Statistica Sinica Preprint No: SS-2022-0343

Title	Semiparametric Cure Regression Models with Informative Case K Interval-Censored Failure Time Data
Manuscript ID	SS-2022-0343
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0343
Complete List of Authors	Yichen Lou, Jianguo Sun and Peijie Wang
Corresponding Authors	Peijie Wang
E-mails	wangpeijie@jlu.edu.cn
Notice: Accepted version subject to English editing.	

Statistica Sinica

SEMIPARAMETRIC CURE REGRESSION MODELS WITH INFORMATIVE CASE K INTERVAL-CENSORED FAILURE TIME DATA

Yichen Lou[1], Jianguo Sun[2], Peijie Wang[1]

[1] *School of Mathematics, Jilin University, Changchun, 130012, China*

[2] *Department of Statistics, University of Missouri, Columbia, Missouri, 65211, U.S.A.*

Abstract: **Contents of the Abstract.**

Interval-censored failure time data occur in many areas and many methods for their analyses have been proposed. In particular, some methods have been developed for the situation with the existence of a cured subgroup or informative censoring. In this paper, we discuss the case where both a cured subgroup and informative censoring exist and a frailty-based semiparametric non-mixture cure model approach is proposed. For inference, a two-step estimation procedure is developed and the resulting estimators of regression parameters are shown to be consistent and asymptotically normal. An extensive simulation study is conducted and indicates that the proposed procedure works well in practice. In addition, the methodology is applied to a set of real data arising from an Alzheimer's disease study.

Key words and phrases: Informative censoring; interval-censored; non-mixture cure model; Bernstein polynomials

1. Introduction

Interval-censored failure time data occur when the failure time of interest is observed only to belong to some intervals instead of being observed exactly (Sun, 2006). It is easy to see that such data can happen naturally in many studies such as clinical trials or periodic follow-up studies and many methods have been developed for their analyses. Interval-censored data can have different forms and one general one is the so-called case K interval-censored data, the focus of this paper, meaning that there exists a sequence of observation times for each study subject. It is apparent that right-censored data can be seen as a special case of interval-censored data (Kalbfleisch and Prentice, 2002).

A typical underlying assumption in standard survival analysis is that if the follow-up is long enough, all subjects will experience the event of interest. However, this assumption may not apply in all cases. With rapid developments in medical and health sciences, it becomes increasingly common that a non-negligible proportion of subjects or patients are expected to be cured or have prolonged disease-free survival. In other words, there exists a cured subgroup. To address this, two types of models are commonly used. One is the two-component mixture cure model and the other is the non-mixture cure model. The former models the non-cured and cured subpopulations separately and treats the cured subjects to have the failure time being infinity (Sy and Taylor, 2000; Mao and Wang, 2010). The latter employs a single model for both subpopulations and has the advantage of giving a uniform model and a relatively easy interpretation (Chen et al., 1999). On the other hand, the inference for the latter may be more complicated and difficult.

In addition to a cured subgroup, dependent or informative interval censoring is another issue that one often has to deal with in the analysis of interval-censored failure time data. It means that the failure time of interest and the censoring mechanism or the observation process are correlated Huang and Wolfe (2002); Wang et al. (2016), and one such example is the periodic follow-up study of certain diseases where study patients do not follow the predetermined visit schedules but instead go to the clinic according to their disease status or how they are feeling. For this, two types of approaches are commonly used, the copula model-based approach and the latent or frailty model-based approach. Among others, Ma et al. (2016), Zhao et al. (2015) and Xu et al. (2019) gave some copula model-based methods under the Cox model, the additive hazards model, and the linear transformation model, respectively. Wang et al. (2016) and Zhang et al. (2007) discussed the use of the frailty model-based approach for the analysis of interval-censored data under the Cox model.

A great deal of literature has been established for the analysis of interval-censored data and in particular, some methods have been proposed for the situation where there exists a cured subgroup. For example, Lam and Xue (2005) and Ma (2009, 2010) developed some estimation procedures under the semiparametric mixture cure model for regression analysis of interval-censored data, while Hu and Xiang (2013) and Liu and Shen (2009) gave some methods for estimation of the non-mixture cure model based on interval-censored data. However, all of the methods above apply only to the case of independent interval censoring. Some authors have also considered the situation where there exists informative interval censoring but not a cured subgroup and they include Wang et al. (2016), Zhang et al. (2007), and Zhao et al. (2015).

There exists little literature for the analysis of interval-censored failure time data in the presence of both a cured subgroup and informative censoring except Wang et al. (2021), who discussed the situation under the mixture cure models. In the following, for the problem, we will present a class of semiparametric non-mixture cure models and propose a two-step spline-based sieve maximum likelihood estimation procedure. The proposed model is quite general and flexible and in the proposed method, the latent variable approach is employed to describe the relationship between the failure time of interest and the observation processes. Compared to some of the existing methods, the proposed approach is more general in that it does not need to impose any distribution assumptions on the latent variable and observation process.

The remainder of the paper is organized as follows. In Section 2, we will first introduce some notation and assumptions that will be used throughout the paper and then describe the structure of the observed data and the resulting likelihood function. The proposed two-step sieve maximum likelihood estimation procedure is presented in Section 3 and the estimators of the resulting regression parameters are shown to be consistent and asymptotically follow a normal distribution. In Section 4, a simulation study is conducted to examine the empirical performance of the proposed method, and the results suggest that it works well in practice. Section 5 applies the proposed approach to a set of real data arising from a study on Alzheimer's disease, and Section 6 contains some concluding remarks.

2. Notation and Assumptions

Consider a failure time study that consists of n independent subjects and let T_i denote the failure time of interest associated with subject i. For subject i, suppose that there exist a vector of covariates denoted by $\boldsymbol{Z}_i$ and a sequence of observation times points denoted by $U_{i0} = 0 < U_{i1} < U_{i2} < \cdots < U_{iK_i}$, where K_i is a random integer, $i = 1, \ldots, n$. Define $\tilde{N}_i(t) = \sum_{j=1}^{K_i} I(U_{ij} \leq t)$ and $\delta_{ij} = I(U_{ij-1} < T_i \leq U_{ij})$, $i = 1, \ldots, n,\ j = 1, \ldots, K_i$. Then $\tilde{N}_i(t)$ is a point process characterizing the ith subject's observation process and jumps only at the observation times. Also for subject i, suppose that there exists a follow-up time denoted by τ_i that is assumed to be independent of T_i and the observed data are given by

$$\left\{ O_i = (\tau_i, U_{ij}, \delta_{ij}, \boldsymbol{Z}_i,\ j = 1, \ldots, K_i),\ i = 1, \ldots, n \right\}.$$

That is, we only have case K interval-censored data.

To describe the covariate effect on T_i, suppose that there exists a latent variable u_i and given $\boldsymbol{Z}_i$ and u_i, the survival function of T_i has the form

$$S_i(t|\boldsymbol{Z}_i, u_i) = P(T_i \geq t|\boldsymbol{Z}_i, u_i) = \exp\left\{ -\exp(\boldsymbol{\beta}_1' \tilde{\boldsymbol{Z}}_i + \beta_2 u_i) F(t) \right\}. \quad (2.1)$$

In the above, $\tilde{\boldsymbol{Z}}_i = (1, \boldsymbol{Z}_i')'$, $F(t)$ is an unspecified cumulative distribution function, and $\boldsymbol{\beta}_1$ and β_2 are unknown regression parameters. That is, T_i follows a non-mixture cure model. In general, covariate may have some effects on the observation process $\tilde{N}_i(t)$ too. To address this, we assume that given $\boldsymbol{Z}_i$ and u_i, $\tilde{N}_i(t)$ satisfies the proportional rate model

$$E\{d\tilde{N}_i(t)|\boldsymbol{Z}_i, u_i\} = \lambda_{0h}(t)\exp(\boldsymbol{\alpha}'\boldsymbol{Z}_i + u_i)dt\,. \tag{2.2}$$

Here $\lambda_{0h}(t)$ denotes a completely unknown continuous baseline rate function and $\boldsymbol{\alpha}$ a vector of regression parameters as $\boldsymbol{\beta}_1$ and β_2. Also it will be assumed that given $\boldsymbol{Z}_i$ and u_i, the failure time of interest T_i and the observation process $\tilde{N}_i(t)$ are independent.

It is well-known that the latent variable approach is a commonly used tool in many areas such as longitudinal data analysis and failure time data analysis to characterize the underlying correlation or relationship among variables. In models (2.1) and (2.2), we use u_i to describe the possible correlation between T_i, the failure time of interest, and the observation times U_{ij}'s or process $\tilde{N}_i(t)$. Correspondingly, the parameter β_2 represents the degree of the correlation and more comments on this are given below.

It is easy to see that as $t \to \infty$, we have that $S(\infty|\boldsymbol{Z}) = \exp\big\{-\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}}_i + \beta_2 u_i)\big\} > 0$, the cure rate of the study population. Note that in general, a mixture cure model consists of a mixture of two separate regression models, one for the probability of the subject being cured and the other for the survival function for the noncured population (Wang et al., 2021). One advantage of model (2.1) over it is that model (2.1) inherits the proportional hazards model structure for the whole population and avoids the indistinguishable influence of covariates in the mixture model. In consequence, the regression parameters have more appealing interpretations.

It is apparent that under models (2.1) and (2.2), the parameter β_2 represents the extent of the association between T_i and $\tilde{N}_i(t)$. The two will be independent if $\beta_2 = 0$. It is worth to emphasize that the main interest here is inference about model (2.1) for which we observe interval-censored data, not model (2.2) for which one observes recurrent event data (Huang and Wang, 2004). We need to consider model (2.2) because of the possible association between T_i and $\tilde{N}_i(t)$ and otherwise, one may carry out a conditional estimation that treats the U_{ij}'s as constants.

