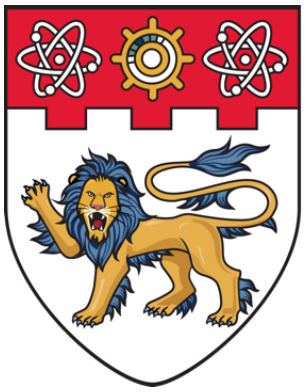


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

EXTENDED SEMIPARAMETRIC MIXTURE CURE MODELS FOR INTERVAL CENSORED DATA

LIU XIAOYU

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

2019

EXTENDED SEMIPARAMETRIC MIXTURE CURE MODELS FOR INTERVAL CENSORED DATA

LIU XIAOYU

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

A thesis submitted to the Nanyang Technological

University in partial fulfilment of the requirement for the

degree of Doctor of Philosophy

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

.... Aug 15, 2019 ....
Date

Liu Xiaoyu.
.....
LIU XIAOYU

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

.... Aug 15, 2019 ....

Date

Prof. Xiang Liming

Authorship Attribution Statement

(A) This thesis does not contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author.

Aug 15, 2019	Liu Xiaoyu.
Date	LIU XIAOYU

Abstract

Analysis of the survival data with a subgroup of cured subjects arising in a clinical trial is commonly performed using a mixture cure model. Existing studies of the two-component mixture cure model assume a logistic regression for the cure probability and a conventional survival model for failure times of susceptible subjects. In this thesis, two extended semiparametric mixture cure models are proposed to analyze interval-censored data in which the failure times are recorded as intervals and there is a subgroup of subjects to be cured. The first proposal is to use the Bayesian doubly semiparametric mixture cure model to incorporate the nonlinear effects of risk factors both in the probability of being cured and the survival risks in the latency stage. The second proposal is based on a generalized accelerated hazards cure model to describe the time-scaled effects in the latency stage.

There are four chapters in the thesis. The first chapter provides a brief introduction to failure time data and interval censoring. It is followed by an outline of commonly used statistical models and inference approaches used in this thesis.

In the second chapter, a more flexible Bayesian doubly semiparametric mixture cure model for interval-censored data is proposed, allowing a combination of linear and nonlinear effects of covariates in both mixture components. A computationally feasible Bayesian estimation procedure is developed, which incorporates two-stage data augmentation with Poisson latent variables to deal with interval-censored data, and monotone splines and polynomial splines to model the cumulative baseline hazard function and nonlinear terms in the model. Simulation results

demonstrate the satisfactory performance of the proposed method in the finite sample cases. The utility of the method is illustrated by the analysis of data from a hypobaric decompression sickness study.

In the third chapter, a generalized accelerated hazards cure model for interval-censored data is proposed, where a general class of accelerated hazard functions is applied to model the failure times of those susceptible subjects in the population. An efficient and easily implemented sieve maximum likelihood approach is developed for estimation in which the unknown cumulative baseline hazard function is approximated by the linear combination of B-spline functions, and a two-step iterative algorithm is used for implementation. The large sample properties of the resulting estimator are established, including the consistency, convergence rate and asymptotic normality. Simulation results demonstrate the satisfactory performance of the method, and its utility is illustrated through the analysis of smoking cessation data.

The last chapter provides conclusions, further discussions and possible directions of future research.

Acknowledgements

I would like to express my very great appreciation to Professor Xiang Liming, my supervisor, for her support, warmly encouragement, constructive suggestions, and patient guidance for this research work. Her conscientious attitude and hard-working inspire me to explore the unknown fields in future research.

I also appreciate the assistance provided by the faculty members. Professor Pan Guangming, Professor Feng Qu, Professor Xing Chaoping, Professor Wang Lilian, Professor Nicolas Privault, and Professor Wu Guohua. Special thanks to Professor Pan Guangming for the suggestions and encouragement as a member of my thesis committee.

I would like to owe my gratitude to all the members of our research group. Dr. Yu Tonghui, Dr. Peng Mengjiao, Dr. Zsolt Szabo, Dr. Wang Chunjie, and Huang Rui. Thank for you all for your support and instructions when I am confused during the research process. Special thanks to Dr. Yu Tonghui for her technical assistance, warmly encouragement, and valuable recommendations in my research work. I am also thankful to the following colleges for their support and help along the time: Zhang Zhixiang, Liu Yiming, Dr. Xu Yanhong, Dr. Zhang Juanyang, Dr. Yuan Chen, Gao Yang, Yu Haiwan, Dr. Zhang Bo, and Dr. Han Xiao. Special mentions to Zhang Zhixiang and Liu Yiming, for your assistance in the theoretical proofs and daily work.

I am really grateful to my family and my friends for inspiring me a lot when I have trouble with research, Yan Hongpan, Lu Bing, Wang Bo and Sun Qiao. Thank you for standing by me and give me emotional support all the time. Special acknowledgment to Lu Bing and Wang Bo, for helping me revise my thesis. My special thanks are also extended to Dr. Zhang Jinbo, who have shown great understanding and taken care of me along the way.

Thank you all for the encouragement, I have a great time during this period, and I owe it all to you!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Semiparametric mixture cure models have been extensively studied in the analysis of survival data in recent years (Sy and Taylor, 2000; Lu et al., 2007; Zhang and Peng, 2009) on account of their abilities to describe a variety of failure mechanisms and assess the covariate effects flexibility. Interestingly, the analysis of the effects of risk factors in semiparametric mixture cure models is more complicated when the failure time data are interval-censored. However, the assumptions used in previous studies of semiparametric mixture cure models are too restrictive under certain circumstances. For example, the nonlinear effects of some time-related covariates both in the cure probability and survival risks cannot be explained by these models (Kim and Jhun, 2008; Ma, 2009; Zhou et al., 2016). Moreover, the intersection of the hazard functions is unable to be explicable by the conventional survival models. Thus, to enable explanation of such features, this thesis describes the development of two extended mixture cure models that incorporate interval-censored data: a Bayesian doubly semiparametric mixture cure model and a generalized accelerated hazards cure model.

In this chapter, an introduction of survival data and relevant distribution functions are given in Section 1.1. In Section 1.2, we give a brief description of

interval-censored data and effective analysis approaches for these. A detailed explanation of conventional semiparametric regression models is provided in Section 1.3, with a focus on the proportional hazards model and the accelerated hazards model. In Section 1.4, the conventional semiparametric cure models are reviewed. In Section 1.5, two commonly used statistical approaches are described: the Bayesian approach and the sieve maximum likelihood approach. In Section 1.6, the introduction is concluded with a thesis outline.

1.1 Failure Time Data

Failure time (or survival time) data, which refer to the occurrence of certain failure events, are common to many fields of research, including biomedical research, public health sciences, economics, and finance studies (Sun, 2007; Lawless, 2011; Kalbfleisch and Prentice, 2011). It is more challenging to handle failure time data than data from other statistical areas due to the existence of censoring. By censoring, we mean that only partial information can be obtained during the data collection process. That is, a portion of the exact failure time data is unavailable, as the subjects may drop out or encounter death by accident. To manage these complex data structures, survival data analysis approaches are required. Here, several commonly used functions in survival analysis are presented to illustrate the current range of approaches for managing censored data. These approaches provide the basis for further research in this thesis.

T is assumed to be a non-negative random variable, representing the failure time of the event of interest. For the sake of convenience, T is assumed to be a continuous random variable, and it needs to be established whether a subject survives beyond a specified time point. The following survival function is hence proposed: $S(t) = P(T > t) = 1 - F(t)$, which describes the probability of a susceptible subject survives beyond time t. The survival function provides a more straightforward

and practical interpretation than the cumulative distribution function $F(t)$.

Furthermore, the failure information of the subject prior to a certain time point is accessible during the survival process. Incorporating this condition, the following hazard function is proposed to more accurately describe the process:

$$\lambda(t) = \lim_{\Delta t \to 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \tag{1.1}$$

which indicates the instant failure risk of a surviving subject at time t. Another effective function frequently utilized is the cumulative hazard function, describing the cumulative risk until time t, which is defined as:

$$\Lambda(t) = \int_0^t \lambda(s) ds. \tag{1.2}$$

From the definitions of the above functions, the following useful relationship can be obtained: $S(t) = \exp(-\Lambda(t))$. This relationship is the primary focus in this thesis, from which any of the other one-to-one relationships can be easily derived.

1.2 Interval Censoring

As mentioned in Section 1.1, censoring is a crucial and distinct feature in survival data, and can be caused by various factors. For example, the subjects in a clinical trial may remain susceptible to the focal event of the trial at the end of the observation period, or die from a disease which is not the target of the trial. Right censoring occurs most commonly, which means the failure time data can only be observed before a specified time point.

This thesis considers another commonly encountered type of censoring: interval censoring, under which the exact failure time is inaccessible, and only the time interval containing the failure time is known. Interval-censored data appear in

clinical trials that involve periodic follow-ups. For example, in the study of hypobaric decompression sickness (HDS) data by the National Aeronautics and Space Administration (NASA) (Conkin and Powell, 2001), the time to the development of HDS was recorded only within pre-scheduled time intervals, without the exact time being observed.

In practice, a critical type of interval censoring commonly occurs is the case II interval-censored data. Let L and R be two observed time points that contain the failure time, satisfying $T \in (L, R]$ and $L \leq R$. Case II interval-censored data derive from a study in which pre-scheduled time points are arranged for the data collection process. Thus, it is possible to find a finite interval $(L, R]$, satisfying $0 < L \leq R < \infty$, that contains the failure time (Huang and Wellner, 1997). The HDS dataset mentioned above is a typical example of case II interval-censored data. In this thesis, the main focus is the investigation of two extended semiparametric models that handle case II interval-censored data.

When estimating functions and analyzing the effects of risk factors in survival analysis, interval-censored data comprising two random variables are more challenging than right censored data. For example, an easily implemented empirical estimation of the survival function in right censored data is given by Kaplan and Meier (1958), but the estimation of the survival function for interval-censored data is more difficult. One option is to utilize the multiple imputation method to impute values from interval-censored data and take these imputed values as right censored data for estimation. However, despite its simplicity, this method uses partial information and may lead to a significant bias.

As the closed form for the nonparametric maximum likelihood estimator (NPMLE) of the survival function cannot be derived from case II interval-censored data, the efficient iterative algorithms are required. The commonly used algorithms include the self-consistency algorithm (Turnbull, 1976), the iterative convex minorant algorithm (Groeneboom and Wellner, 1992), and the hybrid expectation-maximization

(EM) iterative convex minorant (ICM) algorithm (Wellner and Zhan, 1997). In addition, several effective nonparametric approaches exist for estimating the hazard functions of survival data, e.g., the local likelihood approach, the spline-based approach, and the kernel-based approach are commonly used to analyze interval-censored data. For further information, see Tibshirani and Hastie (1987), Kooperberg and Stone (1992), and Lawless (2011).

The theoretical justification of estimators for right censoring has been established based on the utilization of the counting process and martingale theory. As less information is presented in interval-censored data, it is laborious to derive the properties of large samples. Extensive discussions of the broad application of interval-censored data in various situations can be found in Sun (2007). Although the empirical process theory (van der Vaart and Wellner, 1997) has been applied, the theoretical justification of interval censoring lacks elegance. Accordingly, an investigation of theoretical justification for the two semiparametric modeling approaches with case II interval-censored data is provided in this thesis.

1.3 Semiparametric Regression Models

The investigation of covariate effects on the event of interest is an essential topic of study in survival analysis. For example, how breast cancer patients' clinical characteristics affect their survival outcomes have attracted much attention in medical studies (Berkson and Gage, 1952; Brinkley and Haybittle, 1975; Farewell, 1982). Accordingly, semiparametric regression models, in which the underlying distributions of survival data are undetermined, are proposed to assess the effects of risk factors in such situations. In this section, four useful semiparametric regression models are examined, i.e., the proportional hazards model, the proportional odds model, the accelerated failure time model, and the accelerated hazards model.

1.3.1 The proportional hazards model

The proportional hazards (PH) model proposed by Cox (1972) is most commonly used to evaluate the effects of risk factors, due to its straightforward interpretations of coefficients and efficient inference procedures. Let $\boldsymbol{X}$ be a vector of covariates. In terms of the multiplicative form, the hazard function is assumed as:

$$\lambda(t|\boldsymbol{X}) = \lambda_0(t)\exp(\boldsymbol{X}^T\boldsymbol{\beta}), \tag{1.3}$$

where $\lambda_0(t)$ is the unknown baseline hazard function with all covariates being 0, and $\boldsymbol{\beta}$ is the vector of regression coefficients.

The interpretation of the model is illustrated by a two-sample situation, where covariate X is assumed as a binary treatment indicator and takes the value of 1 if a subject receives the treatment, or 0 otherwise. The model demonstrates that the hazard ratio of these two samples does not depend on time t, that is:

$$\frac{\lambda(t|X=1)}{\lambda(t|X=0)} = \exp(\beta). \tag{1.4}$$

In other words, the hazard ratio is a constant regardless of the value of t. $\exp(\beta)$ is the ratio of hazards between two samples at time t, and β is the log relative risk. $\beta > 0$ or $\beta < 0$ indicates the treatment increases or decreases the baseline hazard by a constant of $\exp(\beta)$, and the treatment does not have effects on subjects if $\beta = 0$.

Extensive studies have examined the PH model in the presence of interval-censored data. Finkelstein (1986) utilized the maximum likelihood approach to obtain the estimators, where the baseline hazard function was assumed to be a step function. Satten (1996) proposed a rank-based marginal likelihood approach and obtained the estimator by use of the estimating equation, leaving the baseline hazard function unspecified and regarding it as a nuisance parameter. A similar

estimating equation was utilized by Satten et al. (1998) with imputed failure times. A more efficient multiple imputation approach for the PH model was explored by Pan (2000). Betensky et al. (2002) employed a commonly used local likelihood approach and obtained the estimator by an EM algorithm. Cai and Betensky (2003) developed a penalized likelihood estimation, where linear spline functions were applied for the hazard function. More recently, Zhang et al. (2010) utilized the B-splines approximation and the generalized Rosena's algorithm to obtain the sieve maximum likelihood estimator. The asymptotic properties were established by Huang and Wellner (1995) for the maximum likelihood estimator according to the empirical process (van der Vaart and Wellner, 1997).

Most of the approaches in the previous studies relied on the frequentist inference, and little work has been performed from the Bayesian point of view. Lin et al. (2015) developed an easily implemented two-stage data augmentation method for interval-censored data from a Bayesian perspective, according to the relationship between the PH model and the Poisson process. The PH model has been extended to various situations in the recent studies, where covariates were allowed to depend on time or different baseline hazard functions were applied to subjects.

In the next chapter, a PH-based mixture cure model is explored. Huang et al. (1999) studied the model by incorporating nonlinear covariate effects into the PH model, based on the frequentist approach. In contrast to their method, we investigate the extended PH mixture cure model for interval-censored data, adopting a Bayesian perspective.

1.3.2 The proportional odds model

In practice, the assumption of the constant hazard ratio in the PH model is too restrictive. The hazard ratio can be time-dependent and converges to one as $t \to \infty$ in some situations. Incorporating this possible feature, an alternative proportional odds (PO) model is proposed. Bennett (1983) first introduced the PO model

and investigated the maximum likelihood estimation for right censored data. PO models for right censored and interval-censored data have attracted much attention, such as in works by Dinse and Series (1984), Collett (1994) and Rossini et al. (1996).

Given covariates $\boldsymbol{X}$, the PO model assumes that the odds ratio is a constant at any time t, that is:

$$\frac{S(t|\boldsymbol{X})}{1-S(t|\boldsymbol{X})} = \exp(-\boldsymbol{X}^T\boldsymbol{\beta})\frac{S_0(t)}{1-S_0(t)}, \tag{1.5}$$

where $S_0(t)$ is the baseline survival function and $S(t|\boldsymbol{X})$ is the survival function. Taking the two-sample situation $X = 1$ or 0 as an example, the PO model illustrates that the odds ratio between two samples is a constant $\exp(-\beta)$.

The use of the PO model for interval-censored data has been investigated by various approaches, including the maximum likelihood approach, the sieve maximum likelihood approach, and the conditional likelihood approach. See Murphy et al. (1997), Huang and Rossini (1997), and Rabinowitz et al. (2000) for example. Huang and Wellner (1997) investigated the asymptotic properties of estimators obtained by the maximum likelihood approach.

1.3.3 The accelerated failure time model

In addition to semiparametric regression analysis, the efficiently implemented accelerated failure time (AFT) model is also commonly used, due to the computational convenience of the parametric form. Instead of incorporating covariates in the hazard function, the model directly evaluates the effects of risk factors on the logarithm of failure time T. Specifically, the model assumes that

$$\log T = \boldsymbol{X}^T\boldsymbol{\beta} + W, \tag{1.6}$$

where W is an error variable with an unknown distribution. Different distributions of W correspond to different distributions of T.

Odell et al. (1992) investigated the maximum likelihood approach for the AFT model based on the Weibull distribution. The consistency of the maximum likelihood estimator was given by Huang and Wellner (1997), while the asymptotic normality needs further exploration. Betensky et al. (2001) studied the AFT model with interval-censored data and proposed a computationally simple method of estimating equations, avoiding the use of nonparametric maximum likelihood. To overcome the computational complexity and inefficiency of existing semiparametric estimation approaches in the AFT model, Zeng and Lin (2007) proposed an approximate nonparametric maximum likelihood approach in which a kernel-smoothed profile likelihood was maximized for estimating regression coefficients.

1.3.4 The accelerated hazards model

In the conventional semiparametric regression models mentioned above, the covariates are assumed to have immediate effects when applied. However, this assumption may be inappropriate in reality. For instance, in a screening and intervention study of prediabetes (Rasmussen et al., 2008), subjects were randomly assigned to two groups: the lifestyle intervention group and the control group. Notably, the intervention came into effect gradually, rather than taking effect immediately.

To account for situations in which there may be a lag period before the covariate has an effect on the subjects, an accelerated hazards (AH) model was proposed as an alternative (Chen and Wang, 2000). In the AH model, a time-scaled change is assigned directly to the hazard functions to characterize the change in hazard progression over time. Given covariates $\boldsymbol{X}$, the conditional hazard function of the failure time takes the form of:

$$\lambda(t|\boldsymbol{X}) = \lambda_0(t\exp(\boldsymbol{X}^T\boldsymbol{\beta})), \tag{1.7}$$

where $\exp(\boldsymbol{X}^T\boldsymbol{\beta})$ is the hazard progression time ratio, which measures the effects of risk factors on the underlying hazard progression. Specifically, in the two-sample situation, the treatment will accelerate the hazard progression if $\beta > 0$, decelerate it if $\beta < 0$, or have no effect if $\beta = 0$. Unlike using time-dependent covariates in the conventional model, the AH model is valuable for incorporating covariates with time-scaled effects, such as the treatment effect mentioned above, or the age of an individual.

To relax certain assumptions of survival functions and improve the model flexibility, a general accelerated hazards regression model was proposed by Chen and Jewell (2001) in which the estimators were obtained by estimating equations. The model assumes that the cumulative hazard function is in the form of:

$$\Lambda(t|\boldsymbol{X}, \boldsymbol{Z}) = \Lambda_0(t \exp(\boldsymbol{X}^T\boldsymbol{\beta})) \exp(\boldsymbol{Z}^T\boldsymbol{\gamma}), \tag{1.8}$$

where Λ_0 is the cumulative baseline hazard function and $\boldsymbol{Z}$ is the vector of covariates. The model reverts to the conventional PH model when $\boldsymbol{\beta} = 0$, the AFT model when $\boldsymbol{\gamma} = 0$, and the AH model when $\boldsymbol{\gamma} + \boldsymbol{\beta} = 0$ and $\boldsymbol{X} = \boldsymbol{Z}$.

Extensive research has examined the use of the AH model for right censored data, while few studies have investigated its use for the analysis of interval-censored data. Zhang and Peng (2009) investigated the AH cure model and proposed the EM algorithm for estimation in which the M-step is achieved by a rank-based estimating equation, with the variance estimation provided by a bootstrap method. Zhang et al. (2011) utilized a kernel-smoothed function to approximate the profile likelihood function and applied the bootstrap method for the variance estimation. Zhao et al. (2017) investigated a general class of AH models and employed a sieve maximum likelihood approach for the estimation of parameters.

The estimation procedures in the PH and AFT models require proportionality between the hazard ratio or the failure time, which is unsatisfied in the AH

model. Thus, the methods described above are inappropriate. Moreover, the finite-dimensional parameters are bundled with the infinite-dimensional function Λ_0, which necessitates a more challenging theoretical justification.

In Chapter 3, the utility of the generalized accelerated hazards cure model for the analysis of interval-censored data is explored, and the asymptotic properties of the estimators are provided.

1.4 Semiparametric Cure Models

In classical survival analysis, all subjects are assumed to ultimately fail over a long enough observation time. However, the rapid development of medical technology and quality of health care means that a proportion of patients who receive a medical treatment for a certain disease can often be cured, and these patients (subjects) are regarded as event-free. To account for this new feature, cure rate models are introduced. In clinical trials, a time window is usually set by researchers within which all subjects are expected to experience the study event. Those subjects who are non-susceptible to the study event beyond the window are regarded as having infinite survival times and are said to be cured. The Kaplan-Meier (KM) estimator of the survival function with heavy censoring and a long stable plateau provides empirical evidence for the existence of the cured subgroup.

Two separate groups are assumed to form the population in the cure models: the susceptible and the non-susceptible (cured) groups. The proportion of unsusceptible subjects is defined as the cure fraction. The mixture cure model takes the cure fraction in survival data into account, and enjoys widespread application due to its enabling the convenient interpretation of the effects of risk factors. This is achieved by modeling the population survival as a mixture of two components for susceptible (uncured) and non-susceptible (cured) subgroups, with the incidence taken to represent the probability of a subject being cured and the latency taken

to represent the survival function of susceptible subjects. For further information, see Farewell (1986), Kuk and Chen (1992), Sy and Taylor (2000), Peng and Dear (2000) and Gu et al. (2011) for examples. More recently, Peng and Taylor (2014) reviewed the current research on cure models. The mixture cure model is examined in this thesis due to its clear structure and straightforward interpretation in the presence of interval censoring.

The mixture cure model assumes the marginal survival function of the entire population $S_p(t)$ in the form of:

$$S_p(t|\boldsymbol{X}, \boldsymbol{Z}) = 1 - \pi(\boldsymbol{X}) + \pi(\boldsymbol{X})S(t|\boldsymbol{Z}), \tag{1.9}$$

where $\pi(\boldsymbol{X})$ is the uncured probability of subjects, and $S(t|\boldsymbol{Z})$ is the proper survival function for the susceptible subjects. $\boldsymbol{Z}$ can be overlapped with $\boldsymbol{X}$. The uncured rate $\pi(\boldsymbol{X})$ is often assumed to be associated with covariates $\boldsymbol{X}$ through a log link function.

The cure models for interval-censored data have attracted more attention recently. Many studies have used parametric mixture cure models for interval-censored data with a cure fraction, due to their convenience in computation and theoretical justification. These use various parametric families of survival distributions for uncured subjects, for example, Sparling et al. (2006), Li and Ma (2010), Balakrishnan and Pal (2013) et al.

Semiparametric mixture cure models have also been investigated extensively in recent years to achieve better flexibility. Lam and Xue (2005) considered a semiparametric AFT model for the latency part, and obtained the estimators by the sieve maximum likelihood approach. Kim and Jhun (2008) assumed a piecewise constant hazard function in the PH model for susceptible subjects, and utilized the EM algorithm and multiple imputation for the estimation of parameters and the calculation of the variance matrix, respectively. Ma (2009) investigated the linear

PH model for the survival risks, by assuming a step function for the cumulative hazard function and proposed a maximum likelihood estimation in the case of current status data. A further investigation of the partial linear PH model with mixed interval-censored data was proposed by Ma (2010). Lin et al. (2015) introduced an efficient data augmentation for interval censoring and approximated the cumulative baseline hazard function by a splines approach, with the estimators obtained by Gibbs sampling. More recently, an easily implemented multiple imputation approach was introduced by Zhou et al. (2016) to handle the complex data structure. The extension of the mixture cure models with clustered interval-censored data has been studied by Banerjee and Carlin (2004), Xiang et al. (2011) and Lam and Wong (2014).

Previous studies often focused on parametric and semiparametric regression models with linear covariate effects for the incidence and latency components. In this thesis, nonlinear covariate effects in both components motivated by the HDS dataset are investigated. In addition, the performance of an extended mixture cure model is explored by assuming a generalized AH model for the susceptible subjects.

1.5 Statistical Inference Approaches

The use of semiparametric mixture cure models with interval-censored data has been investigated extensively using various approaches, in attempts to deal with the complex data structure and the unspecified hazard progression. In this section, two approaches used in the following chapters are introduced, namely the Bayesian approach and the sieve maximum likelihood approach.

1.5.1 The Bayesian approach

The Bayesian method has attracted substantial attention for use in the inference of survival models because of its efficient computation framework and straightforward

construction. Here, the further application of the Bayesian method in the presence of interval-censored data is developed. Consider a sample consisting of a total of n subjects and involving interval observations:

$$\mathcal{O} = \{(L_i, R_i], \boldsymbol{X}_i; i = 1, 2 \dots n\}, \tag{1.10}$$

where $(L_i, R_i]$ is the interval containing the failure time T_i and $\boldsymbol{X}_i$ is vector of the covariates of subjects i.

Based on observations, the likelihood function is assumed as $L(\boldsymbol{\theta}|\mathcal{O})$, in which $\boldsymbol{\theta}$ is regarded as a random variable in a Bayesian perspective. The inference of $\boldsymbol{\theta}$ can be obtained from the posterior distribution, which is calculated after assigning a prior distribution. Specifically, assume $f(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$, the posterior distribution of $\boldsymbol{\theta}$ is calculated by the following formula:

$$f(\boldsymbol{\theta}|\mathcal{O}) = \frac{L(\boldsymbol{\theta}|\mathcal{O}) * f(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} L(\boldsymbol{\theta}|\mathcal{O}) * f(\boldsymbol{\theta}) d\boldsymbol{\theta}}, \tag{1.11}$$

where $\boldsymbol{\Theta}$ is the parameter space of $\boldsymbol{\theta}$.

The inference of $\boldsymbol{\theta}$ is easily acquired via the posterior distribution $f(\boldsymbol{\theta}|\mathcal{O})$ given that it has a closed form. This contrasts with the significant effort usually required to obtain the closed form of $f(\boldsymbol{\theta}|\mathcal{O})$, due to possible improper specification of prior distributions and the complex data structure. Thus, one may sample from the posterior distribution and use the sample mean to make an inference. Gibbs sampling, armed rejective sampling, and Markov chain Monte Carlo sampling are conventional sampling techniques used in survival analysis.

The Bayesian approach has received widespread usage and is easily implemented, due to its having a posterior distribution which can be expressed formally as long as one holds information of the prior distribution. However, the computation of the posterior distribution is time-consuming as a result of the complicated structure of the likelihood function, which is caused by different types of censoring

and the improper selection of the prior distribution (Chen et al., 1999; Banerjee and Carlin, 2004; Dey et al., 2012; Lin et al., 2015).

