Statistica Sinica Preprint No: SS-2026-0046	
Title	Simultaneous Estimation and Variable Selection for Linear Transformation Models in the Presence of Functional and Missing Covariates with the Application to Alzheimer's Disease
Manuscript ID	SS-2026-0046
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202026.0046
Complete List of Authors	Yichen Lou, Mingyue Du and Jianguo Sun
Corresponding Authors	Mingyue Du
E-mails	mingydu@jlu.edu.cn
Notice: Accepted author version.	

Simultaneous Estimation and Variable Selection for Linear Transformation Models in the Presence of Functional and Missing Covariates with the Application to Alzheimer's Disease

Yichen Lou[1], Mingyue Du[2*], and Jianguo Sun[3]

[1]School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore.
[2]School of Mathematics, Jilin University, Changchun, China.
[3]Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, China.
[*]Address correspondence to:

Abstract

The linear transformation model is one of the most commonly used classes of models for regression analysis of failure time data. In this paper, we consider its estimation based on interval-censored data in the presence of both functional and missing covariates with the focus on simultaneous estimation and variable selection, a problem for which it does not seem to exist an established approach. For the problem, we first consider estimation and develop a sieve maximum likelihood estimation procedure based on functional principal component analysis and the use of the inverse probability weighting technique to handle the infinite-dimensional nature of functional predictors and missing-at-random covariates, respectively. Then we consider variable selection and propose a penalized estimation approach by employing the minimum approximated information criterion, which eliminates the need for tuning parameter selection that is required for most of the existing methods. The proposed estimators are shown to be consistent and possess the oracle property, and a simulation study is conducted and suggests that the proposed methods work well in practice. Finally, they are applied to an Alzheimer's Disease study that motivated this investigation and the analysis provides some new insights about the roles of high-dimensional imaging and clinical factors.

Keywords: Functional data analysis; Interval-censored data; Penalized likelihood; Transformation model; Variable selection.

1 Introduction

This paper discusses estimation of the linear transformation model (LTM) with the focus on simultaneous estimation and variable selection when one only observes interval-censored failure time data and there exist both functional and missing covariates. Although many methods have been proposed for estimation of the model, one of the most commonly used classes of models for regression analysis of failure time data, it does not seem to exist an established approach for the

situation said above. By interval-censored data, we usually refer to the data where the failure time of interest is known or observed only to lie within an interval rather than being observed exactly. Such data commonly arise in many areas, including economics, epidemiological studies, medical research, and social sciences (Sun, 2006; Sun and Chen, 2022). A more specific and typical area that usually yields interval-censored data is clinical trials or periodic follow-up studies.

This study was motivated by the data arising from the Alzheimer's Disease Neuroimaging Initiative (ADNI), an extensive longitudinal and multi-center study initiated in 2004 (Weiner et al., 2013). A primary objective of ADNI is to identify biomarkers for early detection and monitoring of Alzheimer's disease (AD). Among others, one important outcome is the conversion time from mild cognitive impairment (MCI) to AD, which is critical for tracking disease progression and making timely interventions (Weiner et al., 2015). It is easy to see that due to the periodic follow-up nature, only interval-censored data are available for the AD conversion time (Risacher et al., 2009). In addition to interval censoring, two other issues that make the analysis of the ADNI data difficult are the existences of both functional and missing covariates. Among others, for example, one functional covariate of clinical interest is the hippocampal shape, which is well-known to be associated with AD progression. Furthermore, several important baseline biomarkers have missing values since some participants declined some tests or imaging procedures. More details on the study and these issues will be given below.

Functional covariates, often derived from longitudinal measurements or neuroimaging among others, are increasingly important in survival analysis, especially in precision medicine. It is easy to see that their infinite-dimensional nature demands specialized statistical methods. Existing approaches include B-spline methods (Gellar et al., 2015), Bayesian frameworks (Lee et al., 2015) and kernel-based techniques (Hao et al., 2021; Qu et al., 2016). More recently, the use of functional principal component analysis (FPCA) has allowed dimension reduction by representing functional covariates via the Karhunen–Loève expansion, enabling the reduction of functional Cox models to finite-dimensional ones (Chen et al., 2025; Kong et al., 2018).

Many methods have been developed for regression analysis of failure time data with functional covariates but most of the existing methods are for right-censored data except Shi et al. (2022), who considered interval-censored data under the Cox proportional hazards (PH) model but did not provide theoretical justification. As pointed out by many authors, the analysis of interval-censored data is much more difficult than that of right-censored data since the former has a much more complicated structure than the latter. Also it should be noted that although FPCA can reduce the

functional covariate to a low-dimensional vector, there may exist other high-dimensional covariates. In addition, it is well-known that sometimes the PH model can be too restrictive and thus one may prefer more flexible models such as the class of LTMs, which includes the PH model as a special case and will be focused below.

Variable selection is required and plays a vital role in almost every data analysis field, especially with high-dimensional data. One popular approach for variable selection has been the penalized method, which optimizes an objective function along with a penalty function, and many such procedures have been proposed for failure time data (Du et al., 2025; Fan and Li, 2002; Fan et al., 2020; Li et al., 2020b; Liu et al., 2016; Shi et al., 2014; Zhang et al., 2016; Zhao et al., 2020). In particular, Li et al. (2020b) considered variable selection for LTMs based on interval-censored data and developed a penalized method. However, it does not seem to exist an established approach for the same problem but in the presence of functional covariates.

Missing data occur in many fields and two commonly used methods for their analysis are the imputation approach and the complete-case analysis approach. The former imputes the missing values and then applies the method developed for complete data, while the latter performs the analysis using only the subjects with no missing covariates. Although both are easy to implement, it is well-known that they may be inefficient and yield biased results (Li et al., 2020a; Little and Rubin, 2019; Lou et al., 2024, 2025b). In other words, one may want to develop a more appropriate method and this is especially true for estimation of the LTM based on interval-censored data with both functional and missing covariates, the focus of this paper. In the following, we will assume that missing covariates are missing-at-random.

The contributions of this paper are fourfold assuming that only interval-censored data are available. First in Section 2, we develop a sieve maximum likelihood estimation procedure for the LTM in the presence of functional covariates by using the sieve and FPCA techniques. Secondly in Section 3, we propose a penalized estimation approach to simultaneous variable selection and estimation of the LTM in the presence of functional covariates. For this, we employ the minimum approximated information criterion (MIC) and one key advantage of the proposed approach is that it avoids tuning parameter selection, a difficult issue or task for most penalized methods. Thirdly in Section 4, we generalize the approach proposed in Section 3 to the situation where there also exist missing covariates by using both FPCA and the inverse probability weighting (IPW) techniques. Finally, we establish the asymptotic properties of the resulting estimators by using the empirical process theory among others. To evaluate the empirical performance of the proposed

methods, an extensive simulation study is conducted in Section 5, and in Section 6, we apply the proposed methods to the ADNI data discussed above and obtain some new insights about the roles of high-dimensional imaging and clinical factors. Some discussion and concluding remarks are provided in Section 7.