Under the assumptions above and given the U_{ij}'s and u_i's, it is easy to see that the likelihood function has the form

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, F|u) = \prod_{i=1}^{n} \Bigg\{ \prod_{j=1}^{K_i} \Big\{ &\exp\big\{-\exp(\boldsymbol{\beta}_1' \tilde{\boldsymbol{Z}}_i + \beta_2 u_i) F(U_{ij-1})\big\} \\ &- \exp\big\{-\exp(\boldsymbol{\beta}_1' \tilde{\boldsymbol{Z}}_i + \beta_2 u_i) F(U_{ij})\big\}\Big\}^{\delta_{ij}} \\ \times &\Big\{ \exp\big\{-\exp(\boldsymbol{\beta}_1' \tilde{\boldsymbol{Z}}_i + \beta_2 u_i) F(U_{iK_i})\big\}\Big\}^{1-\sum_{j=1}^{K_i}\delta_{ij}} \Bigg\},\end{aligned}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \beta_2)'$. Note that for each i, only one of the δ_{ij}'s is equal to one with the others being zero. It follows that the likelihood function above can be rewritten as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, F|u) = \prod_{i=1}^{n} \Big\{ & \exp\big\{ -\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}}_i + \beta_2 u_i) F(L_i) \big\} \\ & - \exp\big\{ -\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}}_i + \beta_2 u_i) F(R_i) \big\} \Big\}^{\delta_i} \\ & \times \Big\{ \exp\big\{ -\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}}_i + \beta_2 u_i) F(R_i) \big\} \Big\}^{1-\delta_i}.\end{aligned}$$

Here $\delta_i = \sum_{j=1}^{K_i} \delta_{ij}$ and $(L_i, R_i]$ denotes the smallest interval that brackets T_i such that $L_i = \max\{U_{ij}, U_{ij} < T_i\}$ and $R_i = \min\{U_{ij}, U_{ij} \geq T_i\}$. It is apparent that $L_i = 0$ corresponds to a left-censored observation and $R_i = \infty$ indicates that the observation is right-censored.

To develop the estimation procedure below, let $H(t) = G[F(t)] = -\log\{1 - F(t)\}$, which removes the range restriction on F and is a continuous nondecreasing, nonnegative function. The resulting likelihood function of $\boldsymbol{\beta}$ and $H(\cdot)$ has the form

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, H|u) = \prod_{i=1}^{n} \Big\{ & \exp\big\{ -\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}}_i + \beta_2 u_i) G^{-1}[H(L_i)] \big\} \\ & - \exp\big\{ -\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}}_i + \beta_2 u_i) G^{-1}[H(R_i)] \big\} \Big\}^{\delta_i} \\ & \times \Big\{ \exp\big\{ -\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}}_i + \beta_2 u_i) G^{-1}[H(R_i)] \big\} \Big\}^{1-\delta_i}.\end{aligned} \tag{2.3}$$

In the next section, we will discuss the estimation of regression parameters as well as other parameters.

3. Estimation and Inference Procedure

Now we will discuss the estimation and inference procedure for models (2.1) and (2.2) with the focus on estimation of regression parameters $\boldsymbol{\beta}$. For this, following Wang et al. (2016) and others, we develop a two-step estimation procedure that first estimates model (2.2) or predicts the unknown latent variables u_i's and then estimate model (2.1). In the first step, we employ the approach given in Huang and Wang (2004), which discussed regression analysis of recurrent event data, and in the second step, the likelihood principle will be used.

3.1 Estimation of Model (2.2)

Define $\Lambda_{0h}(t) = \int_0^t \lambda_{0h}(u)du$ and assume that $\Lambda_{0h}(\tau_0) = 1$, where τ_0 denotes the longest follow-up time. For estimation of model (2.2), following Wang et al. (2016), Huang and Wang (2004) and others, we suggest to employ a borrow-strength estimation procedure as follows. Define $N_i(t) = \tilde{N}_i\{\min(t, \tau_i)\}$, $i = 1, \ldots, n$. Then we have that

$$N_i(t) = \int_0^{\min(t,\tau_i)} d\tilde{N}_i(s) = \int_0^t I(\tau_i \geq s) d\tilde{N}_i(s)\,,$$

and

$$E\{N_i(t) I(\tau_i \geq t) | \boldsymbol{Z}_i, u_i\} = \Lambda_{0h}(t) \exp(\boldsymbol{\alpha}' \boldsymbol{Z}_i + u_i) E\{I(\tau_i \geq t) | \boldsymbol{Z}_i, u_i\}\,.$$

Thus

$$\Lambda_{0h}(t) = \frac{E\{N_i(t)I(\tau_i \geq t)\}}{E\{\exp(\boldsymbol{\alpha}'\boldsymbol{Z}_i + u_i)I(\tau_i \geq t)\}}\,.$$

After some simple operations, this suggests that $\Lambda_{0h}(t)$ can be estimated by

$$\hat{\Lambda}_{0h}(t) = \prod_{s_{(l)} > t}\Big(1 - \frac{d_{(l)}}{D_{(l)}}\Big),$$

where the $s_{(l)}$'s denote the ordered and distinct values of the observation times $\{U_{ij}\}$, $d_{(l)}$ is the number of the observation times equal to $s_{(l)}$, and $D_{(l)}$ is the number of observation times satisfying $U_{ij} \leq s_{(l)} \leq \tau_i$ among all subjects.

For estimation of regression parameter $\boldsymbol{\alpha}$, note that

$$E[K_i|\boldsymbol{Z}_i, u_i, \tau_i] = \Lambda_{0h}(\tau_i)\exp(\boldsymbol{\alpha}'\boldsymbol{Z}_i + u_i)\,,$$

which gives

$$E[K_i\Lambda_{0h}^{-1}(\tau_i)|\boldsymbol{Z}_i, \tau_i] = E_{u_i}\{E[K_i\Lambda_{0h}^{-1}(\tau_i)|\boldsymbol{Z}_i, \tau_i, u_i]\} = E[\exp(u_i)]\exp(\boldsymbol{\alpha}'\boldsymbol{Z}_i).$$

This suggests a class of estimating equations

$$\sum_{i=1}^{n} w_i\tilde{\boldsymbol{Z}}_i\Big[K_i\Lambda_{0h}^{-1}(\tau_i) - E\{\exp(u_i)\}\exp(\boldsymbol{\alpha}'\boldsymbol{Z}_i)\Big] = 0$$

for estimation of $\boldsymbol{\alpha}$, where the w_i's are some weights that could depend on $\boldsymbol{Z}_i$'s, τ_i's and Λ_{0h}. Let $\hat{\boldsymbol{\alpha}}_n$ denote the estimation of $\boldsymbol{\alpha}$ given by the estimating equation above. Thus, it is natural to estimate u_i by

$$\hat{u}_i = \log\left\{\frac{K_i}{\hat{\Lambda}_{0h}(\tau_i)\exp(\hat{\boldsymbol{\alpha}}_n'\boldsymbol{Z}_i)}\right\}.$$

3.2 Estimation of Model (2.1)

For estimation of $\boldsymbol{\beta}$, it is natural to replace the u_i by the $\hat{u}_i$ and to maximum the resulting likelihood function $\mathcal{L}(\boldsymbol{\beta}, H|\hat{u})$. However, this may not be easy due to the involvement of the unknown function $H(t)$. To overcome this, by following Zhou et al. (2017), we propose to approximate $H(t)$ using Bernstein polynomials before the maximization of the likelihood function. Specifically, define the parameter space $\Theta = \{\boldsymbol{\theta} = (\boldsymbol{\beta}, H) \in \mathcal{B}\otimes\mathcal{M}\}$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \beta_2)'$, $\mathcal{B} = \{\boldsymbol{\beta} \in \mathbb{R}^{p+2} : ||\boldsymbol{\beta}|| \le M\}$ with M being a positive constant and $\mathcal{M}$ is the collection of all bounded and continuous nondecreasing, nonnegative functions over interval $[a, b]$ with $0 \le a < b < \infty$. Also, define

$$B_k(t, m, t_l, t_u) = \binom{m}{k}\left(\frac{t-a}{b-a}\right)^k\left(1-\frac{t-a}{b-a}\right)^{m-k},$$

Bernstein basis polynomials of degree $m = o(n^v)$ with $v \in (0, 1)$, and the sieve space $\Theta_n = \{\boldsymbol{\theta}_n = (\boldsymbol{\beta}, H_n) \in \mathcal{B}\otimes\mathcal{M}_n\}$, where

$$\mathcal{M}_n = \{H_n(t) = \sum_{k=0}^{m}\phi_k B_k(t, m, t_l, t_u) : 0 \le \phi_0 \le \cdots \le \phi_m, \sum_{0\le k\le m}|\phi_k| \le M_n\}.$$

In practice, v is usually taken to be 1/4.

Note that the positivity and monotonicity constraints imposed on the ϕ_k's can be easily removed by the reparametrization such as reparametrizing the coefficients $\{\phi_0, \ldots, \phi_m\}$ as the cumulative sums of $\{\exp(\phi_0^*), \ldots, \exp(\phi_m^*)\}$. It is clear that the new parameters $\{\phi_0^*, \ldots, \phi_m^*\}$ do not have any constraints, and the range $[a, b]$ is usually taken as the range of the U_{ij}'s. Also note that instead of Bernstein polynomials, one may employ step functions. However, the use of Bernstein polynomial has the optimal shape-preserving property among all approximation polynomials (Carnicer and Pena, 1993). Furthermore, they are easier to work with since they do not require the specification of interior knots.