1.5.2 The sieve maximum likelihood approach

Unlike the Bayesian approach, the maximum likelihood estimation (MLE) for the semiparametric regression models can be obtained directly without the requirement for prior information. The estimation procedure can be demonstrated through the PH model discussed above. Under the assumption of the PH model, the cumulative hazard function is written as:

$$\Lambda(t|\boldsymbol{X}) = \Lambda_0(t|\boldsymbol{X}) \exp(\boldsymbol{X}^T\boldsymbol{\beta}). \tag{1.12}$$

Assume that given $\boldsymbol{X}$, the failure time T and $(L, R]$ are independent, and the joint distribution of $(L, R, \boldsymbol{X})$ is not dependent on $(\boldsymbol{\beta}, \Lambda_0)$. To express the idea clearly, the following indicator variables are introduced: $\delta_{L_i} = I(L_i = 0)$, $\delta_{R_i} = I(R_i = \infty)$ and $\delta_{I_i} = 1 - \delta_{L_i} - \delta_{R_i}$. The likelihood function in terms of $\boldsymbol{\beta}$ and Λ_0 is given as:

$$\begin{aligned} L(\boldsymbol{\beta}, \Lambda_0|\mathcal{O}) = \Pi_i^n &\{\exp[-\Lambda_0(R_i)\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})] - 1\}^{\delta_{L_i}} \\ &\times \{\exp[-\Lambda_0(L_i)\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})] - \exp[-\Lambda_0(R_i)\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})]\}^{\delta_{I_i}} \\ &\times \{\exp[-\Lambda_0(L_i)\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})])\}^{\delta_{R_i}}. \end{aligned} \tag{1.13}$$

The ML estimator $(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0)$ can be obtained by maximizing $L(\boldsymbol{\beta}, \Lambda_0|\mathcal{O})$ over a bounded parameter space, with the nondecreasing and nonnegative constraints for Λ_0. Suppose $\{L_1, R_1, L_2, R_2, ...L_n, R_n\}$ are all end points of the interval-censored observations. We arrange the distinct points of them in ascending order, denoted by $s_{(1)} < \cdots < s_{(m)}$, where m is the total number of such distinct ordered points. It is clear that the ML estimator is determined by Λ_0 through the values of these observations. Huang and Wellner (1995) investigated the case where Λ_0 is assumed to be

a right continuous step function with a jump at distinct time points, and obtained the ML estimator by a Newton-Raphson algorithm. Suppose $(L_i, R_i], i = 1, 2, \ldots n$ are distinct pairs of end points, the number of distinct end points of the interval-censored observations could go up to $2n$. When the unknown function Λ_0 is assumed to be a step function with jumps at distinct points, as the sample size n becomes big, maximizing $L(\boldsymbol{\beta}, \Lambda_0|\mathcal{O})$ with respect to parameters $\boldsymbol{\beta}$ and jumps of Λ_0 at distinct points turns out to be a high-dimensional optimization problem. To manage this possibility, the sieve maximum likelihood approach is proposed.

The sieve maximum likelihood method proposed by Geman and Hwang (1982) as an attractive and efficient estimation approach in semiparametric and nonparametric inference is widely used in survival analysis. To apply this estimation procedure, the infinite-dimensional parameter in the original space is approximated by the finite parameters in the subset space. The subset space is also referred to as the sieve space, and is able to increase with the sample size (Grenander, 1981). As mentioned, the estimation of the infinite-dimensional parameter $\Lambda_0(t)$ will become difficult when vast observation intervals are involved. Therefore, instead of directly maximizing the likelihood function, $\Lambda_0(t)$ is approximated by the linear combination of a sequence of finite-dimensional parameters, such as, for example, $\Lambda_\theta(t)$. The ML estimator of $\boldsymbol{\beta}$ and $\Lambda_0(t)$ are obtained by maximising $L(\boldsymbol{\beta}, \Lambda_0|\mathcal{O})$, which is equivalent to maximising $L(\boldsymbol{\beta}, \Lambda_\theta|\mathcal{O})$ over the sieve space.

The sieve maximum likelihood estimation method has been applied to the partial linear Cox model and the partial linear AFT model for right censored data by Huang and Rossini (1997) and Lam and Xue (2005) respectively. Recently, Zeng et al. (2006) investigated its application to the AH model. The sieve MLE method has been studied for use in PH and PO models with interval-censored data by many researchers; see Rossini et al. (1996), Huang and Rossini (1997), Shen (1998) and Zhang et al. (2010) among others. The approach has also been utilized by Hu and Xiang (2013) in the presence of the cure fraction.

Compared with the conventional maximum likelihood estimation, the sieve maximum likelihood approach avoids the high-dimensional problem in the estimation procedure and achieves a faster convergence rate for estimating the infinite-dimensional parameters (Huang and Rossini, 1997).

1.6 Thesis Outline

The rest of the thesis is organized as follows.

In Chapter 2, a class of doubly semiparametric mixture cure models for interval-censored data is proposed, allowing a combination of linear and nonlinear effects of covariates in both mixture components. To manage challenges in estimation, a computationally feasible Bayesian estimation procedure has been developed, which includes a two-stage Poisson data augmentation for efficiently dealing with interval censoring, monotone splines and B-splines for modeling the cumulative baseline hazard function and nonlinear terms in the model. Our simulation results show that the proposed method has satisfactory performance in the finite sample cases. The utility of the proposed method is demonstrated by the analysis of data from a hypobaric decompression sickness study.

In Chapter 3, a generalized accelerated hazards cure model for interval-censored data is proposed, with bundled parameters involved in the nonparametric components. To manage the challenging aspects of the proposed model, a sieve maximum likelihood estimation is developed, which includes the B-spline functions for the nonparametric components in the survival risks and a two-step iterative algorithm for the estimation procedure. Simulation studies are carried out to evaluate the performance of the finite samples in various situations. The large sample properties, including the consistency, convergence of rate, and the asymptotic normality, are established. For illustrative purposes, the proposed model is applied to smoking cessation data to explore the effects of risk factors both in the latency and

incidence parts of the study.

In Chapter 4, a brief discussion about the proposed models and further research are given.

Chapter 2

Bayesian Doubly Semiparametric Mixture Cure Models with Interval-censored Data

The mixture cure model with interval-censored data, assuming the Cox proportional hazards model as a latency component for the survival risks of interested event and logistic regression as an incidence component for the probability of subjects to be cured, is an important tool to identify risk factors for the probability of being cured and survival of uncured subjects. In the literature, linear predictors are typically incorporated in both mixture components. In some application problems, the linear constraint is not sufficient to explain the effects of covariates on the logit of the cure probability or the log relative risk. For example, in the hypobaric decompression sickness study, the factor AGE may have nonlinear effects due to its time-related feature. Thus it is desirable to accommodate covariate effects nonlinearly in the model.

Besides assuming a semiparametric model in the latency part, a semiparametric incidence model would be more realistic than the popular parametric logistic

model due to unobserved cure statuses of subjects whose survival times are right-censored and thus less information obtained for the incidence part in practice compared with the latency part. There is only limited research in the literature relevant to this topic. Wang et al. (2012) considered a two-component mixture cure model with covariates entering in both latency and incidence components nonparametrically, and developed function estimates using smoothing spline analysis of variance methods. This work is based on right censored data and not applicable to the motivating example due to the complicated structure likelihood caused by interval censoring. For mixture cure models with interval-censored data, Shao et al. (2014) employed nonparametric smoothing for varying coefficients in regression analysis of the probability of cure and log survival times of the uncured subjects. In their model, a parametric accelerated failure time (AFT) model with a known location-scale error distribution is assumed to describe the log survival time of uncured subjects.

In this chapter, we aim to develop a new Bayesian doubly semiparametric mixture cure model for the analysis of interval-censored data in the presence of a cured subgroup, where the latency component is modeled by semiparametric proportional hazards with both linear and nonlinear effects of covariates in the log relative risk and the incidence component assumes a generalized partially linear single-index model for the probability of cure. A Bayesian estimation procedure is developed to facilitate efficient estimation of unknown parameters and nonparametric functions, where a two-stage Poisson data augmentation is utilized to deal with interval censoring, and B-splines and I-splines are used to approximate those unknown functions.

The rest of this chapter is organized as follows. In Section 2.1, the semiparametric model and the likelihood are given for the interval-censored data. In Section 2.2, we introduce the generalized partially linear single-index model on the cure rate and a monotone spline for the susceptible part. Moreover, a two-stage

Poisson data augmentation is applied for the interval-censored data. In addition, we specify the prior and posterior of parameters, following the sampling method of parameters. Simulation results are also presented in Section 2.3. We applied the proposed method to the real data in Section 2.4. Some discussions are conducted in Section 2.5.

2.1 Doubly Semiparametric Mixture Cure Models with Interval-censored Data

Denote the observations from total n subjects by $\mathcal{O} = (A_i, \delta_{L_i}, \delta_{I_i}, \delta_{R_i}, \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i, \boldsymbol{V}_i)$ for $i = 1, 2, \ldots, n$, where $A_i = (L_i, R_i]$ is the interval within which the failure time lies, $\delta_{L_i}, \delta_{I_i}$ and δ_{R_i} are the censoring indicators, with value 1 representing that the failure time of the ith subject is left-, interval- or right-censored, and 0 otherwise. Let $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ (or $\boldsymbol{V}_i$ and $\boldsymbol{W}_i$) be covariate vectors associated with the linear (or nonlinear) parts in regression analysis of the latency and incidence components, respectively. To incorporate the cure fraction, a latent random variable Y_i is introduced, where $Y_i = 1$ indicates the ith subject is susceptible and 0 otherwise. Assume that the probability of being uncured $P(Y_i = 1|\boldsymbol{Z}_i, \boldsymbol{W}_i) = \pi(\boldsymbol{Z}_i, \boldsymbol{W}_i)$, which depends on the covariates $\boldsymbol{Z}_i$ and $\boldsymbol{W}_i$ only. Let $S_p(\,\cdot\,)$ be the survival function for the population. The mixture cure model is defined as follows:

$$S_p(t|\boldsymbol{X}_i, \boldsymbol{V}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i) = \pi(\boldsymbol{Z}_i, \boldsymbol{W}_i) S(t|\boldsymbol{X}_i, \boldsymbol{V}_i) + 1 - \pi(\boldsymbol{Z}_i, \boldsymbol{W}_i), \tag{2.1}$$

where $S(t|\boldsymbol{X}_i, \boldsymbol{V}_i) = P(T > t|Y_i = 1, \boldsymbol{X}_i, \boldsymbol{V}_i)$ is the survival function and covariates $(\boldsymbol{X}_i, \boldsymbol{V}_i)$ in the latency part can be overlapped with $(\boldsymbol{Z}_i, \boldsymbol{W}_i)$ in the incidence part. To incorporate nonlinear covariate effects, we assume a doubly semiparametric mixture cure model, which allows a PH model for the latency part with some covariates $\boldsymbol{V}_i$ entering the model nonparametrically and a generalized partially

linear single-index model for the latency part. That is, the cumulative hazard function of susceptible subjects depends on $\boldsymbol{X}_i$ and $\boldsymbol{V}_i$ in the form of

$$\Lambda(t|\boldsymbol{X}_i, \boldsymbol{V}_i) = \Lambda_0(t) \exp\{\boldsymbol{X}_i^T \boldsymbol{\beta} + \phi(\boldsymbol{V}_i)\}, \tag{2.2}$$

and the probability of being uncured is modeled through a partially linear model

$$\pi(\boldsymbol{Z}_i, \boldsymbol{W}_i) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}, \quad \eta_i = \boldsymbol{Z}_i^T \boldsymbol{\alpha} + \psi(\boldsymbol{W}_i) + \epsilon_i, \quad i = 1, \ldots n, \tag{2.3}$$

where $\Lambda_0(\,\cdot\,)$ is the cumulative baseline hazard function, $\phi(\,\cdot\,)$ and $\psi(\,\cdot\,)$ are unknown smooth functions, $\epsilon_i \sim N(0, \sigma^2)$ is the individual effect, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are regression coefficients to be estimated. The utilisation of the smooth function $\psi(\,\cdot\,)$ makes it possible to incorporate the nonlinear effect in the incidence part. To avoid the curse of dimensionality possibly in covariates $\boldsymbol{W}$, we extend the uncured probability to a generalized partially linear single-index model, that is,

$$\eta_i = \boldsymbol{Z}_i^T \boldsymbol{\alpha} + g(\boldsymbol{W}_i^T \boldsymbol{\xi}) + \epsilon_i, \quad i = 1, \ldots n, \tag{2.4}$$

where the link function $g(\,\cdot\,)$ is an unknown smooth function, which represents the effects of covariate $\boldsymbol{W}_i$ on η_i through a single score $\boldsymbol{W}_i^T \boldsymbol{\xi}$, and thus the interpretation of the impacts of these covariates can be obtained based on the estimated absolute value of the coefficients after standardizing covariates. Amico et al. (2018) employed the single-index model for the uncured probability only, while in the proposed model, both the linear and nonlinear covariate effects are considered in the latency and incidence parts.

Unlike the usual logistic model for the incidence part, additional random individual effects ϵ_i are included in the partially linear predictor η_i of model (2.3) for capturing the lack of fit of the model due to extra variation, possible outliers or unexplained source variation in the observed data. Employing the similar method

of Poon and Wang (2013), we assume that $\epsilon_i \sim N(0, \sigma^2)$. To ensure identifiability, we assume as usual that the index vector $\boldsymbol{\xi}$ satisfies that $\| \boldsymbol{\xi} \| = 1$ with the first element $\boldsymbol{\xi}_1 > 0$, where $\| \cdot \|$ is the Euclidean norm. It is clear that when $\boldsymbol{V}_i$ and $\boldsymbol{W}_i$ are zeros, the proposed model reduces to the conventional mixture cure model and it reduces to the mixture cure model with two nonparametric forms given in Wang et al. (2012), if we leave out the linear part in equation (2.2) and equation (2.3).

Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{\xi}^T, \Lambda_0, \phi, g)^T$ be a set of unknown terms in the model. For presentation simplicity, we use π_i for $\pi(\boldsymbol{Z}_i, \boldsymbol{W}_i)$, $S(t)$ for $S(t|\boldsymbol{X}_i, \boldsymbol{V}_i)$, and $S_p(t)$ for $S_p(t|\boldsymbol{X}_i, \boldsymbol{V}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i)$ in the following context. Let $\mathcal{O}$ be n observed data, the likelihood based on $\mathcal{O}$ is given as follows:

$$\begin{aligned} L(\boldsymbol{\theta}|\mathcal{O}) &= \prod_{i=1}^{n} \{1 - S_p(R_i)\}^{\delta_{L_i}} \{S_p(L_i) - S_p(R_i)\}^{\delta_{I_i}} S_p(L_i)^{\delta_{R_i}} \\ &= \prod_{i=1}^{n} \pi_i^{1-\delta_{R_i}} \{1 - S(R_i)\}^{\delta_{L_i}} \{S(L_i) - S(R_i)\}^{\delta_{I_i}} \{1 - \pi_i + \pi_i S(L_i)\}^{\delta_{R_i}}. \end{aligned} \tag{2.5}$$

We notice that $Y_i = 1$ if $\delta_{R_i} = 0$. The right censored data enter to the likelihood with contribution $1 - \pi_i$ when $Y_i = 0$ and $\pi_i S(L_i)$ when $Y_i = 1$. Using the relation that $(1 - Y_i)\delta_{R_i} = (1 - Y_i)$, the conditional likelihood function in terms of $Y_i's$ is:

$$L(\boldsymbol{\theta}|\mathcal{O}) = \prod_{i=1}^{n} \pi_i^{Y_i} (1 - \pi_i)^{1-Y_i} [\{1 - S(R_i)\}^{\delta_{L_i}} \{S(L_i) - S(R_i)\}^{\delta_{I_i}} S(L_i)^{\delta_{R_i}}]^{Y_i}. \tag{2.6}$$

The complete data likelihood is

$$L(\boldsymbol{\theta}|\mathcal{O}, Y) = \prod_{i=1}^{n} \pi_i^{Y_i} (1 - \pi_i)^{1-Y_i} [\{1 - S(R_i)\}^{\delta_{L_i}} \{S(L_i) - S(R_i)\}^{\delta_{I_i}} S(L_i)^{\delta_{R_i}}]^{Y_i}. \tag{2.7}$$

To handle the missing data, we impute the value of Y from Bernoulli distribution with the parameter $\mathfrak{p}$, given the values of π_i and $S(t_i)$:

$$\mathfrak{p} = P(Y_i = 1|\boldsymbol{X}_i, \boldsymbol{V}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i, T > t_i) = \frac{\pi_i S(t_i)}{1 - \pi_i + \pi_i S(t_i)}. \tag{2.8}$$

In practice, $g(\cdot)$, $\phi(\cdot)$, $\Lambda_0(\cdot)$ are all unknown functions. It is a challenge to directly maximize the complete data likelihood in (2.7) with respect to all parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\xi}$ and functions $g(\cdot)$, $\phi(\cdot)$ and $\Lambda_0(\cdot)$, due to the complex structure. We therefore propose a Bayesian estimation procedure in the next section to tackle the estimation problem.

2.2 Bayesian Inference

Before introducing our Bayesian estimation approach, we first employ splines to approximate unknown functions in a convenient and efficient way and then employ data augmentation to provide a proper likelihood, so that the posterior distributions of parameters can be calculated given prior beliefs.

2.2.1 Splines for nonparametric terms in both parts

In the generalized partially linear single-index model (2.4) fitting the cure fraction π_i, to approximate the unknown function $g(\cdot)$, B-spline functions are adopted over an interval $[a, b]$, where $a = \min_{1\leq i\leq n}\{\boldsymbol{W}_i^T\boldsymbol{\xi}\}$ and $b = \max_{1\leq i\leq n}\{\boldsymbol{W}_i^T\boldsymbol{\xi}\}$. Let $a = d_0 < d_1 \cdots < d_k = b$ be the partition of the interval $[a, b]$, then $g(u)$ can be approximated by:

$$g(u) = \sum_{j=1}^{K_n} B_j(u)\zeta_j = \boldsymbol{B}(u)^T\boldsymbol{\zeta}, \quad u \in [a, b], \tag{2.9}$$

where $\boldsymbol{B}(u) = (B_1(u), B_2(u) \cdots, B_{K_n}(u))^T$ are spline basis functions with order m and $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots \zeta_{K_n})^T$ are the regression coefficients to be estimated, and $K_n = m + k - 1$ is the total number of spline functions. Therefore, the partially linear

predictor η_i in the generalized partially linear single-index model (2.4) becomes:

$$\eta_i = \boldsymbol{Z}_i^T\boldsymbol{\alpha} + \boldsymbol{B}(\boldsymbol{W}_i^T\boldsymbol{\xi})^T\boldsymbol{\zeta} + \epsilon_i = \mathcal{B}_i(\boldsymbol{\xi})^T\boldsymbol{\vartheta} + \epsilon_i, \quad i = 1, \dots n, \tag{2.10}$$

where $\mathcal{B}_i(\boldsymbol{\xi}) = (\boldsymbol{Z}_i^T, B(\boldsymbol{W}_i^T\boldsymbol{\xi})^T)^T$, and $\boldsymbol{\vartheta} = (\boldsymbol{\alpha}^T, \boldsymbol{\zeta}^T)^T$. The linear regression framework makes it easy for us to sample regression coefficients $\boldsymbol{\vartheta}$ and variance σ^2 when computing posterior distributions of parameters. It also explains the use of such a random terms ϵ_i in the predictor of the generalized partially single-index model (2.4).

Similar to the treatment in the incidence part, splines are also utilized to model the unknown smooth function $\phi(\cdot)$ and cumulative baseline hazard function $\Lambda_0(\cdot)$ in the latency part (2.2). Particularly, on the support of $[a_1, b_1]$ with a_1 and b_1 being the minimum and maximum of $V_i, i = 1, 2, \dots n$, $\phi(\cdot)$ can be approximated by B-splines:

$$\phi(h) = \boldsymbol{c}(h)^T\boldsymbol{\varrho}, \quad h \in [a_1, b_1], \tag{2.11}$$

where $\boldsymbol{c}(h) = (c_1(h), \cdots, c_N(h))^T$ are basis functions, and $\boldsymbol{\varrho} = (\varrho_1, \cdots, \varrho_N)^T$ are spline coefficients. Furthermore, we utilize the following monotone splines (Ramsay, 1988) to fit the cumulative baseline hazard function $\Lambda_0(\cdot)$ incorporating its nondecreasing property:

$$\Lambda_0(t) = \sum_{l=1}^{k} \gamma_l I_l(t|d), \tag{2.12}$$

where $I_l(t|d)$ for $l = 1, 2, \dots k$ are the nondecreasing basis functions ranging from 0 to 1 with the degree d to present the smoothness, and γ_l are the corresponding coefficients, which are assumed to be nonnegative to ensure the nondecreasing of $\Lambda_0(t)$. The estimates of parameters γ_l are obtained by sampling from the posterior distributions, which can be calculated easily based on the likelihood function given the spline approximation of Λ_0 in (2.12).

Plugging (2.11) and (2.12) into model (2.2), the cumulative hazard function for susceptible subjects can then be written as:

$$\Lambda(t) = \Lambda_0(t)\exp\{\boldsymbol{X}_i^T\boldsymbol{\beta} + \boldsymbol{c}(\boldsymbol{V}_i)^T\boldsymbol{\varrho}\} = \sum_{l=1}^{k}\gamma_l I_l(t|d)\exp\{\boldsymbol{Q}_i^T\boldsymbol{\rho}\}, \tag{2.13}$$

where $\boldsymbol{Q}_i = (\boldsymbol{X}_i^T, \boldsymbol{c}(\boldsymbol{V}_i)^T)^T$, and $\boldsymbol{\rho} = (\boldsymbol{\beta}^T, \boldsymbol{\varrho}^T)^T$.

When implementing the aforementioned splines, the smoothness of polynomials, the location and the number of knots are essential for the performance of these approximations. In general, quadratic and cubic splines are good enough to provide sufficient smoothness, and the location of knots can be either equally-based or quantile-based or free knots. In the numerical analysis of this study, we use the quantile-based knots and the number of knots are set between 10-15 for the sake of computational efficiency.

2.2.2 Data augmentation for interval-censored data

To facilitate efficient Bayesian estimation, we propose a two-stage Poisson data augmentation that is motivated by the interval-censored data structure and the PH model properties. This method introduces latent Poisson random variables to simplify the distributions in the likelihood along similar lines of Lin et al. (2015). Let $N(t)$ be a latent Poisson process, which is the number of events having occurred till time t with intensity $\Lambda_0(t)\exp\{\boldsymbol{X}^T\boldsymbol{\beta} + \phi(\boldsymbol{V})\}$, and $T = \inf\{t, N(t) > 0\}$ be the time of the first occurrence in the Poisson process. We then have the survival function of a susceptible subject:

$$P(T > t) = P(N(t) = 0) = \exp\{-\Lambda_0(t)\exp(\boldsymbol{X}^T\boldsymbol{\beta} + \phi(\boldsymbol{V}))\}. \tag{2.14}$$

From this equation, the structure of interval-censored data can be expressed by two Poisson processes. Let $t_{i1} = R_i * I_{(\delta_{L_i}=1)} + L_i * I_{(\delta_{I_i}=1)}$ and $t_{i2} = R_i * I_{(\delta_{I_i}=1)} + L_i *$

$I_{(\delta_{R_i}=1)}$. Then $v_i = N(t_{i1})$ and $\omega_i = N(t_{i2}) - N(t_{i1})$ are the number of occurrence till time t_{i1}, and the number of occurrence between times t_{i2} and t_{i1}, respectively. It is clear that both v and w are independent Poisson random variables with $v_i \sim \mathcal{P}\{\Lambda_0(t_{i1})\exp(\boldsymbol{X_i^T\beta}+\phi(\boldsymbol{V_i}))\}$ and $\omega_i \sim \mathcal{P}\{[\Lambda_0(t_{i2})-\Lambda_0(t_{i1})]\exp(\boldsymbol{X}_i^T\boldsymbol{\beta}+\phi(\boldsymbol{V}_i))\}$, respectively, where $\mathcal{P}(a)$ is a Poisson distribution with mean a. For interval-censored data (L_i, R_i), we have

$$\begin{aligned} P(t_{i1} < T \leq t_{i2}) &= P\{N(t_{i1}) = 0, N(t_{i2}) > 0\} \\ &= P(v_i = 0, \omega_i > 0) = S(L_i|X_i, V_i) - S(R_i|X_i, V_i). \end{aligned} \tag{2.15}$$

Similarly, the corresponding probabilities of observing left-censored and right-censored failure times can be obtained when $v_i > 0$ and $(v_i = 0,\ \omega_i = 0)$, respectively. In the following, we denote by $\mathcal{P}(.|\mu)$ the probability mass function of a Poisson distribution with rate parameter μ. In the first stage of data augmentation, the likelihood function with augmented data can be obtained from Eq (2.6) by

$$\begin{aligned} L_1(\boldsymbol{\theta}|\mathcal{O}) = &\prod_{i=1}^{n} \pi_i^{Y_i}(1-\pi_i)^{1-Y_i} \\ &\times\Big(\mathcal{P}\big\{v_i|\Lambda_0(t_{i1})\exp(\boldsymbol{X}_i^T\boldsymbol{\beta}+\phi(\boldsymbol{V}_i))\big\} \\ &\times\mathcal{P}\big\{\omega_i|[\Lambda_0(t_{i2})-\Lambda_0(t_{i1})]\exp(\boldsymbol{X}_i^T\boldsymbol{\beta}+\phi(\boldsymbol{V}_i))\big\}^{\delta_{I_i}+\delta_{R_i}}\Big). \end{aligned} \tag{2.16}$$

Note that integrating the v_is and w_is out, the likelihood (2.16) reduces to the likelihood (2.6).

