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.050	0.044	0.337	0.318	0.926
		$\hat{\beta}_2$	0.013	0.006	0.233	0.222	0.944

Table 4.19: $n = 100, \quad \sigma = 0.8$ and $\quad \gamma = 2$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.018	0.013	0.323	0.292	0.916
		$\hat{\beta}_2$	0.004	-0.015	0.281	0.258	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.930
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.961	-0.955	0.076	0.064	0.0
		$\hat{\beta}_2$	0.177	0.164	0.243	0.244	0.904
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.325	-0.368	0.569	0.374	0.721
		$\hat{\beta}_2$	0.141	0.140	0.325	0.271	0.888
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.054	0.046	0.435	0.406	0.942
		$\hat{\beta}_2$	0.015	0.008	0.310	0.281	0.920
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.021	0.016	0.248	0.232	0.936
		$\hat{\beta}_2$	0.001	-0.008	0.211	0.205	0.938
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.020	0.014	0.339	0.340	0.956
		$\hat{\beta}_2$	0.014	-0.001	0.305	0.302	0.956
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.969	-0.966	0.064	0.057	0.0
		$\hat{\beta}_2$	0.163	0.168	0.186	0.195	0.910
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.328	-0.354	0.450	0.291	0.658
		$\hat{\beta}_2$	0.128	0.124	0.238	0.212	0.878
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.044	0.049	0.332	0.315	0.930
		$\hat{\beta}_2$	0.015	0.013	0.231	0.219	0.952

Table 4.20: $n = 100$, $\sigma = 1.6$ and $\gamma = 2$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.018	0.013	0.323	0.292	0.916
		$\hat{\beta}_2$	0.004	-0.015	0.281	0.258	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.930
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.906	-0.905	0.064	0.055	0.0
		$\hat{\beta}_2$	0.194	0.183	0.253	0.254	0.874
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.214	-0.243	0.518	0.373	0.794
		$\hat{\beta}_2$	0.108	0.109	0.312	0.269	0.904
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.048	0.024	0.425	0.412	0.950
		$\hat{\beta}_2$	0.016	0.007	0.311	0.278	0.922
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.021	0.016	0.248	0.232	0.936
		$\hat{\beta}_2$	0.001	-0.008	0.211	0.205	0.938
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.020	0.014	0.339	0.340	0.956
		$\hat{\beta}_2$	0.014	-0.001	0.305	0.302	0.956
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.902	-0.899	0.054	0.049	0.0
		$\hat{\beta}_2$	0.180	0.185	0.194	0.195	0.876
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.243	-0.229	0.409	0.292	0.766
		$\hat{\beta}_2$	0.097	0.089	0.229	0.211	0.900
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.044	0.045	0.345	0.326	0.920
		$\hat{\beta}_2$	0.013	0.007	0.231	0.217	0.932

Table 4.21: $n = 300$, $\sigma = 0.2$ and $\gamma = 2$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.019	-0.028	0.161	0.164	0.944
		$\hat{\beta}_2$	0.006	0.007	0.146	0.146	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.942
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.994	-0.991	0.048	0.039	0.0
		$\hat{\beta}_2$	0.166	0.169	0.127	0.137	0.796
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.630	-0.635	0.246	0.170	0.108
		$\hat{\beta}_2$	0.202	0.199	0.137	0.125	0..630
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	-0.003	0-.010	0.242	0.233	0.938
		$\hat{\beta}_2$	-0.001	0.009	0.169	0.161	0.938
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.007	-0.019	0.131	0.131	0.948
		$\hat{\beta}_2$	-0.001	-0.002	0.116	0.116	0.952
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	-0.009	-0.016	0.190	0.187	0.960
		$\hat{\beta}_2$	0.008	0.002	0.164	0.166	0.954
	$\hat{\beta}_N$	$\hat{\beta}_1$	-1.001	0.152	0.044	0.034	0
		$\hat{\beta}_2$	-1.007	0.154	0.100	0.110	0.748
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.617	-0.613	0.304	0.219	0.244
		$\hat{\beta}_2$	0.223	0.224	0.178	0.161	0.70
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.007	-0.004	0.190	0.17	0.940
		$\hat{\beta}_2$	-0.006	-0.011	0.128	0.125	0.956