2 Estimation of the LTM in the Presence of Functional Covariates

Consider a failure time study and let T denote the failure time of interest. Suppose that there exist a vector of baseline covariates and a square-integrable functional covariate denoted by $\boldsymbol{Z} \in \mathbb{R}^p$ and $X(\cdot)$, respectively. Also suppose that p may be large and $X_i(\cdot)$ is defined on a compact interval $\mathcal{I} = [0, \tau]$ with mean zero and covariance function $H(s,t) = \mathrm{Cov}(X_i(s), X_i(t))$. In the following, we will assume that given $\boldsymbol{Z}$ and $X(\cdot)$, T follows the LTM

$$\Lambda\{t \mid \boldsymbol{Z}, X(\cdot)\} = G\left(\int_0^t \exp\left[\boldsymbol{Z}^\top \boldsymbol{\beta} + \int_{\mathcal{I}} X(s)\,\gamma(s)\,ds\right] d\Lambda(u)\right) \tag{1}$$

with respect to the cumulative hazard function. In the above, $G(\cdot)$ denotes a known, strictly increasing transformation function, $\Lambda(t)$ an unspecified baseline cumulative hazard function, $\boldsymbol{\beta} \in \mathbb{R}^p$ a vector of regression coefficients, and $\gamma(\cdot)$ a square-integrable coefficient function. Without loss of generality, we will assume that the functional covariate has been centered.

As mentioned above, the model above offers significant flexibility and includes various widely used models as special cases. For instance, it yields the PH model and the proportional odds (PO) model with $G(x) = x$ and $G(x) = \log(1+x)$, respectively. For $G(\cdot)$, two commonly used classes of transformation functions are the Box-Cox transformation and logarithmic transformations, represented by $G(x) = [(1+x)^\rho - 1]/\rho$ and $G(x) = \log(1+\rho x)/\rho$, respectively, with $\rho \geq 0$. Note that as $\rho \to 0$, the former approaches the PO model and the latter approaches the PH model, and the estimation of the LTM is much more difficult than PH or PO model (Li et al., 2020b).

By Mercer's theorem, the covariance function $H(s,t)$ admits the representation $H(s,t) = \sum_{k=1}^{\infty} \lambda_k \psi_k(s)\psi_k(t)$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are the eigenvalues and $\{\psi_k\}_{k=1}^{\infty}$ are the associated orthonormal eigenfunctions. Then the Karhunen–Loève expansion yields

$$X_i(s) = \sum_{k=1}^{\infty} \zeta_{ik}\psi_k(s), \quad \text{where } \zeta_{ik} = \int_{\mathcal{I}} X_i(s)\,\psi_k(s)\,ds.$$

If there would exist an integer K such that $\lambda_k \approx 0$ for $k > K$, then model (1) would reduce to

$$\Lambda\{t \mid \boldsymbol{Z}_i, X_i(\cdot)\} \approx G\left(\Lambda(t)\exp\left[\boldsymbol{Z}_i^\top \boldsymbol{\beta} + \sum_{k=1}^{K} \alpha_k\,\widehat{\zeta}_{ik}\right]\right), \tag{2}$$

where $\alpha_k = \int_{\mathcal{I}} \gamma(s)\psi_k(s)\,ds$. In the following, we will treat the model above as the working model for estimation. Note that model (2) assumes that the relevant information in $X_i(\cdot)$ is adequately captured by the first K functional principal components, which can be selected by thresholding the cumulative variance explained.

Suppose that the study consists of n independent subjects and the observed data have the form $\{(L_i, R_i], \boldsymbol{Z}_i, X_i(\cdot)\}_{i=1}^n$ with $L_i < T_i \leq R_i$. That is, only interval-censored data are available for the T_i's. Define $\widetilde{\boldsymbol{\theta}} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \Lambda)^\top$ and assume independent interval censoring (Sun, 2006). Then the log likelihood function has the form

$$\ell(\widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^n \ell_i(\widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^n \log\left[S\Big\{L_i \mid \boldsymbol{Z}_i, X_i(\cdot)\Big\} - S\Big\{R_i \mid \boldsymbol{Z}_i, X_i(\cdot)\Big\}\right], \tag{3}$$

where $S(t \mid \boldsymbol{Z}_i, X_i(\cdot)) = \exp[-\Lambda\{t \mid \boldsymbol{Z}_i, X_i(\cdot)\}]$. To estimate $\widetilde{\boldsymbol{\theta}}$, it is apparent that a natural approach would be to directly maximize the log function above but this may not be easy due to the involvement of $\Lambda(t)$. Instead by following Zhou et al. (2017) and others, we suggest to employ the sieve maximum likelihood approach that approximates $\Lambda(t)$ using Bernstein polynomials.

More specifically, define the parameter space be $\Theta = \mathcal{B} \otimes \mathcal{M}$, where $\mathcal{B} = \{\boldsymbol{\xi} \equiv (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top \in \mathbb{R}^{p+K} : \|\boldsymbol{\xi}\| \leq M\}$ for some constant M and $\mathcal{M}$ is the set of all non-negative, continuous, non-decreasing functions on $[u, v]$ with $[u, v]$ covering the observed event times. Also define

$$\mathcal{M}_n = \left\{\Lambda_n(t) = \sum_{j=0}^m \phi_j^* B_j(t; m, u, v) : \sum_{j=0}^m |\phi_j^*| \leq M_n,\ 0 \leq \phi_0^* \leq \cdots \leq \phi_m^*\right\},$$

where

$$B_j(t; m, u, v) = \binom{m}{j}\left(\frac{t-v}{u-v}\right)^j\left(1 - \frac{t-v}{u-v}\right)^{m-j}, \quad j = 0, \ldots, m$$

are the Bernstein basis polynomials of degree m with $m = o(n^\nu)$ for some $\nu \in (0, 1)$. It is easy to see that ν characterizes the growth rate of m with respect to n. Then one can estimate $\boldsymbol{\xi}$ and Λ by maximizing $\ell(\boldsymbol{\xi}, \Lambda_n) = \ell(\widetilde{\boldsymbol{\theta}})$ over the sieve space $\Theta_n = \mathcal{B} \otimes \mathcal{M}_n$. Note that the monotonicity constraint above can be easily removed via the reparameterization $\phi_0^* = \exp(\phi_0)$ and $\phi_k^* = \sum_{j=0}^k \exp(\phi_j)$ for $k \geq 1$. Also the degree m can be chosen to satisfy $m = o(n^{1/4})$ (Lou et al., 2025a; Zhou et al., 2017), or alternatively by using an information criterion such as AIC or BIC (Bumham and Anderson, 2002).

3 Simultaneous Estimation and Variable Selection with Functional Covariates

Now we consider simultaneous estimation and variable selection for the LTM (1) with the primary goal of identifying significant nonzero elements in $\boldsymbol{\beta}$. Following Fan and Li (2001) and others, we propose to maximize the penalized log likelihood function

$$\ell(\widetilde{\boldsymbol{\theta}}) - \kappa\, p_{\text{pen}}(|\boldsymbol{\beta}|)\,, \tag{4}$$

where κ denotes a tuning parameter and $p_{\text{pen}}(\cdot)$ a penalty function. For $p_{\text{pen}}(\cdot)$, although many penalty functions can be used, it is well-known that the L_0 penalty is most appealing for its ability to directly penalize model size and produce parsimonious models (Ge et al., 2024; Su et al., 2016; Sun et al., 2019). Using the L_0 penalty, we consider the following objective function

$$H^{\sharp}(\widetilde{\boldsymbol{\theta}}) = \ell(\widetilde{\boldsymbol{\theta}}) - \log(n_0) \sum_{j=1}^{p} \mathbb{I}(\beta_j \neq 0), \tag{5}$$

where n_0 denotes the number of uncensored subjects. As shown in Sun et al. (2019), it avoids computationally intensive tuning while maintaining strong empirical performance.

