REVSTAT – Statistical Journal
Volume 15, Number 2, April 2017, 251-276

IMPROVED PENALTY STRATEGIES in LINEAR REGRESSION MODELS

Authors: Bahadir Yüzbaşi
– Department of Econometrics,
Inonu University, Turkey
b.yzb@hotmail.com

S. Ejaz Ahmed
– Department of Mathematics and Statistics,
Brock University, Canada
sahmed5@brocku.ca

Mehmet Güngör
– Department of Econometrics,
Inonu University, Turkey
mgungor44@gmail.com

Received: March 2016 Revised: January 2017 Accepted: January 2017

Abstract:

- We suggest pretest and shrinkage ridge estimation strategies for linear regression models. We investigate the asymptotic properties of suggested estimators. Further, a Monte Carlo simulation study is conducted to assess the relative performance of the listed estimators. Also, we numerically compare their performance with Lasso, adaptive Lasso and SCAD strategies. Finally, a real data example is presented to illustrate the usefulness of the suggested methods.

Key-Words:

- *Sub-model; Full Model; Pretest and Shrinkage Estimation; Multicollinearity; Asymptotic and Simulation.*

AMS Subject Classification:

- 62J05, 62J07.

1. INTRODUCTION

Consider a linear regression model

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1.1}$$

where y_i's are random responses, $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^\top$ are known vectors, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top$ is a vector denoting unknown coefficients, ε_i's are unobservable random errors and the superscript $(^\top)$ denotes the transpose of a vector or matrix. Further, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top$ has a cumulative distribution function $\mathcal{F}(\boldsymbol{\varepsilon})$; $\mathcal{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathcal{V}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$, where σ^2 is finite and $\boldsymbol{I}_n$ is an identity matrix of dimension $n \times n$. In this paper, we consider that the design matrix has rank p $(p \leq n)$.

It is usually assumed that the explanatory variables are independent of each other in a multiple linear regression model. However, this assumption may not be valid in real life, that is, the independent variables in model may be correlated which cause to multicollinearity problem. In literature, some biased estimations, such as shrinkage estimation, principal components estimation (PCE), ridge estimation, partial least squares (PLS) estimation and Liu-type estimators were proposed to combat this problem. The ridge estimation is proposed by Hoerl and Kennard (1970), and is one of the most effective methods is the most popular one. This estimator has less mean squared error (MSE) than the least squares estimation (LSE) estimation.

The multiple linear regression model is used by data analysts in nearly every field of science and technology as well as economics, econometrics, finance. This model is also used to obtain information about unknown parameters based on sample information and, if available, other relevant information. The other information may be considered as non-sample information (NSI), see Ahmed (2001). This is also known as uncertain prior information (UPI). Such information, which is usually available from previous studies, expert knowledge or researcher's experience, is unrelated to the sample data. The NSI may or may not positively contribute to the estimation procedure. However, it may be advantageous to use the NSI in the estimation process when sample information may be rather limited and may not be completely reliable.

In this study, we consider a linear regression model (1.1) in a more realistic situation when the model is assumed to be sparse. Under this assumption, the vector of coefficients $\boldsymbol{\beta}$ can be partitioned as $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ where $\boldsymbol{\beta}_1$ is the coefficient vector for main effects, and $\boldsymbol{\beta}_2$ is the vector for nuisance effects or insignificant coefficients. We are essentially interested in the estimation of $\boldsymbol{\beta}_1$ when it is reasonable that $\boldsymbol{\beta}_2$ is close to zero. The full model estimation may be subject to high variability and may not be easily interpretable. On the other hand, a sub-model strategy may result with an under-fitted model with large bias. For this reason, we consider pretest and shrinkage strategy to control the magnitude of the bias. Ahmed (2001) gave a detailed definition of shrinkage estimation, and discussed large sample estimation techniques in a regression model. For more recent work on the subject, we refer to Ahmed et al. (2012), Ahmed and Fallahpour (2012), Ahmed (2014a), Ahmed (2014b), Hossain et al. (2016), Gao et al. (2016). Further, for some related work on shrinkage estimation we refer to Prakash and Singh (2009) and Shanubhogue and Al-Mosawi (2010), and others.

In this study, we also consider L_1 type estimators, and compare them with pretest and shrinkage estimators. Yüzbaşı and Ahmed (2015) provided some numerical comparisons of these estimators. The novel aspects of this manuscript, we investigate the asymptotic properties of pretest and shrinkage estimators when the number of observations is larger than the number of covariates.

The paper is organized as following. The full and sub-model estimators based on ridge regression are given in Section 2. The pretest, shrinkage estimators and penalized estimations are also presented in this section. The asymptotic properties of the pretest and shrinkage estimators are given in Section 3. The results of a Monte Carlo simulation study that include a comparison with some penalty estimators are given in Section 4. A real data example is given in Section 5. The concluding remarks are presented in Section 6.

2. ESTIMATION STRATEGIES

The ridge estimator can be obtained from the following model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \;\text{ subject to } \boldsymbol{\beta}^\top\boldsymbol{\beta} \leq \phi,$$

where ϕ is inversely proportional to k, $\mathbf{y} = (y_1, \ldots, y_n)^\top$ and $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$, which is equal to

$$\underset{\boldsymbol{\beta}}{\arg\min} \left\{ \sum\nolimits_{i=1}^{n} \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} \right)^2 + k \sum\nolimits_{j=1}^{p} \beta_j^2 \right\}.$$

It yields

$$\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}} = \left(\mathbf{X}^\top\mathbf{X} + k\boldsymbol{I}_p \right)^{-1} \mathbf{X}^\top \mathbf{y}, \tag{2.1}$$

where $\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}}$ is called a ridge full model estimator and $k \in [0, \infty]$ is tuning ridge parameter. If $k = 0$, then $\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}}$ is the LSE estimator, and $k = \infty$, then $\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}} = \mathbf{0}$. In this study, we select optimal the value of k which minimizes the mean square error of the equation (2.1) via 10-fold cross validation.

We let $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2})$, where $\mathbf{X_1}$ is an $n \times p_1$ sub-matrix containing the regressors of interest and $\mathbf{X_2}$ is an $n \times p_2$ sub-matrix that may or may not be relevant in the analysis of the main regressors. Similarly, $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top \right)^\top$ be the vector of parameters, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have dimensions p_1 and p_2, respectively, with $p_1 + p_2 = p$, $p_i \geq 0$ for $i = 1, 2$.

A sub-model or restricted model is defined as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{subject to } \boldsymbol{\beta}^\top\boldsymbol{\beta} \ \leq \phi \text{ and } \boldsymbol{\beta_2} = \mathbf{0},$$

then we have the following restricted linear regression model

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} \text{ subject to } \boldsymbol{\beta}_1^\top\boldsymbol{\beta}_1 \leq \phi. \tag{2.2}$$

We denote $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ as the full model or unrestricted ridge estimator of $\boldsymbol{\beta}_1$ is given by

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} = \left(\mathbf{X}_1^\top \boldsymbol{M}_2^R \mathbf{X}_1 + k\boldsymbol{I}_{p_1}\right)^{-1} \mathbf{X}_1^\top \boldsymbol{M}_2^{\mathrm{R}}\mathbf{y},$$

where $\boldsymbol{M}_2^{\mathrm{R}} = \boldsymbol{I}_n - \mathbf{X}_2\left(\mathbf{X}_2^\top\mathbf{X}_2 + k\boldsymbol{I}_{p_2}\right)^{-1}\mathbf{X}_2^\top$. For model (2.2), the sub-model or restricted estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ of $\boldsymbol{\beta}_1$ has the form

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} = \left(\mathbf{X}_1^\top\mathbf{X}_1 + k_1\boldsymbol{I}_{p_1}\right)^{-1}\mathbf{X}_1^\top\mathbf{y},$$

where k_1 is ridge parameter for sub-model estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$.

Generally speaking, $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ performs better than $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ when $\boldsymbol{\beta}_2$ is close to zero. However, for $\boldsymbol{\beta}_2$ away from the zero, $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ can be inefficient. But, the estimate $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ is consistent for departure of $\boldsymbol{\beta}_2$ from zero.

The idea of penalized estimation was introduced by Frank and Friedman (1993). They suggested the notion of bridge regression as follows. For a given penalty function $\pi(\cdot)$ and tuning parameter that controls the amount of shrinkage λ, bridge estimators are estimated by minimizing the following penalized least square criterion

$$\sum_{i=1}^{n}\left(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}\right)^2 + \lambda\pi(\beta),$$

where $\pi(\beta)$ is $\sum_{j=1}^{p}|\beta_j|^\gamma$, $\gamma > 0$. This penalty function bounds the L_γ norm of the parameters.