It is easy to maximize the likelihood function in (2.16) due to its form of a production of Poisson probability mass functions. To incorporate the spline approximations and perform Bayesian estimation more efficiently, we further conduct a Poisson data augmentation again in the second stage. Specifically, we decompose each v_i and w_i as the sum of k independent Poisson latent variables, that is

$v_i = \sum_{l=1}^{k} v_{il}$ and $w_i = \sum_{l=1}^{k} w_{il}$, where $v_{il} \sim \mathcal{P}\{\gamma_l I_l(t_{i1}) \exp(X_i^T \boldsymbol{\beta} + \phi(\boldsymbol{V}_i))\}$ and $w_{il} \sim \mathcal{P}\{\gamma_l [I_l(t_{i2}) - I_l(t_{i1})] \exp(X_i^T \boldsymbol{\beta} + \phi(\boldsymbol{V}_i))\}$ for $l = 1, 2 \ldots k$. By the additivity property of Poisson random variables and the spline approximation, the data augmented likelihood function (2.16) becomes

$$\begin{aligned} L_2(\boldsymbol{\theta}|\mathcal{O}) = & \prod_{i=1}^{n} \prod_{l=1}^{k} \pi_i^{Y_i} (1-\pi_i)^{1-Y_i} \\ & \times \Big[\mathcal{P}\big\{ v_{il} | \gamma_l I_l(t_{i2}) \exp(\boldsymbol{Q}_i^T \boldsymbol{\rho}) \big\} \\ & \times \mathcal{P}\big\{ w_{il} | \gamma_l [I_l(t_{i2}) - I_l(t_{i1})] \exp(\boldsymbol{Q}_i^T \boldsymbol{\rho}) \big\}^{\delta_{I_i} + \delta_{R_i}} \Big]^{Y_i}. \end{aligned} \tag{2.17}$$

With the constraint that $\sum_{l=1}^{k} v_{il} = 0$ for left censored data, $\sum_{l=1}^{k} v_{il} = 0$ and $\sum_{l=1}^{k} w_{il} > 0$ for interval-censored data, and $\sum_{l=1}^{k} v_{il} = 0$ and $\sum_{l=1}^{k} w_{il} = 0$ for right censored data. If we treat v_{il} and w_{il} as missing data, the data augmented likelihood function L_2 in equation (2.17) is then a complete data likelihood in the multiplicative form of Poisson probability mass functions and with direct connections to all coefficients in splines and regression coefficients. It provides the basis of the proposed Bayesian estimation procedure in this paper.

2.2.3 Prior specification

In the Bayesian framework, we have to specify priors for unknown parameters, including the single-index vector $\boldsymbol{\xi}$ and regression coefficients $\boldsymbol{\vartheta}$ in the incidence part, regression coefficients $\boldsymbol{\rho}$ and spline coefficients $\boldsymbol{\gamma}$ in the latency part. In particular, we assign the following priors.

- $\boldsymbol{\xi}$: a flat but proper prior is assumed, i.e., $p(\boldsymbol{\xi}) \propto 1$, which is the uniform distribution on the half of the unit circle $\{\boldsymbol{\xi} : \|\boldsymbol{\xi}\| = 1 \text{ with } \xi_1 > 0\}$. A noninformative prior is chosen for $\boldsymbol{\xi}$ with the only conditions that $\boldsymbol{\xi}^T\boldsymbol{\xi}$=1 and $\boldsymbol{\xi}_1 > 0$.

- σ^2, $\boldsymbol{\vartheta}$: we specify an inverse-Gamma prior on σ^2(or Jeffreys's prior)

$$p(\sigma^2) \propto (\sigma^2)^{-r/2-1} \exp\{-s_0^2/(2\sigma^2)\}, \tag{2.18}$$

where r and s_0^2 are two known hyperparameters; and a conditional normal prior on $\boldsymbol{\vartheta}$ with mean 0 and variance matrix ςI_{K+q}, that is

$$p(\boldsymbol{\vartheta}|\sigma^2, \varsigma) \propto (2\pi\varsigma\sigma^2)^{-\frac{K+q}{2}} \exp\left\{ -\frac{\boldsymbol{\vartheta}^T\boldsymbol{\vartheta}}{2\varsigma\sigma^2} \right\}, \tag{2.19}$$

where I_{K+q} is the identity matrix of order $K+q$. Thus the joint prior of $(\boldsymbol{\vartheta}, \sigma^2)$ is in the factorised form of

$$p(\boldsymbol{\vartheta}, \sigma^2) \propto p(\boldsymbol{\vartheta}|\sigma^2)p(\sigma^2). \tag{2.20}$$

This specification of prior distributions makes it possible to integrate out σ^2 and ϑ from the joint posterior distribution, which reduces the dimension of parameters and improves the computational efficiency.

- $\boldsymbol{\rho}$: the conventional independent normal priors are assigned on the regression coefficients in the latency component, for any ρ_j in $\boldsymbol{\rho}$, $\rho_j \sim N(\rho_0, \sigma_\rho^2)$, where $\rho_0 = 0$ and $\sigma_\rho^2 = 100$. The specification of normal prior results in the log-concave posterior distributions for $\boldsymbol{\rho}$, which are easily sampled by adaptive rejection algorithm(ARS) (Gilks and Wild, 1992).

- $\boldsymbol{\gamma}$: to obtain conjugate posterior distributions of $\gamma_l s$, we specify independent exponential priors $\exp(\tau)$ for $\gamma_l s$ and assign a hyper gamma prior $\mathcal{G}(a_\tau, b_\tau)$ on hyperparameter τ with mean a_τ/b_τ, and variance a_τ/b_τ^2. Besides its appealing computational properties in producing conjugate posteriors of parameters $\boldsymbol{\gamma}$ and τ, this prior specification plays the same role as a L_1 penalty on all coefficients $\gamma_l s$ in the penalized likelihood method (Park and Casella, 2008)

with tuning parameter τ, making the coefficients of unnecessary spline basis shrink to zeros.

2.2.4 Sampling for posterior computations

To perform Bayesian estimation, the posterior distributions of parameters are required. Given the augmented joint likelihood in (2.17) and the priors specified for unknown parameters and latent variables, the posterior distributions can be derived, but very complicated under the proposed model. We use the Gibbs sampler to generate samples for each parameter, based on the full posterior distribution. The sampling scheme is given below.

1. Sample $Y_i \sim Bernoulli(\mathfrak{p})$ for those subjects with $\delta_{R_i} = 1$, where $\mathfrak{p}$ is defined as in equation (2.8).

2. Sample v_i, v_{il}, w_i and w_{il} for all $l = 1, 2 \dots k$ and $i = 1, 2 \dots n$. To ensure the condition that $v_i > 0$, we first sample v_i and then sample v_{il} according to the multinomial distribution. If $\delta_{L_i} = 1$, sample

$$\begin{aligned} v_i &\sim \mathcal{P}\{\Lambda_0(t_{i1}) \exp(\boldsymbol{Q}_i^T \boldsymbol{\rho})\}, \\ (v_{i1}, v_{i2}, \dots v_{ik} | v_i) &\sim \mathcal{M}(v_i, \boldsymbol{p}_i), \boldsymbol{p}_i = (p_{i1}, p_{i2} \dots p_{ik}), \\ p_{il} &= \gamma_l I_l \{\sum_{j=1}^{k} \gamma_j I_j(t_{i1})\}^{-1}, l = 1, 2 \dots k, \end{aligned} \tag{2.21}$$

   where $\mathcal{M}(v_i, \boldsymbol{p}_i)$ is a multinomial distribution with index vector v_i and corresponding probabilities $\boldsymbol{p}_i$. For $\delta_{I_i} = 1$, sample $w_i \sim \mathcal{P}\{[\Lambda_0(t_{i2}) - \Lambda_0(t_{i1})] \exp(\boldsymbol{Q}_i^T \boldsymbol{\rho})\}$, and sample w_{il} similar to the way for v_{il}.

3. The full conditional distribution of $\boldsymbol{\rho}$ is proportional to

$$\begin{aligned}&\exp(\sum_{i=1}^{n}\boldsymbol{Q}_i^T\boldsymbol{\rho}[v_i+(\delta_{I_i}+\delta_{R_i})w_i]\\&-\sum_{i=1}^{n}\exp(\boldsymbol{Q}_i^T\boldsymbol{\rho})\{[\Lambda(t_{i2})-\Lambda(t_{i1})](\delta_{I_i}+\delta_{R_i})+\Lambda(t_{i1})\}p(\boldsymbol{\rho}),\end{aligned} \tag{2.22}$$

   which is sampled by the ARS method (Gilks and Wild, 1992) due to the log-concave property.

4. The full conditional distribution of $\boldsymbol{\gamma}$ is a conjugate gamma posterior distribution given the exponential prior. Sample γ_l from the gamma distribution $\mathcal{G}(a_{\gamma_l}, b_{\gamma_l})$, where

$$\begin{aligned}a_{\gamma_l}&=1+\sum_{i=1}^{n}(\delta_{I_i}v_{il}+\delta_{R_i}w_{il}),\\b_{\gamma_l}&=\tau+\sum_{i=1}^{n}\exp(\boldsymbol{Q}_i^T\boldsymbol{\rho})\big\{I(t_{i1})+[I(t_{i2})-I(t_{i1})](\delta_{I_i}+\delta_{R_i})\big\}.\end{aligned} \tag{2.23}$$

5. The hyperparameter τ is sampled from the $\mathcal{G}(a_\tau+k, b_\tau+\sum_{l=1}^{k}\gamma_l)$.

6. For $i=1,2\ldots n$, the full conditional distribution of η_i is:

$$p(\eta_i|Y_i,\boldsymbol{\vartheta},\boldsymbol{\xi},\boldsymbol{\rho},\boldsymbol{\gamma},\sigma^2)\propto\pi_i^{Y_i}(1-\pi_i)^{1-Y_i}\times\exp\left\{-\frac{(\eta_i-\mathcal{B}_i(\xi)^T\boldsymbol{\vartheta})^2}{2\sigma^2}\right\}, \tag{2.24}$$

   where $\pi_i=\frac{\exp(\eta_i)}{1+\exp(\eta_i)}$. The complex structure of posterior distribution leads us to employ the Slice sampler (Damlen et al., 1999) to sample $\boldsymbol{\eta}$. According to the Slice Sampler, sampling from a density function $f(x)\propto\pi(x)l(x)$ is equivalent to that from the density function $f(x,\mu)\propto\pi(x)I(\mu<l(x))$, where μ is a latent variable and x is restricted to the set $A_\mu=\{x:l(x)>\mu\}$. The

joint likelihood of (η_i, μ) is:

$$p(\eta_i, \mu|Y_i, \boldsymbol{\vartheta}, \boldsymbol{\xi}, \sigma^2) \propto \exp\left\{ -\frac{(\eta_i - \mathcal{B}_i(\xi)^T\boldsymbol{\vartheta})^2}{2\sigma^2}\right\} I\{0 < \mu < \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}\}. \tag{2.25}$$

The new state of η_i is sampled by first drawing μ for the uniform distribution on the interval $(0,\ \pi_i^{Y_i}(1-\pi_i)^{1-Y_i})$, and then generating η_i from the normal density $N(\mathcal{B}_i(\xi)^T\boldsymbol{\vartheta}, \sigma^2)$ truncated on the set $\{\eta_i : \pi_i^{Y_i}(1-\pi_i)^{1-Y_i} > \mu\}$.

7. The joint fully conditional posterior of single-index $\boldsymbol{\xi}$, regression coefficients of cure rate $\boldsymbol{\vartheta}$, and the variance σ^2 is

$$\begin{aligned} p(\boldsymbol{\xi}, \boldsymbol{\vartheta}, \sigma^2|\boldsymbol{\eta}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \varsigma,) &\propto \sigma^{-n}\exp\left\{ -\frac{(\eta_i - \mathcal{B}_i(\boldsymbol{\xi})^T\boldsymbol{\vartheta})^2}{2\sigma^2}\right\}(2\pi\sigma^2)^{-\frac{K+q}{2}} \\ &\quad\times\exp\left\{ -\frac{\boldsymbol{\vartheta}^T\boldsymbol{\vartheta}}{2\varsigma\sigma^2} - \frac{s_0^2}{2\sigma^2}\right\}(\sigma^2)^{-\frac{r}{2}-1} \\ &\propto |\Sigma|^{-\frac{1}{2}}(S^2+s_0^2)^{-\frac{n+r}{2}}(S^2+s_0^2)^{\frac{n+r}{2}}(\sigma^2)^{-\frac{n+r}{2}-1}\exp\left\{ -\frac{S^2+s_0^2}{2\sigma^2}\right\} \\ &\quad\times(2\pi\sigma^2)^{-\frac{K+q}{2}}|\Sigma|^{\frac{1}{2}}\exp\{-\frac{1}{2\sigma^2}(\boldsymbol{\vartheta}-\widetilde{\boldsymbol{\vartheta}})^T\Sigma(\boldsymbol{\vartheta}-\widetilde{\boldsymbol{\vartheta}})\}, \end{aligned} \tag{2.26}$$

where

$$\begin{aligned} S^2 &= \boldsymbol{\eta}^T\boldsymbol{\eta} - \widetilde{\boldsymbol{\vartheta}}^T\Sigma\widetilde{\boldsymbol{\vartheta}}, \\ \Sigma &= \boldsymbol{B}(\boldsymbol{\xi})^T\boldsymbol{B}(\boldsymbol{\xi}) + \varsigma^{-1}I_{K+q}, \\ \widetilde{\boldsymbol{\vartheta}} &= \Sigma^{-1}\boldsymbol{B}(\boldsymbol{\xi})^T\boldsymbol{\eta}. \end{aligned} \tag{2.27}$$

Considering the complicated form of posterior distributions, we use the method of composition (Tanner, 1991). From the method of composition, to obtain samples $y \sim \int f(y|x)g(x)dx$, one can first sample x^* from $g(x)$, and then sample from the density $f(y|x^*)$. By treating the marginal posterior

$$p(\boldsymbol{\xi}|\boldsymbol{\eta}, \varsigma, \sigma^2, \boldsymbol{\vartheta}) \propto |\Sigma|^{-\frac{1}{2}}(S^2+s_0^2)^{-\frac{n+r}{2}} \tag{2.28}$$

as our target distribution, we can first sample σ^2 from the conditional inverse-gamma

$$p(\sigma^2|\boldsymbol{\eta}, \boldsymbol{\xi}, \varsigma, \boldsymbol{\vartheta}) = (S^2 + s_0^2)^{\frac{n+r}{2}} (\sigma^2)^{-\frac{n+r}{2}-1} \exp\left\{ - \frac{S^2 + s_0^2}{2\sigma^2} \right\} \tag{2.29}$$

with shape $\frac{n+r}{2}$ and scale $\frac{S^2+s_0^2}{2}$, and then sample $\boldsymbol{\vartheta}$ from the conditional normal distribution $N(\widetilde{\boldsymbol{\vartheta}}, \Sigma)$

$$p(\boldsymbol{\vartheta}|\boldsymbol{\eta}, \boldsymbol{\xi}, \varsigma, \sigma^2) = (2\pi\sigma^2)^{-\frac{K+q}{2}} |\Sigma|^{\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2}(\boldsymbol{\vartheta} - \widetilde{\boldsymbol{\vartheta}})^T \Sigma (\boldsymbol{\vartheta} - \widetilde{\boldsymbol{\vartheta}})\}, \tag{2.30}$$

where S^2, Σ and $\widetilde{\boldsymbol{\vartheta}}$ are defined in (2.27). After sampling $\boldsymbol{\vartheta}$ and σ^2, sample $\boldsymbol{\xi}$ from (2.28) by the Metropolis-Hashing (MH) algorithm. For a given $\boldsymbol{\xi}$, the proposed single-index $\boldsymbol{\xi}^*$ is generated from a Fisher-von Mises distribution $f_i(\boldsymbol{\xi}^*|\boldsymbol{\xi}, \rho_\xi) \propto \exp(\rho_\xi \boldsymbol{\xi}^T \boldsymbol{\xi}^*)$, where $\boldsymbol{\xi}$ is regarded as the modal vector and ρ_ξ is the concentration parameter, acting as turning parameter.

We accept $\boldsymbol{\xi}^*$ with the probability

$$\min\{1, \left(\frac{|\Sigma|}{|\Sigma^*|}\right)^{\frac{1}{2}} \left(\frac{S^2 + s_0^2}{S^{*2} + s_0^2}\right)^{\frac{n+r}{2}}\}. \tag{2.31}$$

With given initial values of the unknown parameters, the iterative algorithm of the proposed Bayesian estimation can be summarised as follows.

- Draw $Y_i, i = 1, 2 \ldots n$ from the Bernoulli distribution with probability parameter given in equation (2.8).
- If $Y_i = 1$
  - sample $v_i s$, $v_{il} s$, $w_i s$ and $w_{il} s$ from equation (2.21),
  - sample $\boldsymbol{\beta}$ by the ARS method from equation (2.22),
  - sample γ_l from $\mathcal{G}(a_{r_l}, b_{r_l})$, where a_{r_l}, b_{r_l} are given in equation (2.23),

– sample τ from $\mathcal{G}(a_\tau + k, b_\tau + \sum_{l=1}^{k} \gamma_l)$.

- Draw $\boldsymbol{\eta}$ from equation (2.25) by the Slice sampler.
- Draw σ from the inverse-gamma distribution, with shape is $\frac{n+r}{2}$ and scale is $\frac{S^2+s_0^2}{2}$.
- Draw $\boldsymbol{\vartheta}$ from $\mathcal{N}(\widetilde{\boldsymbol{\vartheta}}, \Sigma)$, where $\widetilde{\boldsymbol{\vartheta}}$ and Σ are defined in equation (2.30).
- Draw $\boldsymbol{\xi}$ from equation (2.28) by the MH algorithm.

Update $\Lambda(t|x) = \sum_{l=1}^{k} \gamma_l I_l(t|d)$ and $\pi(\boldsymbol{\eta}) = \frac{\exp(\boldsymbol{\eta})}{1+\exp(\boldsymbol{\eta})}$ based on the samplers. Repeating the above sampling procedures for 10000 iterations and the first 2500 iterations are burn-in. Taking the average of rest of the samplers as our estimations of parameters.

2.3 Simulation Studies

The simulation study is carried out to examine the performance of the proposed model. We generate the binary uncured indicator Y_i from Bernoulli($\pi(\eta_i)$) with

$$\pi(\eta_i) = P(Y_i = 1) = \frac{1}{1+\exp(-\eta_i)}, \quad \eta_i = z_i\alpha + g(\boldsymbol{W}_i^T\boldsymbol{\xi}) + \epsilon_i, i = 1, \ldots, n \quad (2.32)$$

and failure times T_i for susceptible subjects (when $Y_i = 1$) from the following partially linear proportional hazards model:

$$\Lambda(t) = \Lambda_0(t)\exp\{x_i\beta + \phi(V_i)\}, \quad (2.33)$$

where covariates $\boldsymbol{W}_i$ in the single-index part of the logistic model (2.32) are bivariate with $W_{i1} \sim U(0,1)$ and $W_{i2} \sim N(0,1)$, $z_i \sim U(-1,1)$, and the individual effects $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. We take the parameters $\xi = (1,1)/\sqrt{2}$, $\alpha = 0.5$, $\sigma_\epsilon = 0.7$ or 1 and specify the single-index function in the forms of $g(u) = 0.5 + \exp\{-4(u-1)^2\}$

and $g(u) = -0.5 + \exp\{-4(u-1)^2\}$. Based on this setting, different intercepts 0.5 and -0.5 in $g(u)$ lead to average cure probabilities of 33% and 54%, respectively. In the latency model (2.33), we set covariate associated with the nonlinear term $V_i \sim U(0.5, 1.5)$, the cumulative hazard function $\Lambda_0(t) = t^2$, the nonlinear function $\phi(v) = \exp(-3 * v)$, and take the parameter $\beta = 0.5$.

The interval-censored observations $(T_{L_i}, T_{R_i}, \delta_{L_i}, \delta_{I_i}, \delta_{R_i})$ of the failure time are generated in the way similar to that in Xiang et al. (2011).

1. Set the censoring time (the length of study) $C_i = 3$.

2. When $Y_i = 0$, we have that subject i is cured and let $T_{L_i} = C_i$ and $\delta_{R_i} = 1$.

3. When $Y_i = 1$, that is subject i is uncured, we compare values of T_i and C_i. If $T_i > C_i$, then we have right censoring and let $\delta_{R_i} = 1$, $T_{L_i} = C_i$ and $T_{R_i} = \infty$. Otherwise, we have interval censoring and let $\delta_{R_i} = 0$.

4. For those subjects with $\delta_{R_i} = 0$, we generate l from the exponential distribution with mean 0.2 and the interval length $len \sim U(0.2, 0.4)$. From $(0, l], (l, l + len], (l + len, l + 2 * len] \ldots (l + k * len, \infty), k = 1, 2, \ldots,$ find $(T_{L_i}, T_{R_i}]$ satisfies that $T_{L_i} < T_i \leq T_{R_i}$. If $T_{L_i} = 0$, we have left censoring and let $\delta_{L_i} = 1$, else let $\delta_{I_i} = 1$.

With the above simulation design, the rates of right censoring in the simulated samples are around 60-65% and 30-35% in total for $g(u) = -0.5 + \exp\{-4(u-1)^2\}$ and $g(u) = 0.5 + \exp\{-4(u-1)^2\}$, respectively.

We repeat the experiment 200 times with sample size of $n = 200$ and 500. The true values of hyperparameters are set to be $\sigma_\beta = 10$, $a_\tau = b_\tau = 1$ and $s_0^2 = r = 1$. We choose the tuning parameters $\iota = \rho_\xi = 10$, and use the quadratic polynomial splines with the degree of freedom $K = 5$ to fit the nonlinear single-index function in the incidence part and the nonlinear terms in the latency part.

We run the Gibbs sampling in each simulation run for 10000 iterations including 2500 burn-in iterations. The estimate of each parameter is obtained by the

sample mean of all samples generated from the corresponding posterior distribution of the parameter. We report simulation results in terms of the average bias (AVE.bias), average standard deviation (Ave.SD), and coverage probability (CP) of 95% credible interval (Let C be the credible interval and θ_j be the sampler in jth iteration, $CP = \sum_{j=1}^{m} I(\theta_j \in C)/m$, where m is the total number of iteration and $I(\cdot)$ is the indicator function.) in Tables 2.1-2.2 for σ^2=0.5 with sample size ranging from 200 to 500, respectively.

As expected, for most parameters the average biases of the proposed estimates increase as censoring rate increases, CPs for under the proposed method are all close to the nominal level of 0.95. When the sample size increases, biases of most proposed estimates decrease. For comparison purpose, we also fit each simulated sample using the usual PH mixture cure (PHMixCure) model in which only linear covariate effects are included in the incidence and latency parts. Corresponding results obtained by R package **GORCure** (Zhou et al., 2018) are also presented in both tables. It can be seen that, in the presence of covariates ξ_i and V with nonlinear effects in both the incidence and latency parts, misspecifying their effects linearly in the PHMixCure model leads to considerably biased parameter estimates with large SDs and consequently poor performance of CPs. Further simulation results for σ^2=1 are summarized in Tables 2.3-2.4 for n=200 and 500, respectively. Findings observed from Tables 2.3-2.4 are similar to those from Tables 2.1-2.2. This indicates that the proposed method performs well regardless values of σ^2.

To examine the performance of spline estimated nonparametric functions, we present the mean estimated baseline survival functions, $\phi(\cdot)$ and $g(\cdot)$ together with their 95% credit intervals in Figure 2.1. The 95% credit intervals for cumulative baseline hazards and $\phi(\cdot)$ include their estimated and true curves, while 95% credit interval for single-index function $g(u)$ covers most part of its true curve except the case when u is too big.

TABLE 2.1: Simulation results for $\sigma^2 = 0.5$ with sample size $n = 200$ based on 200 replications. (AVE.bias for the average of differences between the estimates and the true value, AVE.SD for the average of standard deviations, and CP for the 95% coverage probability.)

		Proposed model			Usual PHMixCure model		
	True value	AVE.bias	AVE.SD	CP	AVE.bias	AVE.SD	CP
	Censoring rate(30-35 %)						
α	0.5	0.025	0.420	0.900	0.128	0.209	0.880
β	0.5	0.014	0.236	0.950	0.045	0.131	0.880
ξ_1	0.707	0.065	0.138	0.950	-0.500	0.316	0.533
ξ_2	0.707	0.044	0.149	0.950	-0.521	0.100	0.000
V	-	-	-	-	-0.200	0.207	0.749
	Censoring rate(60-65 %)						
α	0.5	0.033	0.342	0.900	0.134	0.390	0.914
β	0.5	0.018	0.167	0.950	0.077	0.278	0.886
ξ_1	0.707	0.190	0.146	0.950	-0.558	0.421	0.657
ξ_2	0.707	0.018	0.150	0.950	-0.525	0.109	0.007
V	-	-	-	-	-0.163	0.245	0.833

TABLE 2.2: Simulation results for $\sigma^2 = 0.5$ with sample size $n = 500$ based on 200 replications. (AVE.bias for the average of differences between the estimates and the true value, AVE.SD for the average of standard deviations, and CP for the 95% coverage probability.)

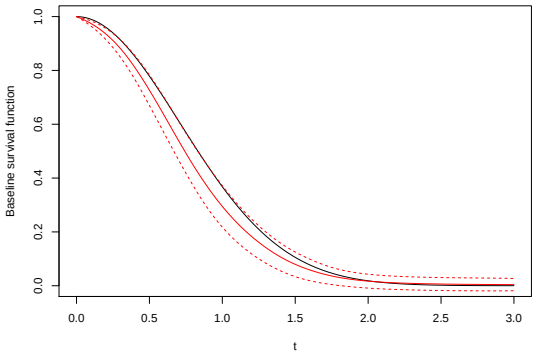
		Proposed model			Usual PHMixCure model		
	True value	AVE.bias	AVE.SD	CP	AVE.bias	AVE.SD	CP
	Censoring rate(30-35 %)						
α	0.5	0.006	0.337	0.920	-0.089	0.176	0.864
β	0.5	0.013	0.162	0.960	0.020	0.147	0.925
ξ_1	0.707	-0.072	0.146	0.970	-0.449	0.330	0.565
ξ_2	0.707	-0.039	0.155	0.970	-0.517	0.093	0.000
V	-	-	-	-	-0.199	0.246	0.782
	Censoring rate (60-65 %)						
α	0.5	0.003	0.192	0.945	-0.077	0.204	0876
β	0.5	-0.026	0.124	0.955	0.027	0.116	0.888
ξ_1	0.707	-0.111	0.147	0.955	-0.513	0.318	0.512
ξ_2	0.707	0.016	0.136	0.955	-0.504	0.318	0.060
V	-	-	-	-	-0.214	0.230	0.735

TABLE 2.3: Simulation results for $\sigma^2 = 1$ with sample size $n = 200$ based on 200 replications. (AVE.bias for the average of differences between the estimates and the true value, AVE.SD for the average of standard deviations, and CP for the 95% coverage probability.)

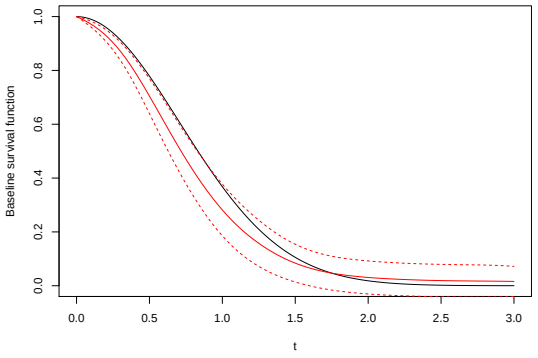
		Proposed model			Usual PHMixCure model		
	True value	AVE.bias	AVE.SD	CP	AVE.bias	AVE.SD	CP
	Censoring rate(30-35 %)						
α	0.5	0.057	0.307	0.835	-0.167	0.497	0.963
β	0.5	0.016	0.165	0.910	0.056	0.179	0.908
ξ_1	0.707	0.068	0.140	0.950	-0.440	0.571	0.780
ξ_2	0.707	0.049	0.153	0.950	-0.500	0.162	0.055
V	-	-	-	-	-0.140	0.319	0.862
	Censoring rate (60-65 %)						
α	0.5	0.020	0.264	0.900	-0.144	0.393	0.938
β	0.5	0.020	0.181	0.935	0.076	0.255	0.904
ξ_1	0.707	0.057	0.149	0.950	-0.498	0.564	0.783
ξ_2	0.707	0.058	0.166	0.950	-0.516	0.174	0.072
V	-	-	-	-	-0.151	0.440	0.914

TABLE 2.4: Simulation results for $\sigma^2 = 1$ with sample size $n = 500$ based on 200 replications. (AVE.bias for the average of differences between the estimates and the true value, AVE.SD for the average of standard deviations, and CP for the 95% coverage probability.)

