entropy

MDPI

Article

Efficient Post-Shrinkage Estimation Strategies in High-Dimensional Cox's Proportional Hazards Models

Syed Ejaz Ahmed [1], Reza Arabi Belaghi [2] and Abdulkhadir Ahmed Hussein [3,*]

[1] Department of Mathematics and Statistics, Brock University, St. Catharines, ON L2S 3A1, Canada; sahmed5@brocku.ca
[2] Department of Energy and Technology, Swedish University of Agricultural Sciences, P.O. Box 7032, 750 07 Uppsala, Sweden; reza.belaghi@slu.se
[3] Department Mathematics & Statistics, University of Windsor, Windsor, ON N9B 3P4, Canada
* Correspondence: ahussein@uwindsor.ca

Abstract: Regularization methods such as LASSO, adaptive LASSO, Elastic-Net, and SCAD are widely employed for variable selection in statistical modeling. However, these methods primarily focus on variables with strong effects while often overlooking weaker signals, potentially leading to biased parameter estimates. To address this limitation, Gao, Ahmed, and Feng (2017) introduced a corrected shrinkage estimator that incorporates both weak and strong signals, though their results were confined to linear models. The applicability of such approaches to survival data remains unclear, despite the prevalence of survival regression involving both strong and weak effects in biomedical research. To bridge this gap, we propose a novel class of post-selection shrinkage estimators tailored to the Cox model framework. We establish the asymptotic properties of the proposed estimators and demonstrate their potential to enhance estimation and prediction accuracy through simulations that explicitly incorporate weak signals. Finally, we validate the practical utility of our approach by applying it to two real-world datasets, showcasing its advantages over existing methods.

Keywords: variable selection; high-dimensional data; Cox proportional hazards model; LASSO; shrinkage estimation; sparse model

MSC: 62J12; 62J07; 62E20; 62F25

Academic Editor: Éloi Bossé

Received: 3 January 2025
Revised: 17 February 2025
Accepted: 19 February 2025
Published: 28 February 2025

Citation: Ahmed, S.E.; Arabi Belaghi, R.; Hussein, A.A. Efficient Post-Shrinkage Estimation Strategies in High-Dimensional Cox's Proportional Hazards Models. *Entropy* **2025**, *27*, 254. https://doi.org/10.3390/e27030254

1. Introduction

High-dimensional data analysis, where the number of covariates frequently exceeds the sample size, has become a central research focus in contemporary statistics (see [1]). The applications of these methods span a broad range of fields, including genomics, medical imaging, signal processing, social science, and financial economics. In particular, high-dimensional regularized Cox regression models have gained traction in survival analysis (e.g., [2–4]), where these techniques help construct parsimonious (sparse) models and can outperform classical selection criteria such as Akaike's information criterion [5] or the Bayesian information criterion [6].

The least absolute shrinkage and selection operation (LASSO) proposed by [7] remains one of the most popular approaches to high-dimensional regression, due to its computational efficiency and its ability to perform variable selection and parameter shrinkage simultaneously. Numerous extensions of LASSO, such as adaptive LASSO [8], elastic

net [9], and scaled LASSO [10], have been developed to further refine estimation and prediction performance. In the context of Cox proportional hazards models, analogous methods—including the LASSO [4,11], the adaptive LASSO [12,13], and smoothly clipped absolute deviation (SCAD; [14])—have been widely examined. Interested readers may also consult [15–18] for recent advancements in high-dimensional Cox regression.

When $p > n$, the focus is often on accurately recovering both the support (i.e., which covariates have nonzero effects) and the magnitudes of the nonzero regression coefficients. Although many penalized inference procedures excel at identifying "strong" signals (i.e., coefficients that are moderately large and thus easily detected), they may fail in adequately accounting for "weak" signals, whose effects may be small but nonzero. To formalize this, one can divide the index set $\{1, \ldots, p_n\}$ into three disjoint subsets as follows: S_1 for strong signals, S_2 for weak signals, and S_{null} for coefficients that are exactly zero. Standard estimation procedures that neglect weak signals risk introducing non-negligible bias, particularly when these weak signals are numerous.

In this paper, we tackle the bias induced by weak signals in high-dimensional Cox regression by adapting the post-selection shrinkage strategy proposed by [19]. Our key contribution is the development of a weighted ridge (WR) estimator, which effectively differentiates small, nonzero coefficients from those that are truly zero. We show that the resulting post-selection estimators dominate submodel estimators derived from standard regularization methods such as LASSO and elastic net. Moreover, under the condition $p_n = O(n^\alpha)$ for some $\alpha > 0$, we establish the asymptotic normality of our post-selection WR estimator, thereby demonstrating its asymptotic efficiency. Through extensive simulations and real data applications, we illustrate that our method achieves substantial improvements in both estimation accuracy and prediction performance.

The remainder of this paper is organized as follows. Section 2 presents the model setup and the proposed post-selection shrinkage estimation procedure. In Section 3, we outline the asymptotic properties of our estimators. Section 4 provides a Monte Carlo simulation study, while Section 5 reports the results of applying our methodology to two real data sets. We conclude in Section 6 with a brief discussion of possible future research directions.

2. Methodology

2.1. Notation and Assumptions

In this section, we state some standard notations and assumptions, used throughout the paper. We use bold upper case letters for matrices and lower case letters for vectors. Moreover, T denotes the matrix transpose and $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix. Design vectors, or columns of $\boldsymbol{X}$, are denoted by $\boldsymbol{X}_j, j = 1, \cdots, p_n$. The index set $\mathcal{M} = \{1, 2, \cdots, p_n\}$ denotes the full model which contains all the potential variables. For a subset $\mathcal{A} \subset \mathcal{M}$, we use $\boldsymbol{\beta}_\mathcal{A}$ for a subvector of $\boldsymbol{\beta}_\mathcal{M}$ indexed by $\mathcal{A}$, and $\boldsymbol{X}_\mathcal{A}$ for a submatrix of $\boldsymbol{X}$ whose columns are indexed by $\mathcal{A}$. For a vector $\mathbf{v} = (v_1, \cdots, v_{p_n})^\mathsf{T}$, we denote $||\mathbf{v}||_2 = \sqrt{\sum_{j=1}^{p_n} v_j^2}$ and $||\mathbf{v}||_1 = \sum_{j=1}^{p_n} |v_j|$. For any square matrix $\boldsymbol{A}$, we let $\Lambda_{\min}(\boldsymbol{A})$ and $\Lambda_{\max}(\boldsymbol{A})$ be the smallest and largest eigenvalues of $\boldsymbol{A}$, respectively. Given $a, b \in \mathbb{R}$, we let $a \vee b$ and $a \wedge b$ denote the maximum and minimum of a and b. For two positive sequences a_n and b_n, $a_n \asymp b_n$, if a_n is in the same order as b_n. We use $I(.)$ to denote the indicator function; $H_\vartheta(.; \Delta)$ denotes the cumulative distribution function (cdf) of a non-central χ^2-distribution with ϑ degrees of freedom and non-centrality parameter Δ. We also use $\xrightarrow{\mathcal{D}}$ to indicate convergence in distribution.

Let $\mathcal{S} \subset \{1, \cdots, p_n\}$ be the set of the indices of nonzero coefficients, with $s = |\mathcal{S}|$ denoting the cardinality of $\mathcal{S}$. We assume that the true coefficient vector $\boldsymbol{\beta}^* = (\beta_1^{*\mathsf{T}}, \cdots, \beta_{p_n}^{*\mathsf{T}})^\mathsf{T}$ is sparse, that is $s < n$. Without loss of generality, we partition the $(n \times p_n)$-matrix $\boldsymbol{X}$ as $\boldsymbol{X} = (\boldsymbol{X}_{\mathcal{S}_1}, \boldsymbol{X}_{\mathcal{S}_2}, \boldsymbol{X}_{\mathcal{S}_{\text{null}}})^\mathsf{T}$, where $\mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathcal{S}_{\text{null}} = \varnothing$, $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_{\text{null}} = \mathcal{M}$ and

$\mathcal{S}_{\text{null}} = \{j : \beta_{0j} = 0\}$. For two matrices $\boldsymbol{X}_{\mathcal{S}_1}$ and $\boldsymbol{X}_{\mathcal{S}_2}$, we define the corresponding sample covariance matrices by

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathcal{S}_1|\mathcal{S}_2} &= \boldsymbol{\Sigma}_{\mathcal{S}_1\mathcal{S}_1} - \boldsymbol{\Sigma}_{\mathcal{S}_1\mathcal{S}_2}\boldsymbol{\Sigma}_{\mathcal{S}_2\mathcal{S}_2}^{-1}\boldsymbol{\Sigma}_{\mathcal{S}_2\mathcal{S}_1},\\ \boldsymbol{\Sigma}_{\mathcal{S}_2|\mathcal{S}_1} &= \boldsymbol{\Sigma}_{\mathcal{S}_2\mathcal{S}_2} - \boldsymbol{\Sigma}_{\mathcal{S}_2\mathcal{S}_1}\boldsymbol{\Sigma}_{\mathcal{S}_1\mathcal{S}_1}^{-1}\boldsymbol{\Sigma}_{\mathcal{S}_1\mathcal{S}_2}.\end{aligned} \tag{1}$$

Let $\boldsymbol{V} = (\boldsymbol{X}_{S_2}, \boldsymbol{X}_{S_{\text{null}}})^{\mathsf{T}}$ be a $p_n - s_1$ submatrix of $\boldsymbol{X}$. Then, another partition can be written as $\boldsymbol{X} = (\boldsymbol{X}_{S_1}, \boldsymbol{V})^{\mathsf{T}}$. Let $\boldsymbol{M}_1 = \boldsymbol{I}_n - \boldsymbol{X}_{S_1}\hat{\boldsymbol{\Sigma}}_{S_1S_1}^{-1}\boldsymbol{X}_{S_1}^{\mathsf{T}}$. Then, $\boldsymbol{V}^{\mathsf{T}}\boldsymbol{M}_1\boldsymbol{V}$ is a $(p_n - s_1) \times (p_n - s_1)$ dimensional singular matrix with rank $k_1 \geq 0$. We denote $\varrho_1 \leq \cdots \leq \varrho_{k_1}$ as all the k_1 positive eigenvalues of $\boldsymbol{V}^{\mathsf{T}}\boldsymbol{M}_1\boldsymbol{V}$.

