December 2013

Flexible Partially Linear Single Index Regression Models for Multivariate Survival Data

Na Lei
The University of Western Ontario

Supervisor
Wenqing He
The University of Western Ontario

Graduate Program in Statistics and Actuarial Sciences

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

FLEXIBLE PARTIALLY LINEAR SINGLE INDEX REGRESSION MODELS FOR MULTIVARIATE SURVIVAL DATA

(Thesis format: Monograph)

by

Na Lei

Graduate Program in Statistical and Actuarial Sciences

A thesis submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

The School of Graduate and Postdoctoral Studies

The University of Western Ontario

London, Ontario, Canada

Abstract

Survival regression models usually assume that covariate effects have a linear form. In many circumstances, however, the assumption of linearity may be violated. The present work addresses this limitation by adding nonlinear covariate effects to survival models. Nonlinear covariates are handled using a single index structure, which allows high-dimensional nonlinear effects to be reduced to a scalar term. The nonlinear single index approach is applied to modeling of survival data with multivariate responses, in three popular models: the proportional hazards (PH) model, the proportional odds (PO) model, and the generalized transformation model. Another extension of the PH and PO model is the handling of the baseline function. Instead of modeling it in a parametric way, which is fairly restrictive, or leaving it unspecified, which makes it impossible to calculate the survival and hazard functions, a weakly parametric approach is used here. As a result, the full likelihood can be applied for inference.

The new developments are realized by adding a number of weakly parametric elements to the standard parametric regression models. The marginal baseline hazard functions are modeled using piecewise constants. Marginal survival functions are combined in using copula models, such as the Clayton model, to incorporate association among the multivariate responses. The nonlinear covariate effect is brought into the model through a smooth function with the single-index structure as the input. The smooth function is modeled using a spline.

The performance of the PH, PO, and transformation models with the proposed extensions is evaluated through extensive simulation studies. The PH and PO models are also applied to a real-world data set. The results suggest that the proposed methods can capture the nonlinear covariate effects well, and that there is benefit to modeling the association between the correlated responses. Individual-level survival or hazard function estimates also provide information of interest to researchers. The proposed transformation model in particular is very promising. Some discussion of

how this model may be further developed is provided.

Keywords: Proportional hazard model, proportional odds model, linear transformation model, spline function, partially linear single index model, Clayton model.

Co-Authorship Statement

All material presented in this thesis was completed under the supervision of Dr. Wenqing He. Dr. He provided valuable insight for the ideas behind the material. The majority of the work associated with implementing this research was done by myself.

Dedicated to the memory of Lanying Ma and Lawrence Wolters

Acknowledgements

I would like to express sincere gratitude to my supervisor, Dr. Wenqing He, for his guidance and support throughout my study at UWO. Especially during the time of conducting this thesis, he gave me very valuable suggestions and showed me the way of scientific thinking. It was truly a pleasure to be his student and learn from him. I also want to thank him for his constant support and understanding in my difficult times. Without him, this thesis would not have been completed.

My appreciation goes to Dr. Xuewen Lu, Dr. Bruce Jones, Dr. Mark Reesor and Dr. Xingfu Zou for being the members of my thesis committee. I appreciate them for taking the time to review my thesis and for giving me very insightful comments and advice.

I would like to extend my gratitude to the rest of the faculty and staff members in the Department of Statistical and Actuarial Sciences at the University of Western Ontario for their academic support throughout my study as a PhD student. They enriched my statistics knowledge and broadened my views when I was taking courses and working on projects with them. Special thanks go to Jennifer, Jane and Lisa, for their kind help whenever I need it in my study and life.

I also want to thank my fellow graduate students. It is my pleasure to know you in those years at UWO. Thank you for your help and companionship.

Deep gratitude goes to my family. To my Mum, for your selfless love toward me until the end of your life; my Dad, for always being there whenever I need you; my husband Mark, for your constant support and love; and my daughter Maria, for the happiness you bring to my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Multivariate survival data arise frequently in health and medical studies. Examples include epidemiological cohort studies, in which members from the same family can have the same disease, and the ages at disease occurrence are collected; and clinical trials, in which multiple event times are recorded for each individual. A common characteristic of these data is that the survival times are correlated and it is not appropriate to model them as independent events. To extract the scientifically useful information from such data, it is appropriate to use multivariate, rather than univariate, survival analysis techniques. On many occasions, the collected survival times arise together with some other health related information, such as the gender, age, smoking status etc., which are called covariates. One important motivation of multivariate survival analysis is to investigate the covariate effects on the survival times.

A variety of survival regression models have been developed over time to study the covariate effects. A large body of literature exists for univariate survival data, while the amount of work on the multivariate case is relatively smaller. The existing multivariate regression models often assume linear relationships when exploring covariate effects. But the linearity assumption may be violated in many circumstances.

We work to loosen this assumption and apply weakly parametric methods in multivariate regression models to explore nonlinear covariate relationships and to allow more flexibility to the modeling.

In Section 1.1, real data examples are introduced to motivate the problems of interest considered in this thesis. In Section 1.2, background literature and related work are reviewed.

1.1 Motivation and Problem Description

Approaches used with multiple survival data vary according to different settings. If the survival times are observed in some specified order, it is referred to as the *longitudinal* or *sequential* setting. If the multiple survival times have no prior ordering, it is called the *parallel* setting. In this thesis, we focus on the parallel setting. In the following two sections, examples of sequential data and parallel data are introduced. In Section 1.1.3, the problems of interest, that will be studied in this thesis, are summarized.

1.1.1 Colon Cancer Data

Moertel et al. (1990) conducted a clinical trial on patients with Duke's stage C colon cancer, a cancer stage at which disease recurrence is common after treatment. The trial evaluated the efficacy of two treatments (a particular drug therapy and a placebo). Two times were measured for each patient: time to cancer recurrence and time from cancer recurrence to death. Because the two times are ordered—recurrence must occur before death—this is an example of sequential data. Also, association can be expected between the two responses, since each pair of times is measured on the same individual.

1.1.2 Busselton Health Study Data

The Busselton Health Study (Knuiman et al., 1994) was a repeated cross-sectional survey that was conducted in the Busselton community in West Australia. From 1966 to 1981 a survey was conducted on adults in the community every three years. Various health-related information was collected, such as demographic variables, general health and lifestyle variables, health history variables, and physical, biochemical, haematological and immunological measurements.

This data set includes the health information from 2306 couples who are adults over 18 years old. The survival experience of the individuals, with survival time defined as age at death, is considered here. This is an example of parallel multivariate survival data, as the survival times of the husband and wife are likely to be associated (because of their similar lifestyles), and there is no prior ordering associated with the times. The censoring rate for female is 80%, and 67% for male. Excluding the censored data, the average survival times for female and male are 75.2 and 74.2 respectively.

The data set has over ten health-related covariates, such as age at the beginning of survey (AGE), body mass index (BMI), total cholesterol (CHOL), alcohol consumption level (DRINKING), smoking status (SMOKE), and history of coronary heart disease (CHD). A very brief look at a portion of the data set is found in Table 1.1. The survival time is variable SURVTIME, and the censoring is recorded using the death indicator DTHCENS.

1.1.3 Problem Description

Parallel multivariate survival data can be roughly classified into two groups. The data can arise from different individuals. For example in the Busselton Health Study data, the survival times of two members of a married couple tend to be associated, as one family likely shares a similar diet and lifestyle. Alternatively, correlated data may come from the same individual. For example, in one population, the ith person can

Table 1.1: Example data set

No.	PAIR	SEX	SURVTIME	DTHCENS	AGE	BMI	CHOL	SMOKE	...
1	1	F	76.3	1	50.4	24.61	6.32	1	...
2	1	M	80.4	0	52.3	27.37	6.13	1	...
3	2	F	65.4	0	40.3	26.39	5.13	0	...
4	2	M	65.6	0	40.5	29.54	5.79	1	...
5	3	F	79.9	1	56.5	39.66	6.92	0	...
6	3	M	78.5	1	66.8	23.63	7.11	1	...
...				...					

have different event times $T_{i1}, \ldots, T_{ik}$ corresponding to different diseases. These times can be associated, as they are influenced by the person's health characteristics. Both groups of data can be viewed as the same type of multivariate data. The subjects that are correlated with each other, such as the people from the same family and the diseases from the same individual, can be viewed as one *cluster*. Both types of parallel data can be dealt with using similar statistical approaches.

In most studies, survival data are collected along with some covariates. Researchers are interested in finding the effects of the covariates on the survival time. Another statistical question of interest is how to include the association of the data in the modeling. In addition, being able to calculate the estimated survival function or hazard function, after including the covariate effect in the model, is often required in research. Regression models can be applied to solve these problems.

Many researchers have worked with various regression models for univariate survival data. Two regression models have drawn a lot of attention. One is the proportional hazards (PH) model, and the other is the proportional odds (PO) model. Letting T be the survival time and $\boldsymbol{x} = (x_1, \ldots, x_p)^T$ be the covariate, a PH model is written as:

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{x}),$$

where $\lambda(t|\boldsymbol{x})$ is the hazard function given the effects of covariate $\boldsymbol{x}$, $\boldsymbol{\beta}$ is a $p \times 1$ vector

of unknown parameters, and $\lambda_0(t)$ is the baseline hazard function when $\boldsymbol{x} = \mathbf{0}$ (Cox, 1972). PH models assume the covariates have a linear effect on the log hazards ratio. This assumption is mainly for mathematical convenience, and it may be violated in many situations. Such a linear structure also appears in the PO model. PO models assume that the covariates have a linear effect on the log survival odds ratio:

$$\frac{S(t|\boldsymbol{x})}{1 - S(t|\boldsymbol{x})} = \exp\{\boldsymbol{\beta}^T\boldsymbol{x}\}\frac{S_0(t)}{1 - S_0(t)},$$

where $S(t|\boldsymbol{x})$ is the survival function given the effects of covariate $\boldsymbol{x}$, $\boldsymbol{\beta}$ is the covariate coefficient with dimension $p \times 1$, and S_0 is the baseline survival function when $\boldsymbol{x} = \mathbf{0}$ (Bennett, 1983).

Various ways have been explored to loosen the linearity assumption. One approach is to assume the covariate effect can be described by a flexible smooth function, rather than a rigid log linear relation. To fit the smooth function, nonparametric methods can be applied. In the application of the nonparametric methods there is one problem that often occurs: the so-called *curse of dimensionality*. This term describes the phenomenon where the volume of the covariate space becomes too large to sample thoroughly with any practical number of observations. In this case, the accuracy of the estimation is often poor. To solve such a problem, a single index model may be used to reduce the dimensions of the covariates. Hardle et al. (1993) generalize the linear covariate effect $\boldsymbol{\beta}^T\boldsymbol{x}$ into $\psi(\boldsymbol{\beta}^T\boldsymbol{x})$, where ψ is an unknown univariate function. This is called a single-index model. The single-index model uses the structure $\boldsymbol{\beta}^T\boldsymbol{x}$ as the input of a smooth function, to reduce the high-dimensional variable $\boldsymbol{x}$ into a scalar, and the curse of dimensionality problem can be avoided.

In this thesis, the above ideas are applied to the modeling of multivariate survival data. Three survival regression models are proposed and explored: the PH model, the PO model and the generalized transformation model (see Section 1.2.2). Nonlinear

covariate effects are added in the model by a smooth single index function, to loosen the linearity assumption while reducing the high dimensions of the nonlinear covariate. Spline functions are used to construct the nonlinear smooth function. Piecewise constants are applied to model the baseline hazard function.

1.2 Literature Review

This section reviews regression methods for modeling multivariate survival data. First, methods for incorporating the association among multivariate survival responses are introduced. Then a literature review on survival regression models is given. Extensions on regression models, such as nonparametric approaches, are summarized at the end.

1.2.1 Modeling Association in Multivariate Survival Data

One important question in multivariate survival data modeling is how to incorporate the association. The common approaches may be distinguished as: marginal models, frailty models and copula models.

The marginal model approach makes inferences about the marginal distributions while treating the dependence among the subjects within one cluster as unspecified. Indeed, it takes into account the dependence only while looking at the variance estimates of the parameters. Therefore, parameters can be estimated from the marginal model using the likelihood by assuming independence among all subjects, which is the product of all the marginal likelihoods over all subjects. This likelihood is called the Independence Working Model (IWM) (Huster et al., 1989).

There are two main issues arising from the IWM. One problem is the consistency of the parameter estimates. It can be demonstrated that under certain conditions the maximum likelihood estimate of the parameters obtained from the IWM is a

consistent estimate in spite of the fact that observations are correlated. The other one is for the appropriate variance estimators of parameters. The variance-covariance structure of the data needs to be taken into account to arrive at good estimators for the variances of the estimated parameters.

Cox proportional hazards model have been considered in the marginal model approach by many researchers. Wei, Lin, and Weissfeld (1989) and Lee, Wei, and Amato (1992) have done such related work. The former allows the baseline hazard functions to be different among the marginal models while the latter assumes a common baseline hazard function. Lin (1994) summarizes and continues the work of both of the previous two methods, and develops simple estimating equations which yield consistent and asymptotically normally distributed estimators. The work also includes robust variance-covariance estimators to account for the intra-class correlation. In recent years, nonparametric methods have been explored in the marginal model. Yu and Lin (2008) use kernel estimating equations to estimate nonparametric covariate effects. They show the nonparametric kernel estimator is consistent for any arbitrary working correlation matrix and its asymptotic variance is minimized by assuming working independence.

The second approach to multivariate modeling is frailty models. Frailty models are random effect models, which account for heterogeneity caused by unmeasured covariates. The term frailty is introduced by Vaupel et al. (1979) in univariate survival models and the model is substantially promoted by its application to multivariate survival data by Clayton (1978) in a study of chronic disease incidence in families. Hougaard (2000) provides broad discussions on this topic. Following the definition from Lawless (2003), the common approach in frailty models is to define a random vector $\boldsymbol{\alpha}_i$ associated with the ith subject, and to assume that (1) given $\boldsymbol{\alpha}_i$ and

covariate $\boldsymbol{x}_i$, the survival times $T_{i1}, \ldots, T_{ik}$ are independent, with the survival function

$$S_{ij}(t_j) = \Pr(T_{ij} \geq t_j | \boldsymbol{x}_i, \boldsymbol{\alpha}_i) \quad j = 1, \ldots, k;$$

and (2) the $\boldsymbol{\alpha}_i$ are independent and identically distributed across all subjects $i = 1, \ldots, n$. Assuming $\boldsymbol{\alpha}_i$ has distribution function $G(\boldsymbol{\alpha}_i; \phi)$, the joint survival function for $T_{i1}, \ldots, T_{ik}$ is

$$S(t_1, \ldots, t_k | \boldsymbol{x}_i) = \int \left[\prod_{j=1}^{k} \Pr(T_{ij} \geq t_j | \boldsymbol{x}_i, \boldsymbol{\alpha}_i) \right] dG(\boldsymbol{\alpha}_i; \phi),$$

where ϕ measures the dependence within the cluster. A common choice of the distribution of $\boldsymbol{\alpha}_i$ is the gamma distribution. More general choices for this distribution are discussed by Hougaard (2000).

Another approach to multivariate modeling is copula models. For simplicity, the bivariate case is used as an example. A copula function is defined as a bivariate function C: $[0,1]^2 \to [0,1]$, which satisfies the following three properties (Shemyakin and Youn, 2006):

1. $C(u, 0) = C(0, u) = 0$ for any $u \in [0, 1]$.

2. $C(u, 1) = C(1, u) = u$ for any $u \in [0, 1]$.

3. For all $0 \leq u_1 \leq u_2 \leq 1$ and $0 \leq v_1 \leq v_2 \leq 1$,

$$C([u_1, v_1] \times [u_2, v_2]) = C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

Therefore, when the arguments of the copula function are univariate survival functions $S_1(t_1) = P(T_1 > t_1)$ and $S_2(t_2) = P(T_2 > t_2)$, the copula function $C(S_1, S_2)$ is a bivariate survival function $S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$, with marginal survival functions $S_1(t_1)$ and $S_2(t_2)$.

Development of the joint survival function can be done through the specification of a parametric copula function $C(u, v; \phi)$ and the specification of the marginal distributions. Parameter ϕ determines the association structure of the data. The marginal survival functions $S_1(t_1)$ and $S_2(t_2)$ can have parametric or semiparametric forms.

One such family of copula models is introduced by Clayton (1978). It has the form

$$S(t_1, t_2) = \left[S_1(t_1)^{-\phi^{-1}} + S_2(t_2)^{-\phi^{-1}} - 1\right]^{-\phi}. \tag{1.1}$$

where $\phi \in (0, \infty)$. The larger the value of ϕ is, the smaller the association among the data is. When $\phi = \infty$, the two survival times are independent. The range of ϕ can be extended to -1 to accommodate some negative association (Lawless, 2003).

Note that for copula models the parameter ϕ only controls the association but does not enter into the marginal distributions, while in frailty models ϕ not only controls the association, but also affects the marginal distributions. A frailty model is more appropriate for designed studies where different treatments or covariate factor levels are assigned to individuals within a cluster.

1.2.2 Univariate Regression Models for Survival Data

The widely used survival regression models are the proportional hazards (PH) model, the accelerated failure time (AFT) model, and the proportional odds (PO) model. In recent years, the linear transformation model has also drawn many researchers' attention. We will discuss each of these models in turn.

Cox (1972) proposes the Cox proportional hazards function with the following form:

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{x}), \tag{1.2}$$

where $\lambda_0(t)$ is the baseline hazard function, $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)^T$ is the vector of explanatory variables for a particular individual, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is the

vector of regression coefficients. The model makes no assumptions on the baseline hazard function $\lambda_0(t)$, but assumes a parametric form for the effect of the predictors on the hazard. Therefore, it is referred to as a semi-parametric model. To make inference, the partial likelihood approach (Cox, 1975) can be applied. The beauty of this model is that the vagueness of the baseline hazard function creates no problem for estimation.

Note that covariates act multiplicatively on the hazard function in the Cox PH model. One assumption of the Cox model is that the hazard of the event in one group is proportional to the hazard in any other. To apply this model properly the proportionality assumption needs to be satisfied.

The PH model can be generalized to have the form

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t)r(\psi(\boldsymbol{x})),$$

where $r(\cdot)$ is a positive function, and $\psi(\boldsymbol{x})$ is usually assumed to be of linear regression form $\psi(\boldsymbol{x}) = \boldsymbol{\beta}^T\boldsymbol{x}$. Conventionally, $r(\cdot) = \exp(\cdot)$ can be applied to satisfy the positivity constraint. When $\lambda_0(t)$ is unspecified, it is referred to as the Cox model. When $\lambda_0(t)$ is assumed to follow a specific distribution, it is called a parametric PH model. More generally, $\lambda_0(t)$ can be specified as a piecewise constants or a spline function (He and Lawless, 2003), in which case the full likelihood approach can be applied.

Another popular model for survival data is the Accelerated Failure Time (AFT) model. The AFT model usually can be written as a log-linear model for failure time T according to:

$$\log T = \mu + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p + \sigma\epsilon.$$

where μ is the intercept, σ is the scale parameter, and ϵ is a random variable assumed to have a particular distribution. The survival function of T can be expressed through

the survival function of ϵ:

$$\begin{aligned} S(t) &= P(T \geq t) \\ &= P\left(\log T \geq \log t\right) \\ &= P\left(\mu + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p + \sigma\epsilon \geq \log t\right) \\ &= P\left(\epsilon \geq \frac{\log t - \mu - \boldsymbol{\alpha}^T \boldsymbol{x}}{\sigma}\right) \\ &= S_\epsilon\left(\frac{\log t - \mu - \boldsymbol{\alpha}^T \boldsymbol{x}}{\sigma}\right) \end{aligned}$$

Another way of expressing the AFT model is

$$S(t|\boldsymbol{x}) = S_0(\eta(\boldsymbol{x})t),$$

where $S_0(\cdot)$ is the baseline survival function, and $\eta(\cdot)$ is an "acceleration factor" which depends on the covariates $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)^T$ through the formula:

$$\eta(\boldsymbol{x}) = \exp(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p).$$

In this model the effect of a covariate is to stretch or shrink the survival curve along the time axis by a constant relative amount $\eta(\boldsymbol{x})$.

In AFT models, it is often necessary to specify the distribution of failure time. The common distributions used for this purpose are Weibull, Log-normal, Log-logistic, and Gamma. Some other approaches have also been discussed by researchers. Wei (1992) reviews some estimation methods applied to the AFT model. For example, an estimation procedure based on rank test statistics, and an approach using least squares principle proposed by Buckley and James (1979) are discussed. Recently, Jin, Lin, Wei, and Ying (2003) and Jin, Lin, and Ying (2006) develop approximations to the rank based estimator and least squares estimator through linear programming.

Resampling procedures are applied for the estimation of the limiting covariance matrix.

A third popular approach for survival data modeling is the Proportional Odds (PO) model, which is introduced by Bennett (1983). This model uses the survival odds ratio as a measure of relative risk in the regression model. It can be expressed as:

$$\frac{S(t|\boldsymbol{x})}{1-S(t|\boldsymbol{x})} = \exp(\boldsymbol{\beta}^T\boldsymbol{x})\frac{S_0(t)}{1-S_0(t)}, \tag{1.3}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)^T$ is the vector of explanatory variables for an individual, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is the vector of regression coefficients, and $S_0(t)$, the baseline survival function, is the survival function for an individual whose explanatory variables all take the value zero.

One important property of PO model is that the hazards ratio converges from the value $\exp(-\boldsymbol{\beta}^T\boldsymbol{x})$ at time $t = 0$, to unity at $t = \infty$. It can be shown that (Collett, 2003)

$$\frac{\lambda(t|\boldsymbol{x})}{\lambda_0(t)} = [1 + (\exp(\boldsymbol{\beta}^T\boldsymbol{x}) - 1)S_0(t)]^{-1}.$$

Therefore, when $t = 0$, $S_0(t) = 1$, and $\frac{\lambda(t|\boldsymbol{x})}{\lambda_0(t)} = \exp(-\boldsymbol{\beta}^T\boldsymbol{x})$; when $t = \infty$, $S_0(t) = 0$, and $\frac{\lambda(t|\boldsymbol{x})}{\lambda_0(t)} = 1$. As a contrast, in the PH model, the hazards ratio remains constant at t. Recall that for the PH model,

$$\frac{\lambda(t|\boldsymbol{x})}{\lambda_0(t)} = \exp(\boldsymbol{\beta}^T\boldsymbol{x}),$$

but this assumption can be unreasonable in certain circumstances. For example, initial effects such as differences in the stage of disease or in treatment can disappear with time. In this case, the property of PO model that the hazards ratio converges to 1 as t increases to ∞ makes more sense.

Bennett (1983) describes how the PO model can be fitted using maximum likelihood estimation. Kirmani and Gupta (2001) explores the structure, implications and

properties of the PO model. Many researchers have discussed various approaches for making inferences on PO model. Cheng et al. (1995) propose a modified V statistic for parameter estimation. Murphy et al. (1997) show the profile likelihood approach of Bennett (1983) results in an asymptotically efficient regression estimator. Yang and Prentice (1999) introduce some weighted empirical odds functions. Several classes of regression estimators such as the psedo-maximum likelihood estimator, martingale residual-based estimators, and minimum distance estimators are derived.

The final model considered here is the linear transformation model. Let T be the survival time and $\boldsymbol{x}$ be the covariate with dimension $p \times 1$. The linear transformation model assumes that

$$H(T) = -\boldsymbol{\beta}^T \boldsymbol{x} + \epsilon, \tag{1.4}$$

where H is an increasing transformation function, ϵ is a random variable with a known distribution, and $\boldsymbol{\beta}$ is the covariate coefficient with dimension $p \times 1$. The proportional hazards model and the proportional odds model are special cases of (1.4) with ϵ following the extreme-value distribution and the standard logistic distribution, respectively.

An alternative way to show that the PH and PO models are the special cases of the transformation model is though the hazard function of ϵ. If we let the hazard function of ϵ have the form

$$\lambda(\epsilon) = \frac{\exp(\epsilon)}{1 + r\exp(\epsilon)},$$

then $r = 0$ corresponds to a PH model, and $r = 1$ corresponds to a PO model. More discussion related to this topic is given in Section 4.1.

Some inference procedures for this model have been proposed by various researchers. Cheng et al. (1995) study a class of generalised estimating equations to examine the covariate effects with censored observations. This method is further developed in Cheng et al. (1997), Fine et al. (1998) and Cai et al. (2000). A key

assumption in their approach is that the censoring variable is independent of the covariates, which makes it possible to use the Kaplan-Meier method to estimate the survival function. Chen et al. (2002) relax this assumption and propose an estimating equation approach to make inference. Zeng and Lin (2006) study nonparametric maximum likelihood estimation in a class of semiparametric transformation models which considers time dependent covariates. More recently, Lu and Zhang (2010) propose a partially linear transformation model by incorporating nonlinear covariate effects, and studied a martingale-based estimating equation approach to make inference.

1.2.3 Extensions on Regression Models

A lot of work has been done on extensions of regression models. As mentioned previously, one of the traditional assumptions of regression models is that the covariates have a linear effect on the log hazards ratio, log survival odds ratio, or other quantity of interest. However, this linearity assumption is mainly for mathematical convenience, and might not be valid in many circumstances. A smooth function $\psi(\cdot)$ has been proposed to generalize the linearity assumption $\boldsymbol{\beta}^T\boldsymbol{x}$. In the exploration of the smooth function, nonparametric or the weakly parametric methods have drawn much attention. Tibshirani and Hastie (1987) use the linear regression method on a defined local neighborhood of each x. This local likelihood estimation approach is applied to find the optimal estimator of the local regression coefficients. Note that this local linear regression model can also be replaced by some other smooth function. Fan, Gijbels, and King (1997) discuss the estimation of the nonparametric covaritate effect in the PH model when the baseline function is parametric and nonparametric. When the baseline function and the covariate effect are both unspecified, nonuniform kernel methods are applied to fit the covariate effect and the inference is based on a local version of partial likelihood. Spline functions have also been considered by some researchers. Kooperberg et al. (1995) explore polynomial spline functions to study the

nonparametric covariate effect in PH models. An automatic procedure involving the maximum likelihood method, stepwise addition, stepwise deletion and BIC is used to select the final model.

