Estimation of the primary hazard ratio in the presence of a secondary covariate with non-proportional hazards

An undergraduate honors thesis submitted to the Department of Mathematics

by

Xinyi Lin

Advisor: Professor Ronghui Xu

June 2007
Abstract

We study the effect of a secondary covariate with non-proportional hazards on the estimation of the regression effect of a primary covariate in the Cox model. This is motivated by epidemiologic studies where estimation of the effect of primary exposure often needs to take into account confounders. The question is whether the simple proportional hazards modelling of the confounders might be sufficient. This is done for two cases: when both covariates are binary and when the primary covariate is binary and the secondary covariate is uniformly distributed. The sign and magnitude of the bias of the primary hazard ratio depends on the non-proportional hazards of the secondary covariate, the strength of the regression effect of the primary covariate itself, and censoring. We summarize the results obtained through both numerical calculations and simulations.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
</tr>
<tr>
<td>1 Introduction</td>
<td>3</td>
</tr>
<tr>
<td>2 Cox proportional hazards model</td>
<td>6</td>
</tr>
<tr>
<td>3 Non-proportional hazards model</td>
<td>11</td>
</tr>
<tr>
<td>II</td>
<td></td>
</tr>
<tr>
<td>4 Theoretical Results</td>
<td>14</td>
</tr>
<tr>
<td>5 Numerical and Simulation Results</td>
<td>19</td>
</tr>
<tr>
<td>Appendix</td>
<td>34</td>
</tr>
<tr>
<td>Bibliography</td>
<td>38</td>
</tr>
</tbody>
</table>
Part I
Chapter 1

Introduction

Scope of Survival Analysis

In survival analysis, interest centers on time to event data. It involves a group (or groups) of subjects for which there is a well-defined point event, often called failure, occurring after a length of time called the failure time. The random variable involved, failure time, is always non-negative and failure can only occur once for any subject. Examples include the time to death after diagnosis of breast cancer, the lifetimes of machine components or the duration of unemployment.

General Features of Survival Data

To define a failure time random variable, three requirements must be fulfilled: a well-defined time origin, a scale for measuring time and a definition of the event. There are three common features of survival data. The first is that individuals do not enter the study at the same time. For example, in a clinical trial comparing the effectiveness of different treatments of cancer, the time origin might be the time of diagnosis, which occurs at various times for different patients. This feature is referred to as staggered entry. The second feature is that when a study closes, some individuals have not experienced failure yet. This might occur when a clinical trial ends after a designated time. The last feature occurs when subjects drop out in the middle of the study. In clinical trials, patients might be lost to follow-up or in other cases, some patients might die from other causes besides the one under study. The second and third features are known as censoring.

Censoring

There are three main types of censoring mechanisms - right censoring, left censoring and interval censoring. Consider a random sample of $n$ subjects with survival times $T_1, T_2, ..., T_n$ and censoring times $C_1, C_2, ..., C_n$ for the individuals $i = 1, 2, ..., n$. Right censoring occurs when the true unobserved event is to the right of our censoring time. That is, all one knows
is that the event has yet to occur at the end of the study. Thus one observes the random variable $X_i = \min(C_i, T_i)$. Left censoring occurs when the true unobserved event is to the left of the censoring time. The random variable observed is $G_i = \max(C_i, T_i)$. For example, in a study of age at which African children learn a task, some already knew the task (left-censored), some learned during the study (exact) and some had not yet learned by the end of the study (right-censored). Interval censoring occurs when the failure time is only known to occur within some interval. One observes $(L_i, R_i)$ where $T_i \in (L_i, R_i)$. Right censoring is the most common mechanism and this thesis will focus only on right censoring.

Additionally, for right (and also left) censoring, there are three types of censoring times: Type I, Type II and random. For a sample of $n$ subjects, Type I censoring occurs when all the $C_i$’s are the same. Type II censoring occurs when $C_i = T(r)$, that is the study is terminated at the failure time of the $r$th subject. In a clinical trial, $r$ has to be pre-determined. Random censoring occurs when $C_i$’s are random variables. One can additionally define an event indicator $\delta_i = I(T_i \leq C_i)$ and an at risk indicator $Y_i(t) = I(X_i \geq t)$. In this thesis, only random censoring will be considered.

**Describing Survival Data**

There are many ways to describe survival data: using the density function $f(t)$, cumulative distribution function $F(t)$, the survival function $S(t)$, the hazard function $\lambda(t)$ or the cumulative hazard function $\Lambda(t)$. The definitions of these functions will be given for a continuous random variable.

\[
f(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t)}{\Delta t} \tag{1.1}
\]

\[
F(t) = P(T < t) \tag{1.2}
\]

\[
S(t) = P(T \geq t) = \int_t^\infty f(u)du \tag{1.3}
\]

\[
\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{1.4}
\]

\[
\Lambda(t) = \int_0^t \lambda(u)du \tag{1.5}
\]

It can be shown that all of these formulations are equivalent; given any one of them, one can easily obtain the rest. Of these five definitions, the first three should be familiar from classical statistical theory. The last two require some explanation. The hazard function gives
the instantaneous hazard rate. It describes the conditional probability rate that a subject will fail in the next interval \( \Delta t \), given that it has not failed at time \( t \). The cumulative hazard function is analogous to the cumulative distribution function.

The survival function \( S(t) \) can be estimated using various nonparametric methods, like Kaplan-Meier (1958) or Life-table estimator. While the survival function gives a much clearer interpretation to the data concerned, there are good reasons why the consideration of the hazard function is a good idea (Cox and Oakes, 1984):

1. It makes more physical sense to consider the instantaneous risk of an individual known to be alive at a certain time.

2. Comparisons of groups of individuals are sometimes made more incisively using the hazard function.

3. It is convenient to model using the hazard function when there is censoring or multiple failures.

4. Comparison made using the exponential distribution is simple as the hazard function is a constant.
Chapter 2

Cox proportional hazards model

In modeling the regression of survival data, a model must accomplish two goals: it must describe the underlying survival time distribution (error component) as well as model how adequately the distribution changes with the covariates (systematic component) (Hosmer and Lemeshow, 1999). The Cox proportional hazards model (Cox 1972, 1975) gives a simple relationship between the hazard function \( \lambda(t) \) and a vector of covariates \( Z \):

\[
\lambda(t|Z) = \lambda_0(t) \exp(\beta'Z) \tag{2.1}
\]

\( \lambda_0(t) \), the baseline hazard function, is an arbitrary non-negative function of time and is the hazard for subjects with all covariates equal to zero. In the modeling of survival data, it is usually treated as a nuisance function as one of the greatest advantages of the Cox model is that the parameters \( \beta \) can be estimated without having to estimate \( \lambda_0(t) \). The term proportional hazards refers to the fact that the ratio of the hazard functions for any two subjects remain proportional over time due to the multiplicative relationship in the model. This model is semi-parametric as it assumes that \( \beta \) is time independent. If a form for \( \lambda_0(t) \) is specified, then the model becomes fully parametric. The proportional hazards assumption is a strong one and the motivation for this thesis comes from its violation.

