Thailand Statistician
July 2023; 21(3): 580-606
http://statassoc.or.th
Contributed paper

Post Improved Estimation and Prediction in the Gamma Regression Model

Pannipa Rintara [a], Syed Ejaz Ahmed [b] and Supranee Lisawadi*[a]
[a] Department of Mathematics and Statistics, Faculty of Science and Technology,
Thammasat University, Pathum Thani, Thailand.
[b] Department of Mathematics and Statistics, Brock University, St. Catharines, Ontario, Canada.
*Corresponding author; e-mail: supranee@mathstat.sci.tu.ac.th

Received: 1 August 2020
Revised: 11 February 2021
Accepted: 26 April 2021

Abstract

In this paper, we consider the estimation of regression coeffcients for a gamma regression model when multicollinearity is present. We suggest pretest and shrinkage estimation strategies based on ridge estimation and compare their performance with some penalty estimators. We investigate the asymptotic properties of the suggested estimators. A Monte Carlo simulation run to evaluate their performance confirmed that the suggested estimators outperformed the unrestricted ridge regression estimator. Finally, a real dataset was analyzed to demonstrate the practicality of the suggested ridge-type and penalty estimators.

Keywords: Ridge, linear shrinkage, shrinkage pretest, LASSO, elastic net.

1. Introduction

Linear regression is the most widely used statistical technique for describing the relationship between predictors and response variables with continuous values, where there is a normal distribution. In areas such as insurance, economics, and medicine, the response variables of interest may be positive, continuous, and right-skewed, so the gamma regression model is often used instead of linear regression. Gamma regression is an important model for generalized linear models (GLMs). Estimating the regression parameters can be done using the maximum likelihood method, which provides unbiased estimators when the predictors are independent. However, in practice it is common for some predictors to be intercorrelated, resulting in multicollinearity. The maximum likelihood estimator (MLE) becomes unstable with extremely high variance, making it inappropriate. To overcome multicollinearity, Hoerl (1962) first proposed ridge regression for linear regression.

In practice, some predictors may or may not influence the response variable and are known respectively as active and inactive predictors. Prediction accuracy can be enhanced by removing inactive predictors from the model. Subspace information can be used to identify these predictors. The source of such information may be the variable selection method, experience of the researcher, or previous studies. Consequently, this information leads to two choices for practitioners, namely a full model with all predictors and a submodel with active predictors. When the subspace information is correct, the submodel estimator can efficiently estimate the regression coefficient, while the full model estimator is inefficient. Conversely, submodel estimator performance degrades while the full model estimator performance is steady. The accuracy of information is typically unknown. Such uncertainties can negatively affect the estimation performance of both the full model and submodel.

The primary motivation for the study comes from the unknown quality of subspace information. To avoid this and improve model prediction, we suggest the pretest and shrinkage estimation strategies, which are a combination of the estimators of the full model and submodel, based ridge regression for the gamma regression model in the presence of multicollinearity. Pretest and shrinkage estimation strategies have been applied to various statistical models, including multiple regression (Ahmed and Yüzbaşı 2016), Poisson regression (Reangsephet et al. 2020b), and gamma regression (Reangsephet et al. 2020a, Mahmoudi et al. 2020). Nevertheless, these previous studies focused on statistical models with uncorrelated predictors. Moreover, their estimators were estimated based on the least squares and maximum likelihood methods. More recently, Algamal and Asar (2020) and Amin et al. (2022) explored estimation based on the gamma model by using other biased estimation methods. Furthermore, the pretest and shrinkage estimators based on ridge regression has been applied to the partially linear model (Yüzbaşı et al. 2020). Ahmed (2014) discussed all the estimators in various contexts. The present literature on parameter estimation with the gamma regression model in the presence of multicollinearity is limited. Therefore, we suggest pretest and shrinkage estimation strategies based on ridge regression for parameter estimation in a gamma regression model when there is multicollinearity and where subspace information is available but there is an unknown degree of uncertainty. We also compare their performance with two penalty estimators, including the least absolute shrinkage and selection operator (LASSO) and elastic net (EN). These select the submodel by shrinking some regression coefficients to zero and provide the estimators of the remaining coefficients.

The plan of this paper is as follows. Section 2 describes the gamma regression model. Section 3 discusses estimation strategies. Section 4 presents the asymptotic results of the suggested estimators. Monte Carlo simulations were conducted to study the performance of the suggested estimators, and the findings reported in Section 5. In Section 6, the estimators are applied to real data. Section 7 provides conclusions for the study. The Appendix gives proofs of the theoretical results.

2. Model Description

In gamma regression, all components of the response variable Y are independent and identically distributed in a gamma distribution $Y_i \sim G(\theta, \theta / \mu_i)$ where $\theta > 0$ is the shape parameter for $i = 1, 2, ..., n.$ The mean and variance of Y_i are $E[Y_i] = \mu_i$ and $V[Y_i] = \mu_i^2 / \theta$. The probability density function is defined as (Cepeda-Cuervo et al. 2016)

$$f(y_i; \theta, \mu_i) = \frac{\theta^\theta}{\mu_i^\theta \Gamma(\theta)} y_i^{\theta-1} e^{-\frac{\theta y_i}{\mu_i}}, \quad y_i > 0,$$

or can be written in the exponential form as

$$f(y_i; \theta, \mu_i) = \exp\left\{ \frac{-\frac{y_i}{\mu_i} + \ln\left(\frac{1}{\mu_i}\right)}{\frac{1}{\theta}} + \theta \ln \theta + (\theta - 1) \ln y_i - \ln \Gamma(\theta) \right\}. \tag{1}$$

Comparing (1) with the general form of the exponential family results in the dispersion parameter $\phi = 1/\theta$ and the canonical link function is the reciprocal function $1/\mu_i = x_i'\boldsymbol{\beta}$. This is the linear combination of the predictors $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)' \in \mathbb{R}^p$ which is a $p \times 1$ vector with unknown regression parameters. For gamma regression, the link function is either used as the canonical function or as a log link function. In this study, we consider the log link function $\ln(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$ since it confirms that $\mu_i > 0$. The corresponding log-likelihood of (1) is then given by

$$\ell(\boldsymbol{\beta}, \theta; y_i) = \sum_{i=1}^{n} \left\{ -\theta\left(y_i e^{-\mathbf{x}_i'\boldsymbol{\beta}} + \mathbf{x}_i'\boldsymbol{\beta}\right) + \theta \ln\theta + (\theta - 1)\ln y_i - \ln\Gamma(\theta) \right\}. \tag{2}$$

The maximum likelihood estimator (MLE) or unrestricted MLE can be obtained by maximizing the log-likelihood in (2) or by solving the following score equation:

$$\frac{\ell(\boldsymbol{\beta}, \theta; y_i)}{\ell(\boldsymbol{\beta})} = \theta \sum_{i=1}^{n} \left\{ y_i e^{-\mathbf{x}_i'\boldsymbol{\beta}} - 1 \right\} \mathbf{x}_i = \mathbf{0}.$$

However, the score equation cannot be directly solved because it is nonlinear in $\boldsymbol{\beta}$, so iteratively reweighted least squares (IRLS) is used to estimate the regression parameters. In the final iteration, the unrestricted MLE of $\boldsymbol{\beta}$ can then be calculated as

$$\hat{\boldsymbol{\beta}}^{\text{UE}} = \left(\mathbf{X}'\hat{\mathbf{W}}_n\mathbf{X}\right)^{-1} \mathbf{X}'\hat{\mathbf{W}}_n\hat{\mathbf{z}}. \tag{3}$$

Here, $\hat{\mathbf{W}}_n = \mathbf{I}$ is the matrix of the weight function, and the ith element of vector $\hat{\mathbf{z}}$ becomes $\hat{z}_i = \mathbf{x}_i'\boldsymbol{\beta} + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}$ and $\hat{\mu}_i = e^{\mathbf{x}_i'\hat{\boldsymbol{\beta}}}$. Next, we assume the regularity conditions as follows:

- $\lim_{n\to\infty} \mathbf{X}'\hat{\mathbf{W}}_n\mathbf{X} = \lim_{n\to\infty} \boldsymbol{C}_n = \boldsymbol{C}$, where $\mathbf{C} = \mathbf{X}'\mathbf{W}\mathbf{X}$ is a finite and positive definite matrix.
- $\max_{1\le i\le n} \mathbf{x}_{ni}^{*\prime}\left(\mathbf{X}'\hat{\mathbf{W}}_n\mathbf{X}\right)\mathbf{x}_{ni}^{*} = o(n)$ as $n \to \infty$, where $\mathbf{x}_{ni}^{*}$ is the i^{th} row of the matrix $\hat{\mathbf{W}}_n^{1/2}\mathbf{X}$.

Then the asymptotic distribution of $\hat{\boldsymbol{\beta}}^{\text{UE}}$ is given by Theorem 1.

Theorem 1 *Under the regularity conditions, as* $n \to \infty$

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{\text{UE}} - \boldsymbol{\beta}\right) \xrightarrow{d} N_p\left(\mathbf{0}, \phi\mathbf{C}^{-1}\right),$$

where $\xrightarrow{d}$ *implies convergence in distribution and* $\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix}$.

3. Estimation Strategies

The subspace information is used to identify the active and inactive predictors, meaning that the design matrix can be partitioned as $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, where $\mathbf{X}_1$ is an $n \times p_1$ submatrix containing the active predictors and $\mathbf{X}_2$ is an $n \times p_2$ submatrix containing the inactive predictors. Similarly, the regression parameter vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ is partitioned into $p_1 \times 1$ and $p_2 \times 1$ subvectors, such that $p_1 + p_2 = p$. With the available subspace information, we are interested in the estimation of the active parameter $\boldsymbol{\beta}_1$ when $\boldsymbol{\beta}_2$ is a known vector. Without loss of generality, $\boldsymbol{\beta}_2$ is set as a zero vector. Hence, the estimators of both the full model and the submodel are dependent on the accuracy of the subspace information.

To describe the suggested estimators, the ridge regression technique in a GLMs is discussed first. For the full model, the unrestricted ridge regression estimator (URRE) of $\boldsymbol{\beta}$ is as follows:

$$\hat{\boldsymbol{\beta}}^{\text{UE}}(\kappa) = \left(\mathbf{X}'\hat{\mathbf{W}}_n\mathbf{X} + \kappa\mathbf{I}_p\right)^{-1}\mathbf{X}'\hat{\mathbf{W}}_n\hat{\mathbf{z}}$$
$$= \mathbf{R}_n(\kappa)\hat{\boldsymbol{\beta}}^{\text{UE}},$$

where $\mathbf{R}_n(\kappa) = [\mathbf{I}_p + \kappa(\mathbf{X}'\hat{\mathbf{W}}_n\mathbf{X})^{-1}]^{-1}$ and $\kappa \geq 0$ is the ridge parameter. If $\kappa = 0,$ then URRE is the unrestricted MLE in (3) and as $\kappa \to \infty,$ URRE becomes zero. Using the inverse of a partitioned matrix, the URRE $\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)$ of $\boldsymbol{\beta}_1$ becomes

$$\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa) = \left(\tilde{\mathbf{X}}_1'\mathbf{M}_2^{\kappa}\tilde{\mathbf{X}}_1 + \kappa\mathbf{I}_{p_1}\right)^{-1}\tilde{\mathbf{X}}_1'\mathbf{M}_2^{\kappa}\tilde{\mathbf{z}},$$

where $\mathbf{M}_2^{\kappa} = \mathbf{I}_n - \tilde{\mathbf{X}}_2[\tilde{\mathbf{X}}_2'\tilde{\mathbf{X}}_2 + \kappa\mathbf{I}_{p_2}]^{-1}\tilde{\mathbf{X}}_2',\ \tilde{\mathbf{X}}_1 = \hat{\mathbf{W}}_n^{1/2}\mathbf{X}_1,\ \tilde{\mathbf{X}}_2 = \hat{\mathbf{W}}_n^{1/2}\mathbf{X}_2,$ and $\tilde{\mathbf{z}} = \hat{\mathbf{W}}_n^{1/2}\hat{\mathbf{z}}.$

If the subspace information about the inactive predictors is available, we know that $\boldsymbol{\beta}_2 = \mathbf{0}.$ We suggest a restricted ridge regression estimator (RRRE) by combining the two approaches underlying restricted maximum likelihood and ridge regression to be the submodel estimator. RRRE is defined in the following form:

$$\hat{\boldsymbol{\beta}}^{\text{RE}}(\kappa) = \mathbf{R}_n(\kappa)\hat{\boldsymbol{\beta}}^{\text{RE}}, \tag{4}$$

where $\hat{\boldsymbol{\beta}}^{\text{RE}}$ is the restricted MLE of $\boldsymbol{\beta}$ which is obtained by maximizing the log-likelihood function in (2) subject to $\boldsymbol{\beta}_2 = \mathbf{0}.$ RRRE $\hat{\boldsymbol{\beta}}_1^{\text{RE}}(\kappa)$ of $\boldsymbol{\beta}_1$ is derived as

$$\hat{\boldsymbol{\beta}}_1^{\text{RE}}(\kappa) = \left[\mathbf{I}_{p_1} + \kappa(\tilde{\mathbf{X}}_1'\mathbf{M}_2^{\kappa}\tilde{\mathbf{X}}_1)^{-1}\right]^{-1}\hat{\boldsymbol{\beta}}_1^{\text{RE}}.$$

Here, $\hat{\boldsymbol{\beta}}_1^{\text{RE}} = (\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{z}}$ is the restricted MLE of $\boldsymbol{\beta}_1.$ RRRE is known to outperform URRE when the subspace information is correct or nearly correct, whereas RRRE becomes inefficient when the subspace information is incorrect, while URRE remains consistent.

