
Faculty of Science

Faculty Publications

“Sparse Estimation Strategies in Linear Mixed Effect Models for High-Dimensional Data Application”

Eugene A. Opoku, Syed Ejaz Ahmed, and Farouk S. Nathoo

2021

© 2021 Opoku et al. This is an open access article distributed under the terms of the Creative Commons Attribution License. <http://creativecommons.org/licenses/by/4.0>

This article was originally published at:

<https://doi.org/10.3390/e23101348>

Citation for this paper:

Opoku, E. A., Ahmed, S. E., & Nathoo, F. S. (2021). Sparse estimation strategies in linear mixed effect models for high-dimensional data application. *entropy*, 23(1348), 1-24. <https://doi.org/10.3390/e23101348>

Sparse Estimation Strategies in Linear Mixed Effect Models for High-Dimensional Data Application

Eugene A. Opoku ^{1,*}, Syed Ejaz Ahmed ² and Farouk S. Nathoo ¹

¹ Department of Mathematics and Statistics, University of Victoria, Victoria, BC V8P 5C2, Canada; nathoo@uvic.ca

² Department of Mathematics and Statistics, Brock University, St. Catharines, ON L2S 3A1, Canada; sahmed5@brocku.ca

* Correspondence: eopoku@uvic.ca

Abstract: In a host of business applications, biomedical and epidemiological studies, the problem of multicollinearity among predictor variables is a frequent issue in longitudinal data analysis for linear mixed models (LMM). We consider an efficient estimation strategy for high-dimensional data application, where the dimensions of the parameters are larger than the number of observations. In this paper, we are interested in estimating the fixed effects parameters of the LMM when it is assumed that some prior information is available in the form of linear restrictions on the parameters. We propose the pretest and shrinkage estimation strategies using the ridge full model as the base estimator. We establish the asymptotic distributional bias and risks of the suggested estimators and investigate their relative performance with respect to the ridge full model estimator. Furthermore, we compare the numerical performance of the LASSO-type estimators with the pretest and shrinkage ridge estimators. The methodology is investigated using simulation studies and then demonstrated on an application exploring how effective brain connectivity in the default mode network (DMN) may be related to genetics within the context of Alzheimer's disease.

Keywords: linear mixed model; ridge estimation; pretest and shrinkage estimation; multicollinearity; asymptotic bias and risk; LASSO estimation; high-dimensional data



Citation: Opoku, E.A.; Ahmed, S.E.; Nathoo, F.S. Sparse Estimation Strategies in Linear Mixed Effect Models for High-Dimensional Data Application. *Entropy* **2021**, *23*, 1348. <https://doi.org/10.3390/e23101348>

Academic Editor: Matteo Convertino

Received: 9 September 2021

Accepted: 12 October 2021

Published: 15 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In many fields such as bio-informatics, physical biology, and epidemiology, the response of interest is represented by repeated measures of some variables of interest that are collected over a specified time period for different independent subjects or individuals. These types of data are commonly encountered in medical research where the responses are subject to various time-dependent and time-constant effects such as pre- and post-treatment types, gender effect, and baseline measures, among others. A widely-used statistical tool in the analysis and modeling of longitudinal and repeated measures data is the linear mixed effects model (LMM) [1,2]. This model provides an effective and flexible way to describe the means and the covariance structures of a response variable after accounting for within subject correlation.

The rapid growth in the size and scope of longitudinal data has created a need for innovative statistical strategies in longitudinal data analysis. Classical methods are based on the assumption that the number of predictors is less than the number of observations. However, there is an increasing demand for efficient prediction strategies for analysis of high-dimensional data, where the number of observed data elements (sample size) are smaller than the number of predictors in a linear model context. Existing techniques that deal with high-dimensional data mostly rely on various penalized estimators. Due to the trade-off between model complexity and model prediction, the statistical inference of model selection becomes an extremely important and challenging problem in high-dimensional data analysis.

Over the years, many penalized regularization approaches have been developed to do variable selection and estimation simultaneously. Among them, the least absolute shrinkage and selection operator (LASSO) is commonly used [3]. It is a useful estimation technique in part due to its convexity and computational efficiency. The LASSO approach is based on an ℓ_1 penalty for regularization of regression parameters. Ref. [4] provides a comprehensive summary of the consistency properties of the LASSO approach. Related penalized likelihood methods have been extensively studied in the literature, see for example [5–10]. The penalized likelihood methods have a close connection to Bayesian procedures. Thus, the LASSO estimate corresponds to a Bayes method that puts a Laplacian (double-exponential) prior on the regression coefficients [11,12].

In this paper, our interest lies in estimating the fixed effect parameters of the LMM using a ridge estimation technique when it is assumed that some prior information is available in the form of potential linear restrictions on the parameters. One possible source of prior information is using a Bayesian approach. An alternative source of prior information may be obtained from previous studies or expert knowledge that search for or assume sparsity patterns.

We consider the problem of fixed effect parameter estimation for LMMs when there exist many predictors relative to the sample size. These predictors may be classified into two groups: sparse and non-sparse. Thus, there are two choices to be considered: a full model with all predictors, and a sub-model that contains only non-sparse predictors. When the sub-model based on available subspace information is true (i.e., the assumed restriction holds), it then provides more efficient statistical inferences than those based on a full model. In contrast, if the sub-model is not true, the estimates could become biased and inefficient. The consequences of incorporating subspace information therefore depend on the quality or reliability of the information being incorporated into the estimation procedure. One way to deal with uncertain subspace information is to use a pretest estimation strategy. The validity of the information is tested before incorporation into a final estimator. Another approach is shrinkage estimation, which shrinks the full model estimator to the sub-model estimator by utilizing subspace information. Besides these estimation strategies, there is a growing literature on simultaneous model selection and estimation. These approaches are known as penalty strategies. By shrinking some regression coefficients toward zero, the penalty methods simultaneously select a sub-model and estimate its regression parameters. Several authors have investigated the pretest, shrinkage, and penalty estimation strategies in partial linear model, Poisson regression model, and Weibull censored regression model [13–15].

To formulate the problem, we suppose that the vector of the fixed effects parameter β in the LMM can be partitioned into two sub-vectors $\beta = (\beta'_1, \beta'_2)'$, where β_1 is the coefficient vector of non-sparse predictors and β_2 is the coefficient vector of sparse predictors. Our interest lies in the estimation of β_1 when β_2 is close to zero. To deal with this problem in the context of low dimensional data, ref. [16] propose an improved estimation strategy using sub-model selection and post-estimation for the LMM. Within this framework, linear shrinkage and shrinkage pretest estimation strategies are developed, which combine full model and sub-model estimators in an effective way as a trade-off between bias and variance. Ref. [17] extend this study by using a likelihood ratio test to develop James–Stein shrinkage and pretest estimation methods based on LMM for longitudinal data. In addition, the non-penalty estimators are compared with several penalty estimators (LASSO, adaptive LASSO and Elastic Net) for best performance.

In most real data situations, there is also the problem of multicollinearity among predictor variables for high-dimensional data. Various biased estimation techniques such as shrinkage estimation, partial least squares estimation [18] and Liu estimators [19] have been implemented to deal with this problem, but the widely used technique is ridge estimation [20]. The ridge estimator overcomes the weakness of the least squares estimator with a smaller mean squared error. To overcome and combat multicollinearity, ref. [21] propose pretest and Stein-type ridge regression estimators for linear and partially linear

models. Furthermore, ref. [22] also develop shrinkage estimation based on Liu regression to overcome multicollinearity in linear models.

Our primary focus is on the estimation and prediction problem for linear mixed effect models when there are many potential predictors that have a weak or no influence on the response of interest. This method simultaneously controls overfitting using general least square estimation with a roughness penalty. We propose pretest and shrinkage estimation strategies using the ridge estimation technique as a base estimator and numerically compare their performance with the LASSO and adaptive LASSO estimators. Our proposed estimation strategy is applied to both high-dimensional and low-dimensional data.

The rest of this article is organized as follows. In Section 2, we present the linear mixed effect model and the proposed estimation techniques. We introduce the full and sub-model estimators based on ridge estimation. Thereafter, we construct the pretest and shrinkage ridge estimators. Section 3 provides the asymptotic bias and risk of these estimators. A Monte Carlo simulation is used to evaluate the performance of the estimators including a comparison with the lasso-type estimators, and the results are reported in Section 4. Section 5 presents a demonstration of the proposed methodology on a high-dimensional resting-state effective brain connectivity and genetic data. We also illustrate the proposed estimation methods in an application to a low-dimensional Amsterdam growth and health study. Section 6 presents a discussion with recommendations.

2. Model and Estimation Strategies

In this section, we present the linear mixed effect model and the proposed estimation strategies.

2.1. Linear Mixed Model

Suppose that we have a sample of N subjects. For the i^{th} subject, we collect the response variable y_{ij} for the j th time, where $i = 1 \dots, n; j = 1 \dots, n_i$ and $N = \sum_{i=1}^n n_i$. Let $\mathbf{Y}_i = (y_{i1}, \dots, y_{in_i})'$ denotes the $n_i \times 1$ vector of responses from the i th subject. Let $\mathbf{X}_i = (x_{i1}, \dots, x_{in_i})'$ and $\mathbf{Z}_i = (z_{i1}, \dots, z_{in_i})'$ be $n_i \times p$ and $n_i \times q$ known fixed-effects and random-effect design matrix for the i th subject of full rank p and q , respectively. The linear mixed effect model [1] for a vector of repeated responses \mathbf{Y}_i on the i th subject is assumed to have the form

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{a}_i + \boldsymbol{\epsilon}_i, \quad (1)$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is the $p \times 1$ vector of unknown fixed-effect parameters or regression coefficients, \mathbf{a}_i is the $q \times 1$ vector of unobservable random effects for the i th subject, assumed to come from a multivariate normal distribution with zero mean and a covariance matrix \mathbf{G} , where \mathbf{G} is an unknown $q \times q$ covariance matrix and $\boldsymbol{\epsilon}_i$ denotes $n_i \times 1$ vector of error terms assumed to be normally distributed with zero mean, covariance matrix $\sigma^2\mathbf{I}_{n_i}$. Further, $\boldsymbol{\epsilon}_i$ are assumed to be independent of the random effects \mathbf{a}_i .

The marginal distribution for the response y_i is normal with mean $\mathbf{X}_i\boldsymbol{\beta}$ and covariance matrix $Cov(\mathbf{Y}_i) = \mathbf{Z}_i\sigma_i^2\mathbf{Z}_i^T + \sigma^2\mathbf{I}_n$. By stacking the vectors, the mixed model can be expressed as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \boldsymbol{\epsilon}$. From the Equation (1), the distribution of the model follows $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, where $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ with covariance, $\mathbf{V} = \sum_{i=1}^n \mathbf{Z}_i\sigma_i^2\mathbf{Z}_i^T + \sigma^2\mathbf{I}_n$.

2.2. Ridge Full Model and Sub-Model Estimator

The generalized least square estimator (GLS) is defined as $\hat{\boldsymbol{\beta}}^{\text{GLS}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$ and the ridge full model estimator can be obtained by introducing a penalized regression

so that $\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \left\{ (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + k\boldsymbol{\beta}^T\boldsymbol{\beta} \right\}$ and

$\hat{\boldsymbol{\beta}}^{\text{Ridge}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$, where $\hat{\boldsymbol{\beta}}^{\text{Ridge}}$ is the ridge full model estimator and $k \in [0, \infty)$ is the tuning parameter. If $k = 0$, $\hat{\boldsymbol{\beta}}^{\text{Ridge}}$ is the GLS estimator and $\hat{\boldsymbol{\beta}}^{\text{Ridge}} = 0$ for k is sufficiently large. We select the value of k using 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 non-sparse predictors and \mathbf{X}_2 is an $n \times p_2$ sub-matrix that contains the sparse predictors. Accordingly, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ 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 is defined as $Y = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \boldsymbol{\epsilon}$ subject to $\boldsymbol{\beta}^T \boldsymbol{\beta} \leq \phi$ and $\boldsymbol{\beta}_2 = \mathbf{0}$ which corresponds to $Y = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}\mathbf{a} + \boldsymbol{\epsilon}$ subject to $\boldsymbol{\beta}_1^T \boldsymbol{\beta}_1 \leq \phi$. The sub-model estimator $\hat{\boldsymbol{\beta}}_1^{\text{RSM}}$ of $\boldsymbol{\beta}_1$ has the form $\hat{\boldsymbol{\beta}}_1^{\text{RSM}} = (\mathbf{X}_1^T \mathbf{V}^{-1} \mathbf{X}_1 + k\mathbf{I})^{-1} \mathbf{X}_1^T \mathbf{V}^{-1} \mathbf{Y}$. We denote $\hat{\boldsymbol{\beta}}_1^{\text{RFM}}$ as the full model ridge estimator of $\boldsymbol{\beta}_1$ and given as

$$\hat{\boldsymbol{\beta}}_1^{\text{RFM}} = (\mathbf{X}_1^T \mathbf{V}^{-1/2} \mathbf{M}_{\mathbf{X}_2} \mathbf{V}^{-1/2} \mathbf{X}_1 + k\mathbf{I})^{-1} \mathbf{X}_1^T \mathbf{V}^{-1/2} \mathbf{M}_{\mathbf{X}_2} \mathbf{V}^{-1/2} \mathbf{Y}, \text{ where } \mathbf{M}_{\mathbf{X}_2} = \mathbf{I} - \mathbf{P} = \mathbf{I} - \mathbf{V}^{-1/2} \mathbf{X}_2 (\mathbf{X}_2 \mathbf{V}^{-1} \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{V}^{-1/2}.$$

2.3. Pretest Ridge Estimation Strategy

Generally, the sub-model estimator will be more efficient than the full model estimator if the information embodied in the imposed linear restrictions is valid, thus $\boldsymbol{\beta}_2$ is close to zero. However, if the information is not valid the sub-model estimator is likely to be more biased and may have a higher risk than the full model estimator. There is, therefore, some doubt as to whether or not to impose the restrictions on the model’s parameter. It is in response to this uncertainty that a statistical test may be used to determine the validity of the proposed restrictions. Accordingly, the procedure to follow in practice is pretest the validity of the restrictions and if the outcome of the pretest suggests that they are correct then the model parameters are estimated incorporating the restrictions. If the pretest rejects the restrictions then the parameters are estimated from the sample information alone. This motivates the consideration of the pretest estimation strategy for the LMM.