Inferences for the proportional hazards model

The full likelihood function is given by

\[
L(\beta) = \prod_{i=1}^{n} \lambda_i(X_i)^{\delta_i} S_i(X_i) \tag{2.2}
\]

Multiplying and dividing by \( \left[ \sum_{j=1}^{n} Y_j(X_i) \lambda_j(X_i) \right]^{\delta_i} \) gives

\[
L(\beta) = \prod_{i=1}^{n} \left[ \frac{\lambda_i(X_i)}{\sum_{j=1}^{n} Y_j(X_i) \lambda_j(X_i)} \right]^{\delta_i} \left[ \sum_{j=1}^{n} Y_j(X_i) \lambda_j(X_i) \right]^{\delta_i} S_i(X_i) \tag{2.3}
\]
The first term in the product is known as the partial likelihood function and Cox argued that it contains all the information about the parameter of interest and conjectured that the resulting parameter estimators would have the same properties as maximum likelihood estimators. This conjecture was later proved to be valid (Tsiatis 1981, Anderson and Gill 1982). Keeping only the first term and using the proportional hazards assumption gives

\[ L(\beta) = \prod_{i=1}^{n} \left[ \frac{\lambda_i(X_i)}{\sum_{j=1}^{n} Y_j(X_i) \lambda_j(X_i)} \right]^{\delta_i} \]

\[ = \prod_{i=1}^{n} \left[ \frac{\lambda_0(X_i) \exp(\beta'Z_i(X_i))}{\sum_{j=1}^{n} Y_j(X_i) \lambda_0(X_i) \exp(\beta'Z_j(X_i))} \right]^{\delta_i} \]

\[ = \prod_{i=1}^{n} \left[ \frac{\exp(\beta'Z_i(X_i))}{\sum_{j=1}^{n} Y_j(X_i) \exp(\beta'Z_j(X_i))} \right]^{\delta_i} \]

(2.4)

The log partial likelihood is given by

\[ l(\beta) = \log \left[ \prod_{i=1}^{n} \left[ \frac{\exp(\beta'Z_i(X_i))}{\sum_{j=1}^{n} Y_j(X_i) \exp(\beta'Z_j(X_i))} \right]^{\delta_i} \right] \]

\[ = \sum_{i=1}^{n} \delta_i \left[ \beta'Z_i(X_i) - \log \left( \sum_{j=1}^{n} Y_j(X_i) \exp(\beta'Z_j(X_i)) \right) \right] \]

\[ = \sum_{i=1}^{n} l_i(\beta) \]

(2.5)

In general, the estimator for \( \beta \) is obtained by taking the partial derivative with respect to \( \beta \) and setting it to zero. In the case of \( k \) covariates, one would obtain a system of \( k \) equations which would have to be solved simultaneously. For simplicity, we will consider the case of a single covariate. The partial likelihood score equation is

\[ U(\beta) = \frac{\partial}{\partial \beta} l(\beta) \]

\[ = \sum_{i=1}^{n} \delta_i \left[ Z_i(X_i) - \frac{\sum_{j=1}^{n} Y_j(X_i) Z_j(X_i) \exp(\beta Z_j(X_i))}{\sum_{j=1}^{n} Y_j(X_i) \exp(\beta Z_j(X_i))} \right] \]

\[ = \sum_{i=1}^{n} \delta_i \left[ Z_i(X_i) - \bar{Z}_i(X_i) \right] \]

(2.6)

\( \bar{Z}_i(X_i) \) is the weighted average of the covariate \( Z \) over all the subjects that are at risk at time \( X_i \), and the weights are the conditional probabilities that contribute to the partial likelihood function. The maximum partial likelihood estimator is found by solving

\[ U(\beta) = 0 \]

(2.7)
The observed information is given by the negative of the second partial derivative of the log partial likelihood.

\[ I(\beta) = -\frac{\partial^2 l(\beta)}{\partial \beta^2} \quad (2.8) \]

The variance of \( \hat{\beta} \) is found by inverting the observed information. An estimator of the variance of \( \hat{\beta} \) is thus given by

\[ \widehat{Var}(\hat{\beta}) = I(\hat{\beta})^{-1} \quad (2.9) \]

The above discussion assumes that there are no tied times. In the event of ties, there are a few methods to modify the partial likelihood to adjust for ties, the details of which will not be discussed. There are three ways where inferences can be made about \( \beta \). The first approach is the Wald test that uses the fact that \( \hat{\beta} \) is asymptotically normally distributed with mean \( \beta \) and variance \( I(\hat{\beta})^{-1} \). Inferences can also be made by using the partial likelihood ratio test or the score test.

**Assessing Model Adequacy**

- **Assessing general fit of Cox Model**
  To assess the fit of the Cox model, Cox-Snell, Schoenfield (scaled or unscaled) or Martingale residuals can be used. The Cox-Snell residual is defined as

\[ \widehat{\Lambda}_i(X_i) = -\log\left[ \hat{S}(X_i|Z_i) \right] \quad (2.10) \]

If the survival time of the \( i^{th} \) individual \( T_i \) has survival function \( S_i(t) \), then the random variable \( S_i(T_i) \) is uniformly distributed on \((0,1)\) and \( \Lambda_i(T_i) = -\log[S(T_i)] \) is exponentially distributed with mean 1. Thus the fit of the model is assessed by checking whether the Cox-Snell residuals are exponentially distributed with mean 1.

The Schoenfeld residuals are defined as

\[ r_i = Z_i(X_i) - \bar{Z}_i(X_i) \quad (2.11) \]

The Schoenfeld residuals are only defined for uncensored observations. Under the assumption of the Cox model, the Schoenfeld residuals are asymptotically uncorrelated and have expectation zero. The adequacy of the Cox model is therefore checked by visually verifying that the plot of the residuals against survival time is centered around zero and shows no trend over time. The scaled Schoenfeld residuals are defined as

\[ r_i^w = m\widehat{Var}(\hat{\beta})r_i \quad (2.12) \]
where $m$ is the observed number of uncensored survival times. The scaled Schoenfeld residuals are used in a similar way as the unscaled residuals except that it has greater diagnostic power than the unscaled residuals (Grambsch and Therneau, 1994).

The Martingale residual for the $i$th individual is defined as

$$M_i = \delta_i - \hat{\Lambda}_i(X_i)$$  \hspace{1cm} (2.13)

It has the interpretation of being the difference between the observed number of deaths and the expected number of deaths based on the fitted model for individual $i$ in the time between 0 and $X_i$. The Martingale residuals have expectation zero and are approximately uncorrelated in large samples.

• Assessing Proportional Hazards assumption

Under the proportional hazards assumption, we have

$$S(t|Z) = \exp\left[-\Lambda(t|Z)\right] = \exp\left[-\int_0^t \lambda(u|Z)du\right] = \exp\left[-\int_0^t \lambda_0(u)\exp(\beta'Z)du\right] = \left[-\int_0^t \lambda_0(u)du\right]^{\exp(\beta'Z)} = [S_0(t)]^{\exp(\beta'Z)}$$  \hspace{1cm} (2.14)

Thus

$$\log\left[-\log S(t|Z)| = \log\left[-\log S_0(t)\right] + \beta'Z$$  \hspace{1cm} (2.15)

Hence, one could obtain the Kaplan-Meier curves for the survival functions and plot $\log\left[-\log S(t|Z)|$ against time for the different values of the covariates. In the case of a continuous covariate, the covariate can be split into categories. If the proportional hazards assumption is not violated, the graphs should be parallel.

An alternative method uses the scaled Schoenfeld residuals. Taking the log of the Cox model gives

$$\log[\lambda(t|Z)] = \log[\lambda_0(t)] + \beta'Z$$  \hspace{1cm} (2.16)

If instead the model has time varying coefficients of the form

$$\beta_j(t) = \beta_j + \gamma_j g_j(t)$$  \hspace{1cm} (2.17)
where $g_j(t)$ is a specified function of time, Grambsch and Therneau (1994) showed that for the $j^{th}$ covariate,

$$E[r_j^w] = \gamma_j g_j(t) \tag{2.18}$$

Thus a plot of the scaled Schoenfeld residuals will give an indication as to whether $\gamma_j$ is zero and if it is not zero, an indication of the form for $g_j(t)$. For a given functional form such as $g_j(t) = \ln(t)$, it is possible to test the hypothesis $\gamma_j = 0$ using the Wald test, partial likelihood ratio test or score test by adding the interaction term $Z_j \ln(t)$ to the Cox model. The advantage of this method is that it can easily be done using the same software that fits a proportional hazards model.
Chapter 3

Non-proportional hazards model

The Cox proportional hazards model described in the previous chapter has wide applications. One of the main assumptions is that the coefficients of regression $\beta$ remains constant with time. When this assumption is violated, the partial likelihood estimator ceases to have a clear interpretation. In this thesis, we consider a non-proportional hazards model of the form

$$\lambda(t|Z(t)) = \lambda_0(t) \exp(\beta(t)'Z(t))$$ (3.1)

Definition 1.

$$S^{(r)}(\beta(t), t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) \exp(\beta(t)'Z(t))Z_i(t)^\otimes r$$ (3.2)

$$s^{(r)}(\beta(t), t) = E[S^{(r)}(\beta(t), t)]$$ (3.3)

for $r = 0, 1, 2$ and the expectations are taken with respect to the true distribution of $(T, C, Z)$.