3.1. Linear shrinkage ridge regression estimator

In practice, the accuracy of the subspace information is unknown, so using either URRE or RRRE as the estimator of $\boldsymbol{\beta}_1$ may not be a good decision. To improve the estimation of $\boldsymbol{\beta}_1,$ we suggest the linear shrinkage ridge regression estimator (LSRRE), which is a linear combination of URRE and RRRE. LSRRE is derived as follows:

$$\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa) = \pi\hat{\boldsymbol{\beta}}_1^{\text{RE}}(\kappa) + (1-\pi)\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa),$$

where $\pi \in [0, 1]$ denotes the shrinkage intensity, which refers to the degree of confidence in the subspace information. If π is close to one, the subspace information is trustworthy. LSRRE then shrinks URRE toward RRRE. If π is close to zero, the subspace information is not trustworthy. LSRRE then shrinks RRRE toward URRE. Hence, RRRE is a special case of LSRRE when $\pi = 1$ and LSRRE is better than URRE in a meaningful parameter space. The value of π may be assigned using the researcher's confidence in the accuracy of the subspace information.

3.2. Shrinkage pretest ridge regression estimator

Another approach to avoiding the unknown accuracy of the subspace information is to test before it is incorporated into the estimation. Ahmed (1992) proposed a shrinkage pretest ridge regression estimator (SPTRRE), a combination between URRE and LSRRE. SPTRRE of $\boldsymbol{\beta}_1$ can be written as

$$\hat{\boldsymbol{\beta}}_1^{\text{SPTE}}(\kappa)=\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)-(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)-\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa))I(\mathcal{L}_n\le\mathcal{L}_{n,\alpha}),$$

or in alternative form

$$\hat{\boldsymbol{\beta}}_1^{\text{SPTE}}(\kappa)=\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)-\pi(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)-\hat{\boldsymbol{\beta}}_1^{\text{RE}}(\kappa))I(\mathcal{L}_n\le\mathcal{L}_{n,\alpha}),$$

where $I(\cdot)$ denotes the indicator function and $\mathcal{L}_n$ is a general test statistic to test the hypothesis $H_0:\boldsymbol{\beta}_2=\mathbf{0}$ against $H_1:\boldsymbol{\beta}_2\neq\mathbf{0}$. The log-likelihood ratio test statistic is suggested (Reangsephet et al. 2020a):

$$\mathcal{L}_n=2\left(\ell(\hat{\boldsymbol{\beta}}^{\text{UE}};y_1,y_2,\ldots,y_n)-\ell(\hat{\boldsymbol{\beta}}^{\text{RE}};y_1,y_2,\ldots,y_n)\right). \quad (5)$$

Under H_0, $\mathcal{L}_n$ follows the chi-square distribution with p_2 degrees of freedom as $n\to\infty$. Thus, $\mathcal{L}_{n,\alpha}$ is the upper α-level critical value of the chi-square distribution with p_2 degrees of freedom.

It is evident that the pretest ridge regression estimator (PTRRE), which follows Bancroft (1944), is a special case of SPTRRE when $\pi=1$. The PTRRE of $\boldsymbol{\beta}_1$ is given as

$$\hat{\boldsymbol{\beta}}_1^{\text{PTE}}(\kappa)=\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)-(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)-\hat{\boldsymbol{\beta}}_1^{\text{RE}}(\kappa))I(\mathcal{L}_n\le\mathcal{L}_{n,\alpha}),$$

It is important to note that SPTRRE and PTRRE are bounded and perform better than URRE in a part of the parameter space.

3.3. Shrinkage ridge regression estimator

SPTRRE has two extreme choices, URRE and LSRRE. We address this limitation by defining the shrinkage ridge regression estimator (SRRE), which combines URRE and LSRRE effectively with the optimal weight. The SRRE of $\boldsymbol{\beta}_1$ is defined

$$\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)=\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa)+(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)-\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa))(1-(p_2-2)\mathcal{L}_n^{-1}),\quad p_2\ge3,$$

or, equivalently,

$$\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)=\hat{\boldsymbol{\beta}}_1^{\text{RE}}(\kappa)+(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)-\hat{\boldsymbol{\beta}}_1^{\text{RE}}(\kappa))(1-\pi(p_2-2)\mathcal{L}_n^{-1}),\quad p_2\ge3,$$

where $\mathcal{L}_n$ is defined in (5). The estimator occasionally suffers from over-shrinkage which occurs when $\mathcal{L}_n$ is small, so that $\mathcal{L}_n<p_2-2$ and the shrinkage factor $1-(\text{p}_2-2)\mathcal{L}_n^{-1}$ becomes negative, resulting in the sign of the estimator reversing. To remedy this, a positive shrinkage estimator is defined by retaining the positive part of the shrinkage factor.

The positive-part shrinkage ridge regression estimator (PSRRE) $\hat{\boldsymbol{\beta}}_1^{\text{PSE}}(\kappa)$ of $\boldsymbol{\beta}_1$ is defined as

$$\hat{\boldsymbol{\beta}}_1^{\text{PSE}}(\kappa)=\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa)-(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)-\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa))(1-(p_2-2)\mathcal{L}_n^{-1})^+,\quad p_2\ge3.$$

Here $a^+=\max(0,a)$. Alternatively, PSRRE can be written in the following form

$$\hat{\boldsymbol{\beta}}_1^{\text{PSE}}(\kappa)=\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)-\pi(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)-\hat{\boldsymbol{\beta}}_1^{\text{RE}}(\kappa))(1-(p_2-2)\mathcal{L}_n^{-1})I(\mathcal{L}_n\le\mathcal{L}_{n,\alpha}),\quad p_2\ge3.$$

When $\pi=1$, SRRE and PSRRE are the original James-Stein approach and it is considered in this study.

As can be seen from these estimators, the value of ridge parameter κ is important. Amin et al. (2020) investigated various techniques for estimating κ under the gamma regression model. The following ridge estimator is suggested in this paper:

$$\kappa = \prod_{j=1}^{p} \left(\frac{1}{m_j} \right)^{1/p},$$

where $m_j = \sqrt{\frac{\hat{\phi}}{\hat{\alpha}_j^2}}$, $\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} \left(\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right)^2$, and $\hat{\alpha}_j$ is the j^{th} element of $\boldsymbol{\gamma}'\hat{\boldsymbol{\beta}}^{\text{UE}}$, where $\boldsymbol{\gamma}$ is the eigenvector of $\mathbf{X}'\hat{\mathbf{W}}_n\mathbf{X}$. They reported that κ had the lowest mean square error.

3.4. Penalty estimators

Penalized methods for GLMs are based on penalized log-likelihood optimization. The coefficient estimates $\hat{\boldsymbol{\beta}}$ can be presented as

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \{ \ell(\boldsymbol{\beta}, \theta; y_i) - \text{P}_\lambda(\boldsymbol{\beta}) \},$$

where $\text{P}_\lambda(\boldsymbol{\beta})$ is the penalty function and λ is a nonnegative tuning parameter. The penalty function is derived as

$$\text{P}_\lambda(\boldsymbol{\beta}) = \frac{1-\lambda}{2} \sum_{j=1}^{p} \boldsymbol{\beta}_j^2 + \lambda \sum_{j=1}^{p} |\boldsymbol{\beta}_j|.$$

This is the least absolute shrinkage and selection operator estimator (LASSO) when $\lambda = 1$ and the elastic net (EN) when $\lambda \in [0, 1]$.

4. Asymptotic Properties

We are interested in the asymptotic behavior of the suggested estimators. To obtain the meaningful asymptotic, a sequence of local alternatives $H_{(n)}$ is considered:

$$H_{(n)} : \boldsymbol{\beta}_2 = \frac{\boldsymbol{\omega}}{\sqrt{n}},$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_{p_2})' \in \mathbb{R}^{p_2}$ is a $p_2 \times 1$ fixed vector. The vector $\boldsymbol{\omega}/\sqrt{n}$ is a measure of the divergence between the null hypothesis $\boldsymbol{\beta}_2 = \mathbf{0}_{p_2}$ and the sequence of local alternatives. Thus, the null hypothesis is a special case of $H_{(n)}$ when $\boldsymbol{\omega} = \mathbf{0}$.

We define the asymptotic distributional bias (ADB) of any estimator $\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa)$ as $ADB(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa)) = \lim_{n\to\infty} \text{E}[\sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa) - \boldsymbol{\beta}_1)]$. The asymptotic distributional risk (ADR) of any estimator $\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa)$ is given as $R(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa)) = tr\left(\mathbf{Q}\Gamma(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa))\right)$, where $tr(\cdot)$ is the trace of a matrix, $\mathbf{Q}$ is a positive-definite matrix and $\Gamma(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa))$ is the asymptotic mean squared error matrix (AMSEM) of any estimator $\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa)$, defined as $\Gamma(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa)) = \lim_{n\to\infty} \text{E}[n(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa) - \boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa) - \boldsymbol{\beta}_1)']$.

As is already known, RRRE and PTRRE are the special cases of LSRRE and SPTRRE, respectively, when $\pi = 1$ so only the asymptotic properties of URRE, LSRRE, SPTRRE, SRRE, and PSRRE are reported.

Theorem 2 *Under the local alternatives $H_{(n)}$ and the regularity conditions, as $n \to \infty$, the ADBs of the suggested estimators become*

$$ADB(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) = -\boldsymbol{\mu}_{11.2},$$
$$ADB(\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa)) = -\boldsymbol{\mu}_{11.2} - \pi\boldsymbol{\delta},$$
$$ADB(\hat{\boldsymbol{\beta}}_1^{\text{SPTE}}(\kappa)) = -\boldsymbol{\mu}_{11.2} - \pi\boldsymbol{\delta}G_{c+4}(\chi^2_{c+2,\alpha};\Delta),$$
$$ADB(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)) = -\boldsymbol{\mu}_{11.2} - c\boldsymbol{\delta}\text{E}[\chi^{-2}_{c+4}(\Delta)],$$
$$ADB(\hat{\boldsymbol{\beta}}_1^{\text{PSE}}(\kappa)) = -\boldsymbol{\mu}_{11.2} - \boldsymbol{\delta}D(\Delta),$$

where $D(\Delta) = c\text{E}[\chi^{-2}_{c+4}(\Delta)] + G_{c+4}(c;\Delta) - c\text{E}[\chi^{-2}_{c+4}(\Delta)I(\chi^2_{c+4}(\Delta) \le c)]$, $c = p_2 - 2$, $\boldsymbol{\delta} = -\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\boldsymbol{\omega}$, $\boldsymbol{\mu}_{11.2} = \kappa_0\mathbf{C}_{11.2}^{-1}(\boldsymbol{\beta} - \boldsymbol{\omega}/\sqrt{n}\,\mathbf{C}_{12}\mathbf{C}_{22}^{-1})$, *and* $\mathbf{C}_{11.2} = \mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}$. $G_q(x;\Delta)$ *denotes the cumulative distribution function of a non-central chi-square distribution with q degrees of freedom and non-central parameter* $\Delta = \phi^{-1}(\boldsymbol{\omega}'\mathbf{C}_{22.1}\boldsymbol{\omega})$ *and* $E[\chi_q^{-2j}(\Delta)] = \int_o^\infty x^{-2j}dG_q(\text{x},\Delta)$ *denotes the expectation of the reciprocal of the non-central chi-square distribution with q degrees of freedom and non-central parameter* Δ.

Proof: See Appendix A.

To make the comparison more meaningful and clear, the ADBs were transformed to the scalar form of the asymptotic distributional quadratic bias (ADQB):

$$ADQB(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa)) = [ADB(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa))]'\mathbf{C}_{11.2}[ADB(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa))].$$

The ADQBs of the suggested estimators are expressed by the following theorem.

Theorem 3 *Under the local alternatives* $H_{(n)}$ *and the regularity conditions, as* $n \to \infty$, *the ADQBs of the suggested estimators become*

$$ADQB(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) = \boldsymbol{\mu}_{11.2}{}'\mathbf{C}_{11.2}\boldsymbol{\mu}_{11.2},$$
$$ADQB(\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa)) = ADQB(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) + 2\pi\boldsymbol{\delta}'\mathbf{C}_{11.2}\boldsymbol{\mu}_{11.2} + \pi^2\boldsymbol{\delta}'\mathbf{C}_{11.2}\boldsymbol{\delta},$$
$$ADQB(\hat{\boldsymbol{\beta}}_1^{\text{SPTE}}(\kappa)) = ADQB(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) + 2\pi\boldsymbol{\delta}'\mathbf{C}_{11.2}\boldsymbol{\mu}_{11.2}G_{c+4}(\chi^2_{c+2,\alpha};\Delta) + \pi^2\boldsymbol{\delta}'\mathbf{C}_{11.2}\boldsymbol{\delta}\left(G_{c+4}(\chi^2_{c+2,\alpha};\Delta)\right)^2,$$
$$ADQB(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)) = ADQB(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) + 2c\boldsymbol{\delta}'\mathbf{C}_{11.2}\boldsymbol{\mu}_{11.2}\text{E}[\chi^{-2}_{c+4}(\Delta)] + c^2\boldsymbol{\delta}'\mathbf{C}_{11.2}\boldsymbol{\delta}\left(\text{E}[\chi^{-2}_{c+4}(\Delta)]\right)^2,$$
$$ADQB(\hat{\boldsymbol{\beta}}_1^{\text{PSE}}(\kappa)) = ADQB(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) + 2\boldsymbol{\delta}'\mathbf{C}_{11.2}\boldsymbol{\mu}_{11.2}D(\Delta) + \boldsymbol{\delta}'\mathbf{C}_{11.2}\boldsymbol{\delta}D^2(\Delta).$$

The proof of this theorem is skipped as it was directly proved by Theorem 2. Clearly, the ADQBs of all estimators are equal under the null hypothesis. For fixed π, it can be observed that the ADQB of SPTRRE is equal to that of LSRRE when $\alpha = 0$. The ADQBs of URRE and SPTRRE are equal when $\alpha = 1$. For a fixed α, the ADQBs of SPTRRE, SRRE, and PSRRE converge on the ADQB of URRE as $\Delta \to \infty$, because $G_{c+4}(\chi^2_{c+2,\alpha};\Delta)$ and $\text{E}[\chi^{-2}_{c+4}(\Delta)]$ are decreasing functions of Δ while the ADQB of LSRRE becomes unbounded. The following theorem concerns the ADR of the estimators.