Table 4.22: $n = 300$, $\sigma = 0.8$ and $\gamma = 2$

C			$mean - \beta_0$	$median - \beta_0$	sse	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.019	-0.028	0.161	0.164	0.944
		$\hat{\beta}_2$	0.006	0.007	0.146	0.146	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.942
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.965	-0.963	0.044	0.129	0.0
		$\hat{\beta}_2$	0.175	0.176	0.036	0.137	0.770
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.399	-0.395	0.325	0.213	0.520
		$\hat{\beta}_2$	0.147	0.145	0.180	0.157	0.808
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.004	0-.006	0.234	0.226	0.942
		$\hat{\beta}_2$	0.000	0.012	0.163	0.159	0.948
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.007	-0.019	0.131	0.131	0.948
		$\hat{\beta}_2$	-0.001	-0.002	0.116	0.116	0.952
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	-0.009	-0.016	0.190	0.187	0.960
		$\hat{\beta}_2$	0.008	0.002	0.164	0.166	0.954
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.972	-0.972	0.039	0.033	0.0
		$\hat{\beta}_2$	0.161	0.164	0.102	0.110	0.716
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.388	-0.399	0.262	0.168	0.412
		$\hat{\beta}_2$	0.127	0.128	0.138	0.122	0.784
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.012	-0.002	0.192	0.180	0.944
		$\hat{\beta}_2$	-0.007	-0.014	0.128	0.126	0.952

Table 4.23: $n = 300, \quad \sigma = 1.6$ and $\quad \gamma = 2$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.019	-0.028	0.161	0.164	0.944
		$\hat{\beta}_2$	0.006	0.007	0.146	0.146	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.942
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.906	-0.908	0.037	0.031	0.0
		$\hat{\beta}_2$	0.194	0.197	0.134	0.137	0.728
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.225	-0.222	0.312	0.232	0.716
		$\hat{\beta}_2$	0.089	0.091	0.176	0.160	0.894
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	-0.005	-0.017	0.242	0.227	0.942
		$\hat{\beta}_2$	0.001	0.013	0.167	0.159	0.944
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.007	-0.019	0.131	0.131	0.948
		$\hat{\beta}_2$	-0.001	-0.002	0.116	0.116	0.952
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	-0.009	-0.016	0.190	0.187	0.960
		$\hat{\beta}_2$	0.008	0.002	0.164	0.166	0.954
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.902	-0.902	0.032	0.105	0.0
		$\hat{\beta}_2$	0.178	0.180	0.105	0.110	0.620
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.208	-0.208	0.248	0.175	0.670
		$\hat{\beta}_2$	0.074	0.077	0.135	0.123	0.902
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.007	-0.010	0.193	0.184	0.938
		$\hat{\beta}_2$	-0.005	-0.009	0.132	0.127	0.946

Table 4.24: $n = 100, \ \ \sigma = 0.2$ and $\ \gamma = 4$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.018	0.013	0.323	0.292	0.916
		$\hat{\beta}_2$	0.004	-0.015	0.281	0.258	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.930
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.852	-0.849	0.042	0.034	0.0
		$\hat{\beta}_2$	0.104	0.090	0.239	0.245	0.952
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.126	-0.127	0.595	0.375	0.845
		$\hat{\beta}_2$	0.081	0.085	0.335	0.273	0.938
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.040	0.021	0.436	0.415	0.954
		$\hat{\beta}_2$	0.021	0.011	0.314	0.285	0.922
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.021	0.016	0.248	0.232	0.936
		$\hat{\beta}_2$	0.001	-0.008	0.211	0.205	0.938
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.020	0.014	0.339	0.340	0.956
		$\hat{\beta}_2$	0.014	-0.001	0.305	0.302	0.956
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.865	-0.864	0.037	0.032	0.0
		$\hat{\beta}_2$	0.092	0.088	0.180	0.195	0.960
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.134	-0.138	0.494	0.290	0.736
		$\hat{\beta}_2$	0.075	0.083	0.250	0.215	0.908
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.023	0.008	0.334	0.334	0.928
		$\hat{\beta}_2$	0.023	0.010	0.232	0.231	0.942