A considerable amount of work has been done to address the curse of dimensionality problem under the univariate data modeling framework when the covariates have high dimension. One approach is to use additive models. Hastie and Tibshirani (1990) propose to use a sum of smooth functions over the components of covariates. That is, instead of using the linear model $\sum_{j=1}^{p} x_{ij}\beta_j$, they suggest using an additive term $\sum_{j=1}^{p} f_j(x_{ij})$, where $x_{i1}, \ldots, x_{ip}$ are covariate values for the ith individual. The $f_j(\cdot)$ are unspecified smooth functions that are estimated using scatter plot smoothers. The other approach that has drawn a lot of attention is the single index model. Hardle et al. (1993) generalize the linear covariate effect $\boldsymbol{\beta}^T\boldsymbol{x}$ into $\psi(\boldsymbol{\beta}^T\boldsymbol{x})$, where $\boldsymbol{x}$ is a $p \times 1$ vector of covariates, $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters, and ψ is an unknown univariate function. This single index model reduces the $p \times 1$ vector of covariates $\boldsymbol{x}$ from higher dimensions to a scalar $\boldsymbol{\beta}^T\boldsymbol{x}$, and then treats the smooth function $\psi(\cdot)$ as a univariate function of $\boldsymbol{\beta}^T\boldsymbol{x}$. Note the scale of $\boldsymbol{\beta}^T\boldsymbol{x}$ in $\psi(\boldsymbol{\beta}^T\boldsymbol{x})$ may be determined arbitrarily, so $\boldsymbol{\beta}$ may be replaced by the unit vector $\boldsymbol{\alpha} = \boldsymbol{\beta}\|\boldsymbol{\beta}\|^{-1}$, where $\|\cdot\|$ denotes the Euclidean metric. Then we can estimate both $\psi(\cdot)$ and $\boldsymbol{\alpha}$ in the model.

Because of the useful characteristic of the single index model that it can reduce the the high dimension of covariates to a scalar, it has drawn a lot of attention from many researchers. Carroll, Fan, Gijbels, and Wand (1997) apply it to the generalized linear model and introduces a more general model which is called "Generalized Partially Linear Single Index Model (GPLSIM)". In this model a response Y is to be predicted by covariates $(\boldsymbol{X}, \boldsymbol{V})$, where $\boldsymbol{X}$ (of length p) is to enter the model linearly and $\boldsymbol{V}$ (of length q) is to enter it nonlinearly. GPLSIM has the following form:

$$g\{\mu(\boldsymbol{x}, \boldsymbol{v})\} = \boldsymbol{\beta}^T\boldsymbol{x} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}), \quad \text{with} \quad \|\boldsymbol{\alpha}\| = 1,$$

where $\mu(\boldsymbol{x}, \boldsymbol{v}) = E(Y)$, g is the link function, ψ is an unknown smooth function, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are coefficients of dimensions q and p. The single index structure $\psi(\boldsymbol{\alpha}^T\boldsymbol{v})$ is added to extend the linear term $\boldsymbol{\beta}^T\boldsymbol{x}$ which the generalized linear model normally has, and makes it more flexible to describe a nonlinear relationship between the covariate $(\boldsymbol{x}, \boldsymbol{v})$ and $g\{\mu(\boldsymbol{x}, \boldsymbol{v})\}$. Kernel methods are used to explore the shape of the smooth function ψ nonparametrically.

In recent years, structures similar to GPLSIM have been applied in some survival models. For example, Lu, Chen, Singh, and Song (2006) introduce the partially linear single index structure under the proportional hazards regression model framework, and define a "Partially Linear Single-Index Survival Model (PLSISM)". More information about this model can be seen in Section 2.1.

The above work mainly focuses on the extensions of regression models for univariate survival data. There is still limited work for regression modeling of multivariate survival data, however, especially on the exploration of nonparametric methods applied in regression models. He and Lawless (2003) use piecewise constant or spline functions to fit the baseline hazard function in either marginal or conditional proportional hazards models. The copula model is proposed to incorporate the association among the multivariate survival data. As the baseline hazard function has a specified form through the parametric approach, full likelihood can be applied to make inference. Yu and Lin (2008) study multivariate survival data through the marginal model approach. They propose the marginal proportional hazards function and use kernel methods in the regression model. They show that the nonparametric kernel estimator is consistent for any arbitrary working correlation matrix and its asymptotic variance is minimized by assuming working independence. These works have given inspiring ideas and thoughts for multivariate survival data modeling.

1.3 Overview of the Thesis

The focus of this thesis is on multivariate survival data regression modeling. Copula functions are proposed to include the association among the multivariate survival data, with covariate effects incorporated through the marginal model. For the marginal distributions, three types of regression models are discussed: (i) the proportional hazards function, (ii) the proportional odds model, (iii) the generalized transformation model. Traditionally, these models assume the covariates have linear effects in the regression part of the model. We relax the linearity assumption and allow both the linear and the nonlinear relationships in the model. A smooth function is used to explore the nonlinear relationship. The single index model is added to reduce the dimensions of the nonlinear covariates into a scalar. Weakly parametric methods are applied to explore the smooth function. In the marginal hazard function, piecewise constants are used to estimate the baseline hazard function. Based on the above setup of the model, the full likelihood function can be applied to make inference.

The remaining chapters of the thesis are organized as follows. In Chapter 2 the partially linear single index proportional hazards regression model for multivariate survival data is investigated. The proposed model is assessed by simulation studies. The model is applied to the real data example, Busselton Health Study, for illustration. In Chapter 3 the partially linear single index proportional odds regression model is explored. Simulation studies and real data analysis are also provided. Chapter 4 examines the partially linear single index generalized transformation model. As the PH model and the PO model are two special cases of the generalized transformation model, the third proposed model is a generalization of the first two. Simulation studies are provided. Some conclusions are discussed at the end of Chapter 4. Discussions and future work are given in Chapter 5.

Chapter 2

Flexible Partially Linear Single Index Proportional Hazard Regression Model for Multivariate Survival Data

2.1 Introduction

The proportional hazards model (1.2) has been a very popular survival model since it was proposed by Cox (1972). One of the assumptions of the PH model is that the covariate $\boldsymbol{x}$ has a linear effect on the log hazards ratio. This assumption is mainly for mathematical convenience, however, may not hold in many situations. Many methods have been proposed to add a smooth function $\psi(\cdot)$ to relax the linearity assumption, letting the covariate effect be written as $\boldsymbol{\beta}^T\boldsymbol{x} + \psi(\boldsymbol{v})$. In this expression $\boldsymbol{x}$ is a vector of covariates that enter the model linearly, and $\boldsymbol{v}$ is a vector of covariates that enter in a nonlinear fashion. This is called a partially linear model.

In the specification of the smooth function $\psi(\cdot)$, nonparametric methods have

drawn much attention. However, if $\boldsymbol{v}$ is high-dimensional covariate vector, the curse of dimensionality problem (described in Section 1.1.3) makes it hard to achieve accurate estimation for covariate effect $\psi(\boldsymbol{v})$ using nonparametric methods.

One approach to solve this issue is through the use of single index models (Hardle et al., 1993). In this approach, a new parameter vector $\boldsymbol{\alpha}$ is introduced, and the inner product $\boldsymbol{\alpha}^T\boldsymbol{v}$ is taken to reduce the covariate effects to a scalar, to summarize covariates first. The smooth function $\psi(\cdot)$ is then taken to be a function of this scalar argument. This particular type of partially linear model, with the covariate effect taking the form $\boldsymbol{\beta}^T\boldsymbol{x} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v})$, is known as a partially linear single index model.

Partially linear single index models have been explored in the literature. For example, Lu, Chen, Singh, and Song (2006) propose the Partially Linear Single-Index Survival Model (PLSISM). The hazard function has the form:

$$\lambda(t|\boldsymbol{x},\boldsymbol{v}) = \lambda_0(t)\exp\{\boldsymbol{\beta}^T\boldsymbol{x} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v})\}, \quad \text{with} \quad \|\boldsymbol{\alpha}\| = 1,$$

where $\lambda_0(t)$ is the baseline hazard function, $\psi(\cdot)$ is the unknown smooth function, $\boldsymbol{x}$ is the linear covariate with dimension p, $\boldsymbol{v}$ is the nonlinear covariate with dimension q, and $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are their corresponding coefficients. The authors use kernel methods to model the smooth function. The profile quasi-likelihood approach is applied to make inference. Sun, Kopciuk, and Lu (2008) also apply a similar partially single index structure in proportional hazards regression. They use spline functions to approximate the smooth function. In their model, the baseline hazard function is not specified, and the partial likelihood approach is applied for inference.

The above work has been done under the univariate response framework. We propose to apply the partially linear single index structure in the marginal hazard function for multivariate data. As has been introduced in Section 1.2.1, a copula model can be used to incorporate the data association among multivariate survival

data. Without loss of generality, a special case of multivariate data–bivariate data–is considered here. Specifically, the Clayton model (Clayton, 1978) is employed to link the marginal survival functions for illustration.

To make inference, two parts of the model still need to be specified. One is the baseline hazard function, and the other one is the smooth function in the single index structure. The objective we consider here is to model these functions without making overly strong parametric assumptions, while also choosing a form that is flexible enough. To do this, two weakly parametric approaches, piecewise constants (He and Lawless, 2003) and spline functions, are proposed to model the baseline hazard functions and the smooth function respectively. This model structure has finite number of parameters, and therefore the full likelihood method can be invoked for inference.

The details of the proposed model framework is introduced in Section 2.2. Simulation studies are conducted in Section 2.3, to assess the performance of the proposed method, to compare the proposed model to the model where the nonlinear covariate effect is ignored, and to evaluate the impact of other factors on the estimation under various scenarios. The Busselton health study is used as an illustrative example. Some conclusions about the proposed model are summarized at the end.

2.2 The Proposed Model

Assume the survival data sample has m clusters. Let T_{ij} and C_{ij} be the failure and censoring times of the jth observation in the ith cluster ($j = 1, \ldots, k$, $i = 1, \ldots, m$). The observed data $(y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{v}_{ij}, \delta_{ij})$ are realizations of the variables $(Y_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{V}_{ij}, \Delta_{ij})$, where $Y_{ij} = \min(T_{ij}, C_{ij})$ is the observed time, and $\Delta_{ij} = I(T_{ij} \leq C_{ij})$ is the censoring indicator. The covariates are $\boldsymbol{X}_{ij}$ and $\boldsymbol{V}_{ij}$, which are assumed to have linear and nonlinear effects, respectively. The survival times T_{ij} within each cluster are assumed

correlated, while the observations from different clusters are assumed to be independent. For illustration purpose, we assume $k = 2$, to limit the survival distribution to be bivariate. The proposed marginal PH function is written as:

$$\lambda_j(t_j|\boldsymbol{x}_j, \boldsymbol{v}_j) = \lambda_{0j}(t_j)\exp\{\boldsymbol{\beta}^T\boldsymbol{x}_j + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_j)\}, \quad j = 1,\ 2,$$

where $\boldsymbol{\beta}$ is the linear covariate coefficient with dimension p and $\boldsymbol{\alpha}$ is the nonlinear covariate coefficient with dimension q, $\psi(\cdot)$ is a smooth function, and $\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_j)$ is the single index model.

To incorporate the association in the model, a copula model is used for multivariate modeling (see Section 1.2.1). More specifically, the Clayton model (Clayton, 1978) is used, and the joint survival function is given by (1.1).

Let $f(t_1, t_2)$ denote the joint density function. The likelihood contributed from the ith cluster is,

$$\begin{aligned} L_i &= f(t_{i1}, t_{i2})^{\delta_{i1}\delta_{i2}} \left[\frac{-\partial S(t_{i1}, t_{i2})}{\partial t_{i1}}\right]^{\delta_{i1}(1-\delta_{i2})} \\ &\quad \times \left[\frac{-\partial S(t_{i1}, t_{i2})}{\partial t_{i2}}\right]^{(1-\delta_{i1})\delta_{i2}} S(t_{i1}, t_{i2})^{(1-\delta_{i1})(1-\delta_{i2})}, \end{aligned} \tag{2.1}$$

which is also shown in Lawless (2003).

To make inference from the model, we need to write this likelihood explicitly as a function of the parameters. To do this, the estimation approaches for the baseline hazard functions $\lambda_{0j}(t_j)$ and the smooth function $\psi(\cdot)$ need to be specified. We propose to use piecewise constant approach (He and Lawless, 2003) to approximate the baseline hazard functions, and to use a spline function approach to approximate the smooth function $\psi(\cdot)$. These details will be given in the following two subsections.

2.2.1 Piecewise Constant Approach for Baseline Function

In modelling the baseline hazard function $\lambda_{0j}(t_j)$, we follow He and Lawless (2003), to have a piecewise constant function approximation. Compared with the parametric PH model, which assumes that the baseline hazard function has a known parametric form, this method relaxes the parametric assumption of $\lambda_{0j}(t_j)$ to allow flexibility in its shape.

Assume that the marginal baseline hazard functions $\lambda_{0j}(t_j)$, $j = 1, 2$, have piecewise constant forms as follows:

$$\lambda_{01}(t_1) = \rho_k, \quad \text{where } t_1 \in A_k = (a_{k-1}, a_k], \ k = 1, \ldots, r,$$

and

$$\lambda_{02}(t_2) = \tau_l, \quad \text{where } y_2 \in B_l = (b_{l-1}, b_l], \ l = 1, \ldots, s,$$

where $0 = a_0 < a_1 < \ldots < a_r = \infty$ and $0 = b_0 < b_1 < \ldots < b_s = \infty$ are prechosen sequences of constants, which are also called cut points. $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_r)^T$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_s)^T$ are unknown positive constants to be estimated. A_k and B_l are the intervals defined by the sequence of cut points. Therefore, the corresponding marginal cumulative hazard functions are the integration over the piecewise constants, which can be written as:

$$\begin{aligned} \Lambda_{01}(t_1) &= \sum_{k=1}^{r} \rho_k u_k(t_1) \\ \Lambda_{02}(t_2) &= \sum_{l=1}^{s} \tau_l w_l(t_2), \end{aligned}$$

where $u_k(t_1) = \max(0, \min(a_k, t_1) - a_{k-1})$ are the length of the intersection of the interval $(0, t_1)$ with the interval A_k, and $w_l(t_2) = \max(0, \min(b_l, t_2) - b_{l-1})$ are the length of the intersection of the interval $(0, t_2)$ with the interval B_l.

How to choose the cut points is an important question. Normally they are chosen based on prior assumptions about the marginal distributions of the survival time. If we know the hazard function slope changes at some points, these points can be set as the cut points. If such changes can not be detected clearly, one can start from a small number of cut points, say 2 or 3, and increase the number to observe their effect. Normally it is desirable to keep roughly the same number of failures within each interval. For data with a high censoring rate, one useful way to choose the cut points is based on the Kaplan-Meier survival probability without considering the covariate effects. That is, use the cut points that give roughly the same survival probability in each interval. In our simulation study and real data analysis, since the censoring rate can be high, we use the strategy based on survival probability. We choose to use four intervals ($r = s = 4$) to approximate the baseline hazard function, which can give reasonably good estimation (Lawless and Zhan, 1998), and keep the number of parameters in the model to a manageable level.

2.2.2 Spline Function Approximation

Spline methods are used to estimate the smooth function $\psi(\cdot)$. M-spline functions and I-spline functions (Ramsay, 1988) are proposed to approximate $\psi(\cdot)$, because of their convenient characteristics. By definition, M-spline basis functions are nonnegative functions. I-spline basis functions are defined as the integration of the corresponding M-spline basis functions, and have the property of being monotone increasing. In the proposed model we would like to model the smooth function $\psi(\cdot)$ and its derivative $\psi'(\cdot)$. Therefore, it is very natural to use M-spline functions to approximate the derivative $\psi'(\cdot)$, and use I-spline functions to approximate the smooth function $\psi(\cdot)$. Note that while the I-spline basis functions are monotone, the I-spline itself will not be monotone unless the coefficients are restricted to be positive. In the proposed model no such restriction is included, because we desire the function $\psi(\cdot)$ to have a

flexible shape.

A spline function is a series of polynomials joined smoothly at some break points. Given an interval $[L, U]$ and the breakpoints $\eta = \{\eta_0, \ldots, \eta_k\}$, where $L = \eta_0 < \cdots < \eta_k = U$, there is one corresponding polynomial function P_j of order m or degree m-1 on each subinterval $[\eta_j, \eta_{j+1})$ called a basis spline function. The spline function is defined as $f(x) = \sum_{j=1}^{k} c_j P_j(x)$, where c_j is the corresponding coefficient for each basis spline function. At the boundaries of the intervals, a certain degree of smoothness is required. The smoothness is defined by the equality of the derivatives of P_j. A popular choice is to let the derivatives be continuous up to order two at the break points, as our human eyes are capable to detect the discontinuity no higher than the second order.

The M-spline and I-spline functions are both defined based on a sequence of "knots". Once the breakpoints for the interval $[L, U]$ are given, $\eta = \{\eta_0, \ldots, \eta_k\}$, the sequence of knots $c = \{c_1, \ldots, c_{k+2m-1}\}$ are constructed as

$$c_1 = \ldots = c_m = \eta_0, \quad c_{k+m} = \ldots = c_{k+2m-1} = \eta_k,$$

and

$$c_{m+j} = \eta_j, \quad j = 1, 2, \ldots, k-1.$$

Let $M_i(x|1, c)$ be the ith M-spline basis function of order m with knot sequence c. These basis functions are defined recursively:

$$M_i(x|1, c) = \begin{cases} \frac{1}{c_{i+1} - c_i} & \text{if } c_i \leq x < c_{i+1}, \\ 0 & \text{otherwise,} \end{cases}$$

and for $m > 1$, :

$$M_i(x|m,t) = \frac{m[(x-c_i)M_i(x|m-1,c) + (c_{i+m}-x)M_{i+1}(x|m-1,c)]}{(m-1)(c_{i+m}-c_i)}.$$

One main characteristic of M-spline function is that the basis spline functions are positive within the interval (c_i, c_{i+m}) and zero otherwise. In total there are $k + m - 1$ of them for spline function of order m. Figure (2.1a) gives one example of a group of cubic M-spline basis function on the interval of (0,1), and the breakpoints (0.3,0.5,0.6). The basis spline functions are M_1, M_2, ..., and M_6. The function $f(x) = 1.2M_1 + 2.0M_2 + 1.2M_3 + 1.2M_4 + 3.0M_5 + 0.0M_6$, is represented by the top dotted line.

I-spline functions are defined based on the M-spline functions. The basis functions are the integration of the corresponding M-spline basis functions:

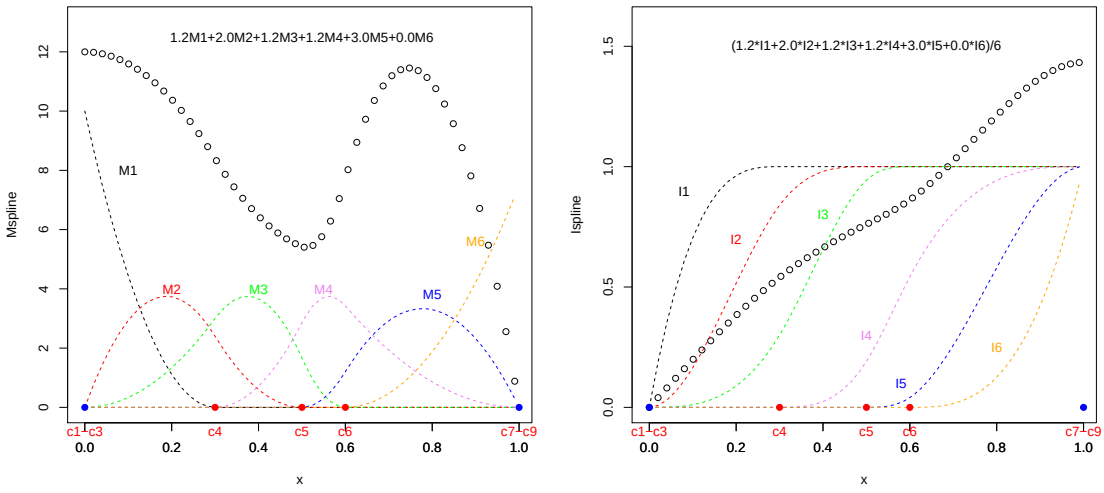
$$I_i(x|m,c) = \int_L^x M_i(u|m,c)du.$$

The definition can be expressed in the following form:

$$I_i(x|m,c) = \begin{cases} 0 & i > j, \\ \sum_{n=i}^{j} (c_{n+m+1} - c_n) M_n(x|m+1,c)/(m+1), & j-m+1 \le i \le j, \\ 1 & i < j-m+1. \end{cases}$$

Figure (2.1b) shows the I-spline function corresponding to Figure (2.1a). The I-spline basis functions I_1, ..., I_6 are shown as dashed lines. They are the integrations of the corresponding M-spline basis functions M_1, ..., M_6. The I-spline function $g(x) = (1.2I_1 + 2.0I_2 + 1.2I_3 + 1.2I_4 + 3.0I_5 + 0.0I_6)/6$, is represented by the dotted line.

We propose to use M-spline functions to fit the derivative of $\psi(\cdot)$, $\psi'(\cdot)$. Let



(a) M-spline (b) I-spline

Figure 2.1: Spline functions

M_j, $j = 1, \ldots, d$, be the M-spline basis functions, and $M(u) = (M_1(u), \ldots, M_d(u))^T$. Then $\psi'(\boldsymbol{\alpha}^T \boldsymbol{v})$ can be written as:

$$\psi'(\boldsymbol{\alpha}^T \boldsymbol{v}) = \sum_{j=1}^{d} \gamma_j M_j(\boldsymbol{\alpha}^T \boldsymbol{v}) = \boldsymbol{\gamma}^T M(\boldsymbol{\alpha}^T \boldsymbol{v}),$$

where γ_j is the coefficient for the jth basis function.

By definition, I-spline functions can be used to fit $\psi(\cdot)$, as the basis function is the integration of the corresponding M-spline basis function. Note that $\psi(0)$ is constrained to be 0. This is because in the regression model $\boldsymbol{\beta}^T \boldsymbol{x} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v})$, we want the intercept to be 0, which implies $\boldsymbol{\beta}^T \boldsymbol{x} = 0$ for $\boldsymbol{x} = \mathbf{0}$ and $\psi(0) = 0$. Using this constraint, $\psi(\boldsymbol{\alpha}^T \boldsymbol{v})$ can be modeled by $\int_0^{\boldsymbol{\alpha}^T \boldsymbol{v}} \psi'(x)dx$, therefore we have:

$$
\begin{aligned}
\psi(\boldsymbol{\alpha}^T\boldsymbol{v}) &= \int_0^{\boldsymbol{\alpha}^T\boldsymbol{v}} \psi'(x)dx \\
&= \int_0^{\boldsymbol{\alpha}^T\boldsymbol{v}} \sum_{j=1}^{d} \gamma_j M_j(x)dx \\
&= \sum_{j=1}^{d} \gamma_j \int_0^{\boldsymbol{\alpha}^T\boldsymbol{v}} M_j(x)dx \\
&= \sum_{j=1}^{d} \gamma_j \int_L^{\boldsymbol{\alpha}^T\boldsymbol{v}} M_j(x)dx - \sum_{j=1}^{d} \gamma_j \int_L^{0} M_j(x)dx \\
&= \sum_{j=1}^{d} \gamma_j \left[I_j(\boldsymbol{\alpha}^T\boldsymbol{v}) - I_j(0)\right] \\
&= \boldsymbol{\gamma}^T \left[I(\boldsymbol{\alpha}^T\boldsymbol{v}) - I(0)\right]
\end{aligned}
$$

Note there are another two constraints in the marginal PH model, and both are related to parameter identifiability. One constraint is, for $\psi(\boldsymbol{\alpha}^T\boldsymbol{v})$ we need $\| \boldsymbol{\alpha} \| = 1$. If the norm of $\boldsymbol{\alpha}$ is not specified, the sets of solutions for $\boldsymbol{\alpha}$ are not unique, and the solutions for $\boldsymbol{\gamma}$ are also not unique. But only with this one constraint, we are still not able to find the unique solution for $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$. The other constraint is to let one particular component of $\boldsymbol{\alpha}$ (for example, its last element) be positive. This is because $\psi^*(-\boldsymbol{\alpha}) = \psi(\boldsymbol{\alpha})$, therefore for ψ^*, $-\boldsymbol{\alpha}$ is the solution, and for ψ, $\boldsymbol{\alpha}$ is the solution. That is, both $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ and $(-\boldsymbol{\alpha}, -\boldsymbol{\gamma})$ are solutions for the smooth function. With these two additional constraints, it is possible to find unique solutions for $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$.

Regarding the strategy of choosing the number of breakpoints, we follow the advice given by Lawless and Zhan (1998), who suggest to use 4-10 intervals. In the simulation study and the real data analysis, in order to satisfy the goal of having good model performance and not too long of a simulation time, 4 intervals (i.e.,3 breakpoints) are chosen in spline functions. Each interval contains roughly the same number of data points.

2.2.3 The Likelihood

Now that the baseline hazard function $\lambda_{0j}(t_j)$ and the smooth function $\psi(\cdot)$ have been fully specified, the likelihood as given in equation (2.1) can be worked out. Using the relationship between the survival function and the hazard function, the marginal survival functions can be written as shown below. Note that in the following expressions we write $S_{ij}(t_{ij})$ rather than $S_{ij}(t_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{v}_{ij})$ for notational simplicity. In the remainder of the thesis we will use these two notations for the survival function interchangeably.

$$S_{i1}(t_{i1}) = \exp\Big[- \Lambda_{01}(t_{i1})\exp\{\boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\}\Big],$$
$$S_{i2}(t_{i2}) = \exp\Big[- \Lambda_{02}(t_{i2})\exp\{\boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\}\Big],$$

where $\Lambda_{01}(t_{i1}) = \sum_{k=1}^{r} \rho_k u_k(t_{i1})$, $\Lambda_{02}(t_{i2}) = \sum_{l=s}^{r} \tau_l w_l(t_{i2})$, and

$$\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1}) = \boldsymbol{\gamma}^T \left[I(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1}) - I(0)\right],$$
$$\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2}) = \boldsymbol{\gamma}^T \left[I(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2}) - I(0)\right].$$

Note, in the proposed model, the linear covariate coefficients and the nonlinear covariate coefficients are assumed to be same for the two members of the same cluster (that is, coefficients $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\gamma}$ are the same for $j = 1$ and $j = 2$). The reason for this is to control the list of the parameters to a manageable level and to avoid long run times in simulations. The natural extension is to assume the linear and nonlinear covariate coefficients are different for the two variates. This extension poses no difficulties if computation time is not a concern, as estimation and inference procedures are unaffected.

