REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000–000
https://doi.org/00.00000/revstat.v00i0.000

Shrinkage Estimation in Sparse Partially Linear Models with Right-Censored Data

Authors: Syed Ejaz Ahmed
– Department of Mathematics and Statistics, Brock University, St. Catharines, Ontario, Canada, L2S 3A1
sahmed5@brocku.ca

Dursun Aydin
– Department of Statistics, Mugla Sitki Kocman University, Mugla, Turkey, 48000
dursun.aydin@mu.edu.tr

Ersin Yilmaz ✉
– Department of Statistics, Mugla Sitki Kocman University, Mugla, Turkey, 48000
ersin.yilmaz@mu.edu.tr

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- This study focuses on the estimation of sparse partially linear models in the presence of right-censored data. A synthetic data transformation is employed to address censoring, while regression coefficients are partitioned to achieve sparsity. The proposed methodology introduces modified shrinkage and pretest estimators, achieved by integrating synthetic data and partitioning techniques. The estimation of the nonparametric component is conducted using smoothing splines. These estimators synthesize synthetic data transformation, smoothing splines, and shrinkage strategies. Comprehensive theoretical explanations and the asymptotic properties of the estimators are presented. Simulation studies, alongside an analysis of a hepatocellular carcinoma dataset, are utilized to demonstrate the approach's efficacy. Additionally, performance is assessed under high-dimensional data settings ($p > n$) through simulation and evaluation on the Norway/Stanford Breast Cancer dataset. The findings indicate that the proposed shrinkage and pretest estimators surpass submodel estimation for both low and high-dimensional data under conditions of right-censoring.

Keywords:

- *right-censored data; sparse partially linear model; shrinkage and pretest strategies; smoothing splines; high-dimensional data.*

AMS Subject Classification:

- 62J07, 62N01.

✉ Corresponding author

1. INTRODUCTION

Consider the standard partially linear model (PLM) expressed as follows:

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + f(t_i) + \varepsilon_i, \tag{1.1}$$

where y_i are the fully observed response variables, $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{ip})$ refers to the ith row consisting of p parametric covariates with $p \leq n$, and $(.)^\top$ represents the transpose operation on a vector or matrix. The vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is of dimension $(p \times 1)$ and contains the regression coefficients. The values of the nonparametric covariate are given by t_i, with the unknown smooth function $f(.)$ depicting its association with y_i. The error terms ε_i are random and are assumed to follow a normal distribution, given by $\varepsilon_i \sim N(0, \sigma^2)$. The model from Equation (1.1) can also be expressed in vector and matrix notation as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f} + \boldsymbol{\varepsilon}. \tag{1.2}$$

Let $\mathbf{y} = (y_1, \ldots, y_n)^\top$ be an $(n \times 1)$ vector representing the response variable, and $\mathbf{X}$ an $(n \times p)$ design matrix. The vector $\mathbf{f} = \{f(t_1), \ldots, f(t_n)\}^\top$ represents an $(n \times 1)$ collection of values of an unknown smooth function, while $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ denotes an $(n \times 1)$ vector of random errors. For ease of interpretation, the nonparametric covariate $t_i \in [0, 1]$ is used, and model (1.2) is specified without an intercept, assuming it is integrated within the nonparametric part. Researchers have concentrated on estimating both the parametric and nonparametric components in model 1.2 with uncensored datasets, notably by Ruppert et al. (2003) and Ahmed (2014). Shrinkage estimation methods for model (1.2) leveraging kernel smoothing have been examined by Ahmed et al. (2007) and Hossain et al. (2009), while Raheem et al. (2012) analyzed shrinkage estimation using a B-spline method. Additional significant studies have been conducted by Phukongtong et al. (2022) and Ahmed et al. (2023).

In various real-world applications, particularly in fields such as medical research, reliability engineering, and actuarial science, the dependent variable is often subject to right-censoring. This occurs when the exact timing of an event (such as death, equipment failure, or policy expiration) cannot be observed for some participants within the study period. For instance, in a clinical trial, some patients may remain alive at the end of the study, leaving their exact survival times unknown. This paper examines right-censored values y_i due to a censoring variable c_i, preventing full data observation. We note $\{\mathbf{x}_i, t_i, z_i, \delta_i\}_{i=1}^n$, involving $z_i = min(y_i, c_i)$ and $\delta_i = I(y_i < c_i)$. The incomplete response variable is z_i, and δ_i shows if the i^{th} observation is censored. If $\delta_i = 0$, then z_i is censored; otherwise, it is not. In model (1.2), the right-censored response vector $n \times 1$ substitutes the fully observed responses $\mathbf{y} = (y_1, \ldots, y_n)$. We assume y_i, c_i, and z_i are i.i.d., essential for further analysis. Ignoring censoring can result in biased and inefficient estimators with larger variances, a loss of statistical power due to reduced effective sample size, and potentially invalid hypothesis tests due to violated assumptions, ultimately rendering the results unreliable. Additionally, censoring complicates model selection and makes the asymptotic analysis of estimators more challenging. Therefore, employing methods that appropriately account for censoring is essential for accurate inference, reliable prediction, valid hypothesis testing, and effective model selection. For censorship given in model (1.2), we apply synthetic data transformation as per Aydin

and Yilmaz (2018) and Yenilmez et al. (2022). Studies by Aydın et al. (2019) and Yenilmez et al. (2022) have also addressed this for model (1.2).

Over the last ten years, there has been noteworthy interest in right-censored data, particularly in selecting key variables and estimating sparse partially linear models. This prevalence underscores the importance of developing statistical methods that can effectively handle the complexities introduced by censoring, ensuring the reliability and validity of research findings in these fields. When considering variable selection, incorporating all covariates into the model (a full model) can lead to complex and hard-to-interpret estimations. Conversely, a model with only a few key covariates (a submodel) might miss out on valuable information from other variables, resulting in biased estimates. Therefore, developing parsimonious models and efficient estimators is critically important for effective variable selection, accurate estimation, and reliable prediction.

This study explores a sparse partially linear model in the context of right-censored data. The principal objective is to employ a shrinkage strategy that integrates a full model (FM) with a submodel (SM) to derive more efficient shrinkage estimators, taking into account the right-censored nature of the data. Altered estimators are proposed based on shrinkage and pretest strategies, utilizing the transformation of synthetic data. These estimators address the complexities associated with the substantial number of predictors in the parametric component of the model (1.2) and the right-censored dependent variable. The nonparametric component is represented using smoothing splines and the penalized least squares method (PLS). It should be emphasized that this paper concentrates on scenarios where the predictor count is low ($p < n$) for the derivation of the estimators and the execution of asymptotic inferences. Nevertheless, simulations and data examples are also presented to illustrate the scenario involving high-dimensional data ($p > n$).

The paper is structured as follows. Section 2 introduces synthetic data transformation as a method to address the censoring issue. Section 3 discusses the comprehensive and submodel estimators, along with shrinkage and pretest estimation methods, to complete the estimation process. In Section 4, theoretical inferences and asymptotic properties of the proposed estimators are discussed. Section 5 provides a detailed explanation of the selection of shrinkage and smoothing parameters utilized in the estimation process. Sections 6 and 7 offer simulation and real data analyses for scenarios where $p < n$. Section 8 presents inferences for the high-dimensional PLM in the context of right-censored data. Lastly, Section 9 delivers the conclusions.

2. SOLUTION OF RIGHT-CENSORED DATA

Consider the probability distribution functions of survival (y_i) and censoring times (c_i), denoted as F and G. For each data point s, the unknown distribution function of y_i is $F(s) = P(y_i \leq s)$, and that of c_i is $G(s) = P(c_i \leq s)$. The model's significance relies on specific assumptions about the response, censoring, and explanatory variables as defined by Stute (1999).

A1: y_i and c_i are identically and independently distributed conditional on the covariates $(\mathbf{x}_i, t_i)$

A2: $P(y_i \leq c_i|y_i, \mathbf{x}_i, t_i) = P(y_i \leq c_i|y_i)$.

Common assumptions in survival analysis include A1 and A2. A1 ensures model accuracy with censored data, and violating it necessitates extra dataset information. A2 allows a connection between $(\mathbf{x}_i, t_i)$ and c_i, presuming covariates offer no extra insight on censoring after the death time is known. For more details, see Stute (1999). Incomplete response variable observations require data transformation since standard methods are inadequate. In the presence of censorship, use observation pairs $\{(z_i, \delta_i),\ i = 1, \ldots, n\}$ instead of just y_i. If G is continuous and known, z_i can be adjusted for an unbiased outcome.

$$y_{iG} = \frac{\delta_i z_i}{1 - G\left(z_i\right)}, i = 1, 2, \ldots, n, \tag{2.1}$$

where y_{iG} shares the mean with y_i. The above assumptions also support $E\left[y_{iG} \mid \mathbf{x}_i, t_i\right] = E\left[y_i \mid \mathbf{x}_i, t_i\right] = \mathbf{x}_i\beta + f\left(t_i\right)$. Note that $\mathbf{y}_G = \left\{y_{iG} = \left(y_{1G} \ldots, y_{nG}\right)^\top\right\}$ represents the vector of transformed response variables. Often, the distribution (G) of the censoring variable is unknown, as described in (2.1), necessitating the use of the Kaplan-Meier estimator as a substitute for G.

$$1 - \hat{G}(s) = \prod_{i=1}^{n} \left(\frac{n-i}{n-i+1}\right)^{\mathrm{I}\left[z_{(i)} \leq s, \delta_{(i)} = 0\right]}, s \geq 0, \tag{2.2}$$

where $z_{(1)} \leq \cdots \leq z_{(n)}$ are ordered values of z_i, and $\delta_{(i)}$ are the ordered values linked with $z_(i)$. When the distribution G is unknown, we apply the following synthetic data transformation:

$$y_{i\hat{G}} = \frac{\delta_i z_i}{1 - \hat{G}\left(z_i\right)}, i = 1, 2, \ldots, n. \tag{2.3}$$

3. SEMIPARAMETRIC ESTIMATORS

When the response variable is censored by a random variable c_i, it's replaced in Model (1.1) by the synthetic response variable $\boldsymbol{y}_{i\hat{G}}$ from (2.3). This section outlines estimating the non-parametric component using the smoothing spline approach for partial residuals and penalized least squares (PLS) based on synthetic responses. Introduce the smoothing matrix S_λ that depends on the smoothing parameter $\lambda > 0$. Let $v_1 < v_2 < \cdots < v_q$ be the unique ordered knot values of the nonparametric component $t_1, t_2, \ldots t_n$, where $q < n$. An $n \times q$ dimensional incidence matrix N expresses the relationship between t_i and v_i. Elements of N are computed as $N_{ij} = 1$ if $t_i = v_j$, and $N_{ij} = 0$ otherwise. The matrix and vector form of PLS for the model (1.1) follows:

$$L\left(\boldsymbol{\beta}, \mathbf{f}\right) = \left\|\mathbf{y}_{\hat{G}} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}\mathbf{f}\right\|_2^2 + n\lambda \int_0^1 \left\{f''(t)\right\}^2 dt, \tag{3.1}$$

where $\mathbf{y}_{\hat{G}} = \left(y_{i\hat{G}}, \ldots, y_{n\hat{G}}\right)$ is the $n \times 1$-dimensional vector of synthetic responses from (2.3). The smoothing parameter $\lambda > 0$ controls the curve's smoothness by regularizing the penalty term $\int_0^1 \left\{f''(t)\right\}^2$.

Simplifying the minimization criterion (3.1), the penalty term $\int_0^1 \left\{f''(t)\right\}^2$ provides a quadratic form $\mathbf{f}^\top\mathbf{K}\mathbf{f}$ where $\mathbf{K} = \mathbf{Q}^\top\mathbf{R}^{-1}\mathbf{Q}$. Note that $\mathbf{Q}$ is $((q-2) \times q)$ and $\mathbf{R}$ is

$((q-2)\times(q-2))$ are tri-diagonal matrices (see Aydın et al., 2019). Based on $\mathbf{K}$, the smoothing matrix is computed as $\mathbf{S}_\lambda = (\mathbf{I}_n - n\lambda\mathbf{K})^{-1}$ where $\mathbf{I}_n$ is an $(n\times n)$-dimensional identity matrix.