2.2. Signal Strength Regularity Conditions

We consider three signal strength assumptions to define three sets of covariates according to their signal strength levels as follows [19]:

(A1) There exists a positive constant c_1, such that $|\beta_j| \gtrsim c_1\sqrt{(\log p)/n}$ for $\forall j \in S_1$;

(A2) The coefficient vector $\boldsymbol{\beta}$ satisfies $||\boldsymbol{\beta}_{S_2}||_2^2 \sim O(n^{\tau})$ for some $0 < \tau < 1$, where $\beta_j \neq 0$ for $\forall j \in S_2$;

(A3) $\beta_j = 0$, for $\forall j \in S_{\text{null}}$.

2.3. Cox Proportional Hazards Model

The proportional hazards (PH) model introduced by [20] is one of the most commonly used approaches for analyzing survival data. In this model, the hazard function for an individual depends on covariates through a multiplicative effect, implying that the ratio of hazards for different individuals remains constant over time. We consider a survival model with a true hazard function $\lambda_0(t \mid \boldsymbol{X})$ for a failure time T, given a covariate vector $\boldsymbol{X} = (X_1, \ldots, X_p)^{\mathsf{T}}$. We let C denote the censoring time and define $Y = \min(T, C)$ and $\delta = I(T \leq C)$. Suppose we have n i.i.d. observations $\{(Y_i, \delta_i, X_i)\}_{i=1}^n$ from this true underlying model, where $\boldsymbol{X} = (X_1, \ldots, X_p)^{\mathsf{T}}$ represents the $n \times p$ design matrix.

The PH model posits that the hazard function for an individual with covariates $\boldsymbol{X}$ is

$$\lambda(t \mid \boldsymbol{X}) \;=\; \lambda_0(t)\,\exp(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\beta}), \tag{2}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathsf{T}}$ is the vector of regression coefficients, and $\lambda_0(t)$ is an unknown baseline hazard function. Because $\lambda_0(t)$ does not depend on $\boldsymbol{X}$, one can estimate $\boldsymbol{\beta}$ by maximizing the partial log-likelihood

$$l(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n}\delta_i\,(\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta}) \;-\; \sum_{i=1}^{n}\delta_i\,\log\Bigg(\sum_{j\in\mathbb{R}(t_i)}\exp(\boldsymbol{x}_j^{\mathsf{T}}\boldsymbol{\beta})\Bigg), \tag{3}$$

where $\delta_i = I(T_i \leq C_i)$ and $\mathbb{R}(t_i) = \{\, j : T_j \geq t_i\}$ is the risk set just prior to t_i. Maximizing $l(\boldsymbol{\beta})$ in (3) with respect to $\boldsymbol{\beta}$ yields the estimator $\hat{\boldsymbol{\beta}}$ for the regression parameters.

2.4. Variable Selection and Estimation

Variable selection can be carried out by minimizing the penalized negative log-partial likelihood as follows:

$$-\,l(\boldsymbol{\beta}) \;+\; \sum_{j=1}^{p_n} P_\lambda(\beta_j), \tag{4}$$

where $P_\lambda(\beta_j)$ is a penalty function applied to each component of $\boldsymbol{\beta}$, and λ is a tuning parameter that controls the magnitude of penalization. We consider the following two popular methods:

1. **LASSO.** The LASSO estimator follows (4) with an L_1-norm penalty,

$$\mathrm{Pen}_\lambda(\beta_j) \;=\; \lambda|\beta_j|.$$

As λ increases, this penalty continuously shrinks the coefficients toward zero, and some coefficients become exactly zero if λ is sufficiently large. The theoretical properties of the LASSO are well studied; see [21] for an extensive review.

2. **Elastic Net (ENet).** The Elastic Net estimator implements (4) with the combined penalty

$$P_\lambda(\beta_j) \;=\; \lambda\Big(\alpha\,|\beta_j| \;+\; (1-\alpha)\,\beta_j^2\Big), \tag{5}$$

where $0 \le \alpha \le 1$. When $\alpha = 1$, this reduces to the LASSO, and when $\alpha = 0$, it becomes Ridge. Combining L_1 and L_2 penalties leverages the benefits of Ridge while still producing sparse solutions. Unlike LASSO, which can select n variables at most, ENet has no such limitation when $p_n > n$.

2.4.1. Variable Selection Procedure for $\mathcal{S}_1$ and $\mathcal{S}_2$

We summarize the variable selection procedure for detecting the strong signals $\mathcal{S}_1$ and the weak signals $\mathcal{S}_2$.

Step 1 (detection of $\mathcal{S}_1$). Obtain a candidate subset $\hat{\mathcal{S}}_1$ of strong signals using a penalized likelihood estimator (PLE). Specifically, consider

$$\hat{\boldsymbol{\beta}}^{\mathrm{PLE}} \;=\; \arg\min_{\boldsymbol{\beta}}\Big\{-\ell_n(\boldsymbol{\beta}) \;+\; \sum_{j=1}^{p_n} P_\lambda\big(\beta_j\big)\Big\}, \tag{6}$$

where $P_\lambda(\beta_j)$ penalizes each β_j, shrinking weak effects toward zero and selecting the strong signals. The tuning parameter $\lambda > 0$ governs the size of the subset $\hat{\mathcal{S}}_1$.

Step 2 (detection of $\mathcal{S}_2$). To identify $\hat{\mathcal{S}}_2$, first solve a penalized regression problem with a ridge penalty only on the variables in $\hat{\mathcal{S}}_1^c$. Formally,

$$\hat{\boldsymbol{\beta}}^{r} \;=\; \arg\min_{\boldsymbol{\beta}}\Big\{-\ell(\boldsymbol{\beta}) \;+\; r_n\,\big\|\boldsymbol{\beta}_{\hat{\mathcal{S}}_1^c}\big\|_2^2\Big\}, \tag{7}$$

where $r_n > 0$ is a tuning parameter controlling the overall strength of regularization for variables in $\hat{\mathcal{S}}_1^c$. We then define a post-selection weighted ridge (WR) estimator $\hat{\boldsymbol{\beta}}^{\mathrm{WR}}$ by

$$\hat{\beta}_j^{\mathrm{WR}} \;=\; \begin{cases} \hat{\beta}_j^{r}, & j \in \hat{\mathcal{S}}_1, \\ \hat{\beta}_j^{r}\, I\big(|\hat{\beta}_j^{r}| \;>\; a_n\big), & j \in \hat{\mathcal{S}}_1^c, \end{cases} \tag{8}$$

where a_n is a thresholding parameter. The set $\hat{\mathcal{S}}_2$ is then

$$\hat{\mathcal{S}}_2 \;=\; \Big\{j \in \hat{\mathcal{S}}_1^c \,:\, \hat{\beta}_j^{\mathrm{WR}} \neq 0,\; 1 \le j \le p\Big\}. \tag{9}$$

We apply this post-selection procedure only if $\big|\hat{\mathcal{S}}_2\big| \;>\; 2$. In particular, we set

$$a_n \;=\; c\,n^{-\kappa}, \qquad 0 < \kappa \;\le\; \tfrac{1}{2}. \tag{10}$$

2.4.2. Post-Selection Shrinkage Estimation

We now propose a shrinkage estimator that combines information from two post-selection estimators, $\hat{\boldsymbol{\beta}}^{\mathrm{RE}}$ and $\hat{\boldsymbol{\beta}}^{\mathrm{WR}}$. Recall that

$$\hat{\boldsymbol{\beta}}^{\mathrm{WR}}_{\hat{\mathcal{S}}_1} = (\hat{\beta}^r_j, j \in \hat{\mathcal{S}}_1)^{\mathsf{T}}, \quad \text{and} \quad \hat{\boldsymbol{\beta}}^{\mathrm{WR}}_{\hat{\mathcal{S}}_2} = \left(\hat{\beta}^r_j \, I(|\hat{\beta}^r_j| > a_n), j \in \hat{\mathcal{S}}_2\right)^{\mathsf{T}}.$$

Define the post-selection shrinkage estimator for $\hat{\mathcal{S}}_1$ as

$$\hat{\boldsymbol{\beta}}^{\mathrm{SE}}_{\hat{\mathcal{S}}_1} = \hat{\boldsymbol{\beta}}^{\mathrm{WR}}_{\hat{\mathcal{S}}_1} - \left(\frac{\hat{s}_2 - 2}{\hat{T}_n}\right)\left(\hat{\boldsymbol{\beta}}^{\mathrm{WR}}_{\hat{\mathcal{S}}_1} - \hat{\boldsymbol{\beta}}^{\mathrm{RE}}_{\hat{\mathcal{S}}_1}\right), \tag{11}$$

where $\hat{s}_2 = |\hat{\mathcal{S}}_2|$, and $\hat{\boldsymbol{\beta}}^{\mathrm{RE}}_{\hat{\mathcal{S}}_1}$ is the restricted estimator obtained by maximizing the partial log-likelihood (3) over the set $\hat{\mathcal{S}}_1$. The term $\hat{T}_n$ is given by

$$\hat{T}_n = (\hat{\boldsymbol{\beta}}^{\mathrm{WR}}_{\hat{\mathcal{S}}_2})^{\mathsf{T}}\left(\boldsymbol{X}^{\mathsf{T}}_{\hat{\mathcal{S}}_2}\boldsymbol{M}_{\hat{\mathcal{S}}_1}\boldsymbol{X}_{\hat{\mathcal{S}}_2}\right)^{-1}\hat{\boldsymbol{\beta}}^{\mathrm{WR}}_{\hat{\mathcal{S}}_2}, \qquad \boldsymbol{M}_{\hat{\mathcal{S}}_1} = \boldsymbol{I}_n - \boldsymbol{X}_{\hat{\mathcal{S}}_1}\hat{\boldsymbol{\Sigma}}^{-1}_{\hat{\mathcal{S}}_1}\boldsymbol{X}^{\mathsf{T}}_{\hat{\mathcal{S}}_1}, \tag{12}$$

using a generalized inverse if $\hat{\boldsymbol{\Sigma}}_{\hat{\mathcal{S}}_1}$ is singular.

To avoid over-shrinking when $\hat{\boldsymbol{\beta}}^{\mathrm{WR}}_{\hat{\mathcal{S}}_1}$ and $\hat{\boldsymbol{\beta}}^{\mathrm{SE}}_{\hat{\mathcal{S}}_1}$ have different signs, we define a *positive* shrinkage estimator via the convex combination

$$\hat{\boldsymbol{\beta}}^{\mathrm{PSE}}_{\hat{\mathcal{S}}_1} = \hat{\boldsymbol{\beta}}^{\mathrm{WR}}_{\hat{\mathcal{S}}_1} - \left(\left(\frac{\hat{s}_2 - 2}{\hat{T}_n}\right) \wedge 1\right)\left(\hat{\boldsymbol{\beta}}^{\mathrm{WR}}_{\hat{\mathcal{S}}_1} - \hat{\boldsymbol{\beta}}^{\mathrm{RE}}_{\hat{\mathcal{S}}_1}\right). \tag{13}$$

This modification is essential to prevent an overly aggressive shrinkage that might reverse the sign of estimates in $\hat{\boldsymbol{\beta}}^{\mathrm{WR}}_{\hat{\mathcal{S}}_1}$.