Theorem 4 *Suppose that* $\mathbf{Q}$ *is a positive-definite matrix, then under the local alternatives* $H_{(n)}$ *and the regularity conditions, as* $n \to \infty$, *the ADRs of the suggested estimators become*

$$
\begin{aligned}
R(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) &= \phi\, tr(\mathbf{QC}_{11.2}^{-1}) + \boldsymbol{\mu}_{11.2}'\mathbf{Q}\boldsymbol{\mu}_{11.2},\\
R(\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa)) &= R(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) - \pi(2-\pi)\, tr(\mathbf{Q\Omega}) + 2\pi\boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\mu}_{11.2} + \pi^2\boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\delta},\\
R(\hat{\boldsymbol{\beta}}_1^{\text{SPTE}}(\kappa)) &= R(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) + 2\pi\boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\mu}_{11.2}G_{c+4}(\chi^2_{c+2,\alpha};\Delta) - \pi(2-\pi)\, tr(\mathbf{Q\Omega})G_{c+4}(\chi^2_{c+2,\alpha};\Delta)\\
&\quad + \boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\delta}\left\{2\pi G_{c+4}(\chi^2_{c+2,\alpha};\Delta) - \pi(2-\pi)G_{c+6}(\chi^2_{c+2,\alpha};\Delta)\right\},\\
R(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)) &= R(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) + 2c\boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\mu}_{11.2}\text{E}[\chi^{-2}_{c+4}(\Delta)] - c\, tr(\mathbf{Q\Omega})\left\{2\text{E}[\chi^{-2}_{c+4}(\Delta)] - c\text{E}[\chi^{-4}_{c+4}(\Delta)]\right\}\\
&\quad + c(c+4)\boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\delta}\text{E}[\chi^{-4}_{c+6}(\Delta)],\\
R(\hat{\boldsymbol{\beta}}_1^{\text{PSE}}(\kappa)) &= R(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)) + 2\boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\mu}_{11.2}\text{E}[(1-c\chi^{-2}_{c+4}(\Delta))I(\chi^2_{c+4}(\Delta)\le c)]\\
&\quad - tr(\mathbf{Q\Omega})\left\{2\text{E}[(1-c\chi^{-2}_{c+4}(\Delta))I(\chi^2_{c+4}(\Delta)\le c)]\right.\\
&\quad \left. + c^2\text{E}[\chi^{-4}_{c+4}(\Delta)I(\chi^2_{c+4}(\Delta)\le c)] + G_{c+4}(c;\Delta)\right\} + \boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\delta}\left\{\text{E}[(1-c\chi^{-2}_{c+4}(\Delta))I(\chi^2_{c+4}(\Delta)\le c)]\right.\\
&\quad \left. - 2\text{E}[(1-c\chi^{-2}_{c+6}(\Delta))I(\chi^2_{c+6}(\Delta)\le c)] - c^2\text{E}[\chi^{-4}_{c+6}(\Delta)I(\chi^2_{c+6}(\Delta)\le c)] + G_{c+6}(c;\Delta)\right\}.
\end{aligned}
$$

Proof: See Appendix B.

It can be seen that the ADR of LSRRE is an unbounded function of Δ. When $\Delta = 0$, all suggested estimators are superior to URRE. Nevertheless, the ADR of SPTRRE approaches the ADR of URRE when $\Delta \to \infty$. It can be seen that $R(\hat{\boldsymbol{\beta}}_1^{\text{PSE}}(\kappa)) \le R(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)) \le R(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa))$ for all values of Δ with $p_2 \ge 3$.

5. Simulation Analysis

A Monte Carlo simulation was conducted to evaluate the performance of the suggested estimators. All calculations were run on R software. The response variable Y_i was generated with sample sizes of n = 100 and 150, using a random number from a gamma distribution $Y_i \sim G(\theta, \theta/\mu_i)$ such that $\mu_i = e^{\mathbf{x}_i'\boldsymbol{\beta}}$. The predictor $\mathbf{x}_i$ was generated using the following formula:

$$x_{ij} = (1-\rho^2)^{1/2} h_{ij} + \rho h_{i(p+1)}, \quad j = 1, 2, \ldots, p.$$

Here, ρ^2 represents the correlation between the predictors and h_{ij} are independent standard normal random numbers. The factors were varied, setting ρ^2 = 0.5, 0.7, 0.9, 0.95, and 0.99, and θ = 0.5 and 1.5. The number of predictors were $(p_1, p_2) \in \{(3, 4), (3, 7), (3, 12), (3, 19)\}$. Three levels of shrinkage intensity were set to π = 0.25, 0.50, and 0.75. The significance level α=0.05. The multicollinearity diagnostic tool for GLMs is a condition number (CN) (Weissfeld and Sereika 1991). If the CN—which is calculated from the ratio of the largest eigenvalue to the smallest eigenvalue of the Fisher information matrix—is greater than 30, then the presence of multicollinearity is implied.

To assess the behavior of the suggested estimators, we first defined $\Delta^* = \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\|$ as the divergence between the simulated model and the submodel under the null hypothesis $\boldsymbol{\beta}_2 = \mathbf{0}_{p_2}$. Here the simulated parameter such that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, such that $\boldsymbol{\beta}_2 = (\Delta^*, \mathbf{0}_{p_2-1}')$, the submodel parameter $\boldsymbol{\beta}^0 = (\boldsymbol{\beta}_1', \mathbf{0}_{p_2})'$, and $\|\cdot\|$ is the Euclidean norm. Two cases were studied. For the first case, it was assumed that $\Delta^*=0$. We then considered the regression coefficient for the simulated model as $\boldsymbol{\beta}=((1.6,-1.2,-0.1),\mathbf{0}_{p_2})'$ and θ=0.5. In the second case, it was assumed that $\Delta^*\geq 0$. The regression coefficient was set to $\boldsymbol{\beta}=((0.75,-0.2,1.68),\Delta^*,\mathbf{0}'_{p_2-1})'$ where θ=1.5.

The simulation was iterated 2,000 times, which was sufficient to provide stable results. The performance of the estimator $\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa)$ was measured by comparing its MSE with that of URRE:

$$RMSE(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa),\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa))=\frac{MSE(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa))}{MSE(\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa))}.$$

Keep in mind that, if RMSE is greater than one, the estimator $\hat{\boldsymbol{\beta}}_1^{\circ}(\kappa)$ is superior to $\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)$.

5.1. Case I: $\Delta^*=0$

In this case, the subspace information $\boldsymbol{\beta}_2=\mathbf{0}$ was assumed to be correct. The penalty estimators can then be expected to estimate $\boldsymbol{\beta}_1$ by selecting the optimal value of λ, which shrinks many elements in $\boldsymbol{\beta}_2$ to exactly zero. We therefore compared the performance of the suggested estimators with the MLE and the penalty estimators LASSO and EN. The cv. HDtweedie function in the HDtweedie package was used to estimate the penalty estimators. The optimal value of λ was selected by minimizing the mean square error using 5-fold cross validation.

Tables 1 and 2 report the results for $(p_1,p_2)\in\{(3,4),(3,7),(3,12),(3,19)\}$, $\theta=0.5$, and $n=$ 100 and 150. The CN was large when ρ^2 and p_2 were large due to the degree of multicollinearity. As expected, it can be concluded that the performance of MLE was worse than the suggested estimators when multicollinearity exists. The RMSEs of the suggested estimators increased when ρ^2 increased from a moderate $(\rho^2=0.5)$ to a high $(\rho^2=0.9)$ level for fixed p_2 and n. However, their RMSEs decreased when ρ^2 increased to a severe $(\rho^2=0.99)$ level, yet the suggested estimators still outperformed URRE. Moreover, the RMSEs of the suggested estimators also increased as p_2 increased for fixed p_1 and n. RRRE outshone all the other estimators for all cases considered in the simulation study. The RMSE of LSRRE and SPTRRE increased as π increased. LSRRE dominated some other estimators when π was large. PTRRE outperformed SRRE and PSRRE when p_1 was close to p_2, but this reversed at large values of p_2. The penalty estimators performed well when p_2 was large, but the suggested estimators performed better than the penalty estimators.

5.2. Case I: $\Delta^*\geq 0$

We considered the case in which the subspace information was assumed to be either correct or incorrect. The penalty estimators were omitted because they do not take advantage of the fact that the regression coefficients of inactive predictor $\boldsymbol{\beta}_2=\mathbf{0}$. Tables 3 and 4 report the results for (p_1,p_2) = $(3,19)$ when $\pi=0.25,\theta=1.5,$ and $n=100$ and 150. To simplify the comparison, Figures 1 and 2 present graphics for all combinations of $p_2,\rho,$ and n.

Table 1 RMSEs of the suggested estimators with respect to $\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)$ when $\Delta^* = 0, \theta = 0.5,$ and $n = 100$

ρ^2	Estimator		Number of inactive predictors (p_2)			
			4	7	12	19
0.5	CN		12.593	20.471	37.262	72.145
	MLE		0.980	0.984	0.982	0.981
	RRRE		1.314	1.436	1.740	2.145
	LSRRE	$\pi = 0.25$	1.117	1.157	1.229	1.313
		$\pi = 0.50$	1.219	1.302	1.468	1.685
		$\pi = 0.75$	1.289	1.404	1.663	2.020
	SPTRRE	$\pi = 0.25$	1.094	1.134	1.197	1.264
		$\pi = 0.50$	1.173	1.252	1.394	1.553
		$\pi = 0.75$	1.226	1.332	1.547	1.792
	PTRRE		1.244	1.355	1.607	1.875
	SRRE		1.139	1.276	1.537	1.904
	PSRRE		1.175	1.327	1.593	1.978
	LASSO		1.026	1.052	1.083	1.149
	EN		1.016	1.034	1.062	1.117
0.7	CN		27.130	45.061	83.418	163.199
	MLE		0.968	0.972	0.971	0.967
	RRRE		1.379	1.515	1.826	2.264
	LSRRE	$\pi = 0.25$	1.137	1.178	1.248	1.332
		$\pi = 0.50$	1.260	1.348	1.516	1.741
		$\pi = 0.75$	1.347	1.473	1.739	2.121
	SPTRRE	$\pi = 0.25$	1.110	1.151	1.214	1.279
		$\pi = 0.50$	1.205	1.288	1.433	1.593
		$\pi = 0.75$	1.269	1.383	1.607	1.859
	PTRRE		1.291	1.412	1.675	1.952
	SRRE		1.161	1.325	1.592	1.985
	PSRRE		1.206	1.381	1.658	2.074
	LASSO		1.033	1.063	1.103	1.176
	EN		1.025	1.046	1.084	1.151
0.9	CN		99.869	168.094	314.304	618.428
	MLE		0.907	0.914	0.905	0.889
	RRRE		1.429	1.573	1.873	2.310
	LSRRE	$\pi = 0.25$	1.152	1.192	1.259	1.340
		$\pi = 0.50$	1.292	1.380	1.543	1.763
		$\pi = 0.75$	1.393	1.522	1.781	2.160
	SPTRRE	$\pi = 0.25$	1.122	1.163	1.223	1.288
		$\pi = 0.50$	1.229	1.315	1.455	1.616
		$\pi = 0.75$	1.303	1.424	1.639	1.898
	PTRRE		1.328	1.460	1.709	1.998
	SRRE		1.177	1.354	1.621	2.015
	PSRRE		1.230	1.418	1.695	2.114
	LASSO		1.023	1.060	1.124	1.212
	EN		1.043	1.076	1.153	1.245

Table 1 (Continued)

ρ^2		Estimator	Number of inactive predictors (p_2)			
			4	7	12	19
0.95	CN		208.994	352.678	660.681	1301.195
	MLE		0.823	0.829	0.807	0.771
	RRRE		1.432	1.564	1.840	2.225
	LSRRE	$\pi = 0.25$	1.153	1.190	1.253	1.327
		$\pi = 0.50$	1.294	1.375	1.528	1.725
		$\pi = 0.75$	1.395	1.514	1.756	2.091
	SPTRRE	$\pi = 0.25$	1.124	1.162	1.220	1.281
		$\pi = 0.50$	1.232	1.311	1.445	1.598
		$\pi = 0.75$	1.306	1.419	1.624	1.868
	PTRRE		1.311	1.455	1.690	1.961
	SRRE		1.182	1.347	1.600	1.956
	PSRRE		1.232	1.414	1.675	2.052
	LASSO		0.998	1.041	1.110	1.197
	EN		1.056	1.098	1.196	1.303
0.99	CN		1082.055	1829.523	3431.927	6762.799
	MLE		0.425	0.410	0.357	0.298
	RRRE		1.304	1.354	1.464	1.555
	LSRRE	$\pi = 0.25$	1.114	1.131	1.165	1.192
		$\pi = 0.50$	1.213	1.248	1.318	1.379
		$\pi = 0.75$	1.280	1.328	1.428	1.515
	SPTRRE	$\pi = 0.25$	1.095	1.116	1.148	1.172
		$\pi = 0.50$	1.175	1.216	1.282	1.333
		$\pi = 0.75$	1.228	1.283	1.377	1.447
	PTRRE		1.246	1.304	1.408	1.479
	SRRE		1.131	1.223	1.350	1.462
	PSRRE		1.178	1.277	1.398	1.507
	LASSO		0.842	0.889	0.937	0.931
	EN		0.959	1.016	1.104	1.122

Table 2 RMSEs of the suggested estimators with respect to $\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)$ when $\Delta^* = 0, \theta = 0.5,$ and $n = 150$