Xu and O'Quigley (2000) also showed that $s^{(1)}(\beta(t), t)/s^{(0)}(\beta, t) = E[Z(t)|T = t]$. For proportional hazards, one will then replace $\beta(t)$ by $\beta$. The following theorem was proved by Struthers and Kalbfleisch (1986):

Theorem 1. The maximum partial likelihood estimator $\hat{\beta}_{PL}$ is a consistent estimator of $\beta^*$ where $\beta^*$ is the unique solution to the equation

$$\int_0^T \left( s^{(1)}(\beta(t), t) - \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} s^{(0)}(\beta(t), t) \right) \lambda_0(t) dt = 0$$ (3.4)

provided that the following two conditions hold:
**Condition 1**

There exists a neighbourhood $B$ of $\beta^*$ such that for each $t < \infty$

$$\sup_{x \in [0,t], \beta \in B} |S^{(0)}(\beta, x) - s^{(0)}(\beta, x)| \to 0$$

in probability as $n \to \infty$, $s^{(0)}(\beta, x)$ is bounded away from zero on $B \times [0,t]$, and $s^{(0)}(\beta, x)$ and $s^{(1)}(\beta, x)$ are bounded on $B \times [0,t]$.

**Condition 2**

For each $t < \infty$, $\int_0^\infty s^{(2)}(\beta, x)dx < \infty$.

The solution $\beta^*$ depends on the unknown censoring mechanism through the term $s^{(0)}(\beta(t), t)$ in (3.4) and hence does not have a clear interpretation under non-proportional hazards. In contrast, Xu and O’Quigley (2000) proposed an alternative estimator which does not depend on the censoring mechanism and thus has a well-defined interpretation even for non-proportional hazards. This estimator can be viewed as an average regression effect and is essentially a weighted average of $\beta(t)$ over time. They also proposed a simple method to estimate this quantity.

Many others have also studied the effects of a misspecified proportional hazards model. Struthers and Kalbfleisch (1986) looked at a two-covariate proportional hazards model and found that if one covariate was missing in the fitted model, the resulting partial likelihood estimator will always underestimate the regression parameter in absolute value in the true model (biased towards 0). Bretagnolle and Huber-Carol (1988) generalized this result to the case when there are more than one covariate remaining and showed that the underestimation still holds for each of the remaining covariates up until some fixed time, which is reasonably long in most practical cases. Ford et al. suggested that the converse may be true, that is the inclusion of a covariate which is not in the true model may bias the regression effect away from zero (1995).
Part II
Chapter 4

Theoretical Results

Problem and Methodology

\[ \lambda(t|Z^{(1)}, Z^{(2)}) = \lambda_0(t) \exp(\beta_1 Z^{(1)} + \beta_2(t) Z^{(2)}) \]  
\hspace{1cm} (4.1) \]

Consider model (4.1) which is a special case of the model in (3.1) where we have two time-independent covariates \( Z^{(1)} \) and \( Z^{(2)} \) that are mutually independent, with corresponding regression coefficients \( \beta_1 \) and \( \beta_2(t) \). The regression coefficient of \( Z^{(1)} \) is time-independent while that of \( Z^{(2)} \) is time-dependent. Given the non-proportional hazards of the second covariate, we are concerned with how this will affect the estimation of \( \beta_1 \) when a proportional hazards model of the form

\[ \lambda(t|Z^{(1)}, Z^{(2)}) = \lambda_0(t) \exp(\beta_1 Z^{(1)} + \beta_2 Z^{(2)}) \]  
\hspace{1cm} (4.2) \]

is fitted. This problem is motivated by epidemiologic studies where estimation of the regression effect of primary exposure often needs to take into account confounders which may follow non-proportional hazards. The exact solutions of the regression coefficients of the Cox model (4.2), \( \beta_1^* \) and \( \beta_2^* \) were obtained by solving (3.4) numerically. Additionally, simulations were run to compare the partial likelihood estimators \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \), which are obtained by solving (2.7), with the exact solutions. In the remainder of this chapter, we state the assumptions, theorems and corollaries that will be used to solve (3.4). The proofs of the theorems are given in the Appendix.
Assumptions

(A1) $T_i, C_i$ and $Z_i$ for $i = 1, 2, ..., n$ are independent and identically distributed with distributions $F_T, F_C$ and $F_Z$ respectively.

(A2) $\lambda_0(t) = 1$

(A3) $\beta_2(t) = k_1$ for $t \leq t_0$ and $\beta_2(t) = k_2$ for $t > t_0$

(A4) $Z$ is time independent.

(A5) Censoring random variable $C$ is independent of $Z$ and $T$.

(A6) $C \sim \text{Uniform}(0, \tau)$

Theorem 2. Under assumptions (A1-4) and model (4.1), the cumulative density function of $T$, condition on $Z$ is given by

$$F_{T|Z}(t) = \begin{cases} 1 - \exp(-\omega_1 t) & , t \leq t_0 \\ 1 - \exp(-\omega_1 t_0 - \omega_2 (t - t_0)) & , t > t_0 \end{cases}$$

(4.3)

where

$$Z = \begin{pmatrix} Z^{(1)} \\ Z^{(2)} \end{pmatrix}$$

(4.4)

$$\omega_1 = \exp(\beta_1 Z^{(1)} + k_1 Z^{(2)})$$

(4.5)

$$\omega_2 = \exp(\beta_1 Z^{(1)} + k_2 Z^{(2)})$$

(4.6)

Theorem 3. Under (A1),(A4)-(A6), the expectation of the at-risk indicator $Y(t)$, condition on $Z$ is given by

$$E[Y(t)|Z] = \left(1 - \frac{t}{\tau}\right) \left(1 - F_{T|Z}(t)\right)$$

(4.7)

Theorem 4. Under (A1)-(A6) and model (4.1), the quantities $s^{(0)}(\beta(t), t)$ and $s^{(1)}(\beta(t), t)$ are given by

$$s^{(0)}(\beta(t), t) = \begin{cases} E_Z \left[\omega_1 \left(1 - \frac{t}{\tau}\right) \exp(-t\omega_1)\right] & , t \leq t_0 \\ E_Z \left[\omega_2 \left(1 - \frac{t}{\tau}\right) \exp(-t_0\omega_1 - (t - t_0)\omega_2)\right] & , t > t_0 \end{cases}$$

(4.8)

$$s^{(1)}_{1}(\beta(t), t) = \begin{cases} E_Z \left[Z^{(1)}\omega_1 \left(1 - \frac{t}{\tau}\right) \exp(-t\omega_1)\right] & , t \leq t_0 \\ E_Z \left[Z^{(1)}\omega_2 \left(1 - \frac{t}{\tau}\right) \exp(-t_0\omega_1 - (t - t_0)\omega_2)\right] & , t > t_0 \end{cases}$$