From the given information, the estimation of the nonparametric component $\mathbf{f} = \{f(t_i)\}_{i=1}^n$ is realized by:

$$\hat{\mathbf{f}}(\boldsymbol{\beta}) = \left(\mathbf{N}^\top\mathbf{N} + n\lambda\mathbf{K}\right)^{-1}\mathbf{N}^\top\left(\mathbf{y}_{\hat{G}} - \mathbf{X}\boldsymbol{\beta}\right) = \mathbf{S}_\lambda\left(\mathbf{y}_{\hat{G}} - \mathbf{X}\boldsymbol{\beta}\right), \tag{3.2}$$

$\mathbf{S}_\lambda$ is a positive-definite matrix determined by the smoothing parameter λ. Equation (3.2) is inactive as $\boldsymbol{\beta}$ is unknown. Post $\boldsymbol{\beta}$ estimates in Section 3.2, they substitute $\boldsymbol{\beta}$ in (3.2). When t_i values are distinct and ordered, $\mathbf{I}_n$ and $\mathbf{S}_\lambda$ simplify to the smoothing matrix $\mathbf{S}_\lambda = (\mathbf{I}_n + n\lambda\mathbf{K})^{-1}$. Thus, with $\mathbf{S}_\lambda$, partial residuals are derived, allowing the minimization criterion (3.1) to be expressed in matrix and vector form,

$$L(\boldsymbol{\beta}) = \left\|\widetilde{\mathbf{y}}_{\hat{G}} - \widetilde{\mathbf{X}}\boldsymbol{\beta}\right\|_2^2 = (\widetilde{\mathbf{y}}_{\hat{G}} - \widetilde{\mathbf{X}}\boldsymbol{\beta})^\top(\widetilde{\mathbf{y}}_{\hat{G}} - \widetilde{\mathbf{X}}\boldsymbol{\beta}), \tag{3.3}$$

where $\tilde{\mathbf{X}} = (\mathbf{I} - \mathbf{S}_\lambda)\,\mathbf{X}$ and $\tilde{\mathbf{y}}_{\hat{G}} = (\mathbf{I} - \mathbf{S}_\lambda)\,\mathbf{y}_{\hat{G}}$ are partial residuals of both parametric covariate and the response variable respectively. Hence, β without sparsity assumption on design matrix $\mathbf{X}$ can be estimated by:

$$\hat{\boldsymbol{\beta}} = \left(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{y}}_{\hat{G}}. \tag{3.4}$$

Outlined below are the requisite conditions and assumptions necessary for deriving asymptotic inferences using the smoothing splines method:

i Smoothness of the Nonparametric Function:

- The unknown smooth function $f(\cdot) \in W^\tau[a,b]$, meaning that $f(\cdot)$ is $(\tau-1)$ times continuously differentiable on $[a,b]$, and its τ-th derivative is square-integrable:

$$\int_a^b \left[f^{(\tau)}(t)\right]^2 dt < \infty.$$

ii Smoothing Splines and Knot Selection:

- Based on (ii), we use natural spline functions of degree $2\tau-1$ with knots at points v_j to estimate $f(t_i)$.
- Note that the derivatives exist only at the knot points v_j.

iii Convergence Rates of the Nonparametric Estimator:

- Considering the degree $2\tau-1=\psi$, the expected squared difference between $f(t_i)$ and its estimator $\hat{f}(t_i)$ satisfies:

$$E\left[\left(f(t_i) - \hat{f}(t_i)\right)^2\right] = O\left(\lambda^2\right) + O\left(\frac{\sigma^2}{n\lambda^{(2\tau-1)/(2\tau)}}\right),$$

  where $\lambda > 0$ is the smoothing parameter, and σ^2 is the variance of the error term.
- The optimal convergence rate is achieved when $\lambda \asymp n^{-1/(2\tau+1)}$, which balances the bias and variance terms.

Given the assumptions stated above, the subsequent asymptotic expression for $\hat{\boldsymbol{\beta}}$ can be formulated as follows:

Theorem 3.1. *Following Gao (1995) Gao's Theorem under the given assumptions, the full model estimator $\hat{\boldsymbol{\beta}}^{\mathrm{FM}}$ in the right-censored partially linear regression with smoothing splines shares this asymptotic distribution:*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{\mathrm{FM}}-\boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma_{\mathit{eff}}^{2}\mathbf{B}^{-1}\right),$$

where $\sigma_{\mathit{eff}}^{2}$ captures error variance σ^2 and variability from right-censoring and smoothing estimation $f(\cdot)$, and $\mathbf{B}$ is the limit of $n^{-1}\mathbf{X}^{\top}\mathbf{X}$.

Proof of Theorem 3.1 is given in appendix.

3.1. Full model and submodel estimation

In this paper's $p \leq n$ settings, model (1.2) is considered sparse. The design matrix is split into $\mathbf{X}=(\mathbf{X_1},\mathbf{X_2})$ where $\mathbf{X}_1$, an $n\times p_1$ dimensional matrix, contains key variables, and $\mathbf{X}_2$, an $n\times p_2$ dimensional matrix, either includes inactive covariates or is sparse when $\boldsymbol{\beta}_2$ holds. The regression coefficient vector is partitioned into $\boldsymbol{\beta}=(\boldsymbol{\beta}_1^{\top},\boldsymbol{\beta}_2^{\top})^{\top}$, with $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ indicating strong (nonzero) and sparse signals with $p_1+p_2=p$. The goal is to estimate pairs $(\boldsymbol{\beta}_1,\mathbf{f})$ when sparsity is met, i.e., $\boldsymbol{\beta}_2=\mathbf{0}$. If not, shrinkage and pretest estimation borrow from the full model, examined in simulations for both low and high dimensions. Based on partitioned matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ and coefficient vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, the objective function (3.3) is adjusted for full and submodel estimates

$$L\left(\boldsymbol{\beta}\right)=\left\|\widetilde{\mathbf{y}}_{\hat{G}}-\widetilde{\mathbf{X}}_1\boldsymbol{\beta}_1-\widetilde{\mathbf{X}}_2\boldsymbol{\beta}_2\right\|_2^2=(\widetilde{\mathbf{y}}_{\hat{G}}-\widetilde{\mathbf{X}}_1\boldsymbol{\beta}_1-\widetilde{\mathbf{X}}_2\boldsymbol{\beta}_2)^{\top}(\widetilde{\mathbf{y}}_{\hat{G}}-\widetilde{\mathbf{X}}_1\boldsymbol{\beta}_1-\widetilde{\mathbf{X}}_2\boldsymbol{\beta}_2). \tag{3.5}$$

Consider the full model (FM) and submodel (SM) estimates. The FM estimator of $\boldsymbol{\beta}_1$, denoted as $\hat{\boldsymbol{\beta}}_1^{FM}$, is obtained by solving (3.5).

$$\begin{aligned}
L\left(\boldsymbol{\beta}\right)&=(\widetilde{\mathbf{y}}_{\hat{G}}-\widetilde{\mathbf{X}}_1\boldsymbol{\beta}_1-\widetilde{\mathbf{X}}_2\boldsymbol{\beta}_2)^{\top}(\widetilde{\mathbf{y}}_{\hat{G}}-\widetilde{\mathbf{X}}_1\boldsymbol{\beta}_1-\widetilde{\mathbf{X}}_2\boldsymbol{\beta}_2)\\
&=\widetilde{\mathbf{y}}_{\hat{G}}^{\top}\widetilde{\mathbf{y}}_{\hat{G}}-2\widetilde{\mathbf{y}}_{\hat{G}}^{\top}\widetilde{\mathbf{X}}_1^{\top}\boldsymbol{\beta}_1-2\widetilde{\mathbf{y}}_{\hat{G}}^{\top}\widetilde{\mathbf{X}}_2\boldsymbol{\beta}_2+2\boldsymbol{\beta}_1^{\top}\widetilde{\mathbf{X}}_1^{\top}\widetilde{\mathbf{X}}_2\boldsymbol{\beta}_2+\boldsymbol{\beta}_1^{\top}\widetilde{\mathbf{X}}_1^{\top}\boldsymbol{\beta}_1\widetilde{\mathbf{X}}_1+\boldsymbol{\beta}_2^{\top}\widetilde{\mathbf{X}}_2^{\top}\boldsymbol{\beta}_2\widetilde{\mathbf{X}}_2\\
\frac{L\left(\boldsymbol{\beta}\right)}{\partial\beta_1}&=-2\widetilde{\mathbf{y}}_{\hat{G}}^{\top}\widetilde{\mathbf{X}}_1+2\widetilde{\mathbf{X}}_1^{\top}\widetilde{\mathbf{X}}_2\boldsymbol{\beta}_2+2\widetilde{\mathbf{X}}_1^{\top}\widetilde{\mathbf{X}}_1\boldsymbol{\beta}_1\\
\frac{L\left(\boldsymbol{\beta}\right)}{\partial\beta_2}&=-2\widetilde{\mathbf{y}}_{\hat{G}}^{\top}\widetilde{\mathbf{X}}_2+2\widetilde{\mathbf{X}}_1^{\top}\widetilde{\mathbf{X}}_2\boldsymbol{\beta}_1+2\widetilde{\mathbf{X}}_2^{\top}\widetilde{\mathbf{X}}_2\boldsymbol{\beta}_2
\end{aligned}$$

after some algebraic operations based on $\frac{L(\boldsymbol{\beta})}{\partial\beta_1}=0$, and $\frac{L(\boldsymbol{\beta})}{\partial\beta_2}=0$ then $\hat{\boldsymbol{\beta}}_1^{FM}$ can be given by:

$$\hat{\boldsymbol{\beta}}_1^{FM}=\left(\widetilde{\mathbf{X}}_1^{\top}\mathbf{M}_1\widetilde{\mathbf{X}}_1\right)^{-1}\widetilde{\mathbf{X}}_1^{\top}\mathbf{M}_1\widetilde{\mathbf{y}}_{\hat{G}} \tag{3.6}$$

where

$$\mathbf{M}_1=\mathbf{I}-\widetilde{\mathbf{X}}_2\left(\widetilde{\mathbf{X}}_2^{\top}\widetilde{\mathbf{X}}_2\right)^{-1}\widetilde{\mathbf{X}}_2^{\top}.$$

After $\boldsymbol{\beta}_1^{FM}$ is obtained, let us assume the sparsity assumption $\boldsymbol{\beta}_2 = 0$ is ensured. Then, the submatrix of sparse predictors $\widetilde{\mathbf{X}}_2$ is removed from the model and submodel estimator $\hat{\boldsymbol{\beta}}_1^{SM}$ is obtained as follows:

$$\hat{\boldsymbol{\beta}}_1^{SM} = \left(\widetilde{\mathbf{X}}_1^{\top}\widetilde{\mathbf{X}}_1\right)^{-1}\widetilde{\mathbf{X}}_1^{\top}\widetilde{\mathbf{y}}_{\hat{G}}. \tag{3.7}$$

The modified semiparametric least squares estimator for the submodel $\hat{\boldsymbol{\beta}}_1^{SM}$ considers $\hat{\boldsymbol{\beta}}_1^{SM}$ superior to $\hat{\boldsymbol{\beta}}_1^{FM}$ if the model is sparse, meaning parameter vector $\boldsymbol{\beta}_2$ is near 0, expressed as $\|\boldsymbol{\beta}_1\|_0 \ll p,$, with $\|\boldsymbol{\beta}_1\|_0$ as nonzero elements in $\boldsymbol{\beta}_1$ and p as total parameters. Significant deviation of sub-vector $\boldsymbol{\beta}_2$ from 0 affects estimation and prediction of β_1. The submodel estimator $\hat{\boldsymbol{\beta}}_1^{SM}$ may be biased, inefficient, and inconsistent, while estimate $\hat{\boldsymbol{\beta}}_1^{FM}$ stays consistent for deviations of $\boldsymbol{\beta}_2$ from 0.

In the partially linear regression model, two main strategies, $\hat{\boldsymbol{\beta}}_1^{FM}$ and $\hat{\boldsymbol{\beta}}_1^{SM}$, exist. A reasonable compromise uses shrinkage and pretest methods as outlined by Ahmed et al. (2007). This paper develops a robust estimator of $\boldsymbol{\beta}_1$ for censored responses, using these methods to focus on key regression parameters while minimizing others.

3.2. Pretest and Shrinkage methods

This section discusses shrinkage and pretest estimation procedures for right-censored partially linear models. These strategies, detailed by Ahmed (2014), Yüzbaşı et al. (2020), and Ahmed et al. (2023), have been adapted with a synthetic data process as the censorship solution is finalized here.

From the information given above, the shrinkage estimator of $\boldsymbol{\beta}_1$ by combining $\hat{\boldsymbol{\beta}}_1^{FM}$ and $\hat{\boldsymbol{\beta}}_1^{SM}$ that obtained by solving minimization criterion given in (3.5), as follows:

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}} = \widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}} + \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)\left(1 - (p_2 - 2)\,\mathcal{T}_n^{-1}\right),\ p_2 \geq 3, \tag{3.8}$$

and positive-part of the given shrinkage estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}$ can be shown by:

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}} = \widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}} + \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)\left(1 - (p_2 - 2)\,\mathcal{T}_n^{-1}\right)^{+}, \tag{3.9}$$

where $\mathcal{T}^{-1}$ is the inverse of the determined distance measure and $\alpha^{+} = max(0, \alpha)$. Here, $\mathcal{T}$ can be expressed based on $\widetilde{\mathbf{U}}_1 = \mathbf{I}_n - \tilde{\mathbf{X}}_1\left(\tilde{\mathbf{X}}_1^{\top}\tilde{\mathbf{X}}_1\right)^{-1}\tilde{\mathbf{X}}_1^{\top}$ and the following estimate of $\boldsymbol{\beta}_2$. Note that $\boldsymbol{\beta}_2$ is estimated using sparse set of partial residuals $\tilde{\mathbf{X}}_2$ with ordinary least squares process

$$\widehat{\boldsymbol{\beta}}_2 = \left(\widetilde{\mathbf{X}}_2^{\top}\widetilde{\mathbf{U}}_1\widetilde{\boldsymbol{X}}_2\right)^{-1}\widetilde{\boldsymbol{X}}_2^{\top}\widetilde{\mathbf{U}}_1\tilde{\boldsymbol{y}}_{\hat{G}}. \tag{3.10}$$

Then, with (3.10), $\mathcal{T}$ is given by:

$$\mathcal{T} = \frac{1}{\widehat{\sigma}_1^2}\left(\widehat{\boldsymbol{\beta}}_2\right)^{\top}\left(\tilde{\mathbf{X}}_2^{\top}\tilde{\mathbf{U}}_1\tilde{\mathbf{X}}_2\right)\widehat{\boldsymbol{\beta}}_2, \tag{3.11}$$

where $\widehat{\sigma}_1^2$ is the variance of the right-censored submodel, calculated as

$$\widehat{\sigma}_1^2 = \frac{1}{n - p_1}\left(\tilde{\mathbf{y}}_{\hat{G}} - \tilde{\mathbf{X}}_1\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)^{\top}\left(\tilde{\mathbf{y}}_{\hat{G}} - \tilde{\mathbf{X}}_1\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right). \tag{3.12}$$

After shrinkage estimation, a Pretest estimate can be performed by testing hypothesis $\mathrm{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ for the model's sparse coefficients. Consequently, the semiparametric pretest estimator for the right-censored model is given by

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}} = \widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}} - \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right) \mathrm{I}\left(\mathcal{T} \leq \nu_{n,\alpha}\right). \tag{3.13}$$

$\nu_{n,\alpha}$ is the critical value based on the distribution of $\mathcal{T}$ at significance level α. As Ahmed et al. (2007) noted, $\mathcal{T}$ follows the $\chi^2_{p_2,\alpha}$ distribution as the sample size n approaches infinity. (3.8)-(3.13), The paper finalizes its estimators through synthetic data transformation, underscoring its contribution. The next section provides statistical inferences and asymptotic properties for a clearer grasp of the estimators: $\hat{\boldsymbol{\beta}}_1^{FM}$, $\hat{\boldsymbol{\beta}}_1^{SM}$, $\hat{\boldsymbol{\beta}}_1^{S}$, $\hat{\boldsymbol{\beta}}_1^{PS}$, and $\hat{\boldsymbol{\beta}}_1^{PT}$, considering the impact of synthetic data.