If the Clayton bivariate model is assumed, the joint survival function has the form

given in Formula 1.1. The log likelihood is written as:

$$\begin{aligned} l &= \sum_{i=1}^{n} \log L_i \\ &= \sum_{i=1}^{n} \Bigg\{ \delta_{i1}\delta_{i2} \log f(t_{i1}, t_{i2}) \\ &\quad + \delta_{i1}(1-\delta_{i2}) \log\Big[-\frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i1}} \Big] \\ &\quad + (1-\delta_{i1})\delta_{i2} \log\Big[-\frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i2}} \Big] \\ &\quad + (1-\delta_{i1})(1-\delta_{i2}) \log S(t_{i1}, t_{i2}) \Bigg\} \end{aligned} \tag{2.2}$$

The entries of the log likelihood function $\frac{-\partial S(t_{i1},t_{i2})}{\partial t_{i1}}$, $\frac{-\partial S(t_{i1},t_{i2})}{\partial t_{i2}}$, and $f(t_{i1}, t_{i2})$ are given as follows:

$$\begin{aligned} \frac{-\partial S(t_{i1}, t_{i2})}{\partial t_{i1}} &= \frac{-\partial \Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1 \Big]^{-\phi}}{\partial t_{i1}} \\ &= -\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1 \Big]^{-\phi-1} \Big[S_{i1}(t_{i1})^{-\phi^{-1}} \Big] \\ &\quad \times \exp\{\boldsymbol{\beta}^T \boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i1})\} \frac{\partial[-\Lambda_{01}(t_{i1})]}{\partial t_{i1}} \end{aligned}$$

$$\begin{aligned} \frac{-\partial S(t_{i1}, t_{i2})}{\partial t_{i2}} &= \frac{-\partial \Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1 \Big]^{-\phi}}{\partial t_{i2}} \\ &= -\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1 \Big]^{-\phi-1} \Big[S_{i2}(t_{i2})^{-\phi^{-1}} \Big] \\ &\quad \times \exp\{\boldsymbol{\beta}^T \boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i2})\} \frac{\partial[-\Lambda_{02}(t_{i2})]}{\partial t_{i2}} \end{aligned}$$

$$
\begin{aligned}
f(t_{i1}, t_{i2}) &= \frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i1} \partial t_{i2}} \\
&= \left[S_{i1}(t_{i1})^{-\phi^{-1}}\right]\left[S_{i2}(t_{i2})^{-\phi^{-1}}\right]\exp\{\boldsymbol{\beta}^T \boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i1})\} \\
&\quad \times \exp\{\boldsymbol{\beta}^T \boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i2})\} \frac{\partial[-\Lambda_{01}(t_{i1})]}{\partial t_{i1}} \\
&\quad \times \frac{\partial[-\Lambda_{02}(t_{i2})]}{\partial t_{i2}} \left[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\right]^{-\phi-2} (1 + \phi^{-1})
\end{aligned}
$$

2.2.4 Parameter Estimates

The likelihood function involves the following parameters:

$$\boldsymbol{\theta} = (\phi, \boldsymbol{\rho}^T, \boldsymbol{\tau}^T, \boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T,$$

where ϕ is the association parameter; $\boldsymbol{\rho}$ and $\boldsymbol{\tau}$ are the parameters for the piecewise constants, and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_r)^T$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_s)^T$; $\boldsymbol{\alpha}$ are the nonlinear covariate parameters with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)^T$; $\boldsymbol{\beta}$ are the linear covariate parameters with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$; and $\boldsymbol{\gamma}$ are the parameters for spline functions with $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)^T$.

Recall, some of the parameters have constraints required by the model, which are summarized as follows:

$$\rho_1, \ldots, \rho_r > 0,$$

$$\tau_1, \ldots, \tau_s > 0,$$

$$\phi > 0,$$

$$\| \boldsymbol{\alpha} \| = 1, \text{ and } \alpha_q > 0,$$

$$\psi(0) = 0.$$

The question now is how to find the maximum likelihood estimates of the parameters, which actually becomes an optimization problem. The constraints will bring some difficulties in finding the maximum likelihood estimators for the parameters in

the process of optimizing the solutions. To avoid such problem, we introduce some new parameters which are transformed from the original parameters. Through the transformation, the new parameters will have no constraints.

For positive parameters ρ_k, τ_l, and ϕ, a log transformation is applied. New parameters ξ_k, ζ_l and ϱ are introduced as given below:

$$\xi_k = \log \rho_k, \quad k = 1, \ldots, r.$$

$$\zeta_l = \log \tau_l, \quad l = 1, \ldots, s,$$

$$\varrho = \log \phi.$$

For $\boldsymbol{\alpha}$, the transformation takes two steps. First we use the trigonometric transformation on $\boldsymbol{\alpha}$ for the constraint $\| \boldsymbol{\alpha} \| = 1$, and make sure the last component of $\boldsymbol{\alpha}$ is positive. Parameter ω is introduced with $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_{q-1})^T$. That is:

$$\begin{aligned}
\alpha_1 &= \sin\omega_1 \sin\omega_2 \cdots \sin\omega_{q-2} \sin\omega_{q-1} \\
\alpha_2 &= \sin\omega_1 \sin\omega_2 \cdots \sin\omega_{q-2} \cos\omega_{q-1} \\
&\ldots \\
\alpha_{q-1} &= \sin\omega_1 \cos\omega_2 \\
\alpha_q &= \cos\omega_1
\end{aligned} \tag{2.3}$$

where $\omega_1, \ldots, \omega_{q-1} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $\omega_1 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ can satisfy $\alpha_q > 0$.

Then the second step of the transformation is to avoid the constraints on $\boldsymbol{\omega}$. A

new parameter $\boldsymbol{\varphi}$ is introduced, where $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_{q-1})^T$:

$$\begin{aligned}
\varphi_1 &= \log \frac{\frac{\pi}{2} + \omega_1}{\frac{\pi}{2} - \omega_1} \Rightarrow \omega_1 = \frac{e^{\varphi_1} - 1}{e^{\varphi_1} + 1} \frac{\pi}{2} \\
\varphi_2 &= \log \frac{\frac{\pi}{2} + \omega_2}{\frac{\pi}{2} - \omega_2} \Rightarrow \omega_2 = \frac{e^{\varphi_2} - 1}{e^{\varphi_2} + 1} \frac{\pi}{2} \\
&\cdots \\
\varphi_{q-1} &= \log \frac{\frac{\pi}{2} + \omega_{q-1}}{\frac{\pi}{2} - \omega_{q-1}} \Rightarrow \omega_{q-1} = \frac{e^{\varphi_{q-1}} - 1}{e^{\varphi_{q-1}} + 1} \frac{\pi}{2}
\end{aligned} \tag{2.4}$$

Through the two steps of transformation the relationship between $\boldsymbol{\alpha}$ and $\boldsymbol{\varphi}$ are as follows:

$$\begin{aligned}
\alpha_1 &= \sin\left[\frac{e^{\varphi_1} - 1}{e^{\varphi_1} + 1} \frac{\pi}{2}\right] \sin\left[\frac{e^{\varphi_2} - 1}{e^{\varphi_2} + 1} \frac{\pi}{2}\right] \cdots \sin\left[\frac{e^{\varphi_{q-1}} - 1}{e^{\varphi_{q-1}} + 1} \frac{\pi}{2}\right] \\
\alpha_2 &= \sin\left[\frac{e^{\varphi_1} - 1}{e^{\varphi_1} + 1} \frac{\pi}{2}\right] \sin\left[\frac{e^{\varphi_2} - 1}{e^{\varphi_2} + 1} \frac{\pi}{2}\right] \cdots \cos\left[\frac{e^{\varphi_{q-1}} - 1}{e^{\varphi_{q-1}} + 1} \frac{\pi}{2}\right] \\
&\cdots \\
\alpha_q &= \cos\left[\frac{e^{\varphi_1} - 1}{e^{\varphi_1} + 1} \frac{\pi}{2}\right]
\end{aligned}$$

As a summary, the new parameters to estimate without constraints are:

$$\boldsymbol{\theta}^* = (\varrho, \boldsymbol{\xi}^T, \boldsymbol{\zeta}^T, \boldsymbol{\varphi}^T, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T,$$

where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_r)^T$, and $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_s)^T$, $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_{q-1})^T$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)^T$. For these unconstrained parameters, a standard optimization algorithm such as Newton-Raphson method can be applied to find the maximum likelihood estimators.

For the purpose of making inference, the variance estimate of the maximum likelihood estimator $\hat{\boldsymbol{\theta}}^*$ can be obtained from the inverse of the observed information matrix $I(\hat{\boldsymbol{\theta}}^*)$. If the second derivatives are inconvenient to get, one can use (He and Lawless,

2003)

$$\widehat{\text{var}}(\hat{\boldsymbol{\theta}}^*) = \left[\sum_{i=1}^{n} \left(\frac{\partial \log L_i}{\partial \boldsymbol{\theta}^*} \right) \cdot \left(\frac{\partial \log L_i}{\partial \boldsymbol{\theta}^*} \right)' \right]^{-1}_{\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}^*}. \tag{2.5}$$

In the simulation study and real data analysis, we use this formula to obtain the variance estimate of $\hat{\boldsymbol{\theta}}^*$.

Once the variance estimate for $\hat{\boldsymbol{\theta}}^*$, $\widehat{\text{var}}(\hat{\boldsymbol{\theta}}^*)$, is found, the delta method can be applied to calculate the variance estimate for $\hat{\boldsymbol{\theta}}$. If we call the map G: $\boldsymbol{\theta}^* \to \boldsymbol{\theta}$, that is $\boldsymbol{\theta} = G(\boldsymbol{\theta}^*)$, by the delta method, $\hat{\boldsymbol{\theta}}$ is asymptotically normal with an estimated asymptotic variance-covariance matrix:

$$\begin{aligned} \text{var}(\hat{\boldsymbol{\theta}}) &= \text{var}(G(\hat{\boldsymbol{\theta}}^*)) \\ &= G'(\hat{\boldsymbol{\theta}}^*)\text{var}(\hat{\boldsymbol{\theta}}^*)G'(\hat{\boldsymbol{\theta}}^*)^T, \end{aligned} \tag{2.6}$$

where G' denotes the derivative of G with respect to $\hat{\boldsymbol{\theta}}^*$. Based on the relationship between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$, G' can be worked out analytically. In the above formula, we substitute $\widehat{\text{var}}(\hat{\boldsymbol{\theta}}^*)$ for $\text{var}(\hat{\boldsymbol{\theta}}^*)$, and then the estimate of $\text{var}(\hat{\boldsymbol{\theta}})$ can be obtained.

2.3 Simulation Studies

In this section, the performance of the proposed model is assessed through three simulation studies. The first simulation study gives the estimates of the parameters of interest, such as the linear covariate parameter β, the nonlinear covariate parameter $\boldsymbol{\alpha}$ and the association parameter ϕ. The estimate of the cumulative baseline hazard function $\Lambda_0(\cdot)$ and the nonlinear function $\psi(\cdot)$ are also given. The second simulation study compares the proposed model with what is referred to as the *PH linear model* here. The PH linear model is the one that has the same structure as the proposed model except that the covariates are all assumed to have linear effects on the log hazards ratio. That is, the single index structure is dropped. Through the comparison,

the advantage of the proposed model that can capture the nonlinear pattern will be shown. The third simulation study assesses the impacts of other factors on the parameter estimates under various scenarios. For example, when some settings change, such as the sample size, censoring rate, and the association among the observed data, the performance of the model can be observed.

The data were generated as follows. The baseline hazard function was set to be a Weibull hazard with scale parameter 1 and shape parameter $p = 1.5$. The true marginal log relative hazards ratio for the survival time T_j conditional on covariates $(X_j, \boldsymbol{V}_j)$ is given by

$$\lambda_j(t_j|x_j, \boldsymbol{v}_j) = \lambda_{0j}(t_j)\exp\left\{\beta x_j + 3\sin(2\boldsymbol{\alpha}^T\boldsymbol{v}_j)\right\}, \quad j = 1, 2,$$

where the trigonometric function $3\sin(2\boldsymbol{\alpha}^T\boldsymbol{v}_j)$ is used as the nonlinear smooth function $\psi(\cdot)$, as has been used previously in the literature (Sun et al., 2008). Figure 2.2b shows that the function $\psi(\cdot)$ has a nonlinear "S" shape. The association parameter ϕ was set to be 0.5 which corresponds to relatively strong dependence between T_1 and T_2. The censoring time C_j was assumed to be independent of the survival time T_j, conditional on $(X_j, \boldsymbol{V}_j)$. It was set at 1.2 and the corresponding censoring rates were about 45% for both variates.

We used the uniform distribution to generate the nonlinear covariates $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, $V_{k1} \sim U(-1, 1)$, $V_{k2} \sim U(-1, 1)$, $k = 1, 2, 3$. The linear covariates (x_1, x_2) were chosen from four sets of combinations (0,0), (0,1), (1,0), (1,1), each with probability one quarter. The nonlinear regression coefficients are $\boldsymbol{\alpha} = (1, 1, 1,)^T/\sqrt{3}$ and the linear regression coefficient is $\beta = 1$.

The bivariate survival times (t_{i1}, t_{i2}), $i = 1, \ldots, n$ were generated through the inverse transform sampling method:

$$t_{i1} = \Big[-\log(u_{i1})\exp\{-(\beta x_{i1} + 3\sin(2(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})))\}\Big]^{1/p},$$

$$t_{i2} = \left[\phi \log(1 - u_{i1}^{-1/\phi} + u_{i1}^{-1/\phi} u_{i2}^{-1/(1+\phi)}) \exp\{-(\beta x_{i2} + 3\sin(2(\boldsymbol{\alpha}^T \boldsymbol{v}_{i2})))\}\right]^{1/p},$$

where $u_{i1} \sim U[0,1]$, $u_{i2} \sim U[0,1]$, and $p = 1.5$ for Weibull margins.

For the estimate of the baseline function, each variate uses a piecewise constant estimate with four pieces. The cut points are chosen to give approximately equal survival probabilities in each interval.

2.3.1 Evaluate the Performance of the Proposed Model

Based on the above settings, the sample size $n = 200$ was considered, and 200 simulation runs were conducted. Table 2.1 gives the estimates for the covariate coefficients $\boldsymbol{\varphi}$ and ϱ (after transformation), and for the original parameters $\boldsymbol{\alpha}$, β and ϕ (before transformation). The notations in the tables are as follows: Bias is the bias of the parameters, SD($\hat{\cdot}$) is the sample standard deviation of the estimates, A{SE($\hat{\cdot}$)} is the average of the model based standard errors (using Equation 2.5), and Cov.prob. is the coverage probability of the 95% confidence interval. From the outputs we can see that the estimates are fairly close to the true values. SD($\hat{\cdot}$) and A{SE($\hat{\cdot}$)} values are similar which shows that the asymptotic variances are reasonable. The coverage probabilities also show the overall performance is reasonably good.

Table 2.1: Estimates for the covariate coefficients $\boldsymbol{\varphi}$ and ϱ (after transformation), and $\boldsymbol{\alpha}$, β and ϕ (before transformation) based on 200 runs

	After transformation			Before transformation				
	φ_1	φ_2	ϱ	α_1	α_2	α_3	β	ϕ
True	1.412	1.099	-0.693	0.577	0.577	0.577	1.000	0.500
Bias	0.003	-0.005	-0.118	-0.002	0.002	-0.001	-0.008	-0.045
SD($\hat{\cdot}$)	0.044	0.040	0.212	0.016	0.017	0.018	0.118	0.100
A{SE($\hat{\cdot}$)}	0.042	0.043	0.235	0.017	0.017	0.017	0.128	0.109
Cov.prob.	0.930	0.950	0.910	0.950	0.950	0.930	0.965	0.850

Table 2.2 gives the average estimates over 200 simulation runs for the cumulative

baseline hazard function $\Lambda_{01}(\cdot)$ at the points (0.02,0.2,0.5), and for the nonlinear function $\psi(\cdot)$ at the points (-0.9,-0.8,-0.7,0.2). The latter four points are not collinear on the true sine function, and therefore are useful to test the model's ability to capture nonlinearity. The estimate are fairly close to the true values. Figure 2.2 gives a visual display of these estimates. The black line represents the true value of $\Lambda_{01}(\cdot)$ and $\psi(\cdot)$, and the red dots are the average estimates at the chosen points over 200 runs. The figures also show the estimate follow the true curve fairly well. Note the estimate for $\psi(x)$ at $x = 0$ is zero based on one of the constraints of the model, which is also shown in the figure. Since the estimate for the cumulative baseline hazard function of the second variate Λ_{02} has similar performance, it is not shown here.

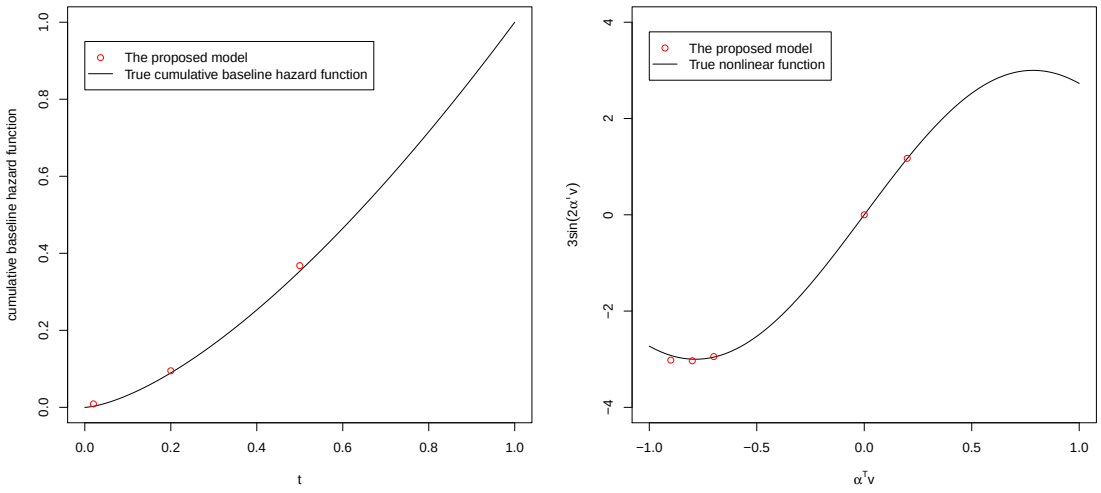
Table 2.2: The estimates for $\Lambda_{01}(\cdot)$ and $\psi(\cdot)$ of one variate based on 200 runs

	x	True	Estimate	SD
	0.020	0.003	0.006	0.002
$\Lambda_{01}(x)$	0.200	0.089	0.095	0.024
	0.500	0.354	0.368	0.061
	-0.900	-2.922	-3.016	0.828
	-0.800	-2.999	-3.030	0.712
$\psi(x)$	-0.700	-3.002	-2.943	0.572
	0.200	1.168	1.172	0.111

2.3.2 Compare the Proposed Model with PH Linear Model

In this section, the proposed model is compared with the PH linear model. As we have explained before, PH linear model assumes all the covariates have linear effects on the log hazards ratio, while the rest of the model is the same as the proposed model. In order to make it easily comparable with the proposed model, the marginal hazard function of the linear model has the following form:

$$\lambda_j(t_j|x_j, \boldsymbol{v}_j) = \lambda_{0j}(t_j)\exp\{\beta x_j + \tilde{\boldsymbol{\alpha}}^T \boldsymbol{v}_j\},\ j = 1, 2,$$



(a) Λ_{01} (b) $\psi(\cdot)$

Figure 2.2: Cumulative baseline hazard function Λ_{01} and nonlinear smooth function $\psi(\cdot)$ with their estimates at the chosen points. Black line represents the true values, and red dots represents the estimates.

where β is the coefficient for linear covariate x_j and $\tilde{\boldsymbol{\alpha}}$ is the coefficient for nonlinear covariate $\boldsymbol{v}_j$. Note that $\tilde{\boldsymbol{\alpha}}$ is not subject to the same unit norm constraint enforced on the $\boldsymbol{\alpha}$ parameters in the proposed model. To make the estimates of $\tilde{\boldsymbol{\alpha}}$ comparable with the estimates from the proposed model, we factor each estimate as $\hat{\tilde{\boldsymbol{\alpha}}} = \hat{b}\hat{\boldsymbol{\alpha}}$, where $\| \hat{\boldsymbol{\alpha}} \| = 1$ and $\hat{b}$ is interpreted as the slope for linear regression on $\hat{\boldsymbol{\alpha}}^T \boldsymbol{v}$.

In this simulation study the same settings were used as in the first simulation study. The parameter estimates are the average of the estimates of 200 simulation runs. The results from both models for $\boldsymbol{\alpha}$, β and ϕ are listed in Table 2.3.

We can see that comparing with the linear model, the parameter estimates of the proposed model are generally closer to the true values, and the standard deviations are smaller than those of the linear model. Similar results also appear in the estimation of the cumulative baseline hazard function Λ_{01} and the nonlinear smooth function $\psi(\cdot)$. In Table 2.4, the estimation performance of the proposed model is better than the linear model. Such comparison gets much clearer in the graph. Figure 2.3 gives

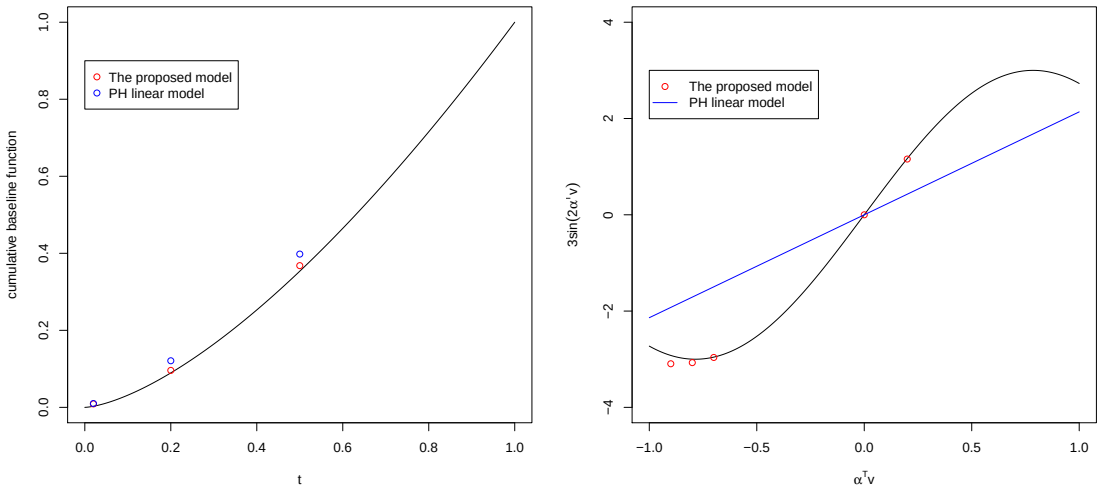
the estimates for Λ_{01} and $\psi(\cdot)$ from both models. In Figure 2.3a, the estimation of Λ_{01} from the linear model at (0.02, 0.2, 0.5), represented by the blue dots, are farther from the true value, the black line, than the proposed model, represented by the red dots. In Figure 2.3b, the black line represents the true nonlinear function $\psi(\boldsymbol{\alpha}^T\boldsymbol{v})$. The linear model can be thought of as a restricted case of the single index model, where $\psi(\cdot)$ is a line passing through the origin. The blue line shows the mean estimate of this line based on the simulation runs. Its slope is equal to the average estimate of $\hat{b}$. The linear model completely misses the "S" curvature while the proposed model can handle the nonlinear pattern quite well as proved by the estimation at (-0.9,-0.8,-0.7,0.2).

Table 2.3: The estimates comparison of two models for the covariate coefficients $\boldsymbol{\alpha}$, β and ϕ based on 200 runs

	α_1	α_2	α_3	β	ϕ
True	0.577	0.577	0.577	1.000	0.500
Estimate (proposed)	0.575	0.577	0.579	0.971	0.455
SD($\hat{\cdot}$)(proposed)	0.017	0.018	0.017	0.100	0.128
Estimate (linear)	0.571	0.576	0.581	0.709	3.060
SD($\hat{\cdot}$) (linear)	0.041	0.042	0.041	0.178	2.226

Table 2.4: The estimates comparison of two models for $\Lambda_{01}(\cdot)$ and $\psi(\cdot)$ based on 200 runs

			Proposed		Linear	
	x	True	Estimate	SD	Estimate	SD
	0.020	0.003	0.006	0.002	0.014	0.004
$\Lambda(x)$	0.200	0.089	0.097	0.024	0.185	0.037
	0.500	0.354	0.375	0.062	0.479	0.069
	-0.900	-2.922	-3.094	0.905	-1.924	0.284
	-0.800	-2.999	-3.071	0.720	-1.710	0.252
$\psi(x)$	-0.700	-3.002	-2.963	0.590	-1.496	0.221
	0.200	1.168	1.158	0.112	0.428	0.063



(a) Λ_{01} (b) $\psi(\cdot)$

Figure 2.3: Compare the proposed model and the PH linear model on cumulative baseline hazard function Λ_{01} and nonlinear smooth function $\psi(\cdot)$. Black line represents the true curve, blue line and dots are the estimates from the PH linear model, and the red line and dots are the estimates from the proposed model.

2.3.3 Assess the Proposed Model Under Various Scenarios

In this section, the proposed model is assessed under different scenarios. The following settings of the models are changed. The shape parameter p of the Weibull baseline function is set as 0.5 and 1.5. The sample size is used as 80 and 200. The censoring rate of 20% and 50% are applied. The strength of the association changes from strong to weak, that is the association parameter ϕ takes values 0.5, 1 and 4.

Table 2.5 and Table 2.6 give the simulation result of 200 runs for $\boldsymbol{\alpha}$, β and ϕ when the Weibull shape parameter p is 0.5 and 1.5 respectively. For both tables, the standard deviations and the average of the standard errors are generally close to each other. In each of the two tables, we can grasp the following patterns. In general, the estimate of sample size 200 is better than sample size 80. The bias of the parameters at the larger sample size is smaller and the standard deviations are also smaller. When the censoring rate increases, the estimation performance decreases

as the standard deviations are increasing. As the strength of association decreases, the estimation performance for ϕ decreases. The bias increases and the standard deviations increases. Especially when $\phi = 4$, we can see the standard deviations and the average standard errors of ϕ are not very close, and some bias appears. It is believed that the poor performance at this setting is due to the likelihood becoming very flat for large ϕ values.

2.4 Real data analysis

2.4.1 Introduction

In this section, the proposed PH model is applied on a real data set collected from the Busselton Health Study (Knuiman et al., 1994). The background information of this data set can be seen in Section 1.1.2.

The response survival times are the time to death (SURVTIME) ($t_{ij},\ j = 1, 2$), where $j = 1$ indicates wife, and $j = 2$ denotes husband. The death indicator is DTHCENS ($\delta_{ij},\ j = 1, 2$). The risk factors of interest to mortality are age at survey (AGE) ($v_{i1j},\ j = 1, 2$), body mass index (BMI) ($v_{i2j},\ j = 1, 2$), cholesterol level (CHOL) ($v_{i3j},\ j = 1, 2$), and the smoking status (SMOKE) ($x_{ij},\ j = 1, 2$), where $x_{ij} = 1$ indicates an individual is a smoker, and $x_{ij} = 0$ indicates an individual never smoked. Among the four risk factors of interest, SMOKE is a discrete variable, which is regarded as the linear effect covariate in the proposed model. The other three factors are continuous variables and are regarded as the nonlinear effect covariates. To avoid the difficulties of the optimization caused by the extreme value of the covariates, standardized continuous covariates are used.

