İNÖNÜ UNIVERSITY
GRADUATE SCHOOL of NATURAL and APPLIED SCIENCES

PENALTY and NON-PENALTY ESTIMATION STRATEGIES for LINEAR and PARTIALLY LINEAR MODELS

BAHADIR YÜZBAŞI

DOCTOR of PHILOSOPHY DISSERTATION
DEPARTMENT of MATHEMATICS

DECEMBER 2014

Name of Thesis	:	Penalty and Non-Penalty Estimation Strategies for Linear and Partially Linear Models
Submitted by	:	Bahadır Yüzbaşı
Date of Defence	:	11.12.2014

Evaluated by the jury the above mentioned thesis has been accepted as a Doctoral Thesis at the Department of Mathematics

Members of Defence Jury

Supervisor	:	**Prof. Dr. Mehmet GÜNGÖR** İnönü University	.....
Co – Supervisor	:	**Prof. Dr. Syed Ejaz AHMED** Brock University	.....
		Prof. Dr. Bayram ŞAHİN İnönü University	.....
		Assoc. Prof. Dr. Mahmut IŞIK Fırat University	.....
		Assist. Prof. Dr. Fatma ZEREN İnönü University	.....
		Assist. Prof. Dr. Hasan SÖYLER İnönü University	.....

Prof. Dr. Alaattin ESEN
Director of Institute

The Word of Honour

I submitted the Doctoral Dissertation ”Penalty and Non-Penalty Estimation Strategies for Linear and Partially Linear Models” titled this work ethic and traditions contrary to a help recourse by me are written and taken advantage of all the resources, both in the text and the reference method in accordance with those shown refers to the formation, so with honour verify.

Bahadır Yüzbaşı

ABSTRACT

PhD Thesis

PENALTY and NON-PENALTY ESTIMATION STRATEGIES
for LINEAR and PARTIALLY LINEAR MODELS

Bahadır Yüzbaşı

İnönü University
Graduate School of Natural and Applied Sciences
Department of Mathematics

123+xi pages

2014

Supervisors : Prof. Dr. Mehmet GÜNGÖR and Prof. Dr. Syed Ejaz AHMED

In this dissertation we obtained pretest ridge regression, shrinkage ridge regression estimators, and compared their performance with penalty estimators in linear and partially linear models. We also investigated asymptotic properties of proposed estimators both analytically and thorough simulation studies.

In Chapter 1, we presented preliminary definitions and theorems which are used at the next two chapters.

In Chapter 2, we defined pretest ridge regression, shrinkage ridge regression and positive shrinkage ridge regression estimators for a multiple linear regression model, and compared their performance with some penalty estimators which are lasso, adaptive lasso and SCAD. Monte Carlo studies were conducted to compare the estimators in two situations: when $p < n$ and when $p > n$. Three real data examples for low-dimensional scenario and two real data examples for high-dimensional scenario are presented to illustrate the usefulness of the suggested methods. Finally, we investigated the asymptotic properties of these estimators analytically.

In Chapter 3, we defined pretest ridge regression, shrinkage ridge regression and positive shrinkage ridge regression estimators for a partially linear regression model. In this model, the nonparametric function is estimated using the smoothing spline method. We also compared the performance of suggested estimators with some penalty estimators which are lasso, adaptive lasso and SCAD. Monte Carlo studies were conducted to compare the estimators in two situations: when $p < n$ and when $p > n$. Finally, we investigated the asymptotic properties of these estimators analytically.

In Chapter 4, it is given conclusions and future work.

KEYWORDS : Ridge Regression, Pretest Estimation, Shrinkage Estimators, Penalty Estimators, Smoothing Spline, High-Dimensional Data.

ÖZET

Doktora Tezi

LİNEER ve KISMİ LİNEER MODELLER İÇİN
CEZALI ve CEZASIZ TAHMİN STRATEJİLERİ

Bahadır Yüzbaşı

İnönü Üniversitesi
Fen Bilimleri Enstitüsü
Matematik Anabilim Dalı

123+xi sayfa

2014

Danışmanlar : Prof. Dr. Mehmet GÜNGÖR ve Prof. Dr. Syed Ejaz AHMED

Bu tezde, lineer ve kısmi lineer modeller için, öntest ridge regresyon ve shrinkage ridge regresyon tahmincilerini elde ettik, ve bunların performanslarını cezalı tahmincilerle karışılaştırdık. Ayrıca önerilen tahmincilerin özelliklerini hem analitik olarak hem de ayrıntılı simülasyon çalışmalarıyla araştırdık.

Bölüm 1 de, gelecek iki bölümde kullanılan temel tanım ve teoremleri verdik.

Bölüm 2 de, çoklu bir lineer regresyon modeli için öntest ridge regresyon, shrinkage ridge regresyon ve pozitif shrinkage ridge regresyon tahmincilerini tanımladık, ve bu tahmincilerin performanslarını bazı cezalı tahmincilerden olan lasso, uyarlamalı lasso ve SCAD tahmincileriyle karşılaştırdık. İki durum, yani $p < n$ ve $p > n$ hallerinde tahmincileri karşılaştırmak için Monte Carlo çalışmaları yapıldı. Önerilen metodların faydalılığını göstermek için, düşük-boyutlu senaryoda üç tane, yüksek-boyutlu senaryoda iki tane gerçek veri örnekleri yapıldı. Son olarak ise, bu tahmincilerin asimtotik özellikleri analitik olarak incelendi.

Bölüm 3 de, bir kısmi lineer regresyon modeli için öntest ridge regresyon, shrinkage ridge regresyon ve pozitif shrinkage ridge regresyon tahmincilerini tanımladık. Bu modelde, splayn düzeltme metodu kullanılarak parametrik olmayan fonksiyon tahmin edildi. Ayrıca, önerilen tahmincilerin performanslarını bazı cezalı tahmincilerden olan lasso, uyarlamalı lasso ve SCAD tahmincileriyle karşılaştırdık. İki durum, yani $p < n$ ve $p > n$ hallerinde tahmincileri karşılaştırmak için Monte Carlo çalışmaları yapıldı. Son olarak ise, bu tahmincilerin asimtotik özellikleri analitik olarak incelendi.

Bölüm 4 de, sonuçlar ve gelecekteki çalışmalar verildi.

ANAHTAR KELİMELER : Ridge Regresyon, Öntest Tahmin, Shrinkage Tahmincileri, Cezalı Tahminciler, Düzleştirme Splayn, Yüksek-Boyutlu Veri.

ACKNOWLEDGEMENTS

My sincere gratitude goes to my advisors Prof. Dr. Mehmet Güngör and Prof. Dr. Syed Ejaz Ahmed for their guidance which has lead to the completion of this dissertation. I am grateful to them for their support during my doctoral studies.

I would like to thank Assoc. Prof. Dr. Dursun Aydın and Assist. Prof. Dr. S.M. Enayetur Raheem, who were always willing to help and give their best suggestions.

Special thanks go to my advisory committee member Prof. Dr. Bayram Şahin, Assoc. Prof. Dr. Mahmut Işık, Assist. Prof. Dr. Fatma Zeren and Assist. Prof. Dr. Hasan Söyler for their endless, important reviewing the dissertation and providing with valuable suggestions.

I would never have been able to finish my dissertation without my wife's contributions during my doctoral thesis and life. I also thank my parents for all their patience, support and prayers.

Finally, I would like to thank The Scientific and Technological Research Council of Turkey (TUBİTAK), which supported my research in Canada.

Bahadır Yüzbaşı
December 11, 2014
Malatya, TURKEY

CONTENTS

LIST of FIGURE

LIST of TABLE

ABBREVIATIONS

ADB	asymptotic distributional bias
AE	auxiliary information
aLasso	adaptive lasso
AIC	akaike information criterion
AQDB	asymptotic quadratic distributional bias
AQDR	asymptotic quadratic distributional risk
BLUE	best linear unbiased estimator
BIC	bayesian information criterion
BSS	best subset selection
CNT	condition number test
FM	full model estimator
LAR	least angle regression
Lasso	least absolute shrinkage and selection operator
MCP	minimax concave penalty
MSE	mean squared error
MVUE	minimum-variance unbiased estimator
NSI	non-sample information
PCE	principal components estimation
PT	pretest estimator
PLM	partially linear model
PLS	penalized least squares
PLSE	partial least squares estimation
RFM	full model ridge regression estimator
RS	shrinkage ridge regression estimator
PSE	positive shrinkage estimator
RSM	submode ridge regression estimator
RMSE	relative mean squared error
SCAD	smoothly clipped absolute deviation
SE	shrinkage estimator
SM	sub-model estimator
SRFM	semiparametric full model ridge regression estimator
SRPS	semiparametric positive shrinkage ridge regression estimator
SRPT	semiparametric pretest ridge regression estimator
SRSM	semiparametric sub-model ridge regression estimator
SRM	semiparametric sub-model estimator
UMVUE	uniformly minimum variance unbiased estimator
UPI	uncertain prior information

LIST OF SYMBOLS

$\boldsymbol{\beta}$	regression parameter vector
p	the number of regression parameters
n	sample size
H_0	null hypothesis
$\mathscr{L}$	test statistic for linear models
$\mathcal{T}$	test statistic for partially linear models
λ	tuning parameter
$\widehat{\boldsymbol{\beta}}^{RFM}$	full model ridge regression estimator
$\widehat{\boldsymbol{\beta}}^{RSM}$	submodel ridge regression estimator
$\widehat{\boldsymbol{\beta}}^{RS}$	shrinkage ridge regression estimator
$\widehat{\boldsymbol{\beta}}^{RPS}$	positive shrinkage ridge regression estimator
$\widehat{\boldsymbol{\beta}}^{RPT}$	pretest ridge regression estimator
$I(\cdot)$	indicator function
$\widehat{\boldsymbol{\beta}}^{SRFM}$	semiparametric full model ridge regression estimator
$\widehat{\boldsymbol{\beta}}^{SRSM}$	semiparametric submodel ridge regression estimator
$\widehat{\boldsymbol{\beta}}^{SRS}$	semiparametric shrinkage ridge regression estimator
$\widehat{\boldsymbol{\beta}}^{SRPS}$	semiparametric positive shrinkage ridge regression estimator
$\widehat{\boldsymbol{\beta}}^{SRPT}$	semiparametric pretest ridge regression estimator
$\boldsymbol{W}$	positive semi-definite weight matrix in the quadratic loss function
$\boldsymbol{\Gamma}$	asymptotic distributional mean square error
$R(\cdot)$	asymptotic distributional quadratic risk of an estimator
K_n	local alternative hypothesis
$\boldsymbol{w}$	a fixed real valued vector in K_n
Δ	non-centrality parameter
Δ^*	a measure of the degree of deviation from the true model
$\boldsymbol{G}(\boldsymbol{y})$	non-degenerate distribution function of $\boldsymbol{y}$
$\square$	the symbol of end of the proof
$^\top$	matrix transpose operator

1 INTRODUCTION

Regression analysis consists of techniques for modelling the relationship between a response variable (also called dependent variable) and one or more explanatory variables (also known as independent variables or predictors) in statistics. It is used by data analysts in nearly every field of science and technology as well as economics, econometrics, finance. Statistical models are used to obtain information about unknown parameters based on sample information and, if available, other relevant information. The other information may be considered as non-sample information (NSI) (Ahmed, 2001). This is also known as uncertain prior information (UPI). Such information, which is usually available from previous studies, expert knowledge or researcher's experience, is unrelated to the sample data.The NSI may or may not positively contribute to the estimation procedure. However, it may be advantageous to use the NSI in the estimation process when sample information may be rather limited and may not be completely reliable.

The NSI can be classified as unknown, known and uncertain. For the three different scenarios, three different estimators, namely the unrestricted estimator (UE), restricted estimator (RE) and pretest estimator (PT) are defined in the literature (see Judge and Bock, 1978; Saleh, 2006).

In literature, (Bancroft, 1944) introduced a pretest test estimation of parameters to estimate the parameters of a model with uncertain prior information. Later, (Stein, 1956) introduced shrinkage estimator or Stein-type estimator which takes a hybrid approach by shrinking the base estimator to a plausible alternative estimator utilizing the NSI. Apart from Stein-type shrinkage estimation strategy, there are penalty estimators which are a class of estimators in the penalized least square family. The penalized estimation strategy performs variable selection and parameters estimation simultaneously. These estimators provide simultaneous variable selection and shrinkage of the coefficients towards zero.

In a multiple linear regression model, it is usually assumed that the explanatory variables are independent. But, in practice, there may be strong or near to strong linear relationships among the explanatory variables. In that case the independence assumptions are no longer valid, which causes the problem of multicollinearity. In the existence of multicollinearity may lead to wide confidence intervals for individ-

ual parameters or linear combination of the parameters and may produce estimates with wrong signs, etc. Therefore, the regression coefficient will be experienced with unduly large sampling variance which affects both inference and prediction. In literature, to overcome this problem, many studies have been made. Some biased estimations, such as Shrinkage Estimation, Principal Components Estimation (PCE), ridge estimation, Partial Least Squares Estimation (PLSE), Liu-type estimator, Almost unbiased estimation were proposed to improve the ordinary least square estimation. Recently, these methods have been considerably applied. Among them, ridge estimation is proposed by (Hoerl and Kennard, 1970a), is one of the most effective methods to solve the problem of multicollinearity and is the most popular one which has many usefulness in application. This estimator has less mean squared error (MSE) than ordinary least squares estimation.