4. STATISTICAL PROPERTIES

This section addresses the bias and variance of semi-parametric estimators for the low-dimensional case $p \leq n$, along with their asymptotic quadratic biases and variances. We also consider censoring's impact on asymptotic inferences.

Remark 4.1. In synthetic data transformation, suppose that the censorship assumptions (A1-A2) are ensured. Then, as $n \longrightarrow \infty$, $E\left[y_{i\hat{G}} \mid \mathbf{x}_i, t_i\right] = E\left[y_{iG} \mid \mathbf{x}_i, t_i\right] = E\left[y_i \mid \mathbf{x}_i, t_i\right] = \mathbf{x}_i\boldsymbol{\beta} + f\left(t_i\right)$ (see Aydin and Yilmaz, 2018 for the proof).

Under regular conditions and Remark 4.1, this subsection offers key insights into the bias, variance, and asymptotic properties of the semiparametric estimators.

4.1. Bias

To simplify notation, let $\widehat{\boldsymbol{\beta}}_1^*$ be the estimator of $\boldsymbol{\beta}_1$ derived from pretest and shrinkage methods. Since these procedures are interconvertible (see Ahmed et al., 2007), inferences are focused on $\widehat{\boldsymbol{\beta}}^*$, which also apply to variance and asymptotic analyses. Thus, the estimator's bias can be determined as follows:

$$\begin{aligned}
\mathrm{Bias}\left(\hat{\beta}_1^*\right) &= E\left[\hat{\beta}_1^*\right] - \beta_1 \\
&= E\left[\hat{\boldsymbol{\beta}}_1^{SM} + (\hat{\boldsymbol{\beta}}_1^{FM} - \hat{\boldsymbol{\beta}}_1^{SM})(1 - (p_2 - 2)\mathcal{T}^{-1})\right] - \boldsymbol{\beta}_1 \\
\text{if } (1 - (p_2 - 2)\mathcal{T}^{-1}) &\text{ notated as } \pi \text{ then,} \\
&= E\left[\hat{\boldsymbol{\beta}}_1^{SM}\right] + \pi E\left[\hat{\boldsymbol{\beta}}_1^{FM}\right] - \pi E\left[\hat{\boldsymbol{\beta}}_1^{SM}\right] - \boldsymbol{\beta}_1 \\
&= \pi E\left[\hat{\boldsymbol{\beta}}_1^{FM}\right] + (1 - \pi)\, E\left[\hat{\boldsymbol{\beta}}_1^{SM}\right] - \boldsymbol{\beta}_1 \\
&= \pi Bias\left[\hat{\boldsymbol{\beta}}_1^{FM}\right] + (1 - \pi)\, Bias\left[\hat{\boldsymbol{\beta}}_1^{SM}\right].
\end{aligned}$$

Thus, $Bias\left(\hat{\boldsymbol{\beta}}_1^*\right)$ depends on individual biases and sparse coefficients. We examine the asymptotic quadratic distributional bias with local alternatives, detailed in Section 4.3.

4.2. Variance

The variance of shrinkage and pretest estimators is mainly affected by the SM and FM estimators and their covariance. Hence, the general formula for $Var(\hat{\boldsymbol{\beta}}_1^*)$ assumes $\pi = (1-(p_2-2)\mathcal{T}^{-1})$ as stated above:

$$Var(\hat{\boldsymbol{\beta}}_1^*) = \pi^2 Var(\hat{\boldsymbol{\beta}}_1^{FM}) + (1-\pi)^2 Var(\hat{\boldsymbol{\beta}}_1^{SM}) + 2\pi(1-\pi)Cov\left(\hat{\boldsymbol{\beta}}_1^{FM}, \hat{\boldsymbol{\beta}}_1^{SM}\right).$$

Further details are provided in the next section for the asymptotic variance of the introduced estimators.

4.3. Asymptotic behaviors

This section presents the asymptotic distributional risk (ADR) for shrinkage and pretest estimators, including the asymptotic distributional bias (ADB) and quadratic distributional bias (AQDB). Inferences about covariance matrices are based on ADR. The estimators emphasize consistency, asymptotic normality, and oracle properties, considering the design matrix $\mathbf{X}$, partial residuals $\tilde{\mathbf{X}}$, and sparsity of true model coefficients $\boldsymbol{\beta}$.

Consider local alternatives as parameter sequences that converge to the null hypothesis at a rate based on sample size n. Essentially, these alternatives approach the null hypothesis $H_0 : \hat{\boldsymbol{\beta}}^* = \boldsymbol{\beta}$ as $n \longrightarrow \infty$. The sequence of local alternatives $\mathcal{M}_n$ is as follows:

$$\mathcal{M}_n : \beta_2 = \frac{\omega}{\sqrt{n}},\ \omega = (\omega_1, \ldots, \omega_{p_2})^\top \in \mathbb{R}^{p_2}. \tag{4.1}$$

Then, the quadratic loss function can be given by:

$$\mathcal{L}\left(\hat{\boldsymbol{\beta}}_1^*\right) = n\left(\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{\beta}_1\right)^\top \mathbf{V}\left(\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{\beta}_1\right), \tag{4.2}$$

where $\mathbf{V}$ is the positive-definite matrix. Accordingly, under local alternatives (4.1), non degenerate asymptotic distribution function for the $\hat{\boldsymbol{\beta}}_1^*$ is shown as:

$$F(\nabla) = \lim_{n\to\infty} P\left(\sqrt{n}\left(\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{\beta}_1\right) \leq \nabla \mid \mathcal{M}_n\right), \tag{4.3}$$

and from that ADR of $\hat{\boldsymbol{\beta}}_1^*$ is defined as follows:

$$\text{ADR}\left(\hat{\boldsymbol{\beta}}_1^*\right) = \text{tr}\left(\mathbf{V}\int_{\mathbb{R}^{p_1}}\int \nabla\nabla^\top dF(\nabla)\right) = \text{tr}(\mathbf{V}\mathcal{R}), \tag{4.4}$$

where $\mathcal{R}$ is the dispersion matrix of the asymptotic distribution function $F(\nabla)$. Hence, ADB of $\hat{\boldsymbol{\beta}}_1^*$ can be derived as:

$$\text{ADB}\left(\boldsymbol{\beta}_1^*\right) = \text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1\right)\right\}. \tag{4.5}$$

Given extra conditions for design matrix's partial residuals,
$\tilde{\mathbf{X}}$: *(i)* $\frac{1}{n}\max_{1\leq j\leq n}\tilde{\mathbf{x}}_j^\top\left(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{x}}_j \to 0$ when $n\to\infty$, *(ii)* For positive definite matrix $\tilde{\mathbf{P}}$,

$\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{X}}^{\top}\tilde{\mathbf{X}} \to \tilde{\boldsymbol{P}}$. Also, some notations are needed to show the $ADB(\hat{\boldsymbol{\beta}}_1^*)$ that are given below:

$$\tilde{\mathbf{P}} = \begin{pmatrix} \tilde{\mathbf{P}}_{11} & \tilde{\mathbf{P}}_{12} \\ \tilde{\mathbf{P}}_{21} & \tilde{\mathbf{P}}_{22} \end{pmatrix}, \Delta = \left(\boldsymbol{\omega}^{\top}\tilde{\mathbf{P}}_{22.1}^{-1}\boldsymbol{\omega}\right)\sigma^{-2}, \tilde{\mathbf{P}}_{22.1} = \tilde{\mathbf{P}}_{22} - \tilde{\mathbf{P}}_{21}\tilde{\mathbf{P}}_{11}^{-1}\tilde{\mathbf{P}}_{12}$$

$$\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\varphi}_1 \\ \boldsymbol{\varphi}_2 \end{pmatrix} = -\lambda_{2n}\tilde{\mathbf{P}}^{-1}\boldsymbol{\beta}, \boldsymbol{\varphi}_{11.2} = \boldsymbol{\varphi}_1 - \tilde{\mathbf{P}}_{12}\tilde{\mathbf{P}}_{22}^{-1}\left((\boldsymbol{\beta}_2 - \boldsymbol{\omega}) - \boldsymbol{\varphi}_2\right),$$

$$\boldsymbol{\theta} = \boldsymbol{\varphi}_{11.2} - \boldsymbol{\delta}, \boldsymbol{\delta} = \tilde{\mathbf{P}}_{11}^{-1}\tilde{\mathbf{P}}_{12}\boldsymbol{\omega}.$$

$\mathcal{D}_v(\kappa, \Delta)$ is the cumulative distribution function of a non-central χ^2 distribution with v degrees of freedom, defined as:

$$\mathrm{E}\left(\chi_v^{-2r}(\Delta)\right) = \int_0^{\infty} \kappa^{-2r} d\mathcal{D}_v(\kappa, \Delta). \tag{4.6}$$

Given these conditions and ensuring Remark 4.1, the ADB of the estimators inspired by Theorem 1 in Yüzbaşı et al. (2020) are described as follows:

$$\begin{aligned}
\mathrm{ADB}\left(\hat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right) &= -\boldsymbol{\varphi}_{11.2}, \\
\mathrm{ADB}\left(\hat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right) &= -\boldsymbol{\theta}, \\
\mathrm{ADB}\left(\hat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right) &= -\boldsymbol{\varphi}_{11.2} - \boldsymbol{\delta}\mathcal{D}_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right), \\
\mathrm{ADB}\left(\hat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right) &= -\boldsymbol{\varphi}_{11.2} - (p_2 - 2)\,\boldsymbol{\delta}\left(\chi^{-2}_{p_2+2}(\Delta)\right), \\
\mathrm{ADB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right) &= -\boldsymbol{\varphi}_{11.2} - \boldsymbol{\delta}\mathcal{D}_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right) \\
&\quad - (p_2 - 2)\,\boldsymbol{\delta}\mathrm{E}\left\{\chi^{-2}_{p_2+2}(\Delta)\mathrm{I}\left(\chi^2_{p_2+2}(\Delta) > p_2 - 2\right)\right\},
\end{aligned} \tag{4.7}$$

For proof of (4.7), refer to Yüzbaşı et al. (2020) regarding the synthetic data transformation in Remark 4.1. The ADBs of the estimators in quadratic form, as mentioned earlier, are presented using matrix AQDB. $\tilde{\mathbf{P}}_{11.2} = \tilde{\mathbf{P}}_{11} - \tilde{\mathbf{P}}_{12}\tilde{\mathbf{P}}_{22}^{-1}\tilde{\mathbf{P}}_{21}$:

$$\mathrm{AQDB}\left(\hat{\boldsymbol{\beta}}_1^*\right) = \mathrm{ADB}\left(\hat{\boldsymbol{\beta}}_1^*\right)^{\top}\tilde{\mathbf{P}}_{11.2}\,\mathrm{ADB}\left(\hat{\boldsymbol{\beta}}_1^*\right), \tag{4.8}$$

and using the definition of AQDB in equation (4.8), the following AQDBs can be obtained:

$$
\begin{aligned}
&\mathrm{AQDB}\left(\hat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)=\boldsymbol{\varphi}_{11.2}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\varphi}_{11.2},\\
&\mathrm{AQDB}\left(\hat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)=\boldsymbol{\theta}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\theta},\\
&\mathrm{AQDB}\left(\hat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right)=\boldsymbol{\varphi}_{11.2}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\varphi}_{11.2}+\boldsymbol{\varphi}_{11.2}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\delta}\mathcal{D}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\qquad+\boldsymbol{\delta}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\varphi}_{11.2}\mathcal{D}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)+\boldsymbol{\delta}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\delta}\mathcal{D}^2_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right),\\
&\mathrm{AQDB}\left(\hat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)=\boldsymbol{\varphi}_{11.2}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\varphi}_{11.2}+(p_2-2)\,\boldsymbol{\varphi}_{11.2}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\delta}\mathbf{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\\
&\qquad+(p_2-2)\,\boldsymbol{\delta}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\varphi}_{11.2}\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\\
&\qquad+(p_2-2)^2\,\boldsymbol{\delta}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\delta}\left[\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\right]^2,\\
&\mathrm{AQDB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right)=\boldsymbol{\varphi}_{11.2}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\varphi}_{11.2}+\left(\boldsymbol{\delta}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\varphi}_{11.2}+\boldsymbol{\varphi}_{11.2}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\delta}\right)\\
&\qquad\times\left[\mathcal{D}_{p_2+2}\left((p_2-2);\Delta\right)\right.\\
&\qquad\left.+(p_2-2)\,\mathrm{E}\left\{\chi^{-2}_{p_2+2}(\Delta)\mathrm{I}\left(\chi^{-2}_{p_2+2}(\Delta)>p_2-2\right)\right\}\right]\\
&\qquad+\boldsymbol{\delta}^{\top}\tilde{\mathbf{P}}_{11.2}\boldsymbol{\delta}\left[\mathcal{D}_{p_2+2}\left((p_2-2);\Delta\right)\right.\\
&\qquad\left.+(p_2-2)\,\mathrm{E}\left\{\chi^{-2}_{p_2+2}(\Delta)\mathrm{I}\left(\chi^{-2}_{p_2+2}(\Delta)>p_2-2\right)\right\}\right]^2.
\end{aligned}
\tag{4.9}
$$