(4.9)
where $E_\omega$ is the expectation taken with respect to $Z$ and

$$s_1^{(1)}(\beta(t), t) = \left( \begin{array}{c} s_1^{(1)}(\beta(t), t) \\ s_2^{(1)}(\beta(t), t) \end{array} \right)$$

(4.11)

**Corollary 1.** If $Z^{(1)}$ has a multinomial distribution with values $a_i$ occurring with probabilities $p_i$ for $1 \leq i \leq m$ and $Z^{(2)}$ follows a multinomial distribution with values $\gamma_j$ occurring with probabilities $q_j$ for $1 \leq j \leq l$, where $\sum_{i=1}^{m} p_i = 1$ and $\sum_{j=1}^{l} q_j = 1$, and the two variables are independent, then the quantities $s_0^{(0)}(\beta(t), t)$ and $s_0^{(1)}(\beta(t), t)$ are given by

$$s_0^{(0)}(\beta(t), t) = \left\{ \begin{array}{ll} \sum_{i=1}^{l} \sum_{i=1}^{m} \beta_i p_i q_j \exp(\beta_i a_i + \gamma_j) (1 - \frac{1}{2}) \exp(-t \exp(\beta_i a_i + \gamma_j)) & , t \leq t_0 \\
\sum_{i=1}^{l} \sum_{i=1}^{m} \beta_i p_i q_j \exp(\beta_i a_i + \gamma_j) (1 - \frac{1}{2}) \exp(-t \exp(\beta_i a_i + \gamma_j)) & , t > t_0 \\
\end{array} \right.$$
Corollary 2. If $Z^{(1)}$ has a multinomial distribution with values $a_i$ occurring with probabilities $p_i$ for $1 \leq i \leq m$, where $\sum_{i=1}^{m} p_i = 1$, and $Z^{(2)}$ is uniformly distributed on $(0,1)$, and the two variables are independent, then the quantities $s^{(0)}(\beta, t)$ and $s^{(1)}(\beta, t)$ are given by

$$s^{(0)}(\beta, t) = \begin{cases} f_0^1 \sum_{i=1}^{m} p_i \exp \left( \beta_1 + k_1 Z^{(2)} \right) (1 - \frac{t}{\tau}) \times \\
\exp \left( -t \exp \left( \beta_1 + k_1 Z^{(2)} \right) \right) dZ^{(2)} , t \leq t_0 \\
\exp \left( -t_0 \exp \left( \beta_1 + k_1 Z^{(2)} \right) \right) - (t - t_0) \exp \left( \beta_1 + k_2 Z^{(2)} \right) dZ^{(2)} , t > t_0 
\end{cases}$$

$$s^{(1)}(\beta, t) = \begin{cases} f_0^1 \sum_{i=1}^{m} p_i \exp \left( \beta_1 + k_1 Z^{(2)} \right) (1 - \frac{t}{\tau}) \times \\
\exp \left( -t \exp \left( \beta_1 + k_1 Z^{(2)} \right) \right) dZ^{(2)} , t \leq t_0 \\
\exp \left( -t_0 \exp \left( \beta_1 + k_1 Z^{(2)} \right) \right) - (t - t_0) \exp \left( \beta_1 + k_2 Z^{(2)} \right) dZ^{(2)} , t > t_0 
\end{cases}$$

Furthermore, the quantities $s^{(0)}(\beta^*, t)$ and $s^{(1)}(\beta^*, t)$ are given by

$$s^{(0)}(\beta^*, t) = \begin{cases} f_0^1 \sum_{i=1}^{m} p_i \exp \left( \beta_1^* + \beta_2^* Z^{(2)} \right) (1 - \frac{t}{\tau}) \times \\
\exp \left( -t \exp \left( \beta_1^* + \beta_2^* Z^{(2)} \right) \right) dZ^{(2)} , t \leq t_0 \\
\exp \left( -t_0 \exp \left( \beta_1^* + \beta_2^* Z^{(2)} \right) \right) - (t - t_0) \exp \left( \beta_1^* + \beta_2^* Z^{(2)} \right) dZ^{(2)} , t > t_0 
\end{cases}$$

$$s^{(1)}(\beta^*, t) = \begin{cases} f_0^1 \sum_{i=1}^{m} p_i \exp \left( \beta_1^* + \beta_2^* Z^{(2)} \right) (1 - \frac{t}{\tau}) \times \\
\exp \left( -t \exp \left( \beta_1^* + \beta_2^* Z^{(2)} \right) \right) dZ^{(2)} , t \leq t_0 \\
\exp \left( -t_0 \exp \left( \beta_1^* + \beta_2^* Z^{(2)} \right) \right) - (t - t_0) \exp \left( \beta_1^* + \beta_2^* Z^{(2)} \right) dZ^{(2)} , t > t_0 
\end{cases}$$

The proofs of Corollary 1 and 2 follow trivially from Theorem 4 and will not be presented. The first part of Corollary 1 and 2 can be easily obtained by taking the expectation with respect to $Z$. The second part can be obtained from the first part with the replacements $\beta \rightarrow \beta_1^*$, $k_1 \rightarrow \beta_2^*$, $k_2 \rightarrow \beta_2^*$ only for the first exponential term. The replacement is not done for the other terms as the expectation is taken with respect to the true distribution of $(T, C, Z)$. 

17
Solution to (3.4) using Newton-Raphson

The exact solutions of the regression coefficients $\beta^* = (\beta_1^*, \beta_2^*)'$ of the Cox model (4.2), with baseline hazard 1, are found by solving the system of two equations

$$\int_0^\tau \left( s_1^{(1)}(\beta(t), t) - \frac{s_1^{(1)}(\beta^*, t)}{s(0)(\beta^*, t)} s^{(0)}(\beta(t), t) \right) dt = 0 \quad (4.12)$$

$$\int_0^\tau \left( s_2^{(1)}(\beta(t), t) - \frac{s_2^{(1)}(\beta^*, t)}{s(0)(\beta^*, t)} s^{(0)}(\beta(t), t) \right) dt = 0 \quad (4.13)$$

We consider two special cases:

A. $Z^{(1)}, Z^{(2)}$ take on values 0 or 1 with probability $\frac{1}{2}$ each.

B. $Z^{(1)}$ as above and $Z^{(2)}$ is a continuous random variable, uniformly distributed on (0,1).

The expressions for $s^{(0)}(\beta(t), t), s^{(0)}(\beta^*, t), s^{(1)}(\beta(t), t)$ and $s^{(1)}(\beta^*, t)$ are given by Corollary 1 and 2 for cases A and B respectively. For case A, the system of equations (4.12) and (4.13) was solved numerically using the Newton-Raphson method, of which the existence and uniqueness of the solution is guaranteed by Theorem 1. These results are then compared to the partial likelihood estimators from simulations. For case B, the system of equations could not be solved using Newton-Raphson and so only simulations were run to study the effect of non-proportional hazards on the estimation of $\beta_1$. The partial likelihood estimators were obtained by averaging over 1000 simulations, with a sample size of 1000 each.
Chapter 5
Numerical and Simulation Results

In this chapter, we summarize both the numerical and simulation results obtained when model (4.2) is fitted when the true model is (4.1).

Case A. $Z^{(1)}, Z^{(2)}$ binary 0 or 1

The results of the simulations and numerical solutions obtained by Newton-Raphson for case A are shown in Figures 5.1-5.5 and Tables 5.1-5.4. In Figures 5.1-5.3 and 5.5, the dotted horizontal line shows the true value of $\beta_1$ (in absolute value for cases where $\beta_1 < 0$), the solid line shows the value of $\beta_1^*$ obtained by solving (4.12) and (4.13) using Newton-Raphson and the points are the average of the partial likelihood estimators from 1000 simulations (sample size $n=1000$). In Figure 5.5, the absolute value is shown. The value of $k_1$ is indicated by the vertical dotted line. Clearly when $k_1 = k_2$, the proportional hazards assumption is not violated and we would have $\beta_1^* = \beta_1$. This corresponds to the common intersection point of the solid line, the dotted horizontal line and the dotted vertical line. For the same values of $\beta_1$, $k_1$, $t_0$ and $\tau$, $k_2$ was increased from -2 to 2 in step size of 0.1. This change in $k_2$ causes the censoring to change and we report the range of levels of censoring for each of the graphs in Figures 5.1-5.3 and 5.5. The lower and upper limits for the reported range always correspond to the cases when $k_2 = 2$ and $k_2 = -2$ respectively.