2.1. Pretest and Shrinkage Ridge Estimation

The pretest is a combination of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ and $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ through an indicator function $I\left(\mathscr{L}_n \leq c_{n,\alpha}\right)$, where $\mathscr{L}_n$ is appropriate test statistic to test $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_A : \boldsymbol{\beta}_2 \neq \mathbf{0}$. Moreover, $c_{n,\alpha}$ is an $\alpha-$level critical value using the distribution of $\mathscr{L}_n$. We define test statistics as follows:

$$\mathscr{L}_n = \frac{n}{\widehat{\sigma}^2}\left(\widehat{\boldsymbol{\beta}}_2^{\mathrm{LSE}}\right)^\top \mathbf{X}_2^\top\boldsymbol{M}_1\mathbf{X}_2\left(\widehat{\boldsymbol{\beta}}_2^{\mathrm{LSE}}\right),$$

where $\widehat{\sigma}^2 = \frac{1}{n-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}})^\top(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}})$ is consistent estimator of σ^2, $\boldsymbol{M}_1 = \boldsymbol{I}_n - \mathbf{X}_1\left(\mathbf{X}_1^\top\mathbf{X}_1\right)^{-1}\mathbf{X}_1^\top$ and $\widehat{\boldsymbol{\beta}}_2^{\mathrm{LSE}} = \left(\mathbf{X}_2^\top\boldsymbol{M}_1\mathbf{X}_2\right)^{-1}\mathbf{X}_2^\top\boldsymbol{M}_1\mathbf{y}$. Under H_0, the test statistic $\mathscr{L}_n$ follows chi-square distribution with p_2 degrees of freedom for large n values. The pretest test ridge regression estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ of $\boldsymbol{\beta}_1$ is defined by

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}} = \widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) I\left(\mathscr{L}_n \leq c_{n,\alpha}\right),$$

where $c_{n,\alpha}$ is an $\alpha-$ level critical value.

The shrinkage or Stein-type ridge regression estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}$ of $\boldsymbol{\beta}_1$ is defined by

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}} = \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} + \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(1 - (p_2 - 2)\mathscr{L}_n^{-1}\right),\ p_2 \geq 3.$$

The estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}$ is general form of the Stein-rule family of estimators, where shrinkage of the base estimator is towards the restricted estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$. The Shrinkage estimator is pulled towards the restricted estimator when the variance of the unrestricted estimator is large. Also, $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}$ is the smooth version of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$.

The positive part of the shrinkage ridge regression estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}$ of $\boldsymbol{\beta}_1$ defined by

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}} = \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} + \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(1 - (p_2 - 2)\mathscr{L}_n^{-1}\right)^{+},$$

where $z^{+} = max(0, z)$.

2.1.1. Lasso strategy

For $\gamma = 1$, we obtain the L_1 penalized least squares estimator, which is commonly known as Lasso (least absolute shrinkage and selection operator).

$$\widehat{\boldsymbol{\beta}}^{\mathrm{Lasso}} = \underset{\boldsymbol{\beta}}{\arg\min}\left\{\sum\nolimits_{i=1}^{n}\left(y_i - \mathbf{x}_i^{\top}\boldsymbol{\beta}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\}.$$

The parameter $\lambda \geq 0$ controls the amount of shrinkage.

2.1.2. Adaptive Lasso strategy

The adaptive Lasso estimator is defined as

$$\widehat{\boldsymbol{\beta}}^{\mathrm{aLasso}} = \underset{\boldsymbol{\beta}}{\arg\min}\left\{\sum\nolimits_{i=1}^{n}\left(y_i - \mathbf{x}_i^{\top}\boldsymbol{\beta}\right)^2 + \lambda\sum_{j=1}^{p}\widehat{\xi}_j\,|\beta_j|\right\},$$

where the weight function is

$$\widehat{\xi}_j = \frac{1}{|\beta_j{}^*|^\gamma};\ \gamma > 0.$$

The $\beta_j{}^*$ a root–n consistent estimator of β. For computational details we refer to Zou (2006).

2.1.3. SCAD strategy

The smoothly clipped absolute deviation (SCAD) is proposed by Fan and Li (2001). Given $a > 2$ and $\lambda > 0$, the SCAD penalty at β is

$$J_\lambda(\beta; a) = \begin{cases} \lambda\,|\beta|\,, & |\beta| \leq \lambda \\ -\left(\beta^2 - 2a\lambda\,|\lambda| + \lambda^2\right)/\left[2(a-1)\right], & \lambda < |\beta| \leq a\lambda \\ (a+1)\lambda^2/2 & |\beta| > a\lambda. \end{cases}$$

Hence, the SCAD estimation is given by

$$\widehat{\boldsymbol{\beta}}^{\text{SCAD}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right)^2 + \lambda \sum_{j=1}^{p} J_\lambda(\beta_j; a) \right\}.$$

For estimation strategies based on $\gamma = 2$, we establish some useful asymptotic results in the following section.

3. ASYMPTOTIC ANALYSIS

Consider a sequence of local alternatives $\{K_n\}$ given by

$$K_n : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{2(n)} = \frac{\boldsymbol{\kappa}}{\sqrt{n}},$$

where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_{p_2})^\top$ is a fixed vector. The asymptotic bias of an estimator $\boldsymbol{\beta}_1^*$ is defined as $\mathcal{B}(\boldsymbol{\beta}_1^*) = \mathcal{E} \lim\limits_{n\to\infty} \{\sqrt{n}\,(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1)\}$, the asymptotic covariance of an estimator $\boldsymbol{\beta}_1^*$ is $\boldsymbol{\Gamma}(\boldsymbol{\beta}_1^*) = \mathcal{E} \lim\limits_{n\to\infty} \left\{ n\,(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1)\,(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1)^\top \right\}$, and by using asymptotic covariance matrix $\boldsymbol{\Gamma}$, the asymptotic risk of an estimator $\boldsymbol{\beta}_1^*$ is given by $\mathcal{R}(\boldsymbol{\beta}_1^*) = \operatorname{tr}(\boldsymbol{W\Gamma})$, where $\boldsymbol{\kappa}$ is a positive definite matrix of weights with dimensions of $p \times p$, and $\boldsymbol{\beta}_1^*$ is one of the suggested estimators.

We consider two regularity conditions as the following to establish the asymptotic properties of the estimators.

(i) $\frac{1}{n} \max\limits_{1\leq i\leq n} \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i \to 0$ as $n \to \infty$, where $\mathbf{x}_i^\top$ is the ith row of $\mathbf{X}$

(ii) $\lim_{n\to\infty} n^{-1}(\mathbf{X}^\top \mathbf{X}) = \boldsymbol{C}$, for finite $\boldsymbol{C}$.

Theorem 3.1. *When $k \neq \infty$, if $k/\sqrt{n} \to \lambda_0 \geq 0$ and $\boldsymbol{C}$ is non-singular, then*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}} - \boldsymbol{\beta}\right) \xrightarrow{d} \left(-\lambda_0 \boldsymbol{C}^{-1}\boldsymbol{\beta},\, \sigma^2 \boldsymbol{C}^{-1}\right).$$

For proof, see Knight and Fu (2000).

Proposition 3.1. *Assuming above regularity conditions (i) and (ii) hold, then, together with Theorem 3.1, under $\{K_n\}$ as $n \to \infty$ we have*

$$\begin{pmatrix} \vartheta_1 \\ \vartheta_3 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} -\boldsymbol{\mu}_{11.2} \\ \boldsymbol{\delta} \end{pmatrix}, \begin{pmatrix} \sigma^2 \boldsymbol{C}_{11.2}^{-1} & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} \end{pmatrix}\right],$$

$$\begin{pmatrix} \vartheta_3 \\ \vartheta_2 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \boldsymbol{\delta} \\ -\boldsymbol{\gamma} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Phi} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \boldsymbol{C}_{11}^{-1} \end{pmatrix}\right],$$

where $\vartheta_1 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \boldsymbol{\beta}_1\right)$, $\vartheta_2 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} - \boldsymbol{\beta}_1\right)$, $\vartheta_3 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)$, $\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{pmatrix}$, $\boldsymbol{\gamma} = \boldsymbol{\mu}_{11.2} + \boldsymbol{\delta}$ *and* $\boldsymbol{\delta} = \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{\omega}$, $\boldsymbol{\Phi} = \sigma^2 \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}$, $\boldsymbol{\mu} = -\lambda_0 \boldsymbol{C}^{-1}\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$ *and* $\boldsymbol{\mu}_{11.2} = \boldsymbol{\mu}_1 - \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\left((\boldsymbol{\beta}_2 - \boldsymbol{\kappa}) - \boldsymbol{\mu}_2\right)$.