While attractive in theory, the L_0 penalty can be computationally intractable in high dimensions. To address this, following Su et al. (2016), we adopt the MIC and propose to maximize

$$H(\widetilde{\boldsymbol{\theta}}) = \ell(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{\omega}[\boldsymbol{\varrho}]}) - \kappa \sum_{j=1}^{p} \omega(\varrho_j), \tag{6}$$

where $\widetilde{\boldsymbol{\theta}}_{\boldsymbol{\omega}[\boldsymbol{\varrho}]} = (\boldsymbol{\omega}[\boldsymbol{\varrho}]\boldsymbol{\varrho}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$, $\boldsymbol{\varrho} = (\varrho_1, \ldots, \varrho_p)^\top$ and $\boldsymbol{\omega}[\boldsymbol{\varrho}] = \text{diag}(\omega(\varrho_1), \ldots, \omega(\varrho_p))$. Here each β_j is reparameterized as $\beta_j = \varrho_j\, \omega(\varrho_j)$, where

$$\omega(\varrho_j) = \frac{\exp(2\, a\, \varrho_j^2 - 1)}{\exp(2\, a\, \varrho_j^2 + 1)}\,,$$

and $a > 0$ controls the sharpness of the approximation. After obtaining $\widehat{\varrho}_j$, one can estimate β_j by $\widehat{\beta}_j = \widehat{\varrho}_j\, \omega(\widehat{\varrho}_j)$.

Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\alpha}}^\top, \widehat{\Lambda})^\top$ denote the estimator of $\widetilde{\boldsymbol{\theta}}$ defined above, and let $\widehat{\gamma}(\cdot) = \sum_{k=1}^{K} \widehat{\alpha}_k \psi_k(\cdot)$ denote the estimator of the coefficient function $\gamma(\cdot)$. To establish the asymptotic properties of $\widehat{\boldsymbol{\theta}}$, let $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^\top, \boldsymbol{\alpha}_0, \Lambda_0)^\top$ denote the true value of $\widetilde{\boldsymbol{\theta}}$, where $\boldsymbol{\alpha}_0$ corresponds to the true coefficient function $\gamma_0(\cdot)$. Moreover, suppose that we can write $\boldsymbol{\beta}_0$ as $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{a0}^\top, \boldsymbol{\beta}_{b0}^\top)^\top$, where $\boldsymbol{\beta}_{a0} \in \mathbb{R}^{s_\beta}$ contains the nonzero elements and $\boldsymbol{\beta}_{b0} = \mathbf{0}$. For a vector $\boldsymbol{v}$, let $\|\boldsymbol{v}\|$ denote the Euclidean norm,

and for a function f, define $\|f\|_{L_2} = \left(\int |f|^2 \, d\mathbb{P}\right)^{1/2}$ and $\|f\|_\infty = \sup_t |f(t)|$. Also define

$$\|\Lambda\|_2^2 = \int \{\Lambda(l)^2 + \Lambda(r)^2\} \, dF(l, r),$$

where $F(l, r)$ denotes the joint distribution of the L_i's and R_i's. The following theorem establishes the consistency and oracle properties of the proposed estimator.

Theorem 1. *Suppose that the regularity conditions given in Web Appendix A of the Supplementary Material hold, $r > 2$, $\nu > 1/(2r)$, and $n_0 = O_p(n)$, where r and ν are defined in Condition 10 and the last paragraph of the previous section, respectively. Then we have that as $n \to \infty$,*

(i) (Consistency). $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \to 0$, $\|\widehat{\gamma}(\cdot) - \gamma_0(\cdot)\|_{L_2} \to 0$, and $\|\widehat{\Lambda} - \Lambda_0\|_2 \to 0$ in probability.

(ii) (Sparsity). With probability tending to 1, $\widehat{\boldsymbol{\beta}}_b = \mathbf{0}$.

(iii) (Asymptotic Normality). $\sqrt{n}(\widehat{\boldsymbol{\beta}}_a - \boldsymbol{\beta}_{a0}) \rightsquigarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is defined in the Supplementary Material.

The proof of the results above is sketched in Web Appendix B of the Supplementary Material. The theorem tells us that the proposed estimator is consistent and possesses the oracle property (Fan and Li, 2001). For the implementation of the proposed method above, we suggest to adopt the two-stage optimization strategy similar to that of Su et al. (2016). Specifically, we first apply the simulated annealing (Bélisle, 1992), a global optimization algorithm that helps to avoid the convergence to suboptimal local maxima and then refine the solution using the BFGS (Broyden–Fletcher–Goldfarb–Shanno) quasi-Newton method implemented in the MATLAB function 'fminunc'. This hybrid approach ensures that the final estimators converge to a stable critical point (Xie et al., 2024). For variance estimation, we suggest to use the Hessian matrix and the numerical study below indicates that it works well in practical situations.

4 Simultaneous Estimation and Variable Selection with Functional and Missing Covariates

In this section, we suppose that the baseline covariate $\boldsymbol{Z}$ consists of two parts and can be written as $\boldsymbol{Z} = (\boldsymbol{W}^\top, \boldsymbol{Q}^\top)^\top$, where $\boldsymbol{W}$ denotes the components that can be fully observed and $\boldsymbol{Q}$ the components subject to missing. That is, we have missing covariates. Also suppose that $\boldsymbol{Q}$ is q-dimensional and let $\boldsymbol{\chi}_i = (\chi_{i1}, \ldots, \chi_{iq})^\top$ denote the missing indicator for subject i with $\chi_{ij} = 1$

if Q_{ij} is observed and $\chi_{ij} = 0$ otherwise. Define $\chi_{i0} = \prod_{j=1}^{q} \chi_{ij}$, indicating whether subject i has complete covariate data. Then the observed data have the form

$$O = \left\{ O_i = \left((L_i, R_i],\, \boldsymbol{W}_i,\, \chi_{i0},\, \boldsymbol{\chi}_i,\, \boldsymbol{\chi}_i \times \boldsymbol{Q}_i,\, X_i(\cdot) \right) \right\}.$$

In the following, we will assume that

$$\mathbb{P}(\chi_{i,j} = 1 \mid L_i, R_i, \boldsymbol{W}_i, \boldsymbol{Q}_i, X_i(\cdot)) = \mathbb{P}(\chi_{i,j} = 1 \mid L_i, R_i, \boldsymbol{W}_i, X_i(\cdot)), \quad j = 1, \ldots, q.$$

That is, we have missing at random (Little and Rubin, 2019; Lou et al., 2024). Define

$$\pi_i(\wp) = \mathbb{P}\left\{ \chi_{i0} = 1 \mid L_i, R_i, \boldsymbol{W}_i, X_i(\cdot); \wp \right\},$$

the probability that subject i has no missing covariates, where $\wp$ denotes a vector of unknown parameters. For $\pi_i(\wp)$, one may employ any parametric model and a commonly used one is the logistic model

$$\pi_i(\wp) = \frac{\exp\left(\boldsymbol{U}_i^\top \wp\right)}{1 + \exp\left(\boldsymbol{U}_i^\top \wp\right)},$$

where $\boldsymbol{U}_i$ may contain part or all components of $\{L_i, R_i, \boldsymbol{W}_i, X_i(\cdot)\}$.