The pretest estimator is a combination of the full model estimator $\hat{\boldsymbol{\beta}}_1^{\text{RFM}}$, and sub-model estimator $\hat{\boldsymbol{\beta}}_1^{\text{RSM}}$, through an indicator function $I(L_n \leq d_{n,\alpha})$, where L_n is an appropriate test statistic to test $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_A : \boldsymbol{\beta}_2 \neq \mathbf{0}$. Moreover, $d_{n,\alpha}$ is an α level critical value based on distribution of L_n under H_0 . We define test statistics based on the log-likelihood ratio test as $L_n = 2\{\ell^*(\hat{\boldsymbol{\beta}}^{\text{RFM}} | \mathbf{Y}) - \ell^*(\hat{\boldsymbol{\beta}}^{\text{RSM}} | \mathbf{Y})\}$.

Under H_0 , the test statistic L_n follows asymptotic chi-square distribution with p_2 degrees of freedom. The pretest test ridge estimator $\hat{\boldsymbol{\beta}}_1^{\text{RPT}}$ of $\boldsymbol{\beta}_1$ is then defined by

$$\hat{\boldsymbol{\beta}}_1^{\text{RPT}} = \hat{\boldsymbol{\beta}}_1^{\text{RFM}} - (\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \hat{\boldsymbol{\beta}}_1^{\text{RSM}})I(L_n \leq d_{n,\alpha}), \quad p_2 \geq 1.$$

2.4. Shrinkage Ridge Estimation Strategy

The pre-test estimator is a discontinuous function of the sub-model $\hat{\boldsymbol{\beta}}_1^{\text{RSM}}$ and full model $\hat{\boldsymbol{\beta}}_1^{\text{RFM}}$, which depends on the hard threshold ($d_{n,\alpha} = \chi_{p_2,\alpha}^2$). We address this limitation by defining the shrinkage ridge estimator based on soft thresholding. The shrinkage ridge estimator (RSE) of $\boldsymbol{\beta}_1$, denoted as $\hat{\boldsymbol{\beta}}_1^{\text{RSE}}$, is defined as

$$\hat{\boldsymbol{\beta}}_1^{\text{RSE}} = \hat{\boldsymbol{\beta}}_1^{\text{RSM}} + (\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \hat{\boldsymbol{\beta}}_1^{\text{RSM}})(1 - (p_2 - 2)L_n^{-1}), \quad p_2 \geq 3.$$

Here, $\hat{\boldsymbol{\beta}}_1^{\text{RSE}}$ is the linear combination of the full model $\hat{\boldsymbol{\beta}}_1^{\text{RFM}}$ and sub-model $\hat{\boldsymbol{\beta}}_1^{\text{RSM}}$ estimates. If $L_n \leq (p_2 - 2)$, then a relatively large weight is placed on $\hat{\boldsymbol{\beta}}_1^{\text{RSM}}$ otherwise, more weight is on $\hat{\boldsymbol{\beta}}_1^{\text{RFM}}$. A setback with $\hat{\boldsymbol{\beta}}_1^{\text{RSE}}$ is that it is not a convex combination of $\hat{\boldsymbol{\beta}}_1^{\text{RFM}}$ and $\hat{\boldsymbol{\beta}}_1^{\text{RSM}}$. This can cause over-shrinkage, which gives the estimator opposite sign of $\hat{\boldsymbol{\beta}}_1^{\text{RFM}}$. This could happen if $(p_2 - 2)L_n^{-1}$ is larger than one. To counter this, we use the positive-part shrinkage ridge estimator (RPS) defined as

$$\hat{\boldsymbol{\beta}}_1^{\text{RPS}} = \hat{\boldsymbol{\beta}}_1^{\text{RSM}} + (\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \hat{\boldsymbol{\beta}}_1^{\text{RSM}})(1 - (p_2 - 2)L_n^{-1})^+, \quad p_2 \geq 3$$

where $(1 - (p_2 - 2)L_n^{-1})^+ = \max(0, 1 - (p_2 - 2)L_n^{-1})$. The RPS estimator will control possible over-shrinking in the RSE estimator.

3. Asymptotic Results

In this section, we derive the asymptotic distributional bias and risk of the estimators considered in Section 2. We examine the properties of the estimators for increasing n and as β_2 approaches the null vector under the sequence of local alternatives defined as

$$K_n : \beta_2 = \beta_{2(n)} = \frac{\kappa}{\sqrt{n}}, \tag{2}$$

where $\kappa = (\kappa_1, \kappa_2, \dots, \kappa_{p_2})' \in \mathbb{R}^{p_2}$ is a fixed vector. The vector $\frac{\kappa}{\sqrt{n}}$ is a measure of how far local alternatives K_n differ from the subspace information $\beta_2 = \mathbf{0}$. In order to evaluate the performance of the estimators, we define the asymptotic distributional bias of the estimator $\hat{\beta}_1^*$ as

$$ADB(\hat{\beta}_1^*) = \lim_{n \rightarrow \infty} E\{\sqrt{n}(\hat{\beta}_1^* - \beta_1)\},$$

In order to compute the risk functions, we first compute the asymptotic covariance of the estimators. The asymptotic covariance of an estimator $\hat{\beta}_1^*$ is expressed as

$$Cov(\hat{\beta}_1^*) = \lim_{n \rightarrow \infty} E\{n(\hat{\beta}_1^* - \beta_1)(\hat{\beta}_1^* - \beta_1)^T\}.$$

Following the asymptotic covariance matrix, we define the asymptotic risk of an estimator $\hat{\beta}_1^*$ as $R(\hat{\beta}_1^*) = \text{tr}(\mathbf{Q}Cov(\hat{\beta}_1^*))$. \mathbf{Q} is a positive definite matrix of weights with dimensions of $p \times p$. We set $\mathbf{Q} = \mathbf{I}$ in this study.

Assumption 1. We make the following two regularity conditions to establish the asymptotic properties of the estimators.

1. $\frac{1}{n} \max_{1 \leq i \leq n} \mathbf{x}_i^T [\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}]^{-1} \mathbf{x}_i \rightarrow \mathbf{0}$ as $n \rightarrow \infty$, where \mathbf{x}_i^T is the i th row of \mathbf{X} .
2. $\mathbf{B}_n = n^{-1} [\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}]^{-1} \rightarrow \mathbf{B}$, for some finite $\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$.

Theorem 1. For $k < \infty$, If $k/\sqrt{n} \rightarrow \lambda_0$ and \mathbf{B} is non-singular, the distribution of the full model ridge estimator, $\hat{\beta}_n^{\text{RFM}}$ is

$$\sqrt{n}(\hat{\beta}_n^{\text{RFM}} - \beta) \xrightarrow{D} \mathcal{N}(-\lambda_0 \mathbf{B}^{-1} \beta, \mathbf{B}^{-1}),$$

where \xrightarrow{D} denotes convergence in distribution.

Proof. See Theorem 2 in [23]. \square

Proposition 1. Assuming the above assumption 1 together with Theorem 1 hold, under the local alternatives K_n , we have

$$\begin{aligned} \begin{pmatrix} \varphi_1 \\ \varphi_3 \end{pmatrix} &\xrightarrow{D} \mathcal{N}\left[\begin{pmatrix} -\mu_{11.2} \\ \delta \end{pmatrix}, \begin{pmatrix} \mathbf{B}_{11.2}^{-1} & \mathbf{\Phi} \\ \mathbf{\Phi} & \mathbf{\Phi} \end{pmatrix}\right], \\ \begin{pmatrix} \varphi_3 \\ \varphi_2 \end{pmatrix} &\xrightarrow{D} \mathcal{N}\left[\begin{pmatrix} \delta \\ -\gamma \end{pmatrix}, \begin{pmatrix} \mathbf{\Phi} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{11}^{-1} \end{pmatrix}\right], \end{aligned}$$

where $\varphi_1 = \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1)$, $\varphi_2 = \sqrt{n}(\hat{\beta}_1^{\text{RSM}} - \beta_1)$, $\varphi_3 = \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}})$, $\gamma = \mu_{11.2} + \delta$, $\delta = \mathbf{B}_{11}^{-1} \mathbf{B}_{12} \kappa$, $\mathbf{\Phi} = \mathbf{B}_{11}^{-1} \mathbf{B}_{12} \mathbf{B}_{22.1}^{-1} \mathbf{B}_{21} \mathbf{B}_{11}^{-1}$, $\mathbf{B}_{22.1} = \mathbf{B}_{22} - \mathbf{B}_{21} \mathbf{B}_{11}^{-1} \mathbf{B}_{12}$, $\mu = -\lambda_0 \mathbf{B}^{-1} \beta = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\mu_{11.2} = \mu_1 - \mathbf{B}_{12} \mathbf{B}_{22}^{-1} ((\beta_2 - \kappa) - \mu_2)$.

Proof. See Appendix A \square

Theorem 2. Under the condition of Theorem 1 and the local alternatives K_n , the ADBs of the proposed estimators are

$$\begin{aligned} ADB(\hat{\beta}_1^{RFM}) &= -\mu_{11.2}, \\ ADB(\hat{\beta}_1^{RSM}) &= -\mu_{11.2} - \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\delta = -\gamma, \\ ADB(\hat{\beta}_1^{RPT}) &= -\mu_{11.2} - \delta\mathbf{H}_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta), \\ ADB(\hat{\beta}_1^{RSE}) &= -\mu_{11.2} - (p_2 - 2)\delta E(\chi_{p_2+2}^{-2}(\Delta)), \\ ADB(\hat{\beta}_1^{RPS}) &= -\mu_{11.2} - \delta\mathbf{H}_{p_2+2}(\chi_{p_2-2}^2; \Delta) - (p_2 - 2)\delta E\{\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{-2} > p_2 - 2)\}, \end{aligned}$$

where $\Delta = \kappa^T\mathbf{B}_{22.1}^{-1}\kappa$, $\mathbf{B}_{22.1} = \mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}$, and $H_v(x; \Delta)$ is the cumulative distribution function of the non-central chi-squared distribution with non-centrality parameter Δ and v degrees of freedom, and $E(\chi_v^{-2j}(\Delta))$ is the expected value of the inverse of a non-central χ^2 distribution with v degrees of freedom and non-centrality parameter Δ ,

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

Proof. See Appendix B.1 \square

Since the ADBs of the estimators are in non-scalar form, we define the following asymptotic quadratic bias (AQDB) of $\hat{\beta}_1^*$ by

$$AQDB(\hat{\beta}_1^*) = \left(ADB(\hat{\beta}_1^*) \right)' \mathbf{B}_{11.2} \left(ADB(\hat{\beta}_1^*) \right),$$

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

Corollary 1. Suppose Theorem 2 holds. Then, under $\{K_n\}$, the AQDBs of the estimators are

$$\begin{aligned} AQDB(\hat{\beta}_1^{RFM}) &= \mu_{11.2}^T \mathbf{B}_{11.2} \mu_{11.2}, \\ AQDB(\hat{\beta}_1^{RSM}) &= \gamma^T \mathbf{B}_{11.2} \gamma, \\ AQDB(\hat{\beta}_1^{RPT}) &= \mu_{11.2}^T \mathbf{B}_{11.2} \mu_{11.2} + \mu_{11.2}^T \mathbf{B}_{11.2} \delta \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) \\ &\quad + \delta^T \mathbf{B}_{11.2} \mu_{11.2} \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) + \delta^T \mathbf{B}_{11.2} \delta \mathbf{H}_{p_2+2}^2(\chi_{p_2}^2; \Delta), \\ AQDB(\hat{\beta}_1^{RSE}) &= \mu_{11.2}^T \mathbf{B}_{11.2} \mu_{11.2} + (p_2 - 2)\mu_{11.2}^T \mathbf{B}_{11.2} \delta E(\chi_{p_2+2}^{-2}(\Delta)) \\ &\quad + (p_2 - 2)\delta^T \mathbf{B}_{11.2} \mu_{11.2} E(\chi_{p_2+2}^{-2}(\Delta)) + (p_2 - 2)^2 \delta^T \mathbf{B}_{11.2} \delta \left(E(\chi_{p_2+2}^{-2}(\Delta)) \right)^2, \\ AQDB(\hat{\beta}_1^{RPS}) &= \mu_{11.2}^T \mathbf{B}_{11.2} \mu_{11.2} + (\delta^T \mathbf{B}_{11.2} \mu_{11.2} + \mu_{11.2}^T \mathbf{B}_{11.2} \delta) [\mathbf{H}_{p_2+2}(p_2 - 2; \Delta) \\ &\quad + (p_2 - 2)E\{\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{-2}(\Delta) > p_2 - 2)\}] + \delta^T \mathbf{B}_{11.2} \delta \left[\mathbf{H}_{p_2+2}(p_2 - 2; \Delta) \right. \\ &\quad \left. + (p_2 - 2)E\{\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{-2}(\Delta) > p_2 - 2)\} \right]^2. \end{aligned}$$