The expressions for bias for listed estimators are:

Theorem 3.2.

$$\begin{aligned}
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) &= -\boldsymbol{\mu}_{11.2} \\
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) &= -\boldsymbol{\gamma} \\
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right) &= -\boldsymbol{\mu}_{11.2} - \boldsymbol{\delta} H_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right), \\
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) &= -\boldsymbol{\mu}_{11.2} - (p_2 - 2)\boldsymbol{\delta}\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right), \\
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right) &= -\boldsymbol{\mu}_{11.2} - \boldsymbol{\delta} H_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right), \\
&\quad -(p_2 - 2)\boldsymbol{\delta}\mathcal{E}\left\{\chi^{-2}_{p_2+2}(\Delta)\, I\left(\chi^2_{p_2+2}(\Delta) > p_2 - 2\right)\right\},
\end{aligned}$$

where $\Delta = \left(\boldsymbol{\kappa}^\top \boldsymbol{C}_{22.1}^{-1}\boldsymbol{\kappa}\right)\sigma^{-2}$, $\boldsymbol{C}_{22.1} = \boldsymbol{C}_{22} - \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}$, *and* $H_v(x, \Delta)$ *is the cumulative distribution function of the non-central chi-squared distribution with non-centrality parameter Δ and v degree of freedom, and*

$$\mathcal{E}\left(\chi_v^{-2j}(\Delta)\right) = \int_0^\infty x^{-2j}\, dH_v(x, \Delta).$$

Proof: See Appendix. ▮

Now, we define the following asymptotic quadratic bias ($\mathcal{QB}$) of an estimator $\boldsymbol{\beta}_1^*$ by converting them into the quadratic form since the bias expression of all the estimators are not in the scalar form.

$$\mathcal{QB}\left(\boldsymbol{\beta}_1^*\right) = \left(\mathcal{B}\left(\boldsymbol{\beta}_1^*\right)\right)^\top \boldsymbol{C}_{11.2}\mathcal{B}\left(\boldsymbol{\beta}_1^*\right),$$

where $\boldsymbol{C}_{11.2} = \boldsymbol{C}_{11} - \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}$.

$$\begin{aligned}
\mathcal{QB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) &= \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2},\\
\mathcal{QB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) &= \boldsymbol{\gamma}^\top \boldsymbol{C}_{11.2}\boldsymbol{\gamma},\\
\mathcal{QB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right) &= \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2} + \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta} H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad +\boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2} H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad +\boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta} H^2_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right),\\
\mathcal{QB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) &= \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2} + (p_2-2)\boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta}\mathcal{E}\left(\chi^{-2}_{p_2+2}\left(\Delta\right)\right)\\
&\quad +(p_2-2)\boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2}\mathcal{E}\left(\chi^{-2}_{p_2+2}\left(\Delta\right)\right)\\
&\quad +(p_2-2)^2\boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta}\left(\mathcal{E}\left(\chi^{-2}_{p_2+2}\left(\Delta\right)\right)\right)^2,\\
\mathcal{QB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right) &= \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2} + \left(\boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2} + \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta}\right)\\
&\quad \cdot\left[H_{p_2+2}\left(p_2-2;\Delta\right)\right.\\
&\quad \left.+(p_2-2)\mathcal{E}\left\{\chi^{-2}_{p_2+2}\left(\Delta\right) I\left(\chi^{-2}_{p_2+2}\left(\Delta\right) > p_2-2\right)\right\}\right]\\
&\quad +\boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta}\left[H_{p_2+2}\left(p_2-2;\Delta\right)\right.\\
&\quad \left.+(p_2-2)\mathcal{E}\left\{\chi^{-2}_{p_2+2}\left(\Delta\right) I\left(\chi^{-2}_{p_2+2}\left(\Delta\right) > p_2-2\right)\right\}\right]^2.
\end{aligned}$$

The $\mathcal{QB}$ of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ is $\boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2}$ and the $\mathcal{QB}$ of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ is an unbounded function of $\boldsymbol{\gamma}^\top \boldsymbol{C}_{11.2}\boldsymbol{\gamma}$. The $\mathcal{QB}$ of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ starts from $\boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2}$ at $\Delta = 0$, and when Δ increases it increases to the maximum point and then decreases to zero. For the $\mathcal{QB}$s of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}$ and $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}$, they similarly start from $\boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2}$, and increase to a point, and then decrease towards zero.

Theorem 3.3. *Under local alternatives and assumed regularity conditions the risks of the estimators are:*

$$\begin{aligned}
\mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) &= \sigma^2 tr\left(\boldsymbol{W}\boldsymbol{C}_{11.2}^{-1}\right) + \boldsymbol{\mu}_{11.2}^\top \boldsymbol{W}\boldsymbol{\mu}_{11.2}\\
\mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) &= \sigma^2 \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{C}_{11}^{-1}\right) + \boldsymbol{\gamma}^\top \boldsymbol{W}\boldsymbol{\gamma}
\end{aligned}$$

$$
\begin{aligned}
\mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right) &= \mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) - 2\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&\quad -\sigma^2\mathrm{tr}(\boldsymbol{W}\boldsymbol{C}_{11.2}^{-1} - \boldsymbol{W}\boldsymbol{C}_{11}^{-1})H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&\quad +\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\left\{2H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) - H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\right\}, \\
\mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) &= \mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) + 2(p_2-2)\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\delta}\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right) \\
&\quad -(p_2-2)\sigma^2\mathrm{tr}\left(\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{W}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\right)\{2\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right) \\
&\quad -(p_2-2)\mathcal{E}\left(\chi^{-4}_{p_2+2}(\Delta)\right)\} \\
&\quad +(p_2-2)\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\{2\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right) \\
&\quad -2\mathcal{E}\left(\chi^{-2}_{p_2+4}(\Delta)\right) - (p_2-2)\mathcal{E}\left(\chi^{-4}_{p_2+4}(\Delta)\right)\}, \\
\mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right) &= \mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) \\
&\quad -2\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\delta}\mathcal{E}\left(\left\{1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta)\right\}I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right) \\
&\quad +(p_2-2)\sigma^2\mathrm{tr}\left(\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{W}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\right) \\
&\quad \cdot\left[2\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\right. \\
&\quad \left.-(p_2-2)\mathcal{E}\left(\chi^{-4}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\right] \\
&\quad -\sigma^2\mathrm{tr}\left(\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{W}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\right)H_{p_2+2}(p_2-2;\Delta) \\
&\quad +\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\left[2H_{p_2+2}(p_2-2;\Delta) - H_{p_2+4}(p_2-2;\Delta)\right] \\
&\quad -(p_2-2)\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\left[2\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\right. \\
&\quad -2\mathcal{E}\left(\chi^{-2}_{p_2+4}(\Delta)I\left(\chi^2_{p_2+4}(\Delta)\le p_2-2\right)\right) \\
&\quad \left.+(p_2-2)\mathcal{E}\left(\chi^{-4}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\right].
\end{aligned}
$$

Proof: See Appendix. ∎

Noting that if $\boldsymbol{C}_{12} = \mathbf{0}$, then all the risks reduce to common value $\sigma^2\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{C}_{11}^{-1}\right) + \boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\mu}_{11.2}$ for all $\boldsymbol{\omega}$. For $\boldsymbol{C}_{12} \neq \mathbf{0}$, the risk of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ remains constant while the risk of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ is an unbounded function of Δ since $\Delta \in [0,\infty)$. The risk of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ increases as Δ moves away from zero, achieves it maximum and then decreases towards the risk of the full model estimator. Thus, it is a bounded function of Δ. The risk of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ is smaller than the risk of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ for some small values of Δ and opposite conclusions holds for rest of the parameter space. It can be seen that $\mathcal{R}(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}) \le \mathcal{R}(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}) \le \mathcal{R}(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}})$, strictly inequality holds for small values of Δ. Thus positive shrinkage is superior to the shrinkage estimator. However, both shrinkage estimators outperform the full model estimator in the entire parameter space induced by Δ. On the other hand, the pretest estimator performs better than the shrinkage estimators when Δ takes small values and outside this interval the opposite conclusion holds.

4. SIMULATION STUDIES

In this section, we conduct a Monte Carlo simulation study. The design matrix is generated to be correlated with different magnitudes. We simulate the response from the following model:

$$y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + ... + x_{pi}\beta_p + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$ with $\sigma^2 = 1$. We generate the design matrix $\mathbf{X}$ from a multivariate normal distribution with mean vector $\boldsymbol{\mu} = \mathbf{0}_{p_1}$ and covariance matrix $\boldsymbol{\Sigma}_x$. Further, we consider the off-diagonal elements of the covariance matrix $\boldsymbol{\Sigma}_x$ are equal to be ρ, which is the coefficient of correlation between any two predictors, with $\rho = 0.25, 0.5, 0.75$. The ratio of the largest eigenvalue to the smallest eigenvalue of matrix $\mathbf{X}^\top\mathbf{X}$ is calculated as the condition number test (CNT) which is helpful in detecting the existence of multicollinearity in the design matrix. If the CNT is larger than 30, then the model may have significant multicollinearity, for which we refer to Belsley (1991).