Weak to moderate positive regression effect of $Z^{(1)}$ ($\beta_1 = 0.5, 1.0$)

For cases where the regression effect was weak or moderate positive (i.e. $\beta_1 = 0.5, 1.0 > 0$), and when $k_1$ and $k_2$ have the same sign (non-crossing hazards) for the covariate $Z^{(2)}$, three observations can be made from Figure 5.1 and Table 5.1-5.2:

• If $k_2 > k_1 \geq 0$, then $\beta_1^* < \beta_1$. That is, if $\beta_2(t)$ increases over time, then the partial likelihood estimator underestimates $\beta_1$. 

19
- If $0 \leq k_2 < k_1$, then $\beta_1^* > \beta_1$. That is, if $\beta_2(t)$ decreases over time, then the partial likelihood estimator overestimates $\beta_1$.
- The absolute value of the bias increases as the absolute difference between $k_1$ and $k_2$ increases.

The case where $k_1$ and $k_2$ have different signs (crossing hazards) is more complex. As can be seen from Figure 5.1, it is possible for $\beta_1^*$ to always underestimate or always over-estimate $\beta_1$ or do a combination of both, depending on the values of $k_1$, $k_2$, $\beta_1$ and censoring.

**Strong positive regression effect of $Z^{(1)}$ ($\beta_1 = 2$)**

The results for cases where the regression effect of $Z^{(1)}$ is strong and positive (i.e. $\beta_1 = 2 > 0$) can be seen from Figure 5.2 and Table 5.3. For the cases where $k_1$ and $k_2$ have different signs (crossing hazards), the bias is typically negative ($\beta_1^* < \beta_1$). For non-crossing hazards, if the regression coefficient of $Z^{(2)}$ decreases ($0 \leq k_2 < k_1$), the bias is typically negative ($\beta_1^* < \beta_1$).

In the few cases where positive bias was observed, the bias was very small. However, if the regression coefficient of $Z^{(2)}$ increases ($0 \leq k_1 < k_2$), the bias can be either positive or negative or both.

**Special case: $k_1 = 0$**

In the case when $k_1 = 0$, that is when the covariate $Z^{(2)}$ has no effect on the hazard up until time $t_0$, regardless of the values of $k_2$, $\beta_1$ and censoring, $\beta_1^*$ always underestimates the true value of $\beta_1$. This is illustrated in Figure 5.3.

**Censoring**

We also studied the effect of censoring on the estimation of $\beta_1$. This was done by varying the values of $t_0$ and $\tau$ while holding the ratio of $t_0$ to $\tau$ fixed at 0.5. An alternative way to vary censoring would have been to hold $t_0$ fixed and vary $\tau$; this was not done, however, because if $t_0$ was held fixed and $\tau$ increased, we would have expected the bias to increase as there would have been more observations that lie in the region $t > t_0$. Thus, we varied both $t_0$ and $\tau$ simultaneously while holding the ratio of $t_0$ to $\tau$ fixed to rule out this effect, so that the change in bias can be attributed to censoring alone. The first two graphs in Figure 5.4 show the case where $\beta_1 = 1$ while the two graphs below it show the case where $\beta_1 = 2$. In the first two graphs the bias is either always positive or always negative, while in the last two graphs, the bias changes sign. In all cases, when censoring decreases, the absolute value of the bias increases until a maximum where it then starts to decrease. In cases where the absolute value of bias decreases to zero at some point, further decrease in censoring will cause the absolute bias to increase (this usually happens when the bias changes sign).
Negative regression effect of $Z^{(1)}$ ($\beta_1 < 0$)

For comparison with the cases of positive regression effect, we also studied cases where the regression effect of $Z^{(1)}$ is negative (for $\beta_1 = -0.5, -1.0, -2.0 < 0$) and the results are plotted in Figure 5.5. The same results as described in the weak to moderate positive regression effect of $Z^{(1)}$ section hold in absolute value for cases where the regression effect is negative, regardless of whether the regression effect is weak, moderate or strong. The results that are described for strong positive regression effect of $Z^{(1)}$ was not observed when the regression effect was negative. In Table 5.4, we compare a few cases where the percentage bias is about the same, but the CI was significantly different. The results show that even when the bias is only around -10%, the CI can drop to as low as 0.371.

Case B. $Z^{(1)}$ binary 0 or 1, $Z^{(2)}$ uniform (0,1)

For comparison, the simulation results for the case where $Z^{(1)}$ is binary 0,1 and $Z^{(2)}$ is uniform (0,1) are shown in Figures 5.6-5.9 and Tables 5.5-5.7. In Figures 5.6-5.9, the dotted line shows the true value of $\beta_1$ (in absolute value for cases where $\beta_1 < 0$) while the results from simulation are shown as dots and are joined by lines. In Figure 5.9, the absolute value is shown. Figure 5.6 and Tables 5.5-5.6 show the case where the regression effect of $Z^{(1)}$ is weak or moderate positive; Figure 5.7 and Table 5.7 show the case where the regression effect of $Z^{(1)}$ is strong positive; Figure 5.8 shows the special case where $k_1 = 0$; Figure 5.9 shows the case where the regression effect of $Z^{(1)}$ is negative. A comparison between Figures 5.1 and 5.6, 5.2 and 5.7, 5.3 and 5.8, 5.5 and 5.9 shows that the trend observed in Case A (both covariates binary 0,1) is also present in Case B ($Z^{(1)}$ binary, $Z^{(2)}$ continuous).
Table 5.1: Weak positive regression effect of $Z^{(1)} (Z^{(2)}$ binary)

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>%cens</th>
<th>$\beta_1^*$</th>
<th>$\beta_2^*$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>CI</th>
<th>%bias</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.1</td>
<td>0</td>
<td>0.5146</td>
<td>0.7010</td>
<td>0.514 (0.068)</td>
<td>0.707 (0.071)</td>
<td>0.938</td>
<td>2.93</td>
<td>0.065</td>
</tr>
<tr>
<td>1.9</td>
<td>0</td>
<td>0</td>
<td>0.4766</td>
<td>1.2066</td>
<td>0.477 (0.063)</td>
<td>1.208 (0.070)</td>
<td>0.946</td>
<td>-4.68</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0</td>
<td>0</td>
<td>0.4992</td>
<td>0.4331</td>
<td>0.502 (0.069)</td>
<td>0.439 (0.070)</td>
<td>0.930</td>
<td>-0.16</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>2.0</td>
<td>24</td>
<td>0.4511</td>
<td>0.7393</td>
<td>0.452 (0.070)</td>
<td>0.739 (0.079)</td>
<td>0.909</td>
<td>-9.79</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>-0.3</td>
<td>23</td>
<td>0.5175</td>
<td>0.9415</td>
<td>0.523 (0.078)</td>
<td>0.951 (0.082)</td>
<td>0.913</td>
<td>3.49</td>
<td>0.073</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>43</td>
<td>0.5056</td>
<td>0.9284</td>
<td>0.507 (0.086)</td>
<td>0.936 (0.085)</td>
<td>0.949</td>
<td>1.11</td>
<td>0.085</td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>41</td>
<td>0.4863</td>
<td>1.1017</td>
<td>0.487 (0.080)</td>
<td>1.104 (0.089)</td>
<td>0.952</td>
<td>-2.74</td>
<td>0.084</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>44</td>
<td>0.5071</td>
<td>0.8918</td>
<td>0.507 (0.087)</td>
<td>0.900 (0.090)</td>
<td>0.943</td>
<td>1.42</td>
<td>0.086</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.2: Moderate positive regression effect of $Z^{(1)} (Z^{(2)}$ binary)