Based on the partitioned variance-covariance matrix $\tilde{\mathbf{P}}$ and its defined sub-matrices, the asymptotic covariance matrices are derived below:

$$
\begin{aligned}
&\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)=\sigma^2\tilde{\mathbf{P}}_{11.2}^{-1}+\boldsymbol{\varphi}_{11.2}\boldsymbol{\varphi}_{11.2}^{\top},\\
&\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)=\sigma^2\tilde{\mathbf{P}}_{11}^{-1}+\boldsymbol{\theta}\boldsymbol{\theta}^{\top},\\
&\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right)=\sigma^2\tilde{\mathbf{P}}_{11.2}^{-1}+\boldsymbol{\varphi}_{11.2}\boldsymbol{\varphi}_{11.2}^{\top}+2\boldsymbol{\varphi}_{11.2}^{\top}\boldsymbol{\delta}\mathcal{D}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\qquad+\sigma^2\left(\tilde{\mathbf{P}}_{11.2}^{-1}-\tilde{\mathbf{P}}_{11}^{-1}\right)\mathcal{D}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\qquad+\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\left[2\mathcal{D}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)-\mathcal{D}_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\right],\\
&\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)=\sigma^2\tilde{\mathbf{P}}_{11.2}^{-1}+\boldsymbol{\varphi}_{11.2}\boldsymbol{\varphi}_{11.2}^{\top}+2\,(p_2-2)\,\delta\boldsymbol{\varphi}_{11.2}^{\top}\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\\
&\qquad-(p_2-2)\,\sigma^2\tilde{\mathbf{P}}_{11}^{-1}\tilde{\mathbf{P}}_{12}\tilde{\mathbf{P}}_{22.1}^{-1}\tilde{\mathbf{P}}_{21}\tilde{\mathbf{P}}_{11}^{-1}\left\{2\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\right.\\
&\qquad\left.-(p_2-2)\,\mathrm{E}\left(\chi^{-4}_{p_2+2}(\Delta)\right)\right\}\\
&\qquad+(p_2-2)\,\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\left\{2\mathrm{E}\left(x^{-2}_{p_2+2}(\Delta)\right)\right.\\
&\qquad\left.-2\mathrm{E}\left(\chi^{-2}_{p_2+4}(\Delta)\right)-(p_2-2)\,\mathrm{E}\left(x^{-4}_{p_2+4}(\Delta)\right)\right\}
\end{aligned}
\tag{4.10}
$$

$$
\begin{aligned}
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right)=&\,\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)\\
&-2\boldsymbol{\delta}\boldsymbol{\varphi}_{11.2}^{\top}\mathrm{E}\left(\left\{1-(p_2-2)\,\chi^{-2}_{p_2+2}(\Delta)\right\}\mathrm{I}\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\\
&+(p_2-2)\,\sigma^2\tilde{\mathbf{P}}_{11}^{-1}\tilde{\mathbf{P}}_{12}\tilde{\mathbf{P}}_{22.1}^{-1}\tilde{\mathbf{P}}_{21}\tilde{\mathbf{P}}_{11}^{-1}\\
&\times\left[2\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\mathrm{I}\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\right.\\
&\left.-(p_2-2)\,\mathrm{E}\left(\chi^{-4}_{p_2+2}(\Delta)\mathrm{I}\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\right]\\
&-\sigma^2\tilde{\mathbf{P}}_{11}^{-1}\tilde{\mathbf{P}}_{12}\tilde{\mathbf{P}}_{22.1}^{-1}\tilde{\mathbf{P}}_{21}\tilde{\mathbf{P}}_{11}^{-1}\mathcal{D}_{p_2+2}\left((p_2-2);\Delta\right)
\end{aligned}
$$

Note that proofs of (4.10) and (4.11) can be found in Yüzbaşı et al. (2020).

Regarding the AQDBs given in (4.9) the following inferences can be made under the assumption $\tilde{\mathbf{P}}_{12} \neq \mathbf{0}$. (i) The $AQDB(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}})$ is a constant and (ii) $AQDB(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}})$ is an unbounded function based on $\boldsymbol{\theta}$ and $\tilde{\mathbf{P}}_{11.2}$. (iii) $ADQB(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}})$ shows increases to its maximum value and then decreases dependent on the increment of $\Delta > 0$. This inference is similar to the $ADQB(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}})$. (iv) Under different Δ values, although $ADQB(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}})$ and $ADQB(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}})$ have close performances, $\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}$ shows slightly better performance than $\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}$ in terms of quadratic bias.

Thus, asymptotic distributional risk (ADR) for the three semiparametric estimators is provided here:

$$
\begin{aligned}
&\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right) = \sigma^2 \operatorname{tr}\left(\mathbf{V}\tilde{\mathbf{P}}_{11.2}^{-1}\right) + \boldsymbol{\varphi}_{11.2}^{\top}\mathbf{V}\boldsymbol{\varphi}_{11.2},\\
&\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right) = \sigma^2 \operatorname{tr}\left(\mathbf{V}\tilde{\mathbf{P}}_{11}^{-1}\right) + \boldsymbol{\theta}^{\top}\mathbf{V}\boldsymbol{\theta},\\
&\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right) = \sigma^2 \operatorname{tr}\left(\mathbf{V}\tilde{\mathbf{P}}_{11.2}^{-1}\right) + \boldsymbol{\varphi}_{11.2}^{\top}\mathbf{V}\boldsymbol{\varphi}_{11.2} - 2\boldsymbol{\varphi}_{11.2}^{\top}\mathbf{V}\boldsymbol{\delta}\mathcal{D}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad - \sigma^2 \operatorname{tr}\left(\mathbf{V}\tilde{\mathbf{P}}_{11.2}^{-1} - \mathbf{V}\tilde{\mathbf{P}}_{11}^{-1}\right)\mathcal{D}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad + \boldsymbol{\delta}^{\top}\mathbf{V}\boldsymbol{\delta}\left\{2\mathcal{D}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) - \mathcal{D}_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\right\},\\
&\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right) = \sigma^2 \operatorname{tr}\left(\mathbf{V}\tilde{\mathbf{P}}_{11.2}^{-1}\right) + \boldsymbol{\varphi}_{11.2}^{\top}\mathbf{V}\boldsymbol{\varphi}_{11.2} + 2\left(p_2-2\right)\boldsymbol{\varphi}_{11.2}^{\top}\mathbf{V}\boldsymbol{\delta}\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\\
&\quad - \left(p_2-2\right)\sigma^2 \operatorname{tr}\left(\tilde{\mathbf{P}}_{21}\tilde{\mathbf{P}}_{11}^{-1}\mathbf{V}\tilde{\mathbf{P}}_{11}^{-1}\tilde{\mathbf{P}}_{12}\tilde{\mathbf{P}}_{22.1}^{-1}\right)\left\{2\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\right.\\
&\quad \left. - \left(p_2-2\right)\mathrm{E}\left(\chi^{-4}_{p_2+2}(\Delta)\right)\right\}\\
&\quad + \left(p_2-2\right)\boldsymbol{\delta}^{\top}\mathbf{V}\boldsymbol{\delta}\left\{2\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\right.\\
&\quad - 2\mathrm{E}\left(\chi^{-2}_{p_2+4}(\Delta)\right) - \left(p_2-2\right)\mathrm{E}\left(\chi^{-4}_{p_2+4}(\Delta)\right)\\
&\quad \left. + \left(p_2-2\right)\mathrm{E}\left(\chi^{-4}_{p_2+2}(\Delta)\mathrm{I}\left(\chi^2_{p_2+2}(\Delta) \leq p_2-2\right)\right)\right]
\end{aligned}
\tag{4.11}
$$

$$
\begin{aligned}
\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right) &= \mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)\\
&\quad - 2\boldsymbol{\phi}_{11.2}^{\top}\boldsymbol{V}\boldsymbol{\delta}\mathrm{E}\left(\left\{1-\left(p_2-2\right)\chi^{-2}_{p_2+2}(\Delta)\right\}\mathrm{I}\left(\chi^2_{p_2+2}(\Delta) \leq p_2-2\right)\right)\\
&\quad + \left(p_2-2\right)\sigma^2 \operatorname{tr}\left(\tilde{\mathbf{P}}_{21}\tilde{\mathbf{P}}_{11}^{-1}\boldsymbol{V}\tilde{\mathbf{P}}_{11}^{-1}\tilde{\mathbf{P}}_{12}\tilde{\mathbf{P}}_{22.1}^{-1}\right)\\
&\quad \times\left[2\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\mathrm{I}\left(\chi^2_{p_2+2}(\Delta) \leq p_2-2\right)\right)\right.\\
&\quad \left. - \left(p_2-2\right)\mathrm{E}\left(\chi^{-4}_{p_2+2}(\Delta)\mathrm{I}\left(\chi^2_{p_2+2}(\Delta) \leq p_2-2\right)\right)\right]\\
&\quad - \sigma^2 \operatorname{tr}\left(\tilde{\mathbf{P}}_{21}\tilde{\mathbf{P}}_{11}^{-1}\boldsymbol{V}\tilde{\mathbf{P}}_{11}^{-1}\tilde{\mathbf{P}}_{12}\tilde{\mathbf{P}}_{22.1}^{-1}\right)\mathcal{D}_{p_2+2}\left(\left(p_2-2\right);\Delta\right)\\
&\quad + \boldsymbol{\delta}^{\top}\boldsymbol{V}\left[2\mathcal{D}_{p_2+2}\left(\left(p_2-2\right);\Delta\right) - \mathcal{D}_{p_2+4}\left(\left(p_2-2\right);\Delta\right)\right]\\
&\quad - \left(p_2-2\right)\boldsymbol{\delta}^{\top}\boldsymbol{V}\boldsymbol{\delta}\left[2\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\mathrm{I}\left(\chi^2_{p_2+2}(\Delta) \leq p_2-2\right)\right)\right.\\
&\quad \left. - 2\mathrm{E}\left(\chi^{-2}_{p_2+4}(\Delta)\mathrm{I}\left(\chi^2_{p_2+4}(\Delta) \leq p_2-2\right)\right).\right.
\end{aligned}
$$

If $\tilde{\mathbf{P}}_{12} = \mathbf{0}$, then under the local alternatives $\mathcal{M}_n$ and for all $\boldsymbol{\omega}$ ADRs given in (4.11) turns into $\sigma^2 \operatorname{tr}\left(\mathbf{V}\tilde{\mathbf{P}}_{11}^{-1}\right) + \boldsymbol{\varphi}_{11.2}^{\top}\mathbf{V}\boldsymbol{\varphi}_{11.2}$. Also, if $\tilde{\mathbf{P}}_{12} \neq 0$, then the following asymptotic analysis can be made:

Asymptotic analysis for ADR: (i) When $\Delta > 0$, then ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)$ becomes unbounded and ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right) \leq$ ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)$. (ii) For different values of ω regarding the ADRs, $\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}$ dominates the $\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}$ and it can be written that ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right) \leq$ ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right) \leq$ ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)$ which is ensured by the following property (iii). (iii) For all $\mathbf{V}$ and ω, ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right) \leq$ ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)$, if the following condition is ensured for the ξ_{max} which denotes the maximum characteristic root:

$$\frac{\operatorname{tr}\left(\tilde{\mathbf{P}}_{21}\tilde{\mathbf{P}}_{11}^{-1}\mathbf{V}\tilde{\mathbf{P}}_{11}^{-1}\tilde{\mathbf{P}}_{12}\tilde{\mathbf{P}}_{22.1}^{-1}\right)}{\xi_{max}\left(\tilde{\mathbf{P}}_{21}\tilde{\mathbf{P}}_{11}^{-1}\mathbf{V}\tilde{\mathbf{P}}_{11}^{-1}\tilde{\mathbf{P}}_{12}\tilde{\mathbf{P}}_{22.1}^{-1}\right)} \geq \frac{p_2+2}{2}.$$

5. OPTIMIZATION OF SMOOTHING PARAMETER

Choosing the smoothing parameter $\lambda > 0$ is vital for efficiency in semiparametric shrinkage strategies. This section outlines the method for selecting λ for smoothing splines. The optimal λ is found using the improved Akaike Information Criterion, $AIC_c(\lambda)$, which corrects finite sample bias for accurate model fit, especially with small sample sizes (n) relative to the number of parameters (p). The formula for $AIC_c(\lambda)$ is:

$$AIC_c\left(\lambda\right) = \log\left(\hat{\sigma}_1^2\right) + 1 + \left[\frac{2(p_1+1)}{n-p_1-2}\right], \tag{5.1}$$

σ_1^2 is defined in (3.12), while p_1 indicates the covariate count for SM or FM based on either penalty function.