For $H_0 : \beta_j = 0$, $j = p_1 + 1, p_1 + 2, \ldots, p$, with $p = p_1 + p_2$, the regression coefficients are set $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top = \left(\boldsymbol{\beta}_1^\top, \mathbf{0}_{p_2}^\top\right)^\top$ with $\boldsymbol{\beta}_1 = (1, 1, 1, 1)^\top$. In order to investigate the behaviour of the estimators, we define $\Delta^* = \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|$, where $\boldsymbol{\beta}_0 = \left(\boldsymbol{\beta}_1^\top, \mathbf{0}_{p_2}^\top\right)^\top$ and $\|\cdot\|$ is the Euclidean norm. We considered Δ values between 0 and 4. If $\Delta^* = 0$, then it means that we will have $\boldsymbol{\beta} = (1, 1, 1, 1, \underbrace{0, 0, \ldots, 0}_{p_2})^\top$ to generated the response under null hypothesis. On the other hand, when $\Delta^* > 0$, say $\Delta^* = 2$, we will have $\boldsymbol{\beta} = (1, 1, 1, 1, 2, \underbrace{0, 0, \ldots, 0}_{p_2-1})^\top$ to generated the response under the local alternative hypotheses. When we increase the number of Δ, it indicates the degree of violation of null hypothesis. In our simulation study, we consider the sample size of $n = 60, 100$. Also, the number of predictor variables: $(p_1, p_2) \in \{(4, 4), (4, 8), (4, 16), (4, 32)\}$. Finally, each realization was repeated 1000 times to calculate the MSE of suggested estimators and $\alpha = 0.05$. All computations were conducted using the statistical package R (R Development Core Team, 2010). The performance of one of the suggested estimator was evaluated by using MSE criterion. Also, the relative mean square efficiency (RMSE) of the $\boldsymbol{\beta}_1^{\blacktriangle}$ to the $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ is indicated by

$$\mathrm{RMSE}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} : \boldsymbol{\beta}_1^{\blacktriangle}\right) = \frac{\mathrm{MSE}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right)}{\mathrm{MSE}\left(\boldsymbol{\beta}_1^{\blacktriangle}\right)},$$

where $\boldsymbol{\beta}_1^{\blacktriangle}$ is one of the listed estimators.

For brevity, we report the results for the values of $n = 60, 100$, $p_1 = 4$, $p_2 = 32$ and $\rho = 0.75$ in Table 1, and we plot the simulation results in Figures 1 and 2.

Table 1: RMSE of estimators for $p_1 = 4$, $p_2 = 32$ and $\rho = 0.75$.

Δ	$n=60$					$n=100$				
	CNT	**RSM**	**RPT**	**RS**	**RPS**	**CNT**	**RSM**	**RPT**	**RS**	**RPS**
0.000		2.240	1.964	2.015	2.158		1.871	1.737	1.749	1.802
0.200		2.152	2.016	1.910	2.074		1.695	1.546	1.623	1.667
0.400		2.099	1.689	1.918	2.082		1.541	1.283	1.508	1.543
0.600		1.621	1.323	1.615	1.708		1.298	1.058	1.387	1.400
0.800		1.396	1.027	1.554	1.589		1.156	0.955	1.392	1.396
1.000	1798.267	1.193	0.908	1.465	1.500	599.313	0.815	0.925	1.209	1.209
1.250		1.037	0.885	1.410	1.410		0.700	0.962	1.217	1.217
1.500		0.798	0.982	1.352	1.352		0.540	0.993	1.111	1.111
1.750		0.628	0.985	1.238	1.238		0.411	1.000	1.078	1.078
2.000		0.586	0.995	1.227	1.227		0.319	1.000	1.060	1.060
4.000		0.198	1.000	1.058	1.058		0.098	1.000	1.018	1.018

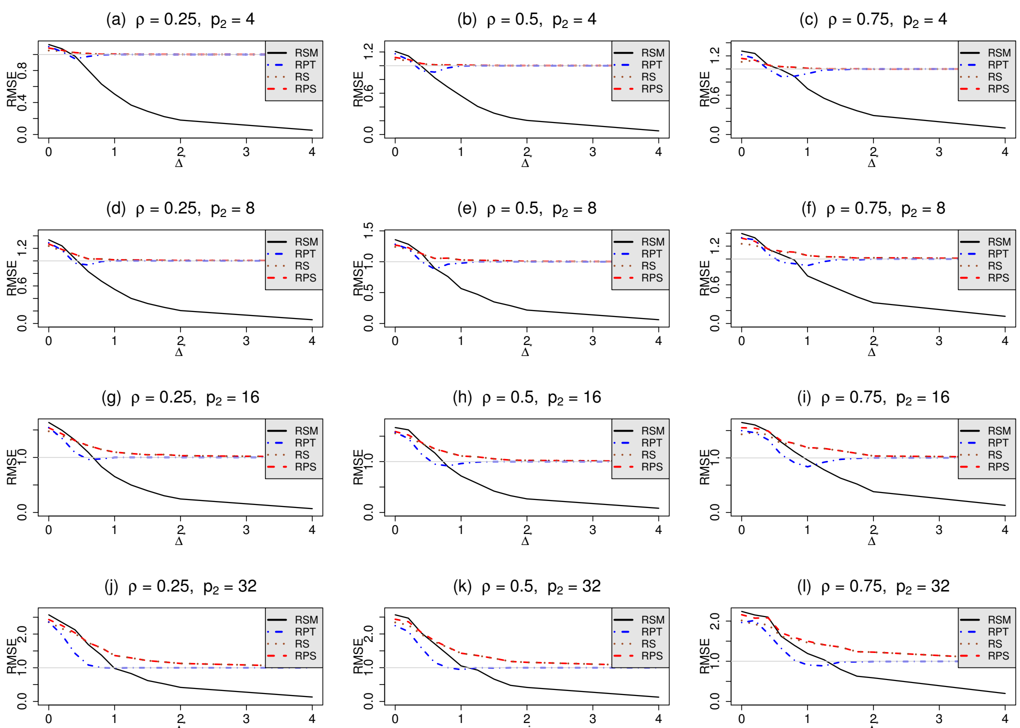

Figure 1: RMSE of the estimators as a function of the non-centrality parameter Δ^* when $n = 60$ and $p_1 = 4$.

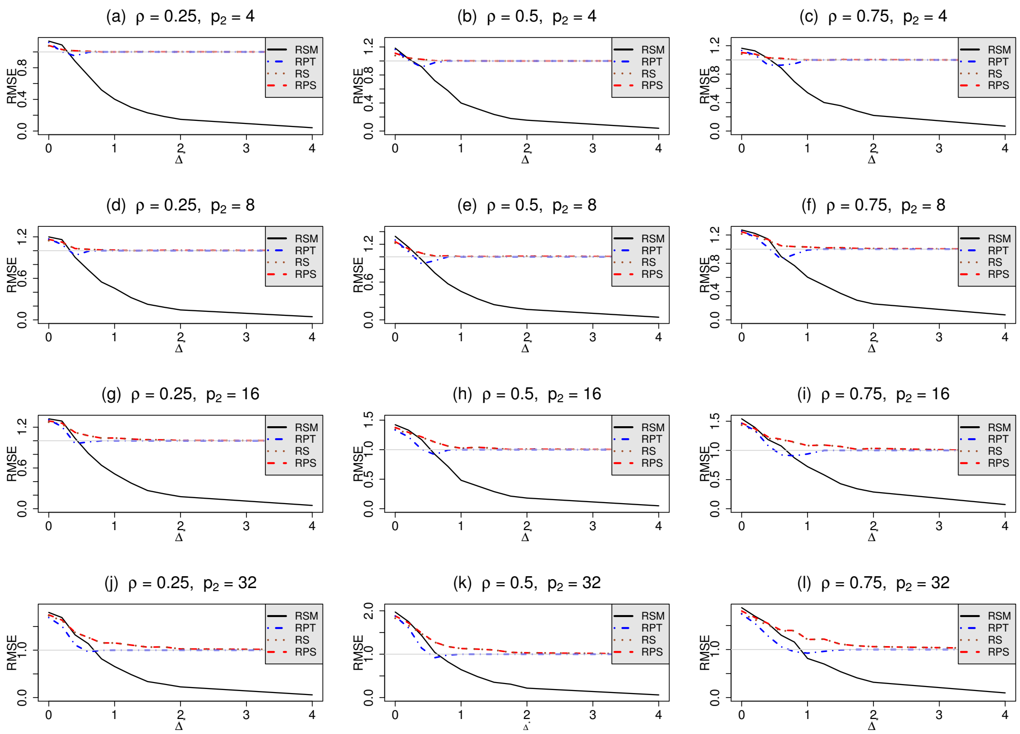

Figure 2: RMSE of the estimators as a function of the non-centrality parameter Δ^* when $n = 100$ and $p_1 = 4$.

The findings can be summarized as follows:

a) When $\Delta^* = 0$, RSM outperforms all listed estimators. In contrast to this, after the small interval near Δ^*, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ decreases and goes to zero.

b) The RPT outperforms RS and RPS in case of $\Delta^* = 0$. However, for large p_2 values while keeping p_1 and n fixed, RPT is less efficient than RPS. When Δ^* is larger than zero, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ decreases, and it remains below 1 for intermediate values of Δ^*, after that the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ increases and approaches one for larger values of Δ^*.

c) Clearly, RPS performs better than RS in the entire parameter space induced by Δ^*. Both shrinkage estimators outshine the full model estimator regardless the correctness of the selected sub-model at hand. This is consistent with the asymptotic theory we presented earlier. Recalling that Δ^* measures the degree of deviation from the assumption on the parameter space, it is clear that one cannot go wrong with the use of shrinkage estimators even if the selected sub-model is not correctly specified. As evident from the table and graphs, if the selected sub-model is correct, that is, $\Delta^* = 0$ then shrinkage estimators are relatively highly efficient than the full model estimator. In other words risk reduction is substantial. On the other hand, the gain slowly diminishes if the sub-model is grossly misspecified. Nevertheless, the shrinkage estimators are at least as good as the full model estimator in terms of risk. Hence, the use of shrinkage estimators make sense in real-life applications when a sub-model cannot be correctly specified, which is the case in most applications.

d) Generally speaking, ridge-type estimators perform better than classical estimator in the presence of multicollinearity among predictors. Our simulation results strongly corroborates to this effect; the RMSE of the ridge-type estimators are increasing function of the amount of multicollinearity.