The analysis, described in detail below, shows there is mild association existing between the female and the male. The estimate of the baseline cumulative hazard function is very similar with the Nelson-Aalen estimator (N-A estimator). The N-A

Table 2.5: Simulation result for Weibull with shape parameter $p = 0.5$ of 200 runs

			α_1			α_2			α_3		
ϕ	P(%)	n	Bias	SD(·̂)	A{SE(·̂)}	Bias	SD(·̂)	A{SE(·̂)}	Bias	SD(·̂)	A{SE(·̂)}
0.5	20%	80	-0.003	0.023	0.020	0.000	0.024	0.020	0.002	0.023	0.020
		200	0.001	0.012	0.010	-0.001	0.013	0.010	0.000	0.013	0.010
	50%	80	0.000	0.054	0.031	-0.003	0.056	0.031	-0.005	0.048	0.031
		200	0.000	0.019	0.018	0.000	0.018	0.018	-0.001	0.019	0.018
1	20%	80	-0.001	0.026	0.023	-0.002	0.026	0.024	0.002	0.026	0.023
		200	-0.001	0.014	0.013	0.001	0.015	0.012	0.000	0.014	0.013
	50%	80	-0.006	0.054	0.035	-0.002	0.054	0.035	0.001	0.048	0.035
		200	-0.003	0.022	0.020	-0.002	0.022	0.020	0.003	0.023	0.020
4	20%	80	-0.001	0.029	0.026	0.002	0.025	0.026	-0.003	0.028	0.026
		200	-0.001	0.017	0.014	0.001	0.018	0.014	-0.001	0.016	0.014
	50%	80	0.005	0.038	0.033	-0.002	0.040	0.033	-0.007	0.040	0.034
		200	-0.004	0.023	0.021	0.000	0.022	0.022	0.003	0.025	0.021

			β			ϕ		
ϕ	P(%)	n	Bias	SD(·̂)	A{SE(·̂)}	Bias	SD(·̂)	A{SE(·̂)}
0.5	20%	80	0.171	0.205	0.180	0.071	0.203	0.180
		200	0.168	0.128	0.093	0.122	0.109	0.106
	50%	80	0.112	0.258	0.265	0.019	0.318	0.289
		200	0.079	0.131	0.138	0.029	0.142	0.137
1	20%	80	0.171	0.216	0.205	0.271	0.748	0.644
		200	0.164	0.120	0.108	0.240	0.313	0.287
	50%	80	0.100	0.277	0.290	0.063	0.816	0.934
		200	0.083	0.162	0.154	0.143	0.421	0.440
4	20%	80	0.191	0.235	0.225	-0.468	1.636	3.360
		200	0.160	0.141	0.121	0.682	1.768	2.787
	50%	80	0.107	0.301	0.275	-1.493	1.579	2.588
		200	0.072	0.179	0.164	-0.326	1.472	3.158

Table 2.6: Simulation result for Weibull with shape parameter $p = 1.5$ of 200 runs

			α_1			α_2			α_3		
ϕ	P(%)	n	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}
0.5	20%	80	-0.002	0.024	0.023	0.003	0.023	0.023	-0.002	0.023	0.023
		200	-0.001	0.011	0.012	0.001	0.011	0.012	0.000	0.011	0.012
	50%	80	0.002	0.035	0.036	-0.003	0.036	0.036	-0.002	0.036	0.036
		200	-0.002	0.019	0.019	0.000	0.019	0.019	0.001	0.019	0.019
1	20%	80	0.000	0.027	0.028	-0.005	0.024	0.028	0.003	0.028	0.029
		200	0.002	0.014	0.016	-0.002	0.016	0.016	-0.001	0.015	0.016
	50%	80	-0.001	0.038	0.038	-0.001	0.039	0.038	-0.001	0.037	0.038
		200	0.000	0.022	0.022	-0.001	0.021	0.022	0.001	0.019	0.022
4	20%	80	-0.002	0.048	0.033	-0.001	0.050	0.033	-0.003	0.042	0.034
		200	-0.002	0.018	0.019	-0.001	0.017	0.019	0.002	0.017	0.019
	50%	80	0.000	0.042	0.042	0.003	0.042	0.042	-0.007	0.043	0.043
		200	0.001	0.024	0.024	-0.005	0.023	0.024	0.003	0.023	0.024

			β			ϕ		
ϕ	P(%)	n	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}
0.5	20%	80	-0.057	0.147	0.175	-0.077	0.122	0.134
		200	-0.074	0.081	0.093	-0.063	0.070	0.076
	50%	80	-0.035	0.231	0.265	-0.091	0.238	0.223
		200	-0.023	0.127	0.139	-0.032	0.131	0.126
1	20%	80	-0.039	0.176	0.214	-0.159	0.429	0.388
		200	-0.056	0.102	0.115	-0.132	0.196	0.199
	50%	80	0.002	0.240	0.285	-0.102	0.702	0.781
		200	-0.013	0.149	0.157	-0.072	0.342	0.360
4	20%	80	-0.030	0.194	0.246	-1.642	1.208	2.118
		200	-0.046	0.115	0.135	-0.474	1.526	2.240
	50%	80	-0.021	0.242	0.319	-2.149	1.056	2.251
		200	-0.020	0.157	0.171	-0.774	1.377	2.826

estimator is a non-parametric estimator of the cumulative hazard function. It has the form (Lawless, 2003):

$$\hat{H}(t) = \sum_{j:t_j \leq t} \frac{d_j}{n_j},$$

where $t_1, \ldots, t_k$ represent the distinct times at which failures are observed, $d_j = \sum I(t_i' = t_j, \delta_j = 1)$ represents the number of deaths at t_j, and $n_j = \sum I(t_i' \geq t_j)$ is the number of individuals at risk at t_j. The covariate effect of each individual on the cumulative hazard function is also clearly shown in the proposed model, compared with N-A estimator which does not consider the covariate effect. From the shape of the smooth function $\psi(\cdot)$ we can see a nonlinear pattern. If we fix two of the nonlinear covariates at their medians and change the third nonlinear covariate, a nonlinear changing pattern of the smooth function $\psi(\cdot)$ can also be observed.

2.4.2 Analysis

Before starting to analyze the data by the proposed model, the assumption of proportional hazards is tested. Using R function `cox.zph` (Therneau, 2013), it is seen that the assumption is reasonable.

To find the interior cut points for the piecewise approach of the baseline hazard function estimation, we use the strategy that is based on the Kaplan-Meier survival probability estimates. The cut points are selected to let the four intervals give roughly the same survival probability differences. As a result the chosen interior cut points are (78.1, 85.2, 90.9) for female and (74.4, 81.5, 87.4) for male. The breakpoints applied in the spline function for the nonlinear function $\psi(\boldsymbol{\alpha}^T \boldsymbol{v})$ estimation divide the survival time range of each variate into four intervals, and each interval contains roughly an equal number of data points.

The parameters to be estimated are, as before, ϕ, $\boldsymbol{\rho}$, $\boldsymbol{\tau}$, $\boldsymbol{\alpha}$, β, $\boldsymbol{\gamma}$. In this case the baseline hazard functions have four constant pieces, so the parameter vectors $\boldsymbol{\rho}$ (for

females) and $\boldsymbol{\tau}$ (for males) have length four each. Parameter $\boldsymbol{\gamma}$ has six elements as required by the spline estimate of the nonlinear component. There are three nonlinear predictors: AGE, BMI, and CHOL, with coefficients α_1, α_2 and α_3, respectively. The single linear predictor is SMOKE, with coefficient β.

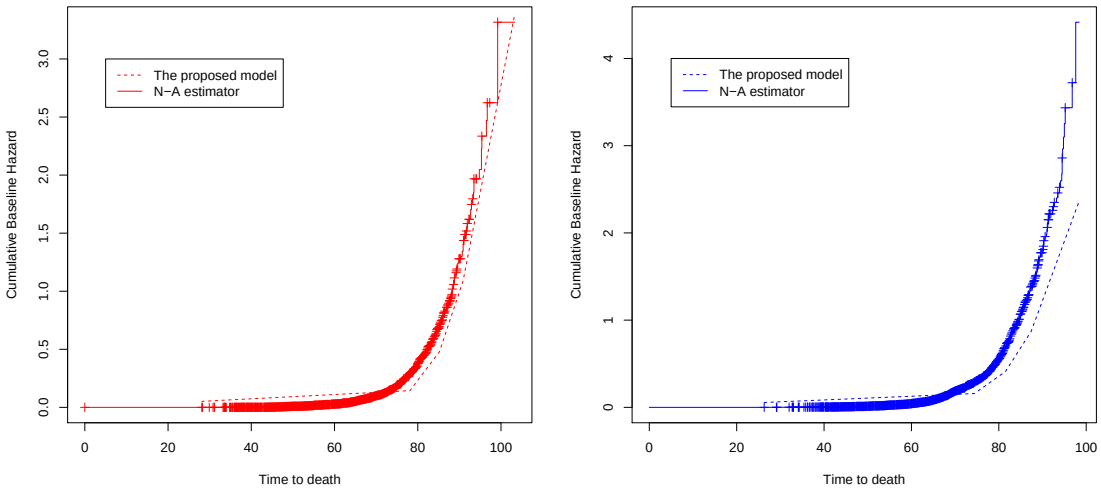
Table 2.7 gives the summary of the parameter estimates. Among the three continuous covariates, age has a dominating influence on survival time. The statistically significant covariates are AGE and SMOKE. The association parameter ϕ is also significant, with the value 4.955, which shows a mild degree of association between female and male.

Table 2.7: The estimates for the covariate coefficients $\boldsymbol{\alpha}$, β and ϕ of the data analysis

	α_1	α_2	α_3	β	ϕ
Estimate	0.999	0.029	0.044	0.356	4.955
SE($\hat{\cdot}$)	0.003	0.051	0.065	0.076	2.017

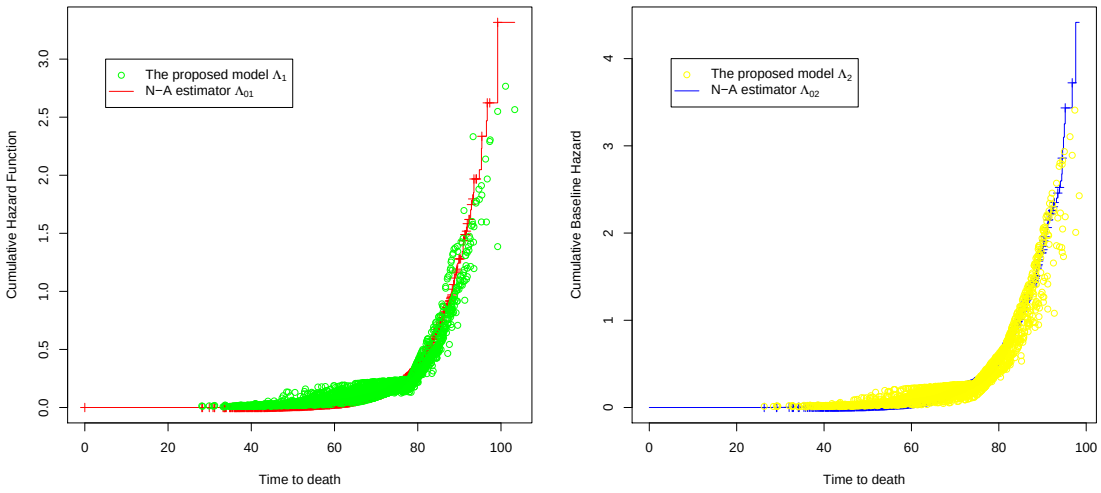
In Figure 2.4, the baseline cumulative hazard function from the proposed model is compared with N-A estimator. Figure 2.4a gives the comparison for female, and Figure 2.4b gives the comparison for male. Both the proposed model and N-A estimator give similar convex shape of the cumulative hazard function.

In Figure 2.5, the N-A estimator of the cumulative hazard function, which does not consider the covariate effect, is compared with the estimate from the proposed model that considers the covariate effect. The lines in the graph represent the N-A estimator, while each point represent the cumulative hazard function for each individual, which are the estimate from the proposed model. Since the covariate effect is not included for N-A estimator, it forms a non-decreasing line. For the proposed model estimate, after adding the covariate effect, the cumulative hazard function is changed by the covariate effect, and the effect is different for each individual. Therefore it can not form a non-decreasing line any more.



(a) Female (b) Male

Figure 2.4: Compare the cumulative baseline hazard function Λ_0 with N-A estimator for female and male.



(a) Female (b) Male

Figure 2.5: Compare the cumulative hazard function Λ considering covariate effects with N-A estimator of Λ_0 for female and male. Dots represent Λ for individuals of the proposed model, and the lines represent N-A estimator of Λ_0.

Figure 2.6 shows the nonlinear function $\psi(\cdot)$ for females. The figure clearly shows the nonlinear relationship between $\psi(\cdot)$ and $\boldsymbol{\alpha}^T\boldsymbol{v}$. As the proposed model assumes the

same parameters of the nonlinear functions for both female and male, the nonlinear function for male has the same shape.

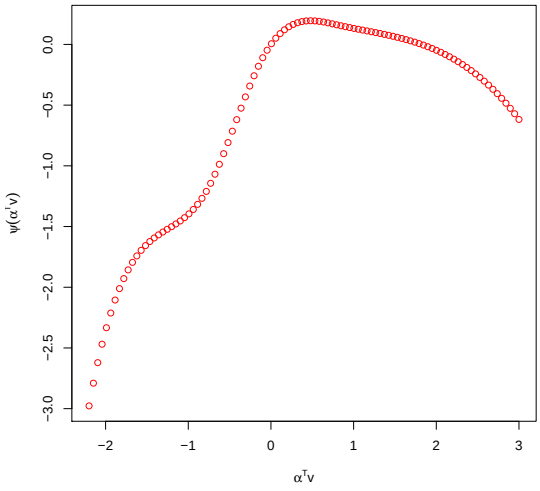
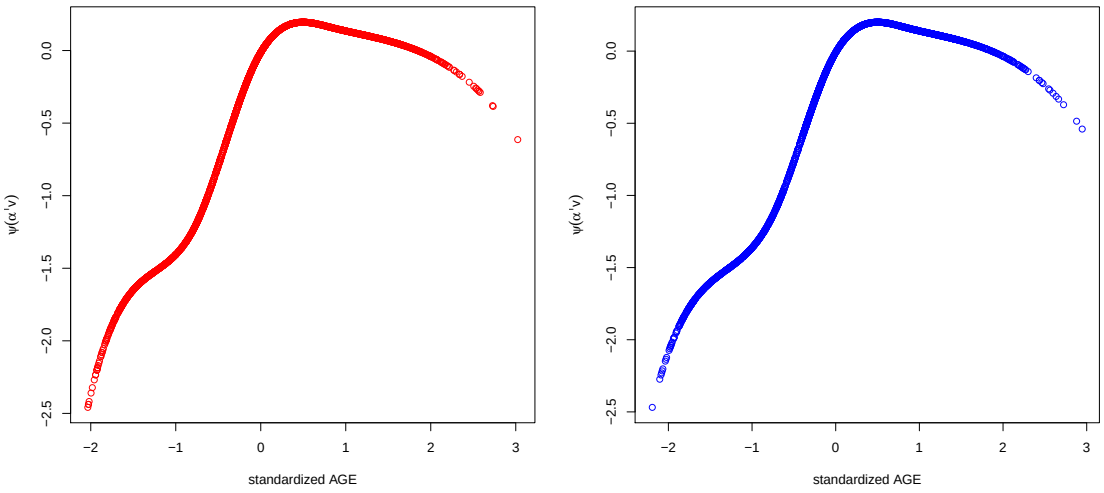


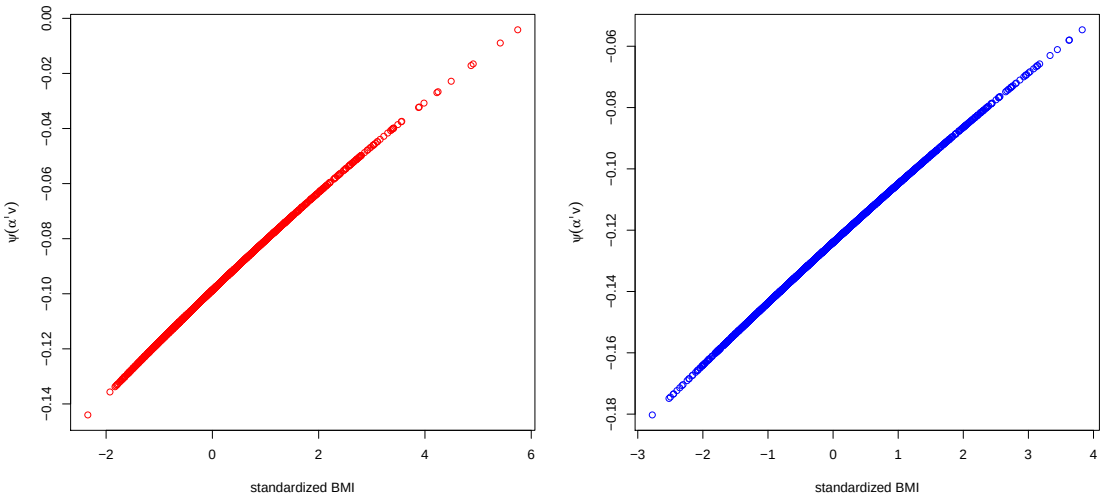
Figure 2.6: Nonlinear function $\psi(\cdot)$ for females

Figure 2.7 to Figure 2.9 show how the nonlinear function $\psi(\cdot)$ changes when one covariate changes while the other two covariates are fixed at their medians, for both females and males. The differences between the females and males are due to the different medians of the fixed covariates for different genders. In Figure 2.7, as age increases, $\psi(\cdot)$ experiences a steeper increasing stage in the beginning and then a relatively more mild decreasing stage roughly after age 50 for female and 54 for male. From the relationship of the death hazard and the nonlinear function $\psi(\cdot)$ in the proposed model, it can be found that the monotonicity of the death hazard is consistent with the nonlinear function $\psi(\cdot)$. Therefore roughly from age 18 to 50 the death hazard is getting bigger, and after 50 the death hazard maintains a fairly high level. For Figure 2.8, as the body mass index increases, $\psi(\cdot)$ increases, which also indicates the death hazard increases. Similarly for Figure 2.9, as the cholesterol level increases, the death hazard increases.



(a) Female (b) Male

Figure 2.7: Nonlinear function $\psi(\cdot)$ as the covariate age changes

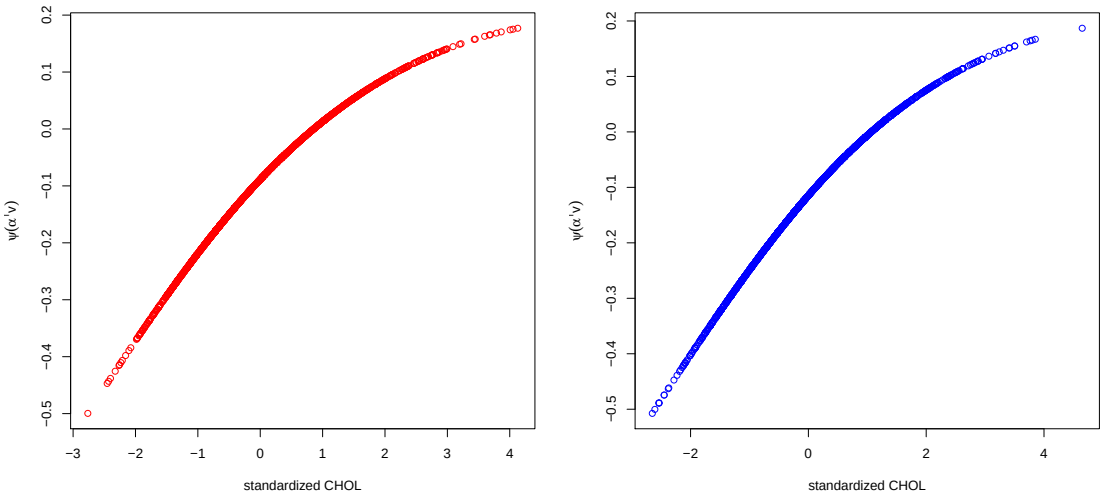


(a) Female (b) Male

Figure 2.8: Nonlinear function $\psi(\cdot)$ as the covariate body mass index changes

2.5 Conclusion

In this chapter, the PH model is extended by adding a nonlinear covariate effect to the log hazards ratio. Spline functions are applied to explore the pattern of the



(a) Female (b) Male

Figure 2.9: Nonlinear function $\psi(\cdot)$ as the covariate cholesterol level changes

smooth function which is used to model the nonlinear effect. A piecewise constant estimator is applied to estimate the baseline hazard function. A copula model is used to accommodate the association in the multivariate data. Therefore, the full likelihood approach can be applied to find the maximized likelihood estimate of the parameters. The major advantage of this proposed model is the ability to flexibly model the covariate effect, either linear or nonlinear. Additionally, it is possible to calculate the survival function and hazard function for each individual.

Through the simulation it was found that the proposed model is able to capture the nonlinear relationship well. When a nonlinear effect exists, this is a clear advantage over the model that only includes linear covariate effects. In the real data analysis, mild association was found to be existing between husbands and wives. The hazard function considering the covariate effects can be estimated for each individual. Nonlinear patterns are clearly shown in the shape of the smooth function $\psi(\cdot)$.

Chapter 3

Flexible Partially Linear Single Index Proportional Odds Regression Model for Multivariate Survival Data

3.1 Introduction

Bennett (1983) introduces the proportional odds model with the following form:

$$\frac{S(t|\boldsymbol{x})}{1-S(t|\boldsymbol{x})} = \exp(\boldsymbol{\beta}^T\boldsymbol{x})\frac{S_0(t)}{1-S_0(t)},$$

where $S(t|\boldsymbol{x})$ is the survival function for an individual given covariate $\boldsymbol{x}$, $\boldsymbol{\beta}$ is the covariate coefficient with dimension $p\times 1$, $\boldsymbol{x}$ is the $p\times 1$ vector of explanatory variables for an individual, and $S_0(t)$ is the baseline survival function when $\boldsymbol{x} = \mathbf{0}$. This model has an important property involving the ratio of the hazard function for an individual

to the baseline hazard function. It can be shown that (Collett, 2003),

$$\frac{\lambda(t|\boldsymbol{x})}{\lambda_0(t)} = [1 + (\exp(\boldsymbol{\beta}^T\boldsymbol{x}) - 1)S_0(t)]^{-1}.$$

That is, when $t = 0$, $\frac{\lambda(t|\boldsymbol{x})}{\lambda_0(t)} = \exp(-\boldsymbol{\beta}^T\boldsymbol{x})$, and when $t = \infty$, $\frac{\lambda(t|\boldsymbol{x})}{\lambda_0(t)} = 1$. This is different from the PH model for which the hazards ratio is constant over time, with form

$$\frac{\lambda(t|\boldsymbol{x})}{\lambda_0(t)} = \exp(\boldsymbol{\beta}^T\boldsymbol{x}).$$

In some situations, such assumption that the hazards ratio is constant in t can be unreasonable. For example, initial effects such as differences in the stage of disease or in treatment can disappear with time. In this case, the property of PO model that the hazards ratio converges to 1 as t increases to infinity makes more sense. Another advantage of the PO model over the PH model is, there is no assumption needed about the proportionality of the hazards ratio, while the latter requires to check it in order to use the model properly.

The PO model has drawn many researchers' attention. Various approaches to inference have been explored. Bennett (1983) describes how the PO model can be fitted using maximum likelihood estimation. Murphy et al. (1997) further demonstrate that the profile likelihood approach of Bennett (1983) results in an asymptotically efficient regression estimator. Cheng et al. (1995) propose a modified V statistic for parameter estimation. Yang and Prentice (1999) introduce weighted empirical odds functions which are solutions of self-consistency equations. From these functions, several classes of new regression estimators, such as the pseudo-maximum likelihood estimator, martingale residual-based estimators, and minimum distance estimators are derived. Some researchers have studied methods that can be applied to both the PH and PO models. Royston and Parmar (2002) talk about the advantages of using parametric assumptions with the PH model and PO model under certain conditions.

Another important topic in this area is generalized transformation models, in which the PH model and the PO model are two special cases in a more general framework. This will be discussed in Chapter 4.

In this chapter, the PO model is applied to the multivariate survival data framework with some extensions. As has been done in Chapter 2, the linearity assumption of the covariate effect on the log survival odds ratio is relaxed by adding a nonlinear relationship through a smooth function. Single index model is applied to reduce the high dimensional covariates to a scalar. Piecewise constants are used to fit the baseline hazard function. A spline function is applied to explore the nonlinear shape of the smooth function.

In section 3.2, the proposed PO model is introduced. Simulation studies are provided in section 3.3. The real data analysis on the Busselton study data is given in section 3.4. Some concluding remarks are noted in last section.

3.2 The Proposed Model

The proposed marginal PO function is written as:

$$\frac{S_j(t_j|\boldsymbol{x}_j,\boldsymbol{v}_j)}{1-S_j(t_j|\boldsymbol{x}_j,\boldsymbol{v}_j)} = \exp\{\boldsymbol{\beta}^T\boldsymbol{x}_j + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_j)\}\frac{S_{0j}(t_j)}{1-S_{0j}(t_j)}, \quad j=1,\ 2,$$

where S_{0j} is the marginal baseline survival function, $\boldsymbol{x}_j$ enters as a linear covariate with the coefficient $\boldsymbol{\beta}$ (dimension $p\times 1$), $\boldsymbol{v}_j$ enters as a nonlinear covariate with the coefficient $\boldsymbol{\alpha}$ (dimension $q\times 1$), and $\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_j)$ is the single index model. Constraints of $\parallel \boldsymbol{\alpha} \parallel = 1$ and $\alpha_q > 0$ are required for identifiability purpose.

Details of the model set up such as piecewise constants for the baseline hazard function and a spline function for the smooth function $\psi(\cdot)$ are described in the following section.