3. Asymptotic Properties

In this section, we study the asymptotic properties of the the post-selection shrinkage estimators for the Cox regression model. To investigate the asymptotic theory, we need the following regularity conditions to be met.

(B1) $p = \exp(\mathcal{O}(n^{\alpha}))$ for some $0 < \alpha < 1$.

(B2) $\varrho_1 = \mathcal{O}(n^{-\eta})$, where $\tau < \eta \leq 1$ for τ in **(A2)**.

(B3) The existence of a positive definite matrix $\boldsymbol{\Sigma}_n$ such that $\lim_{n\to\infty}\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}$, where the eigenvalues of $\boldsymbol{\Sigma}$ satisfy $0 < \kappa_1 < \lambda_{\min}(\boldsymbol{\Sigma}) \leq \lambda_{\max}(\boldsymbol{\Sigma}) < \kappa_2 < \infty$.

(B4) Sparse Riesz condition: For the random design matrix $\boldsymbol{X}$, any $S \subset \mathcal{M}$ with $|S| = q, q \leq p$, and any vector $\mathbf{v} \in \mathbb{R}^q$, there exists $0 < c_* < c^* < \infty$ such that $c_* \leq ||\boldsymbol{X}^{\mathsf{T}}_S\mathbf{v}||^2_2/||\mathbf{v}||^2_2 \leq c^*$ holds with probability tending to 1.

The following theorems will make it easier to compute the asympotic distributional bias (ADB) and asympotic distributional risk (ADR) of the proposed estimators:

Theorem 1. *Suppose that assumptions (***A1***)–(***A3***) and* **(B1)–(B4)** *hold. If we choose* $r_n = c_2 a_n^{-2}(\log\log n)^3\log(n \vee p)$ *for some constant* $c_2 > 0$ *and* a_n *defined in (10) with* $\nu < (\eta - \alpha - \tau)/3$, *then,* $\hat{\mathcal{S}}_2$ *in (9) satisfies*

$$\lim_{n\to\infty} P(\hat{\mathcal{S}}_2 = \mathcal{S}_2 | \hat{\mathcal{S}}_1 = \mathcal{S}_1) = 1, \tag{14}$$

where τ, η, *and* α *are defined in (***A2***), (***B1***), and (***B2***), respectively.*

Theorem 2. *Let* $s_n^2 = \boldsymbol{d}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{d}_n$ *for any* $(p_1 + p_2) \times 1$ *vector* $\boldsymbol{d}_n$ *satisfying* $||\boldsymbol{d}_n||_2^2 \leq 1$. *Suppose assumptions* (**B1**)*–*(**B4**) *hold. Consider a sparse Cox model with a signal strength under* (**A1**)*–*(**A3**)*, and with* $0 < \tau < 1/2$. *Suppose a pre-selected model such as* $S_1 \subset \hat{S}_1 \subset S_1 \cup S_2$ *is obtained with probability 1. If we choose* r_n *in Theorem 1 with* $\nu < \{(\eta - \alpha - \tau)/, 1/4 - \tau/2\}$, *then, we have the asymptotic normality,*

$$n^{1/2} s_n^{-1} \boldsymbol{d}_n^\top (\hat{\boldsymbol{\beta}}^{WR}_{S^c_{null}} - \boldsymbol{\beta}_{S^c_{null}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \tag{15}$$

Asymptotic Distributional Bias and Risk Analysis

In order to compare the estimators, we use the asymptotic distributional bias (ADB) and the asymptotic risk (ADR) expressions of the proposed estimators.

Definition 1. *For any estimator* $\boldsymbol{\beta}_{1n}^\diamond$ *and* p_1*-dimensional vector* $\boldsymbol{d}_{1n}$, *satisfying* $||\boldsymbol{d}_{1n}||_2^2 \leq 1$, *the ADB and ADR of* $\boldsymbol{d}_{1n}^\top \boldsymbol{\beta}_{1n}^\diamond$, *respectively, are defined as*

$$ADB(\boldsymbol{d}_{1n}^\top \boldsymbol{\beta}_{1n}^\diamond) = \lim_{n\to\infty} E[\{n^{1/2} s_{1n}^{-1} \boldsymbol{d}_{1n}^\top (\boldsymbol{\beta}_{1n}^\diamond - \boldsymbol{\beta}_1)\}], \tag{16}$$

$$ADR(\boldsymbol{d}_{1n}^\top \boldsymbol{\beta}_{1n}^\diamond) = \lim_{n\to\infty} E[\{n^{1/2} s_{1n}^{-1} \boldsymbol{d}_{1n}^\top (\boldsymbol{\beta}_{1n}^\diamond - \boldsymbol{\beta}_1)\}^2], \tag{17}$$

where $s_{1n}^2 = \boldsymbol{d}_{1n}^\top \boldsymbol{\Sigma}^{-1}_{\mathcal{S}_1|\mathcal{S}_2} \boldsymbol{d}_{1n}$. *Let* $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_{p_2})^\top \in \mathbb{R}^{p_2}$ *and*

$$\boldsymbol{\Delta}_{\boldsymbol{d}_{1n}} = \frac{\boldsymbol{d}_{1n}^\top (\boldsymbol{\Sigma}_{S_1}^{-1} \boldsymbol{\Sigma}_{S_1 S_2} \boldsymbol{\delta} \boldsymbol{\delta}^\top \boldsymbol{\Sigma}_{S_2 S_1} \boldsymbol{\Sigma}_{S_1}^{-1}) \boldsymbol{d}_{1n}}{\boldsymbol{d}_{1n}^\top (\boldsymbol{\Sigma}_{S_1}^{-1} \boldsymbol{\Sigma}_{S_1 S_2} \boldsymbol{\Sigma}^{-1}_{S_2|S_1} \boldsymbol{\Sigma}_{\mathcal{S}_2 \mathcal{S}_1} \boldsymbol{\Sigma}_{\mathcal{S}_1}^{-1}) \boldsymbol{d}_{1n}}. \tag{18}$$

We have the following theorems on the expression of ADBs and ADRs of the post-selection estimators.

Theorem 3. *Let* $\boldsymbol{d}_{1n}$ *be any* p_1*-dimensional vector satisfying* $0 < ||\boldsymbol{d}_{1n}||_2^2 \leq 1$ *and* $s_{1n}^2 = \boldsymbol{d}_{1n}^\top \boldsymbol{\Sigma}^{-1}_{S_1|S_2} \boldsymbol{d}_{1n}$. *Under the assumptions* (**A1**)-(**A3**)*, we have*

$$ADB(\boldsymbol{d}_{1n}^\top \hat{\boldsymbol{\beta}}_{1n}^{WR}) = 0, \tag{19}$$

$$ADB(\boldsymbol{d}_{1n}^\top \hat{\boldsymbol{\beta}}_{1n}^{RE}) = s_1^{-1} \boldsymbol{d}_2^\top \boldsymbol{\beta}_2, \tag{20}$$

$$ADB(\boldsymbol{d}_{1n}^\top \hat{\boldsymbol{\beta}}_{1n}^{SE}) = (p_2 - 2) s_1^{-1} \boldsymbol{d}_2^\top \boldsymbol{\beta}_2^* E[\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})], \tag{21}$$

$$\begin{aligned} ADB(\boldsymbol{d}_{1n}^\top \hat{\boldsymbol{\beta}}_{1n}^{PSE}) = s_1^{-1} \boldsymbol{d}_2^\top \boldsymbol{\beta}_2^* \Bigg[& (p_2 - 2) \Big\{ E[\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})] + E\left[\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) I(\chi_{p_2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) < (p_2 - 2))\right] \Big\} \\ & - H_{p_2}\Big(p_2 - 2; \boldsymbol{\Delta}_{\boldsymbol{d}_2}\Big) \Bigg], \end{aligned} \tag{22}$$

where $\boldsymbol{d}_{2n} = \boldsymbol{\Sigma}_{S_2 S_1} \boldsymbol{\Sigma}_{\mathcal{S}_1}^{-1} \boldsymbol{d}_{1n}$ *and* $E[\chi_{p_2}^{-2j}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})] = \int_0^\infty x^{-2j} dH_{p_2}(x; \boldsymbol{\Delta}_{\boldsymbol{d}_2})$.

See the Appendix A for a detailed proof.

Theorem 4. *Under the assumptions of Theorem 2, except* (**A2**) *is replaced by* $\beta_j = \delta/\sqrt{n}$, *for* $j \in S_2$, *with* $|\delta_j| < \delta_{\max}$, *for some* $\delta_{\max} > 0$, *we have*

$$ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{WR}) = 1, \tag{23}$$

$$ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{RE}) = 1 + (1-c)^{1/2}[2 + (1-c)^{1/2}(1+2\boldsymbol{\Delta}_1)], \tag{24}$$

$$\begin{aligned} ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{SE}) &= 1 + (1-c)^{1/2}(p_2-2)\Big[(1-c)^{1/2}(p_2-2)\Big\{E[\chi_{p_2+2}^{-4}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})] \\ &+ (s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2)^2 E[\chi_{p_2}^{-4}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})]\Big\} + 2E[\chi_{p_2+2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})]\Big], \end{aligned} \tag{25}$$

$$\begin{aligned} ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{PSE}) &= 1 + (1-c)(p_2-2)^2\Bigg\{E[\chi_{p_2+2}^{-4}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})] + (s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2)^2 E[\chi_{p_2}^{-4}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})] \\ &+ E[\chi_{p_2+2}^{-4}(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) I(\chi_{p_2+2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) < (p_{2n}-2))]\Bigg\} \\ &+ 2(1-c)^{1/2}(p_2-2)\Bigg\{E[\chi_{p_2+2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})] + E[\chi_{p_2+2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) I(\chi_{p_2+2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) < (p_2-2))] \\ &- (p_2-2)E[\chi_{p_2+2}^{-4}(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) I(\chi_{p_2+2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) < (p_2-2))] - (1-c)^{1/2} \\ &\times \Bigg[E[\chi_{p_2+2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) I(\chi_{p_2+2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) < (p_2-2))] \\ &+ (s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2^*)^2 E[\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) I(\chi_{p_2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) < (p_2-2))]\Bigg]\Bigg\}, \\ &+ (1-c)^{1/2}\Bigg[(1-c)^{1/2}\Big(E[\chi_{p_2+2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2})] + (s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2^*)^2 H_{p_2}(p_2-2;\boldsymbol{\Delta}_{\boldsymbol{d}_2})\Big) \\ &+ 2\Big\{H_{p_2}(p_2-2;\boldsymbol{\Delta}_{\boldsymbol{d}_2}) - (p_2-2)E[\chi_{p_2+2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) I(\chi_{p_2+2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2}) < (p_2-2))]\Big\}\Bigg], \end{aligned} \tag{26}$$

where $c = \lim_{n\to\infty} \boldsymbol{d}_{1n}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathcal{S}_1}^{-1}\boldsymbol{d}_{1n}/(\boldsymbol{d}_{1n}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathcal{S}_{11.2}}^{-1}\boldsymbol{d}_{1n}) \leq 1$ *and* $s_{2n}^2 = \boldsymbol{d}_{2n}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathcal{S}_{22.1}}^{-1}\boldsymbol{d}_{2n}$.