Table 4.25: $n = 100$, $\sigma = 0.8$ and $\gamma = 4$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.018	0.013	0.323	0.292	0.916
		$\hat{\beta}_2$	0.004	-0.015	0.281	0.258	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.930
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.845	-0.846	0.041	0.034	0.0
		$\hat{\beta}_2$	0.108	0.093	0.241	0.245	0.946
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.102	-0.127	0.606	0.401	0.894
		$\hat{\beta}_2$	0.118	0.108	0.336	0.271	0.888
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.042	0.034	0.435	0.432	0.946
		$\hat{\beta}_2$	0.074	0.085	0.336	0.275	0.794
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.021	0.016	0.248	0.232	0.936
		$\hat{\beta}_2$	0.001	-0.008	0.211	0.205	0.938
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.020	0.014	0.339	0.340	0.956
		$\hat{\beta}_2$	0.014	-0.001	0.305	0.302	0.956
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.859	-0.857	0.035	0.031	0.0
		$\hat{\beta}_2$	0.096	0.097	0.182	0.195	0.952
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.121	-0.121	0.492	0.300	0.726
		$\hat{\beta}_2$	0.072	0.072	0.250	0.215	0.908
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.034	0.022	0.347	0.318	0.922
		$\hat{\beta}_2$	0.022	0.015	0.235	0.224	0.944

Table 4.26: $n = 100$, $\sigma = 1.6$ and $\gamma = 4$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.018	0.013	0.323	0.292	0.916
		$\hat{\beta}_2$	0.004	-0.015	0.281	0.258	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.930
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.837	-0.834	0.038	0.248	0.0
		$\hat{\beta}_2$	0.117	0.102	0.032	0.245	0.904
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.081	-0.141	0.575	0.390	0.812
		$\hat{\beta}_2$	0.069	0.080	0.327	0.275	0.908
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.039	0.023	0.443	0.478	0.952
		$\hat{\beta}_2$	0.013	0.010	0.326	0.299	0.926
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	0.021	0.016	0.248	0.232	0.936
		$\hat{\beta}_2$	0.001	-0.008	0.211	0.205	0.938
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.020	0.014	0.339	0.340	0.956
		$\hat{\beta}_2$	0.014	-0.001	0.305	0.302	0.956
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.837	-0.834	0.038	0.032	0.0
		$\hat{\beta}_2$	0.117	0.102	0.248	0.243	0.938
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.095	-0.119	0.464	0.292	0.760
		$\hat{\beta}_2$	0.066	0.069	0.246	0.214	0.904
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.043	0.019	0.344	0.319	0.940
		$\hat{\beta}_2$	0.020	0.017	0.233	0.222	0.940

Table 4.27: $n = 300$, $\sigma = 0.2$ and $\gamma = 4$

C			$mean - \beta_0$	$median - \beta_0$	sse	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.019	-0.028	0.161	0.164	0.944
		$\hat{\beta}_2$	0.006	0.007	0.146	0.146	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.942
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.856	-0.854	0.024	0.020	0.0
		$\hat{\beta}_2$	0.097	0.099	0.129	0.138	0.910
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.293	-0.292	0.331	0.213	0.608
		$\hat{\beta}_2$	0.111	0.111	0.181	0.158	0.878
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	-0.013	0-.016	0.237	0.226	0.942
		$\hat{\beta}_2$	0.003	0.009	0.167	0.159	0.944
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.007	-0.019	0.131	0.131	0.948
		$\hat{\beta}_2$	-0.001	-0.002	0.116	0.116	0.952
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	-0.009	-0.016	0.190	0.187	0.960
		$\hat{\beta}_2$	0.008	0.002	0.164	0.166	0.954
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.868	-0.867	0.023	0.018	0.0
		$\hat{\beta}_2$	0.087	0.090	0.100	0.110	0.904
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.310	-0.301	0.284	0.167	0.508
		$\hat{\beta}_2$	0.103	0.096	0.143	0.123	0.814
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	-0.003	-0.011	0.187	0.178	0.940
		$\hat{\beta}_2$	-0.002	-0.005	0.129	0.125	0.946

Table 4.28: $n = 300$, $\sigma = 0.8$ and $\gamma = 4$

C			$mean - \beta_0$	$median - \beta_0$	sse	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.019	-0.028	0.161	0.164	0.944
		$\hat{\beta}_2$	0.006	0.007	0.146	0.146	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.942
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.851	-0.850	0.024	0.019	0.0
		$\hat{\beta}_2$	0.100	0.099	0.128	0.138	0.908
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.218	-0.191	0.335	0.221	0.682
		$\hat{\beta}_2$	0.086	0.081	0.183	0.159	0.892
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	-0.012	0-.014	0.234	0.226	0.942
		$\hat{\beta}_2$	0.004	0.013	0.166	0.160	0.944
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.007	-0.019	0.131	0.131	0.948
		$\hat{\beta}_2$	-0.001	-0.002	0.116	0.116	0.952
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	-0.009	-0.016	0.190	0.187	0.960
		$\hat{\beta}_2$	0.008	0.002	0.164	0.166	0.954
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.868	-0.867	0.023	0.018	0.0
		$\hat{\beta}_2$	0.087	0.090	0.100	0.110	0.904
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.221	-0.217	0.284	0.169	0.614
		$\hat{\beta}_2$	0.076	0.074	0.143	0.123	0.874
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	-0.001	-0.006	0.190	0.179	0.934
		$\hat{\beta}_2$	-0.001	0.001	0.127	0.125	0.950