When $\mathbf{B}_{11.2} = \mathbf{0}$, the AQDB of all estimators are equivalent, and the estimators are therefore asymptotically unbiased. If we assume that $\mathbf{B}_{11.2} \neq \mathbf{0}$, the results for the bias of the estimators can be summarized as follows:

1. The AQDB of $\hat{\beta}_1^{RSM}$ is an unbounded function of $\gamma^T \mathbf{B}_{11.2} \gamma$.

2. The AQDB of $\hat{\beta}_1^{RPT}$ starts from $\mu_{11.2}^T \mathbf{B}_{11.2} \mu_{11.2}$ at $\Delta = 0$, and when Δ increases, it increases to the maximum and then decreases to zero.
3. The characteristics of $\hat{\beta}_1^{RSE}$ and $\hat{\beta}_1^{RPS}$ are similar to $\hat{\beta}_1^{RPT}$. The AQDB of $\hat{\beta}_1^{RSE}$ and $\hat{\beta}_1^{RPS}$ similarly start from $\mu_{11.2}^T \mathbf{B}_{11.2} \mu_{11.2}$ at $\Delta = 0$, and increase to a point, and then decrease towards zero, since $E\{\chi_{p_2+2}^{-2}(\Delta)\}$ is a non-increasing on of Δ .

Theorem 3. Suppose Theorem 1 holds and under the local alternatives K_n , the covariance matrices of the estimators are

$$\begin{aligned}
 Cov(\hat{\beta}_1^{RFM}) &= \mathbf{B}_{11.2}^{-1} + \mu_{11.2} \mu_{11.2}^T, \\
 Cov(\hat{\beta}_1^{RSM}) &= \mathbf{B}_{11}^{-1} + \gamma \gamma^T, \\
 Cov(\hat{\beta}_1^{RPT}) &= \mathbf{B}_{11.2}^{-1} + \mu_{11.2} \mu_{11.2}^T + 2\mu_{11.2}^T \delta \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) - \Phi \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) \\
 &\quad + \delta \delta^T [2\mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) - \mathbf{H}_{p_2+4}(\chi_{p_2}^2; \Delta)], \\
 Cov(\hat{\beta}_1^{RSE}) &= \mathbf{B}_{11.2}^{-1} + \mu_{11.2} \mu_{11.2}^T + 2(p_2 - 2)\mu_{11.2}^T \delta E(\chi_{p_2+2}^{-2}(\Delta)) \\
 &\quad - (p_2 - 2)\Phi \left\{ 2E(\chi_{p_2+2}^{-2}(\Delta)) - (p_2 - 2)E(\chi_{p_2+2}^{-4}(\Delta)) \right\} \\
 &\quad + (p_2 - 2)\delta \delta^T \left\{ -2E(\chi_{p_2+4}^{-2}(\Delta)) + 2E(\chi_{p_2+2}^{-2}(\Delta)) + (p_2 - 2)E(\chi_{p_2+4}^{-4}(\Delta)) \right\}, \\
 Cov(\hat{\beta}_1^{RPS}) &= Cov(\hat{\beta}_1^{RSE}) + 2\delta \mu_{11.2}^T E\left(\left\{1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta)\right\} I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\right) \\
 &\quad - 2\Phi E\left(\left\{1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta)\right\} I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\right) \\
 &\quad - 2\delta \delta^T E\left(\left\{1 - (p_2 - 2)\chi_{p_2+4}^{-2}(\Delta)\right\} I(\chi_{p_2+4}^2(\Delta) \leq p_2 - 2)\right) \\
 &\quad + 2\delta \delta^T E\left(\left\{1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta)\right\} I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\right) \\
 &\quad - (p_2 - 2)^2 \Phi E\left(\chi_{p_2+2}^{-4}(\Delta) I(\chi_{p_2+2,\alpha}^2(\Delta) \leq p_2 - 2)\right) \\
 &\quad - (p_2 - 2)^2 \delta \delta^T E\left(\chi_{p_2+2,\alpha}^{-4}(\Delta) I(\chi_{p_2+2,\alpha}^2(\Delta) \leq p_2 - 2)\right) \\
 &\quad + \Phi \mathbf{H}_{p_2+2}(p_2 - 2; \Delta) + \delta \delta^T \mathbf{H}_{p_2+4}(p_2 - 2; \Delta).
 \end{aligned}$$

Proof. See Appendix B.2. \square

Corollary 2. Under the local alternatives (K_n) and from Theorem 3, the risk of the estimators are obtained as

$$\begin{aligned}
 R[\hat{\beta}_1^{RFM}] &= tr(\mathbf{Q}\mathbf{B}_{11.2}^{-1}) + \mu_{11.2}^T \mathbf{Q} \mu_{11.2}, \\
 R[\hat{\beta}_1^{RSM}] &= tr(\mathbf{Q}\mathbf{B}_{11}^{-1}) + \gamma^T \mathbf{Q} \gamma, \\
 R[\hat{\beta}_1^{RPT}] &= tr(\mathbf{Q}\mathbf{B}_{11.2}^{-1}) + \mu_{11.2}^T \mathbf{Q} \mu_{11.2} + 2\mu_{11.2}^T \mathbf{Q} \delta \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) \\
 &\quad - tr(\mathbf{Q}\Phi) \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) + \delta \mathbf{Q} \delta^T [2\mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) - \mathbf{H}_{p_2+4}(\chi_{p_2}^2; \Delta)], \\
 R[\hat{\beta}_1^{RSE}] &= tr(\mathbf{Q}\mathbf{B}_{11.2}^{-1}) + \mu_{11.2}^T \mathbf{Q} \mu_{11.2} + 2(p_2 - 2)\mu_{11.2}^T \mathbf{Q} \delta E(\chi_{p_2+2}^{-2}(\Delta)) \\
 &\quad - (p_2 - 2)tr(\mathbf{Q}\Phi) [E(\chi_{p_2+2}^{-2}(\Delta)) - (p_2 - 2)E(\chi_{p_2+2}^{-4}(\Delta))] \\
 &\quad + (p_2 - 2)\delta^T \mathbf{Q} \delta [2E(\chi_{p_2+2}^{-2}(\Delta)) - 2E(\chi_{p_2+4}^{-2}(\Delta)) - (p_2 - 2)E(\chi_{p_2+4}^{-4}(\Delta))], \\
 R[\hat{\beta}_1^{RPS}] &= R[\hat{\beta}_1^{RSE}] + 2\delta \mathbf{Q} \mu_{11.2}^T E\left(\{1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta)\} I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\right) \\
 &\quad - 2tr(\mathbf{Q}\Phi) E\left(\{1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta)\} I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\right) \\
 &\quad - 2\delta^T \mathbf{Q} \delta E(\{1 - (p_2 - 2)\chi_{p_2+4}^{-2}(\Delta)\} I(\chi_{p_2+4}^2(\Delta) \leq p_2 - 2)) \\
 &\quad + 2\delta^T \mathbf{Q} \delta E\left(\{1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta)\} I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\right) \\
 &\quad - (p_2 - 2)^2 tr(\mathbf{Q}\Phi) E\left(\chi_{p_2+2}^{-4}(\Delta) I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\right) \\
 &\quad - (p_2 - 2)^2 \delta^T \mathbf{Q} \delta E\left(\chi_{p_2+2}^{-4}(\Delta) I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\right) \\
 &\quad + tr(\mathbf{Q}\Phi) \mathbf{H}_{p_2+2}(p_2 - 2; \Delta) + \delta^T \mathbf{Q} \delta \mathbf{H}_{p_2+4}(p_2 - 2; \Delta).
 \end{aligned}$$

From Theorem 2, when $\mathbf{B}_{12} = \mathbf{0}$, the risks of estimators $\hat{\beta}_1^{RSM}$, $\hat{\beta}_1^{RPT}$, $\hat{\beta}_1^{RSE}$, and $\hat{\beta}_1^{RPS}$ are reduced to common value $tr(\mathbf{Q}\mathbf{B}_{11.2}^{-1}) + \mu_{11.2}^T \mathbf{Q} \mu_{11.2}$, the risk of $\hat{\beta}_1^{RFM}$. If $\mathbf{B}_{12} \neq \mathbf{0}$, the results can be summarized as follows:

1. The risk of $\hat{\beta}_1^{RFM}$ remains constant while the risk of $\hat{\beta}_1^{RSM}$ is an unbounded function of Δ since $\Delta \in [0, \infty)$.
2. The risk of $\hat{\beta}_1^{RPT}$ increases as Δ moves away from zero, achieves it maximum and then decreases towards the risk of the full model estimator.
3. The risk of $\hat{\beta}_1^{RFM}$ is smaller than the risk of $\hat{\beta}_1^{RPT}$ for small values in the neighborhood of Δ and for the rest of the parameter space, $\hat{\beta}_1^{RPT}$ outperforms $\hat{\beta}_1^{RFM}$, thus, $R[\hat{\beta}_1^{RFM}] > R[\hat{\beta}_1^{RPT}]$.
4. Comparing the risks of $\hat{\beta}_1^{RSE}$ and $\hat{\beta}_1^{RFM}$, it can be seen that the estimator $\hat{\beta}_1^{RSE}$ outperforms $\hat{\beta}_1^{RFM}$ that is, $R[\hat{\beta}_1^{RSE}] \leq R[\hat{\beta}_1^{RFM}]$ for all $\Delta \geq 0$.

4. Simulation Studies

In this section, we conduct a simulation study to assess the performance of the suggested estimators for finite samples. The criterion for comparing the performance of any estimator in our study is the mean square error. We simulate the response from the following LMM model

$$Y_i = \mathbf{X}_i \beta + \mathbf{Z}_i \mathbf{a}_i + \epsilon_i, \tag{3}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{n_i})$ with $\sigma^2 = 1$. We generate random effect covariate \mathbf{a}_i from a multivariate normal distribution with zero mean and covariance matrix $\mathbf{G} = 0.5\mathbf{I}_{2 \times 2}$, where

$I_{2 \times 2}$ is 2×2 identity matrix. The design matrix $X_i = (x_{i1}, \dots, x_{in_i})'$ is generated from a n_i -multivariate normal distribution with mean vector and covariance matrix Σ_x . Furthermore, we assume that the off-diagonal elements of the covariance matrix Σ_x are equal to ρ , which is the coefficient of correlation between any two predictors, with $\rho = 0.3, 0.7, 0.9$. The ratio of the largest eigenvalue to the smallest eigen-value of matrix $X^T V^{-1} X$ is calculated as a condition number index (CNI) [24], which assesses the existence of multicollinearity in the design matrix. If the CNI is larger than 30, then the model has significant multicollinearity. Our simulations are based on the linear mixed effects model in Equation (3) with $n = 60$ and 100 subjects.

We consider a situation when the model is assumed to be sparse. In this study, our interest lies in testing the hypothesis $H_0 : \beta_2 = 0$, and our goal is to estimate the fixed effect coefficient β_1 . We partition the fixed effects coefficients as $\beta = (\beta'_1, \beta'_2)' = (\beta'_1, 0_{p_2})'$. The coefficients β_1 and β_2 are p_1 and p_2 dimensional vectors, respectively, with $p = p_1 + p_2$.

In order to investigate the behavior of the estimators, we define $\Delta^* = \|\beta - \beta_0\|$, where $\beta_0 = (\beta_1^T, 0_{p_2})^T$ and $\|\cdot\|$ is the euclidean norm. We considered Δ^* values between 0 and 4. If $\Delta^* = 0$, then we will have $\beta = (1, 1, 1, 1, \underbrace{0, \dots, 0}_{p_2})^T$ to generate the response

under null hypothesis. On the other hand, when $\Delta^* \geq 0$, say $\Delta^* = 4$, we will have $\beta = (1, 1, 1, 1, \underbrace{4, 0, \dots, 0}_{p_2-1})^T$ to generate the response under the local alternative hypothesis.

In our simulation study, we consider the number of fixed effect or predictor variables as $(p_1, p_2) \in \{(5, 40), (5, 500), (5, 1000)\}$. Each realization is repeated 5000 times to obtain consistent results and compute the MSE of suggested estimators with $\alpha = 0.05$.