3.2.1 Piecewise Constant Baseline Hazard Function and Spline Functions

The baseline hazard functions $\lambda_{0j}(t_j)$, $j = 1, 2$, are modeled weakly parametrically by piecewise constant functions. The piecewise constant model is identical to the one used in the previous chapter for the PH baseline function, but it is repeated here for the sake of completeness. Assume that the baseline hazard functions have the following forms:

$$\lambda_{01}(t_1) = \rho_k, \quad \text{where } t_1 \in A_k = (a_{k-1}, a_k], \ k = 1, \ldots, r,$$

and

$$\lambda_{02}(t_2) = \tau_l, \quad \text{where } t_2 \in B_l = (b_{l-1}, b_l], \ l = 1, \ldots, s,$$

where $0 = a_0 < a_1 < \ldots < a_r = \infty$ and $0 = b_0 < b_1 < \ldots < b_s = \infty$ are pre-specified cut points, and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_r)^T$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_s)^T$ are unknown positive constants. Therefore, the corresponding cumulative hazard functions can be written as:

$$\begin{aligned} \Lambda_{01}(t_1) &= \sum_{k=1}^{r} \rho_k u_k(t_1), \\ \Lambda_{02}(t_2) &= \sum_{l=1}^{s} \tau_l w_l(t_2), \end{aligned}$$

where $u_k(t_1) = \max(0, \min(a_k, t_1) - a_{k-1})$ are the length of the intersection of the interval $(0, t_1)$ with the interval A_k, and $w_l(t_2) = \max(0, \min(b_l, t_2) - b_{l-1})$ are the length of the intersection of the interval $(0, t_2)$ with the interval B_l.

A spline is used to model the smooth function $\psi(\cdot)$. Similar to the methods used in the proposed PH model, M-spline and I-spline functions are applied to estimate $\psi'(\cdot)$ and $\psi(\cdot)$ respectively. To review the definitions of M-spline functions and I-spline functions, see Section 2.2.2.

Let M_j, $j = 1, \ldots, d$, be the M-spline basis functions, and $M(u) = (M_1(u), \ldots, M_d(u))^T$. Then $\psi'(\boldsymbol{\alpha}^T\boldsymbol{v})$ can be written as:

$$\psi'(\boldsymbol{\alpha}^T\boldsymbol{v}) = \sum_{j=1}^{d} \gamma_j M_j(\boldsymbol{\alpha}^T\boldsymbol{v}) = \boldsymbol{\gamma}^T M(\boldsymbol{\alpha}^T\boldsymbol{v}),$$

where γ_j is the coefficient for the jth basis function.

As I-spline basis functions are defined as the integration of the corresponding M-spline basis function, I-spline functions are naturally used to fit $\psi(\cdot)$. Note that $\psi(0)$ is constrained to be 0, because the intercept in the regression model is set to be 0, that is, $\boldsymbol{\beta}^T\boldsymbol{x} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}) = 0$ when $\boldsymbol{x} = \boldsymbol{v} = 0$. Using this constraint, $\psi(\boldsymbol{\alpha}^T\boldsymbol{v})$ can be modeled by $\int_0^{\boldsymbol{\alpha}^T\boldsymbol{v}} \psi'(x)dx$. As a result,

$$\begin{aligned}
\psi(\boldsymbol{\alpha}^T\boldsymbol{v}) &= \int_0^{\boldsymbol{\alpha}^T\boldsymbol{v}} \psi'(x)dx \\
&= \int_0^{\boldsymbol{\alpha}^T\boldsymbol{v}} \sum_{j=1}^{d} \gamma_j M_j(x)dx \\
&= \boldsymbol{\gamma}^T \left[I(\boldsymbol{\alpha}^T\boldsymbol{v}) - I(0)\right]
\end{aligned}$$

3.2.2 The Likelihood Function

Based on the above model set up, the full likelihood approach can be applied to obtain parameter estimates. From the marginal PO model, the marginal survival function can be written as:

$$S_{i1}(t_{i1}) = \frac{\exp\{\boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1}) - \Lambda_{01}(t_{i1})\}}{\exp\{\boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1}) - \Lambda_{01}(t_{i1})\} + 1 - \exp\{-\Lambda_{01}(t_{i1})\}},$$

$$S_{i1}(t_{i2}) = \frac{\exp\{\boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2}) - \Lambda_{02}(t_{i2})\}}{\exp\{\boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2}) - \Lambda_{02}(t_{i2})\} + 1 - \exp\{-\Lambda_{02}(t_{i2})\}}.$$

The joint survival function is still set up according to the Clayton model as in (1.1). Let $f(t_1, t_2)$ denote the joint density function. The general form of the log-

likelihood function can be worked out following (2.2). Each entry of the log likelihood function, $f(t_{i1}, t_{i2})$, $-\frac{\partial S(t_{i1},t_{i2})}{\partial t_{i1}}$, and $-\frac{\partial S(t_{i1},t_{i2})}{\partial t_{i2}}$, are given as below:

$$\begin{aligned}\frac{-\partial S(t_{i1}, t_{i2})}{\partial t_{i1}} &= \frac{-\partial\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi}}{\partial t_{i1}} \\ &= -\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi-1}\Big[S_{i1}(t_{i1})^{-\phi^{-1}-1}\Big]\frac{\partial S_{i1}(t_{i1})}{\partial t_{i1}},\end{aligned}$$

where

$$\frac{\partial S_{i1}(t_{i1})}{\partial t_{i1}} = \frac{\exp\{\boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1}) - \Lambda_{01}(t_{i1})\}\left[\frac{-\partial\Lambda_{01}(t_{i1})}{\partial t_{i1}}\right]}{\Big[\exp\{\boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1}) - \Lambda_{01}(t_{i1})\} + 1 - \exp\{-\Lambda_{01}(t_{i1})\}\Big]^2},$$

and

$$\begin{aligned}\frac{-\partial S(t_{i1}, t_{i2})}{\partial t_{i2}} &= \frac{-\partial\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi}}{\partial t_{i2}} \\ &= -\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi-1}\Big[S_{i2}(t_{i2})^{-\phi^{-1}-1}\Big]\frac{\partial S_{i2}(t_{i2})}{\partial t_{i2}},\end{aligned}$$

where

$$\frac{\partial S_{i2}(t_{i2})}{\partial t_{i2}} = \frac{\exp\{\boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2}) - \Lambda_{02}(t_{i2})\}\left[\frac{-\partial\Lambda_{02}(t_{i2})}{\partial t_{i2}}\right]}{\Big[\exp\{\boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2}) - \Lambda_{02}(t_{i2})\} + 1 - \exp\{-\Lambda_{02}(t_{i2})\}\Big]^2}.$$

The resulting joint density is

$$
\begin{aligned}
f(t_{i1}, t_{i2}) &= \frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i1} \partial t_{i2}} \\
&= (1+\phi^{-1})\Big[S_{i1}(t_{i1})^{-\phi^{-1}}\Big]\Big[S_{i2}(t_{i2})^{-\phi^{-1}}\Big] \\
&\quad \times \Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi-2} \\
&\quad \times \frac{\exp\{\boldsymbol{\beta}^T \boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i1}) - \Lambda_{01}(t_{i1})\}\left[\frac{-\partial \Lambda_{01}(t_{i1})}{\partial t_{i1}}\right]}{\Big[\exp\{\boldsymbol{\beta}^T \boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i1}) - \Lambda_{01}(t_{i1})\} + 1 - \exp\{-\Lambda_{01}(t_{i1})\}\Big]^2} \\
&\quad \times \frac{\exp\{\boldsymbol{\beta}^T \boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i2}) - \Lambda_{02}(t_{i2})\}\left[\frac{-\partial \Lambda_{02}(t_{i2})}{\partial t_{i2}}\right]}{\Big[\exp\{\boldsymbol{\beta}^T \boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i2}) - \Lambda_{02}(t_{i2})\} + 1 - \exp\{-\Lambda_{02}(t_{i2})\}\Big]^2}.
\end{aligned}
$$

3.2.3 Parameter Estimates

In this section, the parameters to be estimated are summarized and maximum likelihood method is used to make inference. The parameters to be estimated are:

$$\boldsymbol{\theta} = (\phi, \boldsymbol{\rho}^T, \boldsymbol{\tau}^T, \boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T,$$

where ϕ is the association parameter; $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_r)^T$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_s)^T$ are the parameters for the piecewise constant functions; $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)^T$ are the nonlinear covariate parameters; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ are the linear covariate parameters; and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)^T$ are the parameters for the spline basis functions.

As in the proposed PH model, some parameters have constraints required by the proposed PO model. Parameter transformations are applied for those parameters to avoid difficulties finding the maximum likelihood estimates in the optimization process. The parameter constraints are the same as applied in the previous chapter with the proposed PH model.

In summary, the parameter constraints are:

$$\rho_1, \ldots, \rho_r > 0,$$

$$\tau_1, \ldots, \tau_s > 0,$$

$$\phi > 0,$$

$$\| \boldsymbol{\alpha} \| = 1, \text{ and } \alpha_q > 0,$$

$$\psi(0) = 0.$$

For parameter $\boldsymbol{\rho}$, $\boldsymbol{\tau}$ and ϕ, a logorithmic transformation is applied. That is, let

$$\xi_k = \log \rho_k, \quad k = 1, \ldots, r.$$

$$\zeta_l = \log \tau_l, \quad l = 1, \ldots, s,$$

$$\varrho = \log \phi.$$

For $\boldsymbol{\alpha}$, two steps of transformation are applied. Details about the intermediate steps can be seen in Formulas (2.3) and (2.4). After the transformation, the relationship between $\boldsymbol{\alpha}$ and the new parameter $\boldsymbol{\varphi}$ is:

$$\begin{aligned}
\alpha_1 &= \sin\left[\frac{e^{\varphi_1}-1}{e^{\varphi_1}+1}\frac{\pi}{2}\right]\sin\left[\frac{e^{\varphi_2}-1}{e^{\varphi_2}+1}\frac{\pi}{2}\right]\cdots\sin\left[\frac{e^{\varphi_{q-1}}-1}{e^{\varphi_{q-1}}+1}\frac{\pi}{2}\right] \\
\alpha_2 &= \sin\left[\frac{e^{\varphi_1}-1}{e^{\varphi_1}+1}\frac{\pi}{2}\right]\sin\left[\frac{e^{\varphi_2}-1}{e^{\varphi_2}+1}\frac{\pi}{2}\right]\cdots\cos\left[\frac{e^{\varphi_{q-1}}-1}{e^{\varphi_{q-1}}+1}\frac{\pi}{2}\right] \\
&\cdots \\
\alpha_q &= \cos\left[\frac{e^{\varphi_1}-1}{e^{\varphi_1}+1}\frac{\pi}{2}\right]
\end{aligned} \tag{3.1}$$

After these transformations, the new parameters to estimate with no constraints

are:

$$\boldsymbol{\theta}^* = (\varrho, \boldsymbol{\xi}^T, \boldsymbol{\zeta}^T, \boldsymbol{\varphi}^T, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T,$$

where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_r)^T$, $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_s)^T$, $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_{q-1})^T$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)^T$.

As an unconstrained problem, standard optimization algorithms, such as Newton-Raphson method, can be used to find the the maximum likelihood estimators of $\boldsymbol{\theta}^*$. Similar to the PH case, for the simulation studies and real data analysis, Formula (2.5) is used to calculate the variance estimate of $\hat{\boldsymbol{\theta}}^*$.

After the variance estimate for $\hat{\boldsymbol{\theta}}^*$, $\widehat{\text{var}}(\hat{\boldsymbol{\theta}}^*)$, is found, the delta method is applied to calculate the variance estimate for $\hat{\boldsymbol{\theta}}$. That is, let $\boldsymbol{\theta} = G(\boldsymbol{\theta}^*)$, then by the delta method, $\hat{\boldsymbol{\theta}}$ is asymptotically normal with an estimated asymptotic variance-covariance matrix given in Formula (2.6). Then the estimate of $\text{var}(\hat{\boldsymbol{\theta}}^*)$ can be obtained accordingly.

3.3 Simulation Studies

In this section, three simulation studies are conducted to evaluate the performance of the proposed PO model. The first simulation study gives the estimates of the parameters of interest: nonlinear covariate coefficients $\boldsymbol{\alpha}$, linear covariate coefficient β and the association parameter ϕ. The results show that the estimates of these parameters are fairly close to the true values. The estimates of the baseline cumulative hazard function $\Lambda_{01}(\cdot)$ and the smooth function $\psi(\cdot)$ are also given. In the second simulation study the proposed PO model is compared with the *PO linear model*. The PO linear model refers to the model which has the same structure as the proposed PO model, but with only linear effect assumed between $\log\frac{S}{1-S}$ and $\log\frac{S_0}{1-S_0}$. Results show that the parameter estimates of the PO linear model are not as good as that of the proposed PO model when nonlinear structure exists. The estimates of the

covariate effect from the linear model completely miss the "S" shape of the true nonlinear function while the proposed model can grasp the nonlinear pattern. The third simulation study assesses the proposed model under various scenarios. The summary of the model performance when some parameters change, such as the censoring rate, sample size, and association degree, is given.

In the simulation study, the true marginal function is assumed to have the following proportional odds form:

$$\frac{S_j(t_j|x_j,\boldsymbol{v}_j)}{1-S_j(t_j|x_j,\boldsymbol{v}_j)} = \exp\{\beta x_j + 3\sin(2\boldsymbol{\alpha}^T\boldsymbol{v}_j)\}\frac{S_{0j}(t_j)}{1-S_{0j}(t_j)}, \quad j=1,\ 2,$$

where the trigonometric function $3\sin(2\boldsymbol{\alpha}^T\boldsymbol{v}_j)$ is used as the nonlinear smooth function $\psi(\cdot)$. A Weibull distribution with unit scale parameter and shape parameter $p=1.5$ was used for the baseline survival function. The association parameter ϕ was set to 0.5, corresponding to fairly strong correlation between survival time T_1 and T_2.

The nonlinear covariates $\boldsymbol{V}_1$, $\boldsymbol{V}_2$ were generated by the uniform distribution. That is, $V_{k1} \sim U(-1,1)$, $V_{k2} \sim U(-1,1)$, $k=1,2,3$. The linear covariates (x_1,x_2) were chosen from four sets of combinations (0,0), (0,1), (1,0), (1,1), each with probability one quarter. The nonlinear regression coefficients were set to $\boldsymbol{\alpha}=(1,1,1,)^T/\sqrt{3}$ and the linear regression coefficient is $\beta=1$.

Using the Clayton model the bivariate survival times can be simulated. Times (t_{i1},t_{i2}), $i=1,\ldots,n$, were generated through the inverse transform sampling method, and have the form:

$$t_{i1} = -\log\frac{u_{i1}}{(1-u_{i1})\exp\{\beta x_{i1}+3\sin(2\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\}+u_{i1}},$$

$$t_{i2} = -\log\frac{A}{(1-A)\exp\{\beta x_{i2}+3\sin(2\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\}+A},$$

where

$$A = \left[u_{i1}^{\phi^{-1}} u_{i2}^{(\phi+1)^{-1}} + 1 - u_{i1}^{\phi^{-1}}\right]^{-\phi},$$

and $u_{i1} \sim U[0,1]$, $u_{i2} \sim U[0,1]$. The censoring time C_j was set to be 1.2 and the corresponding censoring rate were about 45% for both of the two variates.

The strategies for choosing the cut points for the piecewise constants and the breakpoints for the spline function are the same as for the proposed PH model. See Section 2.2.1 and Section 2.2.2 for details. We take $r = s = 4$, and $d = 6$. That is, four piecewise constants are used and the spline uses three breakpoints and six basis functions.

3.3.1 Evaluate the Performance of the Proposed Model

Based on the above settings, 200 simulation runs were conducted with sample size 200. The average estimates of the parameters $\boldsymbol{\varphi}$ and ϱ (after transformation) and the original parameters $\boldsymbol{\alpha}$, β and ϕ (before transformation) are given in Table 3.1. It shows that the performance of the proposed model is quite good. The bias of the parameters is small. The standard deviation is similar to the average of the standard errors. The coverage probability is close to the nominal of 95%.

Table 3.1: The estimates for the covariate coefficients $\boldsymbol{\varphi}$ and ϱ (after transformation), and $\boldsymbol{\alpha}$, β and ϕ (before transformation) based on 200 runs

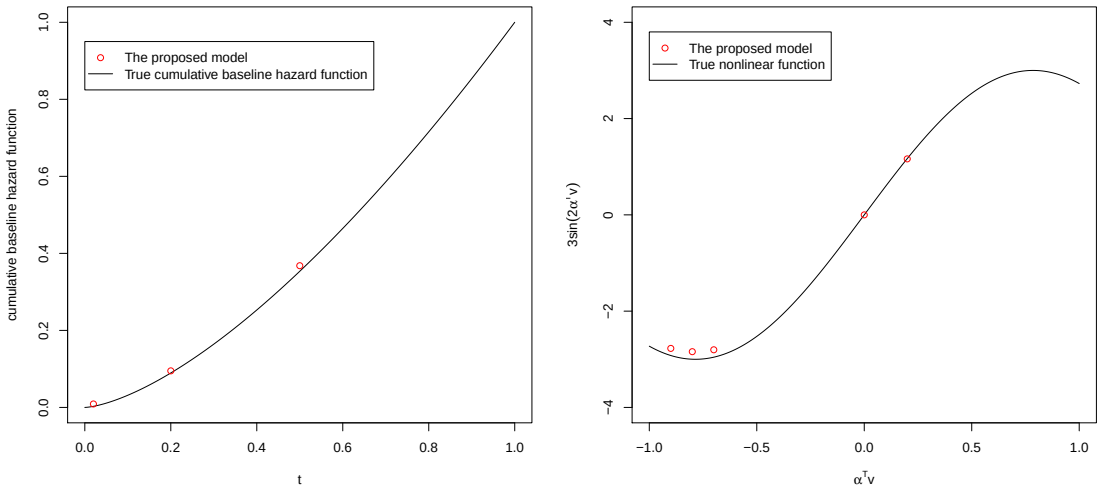
	After transformation			Before transformation				
	φ_1	φ_2	ϱ	α_1	α_2	α_3	β	ϕ
True	1.412	1.099	-0.693	0.577	0.577	0.577	1.000	0.500
Bias	-0.008	-0.001	-0.110	-0.003	-0.002	0.003	-0.035	-0.042
SD($\hat{\cdot}$)	0.057	0.075	0.210	0.027	0.029	0.023	0.178	0.099
A{SE($\hat{\cdot}$)}	0.062	0.065	0.221	0.025	0.025	0.025	0.183	0.103
Cov.prob.	0.950	0.905	0.940	0.940	0.895	0.950	0.940	0.875

Table 3.2 gives the simulation-average estimates for the cumulative baseline hazard function at the points (0.02, 0.2, 0.5), and for the nonlinear function $\psi(\cdot)$ at the points

(-0.9, -0.8, -0.7, 0.2). The average estimates are fairly close to the true values. Figure 3.1 gives a more direct display of the above observations. The red dots represent the estimates at the chosen points. The black lines are the true values of functions Λ_{01} and $\psi(\cdot)$. Clearly, the red dots follow the curve of both functions very well which verifies the good performance of the model. Note that $\psi(x)$ is set to be 0 at $x = 0$ by the model constraints, which is also shown in the figure.

Table 3.2: The estimates for $\Lambda_{01}(\cdot)$ and $\psi(\cdot)$ of one variate based on 200 runs

	x	True	Estimate	SD
$\Lambda_{01}(x)$	0.020	0.003	0.009	0.002
	0.200	0.089	0.095	0.025
	0.500	0.354	0.368	0.072
$\psi(x)$	-0.900	-2.922	-2.774	0.362
	-0.800	-2.999	-2.842	0.342
	-0.700	-3.002	-2.802	0.330
	0.200	1.168	1.163	0.198



(a) Λ_{01} (b) $\psi(\cdot)$

Figure 3.1: Cumulative baseline hazard function Λ_{01} and nonlinear smooth function $\psi(\cdot)$ with the estimates at the chosen points. Black lines represent the true values, and red dots represent the average estimates of selected points.

3.3.2 Compare the Proposed Model with PO Linear Model

In this section, the proposed PO model is compared with the PO linear model. The marginal function of the PO linear model has the following form:

$$\frac{S_j(t_j|x_j, \boldsymbol{v}_j)}{1 - S_j(t_j|x_j, \boldsymbol{v}_j)} = \exp\{\beta^T x_j + \tilde{\boldsymbol{\alpha}}^T \boldsymbol{v}_j\} \frac{S_{0j}(t_j)}{1 - S_{0j}(t_j)}, \quad j = 1,\ 2,$$

where β is the coefficient for linear covariate x_j and $\tilde{\boldsymbol{\alpha}}$ is the coefficient for nonlinear covariate $\boldsymbol{v}_j$. As with the proposed PH model, $\hat{\tilde{\boldsymbol{\alpha}}}$ is factored into $\hat{b}\hat{\boldsymbol{\alpha}}$, where $\hat{\boldsymbol{\alpha}}$ has unit norm.

In this simulation study, the same parameter settings were used as in the first simulation study. The sample size and simulation runs were set as 200. Table 3.3 lists the estimates of parameters $\boldsymbol{\alpha}$, β and ϕ from both models. The proposed PO model has an overall better performance compared to the PO linear model. The parameter average estimates of the proposed model for 200 runs are generally close to the true values, and the standard deviations are also small. As a contrast, the estimates of parameter β and ϕ of the linear model are further away from the true values, and the standard deviations are relatively large. Table 3.4 compares the two models at chosen points for Λ_{01} and $\psi(\cdot)$. Results indicate that the proposed model has better estimates at the chosen points. Figure 3.2 shows these comparisons more clearly. In Figure 3.2b the linear model (the blue line) completely misses the "S" curve of the smooth function from the true model (the black line), while the proposed model (the red dots) can capture the nonlinear pattern. Compared to the PH linear model, however, this linear model chooses a line that approximates the nonlinear function better.

Table 3.3: The estimates comparison of two models for the covariate coefficients $\boldsymbol{\alpha}$, β and ϕ based on 200 runs

	α_1	α_2	α_3	β	ϕ
True	0.577	0.577	0.577	1.000	0.500
Estimate(proposed)	0.574	0.581	0.575	0.964	0.442
SD($\hat{\cdot}$)(proposed)	0.028	0.026	0.027	0.168	0.100
Estimate(linear)	0.573	0.579	0.576	0.852	0.853
SD($\hat{\cdot}$)(linear)	0.036	0.037	0.037	0.175	0.287

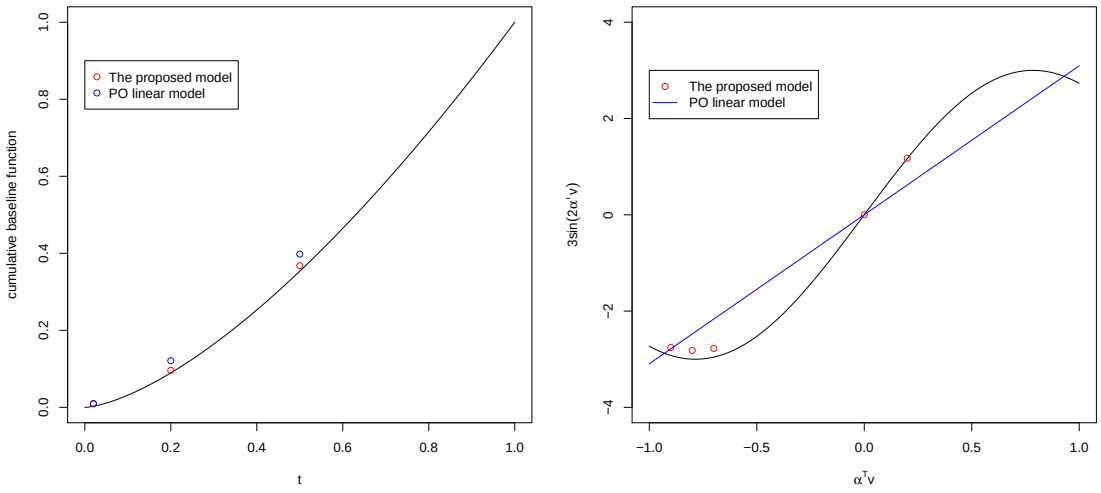
Table 3.4: The estimates comparison of two models for $\Lambda_{01}(\cdot)$ and $\psi(\cdot)$ based on 200 runs

			Proposed		Linear	
	x	True	Estimate	SD	Estimate	SD
	0.020	0.003	0.009	0.002	0.010	0.002
$\Lambda(x)$	0.200	0.089	0.096	0.026	0.112	0.026
	0.500	0.354	0.368	0.067	0.398	0.058
	-0.900	-2.922	-2.758	0.374	-2.786	0.258
	-0.800	-2.999	-2.817	0.355	-2.476	0.229
$\psi(x)$	-0.700	-3.002	-2.776	0.338	-2.166	0.200
	0.200	1.168	1.176	0.179	0.619	0.057

3.3.3 Assess the Proposed Model Under Various Scenarios

In this section, the proposed model is assessed when changing the parameter settings. The shape parameter p of the Weibull baseline function is set to be 1.5. The model is assessed when the following parameters change. The sample size is set as 80 and 200. The censoring rate of 20% and 50% are applied. The association parameter takes values 0.5, 1 and 4, representing the association varying from relatively strong to weak.

Table 3.5 gives the summary of the result of 200 runs. From the results the following patterns are observed. As the sample size increases from 80 to 200, the estimation performance improves. The standard deviations decrease. Censoring rate 20% has



(a) Λ_{01} (b) $\psi(\cdot)$

Figure 3.2: Compare the proposed model and PO linear model on cumulative baseline hazard function Λ_{01} and nonlinear smooth function $\psi(\cdot)$. Black line represents the true curve, blue line and dot are the estimates from the PH linear model, and the red line and dots are the estimates from the proposed model.

a generally better result than censoring rate 45%. As the censoring rate increases, the standard deviations increase. When the value of the association parameter ϕ increases, that is, the association degree drops from strong to weak, the estimation performance of $\hat{\phi}$ decreases. The bias increases and the standard deviations increase as well. When $\phi = 4$, the standard deviations and the average standard errors of ϕ are not so close to each other, and there appears to be some bias. The reason for the poor performance for $\phi = 4$ might be, when the association is weak, the likelihood function is too flat to find a good local optimum for ϕ.

The results for $p = 0.5$ are not shown here. It was found that when the shape parameter of the Weibull baseline function is $p = 0.5$, the estimation performance is not good. The poor performance in this case is likely due to too few pieces being used in the piecewise constant baseline function. It has been observed that when the baseline function is known, the estimation performance is good, even for $p = 0.5$.

Increasing the number of the piecewise constants might solve this problem, although due to time constraints this possibility was not explored.