In this dissertation we discuss prediction problems in two situations. First situation, p is smaller than the number of observations n, often written $p < n$. These problems are called low-dimensional data problems. Second situation, p is much larger than the number of observations n, often written $p > n$. These problems are called high-dimensional data problems. Along with the rapid development and widespread application of computer technology, the collection and storage of high dimensional data has become possible in genomics, financial markets data, data of mobile phone communication, data about DNA in biology.

1.1 Estimation Strategies

In this subsection, we define some estimator based on both full model and sub-model.

1.1.1 Full model estimators

Consider the form of the regression model

$$\boldsymbol{y} = f\left(\boldsymbol{X}, \boldsymbol{\beta}\right) + \boldsymbol{e}, \tag{1.1}$$

where $\boldsymbol{y}$ is $(n \times 1)$ the vector of responses, $\boldsymbol{X}$ is $(n \times p)$ a fixed design matrix, $\boldsymbol{\beta}$ is $(p \times 1)$ an unknown vector of parameters, f denotes a function of the data and

unknown vector of parameters, and $\boldsymbol{e}$ is $(n \times 1)$ the vector of unobservable random errors.

The general forms of the full model estimator (FM) or unrestricted estimators (UE) by $\widehat{\boldsymbol{\beta}}^{UE}$,

$$\widehat{\boldsymbol{\beta}}^{UE} = g\left(\boldsymbol{X}, \boldsymbol{y}\right),$$

where $g\left(\boldsymbol{X}, \boldsymbol{y}\right)$ denotes a function of the data and the response vector.

For given a linear regression model of the form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{1.2}$$

least square estimator (LSE) for fixed p and $n \geq p$ is given by

$$\widehat{\boldsymbol{\beta}}^{LSE} = \left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{y}. \tag{1.3}$$

where the superscript $\left(^{\top}\right)$ denotes the transpose of a vector or matrix and $\widehat{\boldsymbol{\beta}}^{LSE}$ is the best linear unbiased estimator (BLUE) and uniformly minimum-variance unbiased estimator or minimum-variance unbiased estimator (UMVUE or MVUE). However, this estimator may not be efficient in the presents of multicollinearity. Also, it cannot be used when $p > n$.

Multicollinearity is defined as the existence of nearly linear dependency among column vectors of the the design matrix $\boldsymbol{X}$ in the linear model (1.2). The existence of multicollinearity, the ordinary least square estimator can not be properly obtained. To overcome this problem, (Hoerl and Kennard, 1970a) suggested the use of $\left(\boldsymbol{X}^{\top}\boldsymbol{X} + \lambda^{R}\boldsymbol{I}_{p}\right), \lambda^{R} \geq 0$ instead of $\boldsymbol{X}^{\top}\boldsymbol{X}$ in the estimation. Also, the constant $\lambda^{R} \geq 0$ is knows as biasing or ridge parameter. The ridge coefficients are obtained by minimizing the objective function

$$\boldsymbol{e}^{\top}\boldsymbol{e} = \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^{\top}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right) \text{ subject to } \boldsymbol{\beta}^{2} \leq \phi, \tag{1.4}$$

where ϕ is inversely proportional to lambda. Hence, the ridge regression solutions are easily seen to be

$$\widehat{\boldsymbol{\beta}}^{ridge} = \left(\boldsymbol{X}^{\top}\boldsymbol{X} + \lambda^{R}\boldsymbol{I}_{p}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{y}, \tag{1.5}$$

where $\widehat{\boldsymbol{\beta}}^{ridge}$ is ridge estimator and $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. Note that the solution adds a positive constant to the diagonal of $\boldsymbol{X}^\top \boldsymbol{X}$ before inversion. This makes the problem nonsingular, even if $\boldsymbol{X}^\top \boldsymbol{X}$ is not of full rank. Of course, in the equation (1.5), if $\lambda^R = 0$, the least square estimator is obtained.

In situation n is smaller than the number of observations p, the Ridge method may provide a solution for the regression parameters estimation since $\left(\boldsymbol{X}^\top \boldsymbol{X} + \lambda^R \boldsymbol{I}_p\right)^{-1}$ will exist even for $n < p$.

In many situations, there is a priori information is available to the experimenter that parameter space may be belong to a candidate subset, that is it is suspected $\boldsymbol{H\beta} = \boldsymbol{h}$, $\boldsymbol{H}$ is a known matrix with dimension $(q \times p)$ and $\boldsymbol{h}$ is a vector with dimension $(q \times 1)$. In some situations, a subset of the covariates may be considered nuisance such that they are not of main interest, but they must be taken into account in estimating the coefficients of the remaining parameters. This information can be used to form a subset model.

1.1.2 Sub-model estimation strategy

Under the restriction $\boldsymbol{H\beta} = \boldsymbol{h}$, a sub-model estimator of $\boldsymbol{\beta}^*$ can be obtained as follows:

$$\widehat{\boldsymbol{\beta}}^{SM} = \widehat{\boldsymbol{\beta}}^{FM} - g\left(\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{H}\right),$$

where $g\left(\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{H}\right)$ is a function of the data, the response vector, $\boldsymbol{H}$ and $\boldsymbol{h}$.

It is well documented in the literature that is sub-model estimator is highly efficient when the restriction on the parameter space is nearly correct. On the other hand, as normalized distance between $\boldsymbol{H\beta}$ and $\boldsymbol{h}$ increases the $\widehat{\boldsymbol{\beta}}^{SM}$ becomes biased and inefficient relative to $\widehat{\boldsymbol{\beta}}^{FM}$.

1.2 Sparse Linear Models

In Sparse linear models it is assumed that $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^\top$ can be partitioned in to two sub-vectors as $\left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top$, where $\boldsymbol{\beta}_1 = (\beta_1, \beta_2, ..., \beta_{p_1})^\top$ and $\boldsymbol{\beta}_2 = \left(\beta_{p_1+1}, \beta_{p_1+2}, ..., \beta_{p_2}\right)^\top$, $p_1 + p_2 = p$. Usually p_1 is much smaller than p_2 under the assumption of sparsity. In this situation $\boldsymbol{\beta}_1$ is considered to be the coefficient vector for contributing set for prediction, and $\boldsymbol{\beta}_2$ is the vector for nuisance set and

set to be a null vector. Thus, if there is a priori known or suspected that a subset of the covariates do not significantly contribute in the overall prediction of the response variable, they may be left aside and a model without these covariates may be sufficient. Alternatively, this information can be obtained from existing variable selection methods, we called this information as auxiliary information (AE). However, it is important to note that, a subset of the covariates may be considered nuisance such that they are not of main interest, but they must be taken into account in estimating the coefficients of the contributing parameters.

To formulate this situation, the model (1.1) may be written as

$$\boldsymbol{y} = f\left(\boldsymbol{X}_1, \boldsymbol{\beta}_1\right) + f\left(\boldsymbol{X}_2, \boldsymbol{\beta}_2\right) + \boldsymbol{e}, \tag{1.6}$$

where $\boldsymbol{X}$ is partitioned as $\boldsymbol{X}_1$ with dimension $(n \times p_1)$ and $\boldsymbol{X}_2$ with dimension $(n \times p_2)$. In the case $\boldsymbol{H} = (\mathbf{0}, \boldsymbol{I})$, where $\mathbf{0}$ is known matrix with dimension $(p_2 \times p_1)$, $\boldsymbol{I}$ is identity matrix with dimension $(p_2 \times p_2)$, and $\boldsymbol{h} = \mathbf{0}$. In this situation $\boldsymbol{\beta}_2 = \mathbf{0}$, the model (1.6) is reduce to

$$\boldsymbol{y} = f\left(\boldsymbol{X}_1, \boldsymbol{\beta}_1\right) + \boldsymbol{e}. \tag{1.7}$$

Now, the general forms of the sub-model estimator $\widehat{\boldsymbol{\beta}}_1^{SM}$ is

$$\widehat{\boldsymbol{\beta}}_1^{SM} = \widehat{\boldsymbol{\beta}}_1^{FM} - g\left(\boldsymbol{X}, \boldsymbol{\beta}, \mathbf{0}, (\mathbf{0}, \mathbf{I})\right).$$

In other words, based on LSE we have $\widehat{\boldsymbol{\beta}}_1^{SM}$ and $\widehat{\boldsymbol{\beta}}_1^{FM}$, respectively,

$$\widehat{\boldsymbol{\beta}}_1^{SM} = \left(\boldsymbol{X}_1^\top \boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1^\top \boldsymbol{y}$$

and

$$\widehat{\boldsymbol{\beta}}_1^{FM} = \left(\boldsymbol{X}_1^\top \boldsymbol{M}_2 \boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1^\top \boldsymbol{M}_2 \boldsymbol{y},$$

where $\boldsymbol{M}_2 = \boldsymbol{I}_n - \boldsymbol{X}_2 \left(\boldsymbol{X}_2^\top \boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2^\top$, and ridge regression estimators will be

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

and

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

where $\boldsymbol{M}_2^R = \boldsymbol{I}_n - \boldsymbol{X}_2 \left(\boldsymbol{X}_2^\top \boldsymbol{X}_2 + \lambda^R \boldsymbol{I}_{p2}\right)^{-1} \boldsymbol{X}_2^\top$.

However, as stated earlier that sub-model estimators only work better than full model estimators in the neighbourhood of the restriction that $\boldsymbol{\beta}_2 = \mathbf{0}$. Generally speaking, $\widehat{\boldsymbol{\beta}}_1^{SM}$ behaves better than $\widehat{\boldsymbol{\beta}}_1^{FM}$ when $\boldsymbol{\beta}_2$ is close to $\mathbf{0}$. However, for $\boldsymbol{\beta}_2$ away from the pivotal $\mathbf{0}$, $\widehat{\boldsymbol{\beta}}_1^{SM}$ may becomes considerably biased, inefficient, and even possibly inconsistent. On the other hand, the estimate $\widehat{\boldsymbol{\beta}}_1^{FM}$ is consistent for departure of $\boldsymbol{\beta}_2$ from the null vector. Thus, we have two extreme estimation strategies $\widehat{\boldsymbol{\beta}}_1^{FM}$ and $\widehat{\boldsymbol{\beta}}_1^{SM}$ suited best for the models (1.1) and (1.7), respectively. In an effort to strike a compromise between $\widehat{\boldsymbol{\beta}}_1^{FM}$ and $\widehat{\boldsymbol{\beta}}_1^{SM}$ so that the resulting estimator behaves reasonably well relative to $\widehat{\boldsymbol{\beta}}_1^{FM}$ as well as to $\widehat{\boldsymbol{\beta}}_1^{SM}$. We suggest two estimators for the target parameters vector $\boldsymbol{\beta}_1$ of the regression model in (1.1). The first estimator is called the pretest estimator, denoted by $\widehat{\boldsymbol{\beta}}_1^{PT}$. This estimator is a convex combination of $\widehat{\boldsymbol{\beta}}_1^{FM}$ and $\widehat{\boldsymbol{\beta}}_1^{SM}$ through an indicator function $I(\mathcal{D}_n < d_{n,\alpha})$, where $\mathcal{D}_n$ is an appropriate test-function to test the null hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_a : \boldsymbol{\beta_2} \neq \mathbf{0}$. Also, $d_{n,\alpha}$ is an α-level critical value using the distribution of the test statistic $\mathcal{D}_n$. The pretest test estimator selects $\widehat{\boldsymbol{\beta}}_1^{FM}$ or $\widehat{\boldsymbol{\beta}}_1^{SM}$ as per H_0 is accepted or rejected. Keeping in mind that, deciding against H_a does not mean that we have the clear evidence that $\boldsymbol{\beta}_2 = \mathbf{0}$, simply we do have control of the probability of type I error. Our main objective is here to find an efficient estimator $\boldsymbol{\beta}_1$, so in this context we think that we are obtaining a better estimator of $\boldsymbol{\beta}_1$ by setting $\boldsymbol{\beta}_2 = \mathbf{0}$. Thus, $d_{n,\alpha}$ is a threshold that determines a hard thresholding rule, and α may be considered as a tuning parameter. We can construct another estimator based on the James–Stein rule, known as the shrinkage estimator. Interestingly, this shrinkage estimator $\widehat{\boldsymbol{\beta}}_1^{S}$ is a smooth function of the test statistic $\mathcal{D}_n$.