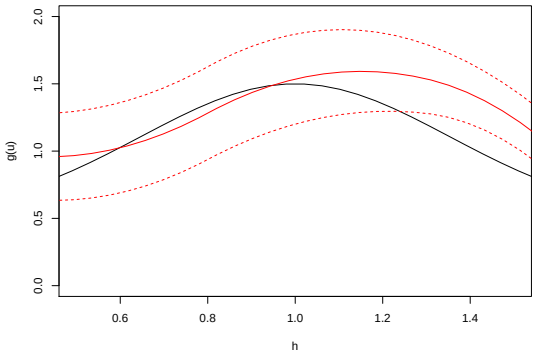
		Proposed model			Usual PHMixCure model		
	True value	AVE.bias	AVE.SD	CP	AVE.bias	AVE.SD	CP
	Censoring rate(30-35 %)						
α	0.5	-0.017	0.264	0.960	-0.114	0.255	0.960
β	0.5	-0.015	0.128	0.955	0.029	0.158	0.913
ξ_1	0.707	0.089	0.153	0.980	-0.538	0.328	0.506
ξ_2	0.707	-0.009	0.140	0.980	-0.525	0.109	0.007
V	-	-	-	-	-0.163	0.245	0.833
	Censoring rate (60-65 %)						
α	0.5	-0.033	0.203	0.945	-0.128	0.209	0.880
β	0.5	-0.016	0.116	0.910	0.045	0.131	0.880
ξ_1	0.707	-0.092	0.152	0.970	-0.499	0.316	0.000
ξ_2	0.707	0.007	0.151	0.970	-0.521	0.207	0.749
V	-	-	-	-	-0.195	0.207	0.749



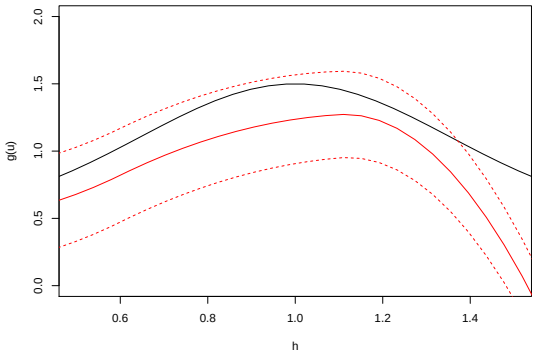
Baseline survival function



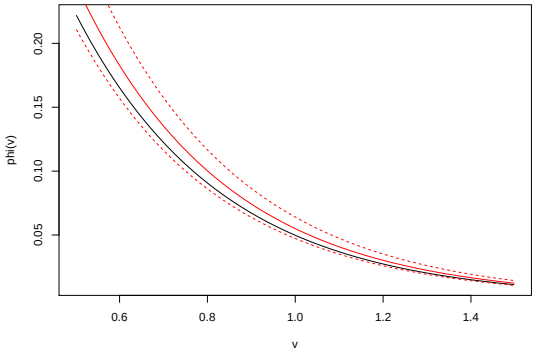
Baseline survival function



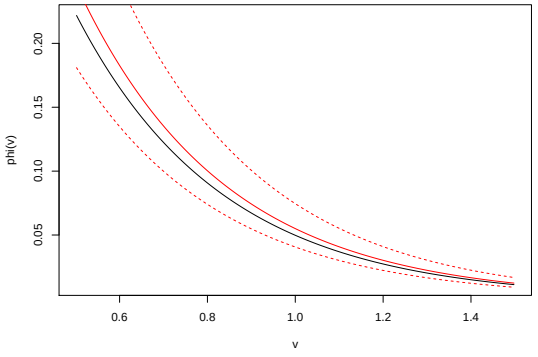
$g(u)$



$g(u)$



$\phi(v)$



$\phi(v)$

FIGURE 2.1: Estimated baseline survival functions and unknown functions $g(u)$ and $\phi(v)$ when $\sigma = 0.7$ and $\beta = 0.5$ with censoring rate of 30-35% in the left panel and 60-65% in the right panel. In each plot, the black solid curve represents the true curve, the red solid curve represents the mean estimated function, and the dotted curves are corresponding 95% credible intervals.

2.4 Real Data Application

For illustration purpose, we apply the proposed method to the analysis of the data from hypobaric decompression sickness study (Conkin and Powell, 2001). A cohort of 549 volunteers who were exposed to the hypobaric environment with age ranging from 20 to 54 was followed during the period of 1983 to 1998. Early analysis had been done by Thompson and Chhikara (2003) and Li and Ma (2010) based on mixture cure models with repeated measures because some patients might experience HDS more than once in the study.

We consider the time to the first experience of HDS for all patients in this section. Out of all 549 observations, 124 observations were interval-censored and the rest 425 observations were right-censored. The NPMLE of survival functions for two groups are given in Figure 2.2 without considering the covariates. For the individuals with ambulatory (NOADYN=1), the survival curve reaches to the bottom at the point 4.75 hr and levels off after that. As for the individuals with lower body dynamic (NOADYN=0), the survival curve remains a constant after the last jump point at 3.8 hr. The long tail of the survival function provides evidence of the existence of individuals who were unsusceptible to HDS. Thus the mixture cure model is reasonable to be considered for analyzing these data. For each subject, several covariates were recorded in the study, such as AGE ranging from 20 to 54 years, an experiment variable TR360 measuring the decompression stress, body dynamic level NOADYN (1=ambulatory, 0=lower body dynamic), and SEX (1=male, 0=female). For convenient computation, continuous covariates AGE and TR360 are all standardised.

We apply the proposed method to assess the effects of these potential covariates on both the risk of HDS and the probability of being disease free. We fit the proposed model with all covariates in both mixture components. Due to its time-varying feature, AGE is included as V in the nonlinear term of the latency

component, while the rest covariates are used as X in the linear term of the model. For the incidence component, two continuous covariates AGE and TR360 are considered in the single-index part and the discrete covariates are put in the linear term following Lei (2013). Table 2.5 reports the Bayesian estimates, the average

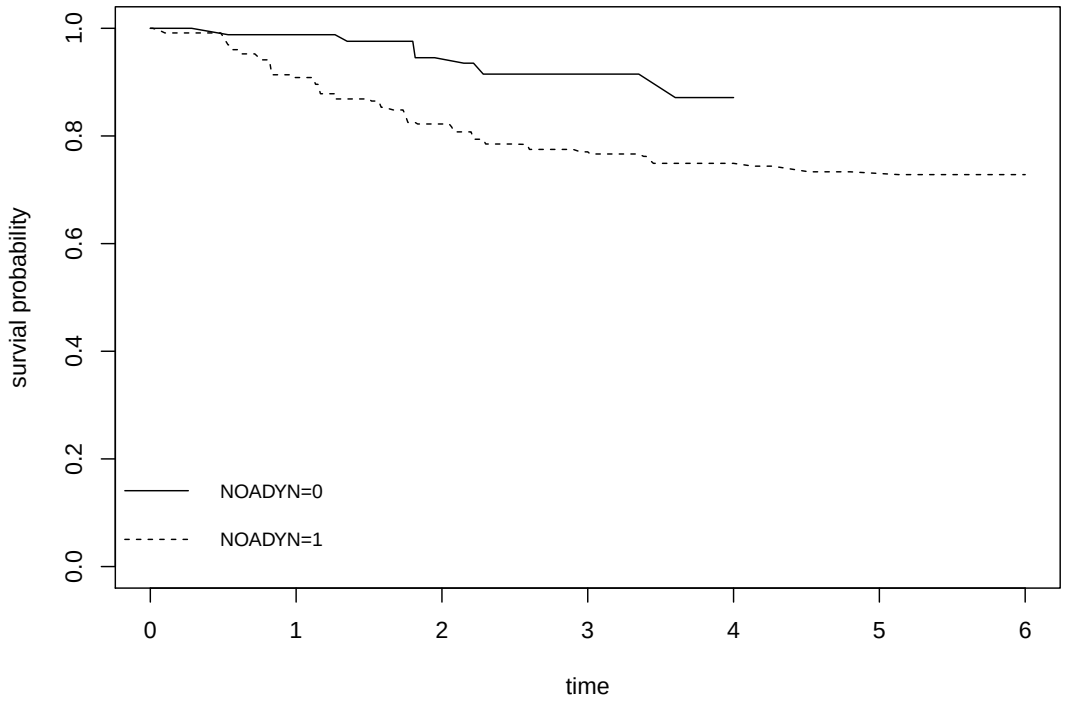


FIGURE 2.2: NPMLE estimated survival curves for two groups of individuals with ambulatory(NOADYN=1) and lower body dynamic(NOADYN=0).

standard deviation and 95% HPD interval of the model parameters. From the incidence model for uncured probability, we observe that male subjects are more likely to experience HDS. This finding is consistent with that provided by Thompson and Chhikara (2003) and Li and Ma (2010). Ambulatory subjects (i.e., NOADYN=1) have less uncured probabilities or high chance to be HDS free than those patients with low body dynamic. Based on the results in the latency part, ambulatory patients tend to have up to 2.7 times higher risk of HDS if they are susceptible. For the uncured subgroup, it is found that there is not a significant influence of TR360

adjusted by other covariates. According to the estimated single-index and non-parametric terms in Figure 2.3, there is strong evidence to support the nonlinear effects of TR360 and AGE.

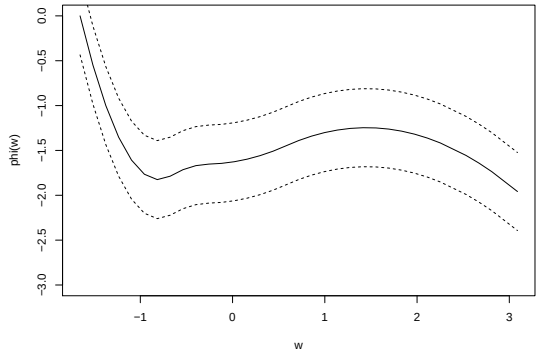
According to the plot of $\phi(w)$, there is a nonlinear effect of AGE. The effects of AGE on the cure part depend on the range of age, and younger subjects tend to have higher uncured probability than the old ones. The estimated function $g(.)$ and its 95% credible interval are presented in Figure 2.3, and it is clearly that $g(.)$ is nonlinear and has a local minimum around -1.5 and maximum around 1.7.

TABLE 2.5: Estimated parameters in the proposed model for the HDS data

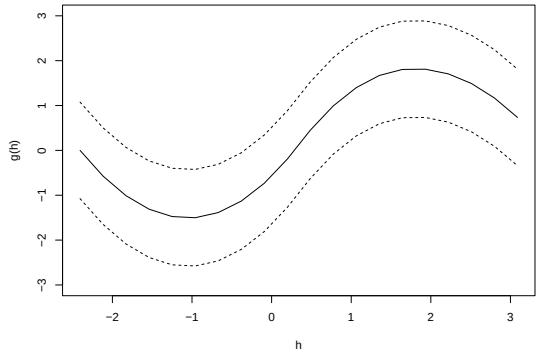
Parameter	Estimate	Ave.SD	95% HPD interval
Survival part			
SEX	0.216	0.269	(-0.314 ,0.749)
NOADYN	1.026	0.283	(0.481, 1.612)
TR360	-0.013	0.140	(-0.281, 0.272)
Cure part			
Intercept	0.558	0.631	(-0.558, 1.889)
NOADYN	-1.152	0.568	(-2.307, -0.188)
SEX	0.330	0.337	(-0.348, 0.963)
ξ_1(TR360)	0.967	0.020	(0.927, 0.994)
ξ_2(AGE)	0.241	0.079	(0.108, 0.374)
σ	0.234	0.017	(0.177, 0.313)

2.5 Discussion

In this chapter, we propose a doubly semiparametric mixture cure survival model with interval-censored data, which provides a more flexible model by allowing the nonlinear effects in both mixture components, and the single-index model employed in the incidence part is efficient to avoid the curse of dimensionality possibly in high-dimensional covariates. Moreover, the proposed Bayesian estimators are efficient and computational convenience since the explicit forms of some parameters can be derived.



(A) Nonlinear part in survival function $\phi(w)$



(B) Nonlinear part in cure rate $g(h)$

FIGURE 2.3: Estimated curve of unknown function (solid curve) and its pointwise 95% credible intervals (dotted curves) in the latency part in (A) and the incidence part in (B).

Extending the proposed model to tackle the clustered interval-censored data by introducing multiple frailties to the baseline hazard function, is possible. Pan et al. (2017) investigated the PH-based multiple frailties for bivariate predictors. Considering the frailties for bivariate predictors, we only need to modified the proposed sampling algorithm, and update the frailty terms by Gibbs sampling. The extension of the proposed model to the time-dependent covariates is straightforward by assuming spline functions to approximate the time-dependent regression coefficients. In addition, Zhang et al. (2019) studied the Cox model with time-dependent covariates for clustered interval-censored data, by assuming a piecewise constant function both for the baseline hazard function and the time-dependent

coefficients. It will be more convenient to employ splines approach for the approximation of unknown functions. To extend our proposed model to time-dependent covariates with clustered interval-censored data is a challenging but valuable work, which is worth further investigations.

Chapter 3

Generalized Accelerated Hazards Cure Models with Interval-censored Data

Most of the existing mixture cure models with interval-censored data assume a conventional survival model, such as the PH, PO, or AFT model, for the susceptible subjects. In some practical cases, none of these conventional assumptions is appropriate for modeling data. For example, the treatment in a clinical trial may take a time lag to have effects on the patients rather than having effects immediately after application. In other words, the hazards of the treatment and control groups are identical at the beginning of the study. As time goes by, the treatment effects on the failure time distribution of susceptible patients can increase. To account for such a feature, the AH model introduced in Chapter 1 is particular appealing.

There are limited studies in the literature to use the AH model for the failure times of uncured subjects. Zhang and Peng (2009) proposed a two-component mixture cure model with the AH model specified in the latency part, and developed an EM algorithm for implementing the maximum likelihood estimation where the M-step was achieved by a rank-based estimating equation. Their work focused

on right censored data and cannot be applied directly to interval-censored data due to the intractable structure of likelihood and inclusion of bundled regression parameters. For interval-censored data without the cure subgroup, Szabo et al. (2019) studied the AH model and proposed a sieve maximum likelihood method for estimation by assuming a spline approximation for the unknown cumulative baseline hazard function. Their proposed sieve ML estimators were shown to be consistent and asymptotically normal.

In this chapter, a generalized AH cure model is proposed for interval-censored data in the presence of a cure fraction. The proposed model is able to capture the time-lag feature of data and allows the PH cure, AFT cure, and AH cure models as its special cases. Thus it provides a unified model framework for the currently available mixture cure models with interval-censored data. We develop a sieve ML estimation procedure, where the B-spline functions are used for approximating the cumulative baseline hazard function. To implement the estimation procedure, a two-step iterative algorithm is developed for the maximization problem with linear inequality constraints. In addition, the existence of the bundled parameters makes it challenging to justify the theoretical properties of the proposed estimator. Using the modern empirical process techniques, asymptotic results of the proposed estimator, such as the consistency, convergence rate and asymptotic normality, are rigorously proved.

The rest of this chapter is arranged as follows. Section 3.1 introduces the generalized accelerated hazards cure model for interval-censored data. The proposed sieve ML estimation procedure, assuming B-spline functions for the cumulative baseline hazard function, is given in Section 3.2. Section 3.3 describes the theoretical properties of the sieve ML estimator in the generalized AH cure model. Simulation studies and real data analysis are presented in Section 3.4 and Section 3.5, respectively. In Section 3.6, a conclusion is drawn regarding the proposed generalized AH cure model. All technical proofs are given in Section 3.7.

3.1 The Generalized Hazards Cure Model with Interval-censored Data

For subject i, we assume that the interested failure time T_i is recorded as a time interval $(L_i, R_i]$ without an exact observation. Define $\delta_{L_i} = I(L_i = 0)$, $\delta_{I_i} = I(0 < L_i < T_i \leq R_i < \infty)$ and $\delta_{R_i} = I(R_i = \infty)$ as the censoring indicators, with value of 1 representing the failure time of the subject is left-, interval- and right-censored, and 0 otherwise. Taking the cure fraction into account, the binary latent random variable Y_i is introduced as a cure status of subject i, where $Y_i = 1$ represents the subject is uncured (susceptible), and 0 otherwise. To assess the cure fraction, the mixture cure model assumes that the population survival function of subject i is in the form of:

$$S_p(t|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i) = \pi(\boldsymbol{W}_i)S(t|\boldsymbol{X}_i, \boldsymbol{Z}_i) + 1 - \pi(\boldsymbol{W}_i), \tag{3.1}$$

where $\pi(\boldsymbol{W}_i) = P(Y_i = 1|\boldsymbol{W}_i)$ is the probability of being uncured and $S(t|\boldsymbol{X}_i, \boldsymbol{Z}_i) = P(T > t|Y_i = 1, \boldsymbol{X}_i, \boldsymbol{Z}_i)$ is the conditional survival function given subject i is uncured. $\boldsymbol{X}_i, \boldsymbol{Z}_i$ and $\boldsymbol{W}_i$ are covariate vectors that may share some common components in practice, with dimensions p_1, p_2 and p_3, respectively.

In view of the binary nature of Y, a simple and popular model for the uncure probability is logistic regression, which connects $\pi(\boldsymbol{W}_i)$ to covariates $\boldsymbol{W}_i$ through a link function:

$$\pi(\boldsymbol{W}_i) = \frac{\exp(\boldsymbol{W}_i^T\boldsymbol{\alpha})}{1 + \exp(\boldsymbol{W}_i^T\boldsymbol{\alpha})}, \tag{3.2}$$

where $\boldsymbol{\alpha}$ is the vector of coefficients and $\boldsymbol{W}_i$ includes 1 for the intercept. Under the generalized AH model (1.8), the conditional survival function is given by:

$$S(t|\boldsymbol{X}_i, \boldsymbol{Z}_i) = \exp\{-\Lambda_0(te^{\boldsymbol{X}_i^T\boldsymbol{\beta}})e^{\boldsymbol{Z}_i^T\boldsymbol{\gamma}}\}, \tag{3.3}$$

where $\Lambda_0(\cdot)$ is the unspecified cumulative baseline hazard function, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are corresponding regression coefficients to be estimated. The mixture cure model with the generalized AH model to assess the survival rate of uncured subjects is referred to as the generalized AH cure (GAHCure) model. The GAHCure model includes the PH cure, AFT cure, and AH cure models as special cases when $\boldsymbol{\beta} = 0$, $\boldsymbol{\gamma} = 0$, and both $\boldsymbol{\gamma} + \boldsymbol{\beta} = 0$ and $X = Z$, respectively.

Based on the observed data $\mathcal{O} = \{(L_i, R_i, \delta_{L_i}, \delta_{I_i}, \delta_{R_i}, \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i), i = 1, 2, \ldots n\}$, the log-likelihood can be written as:

$$
\begin{aligned}
l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda_0 | \mathcal{O}) = \sum_{i=1}^{n} \Bigg(& \delta_{L_i} \log \left[1 - \exp\{-\Lambda_0(R_i e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}) e^{\boldsymbol{Z}_i{}^T \boldsymbol{\gamma}} \} \right] + \delta_{L_i} \log \pi(\boldsymbol{W}_i) \\
& + \delta_{I_i} \log \left[\exp\{-\Lambda_0(L_i e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}) e^{\boldsymbol{Z}_i{}^T \boldsymbol{\gamma}} \} - \exp\{-\Lambda_0(R_i e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}) e^{\boldsymbol{Z}_i^T \boldsymbol{\gamma}} \} \right] \\
& + \delta_{I_i} \log \pi(\boldsymbol{W}_i) + \delta_{R_i} \log \left[1 - \pi(\boldsymbol{W}_i) + \pi(\boldsymbol{W}_i) \exp\{-\Lambda_0(L_i e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}) e^{\boldsymbol{Z}_i{}^T \boldsymbol{\gamma}} \} \right] \Bigg).
\end{aligned}
\tag{3.4}
$$

Estimates of the finite parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ and the unknown function $\Lambda_0(\cdot)$ are obtained directly by maximizing the log-likelihood (3.4), where $\Lambda_0(\cdot)$ is commonly assumed as a non-decreasing step function. The maximization of the log-likelihood depends on the values of $\Lambda_0(\cdot)$ at distinct points of $L_i e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}$ and $R_i e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}, i = 1, 2, \ldots n$. As sample size n increases, the number of such distinct points will substantially increase to $2n$ in the log-likelihood function in (3.4), leading to a high-dimensional optimization problem and intractable computational challenges in the maximum likelihood estimation. Unlike those conventional survival models in which the unspecified function $\Lambda_0(\cdot)$ and regression coefficients $\boldsymbol{\beta}$ are well separated, $\boldsymbol{\beta}$ is bundled into $\Lambda_0(\cdot)$ in the proposed GAHCure model, making the implementation of inference even more challenging. To deal with possible infinite-dimensional parameters to be estimated and obtain efficient estimators, a sieve ML estimation is proposed, where the unknown function $\Lambda_0(\cdot)$ is modeled by a linear combination of B-spline functions.

3.2 Semiparametric Inference Based on Sieve ML Estimation

3.2.1 Sieve maximum likelihood estimator

To facilitate the sieve ML estimation, we employ B-splines to approximate the unknown function $\Lambda_0(\cdot)$, and treat coefficients of the basis functions as parameters to be estimated together with regression coefficients in the proposed model.

The spline-based sieve approach has been investigated in nonparametric and semiparametric models in recent studies (Zhang et al., 2010; Ma et al., 2015; Zhou et al., 2017; Szabo et al., 2019), in which spline functions are constructed over the interval $[\min_{1\leq i\leq n}\{L_i, R_i\},\ \max_{1\leq i\leq n}\{L_i, R_i I(R_i < \infty)\}]$. Because $\boldsymbol{\beta}$ is bundled into $\Lambda_0(\cdot)$ in the GAHCure model, it requires the splines to be constructed over the interval $[\min_{1\leq i\leq n}\{L_i e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}, R_i e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\},\ \max_{1\leq i\leq n}\{L_i e^{\boldsymbol{X}_i^T\boldsymbol{\beta}},\ R_i e^{\boldsymbol{X}_i^T\boldsymbol{\beta}} I(R_i < \infty)\}]$. Note that the end points of such intervals vary with the different values of $\boldsymbol{\beta}$ during the iteration of estimation, yielding more challenges in the estimation procedure. The spline-based sieve approach with bundled parameters has been investigated by Zhao et al. (2017) for right censored data, using cubic B-spline functions to approximate the log baseline hazard function. In the following section, this approach is extended to the interval-censored data setting to overcome the difficulties caused by the bundled parameters and infinite-dimensional unknown functions included in the GAHCure model.

The B-spline functions are constructed based on the stochastic points. Specifically, for a fixed $\boldsymbol{\beta}$, the infinite-dimensional function $\Lambda_0(\cdot)$ is approximated by B-spline functions on the interval $[a^{\boldsymbol{\beta}}, b^{\boldsymbol{\beta}}]$, where $a^{\boldsymbol{\beta}} = \min_{1\leq i\leq n}\{L_i e^{\boldsymbol{X}_i{}^T\boldsymbol{\beta}}\}$ and $b^{\boldsymbol{\beta}} = \max_{1\leq i\leq n}\{R_i e^{\boldsymbol{X}_i{}^T\boldsymbol{\beta}} I(R_i < \infty)\}$. Let $a^{\boldsymbol{\beta}} = d_0 < d_1 \cdots < d_{k_n} < d_{k_n+1} = b^{\boldsymbol{\beta}}$ be a partition on the interval $[a^{\boldsymbol{\beta}}, b^{\boldsymbol{\beta}}]$, which divides the interval into $k_n + 1$ identical subintervals. k_n is a positive integer with $k_n = O(n^v)$ for $0 < v < 0.5$ and

$\max_{1\leq j\leq k_n+1}|d_j - d_{j-1}| = O(n^{-v})$. Define by $D_{k_n}^{\boldsymbol{\beta}}$ the set of all partition points. Due to the non-negative and non-decreasing constraints of the cumulative baseline hazard function, a non-decreasing constraint is imposed on the spline coefficients. Let $\mathcal{S}_n^{\boldsymbol{\beta}}(D_{k_n}^{\boldsymbol{\beta}}, m)$ be the spline space for given $\boldsymbol{\beta}$, defined by:

$$\mathcal{S}_n^{\boldsymbol{\beta}}(D_{k_n}^{\boldsymbol{\beta}}, m) = \{\phi_n : \phi_n(t) = \boldsymbol{c}^T\boldsymbol{B}^{\boldsymbol{\beta}} = \sum_{j=1}^{q_n} c_j B_j^{\boldsymbol{\beta}}(t), \boldsymbol{c}\in\mathcal{C}_{q_n}\}, \tag{3.5}$$

where $q_n = m + k_n$ is the number of spline functions, $\boldsymbol{B}^{\boldsymbol{\beta}} = (B_1^{\boldsymbol{\beta}}, B_2^{\boldsymbol{\beta}}, \ldots, B_{q_n}^{\boldsymbol{\beta}})^T$ is the vector of the basis splines and spline coefficients $\mathcal{C}_{q_n} = \{\boldsymbol{c} = (c_1, \ldots, c_{q_n})^T, 0 \leq c_1 \leq c_2 \leq \ldots, \leq c_{q_n}\}$. The spline functions are all required to be greater than or equal to 0, and the non-decreasing constraint on spline coefficients ensures that every function in $\mathcal{S}_n^{\boldsymbol{\beta}}$ is non-decreasing. Thus, the cumulative baseline hazard function $\Lambda_0(\,\cdot\,)$ can be approximated by some functions in $\mathcal{S}_n^{\boldsymbol{\beta}}(D_{k_n}^{\boldsymbol{\beta}}, m)$. Substituting the function incorporating the spline approximation in the log-likelihood (3.4) affords:

$$\begin{aligned} l(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{c}|\mathcal{O}) = \sum_{i=1}^{n}\Bigg(&\delta_{L_i}\log\left[1-\exp\{-\boldsymbol{c}^T\boldsymbol{B}(R_i e^{\boldsymbol{X}_i^T\boldsymbol{\beta}})e^{\boldsymbol{Z}_i^T\boldsymbol{\gamma}}\}\right] + \delta_{L_i}\log\pi(\boldsymbol{W}_i) \\ &+ \delta_{I_i}\log\left[\exp\{-\boldsymbol{c}^T\boldsymbol{B}(L_i e^{\boldsymbol{X}_i^T\boldsymbol{\beta}})e^{\boldsymbol{Z}_i^T\boldsymbol{\gamma}}\} - \exp\{-\boldsymbol{c}^T\boldsymbol{B}(R_i e^{\boldsymbol{X}_i^T\boldsymbol{\beta}})e^{\boldsymbol{Z}_i^T\boldsymbol{\gamma}}\}\right] \\ &+ \delta_{I_i}\log\pi(\boldsymbol{W}_i) + \delta_{R_i}\log\left[1-\pi(\boldsymbol{W}_i)+\pi(\boldsymbol{W}_i)\exp\{-\boldsymbol{c}^T\boldsymbol{B}(L_i e^{\boldsymbol{X}_i^T\boldsymbol{\beta}})e^{\boldsymbol{Z}_i^T\boldsymbol{\gamma}}\}\right]\Bigg). \end{aligned} \tag{3.6}$$

Denote the parameter space of $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, \Lambda_0)^T$ by $\Theta_n = \mathcal{A}\times\mathcal{B}\times\Gamma\times\mathcal{S}_n$, where $\mathcal{A}, \mathcal{B}$ and Γ are the compact parameter spaces of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and $\mathcal{S}_n$ is the sieve space of Λ_0 with $\mathcal{S}_n = \cup_{\boldsymbol{\beta}\in\mathcal{B}}\mathcal{S}_n^{\boldsymbol{\beta}}$. The sieve ML estimator $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\alpha}}_n^T, \hat{\boldsymbol{\beta}}_n^T, \hat{\boldsymbol{\gamma}}_n^T, \hat{\Lambda}_n)^T$ is obtained by maximizing the log-likelihood $l(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{c}|\mathcal{O})$ over the space $\mathcal{A}\times\mathcal{B}\times\Gamma\times\mathcal{C}_{q_n}$ or equivalently maximizing the log-likelihood $l(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\Lambda_0|\mathcal{O})$ over the parameter space Θ_n.