3.4 Real Data Analysis

3.4.1 Introduction

The proposed PO model is applied on the Busselton Health Study (Knuiman et al., 1994) data. Background on this data set has been reviewed in Sections 1.1.2. As with the analysis in Section 2.4.1, four health-related covariates are included in the model: age (AGE), body mass index (BMI), cholesterol level (CHOL) and smoking status (SMOKE). The first three covariates are continuous variables which are regarded as the nonlinear covariates. The fourth covariate is a discrete variable which is regarded as the linear covariate. Standardization is applied to the covariates to avoid extreme values.

3.4.2 Analysis

The strategy of choosing the cut points for the piecewise constants and the breakpoints of the spline function is the same as used in the proposed PH model (Section 2.2.1 and 2.2.2). The interior cut points are (78.1, 85.2, 90.9) for female and (74.4, 81.5, 87.4) for male. The breakpoints are chosen to divide the survival time range of each variate into four intervals, with each interval containing a roughly equal number of data points.

The parameters to be estimated are ϕ, $\boldsymbol{\rho}$, $\boldsymbol{\tau}$, $\boldsymbol{\alpha}$, β and $\boldsymbol{\gamma}$. The baseline hazard functions have four constant pieces, so the parameter vectors $\boldsymbol{\rho}$ (for females) and $\boldsymbol{\tau}$ (for males) have length four each. The Spline function parameter $\boldsymbol{\gamma}$ has six elements. There are three nonlinear covariates: AGE, BMI and CHOL, with coefficients α_1, α_2 and α_3 respectively. The single linear covariate is SMOKE, with coefficient β.

Table 3.5: Simulation result for Weibull with shape parameter $p = 1.5$ of 200 runs

			α_1			α_2			α_3		
ϕ	P(%)	n	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}
0.5	20%	80	0.000	0.035	0.033	0.000	0.034	0.033	-0.003	0.036	0.033
		200	-0.001	0.020	0.019	0.002	0.021	0.019	-0.001	0.021	0.019
	50%	80	-0.006	0.051	0.045	-0.001	0.051	0.045	0.000	0.050	0.044
		200	-0.001	0.027	0.025	0.000	0.026	0.026	-0.001	0.027	0.026
1	20%	80	-0.004	0.042	0.041	0.005	0.038	0.041	-0.006	0.042	0.042
		200	0.001	0.023	0.025	-0.001	0.024	0.025	-0.001	0.022	0.025
	50%	80	0.000	0.059	0.049	0.005	0.057	0.048	-0.014	0.054	0.050
		200	0.000	0.032	0.030	0.000	0.031	0.030	-0.003	0.032	0.030
4	20%	80	0.000	0.049	0.047	0.004	0.045	0.047	-0.010	0.050	0.048
		200	-0.001	0.028	0.029	0.000	0.028	0.029	0.000	0.030	0.029
	50%	80	-0.001	0.055	0.055	-0.003	0.052	0.055	-0.003	0.051	0.054
		200	-0.001	0.030	0.032	-0.002	0.030	0.032	0.001	0.032	0.032

			β			ϕ		
ϕ	P(%)	n	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}
0.5	20%	80	-0.031	0.204	0.244	-0.083	0.120	0.120
		200	-0.057	0.123	0.137	-0.065	0.063	0.070
	50%	80	-0.005	0.312	0.333	-0.100	0.181	0.180
		200	-0.048	0.173	0.187	-0.058	0.099	0.103
1	20%	80	-0.014	0.283	0.301	-0.143	0.375	0.346
		200	-0.033	0.158	0.173	-0.097	0.194	0.197
	50%	80	-0.005	0.328	0.388	0.294	1.735	1.883
		200	-0.014	0.224	0.221	-0.037	0.394	0.344
4	20%	80	0.056	0.331	0.346	-0.831	1.860	3.257
		200	-0.038	0.159	0.197	-0.103	1.837	2.469
	50%	80	-0.001	0.365	0.423	-1.215	2.037	4.532
		200	-0.016	0.020	0.237	-0.490	1.788	3.069

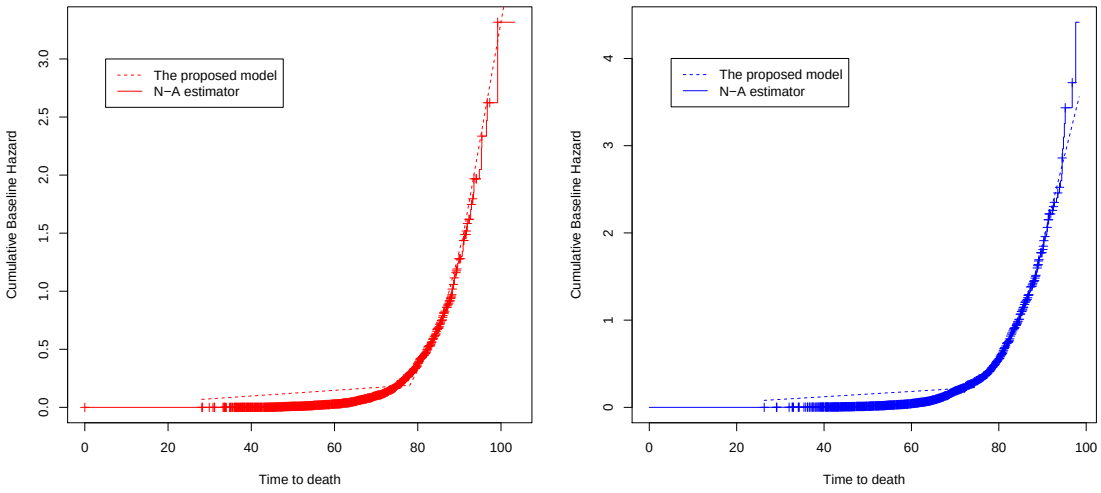
Table 3.6 gives a summary of the parameter estimates. Among the three continuous covariates, AGE plays a dominating role on the time to death, while BMI has almost no effect. The significant covariates are AGE and SMOKE. The association parameter ϕ is significant, with value 5.004, which shows mild association between females and males. These discoveries are consistent with the ones from the proposed PH model.

Table 3.6: The estimates for the covariate coefficients $\boldsymbol{\alpha}$, β and ϕ of the data analysis

	α_1	α_2	α_3	β	ϕ
Estimate	1.000	0.000	0.031	-0.188	5.004
SE($\hat{\cdot}$)	0.001	0.036	0.042	0.099	2.050

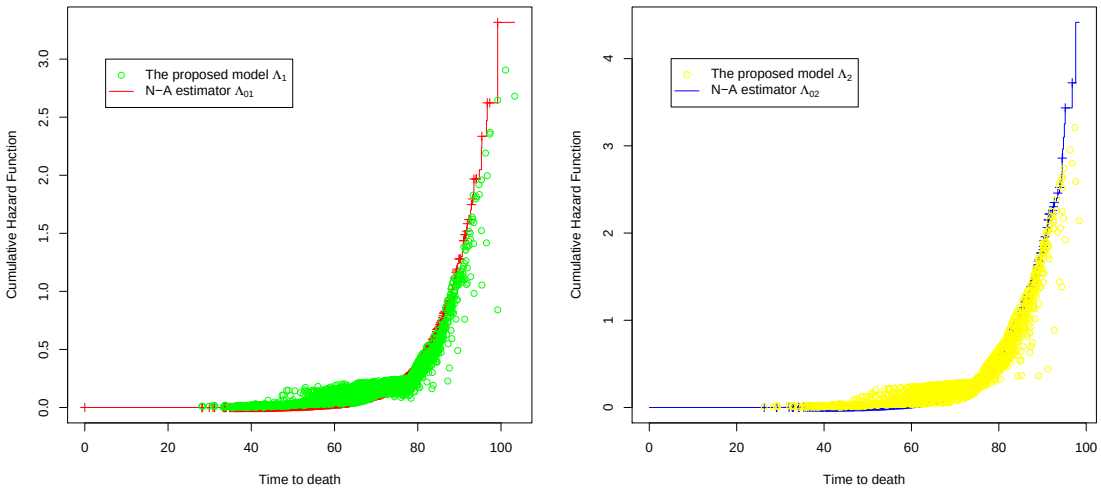
In Figure 3.3, the baseline cumulative hazard function from the proposed model is compared with N-A estimator. The two estimates have very similar shapes. It is noticed that the model estimates seem generally above the N-A line in the flat part. This is because the flat part comes from the first constant piece of the baseline estimate. Figure 3.4 shows the cumulative hazard function after adding the covariate effect, and is compared with N-A estimator of the baseline cumulative hazard function. The cumulative hazard function value for each individual is represented by each dot that are scattered around N-A estimator curve. Note in Figure 3.3, the cumulative baseline hazard function is a non-decreasing line. After including the covariate effect in Figure 3.4, each individual's cumulative hazard function is different, so the points no longer lie on a non-decreasing curve.

Figure 3.5 gives the curve of the nonlinear function $\psi(\cdot)$ for females. The nonlinear relationship between $\psi(\cdot)$ and $\boldsymbol{\alpha}^T\boldsymbol{v}$ is clearly shown by the graph. As the proposed model assumes the same parameters of the nonlinear functions for both female and male, the nonlinear function for males has similar patterns. Therefore it is not shown here.



(a) Female (b) Male

Figure 3.3: Compare the cumulative baseline hazard function Λ_0 of the proposed model with N-A estimator for female and male



(a) Female (b) Male

Figure 3.4: Compare the cumulative hazard function Λ considering covariate effects with N-A estimator of Λ_0 for female and male. Dots represent Λ for individuals of the proposed model, and lines represent N-A estimator of Λ_0.

Figure 3.6 to 3.8 show how the nonlinear function $\psi(\cdot)$ changes when one covariate changes, while the other two covariates are fixed at their medians. The differences

between the females and males are due to the different medians of the fixed covariates for different genders. For Figure 3.6, as age increases, $\psi(\cdot)$ experiences a decreasing stage first and then turns to increase roughly after 49 for female and 52 for male. In Figure 3.7, as body mass index increases, $\psi(\cdot)$ decreases. Similarly in Figure 3.8, as cholesterol level increases, $\psi(\cdot)$ decreases.

Recall the marginal survival function of the proposed model can, after some manipulation, be written as follows, taking females as the example:

$$S_{i1}(y_{i1}) = 1 - \frac{1 - \exp\{-\Lambda_{01}(y_{i1})\}}{\exp\{\boldsymbol{\beta}^T \boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i1}) - \Lambda_{01}(y_{i1})\} + 1 - \exp\{-\Lambda_{01}(y_{i1})\}}.$$

The right hand side of this function has the form $1 - \frac{a}{x+a}$, where $a > 0$ (because $1 - \exp\{-\Lambda_{01}(y_{i1})\} > 0$) and x is a monotone increasing function of the single index structure. Note that $1 - \frac{a}{x+a}$ is itself an increasing function of x, $x > 0$. Hence, when $\psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i1})$ increases, $S_{i1}(y_{i1})$ increases, and when $\psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i1})$ decreases, $S_{i1}(y_{i1})$ decreases. The same relationship holds for $S_{i2}(y_{i2})$ when $\psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i2})$ changes.

Based on the above discussion, Figures 3.6 to 3.8 reveal the following information. In Figure 3.6, when age increases, survival function $S_{i1}(y_{i1})$ decreases roughly before age 50, and increases but stays in relatively low value after age 50. For Figure 3.7, as body mass index increases, $S_{i1}(y_{i1})$ decreases. For Figure 3.8, as cholesterol level increases, $S_{i1}(y_{i1})$ decreases.

3.5 Conclusion

This chapter proposes extensions to the PO model for the regression analysis of multivariate survival data. A single index structure is added to model the nonlinear covariate effect and a spline function is used to estimate the nonlinear relationship in the marginal survival function. The baseline hazard function is modeled by a weakly parametric method. One major advantage of this model is its ability to flexibly model

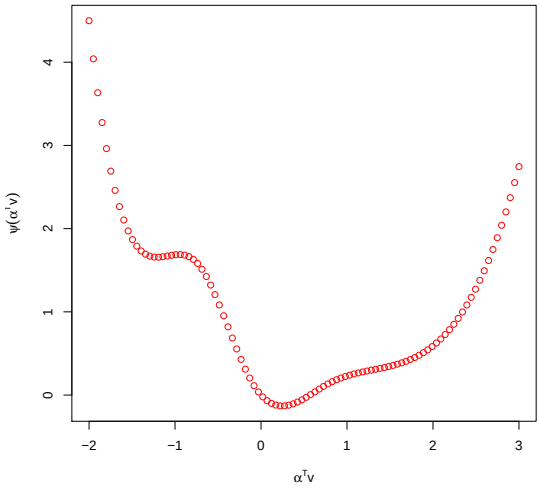
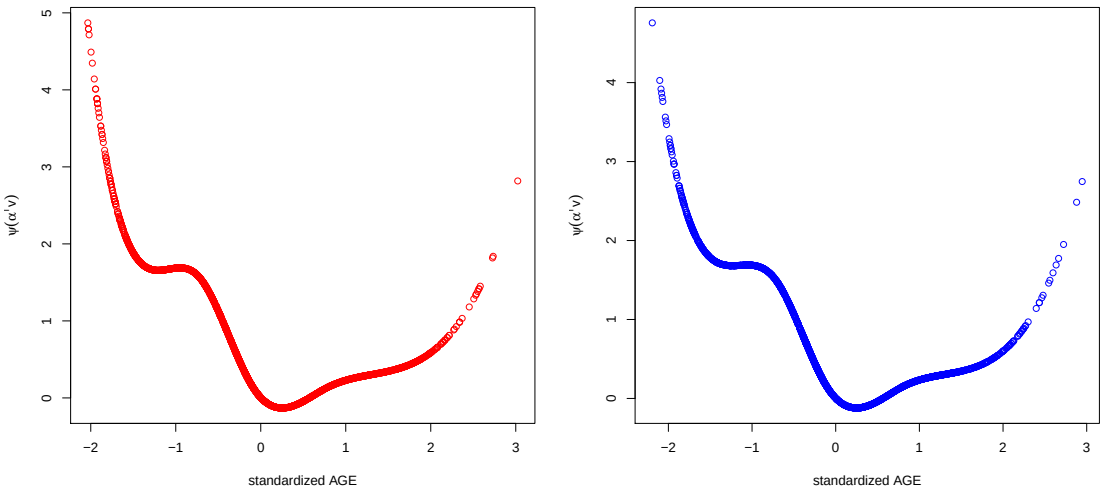


Figure 3.5: Nonlinear function $\psi(\cdot)$ for female

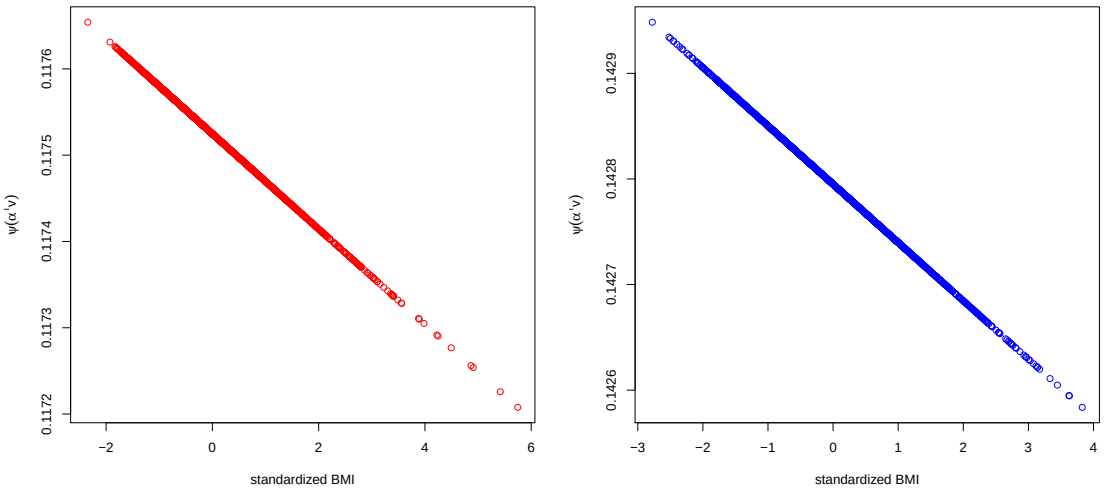


(a) Female (b) Male

Figure 3.6: Nonlinear function $\psi(\cdot)$ as the covariate age changes

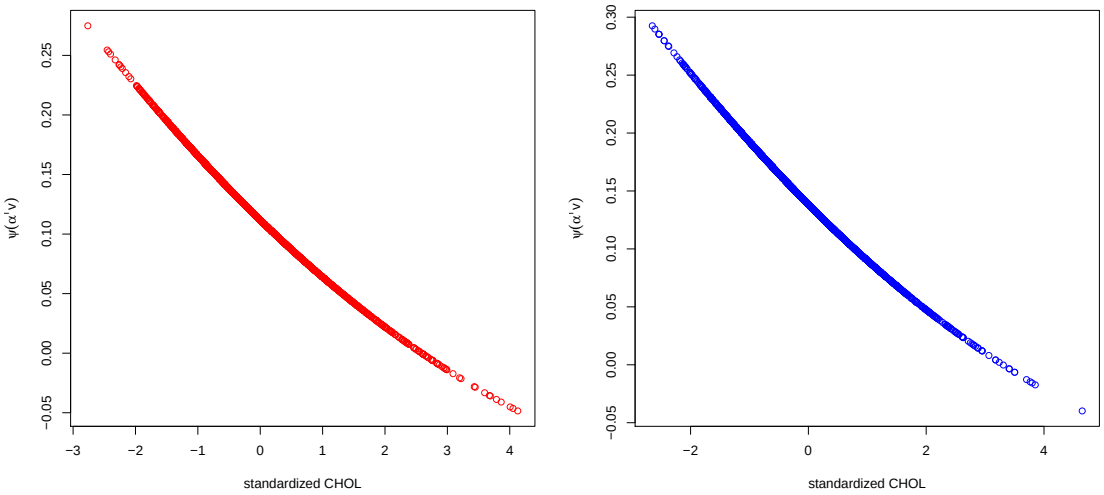
the covariate effect, either linear or nonlinear. The other advantage of this model is the ability to use the full likelihood, making estimation and inference convenient.

The proposed PO model was shown through simulation studies to have better overall performance than the PO linear model when a nonlinear covariate structure exists. The model is able to capture the nonlinear covariate relationship and give good



(a) Female (b) Male

Figure 3.7: Nonlinear function $\psi(\cdot)$ as the covariate body mass index changes



(a) Female (b) Male

Figure 3.8: Nonlinear function $\psi(\cdot)$ as the covariate cholesterol level changes

estimates of the parameters. In the real data analysis, it was found that there is mild association existing between husbands and wives. The estimated baseline cumulative hazard function from the proposed model has a very similar shape with N-A estimator, and the covariate effects can be estimated for each individual. Nonlinear patterns are

clearly shown in the shape of the smooth function $\psi(\cdot)$.

All of these observations indicate that the proposed PO model is a flexible and effective way to deal with multivariate survival data.

Chapter 4

Flexible Partially Linear Single Index Generalized Transformation Regression Model for Multivariate Survival Data

4.1 Introduction

The linear transformation model has gained a lot of attention in recent years due to its flexibility. The two popularly used survival models, the proportional hazards model and the proportional odds model, are special cases of the linear transformation model.

Cheng et al. (1995) provide a nice summary of how the two models are special cases of the transformation model, in the univariate case. Let S be the survival function of T given the covariates $\boldsymbol{X}$. Then the PH model can be written as

$$\log[-\log\{S(t|\boldsymbol{x})\}] = H(t) + \boldsymbol{\beta}^T\boldsymbol{x}, \tag{4.1}$$

where $H(t)$ is an unknown increasing function, and $\boldsymbol{\beta}$ is the coefficients with dimension $p \times 1$. One can show that (4.1) is in fact a PH model by deriving the hazard function corresponding to survival function $S(t)$ by using the relationship $\lambda(t|\boldsymbol{x}) = -\frac{d}{dt}\{\log S(t|\boldsymbol{x})\}$; it has the PH form of equation (1.2). Similarly, the PO model can be written as

$$-\operatorname{logit}\{S(t|\boldsymbol{x})\} = H(t) + \boldsymbol{\beta}^T\boldsymbol{x}, \tag{4.2}$$

where $\operatorname{logit}(x) = \log\{x/(1-x)\}$. It can be easily shown that the survival odds ratio computed from (4.2) has the form of the PO model as in (1.3).

Equation (4.1) and (4.2) can be unified as

$$g\{S(t|\boldsymbol{x})\} = H(t) + \boldsymbol{\beta}^T\boldsymbol{x}, \tag{4.3}$$

where $g(\cdot)$ is a known decreasing function. Moreover, it can be shown that (4.3) is equivalent to the linear transformation model:

$$H(T) = -\boldsymbol{\beta}^T\boldsymbol{x} + \epsilon, \tag{4.4}$$

where ϵ is random error which has the distribution function $F = 1 - g^{-1}$. The PH model and the PO model are special cases of (4.4) when ϵ follows the extreme-value distribution and the standard logistic distribution, respectively.

Equivalently, one can specify the hazard function of ϵ such that the PH and PO models are special cases of the transformation model (Chen et al., 2002). Let ϵ have hazard function parameterized by r, with the form $\lambda(\epsilon) = \exp(\epsilon)/\{1 + r\exp(\epsilon)\}$. The PH and PO models correspond to $r = 0$ (extreme-value hazard) and $r = 1$ (logistic hazard), respectively.

When $r = 0$, it can be shown that the hazard function of t is

$$\lambda(t|\boldsymbol{x}) = \exp(H(t))H'(t)\exp(\boldsymbol{\beta}^T\boldsymbol{x}),$$

which has the form of a PH model with the baseline hazard function

$$\lambda_0(t) = \exp(H(t))H'(t).$$

When $r = 1$, one can show the survival odds of t is

$$\frac{S(t|\boldsymbol{x})}{1 - S(t|\boldsymbol{x})} = \exp(-\boldsymbol{\beta}^T\boldsymbol{x})\exp(-H(t)),$$

which has the form of a PO model with the baseline survival function

$$S_0(t) = \exp(-H(t))/(1 + \exp(-H(t))).$$

Many inference procedures for the transformation model have been introduced in the recent literature. Cheng et al. (1995) propose a general estimation method for linear transformation models with censored data. Their method is further developed in Cheng et al. (1997), Fine et al. (1998) and Cai et al. (2000). A key assumption in their approach is that the censoring variable is independent of the covariates, which makes it possible to use the Kaplan-Meier method to estimate the survival function. Chen et al. (2002) relax this assumption and proposed an estimating equation approach. More recently, the articles by Zeng and Lin (2006, 2007) both consider nonparametric maximum likelihood estimation in a class of semiparametric transformation models. The first article allows time-dependent covariates, while the second one handles recurrent-event data with random effects. Moreover, Lu and Zhang (2010) propose a partially linear transformation model by incorporating nonlinear covariate

effects, and study a martingale estimating equation approach.

We propose to use a partially linear single index transformation model as the marginal function to study multivariate survival data. A smooth function $\psi(\cdot)$ is added to handle the nonlinear covariate effect. The single index model $\psi(\boldsymbol{\alpha}^T\boldsymbol{v})$ is applied to reduce the high dimensional nonlinear covariate $\boldsymbol{v}$ into a scalar. One type of copula model, the Clayton model (Clayton, 1978), is used to incorporate the association among the variates. A spline function approach is applied to model the unknown smooth function.

The transformation model places a restriction on the unknown function $H(t)$, that is, $H(0) = -\infty$ (Chen et al., 2002). It is difficult to satisfy this constraint with the weakly parametric approaches such as the splines used in the previous chapters. Therefore, while the likelihood is worked out in the general case, $H(t)$ is fixed to be $\log(t)$ in the simulation study for simplicity. Section 4.2 introduces the details of the proposed model. Simulation studies are provided in Section 4.3. At the end, some conclusions are summarized.

4.2 The Proposed Model

The proposed marginal transformation model is written as:

$$H_j(T_j|\boldsymbol{x}_j, \boldsymbol{v}_j) = -\boldsymbol{\beta}^T\boldsymbol{x}_j - \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_j) + \epsilon_j, \quad j = 1, 2,$$

where H is an unknown increasing transformation function with $H(0) = -\infty$, and $\boldsymbol{x}_j$ is a linear covariate with the coefficient $\boldsymbol{\beta}$ (dimension $p \times 1$). The single index model $\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_j)$ is applied to model nonlinear effects with the nonlinear covariate $\boldsymbol{v}_j$ and its coefficient $\boldsymbol{\alpha}$ (dimension $q \times 1$). In the general case, the random variable ϵ_j can take any fixed distribution independent of $\boldsymbol{x}_j$ and $\boldsymbol{v}_j$. As discussed above, in this work the distribution of ϵ is parametrized such that $\lambda(\epsilon) = \frac{\exp(\epsilon)}{1+r\exp(\epsilon)}$.

A spline is used to model the smooth function $\psi(\cdot)$. Similar to the methods used in the proposed PH model and PO model, M-spline and I-spline functions are used to estimate $\psi'(\cdot)$ and $\psi(\cdot)$ respectively. To review the details of the spline functions, see Section 2.2.2.