ρ^2		Estimator	Number of inactive predictors (p_2)			
			4	7	12	19
0.5	CN		11.544	17.874	30.755	54.877
	MLE		0.988	0.990	0.990	0.991
	RRRE		1.271	1.402	1.538	1.826
	LSRRE	$\pi = 0.25$	1.101	1.143	1.186	1.243
		$\pi = 0.50$	1.187	1.273	1.366	1.507
		$\pi = 0.75$	1.247	1.367	1.497	1.730
	SPTRRE	$\pi = 0.25$	1.084	1.121	1.164	1.206
		$\pi = 0.50$	1.154	1.227	1.316	1.416
		$\pi = 0.75$	1.202	1.302	1.425	1.583
	PTRRE		1.221	1.330	1.459	1.652
	SRRE		1.112	1.254	1.410	1.664
	PSRRE		1.146	1.295	1.464	1.712
	LASSO		1.038	1.049	1.089	1.108
	EN		1.023	1.024	1.063	1.076

Table 2 (Continued)

ρ^2	Estimator		Number of inactive predictors (p_2)			
			4	7	12	19
0.7	CN		24.920	39.400	68.944	124.399
	MLE		0.982	0.983	0.984	0.984
	RRRE		1.332	1.476	1.620	1.933
	LSRRE	$\pi = 0.25$	1.121	1.165	1.208	1.264
		$\pi = 0.50$	1.228	1.320	1.414	1.559
		$\pi = 0.75$	1.304	1.434	1.570	1.818
	SPTRRE	$\pi = 0.25$	1.097	1.140	1.184	1.224
		$\pi = 0.50$	1.180	1.267	1.359	1.459
		$\pi = 0.75$	1.235	1.358	1.488	1.651
	PTRRE		1.255	1.392	1.528	1.732
	SRRE		1.142	1.295	1.468	1.744
	PSRRE		1.179	1.345	1.530	1.797
	LASSO		1.044	1.064	1.111	1.137
	EN		1.023	1.040	1.082	1.102
0.9	CN		91.806	147.047	259.946	472.056
	MLE		0.947	0.947	0.949	0.948
	RRRE		1.379	1.535	1.679	2.006
	LSRRE	$\pi = 0.25$	1.136	1.181	1.222	1.277
		$\pi = 0.50$	1.259	1.356	1.449	1.594
		$\pi = 0.75$	1.347	1.487	1.623	1.879
	SPTRRE	$\pi = 0.25$	1.110	1.149	1.199	1.235
		$\pi = 0.50$	1.204	1.285	1.393	1.486
		$\pi = 0.75$	1.269	1.384	1.538	1.695
	PTRRE		1.292	1.420	1.582	1.784
	SRRE		1.161	1.328	1.511	1.798
	PSRRE		1.202	1.384	1.577	1.857
	LASSO		1.047	1.075	1.137	1.182
	EN		1.039	1.073	1.128	1.179
0.95	CN		192.124	308.515	546.474	993.553
	MLE		0.895	0.894	0.893	0.887
	RRRE		1.386	1.541	1.679	1.994
	LSRRE	$\pi = 0.25$	1.138	1.183	1.222	1.275
		$\pi = 0.50$	1.263	1.360	1.449	1.589
		$\pi = 0.75$	1.353	1.493	1.624	1.869
	SPTRRE	$\pi = 0.25$	1.112	1.149	1.199	1.233
		$\pi = 0.50$	1.209	1.286	1.393	1.482
		$\pi = 0.75$	1.276	1.385	1.538	1.688
	PTRRE		1.299	1.420	1.583	1.776
	SRRE		1.164	1.331	1.510	1.791
	PSRRE		1.207	1.389	1.578	1.850
	LASSO		1.014	1.045	1.108	1.175
	EN		1.038	1.081	1.152	1.233

Table 2 (Continued)

ρ^2	Estimator		Number of inactive predictors (p_2)			
			4	7	12	19
0.99	CN		994.583	1600.232	2838.809	5165.566
	MLE		0.581	0.561	0.533	0.489
	RRRE		1.327	1.434	1.507	1.677
	LSRRE	$\pi = 0.25$	1.121	1.155	1.180	1.212
		$\pi = 0.50$	1.227	1.298	1.350	1.430
		$\pi = 0.75$	1.301	1.399	1.472	1.605
	SPTRRE	$\pi = 0.25$	1.101	1.131	1.162	1.186
		$\pi = 0.50$	1.187	1.246	1.311	1.369
		$\pi = 0.75$	1.245	1.325	1.415	1.512
	PTRRE		1.264	1.351	1.445	1.570
	SRRE		1.143	1.268	1.384	1.553
	PSRRE		1.186	1.325	1.441	1.598
	LASSO		0.891	0.948	0.980	1.060
	EN		0.994	1.089	1.126	1.226

Table 3 RMSEs of the proposed estimators with respect to $\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)$ when $(p_1, p_2) = (3, 19), \pi = 0.25, \theta = 1.5,$ and $n = 100$

ρ^2	CN	Δ^*	RRRE	LSRRE	SPTRRE	PTRRE	SRRE	PSRRE
0.5	71.907	0.00	1.707	1.218	1.149	1.423	1.544	1.568
		0.15	1.524	1.219	1.123	1.256	1.439	1.462
		0.30	1.096	1.202	1.063	1.017	1.244	1.254
		0.45	0.736	1.170	1.007	0.932	1.126	1.128
		0.60	0.500	1.123	0.998	0.975	1.071	1.071
		0.75	0.351	1.066	1.000	0.995	1.045	1.045
		1.00	0.211	0.955	1.000	1.000	1.027	1.027
		1.25	0.137	0.835	1.000	1.000	1.020	1.020
		1.50	0.095	0.718	1.000	1.000	1.016	1.016
		1.75	0.068	0.611	1.000	1.000	1.014	1.014
		2.00	0.051	0.518	1.000	1.000	1.011	1.011
0.7	162.666	0.00	1.791	1.236	1.162	1.470	1.602	1.629
		0.15	1.656	1.240	1.146	1.344	1.527	1.555
		0.30	1.261	1.230	1.092	1.085	1.329	1.350
		0.45	0.886	1.208	1.035	0.931	1.180	1.188
		0.60	0.621	1.173	1.001	0.920	1.099	1.100
		0.75	0.446	1.129	0.997	0.966	1.058	1.058
		1.00	0.275	1.038	1.000	0.997	1.031	1.031
		1.25	0.183	0.937	1.000	1.000	1.022	1.022
		1.50	0.128	0.832	1.000	1.000	1.019	1.019
		1.75	0.094	0.732	1.000	1.000	1.018	1.018
		2.00	0.071	0.639	1.000	1.000	1.016	1.016

Table 3 (Continued)

ρ^2	CN	Δ^*	RRRE	LSRRE	SPTRRE	PTRRE	SRRE	PSRRE
0.9	616.522	0.00	1.808	1.241	1.165	1.479	1.615	1.645
		0.15	1.821	1.251	1.163	1.453	1.627	1.655
		0.30	1.660	1.256	1.147	1.324	1.540	1.571
		0.45	1.403	1.256	1.116	1.158	1.414	1.441
		0.60	1.137	1.249	1.080	1.019	1.299	1.318
		0.75	0.907	1.238	1.044	0.929	1.211	1.222
		1.00	0.625	1.207	1.004	0.904	1.122	1.124
		1.25	0.443	1.166	0.996	0.954	1.078	1.078
		1.50	0.326	1.116	0.999	0.986	1.058	1.058
		1.75	0.247	1.057	1.000	0.998	1.048	1.048
		2.00	0.193	0.996	1.000	1.000	1.043	1.043
0.95	1,297.303	0.00	1.726	1.224	1.155	1.444	1.558	1.588
		0.50	1.617	1.258	1.141	1.288	1.514	1.545
		1.00	1.020	1.259	1.058	0.960	1.269	1.283
		1.50	0.605	1.230	1.001	0.918	1.146	1.147
		2.00	0.382	1.177	0.998	0.979	1.100	1.100
		2.50	0.258	1.106	1.000	0.998	1.082	1.082
		3.00	0.186	1.024	1.000	1.000	1.072	1.072
		3.50	0.139	0.937	1.000	1.000	1.064	1.064
		4.00	0.108	0.851	1.000	1.000	1.057	1.057
		5.00	0.072	0.697	1.000	1.000	1.043	1.043
		6.00	0.052	0.572	1.000	1.000	1.030	1.030
0.99	6,743.479	0.00	1.260	1.179	1.133	1.189	1.211	1.223
		0.50	1.434	1.261	1.177	1.279	1.331	1.338
		1.00	1.543	1.327	1.192	1.290	1.389	1.398
		1.50	1.537	1.275	1.166	1.215	1.388	1.400
		2.00	1.454	1.200	1.118	1.114	1.354	1.365
		2.50	1.308	1.199	1.067	1.033	1.312	1.318
		3.00	1.145	1.179	1.023	0.986	1.273	1.275
		3.50	0.989	1.142	1.002	0.974	1.238	1.239
		4.00	0.854	1.091	0.997	0.980	1.207	1.207
		5.00	0.651	1.074	0.999	0.997	1.158	1.158
		6.00	0.512	0.952	1.000	1.000	1.120	1.120

Table 4 RMSEs of the proposed estimators with respect to $\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)$ when $(p_1, p_2) = (3, 19), \pi = 0.25, \theta = 1.5,$ and $n = 150$

ρ^2	CN	Δ^*	RRRE	LSRRE	SPTRRE	PTRRE	SRRE	PSRRE
0.5	54.959	0.00	1.636	1.205	1.140	1.385	1.494	1.519
		0.15	1.346	1.200	1.105	1.157	1.333	1.353
		0.30	0.851	1.169	1.019	0.920	1.129	1.133
		0.45	0.523	1.116	0.997	0.959	1.044	1.045
		0.60	0.337	1.046	0.999	0.994	1.017	1.017
		0.75	0.230	0.966	1.000	1.000	1.010	1.010
		1.00	0.135	0.824	1.000	1.000	1.008	1.008
		1.25	0.087	0.688	1.000	1.000	1.009	1.009
		1.50	0.060	0.567	1.000	1.000	1.009	1.009
		1.75	0.043	0.466	1.000	1.000	1.009	1.009
		2.00	0.033	0.384	1.000	1.000	1.008	1.008

Table 4 (Continued)

ρ^2	CN	Δ^*	RRRE	LSRRE	SPTRRE	PTRRE	SRRE	PSRRE
0.7	124.592	0.00	1.726	1.226	1.155	1.437	1.558	1.588
		0.15	1.485	1.225	1.128	1.241	1.425	1.452
		0.30	0.997	1.203	1.054	0.958	1.193	1.206
		0.45	0.637	1.161	1.003	0.883	1.066	1.068
		0.60	0.421	1.104	0.996	0.955	1.015	1.016
		0.75	0.292	1.036	0.999	0.993	1.000	1.000
		1.00	0.175	0.911	1.000	1.000	1.000	1.000
		1.25	0.114	0.785	1.000	1.000	1.000	1.000
		1.50	0.080	0.669	1.000	1.000	1.004	1.004
		1.75	0.059	0.566	1.000	1.000	1.007	1.007
		2.00	0.044	0.479	1.000	1.000	1.009	1.009
0.9	472.821	0.00	1.772	1.237	1.161	1.456	1.593	1.625
		0.15	1.722	1.244	1.156	1.394	1.567	1.598
		0.30	1.447	1.243	1.129	1.207	1.421	1.451
		0.45	1.117	1.233	1.081	1.006	1.264	1.285
		0.60	0.838	1.214	1.034	0.900	1.150	1.159
		0.75	0.631	1.188	1.004	0.865	1.079	1.082
		1.00	0.409	1.130	0.995	0.942	1.023	1.023
		1.25	0.281	1.060	0.999	0.991	1.005	1.005
		1.50	0.202	0.983	1.000	1.000	1.003	1.003
		1.75	0.152	0.904	1.000	1.000	1.006	1.006
		2.00	0.118	0.825	1.000	1.000	1.010	1.010
0.95	995.191	0.00	1.739	1.230	1.157	1.439	1.571	1.603
		0.50	1.371	1.249	1.116	1.140	1.389	1.418
		1.00	0.706	1.218	1.010	0.874	1.125	1.129
		1.50	0.382	1.146	0.996	0.958	1.045	1.045
		2.00	0.232	1.048	1.000	0.996	1.031	1.031
		2.50	0.153	0.940	1.000	1.000	1.033	1.033
		3.00	0.109	0.831	1.000	1.000	1.036	1.036
		3.50	0.081	0.728	1.000	1.000	1.036	1.036
		4.00	0.063	0.636	1.000	1.000	1.035	1.035
		5.00	0.041	0.487	1.000	1.000	1.030	1.030
		6.00	0.029	0.379	1.000	1.000	1.022	1.022
0.99	5,174.282	0.00	1.395	1.143	1.103	1.263	1.320	1.338
		0.50	1.589	1.185	1.123	1.345	1.445	1.462
		1.00	1.581	1.219	1.117	1.258	1.446	1.466
		1.50	1.389	1.213	1.083	1.094	1.376	1.391
		2.00	1.133	1.158	1.041	0.982	1.301	1.308
		2.50	0.899	1.163	1.014	0.939	1.245	1.247
		3.00	0.716	1.160	1.001	0.955	1.208	1.209
		3.50	0.579	1.149	0.999	0.979	1.181	1.181
		4.00	0.476	1.131	1.000	0.994	1.160	1.160
		5.00	0.341	1.079	1.000	1.000	1.126	1.126
		6.00	0.260	1.012	1.000	1.000	1.097	1.097

The results can be summarized as follows:

- When the subspace information was correct or nearly correct, so that Δ^* was at or near zero, RRRE dominated all other estimators as measured by RMSE. As Δ^* moved away from zero, itsRMSE decreased and converged on zero.

- Since LSRRE is a linear combination of RRRE that depends on the shrinkage intensity π, its performance was similar to that of RRRE, in that the RMSEs decreased slowly, converging on zero as Δ^* increased. LSRRE dominated all other estimators in some areas of $\Delta^* > 0$.
- As Δ^* increased from zero, the RMSEs of SPTRRE and PTRRE initially fell. In this phase, PTRRE outperformed SPTRRE. As Δ^* increased, however, SPTRRE began to outperform PTRRE. As Δ^* increased further, the RMSEs of the two estimators converged to one.
- For all combinations of p_2 and ρ, PSRRE dominated SRRE when Δ^* was at or near zero, but their performance became equivalent as Δ^* increased.
- SRRE and PSRRE outperformed all other estimators in some areas of $\Delta^* > 0$.
- Although the degree of multicollinearity increased, the behavior of the suggested estimators was similar.