Based on the simulated data, we calculate the mean square error (MSE) of all the estimators as $MSE(\hat{\beta}) = \frac{1}{5000} \sum_{j=1}^{5000} (\hat{\beta} - \beta)^T (\hat{\beta} - \beta)$, where $\hat{\beta}$ denotes any one of $\hat{\beta}^{RSM}, \hat{\beta}^{RPT}, \hat{\beta}^{RSE}$ and $\hat{\beta}^{RPS}$, in the j th repetition. We use the relative mean squared efficiency (RMSE), or the ratio of MSE for risk performance comparison. The RMSE of an estimator $\hat{\beta}^*$ with respect to the baseline full model ridge estimator $\hat{\beta}_1^{RFM}$ is defined as $RMSE(\hat{\beta}_1^{RFM} : \hat{\beta}_1^*) = \frac{MSE(\hat{\beta}_1^{RFM})}{MSE(\hat{\beta}_1^*)}$, where $\hat{\beta}_1^*$ is one of the suggested estimators under consideration.

4.1. Simulation Results

In this subsection, we present the results from our simulation study. We report the results for $n = 60, 100$ and $p_1 = 5$ with different values of correlation coefficient ρ are shown in Table 1. Furthermore, we plot the RMSEs against Δ^* in Figures 1 and 2. The findings can be summarized as follows:

1. When $\Delta^* = 0$, the sub-model RSM outperforms all other estimators. As $\Delta^* = 0$ moves from zero, the RMSE of the sub-model decreases and goes to zero.
2. The pretest ridge estimator RPT outperforms shrinkage ridge and positive Stein ridge estimators in the case of $\Delta^* = 0$. However, for large number of sparse predictors p_2 while keeping p_1 and n fixed, RPT is less efficient than RPS and RSE. In the case of Δ^* being larger than zero, the RMSE of RPT decreases, and it remains below 1 for immediate values of Δ^* , after that the RMSE of RPT increases and approaches one for larger values of Δ^* .
3. RPS performs better than RSE in the entire parameter space induced by Δ^* as presented in Tables 1 and 2. Similarly, both shrinkage estimators RPS and RSE outperforms the full ridge model estimator irrespective of the corrected sub-model selected. This is consistent with the asymptotic theory presented in Section 3.
4. Δ^* which measures the degree of deviation from the Assumption 1 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 wrongly specified. As evident from Tables 1 and 2, Figures 1 and 2, if the selected sub-model is correct, that is, $\Delta^* = 0$, then the shrinkage estimators are relatively efficient compared with the ridge full model estimator. On the

other hand, if the sub-model is misspecified, the gain slowly diminishes. However, in terms of risk, the shrinkage estimators are at least as good as the full ridge model estimator. Therefore, the use of shrinkage estimators makes sense in application when a sub-model cannot be correctly specified.

5. The RMSE of the ridge-type estimators are an increasing function of the amount of multicollinearity. This indicates that the ridge-type estimators perform better than the classical estimator in the presence of multicollinearity among predictor variables.

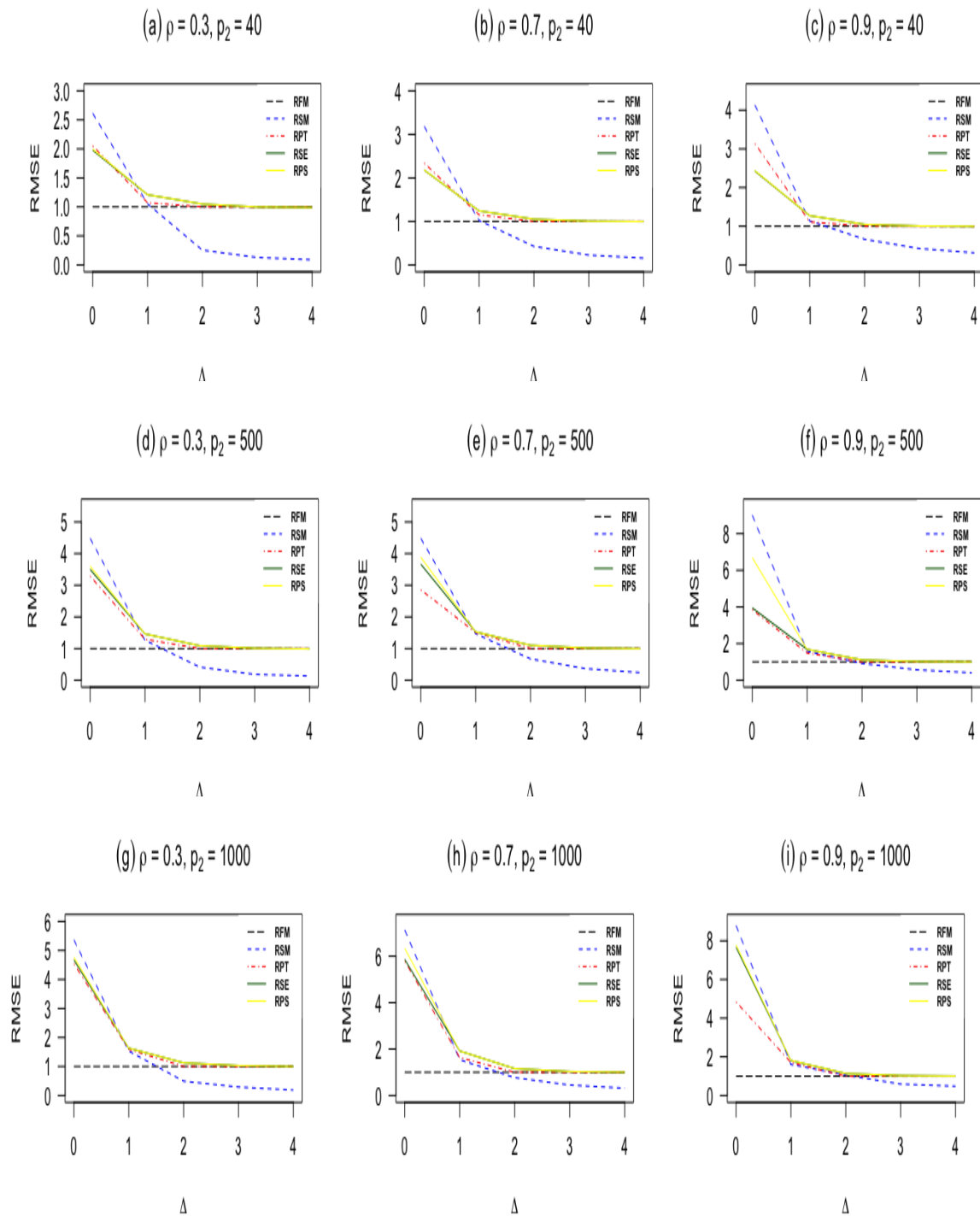


Figure 1. RMSE of estimators as a function of the non-centrality parameter Δ when $n = 60$, and $p_1 = 5$.

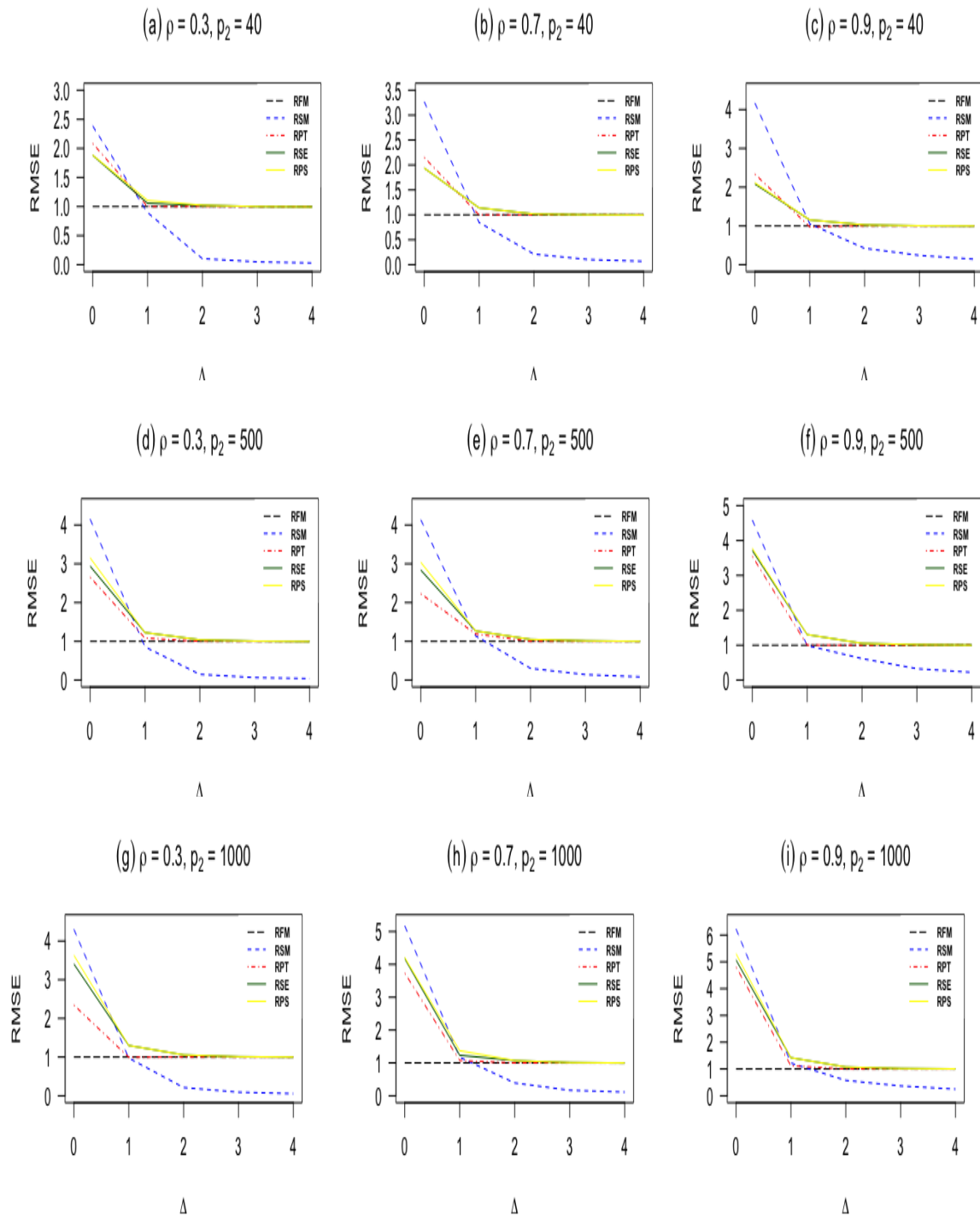


Figure 2. RMSE of estimators as a function of the non-centrality parameter Δ when $n = 100$, and $p_1 = 5$.

Table 1. RMSEs of RSM, RPT, RSE, and RPS estimators with respect to $\hat{\beta}_1^{\text{RFM}}$ when $\Delta \geq 0$ for $p_1 = 5$ and $n = 60$.

ρ	p_2	Δ	CNI	RSM	RPT	RSE	RPS
0.3	40	0	361	2.61	2.07	1.94	1.96
		1		1.05	1.07	1.20	1.25
		2		0.25	0.95	1.04	1.05
		3		0.12	0.98	0.99	1.00
		4		0.08	1.00	1.00	1.00
	500	0	613	4.48	3.29	3.48	1.96
		1		1.26	1.12	1.26	1.29
		2		0.41	0.97	1.08	1.09
		3		0.18	0.99	1.00	1.00
		4		0.13	1.00	1.00	1.00
	1000	0	693	5.36	4.53	4.67	4.71
		1		1.53	1.21	1.35	1.39
2		0.49		1.01	1.13	1.14	
3		0.28		0.99	0.99	0.99	
4		0.10		1.00	1.00	1.00	
0.7	40	0	1352	3.18	2.33	2.17	2.18
		1		1.04	1.11	1.20	1.23
		2		0.42	1.03	1.04	1.04
		3		0.23	0.98	0.99	1.00
		4		0.14	1.00	1.00	1.00
	500	0	1789	4.48	2.76	2.94	3.02
		1		1.08	1.43	1.52	1.53
		2		0.67	1.03	1.07	1.06
		3		0.35	0.98	1.00	1.00
		4		0.19	1.00	1.00	1.00
	1000	0	2134	6.82	5.24	5.30	3.02
		1		1.16	1.32	1.42	1.53
2		0.75		1.10	1.15	1.16	
3		0.39		0.99	1.00	1.00	
4		0.11		1.00	1.00	1.00	

Table 2. RMSEs of RSM, RPT, RSE, and RPS estimators with respect to $\hat{\beta}_1^{\text{RFM}}$ when $\Delta \geq 0$ for $p_1 = 5$, and $n = 100$.