4.2.1 The Likelihood Function

Using the hazard function of ϵ, $\lambda(\epsilon) = \frac{\exp(\epsilon)}{1+r\exp(\epsilon)}$, the survival function and density function of ϵ can be calculated. The transformation model gives the relationship between ϵ_j and T_j, and the marginal survival function of t_j can be found as follows:

$$\begin{aligned} S_{i1}(t_{i1}) &= 1+\Big\{1+r\exp\Big[H(t_{i1})+\boldsymbol{\beta}^T\boldsymbol{x}_{i1}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big]\Big\}^{-\frac{1}{r}} \\ &\quad -\Big\{1+r\exp\Big[H(0)+\boldsymbol{\beta}^T\boldsymbol{x}_{i1}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big]\Big\}^{-\frac{1}{r}}, \end{aligned}$$

$$\begin{aligned} S_{i2}(t_{i2}) &= 1+\Big\{1+r\exp\Big[H(t_{i2})+\boldsymbol{\beta}^T\boldsymbol{x}_{i2}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big]\Big\}^{-\frac{1}{r}} \\ &\quad -\Big\{1+r\exp\Big[H(0)+\boldsymbol{\beta}^T\boldsymbol{x}_{i2}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big]\Big\}^{-\frac{1}{r}}. \end{aligned}$$

The log likelihood function (equation 2.2) can be calculated using the Clayton bivariate model (equation 1.1). The entries of the log likelihood function are written as follows:

$$\begin{aligned} \frac{-\partial S(t_{i1},t_{i2})}{\partial t_{i1}} &= \frac{-\partial\Big[S_{i1}(t_{i1})^{-\phi^{-1}}+S_{i2}(t_{i2})^{-\phi^{-1}}-1\Big]^{-\phi}}{\partial t_{i1}} \\ &= -\Big[S_{i1}(t_{i1})^{-\phi^{-1}}+S_{i2}(t_{i2})^{-\phi^{-1}}-1\Big]^{-\phi-1}\Big[S_{i1}(t_{i1})^{-\phi^{-1}-1}\Big]\frac{\partial S_{i1}(t_{i1})}{\partial t_{i1}}, \end{aligned}$$

where

$$\begin{aligned}\frac{\partial S_{i1}(t_{i1})}{\partial t_{i1}} &= -\Big\{1+r\exp\Big[H(t_{i1})+\boldsymbol{\beta}^T\boldsymbol{x}_{i1}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big]\Big\}^{-\frac{1}{r}-1}\\ &\quad\times\exp\Big[H(t_{i1})+\boldsymbol{\beta}^T\boldsymbol{x}_{i1}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big]H'(t_{i1}),\end{aligned}$$

and

$$\begin{aligned}\frac{-\partial S(t_{i1},t_{i2})}{\partial t_{i2}} &= \frac{-\partial\Big[S_{i1}(t_{i1})^{-\phi^{-1}}+S_{i2}(t_{i2})^{-\phi^{-1}}-1\Big]^{-\phi}}{\partial t_{i2}}\\ &= -\Big[S_{i1}(t_{i1})^{-\phi^{-1}}+S_{i2}(t_{i2})^{-\phi^{-1}}-1\Big]^{-\phi-1}\Big[S_{i2}(t_{i2})^{-\phi^{-1}-1}\Big]\frac{\partial S_{i2}(t_{i2})}{\partial t_{i2}},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial S_{i2}(t_{i2})}{\partial t_{i1}} &= -\Big\{1+r\exp\Big[H(t_{i2})+\boldsymbol{\beta}^T\boldsymbol{x}_{i2}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big]\Big\}^{-\frac{1}{r}-1}\\ &\quad\times\exp\Big[H(t_{i2})+\boldsymbol{\beta}^T\boldsymbol{x}_{i2}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big]H'(t_{i2}).\end{aligned}$$

Thus the joint density of t_{i1} and t_{i2} is

$$\begin{aligned}f(t_{i1},t_{i2}) &= \frac{\partial S(t_{i1},t_{i2})}{\partial t_{i1}\partial t_{i2}}\\ &= \Big[S_{i1}(t_{i1})^{-\phi^{-1}}\Big]\Big[S_{i2}(t_{i2})^{-\phi^{-1}}\Big](1+\phi^{-1})\\ &\quad\times\Big[S_{i1}(t_{i1})^{-\phi^{-1}}+S_{i2}(t_{i2})^{-\phi^{-1}}-1\Big]^{-\phi-2}\\ &\quad\times\Big\{1+r\exp\Big[H(t_{i1})+\boldsymbol{\beta}^T\boldsymbol{x}_{i1}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big]\Big\}^{-\frac{1}{r}-1}\\ &\quad\times\exp\Big[H(t_{i1})+\boldsymbol{\beta}^T\boldsymbol{x}_{i1}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big]H'(t_{i1})\\ &\quad\times\Big\{1+r\exp\Big[H(t_{i2})+\boldsymbol{\beta}^T\boldsymbol{x}_{i2}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big]\Big\}^{-\frac{1}{r}-1}\\ &\quad\times\exp\Big[H(t_{i2})+\boldsymbol{\beta}^T\boldsymbol{x}_{i2}+\psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big]H'(t_{i2}).\end{aligned}$$

The log likelihood function for the case $H(t)=\log(t)$, which the simulation study is based on, is given as the following. Given $\lambda(\epsilon)=\frac{\exp(\epsilon)}{1+r\exp(\epsilon)}$, the marginal survival

functions can be worked out:

$$S_{i1}(t_{i1}) = \Big\{1 + rt_{i1}\exp\Big[\boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big]\Big\}^{-\frac{1}{r}},$$

$$S_{i2}(t_{i2}) = \Big\{1 + rt_{i2}\exp\Big[\boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big]\Big\}^{-\frac{1}{r}}.$$

The entries of the log likelihood function (equation 2.2) are written as follows:

$$\begin{aligned}\frac{-\partial S(t_{i1}, t_{i2})}{\partial t_{i1}} &= \frac{-\partial\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi}}{\partial t_{i1}} \\ &= -\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi-1}\Big[S_{i1}(t_{i1})^{-\phi^{-1}-1}\Big]\frac{\partial S_{i1}(t_{i1})}{\partial t_{i1}},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial S_{i1}(t_{i1})}{\partial t_{i1}} &= -\Big\{1 + rt_{i1}\exp\Big[\boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big]\Big\}^{-\frac{1}{r}-1} \\ &\quad \times\exp\Big[\boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big],\end{aligned}$$

and

$$\begin{aligned}\frac{-\partial S(t_{i1}, t_{i2})}{\partial t_{i2}} &= \frac{-\partial\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi}}{\partial t_{i2}} \\ &= -\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi-1}\Big[S_{i2}(t_{i2})^{-\phi^{-1}-1}\Big]\frac{\partial S_{i2}(t_{i2})}{\partial t_{i2}},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial S_{i2}(t_{i2})}{\partial t_{i1}} &= -\Big\{1 + rt_{i2}\exp\Big[\boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big]\Big\}^{-\frac{1}{r}-1} \\ &\quad \times\exp\Big[\boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big].\end{aligned}$$

Then the joint density of t_{i1} and t_{i2} is

$$\begin{aligned} f(t_{i1}, t_{i2}) &= \frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i1} \partial t_{i2}} \\ &= \Big[S_{i1}(t_{i1})^{-\phi^{-1}}\Big]\Big[S_{i2}(t_{i2})^{-\phi^{-1}}\Big](1+\phi^{-1}) \\ &\quad \times \Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi-2} \\ &\quad \times \Big\{1 + r t_{i1} \exp\Big[\boldsymbol{\beta}^T \boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i1})\Big]\Big\}^{-\frac{1}{r}-1} \\ &\quad \times \exp\Big[\boldsymbol{\beta}^T \boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i1})\Big] \\ &\quad \times \Big\{1 + r t_{i2} \exp\Big[\boldsymbol{\beta}^T \boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i2})\Big]\Big\}^{-\frac{1}{r}-1} \\ &\quad \times \exp\Big[\boldsymbol{\beta}^T \boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i2})\Big]. \end{aligned}$$

It should be noted that, the parameter r appears in the denominator of several parts of the log likelihood function. This will cause some computation difficulty during the optimization search when the true r is zero. This will be able to be detected in the simulation study "Assess model on the generated data from PH and PO model" given in Section 4.3.5. Note that the likelihood is still well defined when $r = 0$, but it has to be worked out separately. This likelihood function is given below. The marginal survival functions are:

$$S_{i1}(t_{i1}) = \exp\Big\{ - t_{i1} \exp\Big[\boldsymbol{\beta}^T \boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i1})\Big]\Big\},$$

$$S_{i2}(t_{i2}) = \exp\Big\{ - t_{i2} \exp\Big[\boldsymbol{\beta}^T \boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i2})\Big]\Big\},$$

The entries of the log likelihood function for the parallel setting are written as:

$$\begin{aligned} \frac{-\partial S(t_{i1}, t_{i2})}{\partial t_{i1}} &= \frac{-\partial \Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi}}{\partial t_{i1}} \\ &= -\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi-1} \Big[S_{i1}(t_{i1})^{-\phi^{-1}-1}\Big] \frac{\partial S_{i1}(t_{i1})}{\partial t_{i1}}, \end{aligned}$$

where

$$\frac{\partial S_{i1}(t_{i1})}{\partial t_{i1}} = -\exp\Big\{ - t_{i1}\exp\Big[\boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big] + \boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big\},$$

and

$$\begin{aligned}\frac{-\partial S(t_{i1}, t_{i2})}{\partial t_{i2}} &= \frac{-\partial\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi}}{\partial t_{i2}} \\ &= -\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi-1}\Big[S_{i2}(t_{i2})^{-\phi^{-1}-1}\Big]\frac{\partial S_{i2}(t_{i2})}{\partial t_{i2}},\end{aligned}$$

where

$$\frac{\partial S_{i2}(t_{i2})}{\partial t_{i2}} = -\exp\Big\{ - t_{i2}\exp\Big[\boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big] + \boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big\}.$$

The joint density of t_{i1} and t_{i2} is

$$\begin{aligned}f(t_{i1}, t_{i2}) &= \frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i1}\partial t_{i2}} \\ &= \Big[S_{i1}(t_{i1})^{-\phi^{-1}}\Big]\Big[S_{i2}(t_{i2})^{-\phi^{-1}}\Big](1+\phi^{-1}) \\ &\quad\times\Big[S_{i1}(t_{i1})^{-\phi^{-1}} + S_{i2}(t_{i2})^{-\phi^{-1}} - 1\Big]^{-\phi-2} \\ &\quad\times\exp\Big\{ - t_{i1}\exp\Big[\boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big] + \boldsymbol{\beta}^T\boldsymbol{x}_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big\} \\ &\quad\times\exp\Big\{ - t_{i2}\exp\Big[\boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big] + \boldsymbol{\beta}^T\boldsymbol{x}_{i2} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i2})\Big\}.\end{aligned}$$

4.2.2 Parameter Estimates

The parameters to be estimated are:

$$\boldsymbol{\theta} = (\phi, \boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, r)^T,$$

where ϕ is the association parameter, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)^T$ are the nonlinear covariate parameters, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ are the linear covariate parameters, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)^T$ are the parameters for spline basis functions, and r is the parameter in the hazard function $\lambda(\epsilon) = \frac{\exp(\epsilon)}{1+r\exp(\epsilon)}$.

Parameters ϕ, $\boldsymbol{\alpha}$ and the smooth function $\psi(\cdot)$ have the same constraints as in the previous two models. The interpretations of the constraints can be found in Section 2.2.4. For the parameter r, since $\lambda(\epsilon) \geq 0$ (that is, $\frac{\exp(\epsilon)}{1+r\exp(\epsilon)} \geq 0$), it is required that $r \geq 0$.

In a summary, the parameter constraints of the proposed transformation model are:

$$\phi > 0,$$

$$\| \boldsymbol{\alpha} \| = 1, \text{ and } \alpha_q > 0,$$

$$\psi(0) = 0,$$

$$r \geq 0.$$

In order to make the optimization search easier, these parameters are transformed into ones without constraints. For parameter ϕ and r, "log" transformation is applied. That is, let

$$\varrho = \log(\phi),$$

$$\kappa = \log(r).$$

For $\boldsymbol{\alpha}$, similarly as what have been done before with the previous two proposed models, two steps of transformation are applied. The relationship between $\boldsymbol{\alpha}$ and the new parameter $\boldsymbol{\varphi}$ is given in (3.1).

After these transformations, the new parameters to estimate with no constraints

are:

$$\boldsymbol{\theta}^* = (\varrho, \boldsymbol{\varphi}^T, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, \kappa)^T,$$

where $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_{q-1})^T$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)^T$. For the unconstrained parameters, Newton-Raphson method is applied to find the maximum likelihood estimators. Similar to the previous proposed models, in the simulation studies, Formula 2.5 and 2.6 are used to calculate the estimate of var($\hat{\boldsymbol{\theta}}^*$).

4.3 Simulation Studies

The results of five different simulation studies are discussed in this section. In all of the simulations, the transformation function is assumed to be $H(t) = \log(t)$. With this specification, $r = 0$ corresponds to a PH model with baseline hazard function $\lambda_0(t) = 1$, which is the exponential distribution. When $r = 1$, the model corresponds to a PO model with the baseline survival function $S_0(t) = 1/(t + 1)$, which is the Lomax distribution (or Pareto type II distribution).

The five simulations are arranged as follows. The first simulation study gives the estimates of the parameters of interest: nonlinear covariate coefficients $\boldsymbol{\alpha}$, linear covariate coefficient β, association parameter ϕ and the parameter r. In the second simulation study the proposed transformation model is compared with the *transformation linear model*. The transformation linear model refers to the model which has the same structure as the proposed model, but assumes only linear covariate effects. Results show that the proposed transformation model can capture the nonlinear pattern better than the transformation linear model, although the transformation linear model gives a good approximation of the nonlinear relationship under the assumption of linearity. The third simulation study summarizes the model performance when some parameters, such as the censoring rate, sample size, and association degree, change. The fourth simulation study assesses the proposed model when parameter

r changes. It finds that the model performs well and stably generally, but when r approaches zero, the model encounters some computational instability. The last simulation study assesses the proposed model on data generated from the nonlinear PH and PO models (that is, using the same data generating processes as in the previous two chapters). Two scenarios are assessed for each of these two cases: one with the true baseline distribution corresponding to $H(t) = \log(t)$, and the other with the baseline distribution misspecified. Results show that misspecification of the function $H(t)$ has a noticeable impact on the model performance.

In all of the simulations, the true marginal function has the following form:

$$\log(T_j|x_j, \boldsymbol{v}_j) = -\beta x_j - 3\sin(2\boldsymbol{\alpha}^T\boldsymbol{v}_j) + \epsilon_j, \quad j = 1, 2,$$

where the trigonometric function $3\sin(2\boldsymbol{\alpha}^T\boldsymbol{v}_j)$ is used as the nonlinear function $\psi(\cdot)$, β is the linear covariate coefficient and $\boldsymbol{\alpha}$ is the nonlinear covariate coefficient. The association parameter ϕ in the Clayton model is set to 0.5, corresponding to fairly strong association between survival times T_1 and T_2.

The nonlinear covariates $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are generated by the uniform distribution. That is, $V_{k1} \sim U(-1, 1)$, $V_{k2} \sim U(-1, 1)$, $k = 1, 2, 3$. The linear covariates (x_1, x_2) are chosen from four sets of combinations (0,0), (0,1), (1,0), (1,1), each with probability one quarter. The nonlinear regression coefficients are set $\boldsymbol{\alpha} = (1, 1, 1,)^T/\sqrt{3}$ and the linear regression coefficient is $\beta = 1$.

Using the relationship between ϵ_j and survival time t_j, and the structure of the Clayton model, the joint distribution of the bivariate survival times can be worked out as a product of the marginal distribution of t_1 and the conditional distribution of $t_2|t_1$. This makes it possible to generate survival times t_{i1}, t_{i2}, $i = 1, \ldots, n$ using the following relations:

$$t_{i1} = \frac{u_{i1}^{-r} - 1}{r\exp\Big\{\beta x_{i1} + \psi(\boldsymbol{\alpha}^T\boldsymbol{v}_{i1})\Big\}},$$

$$t_{i2} = \frac{\left\{\left[1 + (u_{i2}^{-\frac{1}{\phi+1}} u_{i1}^{-\frac{1}{\phi}}) - u_{i1}^{\frac{1}{\phi}}\right]^{r\phi} - 1\right\}}{r\exp\Big\{\beta x_{i2} + \psi(\boldsymbol{\alpha}^T \boldsymbol{v}_{i2})\Big\}},$$

where $u_{i1} \sim U[0,1]$, $u_{i2} \sim U[0,1]$. The censoring time C_j is set to 1 and the corresponding censoring rate is about 45% for both variates.

In the application of spline functions to fit the smooth function, three breakpoints are applied. For how to choose the breakpoints in a spline function see Section 2.2.2.

4.3.1 Evaluate the Performance of the Proposed Model

The parameter r was set to be 1 and 2 and the simulation was conducted for 200 replications for both cases. The estimates for the parameters $\boldsymbol{\varphi}$, ϱ and κ (after transformation) and the original parameters $\boldsymbol{\alpha}$, β, ϕ and r (before transformation) for $r = 1$ and $r = 2$ are given in Table 4.1 and Table 4.2, respectively. We see that, the bias is fairly small for all the parameters, and the sample standard deviations are similar to the average of the model based standard errors. The coverage probabilities of the 95% confidence intervals are all close to the nominal coverage.

Table 4.3 and Table 4.4 give the average estimates over 200 runs for the nonlinear function $\psi(\cdot)$ at the chosen points (-0.9, -0.8, -0.7, 0.2) for $r = 1$ and $r = 2$. The average estimates of the five points are very close to the true values. Figure 4.1a and 4.1b show these results clearly. The black lines represent the true pattern of the nonlinear function $\psi(\cdot)$, and the red dots represent the average estimates at the chosen points. The red dots follow the curve of the function very well which verifies the good performance of the model. Note that $\psi(x)$ is set to be 0 at $x = 0$ by the model constraints and is shown in the figure.

Table 4.1: The estimates for the covariate coefficients $\boldsymbol{\varphi}$, ϱ and κ (after transformation), $\boldsymbol{\alpha}$, β, ϕ and r (before transformation) based on 200 runs ($r = 1$)

	After transformation					
	φ_1	φ_2	ϱ	κ		
True	1.412	1.099	-0.693	0.000		
Bias	-0.001	0.004	-0.021	-0.029		
SD($\hat{\cdot}$)	0.063	0.070	0.229	0.176		
A{SE($\hat{\cdot}$)}	0.058	0.060	0.217	0.184		
Cov.prob.	0.905	0.890	0.920	0.950		
	Before transformation					
	α_1	α_2	α_3	β	ϕ	r
True	0.577	0.577	0.577	1.000	0.500	1.000
Bias	0.000	-0.002	-0.001	-0.021	0.003	-0.013
SD($\hat{\cdot}$)	0.027	0.026	0.026	0.149	0.117	0.172
A{SE($\hat{\cdot}$)}	0.023	0.023	0.023	0.151	0.111	0.179
Cov.prob.	0.900	0.915	0.910	0.955	0.905	0.935

Table 4.2: The estimates for the covariate coefficients $\boldsymbol{\varphi}$, ϱ and κ (after transformation), $\boldsymbol{\alpha}$, β, ϕ and r (before transformation) based on 200 runs ($r = 2$)

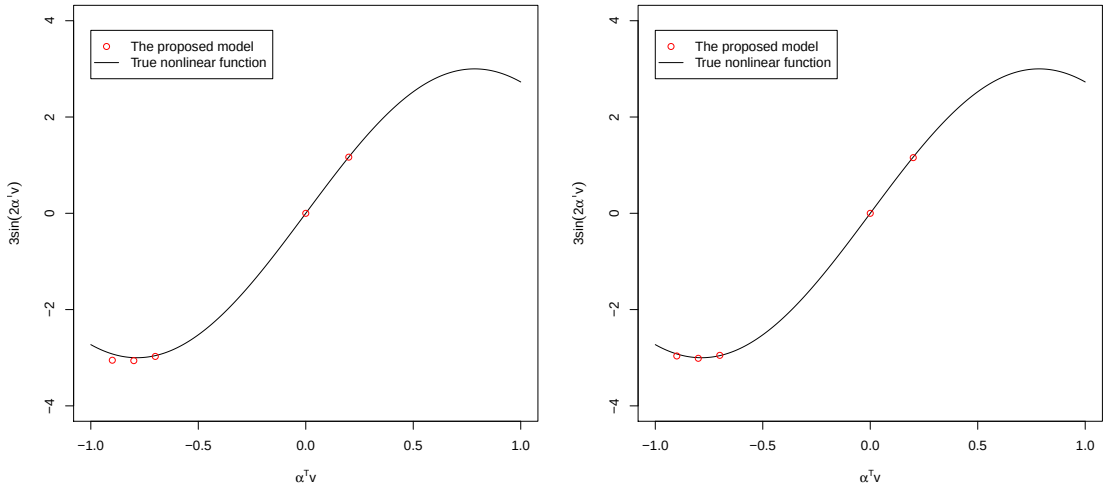
	After transformation					
	φ_1	φ_2	ϱ	κ		
True	1.412	1.099	-0.693	0.000		
Bias	0.002	0.003	-0.031	-0.009		
SD($\hat{\cdot}$)	0.081	0.082	0.198	0.131		
A{SE($\hat{\cdot}$)}	0.074	0.075	0.205	0.135		
Cov.prob.	0.925	0.940	0.950	0.945		
	Before transformation					
	α_1	α_2	α_3	β	ϕ	r
True	0.577	0.577	0.577	1.000	0.500	2.000
Bias	0.000	-0.002	0.000	0.007	-0.006	-0.002
SD($\hat{\cdot}$)	0.033	0.031	0.032	0.189	0.101	0.262
A{SE($\hat{\cdot}$)}	0.029	0.029	0.030	0.193	0.102	0.269
Cov.prob.	0.910	0.910	0.930	0.945	0.935	0.940

Table 4.3: The estimates for $\psi(\cdot)$ of one variate based on 200 runs ($r = 1$)

	x	True	Estimate	SD
$\psi(x)$	-0.900	-2.922	-3.052	0.966
	-0.800	-2.999	-3.060	0.643
	-0.700	-3.002	-2.973	0.493
	0.200	1.168	1.167	0.109

Table 4.4: The estimates for $\psi(\cdot)$ of one variate based on 200 runs ($r = 2$)

	x	True	Estimate	SD
$\psi(x)$	-0.900	-2.922	-2.963	0.509
	-0.800	-2.999	-3.012	0.450
	-0.700	-3.002	-2.951	0.391
	0.200	1.168	1.156	0.129



(a) $r = 1$ (b) $r = 2$

Figure 4.1: Nonlinear smooth function $\psi(\cdot)$. Black lines represent the true pattern of the nonlinear function, and red dots represent the average estimates of selected points from the proposed transformation model.

4.3.2 Compare the Proposed Model with Transformation Linear Model

The proposed transformation model is compared with the transformation linear model.

The marginal function of the transformation linear model has the following form which

does not have the single index structure:

$$\log(T_j|x_j, \boldsymbol{v}_j) = -\beta x_j - \tilde{\boldsymbol{\alpha}}^T \boldsymbol{v}_j + \epsilon_j, \quad j = 1, 2,$$

where β is the coefficient for the covariate x_j and $\tilde{\boldsymbol{\alpha}}$ is the coefficient for nonlinear covariate $\boldsymbol{v}_j$. As in the previous analyses, estimates of $\tilde{\boldsymbol{\alpha}}$ are expressed as $\hat{b}\hat{\boldsymbol{\alpha}}$ with $\| \hat{\boldsymbol{\alpha}} \| = 1$. This makes it easier to compare $\hat{\boldsymbol{\alpha}}$ to the nonlinear coefficient from the proposed transformation model.

The same parameter settings are used as in the first simulation study. The comparison results for two cases $r = 1$ and $r = 2$ are given. Table 4.5 and Table 4.6 give the comparison of the parameters of interest for the proposed model and the transformation linear model when $r = 1$ and $r = 2$ respectively. Tables show the average estimates of the parameters from the proposed model are generally closer to the true values than the linear model, and the standard deviations are smaller as well. Considering the association parameter ϕ, the estimate from the linear model is farther from the truth when $r = 1$ than it is when $r = 2$. Table 4.7 and Table 4.8 show the summary of the average estimates of $\psi(\cdot)$ for the chosen points (-0.9, -0.8, -0.7, 0.2) for $r = 1$ and $r = 2$. Generally the average estimates of the proposed model are closer to the true values. Figure 4.2 gives the clear view of this comparison in the graph. The red dots, the estimates from the proposed model, follow the "S" curve well. The blue line (the linear model) does not follow the "S" curve well, but it does represent a good fit to the curve under the restriction of linearity. In fact, the fit of this line is noticeably better than the fits of the corresponding PO and PH linear models. Further discussion of this observation is given in the Conclusion section.

Table 4.5: The estimates comparison of two models for $\boldsymbol{\alpha}$, β, ϕ and r based on 200 runs ($r = 1$)

	α_1	α_2	α_3	β	ϕ	r
True	0.577	0.577	0.577	1.000	0.500	1.000
Estimate(TR)	0.575	0.577	0.578	0.976	0.517	0.972
SD($\hat{\cdot}$)(TR)	0.026	0.024	0.025	0.146	0.125	0.164
Estimate(linear)	0.574	0.578	0.577	1.093	0.948	1.290
SD($\hat{\cdot}$)(linear)	0.033	0.033	0.034	0.156	0.375	0.201

Table 4.6: The estimates comparison of two models for $\boldsymbol{\alpha}$, β, ϕ and r based on 200 runs ($r = 2$)

	α_1	α_2	α_3	β	ϕ	r
True	0.577	0.577	0.577	1.000	0.500	2.000
Estimate(TR)	0.576	0.573	0.580	1.007	0.491	1.996
SD($\hat{\cdot}$)(TR)	0.031	0.031	0.031	0.180	0.116	0.273
Estimate(linear)	0.573	0.571	0.583	1.085	0.681	2.262
SD($\hat{\cdot}$)(linear)	0.044	0.042	0.044	0.177	0.181	0.295

Table 4.7: The estimates comparison of two models for $\psi(\cdot)$ based on 200 runs ($r = 1$)

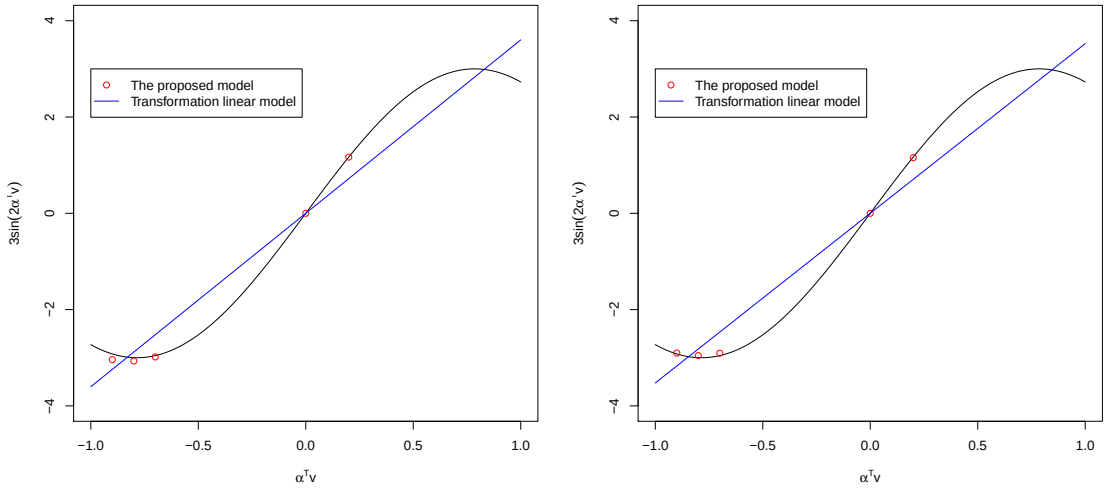
			Proposed		Linear	
	x	True	Estimate	SD	Estimate	SD
	-0.900	-2.922	-3.041	0.585	-3.242	0.215
	-0.800	-2.999	-3.068	0.510	-2.882	0.191
$\psi(x)$	-0.700	-3.002	-2.983	0.429	-2.521	0.167
	0.200	1.168	1.167	0.117	0.720	0.048

4.3.3 Assessment of the Proposed Model Under Various Scenarios

The proposed transformation model is assessed when changing the parameter settings, such as sample size, censoring rate and association parameter ϕ. Sample size is set to either 80 or 200. Censoring rates of 20% and 50% are used. The association

Table 4.8: The estimates comparison of two models for $\psi(\cdot)$ based on 200 runs ($r = 2$)

			Proposed		Linear	
	x	True	Estimate	SD	Estimate	SD
	-0.900	-2.922	-2.901	0.556	-3.171	0.243
	-0.800	-2.999	-2.955	0.460	-2.819	0.216
$\psi(x)$	-0.700	-3.002	-2.905	0.383	-2.467	0.189
	0.200	1.168	1.156	0.137	0.705	0.054



(a) $r = 1$ (b) $r = 2$

Figure 4.2: Compare the proposed transformation model and the transformation linear model on the nonlinear smooth function $\psi(\cdot)$. Black line represents the true pattern of $\psi(\cdot)$, blue line represents the estimate for the linear model, and red dots represent the average estimates for the chosen points.

parameter values of 0.5, 1 and 4 are applied, representing the association decreasing from relatively strong to weak. Table 4.9 gives a summary of the simulation results for 200 runs. As the sample size changes from 80 to 200, the standard deviations of the parameters decreases, which shows the estimates' precision improves. Standard deviations increase when the censoring rate is raised from 20% to 50%, reflecting the loss of information associated with greater censoring. When the association parameter ϕ changes, there is no clear pattern of the estimate for parameters $\boldsymbol{\alpha}$, β and r,

and the estimates all seem very good. But the estimates of ϕ itself differ in quality. As ϕ increases, the estimation performance goes down. Generally $\phi = 1$ has a bigger standard deviations than $\phi = 0.5$. For $\phi = 4$, bias appears, and the standard deviations are relatively big. The values of SD($\hat{\cdot}$) and A{SE($\hat{\cdot}$)} also do not agree with each other. As with proposed PH and PO model, this issue might be caused by the likelihood function being flat when the association among the data is weak.