6. SIMULATION STUDY

This section describes simulation experiments evaluating the semiparametric shrinkage estimators' performance for the right-censored partially linear model under $p \leq n$. The SM estimator uses predictors selected via the Akaike information criterion (AIC) available in the R package 'AICcmodavg' (Mazerolle and Mazerolle, 2017). Details on the simulation design and data generation follow. A key parameter, $\Delta = \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\|_2 \in [0, 2]$, is defined, with $\boldsymbol{\beta}^0$ as the regression coefficient vector under sparsity, expressed as $\boldsymbol{\beta}^0 = \left(\boldsymbol{\beta}_1^\top, \mathbf{0}_{p_2}^\top\right)^\top$ with a $(p_2 \times 1)$ dimensional vector of zeros. The experiments analyze the estimators' behavior as Δ changes. The simulation experiments include three sample sizes ($n = 75, 150, 300$), two censoring levels ($CL = 5\%, 30\%$), two numbers of covariates ($p = 20, 50$), and four Δ values ($\Delta = 0.0, 0.5, 1.0, 2.0$). Each configuration is repeated 500 times, and estimator performances are evaluated using these metrics:

$$\mathrm{ReMSE}\left(\widehat{\boldsymbol{\beta}}_1^{FM}, \widehat{\boldsymbol{\beta}}_1^{*}\right) = \frac{\mathrm{MSE}\left(\widehat{\boldsymbol{\beta}}_1^{FM}\right)}{\mathrm{MSE}\left(\widehat{\boldsymbol{\beta}}_1^{*}\right)}, \tag{6.1}$$

where in ReMSE (Relative Mean Squared Error) $\widehat{\boldsymbol{\beta}}^*$ denotes the estimate of $\boldsymbol{\beta}_1$ by any estimator, while (6.1) assesses parametric component performance. ReMSE serves as a per-

formance measure in our simulation experiments to evaluate the accuracy of the parametric component estimators for the regression coefficients. It is defined by taking the mean squared error (MSE) of a reference estimator which is $\widehat{\boldsymbol{\beta}}_1^{FM}$ here and dividing it by the MSE of any of the introduced estimators ($\widehat{\boldsymbol{\beta}}_1^{*}$), thereby providing a normalized metric that facilitates direct comparisons across different estimators and simulation settings. This ratio enables us to assess the relative performance of each estimator in capturing the true underlying regression parameters within the right-censored partially linear model framework. Note that ReMSE $\left(\widehat{\boldsymbol{\beta}}_1^{FM}, \widehat{\boldsymbol{\beta}}_1^{*}\right) > 1$ indicates that the introduced estimator performs better than the reference estimator, suggesting more accuracy in estimating the regression coefficients. Conversely, ReMSE $\left(\widehat{\boldsymbol{\beta}}_1^{FM}, \widehat{\boldsymbol{\beta}}_1^{*}\right) < 1$ signifies that the reference estimator outperforms the introduced estimator.

The mean squared error (MSE) evaluates the quality of the nonparametric curve estimate. Let $\hat{\mathbf{f}}^*$ be the estimate of $\mathbf{f}$ from any estimator; $MSE(\hat{\mathbf{f}}^*)$ is then calculated by:

$$\text{MSE}(\hat{\mathbf{f}}^*) = n^{-1}\left[\sum_{i=1}^{n} f\left(t_i\right) - \hat{f}^*\left(t_i\right)\right]^2 = n^{-1}(\mathbf{f} - \hat{\mathbf{f}}^*)^\top(\mathbf{f} - \hat{\mathbf{f}}^*). \tag{6.2}$$

By considering the model (1.1), elements are generated as follows:

$$\begin{aligned} &\mathbf{x}_i \sim MN\left[\boldsymbol{\mu}_{p\times 1}, \boldsymbol{\Sigma}_{p\times p}\right]; t_i = 2.5(i-0.5)/n, f\left(t_i\right) = -t_i \sin\left(-t_i^2\right), \\ &\varepsilon_i \sim N\left(0, \sigma_\varepsilon = 0.5\right) \end{aligned} \tag{6.3}$$

and true regression coefficients are generated by the following equation:

$$\beta_j = \begin{cases} 3 & \text{if } j = 1, \ldots, 5 \\ \beta_{\text{s}} & \Delta > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{6.4}$$

If $\Delta = 0$, then $\beta_s = 0$, implying $\boldsymbol{\beta}_2^\top = \mathbf{0}_{p2}^\top$. Hence, sparsity is ensured. If $\Delta > 0$, then $\Delta = \beta_s^2$ by definition of Δ. According to (6.4), the first 5 coefficients are nonzero, with $(p-5)$ approximately or exactly sparse based on Δ. The outcomes from the estimator are detailed in the subsections for parametric and nonparametric components, informed by the simulation design and data generation processes.

6.1. Results of parametric component

Tables 1-3 show results for three sample sizes evaluating shrinkage and pretest estimators. Table 1 includes ReMSE scores for estimated regression parameters using four methods, along with censoring levels (CL), number of parameters (p), and Δ values to observe estimator behavior under various conditions.

Table 1 shows estimator tendencies for small samples. Performance decreases with higher confidence levels (CL) and Δ. Greater model complexity (p) lowers performance. At high sparsity ($\Delta = 0$), $\hat{\boldsymbol{\beta}}_1^{SM}$ is effective, but as sparsity lessens (Δ), shrinkage estimators, particularly $\hat{\boldsymbol{\beta}}_1^{PT}$, excel. $\hat{\boldsymbol{\beta}}_1^{PS}$ performs well against censorship, except in $CL = 30\%$ and $p = 20$, where it matches $\hat{\boldsymbol{\beta}}_1^{PT}$.

		$p=20$				$p=50$			
CL	Δ	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{PS}$	$\hat{\boldsymbol{\beta}}_1^{PT}$	$\hat{\boldsymbol{\beta}}_1^{SM}$	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{PS}$	$\hat{\boldsymbol{\beta}}_1^{PT}$	$\hat{\boldsymbol{\beta}}_1^{SM}$
5%	0	0.995	1.001	1.001	1.051	0.801	0.802	0.769	0.889
	0.5	0.983	0.982	0.999	0.790	0.737	0.737	0.766	0.630
	1	0.788	0.788	0.797	0.621	0.725	0.726	0.762	0.602
	2	0.685	0.689	0.699	0.561	0.522	0.521	0.563	0.423
30%	0	0.901	0.969	0.921	1.077	0.712	0.712	0.803	0.663
	0.5	0.914	0.965	0.924	0.764	0.682	0.682	0.703	0.531
	1	0.711	0.758	0.733	0.455	0.572	0.573	0.600	0.418
	2	0.509	0.554	0.545	0.451	0.377	0.377	0.411	0.212

Table 1: Calculated $ReMSE$ scores when $n = 75$.

An increase in p generally reduces performance, expectedly due to model complexity. $\hat{\boldsymbol{\beta}}_1^{PT}$ handles small sample sizes better than other estimators. Figure 1 illustrates this, showing larger ReMSE variance from SM compared to shrinkage estimators.

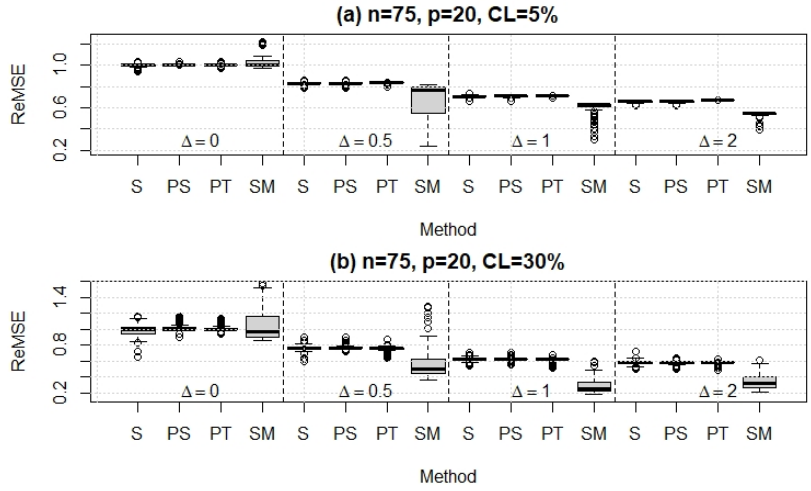

Figure 1: Boxplots of ReMSEs when $n = 75$ for different Δ values.

		$p=20$				$p=50$			
CL	Δ	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{PS}$	$\hat{\boldsymbol{\beta}}_1^{PT}$	$\hat{\boldsymbol{\beta}}_1^{SM}$	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{PS}$	$\hat{\boldsymbol{\beta}}_1^{PT}$	$\hat{\boldsymbol{\beta}}_1^{SM}$
5%	0	0.992	0.999	1.003	0.973	0.847	0.999	0.953	0.886
	0.5	0.985	0.983	0.999	0.944	0.824	0.954	0.907	0.842
	1	0.978	0.981	0.997	0.883	0.731	0.823	0.766	0.723
	2	0.688	0.701	0.708	0.588	0.606	0.694	0.697	0.697
30%	0	0.615	0.771	0.777	0.798	0.562	0.595	0.619	0.666
	0.5	0.691	0.690	0.669	0.355	0.634	0.694	0.582	0.412
	1	0.613	0.627	0.597	0.391	0.508	0.503	0.561	0.309
	2	0.487	0.506	0.460	0.395	0.490	0.521	0.558	0.393

Table 2: Calculated $ReMSE$ scores when $n = 150$.

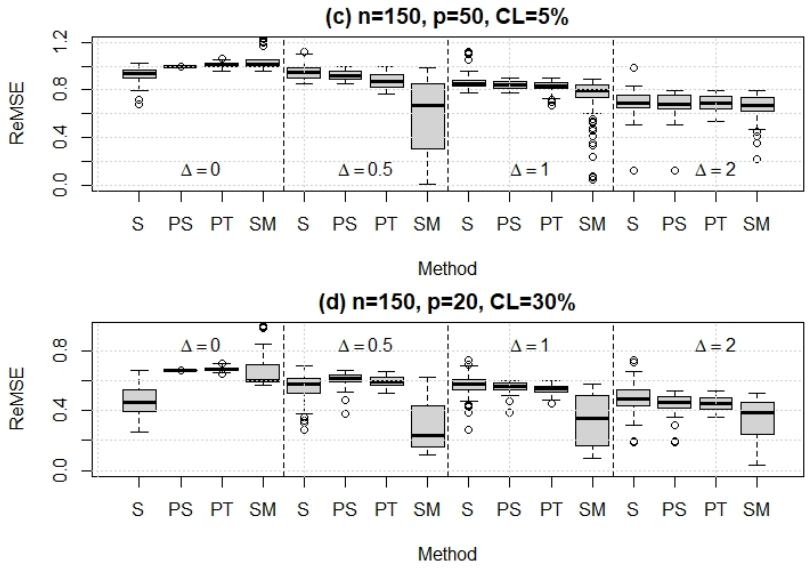

Figure 2: Boxplots of ReMSEs when $n = 150$ for different Δ values.

Table 2 presents ReMSEs for $n = 150$. Thus, the estimator performance is consistently observed with $n = 75$ amid variations in n, CL, and p. Notably, average ReMSEs are more stable, with narrower boxplots in Figure 2 due to large sample sizes.

		$p = 20$				$p = 50$			
CL	Δ	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{PS}$	$\hat{\boldsymbol{\beta}}_1^{PT}$	$\hat{\boldsymbol{\beta}}_1^{SM}$	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{PS}$	$\hat{\boldsymbol{\beta}}_1^{PT}$	$\hat{\boldsymbol{\beta}}_1^{SM}$
5%	0	0.965	1.000	1.001	1.076	0.855	0.855	1.000	0.994
	0.5	0.859	0.865	0.899	0.865	0.729	0.730	0.864	0.790
	1	0.742	0.759	0.784	0.717	0.685	0.685	0.702	0.686
	2	0.551	0.642	0.681	0.580	0.633	0.600	0.662	0.564
30%	0	0.841	1.000	1.011	0.853	0.736	0.800	0.898	0.769
	0.5	0.594	0.698	0.765	0.639	0.437	0.468	0.512	0.363
	1	0.472	0.559	0.559	0.457	0.358	0.401	0.398	0.225
	2	0.466	0.435	0.472	0.293	0.219	0.231	0.243	0.122

Table 3: Calculated $ReMSE$ scores when $n = 300$.

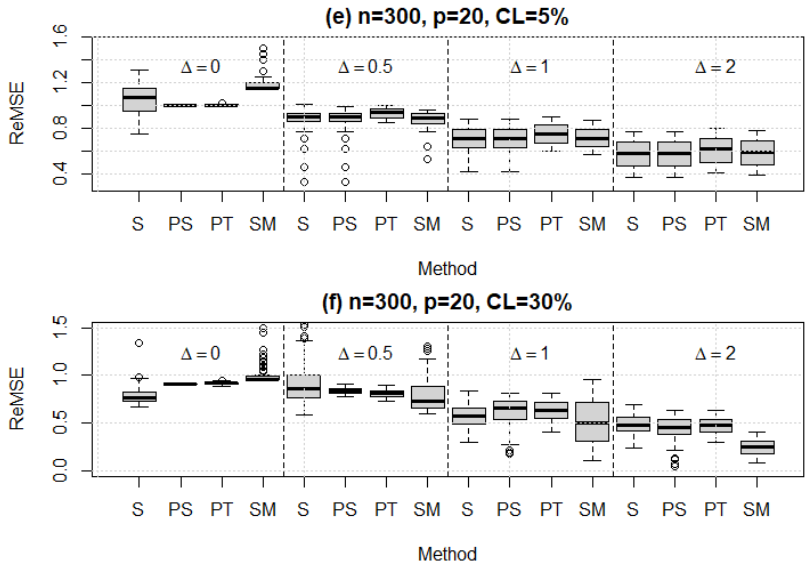

Figure 3: Boxplots of ReMSEs when $n = 300$ for different Δ values.