3.2.2 Computational issues

As previously mentioned, the sieve space of $\Lambda_0(\cdot)$ is indexed by $\boldsymbol{\beta}$ in the GAHCure model, which is different from that in the conventional PH model. The sieve space indexed by $\boldsymbol{\beta}$ is referred to as $\mathcal{S}_n^{\boldsymbol{\beta}}$. With an estimate $\boldsymbol{\beta}'$ of $\boldsymbol{\beta}$, the corresponding sieve space is updated as $\mathcal{S}_n^{\boldsymbol{\beta}'}$. To acquire the sieve ML estimator, the following simulation steps are developed:

Step 1. Choose suitable initial values of $(\hat{\boldsymbol{\alpha}}^{(0)}, \hat{\boldsymbol{\beta}}^{(0)}, \hat{\boldsymbol{\gamma}}^{(0)}, \hat{\boldsymbol{c}}^{(0)})$ and set $k = 0$. In this study, we set all initial values to be 0.

Step 2. At the $(k+1)$th iteration, obtain the estimates $(\hat{\boldsymbol{\alpha}}^{(k+1)}, \hat{\boldsymbol{\beta}}^{(k+1)}, \hat{\boldsymbol{\gamma}}^{(k+1)})$ by maximizing $l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}|\hat{\boldsymbol{c}}^{(k)})$ with respect to $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

Step 3. Obtain the estimate $\hat{\boldsymbol{c}}^{(k+1)}$ by maximizing $l(\boldsymbol{c}|\hat{\boldsymbol{\alpha}}^{(k+1)}, \hat{\boldsymbol{\beta}}^{(k+1)}, \hat{\boldsymbol{\gamma}}^{(k+1)})$ over the constraint parameter space $\mathcal{C}_{q_n}$. Update $\hat{\Lambda}_0(t)^{(k+1)} = \sum_j^{q_n} \hat{c}_j^{(k+1)} B_j^{\hat{\boldsymbol{\beta}}^{(k+1)}}(t)$.

Step 4. Repeat Steps 2 and 3 until the absolute difference of the log-likelihoods between two consecutive iterations is less than 10^{-3}.

The resultant estimator of **Step** 4 is taken as the sieve ML estimator $\hat{\boldsymbol{\theta}}_n$. The numerical optimization of the log-likelihood in the above estimation procedure is implemented by the constrained optimization function **donlp2** in the R package **Rsolnp2**.

Another critical aspect of the implementation is the degree of basis functions. The high order of spline functions offers better smoothness, but may lead to an over-fitting problem. To balance the smoothing and computational burden, quadratic or cubic splines are used to supply the approximation of the cumulative baseline hazard function $\Lambda_0(\cdot)$. Apart from the smoothing condition, the number and location of knots in constructing basis splines should be considered to ensure the accuracy of the approximation. In the proposed method, quantile-based knots are used, and the number of knots is determined by the Bayesian information criterion

(BIC). For the sieve ML estimator $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\alpha}}_n^T, \hat{\boldsymbol{\beta}}_n^T, \hat{\boldsymbol{\gamma}}_n^T, \hat{\Lambda}_n)^T$, the q_n is determined by minimizing the BIC:

$$BIC = -2l(\hat{\boldsymbol{\theta}}_n) + (2q_n)\log n. \tag{3.7}$$

According to the simulation study, it is recommended that quadratic or cubic functions are used to approximate $\Lambda_0(\cdot)$, and the number of knots can range from 3 to 6.

3.3 Asymptotic Properties

For notation simplicity, let $\phi = \log \Lambda_0$. Then the log-likelihood can be written as:

$$\begin{aligned} l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \phi | \mathcal{O}) = \sum_{i=1}^{n} \Bigg(& \delta_{L_i} \log \left[1 - \exp\{-e^{\phi(R_i e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}) + \boldsymbol{Z}_i^T \boldsymbol{\gamma}}\} \right] \\ & + \delta_{I_i} \log \left[\exp\{-e^{\phi(L_i e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}) + \boldsymbol{Z}_i^T \boldsymbol{\gamma}}\} - \exp\{-e^{\phi(R_i e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}) + \boldsymbol{Z}_i^T \boldsymbol{\gamma}}\} \right] \\ & + \delta_{R_i} \log \left[1 - \pi(\boldsymbol{W}_i) + \pi(\boldsymbol{W}_i) \exp\{-e^{\phi(L_i e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}) + \boldsymbol{Z}_i^T \boldsymbol{\gamma}}\} \right] \\ & + (1 - \delta_{R_i}) \log \pi(\boldsymbol{W}_i) \Bigg). \end{aligned} \tag{3.8}$$

Denote the proposed estimator by $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\tau}}_n^T, \hat{\phi}_n)^T$ with $\hat{\tau}_n = (\hat{\boldsymbol{\alpha}}_n^T, \hat{\boldsymbol{\beta}}_n^T, \hat{\boldsymbol{\gamma}}_n^T)^T$, and the corresponding truth by $\boldsymbol{\theta}_0 = (\boldsymbol{\tau}_0^T, \phi_0)^T$ with $\boldsymbol{\tau}_0 = (\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T, \boldsymbol{\gamma}_0^T)^T$. Some additional notations are needed. Let $\boldsymbol{\theta}_1 = (\boldsymbol{\tau}_1^T, \phi_1)^T$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\tau}_2^T, \phi_2)^T$ be two sets of parameters and function to be estimated, and the distance between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ is defined as follows:

$$d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (\| \boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2 \|^2 + \| \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 \|^2 + \| \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2 \|^2 + \| \phi_1 - \phi_2 \|^2_{\Phi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)})^{1/2}, \tag{3.9}$$

where

$$\begin{aligned} \| \phi_1 - \phi_2 \|^2_{\Phi(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)} &= E[(\phi_1(L\exp(\boldsymbol{X}^T\boldsymbol{\beta}_1)) - \phi_2(L\exp(\boldsymbol{X}^T\boldsymbol{\beta}_2)))^2] \\ &+ E[(\phi_1(R\exp(\boldsymbol{X}^T\boldsymbol{\beta}_1)) - \phi_2(R\exp(\boldsymbol{X}^T\boldsymbol{\beta}_2)))^2], \end{aligned} \tag{3.10}$$

and $\Phi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ is the distance between ϕ_1 and ϕ_2, indexed by $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively.

To establish the large sample properties, the following conditions are required to ensure the validity of the given theorems.

C1. The true parameters $\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0$ satisfy that $\boldsymbol{\alpha}_0 \in \mathcal{A}$, $\boldsymbol{\beta}_0 \in \mathcal{B}$, $\boldsymbol{\gamma}_0 \in \Gamma$ and all the parameter spaces $\mathcal{A}, \mathcal{B}, \Gamma$ are compact sets.

C2. The domain of covariates $(\boldsymbol{X}^T, \boldsymbol{Z}^T, \boldsymbol{W}^T)^T$ is a bounded subset of R^d, where $d = p_1 + p_2 + p_3$, and all $E(\boldsymbol{X}\boldsymbol{X}^T)$, $E(\boldsymbol{Z}\boldsymbol{Z}^T)$, $E(\boldsymbol{W}\boldsymbol{W}^T)$ are nonsingular.

C3. Let $\phi \in \Phi$, where Φ is a collection of functions with bounded pth derivatives for $p \geq 2$. The first derivative of ϕ exists and is positive and continuous.

C4. There is a positive integer ξ such that $P(R - L \geq \xi) = 1$, for the finite interval $(L, R]$.

C5. T and $(L, R]$ are conditionally independent given the covariates $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\boldsymbol{W}$.

C6. The joint density of $(T, L, R, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})$ has uniform positive lower and upper bounds in the support region of the joint random variables.

Conditions C1-C6 are common assumptions for the theoretical justification in survival analysis with interval-censored data. Condition C3 is used to guarantee the smoothness of the spline approximation, providing the foundation of employing the result of van der Vaart and Wellner (1997) in the proof of consistency.

Theorem 3.1. *(Identifiability) Assume that Λ_0 is a continuous function with bundled parameters defined in (1.8), then the model is unidentifiable if and only if $\Lambda_0(t) = c_1 t^{c_2}$ for some positive constants c_1 and c_2.*

Theorem 3.2. *(Consistency) Given that conditions C1-C4 hold, we have the estimator* $\hat{\boldsymbol{\theta}}_n$ *converges to* $\boldsymbol{\theta}_0$ *in probability.*

Theorem 3.3. *(Convergence of rate) Let* $k_n = O_p(n^v)$*, where* v *satisfies that* $1/(2p+2) < v < 1/2p$*. Suppose conditions C1-C6 hold, then*

$$d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p\{n^{-\min(pv,(1-v)/2)}\}.$$

Theorem 3.3 implies that if $v = 1/(2p+1)$, the optimal global convergence rate of the estimator can be achieved, which is $d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p\{n^{-p/(1+2p)}\}$. If the nonparametric term is involved, the convergence rate of the proposed estimator $\hat{\boldsymbol{\theta}}_n$ cannot reach $n^{-1/2}$, while the estimator of the parametric term $\hat{\boldsymbol{\tau}}_n$ converges at the rate of $n^{-1/2}$. This theorem can be proved by verifying the conditions of Theorem 1 in Shen and Wong (1994).

Theorem 3.4. *(Asymptotic normality) Suppose that conditions C1-C6 hold and* $1/(2p+2) < v < 1/2p$*, then*

$$n^{1/2}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0) \to N\{0, I^{-1}(\boldsymbol{\tau}_0)\}$$

in distribution.

The proof of Theorem 3.4 with bundled parameters is more complicated than that where the parametric and nonparametric terms are well separated from each other. Involving the bundled parameters in Λ_0, we prove Theorem 3.4 in the same line of that in Ding and Nan (2011). More details of the proof can be found in Section 3.7.

Theorem 3.4 shows that the estimator of $\boldsymbol{\tau}_0$ is semiparametric efficient, with the asymptotic variance matrix $I^{-1}(\boldsymbol{\tau}_0)$. However, as it involves an infinite-dimensional operator and has a complicated form, a consistent estimator is proposed for the variance of $\hat{\boldsymbol{\tau}}_n$, based on the weighted bootstrap method (Ma and Kosorok, 2005).

We draw n observations randomly from the sample data with replacement and the generated samples are used as the bootstrap samples, denoted by $\mathcal{O}_i, i = 1, \ldots n$. The weighted bootstrap estimator $(\tilde{\boldsymbol{\tau}}_n, \tilde{\phi}_n)$ is obtained as follows:

$$(\tilde{\boldsymbol{\tau}}_n, \tilde{\phi}_n) = \underset{(\boldsymbol{\tau}_n^T, \phi_n)^T \in \Theta_n}{\operatorname{argmax}} \sum_{i=1}^{n} w_{ni} l(\boldsymbol{\tau}, \phi, \mathcal{O}_i), \tag{3.11}$$

where the bootstrap weight $(w_{n1}, \ldots, w_{nn}) \sim \text{Multinomial}(n, (n^{-1}, \ldots n^{-1}))$, and $l(\boldsymbol{\tau}, \phi, \mathcal{O}_i)$ is the log-likelihood based on the bootstrap samples $\mathcal{O}_i$. The sample variance matrix of $\tilde{\boldsymbol{\tau}}_n$, calculated by repeating the estimate procedure for M times, is employed as the approximation of $I^{-1}(\boldsymbol{\tau}_0)$. In a similar way to that in Cheng et al. (2010), the consistency of the weighted bootstrap estimator can be proved by showing that $\sqrt{n}(\tilde{\boldsymbol{\tau}}_n - \hat{\boldsymbol{\tau}}_n)$ has the same limiting distribution with $\sqrt{n}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0)$.

3.4 Simulation Studies

Simulation studies are conducted to evaluate the finite sample performance of the proposed estimation procedure. The existing R packages **Splines** and **Rdonlp2** are used to generate the basis spline polynomials and maximize the log-likelihood, respectively.

We generate the interval-censored data with the cure fraction in the same line of the method in Xiang et al. (2011). First, for subject i, $i = 1, 2, \ldots n$, we generate the cure indicator Y_i from the Bernoulli distribution with $P(Y_i = 1) = \frac{\exp(\alpha_0 + W_i\alpha_1)}{1+\exp(\alpha_0+W_i\alpha_1)}$, where the covariate W_i is generated from the uniform distribution $U(0, 1)$. We generate the auxiliary random variable U_i from the exponential distribution with mean 0.02 and the gap of observed event times len_i from the uniform distribution on $[0, 0.02]$. Let $C = 3$ be the length of the study. For cured subjects with $Y_i = 0$, we set $L_i = C$ and $R_i = \infty$. For susceptible subjects with $Y_i = 1$, we generate the failure times T_i for susceptible subjects from the cumulative hazard

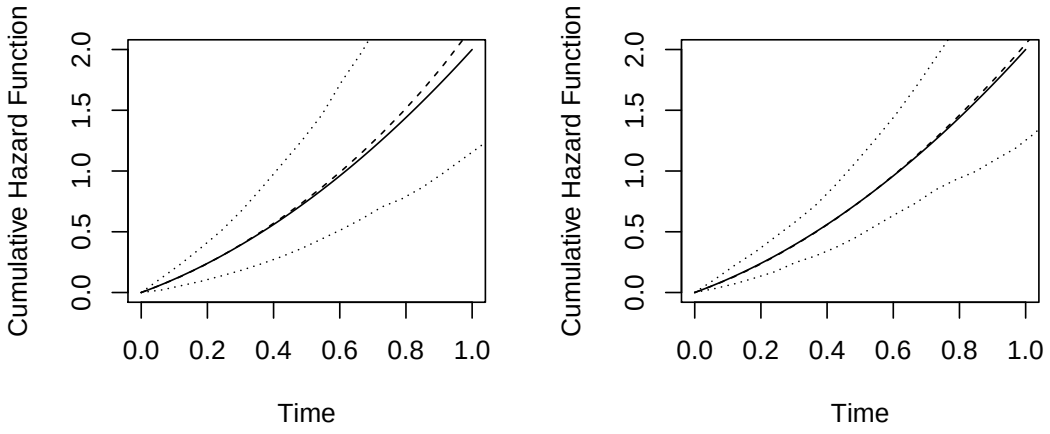
function $\Lambda(t) = \Lambda_0(te^{X_i\beta})e^{Z_i\gamma}$, where $\Lambda_0(t) = (t+0.5)^2 - 0.25$ and $X_i = Z_i = W_i$. If $T_i > C$, then the subject is right censored with $L_i = C$ and $R_i = \infty$. If $T_i < C$ and $T_i < U_i$, the subject is left censored with $L_i = 0$ and $R_i = U_i$. Otherwise, the subject is interval-censored, and we find the interval $[k \cdot len_i + U_i, (k+1) \cdot len_i + U_i]$ which contains T_i. We set $\beta = \gamma = 0.5$, $(\alpha_0, \alpha_1) = (-0.5, 0.3)$ and $(0.5, 0.3)$, yielding the average left censoring and right censoring of (2%, 60%) and (3.5%, 32%), respectively.

In the implementation of the estimation procedure, the degree of splines is set to be 2, corresponding to the quadratic polynomial splines. The quantile-based knots with the number of interior points ranging from 4 to 6 are utilized, and the BIC is used to determine a proper spline setting.

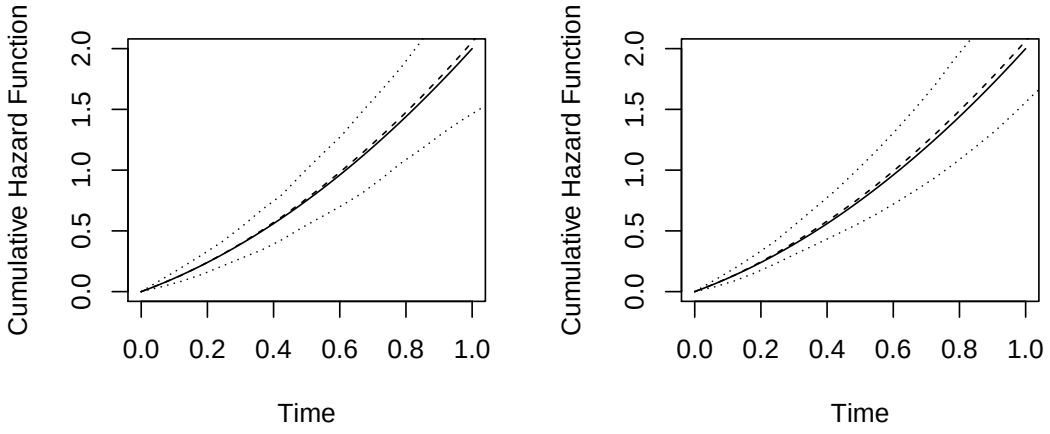
Five hundred replications are generated with the sample size $n = 200$ and $n = 500$. Simulation results are summarized in Table 3.1, including the average bias (AVE.bias), average standard derivation (AVE.SD), and coverage probability (CP) of 95% confidence intervals. The biases of the proposed estimates for most parameters are small, and CPs are close to the nominal level of 0.95. When the sample size increases, the biases and the average standard derivations decrease, and the biases decrease as the censoring level decreases. The estimated cumulative baseline hazard functions for different censoring rates and sample sizes are presented in Figure 3.1, which shows that they go closer to the true cumulative baseline hazard curve as sample size increases, regardless of the censoring rate.

3.5 Real Data Application

As an illustration, we apply the proposed method to interval-censored smoking cessation data. The data have been analyzed by Xiang et al. (2011), where a mixture cure model was fitted using maximum likelihood estimation implemented



n=200



n=500

FIGURE 3.1: Estimated cumulative baseline hazard functions with censoring level 60-65%(left panel) and 30-35%(right panel). The solid line and the dashed line are the true and estimated cumulative hazard functions, respectively. The dotted lines are the pointwise 95% confidence intervals.

n	Censoring level	Parameter	True value	AVE.bias	AVE.SD	CP
	60-65%	α_0	-0.5	0.009	0.193	0.980
		α_1	0.3	0.017	0.293	0.945
		β	0.5	0.017	0.188	0.935
		γ	0.5	-0.023	0.463	0.965
200	30-35%	α_0	0.5	0.031	0.296	0.949
		α_1	0.3	0.052	0.439	0.990
		β	0.5	-0.007	0.139	0.940
		γ	0.5	-0.006	0.355	0.940
	60-65%	α_0	-0.5	0.017	0.122	0.995
		α_1	0.3	0.018	0.159	0.970
		β	0.5	0.002	0.104	0.955
		γ	0.5	-0.005	0.251	0.950
500	30-35%	α_0	0.5	-0.010	0.090	0.935
		α_1	0.3	-0.009	0.164	0.950
		β	0.5	-0.001	0.084	0.955
		γ	0.5	-0.027	0.225	0.955

TABLE 3.1: Simulation results for sample size n=200 and n=500 for Scenario 1 with 200 replications. (AVE.bias for the average of differences between the estimates and the true value, AVE.SD for the average of standard deviations, and CP for the 95% coverage probability.)

by the EM algorithm. More details about the analysis of the data can be found in Banerjee and Carlin (2004) and Yu and Peng (2008).

The data recorded 223 subjects under different conditions who tried to quit smoking, and Murray et al. (1998) provided a full description of this dataset. The research lasted for five years, and each subject involved in the research was observed annually during this period. Thus, the exact relapse time of a subject was only known to lie between two consecutive observation times. In addition, some previous smokers had successfully quit smoking and thus were regarded as cured subjects. The subjects were randomly assigned to one of two groups at the beginning of the study: the smoking intervention (SI) group and the usual care (UC) group, which were also referred to as the treatment group and the control group, respectively. Other available covariates, including a subject's sex (0 for the male, 1 for the female), the number of cigarettes per day, and duration as smokers in years, were

also recorded. The study aimed to explore the intervention effects on the smoking relapse time.

To verify whether the accelerated hazards assumption is suitable for the unsusceptible subjects and the time-scaled effect of treatment type on the hazard function, we present the estimated nonparametric hazard rates without covariates under the intervention and usual care in Figure 3.2. It can be seen that the hazard rates of the two groups are nearly identical at the early stage of the study. However, the hazard rate of subjects in the smoking intervention group is gradually increases to a higher level along with the time. The violation of constant proportionality is aggravated by the apparent crossovers of smoothed hazard functions. It is more reasonable to involve the time-scaled effect of the intervention in the model. We consider the covariate SI/UC in the baseline hazard function and access all covariate effects in the logistic regression and the proportional component in the cumulative hazard function. Table 3.2 reports the estimation results.

In the incidence part, only the covariate duration as a smoker is significant at the 0.05 level. That is, subjects who had smoked for a long time tend to be cured after the intervention. The estimated coefficient (-0.560) of SI/UC indicates that with the intervention, the log odds of uncured subjects decreases by 56%. In other words, the odds of being uncured decreases by $1 - \exp(-0.560) = 43\%$. The estimated coefficient (0.301) of the covariate sex indicates that female smokers are more likely to relapse after quitting smoking than males as the odds of female smokers being uncured is $\exp(0.301) = 1.35$ times higher than that of male smokers. However, both covariate effects are not significant. These results are consistent with the corresponding findings obtained by Banerjee and Carlin (2004) and Xiang et al. (2011).

In the latency part, none of the four covariates is significant at level 0.05, consistent with the findings in Yu and Peng (2008). It indicates that for those who are not successful in quitting, the relapse time has no relationship with these

covariates. In contrast to the findings in Banerjee and Carlin (2004), the estimated coefficient (0.341) of SI/UC indicates that the intervention seems not to be useful for those susceptible subjects to quit smoking because the intervention does not lead to a decrease in the relative risk of relapse. The discrepancy may be due to the fact that covariate SI/UC is considered in the baseline hazard function of the proposed model. For the time-scaled effect of SI/UC, the estimated coefficient shows that the subjects who receive the intervention have decelerated hazard risks. Figure 3.3 shows the estimated baseline hazard function and the estimated cumulative baseline hazard function. It is clear that the subjects suffer a fluctuated risk at the beginning of the study, and then both estimated hazards have a rapid increase as time goes by.

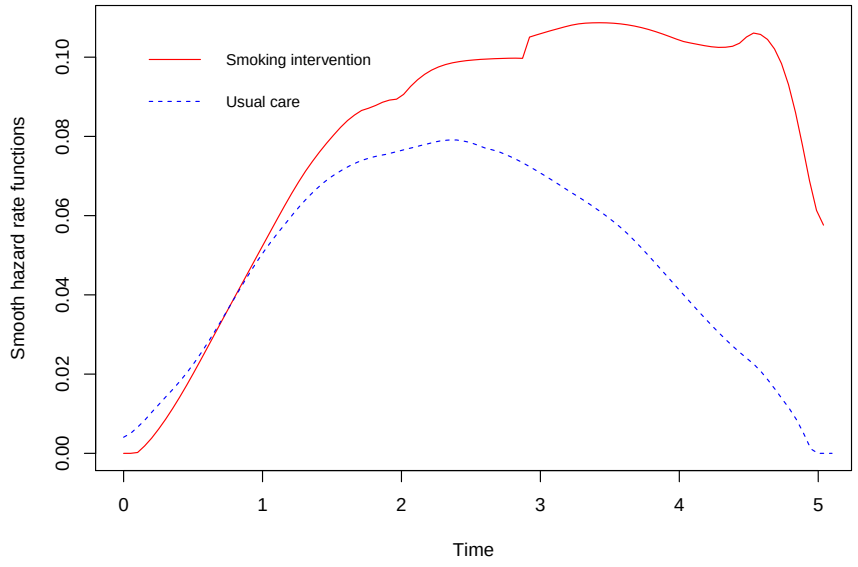


FIGURE 3.2: Smooth hazard rate functions for subjects in the smoking intervention group (solid line) and the usual care group (dashed line), respectively.

	Logistic regression		Survival model	
Parameter	Estimate	SE	Estimate	SE
Intercept	0.660	0.514		
Sex(female=1)	0.301	0.521	0.370	0.598
SI/UC (usual care =0)	-0.560	0.637	0.341	0.630
Cigarettes per day	0.077	0.051	-0.078	0.041
Duration as smoker	-0.095*	0.049	0.082	0.051
SI/UC(in hazard)			-0.042	0.083

$*$p-value < 0.05

TABLE 3.2: Parameter estimates in the GAHCure model for the smoking cessation data with time-scale effects.