ρ	p_2	Δ	CNI	RSM	RPT	RSE	RPS
0.3	40	0	150	2.38	2.09	1.88	1.90
		1		0.89	1.01	1.05	1.08
		2		0.21	0.94	1.01	1.02
		3		0.06	0.94	0.99	1.00
		4		0.02	1.00	1.00	1.00
	500	0	340	4.15	2.65	2.99	3.17
		1		0.87	1.08	1.18	1.21
		2		0.14	0.96	1.03	1.05
		3		0.06	0.99	0.99	1.00
		4		0.03	1.00	1.00	1.00
	1000	0	536	4.30	2.75	3.02	3.08
		1		0.96	1.09	1.13	1.15
2		0.21		0.8	1.03	1.03	
3		0.09		1.00	1.00	1.00	
4		0.04		1.00	1.00	1.00	
0.7	40	0	997	3.27	2.15	2.09	2.11
		1		0.85	1.02	1.09	1.10
		2		0.21	0.98	1.02	1.02
		3		0.06	0.99	0.99	0.99
		4		0.01	1.00	1.00	1.00
	500	0	1589	4.13	2.22	2.35	2.39
		1		1.04	1.19	1.21	1.20
		2		0.30	0.97	1.05	1.05
		3		0.14	1.00	1.00	1.00
		4		0.08	1.00	1.00	1.00
	1000	0	1751	5.17	3.71	4.03	4.09
		1		1.01	1.15	1.24	1.25
2		0.39		1.04	1.07	1.06	
3		0.16		0.99	1.00	1.00	
4		0.11		1.00	1.00	1.00	

4.2. Comparison with LASSO-Type Estimators

We compare our listed estimators with the LASSO and adaptive LASSO estimators. A 10-fold cross-validation is used for selecting the optimal value of the penalty parameters that minimizes the mean square errors for the LASSO-type estimators. The results for $\rho = 0.3, 0.7, 0.9$, $n = 60, 100$, $p_1 = 10$ and $p_2 = 50, 500, 1000, 2000$ are presented in Table 3. We observe the following from Table 3.

1. The performance of the sub-model estimator is the best among all estimators.
2. The pretest ridge estimator performs better than the other estimators. However, for larger values of sparse predictors p_2 the shrinkage estimators outperform the pretest estimator.
3. The performance of the LASSO and aLASSO estimators are comparable when ρ is small. The pretest and shrinkage estimators remain stable for a given value of ρ .
4. For a large number of sparse predictors p_2 , the shrinkage and pretest ridge estimators outperform the lasso-type estimators. This indicates the superiority of the shrinkage estimators over the LASSO-type estimators. Therefore shrinkage estimators are preferable when there is multicollinearity in our predictor variables.

Table 3. RMSEs of estimators with respect to $\hat{\beta}_1^{\text{RFM}}$ when $\Delta = 0$ for $p_1 = 10$.

n	ρ	p_2	CNI	RSM	RPT	RSE	RPS	LASSO	aLASSO
60	0.3	50	35.64	3.31	2.25	1.82	1.95	1.23	1.28
		500	452.76	4.13	3.71	2.61	3.01	1.47	1.52
		1000	1265.34	5.02	4.28	4.61	4.78	1.96	2.15
		2000	4567.56	7.13	5.10	6.18	6.39	2.70	3.06
	0.7	50	61.34	3.52	3.05	2.51	2.55	1.14	1.21
		500	743.17	4.49	3.65	3.41	3.50	1.36	1.58
		1000	2350.89	5.84	4.11	4.32	4.61	1.68	1.95
		2000	6908.39	8.10	5.31	6.24	6.29	1.84	2.02
	0.9	50	120.21	4.21	3.61	3.34	3.35	1.10	1.05
		500	950.98	4.82	3.3.8	3.72	3.73	1.21	1.16
		1000	5892.51	6.35	4.10	5.01	5.13	1.42	1.31
		2000	8352.73	8.51	4.63	5.24	5.38	1.61	1.35
100	0.3	50	31.21	2.91	2.54	2.12	2.23	1.32	1.36
		500	356.64	3.75	3.31	2.84	2.92	1.54	1.61
		1000	975.32	4.25	2.53	3.42	3.61	1.92	2.06
		2000	2764.84	5.61	4.25	4.91	5.08	2.31	2.46
	0.7	50	52.79	3.18	2.61	2.30	2.37	1.28	1.53
		500	578.43	4.28	3.05	3.52	3.59	1.46	2.07
		1000	1281.66	5.10	3.26	3.78	3.82	1.84	2.52
		2000	3498.30	6.12	3.01	4.26	4.33	2.27	2.41
	0.9	50	79.41	4.11	3.41	3.21	3.28	1.28	1.21
		500	681.43	4.35	3.55	3.41	3.50	1.43	1.51
		1000	1470.32	5.82	3.18	4.01	4.14	1.72	1.79
		2000	4105.90	7.04	4.57	5.22	5.32	1.87	1.96

5. Real Data Application

We consider two real data analyses using Amsterdam Growth and Health Data and a genetic and brain network connectivity edge weight data to illustrate the performance of the proposed estimators.

5.1. Amsterdam Growth and Health Data (AGHD)

The AGHD data is obtained from the Amsterdam Growth and Health Study [25]. The goal of this study is to investigate the relationship between lifestyle and health in adolescence into young adulthood. The response variable Y is the total serum cholesterol measured over six time points. There are five covariates: X_1 is the baseline fitness level measured as the maximum oxygen uptake on a treadmill, X_2 is the amount of body fat estimated by the sum of the thickness of four skinfolds, X_3 is a smoking indicator (0 = no, 1 = yes), X_4 is the gender (1 = female, 2 = male), and time measurement as X_5 and subject specific random effects.

A total of 147 subjects participated in the study where all variables were measured at $n_i = 6$ time occasions. In order to apply the proposed methods, firstly, we apply a variable selection based on AIC procedure to select the sub-model. For the AGHD data, we fit a linear mixed model with all the five covariates for both fixed and subject specific random effects by two stage selection procedure for the purpose of choosing both the random and fixed effects. The analysis found X_2 and X_5 to be significant covariates for prediction of the response variable serum cholesterol and the other variables are ignored since they are not significantly important. Based on this information, a sub-model is chosen to be X_2 and X_5 and the full model includes all the covariates. We construct the shrinkage estimators from the full-model and sub-model. In terms of null hypothesis, the restriction can be written as $\beta_2 = (\beta_1, \beta_3, \beta_4) = (0, 0, 0)$ with $p = 5$, $p_1 = 2$ and $p_2 = 3$.

To evaluate the performance of the estimators, we obtain the mean square prediction error (MSPE) using bootstrap samples. We draw 1000 bootstrap samples of the 147 subjects

from the data matrix $\{(Y_{ij}, X_{ij}), i = 1, 2, \dots, 147; j = 1, 2, \dots, 6\}$. We then calculate the relative prediction error (RPE) of $\hat{\beta}_1^*$ with respect to $\hat{\beta}_1^{\text{RFM}}$, the full model estimator. The RPE is defined as

$$\text{RPE}(\hat{\beta}_1^{\text{RFM}} : \hat{\beta}_1^*) = \frac{\text{MSPE}(\hat{\beta}_1^*)}{\text{MSPE}(\hat{\beta}_1^{\text{RFM}})} = \frac{(\mathbf{Y} - \mathbf{X}_1 \hat{\beta}_1^*)'(\mathbf{Y} - \mathbf{X}_1 \hat{\beta}_1^*)}{(\mathbf{Y} - \mathbf{X}_1 \hat{\beta}_1^{\text{RFM}})'(\mathbf{Y} - \mathbf{X}_1 \hat{\beta}_1^{\text{RFM}})},$$

where $\hat{\beta}_1^*$ is one of the listed estimators. If $\text{RPE} < 1$, then $\hat{\beta}_1^*$ outperforms $\hat{\beta}_1^{\text{RFM}}$.

Table 4 reports the estimates, standard error of the non-sparse predictors and RPEs of the estimators with respect to the full model. As expected, the sub-model ridge estimator $\hat{\beta}_1^{\text{RSM}}$ has the minimum RPE because it is computed when the sub-model is correct, that is, $\Delta^* = 0$. It is evident by the RPE values in Table 4 that the shrinkage estimators are superior to the LASSO-type estimators. Furthermore, the positive shrinkage is more efficient than the shrinkage ridge estimator.

Table 4. Estimate, standard error for the active predictors and RPEs of estimators with respect to full-model estimator for the Amsterdam Growth and Health Study data.

	RFM	RSM	RPT	RSE	RPS	LASSO	aLASSO
Estimate(β_2)	0.381	0.395	0.392	0.389	0.390	0.624	0.611
Standard error	0.104	0.102	0.100	0.009	0.008	0.081	0.079
Estimate (β_5)	0.137	0.125	0.131	0.130	0.133	0.101	0.105
Standard error	0.012	0.010	0.009	0.011	0.010	0.013	0.012
RPE	1.000	0.723	0.841	0.838	0.831	0.986	0.973

5.2. Resting-State Effective Brain Connectivity and Genetic Data

This data comprises longitudinal resting-state functional magnetic resonance imaging (rs-fMRI) effective brain connectivity network and genetic study [26] data obtained from a sample of 111 subjects with a total of 319 rs-fMRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The 111 subjects comprise 36 cognitively normal (CN), 63 mild cognitive impairment (MCI) and 12 Alzheimer's Disease (AD) subjects. The response is a network connection between regions of interest estimated from an rs-fMRI scan within the Default Mode Network (DMN), and we observe a longitudinal sequence of such connections for each subject with the number of repeated measurements. The DMN consists of a set of brain regions that tend to be active in resting-state, when a subject is mind wandering with no intended task. For this data analysis, we consider the network edge weight from the left intraparietal cortex to posterior cingulate cortex (LIPC \rightarrow PCC) as our response. The genetic data are single nucleotide polymorphism (SNPs) from non-sex chromosomes, i.e., chromosome 1 to chromosome 22. SNPs with minor allele frequency less than 5% are removed as are SNPs with a Hardy-Weinberg equilibrium p-value lower than 10^{-6} or a missing rate greater than 5%. After preprocessing we are left with 1,220,955 SNPs and the longitudinal rs-fMRI effective connectivity network using the 111 subjects with rs-fMRI data. The response is network edge weight. There are SNPs which are the fixed effects and subject specific random effects.

In order to apply the proposed methods, we use a genome-wide association study (GWAS) for screening the genetic data to 100 SNPs. We implement a second screening by applying multinomial logistic regression to identify a smaller subset of the 100 SNPs that are potentially associated with disease (CN/MCI/AD). This yields a subset of top 10 SNPs. This showed the top 10 SNPs are the most important predictors and the other 90 SNPs are ignored as not significant. We now have two models, which are the full model with all 100 SNPs and sub-model with 10 SNPs selected. Finally, we construct the pretest and shrinkage estimators from the full-model and sub-model.

We draw 1000 bootstrap samples with replacements from the corresponding data matrix $\{(Y_{ij}, \mathbf{X}_{ij}), i = 1, \dots, 111; j = 1, \dots, n_i\}$. We report the RPE of the estimators based on the bootstrap simulation with respect to the full model ridge estimator in Table 5. We observe that the RPE of the sub-model, pretest, shrinkage and positive shrinkage ridge estimators outperforms the full model estimator. Clearly, the sub-model ridge estimator has the smallest RPE since it's computed when the candidate sub-model is correct, i.e., $\Delta = 0$. Both shrinkage ridge estimators outperform the pretest ridge estimator. Particularly, the positive shrinkage performed better than the shrinkage estimator. The performance of both shrinkage and pretest ridge estimators are better than the LASSO-type estimators. Thus, the data analysis is in line with our simulation and theoretical findings.

Table 5. RPEs of estimators.

	RFM	RSM	RPT	RSE	RPS	LASSO	aLASSO
RPE	1.000	0.802	0.947	0.932	0.928	1.051	1.190

6. Conclusions

In this paper, we present efficient estimation strategies for the linear mixed effect model when there exists multicollinearity among predictor variables for high-dimensional data application. We considered the estimation of fixed effects parameters in the linear mixed model when some of the predictors may have a very weak influence on the response of interest. We introduced pretest and shrinkage estimation in our model using the ridge estimation as the reference estimator. In addition, we established the asymptotic properties of the pretest and shrinkage ridge estimators. Our theoretical findings demonstrate that the shrinkage ridge estimators outperform the full model ridge estimator and perform relatively better than the sub-model estimator in a wide range of the parameter space.

Additionally, a Monte Carlo simulation was conducted to investigate and assess the finite sample behavior of proposed estimators when the model is sparse (restrictions on parameters hold). As expected, the sub-model ridge estimator outshines all other estimators when the restrictions hold. However, when this assumption is violated, the shrinkage and pretest ridge estimators outperform the sub-model estimator. Furthermore, when the number of sparse predictors are extremely large relative to the sample size, the shrinkage estimators outperform the pretest ridge estimator. These numerical results are consistent with our asymptotic result. We also assess the relative performance of the LASSO-type estimators with our ridge-type estimators. We observe that the performance of pretest and shrinkage ridge estimators are superior to the LASSO-type estimators when predictors are highly correlated. For our real data application, the shrinkage ridge estimators are superior with the smallest relative prediction error compared to the LASSO-type estimators.

In summary, the results of the data analyses strongly confirm the findings of the simulation study and suggest the use of the shrinkage ridge estimation strategy when no prior information about the parameter subspace is available. The results of our simulation study and real data application are consistent with available results in [27–29].

In our future work, we will focus on other penalty estimators like the Elastic-Net, the minimax concave penalty (MCP), and the smoothly clipped absolute deviation method (SCAD) as estimation strategy in LMM for high-dimensional data. These estimators will be assessed and compared with the proposed ridge-type estimators. Another interesting extension will be integrating two sub-models by incorporating ridge-type estimation strategies in the linear mixed effect models. The goal is to improve the estimation accuracy of the non-sparse set of the fixed effects parameters by combining an over-fitted model estimator with an under-fitted one [27,29]. This approach will include combining two sub-models produced by two different variable selection techniques from the LMM [28].