1.2.1 Pretest estimation strategy

(Bancroft, 1944) introduced the pretest test estimation procedure as one basis for dealing with model-estimator uncertainty, we refer to (Ahmed, 2014). The

pretest estimator (PT) $\widehat{\boldsymbol{\beta}}_1^{PT}$ of $\boldsymbol{\beta}_1$ is defined by

$$\widehat{\boldsymbol{\beta}}_1^{PT} = \widehat{\boldsymbol{\beta}}_1^{FM} - \left(\widehat{\boldsymbol{\beta}}_1^{FM} - \widehat{\boldsymbol{\beta}}_1^{SM}\right) I\left(\mathcal{D}_n < d_{n,\alpha}\right),$$

where $\mathcal{D}_n$ is an appropriate test-function to test the null hypothesis, to be defined later and $d_{n,\alpha}$ is the $100\left(1-\alpha\right)$ percentage point of the test statistic $\mathcal{D}_n$. By design, $\widehat{\boldsymbol{\beta}}_1^{PT}$ selects $\widehat{\boldsymbol{\beta}}_1^{SM}$ when the null hypothesis is not rejected; otherwise, $\widehat{\boldsymbol{\beta}}_1^{FM}$ is selected. Evidently, the variation in $\widehat{\boldsymbol{\beta}}_1^{PT}$ is now controlled effectively, depending on the size α of the test, but the pretest test estimation methodology makes two extreme choices of either $\widehat{\boldsymbol{\beta}}_1^{SM}$ or $\widehat{\boldsymbol{\beta}}_1^{FM}$. It is well documented in the literature that the pretest test procedures are not uniformly superior to benchmark estimator, even though they may improve on classical procedures.

For this reason. we suggest an estimation strategy based on James-Stein rule. (Stein, 1956; James and Stein, 1961) showed the inadmissibility of the maximum likelihood estimator when estimating a multivariate mean vector under quadratic loss. (Scolve et al., 1978) proved the non-optimality of the pretest test estimator in multi-parametric situations.

Similarly, the pretest test ridge regression estimator (RPT) $\widehat{\boldsymbol{\beta}}_1^{RPT}$ of $\boldsymbol{\beta}_1$ is

$$\widehat{\boldsymbol{\beta}}_1^{RPT} = \widehat{\boldsymbol{\beta}}_1^{RFM} - \left(\widehat{\boldsymbol{\beta}}_1^{RFM} - \widehat{\boldsymbol{\beta}}_1^{RSM}\right) I\left(\mathcal{D}_n < d_{n,\alpha}\right).$$

1.2.2 Shrinkage estimation strategy

Following (Ahmed, 2001, 2014), we define shrinkage estimator (SE) $\widehat{\boldsymbol{\beta}}_1^{S}$ of $\boldsymbol{\beta}_1$ is defined by

$$\widehat{\boldsymbol{\beta}}_1^{S} = \widehat{\boldsymbol{\beta}}_1^{SM} + \left(\widehat{\boldsymbol{\beta}}_1^{FM} - \widehat{\boldsymbol{\beta}}_1^{SM}\right)\left(1 - c\mathcal{D}_n^{-1}\right),$$

where c is an optimum constant that minimizes the risk. The estimator $\widehat{\boldsymbol{\beta}}_1^{S}$ can be viewed as a member of general form of the Stein-rule family of estimators. However, in this case the shrinkage of the benchmark estimator is towards the sub-model estimator $\widehat{\boldsymbol{\beta}}_1^{SM}$. The shrinkage estimator is pulled towards the sub-model estimator when the variance of the least squares estimator is large, and pulled towards the full model estimator when the alternative estimator has high variance, high bias, or is more highly correlated with the least squares estimator. It is seen that $\widehat{\boldsymbol{\beta}}_1^{S}$ is the

smooth version of $\widehat{\boldsymbol{\beta}}_1^{PT}$, since $\mathcal{D}_n$ is smooth function. Using the terminology of (Donoho and Johnstone, 1998), $\widehat{\boldsymbol{\beta}}_1^{PT}$ and $\widehat{\boldsymbol{\beta}}_1^{S}$ are based on hard and smooth thresholdings, respectively. Generally speaking, $\widehat{\boldsymbol{\beta}}_1^{S}$ adapts to the magnitude of $\mathcal{D}_n$ and tends to $\widehat{\boldsymbol{\beta}}_1^{FM}$ as $\mathcal{D}_n$ wonders to ∞ and to $\widehat{\boldsymbol{\beta}}_1^{SM}$ as $\mathcal{D}_n \to c$. Similar findings hold for $\widehat{\boldsymbol{\beta}}_1^{PT}$. In passing we would like to remark here that the Steinian strategy is similar in spirit to the model-averaging procedures, Bayesian or otherwise; (see Bickel, 1984; Hoeting et al., 1999, 2002; Burhnam and Anderson, 2002).

Sometimes it would be one problem with SE is that its components may have a different sign the coordinates of $\widehat{\boldsymbol{\beta}}^{UE}$ if $c\mathcal{D}_n^{-1}$ is larger than unity. Practically, the change of sign would affect its interpretability. But, this does not adversely affect the risk performance of SE. To overcome this problem, a positive Shrinkage estimator (PSE) has been defined by retaining the positive part of the SE. A positive shrinkage estimator (PSE) $\widehat{\boldsymbol{\beta}}_1^{PS}$ of $\boldsymbol{\beta}_1$ is defined by

$$\widehat{\boldsymbol{\beta}}^{PS} = \widehat{\boldsymbol{\beta}}_1^{SM} + \left(\widehat{\boldsymbol{\beta}}_1^{FM} - \widehat{\boldsymbol{\beta}}_1^{SM}\right)\left(1 - c\mathcal{D}_n^{-1}\right)^{+},$$

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

In the same sprit, we define the shrinkage ridge regression estimator (RS) and its positive part estimator (RPS) $\widehat{\boldsymbol{\beta}}_1^{RS}$ and $\widehat{\boldsymbol{\beta}}_1^{RPS}$ of $\boldsymbol{\beta}_1$, respectively, as follows:

$$\widehat{\boldsymbol{\beta}}_1^{RS} = \widehat{\boldsymbol{\beta}}_1^{RSM} + \left(\widehat{\boldsymbol{\beta}}_1^{RFM} - \widehat{\boldsymbol{\beta}}_1^{RSM}\right)\left(1 - c\mathcal{D}_n^{-1}\right)$$

and

$$\widehat{\boldsymbol{\beta}}^{RPS} = \widehat{\boldsymbol{\beta}}_1^{RSM} + \left(\widehat{\boldsymbol{\beta}}_1^{RFM} - \widehat{\boldsymbol{\beta}}_1^{RSM}\right)\left(1 - c\mathcal{D}_n^{-1}\right)^{+}.$$

1.3 Penalized Estimation

We now briefly discuss penalized estimation. These estimator are members of the penalized least squares (PLS) family, and are suitable for both low-dimensional and high-dimensional data analysis. A popular version of the PLS is known as Tikhonov regularization (Tikhonov, 1963). (Ahmed et al., 2010; Ahmed, 2014) pointed out that PLS estimation provides a generalization of both nonparametric least squares and weighted projection estimators,

The idea of penalized estimation was introduced by (Frank and Friedman, 1993).

They suggested the notion of bridge regression as follows. For a given penalty function $\pi(\cdot)$ and tuning parameter that controls the amount of shrinkage λ, bridge estimators are estimated by minimizing the following PLS criterion

$$\sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2 + \lambda \pi(\beta), \tag{1.8}$$

where y_i's are responses and $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})^\top$ are data matrix points. The general form of penalty function can be defined as follows:

$$\pi(\beta) = \sum_{j=1}^{p} |\beta_j|^\gamma, \ \gamma > 0. \tag{1.9}$$

We observe that the penalty function in (1.9) bounds the L_γ norm of the parameters in the given model as $\sum_{j=1}^{p} |\beta_j|^\gamma \leq \rho$, where ρ is the tuning parameter that controls the amount of shrinkage.

1.3.1 L_2 penalty strategy

Interestingly when $\gamma = 2$, an important member of the penalized least squares (PLS) family is ridge estimators which is introduced by (Hoerl and Kennard, 1970a)

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

It can be easily verify that the solution of above equation is

$$\widehat{\boldsymbol{\beta}}^{ridge} = \left(\boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I}_p\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}.$$

Thus, ridge regression can be also viewed as a penalty estimator and it will provide regression parameters estimation when $p > n$.

Interestingly, when $\gamma \leq 1$, the estimators minimizing (1.8) have the potentially attractive properties of forcing weak or moderate regression coefficients exactly exactly zero. For example for $\gamma = 1$, it performs variable selection and parameter estimation simultaneously. This special case was proposed by (Tibshirani, 1996), we briefly describe this procedure in the following subsection.

1.3.2 Lasso strategy

For $\gamma = 1$, we obtain the L_1 penalized least squares estimator or commonly known as LASSO (Least Absolute Shrinkage and Selection Operator) which is introduced by (Tibshirani, 1996). Because LASSO is an acronym, we use lasso and LASSO interchangeably in this thesis. Lasso estimators are obtained as follows

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

The parameter $\lambda \geq 0$ controls the amount of shrinkage. However, this procedure is not oracle.

To resolve this issue, (Zou, 2006) introduced adaptive lasso (aLasso) and (Fan and Li, 2001) defined the smoothly clipped absolute penalty (SCAD).

1.3.3 Adaptive lasso strategy

The adaptive lasso estimator $\boldsymbol{\beta}^{aLasso}$ is defined as

$$\widehat{\boldsymbol{\beta}}^{aLasso} = \underset{\boldsymbol{\beta}}{\arg\min} \left\{ \sum\nolimits_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2 + \lambda \sum_{j=1}^{p} \widehat{\xi}_j |\beta_j| \right\}, \tag{1.10}$$

where the wight function is

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

The $\boldsymbol{\beta}^*$ is a consistent estimator of $\boldsymbol{\beta}$. For example, least square estimator can be used as a starting point. The adaptive lasso is essentially an L_1 penalization method and the estimates in (1.10) can be solved by LARS algorithm, (Efron et al., 2004). For computation details we refer to (Zou, 2006).

1.3.4 SCAD strategy

The smoothly clipped absolute deviation (SCAD) is proposed by (Fan and Li, 2001). The L_γ and hard thresholding penalty functions do not simultaneously satisfy the mathematical conditions for unbiasedness, sparsity, and continuity. The

continuous differentiable penalty function defined by

$$J_{\alpha,\lambda}(x) = \lambda \left\{ I(|x| \leq \lambda) + \frac{(\alpha\lambda - |x|)_+}{(\alpha - 1)\lambda} I(|x| > \lambda) \right\}, \quad x \geq 0, \tag{1.11}$$

for some $\alpha > 2$ and $\lambda > 0$. SCAD penalty is a symmetric and a quadratic spline on $(0, \infty]$ with knots λ and $\alpha\lambda$. if $\alpha = \infty$, the expression (1.11) is equivalent to the L_1 penalty in LASSO. The SCAD estimation is given by

$$\widehat{\boldsymbol{\beta}}^{SCAD} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2 + \lambda \sum_{j=1}^{p} J_{\alpha,\lambda} \left\|\beta_j\right\|_1 \right\},$$

where $\|\cdot\|_1$ denotes L_1 norm.

The penalty estimation can be viewed as variable selection and parameter estimation (for remaining parameters) specially when $p > n$. However, there are other technique available for variable selection. We need to select a subset of the covariates from full model, because the shrinkage estimation method combines unrestricted and restricted estimators. In this framework, If prior information about a subset of the covariates is available, then the estimates can be obtained by incorporating the available information in the estimation process. But, if we have not prior information, one might do usual variable selection to select the best subsets. In literature, there are a lot of subset selection methods available such as the Akaike information criterion (AIC), Mallows C_p, stepwise, backward and forward selection procedures, cross-validation and the Bayesian information criterion (BIC) among other.

1.3.5 AIC and BIC

AIC is given by (Akaike, 1977)

$$\text{AIC} = -2(\text{log maximized likelihood}) + 2(\text{number of parameters}).$$

Another seemingly related criteria is BIC or Schwarz criterion was developed (Schwarz, 1978), which is

$$\text{BIC} = -2(\text{log maximized likelihood}) + (\text{number of parameters}) \log n.$$

1.3.6 Best subset selection (BSS)

Best subset regression finds for each $k \in (0, 1, 2, .., p)$ the subset of size k that gives smallest residual sum of squares, (Hastie et al., 2009). An efficient algorithm is the leaps and bounds procedure in R package, (Furnival and Wilson, 1974). To obtain a sub-model by using BSS search all possible models and take the one with highest adjusted R^2 or lowest C_p.