Table 4.9: Simulation result for different settings of censoring rate, sample size, association parameter ϕ

			α_1			α_2			α_3		
ϕ	P(%)	n	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}
0.5	20%	80	0.003	0.034	0.032	-0.004	0.035	0.032	-0.002	0.037	0.032
		200	-0.002	0.021	0.019	0.001	0.021	0.019	0.000	0.020	0.019
	50%	80	-0.005	0.048	0.041	0.005	0.046	0.041	-0.005	0.046	0.041
		200	0.001	0.026	0.025	-0.002	0.030	0.025	-0.002	0.029	0.025
1	20%	80	0.004	0.043	0.037	0.001	0.042	0.037	-0.010	0.046	0.038
		200	0.001	0.026	0.023	0.003	0.026	0.023	-0.005	0.026	0.023
	50%	80	-0.009	0.057	0.046	0.002	0.052	0.045	-0.001	0.058	0.046
		200	-0.005	0.034	0.028	0.004	0.030	0.028	-0.001	0.032	0.028
4	20%	80	-0.005	0.048	0.042	0.009	0.048	0.042	-0.010	0.046	0.042
		200	-0.002	0.029	0.026	-0.003	0.031	0.026	0.002	0.034	0.026
	50%	80	0.003	0.052	0.047	-0.004	0.051	0.048	-0.006	0.051	0.047
		200	0.000	0.034	0.029	-0.003	0.032	0.030	0.000	0.033	0.030

			β			ϕ			r		
ϕ	P(%)	n	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}	Bias	SD($\hat{\cdot}$)	A{SE($\hat{\cdot}$)}
0.5	20%	80	0.005	0.213	0.211	0.040	0.219	0.163	-0.047	0.247	0.234
		200	-0.015	0.122	0.125	-0.002	0.086	0.081	-0.005	0.142	0.142
	50%	80	-0.003	0.256	0.265	0.004	0.254	0.235	-0.098	0.302	0.316
		200	-0.002	0.165	0.159	-0.003	0.118	0.117	-0.006	0.192	0.195
1	20%	80	-0.013	0.237	0.247	0.060	0.421	0.435	-0.050	0.225	0.234
		200	-0.015	0.141	0.149	-0.003	0.221	0.218	-0.016	0.143	0.140
	50%	80	-0.041	0.297	0.302	0.015	0.607	0.677	-0.007	0.302	0.330
		200	0.001	0.172	0.177	0.033	0.351	0.349	-0.042	0.174	0.190
4	20%	80	0.028	0.264	0.276	-0.385	2.059	3.727	-0.042	0.219	0.227
		200	-0.002	0.170	0.167	0.129	1.844	2.479	-0.015	0.130	0.134
	50%	80	-0.011	0.294	0.315	-0.705	2.050	5.401	-0.089	0.324	0.311
		200	-0.025	0.195	0.189	-0.303	1.927	3.198	-0.026	0.191	0.186

4.3.4 Assessment of the Proposed Method with Various r Values

In this section, the goal is to see how the model performs when the parameter r takes different values. The value of r is set to be 0.1, 0.5 and 5. In order to make the comparison clearer, $r = 1$ and 2 are also included. Other than r, the simulation settings are all the same as those of Section 4.3.1. Table 4.10 gives the summary of the parameter estimate over 200 simulation runs.

For all but the smallest r value, that is when r is 0.5, 1, 2 and 5, the estimate of the parameters are all satisfactory. Recall, the case $r = 1$ corresponds to the PO model.

When $r = 0.1$, 10 out of 200 runs of the simulation has the estimate for r very small (for example, the estimate of one run is approximately 10^{-47}), and the variance matrix is numerically singular. It is likely that in these runs the optimization search encountered difficulty, and did not converge to a local maximum. Experience fitting the model with even smaller values of r showed that as r gets closer to zero, this problem becomes more prevalent. Figure 4.3 shows the log likelihood profile of r for a data set exhibiting this problem. It supports the claim that optimization becomes difficult when r approaches zero. When r is less than 0.01, the log likelihood function becomes very flat, which makes it very difficult to achieve a unique local maximum for r. Note that in the figure r is plotted on a logarithmic scale.

4.3.5 Assessment of Model on the Generated Data from the Nonlinear PH and PO model

Recall that the PH model and PO model are the special cases of the transformation model. That is, in the hazard function of ϵ, $\lambda(\epsilon) = \exp(\epsilon)/\{1 + r\exp(\epsilon)\}$, $r = 0$ corresponds to PH model, and $r = 1$ corresponds to PO model. It is interesting

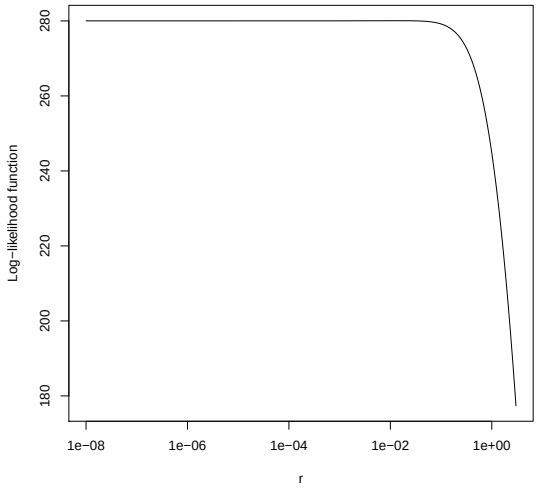


Figure 4.3: The log likelihood profile of r

Table 4.10: Simulation result for different r

	α_1			α_2			α_3		
True r	Bias	SD(·)	A{SE(·)}	Bias	SD(·)	A{SE(·)}	Bias	SD(·)	A{SE(·)}
0.1	0.002	0.017	0.017	0.000	0.020	0.017	-0.003	0.020	0.017
0.5	-0.002	0.023	0.020	0.000	0.024	0.020	0.001	0.024	0.020
1	-0.001	0.025	0.023	-0.002	0.028	0.023	0.001	0.025	0.023
2	0.003	0.030	0.029	-0.002	0.030	0.029	-0.004	0.031	0.029
5	-0.002	0.047	0.044	0.001	0.044	0.044	-0.004	0.044	0.044
	β			ϕ			r		
True r	Bias	SD(·)	A{SE(·)}	Bias	SD(·)	A{SE(·)}	Bias	SD(·)	A{SE(·)}
0.1	0.000	0.101	0.106	0.027	0.116	0.124	0.007	0.078	0.106
0.5	-0.014	0.129	0.127	0.009	0.111	0.112	-0.004	0.125	0.132
1	-0.005	0.174	0.153	0.000	0.108	0.110	0.000	0.191	0.182
2	0.006	0.185	0.191	-0.005	0.094	0.102	-0.008	0.260	0.267
5	0.006	0.301	0.282	0.003	0.101	0.101	-0.038	0.523	0.536

to see how the proposed model performs when it is fit to data generated from the nonlinear PH and PO models (data generated in the same manner as in Sections 2.3 and 3.3). Here, we design a simulation study to assess the following four models:

- Model 1: Data is generated from a nonlinear PH model with baseline exponential distribution.

- Model 2: Data is generated from a nonlinear PH model with baseline Weibull

distribution, the scale parameter is 1, and the shape parameter is $p = 1.5$.

- Model 3: Data is generated from a nonlinear PO model with baseline Lomax distribution (or Pareto type II distribution), the shape parameter is 1, and the location parameter is 1.

- Model 4: Data is generated from a nonlinear PO model with baseline Weibull distribution, the scale parameter is 1, and the shape parameter is $p = 1.5$.

The first two models assess the proposed transformation model using the generated data from nonlinear PH model, and the last two models assess the proposed model using the generated data from nonlinear PO model. The difference between Model 1 and Model 2 is the baseline function. Model 1 uses the exponential distribution, which is the correct baseline distribution for the transformation model with $H(t) = \log(t)$ when $r = 0$, and Model 2 uses a Weibull distribution, a misspecified baseline function for the transformation model when $r = 0$. Similarly, the difference between Model 3 and Model 4 is also the baseline function. Model 3 has the Lomax baseline function, which is the correct baseline distribution for the transformation model when $r = 1$, and Model 4 has a Weibull distribution, a misspecified baseline function for the transformation model when $r = 1$. The parameter settings are the same as in Section 4.3.1.

It should be pointed out here that, as discussed in Section 4.2.1, the proposed transformation model has difficulty handling data generated from the PH model, where $r = 0$. This is due to the fact that the parameter r occurs in the denominator of several parts in the log likelihood function, which causes computation problems during the optimization search.

Table 4.11 summarizes the performance of the proposed model after fitting the data generated from the four different true models. For Model 1, the average estimate of r is 0.04, close to zero and indicates the data might come from a PH model. The

Table 4.11: Simulation result for assessing the proposed transformation model on the generated data from nonlinear PH or PO model

	α_1			α_2			α_3		
Model	Bias	SD(·)	A{SE(·)}	Bias	SD(·)	A{SE(·)}	Bias	SD(·)	A{SE(·)}
Model 1	0.000	0.020	0.015	0.001	0.019	0.016	-0.001	0.018	0.016
Model 2	-0.015	0.138	-	-0.006	0.066	-	-0.002	0.055	-
Model 3	0.001	0.023	0.023	-0.001	0.025	0.024	-0.001	0.025	0.024
Model 4	0.001	0.030	0.027	0.002	0.027	0.027	-0.006	0.029	0.027

	β			ϕ			r		
Model	Bias	SD(·)	A{SE(·)}	Bias	SD(·)	A{SE(·)}	Mean	SD(·)	A{SE(·)}
Model 1	0.009	0.088	0.098	0.009	0.112	0.113	0.040	0.049	0.094
Model 2	-0.021	0.667	-	-0.159	0.099	-	0.000	0.000	-
Model 3	-0.002	0.146	0.152	-0.001	0.100	0.109	0.979	0.168	0.179
Model 4	-1.574	0.100	0.103	-0.098	0.086	0.085	0.039	0.060	0.095

estimates of the rest of the parameters are reasonably good. For Model 2, the majority of the runs have numerical problems when calculating the variance estimate. In those runs, the estimate of r is very small, for example, on the scale of 10^{-300}. This causes the matrices in the variance calculation to be computationally singular. Therefore only the Bias and SD(·) can be reported. The estimates of the covariates, $\boldsymbol{\alpha}$ and β, are reasonable. The average estimate of r is approximately 0, which gives a good indication that the data might come from a PH model. Model 3 corresponds to a correctly-specified transformation model (with PO structure and $r = 1$). The estimates of the parameters are very good. The mean of r, 0.979, strongly suggests the data come from a PO model. As a contrast, the result of Model 4 is not as good. The estimates of $\boldsymbol{\alpha}$ and ϕ are good, but β has a bias. More importantly, the estimate of r, 0.039, would not suggest that the model has a PO structure. It seems the estimate of r is changing to compensate for the wrong choice of the function $H(t)$ in the model, and gives a confusing result. In this case, the estimate of r is somewhat meaningless and fails to provide clear evidence of what model the data come from. Comparing Model 1 and 2, Model 3 and 4, it is found that, misspecification of the function $H(t)$ has a noticeable impact on the model performance.

4.4 Conclusion

In this chapter we propose to use the partially linear single index transformation model as the marginal function to model multivariate survival data. The proposed transformation model has the flexibility to include nonlinear covariate effects, and nonlinear patterns are well displayed by the estimates. The model can accommodate a large class of survival relationships, including both the PH and PO models as special cases.

Besides the advantage of the proposed model, the transformation linear model gives a good estimate of the nonlinear covariate effects, within the limited class of linear relationships. It has better performance than the PH linear model or the PO linear model when nonlinear structure exists. The reason for this improvement is likely additional flexibility introduced by the extra parameter r, which does not appear in the PH or PO linear models. This parameter allows the transformation linear model to adapt itself better to the true nonlinear form.

The model performance was assessed for different values of r, and it was found that when r is greater than approximately 0.5, the model estimates are good. As r approaches zero, the optimization search is more likely to have difficulty finding a good solution. This has been shown through Figure 4.3. Experience suggests that this problem becomes worse as r gets closer to zero.

Fixing $H(t) = \log(t)$ is one restriction of the model as it is presented in the simulation here. The simulation study found that misspecification of the function $H(t)$ can have a noticeable impact on the model performance. To make the model more flexible, it is of interest to try some nonparametric or weakly parametric approaches for estimating $H(t)$. More discussion of this issue is given Section 5.2.

Chapter 5

Discussion and Future Work

The preceding chapters have presented three proposed models. The main findings about these models are summarized in Section 5.1, along with some discussion of computational issues that were encountered. Section 5.2 describes some avenues for future work.

5.1 Discussion

This thesis considers three survival regression models for the marginal distributions of multivariate survival data: a proportional hazards model, a proportional odds model and a transformation model. The traditional linear covariate effect assumption is relaxed by adding nonlinear covariate effects through a smooth single index function. Splines are used to model the single index function in a weakly parametric way. One type of copula model, the Clayton model, is applied to incorporate the association among the variates as an illustration. While only the bivariate case was considered here, it is not difficult to extend the model to higher dimensional survival data. In the proposed PH and PO model, piecewise constants are applied to estimate the baseline hazard function. As a result, inference can be conducted based on the full likelihood. For the proposed transformation model, the unknown increasing function $H(t)$ is fixed

to be known and set as $\log(t)$ in the simulation, therefore, full likelihood approach can be applied as well. In simulation studies, the proposed models demonstrated the advantage of capturing the nonlinear covariate effects over the linear models when such nonlinearities exist.

It was also found that the transformation linear model has better performance than the PH or PO linear models. In the simulation study comparing the transformation linear model with the proposed transformation model, despite the advantage of the proposed model to capture the nonlinear covariate effect, the linear model gives a good estimate under the linearity restriction. For PH and PO linear models, the advantage handling nonlinearity through the proposed PH and PO models is more dramatic. The reasons for this phenomenon might be the flexibility introduced by r in the transformation model. It is suspected that r can adjust its own value to compensate for the misspecification of the linear relationship in the regression model.

While working through the simulation studies and the real data analyses, some computational issues were encountered:

- Various optimization methods available in R have been explored, such as R function `optim` (R Core Team, 2013) which includes a Nelder-Mead method, a quasi-Newton method, and a variant of simulated annealing, among others. Ultimately, the R function `nlm` (which uses the Newton-Raphson method) was found to be more stable and reliable and was therefore used throughout.

- The simulation can be sensitive to the starting values of the parameters for the three proposed models while searching for the maximum likelihood estimates. To address this problem, the corresponding linear models are used to find the estimates first, and these estimates are used as the starting values.

5.2 Future Work

There are some issues inherent in the models considered here. Exploring these issues could become interesting future work.

For modeling of the baseline hazard function using the piecewise constant approach, the effect of changing the number of pieces has not been explored. Increasing the number of the pieces could possibly make the baseline hazard function estimate more accurate, at the cost of more computing time. How to balance these two factors and find a best result is an interesting topic.

One limitation in the linear transformation model is that, under the assumption $H(t) = \log(t)$ and $\lambda(\epsilon) = \frac{\exp(\epsilon)}{1+r\exp(\epsilon)}$, $r = 0$ corresponds to the PH model with the baseline exponential distribution, which is a fairly small family of distributions. To make a linear transformation model that can correspond to a larger class of PH models, the transformation can be generalized into the form

$$H(T) = -\boldsymbol{\beta}^T \boldsymbol{x} + \sigma\epsilon, \tag{5.1}$$

instead of the form (4.4). In this case, $r = 0$ corresponds to the PH model with the baseline Weibull distribution, which has more flexibility. Similarly, $r = 1$ corresponds to the PO model with the baseline Dagum distribution.

Another limitation of the work is that the unknown increasing function $H(t)$ in the transformation model is fixed to be $\log(t)$ in the numerical implementation. The main reason for this is that $H(0)$ is required to be $-\infty$ (Chen et al., 2002), and typical spline functions have difficulty to satisfy such a restriction. An interesting topic to pursue is to explore more approaches to relax $H(t)$ from a fixed parametric form. Including a weakly parametric estimator of $H(t)$ would greatly improve the flexibility and adaptability of the transformation model. Another possible equivalent way is to model the hazard function of ϵ through some nonparametric or weakly

parametric approach. While these ideas still require refinement, this extension of the transformation model is viewed as a most promising avenue for future research.

Bibliography

Bennett, S. (1983), "Analysis of Survival Data by the Proportional Odds Model," *Statistics in Medicine*, 2, 273–277.

Buckley, J. and James, I. (1979), "Linear regression with censored data," *Biometrika*, 66, 429–436.

Cai, T., Wei, L., and Wilcox, M. (2000), "Semiparametric regression analysis for clustered failure time data," *Biometrika*, 87, 867–878.

Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997), "Generalized partially linear single-index models," *Journal of the American Statistical Association*, 92, 477–489.

Chen, K., Jin, Z., and Ying, Z. (2002), "Semiparametric analysis of transformation models with censored data," *Biometrika*, 89, 659–668.

Cheng, S., Wei, L., and Ying, Z. (1995), "Analysis of transformation models with censored data," *Biometrika*, 82, 835–845.

— (1997), "Predicting survival probabilities with semiparametric transformation models," *Journal of the American Statistical Association*, 92, 227–235.

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

Collett, D. (2003), *Modelling survival data in medical research*, vol. 57, Chapman & Hall/CRC.

Cox, D. R. (1972), "Regression models and life-tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, 187–220.

— (1975), "Partial likelihood," *Biometrika*, 62, 269–276.

Fan, J., Gijbels, I., and King, M. (1997), "Local likelihood and local partial likelihood in hazard regression," *The Annals of Statistics*, 25, 1661–1690.

Fine, J., Ying, Z., and Wei, L. (1998), "On the linear transformation model for censored data," *Biometrika*, 85, 980–986.

Hardle, W., Hall, P., and Ichimura, H. (1993), "Optimal smoothing in single-index models," *The annals of Statistics*, 21, 157–178.

Hastie, T. and Tibshirani, R. (1990), "Exploring the nature of covariate effects in the proportional hazards model," *Biometrics*, 1005–1016.

He, W. and Lawless, J. F. (2003), "Flexible maximum likelihood methods for bivariate proportional hazards models," *Biometrics*, 59, 837–848.

Hougaard, P. (2000), *Analysis of Multivariate Survival Data*, Springer, New York.

Huster, W. J., Brookmeyer, R., and Self, S. G. (1989), "Modelling paired survival data with covariates," *Biometrics*, 145–156.

Jin, Z., Lin, D., Wei, L., and Ying, Z. (2003), "Rank-based inference for the accelerated failure time model," *Biometrika*, 90, 341–353.

Jin, Z., Lin, D., and Ying, Z. (2006), "On least-squares regression with censored data," *Biometrika*, 93, 147–161.

Kirmani, S. and Gupta, R. C. (2001), "On the proportional odds model in survival analysis," *Annals of the Institute of Statistical Mathematics*, 53, 203–216.

Knuiman, M., Cullen, K., Bulsara, M., Welborn, T., and Hobbs, M. (1994), "Mortality trends, 1965 to 1989, in Busselton, the site of repeated health surveys and interventions," *Australian Journal of Public Health*, 18, 129–135.

Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995), "Hazard regression," *Journal of the American Statistical Association*, 90, 78–94.

Lawless, J. and Zhan, M. (1998), "Analysis of interval-grouped recurrent-event data using piecewise constant rate functions," *Canadian Journal of Statistics*, 26, 549–565.

Lawless, J. F. (2003), *Statistical Models and Methods for Lifetime Data*, Wiley-Interscience.

Lee, E. W., Wei, L., and Amato, D. A. (1992), "Cox-type regression analysis for large numbers of small groups of correlated failure time observations," *Survival analysis: State of the art*, 211, 237–247.

Lin, D. (1994), "Cox regression analysis of multivariate failure time data: the marginal approach," *Statistics in medicine*, 13, 2233–2247.

Lu, W. and Zhang, H. H. (2010), "On estimation of partially linear transformation models," *Journal of the American Statistical Association*, 105, 683–691.

Lu, X., Chen, G., Singh, R. S., and Song, X. K. P. (2006), "A class of partially linear single-index survival models," *Canadian Journal of Statistics*, 34, 97–112.

Moertel, C. G., Fleming, T. R., Macdonald, J. S., Haller, D. G., Laurie, J. A., Goodman, P. J., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., et al. (1990), "Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma," *New England Journal of Medicine*, 322, 352–358.

Murphy, S., Rossini, A., and Van der Vaart, A. (1997), "Maximum likelihood estimation in the proportional odds model," *Journal of the American Statistical Association*, 92, 968–976.

R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. (1988), "Monotone regression splines in action," *Statistical Science*, 425–441.

Royston, P. and Parmar, M. K. (2002), "Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects," *Statistics in medicine*, 21, 2175–2197.

Shemyakin, A. E. and Youn, H. (2006), "Copula models of joint last survivor analysis," *Applied Stochastic Models in Business and Industry*, 22, 211–224.

Sun, J., Kopciuk, K. A., and Lu, X. (2008), "Polynomial spline estimation of partially linear single-index proportional hazards regression models," *Computational Statistics & Data Analysis*, 53, 176–188.

Therneau, T. M. (2013), *A Package for Survival Analysis in S*. R package version 2.37-4.

Tibshirani, R. and Hastie, T. (1987), "Local likelihood estimation," *Journal of the American Statistical Association*, 82, 559–567.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979), "The impact of heterogeneity in individual frailty on the dynamics of mortality," *Demography*, 16, 439–454.

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

Wei, L.-J., Lin, D. Y., and Weissfeld, L. (1989), "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions," *Journal of the American Statistical Association*, 84, 1065–1073.

Yang, S. and Prentice, R. L. (1999), "Semiparametric inference in the proportional odds regression model," *Journal of the American Statistical Association*, 94, 125–136.

Yu, Z. and Lin, X. (2008), "Nonparametric regression using local kernel estimating equations for correlated failure time data," *Biometrika*.

Zeng, D. and Lin, D. (2006), "Efficient estimation of semiparametric transformation models for counting processes," *Biometrika*, 93, 627–640.

— (2007), "Semiparametric transformation models with random effects for recurrent events," *Journal of the American Statistical Association*, 102, 167–180.

VITA

Name

Na Lei

Education

DOCTOR OF PHILOSOPHY in Biostatistics, December 2013
The University of Western Ontario
London, Ontario, Canada
Supervisor: Wenqing He

MASTER OF SCIENCE in Statistics, August 2007
Simon Fraser University
Burnaby, British Columbia, Canada
Supervisor: Randy R. Sitter

MASTER OF SCIENCE in Statistics, July 2004
University of Science & Technology of China
Hefei, Anhui, China
Supervisor: Zhaoben Fang

BACHELOR OF SCIENCE in Economics, July 2001
University of Science & Technology of China
Hefei, Anhui, China

Employment

Teaching Assistant, 8 semesters Fall 2007-Winter 2013
Department of Statistical and Actuarial Sciences, The University of Western Ontario

Research Assistant, 3 semesters Summer 2008-Summer 2011
Department of Statistical and Actuarial Sciences, The University of Western Ontario

Statistical Consultant, 2 semesters Fall 2008-Winter 2011
Social Science Network and Data Services, The University of Western Ontario

Teaching Assistant, 5 semesters Fall 2004-Winter 2007
Department of Statistics and Actuarial Science, Simon Fraser University

Research Assistant, 2 semesters Summer 2005-Summer 2007
Department of Statistics and Actuarial Science, Simon Fraser University

Statistical Analyst in a co-op job, June 2006-August 2006
BC Cancer Research Center

Research Assistant, June 2002-January 2003
Department of Finance, Chinese University of HongKong

Scholarships and Awards

1. Western Graduate Research Scholarship
   The University of Western Ontario, September 2007-December 2013

2. Graduate Fellowship Award
   Simon Fraser University, May 2006

3. Outstanding Student of Business School
   University of Science & Technology of China, December 2003

4. Excellent Thesis Award
   University of Science & Technology of China, July 2001

5. Academic Excellence Scholarship in three consecutive academic years
   University of Science & Technology of China, 1998-2000

Publications

1. Zhaoben Fang, Na Lei and Yong Li (2004). Data Mining in Auto Credit Market of China, China Statistics (Chinese), 271, 24-25.

2. Shuguang Zhang, Na Lei and Jun Ma (2003). Estimation of Long Range Parameters of the Chinese B-Shares Index and its Applications, Operations and Management (Chinese) 45, 61-64.

Conference Presentations

Na Lei and Wenqing He (2010). Weakly Parametric PH Regression Model for Multivariate Survival Data. *The Annual General Meeting of the Statistical Society of Canada*, Quebec City, QC, Canada, May 23-26.

Membership to Professional Society

Statistical Society of Canada