It can be observed that the theoretical results are different from Theorem 3 of [19]. Ref. [19] considered the ADR of PSE estimations for the linear model. In contrast, our Theorems 3 and 4 are used for the PSE with the Cox proportional hazards model, which are feasible estimations. From Theorem 4, we can compare the ADRs of the estimators.

Corollary 1. *Under the assumptions in Theorem 4, we have*

1. *If* $||\delta||_2^2 \leq 1$, *then* $ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{PSE}) \leq ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{SE}) \leq ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{WR})$;
2. *If* $||\delta||_2^2 = o(1)$ *and* $p_2 \to \infty$, *then* $ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{RE}) < ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{PSE}) \leq ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{WR})$ *for* $\boldsymbol{\delta} = 0$.

Corollary 1 shows that the performance of the post-selection PSE is closely related to the RE. On the ond hand, if $\hat{s}_1 \subset S_1 \cup S_2$ and $(S_1 \cup S_2) \cap \hat{S}_1^c$ are large, then the post-selection PSE tends to dominate the RE. Further, if a variable selection method generates the right submodel and $||\delta||_2^2 = o(1)$, that is, $\lim_{n\to\infty} \hat{S}_1 = S_1 \cup S_2$, then, a post-selection likelihood estimator $\hat{\boldsymbol{\beta}}_{1n}^{RE}$ is the most efficient one compared with all other post-selection estimators.

Remark 1. *The simultaneous variable selection and parameter estimation may not lead to a good estimation strategy when weak signals co-exist with zero signals. Even though the selected candidate subset models can be provided by some existing variable selection techniques when* $p > n$, *the prediction performance can be improved by the post-selection shrinkage strategy, especially when an under-fitted subset model is selected by an aggressive variable selection procedure.*

4. Simulation Study

In this section, we present a simulation study designed to compare the quadratic risk performance of the proposed estimators under the Cox regression model. Each row of the design matrix $\boldsymbol{X}$ is generated i.i.d. from a $N(\mathbf{0}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\Sigma}$ follows an autoregressive covariance structure, as follows:

$$\boldsymbol{\Sigma}_{jj'} \;=\; 0.5^{|j-j'|}, \qquad 1 \le j, j' \le p.$$

In this setup, we consider the following true regression coefficients:

$$\boldsymbol{\beta} \;=\; \Big(\underbrace{8,\,9,\,10}_{\mathcal{S}_1},\; \underbrace{1,\,0.8,\,0.5,\,\underbrace{0.2,\ldots,0.2}_{p_2-p_1}}_{\mathcal{S}_2},\; \underbrace{0,\,0,\,0,\ldots,0}_{p-p_1-p_2}\Big)^{\mathsf{T}}, \tag{27}$$

where the subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ correspond to strong and weak signals, respectively. The true survival times $\boldsymbol{Y}$ are generated from an exponential distribution with parameter $\boldsymbol{X\beta}$. Censoring times are drawn from a Uniform$(0, c)$ distribution, where c is chosen to achieve the desired censoring rate. We consider censoring rates of 15% and 25%, and we explore sample sizes $n = 100, 300, 400$.

We compare the performance of our proposed estimators against two well-known penalized likelihood methods, namely, LASSO and Elastic Net (ENet). We employ the R package glmnet to fit these penalized methods and choose the tuning parameters via cross-validation. For each combination of n and p, we run 1000 Monte Carlo simulations. Let $\boldsymbol{\beta}_{1n}^{\diamond}$ denote either $\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{PSE}}$ or $\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{RE}}$ after variable selection. We assess the performance using the relative mean squared error (RMSE) with respect to $\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{WR}}$ as follows:

$$\mathrm{RMSE}\big(\boldsymbol{\beta}_{1n}^{\diamond}\big) \;=\; \frac{E\,\big\|\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{WR}} - \boldsymbol{\beta}\big\|_2^2}{E\,\big\|\boldsymbol{\beta}_{1n}^{\diamond} - \boldsymbol{\beta}\big\|_2^2}.$$

An $\mathrm{RMSE}(\boldsymbol{\beta}_{1n}^{\diamond}) > 1$ indicates that $\boldsymbol{\beta}_{1n}^{\diamond}$ outperforms $\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{WR}}$, and a larger RMSE signifies a stronger degree of superiority over $\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{WR}}$.

Table 1 presents the relative mean squared error (RMSE) values for different regression methods—LASSO and Elastic Net (ENet)—under varying sample sizes (n), number of predictors (p), and censoring percentages (15% and 25%). The RMSE values are averaged over 1000 simulation runs. The table compares three estimators, $\hat{\beta}_{S_1}^{PLE}$, $\hat{\beta}_{S_1}^{RE}$, and $\hat{\beta}_{S_1}^{PSE}$, providing insight into their performance under different settings.

Table 1. Simulated relative mean squared error (RMSE) across different values of p and n, averaged over $N = 1000$ simulation runs.

			Censoring Percentage					
			15%			**25%**		
n	p	**Method**	$\hat{\boldsymbol{\beta}}_{\widehat{\mathbf{S}}_1}^{\mathbf{PLE}}$	$\hat{\boldsymbol{\beta}}_{\widehat{\mathbf{S}}_1}^{\mathbf{RE}}$	$\hat{\boldsymbol{\beta}}_{\widehat{\mathbf{S}}_1}^{\mathbf{PSE}}$	$\hat{\boldsymbol{\beta}}_{\widehat{\mathbf{S}}_1}^{\mathbf{PLE}}$	$\hat{\boldsymbol{\beta}}_{\widehat{\mathbf{S}}_1}^{\mathbf{RE}}$	$\hat{\boldsymbol{\beta}}_{\widehat{\mathbf{S}}_1}^{\mathbf{PSE}}$
100	300	LASSO	1.04	1.66	1.19	1.08	1.96	1.23
		ENet	1.07	1.45	1.23	0.92	1.44	1.06
	400	LASSO	1.03	1.10	1.60	0.96	1.03	1.98
		ENet	0.90	1.00	1.45	0.89	0.98	1.36
	500	LASSO	1.08	1.13	1.66	0.98	1.05	1.37
		ENet	0.96	1.01	1.03	0.95	1.00	1.22

Table 1. *Cont.*

			Censoring Percentage					
			15%			25%		
n	p	Method	$\hat{\beta}^{\mathrm{PLE}}_{\widehat{S}_1}$	$\hat{\beta}^{\mathrm{RE}}_{\widehat{S}_1}$	$\hat{\beta}^{\mathrm{PSE}}_{\widehat{S}_1}$	$\hat{\beta}^{\mathrm{PLE}}_{\widehat{S}_1}$	$\hat{\beta}^{\mathrm{RE}}_{\widehat{S}_1}$	$\hat{\beta}^{\mathrm{PSE}}_{\widehat{S}_1}$
300	300	LASSO	0.85	1.60	0.98	0.92	1.64	1.06
		ENet	0.96	1.37	1.08	0.87	1.46	1.00
	350	LASSO	0.83	0.99	1.01	0.90	1.07	1.17
		ENet	0.85	0.99	1.52	0.87	1.02	1.56
	400	LASSO	0.90	1.03	1.25	0.81	0.95	1.73
		ENet	0.90	1.04	1.25	0.76	0.89	1.41
400	400	LASSO	0.99	1.52	1.12	0.82	1.50	0.94
		ENet	0.91	1.29	1.02	0.83	1.26	0.99
	450	LASSO	0.83	1.00	1.13	0.84	0.94	1.61
		ENet	0.92	1.05	1.46	0.85	0.99	1.79
	500	LASSO	0.89	0.93	1.83	0.81	0.93	1.90
		ENet	0.82	0.93	1.38	0.82	0.95	1.75

Figures 1 and 2 visualize the RMSE trends for different values of p when comparing LASSO (Figure 1) and ENet (Figure 2) against the proposed estimators (RE and PSE). The plots indicate how RMSE varies as p increases for different sample sizes (n) and censoring levels.

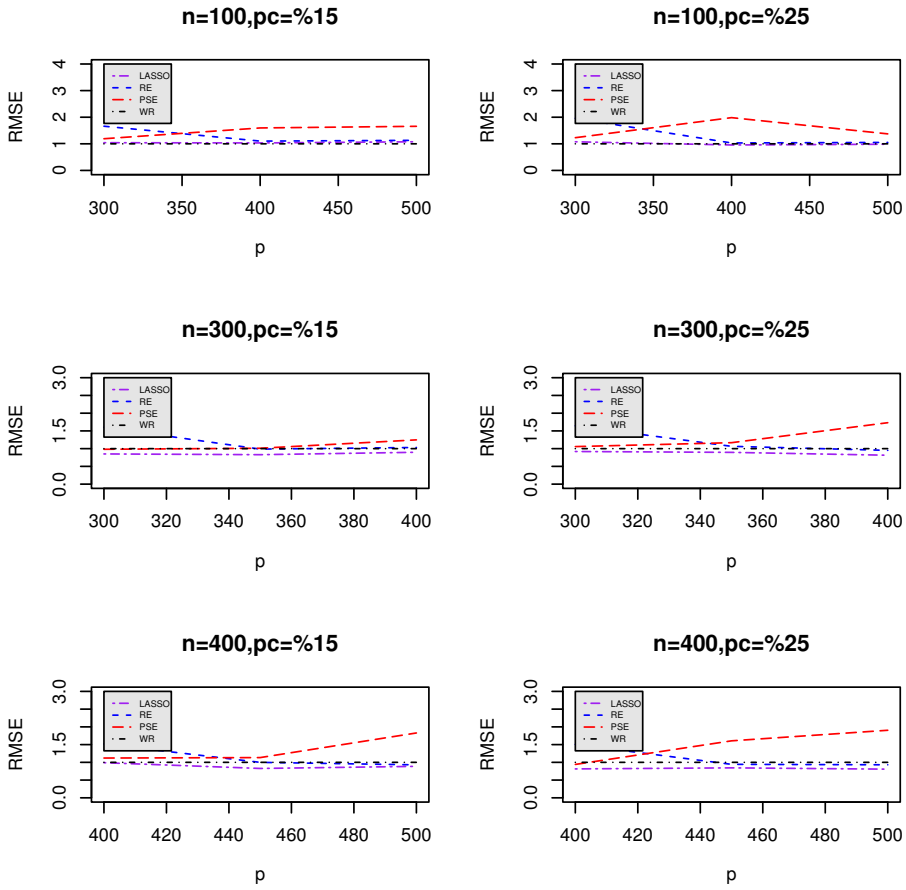

Figure 1. Relative mean squared error (RMSE) of the proposed estimators compared to LASSO for different n and p.