For variable selection and estimation, as discussed above, one naive and simple approach is complete-case analysis, meaning to employ the objective function

$$H_{\mathrm{CC}}^{\sharp}(\widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \chi_{i0}\, \ell_i(\widetilde{\boldsymbol{\theta}}) - \log(\widetilde{n}_0) \sum_{j=1}^{p} \mathbb{I}(\beta_j \neq 0) \tag{7}$$

corresponding to that given in equation (5). In the above, $\ell_i(\widetilde{\boldsymbol{\theta}})$ denotes $\ell(\widetilde{\boldsymbol{\theta}})$ defined above but based only the data from subject i and $\widetilde{n}_0$ the number of uncensored observations with complete data. Also as discussed above, this may result in significant information loss and bias. Correspondingly, by using the IPW technique and following the discussion in the previous section, we propose to maximize the weighted penalized log likelihood function

$$H_{\mathrm{IPW}}^{\sharp}(\widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \frac{\chi_{i0}}{\pi_i(\widehat{\wp})}\, \ell_i(\widetilde{\boldsymbol{\theta}}) - \log(\widetilde{n}_0) \sum_{j=1}^{p} \omega(\varrho_j) \tag{8}$$

corresponding to the objective function given in (6), where $\widehat{\wp}$ denotes a consistent estimator of $\wp$. It is easy to see that this approach leverages both complete and partially observed covariates, offering reduced bias compared to complete-case analysis.

In the following, for notation simplicity, we will use the same $\widehat{\boldsymbol{\theta}}$ defined in the previous section to denote the IPW estimator defined above. Also suppose that $\boldsymbol{\beta}$ can be written as two parts as

before and let $\widetilde{n}$ denote the number of subjects with no missing covariates. As with no missing covariate situation, we establish the asymptotic properties of the IPW estimator.

Theorem 2. *Suppose that the conditions described in Theorem 1 and the additional conditions given in Web Appendix B of the Supplementary Material hold and* $\widetilde{n}_0 = O_p(\widetilde{n})$. *Then we have that as* $\widetilde{n} \to \infty$,

(i) (Consistency). $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \to 0$, $\|\widehat{\gamma}(\cdot) - \gamma_0(\cdot)\|_{L_2} \to 0$, *and* $\|\widehat{\Lambda} - \Lambda_0\|_2 \to 0$ *in probability.*

(ii) (Sparsity). With probability tending to one, $\widehat{\boldsymbol{\beta}}_b = \mathbf{0}$.

(iii) (Asymptotic Normality). $\sqrt{n}(\widehat{\boldsymbol{\beta}}_a - \boldsymbol{\beta}_{a0}) \rightsquigarrow \mathcal{N}(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}})$, *where* $\widetilde{\boldsymbol{\Sigma}}$ *is defined in Section 3 of the Supplementary Material.*

The proof of the theorem above is sketched in Web Appendix B of the Supplementary Material and it indicates that the proposed IPW estimator is also consistent and possesses the oracle property. For the implementation of the estimation procedure above, the same optimization strategy described in the previous section can be used, and the numerical study below indicates that it works well. For variance estimation, note that the estimation procedure here involves two steps and the uncertainty from the first step, modeling the missing mechanism, should be properly accounted for. To this end, we recommend using the nonparametric bootstrap procedure and the numerical results below suggest that it works well practically.

5 A Simulation Study

An extensive simulation study was conducted to evaluate the empirical performance of the estimation procedures proposed in the previous sections. In the study, we first generated the baseline covariates $\boldsymbol{Z}_i$'s from the multivariate normal distribution with mean zero, unit variance and the correlation $\text{corr}(Z_{ij}, Z_{ik}) = \exp(-|j-k|)$. For the functional covariate $X_i(s)$, it was generated based on

$$X_{i0}(s_{ij}) = \widetilde{\alpha}_{i1}\psi_1(s_{ij}) + \widetilde{\alpha}_{i2}\psi_2(s_{ij}) + \epsilon_{ij}\,, \quad i = 1, \ldots, n$$

with the design points s_{ij}'s being the uniform grid over $[0, 1]$ with spacing 0.001. In the above, $\psi_1(s) = \sqrt{2}\sin(2\pi s)$, $\psi_2(s) = \sqrt{2}\cos(2\pi s)$, and the $\widetilde{\alpha}_{i1}$'s and $\widetilde{\alpha}_{i2}$'s were generated independently from Normal$(0, 3^2)$ and Normal$(0, 1^2)$, respectively. Furthermore, the noise errors ϵ_{ij}'s were drawn from Normal$(0, \varepsilon^2)$ with $\varepsilon = 0.1$ or 0.5, corresponding to low or moderate measurement error, respectively, and 25% of the grid points s_{ij}'s were randomly selected to serve as the observation

time points. Given the covariates above, the failure times T_i's were generated from model (1) with $\Lambda(t) = 0.5\, t^{1.2}$ and $G(x) = \log(1 + \rho x)/\rho$.

For the generation of the observed interval-censored data or censoring intervals, it was assumed that for each subject, there exists a sequence of examination or observation times with the inter-visit times generated from Uniform$(0.1, 0.3)$ and the maximum follow-up time τ, which can be used to control the right censoring rate (CR). Then the observed interval $(L_i, R_i]$ was constructed with L_i taken to be the last observation time before T_i and R_i the first observation time after T_i. For the results given below, we set the true regression coefficients to be $\boldsymbol{\beta}_0 = (0.7,\ 0.7,\ 0.7,\ \mathbf{0}_{p-3})^\top$ and $\boldsymbol{\alpha}_0 = (0.5,\ 0.8)^\top$ and $\tau = 2.5$ and 5, corresponding to approximate 50% and 35% CR, respectively. Also we took $n = 200$ or 400, $p = 10$ or 30, and $m = 3$ with 200 replications.

Table 1 and Table S1 in Web Appendix C of the Supplementary Material present the simulation results given by the estimation procedure proposed in Section 3 with no missing covariates, $\varepsilon = 0.1$ and $\rho = 0$ or 1, corresponding to the PH and PO models, respectively. In Table 1, we provide the results on the variable selection, while in Table S1 of Web Appendix C of the Supplementary Material, the results on estimation of the regression coefficients $\boldsymbol{\beta}$ and the functional coefficient $\boldsymbol{\alpha}$ are reported. Note that here for simplicity, we fixed the number of components $K = 2$. We did consider the approach that selected K adaptively based on the pre-specified proportion of variance explained or BIC and found that $K = 2$ typically captures more than 95% of the total variability.

In Table 1, we calculated the average number of true positives or correctly identified nonzero coefficients (TP), the number of the false positives or incorrectly identified nonzero coefficients (FP), and the median mean squared error (MMSE) for estimating the regression parameter along with its standard deviation (STD). For comparison, we also applied and include the results in the table given by two other methods, the sieve maximum likelihood estimation (MLE) described in Section 2 and the oracle estimation that assumes that the true model is known. One can see from the table that the proposed method consistently yielded TP values close to the true number of nonzero coefficients and MMSE values that are comparable to those given by the oracle estimation and substantially lower than those given by the MLE.