Author Contributions: Conceptualization, E.A.O. and S.E.A.; methodology, E.A.O. and F.S.N.; formal analysis, E.A.O.; writing—original draft preparation, E.A.O.; writing—review and editing, E.A.O., S.E.A. and F.S.N.; supervision, F.S.N. and S.E.A.; funding acquisition, F.S.N. and S.E.A. All authors have read and agreed to the published version of the manuscript.

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here <https://pubmed.ncbi.nlm.nih.gov/22434862/> (accessed on 20 April 2021).

Acknowledgments: Research is supported by the Visual and Automated Disease Analytics (VADA) graduate training program.

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

Appendix A

Proof of Proposition 1. The asymptotic relationship between the sub-model and full model estimators of β_1 , we use the argument and equation: $\hat{Y} = Y - X_2 \hat{\beta}_2^{\text{RFM}}$, where

$$\begin{aligned}\hat{\beta}_1^{\text{RFM}} &= \arg \min_{\beta_1} \left\{ (\hat{Y} - X_1 \beta_1)^T V^{-1} (\hat{Y} - X_1 \beta_1) + \lambda \|\beta_1\|^2 \right\} \\ &= [X_1^T V^{-1} X_1 + \lambda I_{p_1}]^{-1} X_1^T V^{-1} \hat{Y} \\ &= [X_1^T V^{-1} X_1 + \lambda I_{p_1}]^{-1} X_1^T V^{-1} Y - [X_1^T V^{-1} X_1 + \lambda I_{p_1}]^{-1} X_1^T V^{-1} X_2 \hat{\beta}_2^{\text{RFM}} \\ &= \hat{\beta}_1^{\text{RSM}} - [X_1 V^{-1} X_1 + \lambda I_{p_1}]^{-1} X_1^T V^{-1} X_2 \hat{\beta}_2^{\text{RFM}} \\ &= \hat{\beta}_1^{\text{RSM}} - B_{11}^{-1} B_{12} \hat{\beta}_2^{\text{RFM}}\end{aligned}$$

From Theorem 1, we partition $\sqrt{n}(\hat{\beta}^{\text{RFM}} - \beta)$ as $\sqrt{n}(\hat{\beta}^{\text{RFM}} - \beta) = (\sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1), \sqrt{n}(\hat{\beta}_2^{\text{RFM}} - \beta_2))$. We obtain $\sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1) \xrightarrow{D} \mathcal{N}_{p_1}(-\mu_{11.2}, B_{11.2}^{-1})$, where $B_{11.2}^{-1} = B_{11} - B_{12} B_{22}^{-1} B_{21}$. We have shown that $\hat{\beta}_1^{\text{RSM}} = \hat{\beta}_1^{\text{RFM}} + B_{11}^{-1} B_{12} \hat{\beta}_2^{\text{RFM}}$. Using this expression and under the local alternative $\{K_n\}$, we obtain the following expressions

$$\begin{aligned}\varphi_2 &= \sqrt{n}(\hat{\beta}_1^{\text{RSM}} - \beta_1) \\ &= \sqrt{n}(\hat{\beta}_1^{\text{RFM}} + B_{11}^{-1} B_{12} \hat{\beta}_2^{\text{RFM}} - \beta_1) \\ &= \varphi_1 + B_{11}^{-1} B_{12} \sqrt{n} \hat{\beta}_2^{\text{RFM}}, \\ \varphi_3 &= \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}}) \\ &= \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1) - \sqrt{n}(\hat{\beta}_1^{\text{RSM}} - \beta_1) \\ &= \varphi_1 - \varphi_2.\end{aligned}$$

Since φ_2 and φ_3 are linear functions of φ_1 , as $n \rightarrow \infty$, they are also asymptotically normally distributed. Their mean vectors and covariance matrices are as follows:

$$\begin{aligned}
 E(\varphi_1) &= E\left(\sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1)\right) = -\mu_{11.2} \\
 E(\varphi_2) &= E\left(\varphi_1 + \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\sqrt{n}\hat{\beta}_2^{\text{RFM}}\right) \\
 &= E(\varphi_1) + \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\sqrt{n}E(\hat{\beta}_2^{\text{RFM}}) \\
 &= -\mu_{11.2} + \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\kappa = -(\mu_{11.2} - \delta) = -\gamma \\
 E(\varphi_3) &= E(\varphi_1 - \varphi_2) = -\mu_{11.2} - (-\mu_{11.2} - \delta) = \delta \\
 \text{Var}(\varphi_1) &= \mathbf{B}_{22.1}^{-1} \\
 \text{Var}(\varphi_2) &= \text{Var}\left(\varphi_1 + \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\sqrt{n}\hat{\beta}_2^{\text{RFM}}\right) \\
 &= \text{Var}(\varphi_1) + \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\mathbf{B}_{22.1}^{-1}\mathbf{B}_{21}\mathbf{B}_{11}^{-1} \\
 &\quad + 2\text{Cov}\left[\sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1), \sqrt{n}(\hat{\beta}_2^{\text{RFM}} - \beta_2)\right] (\mathbf{B}_{11}^{-1}\mathbf{B}_{12})^T \\
 &= \mathbf{B}_{22.1}^{-1} - \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\mathbf{B}_{22.1}^{-1}\mathbf{B}_{21}\mathbf{B}_{11}^{-1} = \mathbf{B}_{11}^{-1} \\
 \text{Var}(\varphi_3) &= \text{Var}\left(\sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}})\right) \\
 &= \text{Var}\left(\sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RFM}} - \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\hat{\beta}_2^{\text{RFM}})\right) \\
 &= \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\text{Var}\left[\sqrt{n}\hat{\beta}_2^{\text{RFM}}\right] (\mathbf{B}_{11}^{-1}\mathbf{B}_{12})^T \\
 &= \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\mathbf{B}_{22.1}^{-1}\mathbf{B}_{21}\mathbf{B}_{11}^{-1} = \Phi \\
 \text{Cov}(\varphi_1, \varphi_3) &= \text{Cov}\left[\sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1), \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}})\right] \\
 &= \text{Var}\left(\sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1)\right) - \text{Cov}\left[\sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1), \sqrt{n}(\hat{\beta}_1^{\text{RSM}} - \beta_1)\right] \\
 &= \text{Var}(\varphi_1) - \text{Cov}\left[\sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1), \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1) + \sqrt{n}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\hat{\beta}_2^{\text{RFM}}\right] \\
 &= \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\mathbf{B}_{22.1}^{-1}\mathbf{B}_{21}\mathbf{B}_{11}^{-1} = \Phi \\
 \\
 \text{Cov}(\varphi_2, \varphi_3) &= \text{Cov}\left[\sqrt{n}(\hat{\beta}_1^{\text{RSM}} - \beta_1), \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}})\right] \\
 &= \text{Cov}\left[\sqrt{n}(\hat{\beta}_1^{\text{RSM}} - \beta_1), \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1)\right] - \text{Var}\left(\sqrt{n}(\hat{\beta}_1^{\text{RSM}} - \beta_1)\right) \\
 &= \mathbf{B}_{11.2}^{-1} - \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\mathbf{B}_{22.1}^{-1}\mathbf{B}_{21}\mathbf{B}_{11}^{-1} - \mathbf{B}_{11}^{-1} \\
 &= \mathbf{B}_{11.2}^{-1} - (\mathbf{B}_{11.2}^{-1} - \mathbf{B}_{11}^{-1}) - \mathbf{B}_{11}^{-1} = \mathbf{0}
 \end{aligned}$$

Therefore, the asymptotic distributions of the vectors φ_2 and φ_3 are obtained as follows:

$$\begin{aligned}
 \varphi_2 &= \sqrt{n}(\hat{\beta}_1^{\text{RSM}} - \beta_1) \xrightarrow{D} \mathcal{N}_{p_1}(-\gamma, \mathbf{B}_{11}^{-1}) \\
 \varphi_3 &= \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}}) \xrightarrow{D} \mathcal{N}_{p_1}(\delta, \Phi)
 \end{aligned}$$

□

Appendix B

We next introduce the lemmas given in [30] to aid with the proof of the bias and covariance of the estimators.

Lemma A1. Let $\mathbf{V} = (V_1, V_2, \dots, V_p)^T$ be a p -dimensional normal vector distributed as $\mathcal{N}_p(\boldsymbol{\mu}_v, \boldsymbol{\Sigma}_p)$, then for a measurable function Ψ , we have

$$\begin{aligned} E[\mathbf{V}\Psi(\mathbf{V}^T\mathbf{V})] &= \boldsymbol{\mu}_v E[\Psi\chi_{p+2}^2(\Delta)] \\ E[\mathbf{V}\mathbf{V}^T\Psi(\mathbf{V}^T\mathbf{V})] &= \boldsymbol{\Sigma}_p E[\Psi\chi_{p+2}^2(\Delta)] + \boldsymbol{\mu}_v\boldsymbol{\mu}_v^T E[\Psi\chi_{p+4}^2(\Delta)] \end{aligned}$$

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

Appendix B.1

Proof of Theorem 2.

$$\begin{aligned} \text{ADB}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}}) &= E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \boldsymbol{\beta}_1)\right\} \\ &= -\boldsymbol{\mu}_{11.2}. \end{aligned}$$

$$\begin{aligned} \text{ADB}(\hat{\boldsymbol{\beta}}_1^{\text{RSM}}) &= E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RSM}} - \boldsymbol{\beta}_1)\right\} \\ &= E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\hat{\boldsymbol{\beta}}_2^{\text{RFM}} - \boldsymbol{\beta}_1)\right\} \\ &= E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \boldsymbol{\beta}_1)\right\} - E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\hat{\boldsymbol{\beta}}_2^{\text{RFM}})\right\} \\ &= -\boldsymbol{\mu}_{11.2} - E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\hat{\boldsymbol{\beta}}_2^{\text{RFM}})\right\} \\ &= -\boldsymbol{\mu}_{11.2} - \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\boldsymbol{\kappa} = -(\boldsymbol{\mu}_{11.2} + \boldsymbol{\delta}) = -\boldsymbol{\gamma}. \end{aligned}$$

Using Lemma 1,

$$\begin{aligned} \text{ADB}(\hat{\boldsymbol{\beta}}_1^{\text{RPT}}) &= E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RPT}} - \boldsymbol{\beta}_1)\right\} \\ &= E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - (\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \hat{\boldsymbol{\beta}}_1^{\text{RSM}})I(L_n \leq d_{n,\alpha}) - \boldsymbol{\beta}_1)\right\} \\ &= E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \boldsymbol{\beta}_1)\right\} - E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \hat{\boldsymbol{\beta}}_1^{\text{RSM}})I(L_n \leq d_{n,\alpha})\right\} \\ &= -\boldsymbol{\mu}_{11.2} - E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \hat{\boldsymbol{\beta}}_1^{\text{RSM}})I(L_n \leq d_{n,\alpha})\right\} \\ &= -\boldsymbol{\mu}_{11.2} - \boldsymbol{\delta}\mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta). \end{aligned}$$

$$\begin{aligned} \text{ADB}(\hat{\boldsymbol{\beta}}_1^{\text{RSE}}) &= E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RSE}} - \boldsymbol{\beta}_1)\right\} \\ &= E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - (\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \hat{\boldsymbol{\beta}}_1^{\text{RSM}})(p_2 - 2)L_n^{-1} - \boldsymbol{\beta}_1)\right\} \\ &= E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \boldsymbol{\beta}_1)\right\} - E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \hat{\boldsymbol{\beta}}_1^{\text{RSM}})(p_2 - 2)L_n^{-1}\right\} \\ &= -\boldsymbol{\mu}_{11.2} - E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}_1^{\text{RFM}} - \hat{\boldsymbol{\beta}}_1^{\text{RSM}})(p_2 - 2)L_n^{-1}\right\} \\ &= -\boldsymbol{\mu}_{11.2} - (p_2 - 2)\boldsymbol{\delta}E(\chi_{p_2+2}^{-2}(\Delta)). \end{aligned}$$

$$\begin{aligned}
 \text{ADB}(\hat{\beta}_1^{\text{RPS}}) &= E\left\{ \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\beta}_1^{\text{RPS}} - \beta_1) \right\} \\
 &= E\left\{ \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\beta}_1^{\text{RSM}} + (\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}})(1 - (p_2 - 2)L_n^{-1})I(L_n > p_2 - 2) - \beta_1) \right\} \\
 &= E\left\{ \sqrt{n}[\hat{\beta}_1^{\text{RSM}} + (\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}})(1 - I(L_n \leq p_2 - 2)) \right. \\
 &\quad \left. - (\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}})(p_2 - 2)L_n^{-1}I(L_n > p_2 - 2) - \beta_1] \right\} \\
 &= E\left\{ \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1) \right\} - E\left\{ \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}})(p_2 - 2)I(L_n \leq p_2 - 2) \right\} \\
 &\quad - E\left\{ \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}})(p_2 - 2)L_n^{-1}I(L_n > p_2 - 2) \right\} \\
 &= -\mu_{11.2} - \delta \mathbf{H}_{p_2+2}(\chi_{p_2-2}^2; \Delta) - (p_2 - 2)\delta E\{\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{-2} > p_2 - 2)\}. \\
 &\quad \square
 \end{aligned}$$