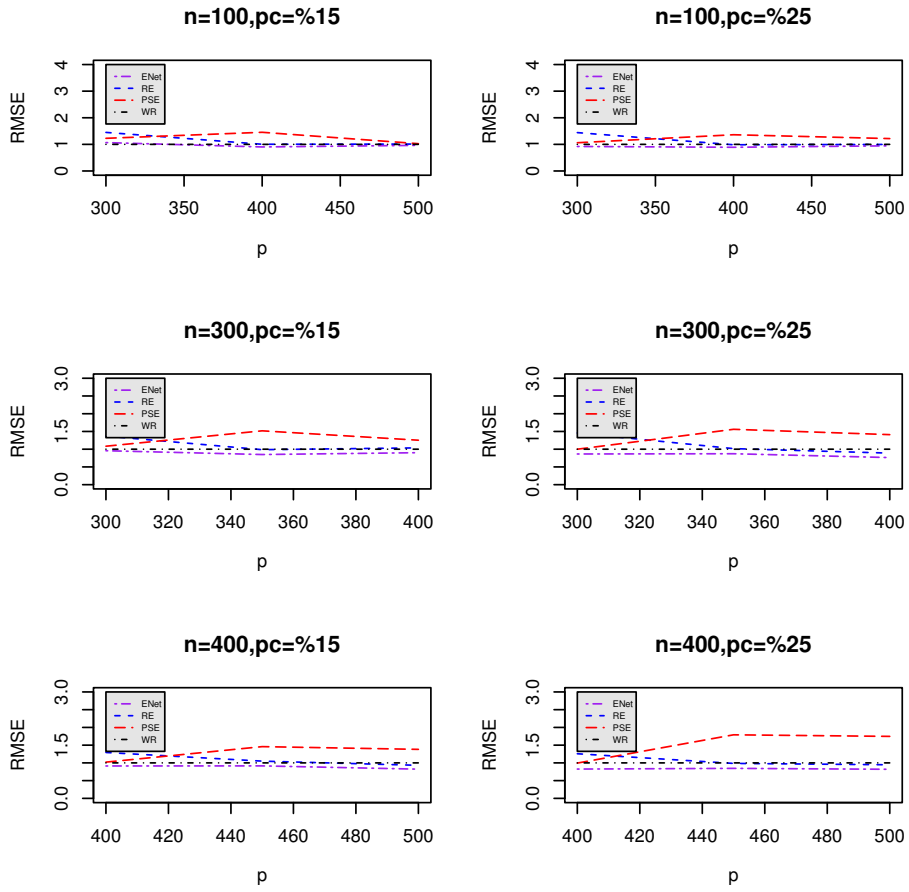

Figure 2. Relative mean squared error (RMSE) of the proposed estimators compared to Elastic Net for different n and p.

Key Observations and Insights

1. **Superior performance of post-selection estimators:** Across all combinations of n and p, the post-selection estimators ($\hat{\beta}_{S_1}^{RE}$ and $\hat{\beta}_{S_1}^{PSE}$) consistently demonstrate lower RMSEs compared to LASSO and ENet. This suggests that these estimators provide better predictive accuracy and stability.
2. **Impact of censoring percentage:**
   - When the censoring percentage increases from 15% to 25%, the RMSE values tend to increase across all methods, indicating the expected loss of predictive power due to increased censoring.
   - However, the post-selection estimators maintain a more stable RMSE trend, demonstrating their robustness in handling censored data.
3. **Effect of increasing predictors (p):**
   - As p increases, the RMSE for LASSO and ENet tends to rise, particularly under higher censoring rates.
   - This trend suggests that LASSO and ENet struggle with larger feature spaces, likely due to their tendency to aggressively shrink weaker covariates.
   - In contrast, the post-selection estimators show relatively stable RMSE behavior, indicating their ability to retain relevant information even in high-dimensional settings.
4. **Impact of sample size (n) on RMSE stability:**
   - Larger sample sizes (n) generally lead to lower RMSE values across all methods.

- However, the gap between LASSO/ENet and the post-selection estimators remains consistent, reinforcing the advantage of the proposed methods even with more data.

5. **Comparing LASSO and ENet:**
   - ENet generally has lower RMSE values than LASSO, particularly for small sample sizes, indicating its advantage in balancing feature selection and regularization.
   - However, ENet still underperforms compared to post-selection estimators, suggesting that the additional shrinkage adjustments help mitigate underfitting issues.

To further compare the sparsity of the coefficient estimators, we also measure the False Positive Rate (FPR), as follows:

$$\mathrm{FPR}(\hat{\boldsymbol{\beta}}) = \frac{\left|\{j : \hat{\beta}_j \neq 0 \wedge \beta_j = 0\}\right|}{\left|\{j : \beta_j = 0\}\right|}. \tag{28}$$

A higher FPR indicates that more non-informative variables are incorrectly included in the model, thereby complicating interpretation [22]. When $\boldsymbol{\beta}$ does not contain any zero components, the FPR is undefined. Table 2 compares the performance of LASSO and Elastic Net (ENet) in selecting variables in a high-dimensional Cox model under 15% and 25% censoring. As sample size (n) increases, both methods select more variables, but false positive rates (FPR) also rise, especially for ENet. LASSO is more conservative, selecting fewer variables with a lower FPR, while ENet selects more but at the cost of higher false discoveries. Higher censoring (25%) slightly increases FPR, reducing selection accuracy. Overall, LASSO offers better false positive control, whereas ENet captures more variables but with increased risk of selecting irrelevant ones.

Table 2. Average number of selected predictors ($\hat{S}_1$) and false positive rate (FPR) across different values of n and p, averaged over $N = 1000$ simulation runs.

			Censoring Percentage			
			%15		**%25**	
n	***p***	**Method**	**Average $\hat{S}_1$**	**FPR**	**Average $\hat{S}_1$**	**FPR**
100	300	LASSO	6.1	0.063	6.4	0.056
		ENet	6.2	0.063	6.6	0.052
	400	LASSO	4.9	0.072	5.2	0.085
		ENet	5.1	0.072	4.8	0.075
	500	LASSO	5.6	0.039	12.6	0.043
		ENet	4.9	0.039	4.0	0.033
300	300	LASSO	13.4	0.209	13.8	0.223
		ENet	12.9	0.209	16.3	0.282
	350	LASSO	15.6	0.202	15.8	0.208
		ENet	15.7	0.202	22.6	0.279
	400	LASSO	14.5	0.137	13.7	0.155
		ENet	13.5	0.137	14.2	0.173
400	400	LASSO	14.1	0.163	15.8	0.171
		ENet	14.2	0.163	20.4	0.212
	450	LASSO	18.4	0.217	23.5	0.24
		ENet	19.1	0.217	30.1	0.263
	500	LASSO	13.6	0.150	13.3	0.158
		ENet	13.3	0.150	13.6	0.158

5. Real Data Example

In this section, we illustrate the practical utility of our proposed methodology on two different high-dimensional datasets.

5.1. Example 1

We first apply our method to a gene expression dataset comprising $n = 614$ breast cancer patients, each with $p = 1490$ genes. All patients received anthracycline-based chemotherapy. Among these 614 individuals, there were 134 (21%) censored observations, and the mean time to treatment response was approximately 2.98 years. Using biological pathways to identify important genes, Ref. [23] previously selected 29 genes and reported a maximum area under the receiver operating characteristic curve (AUC) of about 62%. This relatively low AUC suggests limited predictive power when only using these 29 genes.

To improve upon these findings, we begin by performing an initial noise-reduction procedure on the data. This step helps remove potential outliers and irrelevant features, thereby enhancing the quality of the subsequent variable selection and estimation processes. We applied LASSO and Elastic Net (ENet) for gene selection. The results show that LASSO selected 14 genes, whereas ENet selected 12 genes. We then applied the proposed post-selection shrinkage estimators introduced in Section 2 to evaluate their performance compared to standard methods such as LASSO and Elastic Net. Table 3 shows the estimated coefficients from different estimators, along with the AUC at the bottom. It is evident that the PSE estimate has slightly improved the prediction performance.

Table 3. Estimated coefficients using the LASSO and ENet method for example 1.

Gen ID	LASSO			ENet		
	$\hat{\beta}^{\text{LASSO}}_{\hat{S}_1}$	$\hat{\beta}^{\text{RE}}_{\hat{S}_1}$	$\hat{\beta}^{\text{PSE}}_{\hat{S}_1}$	$\hat{\beta}^{\text{ENet}}_{\hat{S}_1}$	$\hat{\beta}^{\text{RE}}_{\hat{S}_1}$	$\hat{\beta}^{\text{PSE}}_{\hat{S}_1}$
18	−0.02	0.26	0.21	−0.03	0.07	0.20
97	0.01	0.27	0.00	0.01	0.26	0.01
101	0.05	0.19	0.13	0.05	0.27	0.12
128	–	–	–	−0.01	–	–
232	0.04	−0.42	−0.28	0.04	0.20	−0.25
342	0.15	−0.42	−0.13	0.14	−0.39	−0.10
369	−0.09	−0.05	0.04	−0.08	−0.40	−0.12
408	−0.01	–	–	−0.01	−0.09	0.03
410	0.03	−0.26	−0.15	0.03	−0.06	−0.14
445	–	–	–	−0.00	–	–
468	0.14	0.08	0.02	0.13	−0.26	−0.01
660	−0.00	–	–	−0.00	–	–
731	−0.08	0.09	0.06	−0.08	0.06	0.06
810	−0.04	−0.08	−0.09	0.01	0.09	−0.09
907	–	–	–	0.01	–	–
934	−0.00	–	–	−0.00	–	–
952	–	–	–	−0.01	–	–
961	−0.05	—	–	−0.05	−0.08	0.20
1212	–	–	–	−0.00	–	–
AUC	0.62	0.63	0.65	0.63	0.64	0.66

5.2. Example 2

We now consider the diffuse large B-cell lymphoma (DLBCL) dataset of [24], which is also high-dimensional. This dataset was used as a primary example to illustrate the effectiveness of our proposed dimension-reduction method. It consists of measurements on 7399 genes obtained from 240 patients via customized cDNA microarrays (lymphochip).