To assess the performance on estimation, in Table S1 in Web Appendix C of the Supplementary Material and for each of the nonzero components of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, we calculated the average bias (Bias), the empirical standard error (SSE), the average of the estimated standard errors (SEE), and the 95% empirical coverage probability (CP). They suggest that the proposed estimator seems to be unbiased and the variance estimate is accurate. Also the coverage probabilities are close to the

nominal level, indicating that the normal approximation to the distribution of the estimator is appropriate. Table S2 in Web Appendix C of the Supplementary Material presents the simulation results on variable selection given by the estimation procedure proposed in Section 3 with $\varepsilon = 0.5$. They are similar to those given above except that FP is slightly larger than before due to higher noise and again indicate that the proposed method seems to perform well. The results on the parameter estimation are also similar to those given above and omitted.

In the above, we considered the situation with no missing covariates. For the situation with missing covariates, it was assumed that the covariate Z_{i3} may be missing according to

$$\mathbb{P}(\chi_i = 1 \mid O_i) = \left\{1 + \exp\left(-\wp_0 + Z_{i1} - Z_{i2} - Z_{i4} + 0.5Z_{ip} + 0.5L_i + 0.5R_i\right)\right\}^{-1}.$$

Here $\chi_i = 1$ indicates that Z_{i3} is observed and the intercept $\wp_0$ was determined to yield approximately 30% missingness. Table 2 and Table S3 in Web Appendix C of the Supplementary Material present the results given by the IPW estimation procedure proposed in Section 4 with $n = 400$, $p = 10$ and all of the other settings being the same as above. The former gives the results on variable selection and the latter contains the results on parameter estimation. For comparison, we also applied the complete-case analysis method and include the obtained results in the tables.

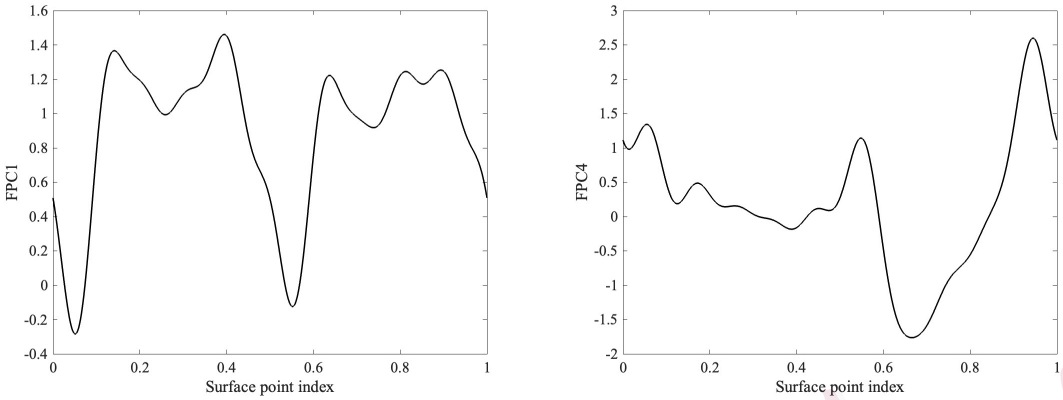
One can see from Table 2 that the proposed IPW estimation procedure and the complete-case analysis gave similar results in terms of TP and MMSE but the former tended to yield larger FP than the latter maybe due to the inclusion of non-informative covariates in the logistic model estimation for the π_i's. The major difference between the two methods is in the parameter estimation. It is apparent from Table S3 in Web Appendix C of the Supplementary Material that the IPW estimator seems to be unbiased and its variance estimation procedure and CP appear appropriate. In contrast, the complete-case estimator seems to be biased and has poor CP. These findings suggest that one may be careful to apply the complete-case analysis since it can yield misleading inferences. Note that for the variance estimation above, 30 bootstrap samples were used to save the computation time. Nevertheless the results seem to be reasonable and stable and to further investigate this, we also tried 100 bootstrap samples and present the results in Web Appendix C of the Supplementary Material. One can see that the results are similar and suggest that 30 is sufficient for the current setting.

6 Analysis of the ADNI study

In this section, we apply the estimation procedures proposed in the previous sections to the ADNI described above. As mentioned before, it is a longitudinal study with a primary objective of identifying various biomarkers for early detection and monitoring of AD. The participants in the study were examined intermittently and classified based on their cognitive conditions into three groups, CN, MCI, and AD. Among others, one variable of interest is the time to the AD conversion and due to the nature of the study, only interval-censored data are available for the occurrence time of the AD conversion, the failure time of interest here. For the analysis here, following Han et al. (2017) and Li et al. (2017, 2020b), we will focus on the participants in the MCI group and are interested in identifying the factors that had significant effects on the risk of developing the AD.

The study includes three types of baseline covariates or factors, demographic, clinical, and imaging-based ones plus functional covariates. The demographic factors consist of age, gender, race, and years of education (EDU). The clinical ones include the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS13), the delayed word recall subscore of ADAS (ADASQ4), the Functional Assessment Questionnaire score (FAQ), the Mini-Mental State Examination (MMSE), the Rey Auditory Verbal Learning Test immediate recall score (RAVLTI) and learning ability score (RAVLTL), the Trails B score (TRABS), and the digit symbol substitution test score (DIGITS). The imaging-based factors, derived from MRI scans, are the hippocampal volume (HIPP), the middle temporal gyrus volume (MIDTEMP), the entorhinal cortex volume (ENTORH), and the fusiform gyrus volume (FUSIF). For the functional covariate, we are interested in the hippocampal shape, for which the radial distance measurements were extracted from approximately 30,000 surface points across both hippocampi (Kong et al., 2018). These distances, defined from each vertex to the medial axis, summarize both size and local shape. As mentioned above, there exist some baseline biomarkers or factors that have missing values since some participants declined some tests or imaging procedures. More specifically, 82 patients had missing values on the factors including HIPP, MIDTEMP, ENTORH, and FUSIF.

Table 3 presents the estimation results on the baseline factors given by the estimation procedure proposed in Section 4, which is referred to as IPW, and they include the estimated effects (EST) and standard errors (SE) along with the p-values for testing the effect to be zero (PVAL). Here we considered the PH and PO models and retained eight FPC's, which explained 90% of the total



(a) FPC1 (b) FPC4

Figure 1: Estimated FPCs of the hippocampal radial distance profile.

variability. In addition, the standard errors were obtained based on 100 bootstrap samples. For comparison, we also applied the complete-case analysis method, which is referred to as Complete-Case, and the IPW and Complete-Case but omitting the functional covariate. The latter two methods are referred to as IPW-Naive and Complete-Case-Naive in the table and are equivalent to applying the IPW and Complete-Case under the restriction $\alpha_1 = \cdots = \alpha_K = 0$.