Under light censorship ($CL = 5\%$), methods perform similarly even if $p = 50$. Under heavy censorship, instability occurs, notably with $\hat{\boldsymbol{\beta}}_1^S$ and $\hat{\boldsymbol{\beta}}_1^{SM}$, likely due to synthetic data transformation increasing data and model variance. Inferences for $n = 75$ and $n = 300$ apply to the ReMSEs in Table 3 and Figure 3, with more stable, higher ReMSEs. Tables and figures show estimator performance declines as Δ values rise under right-censored data.

6.2. Results of nonparametric component

This section presents results for the nonparametric component of the right-censored model using shrinkage estimators. Table 4 shows that these estimators generally outperform $\hat{\mathbf{f}}^{SM}$ in most cases. The differences are clearer when $n = 75$. For larger n, performances are similar, as expected. With $CL = 5\%$, as with the parametric component, $\hat{\mathbf{f}}^{PS}$ and $\hat{\mathbf{f}}^{PT}$ perform best, followed by $\hat{\mathbf{f}}^{S}$. Under heavy censoring ($CL = 30\%$), $\hat{\mathbf{f}}^{SM}$ is the most sensitive. Unlike in the parametric estimation, $\hat{\mathbf{f}}^{S}$ handles censorship better, especially for larger n and p.

		$p = 20$				$p = 50$			
n	CL	$\hat{\mathbf{f}}^S$	$\hat{\mathbf{f}}^{PS}$	$\hat{\mathbf{f}}^{PT}$	$\hat{\mathbf{f}}^{SM}$	$\hat{\mathbf{f}}^S$	$\hat{\mathbf{f}}^{PS}$	$\hat{\mathbf{f}}^{PT}$	$\hat{\mathbf{f}}^{SM}$
$n = 75$	5%	0.056	0.045	0.024	0.136	0.195	0.175	0.167	0.195
	30%	0.405	0.620	0.497	0.857	0.415	0.415	0.455	0.734
$n = 150$	5%	0.023	0.023	0.025	0.024	0.184	0.181	0.177	0.181
	30%	0.105	0.098	0.097	0.098	0.201	0.197	0.195	0.195
$n = 300$	5%	0.003	0.003	0.002	0.011	0.012	0.012	0.013	0.013
	30%	0.043	0.045	0.047	0.048	0.061	0.086	0.081	0.091

Table 4: MSE scores of $\hat{f}(t_i)$ obtained from the introduced estimators for all simulation configurations

Figures 4-5 show how censorship level (CL) and the number of variables (p) affect the fitted curves. Table 4 lists MSE values for the estimated functions. In Figure 4, CL negatively influences curves, with panel (b) deviating towards the x-axis compared to panel (a), aligning synthetic with censored data and causing curves to approach zero.

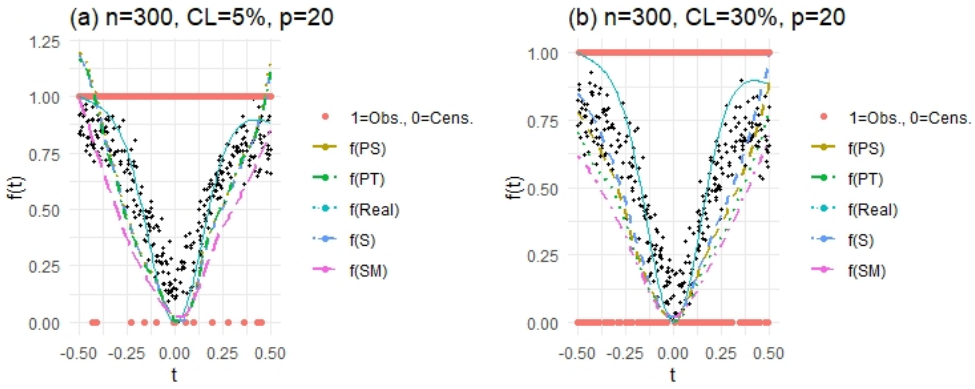

Figure 4: Fitted curves obtained from the three estimators $\hat{\mathbf{f}}^{S}$, $\hat{\mathbf{f}}^{PS}$, $\hat{\mathbf{f}}^{PT}$, and $\hat{\mathbf{f}}^{SM}$ for two simulation configurations to show the effect of the censoring level CL.

Figure 5 illustrates fitted curves for $p = 20$ in panel (a) and a complex model with $p = 50$ in panel (b). Although model complexity affects shrinkage estimator performance under censorship in $p < n$, it minimally impacts the fitted curves.

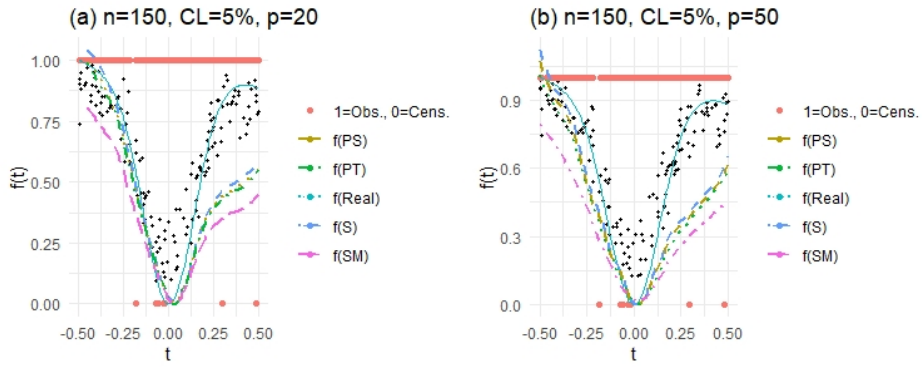

Figure 5: Fitted curves obtained from the three estimators $\hat{\mathbf{f}}^{S}$, $\hat{\mathbf{f}}^{PS}$, $\hat{\mathbf{f}}^{PT}$, and $\hat{\mathbf{f}}^{SM}$ for two simulation configurations to show the effect of the number of parametric covariates p.

7. REAL-DATA EXAMPLE

We model the right-censored Hepatocellular Carcinoma dataset using modified shrinkage estimators, comparing parametric and nonparametric components as in Section 6. The dataset from Li et al. (2014) examines CXCL17 gene expression in hepatocellular carcinoma with 227 data points and 48 variables. Due to insufficient data, 18 variables are excluded.

Analysis uses 18 explanatory variables for the parametric part; the nonparametric part is represented by the *Age* variable due to its relationship with the log of overall survival time (OS). The data includes 84 right-censored OS points, matching heavy censoring levels from simulation configurations $CL = 37\%$-$CL = 30\%$. Results appear in Table 5 and Figure 6. Table 5 presents ReMSEs, MSEs of curves, model variances, and the number of covariates

n	p_1^{FM}	p_1^{SM}		S	PS	PT	SM	FM
227	18	14	$ReMSE(\hat{\beta})$	0.863	1.000	0.701	0.885	-
			$MSE(\hat{f})$	3.926	3.389	3.354	4.834	3.914
			$\hat{\sigma}_1^2$	1.729	1.493	1.212	2.298	1.493

Table 5: Overall results for Hepatocellular dataset.

(p_1^{FM}, p_1^{SM}) for full and submodels. Shrinkage estimators, particularly $\boldsymbol{\beta}_1^S$ and $\boldsymbol{\beta}_1^{PS}$, show higher efficiency, confirmed by $MSE(f)$ with smaller values. Performance differs between SM-based models and shrinkage estimators. Figure 6 shows fitted curves with synthetic and censored data in panel (a) and MSE bar plots in panel (b). In panel (a), $\hat{\mathbf{f}}^{SM}$ skews more towards zeros due to censorship, as Table 5 confirms. Conversely, $\hat{\mathbf{f}}^{PS}$ and $\hat{\mathbf{f}}^{PT}$ better resist censorship, contributing distinctively.

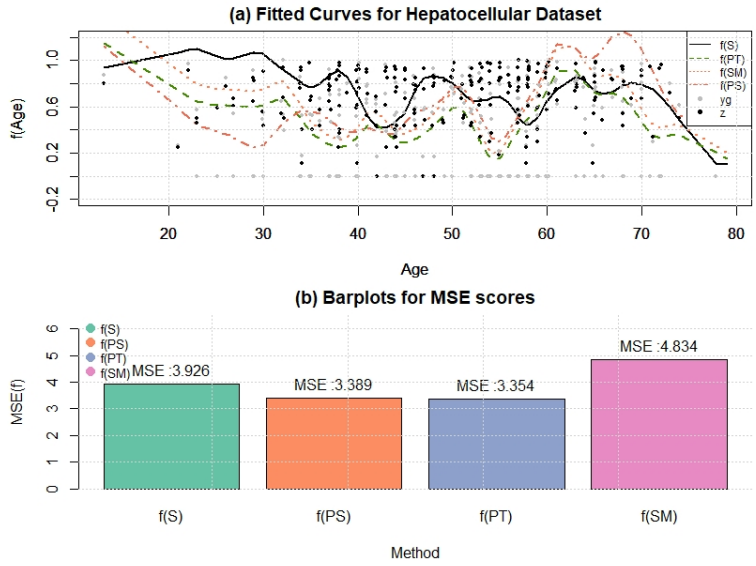

Figure 6: Panel (a): Estimated curves based on the three estimators. Panel (b): Barplot for the MSE scores of estimated functions.

8. HIGH-DIMENSIONAL MODEL ESTIMATION

This section explores numerical studies through a simulation and a real data example to assess modified shrinkage and pretest estimators in high-dimensional, right-censored scenarios. Due to the high dimensionality, suitable FM and SM choices are needed, employing penalty functions as discussed in Section 1. Lasso penalty is used for $\boldsymbol{\beta}_1^{FM}$, while the adaptive Lasso (aL) is used for $\boldsymbol{\beta}_1^{SM}$ due to its harsher penalty and fewer selected predictors.

Subsequent subsections present the simulation study and real data example.

8.1. Simulation study

The simulation design follows 6 for $p > n$, using three sample sizes ($n = 75, 150, 300$), two censoring levels ($CL = 5\%, 30\%$), two covariate counts ($p = 500, 1000$), and four Δ values ($\Delta = 0.0, 0.5, 1.0, 2.0$). Configurations are repeated 500 times, evaluated on ReMSE (6.1) and MSE (6.2). Data generation follows equation (6.3), with regression coefficients by equation (6.4) for Δ and p. Figure 7 illustrates regression coefficients for Δ values, representing exactly and approximately sparse models.

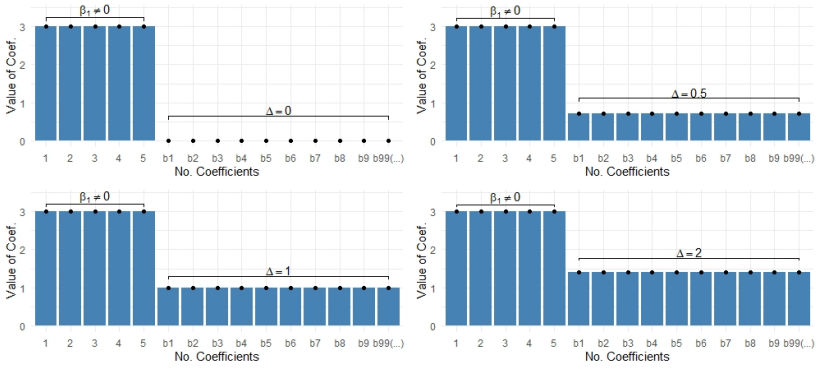

Figure 7: Generated β_j's for different Δ values when $n = 75$, $CL = 5\%, p = 500$.

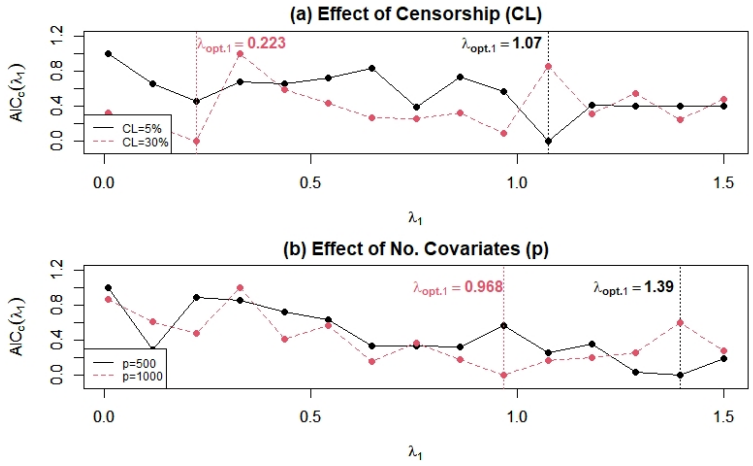

Figure 8: Selection of smoothing parameter λ and illustration of both effect of censorship levels when n=75 and effect of number of covariates (p) when n=150 on the selection of the smoothing parameter λ.

Figure 8 shows λ selection examining the effects of censorship level (CL in Panel (a)) and the number of explanatory variables (p in Panel (b)). In Panel (a), increased CL signif-

icantly reduces λ, as expected from the data structure. Larger p values also lead to smaller λ.