1.4 Review of Literature

The use of *uncertain prior information* (UPI) or *non sample information* (NSI) have been common in the statistical inference of *conditionally specified models* for the past few decades. UPI is usually integrated in a model at hand with an assumed restriction on the model parameters, resulting in candidate sub-models. The restriction on the parameters are usually obtained through information from past data. Nowadays the assumption of sparsity is a common one, when models are sparse sub-models can be obtained through variable selection methods, such as BIC, AIC, and BSS, among others. A sub-model is more attractive and useful from practitioners perspectives. When a given sub-model is correct one then it provides an efficient statistical inferences than inference based on a full model. However, the estimators based on a sub-model may become considerably biased and inefficient. In past two decades, simultaneous variable selection and estimation of sub-model parameters has become popular. Not to mention, such procedures may also give biased estimators of the given sub-model parameters. To deal with model uncertainty, pretest and shrinkage estimation strategies existed in the literature for more than five decades which incorporate uncertain prior information into the estimation procedure. (Bancroft, 1944) suggested the *pretest estimation* (PE) strategy. The pretest estimation approach uses a test to decide the estimator based on either the sub-model or a full model. (Bancroft, 1944) suggested two problems on pretesting strategy. One is a data pooling problem from various sources based on a pretest approach and other one is related to simultaneous model selection and pretest estimation problem in regression model. A lot of research work has been done using the pretest estimation procedure in a host of applications. (Stein, 1956) and (James and Stein, 1961) suggested shrinkage estimation strategy. The shrinkage strategy may be re-

garded as a smooth version of the (Bancroft, 1944) pretest estimation procedure. Since its inception, both shrinkage and pretest estimation methods has received a lot of attention from researchers and practitioners. It has been demonstrated that shrinkage estimation strategy dominates the classical estimators in many situations. During the past three decades, Ahmed and his co-researchers, among others, have demonstrated that shrinkage estimators outclass the traditional estimators in terms of *mean squared error* (MSE) for a host of statistical models. A detailed description of shrinkage estimation and large sample estimation techniques in a regression model can be found in (Ahmed, 2014). As stated earlier the penalty estimators are members of the penalized least squares family. Commonly used penalty estimators includes the least absolute and shrinkage operator (LASSO), adaptive LASSO, group LASSO, the smoothly clipped absolute deviation (SCAD), and minimax concave penalty (MCP), among others. The procedure performs variable selection and shrinkage simultaneously. The procedure selects a sub-model and estimates the regression parameters in the sub-model. This technique is effective when the model is sparse, and also applicable when number of predictors (p) is greater than the number of observations (n). It is important to note that the output of penalty estimation resembles shrinkage methods as it shrinks and selects the variables simultaneously. However, there is an important difference in how the shrinking works in penalty estimation when compared to the shrinkage estimation. The penalty estimator shrinks the full model estimator towards zero and depending on the value of the tuning or penalty parameter, it sets some coefficients to zero exactly. Thus, penalty procedure does variable selection automatically by treating all the variables equally. It does not single out nuisance covariates, or for that matter, the UPI, for special scrutiny as to their usefulness in estimating the coefficients of the active predictors. However, SCAD, MCP, and adaptive LASSO, on the other hand, are able to pick the right set of variables while shrinking the coefficients of the regression model. Lasso-type regularization approaches have some advantages of generating a parsimony sparse model, but are not able to separate covariates with small contribution and covariates with no contributions. This could be a serious problem if there was a large number of covariates with small contributions and were forced to shrink towards zero. In the reviewed published studies on high dimensional data analysis, it has

been assumed that the signals and noises are well separated. This presentation is this thesis fundamental, because pretest and shrinkage strategies may be appropriate for model selection problems and there is a growing interest in this area to fill the gap between two competitive strategies when either p is increasing with n or $p > n$. The goal of this thesis is to discuss some of the issues involved in the estimation of the parameters in two linear models that may be over-parameterized by including too many variables in the model using penalty and non-penalty estimation strategies. For example, in genomics research it is common practise to test a given subset of genetic markers in association with disease (Zeggini et al., 2007). Here the subset is found in a certain population by doing genome wide association studies. The subset is then tested for disease association in a new population. In this new population it is possible that genetic markers not found in the first population are associated with disease. Pretest and shrinkage strategies are used to trade-off between bias and efficiency in a data adaptive way in order to provide efficient and useful solutions to problems in a host applications. We summarize some of the important works found in the reviewed literature pertaining to this dissertation. The pretest and shrinkage estimation techniques have received considerable attention from the researchers since their introduction. Small sample and asymptotic properties of shrinkage and pretest estimators using quadratic loss function, and their dominance over the classical estimators have been documented in numerous studies in the literature. Since 1987, Ahmed and his co-researchers are among others who have analytically demonstrated that shrinkage estimators outshine the classical estimator. (Ahmed, 1997, 2014) gave a detailed description of shrinkage estimation, and discusse large sample estimation in a regression model with non-normal error terms. A review of shrinkage and some penalty estimators can be found in (van Houwelingen, 2001). In their work, the shrinkage, pretest estimator, ridge regression, LASSO and the Garotte estimators are discussed from a Bayesian framework. An application of empirical Bayes shrinkage estimation is in (Castner and Schirm, 2003). For fixed n and p, (Khan and Ahmed, 2003) considered the problem of estimating the coefficient vector of a multiple regression model when it is a priori suspected that the parameter vector may belong to a reduced space. They established that the positive-part of shrinkage estimator dominated the least square

estimator.

For fixed p, a comparison of penalty and non-penalty estimators in PLM can be found in (Ahmed et al., 2007). There has been no study in the reviewed literature to compare the risk properties of non-penalty and penalty estimators in the context of PLM model, when p may be greater than n and using ridge estimator as a full model estimator. In this dissertation, based on a ridge regression we construct high dimensional shrinkage estimators. We compare high dimensional non-penalty estimators and in linear and partially linear models with some penalty estimators.

1.5 Objective, Contribution and Organization of the Thesis

There are four main objectives of this dissertation:

(i) to construct a high dimensional shrinkage estimators using ridge regression for linear and partially linear models, respectively.

(ii) to investigate the relative performance of high dimensional shrinkage estimator to ridge regression estimator.

(iii) implement the penalty estimators for these models for simultaneous variable selection and regression parameters estimation.

(iv) Critically assess the relative properties of penalty and non-penalty estimators.

The properties of pretest and shrinkage estimators for fixed dimension have been extensively investigated and well document in reviewed literature, for example we refer to (Ahmed, 2001, 2002; Ahmed et al., 2010; Nkurunziza and Ahmed, 2011; Ahmed and Raheem, 2012; Fallahpour and Ahmed, 2013; Ahmed, 2014). However, there is little available for high dimensional shrinkage estimators, we refer to (Ahmed, 2014). For this reason we propose high dimensional shrinkage estimators using ridge regression estimator as a full model estimator. These estimators can be classified as non-penalty estimators. On the other hand, since the inception of LASSO (Tibshirani, 1996) there has been a bulk amount of research in penalty estimation techniques has been done. A little work has been done to compare the relative performance of non-penalty and penalty estimators. For fixed p, (Ahmed et al., 2007) are the first to compare shrinkage and LASSO estimates in the context of a PLM. (Raheem et al., 2012) compares shrinkage and LASSO in linear regres-

sion models. Therefore, in this thesis, we first propose shrinkage estimators for $p > n$ using ridge regression estimate as a benchmark estimator. Then compare the performance and penalty and non-penalty estimators in multiple linear regression. For illustration purpose, real data examples are give by calculating average prediction error through cross validation. Monte Carlo study is conducted with low- and high-dimensional data to compare the predictive performance of non-penalty estimators with those of LASSO, ALASSO, and SCAD estimators. Further, we extend high dimensional shrinkage estimators for a PLM. In particular, we wish to study the the suitability of smoothing spline basis function in estimating the non-parametric component in a PLM to obtain ridge regression estimates. Some useful procedures for simultaneous sub-model selection are developed and implemented to obtain the parameter estimates after incorporating the smooth spline bases at the model at hand. We will also investigate and compare the performance of the non-penalty estimators with some penalty estimators. The dissertation is divided into four chapters. Chapter 1 introduces various shrinkage and pretest estimators along with penalty estimators, namely, the least absolute penalty and shrinkage operator (LASSO), adaptive LASSO (aLASSO), and the smoothly clipped absolute deviation (SCAD) have been defined. In Chapter 2, we present least square, ridge, penalty, shrinkage and pretest estimation strategies in a multiple regression model. Our full and sub-model estimators are based on ridge regression. We propose and construct high dimensional shrinkage estimator using ridge regression estimator. Asymptotic bias and MSE expressions of the non-penalty estimators have been derived and the performance of theses estimators are compared with the least square and ridge estimators, respectively. Further, some numerical results are showcased using real data examples and through Monte Carlo simulation experiments. Several penalty estimators such as LASSO, adaptive LASSO, and SCAD estimators have been throughly discussed for high dimensional data. Monte Carlo simulation studies have been used to compare the performance of some non-penalty and penalty estimators. We consider parameter estimation of a partially linear regression model by using smoothing spline method in chapter 3. We construct new estimation strategies based on ridge regression for high dimensional datasets. Using ridge regression estimator we construct high dimensional shrinkage estimators. The performance of

some penalty estimators namely LASSO, adaptive LASSO, and SCAD estimators are also studied for this model. The bias and risk properties of the estimators are throughly investigated using asymptotic distributional bias and risk, respectively. We conduct an extensive Monte Carlo simulation studies to examine the relative performance of these estimators. Finally, concluding remarks and an outline for some open problems for future research are presented in Chapter 4.

2 PENALTY and NON-PENALTY ESTIMATION STRATEGIES in LINEAR MODELS

2.1 Introduction

In this chapter, we establish non-penalty estimation based on ridge regression, and compare their performance with penalty estimators. We consider two situation to estimate the suggested estimators: firstly, we obtain penalty estimators for low-dimensional data. Lastly, it is acquired for high-dimensional data.

2.2 Organization of the Chapter

We dimidiate this chapter. In the first half, we demonstrate shrinkage ridge regression estimation strategies when the number of experimental units is larger than the number of covariates ($n > p$). To this end, we present full model estimation by using ridge regression, and sub-model estimation in the following two subsection. Asymptotic properties of ridge regression full model and sub-model estimators, pretest ridge regression, shrinkage ridge regression and positive shrinkage ridge regression estimators are obtained in Section 2.5. Monte Carlo simulation studies are obtained to compare non-penalty estimators with penalty estimators in Section 2.6. Also, applications of the suggested estimators and penalty estimators are demonstrated with three real data in Section 2.8. In the second half of this chapter, we suggest test statistic, and present penalty estimators when the number of experimental units is smaller than the number of covariates ($n < p$). In Section 2.10, Application of these estimators is illustrated with two real data examples. Finally, in Section 2.11, performance of penalty and non-penalty is compared using a Monte Carlo experiment.

2.3 Full Model Estimation

Consider a regression linear model

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{2.1}$$

where $y_i^\top s$ are responses, $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})^\top$ is design points, $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^\top$ is vector denoting unknown coefficients, ε_i's are unobservable random errors and the superscript $\left(^\top\right)$ denotes the transpose of a vector or matrix. Further, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^\top$ has a cumulative distribution function $F(\boldsymbol{\varepsilon})$; $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $Var(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$, where σ^2 is finite and $\boldsymbol{I}_n$ is an identity matrix of dimension $n \times n$. It is observed from model (2.1) that the ordinary least square estimator (OLS) of $\boldsymbol{\beta}$ depends mostly on the characteristics of the matrix $\boldsymbol{X}^\top \boldsymbol{X}$, $\left(\boldsymbol{X} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{..., x}_n)^\top\right)$. If there exists multicollinearity in $\boldsymbol{X}$ (that is $\boldsymbol{X}^\top \boldsymbol{X}$ is ill-conditioned), then the OLS estimator produces unduly large sample variances. In practice, the researchers often encounter the problem of multicollinearity. In literature, there are various methods available to deal with this problem. Among them, ridge regression is the most popular one. To solve the problem of multicollinearity, (Hoerl and Kennard, 1970a) suggested the use of $\left(\boldsymbol{X}^\top \boldsymbol{X} + \lambda^R \boldsymbol{I}_p\right)$, $\lambda^R \geq 0$ is known as ridge parameter. The ridge estimator can be is obtained from the following

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

where ϕ is inversely proportional to λ^R, which is equal to

$$\arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2 + \lambda^R \sum\nolimits_{j=1}^{p} \beta_j^2 \right\}. \tag{2.3}$$

It yields

$$\widehat{\boldsymbol{\beta}}^{RFM} = \left(\boldsymbol{X}^\top \boldsymbol{X} + \lambda^R \boldsymbol{I}_p\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}, \tag{2.4}$$

where $\widehat{\boldsymbol{\beta}}^{RFM}$ is called ridge estimator and $\boldsymbol{y} = (y_1, y_2, ..., y_n)^\top$. If $\lambda^R = 0$, then $\widehat{\boldsymbol{\beta}}^{RFM}$ is OLS estimator, and $\lambda^R = \infty$, then $\widehat{\boldsymbol{\beta}}^{RFM} = \mathbf{0}$.

In literature, Ridge regression methods have been considered by many researchers, beginning with (Hoerl and Kennard, 1970a,b), and followed by (Gibbons, 1981), (Vinod and Ullah, 1981), (Sarkar, 1992), (Saleh and Kibria, 1993), (Gruber, 1998), (Singh and Tracy, 1999), (Tabatabaey, 1995), (Wencheko, 2000), (Inoue, 2001), (Montgomery, 2001), (Kibria, 1996a,b, 2003, 2004), (Saleh, 2006), (Kibria and Saleh, 1993, 2003, 2004a,b).

2.4 Sub-Model Estimation

(Najarian et al., 2013) considered some restricted ridge regression estimators under a linear restriction $\boldsymbol{R\beta} = \boldsymbol{r}$ and provided a simulation analysis. However, it is well documented in the literature that restricted or sub-model estimator are biased and inefficient when linear restriction may not hold. For this reason we consider pretest and shrinkage strategy to control the magnitude of the bias. (Ahmed, 2001) gave a detailed definition of shrinkage estimation, and discussed large sample estimation techniques in a regression model with non-normal errors. Theoretical comparison and relationship of restricted ridge estimator with related methods are described in (Kaçıranlar et al., 2011).