Over the sieve space Θ_n, the log of the likelihood function given in (2.3) can be written as

$$\begin{aligned}\ell(\boldsymbol{\beta}, H_n|u) = \sum_{i=1}^{n} \delta_i \log \Big\{ & \exp\big\{ -\exp(\boldsymbol{\beta}_1' \tilde{\boldsymbol{Z}}_i + \beta_2 u_i) G^{-1}[H_n(L_i)]\big\} \\ & - \exp\big\{ -\exp(\boldsymbol{\beta}_1' \tilde{\boldsymbol{Z}}_i + \beta_2 u_i) G^{-1}[H_n(R_i)]\big\}\Big\} \\ & + (1-\delta_i) \log \Big\{ \exp\big\{ -\exp(\boldsymbol{\beta}_1' \tilde{\boldsymbol{Z}}_i + \beta_2 u_i) G^{-1}[H_n(R_i)]\big\}\Big\}.\end{aligned}$$

For estimation of $\boldsymbol{\theta}$, we propose to use the sieve maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}_n, \hat{H}_n)$ defined as the value that maximizes the log-likelihood function above with the u_i's replaced by the $\hat{u}_i$'s, which has the form

$$\begin{aligned}\ell(\boldsymbol{\beta}, H_n|\hat{u}) = \sum_{i=1}^{n} \delta_i \log \Big\{ & \exp\big\{ - \exp(\boldsymbol{\beta}_1' \tilde{\boldsymbol{Z}}_i + \beta_2 \hat{u}_i) G^{-1}[H_n(L_i)] \big\} \\ & - \exp\big\{ - \exp(\boldsymbol{\beta}_1' \tilde{\boldsymbol{Z}}_i + \beta_2 \hat{u}_i) G^{-1}[H_n(R_i)] \big\} \Big\} \\ & + (1-\delta_i) \log \Big\{ \exp\big\{ - \exp(\boldsymbol{\beta}_1' \tilde{\boldsymbol{Z}}_i + \beta_2 \hat{u}_i) G^{-1}[H_n(R_i)] \big\} \Big\} . \end{aligned} \tag{3.4}$$

Given $\hat{H}_n$, one can estimate F by $\hat{F}_n(t) = G^{-1}[H_n(t)]$ for $t < b$ and $\hat{F}_n(t) = 1$ for $t \geq b$.

To see the validity of the estimation procedure above and establish the asymptotic properties of the proposed estimators $\hat{\boldsymbol{\beta}}_n$, first note that Huang and Wang (2004) has shown that $\hat{\Lambda}_{0h}(t)$ and $\hat{\boldsymbol{\alpha}}_n$ are consistent and possess the asymptotical normality. Hence, one can treat the working procedure above as the one as if the u_i's were observed. However, it should be noted that the method and results given in Huang and Wang (2004) only apply to recurrent event data or the data on the $\tilde{N}_i(t)$'s, not the observed interval-censored data on the T_i's. Nevertheless, in the Appendix, we will show that under some regularity conditions, $\hat{\boldsymbol{\beta}}_n$ is consistent and $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ converges to the multivariate normal distribution with mean zero, where $\boldsymbol{\beta}_0$ denotes the true value of $\boldsymbol{\beta}$.

For making inference about the parameter $\hat{\boldsymbol{\beta}}_n$, it is obvious that we need to estimate the covariance matrix of $\hat{\boldsymbol{\beta}}_n$. For this, it will be difficult to derive a consistent estimator as it will be seen from the Appendix that the asymptotic covariance does not have an explicit expression. Thus instead we suggest to employ the weighted bootstrap method of Ma and Kosorok (2005). More specifically, let $\{q_1, \ldots, q_n\}$ denote n independent realizations of a bounded positive random variable q satisfying $\mathrm{E}(q) = 1$ and $\mathrm{Var}(q) = 1$. Given the q_i's, define the weighted sieve maximum log-likelihood estimators $\hat{\boldsymbol{\theta}}_n^* = \{\hat{\boldsymbol{\beta}}_n^*, \hat{H}_n^*\}$ as follows:

$$(\hat{\boldsymbol{\beta}}_n^*, \hat{H}_n^*) = \underset{(\boldsymbol{\beta}, H_n)\in\Theta_n}{\arg\min} \sum_{i=1}^{n} q_i \ell(\boldsymbol{\beta}, H_n|\hat{u}) \,.$$

Then if we generate the above process B samples and obtain the corresponding estimators $\hat{\boldsymbol{\beta}}_n^*$'s, we can estimate the covariance matrix of $\hat{\boldsymbol{\beta}}_n$ by the sample covariance matrix of the $\hat{\boldsymbol{\beta}}_n^*$'s. The numerical study below indicates that it works well and for the determination of $\hat{\boldsymbol{\theta}}_n$, the Quasi-Newton algorithm built in the function *fminunc* in Matlab will be used.

4. A Simulation Study

In this section, we present some results obtained from an extensive simulation study conducted to assess the empirical performance of the estimation procedure proposed in the previous sections under different practical situations. In the study, we considered a two-dimensional vector of covariates $\boldsymbol{Z}_i = (Z_{1i}, Z_{2i})'$ with the Z_{1i}'s following the Bernoulli distribution with the success probability of 0.5 and the Z_{2i}'s generated from the standard normal distribution. To generate the simulated data, we first generated $u_i^* = \exp(u_i)$ from the gamma distribution with mean 4 and variance 8 and assumed that the τ_i's follow the uniform distribution over the interval $[9, 10]$. For the observation process, given $\boldsymbol{Z}_i$ and u_i, we generated K_i, the number of real observation times for subject i, from the Poisson distribution with the mean

$$\Lambda_{ih}(\tau_i|\boldsymbol{Z}_i, u_i) = \frac{\tau_i \exp(\boldsymbol{\alpha}'\boldsymbol{Z}_i + u_i)}{10},$$

which corresponds to model (2.2) with $\lambda_{0h} = 1/10$. Then the observation times $(U_{i1}, \ldots, U_{iK_i})$ were set to be the order statistics of a random sample of size K_i from the uniform distribution over $(0, \tau_i)$, $i = 1, \ldots, n$.

To generate the true failure times T_i's, note that as pointed out by Liu and Shen (2009), under model (2.1), the overall survival function $S(t|\boldsymbol{Z}, u)$ can be written as

$$S(t|\boldsymbol{Z}, u) = \pi(\boldsymbol{Z}, u) + \{1 - \pi(\boldsymbol{Z}, u)\}S^*(t|\boldsymbol{Z}, u)$$

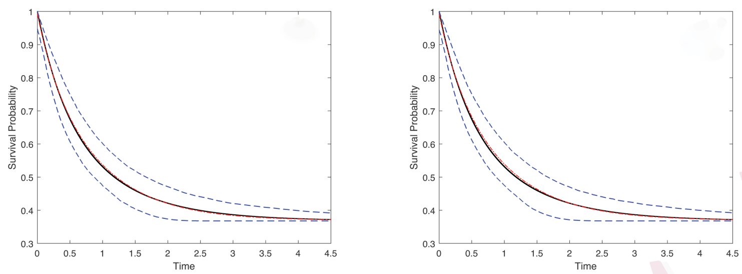
in the form of the mixture cure model. In the above, $\pi(\boldsymbol{Z}, u) = \exp\{-\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}} + \beta_2 u)\}$, which can be regarded as the covariate-specific cure probability, and

$$S^*(t|\boldsymbol{Z}, u) = \frac{\exp\big\{-\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}} + \beta_2 u)F(t)\big\} - \exp\big\{-\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}} + \beta_2 u)\big\}}{1 - \exp\big\{-\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}} + \beta_2 u)\big\}}.$$

Also note that the conditional survival function $S^*(t|\boldsymbol{Z}, u)$ is a proper survival function. Thus to generate the failure time T under (2.1), one can first generate a random variable from the Bernoulli distribution with the probability of success $\pi(\boldsymbol{Z}, u)$ for given $\boldsymbol{Z}$ and u. Then if it is 1, set $T = \infty$ and otherwise, generate $T = T^*$ from $S^*(t|\boldsymbol{Z}, u)$. The results given below are based on $n = 200$ or 400, $m = 4$, and $B = 100$ with 1000 replications.

Table 1 presents the results on the estimation of the regression parameters $\boldsymbol{\beta}_1$ and β_2 given by the proposed method with $F(t) = 1 - \exp(-t)$ and the true values $\boldsymbol{\beta}_1 = (\beta_{1,0}, \beta_{1,1}, \beta_{1,2})' = (0.5, 1, -0.5)'$, $\beta_2 = -0.2$, 0 or 0.2, and $\boldsymbol{\alpha}_0 = (0.2, 0.2)'$, $(0, 0)'$ or $(-0.2, 0.2)'$. They include the estimated bias (Bias) given by the average of the estimates minus the true value, the sample standard error (SSE) of the estimates, the average of the estimated standard errors (ESE), and the 95% empirical coverage probability (CP). One can see from the table that the proposed estimator seems to be unbiased and the variance estimation procedure appears to work well. In addition, the results on the empirical coverage probabilities suggest that the normal approximation to the distribution of the estimator appears to be appropriate, and as expected, the results became better when the sample size increased.