Appendix B.2

In order to compute the risk functions, we first compute the asymptotic covariance of the estimators. The asymptotic covariance of an estimator $\hat{\beta}_1^*$ is expressed as

$$\text{Cov}(\hat{\beta}_1^*) = \lim_{n \rightarrow \infty} E\{n(\hat{\beta}_1^* - \beta_1)(\hat{\beta}_1^* - \beta_1)^T\}.$$

Proof of Theorem 3. We first start by computing the asymptotic covariance of the estimator $\hat{\beta}_1^{\text{RFM}}$ as:

$$\begin{aligned}
 \text{Cov}(\hat{\beta}_1^{\text{RFM}}) &= E\left\{ \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1)\sqrt{n}(\hat{\beta}_1^{\text{RFM}} - \beta_1)^T \right\} \\
 &= E(\varphi_1\varphi_1^T) = \text{Cov}(\varphi_1\varphi_1^T) + E(\varphi_1)E(\varphi_1^T) \\
 &= \mathbf{B}_{11.2}^{-1} + \mu_{11.2}\mu_{11.2}^T.
 \end{aligned}$$

Furthermore, similarly, the asymptotic covariance of the estimator $\hat{\beta}_1^{\text{RSM}}$ is obtained as:

$$\begin{aligned}
 \text{Cov}(\hat{\beta}_1^{\text{RSM}}) &= E\left\{ \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\beta}_1^{\text{RSM}} - \beta_1)\sqrt{n}(\hat{\beta}_1^{\text{RSM}} - \beta_1)^T \right\} \\
 &= E(\varphi_2\varphi_2^T) = \text{Cov}(\varphi_2\varphi_2^T) + E(\varphi_2)E(\varphi_2^T) \\
 &= \mathbf{B}_{11}^{-1} + \gamma\gamma^T.
 \end{aligned}$$

The asymptotic covariance of the estimator $\hat{\beta}_1^{\text{RPT}}$ is obtained as:

$$\begin{aligned}
 \text{Cov}(\hat{\beta}_1^{\text{RPT}}) &= E\left\{ \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\beta}_1^{\text{RPT}} - \beta_1)\sqrt{n}(\hat{\beta}_1^{\text{RPT}} - \beta_1)^T \right\} \\
 &= E\left\{ \lim_{n \rightarrow \infty} n[(\hat{\beta}_1^{\text{RFM}} - \beta_1) - (\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}})I(L_n \leq d_{n,\alpha})] \right. \\
 &\quad \left. [(\hat{\beta}_1^{\text{RFM}} - \beta_1) - (\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}})I(L_n \leq d_{n,\alpha})]^T \right\} \\
 &= E\left\{ [\varphi_1 - \varphi_3I(L_n \leq d_{n,\alpha})][\varphi_1 - \varphi_3I(L_n \leq d_{n,\alpha})]^T \right\} \\
 &= E\left\{ \varphi_1\varphi_1^T - 2\varphi_3\varphi_1^T I(L_n \leq d_{n,\alpha}) + \varphi_3\varphi_3^T I(L_n \leq d_{n,\alpha}) \right\}
 \end{aligned}$$

Thus, we need to find $E\{\varphi_1\varphi_1^T\}$, $E\{\varphi_3\varphi_1^T I(L_n \leq d_{n,\alpha})\}$ and $E\{\varphi_3\varphi_3^T I(L_n \leq d_{n,\alpha})\}$. The first term is $E\{\varphi_1\varphi_1^T\} = \mathbf{B}_{11.2}^{-1} + \mu_{11.2}\mu_{11.2}^T$. Using Lemma 1, the third term is computed as:

$$E\{\varphi_3\varphi_3^T I(L_n \leq d_{n,\alpha})\} = \Phi \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) + \delta \delta^T \mathbf{H}_{p_2+4}(\chi_{p_2}^2; \Delta).$$

The second term $E\{\varphi_3\varphi_1^T I(L_n \leq d_{n,\alpha})\}$ can be computed from normal theory as

$$\begin{aligned}
 E\left\{\varphi_3\varphi_1^T I(L_n \leq d_{n,\alpha})\right\} &= E\left\{E(\varphi_3\varphi_1^T I(L_n \leq d_{n,\alpha})|\varphi_3)\right\} = E\left\{\varphi_3 E(\varphi_1^T I(L_n \leq d_{n,\alpha})|\varphi_3)\right\} \\
 &= E\left\{\varphi_3[-\mu_{11.2} + (\varphi_3 - \delta)]^T I(L_n \leq d_{n,\alpha})\right\} \\
 &= -E\left\{\varphi_3\mu_{11.2} I(L_n \leq d_{n,\alpha})\right\} + E\left\{\varphi_3(\varphi_3 - \delta)^T I(L_n \leq d_{n,\alpha})\right\} \\
 &= -\mu_{11.2}^T E\left\{\varphi_3 I(L_n \leq d_{n,\alpha})\right\} + E\left\{\varphi_3\varphi_3^T I(L_n \leq d_{n,\alpha})\right\} \\
 &\quad - E\left\{\varphi_3\delta^T I(L_n \leq d_{n,\alpha})\right\} \\
 &= -\mu_{11.2}^T \delta \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) + \left\{Cov(\varphi_3\varphi_3^T) \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta)\right. \\
 &\quad \left.+ E(\varphi_3)E(\varphi_3^T) \mathbf{H}_{p_2+4}(\chi_{p_2}^2; \Delta) - \delta\delta^T \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta)\right\} \\
 &= -\mu_{11.2}^T \delta \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) + \Phi \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) + \delta\delta^T \mathbf{H}_{p_2+4}(\chi_{p_2}^2; \Delta) \\
 &\quad - \delta\delta^T \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta)
 \end{aligned}$$

Putting all the terms together and simplifying, we obtain

$$\begin{aligned}
 Cov(\hat{\beta}_1^{RPT}) &= \mu_{11.2}\mu_{11.2}^T + 2\mu_{11.2}^T \delta \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) + \mathbf{B}_{11.2}^{-1} - \Phi \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) - \delta\delta^T \mathbf{H}_{p_2+4}(\chi_{p_2}^2; \Delta) \\
 &\quad + 2\delta\delta^T \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) \\
 &= \mathbf{B}_{11.2}^{-1} + \mu_{11.2}\mu_{11.2}^T + 2\mu_{11.2}^T \delta \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) - \Phi \mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) \\
 &\quad + \delta\delta^T [2\mathbf{H}_{p_2+2}(\chi_{p_2}^2; \Delta) - \mathbf{H}_{p_2+4}(\chi_{p_2}^2; \Delta)].
 \end{aligned}$$

The asymptotic covariance of the estimator $\hat{\beta}_1^{RSE}$ can be obtained as

$$\begin{aligned}
 Cov(\hat{\beta}_1^{RSE}) &= E\left\{\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\beta}_1^{RSE} - \beta_1)\sqrt{n}(\hat{\beta}_1^{RSE} - \beta_1)^T\right\} \\
 &= E\left\{\lim_{n \rightarrow \infty} n[(\hat{\beta}_1^{RFM} - \beta_1) - (\hat{\beta}_1^{RFM} - \hat{\beta}_1^{RSM})(p_2 - 2)L_n^{-1}] \right. \\
 &\quad \left. [(\hat{\beta}_1^{RFM} - \beta_1) - (\hat{\beta}_1^{RFM} - \hat{\beta}_1^{RSM})(p_2 - 2)L_n^{-1}]^T\right\} \\
 &= E\left\{[\varphi_1 - \varphi_3(p_2 - 2)L_n^{-1}][\varphi_1 - \varphi_3(p_2 - 2)L_n^{-1}]^T\right\} \\
 &= E\left\{\varphi_1\varphi_1^T - 2(p_2 - 2)\varphi_3\varphi_1^T L_n^{-1} + (p_2 - 2)^2\varphi_3\varphi_3^T L_n^{-2}\right\}
 \end{aligned}$$

We need to compute $E\{\varphi_3\varphi_3^T L_n^{-2}\}$ and $E\{\varphi_3\varphi_1^T L_n^{-1}\}$. By using Lemma 1, the first term is obtained as follows:

$$E\{\varphi_3\varphi_3^T L_n^{-2}\} = \Phi E(\chi_{p_2+2}^{-4}(\Delta)) + \delta\delta^T E(\chi_{p_2+4}^{-4}(\Delta)).$$

The second term is computed from normal theory

$$\begin{aligned}
 E\left\{\varphi_3\varphi_1^T L_n^{-1}\right\} &= E\left\{E(\varphi_3\varphi_1^T L_n^{-1}|\varphi_3)\right\} = E\left\{\varphi_3 E(\varphi_1^T L_n^{-1}|\varphi_3)\right\} \\
 &= E\left\{\varphi_3[-\mu_{11.2} + (\varphi_3 - \delta)]^T L_n^{-1}\right\} \\
 &= -E\left\{\varphi_3\mu_{11.2} L_n^{-1}\right\} + E\left\{\varphi_3(\varphi_3 - \delta)^T L_n^{-1}\right\} \\
 &= -\mu_{11.2}^T E\left\{\varphi_3 L_n^{-1}\right\} + E\left\{\varphi_3\varphi_3^T L_n^{-1}\right\} - E\left\{\varphi_3\delta^T L_n^{-1}\right\}
 \end{aligned}$$

From above, we can find $E\{\varphi_3 \delta^T L_n^{-1}\} = \delta \delta^T E(\chi_{p_2+2}^{-2}(\Delta))$ and $E\{\varphi_3 L_n^{-1}\} = \delta E(\chi_{p_2+2}^{-2}(\Delta))$. Putting these terms together and simplifying, we obtain

$$\begin{aligned} \text{Cov}(\hat{\beta}_1^{\text{RSE}}) &= \mathbf{B}_{11.2}^{-1} + \boldsymbol{\mu}_{11.2} \boldsymbol{\mu}_{11.2}^T + 2(p_2 - 2) \boldsymbol{\mu}_{11.2}^T \delta E(\chi_{p_2+2}^{-2}(\Delta)) \\ &\quad - (p_2 - 2) \boldsymbol{\Phi} \left\{ 2E(\chi_{p_2+2}^{-2}(\Delta)) - (p_2 - 2)E(\chi_{p_2+2}^{-4}(\Delta)) \right\} \\ &\quad + (p_2 - 2) \delta \delta^T \left\{ -2E(\chi_{p_2+4}^{-2}(\Delta)) + 2E(\chi_{p_2+2}^{-2}(\Delta)) + (p_2 - 2)E(\chi_{p_2+4}^{-4}(\Delta)) \right\}. \end{aligned}$$

Since $\hat{\beta}_1^{\text{RPS}} = \hat{\beta}_1^{\text{RSE}} - (\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}}) \{1 - (p_2 - 2)L_n^{-1}\} \mathbf{I}(L_n \leq p_2 - 2)$.

We derive the covariance of the estimator $\hat{\beta}_1^{\text{RPS}}$ as follows.

$$\begin{aligned} \text{Cov}(\hat{\beta}_1^{\text{RPS}}) &= E \left\{ \lim_{n \rightarrow \infty} \sqrt{n} (\hat{\beta}_1^{\text{RPS}} - \beta_1) \sqrt{n} (\hat{\beta}_1^{\text{RPS}} - \beta_1)^T \right\} \\ &= E \left\{ \lim_{n \rightarrow \infty} \sqrt{n} (\hat{\beta}_1^{\text{RSE}} - \beta_1) - \sqrt{n} (\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}}) \{1 - (p_2 - 2)L_n^{-1}\} \mathbf{I}(L_n \leq p_2 - 2) \right. \\ &\quad \times \left. \left[\sqrt{n} (\hat{\beta}_1^{\text{RSE}} - \beta_1) - \sqrt{n} (\hat{\beta}_1^{\text{RFM}} - \hat{\beta}_1^{\text{RSM}}) \{1 - (p_2 - 2)L_n^{-1}\} \mathbf{I}(L_n \leq p_2 - 2) \right]^T \right\} \\ &= E \left\{ \lim_{n \rightarrow \infty} \sqrt{n} (\hat{\beta}_1^{\text{RSE}} - \beta_1) \sqrt{n} (\hat{\beta}_1^{\text{RSE}} - \beta_1)^T - 2\boldsymbol{\varphi}_3 \sqrt{n} (\hat{\beta}_1^{\text{RSE}} - \beta_1)^T \{1 - (p_2 - 2)L_n^{-1}\} \mathbf{I}(L_n \leq p_2 - 2) \right. \\ &\quad \left. + \boldsymbol{\varphi}_3 \boldsymbol{\varphi}_3^T \{1 - (p_2 - 2)L_n^{-1}\}^2 \mathbf{I}(L_n \leq p_2 - 2) \right\} \\ &= \text{Cov}(\hat{\beta}_1^{\text{RSE}}) - 2E \left\{ \lim_{n \rightarrow \infty} \boldsymbol{\varphi}_3 \sqrt{n} (\hat{\beta}_1^{\text{RSE}} - \beta_1)^T \{1 - (p_2 - 2)L_n^{-1}\}^2 \mathbf{I}(L_n \leq p_2 - 2) \right\} \\ &\quad + E \left\{ \lim_{n \rightarrow \infty} \boldsymbol{\varphi}_3 \boldsymbol{\varphi}_3^T \{1 - (p_2 - 2)L_n^{-1}\}^2 \mathbf{I}(L_n \leq p_2 - 2) \right\} \\ &= \text{Cov}(\hat{\beta}_1^{\text{RSE}}) - 2E \left\{ \lim_{n \rightarrow \infty} \boldsymbol{\varphi}_3 \boldsymbol{\varphi}_1^T \{1 - (p_2 - 2)L_n^{-1}\} \mathbf{I}(L_n \leq p_2 - 2) \right\} \\ &\quad + 2E \left\{ \lim_{n \rightarrow \infty} \boldsymbol{\varphi}_3 \boldsymbol{\varphi}_3^T (p_2 - 2)L_n^{-1} \{1 - (p_2 - 2)L_n^{-1}\} \mathbf{I}(L_n \leq p_2 - 2) \right\} \\ &\quad + E \left\{ \lim_{n \rightarrow \infty} \boldsymbol{\varphi}_3 \boldsymbol{\varphi}_3^T \{1 - (p_2 - 2)L_n^{-1}\}^2 \mathbf{I}(L_n \leq p_2 - 2) \right\} \\ &= \text{Cov}(\hat{\beta}_1^{\text{RSE}}) - 2E \left\{ \lim_{n \rightarrow \infty} \boldsymbol{\varphi}_3 \boldsymbol{\varphi}_1^T \{1 - (p_2 - 2)L_n^{-1}\} \mathbf{I}(L_n \leq p_2 - 2) \right\} \\ &\quad - E \left\{ \lim_{n \rightarrow \infty} \boldsymbol{\varphi}_3 \boldsymbol{\varphi}_3^T (p_2 - 2)^2 L_n^{-2} \mathbf{I}(L_n \leq p_2 - 2) \right\} + E \left\{ \lim_{n \rightarrow \infty} \boldsymbol{\varphi}_3 \boldsymbol{\varphi}_3^T \mathbf{I}(L_n \leq p_2 - 2) \right\} \end{aligned}$$