On the proposed IPW estimation procedure, one can see from Table 3 that it gave consistent results between the PH and PO models and identified three significant predictors, ADAS13, FAQ and MIDTEMP, for the AD conversion. In particular, it suggests that the higher FAQ scores were positively associated with the increased conversion risk and the higher MIDTEMP values were negatively associated with risk. In contrast, the results from either the complete-case analysis method or the other two naive methods are less consistent. Furthermore, the complete-case analysis identified more significant factors and in particular, it selected AGE as a significant predictor and indicates that it was negatively significantly associated with the AD conversion. This implies that the older patients had a lower risk of AD conversion, which is apparently counterintuitive and biologically implausible, likely due to the bias introduced by excluding the subjects with missing data. Compared to both the proposed method and the complete-case analysis, the methods ignoring the functional covariate selected more factors and also identified more significant factors but missed the two commonly recognized factors, FAQ and MIDTEMP.

Table 4 gives the results on estimation of the coefficients associated with the FPC scores and the p-values for testing each of them being zero, and they consistently suggest that FPC1 was significant. In addition, under the PH model, the proposed method additionally identified FPC4 as a significant predictor. To better understand the nature of these components, we visualized the

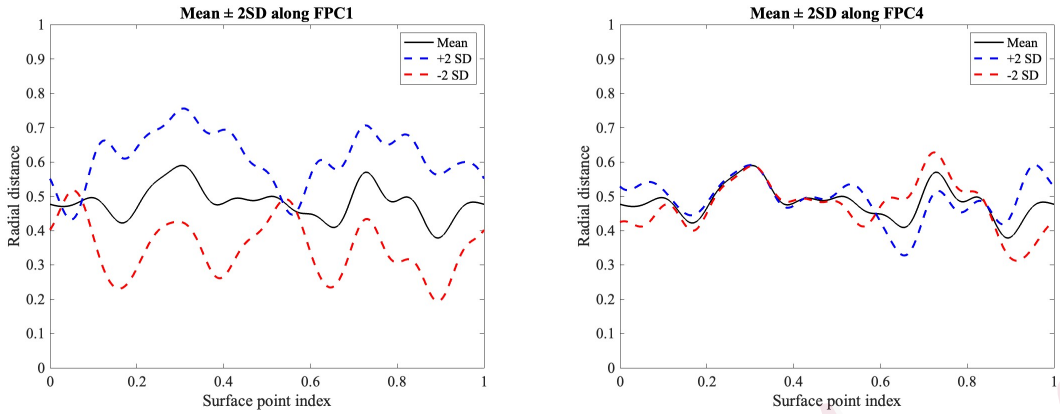


Figure 2: The shape variability along the first and fourth FPCs. Each panel shows the mean hippocampal radial distance profile (black) and its ±2 standard deviation perturbations along the corresponding FPC (blue and red).

estimated eigenfunctions in Figure 1. The FPC1 exhibits smooth, large-scale variation over the surface domain, while FPC4 shows more localized, higher-frequency oscillations, suggesting it may capture finer-scale anatomical differences. To further illustrate these modes of variation, Figure 2 presents the mean radial profile and its perturbations by ±2 standard deviations along FPC1 and FPC4. The deformations along FPC1 reflect broad and coherent shifts in shape, consistent with global hippocampal atrophy or enlargement. In contrast, FPC4 reveals more regionally concentrated deviations, pointing to potential localized shape differences across individuals. While these components do not directly map onto anatomical subfields, their statistical significance suggests that both global and regional hippocampal morphology may contain meaningful prognostic information for identifying patients at higher risk of AD conversion (Gentreau et al., 2023; Macdonald et al., 2013).

7 Discussion and Concluding Remarks

This paper considered regression analysis of interval-censored failure time data in the presence of both functional and missing covariates with the focus on simultaneous estimation and variable selection. As discussed above, many methods have been developed for regression analysis of failure time data with functional covariates but most of them apply only to right-censored data. To our knowledge, the proposed methods are the first that allow interval censoring in addition to missing covariates. Also as pointed out before, a key advantage of the proposed methods is that they circumvent the need for tuning parameter selection, a common challenge in penalization-based approaches. The oracle property of the resulting estimators was established and the extensive sim-

ulation study demonstrated their strong finite-sample performances. In addition, the application of the proposed method to the ADNI provided some new insights about the risk factors for AD progression.

It is worth to point out that for simplicity, in the preceding sections, we only considered one functional covariate and it is apparent that in practice, there may exist more than one functional covariate. It is straightforward to generalize the proposed methods to the latter situation. In the proposed estimation procedures, Bernstein polynomials were employed to approximate the unknown baseline cumulative hazard function. As an alternative, one may apply other smooth functions like splines and the proposed methods are still valid. A main advantage of Bernstein polynomials is that they naturally guarantee the required monotonicity and positivity. Also it is worth to note that although the LTM (1) is quite flexible, sometimes one may still prefer some other models such as the additive hazards or accelerated failure time model. The ideas discussed above should apply to these models and one can develop similar estimation procedures.

There exist several directions for future research related to the proposed methodology. One is that in the above, we have assumed independent or non-informative interval censoring and it is apparent that sometimes this may not hold or the interval censoring may be dependent or informative (Sun, 2006; Sun and Chen, 2022). For the latter situation, one commonly used approach is to jointly model the failure time of interest and censoring mechanism or process together using latent variables or copula functions. Although the general strategy used above should be applicable, such extensions would introduce significant computational challenges among others. Another assumption used above is that the effects of covariates are linear and in practice, nonlinear effects may be present. Thus it would be useful to generalize the proposed methods to accommodate nonlinear covariate effects. In addition, the focus above has been on an univariate failure outcome of interest and one may want to generalize the proposed estimation procedures to allow multivariate failure outcomes of interest.

Supporting Information

Supplementary Material contains the required regularity conditions and the proofs of Theorems 1 and 2 as well as some extra numerical results.

Acknowledgements

The authors wish to thank the Co-Editor, Dr. John Stufken, the Associate Editor and a reviewer for their many insightful and valuable comments and suggestions that greatly improved the paper. The research was partly supported by Scientific Research Foundation of the Education Department of Jilin Province (Grant No. JJKH20261483KJ) to the second author.

References

Bélisle, C. J. (1992). Convergence theorems for a class of simulated annealing algorithms on rd. *Journal of Applied Probability*, 29(4):885–895.

Bumham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. Springer.

Chen, X., Liu, H., Men, J., and You, J. (2025). High-dimensional partially linear functional cox models. *Biometrics*, 81(1):ujae164.

Du, M., Lou, Y., and Sun, J. (2025). Estimation and variable selection for interval-censored failure time data with random change point and application to breast cancer study. *Journal of the American Statistical Association*, 120(552):2276–2287.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Fan, J. and Li, R. (2002). Variable selection for cox's proportional hazards model and frailty model. *The Annals of Statistics*, 30(1):74–99.

Fan, J., Li, R., Zhang, C.-H., and Zou, H. (2020). *Statistical Foundations of Data Science*. Chapman and Hall/CRC.

Ge, L., Hu, T., and Li, Y. (2024). Simultaneous variable selection and estimation in semiparametric regression of mixed panel count data. *Biometrics*, 80(1):ujad041.

Gellar, J. E., Colantuoni, E., Needham, D. M., and Crainiceanu, C. M. (2015). Cox regression models with functional covariates for survival data. *Statistical Modelling*, 15(3):256–278.