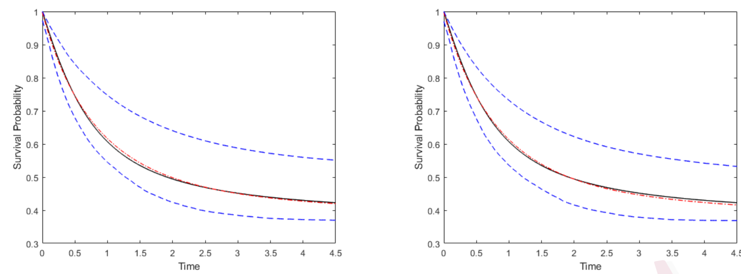
To see the possible effect of the function $F(t)$ on the proposed estima-



(a) $\boldsymbol{\alpha} = (0.2, 0.2)'$. (b) $\boldsymbol{\alpha} = (-0.2, -0.2)'$.

Figure 1: Estimated baseline survival curves along with the true curve regarding Table 1: the true function (solid), the estimated function (dash-dot), and the 95% confidence bands (dashed).

tion procedure, Table 2 gives the results obtained with $F(t) = 2\arctan(t)/\pi$, $n = 400$, the true values $\boldsymbol{\alpha}_0 = (0.2, 0.2)'$ or $(-0.2, 0.2)'$, and the other set-ups being the same as in Table 1. It is apparent that they gave similar conclusions as above. To assess the performance of the proposed estimation procedure with respect to the estimation of the baseline survival function, Figures 1 and 2 display the averages of the estimated baseline survival functions corresponding to the cases in Tables 1 and 2 with $n = 400$, $\boldsymbol{\beta}_1 = (\beta_{1,0}, \beta_{1,1}, \beta_{1,2})' = (0.5, 1, -0.5)'$, $\beta_2 = 0.2$, and $\boldsymbol{\alpha} = (0.2, 0.2)'$ or $(-0.2, -0.2)'$, respectively. For comparison, the true curve is also included along with the 95% empirical confidence bands and they again suggest that the proposed approach appears to work well.



(a) $\boldsymbol{\alpha} = (0.2, 0.2)'$. (b) $\boldsymbol{\alpha} = (-0.2, -0.2)'$.

Figure 2: Estimated baseline survival curves along with the true curve regarding Table 2: the true function (solid), the estimated function (dash-dot), and the 95% confidence bands (dashed).

Furthermore, we compared the proposed estimator to the sieve maximum likelihood estimator given by

$$\ell_{\mathrm{IND}}(\boldsymbol{\beta}, H_n) = \sum_{i=1}^{n} \log\Big\{S_{\mathrm{IND}}(L_i|\boldsymbol{Z}_i) - S_{\mathrm{IND}}(R_i|\boldsymbol{Z}_i)\Big\} + (1-\delta_i)\log\Big\{S_{\mathrm{IND}}(R_i|\boldsymbol{Z}_i)\Big\}$$

obtained under the independent interval censoring assumption. In the above, $S_{\mathrm{IND}}(t|\boldsymbol{Z}_i) = \exp\big\{-\exp(\boldsymbol{\beta}_1'\tilde{\boldsymbol{Z}}_i)G^{-1}[H_n(t)]\big\}$. Table 3 presents the results on the estimation of regression parameter $\boldsymbol{\beta}$ given by both the proposed method and the above independent censoring approach based on the simulated data with the true value $\boldsymbol{\beta}_1 = (\beta_{1,0}, \beta_{1,1}, \beta_{1,2})' = (0.5, 1, 0.5)'$, $\beta_2 = 0$, 0.2, 0.5, 0.7, or 1, and $\boldsymbol{\alpha}_0 = (\alpha_{0,1}, 0.5)'$ where $\alpha_{0,1} = \beta_2$. The other settings are being the same as for $n = 400$ in Table 1. Note that the dependence between the failure time and the observed process becomes stronger as the values of $\alpha_{0,1}$ and β_2 increase. Here for the independent censoring approach, we still use the weighted bootstrap procedure to estimate the covariance matrix of $\boldsymbol{\beta}$. Based on the results in Table 3, ignoring the informative censoring may lead to the wrong estimate when there is a correlation between the failure time and the observation process, and the bias increases as the dependence increases. Moreover, it shows that the proposed method can be applied to the independent censoring assumption if $\beta_2 = 0$.

To assess the possible effect of the observation process on the proposed estimation procedure, we repeated the study above by generating the observation times from two different renewal processes that assume that the gap time between two observation times follows either the gamma distribution or the uniform distribution. More specifically, for subject i and given $\boldsymbol{Z}_i$ and u_i, we repeatedly generated the gap times $10\exp(-\boldsymbol{\alpha}'\boldsymbol{Z}_i - u_i)V_i$ with V_i from either the gamma distribution with mean 1 and variance 0.5 or the uniform distribution over [0, 2] until the summation of the generated gap times being larger than τ_i. Table 4 presents the results on estimating regression parameters obtained by the proposed method with $n = 400$ and the other set-ups being the same as in Table 1. It can be seen that they are similar to those given Table 1 and thus suggest that the proposed method seems to be robust with respect to the observation process. We also considered other set-ups and obtained similar results and conclusions. In particular, we investigated the first estimation step on model (2.2) and the results indicated that both the first step and overall estimation performed well.

5. An Application

In this section, we apply the methodology proposed in the previous sections to the data arising from the Alzheimer's Disease Neuroimaging Initiative (ADNI) described above. It is a longitudinal study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's disease (AD). The study subjects in ADNI can have three stages, cognitively normal (CN), mild cognitive impairment (MCI) and AD. Due to the nature of the study, the occurrence time of either conversion is only known to be between the last observation time when the conversion had not occurred and the first observation time when it had already occurred. In other words, only interval-censored data are available for the conversion times.

In the analysis below, we will consider all participants in the MCI group with complete covariate information with the focus on the relationship between the time from the baseline visit date to the AD conversion and the seven covariates by following the existing literature (Li et al., 2020; Du et al., 2021; Wu et al., 2023). This gives a set of data consisting of 795 subjects with 303 who converted to AD during the study. The seven covariates are the two AD assessment scale test results (ADAS11, ADAS13), middle temporal gyrus volume (Midtemp), and rey auditory verbal learning test score of immediate recall (RAVLT.i) along with three baseline covariates Age, Gender and a gene that has been found to have a significant effect on AD, the ApoE4. Note that the literature has suggested that ADAS13, Midtemp and RAVLT.i are the most important clinical and demographic factors associated with AD conversion based on the individual variable analysis, and there is medical evidence that the people with ApoE4 had a higher rate of AD conversion (Association, 2019; Safieh et al., 2019; Li et al., 2020; Du et al., 2021; Wu et al., 2023).

To apply the proposed method, we first check if there exists a cured subgroup. For this, we obtained the nonparametric maximum likelihood estimator of the survival function given by Turnbull's self-consistency algorithm and present it in Figure 3. One can see that there appears to exist a plateau around 12.5 years, suggesting that a fraction of the population may never experience the event or there is a cured subgroup. Thus the proposed

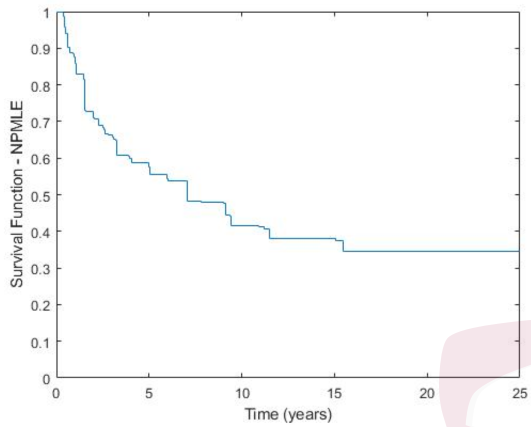


Figure 3: Turnbull survival curve, taking interval-censoring into account.

methodology should be used. For its application, we will let τ_i represent the last recorded visit time for subject i.

Table 5 presents the estimated covariate effects given by the proposed estimation procedure with $m = 5$. For comparison, we also obtained and include in the table the results given by the method ignored the informative censoring as in the simulation study. The table includes the proposed estimates (Est), the estimated standard errors (SE), and the p-values (p-value) for testing each covariate regression parameter being zero. We also tried different values for m and give the results in Table 6.

One can see from the tables that it seems that there existed a significantly negative association between the failure time and the observation process and both methods yielded similar results on the covariates except

Age. They indicate that ADAS13, Midtemp, and RAVLT.i were significant predictors of AD conversion. More specifically, the results suggest that AD conversion was negatively correlated with Midtemp and RAVLT.i but positively related to ADAS13. For the covariate Age, the proposed method found no significant effect on AD conversion, while ignoring informative censoring suggests a different result. One possible reason for this is that the latter overestimated the covariate effect by ignoring the informative censoring.

6. Discussion and Concluding Remarks

In this paper, we discussed regression analysis of case K interval-censored failure time data in the presence of both a cured subgroup and informative censoring or observation process. For the problem, a non-mixture cure model was presented and in the model, the latent variable approach was employed to characterize the correlation between the failure time of interest and the observation process. Unlike many existing methods, the proposed approach does not impose any distribution assumptions on the latent variable. For inference, a two-step estimation procedure was proposed and the asymptotic properties of the resulting estimators were established. Furthermore, a simulation study was performed and suggested that the proposed method works well for practical situations.

There exist several directions for future research related to the method proposed above. One is that in the proposed estimation procedure, we have assumed that the model has the proportional hazards model structure. It is apparent that in some situations, some other model structures may be more appropriate and thus it would be useful to generalize the proposed approach to these situations. Note that although the proposed method is flexible and can be easily implemented, it may be less efficient if the distribution of the latent variable is known or known up to some parameters. For the situation, one may want to directly maximize the observed likelihood function to derive a more efficient estimation procedure although it may not be easy or straightforward. A third direction is that in the previous sections, we have focused on univariate interval-censored data and it is well-known that sometimes one may face multivariate interval-censored data. Thus it would be helpful to generalize the proposed estimation procedure to the latter situation.