From all the results, no estimator uniformly outperformed the others for the entire parameter space. Nonetheless, PTRRE, PSRRE are quite robust although there is multicollinearity and the subspace information is uncertain. While PSRRE has a requirement $p_2 \geq 3$, PTRRE does not have such a requirement. Thus, when $p_2 \geq 3$, PSRRE should be used, otherwise PTRRE should be used.

6. Application to Real Data

To investigate the practical use of our approach, the suggested estimators were applied to a new car and truck dataset. This dataset is available from the Journal of Statistics Education (JSE) website (http://jse.amstat.org/jse_data_archive.htm). The goal was to build a model to predict the retail price in ten thousand US dollars from n=428 observations. The list of variables associated with this dataset are given in Table 5.

Table 5 List of variables

Variable	Description	Remarks
Response variable		
price	Retail price in ten thousand US dollars	Numeric
Predictors		
sport	Sport car	1 = Yes and 0 = No
sportVe	Sport utility vehicle	1 = Yes and 0 = No
wagon	Wagon	1 = Yes and 0 = No
van	Minivan	1 = Yes and 0 = No
adrive	All-wheel drive	1 = Yes and 0 = No
rdrive	Rear-wheel drive	1 = Yes and 0 = No
cost	Dealer cost in US dollars	Numeric
engine	Engine size	Numeric
cylinder	Number of cylinders	Numeric
power	Horsepower	Numeric
city	City miles per gallon	Numeric
highway	Highway mile per gallon	Numeric
weight	Weight in pounds	Numeric
wheel	Wheelbase in inches	Numeric
length	Length in inches	Numeric
width	Width in inches	Numeric

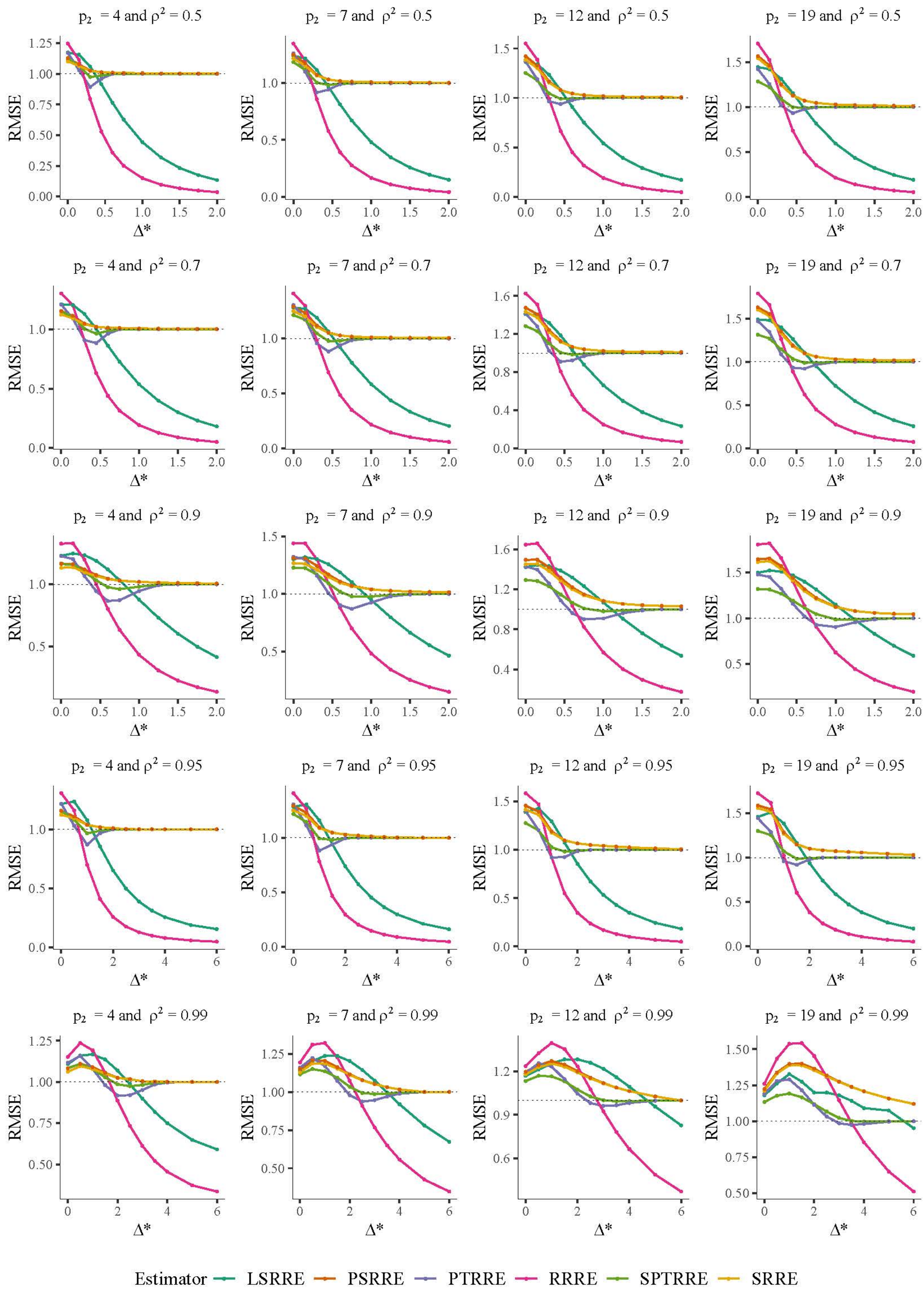

Figure 1 RMSEs of the proposed estimators with respect to $\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)$ as a function of Δ^* when $\pi = 0.50$, $\alpha = 0.05$, $\theta = 1.5$, and $n = 100$

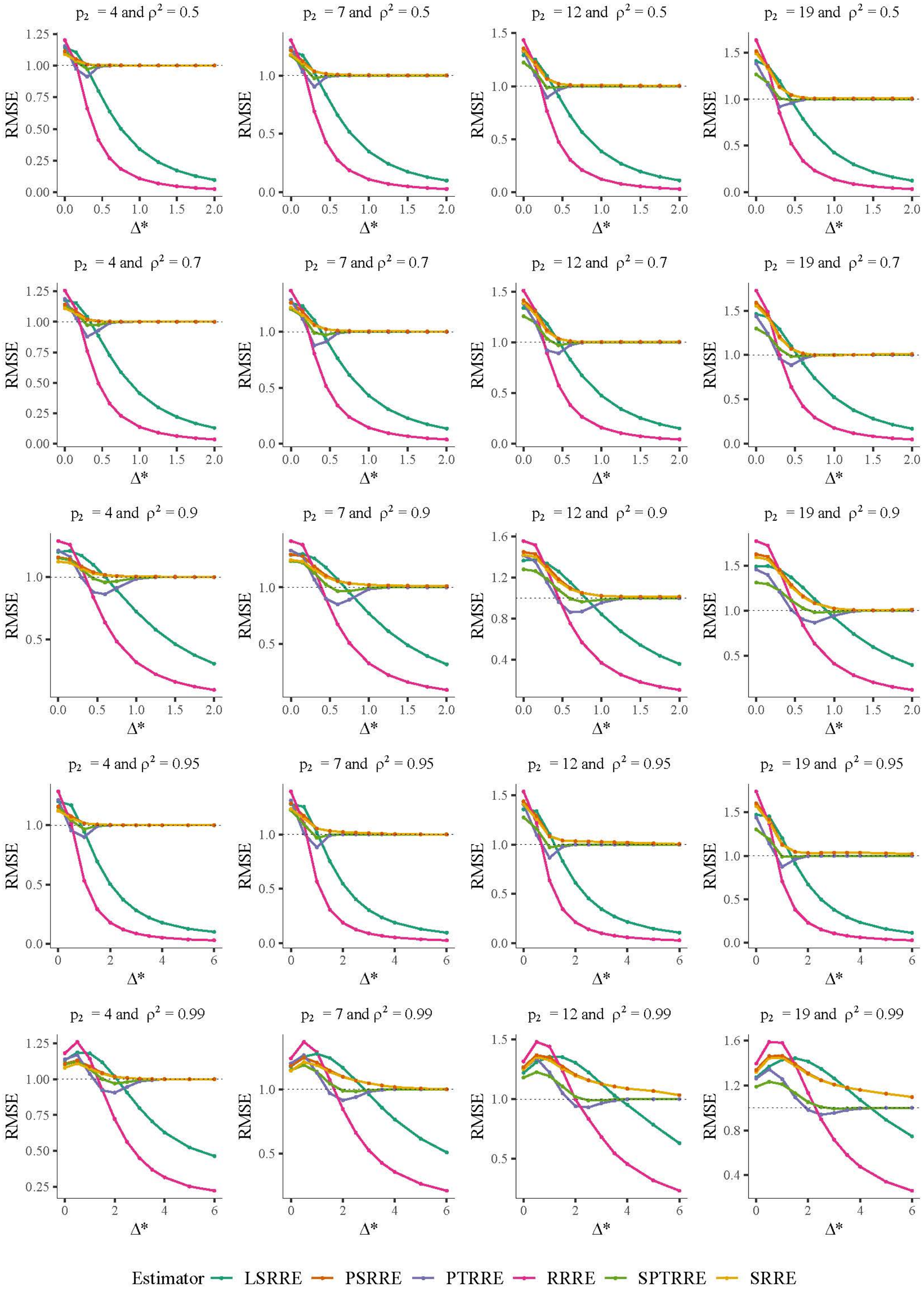

Figure 2 RMSEs of the proposed estimators with respect to $\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)$ as a function of Δ^* when $\pi = 0.50$, $\alpha = 0.05$, $\theta = 1.5$, and $n = 150$

The response variable was positive, continuous, and right skewed, meaning that this could be treated as a gamma distribution with an estimated shape parameter $\theta = 4.065$ based on a chi-square test 17.116 with a p-value of 0.145. The CN value of the Fisher information matrix was 2,167.947, implying that multicollinearity was present. To overcome this, the suggested estimators should be applied to this dataset.

In real applications, subspace information is unavailable concerning which predictors affect the response variable. Instead, the active predictors were identified using the variable selection method based on AIC, BIC, LASSO, and EN. The variable selection results to build the submodels are given in Table 6. Therefore, three different submodels were considered.

To evaluate the performance of the suggested estimators, the resampling bootstrap method was used to estimate parameter $\boldsymbol{\beta}_1$. We drew $m = 300$ bootstrap rows from the dataset with replacement and $N = 2,000$ replications. Moreover, the accuracy of the subspace information was unknown, so we conservatively selected $\pi = 0.50$. The relative prediction error was calculated to compare the performance of the suggested estimators (Reangsephet et al. 2020b):

$$RPE(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa), \hat{\boldsymbol{\beta}}_1^{\circ}(\kappa)) = \frac{\text{Simulated} \sum_{i=1}^{m} \left(y_i - \exp(\boldsymbol{x}_i' \hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa) \right)^2}{\text{Simulated} \sum_{i=1}^{m} \left(y_i - \exp(\boldsymbol{x}_i' \hat{\boldsymbol{\beta}}_1^{\circ}(\kappa) \right)^2}.$$

Table 6 Variable selection results

Variable selection method	p_1	p_2	Active predictors
AIC	13	3	sport, sportiVe, van, adrive, rdrive, cost, engine, power, city, highway, weight, wheel, length
BIC	9	7	sport, sportiVe, van, adrive, rdrive, cost, engine, power, weight
LASSO	1	15	cost
EN	1	15	cost

Table 7 RPEs of suggested estimators with respect to URRE for new car and truck data

Submodel	Estimator							
	RRRE	LSRRE	SPTRRE	PTRRE	SRRE	PSRRE	LASSO	EN
AIC	1.448	1.524	1.123	1.090	1.129	1.140	1.214	1.325
BIC	1.001	1.022	1.004	1.003	1.012	1.013	1.015	1.053
LASSO/EN	0.524	1.063	1.000	1.000	1.023	1.023	0.908	1.023

As can be seen in Table 7, the subspace information given by AIC was reliable, having the highest RPEs of the suggested estimators. LSRRE was more efficient than RRRE and also performed better than all other estimators. Moreover, SPTRRE outperformed PTRRE, and PSRRE outperformed SRRE. These results are consistent with the simulation results when the subspace information was nearly correct.

The subspace information provided by BIC was unreliable since it had poor prediction accuracy. RRRE was less efficient than LSRRE. The RPEs of PTRRE and SPTRRE were close to one and the performance of SRRE was close to that of PSRRE. Likewise, the submodel provided by LASSO/EN was also unreliable. The performance of RRRE was profoundly poor. The efficiency of PTRRE and SPTRRE were equal to that of URRE, and the efficiency of SRRE was identical to PSRRE. These results were consistent with the simulation results when the subspace information was far from correct.

When comparing the penalized maximum likelihood estimators, the EN estimator outperformed the LASSO estimators. The penalized maximum likelihood estimators performed better than URRE when the subspace information was given by AIC. In contrast, the RPEs of the LASSO and EN estimators were low when the subspace information was provided by BIC and LASSO/EN. Hence, regardless of subspace information accuracy, the shrinkage estimation strategy was robust, which is consistent with the theoretical and numerical results.

7. Conclusions

In this study, we analyzed ridge regression estimators as alternative estimators when multi-collinearity exists under a gamma regression model. We then suggested pretest and shrinkage estimation strategies to improve efficiency when subspace information on inactive predictors is available. We examined the performance of the suggested estimators from the asymptotic properties and provided numerical results. We also compared the suggested estimators with two penalty estimators LASSO and EN. Finally, we applied the suggested estimators to real data.