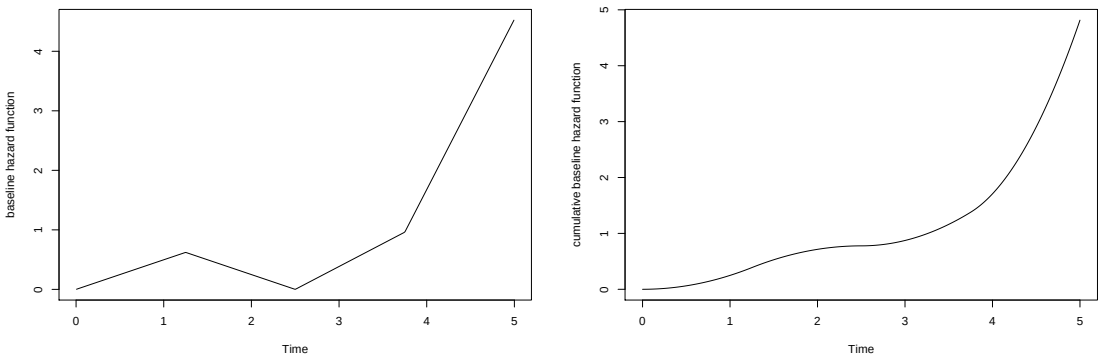


FIGURE 3.3: Estimated baseline hazard function (left panel) and cumulative hazard function (right panel) for the smoking cessation data.

3.6 Conclusion

In this chapter, a more flexible GAHCure model is proposed for analysis of interval-censored data with a cured subgroup. It is capable of evaluating the time-scaled effects and the proportional risk factors in the latency part. Theoretical properties of the proposed sieve ML estimator have been provided in the presence of the bundled parameters.

Farrington (2000) and Peng and Taylor (2017) investigated model checking methods based on residuals for interval-censored data and the mixture cure model, respectively. Following their works, adopting the residual-based approach to check the goodness-of-fit of the proposed GAHCure model with interval-censored data,

is possible. Model diagnosis for the mixture cure model in the presence of interval-censored data is a valuable topic but beyond the scope of current studies. A further investigation of this topic is in progress.

3.7 Technical Proofs

3.7.1 Proof of Theorem 3.1

Suppose $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda_0)$ and $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\gamma}^*, \Lambda_0^*)$ are two sets of parameters satisfying (3.4). To justify the identifiability of the model, it is sufficient to show that $l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda_0|\mathcal{O}) = l(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\gamma}^*, \Lambda_0^*|\mathcal{O})$ implies $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda_0) = (\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\gamma}^*, \Lambda_0^*)$.

First, we prove the identifiability of the incidence part. It is clear that the proposed model satisfies the constant-sum property according to Proposition 1 of Oller et al. (2004). Theorem 1 of Oller et al. (2007) indicates that, for the cure models with interval censoring, $l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda_0|\mathcal{O}) = l(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\gamma}^*, \Lambda_0^*|\mathcal{O})$ implies $S_p(t|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda_0) = S_p^*(t|\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\gamma}^*, \Lambda_0^*)$ for almost every t. Along the same lines of the method used by Li et al. (2001), we suppose $S_p(t|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda_0) = S_p^*(t|\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\gamma}^*, \Lambda_0^*)$. After rearranging the equation (3.1), we have the ratio

$$\frac{\pi(\boldsymbol{W})}{\pi^*(\boldsymbol{W})} = \frac{1 - S^*(t|\boldsymbol{X}, \boldsymbol{Z})}{1 - S(t|\boldsymbol{X}, \boldsymbol{Z})}. \tag{3.12}$$

The left hand side of the equation above depends on $\boldsymbol{W}$ only, where the right hand side depends on $\boldsymbol{X}$, $\boldsymbol{Z}$ and t. Hence, the ratio in (3.12) should be a constant free of $\boldsymbol{X}$, $\boldsymbol{Z}$,$\boldsymbol{W}$ or t, denoted by c. Then

$$S^*(t|\boldsymbol{X}, \boldsymbol{Z}) = 1 - c + cS(t|\boldsymbol{X}, \boldsymbol{Z}), \tag{3.13}$$

$$\pi^*(\boldsymbol{W}) = \pi(\boldsymbol{W})/c. \tag{3.14}$$

Without loss of generality, assume W is one-dimensional and $\alpha_2 \neq 0$. The equation $\frac{\exp(\alpha_1^*+W\alpha_2^*)}{1+\exp(\alpha_1^*+W\alpha_2^*)} = \frac{\exp(\alpha_1+W\alpha_2)}{[1+\exp(\alpha_1+W\alpha_2)]c}$ implies that $\alpha_1^* + W\alpha_2^* = -\ln[c-1+c\exp(-\alpha_1 - W\alpha_2)]$ for all possible W. If $c \neq 1$, there exist at least 3 different values W_k of W, providing three equations $\alpha_1^* + W_k\alpha_2^* = -\ln[c-1+c\exp(-\alpha_1 - W_k\alpha_2)]$, for $k = 1, 2, 3$. Then there is no root of α_1^* and α_2^* to make these equations hold. Thus $c = 1$, and $S_p(t|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda_0)$ is uniquely expressed by $\pi(\boldsymbol{W})$ and $S(t|\boldsymbol{X}, \boldsymbol{Z})$.

Next, we prove the identifiability of the conditional survival function. To this end, it is required to justify that $S(t|\boldsymbol{X}, \boldsymbol{Z}) = S^*(t|\boldsymbol{X}, \boldsymbol{Z})$ implies $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda_0) = (\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*, \Lambda_0^*)$. After the simple calculation, we can obtain that

$$\lambda_0(te^{\boldsymbol{X}^T\boldsymbol{\beta}})e^{\boldsymbol{X}^T\boldsymbol{\beta}+\boldsymbol{Z}^T\boldsymbol{\gamma}} = \lambda_0^*(te^{\boldsymbol{X}^T\boldsymbol{\beta}^*})e^{\boldsymbol{X}^T\boldsymbol{\beta}^*+\boldsymbol{Z}^T\boldsymbol{\gamma}^*}. \tag{3.15}$$

By Proportion 1 of Chen and Jewell (2001), the above model is unidentifiable unless $\lambda_0(t) = c_1 t^{c_2}$ for some constants c_1 and c_2, corresponding to the Weibull distribution. In other words, it is identifiable as long as the baseline distribution is not Weibull.

3.7.2 Proof of Theorem 3.2

The consistency of estimators was established by Vaart and Wellner (1996) and Van der Vaart (1998), using the empirical process theorem. Following their works, we prove the consistency of $\hat{\boldsymbol{\theta}}_n$. Employing the notations of Vaart and Wellner (1996), we denote $Pf = \int f(x)dP$ as the expectation of the function $f(x)$, and $P_n f = \frac{1}{n}\sum_{i=1}^n f(X_i)$ as the empirical process indexed by $f(X)$. Let $M(\boldsymbol{\theta}) = Pl(\boldsymbol{\theta}|\mathcal{O})$, $M_n(\boldsymbol{\theta}) = P_n l(\boldsymbol{\theta}|\mathcal{O})$, and let C denote a constant that may take different values. For all $\boldsymbol{\theta} \in \Theta_n$, $M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta}) = P_n l(\boldsymbol{\theta}|\mathcal{O}) - Pl(\boldsymbol{\theta}|\mathcal{O}) = (P_n - P)l(\boldsymbol{\theta}|\mathcal{O})$.

The consistency of $\hat{\boldsymbol{\theta}}_n$ can be established by checking three conditions of Theorem 5.7 in Van der Vaart (1998). The first condition we need to verify is

$$\sup_{\boldsymbol{\theta}\in\Theta_n} |M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| \xrightarrow{p} 0.$$

In fact, it can be proved by constructing ϵ-brackets. Let $\mathcal{L}_1 = \{l(\boldsymbol{\theta}|\mathcal{O}) : \boldsymbol{\theta} \in \Theta_n\}$. The result of Shen and Wong (1994) implies that, for any $\epsilon > 0$, there exists a set of brackets $\{[\phi_i^L, \phi_i^U] : i = 1, 2, \dots \lceil (1/\epsilon)^{Cq_n} \rceil\}$ satisfying the following properties. For any $\phi \in \mathcal{S}_n$, there exists i such that $\phi_i^L(t) \le \phi(t) \le \phi_i^U(t)$ for all t and $P_n\|\phi_i^U - \phi_i^L\| \le \epsilon$. Condition C1 implies that all parameter sets $\mathcal{A}, \mathcal{B}$ and Γ are compact, thus they can be covered by $\lceil C(1/\epsilon)^{p_1} \rceil$, $\lceil C(1/\epsilon)^{p_2} \rceil$, and $\lceil C(1/\epsilon)^{p_3} \rceil$ balls with radius ϵ, respectively. Namely, for any $\boldsymbol{\alpha} \in \mathcal{A}$, $\boldsymbol{\beta} \in \mathcal{B}$, and $\boldsymbol{\gamma} \in \Gamma$, there exist $1 \le l \le \lceil C(1/\epsilon)^{p_1} \rceil$, $1 \le s \le \lceil C(1/\epsilon)^{p_2} \rceil$ and $1 \le k \le \lceil C(1/\epsilon)^{p_3} \rceil$ such that $\|\boldsymbol{\alpha}_l - \boldsymbol{\alpha}\| \le \epsilon$, $\|\boldsymbol{\beta}_s - \boldsymbol{\beta}\| \le \epsilon$ and $\|\boldsymbol{\gamma}_k - \boldsymbol{\gamma}\| \le \epsilon$. Moreover, because the covariates $\boldsymbol{X}, \boldsymbol{Z}$, and $\boldsymbol{W}$ are bounded by condition C2, we have $\|\boldsymbol{W}^T\boldsymbol{\alpha}_l - \boldsymbol{W}^T\boldsymbol{\alpha}\| \le C\epsilon$ and $\|\boldsymbol{Z}^T\boldsymbol{\gamma}_k - \boldsymbol{Z}^T\boldsymbol{\gamma}\| \le C\epsilon$, leading to $\boldsymbol{W}^T\boldsymbol{\alpha} \in [\boldsymbol{W}^T\boldsymbol{\alpha}_l - C\epsilon, \boldsymbol{W}^T\boldsymbol{\alpha}_l + C\epsilon]$ and $\boldsymbol{Z}^T\boldsymbol{\gamma} \in [\boldsymbol{Z}^T\boldsymbol{\gamma}_k - C\epsilon, \boldsymbol{Z}^T\boldsymbol{\gamma}_k + C\epsilon]$, respectively. By the mean value theorem, there exists $\zeta \in [1, 1 + C\epsilon]$ such that $e^{\boldsymbol{X}^T\boldsymbol{\beta}_s} \in [e^{\boldsymbol{X}^T\boldsymbol{\beta}_s}\zeta^{-1}, e^{\boldsymbol{X}^T\boldsymbol{\beta}_s}\zeta]$ for some small enough ϵ. For $\mathcal{L}_1 = \{l(\boldsymbol{\theta}|\mathcal{O}) : \boldsymbol{\theta} \in \Theta_n\}$, we now construct the brackets as follows.

Let

$$\begin{aligned}
l^L_{i,l,s,k}(\mathcal{O}) = {} & \delta_{L_i} \log \left[1 - \exp\{-e^{\phi_i^L(Re^{\boldsymbol{X}^T\boldsymbol{\beta}_s}\zeta^{-1}) + \boldsymbol{Z}^T\boldsymbol{\gamma}_k - C\epsilon}\}\right] \\
& + \delta_{I_i} \log \left[\exp\{-e^{\phi_i^U(Le^{\boldsymbol{X}^T\boldsymbol{\beta}_s}\zeta^{-1}) + \boldsymbol{Z}^T\boldsymbol{\gamma}_k + C\epsilon}\} - \exp\{-e^{\phi_i^L(Re^{\boldsymbol{X}^T\boldsymbol{\beta}_s}\zeta^{-1}) + \boldsymbol{Z}^T\boldsymbol{\gamma}_k - C\epsilon)}\}\right] \\
& + \delta_{R_i} \log \left[\frac{1}{1 + e^{\boldsymbol{W}^T\boldsymbol{\alpha}_l + C\epsilon}} + \frac{e^{\boldsymbol{W}^T\boldsymbol{\alpha}_l - C\epsilon}}{1 + e^{\boldsymbol{W}^T\boldsymbol{\alpha}_l - C\epsilon}} \exp\{-e^{\phi_i^L(Re^{\boldsymbol{X}^T\boldsymbol{\beta}_s}\zeta^{-1}) + \boldsymbol{Z}^T\boldsymbol{\gamma}_k + C\epsilon}\}\right]\Bigg) \\
& + (1 - \delta_{R_i}) \log \frac{e^{\boldsymbol{W}^T\boldsymbol{\alpha}_l - C\epsilon}}{1 + e^{\boldsymbol{W}^T\boldsymbol{\alpha}_l - C\epsilon}},
\end{aligned}$$

and

$$
\begin{aligned}
l^U_{i,l,s,k}(\mathcal{O}) = \delta_{L_i} \log &\left[1 - \exp\{-e^{\phi^L_i(Re^{\boldsymbol{X}^T\boldsymbol{\beta}_s}\zeta)+\boldsymbol{Z}^T\boldsymbol{\gamma}_k+C\epsilon}\}\right] \\
&+ \delta_{I_i} \log \left[\exp\{-e^{\phi^U_i(Le^{\boldsymbol{X}^T\boldsymbol{\beta}_s}\zeta)+\boldsymbol{Z}^T\boldsymbol{\gamma}_k-C\epsilon}\} - \exp\{-e^{\phi^L_i(Re^{\boldsymbol{X}^T\boldsymbol{\beta}_s}\zeta)+\boldsymbol{Z}^T\boldsymbol{\gamma}_k+C\epsilon)}\}\right] \\
&+ \delta_{R_i} \log \left[\frac{1}{1+e^{\boldsymbol{W}^T\boldsymbol{\alpha}_l-C\epsilon}} + \frac{e^{\boldsymbol{W}^T\boldsymbol{\alpha}_l+C\epsilon}}{1+e^{\boldsymbol{W}^T\boldsymbol{\alpha}_l+C\epsilon}} \exp\{-e^{\phi^L_i(Re^{\boldsymbol{X}^T\boldsymbol{\beta}_s}\zeta)+\boldsymbol{Z}^T\boldsymbol{\gamma}_k-C\epsilon}\}\right]\Bigg) \\
&+ (1-\delta_{R_i}) \log \frac{e^{\boldsymbol{W}^T\boldsymbol{\alpha}_l+C\epsilon}}{1+e^{\boldsymbol{W}^T\boldsymbol{\alpha}_l+C\epsilon}}.
\end{aligned}
$$

For any $l(\boldsymbol{\theta}|\mathcal{O}) \in \mathcal{L}_1$, there exists (i,l,s,k), where $i \leq (1/\epsilon)^{Cq_n}$, $l \leq C(1/\epsilon)^{p_1}$, $s \leq C(1/\epsilon)^{p_2}$ and $k \leq C(1/\epsilon)^{p_3}$, such that the bracket $[l^L_{i,l,s,k}, l^U_{i,l,s,k}]$ covers $l(\boldsymbol{\theta}|\mathcal{O})$. After some calculations, we can obtain the length $|l^U_{i,l,s,k} - l^L_{i,l,s,k}| \leq C\epsilon$. Let $N_{[]}(\epsilon, \mathcal{L}_1, L_1(P_n))$ and $N(\epsilon, \mathcal{L}_1, L_1(P_n))$ be the ϵ-bracketing number and the corresponding ϵ-covering number of $\mathcal{L}_1$ with respect to $L_1(P_n)$ norm, respectively. $N_{[]}(\epsilon, \mathcal{L}_1, L_1(P_n)) \leq (1/\epsilon)^{Cq_n}(1/\epsilon)^{p_1}(1/\epsilon)^{p_2}(1/\epsilon)^{p_3} = (1/\epsilon)^{Cq_n+d}$ is bounded. Since $N(\epsilon, \mathcal{L}_1, L_1(P_n)) \leq N_{[]}(2\epsilon, \mathcal{L}_1, L_1(P_n))$, we have $\sup_{\boldsymbol{\theta}\in\Theta_n} |M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| \xrightarrow{p} 0$ by Theorem 2.4.3 of Vaart and Wellner (1996).

The second condition is that $\sup_{\boldsymbol{\theta}:d(\boldsymbol{\theta},\boldsymbol{\theta_0}>\epsilon)} M(\boldsymbol{\theta}) < M(\theta_0)$, which can be verified as follows. Note that the Gibbs inequality implies that $\sup_{\boldsymbol{\theta}:d(\boldsymbol{\theta},\boldsymbol{\theta_0}>\epsilon)} M(\boldsymbol{\theta}) \leq M(\theta_0)$ for all $\boldsymbol{\theta} \in \Theta_n$. If $\sup_{\boldsymbol{\theta}:d(\boldsymbol{\theta},\boldsymbol{\theta_0}>\epsilon)} M(\boldsymbol{\theta}) = M(\theta_0)$ holds for some $\boldsymbol{\theta} \in \Theta_n$, then there exists a sequence θ_m such that $M(\boldsymbol{\theta}_m) \to \sup_{\boldsymbol{\theta}:d(\boldsymbol{\theta},\boldsymbol{\theta_0}>\epsilon)} M(\boldsymbol{\theta}) = M(\theta_0)$ and $d(\boldsymbol{\theta}_m, \boldsymbol{\theta_0}) > \epsilon$. As $\boldsymbol{\tau}_0 = (\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T, \boldsymbol{\gamma}_0^T)^T$ is a compact set and the coefficients of spline functions are bounded, there exists a subsequence $\boldsymbol{\theta}_{m'}$ of $\boldsymbol{\theta}_m$, converging to $\boldsymbol{\theta}_{m0}$. Since $M(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$, $M(\boldsymbol{\theta}_{m0}) = M(\boldsymbol{\theta}_0)$ and consequently $\boldsymbol{\theta}_{m0} = \boldsymbol{\theta}_0$ according to the identifiability of the proposed model. However, $\boldsymbol{\theta}_{m'}$ does not converge to $\boldsymbol{\theta}_0$ due to the fact $d(\boldsymbol{\theta}_{m'}, \boldsymbol{\theta_0}) > \epsilon$. This conflicts with the aforementioned result that $\boldsymbol{\theta}_{m'}$ converges to $\boldsymbol{\theta}_{m_0}$. Therefore, we obtain that $\sup_{\boldsymbol{\theta}:d(\boldsymbol{\theta},\boldsymbol{\theta_0}>\epsilon)} M(\boldsymbol{\theta}) < M(\boldsymbol{\theta}_0)$.

The third condition we need to verify is that $M_n(\hat{\boldsymbol{\theta}}_n) \geq M_n(\boldsymbol{\theta}_0) - o_p(1)$. Let $\boldsymbol{\theta}_{0,n} = (\boldsymbol{\tau}_0^T, \phi_{0,n})^T$, we have

$$\begin{aligned} M_n(\hat{\boldsymbol{\theta}}_n) - M_n(\boldsymbol{\theta}_0) &= M_n(\hat{\boldsymbol{\theta}}_n) - M_n(\boldsymbol{\theta}_{0,n}) + M_n(\boldsymbol{\theta}_{0,n}) - M_n(\boldsymbol{\theta}_0) \\ &\geq P_n l(\boldsymbol{\theta}_{0,n}|\mathcal{O}) - P_n l(\boldsymbol{\theta}_0|\mathcal{O}) \\ &= (P_n - P)\{l(\boldsymbol{\theta}_{0,n}|\mathcal{O}) - l(\boldsymbol{\theta}_0|\mathcal{O})\} + P\{l(\boldsymbol{\theta}_{0,n}|\mathcal{O}) - l(\boldsymbol{\theta}_0|\mathcal{O})\}. \end{aligned}$$

Firstly, we show that $P\{l(\boldsymbol{\theta}_{0,n}|\mathcal{O}) - l(\boldsymbol{\theta}_0|\mathcal{O})\} = Pl(\boldsymbol{\theta}_{0,n}|\mathcal{O}) - Pl(\boldsymbol{\theta}_0|\mathcal{O}) = -o_p(1)$. The arguments of Lu et al. (2007) show that there is a function $\phi_{0,n}$ such that $\|\phi_{0,n} - \phi_0\| \leq Cq_n^p = O(n^{-np})$. By the dominated convergence theorem, it is easy to see that

$$P\{l(\boldsymbol{\theta_{0,n}}|\mathcal{O}) - l(\boldsymbol{\theta_0}|\mathcal{O})\} \leq C\|\phi_{0,n}(\cdot, \beta_0) - \phi_0(\cdot, \beta_0)\|_2^2 \leq C\|\phi_{0,n}(\cdot, \beta_0) - \phi_0(\cdot, \beta_0)\|_\infty^2.$$

As $n \to \infty$, we have $P\{l(\boldsymbol{\theta_{0,n}}|\mathcal{O}) - l(\boldsymbol{\theta_0}|\mathcal{O})\} \to 0$, thus $P\{l(\boldsymbol{\theta_{0,n}}|\mathcal{O}) - l(\boldsymbol{\theta_0}|\mathcal{O})\} = -o_p(1)$.

Secondly, we show that $(P_n - P)\{l(\boldsymbol{\theta}_{0,n}|\mathcal{O}) - l(\boldsymbol{\theta}_0|\mathcal{O})\} = o_p(n^{-1/2})$. Define $\mathcal{L}_2 = \{l(\boldsymbol{\tau}_0, \phi|\mathcal{O}) - l(\boldsymbol{\tau}_0, \phi_0|\mathcal{O}), \phi \in \mathcal{S}_n, \|\phi - \phi_0\| \leq Cn^{-pv}\}$. Clearly $l(\boldsymbol{\tau}_0, \phi_{0,n}|\mathcal{O}) - l(\boldsymbol{\tau_0}, \phi_0|\mathcal{O}) \in \mathcal{L}_2$. We construct a set of brackets $\{l_{i,l,s,k}^L(\boldsymbol{\tau}_0, \phi|\mathcal{O}) - l_{i,l,s,k}(\boldsymbol{\tau}_0, \phi_0|\mathcal{O}), l_{i,l,s,k}^U(\boldsymbol{\tau}_0, \phi|\mathcal{O}) - l_{i,l,s,k}(\boldsymbol{\tau}_0, \phi_0|\mathcal{O})\}$. Similar to the proof of bounded bracket number of $\mathcal{L}_1$, we can prove that the ϵ- bracketing number of brackets is also bounded by $(1/\epsilon)^{Cq_n}$. Moreover, the bracketing integral

$$J_{[]}(\delta, \mathcal{L}_2, L_2(P)) = \int_0^\delta \sqrt{\log N(\epsilon, \mathcal{L}_2, L_2(P))} d\epsilon \leq \int_0^\delta \sqrt{Cq_n \log 1/\epsilon} d\epsilon < \infty,$$

which is finite by the finite-value bracketing integral defined in Van der Vaart (1998)(P270). Thus $\mathcal{L}_2$ is a P-Donsker by Theorem 19.5 in Van der Vaart (1998).

According to Corollary 2.3.12 in Vaart and Wellner (1996), we have

$$(P_n - P)\{l(\boldsymbol{\theta}_{0,n}|\mathcal{O}) - l(\boldsymbol{\theta}_0|\mathcal{O})\} = o_p(n^{-1/2}).$$

Therefore,

$$M_n(\hat{\boldsymbol{\theta}}_n) - M_n(\boldsymbol{\theta}) \geq o_p(n^{-1/2}) - o_p(1) = -o_p(1).$$

Now, three conditions of Theorem 5.7 in Van der Vaart (1998) are verified, and thus we can obtain that $d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) \to 0$ in probability.

3.7.3 Proof of Theorem 3.3

The rate of convergence can be verified by checking the conditions of Theorem 3.4.1 in Vaart and Wellner (1996).

The first condition is that for every n, and an arbitrary δ with $\delta > \delta_n = n^{-pv}$,

$$\sup_{\delta/2 < d(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n}) < \delta, \boldsymbol{\theta} \in \Theta} \left(M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_{0,n}) \right) \leq -c\delta^2.$$

From the proof of Theorem 2, we have $d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{0,n}) = O(n^{-pv})$ and $M(\boldsymbol{\theta}_0) - M(\boldsymbol{\theta}_{0,n}) \leq Cd^2(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{0,n}) \leq O(n^{-2pv})$. Moreover, $M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_0) = Pl(\boldsymbol{\theta}|\mathcal{O}) - Pl(\boldsymbol{\theta}_0|\mathcal{O}) \leq -Cd^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \leq -C\delta^2$ for every $\boldsymbol{\theta}$ in the neighbourhood of $\boldsymbol{\theta}_0$. Then, for a large enough n, we have

$$\begin{aligned} M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_{0,n}) &= M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_0) + M(\boldsymbol{\theta}_0) - M(\boldsymbol{\theta}_{0,n}) \\ &\leq -C\delta^2 + Cn^{-2pv} \\ &\leq -c\delta^2, \end{aligned}$$

which verifies the first condition in Theorem 3.4.1 of Vaart and Wellner (1996).

The second condition needed to be verified is that there exists a proper function $\phi(\cdot)$ such that

$$E[\sup_{\delta/2<d(\boldsymbol{\theta},\boldsymbol{\theta}_{0,n})<\delta} \sqrt{n}(M_n - M)(\boldsymbol{\theta} - \boldsymbol{\theta}_{0,n})] < c\frac{\phi(\delta)}{\sqrt{n}}$$

holds for small enough δ, where $\delta \to \phi(\delta)/\delta^{\vartheta}$ is a decreasing function of δ for some $\vartheta < 2$, and for $r_n \leq \delta_n^{-1} = n^{pv}$, the function $\phi(\cdot)$ satisfies $r_n^2\phi(1/r_n) \leq c\sqrt{n}$ for every n. Define a class $\mathcal{L}_\delta = \{l(\boldsymbol{\theta}|\mathcal{O}) - l(\boldsymbol{\theta}_{0,n}|\mathcal{O}) : \boldsymbol{\theta} \in \Theta_n, \delta/2 < d(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n}) < \delta\}$. Using the similar arguments used in Theorem 3.2 regrading $\mathcal{L}_2$ and Lemma 0.6 of Wu et al. (2012), we can prove that the ϵ-bracketing number of $\mathcal{L}_\delta$ is bounded by $(\delta/\epsilon)^{Cq_n}$. Since L_2-norm is bounded by $\| \quad \|_\infty$ norm, we have

$$N_{[]}\{\epsilon, \mathcal{L}_\delta, L_2(P)\} \leq N_{[]}\{\epsilon, \mathcal{L}_\delta, \| \quad \|_\infty\} \leq (\delta/\epsilon)^{Cq_n}.$$

Under conditions C1-C5, L_δ is uniformly bounded and thus we have $P\{l(\boldsymbol{\theta}|\mathcal{O}) - l(\boldsymbol{\theta}_{0,n}|\mathcal{O})\}^2 \leq Cd^2(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n}) \leq C\delta^2$. Moreover,

$$\begin{aligned} J_{[]}\{\epsilon, \mathcal{L}_\delta, L_2(P)\} &= \int_0^\delta [1 + \log N_{[]}\{\epsilon, \mathcal{L}_\delta, L_2(P)\}]^{1/2} d\epsilon \\ &\leq \int_0^\delta \sqrt{1 + Cq_n \log(\delta/\epsilon)} d\epsilon \\ &\leq \int_0^\delta Cq_n^{1/2}(\delta/\epsilon)^{1/2} d\epsilon = Cq_n^{1/2}\delta. \end{aligned}$$

By applying Lemma 3.4.2 of Van der Vaart (1998), we have

$$\begin{aligned} E\|P_n - P\|_{L_\delta} &\leq CJ_{[]}\{\epsilon, \mathcal{L}_\delta, L_2(P)\}[1 + \frac{J_{[]}\{\epsilon, \mathcal{L}_\delta, L_2(P)\}}{\delta^2 n^{1/2}}] \\ &\leq C\phi(\delta), \end{aligned}$$

with $\phi(\delta) = qn^{1/2}\delta + q_n/n^{1/2}$. It is easy to see that $\phi(\delta)/\delta$ is a decreasing function of δ. If $r_n = q_n^{-1/2}n^{1/2}$, we have $r_n^2\phi(1/r_n) = r_n q_n^{1/2} + r_n^2 q_n^{1/2}/n^{1/2} = n^{1/2}$.