In this study, we consider a linear regression model (2.1) in a more realistic situation when model is assumed to be sparse. Under this assumption, the vector of coefficients $\boldsymbol{\beta}$ can be partitioned as $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ where $\boldsymbol{\beta}_1$ is the coefficient vector for main effects, and $\boldsymbol{\beta}_2$ is the vector for nuisance effects or insignificant coefficients. We are essentially interested in the estimation of $\boldsymbol{\beta}_1$ when it is reasonable that $\boldsymbol{\beta}_2$ is close to zero.

A sub-model or restricted model with a general restriction is defined as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad \text{subject to } \boldsymbol{\beta}^\top \boldsymbol{\beta} \leq \phi \text{ and } \boldsymbol{R\beta} = \boldsymbol{r}, \tag{2.5}$$

where $\boldsymbol{R}$ is an $s \times p$ restriction matrix, and $\boldsymbol{r}$ is an $s \times 1$ vector of constants.

In this study, we let $\boldsymbol{X} = (\boldsymbol{X_1}, \boldsymbol{X_2})$, where $\boldsymbol{X_1}$ is an $n \times p_1$ sub-matrix containing the regressors of interest and $\boldsymbol{X_2}$ is an $n \times p_2$ sub-matrix that may or may not be relevant in the analysis of the main regressors. Similarly, $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top$ be the vector of parameters, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have dimensions p_1 and p_2, respectively, with $p_1 + p_2 = p$, $p_i \geq 0$ for $i = 1, 2$. We are constitutively interested in the estimation of $\boldsymbol{\beta}_1$ when it is suspected that $\boldsymbol{\beta}_2$ is close to $\mathbf{0}$. In order to obtain relevant hypothesis, we consider $\boldsymbol{R} = [\mathbf{0}, \boldsymbol{I}]$, and $\boldsymbol{r} = \mathbf{0}$, where $\mathbf{0}$ is a $p_2 \times p_1$ matrix of zeroes and $\boldsymbol{I}$ is the identity matrix. Then we call $\widehat{\boldsymbol{\beta}}_1^{RFM}$ the full model or unrestricted ridge estimator of $\boldsymbol{\beta}_1$. if $\boldsymbol{\beta}_2 = \mathbf{0}$, then we have the following restricted

linear regression model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} \text{ subject to } \boldsymbol{\beta}_1^\top\boldsymbol{\beta}_1 \leq \phi. \tag{2.6}$$

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

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

where $\boldsymbol{M}_2^R = \boldsymbol{I}_n - \boldsymbol{X}_2\left(\boldsymbol{X}_2^\top\boldsymbol{X}_2 + \lambda^R\boldsymbol{I}_{p_2}\right)^{-1}\boldsymbol{X}_2^\top$. For model (2.6), the sub-model or restricted estimator $\widehat{\boldsymbol{\beta}}_1^{RSM}$ of $\boldsymbol{\beta}_1$ has the form

$$\widehat{\boldsymbol{\beta}}_1^{RSM} = \left(\boldsymbol{X}_1^\top\boldsymbol{X}_1 + \lambda_1^R\boldsymbol{I}_{p_1}\right)^{-1}\boldsymbol{X}_1^\top\boldsymbol{y}.$$

Generally speaking, $\widehat{\boldsymbol{\beta}}_1^{RSM}$ performs better than $\widehat{\boldsymbol{\beta}}_1^{RFM}$ when $\boldsymbol{\beta}_2$ close to zero. However, for $\boldsymbol{\beta}_2$ away from the zero, $\widehat{\boldsymbol{\beta}}_1^{RSM}$ can be inefficient. But, the estimate $\widehat{\boldsymbol{\beta}}_1^{RFM}$ is consistent for departure of $\boldsymbol{\beta}_2$ from zero. We also consider pretest and shrinkage estimators for model (2.2). The pretest test estimator was introduced by (Bancroft, 1944). This estimator is a combination of $\widehat{\boldsymbol{\beta}}_1^{RFM}$ and $\widehat{\boldsymbol{\beta}}_1^{RSM}$ through an indicator function $I\left(\mathscr{L}_n \leq c_{n,\alpha}\right)$, where $\mathscr{L}_n$ is appropriate test statistic to test $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_A : \boldsymbol{\beta}_2 \neq \mathbf{0}$. Moreover, $c_{n,\alpha}$ is an $\alpha-$level critical value using the distribution of $\mathscr{L}_n$. We define test statistics as follows:

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

where

$$\widehat{\sigma}^2 = \frac{1}{n-p}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{RFM})^\top(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{RFM}),$$

and $\boldsymbol{M}_1 = \boldsymbol{I}_n - \boldsymbol{X}_1\left(\boldsymbol{X}_1^\top\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1^\top$, $\widehat{\boldsymbol{\beta}}_2^{FM} = \left(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y}$. Under H_0, the test statistic $\mathscr{L}_n$ follows chi-square distribution with p_2 degrees of freedom for large n values. Now, we can choose an $\alpha-$level critical value $c_{n,\alpha}$ and introduce pretest test ridge regression estimator of $\boldsymbol{\beta}_1$ defined by

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

Hence, $\widehat{\boldsymbol{\beta}}_1^{RPT}$ choose $\widehat{\boldsymbol{\beta}}_1^{RSM}$ when H_0 is tenable; otherwise, $\widehat{\boldsymbol{\beta}}_1^{RFM}$ is chosen.

The shrinkage or stein-type ridge regression estimator of $\boldsymbol{\beta}_1$ defined by

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

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

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

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

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

2.5 Asymptotic Analysis

In this section, we present the expressions for asymptotic distributional bias (ADB), asymptotic quadratic distributional bias (AQDB), asymptotic distributional covariance matrices and asymptotic distributional risks (ADR) of the listed estimators.

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

$$K_n : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{2(n)} = \frac{\boldsymbol{w}}{\sqrt{n}}, \quad \boldsymbol{w} = (w_1, w_2, ..., w_{p_2})^{\top} \in \mathbb{R}^{p_2},$$

Thus, for $\boldsymbol{w} = \mathbf{0}$ it reduces to H_0.

To study the asymptotic quadratic risks of $\widehat{\boldsymbol{\beta}}_1^{RFM}$, $\widehat{\boldsymbol{\beta}}_1^{RSM}$, $\widehat{\boldsymbol{\beta}}_1^{RPT}$, $\widehat{\boldsymbol{\beta}}_1^{RS}$ and $\widehat{\boldsymbol{\beta}}_1^{RPS}$, we may define a quadratic loss function using a positive definite matrix (p.d.m) $\boldsymbol{W}$, by

$$L\left(\boldsymbol{\beta}_1^{*} - \boldsymbol{\beta}_1\right) = n\left(\boldsymbol{\beta}_1^{*} - \boldsymbol{\beta}_1\right)^{\top} \boldsymbol{W}\left(\boldsymbol{\beta}_1^{*} - \boldsymbol{\beta}_1\right),$$

where $\boldsymbol{\beta}_1^{*}$ is anyone of suggested estimators. Now, under $\{K_n\}$, we can write the asymptotic distribution function of $\boldsymbol{\beta}_1^{*}$ as

$$F\left(\boldsymbol{x}\right) = \lim_{n\to\infty} P\left(\sqrt{n}\left(\boldsymbol{\beta}_1^{*} - \boldsymbol{\beta}_1\right) \leq \boldsymbol{x} | K_n\right),$$

where $F(\boldsymbol{x})$ is non degenerate. Then ADR of $\boldsymbol{\beta}_1^*$ is defined as follows:

$$R(\boldsymbol{\beta}_1^*) = \text{tr}\left(\boldsymbol{W}\int_{\mathbb{R}^{p_1}}\int \boldsymbol{x}\boldsymbol{x}^\top dF(\boldsymbol{x})\right) = \text{tr}(\boldsymbol{W}\boldsymbol{V}),$$

where $\boldsymbol{V}$ is the dispersion matrix for the distribution $F(\boldsymbol{x})$.

Asymptotic distributional bias of an estimator $\boldsymbol{\beta}_1^*$ is defined as

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

First, let $H_v(x,\Delta)$ be the cumulative distribution function of the non-central chi-squared distribution with non-centrality parameter Δ and v degree of freedom, then

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

Assumption 1. We make the following two regularity conditions

(i) $\frac{1}{n}\max_{1\le i\le n}\boldsymbol{x}_i^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{x}_i \to 0$ as $n\to\infty$, where $\boldsymbol{x}_i^\top$ is the ith row of $\boldsymbol{X}$,

(ii) $\boldsymbol{Q}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}^\top\boldsymbol{X} \to \boldsymbol{Q}$.

We assume that the sub-matrices of $\boldsymbol{Q}$ satisfy the following equations

$$\lim_{n\to\infty}\frac{1}{n}\left(\boldsymbol{X}_1^\top\boldsymbol{X}_1 + \lambda^R\boldsymbol{I}_{p_1}\right) = \lim_{n\to\infty}\frac{1}{n}\boldsymbol{X}_1^\top\boldsymbol{X}_1 = \boldsymbol{Q}_{11}, \tag{2.7}$$

$$\lim_{n\to\infty}\frac{1}{n}\boldsymbol{X}_1^\top\boldsymbol{X}_2 = \boldsymbol{Q}_{12}, \tag{2.8}$$

$$\lim_{n\to\infty}\frac{1}{n}\boldsymbol{X}_2^\top\boldsymbol{X}_1 = \boldsymbol{Q}_{21}, \tag{2.9}$$

$$\lim_{n\to\infty}\frac{1}{n}\left(\boldsymbol{X}_2^\top\boldsymbol{X}_2 + \lambda^R\boldsymbol{I}_{p_2}\right) = \lim_{n\to\infty}\frac{1}{n}\boldsymbol{X}_2^\top\boldsymbol{X}_2 = \boldsymbol{Q}_{22}, \tag{2.10}$$

where $\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{pmatrix}$.

Theorem 2. If $\lambda^R/\sqrt{n} \to \lambda_0 \ge 0$ and $\boldsymbol{Q}$ is non-singular, then

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{RFM} - \boldsymbol{\beta}\right) \xrightarrow{\mathscr{D}} \left(-\lambda_0\boldsymbol{Q}^{-1}\boldsymbol{\beta}, \sigma^2\boldsymbol{Q}^{-1}\right).$$

Proof. (Knight and Fu, 2000, see). □

Theorem 3. Let $\boldsymbol{X}$ be $q-$dimensional normal vector distributed as $N_q\left(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_q\right)$, then, for a measurable function of of φ, we have

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

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

Proof. It can be found in Judge and Bock (1978) □

Theorem 4. Let $\boldsymbol{X} = \left(\boldsymbol{X}_1^\top \vdots \boldsymbol{X}_2^\top\right)^\top$ be distributed $N_p\left(\boldsymbol{\eta}, \boldsymbol{\Sigma}\right)$ with

$$\boldsymbol{\eta} = \begin{pmatrix}\boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2\end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Then the conditional distribution of $\boldsymbol{X}_1$, given $\boldsymbol{X}_2 = \boldsymbol{x}_2$, is normal with

$$\boldsymbol{\eta}_{11.2} = \boldsymbol{\eta}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{x}_2 - \boldsymbol{\eta}_2\right) \text{ and } \boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

Proof. Johnson and Wichern (see 2001). □

Now, under the assumed regularity conditions, theorems and equations above, and under $\{K_n\}$, the $ADBs$ of the estimators are given as follows:

Theorem 5.

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

where $\Delta = \left(\boldsymbol{w}^{\intercal}\boldsymbol{Q}_{22.1}^{-1}\boldsymbol{w}\right)\sigma^{-2}$, $\boldsymbol{Q}_{22.1} = \boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}$, $\boldsymbol{\mu} = -\lambda_0\boldsymbol{Q}^{-1}\boldsymbol{\beta} = \begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}$, $\boldsymbol{\mu}_{11.2} = \boldsymbol{\mu}_1 - \boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\left((\boldsymbol{\beta}_2 - \boldsymbol{w}) - \boldsymbol{\mu}_2\right)$, $\boldsymbol{\gamma} = \boldsymbol{\mu}_{11.2} + \boldsymbol{\delta}$ and $\boldsymbol{\delta} = \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\omega}$.

Proof. See Appendix A. □

The bias expressions for all the estimators are not in scaler form, we also convert them to quadratic forms. Thus, we define $AQDB$ of an estimator $\boldsymbol{\beta}_1^*$ is defined as

$$AQDB\left(\boldsymbol{\beta}_1^*\right) = \left(ADB\left(\boldsymbol{\beta}_1^*\right)\right)^{\top}\boldsymbol{Q}_{11.2}\left(ADB\left(\boldsymbol{\beta}_1^*\right)\right), \tag{2.11}$$

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

Considering equation (2.11), we present the $AQDB$ of the estimators, under $\{K_n\}$, in the following theorem.

Theorem 6.

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

Proof. See Appendix A. □

In the results of the above Theorem, for $\boldsymbol{Q}_{12}=\mathbf{0}$, because of $\boldsymbol{\delta}=\mathbf{0}$, $\boldsymbol{\gamma}=\boldsymbol{\mu}_{11.2}$ and $\boldsymbol{Q}_{11.2}=\boldsymbol{Q}_{11}$, all the $AQDBs$ reduce to common value $\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{Q}_{11}\boldsymbol{\mu}_{11.2}$ for all $\boldsymbol{\omega}$. Following this, If we suppose that $\boldsymbol{Q}_{12}\neq\mathbf{0}$, it is given following results

(i) The $AQDB$ of $\widehat{\boldsymbol{\beta}}_1^{RFM}$ is an unbounded function of $\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{Q}_{11.2}\boldsymbol{\mu}_{11.2}$.