The numerical results confirmed that, when the subspace information was correct, the suggested estimators outperformed URRE. RRRE was superior to all other estimators when Δ^* was close to or equal to zero, but its efficiency converged on zero if the subspace information was incorrect. LSRRE performed similarly to RRRE, but its convergence on zero depended on the shrinkage intensity π. The shrinkage estimation strategy outperformed the pretest estimation strategy when p_2 was large. The penalty estimators were competitive with the suggested estimators when ρ and p_2 were large. Finally, the real data analysis produced results consistent with the theoretical and numerical results. We recommend the use of ridge-type pretest and shrinkage estimation strategies in the presence of multicollinearity.

Acknowledgments

The research of Professor S. Ejaz Ahmed was supported by the Natural Sciences and the Engineering Research Council (NSERC) of Canada and Thammasat University under the Bualuang ASEAN Chair Professorship grants. This study was supported by Thammasat University Research Fund, Contract No.TUFT 31/2564.

References

Ahmed SE. Shrinkage preliminary test estimation in multivariate normal distributions. J Stat Comput Simul. 1992; 43(3-4): 177-195.

Ahmed SE. Penalty, shrinkage and pretest strategies: variable selection and estimation. New York: Springer; 2014.

Ahmed SE, Yüzbaşı B. Big data analytics: integrating penalty strategies. Int J Manag Sci Eng Manag. 2016; 11(2): 105-115.

Algamal ZY, Asar Y. Liu-type estimator for the gamma regression model. Commun Stat Simul Comput. 2020; 49(8): 2035-2048.

Amin M, Qasim M, Amanullah M, Afzal S. Performance of some ridge estimators for the gamma regression model. Stat Pap. 2020; 61(3): 997-1026.

Amin M, Qasim M, Yasin A, Amanullah M. Almost unbiased ridge estimator in the gamma regression model. Commun Stat Simul Comput. 2022; 51(7): 3830-3850.

Bancroft TA. On biases in estimation due to the use of preliminary tests of significance. Ann Math Stat. 1944; 15(2): 190-204.

Cepeda-Cuervo E, Corrales M, Cifuentes-Amada MV, Zarate H. On Gamma Regression Residuals. J Iran stat soc. 2016; 15(1): 29-44.

Hoerl AE. Application of ridge analysis to regression problems. Chem Eng Progr. 1962; 58(3): 54-59.

Johnson RA, Wichern DW. Applied multivariate statistical analysis. New Jersey: Pearson Education, Inc; 2002.

Judge GG, Bock ME. The statistical implications of pre-test and Stein-rule estimators in econometrics. New York: North-Holland Pub. Co; 1978.

Knight K, Fu W. Asymptotics for lasso-type estimators. Ann Stat. 2000; 28(5): 1356-1378.

Mahmoudi A, Arabi Belaghi R, Mandal S. A comparison of preliminary test, Stein-type and penalty estimators in gamma regression model. J Stat Comput Simul. 2020; 90(17): 3051-3079.

Reangsephet O, Lisawadi S, Ahmed SE. Adaptive Estimation Strategies in Gamma Regression Model. J Stat Theory Pract. 2020a; 14: 8.

Reangsephet O, Lisawadi S, Ahmed SE. Post Selection Estimation and Prediction in Poisson Regression Model. Thail Stat. 2020b; 18(2): 176-195.

Weissfeld LA, Sereika SM. A multicollinearity diagnostic for generalized linear models. Commun Stat Theory Methods. 1991; 20(4): 1183-1198.

Yüzbaşı B, Ahmed SE, Aydn D. Ridge-type pretest and shrinkage estimations in partially linear models. Stat Pap. 2020; 61(2): 869-898.

Appendices

To obtain the expressions for the asymptotic properties of the suggested estimators in Theorems 2 and 4, we assume that the submatrices of the Fisher information matrix **C** satisfy the following equations:

$$\lim_{n\to\infty}\frac{1}{n}\left(\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1+\kappa\mathbf{I}\right)=\lim_{n\to\infty}\frac{1}{n}\left(\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1\right)=\mathbf{C}_{11}, \tag{6}$$

$$\lim_{n\to\infty}\frac{1}{n}\left(\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_2\right)=\mathbf{C}_{12},$$

$$\lim_{n\to\infty}\frac{1}{n}\left(\tilde{\mathbf{X}}_2'\tilde{\mathbf{X}}_1\right)=\mathbf{C}_{21},$$

$$\lim_{n\to\infty}\frac{1}{n}\left(\tilde{\mathbf{X}}_2'\tilde{\mathbf{X}}_2+\kappa\mathbf{I}\right)=\lim_{n\to\infty}\frac{1}{n}\left(\tilde{\mathbf{X}}_2'\tilde{\mathbf{X}}_2\right)=\mathbf{C}_{22}. \tag{7}$$

We present the following three lemmas which will enable us to derive the results of the asymptotic properties of the suggested estimators**.**

Lemma 1 *If* $\kappa/\sqrt{n}\to\kappa_0\geq 0$ *and* **C** *is nonsingular then*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)-\boldsymbol{\beta})\sim N(-\kappa_0\mathbf{C}^{-1}\boldsymbol{\beta},\phi\mathbf{C}^{-1}).$$

Proof: The details of the proof can be found in Knight and Fu (2000).

Lemma 2 *Let* $\mathbf{Z}$ *be a q-dimensional normal vector distributed as* $N(\boldsymbol{\mu}_Z, \boldsymbol{\Sigma}_Z)$. *Then for a measureable function of* φ *we have*

$$\mathrm{E}[\mathbf{Z}\varphi(\mathbf{Z}'\mathbf{Z})] = \boldsymbol{\mu}_Z \mathrm{E}[\varphi\chi^2_{q+2}(\Delta)],$$

$$\mathrm{E}[\mathbf{Z}\mathbf{Z}'\varphi(\mathbf{Z}'\mathbf{Z})] = \boldsymbol{\Sigma}_Z \mathrm{E}[\varphi\chi^2_{q+2}(\Delta)] + \boldsymbol{\mu}_Z\boldsymbol{\mu}'_Z \mathrm{E}[\varphi\chi^2_{q+4}(\Delta)],$$

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

Proof: The proof can be found in Judge and Bock (1978).

Lemma 3 *Let* $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2]'$ *be distributed as* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *with* $\boldsymbol{\mu} = [\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2]'$, $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$, *and* $|\boldsymbol{\Sigma}| > 0$. *The conditional distribution of* $\boldsymbol{X}_1$, given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$, *is normal and has*

$$\mathrm{E}[\boldsymbol{X}_1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2] = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) \text{ and } \mathrm{V}[\boldsymbol{X}_1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2] = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

Proof: The proof can be found in Johnson and Wichern (2002).

Proposition 1 *Let* $\boldsymbol{A}_{n1} = \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa) - \boldsymbol{\beta}_1)$, $\boldsymbol{A}_{n2} = \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{RE}}(\kappa) - \boldsymbol{\beta}_1)$, *and* $\boldsymbol{A}_{n3} = \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa) - \hat{\boldsymbol{\beta}}_1^{\mathrm{RE}}(\kappa))$. *Under the local alternative* $H_{(n)}$, *regularity conditions, and Lemma 1, as* $n \to \infty$, *we have the joint distributions:*

$$\begin{pmatrix} \boldsymbol{A}_{n1} \\ \boldsymbol{A}_{n2} \\ \boldsymbol{A}_{n3} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \boldsymbol{A}_3 \end{pmatrix} \sim N\left[\begin{pmatrix} -\boldsymbol{\mu}_{11.2} \\ -\boldsymbol{\mu}_{11.2} - \boldsymbol{\delta} \\ \boldsymbol{\delta} \end{pmatrix}, \begin{pmatrix} \phi\boldsymbol{C}_{11.2}^{-1} & \phi\boldsymbol{C}_{11}^{-1} & \boldsymbol{\Omega} \\ \phi\boldsymbol{C}_{11}^{-1} & \phi\boldsymbol{C}_{11}^{-1} & \mathbf{0} \\ \boldsymbol{\Omega} & \mathbf{0} & \boldsymbol{\Omega} \end{pmatrix}\right],$$

where $\boldsymbol{\mu}_{11.2} = \kappa_0\mathbf{C}_{11.2}^{-1}(\boldsymbol{\beta} - \boldsymbol{\omega}/\sqrt{n}\,\mathbf{C}_{12}\mathbf{C}_{22}^{-1})$, $\boldsymbol{\delta} = -\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\boldsymbol{\omega}$, $\boldsymbol{\Omega} = \phi(\mathbf{C}_{11.2}^{-1} - \mathbf{C}_{11}^{-1})$ *and* $\xrightarrow{d}$ *represents the convergence in distribution.*

Proof: Under Lemmas 1, we obtain

$$\boldsymbol{A}_{n1} \xrightarrow{d} \boldsymbol{A}_1 \sim N(-\boldsymbol{\mu}_{11.2}, \phi\boldsymbol{C}_{11.2}^{-1}).$$

From (4), we know that

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{\mathrm{RE}}(\kappa) &= \mathbf{R}_n(\kappa)\hat{\boldsymbol{\beta}}^{\mathrm{RE}} \\ &= \hat{\boldsymbol{\beta}}^{\mathrm{UE}}(\kappa) - (\mathbf{X}'\hat{\mathbf{W}}_n\mathbf{X} + \kappa\mathbf{I})^{-1}\mathbf{H}'[\mathbf{H}(\mathbf{X}'\hat{\mathbf{W}}_n\mathbf{X})^{-1}\mathbf{H}']^{-1}\mathbf{H}\hat{\boldsymbol{\beta}}^{\mathrm{UE}}. \end{aligned}$$

Using the partitioned matrix formula, we obtain the relationship between the $\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)$ and $\hat{\boldsymbol{\beta}}_1^{\mathrm{RE}}(\kappa)$ as follows:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_1^{\mathrm{RE}}(\kappa) = \hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa) &+ (\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1 + \kappa\mathbf{I})^{-1}\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_2\left(\tilde{\mathbf{X}}_2'\tilde{\mathbf{X}}_2 - \tilde{\mathbf{X}}_2'\tilde{\mathbf{X}}_1(\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1 + \kappa\mathbf{I})^{-1}\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_2 + \kappa\mathbf{I}\right)^{-1} \\ &\times\left(\tilde{\mathbf{X}}_2'\tilde{\mathbf{X}}_2 - \tilde{\mathbf{X}}_2'\tilde{\mathbf{X}}_1(\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1)^{-1}\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_2\right)\hat{\boldsymbol{\beta}}_2^{\mathrm{UE}}, \end{aligned}$$

which is a linear function of $\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)$ and $\hat{\boldsymbol{\beta}}_2^{\mathrm{UE}}(\kappa)$. Hence, by Slutky's Theorem, Theorem 1, and (6) and (7) obtain that $\boldsymbol{A}_2 \sim N_p(\boldsymbol{\mu}_{A_2}, \boldsymbol{\Sigma}_{A_2})$ such that

$$\begin{aligned}
\boldsymbol{\mu}_{A_2} &= \lim_{n\to\infty} E[\sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{RE}}(\kappa)-\boldsymbol{\beta}_1)],\\
&= \lim_{n\to\infty} E[\sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)+\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\hat{\boldsymbol{\beta}}_2^{\mathrm{UE}}-\boldsymbol{\beta}_1)],\\
&= \lim_{n\to\infty} E[\sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)-\boldsymbol{\beta}_1)]+\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\lim_{n\to\infty} E[\sqrt{n}(\hat{\boldsymbol{\beta}}_2^{\mathrm{UE}}-\boldsymbol{\beta}_2+\boldsymbol{\beta}_2)],\\
&= E[\boldsymbol{A}_1+\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{\omega}],\\
&= -\boldsymbol{\mu}_{11.2}+\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{\omega}=-\boldsymbol{\mu}_{11.2}-\boldsymbol{\delta}.
\end{aligned}$$

Hence, $\boldsymbol{A}_2 \sim N_p(-\boldsymbol{\mu}_{11.2}-\boldsymbol{\delta}, \phi\boldsymbol{C}_{11}^{-1})$.

Next, $\boldsymbol{A}_{n3}$ is a linear function of $\boldsymbol{A}_{n1}$ and $\boldsymbol{A}_{n2}$. We then have

$$\begin{aligned}
\boldsymbol{\mu}_{A_3} &= \lim_{n\to\infty} E[\sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)-\hat{\boldsymbol{\beta}}_1^{\mathrm{RE}}(\kappa))],\\
&= \lim_{n\to\infty} E[\sqrt{n}(-\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\hat{\boldsymbol{\beta}}_2^{\mathrm{UE}})],\\
&= -\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{\omega}=\boldsymbol{\delta},\\
\boldsymbol{\Sigma}_{A_3} &= \lim_{n\to\infty} V[\sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)-\hat{\boldsymbol{\beta}}_1^{\mathrm{RE}}(\kappa))],\\
&= \lim_{n\to\infty} V[\sqrt{n}(-\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\hat{\boldsymbol{\beta}}_2^{\mathrm{UE}})],\\
&= \phi\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1},\\
&= \phi(\boldsymbol{C}_{11.2}^{-1}-\boldsymbol{C}_{11}^{-1})=\boldsymbol{\Omega}.
\end{aligned}$$

Finally, we derive the covariance matrix between $\boldsymbol{A}_1, \boldsymbol{A}_2,$ and $\boldsymbol{A}_3$ as follows

$$\begin{aligned}
Cov[\boldsymbol{A}_1,\boldsymbol{A}_2] &= \phi\boldsymbol{C}_{11}^{-1},\\
Cov[\boldsymbol{A}_1,\boldsymbol{A}_3] &= Cov[\boldsymbol{A}_1,\boldsymbol{A}_1-\boldsymbol{A}_2]=\boldsymbol{\Omega},\\
Cov[\boldsymbol{A}_2,\boldsymbol{A}_3] &= Cov[\boldsymbol{A}_2,\boldsymbol{A}_1-\boldsymbol{A}_2]=\boldsymbol{0}.
\end{aligned}$$