Table 4.29: $n = 300$, $\sigma = 1.6$ and $\gamma = 4$

C			$mean - \beta_0$	$median - \beta_0$	se	$mean(\hat{se})$	cp
50%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.019	-0.028	0.161	0.164	0.944
		$\hat{\beta}_2$	0.006	0.007	0.146	0.146	0.956
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	0.026	0.024	0.234	0.233	0.942
		$\hat{\beta}_2$	0.017	0.001	0.217	0.209	0.942
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.839	-0.837	0.022	0.018	0.0
		$\hat{\beta}_2$	0.115	0.118	0.130	0.138	0.884
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.119	-0.099	0.333	0.217	0.760
		$\hat{\beta}_2$	0.053	0.052	0.182	0.157	0.908
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	0.008	0.009	0.236	0.226	0.946
		$\hat{\beta}_2$	0.004	0.012	0.165	0.159	0.950
20%	$\hat{\beta}_F$	$\hat{\beta}_1$	-0.007	-0.019	0.131	0.131	0.948
		$\hat{\beta}_2$	-0.001	-0.002	0.116	0.116	0.952
	$\hat{\beta}_{CC}$	$\hat{\beta}_1$	-0.009	-0.016	0.190	0.187	0.960
		$\hat{\beta}_2$	0.008	0.002	0.164	0.166	0.954
	$\hat{\beta}_N$	$\hat{\beta}_1$	-0.846	-0.845	0.019	0.017	0.0
		$\hat{\beta}_2$	0.105	0.106	0.102	0.110	0.868
	$\hat{\beta}_{ZW}$	$\hat{\beta}_1$	-0.114	-0.115	0.273	0.182	0.724
		$\hat{\beta}_2$	0.045	0.044	0.141	0.125	0.912
	$\hat{\beta}_{EPL}$	$\hat{\beta}_1$	-0.003	-0.017	0.189	0.182	0.942
		$\hat{\beta}_2$	-0.001	-0.003	0.129	0.127	0.952

CHAPTER 5: REAL DATA ANALYSIS

We apply the proposed approach to the primary biliary cirrhosis(PBC) data from the Mayo Clinic trial which was conducted between 1974 and 1984. PBC is a rare and fatal chronic liver disease in which the bile ducts in the liver become inflamed and damaged and, ultimately, destroyed. The cause of this disease is unknown. It develops over time and may cause the liver to stop working completely. There was a total of 424 patients who met the eligibility criteria for the randomized, placebo-controlled study of treatment of PBC with drug D-penicillamine. Complete data were collected on the first 312 cases who participated in the randomized trial. The remaining 112 cases did not participate in the clinical trial but some basic measurements on them were recorded to be followed for survival. Six of these cases were lost to follow-up shortly after diagnosis, so the data here are on an additional 106 cases as well as the 312 randomized participants. A detailed description about this dataset and the covariates recorded can be found in Dickson et al. [6] and Markus et al. [27].

The PBC data can be used to estimate a survival distribution, test for differences between two groups and estimate covariate effects via a regression model. The variables involved in our specfic analysis include:

(1) id: case number;

(2) days: number of days between registration and the earlier of death, transplantation, or study analysis time;

(3) status: status of censoring;

(4) chol: serum cholesterol (inmg/dl);

(5) bili: serum bilirubin (in mg/dl) and

(6) Age: age in days.