Acknowledgments

The authors wish to thank the Associate Editor and a reviewer for their many helpful and useful comments and suggestions that greatly improved the paper. This work was partially supported by the Natural Science Foundation of Jilin Province (Grant No. 20230101002JC).

Conflict of interest

The authors declare no potential conflict of interests.

A. Proof of the asymptotic properties

In this appendix, we will sketch the proof of the asymptotic normality of $\hat{\boldsymbol{\beta}}_n$. The letter C represents a constant, and it does not necessarily represent the same value each time here and in the proofs. We denote by $||\boldsymbol{a}||$ the Euclidean norm of a vector $\boldsymbol{a}$, $||f||_\infty = \sup_t |f(t)|$ the supremum norm of a function f, and $||g(X)||_2 = (\int g^2 dP)^{1/2}$ the $L^2(P)$ norm of a function g for X being distributed according to the probability measure P. For any $\boldsymbol{\theta}^i \in \Theta = \mathcal{B} \times \mathcal{H}$, $i = 1, 2$, we define a distance

$$d(\boldsymbol{\theta}^1, \boldsymbol{\theta}^2) = \left(||\boldsymbol{\beta}^1 - \boldsymbol{\beta}^2||^2 + ||H^1 - H^2||_2^2\right)^{1/2},$$

where $||H^1 - H^2||_2^2 = ||H^1(L) - H^2(L)||_2^2 + ||H^1(R) - H^2(R)||_2^2$. Before proving the theorems, we first describe the regularity conditions needed as follows:

Condition 1. (i) $\boldsymbol{\beta}_0$ is an interior point of $\mathcal{B}$ and $\mathcal{B}$ is compact subset of $\mathbb{R}^{p+2}$. (ii) The distribution of $\boldsymbol{Z}$ has bounded support. (iii) The latent variable u are bounded and satisfy $E\{\exp(u)|\boldsymbol{Z}\} = E\{\exp(u)\}$. Moreover, if $\tilde{\boldsymbol{\gamma}}_1\boldsymbol{Z}+\tilde{\gamma}_2 u+h(t) = 0$ with probability 1, then $\tilde{\boldsymbol{\gamma}}_1 = 0$, $\tilde{\gamma}_2 = 0$ and $h(t) = 0$ for $t \in [a, b]$.

Condition 2. The non-decreasing function $H_0 = G(F_0)$ belongs to $\mathcal{H}$ where $\mathcal{H} = \{H : H \in C^k[a, b],\ |H^{(k)}(t_1) - H^{(k)}(t_2)| \leq C|t_1 - t_2|^\gamma,$ for any $t_1, t_2 \in [a, b]\}$, where k is a non-negative integer, $\gamma \in (0, 1]$, $r = k + \gamma > 0.5$.

Condition 3. For the following-up time τ and latent variable u, we have: (i) $P(\tau \geq \tau_0, \exp(u) > 0) > 0$; (ii) $P(\tau > \tau_\epsilon) = 1$, where $\tau_\epsilon = \inf\{t : \Lambda_{0h}(t) > \epsilon\}$ for some $\epsilon > 0$, and $E\{\tilde{N}(\tau)^2\} < \infty$; (iii) $P(\tau = \infty, T = \infty|\boldsymbol{Z}) > 0$.

Condition 4. For the latent variable u, the variance of $\exp(u)$ is bounded, and there exists a positive small constant $\epsilon > 0$ such that $\exp(u) > \epsilon$ almost surely.

Condition 5. The function $Q(s) = E[\exp(u)I(\tau \geq s)]$ is a continuous function for $s \in [0, \tau_0]$.

Now, we are ready to prove the consistency and the asymptotic normality of $\hat{\boldsymbol{\beta}}_n$. The consistency can be established by using Theorem 5.7 of Van der Vaart (2000). To this end, we need to check the following three conditions:

(i) $\sup_{\boldsymbol{\theta}\in\Theta_n} |M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| \to 0$;

(ii) $\sup_{\boldsymbol{\theta}:d(\boldsymbol{\theta},\boldsymbol{\theta}_0)>\epsilon} M(\boldsymbol{\theta}) < M(\boldsymbol{\theta}_0)$;

(iii) $M_n(\hat{\boldsymbol{\theta}}_n) \geq M_n(\boldsymbol{\theta}_0) - o_p(1)$;

where $M(\boldsymbol{\theta}) = P\ell(\boldsymbol{\theta}|u)$ and $M_n(\boldsymbol{\theta}) = P_n\ell(\boldsymbol{\theta}|\hat{u})$.

For condition (i), first note that $|M_n(\boldsymbol{\theta})-M(\boldsymbol{\theta})| \leq |P_n\ell((\boldsymbol{\theta}|\hat{u})-P\ell(\boldsymbol{\theta}|\hat{u})|+|P\ell(\boldsymbol{\theta}|\hat{u})-P\ell(\boldsymbol{\theta}|u)|$. Denote $\boldsymbol{\eta} = (\boldsymbol{\alpha}, \Lambda_{0h})$ and $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\alpha}}, \hat{\Lambda}_{0h})$, For the second part of the right side, based on the consistency of $\hat{\boldsymbol{\eta}}$ established by Wang et al. (2001) and Condition 1, by using the delta method, we can show that

$$\begin{aligned}|P\ell(\boldsymbol{\theta}|u) - P\ell(\boldsymbol{\theta}|\hat{u})| &= P|\ell\{\boldsymbol{\theta}|u(\boldsymbol{\eta})\} - \ell\{\boldsymbol{\theta}|u(\hat{\boldsymbol{\eta}})\}| \\ &\leq P|\dot{\ell}\{\boldsymbol{\theta}|u(\hat{\boldsymbol{\eta}})\}|\big[\dot{u}(\hat{\gamma})||\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}||\big] + o_p(1)\,,\end{aligned}$$

and thus $\sup_{\boldsymbol{\theta}\in\Theta_n} |P\ell(\boldsymbol{\theta}|\hat{u}) - P\ell(\boldsymbol{\theta}|u)| \to 0$. The first part of the right side simplifies to a fundamental interval-censored problem with u replaced by $\hat{u}$, and similar arguments can be found in Zhou et al. (2017). Specifically, consider the class of functions $\mathcal{F}_n = \{\ell(\boldsymbol{\theta}|\hat{u}) : \boldsymbol{\theta} \in \Theta_n\}$. Then, by Corollary 3.2 in Arcones and Giné (1993) and Lemma 2.5 of Van de Geer (2000), one can show that the covering number of $\mathcal{F}_n$ satisfies $\log N(\epsilon, \mathcal{F}_n, L_1(P_n))/n \to 0$, and thus, $\sup_{\boldsymbol{\theta}\in\Theta_n} |P_n\ell((\boldsymbol{\theta}|\hat{u}) - P\ell(\boldsymbol{\theta}|\hat{u})| \to 0$. Hence, combining the above conclusion, we have (i) holds.

For condition (ii), first note that the Gibbs inequality implies that $\sup_{\boldsymbol{\theta}:d(\boldsymbol{\theta},\boldsymbol{\theta}_0)>\epsilon} M(\boldsymbol{\theta}) \leq M(\boldsymbol{\theta}_0)$. Assume that $\sup_{\boldsymbol{\theta}:d(\boldsymbol{\theta},\boldsymbol{\theta}_0)>\epsilon} M(\boldsymbol{\theta}) = M(\boldsymbol{\theta}_0)$.

Then there exists a sequence $\boldsymbol{\theta}_m$ such that $M(\boldsymbol{\theta}_m) \to \sup_{\boldsymbol{\theta}:d(\boldsymbol{\theta},\boldsymbol{\theta}_0)>\epsilon} M(\boldsymbol{\theta})$ and $d(\boldsymbol{\theta}_m, \boldsymbol{\theta}_0) > \epsilon$. By Condition 1, $\mathcal{B}$ is compact, and the sieve coefficients are bounded, there exists a subsequence $\boldsymbol{\theta}_{\tilde{m}}$ 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, $\boldsymbol{\theta}_{\tilde{m}}$ does not converge to $\boldsymbol{\theta}_0$ due to the fact $d(\boldsymbol{\theta}_{\tilde{m}}, \boldsymbol{\theta}_0) > \epsilon$. This conflicts with the aforementioned results that $\boldsymbol{\theta}_{\tilde{m}}$ converges to $\boldsymbol{\theta}_{m0}$. Therefore, (ii) holds.