Appendix A: Proof of Theorem 2

Proof: Using Proposition 1, we have $ADB(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa))=-\boldsymbol{\mu}_{11.2}$. Next, the ADB of LSRRE is considered

$$\begin{aligned}
\boldsymbol{A}_{n4} &= \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{LSE}}(\kappa)-\boldsymbol{\beta}_1),\\
&= \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)-\boldsymbol{\beta}_1)-\pi\sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)-\hat{\boldsymbol{\beta}}_1^{\mathrm{RE}}(\kappa)),\\
&= \boldsymbol{A}_{n1}-\pi\boldsymbol{A}_{n3},
\end{aligned}$$

which is a linear function of $\boldsymbol{A}_{n1}$ and $\boldsymbol{A}_{n3}$. By Slutsky's Theorem and as $n\to\infty$, we obtain

$$\begin{aligned}
ADB(\hat{\boldsymbol{\beta}}_1^{\mathrm{LSE}}(\kappa)) &= \lim_{n\to\infty} \mathrm{E}[\boldsymbol{A}_{n4}],\\
&= \lim_{n\to\infty} \mathrm{E}[\boldsymbol{A}_{n1}-\pi\boldsymbol{A}_{n3}],\\
&= -\boldsymbol{\mu}_{11.2}-\pi\boldsymbol{\delta}.
\end{aligned}$$

The ADB of SPTRRE is as follows

$$\begin{aligned}
\boldsymbol{A}_{n5} &= \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{SPTE}}(\kappa)-\boldsymbol{\beta}_1),\\
&= \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)-\pi(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)-\hat{\boldsymbol{\beta}}_1^{\mathrm{RE}}(\kappa))I(\mathcal{L}_n\le\mathcal{L}_{p_2,\alpha})-\boldsymbol{\beta}_1),\\
&= \boldsymbol{A}_{n1}-\pi\boldsymbol{A}_{n3}I(\mathcal{L}_n\le\mathcal{L}_{c+2,\alpha}),
\end{aligned}$$

which is a linear function of $\boldsymbol{A}_{n1}$ and $\boldsymbol{A}_{n3}$. By Slutsky's Theorem and as $n\to\infty$, we obtain

$$ADB(\hat{\boldsymbol{\beta}}_1^{\text{SPTE}}(\kappa)) = \lim_{n\to\infty} \mathrm{E}[\boldsymbol{A}_{n5}],$$
$$= \lim_{n\to\infty} \mathrm{E}[\boldsymbol{A}_{n1} - \pi \boldsymbol{A}_{n3} I(\mathcal{L}_n \le \mathcal{L}_{c+2,\alpha})],$$
$$= -\boldsymbol{\mu}_{11.2} - \pi \boldsymbol{\delta} G_{c+4}(\chi^2_{c+2,\alpha};\Delta).$$

The ADB of SRRE is

$$\boldsymbol{A}_{n6} = \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa) - \boldsymbol{\beta}_1),$$
$$= \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa) - c(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa) - \hat{\boldsymbol{\beta}}_1^{\text{RE}}(\kappa))\mathcal{L}_n^{-1} - \boldsymbol{\beta}_1),$$
$$= \boldsymbol{A}_{n1} - c\boldsymbol{A}_{n3}\mathcal{L}_n^{-1},$$

which is a linear function of $\boldsymbol{A}_{n1}$ and $\boldsymbol{A}_{n3}$. By Slutsky's Theorem and as $n \to \infty$, we obtain

$$ADB(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)) = \lim_{n\to\infty} \mathrm{E}[\boldsymbol{A}_{n6}],$$
$$= \lim_{n\to\infty} \mathrm{E}[\boldsymbol{A}_{n1} - c\boldsymbol{A}_{n3}\mathcal{L}_n^{-1}],$$
$$= -\boldsymbol{\mu}_{11.2} - c\boldsymbol{\delta} E[\chi^{-2}_{c+4}(\Delta)].$$

Finally, the ADB of PSRRE is considered such that

$$\boldsymbol{A}_{n7} = \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{PSE}}(\kappa) - \boldsymbol{\beta}_1),$$
$$= \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa) - (\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa) - \hat{\boldsymbol{\beta}}_1^{\text{RE}}(\kappa))(1 - c\mathcal{L}_n^{-1}) I(\mathcal{L}_n \le c) - \boldsymbol{\beta}_1),$$
$$= \boldsymbol{A}_{n6} - \boldsymbol{A}_{n3}(1 - c\mathcal{L}_n^{-1}) I(\mathcal{L}_n \le c),$$

which is a linear function of $\boldsymbol{A}_{n6}$ and $\boldsymbol{A}_{n3}$. By Slutsky's Theorem and as $n \to \infty$, we obtain

$$ADB(\hat{\boldsymbol{\beta}}_1^{\text{PSE}}(\kappa)) = \lim_{n\to\infty} \mathrm{E}[\boldsymbol{A}_{n7}],$$
$$= \lim_{n\to\infty} \mathrm{E}[\boldsymbol{A}_{n6} - \boldsymbol{A}_{n3}(1 - c\mathcal{L}_n^{-1}) I(\mathcal{L}_n \le c)],$$
$$= ADB(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)) - E[\boldsymbol{A}_{n3}]\,\mathrm{E}[(1 - c\chi^{-2}_{c+4}(\Delta)) I(\chi^2_{c+4}(\Delta) \le c)],$$
$$= -\boldsymbol{\mu}_{11.2} - c\boldsymbol{\delta} E[\chi^{-2}_{c+4}(\Delta)] - \boldsymbol{\delta} G_{c+4}(c;\Delta) + c\boldsymbol{\delta} E[\chi^{-2}_{c+4}(\Delta) I(\chi^2_{c+4}(\Delta) \le c)],$$
$$= -\boldsymbol{\mu}_{11.2} - \boldsymbol{\delta}\left\{cE[\chi^{-2}_{c+4}(\Delta)] + G_{c+4}(c;\Delta) - cE[\chi^{-2}_{c+4}(\Delta) I(\chi^2_{c+4}(\Delta) \le c)]\right\},$$
$$= -\boldsymbol{\mu}_{11.2} - \boldsymbol{\delta} D(\Delta).$$

Appendix B: Proof of Theorem 4

To prove the asymptotic distributional risk of the estimators, we first derive the asymptotic mean squared error matrix (AMSEM):

$$\Gamma(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) = \lim_{n\to\infty} E[n(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa) - \boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa) - \boldsymbol{\beta}_1)'],$$
$$= \lim_{n\to\infty} E[\boldsymbol{A}_{n1}\boldsymbol{A}'_{n1})],$$
$$= V[\boldsymbol{A}_1] + \boldsymbol{\mu}_{A_1}\boldsymbol{\mu}'_{A_1},$$
$$= \phi \boldsymbol{C}^{-1}_{11.2} + \boldsymbol{\mu}_{11.2}\boldsymbol{\mu}'_{11.2},$$
$$\Gamma(\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa)) = \lim_{n\to\infty} E[n(\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa) - \boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_1^{\text{LSE}}(\kappa) - \boldsymbol{\beta}_1)'],$$
$$= \lim_{n\to\infty} \mathrm{E}[(\boldsymbol{A}_{n1} - \pi\boldsymbol{A}_{n3})(\boldsymbol{A}_{n1} - \pi\boldsymbol{A}_{n3})'],$$
$$= \Gamma(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) - 2\pi\left\{\mathrm{Cov}[\boldsymbol{A}_1, \boldsymbol{A}_3] + \boldsymbol{\mu}_{A_1}\boldsymbol{\mu}'_{A_3}\right\} + \pi^2\left\{\mathrm{V}[\boldsymbol{A}_3] + \boldsymbol{\mu}_{A_3}\boldsymbol{\mu}'_{A_3}\right\},$$
$$= \Gamma(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) - \pi(2-\pi)\boldsymbol{\Omega} + 2\pi\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}' + \pi^2\boldsymbol{\delta}\boldsymbol{\delta}',$$

$$\Gamma(\hat{\boldsymbol{\beta}}_1^{\text{SPTE}}(\kappa)) = \lim_{n\to\infty} \text{E}[n(\hat{\boldsymbol{\beta}}_1^{\text{SPTE}}(\kappa)-\boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_1^{\text{SPTE}}(\kappa)-\boldsymbol{\beta}_1)'],$$
$$= \lim_{n\to\infty} \text{E}[(\boldsymbol{A}_{n1}-\pi \boldsymbol{A}_{n3} I(\mathcal{L}_n \le \mathcal{L}_{c+2,\,\alpha}))(\boldsymbol{A}_{n1}-\pi \boldsymbol{A}_{n3} I(\mathcal{L}_n \le \mathcal{L}_{c+2,\,\alpha}))'],$$
$$= \text{E}[\boldsymbol{A}_1\boldsymbol{A}_1'] - 2\pi \underbrace{\text{E}[\boldsymbol{A}_1\boldsymbol{A}_3' I(\chi^2_{c+2}(\Delta) \le \chi^2_{c+2,\,\alpha})]}_{E_1} + \pi^2\,\text{E}[\boldsymbol{A}_3\boldsymbol{A}_3' I(\chi^2_{c+2}(\Delta) \le \chi^2_{c+2,\,\alpha})].$$

Applying the law of iterated expectations with Lemmas 2 and 3 gives

$$E_1 = \text{E}\Big[\text{E}[\boldsymbol{A}_1 \mid \boldsymbol{A}_3]\boldsymbol{A}_3' I(\chi^2_{c+2}(\Delta) \le \chi^2_{c+2,\,\alpha})\Big],$$
$$= \text{E}[(-\boldsymbol{\mu}_{11.2} + \boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}(\boldsymbol{A}_3 - \boldsymbol{\delta}))\boldsymbol{A}_3' I(\chi^2_{c+2}(\Delta) \le \chi^2_{c+2,\,\alpha})],$$
$$= -\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}'\,\text{E}[I(\chi^2_{c+4}(\Delta) \le \chi^2_{c+2,\,\alpha})] + \boldsymbol{\Omega}\,\text{E}[I(\chi^2_{c+4}(\Delta) \le \chi^2_{c+2,\,\alpha})] + \boldsymbol{\delta}\boldsymbol{\delta}'\,\text{E}[I(\chi^2_{c+6}(\Delta) \le \chi^2_{c+2,\,\alpha})]$$
$$-\boldsymbol{\delta}\boldsymbol{\delta}'\,\text{E}[I(\chi^2_{c+4}(\Delta) \le \chi^2_{c+2,\,\alpha})],$$
$$= -\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}' G_{c+4}(\chi^2_{c+2,\,\alpha};\Delta) + \boldsymbol{\Omega} G_{c+4}(\chi^2_{c+2,\,\alpha};\Delta) + \boldsymbol{\delta}\boldsymbol{\delta}' G_{c+6}(\chi^2_{c+2,\,\alpha};\Delta) - \boldsymbol{\delta}\boldsymbol{\delta}' G_{c+4}(\chi^2_{c+2,\,\alpha};\Delta).$$

Substituting the results of E_1 into $\Gamma(\hat{\boldsymbol{\beta}}_1^{\text{SPTE}}(\kappa))$, we get

$$\Gamma(\hat{\boldsymbol{\beta}}_1^{\text{SPTE}}(\kappa)) = \Gamma(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) + 2\pi\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}' G_{c+4}(\chi^2_{c+2,\,\alpha};\Delta) - 2\pi\boldsymbol{\Omega} G_{c+4}(\chi^2_{c+2,\,\alpha};\Delta) - 2\pi\boldsymbol{\delta}\boldsymbol{\delta}' G_{c+6}(\chi^2_{c+2,\,\alpha};\Delta)$$
$$+2\pi\boldsymbol{\delta}\boldsymbol{\delta}' G_{c+4}(\chi^2_{c+2,\,\alpha};\Delta) + \pi^2\boldsymbol{\Omega} G_{c+4}(\chi^2_{c+2,\,\alpha};\Delta) + \pi^2\boldsymbol{\delta}\boldsymbol{\delta}' G_{c+6}(\chi^2_{c+2,\,\alpha};\Delta),$$
$$= \Gamma(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) + 2\pi\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}' G_{c+4}(\chi^2_{c+2,\,\alpha};\Delta) - \pi(2-\pi)\boldsymbol{\Omega} G_{c+4}(\chi^2_{c+2,\,\alpha};\Delta)$$
$$+\boldsymbol{\delta}\boldsymbol{\delta}'\left\{2\pi G_{c+4}(\chi^2_{c+2,\,\alpha};\Delta) - \pi(2-\pi) G_{c+6}(\chi^2_{c+2,\,\alpha};\Delta)\right\},$$

$$\Gamma(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)) = \lim_{n\to\infty} \text{E}[n(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)-\boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)-\boldsymbol{\beta}_1)'],$$
$$= \lim_{n\to\infty} \text{E}[(\boldsymbol{A}_{n1} - c\boldsymbol{A}_{n3}\mathcal{L}_n^{-1}))(\boldsymbol{A}_{n1} - c\boldsymbol{A}_{n3}\mathcal{L}_n^{-1})'],$$
$$= \text{E}[\boldsymbol{A}_1\boldsymbol{A}_1'] - 2c\underbrace{\text{E}[\boldsymbol{A}_1\boldsymbol{A}_3'\chi^{-2}_{c+2}(\Delta)]}_{E_2} + c^2\,\text{E}[\boldsymbol{A}_3\boldsymbol{A}_3'\chi^{-4}_{c+2}(\Delta)].$$