8.1.1. Results of the parametric component

In this section, we present results on $\hat{\boldsymbol{\beta}}$. Refer to Tables 6-7 and Figures 9-10 for an overview. Table 6 shows $ReMSE$ scores for $n = 75$ across different Δ values. Notably, estimator magnitudes decrease with higher Δ values, confirming initial hypotheses. We examine this performance decline, comparing estimators. $\hat{\boldsymbol{\beta}}^{SM}$ declines fastest with higher Δ values, while $\hat{\boldsymbol{\beta}}^{PS}$ and $\hat{\boldsymbol{\beta}}^{PT}$ show slower declines in $ReMSE$ scores, along with $\hat{\boldsymbol{\beta}}^{S}$. This analysis highlights the estimators' robustness and efficacy.

		$p = 500$				$p = 1000$			
CL	Δ	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{PS}$	$\hat{\boldsymbol{\beta}}_1^{PT}$	$\hat{\boldsymbol{\beta}}_1^{SM}$	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{PS}$	$\hat{\boldsymbol{\beta}}_1^{PT}$	$\hat{\boldsymbol{\beta}}_1^{SM}$
5%	0	0.977	1.030	1.006	0.697	0.824	0.818	0.822	0.997
	0.5	0.945	1.029	1.006	0.622	0.797	0.779	0.708	0.631
	1	0.931	0.963	0.997	0.588	0.754	0.679	0.605	0.538
	2	0.413	0.849	0.683	0.061	0.173	0.483	0.372	0.054
30%	0	0.852	0.818	1.019	1.435	0.462	0.699	1.003	0.513
	0.5	0.762	0.736	0.995	0.507	0.348	0.642	0.993	0.438
	1	0.531	0.675	0.989	0.438	0.345	0.507	0.943	0.323
	2	0.244	0.538	0.979	0.091	0.090	0.341	0.930	0.109

Table 6: Calculated $ReMSE$ scores when $n = 75$.

Censoring negatively impacts all estimators, with $n = 75$ showing resilience order as $\hat{\boldsymbol{\beta}}^{PT}$, $\hat{\boldsymbol{\beta}}^{PS}$, and $\hat{\boldsymbol{\beta}}^{S}$. Shrinkage and pretest strategies benefit performance despite censoring. With more explanatory variables (p), estimation performance declines as complexity increases. $\hat{\boldsymbol{\beta}}^{PT}$ is similarly affected, yet provides the best performance.

Table 7 highlights patterns where $\hat{\boldsymbol{\beta}}^{PS}$ and $\hat{\boldsymbol{\beta}}^{PT}$ remain stable as sample size (n) grows, showing resilience to censoring and p variations. The performance gap among estimators decreases unexpectedly. Δ's influence on shrinkage and pretest strategies adapts to changes in Δ, unlike its negative effect on $\hat{\boldsymbol{\beta}}^{SM}$. These findings support earlier analyses, confirming robustness and reliability.

Figures 9-10 confirm the table insights. In Figure 9, boxplots show $ReMSE$ values for $n = 75$ drop as Δ increases, with $\hat{\boldsymbol{\beta}}^{S}$ and $\hat{\boldsymbol{\beta}}^{PT}$ declining less than $\hat{\boldsymbol{\beta}}^{SM}$. When $\Delta = 0$, $\hat{\boldsymbol{\beta}}^{SM}$ meets expectations. Figure 10 shows performances converge as sample size n grows. Figure 10 highlights estimation variances due to model complexity. Panels (a-d) for $CL = 5\%$ and $CL = 30\%$ in each plot show censoring effects. $\hat{\boldsymbol{\beta}}^{PT}$ and $\hat{\boldsymbol{\beta}}^{PS}$ perform more robustly than $\hat{\boldsymbol{\beta}}^{SM}$. Despite close results, $\hat{\boldsymbol{\beta}}^{PT}$ consistently outperforms, confirmed by detailed figure analysis.

		$p = 500$				$p = 1000$			
CL	Δ	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{PS}$	$\hat{\boldsymbol{\beta}}_1^{PT}$	$\hat{\boldsymbol{\beta}}_1^{SM}$	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{PS}$	$\hat{\boldsymbol{\beta}}_1^{PT}$	$\hat{\boldsymbol{\beta}}_1^{SM}$
5%	0	1.078	1.131	1.000	0.848	0.996	0.896	0.993	0.836
	0.5	0.980	1.029	1.000	0.808	0.992	0.879	0.987	0.769
	1	0.971	1.018	1.000	0.737	0.965	0.869	0.984	0.719
	2	0.958	0.991	0.996	0.201	0.009	0.816	0.491	0.103
30%	0	1.333	1.058	1.000	0.859	1.011	1.091	0.958	0.818
	0.5	1.014	1.022	1.000	0.846	0.991	1.012	0.945	0.759
	1	1.002	0.977	1.000	0.789	0.964	1.011	0.925	0.700
	2	1.010	0.974	1.000	0.609	0.260	0.850	0.257	0.259

Table 7: Calculated $ReMSE$ scores when $n = 300$.

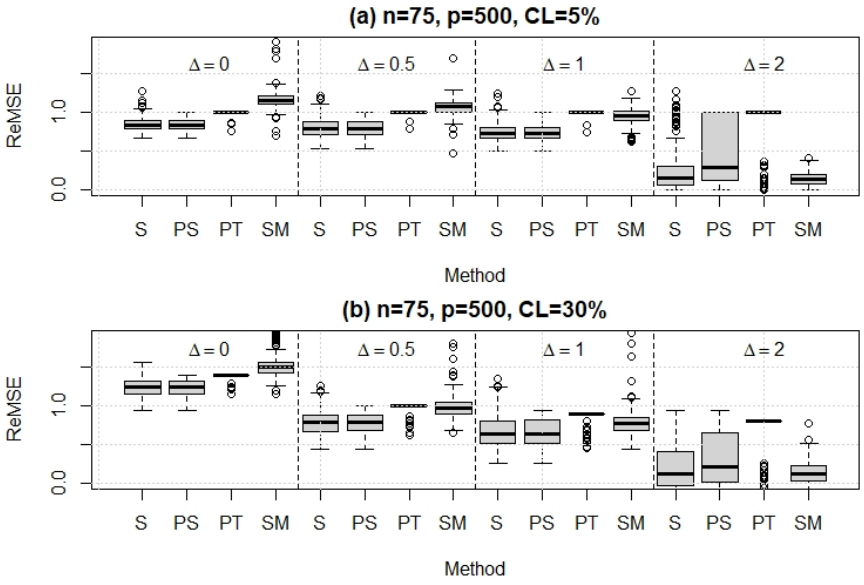

Figure 9: Boxplots of obtained $ReMSE$ scores for the three estimators $\hat{\boldsymbol{\beta}}_1^S$, $\hat{\boldsymbol{\beta}}_1^{PS}$, $\hat{\boldsymbol{\beta}}_1^{PT}$ and $\hat{\boldsymbol{\beta}}_1^{SM}$ for different values of Δ when $n = 75$.

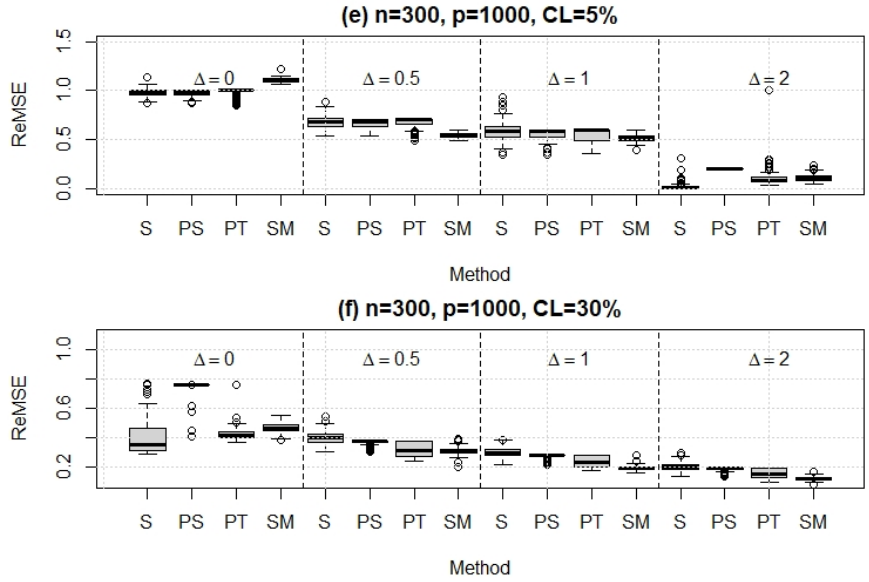

Figure 10:Boxplots of obtained ReMSE scores for the three estimators when $n = 300$.

8.1.2. Results of the nonparametric component

Table 8 and Figures 11-12 show nonparametric component estimation results. MSE values for estimated functions in Table 8 reveal the expected negative impact of censoring on all three estimators' prediction performances. A noticeable increase in MSE values occurs with $p = 1000$. However, as sample sizes increase to $n = 150$ and $n = 300$, these negative effects lessen, aligning the performances.

		$p = 500$				$p = 1000$			
n	CL	$\hat{\mathbf{f}}^{S}$	$\hat{\mathbf{f}}^{PS}$	$\hat{\mathbf{f}}^{PT}$	$\hat{\mathbf{f}}^{SM}$	$\hat{\mathbf{f}}^{S}$	$\hat{\mathbf{f}}^{PS}$	$\hat{\mathbf{f}}^{PT}$	$\hat{\mathbf{f}}^{SM}$
$n = 75$	5%	0.043	0.043	0.051	0.056	0.335	0.320	0.310	0.633
	30%	0.447	0.447	0.191	0.226	0.904	0.728	0.740	0.727
$n = 150$	5%	0.017	0.016	0.007	0.053	0.261	0.249	0.283	0.279
	30%	0.099	0.077	0.081	0.121	0.351	0.179	0.206	0.301
$n = 300$	5%	0.005	0.005	0.003	0.010	0.014	0.014	0.012	0.015
	30%	0.075	0.075	0.066	0.081	0.079	0.099	0.076	0.219

Table 8: MSE scores of estimated nonparametric component for all possible simulation configurations.

Examining these values reveals key insights. For $n = 75$, $\hat{\mathbf{f}}^{PS}$ and $\hat{\mathbf{f}}^{S}$ are notably affected by censoring when $p = 500$. In contrast, $\hat{\mathbf{f}}^{SM}$ with $p = 1000$ shows higher MSE values when $CL = 30\%$, illustrating the shrinkage strategy's benefit with censored data. Across simulations, $\hat{\mathbf{f}}^{PT}$ consistently excels in estimating the non-parametric component.

Table 8 outcomes are visually shown in Figures 11-12 with panels (a) and (b). Figure 11 illustrates how sample size affects estimator performance. In panel (a), $n = 75$ with low censoring ($CL = 5\%$) trends toward zero, indicating failure in representation, while panel

(b) shows $n = 150$'s improved representation under similar conditions. Figure 12 examines censoring level effects on estimation performance. In panel (a), low censoring shifts $\hat{\mathbf{f}}^S$ and $\hat{\mathbf{f}}^{SM}$ toward zero, but $\hat{\mathbf{f}}^{PT}$ and $\hat{\mathbf{f}}^{PS}$ are less impacted. At $CL = 30\%$, representation by all estimators declines significantly.

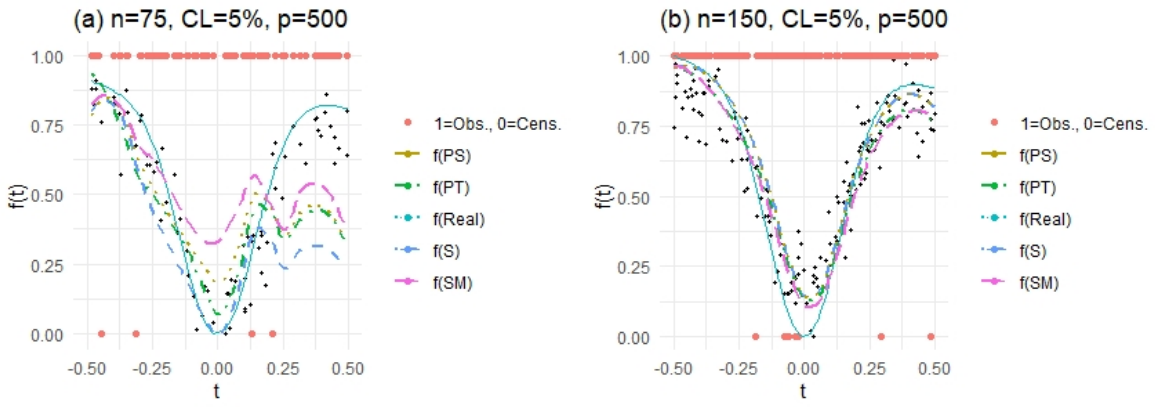

Figure 11:Fitted curves obtained from the three estimators $\hat{\mathbf{f}}^S$, $\hat{\mathbf{f}}^{PS}$, $\hat{\mathbf{f}}^{PT}$, and $\hat{\mathbf{f}}^{SM}$ for two simulation configurations to show the effect of the sample size n.