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>%cens</th>
<th>$\beta_1^*$</th>
<th>$\beta_2^*$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>CI</th>
<th>%bias</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.1</td>
<td>0</td>
<td>1.0144</td>
<td>0.7624</td>
<td>1.021 (0.072)</td>
<td>0.773 (0.073)</td>
<td>0.935</td>
<td>1.44</td>
<td>0.071</td>
</tr>
<tr>
<td>1.9</td>
<td>0</td>
<td>0</td>
<td>0.9680</td>
<td>1.1571</td>
<td>0.967 (0.067)</td>
<td>1.162 (0.070)</td>
<td>0.920</td>
<td>-3.20</td>
<td>0.069</td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0</td>
<td>0</td>
<td>1.0007</td>
<td>0.5530</td>
<td>1.001 (0.070)</td>
<td>0.556 (0.068)</td>
<td>0.955</td>
<td>0.07</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>2</td>
<td>21</td>
<td>0.9177</td>
<td>0.5974</td>
<td>0.921 (0.072)</td>
<td>0.595 (0.074)</td>
<td>0.817</td>
<td>-8.23</td>
<td>0.075</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>-0.6</td>
<td>19</td>
<td>1.0216</td>
<td>0.9818</td>
<td>1.018 (0.075)</td>
<td>0.985 (0.080)</td>
<td>0.941</td>
<td>2.16</td>
<td>0.076</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>37</td>
<td>1.0097</td>
<td>0.9487</td>
<td>1.013 (0.083)</td>
<td>0.952 (0.087)</td>
<td>0.953</td>
<td>0.97</td>
<td>0.085</td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>35</td>
<td>0.9808</td>
<td>1.0725</td>
<td>0.984 (0.082)</td>
<td>1.074 (0.083)</td>
<td>0.960</td>
<td>-1.92</td>
<td>0.083</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>38</td>
<td>1.0130</td>
<td>0.9221</td>
<td>1.017 (0.087)</td>
<td>0.931 (0.088)</td>
<td>0.943</td>
<td>1.30</td>
<td>0.085</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tables 5.1-5.4: $\lambda_0(t) = 1$, $\beta_2(t) = k_1$ for $t \leq t_0$ and $k_2$ otherwise, $Z^{(1)}, Z^{(2)}$ independent binary 0,1 with probability $\frac{1}{2}$ each, uniform $(0, \tau)$ censoring, $t_0 = 0.5$. Empirical standard errors from simulations are shown in parentheses. Sample size 1000 with 1000 simulations each. CI represents the proportion of 95% confidence intervals (computed based on proportional hazards assumption) of $\hat{\beta}_1$ that contains $\beta_1$. %bias computed using $\%bias = \frac{\hat{\beta}_1 - \beta_1}{\beta_1} \times 100$. SE gives the average over the standard errors associated with the estimation of $\hat{\beta}_1$ based on the Cox model.
### Table 5.3: Strong positive regression effect of $Z^{(1)}$ ($Z^{(2)}$ binary)

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>%cens</th>
<th>$\hat{\beta}_1^*$</th>
<th>$\hat{\beta}_2^*$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>CI</th>
<th>%bias</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0.1</td>
<td>0</td>
<td>1.9516</td>
<td>0.7968</td>
<td>1.953 (0.076)</td>
<td>0.802 (0.068)</td>
<td>0.935</td>
<td>-2.42</td>
<td>0.085</td>
</tr>
<tr>
<td>1.9</td>
<td>0</td>
<td>2.0</td>
<td>2.0186</td>
<td>1.1279</td>
<td>2.024 (0.089)</td>
<td>1.130 (0.070)</td>
<td>0.937</td>
<td>0.93</td>
<td>0.085</td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0</td>
<td>1.9071</td>
<td>0.6216</td>
<td>1.911 (0.076)</td>
<td>0.627 (0.066)</td>
<td>0.844</td>
<td>-4.64</td>
<td>0.085</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>30</td>
<td>1.9961</td>
<td>0.9663</td>
<td>2.001 (0.088)</td>
<td>0.968 (0.078)</td>
<td>0.964</td>
<td>-0.20</td>
<td>0.093</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>28</td>
<td>2.0045</td>
<td>1.0523</td>
<td>2.013 (0.090)</td>
<td>1.056 (0.082)</td>
<td>0.952</td>
<td>0.23</td>
<td>0.093</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>30</td>
<td>1.9940</td>
<td>0.9500</td>
<td>1.998 (0.092)</td>
<td>0.953 (0.080)</td>
<td>0.948</td>
<td>-0.30</td>
<td>0.093</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5.4: Comparison with negative regression effect of $Z^{(1)}$ ($Z^{(2)}$ binary)

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>%cens</th>
<th>$\hat{\beta}_1^*$</th>
<th>$\hat{\beta}_2^*$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>CI</th>
<th>%bias</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>2.0</td>
<td>24</td>
<td>0.4511</td>
<td>0.7393</td>
<td>0.452 (0.070)</td>
<td>0.739 (0.079)</td>
<td>0.909</td>
<td>-9.79</td>
<td>0.074</td>
</tr>
<tr>
<td>-0.5</td>
<td>1</td>
<td>-1.8</td>
<td>0</td>
<td>-0.4518</td>
<td>-0.1439</td>
<td>-0.451 (0.065)</td>
<td>-0.140 (0.070)</td>
<td>0.898</td>
<td>-9.63</td>
<td>0.065</td>
</tr>
<tr>
<td>-2</td>
<td>1</td>
<td>-1.6</td>
<td>0</td>
<td>-1.8118</td>
<td>-0.5199</td>
<td>-1.812 (0.084)</td>
<td>-0.518 (0.073)</td>
<td>0.371</td>
<td>-9.41</td>
<td>0.082</td>
</tr>
</tbody>
</table>

### Table 5.5: Weak positive regression effect of $Z^{(1)}$ ($Z^{(2)}$ continuous)

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>%cens</th>
<th>$\hat{\beta}_1^*$</th>
<th>$\hat{\beta}_2^*$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>CI</th>
<th>%bias</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.1</td>
<td>0</td>
<td>0.504 (0.066)</td>
<td>0.700 (0.119)</td>
<td>0.948</td>
<td>0.82</td>
<td>0.065</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>0</td>
<td>0.490 (0.064)</td>
<td>1.266 (0.116)</td>
<td>0.949</td>
<td>-2.08</td>
<td>0.065</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0</td>
<td>0.504 (0.067)</td>
<td>0.406 (0.113)</td>
<td>0.943</td>
<td>0.86</td>
<td>0.065</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>0.504 (0.064)</td>
<td>1.266 (0.116)</td>
<td>0.949</td>
<td>-2.08</td>
<td>0.065</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>2</td>
<td>0.502 (0.087)</td>
<td>0.932 (0.145)</td>
<td>0.950</td>
<td>0.35</td>
<td>0.085</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>0</td>
<td>0.502 (0.087)</td>
<td>0.932 (0.145)</td>
<td>0.950</td>
<td>0.35</td>
<td>0.085</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>45</td>
<td>0.507 (0.086)</td>
<td>0.901 (0.146)</td>
<td>0.952</td>
<td>1.32</td>
<td>0.086</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tables 5.5-5.7: $\lambda_0(t) = 1$, $\beta_2(t) = k_1$ for $t \leq t_0$ and $k_2$ otherwise, $Z^{(1)}$ independent binary 0,1 with probability $\frac{1}{2}$ each, $Z^{(2)}$ uniform (0,1), uniform (0, $\tau$) censoring, $t_0 = 0.5$. Empirical standard errors from simulations are shown in parentheses. Sample size 1000 with 1000 simulations each. CI represents the proportion of 95% confidence intervals (computed based on proportional hazards assumption) of $\hat{\beta}_1$ that contains $\beta_1$. %bias computed using $\%\text{bias} = \frac{\hat{\beta}_1 - \beta_1}{\beta_1} \times 100$. SE gives the average over the standard errors associated with the estimation of $\hat{\beta}_1$ based on the Cox model.
Table 5.6: Moderate positive regression effect of $Z^{(1)}$ ($Z^{(2)}$ continuous)