4.1. Comparison Lasso, aLasso and SCAD

For comparison purposes, we considered $n = 50, 75$, $p_2 = 5, 9, 15, 20$ and $p_1 = 5$ at $\Delta^* = 0$. Here, we used *cv.glmnet* function in *glmnet* package in R for Lasso and aLasso, and *cv.ncvreg* function in *ncvreg* package for SCAD method. The weights for aLasso are obtained from the 10-fold CV Lasso. Results are presented in Table 2.

Table 2: CNT and RMSE of estimators for $p_1 = 5$.

n	ρ	p_2	**CNT**	**LSE**	**RSM**	**RPT**	**RS**	**RPS**	**Lasso**	**aLasso**	**SCAD**
50	0.25	5	10.88	0.90	2.63	2.54	1.61	1.83	1.30	1.46	1.39
		9	20.36	0.85	3.84	3.50	2.34	2.78	1.54	1.92	1.89
		15	46.75	0.76	5.51	4.12	2.85	4.17	1.92	2.56	2.77
		20	90.66	0.65	7.32	5.36	4.24	5.44	2.25	3.18	3.14
	0.5	5	29.39	0.85	2.75	2.39	1.04	1.93	1.22	1.29	1.16
		9	53.39	0.77	4.10	3.32	2.22	2.94	1.46	1.67	1.38
		15	126.44	0.67	5.93	4.69	3.06	4.38	1.76	2.13	1.65
		20	245.60	0.54	8.07	5.77	4.13	5.70	1.99	2.38	1.94
	0.75	5	79.58	0.71	3.93	2.80	1.66	2.12	1.01	0.93	0.73
		9	156.47	0.64	4.75	3.14	2.07	3.01	1.18	1.03	0.73
		15	385.42	0.48	6.50	4.18	2.72	4.41	1.35	1.26	0.83
		20	718.83	0.39	8.94	4.64	3.40	5.91	1.50	1.25	0.83
75	0.25	5	8.90	0.94	2.20	1.97	1.53	1.64	1.25	1.53	1.48
		9	15.12	0.91	3.44	2.96	2.25	2.68	1.60	2.12	2.09
		15	28.11	0.85	5.54	3.26	3.63	3.88	2.05	3.13	2.99
		20	43.77	0.78	7.15	4.11	4.94	5.47	2.63	4.13	3.78
	0.5	5	22.77	0.88	2.59	2.11	1.45	1.79	1.25	1.45	1.19
		9	38.33	0.86	4.03	2.95	2.32	2.73	1.56	1.96	1.59
		15	77.16	0.78	5.79	4.34	3.42	4.35	1.97	2.71	2.45
		20	122.80	0.72	7.30	5.52	4.13	5.50	2.28	3.15	2.75
	0.75	5	65.35	0.80	3.21	2.63	1.33	2.02	1.13	1.11	0.89
		9	113.78	0.76	5.27	3.67	2.30	3.32	1.42	1.52	1.17
		15	225.06	0.66	6.81	4.24	3.80	4.68	1.61	1.71	1.20
		20	359.89	0.57	7.59	5.52	4.11	5.58	1.82	1.91	1.41

Not surprisingly, the performance of the sub-model estimator is the best. The pretest estimator also performs better than other estimators. However, the performance of RPS is better than RPT for larger values of p_2. The performance LSE estimator is worse than listed estimators since the designed matrix is ill-conditioned. The performance of the Lasso, aLasso and SCAD are comparable when ρ is small. On the other hand, pretest and shrinkage estimators remain stable for a given value of ρ. Also, for large values of p_2, the shrinkage and pretest estimators indicate their superiority over L_1 penalty estimators. Thus, we recommend using shrinkage estimators in the presence of multicollinearity.

5. APPLICATION

We use the air pollution and mortality rate data by McDonald and Schwing (1973). This data includes $p = 15$ measurements on mortality rate and explanatory variables, which are air-pollution, socio-economic and meteorological, for $n = 60$ US cities in 1960. The data are freely available from Carnegie Mellon University's StatLib (`http://lib.stat.cmu.edu/datasets/`). In Table 3, we listed variables. Also, the CNT value is calculated as 882.081.574 which implies the existence of multicollinearity in the data set.

Table 3: Lists and Descriptions of Variables

Variables	**Descriptions**
Dependent Variable	
mort	Total age-adjusted mortality rate per 100.000
Covariates	
Air-Pollution	
prec	Average annual precipitation in inches
jant	Average January temperature in degrees F
jult	Average July temperature in degrees F
humid	Annual average % relative humidity at 1pm
Socio-Economic	
ovr65	% of 1960 SMSA population aged 65 or older
popn	Average household size
educ	Median school years completed by those over 22
hous	% of housing units which are sound & with all facilities
dens	Population per sq. mile in urbanized areas, 1960
nonw	% non-white population in urbanized areas, 1960
wwdrk	% employed in white collar occupations
poor	% of families with income < 3000
Meteorological	
hc	Relative hydrocarbon pollution potential of hydrocarbons
nox	Relative hydrocarbon pollution potential of nitric oxides
so2	Relative hydrocarbon pollution potential of sulphur dioxides

In order to apply the proposed methods, we use two step approach since the prior information is not available here. In the first step, one might do usual variable selection to select the best sub-model. We use the Best Subset Selection (BSS). It showed that prec, jant, jult, educ, dens, nonw, hc and nox are the most important covariates for prediction of the response variable mort and the other variables may be ignored since they are not significantly important. In the second step, we have two model which are the full model with all the covariates and the sub-model with covariates via the BSS. Finally, we construct the shrinkage techniques from the full-model and the sub-model. We fit the full and sub-model which are given in Table 4.

Table 4: Fittings of full and sub-models

Models	**Formulas**
Full model	log(mort) = $\beta_0 + \beta_1$prec $+\beta_2$jant$+\beta_3$jult$+\beta_4$ovr65$+\beta_5$popn$+\beta_6$educ$+\beta_7$hous $+\beta_8$dens$+\beta_9$nonw$+\beta_{10}$wwdrk$+\beta_{11}$poor$+\beta_{12}$hc$+\beta_{13}$nox$+\beta_{14}$so2$+\beta_{15}$humid
Sub-Model	log(mort) = $\beta_0 + \beta_1$prec $+\beta_2$jant$+\beta_3$jult$+\beta_6$educ$+\beta_8$dens$+\beta_9$nonw$+\beta_{12}$hc$+\beta_{13}$nox

To evaluate the performance of the suggested estimators, we calculate the predictive error (PE) of an estimator. Furthermore, we define the relative predictive error (RPE) of $\widehat{\boldsymbol{\beta}}^*$ in terms of the full model ridge regression estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}}$ to easy comparison, is evaluated by as follows

$$\mathrm{RPE}\left(\widehat{\boldsymbol{\beta}}^*\right) = \frac{\mathrm{PE}(\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}})}{\mathrm{PE}(\widehat{\boldsymbol{\beta}}^*)},$$

where $\widehat{\boldsymbol{\beta}}^*$ can be any of the listed estimators. If the RPE is larger than one, it indicates the superior to RFM.

Our results are based on 2500 case resampled bootstrap samples. Since there is no noticeable variation for larger number of replications, we did not consider further values. The average prediction errors were calculated via 10-fold CV for each bootstrap replicate. The predictors were first standardized to have zero mean and unit standard deviation before fitting the model. Figure 3 shows that prediction errors of estimators. As expected, the RSM has the smallest prediction error since the suggested sub-model is correct. Also, the Lasso, aLasso and SCAD have higher prediction error than the suggested techniques.

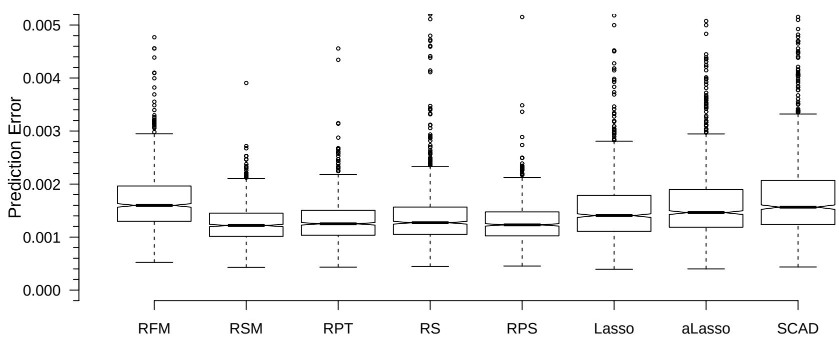

Figure 3: Prediction errors of listed estimators based on bootstrap simulation

Table 5: Estimate (first row) and standard error (second row) for significant coefficients for the air pollution and mortality rate data. The RPE column gives the relative efficiency based on bootstrap simulation with respect to the RFM.