(ii) The $AQDB$ of $\widehat{\boldsymbol{\beta}}_1^{RSM}$ is an unbounded function of $\boldsymbol{\gamma}^{\top}\boldsymbol{Q}_{11.2}\boldsymbol{\gamma}$.

(iii) The $AQDB$ of $\widehat{\boldsymbol{\beta}}_1^{RPT}$ starts from $\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{Q}_{11.2}\boldsymbol{\mu}_{11.2}$ at $\Delta=0$. For $\Delta>0$, it goes up to a point, and so goes down towards zero.

(iv) Similarly, the $AQDB$ of $\widehat{\boldsymbol{\beta}}_1^{RS}$ starts from $\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{Q}_{11.2}\boldsymbol{\mu}_{11.2}$ at $\Delta=0$, and it goes up to a point then goes down towards zero for non-zero Δ values because of $E\left(\chi_{p_2+2}^{-2}\left(\Delta\right)\right)$ being a non increasing log convex function of Δ. Lastly, for all Δ values, the demeanour of the $AQDB$ of $\widehat{\boldsymbol{\beta}}_1^{RPS}$ is nearly the same $\widehat{\boldsymbol{\beta}}_1^{RS}$, but the quadratic bias curve of $\widehat{\boldsymbol{\beta}}_1^{RPS}$ stays on below the curve of $\widehat{\boldsymbol{\beta}}_1^{RS}$.

Under a sequence $\{K_n\}$, the asymptotic covariance matrices of the estimators are given by following theorem.

Theorem 7.

$$
\begin{aligned}
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{RFM}\right) &= \sigma^2\boldsymbol{Q}_{11.2}^{-1}+\boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^{\top},\\
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{RSM}\right) &= \sigma^2\boldsymbol{Q}_{11}^{-1}+\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top},\\
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{RPT}\right) &= \sigma^2\boldsymbol{Q}_{11.2}^{-1}+\boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^{\top}+2\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{\delta}H_{p_2+2}\left(\chi_{p_2,\alpha}^2;\Delta\right)\\
&\quad+\sigma^2\boldsymbol{Q}_{12}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}H_{p_2+2}\left(\chi_{p_2,\alpha}^2;\Delta\right)\\
&\quad+\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\left[2H_{p_2+2}\left(\chi_{p_2,\alpha}^2;\Delta\right)-H_{p_2+4}\left(\chi_{p_2,\alpha}^2;\Delta\right)\right],\\
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{RS}\right) &= \sigma^2\boldsymbol{Q}_{11.2}^{-1}+\boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^{\top}+2(p_2-2)\boldsymbol{\delta}\boldsymbol{\mu}_{11.2}^{\top}E\left(\chi_{p_2+2}^{-2}\left(\Delta\right)\right)\\
&\quad-(p_2-2)\sigma^2\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{Q}_{22.1}^{-1}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\{2E\left(\chi_{p_2+2}^{-2}\left(\Delta\right)\right)\\
&\quad-(p_2-2)E\left(\chi_{p_2+2}^{-4}\left(\Delta\right)\right)\}\\
&\quad+(p_2-2)\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\{2E\left(\chi_{p_2+2}^{-2}\left(\Delta\right)\right)\\
&\quad-2E\left(\chi_{p_2+4}^{-2}\left(\Delta\right)\right)-(p_2-2)E\left(\chi_{p_2+4}^{-4}\left(\Delta\right)\right)\},
\end{aligned}
$$

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

Proof. See Appendix A. □

Finally, we obtain the ADRs of the estimators under $\{K_n\}$ as follows:

Theorem 8.

$$
\begin{aligned}
R\left(\widehat{\boldsymbol{\beta}}_1^{RFM}\right) &= \sigma^2 tr\left(\boldsymbol{W}\boldsymbol{Q}_{11.2}^{-1}\right)+\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\mu}_{11.2}, \\
R\left(\widehat{\boldsymbol{\beta}}_1^{RSM}\right) &= \sigma^2\text{tr}\left(\boldsymbol{W}\boldsymbol{Q}_{11}^{-1}\right)+\boldsymbol{\gamma}^{\top}\boldsymbol{W}\boldsymbol{\gamma}, \\
R\left(\widehat{\boldsymbol{\beta}}_1^{RPT}\right) &= R\left(\widehat{\boldsymbol{\beta}}_1^{RFM}\right)-2\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\delta}H_{p_2+2}\left(\chi_{p_2,\alpha}^{2};\Delta\right) \\
&\quad -\sigma^2\text{tr}\left(\boldsymbol{W}\boldsymbol{Q}_{12}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\right)H_{p_2+2}\left(\chi_{p_2,\alpha}^{2};\Delta\right) \\
&\quad +\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\left\{2H_{p_2+2}\left(\chi_{p_2,\alpha}^{2};\Delta\right)-H_{p_2+4}\left(\chi_{p_2,\alpha}^{2};\Delta\right)\right\}, \\
R\left(\widehat{\boldsymbol{\beta}}_1^{RS}\right) &= R\left(\widehat{\boldsymbol{\beta}}_1^{RFM}\right)+2(p_2-2)\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\delta}E\left(\chi_{p_2+2}^{-2}(\Delta)\right) \\
&\quad -(p_2-2)\sigma^2\text{tr}\left(\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{W}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{Q}_{22.1}^{-1}\right)\left\{2E\left(\chi_{p_2+2}^{-2}(\Delta)\right)\right. \\
&\quad \left.-(p_2-2)E\left(\chi_{p_2+2}^{-4}(\Delta)\right)\right\} \\
&\quad +(p_2-2)\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\{2E\left(\chi_{p_2+2}^{-2}(\Delta)\right) \\
&\quad -2E\left(\chi_{p_2+4}^{-2}(\Delta)\right)-(p_2-2)E\left(\chi_{p_2+4}^{-4}(\Delta)\right)\},
\end{aligned}
$$

$$
\begin{aligned}
R\left(\widehat{\boldsymbol{\beta}}_1^{RPS}\right) &= R\left(\widehat{\boldsymbol{\beta}}_1^{RS}\right) \\
&\quad -2\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\delta}E\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\}I\left(\chi_{p_2+2}^{2}(\Delta)\leq p_2-2\right)\right) \\
&\quad +(p_2-2)\sigma^2\text{tr}\left(\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{W}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{Q}_{22.1}^{-1}\right) \\
&\quad \cdot\left[2E\left(\chi_{p_2+2}^{-2}(\Delta)I\left(\chi_{p_2+2}^{2}(\Delta)\leq p_2-2\right)\right)\right. \\
&\quad \left.-(p_2-2)E\left(\chi_{p_2+2}^{-4}(\Delta)I\left(\chi_{p_2+2}^{2}(\Delta)\leq p_2-2\right)\right)\right] \\
&\quad -\sigma^2\text{tr}\left(\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{W}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{Q}_{22.1}^{-1}\right)H_{p_2+2}(p_2-2;\Delta) \\
&\quad +\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\left[2H_{p_2+2}(p_2-2;\Delta)-H_{p_2+4}(p_2-2;\Delta)\right] \\
&\quad -(p_2-2)\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\left[2E\left(\chi_{p_2+2}^{-2}(\Delta)I\left(\chi_{p_2+2}^{2}(\Delta)\leq p_2-2\right)\right)\right. \\
&\quad -2E\left(\chi_{p_2+4}^{-2}(\Delta)I\left(\chi_{p_2+4}^{2}(\Delta)\leq p_2-2\right)\right) \\
&\quad \left.+(p_2-2)E\left(\chi_{p_2+2}^{-4}(\Delta)I\left(\chi_{p_2+2}^{2}(\Delta)\leq p_2-2\right)\right)\right].
\end{aligned}
$$

Proof. See Appendix A. □

In the results of the above Theorem, if $\boldsymbol{Q}_{12}=\mathbf{0}$, then $\boldsymbol{\delta}=\mathbf{0}$, $\boldsymbol{\gamma}=\boldsymbol{\mu}_{11.2}$ and $\boldsymbol{Q}_{11.2}=\boldsymbol{Q}_{11}$, all the $ADRs$ reduce to common value $\sigma^2\text{tr}\left(\boldsymbol{W}\boldsymbol{Q}_{11}^{-1}\right)+\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\mu}_{11.2}$ for all $\boldsymbol{\omega}$. On the other hand, if we suppose that $\boldsymbol{Q}_{12}\neq\mathbf{0}$, then it is given following results:

(i) When Δ moves away from 0, the risk of $R\left(\widehat{\boldsymbol{\beta}}_1^{RSM}\right)$ becomes unbounded. Moreover, the risk of $\widehat{\boldsymbol{\beta}}_1^{RPT}$ is asymptotically outstanding $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for all values of $\Delta\geq 0$, that is $R\left(\widehat{\boldsymbol{\beta}}_1^{RPT}\right)\leq R\left(\widehat{\boldsymbol{\beta}}_1^{RFM}\right)$.

(ii)Under $\{K_n\}$, for all $\boldsymbol{W}$ and $\boldsymbol{\omega}$, it can be shown by using the following equation that $R\left(\widehat{\boldsymbol{\beta}}_1^{RS}\right)\leq R\left(\widehat{\boldsymbol{\beta}}_1^{RFM}\right)$.

$$
\frac{\text{tr}\left(\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{W}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{Q}_{22.1}^{-1}\right)}{ch_{max}\left(\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{W}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{Q}_{22.1}^{-1}\right)}\geq\frac{p_2+2}{2}
$$

where $ch_{max}(.)$ is the maximum characteristic root.

(iii) For comparing $\widehat{\boldsymbol{\beta}}_1^{RPS}$ and $\widehat{\boldsymbol{\beta}}_1^{RS}$, we observe that the the risk of $\widehat{\boldsymbol{\beta}}_1^{RPS}$ dominates $\widehat{\boldsymbol{\beta}}_1^{RS}$ for all the values of $\boldsymbol{\omega}$. Moreover, with result (ii), we have $R\left(\widehat{\boldsymbol{\beta}}_1^{RPS}\right)\leq R\left(\widehat{\boldsymbol{\beta}}_1^{RS}\right)\leq R\left(\widehat{\boldsymbol{\beta}}_1^{RFM}\right)$.

To illustrate the properties of the theoretical results, we carried on a simulation study and reported in the next subsection, to compare the performance of the suggested estimators.

2.6 Simulation Studies

We perform Monte Carlo simulation experiments to examine the risk performance of the suggested estimators. We simulate the response from the following model:

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

where $\boldsymbol{x}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_x)$ and ε_i are i.i.d. $N(0, 1)$. We also define $\boldsymbol{\Sigma}_x$ that is positive definite covariance matrix. To generate $\boldsymbol{\Sigma}_x$, we define ρ which is coefficient of correlation $\boldsymbol{X}$ and use three values of correlation which are low($\rho = 0.25$), medium($\rho = 0.5$) and high($\rho = 0.75$). The condition number test (CNT) is performed to test the multicollinearity, which is defined as the ratio of the largest eigenvalue to the smallest eigenvalue of matrix $\boldsymbol{C}$. If CNT is larger than 30, then it implies the existence of multicollinearity in the data set.

For $H_0 : \beta_j = 0$, $j = p_1 + 1, p_1 + 2, ..., p$, with $p = p_1 + p_2$, so that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (\boldsymbol{\beta}_1, \mathbf{0})$ with $\boldsymbol{\beta}_1 = (1, 1, 1, 1, 1)$. We use $\alpha = 0.05$.

Based on 3000 realizations, we calculated bias of the estimators. We define $\Delta^* = \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|$, where $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_1, \mathbf{0})$, and $\|\cdot\|$ is the Euclidean norm. Further, to investigate the performance of the estimators for $\Delta^* > 0$, more datasets were generated from those distributions.

The performance of an estimator $\boldsymbol{\beta}_1$ was evaluated by calculating its mean squared error (MSE) criterion. We numerically calculated the RMSE of $\widehat{\boldsymbol{\beta}}_1^{RSM}$, $\widehat{\boldsymbol{\beta}}_1^{RPT}$, $\widehat{\boldsymbol{\beta}}_1^{RS}$ and $\widehat{\boldsymbol{\beta}}_1^{RPS}$, with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$. The relative mean square efficiency (RMSE) of the $\boldsymbol{\beta}_1^{\blacktriangle}$ to the unrestricted least squares estimator $\widehat{\boldsymbol{\beta}}_1^{RFM}$ is indicated by

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

where $\boldsymbol{\beta}_1^{\blacktriangle}$ is one of the listed estimators. The amount by which a RMSE is larger than one indicates the degree of superiority of the estimator $\boldsymbol{\beta}_1^{\blacktriangle}$ over $\widehat{\boldsymbol{\beta}}_1^{RFM}$. All computations were conducted using the R statistical system (R Development Core Team , 2010).

We report the findings which are summarized as follows, (see Tables 2.1–2.7 and Figures 2.1 – 2.3).

Table 2.1: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$, $n = 30, 50$ and $p_2 = 5, 10, 20$ values when $\Delta^* = 0$.