Each patient's survival time was recorded, ranging from 0 to 21.8 years; 127 patients had died (uncensored) and 95 were alive (censored) at the end of the study. Additional details on the dataset can be found in [24].

To obtain the post-selection shrinkage estimators, we first selected candidate subsets using two variable selection approaches—LASSO and Elastic Net (ENet). All tuning parameters were chosen via 10-fold cross-validation. Table 4 shows the estimated coefficients from both LASSO and ENet for the setting $p = 6800$. The AUC results indicate that $\hat{\boldsymbol{\beta}}^{\text{PSE}}_{\hat{S}_1}$ generally outperforms $\hat{\boldsymbol{\beta}}^{\text{RE}}_{\hat{S}_1}$ and $\hat{\boldsymbol{\beta}}^{\text{PLE}}_{\hat{S}_1}$ for both LASSO and ENet procedures. Notably, the ENet-based estimators appear more robust than those obtained via LASSO, underscoring the value of combining L_1 and L_2 penalties in high-dimensional survival analysis.

Table 4. Estimated coefficients using the LASSO and ENet method for example 2.

Gen ID	LASSO			ENet		
	$\hat{\boldsymbol{\beta}}^{\text{LASSO}}_{\hat{S}_1}$	$\hat{\boldsymbol{\beta}}^{\text{RE}}_{\hat{S}_1}$	$\hat{\boldsymbol{\beta}}^{\text{PSE}}_{\hat{S}_1}$	$\hat{\boldsymbol{\beta}}^{\text{ENet}}_{\hat{S}_1}$	$\hat{\boldsymbol{\beta}}^{\text{RE}}_{\hat{S}_1}$	$\hat{\boldsymbol{\beta}}^{\text{PSE}}_{\hat{S}_1}$
95	0.02	–	–	−0.34	–	–
112	0.06	0.71	0.70	−0.00	−0.13	−0.08
173	−0.63	–	–	0.68	–	–
205	–	–	–	–	–	–
551	1.60	1.69	1.57	−0.11	−0.28	−0.20
1377	−0.22	−0.84	−0.80	−0.09	−0.16	−0.12
1526	0.41	0.67	0.56	0.02	–	–
1543	−0.43	−0.79	−0.77	0.40	0.75	0.69
2003	−0.11	–	–	1.10	–	–
2025	0.18	0.90	0.78	1.04	1.22	1.07
2439	–	–	–	−0.01	−0.14	−0.12
2705	−0.85	–	–	0.36	0.77	0.61
2973	0.59	1.23	0.99	−0.63	−1.12	−0.81
3240	1.13	–	–	0.03	–	–
3598	−0.22	−0.59	−0.54	0.29	0.55	0.49
3882	0.13	0.40	0.39	−0.06	−0.20	−0.15
4015	0.34	0.81	0.76	−0.08	−0.13	−0.12
4186	−0.50	−0.72	−0.53	–	–	–
4357	0.09	–	–	−0.59	−0.70	−0.65
4662	0.70	0.90	0.83	0.21	0.60	0.38
5131	0.54	0.80	0.71	0.01	0.01	0.01
5222	−0.15	−0.38	−0.26	1.24	1.67	1.34
5541	–	–	–	−0.52	−0.72	−0.68
5577	0.39	0.86	0.70	−0.73	−0.97	−0.80
5778	−0.62	–	–	−0.09	–	–
5808	–	–	–	0.35	0.55	0.46
5951	–	–	–	1.29	2.12	1.70
6103	–	–	–	−0.63	−0.80	−0.76
6254	–	–	–	0.25	0.56	0.48
6493	–	–	–	0.65	–	–
6510	–	–	–	0.86	1.09	0.99
AUC	0.71	0.71	0.73	0.72	0.72	0.74

6. Conclusions

In this paper, we proposed high-dimensional post-selection shrinkage estimators for Cox's proportional hazards models based on the work of [19]. We investigated the asymptotic risk properties of these estimators in relation to the risks of the subset candidate model, as well as the LASSO and ENet estimators. Our results indicate that the new estimators per-

form particularly well when the true model contains weak signals. The proposed strategy is also conceptually intuitive and computationally straightforward to implement.

Our theoretical analysis and simulation studies demonstrate that the post-selection shrinkage estimator exhibits superior performance relative to LASSO and ENet, in part because it mitigates the loss of efficiency often associated with variable selection. As a powerful tool for producing interpretable models, sparse modeling via penalized regularization has become increasingly popular for high-dimensional data analysis. Our post-selection shrinkage estimator preserves model interpretability while enhancing predictive accuracy compared to existing penalized regression techniques. Furthermore, two real-data examples illustrate the practical advantages of our method, confirming that its performance is robust and potentially valuable for a range of high-dimensional applications.

Author Contributions: Conceptualization, R.A.B. and S.E.A.; methodology, R.A.B., S.E.A. and A.A.H. All authors have read and agreed to the published version of the manuscript

Funding: This research was funded by the Natural Sciences and the Engineering Research Council of Canada (NSERC).

Data Availability Statement: All data that are used in this study is publicly available

Acknowledgments: The research is supported by the Natural Sciences and the Engineering Research Council of Canada (NSERC).

Conflicts of Interest: The authors declare no conflicts of interest.

Nomenclature

Symbol	Description
General Notation	
n	Sample size (number of observations)
p	Number of covariates (predictor variables)
$\mathbb{R}$	Set of real numbers
$\mathbb{P}$	Probability measure
$\mathbb{E}$	Expectation operator
$\mathbb{I}(\cdot)$	Indicator function
Regression and Estimators	
β	Regression coefficient vector
$\hat{\beta}$	Estimated regression coefficients
λ	Regularization parameter (for LASSO/ENet)
$\hat{S}_1$	Selected subset of variables
d_{1n}	p_1-dimensional vector in the selection model
β°_{1n}	Selected regression coefficient estimator
$\hat{\beta}_{WR}$	Weighted Ridge (WR) estimator
Survival Analysis Notation	
$\mathcal{L}(\beta)$	Cox proportional hazards likelihood function
$\mathcal{D}$	Dataset containing observations
X	Covariate matrix
Y	Response variable (time-to-event outcome)
$h(t)$	Hazard function at time t
$\hat{h}(t)$	Estimated hazard function
$\Lambda(t)$	Cumulative hazard function

Evaluation Metrics

RMSE	Root Mean Squared Error
FPR	False Positive Rate
AUC	Area Under the Curve (for classification models)

Methods and Models

LASSO	Least Absolute Shrinkage and Selection Operator
ENet	Elastic Net
Cox-PH	Cox Proportional Hazards Model
WR	Weighted Ridge estimator
PSE	Post-selection Shrinkage Estimator
RE	Restricted Estimator

Appendix A. Proofs

The technical proofs of Theorems 3 and 4 are included in this section.

Proof of Theorem 3. Here, we provide the proof of the ADB expressions of the proposed estimators. Based on Theorem 2, it is clear that

$$\lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)\right] = E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)\right\}\right] = E[\mathcal{Z}] = 0,$$

where $\mathcal{Z} \sim \mathcal{N}(0,1)$. Then,

$$\begin{aligned}
ADB(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{RE}) &= \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{RE}-\boldsymbol{\beta}_1\right)\right] \\
&= \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left\{(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1)-(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE})\right\}\right] \\
&= \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)\right] - \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\right] \\
&= ADB(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{WR}) - \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\right] \\
&= \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\right] \\
&= \lim_{n\to\infty} (s_{2n}/s_{1n}) E\left[n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\right] \\
&= (s_2/s_1)s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2 = s_1^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2
\end{aligned}$$

where $\boldsymbol{d}_{2n} = \Sigma_{\mathcal{S}_2\mathcal{S}_1}\Sigma_{\mathcal{S}_1}^{-1}\boldsymbol{d}_{1n}$, $\boldsymbol{d}_{1n}^{\mathsf{T}}(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}) = -\boldsymbol{d}_1^{\mathsf{T}}\Sigma_{\mathcal{S}_1}^{-1}\Sigma_{\mathcal{S}_2\mathcal{S}_1}\hat{\boldsymbol{\beta}}_{2n}^{WR} = -\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}$ and $s_{2n}^2 = \boldsymbol{d}_2^{\mathsf{T}}\Sigma_{\mathcal{S}_2|\mathcal{S}_1}^{-1}\boldsymbol{d}_{2n}$.

Now, we compute the ADB of $\hat{\boldsymbol{\beta}}_{1n}^{SE}$ as follows

$$\begin{aligned}
ADB(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{SE}) &= \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{SE}-\boldsymbol{\beta}_1\right)\right] \\
&= \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-[(p_{2n}-2)\hat{T}_n^{-1}](\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE})-\boldsymbol{\beta}_1\right)\right] \\
&= \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)\right] \\
&\quad - \lim_{n\to\infty}(p_{2n}-2)E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1}^{RE}\right)\hat{T}_n^{-1}\right] \\
&= E[\mathcal{Z}]-(p_2-2)E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE})T_n^{-1}\right\}\right] \\
&= (p_2-2)E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}T_n^{-1}\right\}\right] \\
&= (p_2-2)(s_2/s_1)E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}T_n^{-1}\right\}\right] \\
&= (p_2-2)s_1^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2 E\left[\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})\right].
\end{aligned}$$