For condition (iii), define $\boldsymbol{\theta}_{0,n} = (\boldsymbol{\beta}_0, H_n)$. We have $M_n(\hat{\boldsymbol{\theta}}_n) - M_n(\boldsymbol{\theta}_0) \geq (P_n - P)\{\ell(\boldsymbol{\theta}_{0,n}|\hat{u}) - \ell(\boldsymbol{\theta}_0|\hat{u})\} - P\{\ell(\boldsymbol{\theta}_0|\hat{u}) - \ell(\boldsymbol{\theta}_{0,n}|\hat{u})\}$. By Theorem 1.6.2 of Lorentz (1986), one can easily show that $P\{\ell(\boldsymbol{\theta}_0|\hat{u}) - \ell(\boldsymbol{\theta}_{0,n}|\hat{u})\} \leq Cd^2(\boldsymbol{\theta}_{0,n} - \boldsymbol{\theta}_0)$, and thus $P\{\ell(\boldsymbol{\theta}_0|\hat{u}) - \ell(\boldsymbol{\theta}_{0,n}|\hat{u})\} = o_p(1)$. Secondly, we show that $(P_n - P)\{\ell(\boldsymbol{\theta}_{0,n}|\hat{u}) - \ell(\boldsymbol{\theta}_0|\hat{u})\} = o_p(n^{-1/2})$. Define $\mathcal{F}_{\tilde{n}} = \{\ell(\boldsymbol{\theta}|\hat{u}) - \ell(\boldsymbol{\theta}_{0,\tilde{n}}|\hat{u}) : \boldsymbol{\theta} \in \Theta_n, ||\boldsymbol{\theta}_0 - \boldsymbol{\theta}_{0,\tilde{n}}|| \leq Cn^{-kv/2}\}$. Clearly, $\ell(\boldsymbol{\theta}_0|\hat{u}) - \ell(\boldsymbol{\theta}_{0,n}|\hat{u}) \in \mathcal{F}_{\tilde{n}}$. Similarly, we can prove that the ϵ bracketing number is also bounded by $C(1/\epsilon)^{m+1}$ and the bracketing integral (p.270 of Van de Geer (2000))

$$J_{[]}(\delta, \mathcal{F}_{\tilde{n}}, L_2(P)) \leq \int_0^\delta \sqrt{C(m+1)\log 1/\epsilon} < \infty\,.$$

leading $F_{\tilde{n}}$ as a P-Donsker by Theorem 19.5 in Van de Geer (2000). According to Corollary 2.3.12 in Van Der Vaart and Wellner (1996), we have $(P_n - P)\{\ell(\boldsymbol{\theta}_{0,n}|\hat{u}) - \ell(\boldsymbol{\theta}_0|\hat{u})\} = o_p(n^{-1/2})$. Hence (iii) holds, and it thus follows from Theorem 5.7 of Van der Vaart (2000) that $\hat{\boldsymbol{\beta}}_n$ is consistent.

To prove the asymptotic normality of $\hat{\boldsymbol{\beta}}_n$, it is apparent that it is sufficient to prove that the working score function $\dot{\ell}(\boldsymbol{\beta}, H_n|\hat{u})$ can be written as the summation of n independent and identically distributed mean zero random variables plus some negligible errors. For this, note that we can rewrite $\dot{\ell}(\boldsymbol{\beta}, H_n|\hat{u})$ as

$$\dot{\ell}(\boldsymbol{\beta}, H_n|\hat{u}) = I + II,$$

where $I = \dot{\ell}(\boldsymbol{\beta}, H_n|u)$ and $II = \dot{\ell}(\boldsymbol{\beta}, H_n|\hat{u}) - \dot{\ell}(\boldsymbol{\beta}, H_n|u)$. By following the proof of Theorem 4 of Hu and Xiang (2013), the efficient score vector for $\boldsymbol{\beta}$ is $\dot{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \dot{\ell}_H(\boldsymbol{\theta})[h^*]$, where $\dot{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ is the score for $\boldsymbol{\beta}$ and $h^* \in \mathbb{R}^{p+2}$ such that it is orthogonal to $\dot{\ell}(\boldsymbol{\theta})[h]$ in $L_2^0(P)$. In this way, we can see that the first term I can be written as the summation of n independent and identically distributed mean zero random variables plus some negligible errors.

To show that the second term II can also be written as the summation of n independent and identically distributed mean zero random variables plus some negligible errors, first note that $\dot{\ell}(\boldsymbol{\beta}, H_n|\hat{u})$ is a continuous function of $\hat{u}$ and $\hat{u}$ is the function of $\hat{\Lambda}_{0h}(\tau)$ and $\hat{\boldsymbol{\alpha}}_n$. Define

$$\ddot{\ell}_{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, H_n|u_i(\Lambda_{0h}(\tau_i), \boldsymbol{\alpha})\} = \frac{\partial}{\partial \boldsymbol{\alpha}}\dot{\ell}\{\boldsymbol{\beta}, H_n|u_i(\Lambda_{0h}(\tau_i), \boldsymbol{\alpha})\},$$

and

$$\ddot{\ell}_{\Lambda_{0h}}\{\boldsymbol{\beta}, H_n|u_i(\Lambda_{0h}(\tau_i), \boldsymbol{\alpha})\} = \frac{\partial}{\partial s}\dot{\ell}\{\boldsymbol{\beta}, H_n|u_i(s, \boldsymbol{\alpha})\}|_{s=\Lambda_{0h}(\tau_i)},$$

then by using the multivariate Taylor expansion, we can obtain that

$$\begin{aligned} II &= \sum_{i=1}^{n} \big[\dot{\ell}\{\boldsymbol{\beta}, H_n | u(\hat{\Lambda}_{0h}(\tau_i), \hat{\boldsymbol{\alpha}}_n)\} - \dot{\ell}\{\boldsymbol{\beta}, H_n | u(\Lambda_{0h}(\tau_i), \boldsymbol{\alpha})\} \big] \\ &= \sum_{i=1}^{n} \big[\ddot{\ell}_{\Lambda_{0h}}\{\boldsymbol{\beta}, H_n | u_i(\Lambda_{0h}(\tau_i), \boldsymbol{\alpha})\}\{\hat{\Lambda}_{0h}(\tau_i) - \Lambda_{0h}(\tau_i)\} \\ &\qquad + \ddot{\ell}_{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, H_n | u_i(\Lambda_{0h}(\tau_i), \boldsymbol{\alpha})\}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}) \big] + o_p(1)\,. \end{aligned}$$

To further investigate the equality above, note that by Huang and Wang (2004), we have

$$\hat{\Lambda}_{0h}(t) - \Lambda_{0h}(t) = \frac{1}{n}\Lambda_{0h}(t) \sum_{i=1}^{n} b_{ih}(t) + o_p(n^{-1/2})$$

for $\inf\{t : \Lambda_{0h}(t) > 0\} < t < \tau_0$, where

$$b_{ih}(t) = \sum_{j=1}^{K_i} \int_t^{\tau_0} \frac{I(U_{ij} \le s \le \tau_i) d\tilde{Q}(s)}{R^2(s)} - \frac{I(t \le U_{ij} \le \tau_0)}{R(U_{ij})}\,,$$

$$\tilde{Q}(t) = \int_0^t Q(s) d\Lambda_{0h}(s)\,, \text{ and } R(t) = Q(t)\Lambda_{0h}(t)\,.$$

Also, we have

$$\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha} = \frac{1}{n} \sum_{i=1}^{n} f_{ih} + o_p(n^{-1/2})\,,$$

where f_{ih} is the vector function $\big(E[-\partial f_i / \partial \boldsymbol{\alpha}]\big)^{-1} f_i$, and

$$f_i = -\int \frac{w \tilde{\boldsymbol{Z}} b_{ih}(\tau) d\mathcal{P}(w, \tilde{\boldsymbol{Z}}, K, \tau)}{\Lambda_{0h}(\tau)} + w_i \tilde{\boldsymbol{Z}}_i \{K_i \Lambda_{0h}^{-1}(\tau_i) - E[\exp(u_i)] \exp(\boldsymbol{\alpha}' \boldsymbol{Z}_i)\}$$

with $\mathcal{P}(\cdot)$ denoting the joint distribution of the underlying variables. These yields that

$$
\begin{aligned}
II = \sum_{i=1}^{n} E_j \big[& \ddot{\ell}_{\Lambda_{0h}}\{\boldsymbol{\beta}, H_n | u_i(\Lambda_{0h}(\tau_i), \boldsymbol{\alpha})\}\Lambda_{0h}(\tau_i) b_{jh}(\tau_i) \\
& + \ddot{\ell}_{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, H_n | u_i(\Lambda_{0h}(\tau_i), \boldsymbol{\alpha})\} f_{jh}\big] + o_p(n^{-1/2}) \\
= \sum_{i=1}^{n} & d_i(\boldsymbol{\beta}, H_n) + o_p(n^{-1/2}) \,.
\end{aligned}
$$

This shows that the working score function $\dot{\ell}(\boldsymbol{\beta}, H_n|\hat{u})$ can be written as the summation of n independent and identically distributed mean zero random variables plus some negligible errors. It thus follows from the Taylor series expansion that as $n \to \infty$, $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ converges in distribution to a mean zero normal random variable.

References

Arcones, M. A. and E. Giné (1993). Limit theorems for u-processes. *The Annals of Probability 21*(3), 1494–1542.

Association, A. (2019). 2019 alzheimer's disease facts and figures. *Alzheimer's & Dementia 15*(3), 321–387.

Carnicer, J. M. and J. M. Pena (1993). Shape preserving representations and optimality of the bernstein basis. *Advances in Computational Mathematics 1*(2), 173–196.

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

Du, M., H. Li, and J. Sun (2021). Regression analysis of censored data with nonignorable missing covariates and application to alzheimer disease. *Computational Statistics & Data Analysis 157*, 107157.

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

Huang, C.-Y. and M.-C. Wang (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association 99*(468), 1153–1165.

Huang, X. and R. A. Wolfe (2002). A frailty model for informative censoring. *Biometrics 58*(3), 510–520.

Kalbfleisch, J. D. and R. L. Prentice (2002). *The statistical analysis of failure time data*. John Wiley & Sons.

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

Li, S., Q. Wu, and J. Sun (2020). Penalized estimation of semiparametric transformation models with interval-censored data and application to alzheimer's disease. *Statistical Methods in Medical Research 29*(8), 2151–2166.

Liu, H. and Y. Shen (2009). A semiparametric regression cure model for interval-censored data. *Journal of the American Statistical Association 104*(487), 1168–1178.

Lorentz, G. G. (1986). *Bernstein polynomials*. New York: Chelsea Publishing Co.

Ma, L., T. Hu, and J. Sun (2016). Cox regression analysis of dependent interval-censored failure time data. *Computational Statistics & Data Analysis 103*, 79–90.