	(Const.)	**prec**	**jant**	**jult**	**educ**	**dens**	**nonw**	**hc**	**nox**	**RPE**
RFM	6.846	0.013	-0.010	-0.002	-0.008	0.010	0.019	-0.007	0.009	1.000
	0.005	0.005	0.006	0.005	0.006	0.004	0.007	0.015	0.017	
RSM	6.845	0.016	-0.018	-0.012	-0.013	0.015	0.042	-0.080	0.081	1.336
	0.005	0.007	0.007	0.006	0.006	0.004	0.007	0.026	0.022	
RPT	6.845	0.016	-0.018	-0.011	-0.013	0.015	0.040	-0.072	0.073	1.288
	0.005	0.007	0.007	0.006	0.007	0.004	0.009	0.029	0.026	
RS	6.845	0.017	-0.017	-0.010	-0.012	0.015	0.038	-0.063	0.065	1.160
	0.005	0.007	0.007	0.007	0.008	0.005	0.012	0.039	0.037	
RPS	6.845	0.016	-0.016	-0.009	-0.012	0.014	0.035	-0.055	0.057	1.316
	0.005	0.006	0.006	0.006	0.007	0.004	0.008	0.025	0.023	
Lasso	6.845	0.019	-0.019	-0.008	-0.012	0.014	0.035	-0.029	0.032	1.060
	0.005	0.009	0.009	0.007	0.011	0.006	0.011	0.046	0.049	
aLasso	6.845	0.022	-0.022	-0.013	-0.012	0.015	0.039	-0.037	0.040	0.965
	0.006	0.010	0.009	0.008	0.012	0.006	0.012	0.050	0.053	
SCAD	6.845	0.019	-0.022	-0.010	-0.014	0.014	0.039	-0.035	0.038	0.897
	0.006	0.012	0.011	0.009	0.013	0.007	0.014	0.052	0.055	

Table 5 reveals that the RPE of the sub-model estimator, pretest, shrinkage and positive part of shrinkage estimators outperform the full model estimator. On the other hand, the sub-model estimator has the highest RPE since it is computed based on the assumption that the selected sub-model is the true model. As expected due to the presence of multicollinearity, the performance of both ridge-type shrinkage and pretest estimators is good and better than estimators based on L_1 criteria. Thus, the data analysis corroborates with our simulation and theoretical findings.

6. CONCLUSIONS

In this study we assessed the performance of least squares, pretest ridge, shrinkage ridge and L_1 estimators when predictors are correlated. We established the asymptotic properties of the pretest ridge and shrinkage ridge estimators. We demonstrated that shrinkage ridge estimators outclass the full model estimator and relatively perform better than sub-model estimator in a wide range of the parameter space. We conducted a Monte Carlo simulation to investigate the behavior of proposed estimators when a selected sub-model may or may not be a true model. Not surprisingly, the sub-model ridge regression estimator outshines all other estimators when the selected sub-model is the true one. However, when this assumption is violated, the performance of the sub-model estimator is profoundly poor. Further, the shrinkage estimators outperform pretest ridge estimators when p_2 is large. Our asymptotic theory is well supported by numerical analysis.

We also analyze the relative performance Lasso, adaptive Lasso and SCAD with other listed estimators. We observe that the performance of pretest and shrinkage ridge regression estimators are superior to L_1 estimators when predictors are highly correlated. The result of a data analysis is very consistent with theoretical and simulated analysis. In conclusion, we suggest to use ridge-type shrinkage estimators when the design matrix is ill-conditioned. The result of this paper are general in nature and consistent with the available results in the reviewed literature. Further, the result of this paper maybe extended to host of models and applications.

APPENDIX

By using $\widetilde{\boldsymbol{y}} = \mathbf{y} - \mathbf{X}_2\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}$

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} &= \underset{\boldsymbol{\beta_1}}{\arg\min}\left\{\|\widetilde{\boldsymbol{y}} - \mathbf{X}_1\boldsymbol{\beta}_1\| + k\|\boldsymbol{\beta}_1\|^2\right\} \\
&= \left(\mathbf{X}_1^\top\mathbf{X}_1 + k\boldsymbol{I}_{p_1}\right)^{-1}\mathbf{X}_1^\top\widetilde{\boldsymbol{y}} \\
&= \left(\mathbf{X}_1^\top\mathbf{X}_1 + k\boldsymbol{I}_{p_1}\right)^{-1}\mathbf{X}_1^\top\mathbf{y} - \left(\mathbf{X}_1^\top\mathbf{X}_1 + k\boldsymbol{I}_{p_1}\right)^{-1}\mathbf{X}_1^\top\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}} \\
&= \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} - \left(\mathbf{X}_1^\top\mathbf{X}_1 + k\boldsymbol{I}_{p_1}\right)^{-1}\mathbf{X}_1^\top\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}. \qquad (6.1)
\end{aligned}$$

Using the equation (6.1), under local alternative $\{K_n\}$, $\boldsymbol{\Phi}$ is derived as follows:

$$\begin{aligned}
\boldsymbol{\Phi} &= Cov\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) \\
&= \mathcal{E}\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)^{\top}\right] \\
&= \mathcal{E}\left[\left(\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}\right)\left(\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}\right)^{\top}\right] \\
&= \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\mathcal{E}\left[\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}\left(\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}\right)^{\top}\right]\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1} \\
&= \sigma^2\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1} = \sigma^2(\boldsymbol{C}_{11.2}^{-1} - \boldsymbol{C}_{11}^{-1})
\end{aligned}$$

Lemma 6.1. *Let* $\mathbf{X}$ *be* $q-$*dimensional normal vector distributed as* $\mathcal{N}\left(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_q\right)$, *then, for a measurable function of* φ, *we have*

$$\begin{aligned}
\mathcal{E}\left[\mathbf{X}\varphi\left(\mathbf{X}^{\top}\mathbf{X}\right)\right] &= \boldsymbol{\mu}_x\mathcal{E}\left[\varphi\chi^2_{q+2}\left(\Delta\right)\right] \\
\mathcal{E}\left[\boldsymbol{X}\boldsymbol{X}^{\top}\varphi\left(\mathbf{X}^{\top}\mathbf{X}\right)\right] &= \boldsymbol{\Sigma}_q\mathcal{E}\left[\varphi\chi^2_{q+2}\left(\Delta\right)\right] + \boldsymbol{\mu}_x\boldsymbol{\mu}_x^{\top}\mathcal{E}\left[\varphi\chi^2_{q+4}\left(\Delta\right)\right]
\end{aligned}$$

where $\chi^2_v\left(\Delta\right)$ *is a non-central chi-square distribution with* v *degrees of freedom and non-centrality parameter* Δ.

Proof: It can be found in Judge and Bock (1978) ∎

Proof of Theorem 3.2: $\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) = -\boldsymbol{\mu}_{11.2}$ is provided by Proposition 3.1, and

$$\begin{aligned}
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} - \boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}} - \boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \boldsymbol{\beta}_1\right)\right\} - \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2} - \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{\omega} = -\left(\boldsymbol{\mu}_{11.2} + \boldsymbol{\delta}\right) = -\boldsymbol{\gamma}.
\end{aligned}$$

Hence, by using Lemma 6.1, it can be written as follows:

$$\begin{aligned}
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}} - \boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)I\left(\mathscr{L}_n \le c_{n,\alpha}\right) - \boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \boldsymbol{\beta}_1\right)\right\} \\
&\quad -\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)I\left(\mathscr{L}_n \le c_{n,\alpha}\right)\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2} - \boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right).
\end{aligned}$$

$$
\begin{aligned}
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}-\boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\,\mathscr{L}_n^{-1}-\boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)\right\} \\
&\quad -\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\,\mathscr{L}_n^{-1}\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2}-(p_2-2)\,\boldsymbol{\delta}\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right).
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}-\boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}+\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(1-(p_2-2)\,\mathscr{L}_n^{-1}\right)\right. \\
&\quad \left. I\left(\mathscr{L}_n>p_2-2\right)-\boldsymbol{\beta}_1)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left[\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}+\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(1-I\left(\mathscr{L}_n\le p_2-2\right)\right)\right.\right. \\
&\quad \left.\left. -\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\,\mathscr{L}_n^{-1}I\left(\mathscr{L}_n>p_2-2\right)-\boldsymbol{\beta}_1\right]\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)\right\} \\
&\quad -\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)I\left(\mathscr{L}_n\le p_2-2\right)\right\} \\
&\quad -\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\,\mathscr{L}_n^{-1}I\left(\mathscr{L}_n>p_2-2\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2}-\boldsymbol{\delta}H_{p_2+2}\left(p_2-2;(\Delta)\right) \\
&\quad -\boldsymbol{\delta}\,(p_2-2)\,\mathcal{E}\left\{\chi_{p_2+2}^{-2}(\Delta)\,I\left(\chi_{p_2+2}^{2}(\Delta)>p_2-2\right)\right\}.
\end{aligned}
$$