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>%cens</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>CI</th>
<th>%bias</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.1</td>
<td>0</td>
<td>1.009 (0.070)</td>
<td>0.766 (0.116)</td>
<td>0.946</td>
<td>0.93</td>
<td>0.070</td>
</tr>
<tr>
<td>1.9</td>
<td>0</td>
<td>0.983 (0.069)</td>
<td>1.206 (0.115)</td>
<td>0.946</td>
<td>-1.68</td>
<td>0.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0</td>
<td>1.000 (0.073)</td>
<td>0.532 (0.118)</td>
<td>0.940</td>
<td>0.01</td>
<td>0.070</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>2</td>
<td>20</td>
<td>0.967 (0.077)</td>
<td>0.676 (0.129)</td>
<td>0.928</td>
<td>-3.32</td>
<td>0.076</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>-0.6</td>
<td>20</td>
<td>1.008 (0.078)</td>
<td>0.982 (0.125)</td>
<td>0.937</td>
<td>0.84</td>
<td>0.075</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>37</td>
<td>1.010 (0.086)</td>
<td>0.950 (0.142)</td>
<td>0.947</td>
<td>0.95</td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>34</td>
<td>0.996 (0.083)</td>
<td>1.079 (0.137)</td>
<td>0.945</td>
<td>-0.37</td>
<td>0.083</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>38</td>
<td>1.005 (0.085)</td>
<td>0.922 (0.138)</td>
<td>0.953</td>
<td>0.50</td>
<td>0.085</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.7: Strong positive regression effect of $Z^{(1)}$ ($Z^{(2)}$ continuous)

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>%cens</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>CI</th>
<th>%bias</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0.1</td>
<td>0</td>
<td>1.985 (0.084)</td>
<td>0.811 (0.115)</td>
<td>0.950</td>
<td>-0.76</td>
<td>0.087</td>
</tr>
<tr>
<td>1.9</td>
<td>0</td>
<td>2.021 (0.090)</td>
<td>1.167 (0.115)</td>
<td>0.938</td>
<td>1.05</td>
<td>0.087</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0</td>
<td>1.968 (0.083)</td>
<td>0.617 (0.113)</td>
<td>0.945</td>
<td>-1.62</td>
<td>0.086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>30</td>
<td>2.004 (0.090)</td>
<td>0.968 (0.135)</td>
<td>0.959</td>
<td>0.21</td>
<td>0.094</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>28</td>
<td>2.005 (0.093)</td>
<td>1.059 (0.133)</td>
<td>0.955</td>
<td>0.23</td>
<td>0.094</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>30</td>
<td>2.002 (0.092)</td>
<td>0.959 (0.135)</td>
<td>0.959</td>
<td>0.08</td>
<td>0.094</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5.1: Estimation of $\beta_1$ when regression effect of $Z^{(1)}$ is weak or moderate positive ($Z^{(2)}$ binary).
Figure 5.2: Estimation of $\beta_1$ when regression effect of $Z^{(1)}$ is strong positive ($Z^{(2)}$ binary).
Figure 5.3: Estimation of \( \beta_1 \) for the case when \( k_1 = 0 \) (\( Z^{(2)} \) binary).
Figure 5.4: Effect of censoring on estimation of $\beta_1$ ($Z^{(2)}$ binary).
Figure 5.5: Estimation of $\beta_1$ when regression effect of $Z^{(1)}$ is negative ($Z^{(2)}$ binary).
Figure 5.6: Estimation of $\beta_1$ when regression effect of $Z^{(1)}$ is weak or moderate positive ($Z^{(2)}$ continuous).
Figure 5.7: Estimation of $\beta_1$ when regression effect of $Z^{(1)}$ is strong positive ($Z^{(2)}$ continuous).
Figure 5.8: Estimation of $\beta_1$ for the case when $k_1 = 0$ ($Z^{(2)}$ continuous).
Figure 5.9: Estimation of $\beta_1$ when regression effect of $Z^{(1)}$ is negative ($Z^{(2)}$ continuous).
Appendix

Proof of Theorem 2. Under model (4.1) and assumptions (A1-4), the hazard function condition on $Z$ is given by

$$
\lambda(t|Z) = \begin{cases} 
\exp (\beta_1 Z^{(1)} + k_1 Z^{(2)}) & , t \leq t_0 \\
\exp (\beta_1 Z^{(1)} + k_2 Z^{(2)}) & , t > t_0 
\end{cases}
$$

The cumulative hazard function is then found by integrating over the hazard function:

$$
\Lambda(t|Z) = \int_0^t \lambda(u|Z)du
$$

For $t \leq t_0$,

$$
\Lambda(t|Z) = \int_0^t \exp (\beta_1 Z^{(1)} + k_1 Z^{(2)}) du \\
= \exp (\beta_1 Z^{(1)} + k_1 Z^{(2)}) t \\
= \omega_1 t
$$

For $t > t_0$,

$$
\Lambda(t|Z) = \int_0^{t_0} \lambda(u|Z)du + \int_{t_0}^t \lambda(u|Z)du \\
= \int_0^{t_0} \exp (\beta_1 Z^{(1)} + k_1 Z^{(2)}) du + \int_{t_0}^t \exp (\beta_1 Z^{(1)} + k_2 Z^{(2)}) du \\
= \exp (\beta_1 Z^{(1)} + k_1 Z^{(2)}) t_0 + \exp (\beta_1 Z^{(1)} + k_2 Z^{(2)}) (t - t_0) \\
= \omega_1 t_0 + \omega_2 (t - t_0)
$$

Since

$$
S(t) = \exp (-\Lambda(t))
$$

and

$$
F(t) = 1 - S(t)
$$
thus
\[ F_{T|Z}(t) = 1 - \exp(-\Lambda(t|Z)) \]
\[ = \begin{cases} 
1 - \exp(-\omega_1 t), & t \leq t_0 \\
1 - \exp(-\omega_1 t_0 - \omega_2(t - t_0)), & t > t_0 
\end{cases} \]

Proof of Theorem 3. The at-risk indicator is defined as
\[ Y(t) = \begin{cases} 
1, & X \geq t \\
0, & X < t 
\end{cases} \]
\[ E[Y(t)|Z] = 1 \times P(X \geq t|Z) + 0 \times P(X < t|Z) \]
\[ = P(X \geq t|Z) \]
\[ = P(\min(C, T) \geq t|Z) \]
\[ = P(C \geq t, T \geq t|Z) \]
\[ = P(C \geq t|Z)P(T \geq t|Z) \]
\[ = (1 - F_{C|Z}(t)) (1 - F_{T|Z}(t)) \]

where we have used (A5), the independence of \( C \) and \( T \) to obtain the second last line. Under (A5) and (A6), since \( C \) is uniformly distributed on \((0, \tau)\) and independent of \( Z \),
\[ F_{C|Z}(t) = F_C(t) = \frac{t}{\tau} \]