Gentreau, M., Maller, J. J., Meslin, C., Cyprien, F., Lopez-Castroman, J., and Artero, S. (2023). Is hippocampal volume a relevant early marker of dementia? *The American Journal of Geriatric Psychiatry*, 31(11):932–942.

Han, X., Zhang, Y., Shao, Y., and Initiative, A. D. N. (2017). Application of concordance probability estimate to predict conversion from mild cognitive impairment to alzheimer's disease. *Biostatistics & Epidemiology*, 1(1):105–118.

Hao, M., Liu, K.-y., Xu, W., and Zhao, X. (2021). Semiparametric inference for the functional cox model. *Journal of the American Statistical Association*, 116(535):1319–1329.

Kong, D., Ibrahim, J. G., Lee, E., and Zhu, H. (2018). Flcrm: Functional linear cox regression model. *Biometrics*, 74(1):109–117.

Lee, E., Zhu, H., Kong, D., Wang, Y., Giovanello, K. S., and Ibrahim, J. G. (2015). Bflcrm: A bayesian functional linear cox regression model for predicting time to conversion to alzheimer's disease. *The Annals of Applied Statistics*, 9(4):2153–2178.

Li, H., Zhang, H., Zhu, L., Li, N., and Sun, J. (2020a). Estimation of the additive hazards model with interval-censored data and missing covariates. *Canadian Journal of Statistics*, 48(3):499–517.

Li, K., Chan, W., Doody, R. S., Quinn, J., Luo, S., and Initiative, A. D. N. (2017). Prediction of conversion to alzheimer's disease with longitudinal measures and time-to-event data. *Journal of Alzheimer's Disease*, 58(2):361–371.

Li, S., Wu, Q., and Sun, J. (2020b). 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.

Little, R. J. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Liu, H., Yao, T., and Li, R. (2016). Global solutions to folded concave penalized nonconvex learning. *The Annals of Statistics*, 44(2):629–659.

Lou, Y., Ma, Y., and Du, M. (2024). A new and unified method for regression analysis of interval-censored failure time data under semiparametric transformation models with missing covariates. *Statistics in Medicine*, 43(11):2062–2082.

Lou, Y., Ma, Y., Sun, J., Wang, P., and Ye, Z. (2025a). Instrumental variable estimation of complier casual treatment effects with interval-censored competing risks data. *Biometrics*, 81(1):ujaf010.

Lou, Y., Ma, Y., Xiang, L., and Sun, J. (2025b). A multiple imputation approach for flexible modelling of interval-censored data with missing and censored covariates. *Computational Statistics & Data Analysis*, 209:108177.

Macdonald, K. E., Bartlett, J. W., Leung, K. K., Ourselin, S., Barnes, J., et al. (2013). The value of hippocampal and temporal horn volumes and rates of change in predicting future conversion to ad. *Alzheimer Disease & Associated Disorders*, 27(2):168–173.

Qu, S., Wang, J.-L., and Wang, X. (2016). Optimal estimation for the functional cox model. *The Annals of Statistics*, 44(4):1708–1738.

Risacher, S. L., Saykin, A. J., Wes, J. D., Shen, L., Firpi, H. A., and McDonald, B. C. (2009). Baseline mri predictors of conversion from mci to probable ad in the adni cohort. *Current Alzheimer Research*, 6(4):347–361.

Shi, H., Ma, D., Faisal Beg, M., and Cao, J. (2022). A functional proportional hazard cure rate model for interval-censored data. *Statistical Methods in Medical Research*, 31(1):154–168.

Shi, Y.-Y., Cao, Y.-X., Jiao, Y.-L., and Liu, Y.-Y. (2014). Sica for cox's proportional hazards model with a diverging number of parameters. *Acta Mathematicae Applicatae Sinica, English Series*, 30(4):887–902.

Su, X., Wijayasinghe, C. S., Fan, J., and Zhang, Y. (2016). Sparse estimation of cox proportional hazards models via approximated information criteria. *Biometrics*, 72(3):751–759.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer.

Sun, J. and Chen, D.-G. (2022). *Emerging Topics in Modeling Interval-Censored Survival Data*. Springer.

Sun, L., Li, S., Wang, L., and Song, X. (2019). Variable selection in semiparametric nonmixture cure model with interval-censored failure time data: An application to the prostate cancer screening study. *Statistics in Medicine*, 38(16):3026–3039.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., Donohue, M. C., Green, R. C., Harvey, D., Jack Jr, C. R., et al. (2015). Impact of the alzheimer's disease neuroimaging initiative, 2004 to 2014. *Alzheimer's & Dementia*, 11(7):865–884.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Liu, E., et al. (2013). The alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer's & Dementia*, 9(5):e111–e194.

Xie, M., Hu, T., and Zhou, J. (2024). Transfer learning under the cox model with interval-censored data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 17(2):e11680.

Zhang, X., Wu, Y., Wang, L., and Li, R. (2016). Variable selection for support vector machines in moderately high dimensions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(1):53–76.

Zhao, H., Wu, Q., Li, G., and Sun, J. (2020). Simultaneous estimation and variable selection for interval-censored data with broken adaptive ridge regression. *Journal of the American Statistical Association*, 115(529):204–216.

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.

Table 1: Simulation results on variable selection with no missing covariates and $\varepsilon = 0.1$.

				$n = 200$				$n = 400$			
Model	p	CR	Method	TP	FP	MMSE	STD	TP	FP	MMSE	STD
PH	10	35%	Oracle	3	0	0.0252	0.0373	3	0	0.0110	0.0127
			Proposed	3	0.035	0.0266	0.0441	3	0.030	0.0110	0.0153
			MLE	—	—	0.1011	0.0716	—	—	0.0414	0.0267
		50%	Oracle	3	0	0.0288	0.0395	3	0	0.0128	0.0160
			Proposed	3	0.040	0.0336	0.0419	3	0.020	0.0130	0.0174
			MLE	—	—	0.1235	0.0837	—	—	0.0470	0.0325
	30	35%	Oracle	3	0	0.0256	0.0521	3	0	0.0130	0.0199
			Proposed	3	0.050	0.0272	0.0615	3	0.040	0.0143	0.0212
			MLE	—	—	0.4646	0.2572	—	—	0.1669	0.0625
		50%	Oracle	3	0	0.0316	0.0711	3	0	0.0170	0.0219
			Proposed	3	0.045	0.0347	0.0732	3	0.045	0.0174	0.0247
			MLE	—	—	0.5829	0.3498	—	—	0.2025	0.0762
PO	10	35%	Oracle	3	0	0.0618	0.0700	3	0	0.0237	0.0284
			Proposed	2.930	0.030	0.0625	0.1201	3	0.030	0.0241	0.0332
			MLE	—	—	0.2221	0.1232	—	—	0.0955	0.0543
		50%	Oracle	3	0	0.0632	0.0678	3	0	0.0284	0.0368
			Proposed	2.860	0.045	0.0709	0.1448	3	0.040	0.0322	0.0407
			MLE	—	—	0.2501	0.1405	—	—	0.1090	0.0647
	30	35%	Oracle	3	0	0.0590	0.0759	3	0	0.0290	0.0307
			Proposed	2.980	0.005	0.0605	0.0896	3	0.020	0.0293	0.0328
			MLE	—	—	0.8522	0.3940	—	—	0.3608	0.1189
		50%	Oracle	3	0	0.0637	0.0915	3	0	0.0372	0.0376
			Proposed	2.980	0.050	0.0665	0.1129	3	0.020	0.0375	0.0465
			MLE	—	—	1.0204	0.5338	—	—	0.4268	0.1493

Table 2: Simulation results on variable selection with missing covariates.