With the sieve ML estimator $\hat{\boldsymbol{\theta}}_n$, we have $P_n\{l(\hat{\boldsymbol{\theta}}_n|\mathcal{O}) - l(\boldsymbol{\theta}_{0,n}|\mathcal{O})\} \geq 0$, and $d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{0,n}) \leq d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) + d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{0,n}) \to 0$ in probability. By Theorem 3.4.1 in Van der Vaart (1998), $r_n d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{0,n}) = O_p(1)$. Using the fact that $d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{0,n}) = O(n^{-pv})$, we obtain

$$\begin{aligned} d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) &= O_p(q_n^{1/2} n^{-1/2} + C n^{-pv}) \\ &= O_p(n^{-(1-v)/2} + n^{-pv}). \end{aligned}$$

3.7.4 Proof of Theorem 3.4

The asymptotic normality of the sieve ML estimator with bundled parameters has been established in semiparametric models by Ding and Nan (2011). We prove the asymptotic normality of the proposed estimator by verifing the conditions of their Theorem 2.1.

Let Φ^p be a class of bounded functions ϕ on $[a, b]$, and define

$$\mathcal{H}^p = \big\{\xi(\,\cdot\,, \boldsymbol{\beta}) = \phi(\psi(t, X, \boldsymbol{\beta})), \phi \in \Phi^p, t \in [a, b], X \in \mathcal{X}, \boldsymbol{\beta} \in \mathcal{B}\big\},$$

which is the space of nonparametric terms involving parameters $\boldsymbol{\beta}$. We apply the chain rule to the composite function ξ with $\psi(t, X, \boldsymbol{\beta}) = te^{\boldsymbol{X}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}$. Then $\xi(t, \boldsymbol{X}, \boldsymbol{\beta}_0) = \phi(t)$.

Define the direction space as:

$$\mathcal{H} = \big\{h : h(\,\cdot\,, \boldsymbol{\beta}) = \frac{\partial \xi_\eta(\,\cdot\,, \boldsymbol{\beta})}{\partial \eta}|_{\eta=0} = \omega(\psi(\,\cdot\,, \boldsymbol{\beta})), \xi_\eta \in \mathcal{H}^p\big\}.$$

In the following we verify that the following assumptions required by Theorem 2.1 of Ding and Nan (2011) hold.

(B1) $d(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}, \boldsymbol{\theta}_{\mathbf{0}}) = O_p(n^{-\varrho})$ for some $\varrho > 0$.

(B2) $Pl'_{\boldsymbol{\tau}}(\boldsymbol{\tau}_{\mathbf{0}}, \xi_0(\,\cdot\,, \boldsymbol{\beta}_{\mathbf{0}})|\mathcal{O}) = 0$ and $Pl'_{\xi}(\boldsymbol{\tau}_{\mathbf{0}}, \xi_0(\,\cdot\,, \boldsymbol{\beta}_{\mathbf{0}})|\mathcal{O})[h] = 0$ for all $h \in \mathcal{H}$.

(B3) There exists a least favourable direction $\boldsymbol{h}^*$with dimension $d = p_1 + p_2 + p_3$, where $h_j^* \in \mathcal{H}$ for $j = 1, \ldots, d$ such that

$$Pl''_{\boldsymbol{\tau}\xi}(\boldsymbol{\tau_0}, \xi_0(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[h] - Pl''_{\xi\xi}(\boldsymbol{\tau_0}, \xi_0(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}^*, h] = 0.$$

In addition, the matrix $P\big\{l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O}) - l''_{\xi\boldsymbol{\tau}}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}^*]\big\}$ is non-singular.

(B4) $P_n l'_{\boldsymbol{\tau}}(\hat{\boldsymbol{\tau}}_{\boldsymbol{n}}, \hat{\xi}_n(\cdot, \hat{\boldsymbol{\beta}}_{\boldsymbol{n}})|\mathcal{O}) = o_p(n^{-1/2})$ and $P_n l'_{\xi}(\hat{\boldsymbol{\tau}}_{\boldsymbol{n}}, \hat{\xi}_n(\cdot, \hat{\boldsymbol{\beta}}_{\boldsymbol{n}})|\mathcal{O})[\boldsymbol{h}^*] = o_p(n^{-1/2})$.

(B5) For some $c > 0$, let $G_n = n^{1/2}(P_n - P)$,

$$\sup_{d(\boldsymbol{\theta}, \boldsymbol{\theta_0}) \leq cn^{-\varrho}, \boldsymbol{\theta} \in \Theta_n} \big| G_n l'_{\boldsymbol{\tau}}(\boldsymbol{\tau}, \xi(\cdot, \boldsymbol{\beta})|\mathcal{O}) - G_n l'_{\tau}(\boldsymbol{\tau_0}, \xi_0(\cdot, \boldsymbol{\beta_0})|\mathcal{O}) \big| = o_p(1),$$

and

$$\sup_{d(\boldsymbol{\theta}, \boldsymbol{\theta_0}) \leq cn^{-\varrho}, \boldsymbol{\theta} \in \Theta_n} \big| G_n l'_{\xi}(\boldsymbol{\tau}, \xi(\cdot, \boldsymbol{\beta})|\mathcal{O})[\boldsymbol{h}^*(\cdot, \boldsymbol{\beta})] - G_n l'_{\xi}(\boldsymbol{\tau_0}, \xi_0(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}^*(\cdot, \boldsymbol{\beta}_0)] \big| = o_p(1).$$

(B6) For some $\iota > 1$, which satisfies $\iota\varrho > 1/2$, and for $\boldsymbol{\theta}$ in neighbourhood of $\boldsymbol{\theta_0} : \{\boldsymbol{\theta}, d(\boldsymbol{\theta}, \boldsymbol{\theta_0}) \leq cn^{-\varrho}, \boldsymbol{\theta} \in \Theta_n\}$,

$$\begin{aligned} &Pl'_{\boldsymbol{\tau}}(\boldsymbol{\tau}, \xi(\cdot, \boldsymbol{\beta})|\mathcal{O}) - Pl'_{\boldsymbol{\tau}}(\boldsymbol{\tau_0}, \xi_0(\cdot, \boldsymbol{\beta_0})|\mathcal{O}) - Pl''_{\boldsymbol{\tau\tau}}(\boldsymbol{\tau_0}, \xi_0(\cdot, \boldsymbol{\beta_0})|\mathcal{O})(\boldsymbol{\tau} - \boldsymbol{\tau_0}) \\ &- Pl''_{\boldsymbol{\tau}\xi}(\boldsymbol{\tau_0}, \xi_0(\cdot, \boldsymbol{\beta_0})|\mathcal{O})|\xi(\cdot, \boldsymbol{\beta}) - \xi_0(\cdot, \boldsymbol{\beta})| \\ &= O(d^{\iota}(\boldsymbol{\theta}, \boldsymbol{\theta_0})), \end{aligned}$$

and

$$
\begin{aligned}
&Pl'_{\xi}(\boldsymbol{\tau}, \xi(\cdot, \boldsymbol{\beta})|\mathcal{O})[\boldsymbol{h}^*(\cdot, \boldsymbol{\beta})] - Pl'_{\xi}(\boldsymbol{\tau_0}, \xi_0(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}^*(\cdot, \boldsymbol{\beta}_0)] \\
&- Pl''_{\xi\boldsymbol{\tau}}(\boldsymbol{\tau_0}, \xi_0(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}^*(\cdot, \boldsymbol{\beta}_0)](\boldsymbol{\tau} - \boldsymbol{\tau_0}) \\
&- Pl''_{\xi\xi}(\boldsymbol{\tau_0}, \xi_0(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}^*(\cdot, \boldsymbol{\beta}_0), \xi(\cdot, \boldsymbol{\beta}) - \xi_0(\cdot, \boldsymbol{\beta_0})] \\
&= O(d^{\iota}(\boldsymbol{\theta}, \boldsymbol{\theta_0})).
\end{aligned}
$$

Before justifying the above conditions, we give a discussion about them. The convergence rate in assumption B1 is required prior to obtaining the asymptotic normality. Assumption B2 is a common assumption in the maximum likelihood estimation, which evaluates the score function at true values. Assumption B3 is standard in the maximum likelihood theory, which is used to find the least favourite direction $\boldsymbol{h}^*$ and provides the non-singular information matrix along the direction. The score function of the estimator at the sample level is assessed by assumption B4. The stochastic equicontinuities in assumption B5 are verified by employing the Donsker property and the maximal inequality. Taylor expansion is used to verify assumption B6.

First, assumption B1 holds with $\varrho = \min((1-\upsilon)/2, p\upsilon)$ by the proof of the convergence rate. Assumption B2 is true due to the property of the zero-mean score function.

Consider assumption B3, For presentation simplicity, we use π for $\pi(\boldsymbol{W}) = \frac{\exp(\boldsymbol{W}^T\boldsymbol{\alpha})}{1+\exp(\boldsymbol{W}^T\boldsymbol{\alpha})}$, $S(t)$ for $S(t|X, Z) = \exp\{-\exp[\phi(t\boldsymbol{X}^T\boldsymbol{\beta}) + \boldsymbol{Z}^T\boldsymbol{\gamma}]\}$ and $S_p(t)$ for $S_p(t|\boldsymbol{X}, \boldsymbol{Z},) = \pi(\boldsymbol{W})S(t|\boldsymbol{X}, \boldsymbol{Z}) + 1 - \pi(\boldsymbol{W})$. The log-likelihood based on one observation is

$$
l(\boldsymbol{\theta}|\mathcal{O}) = \delta_L \log(1 - S_p(R)) + \delta_I \log(S_p(L)) - S_p(R)) + \delta_R \log(S_p(L)). \qquad (3.16)
$$

Denote

$$
\begin{aligned}
S'_{\boldsymbol{\alpha}}(t) &= \frac{\partial S_p(t)}{\partial \boldsymbol{\alpha}} = \pi(1-\pi)\cdot(S(t)-1)\cdot \boldsymbol{W},\\
S'_{\boldsymbol{\beta}}(t) &= \frac{\partial S_p(t)}{\partial \boldsymbol{\beta}} = S(t)\log(S(t))\cdot \phi'(te^{\boldsymbol{X}^T\boldsymbol{\beta}})\cdot te^{\boldsymbol{X}^T\boldsymbol{\beta}}\cdot \boldsymbol{X},\\
S'_{\boldsymbol{\gamma}}(t) &= \frac{\partial S_p(t)}{\partial \boldsymbol{\gamma}} = S(t)\log(S(t))\cdot \boldsymbol{Z}^T\boldsymbol{\gamma}\cdot \boldsymbol{Z},\\
S'_{\boldsymbol{\xi}}(t)(h) &= \frac{\partial S_p(t)}{\partial \eta}|_{\eta=0} = \pi S(t)\log(S(t))\cdot h(te^{\boldsymbol{X}^T\boldsymbol{\beta}}).
\end{aligned}
$$

Let $D_1(S) = \frac{1}{1-S_p(R)} + \frac{1}{S_p(L)-S_p(R)}$ and $D_2(S) = \frac{1}{S_p(L)} + \frac{1}{S_p(L)-S_p(R)}$, then the score functions of the parametric and nonparametric components are:

$$
l_1(\boldsymbol{\theta}|\mathcal{O}) = (l_{\boldsymbol{\alpha}}^T, l_{\boldsymbol{\beta}}^T, l_{\boldsymbol{\gamma}}^T)^T = \begin{pmatrix} D_1(S)\cdot S'_{\boldsymbol{\alpha}}(R) + D_2(S)\cdot S'_{\boldsymbol{\alpha}}(L) \\ D_1(S)\cdot S'_{\boldsymbol{\beta}}(R) + D_2(S)\cdot S'_{\boldsymbol{\beta}}(L) \\ D_1(S)\cdot S'_{\boldsymbol{\gamma}}(R) + D_2(S)\cdot S'_{\boldsymbol{\gamma}}(L) \end{pmatrix},
$$

and

$$
l_2(\boldsymbol{\theta}|\mathcal{O})[\boldsymbol{h}] = D_1(S)\cdot S'_{\boldsymbol{\xi}}(R)(h) + D_2(S)\cdot S'_{\boldsymbol{\xi}}(L)(\boldsymbol{h}).
$$

The efficient score functions of the parametric components are

$$
\begin{aligned}
l^*(\boldsymbol{\theta}|\mathcal{O}) &= l_1(\boldsymbol{\theta}|\mathcal{O}) - l_2(\boldsymbol{\theta}|\mathcal{O})[\boldsymbol{h}^*]\\
&= \begin{pmatrix} D_1(S)\cdot[S'_{\boldsymbol{\alpha}}(R) - S'_{\boldsymbol{\xi}}(R)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})] + D_2(S)\cdot[S'_{\boldsymbol{\alpha}}(L) - S'_{\boldsymbol{\xi}}(L)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})] \\ D_1(S)\cdot[S'_{\boldsymbol{\beta}}(R) - S'_{\boldsymbol{\xi}}(R)(\boldsymbol{h}^*_{\boldsymbol{\beta}})] + D_2(S)\cdot[S'_{\boldsymbol{\beta}}(L) - S'_{\boldsymbol{\xi}}(L)(\boldsymbol{h}^*_{\boldsymbol{\beta}})] \\ D_1(S)\cdot[S'_{\boldsymbol{\gamma}}(R) - S'_{\boldsymbol{\xi}}(R)(\boldsymbol{h}^*_{\boldsymbol{\gamma}})] + D_2(S)\cdot[S'_{\boldsymbol{\gamma}}(L) - S'_{\boldsymbol{\xi}}(L)(\boldsymbol{h}^*_{\boldsymbol{\gamma}})] \end{pmatrix},
\end{aligned}
$$

where $\boldsymbol{h}^* = (\boldsymbol{h}^*_{\boldsymbol{\alpha}}, \boldsymbol{h}^*_{\boldsymbol{\beta}}, \boldsymbol{h}^*_{\boldsymbol{\gamma}})$ satisfies that $E_0\{l^*(\boldsymbol{\theta}|\mathcal{O})\cdot l_2(\boldsymbol{\theta}|\mathcal{O})[\boldsymbol{h}]\} = 0$. For the sake of convenience, we only give a specified description for the calculation of $\boldsymbol{h}^*_{\boldsymbol{\alpha}}$. Denote

$D(S) = D_1(S) \cdot S'_{\boldsymbol{\xi}}(R)(\boldsymbol{h}) + D_2(S) \cdot S'_{\boldsymbol{\xi}}(L)(\boldsymbol{h})$, we have that

$$
\begin{aligned}
&E_0\Bigg[\Big\{D_1(S) \cdot [S'_{\boldsymbol{\alpha}}(R) - S'_{\boldsymbol{\xi}}(R)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})] + D_2(S) \cdot [S'_{\boldsymbol{\alpha}}(L) - S'_{\boldsymbol{\xi}}(L)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})]\Big\} \cdot D(S)\Bigg] \\
&= E_0\Bigg[D_1(S)D(S) \cdot [S'_{\boldsymbol{\alpha}}(R) - S'_{\boldsymbol{\xi}}(R)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})]\Bigg] + E_0\Bigg[D_2(S)D(S) \cdot [S'_{\boldsymbol{\alpha}}(L) - S'_{\boldsymbol{\xi}}(L)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})]\Bigg].
\end{aligned}
$$

Denote E_{J1} as the expectation with the density function $C_{J1}^{-1} D_1(S)D(S) \cdot l(\boldsymbol{\theta}|\mathcal{O})$, where $C_{J1} = E_0[D_1(S)D(S)]$, and denote E_{J2} as the expectation with the density function $C_{J2}^{-1} D_2(S)D(S) \cdot l(\boldsymbol{\theta}|\mathcal{O})$, where $C_{J2} = E_0[D_2(S)D(S)]$. Thus,

$$
\begin{aligned}
&E_0\Bigg[D_1(S)D(S) \cdot [S'_{\boldsymbol{\alpha}}(R) - S'_{\boldsymbol{\xi}}(R)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})]\Bigg] + E_0\Bigg[D_2(S)D(S) \cdot [S'_{\boldsymbol{\alpha}}(L) - S'_{\boldsymbol{\xi}}(L)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})]\Bigg] \\
&= E_{J1}[S'_{\boldsymbol{\alpha}}(R) - S'_{\boldsymbol{\xi}}(R)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})] + E_{J2}[S'_{\boldsymbol{\alpha}}(L) - S'_{\boldsymbol{\xi}}(L)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})].
\end{aligned}
$$

To solve the equation $E_{J1}[S'_{\boldsymbol{\alpha}}(R) - S'_{\boldsymbol{\xi}}(R)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})] + E_{J2}[S'_{\boldsymbol{\alpha}}(L) - S'_{\boldsymbol{\xi}}(L)(\boldsymbol{h}^*_{\boldsymbol{\alpha}})] = 0$, we obtain that $\boldsymbol{h}^*_{\boldsymbol{\alpha}}(t, x, \boldsymbol{\beta}_0) = \frac{S'_{\boldsymbol{\alpha}}(t)}{S'_{\boldsymbol{\xi}}(t)} = \frac{(1-\pi) \cdot (S(t)-1) \cdot \boldsymbol{W}}{S(t)\log(S(t))}$. Following the same procedure, we can obtain that one choice for $\boldsymbol{h}^*$ is

$$
\begin{aligned}
\boldsymbol{h}^*_{\boldsymbol{\alpha}}(t, x, \boldsymbol{\beta}_0) &= \frac{S'_{\boldsymbol{\alpha}}(t)}{S'_{\boldsymbol{\xi}}(t)} = \frac{(1-\pi) \cdot (S(t)-1) \cdot \boldsymbol{W}}{S(t)\log(S(t))}, \\
\boldsymbol{h}^*_{\boldsymbol{\beta}}(t, x, \boldsymbol{\beta}_0) &= \frac{S'_{\boldsymbol{\beta}}(t)}{S'_{\boldsymbol{\xi}}(t)} = \frac{\phi'(te^{\boldsymbol{X}^T\boldsymbol{\beta}}) \cdot te^{\boldsymbol{X}^T\boldsymbol{\beta}} \cdot \boldsymbol{X}}{\pi}, \\
\boldsymbol{h}^*_{\boldsymbol{\gamma}}(t, x, \boldsymbol{\beta}_0) &= \frac{S'_{\boldsymbol{\gamma}}(t)}{S'_{\boldsymbol{\xi}}(t)} = \frac{\boldsymbol{Z}^T\boldsymbol{\gamma} \cdot \boldsymbol{Z}}{\pi S(t)\log(S(t))}.
\end{aligned}
$$

The following equalities can be verified based on the zero-mean property of the score function:

$$
\begin{aligned}
&Pl''_{\boldsymbol{\tau}\xi}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}] = -P\{l'_{\boldsymbol{\tau}}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O}) l'^T_{\xi}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}]\}, \\
&Pl''_{\xi\boldsymbol{\tau}}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}] = -P\{l'_{\xi}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}] l'^T_{\boldsymbol{\tau}}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})\}, \\
&Pl''_{\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O}) = -P\{l'_{\boldsymbol{\tau}}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O}) l'^T_{\boldsymbol{\tau}}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})\}, \\
&Pl''_{\xi\xi}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}_1, \boldsymbol{h}_2] = -P\{l'_{\xi}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}_1] l'^T_{\xi}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h}_2]\}.
\end{aligned}
$$

Together with the fact that

$$Pl''_{\boldsymbol{\tau}\xi}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h^*}] - Pl''_{\xi\xi}(\boldsymbol{\tau_0}, \xi(\cdot, \boldsymbol{\beta_0})|\mathcal{O})[\boldsymbol{h^*}, \boldsymbol{h^*}] = 0,$$

we can show that the matrix

$$\begin{aligned}
&P\big\{ - l''_{\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta_0}|\mathcal{O}) + l''_{\xi\boldsymbol{\tau}}(\boldsymbol{\theta_0}|\mathcal{O})[\boldsymbol{h}^*] + l''_{\boldsymbol{\tau}\xi}(\boldsymbol{\theta_0}|\mathcal{O})[\boldsymbol{h}^*] - l''_{\xi\xi}(\boldsymbol{\theta_0}|\mathcal{O})[\boldsymbol{h}^*, \boldsymbol{h}^*]\big\} \\
&= P\{l'_{\boldsymbol{\tau}}(\boldsymbol{\theta_0}|\mathcal{O}) l'^T_{\boldsymbol{\tau}}(\boldsymbol{\theta_0})|\mathcal{O}) - l'_{\xi}(\boldsymbol{\theta_0}|\mathcal{O})[\boldsymbol{h}^*] l'^T_{\boldsymbol{\tau}}(\boldsymbol{\theta_0}|\mathcal{O}) \\
&- l'_{\boldsymbol{\tau}}(\boldsymbol{\theta_0}|\mathcal{O}) l'^T_{\xi}(\boldsymbol{\theta_0})|\mathcal{O})[\boldsymbol{h}^*] + l'_{\boldsymbol{\tau}}(\boldsymbol{\theta_0}|\mathcal{O})[\boldsymbol{h}^*] l'^T_{\xi}(\boldsymbol{\theta_0}|\mathcal{O})[\boldsymbol{h}^*]\} \\
&= P\{l'_{\boldsymbol{\tau}}(\boldsymbol{\theta_0}|\mathcal{O}) - l'_{\xi}(\boldsymbol{\theta_0}|\mathcal{O})[\boldsymbol{h}^*]\}^{\otimes 2} \\
&= Pl^*_{\boldsymbol{\tau_0}}(\mathcal{O})^{\otimes 2}
\end{aligned}$$

is non-singular.

For assumption B4, we have $P_n l'_{\boldsymbol{\tau}}(\hat{\boldsymbol{\tau}}_n, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_n)|\mathcal{O}) \equiv 0$ since $\hat{\boldsymbol{\theta}}_n$ is the sieve ML estimator. Another condition we need to verify is $P_n l'_{\xi}(\hat{\boldsymbol{\tau}}_n, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_n)|\mathcal{O})[\boldsymbol{h}^*] = o_p(n^{-1/2})$. It is clear that $P_n l'_{\xi}(\hat{\boldsymbol{\tau}}_n, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_{\boldsymbol{n}})|\mathcal{O})[\boldsymbol{h}^*_n] = 0$, which is the directional derivative for $l(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}|\mathcal{O})$ along $\boldsymbol{h}^*_n$ at $\hat{\phi}_n$. To verify $P_n l'_{\xi}(\hat{\boldsymbol{\tau}}_n, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_n)|\mathcal{O})[\boldsymbol{h}^*] = o_p(n^{-1/2})$, it is sufficient to prove $P_n l'_{\xi}(\hat{\boldsymbol{\tau}}_n, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_n)|\mathcal{O})[h^*_j] = o_p(n^{-1/2})$ for any jth component of $\boldsymbol{h}^*$. For the component h^*_j, there is a $h^*_{j,n} \in \mathcal{H}_n$ satisfying that $\| h^*_j - h^*_{j,n} \| = O(n^{-v})$, thus we have $Pl'_{\xi}(\hat{\boldsymbol{\tau}}_0, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_0)|\mathcal{O})[h^*_j - h^*_{j,n}] = 0$. We rewrite $P_n l'_{\xi}(\hat{\boldsymbol{\tau}}_n, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_n)|\mathcal{O})[h^*_j] = I_{1,n} + I_{2,n}$, where

$$I_{1,n} = (P_n - P) l'_{\xi}(\hat{\boldsymbol{\tau}}_n, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_n)|\mathcal{O})[h^*_j - h^*_{j,n}],$$

and

$$I_{2,n} = P\big\{l'_{\xi}(\hat{\boldsymbol{\tau}}_n, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_n)|\mathcal{O})[h^*_j - h^*_{j,n}] - l'_{\xi}(\hat{\boldsymbol{\tau}}_0, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_0)|\mathcal{O})[h^*_j - h^*_{j,n}]\big\}.$$

Let $\mathcal{L}_3 = \{l'_{\xi}(\boldsymbol{\theta}|\mathcal{O})[h^*_j - h_j] : \boldsymbol{\theta} \in \Theta_n, h_j \in \mathcal{H}_n, \text{ and } \| h^*_j - h_j \| \leq n^{-v}\}$. It

can be proved that the ϵ-bracketing number of $\mathcal{L}_3$ is bounded by $C(1/\epsilon)^d(1/\epsilon)^{Cq_n}$, utilising similar arguments regrading $\mathcal{L}_1$. Since $l'_\xi(\hat{\boldsymbol{\tau}}_n, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_n)|\mathcal{O})\ [h_j^* - h_{j,n}^*] \in \mathcal{L}_3$, and $P\big\{l'_\xi(\hat{\boldsymbol{\tau}}_n, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_n)|\mathcal{O})[h_j^* - h_{j,n}^*]\big\}^2 \to 0$ as $n \to \infty$, we have $\mathcal{L}_3$ is a P-Donsker. According to Corollary 2.3.12 of Vaart and Wellner (1996), we have $I_{1,n} = o_p(n^{-1/2})$.

By algebraic calculations and the Cauchy-Schwarz inequality, we have that

$$\begin{aligned} I_{2,n} &= P\big\{l'_\xi(\hat{\boldsymbol{\tau}}_n, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_n)|\mathcal{O})[h_j^* - h_{j,n}^*] - l'_\xi(\hat{\boldsymbol{\tau}}_0, \hat{\xi}(\cdot, \hat{\boldsymbol{\beta}}_0)|\mathcal{O})[h_j^* - h_{j,n}^*]\big\} \\ &\le Cd(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}, \boldsymbol{\theta}_0) \parallel h_j^* - h_{j,n}^* \parallel_\infty = O(n^{-\min(pv,(1-v)/2)} n^{-v}) \\ &= O(n^{-\min((p+1)v,(1+v)/2)}) = o_p(n^{-1/2}). \end{aligned}$$

Hence assumption B4 holds.