Thus,
\[ E[Y(t)|Z] = \left(1 - \frac{t}{\tau}\right) \left(1 - F_{T|Z}(t)\right) \]

Proof of Theorem 4. From Definition 1 and assumption (A4), we have
\[ S^{(0)}(\beta(t), t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) \exp\left(\beta_i(t)Z_i^{(1)} + \beta_2(t)Z_i^{(2)}\right) \]
\[ S^{(1)}(\beta(t), t) = \begin{pmatrix} 
S^{(1)}_1(\beta(t), t) \\
S^{(1)}_2(\beta(t), t) 
\end{pmatrix} \]
\[ = \begin{pmatrix} 
\frac{1}{n} \sum_{i=1}^{n} Y_i(t) \exp\left(\beta_i(t)Z_i^{(1)} + \beta_2(t)Z_i^{(2)}\right)Z_i^{(1)} \\
\frac{1}{n} \sum_{i=1}^{n} Y_i(t) \exp\left(\beta_i(t)Z_i^{(1)} + \beta_2(t)Z_i^{(2)}\right)Z_i^{(2)} 
\end{pmatrix} \]

35
\[ s^{(0)}(\beta(t), t) = E[S^{(0)}(\beta(t), t)] \]

\[ s^{(1)}(\beta(t), t) = E[S^{(1)}(\beta(t), t)] = \left( \begin{array}{c}
E[S^{(1)}_1(\beta(t), t)] \\
E[S^{(1)}_2(\beta(t), t)]
\end{array} \right) \]

Using (A1), since the random sample is independent and identically distributed

\[
E[S^{(0)}(\beta(t), t)|Z] = E\left[\frac{1}{n} \sum_{i=1}^{n} Y_i(t) \exp \left( \beta_1(t)Z_i^{(1)} + \beta_2(t)Z_i^{(2)} \right) | Z \right]
= \exp \left( \beta_1(t)Z^{(1)} + \beta_2(t)Z^{(2)} \right) E[Y(t)|Z]
\]

\[
E[S^{(1)}(\beta(t), t)|Z] = \left( \begin{array}{c}
E\left[\frac{1}{n} \sum_{i=1}^{n} Y_i(t) \exp \left( \beta_1(t)Z_i^{(1)} + \beta_2(t)Z_i^{(2)} \right) Z_i^{(1)} | Z \right] \\
E\left[\frac{1}{n} \sum_{i=1}^{n} Y_i(t) \exp \left( \beta_1(t)Z_i^{(1)} + \beta_2(t)Z_i^{(2)} \right) Z_i^{(2)} | Z \right]
\end{array} \right)
= \left( \begin{array}{c}
\exp \left( \beta_1(t)Z^{(1)} + \beta_2(t)Z^{(2)} \right) Z^{(1)} E[Y(t)|Z] \\
\exp \left( \beta_1(t)Z^{(1)} + \beta_2(t)Z^{(2)} \right) Z^{(2)} E[Y(t)|Z]
\end{array} \right)
\]

Under (A1), (A4)-(A6), we can use the result from Theorem 3:

\[
E[S^{(0)}(\beta(t), t)|Z] = \exp \left( \beta_1(t)Z^{(1)} + \beta_2(t)Z^{(2)} \right) E[Y(t)|Z]
= \exp \left( \beta_1(t)Z^{(1)} + \beta_2(t)Z^{(2)} \right) \left( 1 - \frac{t}{\tau} \right) \left( 1 - F_{T|Z}(t) \right)
\]

\[
E[S^{(1)}(\beta(t), t)|Z] = \left( \begin{array}{c}
\exp \left( \beta_1(t)Z^{(1)} + \beta_2(t)Z^{(2)} \right) Z^{(1)} E[Y(t)|Z] \\
\exp \left( \beta_1(t)Z^{(1)} + \beta_2(t)Z^{(2)} \right) Z^{(2)} E[Y(t)|Z]
\end{array} \right)
= \left( \begin{array}{c}
\exp \left( \beta_1(t)Z^{(1)} + \beta_2(t)Z^{(2)} \right) Z^{(1)} \left( 1 - \frac{1}{t} \right) \left( 1 - F_{T|Z}(t) \right) \\
\exp \left( \beta_1(t)Z^{(1)} + \beta_2(t)Z^{(2)} \right) Z^{(2)} \left( 1 - \frac{1}{t} \right) \left( 1 - F_{T|Z}(t) \right)
\end{array} \right)
\]

This result can be simplified further by using the result from Theorem 2, which holds under (A1)-(A4) and model (4.1).

\[
E[S^{(0)}(\beta(t), t)|Z] = \exp \left( \beta_1 Z^{(1)} + \beta_2 Z^{(2)} \right) \left( 1 - \frac{t}{\tau} \right) \left( 1 - F_{T|Z}(t) \right)
= \left\{ \begin{array}{ll}
\omega_1 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t\omega_1 \right) , & t \leq t_0 \\
\omega_2 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t_0\omega_1 - (t - t_0)\omega_2 \right) , & t > t_0
\end{array} \right.
\]

\[
E[S^{(1)}_1(\beta(t), t)|Z] = \exp \left( \beta_1 Z^{(1)} + \beta_2 Z^{(2)} \right) Z^{(1)} \left( 1 - F_{T|Z}(t) \right)
= \left\{ \begin{array}{ll}
Z^{(1)}\omega_1 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t\omega_1 \right) , & t \leq t_0 \\
Z^{(1)}\omega_2 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t_0\omega_1 - (t - t_0)\omega_2 \right) , & t > t_0
\end{array} \right.
\]

\]

36
\[ E \left[ S_2^{(1)}(\beta(t), t) | Z \right] = \exp \left( \beta_1 Z^{(1)} + \beta_2(t) Z^{(2)} \right) \left( 1 - \frac{t}{\tau} \right) Z^{(2)} (1 - F_{T|Z}(t)) \]
\[
= \begin{cases} 
Z^{(2)} \omega_1 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t \omega_1 \right), & t \leq t_0 \\
Z^{(2)} \omega_2 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t_0 \omega_1 - (t - t_0) \omega_2 \right), & t > t_0 
\end{cases}
\]

The unconditional expectation of \( S^{(0)} \) and \( S^{(1)} \) is then given by

\[
s^{(0)}(\beta(t), t) = E \left[ S^{(0)}(\beta(t), t) \right] = E_Z \left[ E \left[ S^{(0)}(\beta(t), t) | Z \right] \right] \]
\[
= \begin{cases} 
E_Z [\omega_1 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t \omega_1 \right)], & t \leq t_0 \\
E_Z [\omega_2 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t_0 \omega_1 - (t - t_0) \omega_2 \right)], & t > t_0 
\end{cases}
\]

\[
s^{(1)}_1(\beta(t), t) = E \left[ S^{(1)}_1(\beta(t), t) \right] = E_Z \left[ E \left[ S^{(1)}_1(\beta(t), t) | Z \right] \right] \]
\[
= \begin{cases} 
E_Z [Z^{(1)} \omega_1 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t \omega_1 \right)], & t \leq t_0 \\
E_Z [Z^{(1)} \omega_2 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t_0 \omega_1 - (t - t_0) \omega_2 \right)], & t > t_0 
\end{cases}
\]

\[
s^{(1)}_2(\beta(t), t) = E \left[ S^{(1)}_2(\beta(t), t) \right] = E_Z \left[ E \left[ S^{(1)}_2(\beta(t), t) | Z \right] \right] \]
\[
= \begin{cases} 
E_Z [Z^{(2)} \omega_1 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t \omega_1 \right)], & t \leq t_0 \\
E_Z [Z^{(2)} \omega_2 \left( 1 - \frac{t}{\tau} \right) \exp \left( -t_0 \omega_1 - (t - t_0) \omega_2 \right)], & t > t_0 
\end{cases}
\]