Using the law of iterated expectation, we have

$$E_2 = \text{E}\Big[\text{E}[\boldsymbol{A}_1 \mid \boldsymbol{A}_3]\boldsymbol{A}_3'\chi^{-2}_{c+2}(\Delta)\Big],$$
$$= \text{E}[(-\boldsymbol{\mu}_{11.2} + \boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}(\boldsymbol{A}_3 - \boldsymbol{\delta}))\boldsymbol{A}_3'\chi^{-2}_{c+2}(\Delta)],$$
$$= \text{E}[-\boldsymbol{\mu}_{11.2}\boldsymbol{A}_3'\chi^{-2}_{c+2}(\Delta)] + \text{E}[\boldsymbol{A}_3\boldsymbol{A}_3'\chi^{-2}_{c+2}(\Delta)] - \boldsymbol{\delta}\text{E}[\boldsymbol{A}_3'\chi^{-2}_{c+2}(\Delta)],$$
$$= -\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}'\,\text{E}[\chi^{-2}_{c+4}(\Delta)] + \boldsymbol{\Omega}\,\text{E}[\chi^{-2}_{c+4}(\Delta)] + \boldsymbol{\delta}\boldsymbol{\delta}'\,\text{E}[\chi^{-2}_{c+6}(\Delta)] - \boldsymbol{\delta}\boldsymbol{\delta}'\,\text{E}[\chi^{-2}_{c+4}(\Delta)].$$

Replacing E_2 in $\Gamma(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa))$, we obtain

$$\Gamma(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)) = \Gamma(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) + 2c\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}'\,\text{E}[\chi^{-2}_{c+4}(\Delta)] - 2c\boldsymbol{\Omega}\,\text{E}[\chi^{-2}_{c+4}(\Delta)] - 2c\boldsymbol{\delta}\boldsymbol{\delta}'\,\text{E}[\chi^{-2}_{c+6}(\Delta)]$$
$$+2c\boldsymbol{\delta}\boldsymbol{\delta}'\,\text{E}[\chi^{-2}_{c+4}(\Delta)] + c^2\boldsymbol{\Omega}\,\text{E}[\chi^{-4}_{c+4}(\Delta)] + c^2\boldsymbol{\delta}\boldsymbol{\delta}'\,\text{E}[\chi^{-4}_{c+4}(\Delta)],$$
$$= \Gamma(\hat{\boldsymbol{\beta}}_1^{\text{UE}}(\kappa)) + 2c\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}'\,\text{E}[\chi^{-2}_{c+4}(\Delta)] - c\boldsymbol{\Omega}\left\{2\,\text{E}[\chi^{-2}_{c+4}(\Delta)] - c\,\text{E}[\chi^{-4}_{c+4}(\Delta)]\right\}$$
$$+c\boldsymbol{\delta}\boldsymbol{\delta}'\left\{-2\,\text{E}[\chi^{-2}_{c+6}(\Delta)] + 2\,\text{E}[\chi^{-2}_{c+4}(\Delta)] + c\,\text{E}[\chi^{-4}_{c+6}(\Delta)]\right\}.$$

Using the identity $\text{E}[\chi^{-2}_{c+4}(\Delta)] - \text{E}[\chi^{-2}_{c+6}(\Delta)] = 2\,\text{E}[\chi^{-4}_{c+6}(\Delta)]$. Therefore, $\Gamma(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa))$ is obtained as

$$\Gamma(\hat{\boldsymbol{\beta}}_1^{\mathrm{SE}}(\kappa)) = \Gamma(\hat{\boldsymbol{\beta}}_1^{\mathrm{UE}}(\kappa)) + 2c\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}'\,\mathrm{E}[\chi_{c+4}^{-2}(\Delta)] - c\boldsymbol{\Omega}\left\{2\,\mathrm{E}[\chi_{c+4}^{-2}(\Delta)] - c\,\mathrm{E}[\chi_{c+4}^{-4}(\Delta)]\right\}$$
$$+ c(c+4)\boldsymbol{\delta}\boldsymbol{\delta}'\,\mathrm{E}[\chi_{c+6}^{-4}(\Delta)],$$

$$\Gamma(\hat{\boldsymbol{\beta}}_1^{\mathrm{PSE}}(\kappa)) = \lim_{n\to\infty}\mathrm{E}[n(\hat{\boldsymbol{\beta}}_1^{\mathrm{PSE}}(\kappa) - \boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_1^{\mathrm{PSE}}(\kappa) - \boldsymbol{\beta}_1)'],$$
$$= \lim_{n\to\infty}\mathrm{E}[(\boldsymbol{A}_{n6} - \boldsymbol{A}_{n3}(1 - c\mathcal{L}_n^{-1})I(\mathcal{L}_n \le c))(\boldsymbol{A}_{n6} - \boldsymbol{A}_{n3}(1 - c\mathcal{L}_n^{-1})I(\mathcal{L}_n \le c))'],$$
$$= \lim_{n\to\infty}\mathrm{E}[\boldsymbol{A}_{n6}\boldsymbol{A}_{n6}'] - 2\lim_{n\to\infty}\mathrm{E}[\boldsymbol{A}_{n6}\boldsymbol{A}_{n3}'(1 - c\mathcal{L}_n^{-1})I(\mathcal{L}_n \le c)]$$
$$+ \lim_{n\to\infty}\mathrm{E}[\boldsymbol{A}_{n3}\boldsymbol{A}_{n3}'(1 - c\mathcal{L}_n^{-1})I(\mathcal{L}_n \le c)],$$
$$= \Gamma(\hat{\boldsymbol{\beta}}_1^{\mathrm{SE}}(\kappa)) - 2\lim_{n\to\infty}\mathrm{E}[(\boldsymbol{A}_{n1} - c\boldsymbol{A}_{n3}\mathcal{L}_n^{-1})\boldsymbol{A}_{n3}'(1 - c\mathcal{L}_n^{-1})I(\mathcal{L}_n \le c)]$$
$$+ \lim_{n\to\infty}\mathrm{E}[\boldsymbol{A}_{n3}\boldsymbol{A}_{n3}'(1 - 2c\mathcal{L}_n^{-1} + c^2\mathcal{L}_n^{-1})I(\mathcal{L}_n \le c)],$$
$$= \Gamma(\hat{\boldsymbol{\beta}}_1^{\mathrm{SE}}(\kappa)) - 2\,\mathrm{E}[(\boldsymbol{A}_1 - c\boldsymbol{A}_3\chi_{c+2}^{-2}(\Delta))\boldsymbol{A}_3'(1 - c\chi_{c+2}^{-2}(\Delta))I(\chi_{c+2}^{2}(\Delta) \le c)]$$
$$+ \mathrm{E}[\boldsymbol{A}_3\boldsymbol{A}_3'I(\chi_{c+2}^{2}(\Delta) \le c)] - 2c\,\mathrm{E}[\boldsymbol{A}_3\boldsymbol{A}_3'\chi_{c+2}^{-2}(\Delta)I(\chi_{c+2}^{2}(\Delta) \le c)]$$
$$+ c^2\,\mathrm{E}[\boldsymbol{A}_3\boldsymbol{A}_3'\chi_{c+2}^{-4}(\Delta)I(\chi_{c+2}^{2}(\Delta) \le c)],$$

$$\Gamma(\hat{\boldsymbol{\beta}}_1^{\mathrm{PSE}}(\kappa)) = \Gamma(\hat{\boldsymbol{\beta}}_1^{\mathrm{SE}}(\kappa)) - 2\,\mathrm{E}[(\boldsymbol{A}_1\boldsymbol{A}_3'(1 - c\chi_{c+2}^{-2}(\Delta))I(\chi_{c+2}^{2}(\Delta) \le c)]$$
$$+ 2c\,\mathrm{E}[\boldsymbol{A}_3\boldsymbol{A}_3'\chi_{c+2}^{-2}(\Delta)(1 - c\chi_{c+2}^{-2}(\Delta))I(\chi_{c+2}^{2}(\Delta) \le c)]$$
$$+ \mathrm{E}[\boldsymbol{A}_3\boldsymbol{A}_3'I(\chi_{c+2}^{2}(\Delta) \le c)] - 2c\,\mathrm{E}[\boldsymbol{A}_3\boldsymbol{A}_3'\chi_{c+2}^{-2}(\Delta)I(\chi_{c+2}^{2}(\Delta) \le c)]$$
$$+ c^2\,\mathrm{E}[\boldsymbol{A}_3\boldsymbol{A}_3'\chi_{c+2}^{-4}(\Delta)I(\chi_{c+2}^{2}(\Delta) \le c)],$$
$$= \Gamma(\hat{\boldsymbol{\beta}}_1^{\mathrm{SE}}(\kappa)) - 2\underbrace{\mathrm{E}[(\boldsymbol{A}_1\boldsymbol{A}_3'(1 - c\chi_{c+2}^{-2}(\Delta))I(\chi_{c+2}^{2}(\Delta) \le c)]}_{E_3}$$
$$- c^2\,\mathrm{E}[\boldsymbol{A}_3\boldsymbol{A}_3'\chi_{c+2}^{-4}(\Delta)I(\chi_{c+2}^{2}(\Delta) \le c)] + \mathrm{E}[\boldsymbol{A}_3\boldsymbol{A}_3'I(\chi_{c+2}^{2}(\Delta) \le c)].$$

Applying the law of iterated expectations with Lemmas 2 and 3 gives

$$E_3 = \mathrm{E}\Big[\mathrm{E}[\boldsymbol{A}_1 \mid \boldsymbol{A}_3]\boldsymbol{A}_3'(1 - c\chi_{c+2}^{-2}(\Delta))I(\chi_{c+2}^{2}(\Delta) \le c)\Big],$$
$$= \mathrm{E}[(-\boldsymbol{\mu}_{11.2} + \boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}(\boldsymbol{A}_3 - \boldsymbol{\delta}))\boldsymbol{A}_3'(1 - c\chi_{c+2}^{-2}(\Delta))I(\chi_{c+2}^{2}(\Delta) \le c)],$$
$$= \mathrm{E}[-\boldsymbol{\mu}_{11.2}\boldsymbol{A}_3'(1 - c\chi_{c+2}^{-2}(\Delta))I(\chi_{c+2}^{2}(\Delta) \le c)] + \mathrm{E}[\boldsymbol{A}_3\boldsymbol{A}_3'(1 - c\chi_{c+2}^{-2}(\Delta))I(\chi_{c+2}^{2}(\Delta) \le c)]$$
$$- \boldsymbol{\delta}\mathrm{E}[\boldsymbol{A}_3'(1 - c\chi_{c+2}^{-2}(\Delta))I(\chi_{c+2}^{2}(\Delta) \le c)],$$
$$= -\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}'\,\mathrm{E}[(1 - c\chi_{c+4}^{-2}(\Delta))I(\chi_{c+4}^{2}(\Delta) \le c)] + \boldsymbol{\Omega}\,\mathrm{E}[(1 - c\chi_{c+4}^{-2}(\Delta))I(\chi_{c+4}^{2}(\Delta) \le c)]$$
$$+ \boldsymbol{\delta}\boldsymbol{\delta}'\,\mathrm{E}[(1 - c\chi_{c+6}^{-2}(\Delta))I(\chi_{c+6}^{2}(\Delta) \le c)] - \boldsymbol{\delta}\boldsymbol{\delta}'\,\mathrm{E}[(1 - c\chi_{c+4}^{-2}(\Delta))I(\chi_{c+4}^{2}(\Delta) \le c)].$$

Substituting the result of E3 into $\Gamma(\hat{\boldsymbol{\beta}}_1^{\mathrm{PSE}}(\kappa))$, we obtain

$$\begin{aligned}
\Gamma(\hat{\boldsymbol{\beta}}_1^{\text{PSE}}(\kappa)) &= \Gamma(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)) + 2\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}'\,\mathrm{E}[(1-c\chi_{c+4}^{-2}(\Delta))I(\chi_{c+4}^{2}(\Delta)\le c)] \\
&\quad -2\boldsymbol{\Omega}\,\mathrm{E}[(1-c\chi_{c+4}^{-2}(\Delta))I(\chi_{c+4}^{2}(\Delta)\le c)] - 2\boldsymbol{\delta\delta}'\,\mathrm{E}[(1-c\chi_{c+6}^{-2}(\Delta))I(\chi_{c+6}^{2}(\Delta)\le c)] \\
&\quad +2\boldsymbol{\delta\delta}'\,\mathrm{E}[(1-c\chi_{c+4}^{-2}(\Delta))I(\chi_{c+4}^{2}(\Delta)\le c)] - c^2\boldsymbol{\Omega}\,\mathrm{E}[\chi_{c+4}^{-4}(\Delta)I(\chi_{c+4}^{2}(\Delta)\le c)] \\
&\quad -c^2\boldsymbol{\delta\delta}'\,\mathrm{E}[\chi_{c+6}^{-4}(\Delta)I(\chi_{c+6}^{2}(\Delta)\le c)] + \boldsymbol{\Omega}\,\mathrm{E}[I(\chi_{c+4}^{2}(\Delta)\le c)] \\
&\quad +\boldsymbol{\delta\delta}'\,\mathrm{E}[I(\chi_{c+5}^{2}(\Delta)\le c)], \\
&= \Gamma(\hat{\boldsymbol{\beta}}_1^{\text{SE}}(\kappa)) + 2\boldsymbol{\mu}_{11.2}\boldsymbol{\delta}'\,\mathrm{E}[(1-c\chi_{c+4}^{-2}(\Delta))I(\chi_{c+4}^{2}(\Delta)\le c)] \\
&\quad -\boldsymbol{\Omega}\left\{\begin{array}{l} 2\,\mathrm{E}[(1-c\chi_{c+4}^{-2}(\Delta))I(\chi_{c+4}^{2}(\Delta)\le c)] \\ +c^2\,\mathrm{E}[\chi_{c+4}^{-4}(\Delta)I(\chi_{c+4}^{2}(\Delta)\le c)] + G_{c+4}(c;\Delta) \end{array}\right\} \\
&\quad +\boldsymbol{\delta\delta}'\left\{\begin{array}{l} \mathrm{E}[(1-c\chi_{c+4}^{-2}(\Delta))I(\chi_{c+4}^{2}(\Delta)\le c)] - 2\,\mathrm{E}[(1-c\chi_{c+6}^{-2}(\Delta))I(\chi_{c+6}^{2}(\Delta)\le c)] \\ -c^2\,\mathrm{E}[\chi_{c+6}^{-4}(\Delta)I(\chi_{c+6}^{2}(\Delta)\le c)] + G_{c+6}(c;\Delta) \end{array}\right\}.
\end{aligned}$$

The ADRs of the suggested estimators can be proved using the AMSEMs.