Finally, we obtain the ADB of $\hat{\boldsymbol{\beta}}_{1n}^{PSE}$,

$$\begin{aligned}
ADB(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{PSE}) &= \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{PSE}-\boldsymbol{\beta}_1\right)\right] \\
&= \lim_{n\to\infty} E\Bigg[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\Big\{\hat{\boldsymbol{\beta}}_{1n}^{SE}+[1-(p_{2n}-2)\hat{T}_n^{-1}](\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}) \\
&\quad \times I(\hat{T}_n<(p_{2n}-2))-\boldsymbol{\beta}_1\Big\}\Bigg] \\
&= \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{SE}-\boldsymbol{\beta}_1\right)\right] \\
&\quad + \lim_{n\to\infty} E\left[n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}[1-(p_{2n}-2)\hat{T}_n^{-1}]\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)I\left(\hat{T}_n<(p_{2n}-2)\right)\right] \\
&\quad + E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}(\hat{\boldsymbol{\beta}}_{1n}^{SE}-\boldsymbol{\beta}_1)\right\}\right]
\end{aligned}$$

$$
\begin{aligned}
&+E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE})I(\hat{T}_n<(p_{2n}-2))\right\}\right]\\
&-(p_2-2)E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE})T_n^{-1}I(\hat{T}_n<(p_{2n}-2))\right\}\right]\\
&=ADB(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{SE})-E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}I(\hat{T}_n<(p_{2n}-2))\right\}\right]\\
&+(p_2-2)E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\hat{T}_n^{-1}I(\hat{T}_n<(p_{2n}-2))\right\}\right]\\
&=ADB(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{SE})-(s_2/s_1)E\left[\mathcal{Z}I\big(\chi^2_{P_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\big)\right]\\
&-s_1^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2H_{p_2}(p_2-2;\boldsymbol{\Delta}_{\boldsymbol{d}_2})\\
&+(p_2-2)(s_2/s_1)E\left[\mathcal{Z}\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I\big(\chi^2_{P_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\big)\right]\\
&+(p_2-2)s_1^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2E\left[\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I\big(\chi^2_{P_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\big)\right]\\
&=ADB(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{SE})-s_1^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2H_{p_2}(p_2-2;\boldsymbol{\Delta}_{\boldsymbol{d}_2})\\
&+(p_2-2)s_1^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2E\left[\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I\big(\chi^2_{p_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\big)\right]\\
&=s_1^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2\Bigg[(p_2-2)\Big\{E\big[\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})\big]+E\big[\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I\big(\chi^2_{P_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\big)\big]\Big\}\\
&-H_{p_2}(p_2-2;\boldsymbol{\Delta}_{\boldsymbol{d}_2})\Bigg].
\end{aligned}
$$

□

Proof of Theorem 4. We provide the proof of the ADR expressions of the proposed estimators. It is clear that

$$
\lim_{n\to\infty}E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)\right\}^2\right]=E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)\right\}^2\right]=E[\mathcal{Z}^2]=1,
$$

where $\mathcal{Z}\sim\mathcal{N}(0,1)$. Then,

$$
\begin{aligned}
ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{RE})&=\lim_{n\to\infty}E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{RE}-\boldsymbol{\beta}_1\right)\right\}^2\right]\\
&=\lim_{n\to\infty}s_{1n}^{-2}E\left[\left\{n^{1/2}\boldsymbol{d}_{1n}^{\mathsf{T}}\left[\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)-\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\right]\right\}^2\right]\\
&=\lim_{n\to\infty}E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)\right\}^2\right]+\lim_{n\to\infty}E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\right\}^2\right]\\
&-2\lim_{n\to\infty}E\left[ns_{1n}^{-2}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\Big)^{\mathsf{T}}\boldsymbol{d}_{1n}\right]\\
&=I_1+I_2+I_3.
\end{aligned}
$$

From (23), we have $I_1=\lim_{n\to\infty}E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1)\right\}^2\right]=1$. Furthermore,

$$I_2 = \lim_{n\to\infty} s_{1n}^{-2} E\left[n^{1/2}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\right]^2$$
$$= \lim_{n\to\infty} (s_{2n}^2/s_{1n}^2) E\left[n^{1/2} s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\right]^2.$$

Since $s_{2n}^2/s_{1n}^2 \to 1-c$, then,

$$I_2 = (1-c)\lim_{n\to\infty} E\left[\chi_1^2(\boldsymbol{\Delta}_{d_{2n}})\right] = (1-c)(1+2\boldsymbol{\Delta}_{d_2}).$$

Furthermore,

$$I_3 = -2\lim_{n\to\infty} E\left[n s_{1n}^{-2}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \boldsymbol{\beta}_1\right)^{\mathsf{T}}\boldsymbol{d}_{1n}\right]$$
$$= 2\lim_{n\to\infty} (s_{2n}/s_{1n}) E\left[n^{1/2} s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR} n^{1/2} s_{1n}^{-1}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \boldsymbol{\beta}_1\right)^{\mathsf{T}}\boldsymbol{d}_{1n}\right]$$
$$= 2(1-c)^{1/2}.$$

Now, we investigate (25). By using Equation (17), we have

$$ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{SE}) = \lim_{n\to\infty} E\left[\left\{n^{1/2} s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{SE} - \boldsymbol{\beta}_1\right)\right\}^2\right]$$
$$= \lim_{n\to\infty} E\left[n^{1/2} s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left\{\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \boldsymbol{\beta}_1\right) - [(p_{2n}-2)/\hat{T}_n]\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\right\}\right]^2$$
$$= \lim_{n\to\infty} E\left[n^{1/2} s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \boldsymbol{\beta}_1\right)\right]^2$$
$$+ \lim_{n\to\infty} E\left[n^{1/2} s_{1n}^{-1}[(p_{2n}-2)T_n^{-1}]\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\right]^2$$
$$- 2\lim_{n\to\infty} E\left[n s_{1n}^{-2}[(p_{2n}-2)\hat{T}_n^{-1}]\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \boldsymbol{\beta}_1\right)^{\mathsf{T}}\boldsymbol{d}_{1n}\right]$$
$$= J_1 + J_2 + J_3.$$

Again, $J_1 = \lim_{n\to\infty} E\left[\left\{n^{1/2} s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \boldsymbol{\beta}_1\right)\right\}^2\right] = 1$. Then, we have

$$J_2 = \lim_{n\to\infty} E\left[n^{1/2} s_{1n}^{-1}[(p_{2n}-2)\hat{T}_n^{-1}]\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR} - \hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\right]^2$$
$$= \lim_{n\to\infty} (p_{2n}-1)^2 E\left[n^{1/2} s_{1n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR} T_n^{-1}\right]^2$$
$$= (s_2^2/s_1^2)(p_2-1)^2 E\left[\lim_{n\to\infty}\left\{n^{1/2} s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\hat{T}_n^{-1}\right\}\right]^2$$
$$= (s_2^2/s_1^2)(p_2-1)^2\left\{E[\mathcal{Z}^2\chi_{p_2}^{-4}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})] + (s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2)^2 E[\chi_{p_2}^{-4}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})]\right\}$$
$$= (1-c)(p_2-2)^2\left\{E[\chi_{p_2+2}^{-4}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})] + (s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2)^2 E[\chi_{p_2}^{-4}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})]\right\},$$

and

$$\begin{aligned}
J_3 &= -2\lim_{n\to\infty} E\left[ns_{1n}^{-2}\left[(p_{2n}-2)\hat{T}_n^{-1}\right]\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)^{\mathsf{T}}\boldsymbol{d}_{1n}\right] \\
&= 2\lim_{n\to\infty}(s_{2n}/s_{1n})(p_{2n}-2)E\left[n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}s_{1n}^{-1}(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1)^{\mathsf{T}}\hat{T}_n^{-1}\right] \\
&= 2(1-c)^{1/2}(p_2-2)\left\{E\left[\mathcal{Z}^2\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})\right]+s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2 E\left[\mathcal{Z}\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})\right]\right\} \\
&= 2(1-c)^{1/2}(p_2-2)E\left[\chi_{p_2+2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})\right].
\end{aligned}$$

Finally, we compute the ADR of $\hat{\boldsymbol{\beta}}_{1n}^{PSE}$ as follows:

$$\begin{aligned}
ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{PSE}) &= \lim_{n\to\infty} E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{PSE}-\boldsymbol{\beta}_1\right)\right\}^2\right] \\
&= \lim_{n\to\infty} E\Bigg[\Big\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\Big[\left(\hat{\boldsymbol{\beta}}_{1n}^{SE}-\boldsymbol{\beta}_1\right) \\
&\quad +\left[1-(p_{2n}-2)\hat{T}_n^{-1}\right]\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)I(\hat{T}_n<(p_{2n}-2))\Big]\Big\}^2\Bigg] \\
&= \lim_{n\to\infty} E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{SE}-\boldsymbol{\beta}_1\right)\right\}^2\right] \\
&+ \lim_{n\to\infty} E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left[1-(p_{2n}-2)\hat{T}_n^{-1}\right]\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)I(\hat{T}_n<(p_{2n}-2))\right\}^2\right] \\
&+ 2\lim_{n\to\infty} E\Bigg[n^{1/2}s_{1n}^{-2}\boldsymbol{d}_{1n}^{\mathsf{T}}\left[1-(p_{2n}-2)\hat{T}_n^{-1}\right]\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\left(\hat{\boldsymbol{\beta}}_{1n}^{SE}-\boldsymbol{\beta}_1\right)^{\mathsf{T}} \\
&\quad \times I(\hat{T}_n<(p_{2n}-2))\boldsymbol{d}_{1n}\Bigg] \\
&= ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{SE}) + \lim_{n\to\infty} E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)I(\hat{T}_n<(p_{2n}-2))\right\}^2\right] \\
&+ \lim_{n\to\infty}(p_{2n}-2)^2E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\hat{T}_n^{-1}I(\hat{T}_n<(p_{2n}-2))\right\}^2\right] \\
&- 2\lim_{n\to\infty}(p_{2n}-2)E\left[\left\{ns_{1n}^{-2}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)^2\hat{T}_n^{-1}I(\hat{T}_n<(p_{2n}-2))\boldsymbol{d}_{1n}\right\}\right] \\
&+ 2\lim_{n\to\infty} E\left[\left\{ns_{1n}^{-2}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\left(\hat{\boldsymbol{\beta}}_{1n}^{SE}-\boldsymbol{\beta}_1\right)^{\mathsf{T}}I(\hat{T}_n<(p_{2n}-2))\boldsymbol{d}_{1n}\right\}\right] \\
&- 2\lim_{n\to\infty}(p_{2n}-2)E\Bigg[\Big\{ns_{1n}^{-2}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\left(\hat{\boldsymbol{\beta}}_{1n}^{SE}-\boldsymbol{\beta}_1\right)^{\mathsf{T}} \\
&\quad \times\hat{T}_n^{-1}I(\hat{T}_n<(p_{2n}-2))\boldsymbol{d}_{1n}\Big\}\Bigg] \\
&= ADR(\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{1n}^{SE}) + K_1 + K_2 + K_3 + K_4 + K_5,
\end{aligned}$$

where

$$
\begin{aligned}
K_1 &= \lim_{n\to\infty} E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)I\left(\hat{T}_n<(p_{2n}-2)\right)\right\}^2\right]\\
&= \lim_{n\to\infty} E\left[\left\{n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}I\left(\hat{T}_n<(p_{2n}-2)\right)\right\}^2\right]\\
&= \lim_{n\to\infty}(s_{2n}/s_{1n})^2E\left[\left\{n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}I\left(\hat{T}_n<(p_{2n}-2)\right)\right\}^2\right]\\
&= (s_2/s_1)^2\left\{E\left[\mathcal{Z}^2I\left(\chi^2_{p_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\right)\right]+(s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2)^2H_{p_2}(p_2-2;\boldsymbol{\Delta}_{\boldsymbol{d}_2})\right\}\\
&= (1-c)\left\{E\left[\chi^2_{p_2+2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})\right]+(s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2)^2H_{p_2}(p_2-2;\boldsymbol{\Delta}_{\boldsymbol{d}_2})\right\},
\end{aligned}
$$