We first compute the last term in the equation above $E\{\boldsymbol{\varphi}_3 \boldsymbol{\varphi}_3^T \mathbf{I}(L_n \leq p_2 - 2)\}$ as $E\{\boldsymbol{\varphi}_3 \boldsymbol{\varphi}_3^T \mathbf{I}(L_n \leq p_2 - 2)\} = \boldsymbol{\Phi} \mathbf{H}_{p_2+2}(p_2 - 2; \Delta) + \delta \delta^T \mathbf{H}_{p_2+4}(p_2 - 2; \Delta)$. Using Lemma 1 and from the normal theory, we find,

$$\begin{aligned}
& E\left\{\boldsymbol{\varphi}_3\boldsymbol{\varphi}_1^T\{1-(p_2-2)L_n^{-1}\}I(L_n\leq p_2-2)\right\} \\
&= E\left\{E(\boldsymbol{\varphi}_3\boldsymbol{\varphi}_1^T\{1-(p_2-2)L_n^{-1}\}I(L_n\leq p_2-2)|\boldsymbol{\varphi}_3)\right\} \\
&= E\left\{\boldsymbol{\varphi}_3E(\boldsymbol{\varphi}_1^T\{1-(p_2-2)L_n^{-1}\}I(L_n\leq p_2-2)|\boldsymbol{\varphi}_3)\right\} \\
&= E\left\{\boldsymbol{\varphi}_3[\boldsymbol{\mu}_{11.2}+(\boldsymbol{\varphi}_3-\boldsymbol{\delta})]^T\{1-(p_2-2)L_n^{-1}\}I(L_n\leq p_2-2)\right\} \\
&= -\boldsymbol{\mu}_{11.2}E\left(\boldsymbol{\varphi}_3\{1-(p_2-2)L_n^{-1}\}I(L_n\leq p_2-2)\right) \\
&+ E\left(\boldsymbol{\varphi}_3\boldsymbol{\varphi}_3^T\{1-(p_2-2)L_n^{-1}\}I(L_n\leq p_2-2)\right) \\
&- E\left(\boldsymbol{\varphi}_3\boldsymbol{\delta}^T\{1-(p_2-2)L_n^{-1}\}I(L_n\leq p_2-2)\right) \\
&= -\boldsymbol{\delta}\boldsymbol{\mu}_{11.2}^TE\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\}I\left(\chi_{p_2+2}^{-2}(\Delta)\leq p_2-2\right)\right) \\
&+ \boldsymbol{\Phi}E\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\}I\left(\chi_{p_2+2}^{-2}(\Delta)\leq p_2-2\right)\right) \\
&+ \boldsymbol{\delta}\boldsymbol{\delta}^TE\left(\left\{1-(p_2-2)\chi_{p_2+4}^{-2}(\Delta)\right\}I\left(\chi_{p_2+4}^{-2}(\Delta)\leq p_2-2\right)\right) \\
&- \boldsymbol{\delta}\boldsymbol{\delta}^TE\left(\left\{1-(p_2-2)\chi_{p_2+4}^{-2}(\Delta)\right\}I\left(\chi_{p_2+4}^{-2}(\Delta)\leq p_2-2\right)\right).
\end{aligned}$$

$$\begin{aligned}
E\left\{\boldsymbol{\varphi}_3\boldsymbol{\varphi}_3^T(p_2-2)^2L_n^{-2}I(L_n\leq p_2-2)\right\} &= (p_2-2)^2\boldsymbol{\Phi}E\left(\chi_{p_2+2}^{-4}(\Delta)I\left(\chi_{p_2+2}^2(\Delta)\leq p_2-2\right)\right) \\
&+ (p_2-2)^2\boldsymbol{\delta}\boldsymbol{\delta}^TE\left(\chi_{p_2+2}^{-4}(\Delta)I\left(\chi_{p_2+2}^2(\Delta)\leq p_2-2\right)\right)
\end{aligned}$$

Putting all the terms together, we obtain

$$\begin{aligned}
\text{Cov}(\hat{\boldsymbol{\beta}}_1^{\text{RPS}}) &= \text{Cov}(\hat{\boldsymbol{\beta}}_1^{\text{RSE}}) + 2\boldsymbol{\delta}\boldsymbol{\mu}_{11.2}^TE\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\}I\left(\chi_{p_2+2}^2(\Delta)\leq p_2-2\right)\right) \\
&- 2\boldsymbol{\Phi}E\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\}I\left(\chi_{p_2+2}^2(\Delta)\leq p_2-2\right)\right) \\
&- 2\boldsymbol{\delta}\boldsymbol{\delta}^TE\left(\left\{1-(p_2-2)\chi_{p_2+4}^{-2}(\Delta)\right\}I\left(\chi_{p_2+4}^2(\Delta)\leq p_2-2\right)\right) \\
&+ 2\boldsymbol{\delta}\boldsymbol{\delta}^TE\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\}I\left(\chi_{p_2+2}^2(\Delta)\leq p_2-2\right)\right) \\
&- (p_2-2)^2\boldsymbol{\Phi}E\left(\chi_{p_2+2}^{-4}(\Delta)I\left(\chi_{p_2+2}^2(\Delta)\leq p_2-2\right)\right) \\
&- (p_2-2)^2\boldsymbol{\delta}\boldsymbol{\delta}^TE\left(\chi_{p_2+2}^{-4}(\Delta)I\left(\chi_{p_2+2}^2(\Delta)\leq p_2-2\right)\right) \\
&+ \boldsymbol{\Phi}\mathbf{H}_{p_2+2}(p_2-2;\Delta) + \boldsymbol{\delta}\boldsymbol{\delta}^T\mathbf{H}_{p_2+4}(p_2-2;\Delta).
\end{aligned}$$

□

References

1. Laird, N.M.; Ware, J.H. Random-effects models for longitudinal data. *Biometrics* **1982**, *38*, 963–974. [[CrossRef](#)]
2. Longford, N. Regression analysis of multilevel data with measurement error. *Br. J. Math. Stat. Psychol.* **1993**, *46*, 301–311. [[CrossRef](#)]
3. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [[CrossRef](#)]

4. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [[CrossRef](#)]
5. Tran, M.N. The loss rank criterion for variable selection in linear regression analysis. *Scand. J. Stat.* **2011**, *38*, 466–479. [[CrossRef](#)]
6. Huang, J.; Ma, S.; Zhang, C.H. Adaptive Lasso for sparse high-dimensional regression models. *Stat. Sin.* **2008**, *18*, 1603–1618.
7. Kim, Y.; Choi, H.; Oh, H.S. Smoothly clipped absolute deviation on high dimensions. *J. Am. Stat. Assoc.* **2008**, *103*, 1665–1673. [[CrossRef](#)]
8. Wang, H.; Leng, C. Unified LASSO estimation by least squares approximation. *J. Am. Stat. Assoc.* **2007**, *102*, 1039–1048. [[CrossRef](#)]
9. Yuan, M.; Lin, Y. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2006**, *68*, 49–67. [[CrossRef](#)]
10. Leng, C.; Lin, Y.; Wahba, G. A note on the lasso and related procedures in model selection. *Stat. Sin.* **2006**, *16*, 1273–1284.
11. Park, T.; Casella, G. The bayesian lasso. *J. Am. Stat. Assoc.* **2008**, *103*, 681–686. [[CrossRef](#)]
12. Greenlaw, K.; Szefer, E.; Graham, J.; Lesperance, M.; Nathoo, F.S.; Initiative, A.D.N. A Bayesian group sparse multi-task regression model for imaging genetics. *Bioinformatics* **2017**, *33*, 2513–2522. [[CrossRef](#)] [[PubMed](#)]
13. Ahmed, S.E.; Nicol, C.J. An application of shrinkage estimation to the nonlinear regression model. *Comput. Stat. Data Anal.* **2012**, *56*, 3309–3321. [[CrossRef](#)]
14. Ahmed, S.E.; Raheem, S.E. Shrinkage and absolute penalty estimation in linear regression models. *Wiley Interdiscip. Rev. Comput. Stat.* **2012**, *4*, 541–553. [[CrossRef](#)]
15. Lisawadi, S.; Kashif Ali Shah, M.; Ejaz Ahmed, S. Model selection and post estimation based on a pretest for logistic regression models. *J. Stat. Comput. Simul.* **2016**, *86*, 3495–3511. [[CrossRef](#)]
16. Ahmed, S.E.; Opoku, E.A. Submodel selection and post-estimation of the linear mixed models. In *Proceedings of the Tenth International Conference on Management Science and Engineering Management*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 633–646.
17. Raheem, S.E.; Ahmed, S.E.; Doksum, K.A. Absolute penalty and shrinkage estimation in partially linear models. *Comput. Stat. Data Anal.* **2012**, *56*, 874–891. [[CrossRef](#)]
18. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17. [[CrossRef](#)]
19. Liu, K. Using Liu-type estimator to combat collinearity. *Commun. Stat.-Theory Methods* **2003**, *32*, 1009–1020. [[CrossRef](#)]
20. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [[CrossRef](#)]
21. Yüzbaşı, B.; Ejaz Ahmed, S. Shrinkage and penalized estimation in semi-parametric models with multicollinear data. *J. Stat. Comput. Simul.* **2016**, *86*, 3543–3561. [[CrossRef](#)]
22. Yüzbaşı, B.; Ahmed, S.E.; Güngör, M. Improved penalty strategies in linear regression models. *REVSTAT J.* **2017**, *15*, 251–276.
23. Knight, K.; Fu, W. Asymptotics for lasso-type estimators. *Ann. Stat.* **2000**, *28*, 1356–1378.
24. Belsley, D.A. *Conditioning Diagnostics: Collinearity and Weak Data in Regression*; Number 519.536 B452; Wiley: Hoboken, NJ, USA, 1991.
25. Twisk, J.; Kemper, H.; Mellenbergh, G. Longitudinal development of lipoprotein levels in males and females aged 12–28 years: The Amsterdam Growth and Health Study. *Int. J. Epidemiol.* **1995**, *24*, 69–77. [[CrossRef](#)] [[PubMed](#)]
26. Nie, Y.; Opoku, E.; Yasmin, L.; Song, Y.; Wang, J.; Wu, S.; Scarapicchia, V.; Gawryluk, J.; Wang, L.; Cao, J.; et al. Spectral dynamic causal modelling of resting-state fMRI: An exploratory study relating effective brain connectivity in the default mode network to genetics. *Stat. Appl. Genet. Mol. Biol.* **2020**, *19*. [[CrossRef](#)]
27. Ahmed, S.E.; Kim, H.; Yıldırım, G.; Yüzbaşı, B. High-Dimensional Regression Under Correlated Design: An Extensive Simulation Study. In *International Workshop on Matrices and Statistics*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 145–175.
28. Ejaz Ahmed, S.; Yüzbaşı, B. Big data analytics: Integrating penalty strategies. *Int. J. Manag. Sci. Eng. Manag.* **2016**, *11*, 105–115. [[CrossRef](#)]
29. Ahmed, S.E.; Yüzbaşı, B. High dimensional data analysis: Integrating submodels. In *Big and Complex Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 285–304.
30. Judge, G.G.; Bock, M.E. *The Statistical Implication of Pre-Test and Steinrule Estimators in Econometrics*; Elsevier: Amsterdam, The Netherlands, 1978.