∎

Proof of Theorem 3.3: Firstly, the asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ is given by

$$
\begin{aligned}
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)^{\top}\right\} \\
&= \mathcal{E}\left(\vartheta_1\vartheta_1^{\top}\right)=Cov\left(\vartheta_1\vartheta_1^{\top}\right)+\mathcal{E}\left(\vartheta_1\right)\mathcal{E}\left(\vartheta_1^{\top}\right)=\sigma^2\boldsymbol{C}_{11.2}^{-1}+\boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^{\top}.
\end{aligned}
$$

The asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ is given by

$$
\begin{aligned}
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}-\boldsymbol{\beta}_1\right)^{\top}\right\} \\
&= \mathcal{E}\left(\vartheta_2\vartheta_2^{\top}\right)=Cov\left(\vartheta_2\vartheta_2^{\top}\right)+\mathcal{E}\left(\vartheta_2\right)\mathcal{E}\left(\vartheta_2^{\top}\right)=\sigma^2\boldsymbol{C}_{11}^{-1}+\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top},
\end{aligned}
$$

The asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ is given by

$$\begin{aligned}
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}-\boldsymbol{\beta}_1\right)^\top\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty} n\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right]\right. \\
&\quad \left.\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right]^\top\right\} \\
&= \mathcal{E}\left\{\left[\vartheta_1-\vartheta_3 I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right]\left[\vartheta_1-\vartheta_3 I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right]^\top\right\} \\
&= \mathcal{E}\left\{\vartheta_1\vartheta_1^\top-2\vartheta_3\vartheta_1^\top I\left(\mathscr{L}_n\le c_{n,\alpha}\right)+\vartheta_3\vartheta_3^\top I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}.
\end{aligned}$$

Considering,

$$\begin{aligned}
&\mathcal{E}\left\{\vartheta_3\vartheta_1^\top I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\} \\
&= \mathcal{E}\left\{\mathcal{E}\left(\vartheta_3\vartheta_1^\top I\left(\mathscr{L}_n\le c_{n,\alpha}\right)|\vartheta_3\right)\right\} = \mathcal{E}\left\{\vartheta_3\mathcal{E}\left(\vartheta_1^\top I\left(\mathscr{L}_n\le c_{n,\alpha}\right)|\vartheta_3\right)\right\} \\
&= \mathcal{E}\left\{\vartheta_3\left[-\mu_{11.2}+\left(\vartheta_3-\boldsymbol{\delta}\right)\right]^\top I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\} \\
&= -\mathcal{E}\left\{\vartheta_3\boldsymbol{\mu}_{11.2}^\top I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}+\mathcal{E}\left\{\vartheta_3\left(\vartheta_3-\boldsymbol{\delta}\right)^\top I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2}^\top\mathcal{E}\left\{\vartheta_3 I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}+\mathcal{E}\left\{\vartheta_3\vartheta_3^\top I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\} \\
&\quad -\mathcal{E}\left\{\vartheta_3\boldsymbol{\delta}^\top I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}
\end{aligned}$$

and based on Lemma 6.1, we have

$$\begin{aligned}
&\mathcal{E}\left\{\vartheta_3\vartheta_1^\top I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2}^\top\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)+\left\{Cov(\vartheta_3\vartheta_3^\top)H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\right. \\
&\quad \left.+\mathcal{E}\left(\vartheta_3\right)\mathcal{E}\left(\vartheta_3^\top\right)H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)-\boldsymbol{\delta}\boldsymbol{\delta}^\top H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2}^\top\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)+\boldsymbol{\Phi}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&\quad +\boldsymbol{\delta}\boldsymbol{\delta}^\top H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)-\boldsymbol{\delta}\boldsymbol{\delta}^\top H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right),
\end{aligned}$$

then,

$$\begin{aligned}
&\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right) \\
&= \boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^\top+2\boldsymbol{\mu}_{11.2}^\top\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)+\sigma^2\boldsymbol{C}_{11.2}^{-1}-\boldsymbol{\Phi}H_{p_2+2}\left(\chi^2_{p_2,\alpha};(\Delta)\right) \\
&\quad -\boldsymbol{\delta}\boldsymbol{\delta}^\top H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)+2\boldsymbol{\delta}\boldsymbol{\delta}^\top H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&= \sigma^2\boldsymbol{C}_{11.2}^{-1}+\boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^\top+2\boldsymbol{\mu}_{11.2}^\top\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&\quad +\sigma^2(\boldsymbol{C}_{11.2}^{-1}-\boldsymbol{C}_{11}^{-1})H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&\quad +\boldsymbol{\delta}\boldsymbol{\delta}^\top\left[2H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)-H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\right].
\end{aligned}$$

The asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}$ is given by

$$\begin{aligned}
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}-\boldsymbol{\beta}_1\right)^\top\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty} n\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\mathscr{L}_n^{-1}\right]\right. \\
&\quad \left.\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\mathscr{L}_n^{-1}\right]^\top\right\} \\
&= \mathcal{E}\left\{\vartheta_1\vartheta_1^\top - 2(p_2-2)\vartheta_3\vartheta_1^\top\mathscr{L}_n^{-1} + (p_2-2)^2\vartheta_3\vartheta_3^\top\mathscr{L}_n^{-2}\right\}.
\end{aligned}$$

Considering,

$$\begin{aligned}
\mathcal{E}\left\{\vartheta_3\vartheta_1^\top\mathscr{L}_n^{-1}\right\} &= \mathcal{E}\left\{\mathcal{E}\left(\vartheta_3\vartheta_1^\top\mathscr{L}_n^{-1}|\vartheta_3\right)\right\} = \mathcal{E}\left\{\vartheta_3\mathcal{E}\left(\vartheta_1^\top\mathscr{L}_n^{-1}|\vartheta_3\right)\right\} \\
&= \mathcal{E}\left\{\vartheta_3\left[-\boldsymbol{\mu}_{11.2}+(\vartheta_3-\boldsymbol{\delta})\right]^\top\mathscr{L}_n^{-1}\right\} \\
&= -\mathcal{E}\left\{\vartheta_3\boldsymbol{\mu}_{11.2}^\top\mathscr{L}_n^{-1}\right\} + \mathcal{E}\left\{\vartheta_3(\vartheta_3-\boldsymbol{\delta})^\top\mathscr{L}_n^{-1}\right\} \\
&= -\boldsymbol{\mu}_{11.2}^\top\mathcal{E}\left\{\vartheta_3\mathscr{L}_n^{-1}\right\} + \mathcal{E}\left\{\vartheta_3\vartheta_3^\top\mathscr{L}_n^{-1}\right\} - \mathcal{E}\left\{\vartheta_3\boldsymbol{\delta}^\top\mathscr{L}_n^{-1}\right\}
\end{aligned}$$

by using Lemma 6.1, we have

$$\begin{aligned}
\mathcal{E}\left\{\vartheta_3\vartheta_1^\top\mathscr{L}_n^{-1}\right\} &= -\boldsymbol{\mu}_{11.2}^\top\boldsymbol{\delta}\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right) + \left\{Cov(\vartheta_3\vartheta_3^\top)\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)\right. \\
&\quad \left.+\mathcal{E}(\vartheta_3)\mathcal{E}\left(\vartheta_3^\top\right)\mathcal{E}\left(\chi_{p_2+4}^{-2}(\Delta)\right) - \boldsymbol{\delta}\boldsymbol{\delta}^\top H_{p_2+2}\left(\chi_{p_2,\alpha}^2;\Delta\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2}^\top\boldsymbol{\delta}\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right) + \boldsymbol{\Phi}\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right) \\
&\quad +\boldsymbol{\delta}\boldsymbol{\delta}^\top\mathcal{E}\left(\chi_{p_2+4}^{-2}(\Delta)\right) - \boldsymbol{\delta}\boldsymbol{\delta}^\top\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right).
\end{aligned}$$