Ma, S. (2009). Cure model with current status data. *Statistica Sinica 19*(1), 233–249.

Ma, S. (2010). Mixed case interval censored data with a cured subgroup. *Statistica Sinica 20*(3), 1165–1181.

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

Mao, M. and J.-L. Wang (2010). Semiparametric efficient estimation for a class of generalized proportional odds cure models. *Journal of the American Statistical Association 105*(489), 302–311.

Safieh, M., A. D. Korczyn, and D. M. Michaelson (2019). Apoe4: an emerging therapeutic target for alzheimer's disease. *BMC Medicine 17*(1), 1–17.

Sun, J. (2006). *The statistical analysis of interval-censored failure time data*. Springer.

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

Van de Geer, S. (2000). *Empirical Processes in M-estimation*, Volume 6. Cambridge University Press.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge University Press.

Van Der Vaart, A. W. and J. Wellner (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer Science & Business Media.

Wang, L., C. S. McMahan, M. G. Hudgens, and Z. P. Qureshi (2016). A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics 72*(1), 222–231.

Wang, M.-C., J. Qin, and C.-T. Chiang (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association 96*(455), 1057–1065.

Wang, P., H. Zhao, and J. Sun (2016). Regression analysis of case k interval-censored failure time data in the presence of informative censoring. *Biometrics 72*(4), 1103–1112.

Wang, S., D. Xu, C. Wang, and J. Sun (2021). Semiparametric analysis of case k interval-censored failure time data in the presence of a cured subgroup and informative censoring. *Journal of Statistical Computation and Simulation 91*(10), 2018–2033.

Wu, Y., H. Zhao, and J. Sun (2023). Group variable selection for the cox model with interval-censored failure time data. *Biometrics*.

Xu, D., S. Zhao, T. Hu, and J. Sun (2019). Regression analysis of informatively interval-censored failure time data with semiparametric linear transformation model. *Journal of Nonparametric Statistics 31*(3), 663–679.

Zhang, Z., L. Sun, J. Sun, and D. M. Finkelstein (2007). Regression analysis of failure time data with informative interval censoring. *Statistics in Medicine 26*(12), 2533–2546.

Zhao, S., T. Hu, L. Ma, P. Wang, and J. Sun (2015). Regression analysis of interval-censored failure time data with the additive hazards model in the presence of informative censoring. *Statistics and its Interface 8*(3), 367–377.

Zhou, Q., T. Hu, and J. Sun (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.

Yichen Lou

E-mail: louyc19@mails.jlu.edu.cn

Jianguo Sun

E-mail: sunj@missouri.edu

Peijie Wang

E-mail: wangpeijie@jlu.edu.cn

Table 1: Simulation results on estimation of $\boldsymbol{\beta}$ with $F(t) = 1 - \exp(-t)$.

		$n = 200$				$n = 400$			
$\boldsymbol{\alpha}$	Parameter	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
$(0.2, 0.2)$	$\beta_{1,0} = 0.5$	0.0619	0.2203	0.2111	0.933	0.0487	0.1416	0.1428	0.944
	$\beta_{1,1} = 1$	0.0601	0.2547	0.2387	0.939	0.0286	0.1668	0.1612	0.945
	$\beta_{1,2} = -0.5$	-0.0156	0.1321	0.1224	0.939	-0.0061	0.0864	0.0833	0.949
	$\beta_2 = 0.2$	-0.0334	0.1524	0.1351	0.915	-0.0378	0.0965	0.0933	0.918
	$\beta_{1,0} = 0.5$	0.0226	0.2058	0.2043	0.946	0.0081	0.1453	0.1400	0.939
	$\beta_{1,1} = 1$	0.0553	0.2275	0.2185	0.943	0.0262	0.1528	0.1495	0.948
	$\beta_{1,2} = -0.5$	-0.0258	0.1245	0.1164	0.940	-0.0138	0.8336	0.0799	0.941
	$\beta_2 = 0$	0.0002	0.1377	0.1291	0.930	0.0001	0.0933	0.0898	0.942
	$\beta_{1,0} = 0.5$	-0.0431	0.2005	0.1990	0.945	-0.0505	0.1352	0.1382	0.936
	$\beta_{1,1} = 1$	0.0331	0.2215	0.2064	0.945	0.0059	0.1486	0.1436	0.948
	$\beta_{1,2} = -0.5$	-0.0219	0.1232	0.1101	0.914	-0.0081	0.0784	0.0771	0.944
	$\beta_2 = -0.2$	0.0505	0.1328	0.1277	0.917	0.0410	0.0915	0.0892	0.917
$(0, 0)$	$\beta_{1,0} = 0.5$	0.0749	0.2282	0.2140	0.918	0.0513	0.1466	0.1440	0.925
	$\beta_{1,1} = 1$	0.0594	0.2699	0.2440	0.935	0.0169	0.1666	0.1626	0.951
	$\beta_{1,2} = -0.5$	-0.0346	0.1373	0.1249	0.933	-0.0048	0.0855	0.0840	0.954
	$\beta_2 = 0.2$	-0.0387	0.1489	0.1394	0.927	-0.0365	0.0996	0.0959	0.929
	$\beta_{1,0} = 0.5$	0.0223	0.2303	0.2055	0.933	0.0062	0.1447	0.1414	0.944
	$\beta_{1,1} = 1$	0.0620	0.2425	0.2235	0.932	0.0268	0.1573	0.1522	0.957
	$\beta_{1,2} = -0.5$	-0.0311	0.1262	0.1161	0.928	-0.0122	0.0841	0.0798	0.940
	$\beta_2 = 0$	0.0007	0.1443	0.1335	0.942	-0.0007	0.0923	0.0920	0.942
	$\beta_{1,0} = 0.5$	-0.0463	0.2129	0.2007	0.931	-0.0436	0.1398	0.1412	0.935
	$\beta_{1,1} = 1$	0.0284	0.2340	0.2068	0.909	-0.0007	0.1457	0.1435	0.948
	$\beta_{1,2} = -0.5$	-0.0210	0.1217	0.1112	0.934	-0.0050	0.0814	0.0767	0.934
	$\beta_2 = -0.2$	0.0478	0.1397	0.1302	0.920	0.0478	0.0925	0.0911	0.909
$(-0.2, -0.2)$	$\beta_{1,0} = 0.5$	0.0652	0.2292	0.2314	0.925	0.0496	0.1506	0.1477	0.933
	$\beta_{1,1} = 1$	0.0697	0.3102	0.2497	0.936	0.0205	0.1752	0.1665	0.944
	$\beta_{1,2} = -0.5$	-0.0432	0.1354	0.1241	0.930	-0.0147	0.0880	0.0845	0.939
	$\beta_2 = 0.2$	-0.0302	0.1591	0.1445	0.918	-0.0363	0.1001	0.0991	0.927
	$\beta_{1,0} = 0.5$	0.0303	0.2184	0.2116	0.936	0.0058	0.1490	0.1446	0.944
	$\beta_{1,1} = 1$	0.0619	0.2386	0.2270	0.934	0.0278	0.1614	0.1531	0.943
	$\beta_{1,2} = -0.5$	-0.0309	0.1292	0.1167	0.927	-0.0115	0.0825	0.0797	0.941
	$\beta_2 = 0$	-0.0067	0.1467	0.1366	0.930	0.0027	0.1004	0.0953	0.934
	$\beta_{1,0} = 0.5$	-0.0461	0.2101	0.2048	0.930	-0.0454	0.1481	0.1435	0.930
	$\beta_{1,1} = 1$	0.0371	0.2221	0.2079	0.936	0.0133	0.1520	0.1444	0.938
	$\beta_{1,2} = -0.5$	-0.0131	0.1183	0.1102	0.939	-0.0052	0.0817	0.0765	0.940
	$\beta_2 = -0.2$	0.0514	0.1419	0.1341	0.910	0.0463	0.0973	0.0937	0.919

Table 2: Simulation results on estimation of $\boldsymbol{\beta}$ with $F(t) = 2\arctan(t)/\pi$ and $n = 400$.

	$\boldsymbol{\alpha} = (-0.2, -0.2)$				$\boldsymbol{\alpha} = (0.2, 0.2)$			
Parameter	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
$\beta_{1,0} = 0.5$	0.0332	0.1846	0.2134	0.964	0.0509	0.2028	0.1983	0.947
$\beta_{1,1} = 1$	0.0078	0.1582	0.1523	0.946	0.0066	0.1562	0.1482	0.939
$\beta_{1,2} = -0.5$	-0.0042	0.0822	0.0784	0.942	-0.0061	0.0854	0.0786	0.926
$\beta_2 = 0.2$	-0.0384	0.0997	0.0939	0.915	-0.0353	0.0906	0.0887	0.924
$\beta_{1,0} = 0.5$	-0.0036	0.1786	0.1883	0.959	0.0048	0.1703	0.1781	0.957
$\beta_{1,1} = 1$	0.0074	0.1414	0.1433	0.964	0.0161	0.1458	0.1413	0.944
$\beta_{1,2} = -0.5$	-0.0106	0.0752	0.0756	0.952	-0.0117	0.0773	0.0760	0.951
$\beta_2 = 0$	0.0035	0.0933	0.0914	0.939	-0.0014	0.0896	0.0867	0.947
$\beta_{1,0} = 0.5$	-0.0573	0.1790	0.1786	0.939	-0.0518	0.1678	0.1705	0.936
$\beta_{1,1} = 1$	0.0108	0.1369	0.1392	0.949	0.0008	0.1399	0.1385	0.949
$\beta_{1,2} = -0.5$	0.0012	0.0757	0.0737	0.945	-0.0043	0.0793	0.0741	0.934
$\beta_2 = -0.2$	0.0447	0.0901	0.0912	0.906	0.0446	0.0900	0.0869	0.910

Table 3: Simulation results on comparison with ignoring informative censoring.