Assumption B5 can be verified by P-Donsker classes. Define the class $\mathcal{L}_4(\eta) = \{l'_{\boldsymbol{\tau}}(\boldsymbol{\tau}, \xi(\cdot, \boldsymbol{\beta})|\mathcal{O}) - l'_{\boldsymbol{\tau}_0}(\boldsymbol{\tau}_0, \xi(\cdot, \boldsymbol{\beta}_0)|\mathcal{O}) : \boldsymbol{\theta} \in \Theta_n, d(\boldsymbol{\theta}, \boldsymbol{\theta_0}) \le \eta\}$. The ϵ-bracketing number for $\mathcal{L}_4(\eta)$ is bounded by $(\eta/\epsilon)^{Cq_n+d}$. We choose $\eta = O(n^{-\min(2v,(1-v)/2)})$. Since $l'_{\boldsymbol{\tau}}(\boldsymbol{\tau}, \xi(\cdot, \boldsymbol{\beta})|\mathcal{O}) - l'_{\boldsymbol{\tau}}(\boldsymbol{\tau}_0, \xi(\cdot, \boldsymbol{\beta}_0)|\mathcal{O}) \in \mathcal{L}_4(\eta)$, the convergence rate of $\hat{\boldsymbol{\theta}}_n$ leads to that as $n \to \infty$

$$P\big\{l'_{\boldsymbol{\tau}}(\boldsymbol{\tau}, \xi(\cdot, \boldsymbol{\beta})|\mathcal{O}) - l'_{\boldsymbol{\tau}}(\boldsymbol{\tau}_0, \xi(\cdot, \boldsymbol{\beta}_0)|\mathcal{O})\big\}^2 \to 0.$$

Hence, $\mathcal{L}_4(\eta)$ is a P-Donsker. Similarly, $\mathcal{L}_5(\eta) = \{l'_\xi(\boldsymbol{\tau}, \xi(\cdot, \boldsymbol{\beta})|\mathcal{O}) - l'_\xi(\boldsymbol{\tau}_0, \xi(\cdot, \boldsymbol{\beta}_0)|\mathcal{O}) : \boldsymbol{\theta} \in \Theta_n, d(\theta, \boldsymbol{\theta_0}) \le \eta\}$ is a P-Donsker and for any $\rho(\boldsymbol{\theta}, \xi; z) \in \mathcal{L}_5(\eta)$, we can prove that $P\rho^2 \to 0$. Thus assumption B5 can be verified.

Assumption B6 is verified by the Taylor expansion. The Taylor expansion for $l'_{\boldsymbol{\tau}}(\boldsymbol{\theta}|\mathcal{O})$ is

$$\begin{aligned} l'_{\boldsymbol{\tau}}(\boldsymbol{\theta}|\mathcal{O}) &= l'_{\tau}(\boldsymbol{\theta_0}|\mathcal{O}) + l''_{\tau\tau}(\widetilde{\boldsymbol{\theta}}|\mathcal{O})(\boldsymbol{\tau}-\boldsymbol{\tau}_0) + l''_{\boldsymbol{\tau}\xi}(\widetilde{\boldsymbol{\theta}}|\mathcal{O})[\xi(\cdot,\boldsymbol{\beta}) - \xi_0(\cdot,\boldsymbol{\beta}_0)] \\ &= l'_{\tau}(\boldsymbol{\theta_0}|\mathcal{O}) + l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\theta_0}|\mathcal{O})(\boldsymbol{\tau}-\boldsymbol{\tau}_0) + l''_{\boldsymbol{\tau}\xi}(\boldsymbol{\theta_0}|\mathcal{O})[\xi(\cdot,\boldsymbol{\beta}) - \xi_0(\cdot,\boldsymbol{\beta}_0)] \\ &+ \{l''_{\boldsymbol{\tau\tau}}(\widetilde{\boldsymbol{\theta}}|\mathcal{O})(\boldsymbol{\tau}-\boldsymbol{\tau}_0) - l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\theta_0}|\mathcal{O})(\boldsymbol{\tau}-\boldsymbol{\tau}_0)\} \\ &+ \{l''_{\boldsymbol{\tau}\xi}(\widetilde{\boldsymbol{\theta}}|\mathcal{O})[\xi(\cdot,\boldsymbol{\beta}) - \xi_0(\cdot,\boldsymbol{\beta}_0)] - l''_{\boldsymbol{\tau}\xi}(\boldsymbol{\theta_0}|\mathcal{O})[\xi(\cdot,\boldsymbol{\beta}) - \xi_0(\cdot,\boldsymbol{\beta}_0)]\}, \end{aligned}$$

where $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\tau}}, \widetilde{\xi}(\cdot,\widetilde{\boldsymbol{\beta}}))$ is the point between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. Thus

$$\begin{aligned} &P\{l'_{\boldsymbol{\tau}}(\boldsymbol{\theta}|\mathcal{O}) - l'_{\boldsymbol{\tau}}(\boldsymbol{\theta_0}|\mathcal{O}) - l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\theta_0}|\mathcal{O})(\boldsymbol{\tau}-\boldsymbol{\tau}_0) - l''_{\boldsymbol{\tau}\xi}(\boldsymbol{\theta_0}|\mathcal{O})[\xi(\cdot,\boldsymbol{\beta}) - \xi_0(\cdot,\boldsymbol{\beta}_0)]\} \\ &= P\{[l''_{\boldsymbol{\tau\tau}}(\widetilde{\boldsymbol{\theta}}|\mathcal{O}) - l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\theta_0}|\mathcal{O})](\boldsymbol{\tau}-\boldsymbol{\tau}_0)\} \\ &+ P\{l''_{\boldsymbol{\tau}\xi}(\widetilde{\boldsymbol{\theta}}|\mathcal{O})[\xi(\cdot,\boldsymbol{\beta}) - \xi_0(\cdot,\boldsymbol{\beta})] - l''_{\boldsymbol{\tau}\xi}(\boldsymbol{\theta_0}|O)[\xi(\cdot,\boldsymbol{\beta}) - \xi_0(\cdot,\boldsymbol{\beta})]\}. \end{aligned}$$

Under conditions C3, the first and second derivatives of log-likelihood $l(\boldsymbol{\tau}, \xi(\cdot,\boldsymbol{\beta})|\mathcal{O})$ are bounded and continuous with respect to $\boldsymbol{\tau} \in \mathcal{T}$ and $\xi(\cdot,\boldsymbol{\beta}) \in \mathcal{H}^p$. Denote $\mathcal{L}_6(\eta) = \{l''_{\boldsymbol{\tau\tau}}(\widetilde{\boldsymbol{\theta}}|\mathcal{O}) - l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\theta_0}|\mathcal{O}) : \widetilde{\boldsymbol{\theta}} \in \Theta_n, d(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta_0}) \leq \eta\}$. By choosing $\eta = n^{-\min(pv,(1-v)/2)}$, we can verify that $\mathcal{L}_6(\eta)$ is a P-Donsker. Moreover, for any $0 < \epsilon < 1/2 - \min(pv, (1-v)/2)$, we have

$$\begin{aligned} &P\{[l''_{\boldsymbol{\tau\tau}}(\widetilde{\boldsymbol{\theta}}|\mathcal{O}) - l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\theta_0}|\mathcal{O})](\boldsymbol{\tau}-\boldsymbol{\tau}_0)\} \\ &= n^{-\min(pv,(1-v)/2)+\epsilon} P\left\{[l''_{\boldsymbol{\tau\tau}}(\widetilde{\boldsymbol{\theta}}|\mathcal{O}) - l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\theta_0}|\mathcal{O})]\frac{\boldsymbol{\tau}-\boldsymbol{\tau}_0}{n^{-\min(pv,(1-v)/2)+\epsilon}}\right\}. \end{aligned}$$

Since $d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta_0}) = O(n^{-\min(pv,(1-v)/2)})$ and

$$P\left\{[l''_{\boldsymbol{\tau\tau}}(\widetilde{\boldsymbol{\theta}}|\mathcal{O}) - l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\theta_0}|\mathcal{O})]\frac{\boldsymbol{\tau}-\boldsymbol{\tau}_0}{n^{-\min(pv,(1-v)/2)+\epsilon}}\right\}^2 \to 0,$$

together with the result of $\mathcal{L}_6$ being a Donsker, we can obtain that $P\Big\{[l''_{\boldsymbol{\tau\tau}}(\widetilde{\boldsymbol{\theta}}|\mathcal{O}) -$

$l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\theta_0}|\mathcal{O})]\frac{\boldsymbol{\tau}-\boldsymbol{\tau}_0}{n^{-\min(pv,(1-v)/2)+\epsilon}}\Big\} = o_p(n^{-1/2})$, by using Corollary 2.3.12 of Vaart and Wellner (1996) again. Thus

$$P\{[l''_{\boldsymbol{\tau\tau}}(\widetilde{\boldsymbol{\theta}}|\mathcal{O})-l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\theta_0}|\mathcal{O})](\boldsymbol{\tau}-\boldsymbol{\tau}_0)\} = O_p(n^{-\min(pv,(1-v)/2)+\epsilon}n^{-1/2}) = O_p(n^{-2\min(pv,(1-v)/2)}).$$

In a similar way, we can prove that

$$P\{l''_{\boldsymbol{\tau}\xi}(\widetilde{\boldsymbol{\theta}}|\mathcal{O})[\xi(\cdot,\boldsymbol{\beta})-\xi_0(\cdot,\boldsymbol{\beta}_0)]-l''_{\tau\xi}(\boldsymbol{\theta_0}|O)[\xi(\cdot,\boldsymbol{\beta})-\xi_0(\cdot,\boldsymbol{\beta}_0)]\} = O_p(n^{-2\min(pv,(1-v)/2)}).$$

Therefore, we have

$$P\{l'_{\boldsymbol{\tau}}(\boldsymbol{\theta}|\mathcal{O})-l'_{\boldsymbol{\tau}}(\boldsymbol{\theta_0}|\mathcal{O})-l''_{\boldsymbol{\tau\tau}}(\boldsymbol{\theta_0}|\mathcal{O})(\boldsymbol{\tau}-\boldsymbol{\tau}_0)-l''_{\boldsymbol{\tau}\xi}(\boldsymbol{\theta_0}|\mathcal{O})[\xi(\cdot,\boldsymbol{\beta})-\xi_0(\cdot,\boldsymbol{\beta}_0)]\} = O(n^{-\iota\varrho}),$$

where $\varrho = \min(pv,(1-v)/2)$, $\iota = 2 > 1$, and $\frac{1}{2} < 1-\frac{1}{1+p} < 2pv < 1, 1-\frac{1}{2p} < 1-v < 1-\frac{1}{2(1+p)}$ based on the restriction that $p \geq 2$. Thus the first assumption of B6 holds. The second assumption can be verified by employing the same techniques, and we omit the proof of the second condition here.

Till now, we have verified all six assumptions. By Theorem 2.1 of Ding and Nan (2011), we can obtain the asymptotical normality of the proposed estimators. That is

$$n^{1/2}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0) = A^{-1}\sqrt{n}\mathbb{P}_n l^*(\boldsymbol{\tau}_0, \xi_0(\cdot,\boldsymbol{\beta}_0)|\mathcal{O}) + o_p(1) \xrightarrow{d} N(0, I(\boldsymbol{\tau}_0)^{-1}),$$

where $l^*(\boldsymbol{\tau}_0, \xi_0(\cdot,\boldsymbol{\beta}_0)|\mathcal{O}) = l'_{\boldsymbol{\tau}}(\boldsymbol{\beta}_0, \xi_0(\cdot,\boldsymbol{\beta}_0)|\mathcal{O}) - l'_{\xi}(\boldsymbol{\tau}_0, \xi_0(\cdot,\boldsymbol{\beta}_0)|\mathcal{O})[\boldsymbol{h}^*]$ is the efficient score function for $\boldsymbol{\tau}$ and $A = P\{l^*(\boldsymbol{\tau}_0, \xi_0(\cdot,\boldsymbol{\beta}_0)|\mathcal{O})^{\otimes 2}\}$. Since $I(\boldsymbol{\tau}_0) = P\{l^*(\boldsymbol{\tau}_0, \xi_0(\cdot,\boldsymbol{\beta}_0)|\mathcal{O})^{\otimes 2}\}$, we have the $n^{1/2}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0)$ converges to a normal distribution with covariance matrix $I(\boldsymbol{\tau}_0)^{-1}$.

Chapter 4

Discussion and Future Research

The primary contribution of this thesis is extending the conventional semiparametric mixture cure models to incorporate nonlinear covariate effects in both the latency and incidence parts and allow time-scaled effects on the hazard function. This leads to a flexible modeling framework for survival data in the presence of a cure fraction and interval censoring. We developed the estimation procedures from both the Bayesian and frequency perspectives and provided the theoretical justification for the proposed sieve maximum likelihood estimators using empirical process techniques, which was challenging in the setting of interval-censored data because of the bundled parameters.

First, semiparametric models were specified for the incidence and latency parts in a PH-based mixture cure model. A Bayesian approach was employed for inference of this doubly semiparametric mixture cure model. The use of spline approximations and a two-stage data augmentation enabled conjugate posterior distributions of parameters to be obtained when proper prior distributions were specified. To reduce the computation burden, the rejection algorithm, the Metropolis algorithm, and the composition method were utilized.

Next, a general class of AH models for the survival function in the latency part was considered. The sieve maximum likelihood estimation approach was adopted

for the estimation. The most difficult part of the sieve maximum likelihood estimation was to deal with the infinite-dimensional unknown cumulative baseline hazard function and the bundled parameters, which were treated using the B-spline approximation and the proposed two-step iterative algorithm. Simulation results and real data analyses were provided to illustrate the effectiveness and practical utilities of two proposed models.

There are some limitations of the proposed methods in this thesis. For the Bayesian doubly semiparametric mixture cure model, the proposed inference method is somehow computationally intensive. Kim and Pavlovic (2018) proposed variational inference algorithms for the PH model with right censored data, which reduce the complexity of inference and improve scalability to large datasets. Motivating by their work, to develop a variational Bayesian method for the mixture cure model with interval-censored data would have potentials to improve computational efficiency of the inference procedure proposed in Chapter 2. It is worthwhile for study further.

The primary challenge with the generalized accelerated cure model developed in Chapter 3 is to assess its goodness-of-fit. The residual-based approach, explored by Farrington (2000) and Peng and Taylor (2017), may be extended to evaluate the goodness of model fitting of the GAHCure model in the presence of interval-censored data. This will be another direction of our future research.

Extending the proposed approaches to model clustered interval-censored data and time-varying regression coefficients should be possible, by introducing a frailty term in the model and using the splines method to approximate the time-dependent coefficients, respectively. The implementation of statistical inference will be similar to the proposed algorithms, while the updates of the frailty term and regression coefficients are more complicated. We will put our effort into such extensions of our work in future research.

Bibliography

Amico, M., Van Keilegom, I., and Legrand, C. (2018). The single-index/cox mixture cure model. *Biometrics*. 22

Balakrishnan, N. and Pal, S. (2013). Lognormal lifetimes and likelihood-based inference for flexible cure rate models based on com-poisson family. *Computational Statistics & Data Analysis*, 67:41–67. 12

Banerjee, S. and Carlin, B. P. (2004). Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics*, 60(1):268–275. 13, 15, 58, 59, 60

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in medicine*, 2(2):273–277. 7

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47(259):501–515. 5

Betensky, R. A., Lindsey, J. C., Ryan, L. M., and Wand, M. (2002). A local likelihood proportional hazards model for interval censored data. *Statistics in Medicine*, 21(2):263–275. 7

Betensky, R. A., Rabinowitz, D., and Tsiatis, A. A. (2001). Computationally simple accelerated failure time regression for interval censored data. *Biometrika*, 88(3):703–711. 9

Brinkley, D. and Haybittle, J. (1975). The curability of breast cancer. *The Lancet*, 306(7925):95–97. 5

Cai, T. and Betensky, R. A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics*, 59(3):570–579. 7

Chen, M.-H., Ibrahim, J. G., and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94(447):909–919. 15

Chen, Y. Q. and Jewell, N. P. (2001). On a general class of semiparametric hazards regression models. *Biometrika*, 88(3):687–702. 10, 63

Chen, Y. Q. and Wang, M.-C. (2000). Analysis of accelerated hazards models. *Journal of the American Statistical Association*, 95(450):608–618. 9

Cheng, G., Huang, J. Z., et al. (2010). Bootstrap consistency for general semiparametric m-estimation. *The Annals of Statistics*, 38(5):2884–2915. 55

Collett, D. (1994). *Modelling survival data in medical research*. Chapman and Hall/CRC. 8

Conkin, J. and Powell, M. R. (2001). Lower body adynamia as a factor to reduce the risk of hypobaric decompression sickness. *Aviation, space, and environmental medicine*, 72(3):202–214. 4, 40

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202. 6

Damlen, P., Wakefield, J., and Walker, S. (1999). Gibbs sampling for bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):331–344. 31

Dey, D. D., Müller, P., and Sinha, D. (2012). *Practical nonparametric and semiparametric Bayesian statistics*, volume 133. Springer Science & Business Media. 15

Ding, Y. and Nan, B. (2011). A sieve m-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *Annals of statistics*, 39(6):2795. 54, 69, 77

Dinse, G. E. and Series, S. L. o. t. R. S. S. (1984). Regression Analysis of Tumour Prevalence Data. 8

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, pages 1041–1046. 5

Farewell, V. T. (1986). Mixture models in survival analysis: Are they worth the risk? *Canadian Journal of Statistics*, 14(3):257–262. 12

Farrington, C. (2000). Residuals for proportional hazards models with interval-censored survival data. *Biometrics*, 56(2):473–482. 61, 79

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, 42(4):845–854. 6

Geman, S. and Hwang, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *The Annals of Statistics*, pages 401–414. 16

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *Applied Statistics*, pages 337–348. 29, 31

Grenander, U. (1981). Abstract inference. Technical report. 16

Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*, volume 19. Springer Science & Business Media. 4

Gu, Y., Sinha, D., and Banerjee, S. (2011). Analysis of cure rate survival data under proportional odds model. *Lifetime data analysis*, 17(1):123–134. 12

Hu, T. and Xiang, L. (2013). Efficient estimation for semiparametric cure models with interval-censored data. *Journal of Multivariate Analysis*, 121:139–151. 16

Huang, J. et al. (1999). Efficient estimation of the partly linear additive cox model. *The annals of Statistics*, 27(5):1536–1563. 7

Huang, J. and Rossini, A. (1997). Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*, 92(439):960–967. 8, 16, 17

Huang, J. and Wellner, J. A. (1995). Efficient estimation for the proportional hazards model with" case 2" interval censoring. 7, 15

Huang, J. and Wellner, J. A. (1997). Interval censored survival data: a review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics*, pages 123–169. Springer. 4, 8, 9

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons. 2

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481. 4

Kim, M. and Pavlovic, V. (2018). Variational inference for gaussian process models for survival analysis. In *UAI*, pages 435–445. 79

Kim, Y.-J. and Jhun, M. (2008). Cure rate model with interval censored data. *Statistics in medicine*, 27(1):3–14. 1, 12

Kooperberg, C. and Stone, C. J. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, 1(4):301–328. 5

Kuk, A. Y. and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79(3):531–541. 12

Lam, K. and Wong, K.-Y. (2014). Semiparametric analysis of clustered interval-censored survival data with a cure fraction. *Computational Statistics & Data Analysis*, 79:165–174. 13

Lam, K. and Xue, H. (2005). A semiparametric regression cure model with current status data. *Biometrika*, 92(3):573–586. 12, 16

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons. 2, 5

Lei, N. (2013). Flexible partially linear single index regression models for multivariate survival data. 41

Li, C.-S., Taylor, J. M., and Sy, J. P. (2001). Identifiability of cure models. *Statistics & Probability Letters*, 54(4):389–395. 62

Li, J. and Ma, S. (2010). Interval-censored data with repeated measurements and a cured subgroup. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(4):693–705. 12, 40, 41

Lin, X., Cai, B., Wang, L., and Zhang, Z. (2015). A bayesian proportional hazards model for general interval-censored data. *Lifetime data analysis*, 21(3):470–490. 7, 13, 15, 26

Lu, M., Zhang, Y., and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika*, 94(3):705–718. 1, 66

Ma, L., Hu, T., and Sun, J. (2015). Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika*, 102(3):731–738. 49

Ma, S. (2009). Cure model with current status data. *Statistica Sinica*, pages 233–249. 1, 12

Ma, S. (2010). Mixed case interval censored data with a cured subgroup. *Statistica Sinica*, pages 1165–1181. 13

Ma, S. and Kosorok, M. R. (2005). Robust semiparametric m-estimation and the weighted bootstrap. *Journal of Multivariate Analysis*, 96(1):190–217. 54

Murphy, S., Rossini, A., and van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, 92(439):968–976. 8

Murray, R. P., Anthonisen, N. R., Connett, J. E., Wise, R. A., Lindgren, P. G., Greene, P. G., Nides, M. A., Group, L. H. S. R., et al. (1998). Effects of multiple attempts to quit smoking and relapses to smoking on pulmonary function. *Journal of clinical epidemiology*, 51(12):1317–1326. 58

Odell, P. M., Anderson, K. M., and D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a weibull-based accelerated failure time model. *Biometrics*, pages 951–959. 9

Oller, R., Gómez, G., and Calle, M. L. (2004). Interval censoring: model characterizations for the validity of the simplified likelihood. *Canadian Journal of Statistics*, 32(3):315–326. 62

Oller, R., Gómez, G., and Calle, M. L. (2007). Interval censoring: identifiability and the constant-sum property. *Biometrika*, 94(1):61–70. 62

Pan, C., Cai, B., and Wang, L. (2017). Multiple frailty model for clustered interval-censored data with frailty selection. *Statistical methods in medical research*, 26(3):1308–1322. 43

Pan, W. (2000). A multiple imputation approach to cox regression with interval-censored data. *Biometrics*, 56(1):199–203. 7

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686. 29

Peng, Y. and Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, 56(1):237–243. 12

Peng, Y. and Taylor, J. M. (2014). Cure models. *Handbook of Survival Analysis*, pages 113–134. 12

Peng, Y. and Taylor, J. M. (2017). Residual-based model diagnosis methods for mixture cure models. *Biometrics*, 73(2):495–505. 61, 79

Poon, W.-Y. and Wang, H.-B. (2013). Bayesian analysis of generalized partially linear single-index models. *Computational Statistics and Data Analysis*, 68:251–261. 23

Rabinowitz, D., Betensky, R. A., and Tsiatis, A. A. (2000). Using conditional logistic regression to fit proportional odds models to interval censored data. *Biometrics*, 56(2):511–518. 8

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical science*, pages 425–441. 25

Rasmussen, S., Glümer, C., Sandbaek, A., Lauritzen, T., and Borch-Johnsen, K. (2008). Determinants of progression from impaired fasting glucose and impaired glucose tolerance to diabetes in a high-risk screened population: 3 year follow-up in the addition study, denmark. *Diabetologia*, 51(2):249–257. 9

Rossini, A. J., Statistical, A. T. J. o. t. A., and 1996 (1996). A semiparametric proportional odds regression model for the analysis of current status data. *Taylor & Francis*. 8, 16

Satten, G. A. (1996). Rank-based inference in the proportional hazards model for interval censored data. *Biometrika*, 83(2):355–370. 6

Satten, G. A., Datta, S., and Williamson, J. M. (1998). Inference based on imputed failure times for the proportional hazards model with interval-censored data. *Journal of the American Statistical Association*, 93(441):318–327. 7

Shao, F., Li, J., Ma, S., and Lee, M.-L. T. (2014). Semiparametric varying-coefficient model for interval censored data with a cured proportion. *Statistics in Medicine*, 33(10):1700–1712. 20

Shen, X. (1998). Propotional odds regression and sieve maximum likelihood estimation. *Biometrika*, 85(1):165–177. 16

Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *The Annals of Statistics*, pages 580–615. 54, 64

Sparling, Y. H., Younes, N., Lachin, J. M., and Bautista, O. M. (2006). Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics*, 7(4):599–614. 12

Sun, J. (2007). *The statistical analysis of interval-censored failure time data*. Springer Science & Business Media. 2, 5

Sy, J. P. and Taylor, J. M. (2000). Estimation in a cox proportional hazards cure model. *Biometrics*, 56(1):227–236. 1, 12

Szabo, Z., Liu, X., and Xiang, L. (2019). Semiparametric sieve maximum likelihood estimation for accelerated hazards model with interval-censored data. *Journal of Statistical Planning and Inference*. 46, 49

Tanner, M. A. (1991). *Tools for statistical inference*, volume 3. Springer. 32

Thompson, L. A. and Chhikara, R. (2003). A bayesian cure rate model for repeated measurements and interval censoring. *Proceedings of JSM*. 40, 41

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, 82(398):559–567. 5

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 290–295. 4

Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer. 63, 65, 67, 75, 77

Van der Vaart, A. (1998). Asymptotic statistics. 63, 64, 66, 67, 68, 69

van der Vaart, A. W. and Wellner, J. A. (1997). Weak convergence and empirical processes with applications to statistics. *Journal of the Royal Statistical Society-Series A Statistics in Society*, 160(3):596–608. 5, 7, 53

Wang, L., Du, P., and Liang, H. (2012). Two-component mixture cure rate model with spline estimated nonparametric components. *Biometrics*, 68(3):726–735. 20, 23

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association*, 92(439):945–959. 5

Wu, Y., Zhang, Y., et al. (2012). Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data. *The Annals of Statistics*, 40(3):1609–1636. 68

Xiang, L., Ma, X., and Yau, K. K. (2011). Mixture cure model with random effects for clustered interval-censored survival data. *Statistics in medicine*, 30(9):995–1006. 13, 35, 55, 56, 59

Yu, B. and Peng, Y. (2008). Mixture cure models for multivariate survival data. *Computational Statistics & Data Analysis*, 52(3):1524–1532. 58, 59

Zeng, D. and Lin, D. (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association*, 102(480):1387–1396. 9

Zeng, D., Yin, G., and Ibrahim, J. G. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, 101(474):670–684. 16

Zhang, J. and Peng, Y. (2009). Accelerated hazards mixture cure model. *Lifetime data analysis*, 15(4):455. 1, 10, 45

Zhang, J., Peng, Y., and Zhao, O. (2011). A new semiparametric estimation method for accelerated hazard model. *Biometrics*, 67(4):1352–1360. 10

Zhang, Y., Hua, L., and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the cox model with interval-censored data. *Scandinavian Journal of Statistics*, 37(2):338–354. 7, 16, 49

Zhang, Y., Wang, X., and Zhang, B. (2019). Bayesian approach for clustered interval-censored data with time-varying covariate effects. *Statistics and Its Interface*, 12(3):457–465. 43

Zhao, X., Wu, Y., Yin, G., et al. (2017). Sieve maximum likelihood estimation for a general class of accelerated hazards models with bundled parameters. *Bernoulli*, 23(4B):3385–3411. 10, 49

Zhou, J., Zhang, J., and Lu, W. (2018). Computationally efficient estimation for the generalized odds rate mixture cure model with interval-censored data. *Journal of Computational and Graphical Statistics*, 27(1):48–58. 36

Zhou, J., Zhang, J., McLain, A. C., and Cai, B. (2016). A multiple imputation approach for semiparametric cure model with interval censored data. *Computational Statistics & Data Analysis*, 99:105–114. 1, 13

Zhou, Q., Hu, T., and Sun, J. (2017). A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *Journal of the American Statistical Association*, 112(518):664–672. 49