For this dataset, we wanted to analyze the effect of patients' serum cholesterol and age on the survival of the patients. This type of failure time data can be modeled by the Cox Proportional hazards models with an unknown baseline hazard function but special techniques are required when we have missing data. About 31% outcomes of cholesterol were missing in the data set. If we remove those observations, we may get biased estimates. In such situation we wanted to use the information from auxiliary covariates if available. To choose the auxiliary covariates, we performed preliminary statistical analysis on the available covariates. We observed that the outcomes of serum bilirubin were completely obtained with no missing values and we found that a significant correlation (0.4490) exists between serum cholesterol and bilirubin. We performed a Cox regression analysis to explore whether bilirubin has some additional effect on the hazard of failure. The results obtained are shown in the following table. We observed that, the estimates of the coefficients and their standard error estimates

Table 5.1: Regression Analysis of Primary Biliary Cirrhosis (PBC) data

	Method	Variable	Estimates of the Parameter	Standard Error	95% Confidence Interval
$logbili < 1.6$	CC	logchol	0.271	0.393	(-0.499, 1.040)
		age	0.055	0.012	(0.031, 0.079)
$logbili \geq 1.6$	ZW	logchol	-0.635	0.345	(-1.312, 0.042)
		age	-0.005	0.016	(-0.037, 0.027)

are quite different for both the situations and the 95% confidence intervals for the coefficient of age are nonoverlapping. Though there is a significant correlation between bilirubin and cholesterol, from the above analysis, we can conclude that serum bilirubin has some additional effect on the hazard of failure given the available information on serum cholesterol. That means it may not be a true surrogate for cholesterol. Hence, our proposed method can be applied to this dataset by considering serum bilirubin as an informative auxiliary covariate. In a preliminary step, we take the logarithmic transformation of cholesterol and bilirubin as suggested in the clinical literature. The following table displays the analysis results based on the the CC method, the ZW method proposed by Zhou and Wang(2000) and the proposed EPL method. The CC method uses only 284 complete-case observations and the other two methods use all 418 observations. Variable "logchol" denote the logarithm of cholesterol. The estimates of the coefficients and their standard errors are given in the Table 5.2.

Table 5.2: Regression Analysis of Primary Biliary Cirrhosis (PBC) data

Method	Variable	Estimates of the Parameter	Standard Error	95% Confidence Interval
CC	logchol	0.853	0.214	(0.432, 1.273)
	age	0.048	0.010	(0.029, 0.067)
ZW	logchol	1.054	0.168	(0.726, 1.383)
	age	0.046	.008	(0.032, 0.061)
EPL	logchol	0.871	0.212	(0.456, 1.287)
	age	0.043	0.007	(0.029, 0.058)

The regression analysis confirm that both serum cholesterol and age are significantly related to the time to event. The estimates of the regression parameters from the EPL method are very close to those obtained from the CC method. Note that there is a discrepancy between the estimates for "logchol"from complete data and Zhou and Wangs estimate which could be due to the fact that the latter method does not consider the additional effect contributed by the auxiliary covariate. The variance estimate for $\hat{\beta}_{EPL}$ is calculated using the proposed estimator $\hat{\Omega}(\beta)$. We observe that the estimated standard errors from the EPL and ZW methods are smaller than those from the CC method. The ratio of estimated standard errors for $\hat{\beta}_{EPL}$ and $\hat{\beta}_{ZW}$ relative to $\hat{\beta}_{CC}$ are(0.991, 0.7) and (0.785, 0.8) respectively. The 95% confidence interval for the regression parameter of "logchol" from the CC method, ZW method and EPL methods are (0.432, 1.273), (0.726, 1.383) and (0.456, 1.287) respectively. In our simulations we observed that the standard errors of the estimates were underestimated by ZW method (2000) method when the auxiliary variable was informative. In the real data analysis also the standard error estimate for serum cholesterol might be underestimated. In addition, the computation burden for our method is much less than that for Zhou and Wang's(2000) method. This is because the latter needs to run two-dimensional smoothing and the former just runs one-dimensional smoothing. This is also true for simulations.

CHAPTER 6: CONCLUSION AND FUTURE WORK

In this dissertation we have studied the proposed partial likelihood method (EPL)for dealing with missing and auxiliary covariates in failure time data. We compared the proposed method with several others methods. We assumed that the auxiliary covariate Z is continuous. In our model the auxiliary covariate W is assumed to be informative about the hazard of failure conditional on X, where X is the exposure variable which is missing for some of the subjects in the study cohort. We discussed the asymptotic properties of the proposed estimator. We have shown that the proposed estimator $\hat{\beta}_{EPL}$ is consistent for the parameter β and is asymptotically normally distributed. We also derived the consistent estimator of the asymptotic covariance of $\hat{\beta}_{EPL}$.