			Proposed Method				Igonring informative censoring			
β_2	α	Parameter	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
0.0	(0.0,0.5)	$\beta_{1,0} = 0.5$	0.0127	0.1446	0.1422	0.947	0.0105	0.0913	0.0921	0.949
		$\beta_{1,1} = 1$	0.0258	0.1583	0.1469	0.938	0.0212	0.1575	0.1467	0.937
		$\beta_{1,2} = 0.5$	0.0149	0.0790	0.0750	0.937	0.0124	0.0787	0.0748	0.941
0.2	(0.2,0.5)	$\beta_{1,0} = 0.5$	0.0506	0.1419	0.1433	0.946	0.2290	0.0983	0.0961	0.345
		$\beta_{1,1} = 1$	0.0231	0.1618	0.1563	0.938	0.0130	0.1621	0.1574	0.942
		$\beta_{1,2} = 0.5$	0.0176	0.0837	0.0781	0.931	0.0024	0.0816	0.0776	0.937
0.5	(0.5,0.5)	$\beta_{1,0} = 0.5$	0.0562	0.1515	0.1495	0.938	0.4492	0.1105	0.1057	0.008
		$\beta_{1,1} = 1$	0.0128	0.1840	0.1778	0.927	-0.0904	0.1914	0.1741	0.878
		$\beta_{1,2} = 0.5$	0.0258	0.0962	0.0868	0.931	-0.0441	0.0901	0.0822	0.876
0.7	(0.7,0.5)	$\beta_{1,0} = 0.5$	0.0487	0.1602	0.1639	0.937	0.5176	0.1141	0.1074	0.001
		$\beta_{1,1} = 1$	0.0083	0.2125	0.1933	0.924	-0.1949	0.1903	0.1799	0.760
		$\beta_{1,2} = 0.5$	0.0213	0.1040	0.0932	0.939	0.0971	0.0901	0.0834	0.724
1.0	(1.0,0.5)	$\beta_{1,0} = 0.5$	0.0191	0.1812	0.1837	0.950	0.5558	0.1087	0.1062	0.000
		$\beta_{1,1} = 1$	0.0073	0.2256	0.2104	0.923	-0.3455	0.1962	0.1784	0.476
		$\beta_{1,2} = 0.5$	0.0033	0.1177	0.1009	0.906	-0.1759	0.0922	0.0822	0.412

Table 4: Simulation results on $\boldsymbol{\beta}$ with the renewal observation processes.

		Gamma Gap				Uniform Gap			
$\boldsymbol{\alpha}$	Parameter	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
$(0.2, 0.2)$	$\beta_{1,0} = 0.5$	0.0497	0.1524	0.1428	0.921	0.0498	0.1477	0.1418	0.921
	$\beta_{1,1} = 1$	0.0328	0.1814	0.1697	0.935	0.0247	0.1808	0.1718	0.942
	$\beta_{1,2} = -0.5$	-0.0125	0.0945	0.0876	0.930	-0.0062	0.0902	0.0875	0.941
	$\beta_2 = 0.2$	-0.0244	0.1048	0.0986	0.922	-0.0143	0.1044	0.0983	0.930
	$\beta_{1,0} = 0.5$	0.0206	0.1403	0.1374	0.942	0.0068	0.1414	0.1353	0.939
	$\beta_{1,1} = 1$	0.0242	0.1620	0.1554	0.930	0.0347	0.1620	0.1568	0.945
	$\beta_{1,2} = -0.5$	-0.0124	0.0872	0.0829	0.934	-0.0098	0.0838	0.0823	0.950
	$\beta_2 = 0$	-0.0064	0.0984	0.0936	0.934	0.0025	0.0984	0.0936	0.938
	$\beta_{1,0} = 0.5$	-0.0444	0.1386	0.1348	0.924	-0.0448	0.1381	0.1331	0.920
	$\beta_{1,1} = 1$	0.0103	0.1571	0.1474	0.934	0.0134	0.1516	0.1473	0.948
	$\beta_{1,2} = -0.5$	-0.0051	0.0843	0.0789	0.930	-0.0113	0.0798	0.0793	0.958
	$\beta_2 = -0.2$	0.0336	0.0957	0.0921	0.917	0.0260	0.0967	0.0925	0.927
$(0, 0)$	$\beta_{1,0} = 0.5$	0.0483	0.1517	0.1464	0.924	0.0523	0.1456	0.1415	0.930
	$\beta_{1,1} = 1$	0.0306	0.1919	0.1742	0.932	0.0360	0.1823	0.1759	0.950
	$\beta_{1,2} = -0.5$	-0.0159	0.0936	0.0886	0.938	-0.0155	0.1005	0.0888	0.920
	$\beta_2 = 0.2$	-0.0193	0.1087	0.1019	0.942	-0.0112	0.1081	0.1026	0.934
	$\beta_{1,0} = 0.5$	0.0161	0.1458	0.1398	0.932	0.0108	0.1421	0.1372	0.937
	$\beta_{1,1} = 1$	0.0280	0.1669	0.1576	0.939	0.0232	0.1595	0.1582	0.943
	$\beta_{1,2} = -0.5$	-0.0138	0.0830	0.0825	0.947	-0.0154	0.0855	0.0832	0.940
	$\beta_2 = 0$	-0.0027	0.1003	0.0963	0.937	0.0022	0.1007	0.0959	0.935
	$\beta_{1,0} = 0.5$	-0.0503	0.1409	0.1370	0.916	-0.0408	0.1358	0.1345	0.933
	$\beta_{1,1} = 1$	0.0171	0.1526	0.1477	0.942	0.0155	0.1553	0.1481	0.937
	$\beta_{1,2} = -0.5$	-0.0103	0.0783	0.0788	0.953	-0.0083	0.0812	0.0796	0.947
	$\beta_2 = -0.2$	0.0339	0.0962	0.0939	0.913	0.0238	0.0983	0.0942	0.926
$(-0.2, -0.2)$	$\beta_{1,0} = 0.5$	0.0544	0.1583	0.1487	0.923	0.0476	0.1538	0.1460	0.933
	$\beta_{1,1} = 1$	0.0252	0.1957	0.1799	0.940	0.0308	0.1949	0.1804	0.947
	$\beta_{1,2} = -0.5$	-0.0213	0.0905	0.0896	0.946	-0.0215	0.0934	0.0896	0.941
	$\beta_2 = 0.2$	-0.0195	0.1153	0.1071	0.934	-0.0080	0.1141	0.1064	0.931
	$\beta_{1,0} = 0.5$	0.0111	0.1463	0.1442	0.935	0.0128	0.1488	0.1408	0.929
	$\beta_{1,1} = 1$	0.0237	0.1698	0.1596	0.936	0.0274	0.1659	0.1598	0.942
	$\beta_{1,2} = -0.5$	-0.0069	0.0855	0.0825	0.936	-0.0138	0.0882	0.0828	0.937
	$\beta_2 = 0$	0.0019	0.1000	0.1000	0.952	-0.0020	0.1049	0.1001	0.938
	$\beta_{1,0} = 0.5$	-0.0351	0.1410	0.1384	0.941	-0.0393	0.1446	0.1376	0.924
	$\beta_{1,1} = 1$	0.0235	0.1616	0.1480	0.935	0.0206	0.1524	0.1496	0.941
	$\beta_{1,2} = -0.5$	-0.0037	0.0803	0.0787	0.947	-0.0067	0.0839	0.0790	0.937
	$\beta_2 = -0.2$	0.0244	0.1012	0.0977	0.933	0.0260	0.1019	0.0973	0.926

Table 5: The estimated covariate effects on the AD conversion.

	Proposed method			Igonring informative censoring		
Factors	Est	SE	p-value	Est	SE	p-value
$\beta_{1,0}$	4.7996	1.1245	0.0000	3.7530	1.0094	0.0002
Age	-0.0140	0.0085	0.0995	-0.0163	0.0082	0.0463
Gender	-0.1097	0.1416	0.4387	-0.0973	0.1379	0.4803
ADAS11	-0.0702	0.0385	0.0678	-0.0667	0.0409	0.1033
ADAS13	0.1583	0.0294	0.0000	0.1538	0.0312	0.0000
RAVLT.i	-0.0388	0.0096	0.0001	-0.0400	0.0095	0.0000
MidTemp	-0.0163	0.0025	0.0000	-0.0160	0.0024	0.0000
ApoE4	0.4431	0.0815	0.0000	0.4202	0.0768	0.0000
β_2	-0.5028	0.1437	0.0005	—	—	—

Table 6: The estimated covariate effects with different m.

	$m = 3$			$m = 7$		
Factors	Est	SE	p-value	Est	SE	p-value
$\beta_{1,0}$	4.9745	1.0990	0.0001	4.9645	1.0165	0.0000
Age	-0.0140	0.0083	0.0902	-0.0149	0.0079	0.0588
Gender	-0.1051	0.1346	0.4349	-0.1162	0.1275	0.3621
ADAS11	-0.0704	0.0393	0.0732	-0.0684	0.0400	0.0875
ADAS13	0.1591	0.0308	0.0000	0.1570	0.0296	0.0000
RAVLT.i	-0.0389	0.0099	0.0001	-0.0394	0.0090	0.0000
MidTemp	-0.0164	0.0024	0.0000	-0.0165	0.0025	0.0000
ApoE4	0.4472	0.0845	0.0000	0.4410	0.0801	0.0000
β_2	-0.5213	0.1588	0.0010	-0.5101	0.1370	0.0002