$$
\begin{aligned}
K_2 &= \lim_{n\to\infty} E\left[\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)(p_{2n}-2)\hat{T}_n^{-1}I\left(\hat{T}_n<(p_{2n}-2)\right)\right\}^2\right]\\
&= \lim_{n\to\infty}(p_{2n}-2)^2(s_{1n}/s_{2n})^2E\left[\left\{n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\hat{T}_n^{-1}I\left(\hat{T}_n<(p_{2n}-2)\right)\right\}^2\right]\\
&= (p_2-2)^2(s_2/s_1)^2E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\hat{T}_n^{-1}I\left(\hat{T}_n<(p_{2n}-2)\right)\right\}^2\right]\\
&= (p_2-2)^2(1-c)E\left[\mathcal{Z}^2\chi^{-4}_{p_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I\left(\chi^2_{p_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\right)\right]\\
&= (p_2-2)^2(1-c)E\left[\chi^{-4}_{p_2+2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I\left(\chi^2_{p_2+2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
K_3 &= -2\lim_{n\to\infty}(p_{2n}-2)E\left[\left\{ns_{1n}^{-2}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)^2\hat{T}_n^{-1}I\left(\hat{T}_n<(p_{2n}-2)\right)\boldsymbol{d}_{1n}\right\}\right]\\
&= -2(p_2-2)(s_2/s_1)^2E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\right\}^2\hat{T}_n^{-1}I\left(\hat{T}_n<(p_{2n}-2)\right)\right]\\
&= -2(p_2-2)(1-c)\left\{E\left[\mathcal{Z}^2\chi^{-2}_{p_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I\left(\chi^2_{p_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\right)\right]\right.\\
&\left.+(s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2)^2E\left[\chi^{-2}_{p_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I\left(\chi^2_{p_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\right)\right]\right\}\\
&= -2(p_2-2)(1-c)\left\{E\left[\chi^{-2}_{p_2+2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I\left(\chi^2_{p_2+2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\right)\right]\right.\\
&\left.+(s_2^{-1}\boldsymbol{d}_2^{\mathsf{T}}\boldsymbol{\beta}_2)^2E\left[\chi^{-2}_{p_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I\left(\chi^2_{p_2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2)\right)\right]\right\},
\end{aligned}
$$

$$
\begin{aligned}
K_4 &= 2\lim_{n\to\infty} E\left[\left\{ns_{1n}^{-2}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\left(\hat{\boldsymbol{\beta}}_{1n}^{SE}-\boldsymbol{\beta}_1\right)^{\mathsf{T}} I(\hat{T}_n<(p_{2n}-2))\boldsymbol{d}_{1n}\right\}\right] \\
&= 2\lim_{n\to\infty}(s_{2n}/s_{1n})E\left[\left\{n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\right\}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{SE}-\boldsymbol{\beta}_1\right)\right\}^{\mathsf{T}} I(\hat{T}_n<(p_{2n}-2))\right] \\
&= 2(s_2/s_1)E\Bigg[\left\{n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\right\}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)\right. \\
&\quad \left.-(p_{2n}-2)\hat{T}_n^{-1}I(\hat{T}_n<(p_{2n}-2))\right\}^{\mathsf{T}} I(\hat{T}_n<(p_{2n}-2))\Bigg] \\
&= 2(1-c)^{1/2}\Bigg\{E\left[\lim_{n\to\infty}\left\{n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\right\}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)I(\hat{T}_n<(p_{2n}-2))\right\}^{\mathsf{T}}\right] \\
&\quad -(p_2-2)E\left[\left\{n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\right\}\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)\hat{T}_n^{-1}I(\hat{T}_n<(p_{2n}-2))\right\}^{\mathsf{T}}\right]\Bigg\} \\
&= 2(1-c)^{1/2}\left\{E\left[\mathcal{Z}^2 I(\chi_{p_2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2))\right]-(p_2-2)E\left[\mathcal{Z}^2\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I(\chi_{p_2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2))\right]\right\} \\
&= 2(1-c)^{1/2}\left\{H_{p_2+2}(p_2-2;\boldsymbol{\Delta}_{\boldsymbol{d}_2})-(p_2-2)E\left[\chi_{p_2+2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I(\chi_{p_2+2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2))\right]\right\}
\end{aligned}
$$

and

$$
\begin{aligned}
K_5 &= -2\lim_{n\to\infty}(p_{2n}-2)E\left[\left\{ns_{1n}^{-2}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\left(\hat{\boldsymbol{\beta}}_{1n}^{SE}-\boldsymbol{\beta}_1\right)^{\mathsf{T}}\hat{T}_n^{-1}I(\hat{T}_n<(p_{2n}-2))\boldsymbol{d}_{1n}\right\}\right] \\
&= 2(p_2-2)(s_2/s_1)E\Bigg[\lim_{n\to\infty}\left\{n^{1/2}s_{2n}^{-1}\boldsymbol{d}_{2n}^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{2n}^{WR}\hat{T}_n^{-1}I(\hat{T}_n<(p_{2n}-2))\right\} \\
&\quad \times\left\{n^{1/2}s_{1n}^{-1}\boldsymbol{d}_{1n}^{\mathsf{T}}\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\boldsymbol{\beta}_1\right)-(p_{2n}-2)\left(\hat{\boldsymbol{\beta}}_{1n}^{WR}-\hat{\boldsymbol{\beta}}_{1n}^{RE}\right)\hat{T}_n^{-1}I(\hat{T}_n<(p_{2n}-2))\right\}^{\mathsf{T}}\Bigg] \\
&= 2(p_2-2)(s_2/s_1)\Bigg\{E\left[\mathcal{Z}^2\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I(\chi_{p_2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2))\right] \\
&\quad -(p_2-2)E\left[\mathcal{Z}^2\chi_{p_2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I(\chi_{p_2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2))\right]\Bigg\} \\
&= 2(p_2-2)(1-c)^{1/2}\Bigg\{E\left[\chi_{p_2+2}^{-2}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I(\chi_{p_2+2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2))\right] \\
&\quad -(p_2-2)E\left[\chi_{p_2+2}^{-4}(\boldsymbol{\Delta}_{\boldsymbol{d}_2})I(\chi_{p_2+2}^2(\boldsymbol{\Delta}_{\boldsymbol{d}_2})<(p_2-2))\right]\Bigg\}.
\end{aligned}
$$

□

References

1. Ahmed, S.E. *Big and Complex Data Analysis: Methodologies and Applications*; Springer: Cham, Switzerland, 2017.
2. Bradic, J.; Fan, J.; Jiang, J. Regularization for Cox's proportional hazards model with n_p-dimensionality. *Ann. Stat.* **2011**, *39*, 3092–3120. [CrossRef] [PubMed]
3. Bradic, J.; Song, R. Structured estimation for the nonparametric Cox model. *Electron. J. Stat.* **2015**, *9*, 492–534. [CrossRef]
4. Gui, J.; Li, H. Penalized Cox regression analysis in the high-dimensional and low-sample size settings with applications to microarray gene expression data. *Bioinformatics* **2005**, *21*, 3001–3008. [CrossRef]
5. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]
6. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]

7. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [CrossRef]
8. Zou, H. The adaptive Lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [CrossRef]
9. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 301–320. [CrossRef]
10. Sun, T.; Zhang, C.H. Scaled sparse linear regression. *Biometrika* **2012**, *99*, 879–898. [CrossRef]
11. Tibshirani, R. The lasso method for variable selection in the Cox model. *Stat. Med.* **1997**, *16*, 385–395. [CrossRef]
12. Zhang, H.; Lu, W. Adaptive lasso for Cox's proportional hazards model. *Biometrika* **2007**, *94*, 691–703. [CrossRef]
13. Zou, H. A note on path-based variable selection in the penalized proportional hazards model. *Biometrika* **2008**, *95*, 241–247. [CrossRef]
14. Fan, J.; Li, R. Variable selection for Cox's proportional hazards model and frailty model. *Ann. Stat.* **2002**, *6*, 74–99. [CrossRef]
15. Hong, H.; Li, Y. Feature selection of ultrahigh-dimensional covariates with survival outcomes: A selective review. *Appl. Math. Ser. B* **2017**, *32*, 379–396. [CrossRef] [PubMed]
16. Hong, H.; Zheng, Q.; Li, Y. Forward regression for Cox models with high-dimensional covariates. *J. Multivar. Anal.* **2019**, *173*, 268–290. [CrossRef]
17. Hong, H.; Chen, X.; Kang, J.; Li, Y. The Lq-norm learning for ultrahigh-dimensional survival data: An integrative framework. *Stat. Sin.* **2020**, *30*, 1213–1233. [CrossRef]
18. Ahmed, S.E.; Ahmed, F.; Yüzbaşı, B. *Post-Shrinkage Strategies in Statistical and Machine Learning for High Dimensional Data*; Chapman and Hall/CRC: New York, NY, USA, 2023. [CrossRef]
19. Gao, X.; Ahmed, S.E.; Feng, Y. Post selection shrinkage estimation for high-dimensional data analysis. *Appl. Stoch. Models Bus. Ind.* **2017**, *33*, 97–120. [CrossRef]
20. Cox, D.R. Regression models and life-tables (with discussion). *J. R. Stat. Soc. Ser. B* **1972**, *34*, 187–220. [CrossRef]
21. Buhlmann, P.; van de Geer, S. *Statistics for High-Dimensional Data: Methods, Theory and Applications*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2011.
22. Kurnaz, S.F.; Hoffmann, I.; Filzmoser, P. Robust and sparse estimation methods for high-dimensional linear and logistic regression. *J. Chemom. Intell. Lab. Syst.* **2018**, *172*, 211–222. [CrossRef]
23. Belhechmi, S.; Bin, R.D.; Rotolo, F.; Michiels, S. Accounting for grouped predictor variables or pathways in high-dimensional penalized Cox regression models. *BMC Bioinform.* **2020**, *21*, 277. [CrossRef] [PubMed]
24. Rosenwald, A.; Wright, G.; Chan, W.C.; Connors, J.M.; Campo, E.; Fisher, R.I.; Gascoyne, R.D.; Muller-Hermelink, H.K.; Smel, E.B.; Giltnane, J.M.; et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **2002**, *25*, 1937–1947. [CrossRef] [PubMed]