(n, p_2)	ρ	CNT	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPT}$	$\widehat{\boldsymbol{\beta}}_1^{RS}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$
(30,5)	0.25	16.88	0.79	2.59	2.13	1.41	1.85
	0.5	46.00	0.72	1.58	1.51	1.27	1.38
	0.75	132.29	0.53	1.89	1.73	1.25	1.50
(30,10)	0.25	54.99	0.59	4.57	3.07	2.30	3.22
	0.5	142.16	0.70	2.21	1.94	1.32	1.90
	0.75	413.56	0.49	2.34	2.03	1.55	1.99
(30,20)	0.25	1195.88	0.20	8.50	4.61	4.16	5.76
	0.5	3278.64	0.25	3.26	2.66	2.07	2.85
	0.75	9467.82	0.15	3.40	2.59	2.22	2.99
(50,5)	0.25	11.01	0.92	2.36	2.12	1.38	1.70
	0.5	28.84	0.85	1.42	1.36	1.17	1.27
	0.75	82.28	0.71	1.85	1.68	1.27	1.49
(50,10)	0.25	24.10	0.82	4.22	3.10	2.30	3.15
	0.5	62.22	0.90	1.81	1.66	1.43	1.63
	0.75	183.81	0.75	2.33	2.06	1.51	1.97
(50,20)	0.25	91.49	0.63	6.53	4.59	3.51	4.79
	0.5	245.86	0.89	2.56	2.33	1.94	2.38
	0.75	723.15	0.73	3.14	2.70	1.99	2.84

For convenience, in Table 2.1, we also present the reports to see how the coefficient of ρ effects the performance of estimators as $\Delta^* = 0$.

As it can be seen that in Table 2.1, not surprisingly, when the ρ is large, the CNT is large. This is because of the magnitude of multicollinearity. When it comes to the performance of the ridge regression estimators as there exists the multicollinearity, it can be safely concluded that the performance of ridge regression estimators outshines the performance ordinary least square estimator (OLS). In Table 2.1, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{FM}$ is smaller than one for all simulation results. It means that the performance of the OLS is not well compared to the ridge regression estimator. We also indicate that the CNT increases when the number of p_2 increase for fix p_1 and n values.

Table 2.2: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$, $n = 30$, $p_2 = 5, 10, 20$ values when $\rho = 0.25$.

(n, p_2)	CNT	Δ^*	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPT}$	$\widehat{\boldsymbol{\beta}}_1^{RS}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$
	16.88	0.00	0.79	2.59	2.13	1.41	1.85
	17.53	0.25	0.80	2.07	1.47	1.32	1.54
	17.95	0.50	0.78	1.42	0.98	1.19	1.29
	17.78	0.75	0.79	0.93	0.85	1.16	1.16
(30,5)	18.66	1.00	0.81	0.59	0.95	1.08	1.08
	17.75	1.25	0.79	0.43	0.97	1.04	1.04
	18.57	1.50	0.83	0.33	1.00	1.02	1.02
	17.86	2.00	0.86	0.22	1.00	1.01	1.01
	18.83	4.00	0.96	0.06	1.00	0.99	0.99
	54.99	0.00	0.59	4.57	3.07	2.30	3.22
	54.83	0.25	0.59	2.81	2.34	1.81	2.29
	54.62	0.50	0.62	2.55	1.41	1.69	2.05
	53.47	0.75	0.65	1.63	0.97	1.53	1.55
(30,10)	56.59	1.00	0.67	0.97	0.91	1.30	1.30
	54.36	1.25	0.68	0.74	0.96	1.21	1.21
	50.85	1.50	0.71	0.58	1.00	1.15	1.15
	58.59	2.00	0.73	0.36	1.00	1.06	1.06
	55.04	4.00	0.91	0.13	1.00	1.01	1.01
	1195.88	0.00	0.20	8.50	4.61	4.16	5.76
	1346.02	0.25	0.16	7.11	4.16	4.01	5.52
	935.51	0.50	0.21	4.41	2.00	3.05	3.41
	1041.12	0.75	0.21	3.49	1.63	2.77	2.91
(30,20)	1219.49	1.00	0.15	2.11	1.00	1.97	1.99
	1130.68	1.25	0.19	1.66	0.99	1.73	1.73
	1364.32	1.50	0.27	1.85	1.00	1.45	1.45
	1221.34	2.00	0.23	0.89	1.00	1.25	1.25
	1159.07	4.00	0.35	0.38	1.00	1.06	1.06

The findings in the Tables 2.2–2.7 may be summarized as follows.

(i) In situation $\Delta^* = 0$, RSM outperforms all the other estimators, because the RMSE of $\widehat{\boldsymbol{\beta}}_1^{RSM}$ is the biggest. In contrast to this, after the small interval near $\Delta^* = 0$, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{RSM}$ decreases and goes to one.

(ii) For large p_2 values in situation fix p_1 and n values, as the CNT increase, whereas the RMSE of $\widehat{\boldsymbol{\beta}}_1^{FM}$ decrease, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{RSM}$ increase.

Table 2.3: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$, $n = 50$, $p_2 = 5, 10, 20$ values when $\rho = 0.25$.

(n, p_2)	CNT	Δ^*	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPT}$	$\widehat{\boldsymbol{\beta}}_1^{RS}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$
	11.01	0.00	0.92	2.36	2.12	1.38	1.70
	11.85	0.25	0.90	1.81	1.36	1.32	1.45
	11.24	0.50	0.88	1.10	0.85	1.18	1.18
	12.28	0.75	0.90	0.59	0.94	1.08	1.08
(50,5)	11.45	1.00	0.91	0.39	1.00	1.04	1.04
	11.26	1.25	0.89	0.24	1.00	1.01	1.01
	11.60	1.50	0.90	0.19	1.00	1.01	1.01
	11.88	2.00	0.94	0.11	1.00	1.00	1.00
	11.65	4.00	0.98	0.03	1.00	1.00	1.00
	24.10	0.00	0.82	4.22	3.10	2.30	3.15
	23.66	0.25	0.84	3.04	1.79	2.09	2.29
	22.76	0.50	0.85	1.92	1.05	1.49	1.65
	22.76	0.75	0.83	0.99	0.89	1.34	1.34
(50,10)	23.55	1.00	0.86	0.65	1.00	1.19	1.19
	23.88	1.25	0.83	0.40	1.00	1.11	1.11
	24.02	1.50	0.83	0.31	1.00	1.07	1.07
	23.87	2.00	0.88	0.20	1.00	1.03	1.03
	25.01	4.00	0.95	0.06	1.00	1.01	1.01
	91.49	0.00	0.63	6.53	4.59	3.51	4.79
	86.69	0.25	0.65	5.45	3.07	3.10	4.05
	87.81	0.50	0.63	3.29	1.40	2.62	2.71
	87.54	0.75	0.67	1.81	0.98	1.95	1.95
(50,20)	89.98	1.00	0.64	1.23	0.99	1.55	1.55
	90.98	1.25	0.69	0.91	1.00	1.35	1.35
	90.93	1.50	0.70	0.62	1.00	1.22	1.22
	89.79	2.00	0.73	0.41	1.00	1.11	1.11
	89.86	4.00	0.89	0.13	1.00	1.02	1.02

(iii) The RPT outperforms shrinkage ridge regression estimators in case $\Delta^* = 0$ and values which are p_1 and p_2 close to each other. However, for large p_2 values in situation fix p_1 and n values, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{RPT}$ is larger than the RMSE of $\widehat{\boldsymbol{\beta}}_1^{RS}$ and is smaller than the RMSE of the RMSE of $\widehat{\boldsymbol{\beta}}_1^{RPS}$. As Δ^* is larger than zero, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{RPT}$ decreases, and it remains below 1 when Δ^* change around 0.5 to 1, after that it increases and approaches one as Δ^* is larger than 1.

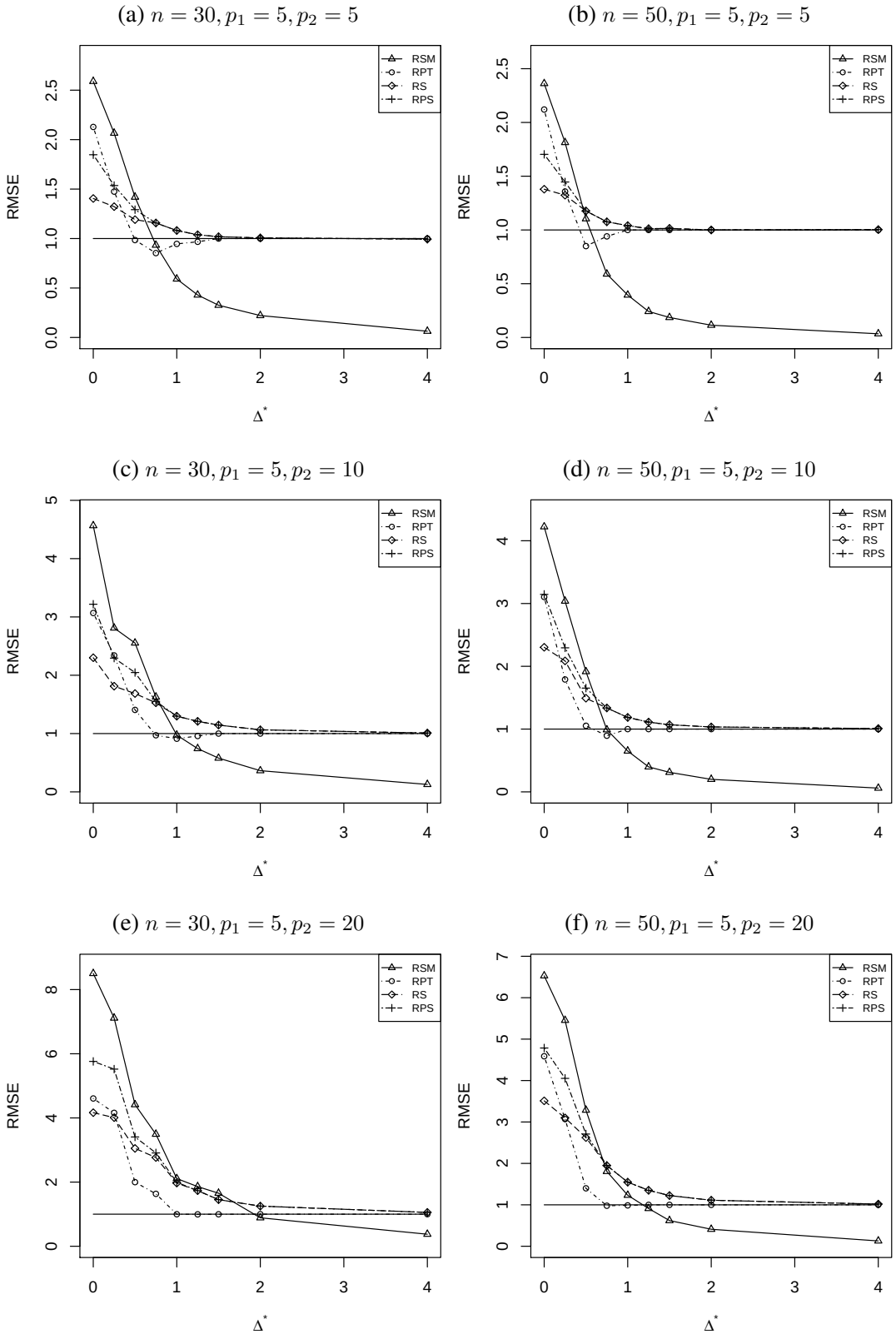


Figure 2.1: Relative efficiency of the estimators as a function of the non-centrality parameter Δ^* for various sample size $n = 30$ and 50, and $p_2 = 5, 10$ and 20 when $\rho = 0.25$.

(iv) Comparing the RS and the RPS shows that the RMSE of $\widehat{\boldsymbol{\beta}}_1^{RPS}$ is larger

than the RMSE of $\widehat{\boldsymbol{\beta}}_1^{RS}$, which implies the better performance of the RPS over the PS.

In Figure 2.1, we plot the values which are shown in Table 2.2 and Table 2.3. In generally, we observe that the RMSE curve of all estimators decreases very quickly as Δ^* changes from 0 to 1. When Δ^* is around and close to 1, the RMSE curve of $\widehat{\boldsymbol{\beta}}_1^{RSM}$ remain below the line one. After that, it goes to zero. For the RMSE curve of $\widehat{\boldsymbol{\beta}}_1^{RPT}$, it goes down the line one as Δ^* values are around one, then it goes up and approaches one gradually. The RMSE curves of $\widehat{\boldsymbol{\beta}}_1^{RS}$ and $\widehat{\boldsymbol{\beta}}_1^{RPS}$ decrease step by step and approach one step by step.

Table 2.4: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$, $n = 30$ and $p_2 = 5, 10, 20$ values when $\rho = 0.5$.