Then,

$$\begin{aligned}
&\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) \\
&= \sigma^2\boldsymbol{C}_{11.2}^{-1} + \boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^\top + 2(p_2-2)\boldsymbol{\mu}_{11.2}^\top\boldsymbol{\delta}\mathcal{E}\left(\chi_{p_2+2,\alpha}^{-2}(\Delta)\right) \\
&\quad -(p_2-2)\boldsymbol{\Phi}\left\{2\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right) - (p_2-2)\mathcal{E}\left(\chi_{p_2+2}^{-4}(\Delta)\right)\right\} \\
&\quad +(p_2-2)\boldsymbol{\delta}\boldsymbol{\delta}^\top\left\{-2\mathcal{E}\left(\chi_{p_2+4}^{-2}(\Delta)\right) + 2\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right) + (p_2-2)\mathcal{E}\left(\chi_{p_2+4}^{-4}(\Delta)\right)\right\}.
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathbf{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty} n\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}-\boldsymbol{\beta}_1\right)\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}-\boldsymbol{\beta}_1\right)^\top\right\} \\
&= \mathbf{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) - 2\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}-\boldsymbol{\beta}_1\right)^\top\right.\right. \\
&\quad \times \left.\left.\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right]\right\} \\
&\quad +\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)^\top\right.\right. \\
&\quad \times \left.\left.\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\}^2 I\left(\mathscr{L}_n \le p_2-2\right)\right]\right\} \\
&= \mathbf{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) - 2\mathcal{E}\left\{\vartheta_3\vartheta_1^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad +2\mathcal{E}\left\{\vartheta_3\vartheta_3^\top (p_2-2)\mathscr{L}_n^{-1} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad -2\mathcal{E}\left\{\vartheta_3\vartheta_3^\top (p_2-2)^2\mathscr{L}_n^{-2} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad +\mathcal{E}\left\{\vartheta_3\vartheta_3^\top I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad -2\mathcal{E}\left\{\vartheta_3\vartheta_3^\top (p_2-2)\mathscr{L}_n^{-1} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad +\mathcal{E}\left\{\vartheta_3\vartheta_3^\top (p_2-2)^2\mathscr{L}_n^{-2} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&= \mathbf{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) - 2\mathcal{E}\left\{\vartheta_3\vartheta_1^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad -\mathcal{E}\left\{\vartheta_3\vartheta_3^\top (p_2-2)^2\mathscr{L}_n^{-2} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad +\mathcal{E}\left\{\vartheta_3\vartheta_3^\top I\left(\mathscr{L}_n \le p_2-2\right)\right\}.
\end{aligned}$$

Considering,

$$\begin{aligned}
&\mathcal{E}\left\{\vartheta_3\vartheta_1^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&= \mathcal{E}\left\{\mathcal{E}\left(\vartheta_3\vartheta_1^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)|\vartheta_3\right)\right\} \\
&= \mathcal{E}\left\{\vartheta_3\mathcal{E}\left(\vartheta_1^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)|\vartheta_3\right)\right\} \\
&= \mathcal{E}\left\{\vartheta_3\left[-\boldsymbol{\mu}_{11.2}+(\vartheta_3-\boldsymbol{\delta})\right]^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2}\mathcal{E}\left(\vartheta_3\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right) \\
&\quad +\mathcal{E}\left(\vartheta_3\vartheta_3^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right) \\
&\quad -\mathcal{E}\left(\vartheta_3\boldsymbol{\delta}^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right) \\
&= -\boldsymbol{\delta}\boldsymbol{\mu}_{11.2}^\top\boldsymbol{E}\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\} I\left(\chi_{p_2+2}^{2}(\Delta) \le p_2-2\right)\right) \\
&\quad \mathbf{\Phi}\mathcal{E}\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\} I\left(\chi_{p_2+2}^{2}(\Delta) \le p_2-2\right)\right) \\
&\quad +\boldsymbol{\delta}\boldsymbol{\delta}^\top\mathcal{E}\left(\left\{1-(p_2-2)\chi_{p_2+4}^{-2}(\Delta)\right\} I\left(\chi_{p_2+4}^{2}(\Delta) \le p_2-2\right)\right) \\
&\quad -\boldsymbol{\delta}\boldsymbol{\delta}^\top\mathcal{E}\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\} I\left(\chi_{p_2+2}^{2}(\Delta) \le p_2-2\right)\right),
\end{aligned}$$

we have

$$
\begin{aligned}
&\mathbf{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right)\\
&= \mathbf{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) + 2\boldsymbol{\delta}\boldsymbol{\mu}_{11.2}^{\top}\mathcal{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\} I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad -2\mathbf{\Phi}\mathcal{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\} I\left(\chi_{p_2+2}^{-2}(\Delta)\le p_2-2\right)\right)\\
&\quad -2\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\mathcal{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+4}^{-2}(\Delta)\right\} I\left(\chi_{p_2+4}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad +2\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\mathcal{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\} I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad -(p_2-2)^2\,\mathbf{\Phi}\mathcal{E}\left(\chi_{p_2+2,\alpha}^{-4}(\Delta)\, I\left(\chi_{p_2+2,\alpha}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad -(p_2-2)^2\,\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\mathcal{E}\left(\chi_{p_2+4}^{-4}(\Delta)\, I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad +\mathbf{\Phi}H_{p_2+2}(p_2-2;\Delta) + \boldsymbol{\delta}\boldsymbol{\delta}^{\top}H_{p_2+4}(p_2-2;\Delta)\\
&= \mathbf{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) + 2\boldsymbol{\delta}\boldsymbol{\mu}_{11.2}^{\top}\mathcal{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\} I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad +(p_2-2)\,\sigma^2\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\\
&\quad \times\Big[2\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\, I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad -(p_2-2)\,\mathcal{E}\left(\chi_{p_2+2}^{-4}(\Delta)\, I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\Big]\\
&\quad -\sigma^2\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}H_{p_2+2}(p_2-2;\Delta)\\
&\quad +\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\left[2H_{p_2+2}(p_2-2;\Delta)-H_{p_2+4}(p_2-2;\Delta)\right]\\
&\quad -(p_2-2)\,\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\Big[2\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\, I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad -2\mathcal{E}\left(\chi_{p_2+4}^{-2}(\Delta)\, I\left(\chi_{p_2+4}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad +(p_2-2)\,\mathcal{E}\left(\chi_{p_2+2}^{-4}(\Delta)\, I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\Big].
\end{aligned}
$$

∎

Now, the proof of Theorem 3.3 can be easily obtained by following the definition of asymptotic risk.

ACKNOWLEDGMENTS

The authors thank the Associate Editor and two anonymous referees for their valuable comments that improved the paper. Bahadır Yüzbaşı was supported by The Scientific and Research Council of Turkey under grant Tubitak-Bideb-2214/A during this study at Brock University in Canada, and S. Ejaz Ahmed is supported by the Natural Sciences and the Engineering Research Council of Canada (NSERC).

REFERENCES

[1] Ahmed, S.E. (2001). *Shrinkage estimation of regression coefficients from censored data with multiple observations. In Ahmed, S. E. and Reid, N., (Eds.),* Lecture Notes in Statistics, 148, 103–120. Springer-Verlag, New York.

[2] Ahmed, S.E. (2014a). *Penalty, Shrinkage and Pretest Strategies: Variable Selection and Estimation*, Springer, New York.

[3] Ahmed, S.E. (Ed.) (2014b). *Perspectives on Big Data Analysis: Methodologies and Applications*, (Vol. 622). American Mathematical Society.

[4] Ahmed, S.E.; Hossain, S. and Doksum, K.A. (2012). Lasso and shrinkage estimation in Weibull censored regression models, *Journal of Statistical Inference and Planning*, **12**, 1273–1284.

[5] Ahmed, S.E. and Fallahpour, S. (2012). Shrinkage Estimation Strategy in Quasi-Likelihood Models, *Statistics & Probability Letters*, **82**, 2170–2179.

[6] Belsley, D.A. (1991). *Conditioning diagnostics*, Encyclopedia of Statistical Sciences. 2.

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

[8] Frank, I.E. and Friedman, J.H. (1993). A statistical view of some chemometrics regression tools, *Technometrics*, **35**,109–148.

[9] Gao, X.; Ahmed, S.E. and Feng, Y. (2016) Post selection shrinkage estimation for high-dimensional data analysis. *Applied Stochastic Models in Business and Industry*, doi: 10.1002/asmb.2193.

[10] Hoerl, A.E. and Kennard, R.W. (1970). Ridge Regression: Biased estimation for non-orthogonal problems, *Technometrics*, **12**, 69–82.

[11] Hossain, S.; Ahmed, S.E.; and Doksum, K.A. (2015). Shrinkage, pretest, and penalty estimators in generalized linear models. *Statistical Methodology*, **24**, 52–68.

[12] Judge, G.G. and Bock, M.E. (1978). *The Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics*, North Holland, Amsterdam.

[13] Knight, K. and Fu, W. (2000). Asymptotics for Lasso-Type Estimators, *The Annals of Statistics*, **28**(5), 1356–1378.

[14] McDonald, G.C.. and Schwing, R.C. (1973). Instabilities of regression estimates relating air pollution to mortality. *Technometrics*, **15**(3), 463–481.

[15] Prakash, G. and Singh, D.C. (2009). A Bayesian shrinkage approach in Weibull type-II censored data using prior point information, *REVSTAT*, **7**(2), 171–187.

[16] R Development Core Team (2010). *A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0.

[17] Shanubhogue, A. and Al-Mosawi, R.R. (2010). On estimation following subset selection from truncated Poisson distributions under Stein loss function, *REVSTAT*, **8**(1), 1–20.

[18] Yüzbaşı, B. and Ahmed, S.E. (2015). *Shrinkage Ridge Regression Estimators in High-Dimensional Linear Models*, In Proceedings of the Ninth International Conference on Management Science and Engineering Management, Springer Berlin Heidelberg, 793 – 807

[19] Zou, H. (2006). The adaptive Lasso and its oracle properties, *Journal of the American Statistical Association*, **101**(456), 1418–1429.