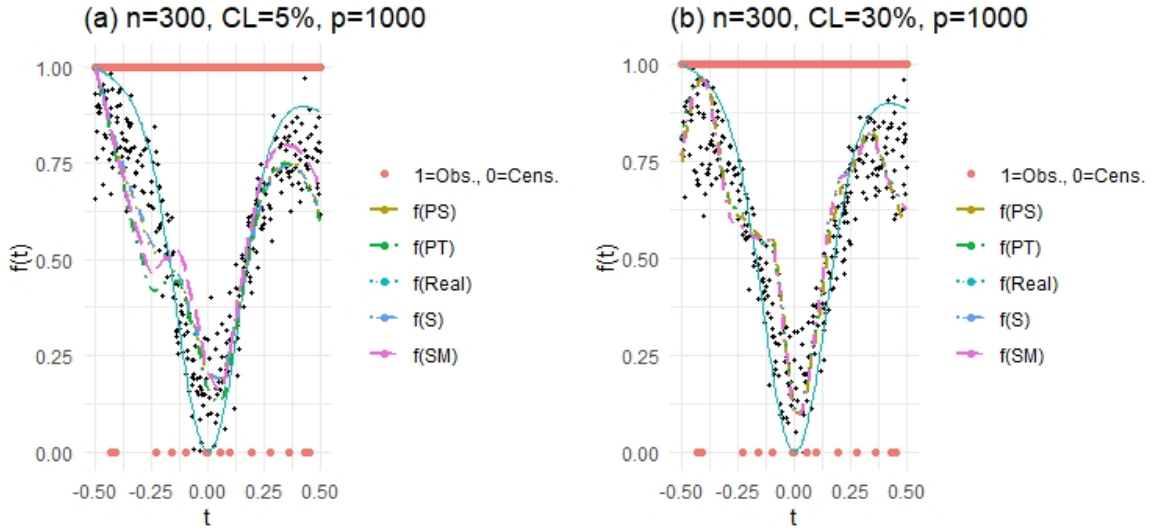

Figure 12:Fitted curves obtained from the three estimators $\hat{\mathbf{f}}^S$, $\hat{\mathbf{f}}^{PS}$, $\hat{\mathbf{f}}^{PT}$, and $\hat{\mathbf{f}}^{SM}$ for two simulation configurations to show the effect of the censoring level CL.

8.2. Real data example

We analyze the right-censored semiparametric regression model using the NSBC dataset, which includes gene expressions from 115 tumors, as reported by Sørlie et al. (2003). Of the

115 patients, 38 experienced an event, yielding a censoring rate of $CL = 33\%$. The model's nonparametric part uses the univariate variable t_i from the 393^{rd} gene expression column, while the parametric part involves 548 explanatory variables for estimating survival times. The defined model for high-dimensional data is:

$$survtime_{i\hat{G}} = \mathbf{x}_i\boldsymbol{\beta} + f(t_i) + \varepsilon_{i\hat{G}},\ i = 1, \ldots, 115. \tag{8.1}$$

Here, $\mathbf{x}_i = \left\{(x_{i1}, \ldots, x_{ip})^\top,\ i = 1, 2, \ldots, n\right\}$ represents the vectors of the high-dimensional design matrix, and $survtime_{i\hat{G}}$ denotes the synthetic response variable. Model (8.1) estimation results are shown in Table 9 and Figure 13. The smoothing parameter is set as $\lambda = 0.735$ by the AIC_c criterion, while shrinkage parameters are $\lambda_s^{Lasso} = 0.0075$ and $\lambda_s^{aLas} = 0.0272$, obtained through k-cross-validation with the "glmnet" package in R.

Table 9 summarizes NSBC data analysis. $ReMSE$ values show $\hat{\boldsymbol{\beta}}^{PT}$ and $\hat{\boldsymbol{\beta}}^{PS}$ outperform p^{FM}, with $\hat{\boldsymbol{\beta}}^{S}$ following closely, matching simulation results. $\hat{\boldsymbol{\beta}}^{SM}$ underperforms in parametric estimation. For $f(t_i)$, $\hat{\mathbf{f}}^{PT}$ has the lowest MSE score, while $\hat{\mathbf{f}}^{FM}$ does well but is complex. Despite similar $\hat{\sigma}_1^2$ values, shrinkage and pretest strategies reduce model variances.

n	p_1^{FM}	p_1^{SM}		S	PS	PT	SM	FM
115	36	4	$ReMSE(\hat{\beta})$	0.977	1.000	1.000	0.845	-
			$MSE(\hat{f})$	0.167	0.166	0.166	0.196	0.183
			$\hat{\sigma}_1^2$	0.020	0.019	0.019	0.022	0.020

Table 9: Overall results for NSBC dataset.

Figure 13 shows function estimates for the non-parametric component (panel (a)) and bar plots for MSE values (panel (b)). Panel (a) demonstrates that $\hat{\mathbf{f}}^{PS}$ and $\hat{\mathbf{f}}^{PT}$ better represent the data than $\hat{\mathbf{f}}^{SM}$, confirming Table 9 results. Gray shading in panel (a) highlights synthetic responses, emphasizing censoring effects. Panel (b) clearly shows estimator differences.

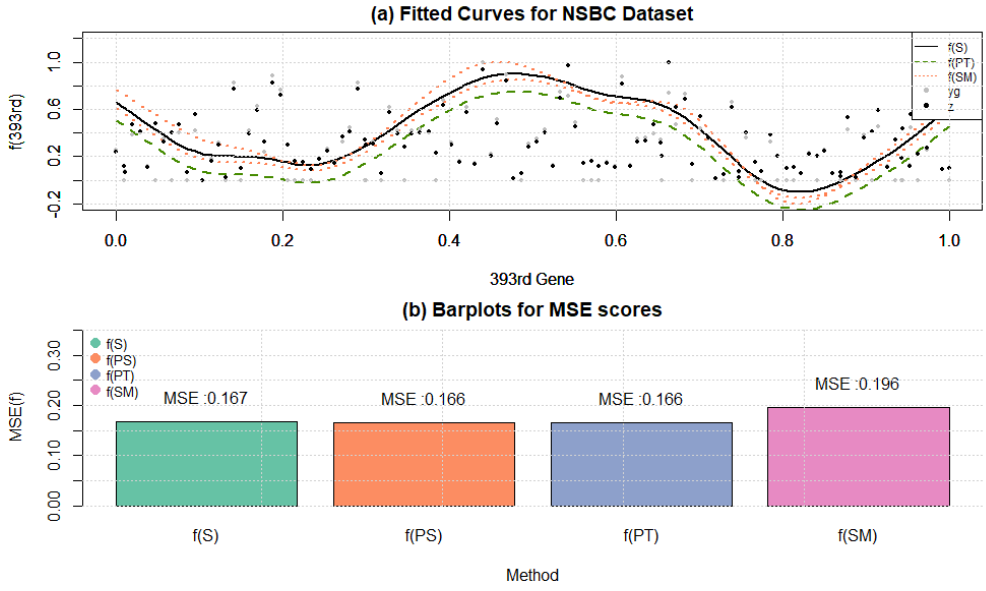

Figure 13:Panel (a): Estimated curves based on the three estimators. Panel (b): Barplot for the MSE scores of estimated functions.

9. CONCLUSIONS

This paper examines modified estimators for right-censored, sparse semiparametric regression models, offering theoretical inferences for both $p < n$ and $p > n$ cases. Using synthetic data transformation, shrinkage, and pretest strategies, it presents a robust approach for right-censored data. The estimators perform consistently across low and high-dimensional datasets. To address high dimensionality, the Lasso function for FM and the adaptive Lasso for SM are used. Simulations confirm their efficacy in complex or less sparse conditions ($\Delta > 0$). The effects of censoring level (CL) and parametric covariate count (p) are analyzed, showing strong performance of pretest and shrinkage estimators. Empirical analysis with Hepatocellular Carcinoma data ($p < n$) and NSBC dataset ($p > n$) further supports these findings, with reliable results from the introduced estimators.

This paper presents innovative estimators that provide a basis for advancing right-censored sparse semiparametric regression models in both low and high dimensions. These tools will be crucial for future data analysis as dataset complexity increases.

Data availability statement: There are two real data examples in the paper. The Hepatocellular Carcinoma dataset is publicly available in asaur package of R. To reach the NSBC Dataset, see supplementary materials of Sørlie et al. (2003).

ACKNOWLEDGMENTS

The research of S. Ejaz Ahmed was supported by the Natural Sciences and the Engineering Research Council (NSERC) of Canada.

REFERENCES

Ahmed, S. E. (2014). *Penalty, shrinkage and pretest strategies: variable selection and estimation*. Springer, New York.

Ahmed, S. E., Ahmed, F., and Yüzbaşı, B. (2023). *Post-Shrinkage Strategies in Statistical and Machine Learning for High Dimensional Data*. CRC Press.

Ahmed, S. E., Doksum, K. A., Hossain, S., and You, J. (2007). Shrinkage, pretest, and absolute penalty estimators in partially linear models. *Australian & New Zealand Journal of Statistics*, 49(4):435–454.

Aydin, D. and Yilmaz, E. (2018). Modified estimators in semiparametric regression models with right-censored data. *Journal of Statistical Computation and Simulation*, 88(8):1470–1498.

Aydın, D., Ahmed, S. E., and Yılmaz, E. (2019). Estimation of semiparametric regression model with right-censored high-dimensional data. *Journal of Statistical Computation and Simulation*, 89(6):985–1004.

Gao, J. (1995). Asymptotic theory for partly linear models. *Communications in Statistics - Theory and Methods*, 24(8):1985–2009.

Hossain, S., Doksum, K. A., and Ahmed, S. E. (2009). Positive shrinkage, improved pretest, and absolute penalty estimators in partially linear models. *Linear Algebra and Its Applications*, 430(10):2749–2761.

Li, L., Yan, J., Xu, J., Liu, C.-Q., Zhen, Z.-J., Chen, H.-W., Ji, Y., Wu, Z.-P., Hu, J.-Y., Zheng, L., et al. (2014). Cxcl17 expression predicts poor prognosis and correlates with adverse immune infiltration in hepatocellular carcinoma. *PloS one*, 9(10):e110064.

Mazerolle, M. J. and Mazerolle, M. M. J. (2017). Package 'aiccmodavg'. R package.

Phukongtong, S., Lisawadi, S., and Ahmed, S. E. (2022). Penalty, post pretest and shrinkage strategies in a partially linear model. *Communications in Statistics - Simulation and Computation*, 51(10):6004–6025.

Raheem, S. E., Ahmed, S. E., and Doksum, K. A. (2012). Absolute penalty and shrinkage estimation in partially linear models. *Computational Statistics & Data Analysis*, 56(4):874–891.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, New York.

Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica*, 9:1089–1102.

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lønning, P. E., Brown, P. O., Børresen-Dale, A. L., and Botstein, D. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences*, 100(14):8418–8423.

Yenilmez, I., Yilmaz, E., Kantar, Y. M., and Aydin, D. (2022). Comparison of parametric and semiparametric models with randomly right-censored data by weighted estimators: Two applications in colon cancer and hepatocellular carcinoma datasets. *Statistical Methods in Medical Research*, 31(2):372–387.

Yüzbaşı, B., Arashi, M., and Ejaz Ahmed, S. (2020). Shrinkage estimation strategies in generalised ridge regression models: Low/high-dimension regime. *International Statistical Review*, 88(1):229–251.

Appendices

A1. Proof of Theorem 3.1

Under the assumptions (i-iii) provided in Section 3, we can proceed as follows: Define the synthetic responses adjusted for right-censoring as:

$$y_{i\hat{G}} = \frac{\delta_i z_i}{1 - \hat{G}(z_i)}, \quad \text{where} \quad z_i = \min(y_i, c_i), \quad \delta_i = I(y_i \leq c_i),$$

and $\hat{G}(z_i)$ is the Kaplan-Meier estimator of the censoring distribution evaluated at z_i.

The full model estimator $\hat{\boldsymbol{\beta}}^{\mathrm{FM}}$ is obtained by minimizing the penalized least squares criterion incorporating smoothing splines for the nonparametric component:

$$\hat{\boldsymbol{\beta}}^{\mathrm{FM}} = \left(\widetilde{\mathbf{X}}_1^\top \widetilde{\mathbf{X}}_1\right)^{-1} \widetilde{\mathbf{X}}_1^\top \widetilde{\mathbf{y}}_{\hat{G}},$$

where $\widetilde{\mathbf{X}}_1$ includes the design matrix for the parametric covariates adjusted for the smoothing spline estimates of $f(t_i)$, and $\widetilde{\mathbf{y}}_{\hat{G}}$ are the transformed responses. Under regularity conditions, as $n \to \infty$,

$$\frac{1}{n}\widetilde{\mathbf{X}}_1^\top \widetilde{\mathbf{X}}_1 \xrightarrow{p} \mathbf{B},$$

where $\mathbf{B}$ is a positive definite matrix.

The sum $\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_1^\top \widetilde{\epsilon}_{\hat{G}}$ converges in distribution to a multivariate normal distribution:

$$\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_1^\top \widetilde{\epsilon}_{\hat{G}} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2_{\text{eff}}\mathbf{B}\right),$$

where $\widetilde{\epsilon}_{\hat{G}} = \widetilde{\mathbf{y}}_{\hat{G}} - \widetilde{\mathbf{X}}_1\boldsymbol{\beta}$ and σ^2_{eff} encapsulates the additional variability from censoring and smoothing.

Combining the above results, we have:

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{\text{FM}} - \boldsymbol{\beta}\right) = \left(\frac{1}{n}\widetilde{\mathbf{X}}_1^\top \widetilde{\mathbf{X}}_1\right)^{-1} \frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_1^\top \widetilde{\epsilon}_{\hat{G}} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2_{\text{eff}}\mathbf{B}^{-1}\right).$$

Thus, under assumptions (i)-(viii), the estimator $\hat{\boldsymbol{\beta}}^{\text{FM}}$ is asymptotically normal with mean $\boldsymbol{\beta}$ and covariance matrix $\sigma^2_{\text{eff}}\mathbf{B}^{-1}$.