In the simulation study, we investigated the finite sample performance of the proposed estimator and compared the performance with several existing methods. It was observed that in most practical scenarios our estimator performs favorably and it is more efficient than the Cox partial likelihood estimator based only on the validation set. It was also found that the proposed method performs much better than Zhou and Wang's estimator when auxiliary covariate W is informative about the hazard of failure given X. In real life we often have auxiliary covariates which may not be true surrogates for X. This demonstrates advantages of our estimator.

A brief description of the nice properties of our proposed estimator $\hat{\beta}_{EPL}$ is given below:

(a) The proposed method allows W to be high dimensional and to be informative in the sense that, conditional X, it provide additional information about hazard of failure.

(c) This method utilizes the information about β in the non validation set since the partial likelihood includes all the individuals in the cohort.

(e) The validation set in our model is taken as a simple random sample from the cohort. This model can be extended to different sampling schemes like stratified sampling or outcome dependent sampling, which is under investigation in another project.

(f) The method is computationally straight forward and the computation time is much less than Zhou and Wang's method.

(g) The problem of curse of dimensionality has been partially removed.

(h) As illustrated in the simulation study, Zhou and Wang's estimator is consistent for different values of σ when $\gamma = 0$. In finite samples their method is more efficient for very small values of σ but for higher values of σ both methods are almost equally efficient and their estimator has more bias than our estimator. Also for situations where $\gamma \neq 0$, the consistency property of their estimator does not hold whereas our estimator remains to be consistent.

CONCERNS:

We have few concerns with the proposed method.

1. The proposed estimator will not perform well if the dimension of Z is high. In such situations we can introduce some additive structure.

2. We used the same bandwidth as suggested by Zhou and Wang (2000) in our estimation. Though it performs reasonably well, it would be worthwhile to consider a bandwidth selection criteria like generalized cross-validation.

3. It is desirable to increase the efficiency of the estimation. In future, We can consider modifying our model by including a suitable weight in the partial likelihood score equation.

FUTURE RESEARCH:

In our study, the sampling scheme is simple random sampling. We would like to extend our method for outcome dependent sampling (ODS) which is a cost effective sampling strategy. In the ODS design, one observes the exposure with a probability, maybe unknown, depending on the outcome.

Also another extension of our method is multivariate failure time data which arise in many contexts. In that case our model can be modified to a stratified model.

REFERENCES

[1] Andersen, P. K. and Gill (1982) Cox's Regression Model for Counting Processes: a large sample study. *Annals of Statistics*, **10**, 1100-1120.

[2] Breslow, N. E. (1972) Contribution to Discussion of Paper by D.R. Cox. *Journal of Royal Statistical Society B*, **34**, 216-217.

[3] Breslow, N. E. (1974) Covariance Analysis of Censored Survival Data. *Biometrics*, **30**, 89-99.

[4] Chen, Y.-H. and Chen and Little, R.J.A. (1999) Proportional Hazards Regression with Missing Covariates. *Journal of Royal Statistical Society B*, **94**, 896908.

[5] Dempster, A.P., Liard, N.M. and Rubin, D.B. (1977) Maximun Likelihood from Incomplete Data via the EM Algorithm. *Journal of Royal Statistical Society B*, **39**, 1-38.

[6] Dickson, E.R., Grambsch, P.M., Fleming, T.R, Fisher, L.D., and Langworthy, A. (1989) Prognosis in Primary Biliary Cirrhosis: Model for Decision Making. *Hepatology*, **10**, 17.

[7] Cox, D. R. (1972) Regression Models and Life-tables (with discussion). *Journal of Royal Statistical Society B*, **34**, 187-220.

[8] Eubank, R. L. (1988) Smoothing Spline and Nonparametric Regression. New York: Dekker.

[9] Efron, B. (1977) Efficiency of Cox's Likelihood Function for Censored Data. *Journal of American Statistical Association*, **72**, 557-565.

[10] Fan, J. and Gijbels, I. (1996) *Local Polynomial Modeling and Its Applications.* Chapman and Hall, London.

[11] Fan, J. and Yao, Q. (2005) *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer-Verlag.

[12] Fisher, Lloyd D., Lin, D. Y. (1999) Time-dependent Covariates in the Cox Proportional Hazards Model. *Annu. Rev. Public Health*, **20**, 145-157.

[13] Fleming, T. R. and Harrington, D. P. (1991) *Counting Process and Survival Analysis.* New York: Wiley.