				Complete-Case Analysis				IPW Analysis			
Model	ε	CR	Method	TP	FP	MMSE	STD	TP	FP	MMSE	STD
PH	0.1	35%	Oracle	3	0	0.0569	0.0439	3	0	0.0430	0.0385
			Proposed	3	0.035	0.0584	0.0449	3	0.225	0.0531	0.0486
		50%	Oracle	3	0	0.0403	0.0318	3	0	0.0434	0.0433
			Proposed	3	0.040	0.0425	0.0344	3	0.185	0.0544	0.0567
	0.5	35%	Oracle	3	0	0.0569	0.0439	3	0	0.0430	0.0387
			Proposed	3	0.040	0.0588	0.0453	3	0.245	0.0538	0.0507
		50%	Oracle	3	0	0.0404	0.0319	3	0	0.0432	0.0431
			Proposed	2.995	0.055	0.0453	0.0485	3	0.180	0.0538	0.0559
PO	0.1	35%	Oracle	3	0	0.1095	0.0726	3	0	0.0947	0.0976
			Proposed	2.935	0.07	0.1352	0.1231	2.970	0.350	0.1385	0.1410
		50%	Oracle	3	0	0.0890	0.0676	3	0	0.0939	0.0854
			Proposed	2.945	0.07	0.1135	0.1154	2.965	0.215	0.1281	0.1304
	0.5	35%	Oracle	3	0	0.1095	0.0724	3	0	0.0943	0.0954
			Proposed	2.935	0.075	0.1350	0.1214	2.970	0.320	0.1333	0.1510
		50%	Oracle	3	0	0.0890	0.0674	3	0	0.0938	0.0848
			Proposed	2.940	0.07	0.1152	0.1185	2.970	0.245	0.1294	0.1310

Table 3: Analysis results on the baseline factors for the ADNI.

		Complete-Case-Naive			IPW-Naive			Complete-Case			IPW		
Model	FACTOR	EST	SE	PVAL	EST	SE	PVAL	EST	SE	PVAL	EST	SE	PVAL
PH	AGE	—	—	—	—	—	—	-0.542	0.105	0.000	—	—	—
	GENDER	—	—	—	-1.200	0.349	0.001	—	—	—	—	—	—
	EDU	0.113	0.035	0.001	0.119	0.090	0.185	—	—	—	—	—	—
	RACE	1.401	0.459	0.002	—	—	—	—	—	—	—	—	—
	ADAS13	0.202	0.027	0.000	0.225	0.096	0.019	0.054	0.025	0.032	0.073	0.035	0.037
	ADASQ4	—	—	—	—	—	—	—	—	—	—	—	—
	FAQ	0.142	0.019	0.000	0.133	0.074	0.074	0.061	0.019	0.001	0.066	0.028	0.017
	RAVLTI	-0.049	0.018	0.005	-0.077	0.035	0.025	-0.047	0.014	0.001	-0.030	0.018	0.090
	RAVLTL	-0.141	0.047	0.003	—	—	—	—	—	—	—	—	—
	MMSE	—	—	—	—	—	—	—	—	—	—	—	—
	TRABS	0.002	0.002	0.193	—	—	—	—	—	—	—	—	—
	DIGITS	-0.034	0.012	0.004	-0.032	0.043	0.466	—	—	—	—	—	—
	HIPP	-0.059	0.016	0.000	-0.070	0.042	0.097	—	—	—	0.016	0.017	0.356
	MIDTEMP	-0.031	0.006	0.000	-0.021	0.018	0.231	-0.014	0.005	0.002	-0.015	0.006	0.008
	ENTORH	—	—	—	0.001	0.004	0.794	-0.003	0.002	0.101	-0.003	0.002	0.188
	FUSIF	-0.016	0.006	0.004	-0.011	0.021	0.597	—	—	—	—	—	—
PO	AGE	—	—	—	—	—	—	-0.891	0.153	0.000	—	—	—
	GENDER	—	—	—	—	—	—	—	—	—	—	—	—
	EDU	—	—	—	—	—	—	—	—	—	—	—	—
	RACE	3.141	0.709	0.000	3.141	1.177	0.008	—	—	—	—	—	—
	ADAS13	0.142	0.042	0.001	0.153	0.122	0.212	—	—	—	0.085	0.043	0.046
	ADASQ4	—	—	—	—	—	—	—	—	—	—	—	—
	FAQ	0.237	0.035	0.000	0.240	0.129	0.062	0.112	0.029	0.000	0.123	0.052	0.019
	RAVLTI	-0.107	0.024	0.000	-0.110	0.081	0.175	-0.071	0.020	0.000	-0.044	0.025	0.082
	RAVLTL	—	—	—	—	—	—	—	—	—	—	—	—
	MMSE	—	—	—	—	—	—	—	—	—	—	—	—
	TRABS	0.001	0.003	0.590	—	—	—	—	—	—	—	—	—
	DIGITS	—	—	—	—	—	—	-0.019	0.015	0.214	—	—	—
	HIPP	—	—	—	—	—	—	—	—	—	0.014	0.020	0.479
	MIDTEMP	-0.034	0.007	0.000	-0.035	0.019	0.063	-0.017	0.006	0.007	-0.017	0.008	0.034
	ENTORH	-0.009	0.003	0.001	-0.009	0.008	0.268	-0.005	0.002	0.021	-0.003	0.003	0.176
	FUSIF	0.007	0.009	0.391	0.009	0.013	0.523	—	—	—	—	—	—

Table 4: Analysis results on the functional covariate for the ADNI.

		Complete-Case Analysis			IPW Analysis		
Model	α	EST	SE	PVAL	EST	SE	PVAL
PH	α_1	-0.3009	0.1301	0.021	-0.3084	0.1340	0.021
	α_2	0.1467	0.0908	0.106	0.0702	0.1197	0.558
	α_3	-0.1697	0.1092	0.120	-0.1709	0.1346	0.204
	α_4	0.2159	0.0934	0.021	0.2126	0.1080	0.049
	α_5	-0.0314	0.0838	0.708	0.0005	0.1008	0.996
	α_6	0.1080	0.1039	0.299	-0.0371	0.1092	0.734
	α_7	0.0254	0.0926	0.784	0.0275	0.1142	0.810
	α_8	0.0302	0.1027	0.769	-0.0077	0.1231	0.950
PO	α_1	-0.4140	0.1616	0.010	-0.3394	0.1513	0.025
	α_2	0.2646	0.1298	0.042	0.1654	0.1492	0.268
	α_3	-0.2532	0.1472	0.085	-0.2254	0.1771	0.203
	α_4	0.1615	0.1260	0.200	0.2132	0.1368	0.119
	α_5	-0.0968	0.1261	0.443	0.1176	0.1505	0.435
	α_6	0.1655	0.1392	0.235	-0.0320	0.1658	0.847
	α_7	0.1117	0.1338	0.404	-0.0561	0.1581	0.723
	α_8	-0.0687	0.1470	0.640	0.0298	0.1489	0.841