[14] Foutz, R. V. (1977) On the Unique Consistent Solution to the Likelihood Equations. *Journal of American Statistical Association*, **72**, 147-148.

[15] Glasser, M. (1967) Exponential Survival with Covariance. *Annals of Statistics*, **9**, 861-869.

[16] Gong, G. and samaniego, F.J. (1981) Pseudo Maximum Likelihood Estimation: Theory and Applications. *Journal of American Statistical Association*, **62**, 561-568.

[17] Härdle, W. (1990) *Applied Nonparametric Regression*. London: Cambridge University Press.

[18] Hughes, M. D. (1993) Regression Dilution in the Proportional Hazards Model. *Biometrics*, **49**, 1056-1066.

[19] Jiang, J. and Mack, Y. P. (2001) Robust Local Polynomial Regression for Dependent Data. *Statistica Sinica*, **11**, 705-722.

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

[21] Kulich, M. and Lin, D. Y. (2000) Additive Hazards Regression with Covariate Measurement Error. *Journal of American Statistical Association*, **95**, 238-248.

[22] Lehmann, E.L. (1983) Theory of Point Estimation. *J. Wiley and Sons, New York*.

[23] Lin, D. Y. and Ying, Z. (1993) Cox Regression with Incomplete Covariate Measurements. *Journal of American Statistical Association*, **88**, 1341-1349.

[24] Lipsitz, S. and Ibrahim, J. G. (1996) Using the E-M algorithm for Survival Data with Incomplete Categorical Covariates. *Lifetime Data Analysis*, **2**, 5-14.

[25] Liu, Y., Wu, Y., and Zhou, H. (2010) Multivariate Failure Times Regression with a Continuous Auxiliary Covariate. *Journal of Multivariate Analysis*, **101**, 679-691.

[26] Louis, T.(1982) Finding the Observed Information Matrix when using the EM Algorithm. *Journal of Royal Statistical Society B*, **44,2**, 226-233.

[27] Markus, B.H., Dickson, E.R., Grambsch, P.M, Fleming, T.R., Mazzaferro, V., Klintmalm, G.B., Wiesner, R.H., Van Thiel, D.H. and Starzl, T.E. (1989) Efficiency of Liver Transplantation in Patients with Primary Biliary Cirrhosis, N. Engl. *J. Med.* **320**, 17091713.

[28] McCullagh, P. and Nelder, J.A. (1989) Generalized Linear Models, Second Edition. *Chapman and Hall, London*.

[29] Oaks, D. (1977) The Asymptotic Information in Censored Survival Data. *Biometrika*, **64**, 441-448.

[30] Nadaraya, E. A. (1964) On Estimating Regression. *Theory Probab. Applic.*, **10**, 186-190.

[31] Pepe, M. S. and Fleming, T.R. (1991) A Nonparametric Method for Dealing with Mismeasured Covariate data. *Journal of American Statistical Association* **86, 413**, 108-113.

[32] Pepe, M. S., Self, S. G. and Prentice, R. L. (1989) Further Results on Covariate Measurement Errors in Cohort Studies with Time to Response Data. *Statist. Med.*, **8**, 1167-1178.

[33] Prentice, R. L. (1982) Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model. *Biometrika*, **69**, 331-342.

[34] Prentice, R. L. and Self, S.G. (1983) Asymptotic Distribution Theory for Cox-type Regression Models with General Relative Risk Form. *Annals of Statistics*, **11**, 804-813.

[35] Rubin, D. B. (1976) Inference and Missing Data. *Biometrika*, **63**, 581-592.

[36] Rudin, W. (1964) PrincipleS of Mathematical Analysis, New York: McGraw-Hill Book Co.

[37] Silverman, B. W. (1978) Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and its Derivatives. *Annals of Statistics*, **6**, 177-184.

[38] Watson, G. S. (1964) Smooth Regression Analysis. *Sankhya A*, **26**, 359-372.

[39] Zhou, H. (1992) Auxiliary and Missing Covariate Problems in Failure Time Regression Analysis, Ph.D. Thesis, University of Washington.

[40] Zhou, H. and Pepe, M. S. (1995) Auxiliary Covariate Data in Failure Time Regression Analysis. *Biometrika*, **82**, 139-149.

[41] Zhou, H. and Wang, C.-Y. (2000) Failure Time Regression with Continuous Covariates Measured with Error. *Journal of Royal Statistical Society B*, **62**, 657-665.