(n, p_2)	CNT	Δ^*	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPT}$	$\widehat{\boldsymbol{\beta}}_1^{RS}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$
	46.00	0.00	0.72	1.58	1.51	1.27	1.38
	45.70	0.25	0.70	1.49	1.35	1.22	1.31
	45.38	0.50	0.68	1.30	1.05	1.12	1.19
	45.91	0.75	0.66	1.08	0.88	1.07	1.12
(30,5)	44.95	1.00	0.68	0.93	0.88	1.10	1.10
	47.16	1.25	0.66	0.77	0.94	1.07	1.07
	44.98	1.50	0.67	0.59	0.96	1.04	1.04
	45.70	2.00	0.70	0.43	1.00	1.02	1.02
	45.53	4.00	0.86	0.20	1.00	1.01	1.01
	142.16	0.00	0.70	2.21	1.94	1.32	1.90
	139.70	0.25	0.67	1.99	1.63	1.34	1.74
	140.85	0.50	0.66	1.79	1.34	1.33	1.62
	145.26	0.75	0.62	1.43	0.98	1.37	1.43
(30,10)	142.79	1.00	0.60	1.14	0.88	1.30	1.32
	141.20	1.25	0.63	0.97	0.90	1.26	1.27
	143.29	1.50	0.62	0.82	0.96	1.21	1.21
	142.76	2.00	0.64	0.56	1.00	1.12	1.12
	140.81	4.00	0.76	0.23	1.00	1.02	1.02
	3278.64	0.00	0.25	3.26	2.66	2.07	2.85
	3254.63	0.25	0.22	3.12	2.49	2.11	2.79
	3177.72	0.50	0.21	2.74	1.80	2.10	2.54
	3227.72	0.75	0.20	2.34	1.27	2.03	2.29
(30,20)	3227.38	1.00	0.22	1.85	0.97	1.96	2.00
	3132.70	1.25	0.22	1.63	0.92	1.79	1.82
	3062.46	1.50	0.21	1.29	0.94	1.65	1.65
	3161.33	2.00	0.24	0.95	0.99	1.45	1.45
	3318.83	4.00	0.29	0.37	1.00	1.08	1.08

In Table 2.4 and Table 2.5, we present simulated relative efficiency with respect

to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$, $n = 30, 50$ and $p_2 = 5, 10, 20$ values when the coefficient of correlation level is medium. As it can be understood that from Tables the estimators exhibit similar behaviour in the previous Tables 2.2 – 2.3.

Table 2.5: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$, $n = 50$ and $p_2 = 5, 10, 20$ values when $\rho = 0.5$.

(n, p_2)	CNT	Δ^*	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPT}$	$\widehat{\boldsymbol{\beta}}_1^{RS}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$
	28.84	0.00	0.85	1.42	1.36	1.17	1.27
	28.62	0.25	0.84	1.37	1.21	1.12	1.22
	28.66	0.50	0.83	1.17	0.96	1.11	1.12
	28.70	0.75	0.82	0.94	0.92	1.08	1.08
(50,5)	28.98	1.00	0.81	0.75	0.98	1.05	1.05
	28.38	1.25	0.81	0.59	0.99	1.04	1.04
	28.55	1.50	0.82	0.48	1.00	1.03	1.03
	28.67	2.00	0.83	0.33	1.00	1.01	1.01
	28.47	4.00	0.93	0.12	1.00	1.00	1.00
	62.22	0.00	0.90	1.81	1.66	1.43	1.63
	62.52	0.25	0.88	1.68	1.43	1.34	1.53
	63.27	0.50	0.86	1.42	1.06	1.32	1.38
	64.50	0.75	0.86	1.15	0.91	1.28	1.28
(50,10)	63.27	1.00	0.86	0.93	0.96	1.23	1.23
	62.70	1.25	0.82	0.70	1.00	1.15	1.15
	63.08	1.50	0.81	0.57	1.00	1.10	1.10
	62.62	2.00	0.83	0.38	1.00	1.06	1.06
	63.49	4.00	0.90	0.14	1.00	1.01	1.01
	245.86	0.00	0.89	2.56	2.33	1.94	2.38
	246.16	0.25	0.90	2.51	1.96	1.86	2.30
	245.47	0.50	0.88	2.18	1.35	1.88	2.10
	247.84	0.75	0.81	1.64	0.96	1.78	1.80
(50,20)	245.06	1.00	0.79	1.27	0.92	1.62	1.62
	248.31	1.25	0.79	1.05	0.97	1.50	1.50
	245.21	1.50	0.76	0.80	0.99	1.38	1.38
	247.05	2.00	0.74	0.52	1.00	1.21	1.21
	246.46	4.00	0.80	0.18	1.00	1.04	1.04

In Figure 2.2, we plot the values which are shown in Table 2.4 and Table 2.5. The difference between Figure 2.1 and Figure 2.2 is just size of multicollinearity. It can be seen that from Figure 2.2 the results have smaller the RMSE values of estimators compare to in Figure 2.1. This is because the $\rho = 0.5$ changes .25 to .5, that is the level of multicollinearity increases.

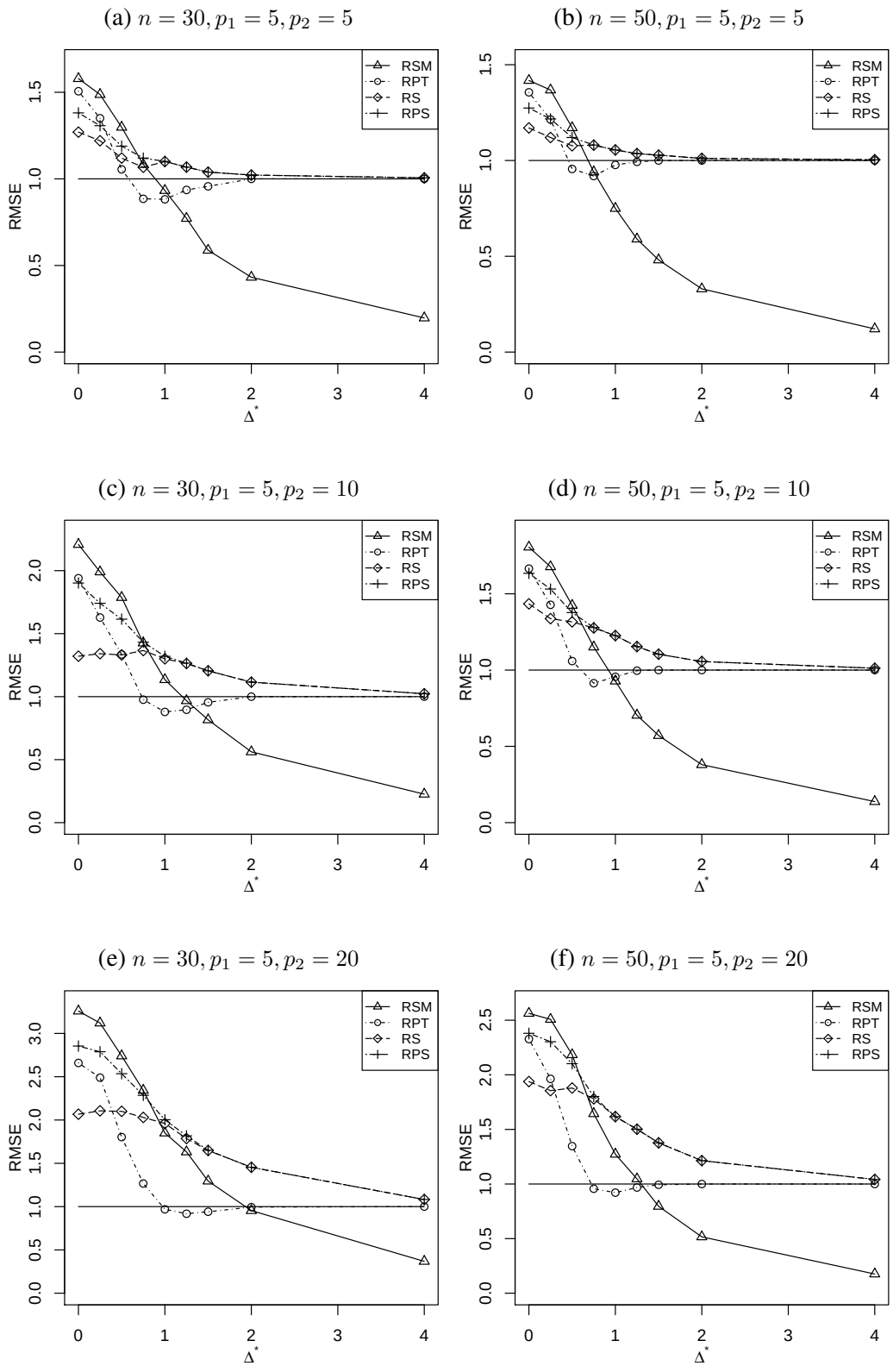


Figure 2.2: Relative efficiency of the estimators as a function of the non-centrality parameter Δ^* for various sample size $n = 30$ and 50, and $p_2 = 5, 10$ and 20 when $\rho = 0.5$.

In the following Table 2.6 and Table 2.7, we present simulated relative efficiency

with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$, $n = 30, 50$ and $p_2 = 5, 10, 20$ values when the coefficient of correlation level is high. As it can be seen that from Tables the estimators exhibit similar behaviour in the previous Tables 2.2 – 2.3.

Table 2.6: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$, $n = 30$, $p_2 = 5, 10, 20$ values when $\rho = 0.75$.

(n, p_2)	CNT	Δ^*	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPT}$	$\widehat{\boldsymbol{\beta}}_1^{RS}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$
	132.29	0.00	0.53	1.89	1.73	1.25	1.50
	130.92	0.25	0.51	1.67	1.46	1.17	1.40
	129.56	0.50	0.51	1.51	1.24	1.11	1.30
	126.95	0.75	0.49	1.33	0.99	1.06	1.22
(30,5)	132.68	1.00	0.46	1.08	0.85	1.03	1.14
	132.76	1.25	0.48	0.94	0.83	1.12	1.13
	130.28	1.50	0.50	0.77	0.83	1.06	1.07
	127.97	2.00	0.52	0.60	0.95	1.04	1.04
	130.36	4.00	0.73	0.29	1.00	1.01	1.01
	413.56	0.00	0.49	2.34	2.03	1.55	1.99
	417.52	0.25	0.47	2.37	1.96	1.42	2.02
	419.77	0.50	0.46	2.16	1.59	1.35	1.86
	418.01	0.75	0.48	1.84	1.29	1.33	1.69
(30,10)	421.42	1.00	0.46	1.48	0.94	1.40	1.49
	408.70	1.25	0.45	1.22	0.82	1.34	1.39
	418.11	1.50	0.44	1.04	0.81	1.30	1.31
	413.94	2.00	0.45	0.79	0.92	1.24	1.24
	409.90	4.00	0.62	0.34	1.00	1.05	1.05
	9467.82	0.00	0.15	3.40	2.59	2.22	2.99
	9193.34	0.25	0.16	3.34	2.49	2.07	2.92
	8769.65	0.50	0.16	3.12	2.22	2.03	2.84
	9489.95	0.75	0.15	2.57	1.55	1.94	2.46
(30,20)	9517.39	1.00	0.17	2.34	1.18	1.94	2.33
	8769.05	1.25	0.15	1.84	0.93	1.84	2.02
	9042.83	1.50	0.17	1.73	0.86	1.94	1.99
	9486.12	2.00	0.16	1.10	0.88	1.66	1.66
	9143.66	4.00	0.19	0.46	1.00	1.17	1.17

Table 2.7: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$, $n = 50$, $p_2 = 5, 10, 20$ values when $\rho = 0.75$.

(n, p_2)	CNT	Δ^*	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPT}$	$\widehat{\boldsymbol{\beta}}_1^{RS}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$
	82.28	0.00	0.71	1.85	1.68	1.27	1.49
	80.50	0.25	0.70	1.77	1.53	1.20	1.44
	81.71	0.50	0.68	1.49	1.14	1.18	1.27
	81.29	0.75	0.67	1.27	0.94	1.16	1.20
(50,5)	81.27	1.00	0.64	0.97	0.87	1.11	1.11
	80.85	1.25	0.66	0.83	0.93	1.08	1.08
	80.85	1.50	0.67	0.67	0.99	1.06	1.06
	81.72	2.00	0.69	0.51	1.00	1.03	1.03
	80.93	4.00	0.87	0.20	1.00	1.00	1.00
	183.81	0.00	0.75	2.33	2.06	1.51	1.97
	184.57	0.25	0.74	2.17	1.77	1.31	1.87
	187.07	0.50	0.72	1.88	1.37	1.40	1.67
	180.56	0.75	0.72	1.58	1.03	1.37	1.52
(50,10)	182.67	1.00	0.69	1.28	0.88	1.37	1.40
	183.54	1.25	0.70	1.03	0.90	1.31	1.32
	180.83	1.50	0.68	0.82	0.96	1.24	1.24
	183.44	2.00	0.67	0.57	1.00	1.14	1.14
	183.47	4.00	0.81	0.22	1.00	1.03	1.03
	723.15	0.00	0.73	3.14	2.70	1.99	2.84
	725.40	0.25	0.72	3.08	2.52	2.04	2.81
	730.61	0.50	0.72	2.78	1.87	2.05	2.62
	730.78	0.75	0.70	2.15	1.24	1.93	2.20
(50,20)	748.46	1.00	0.67	1.80	0.92	1.96	2.04
	722.69	1.25	0.64	1.40	0.87	1.80	1.82
	727.83	1.50	0.63	1.17	0.92	1.71	1.71
	721.98	2.00	0.61	0.73	0.98	1.42	1.42
	729.73	4.00	0.67	0.27	1.00	1.10	1.10

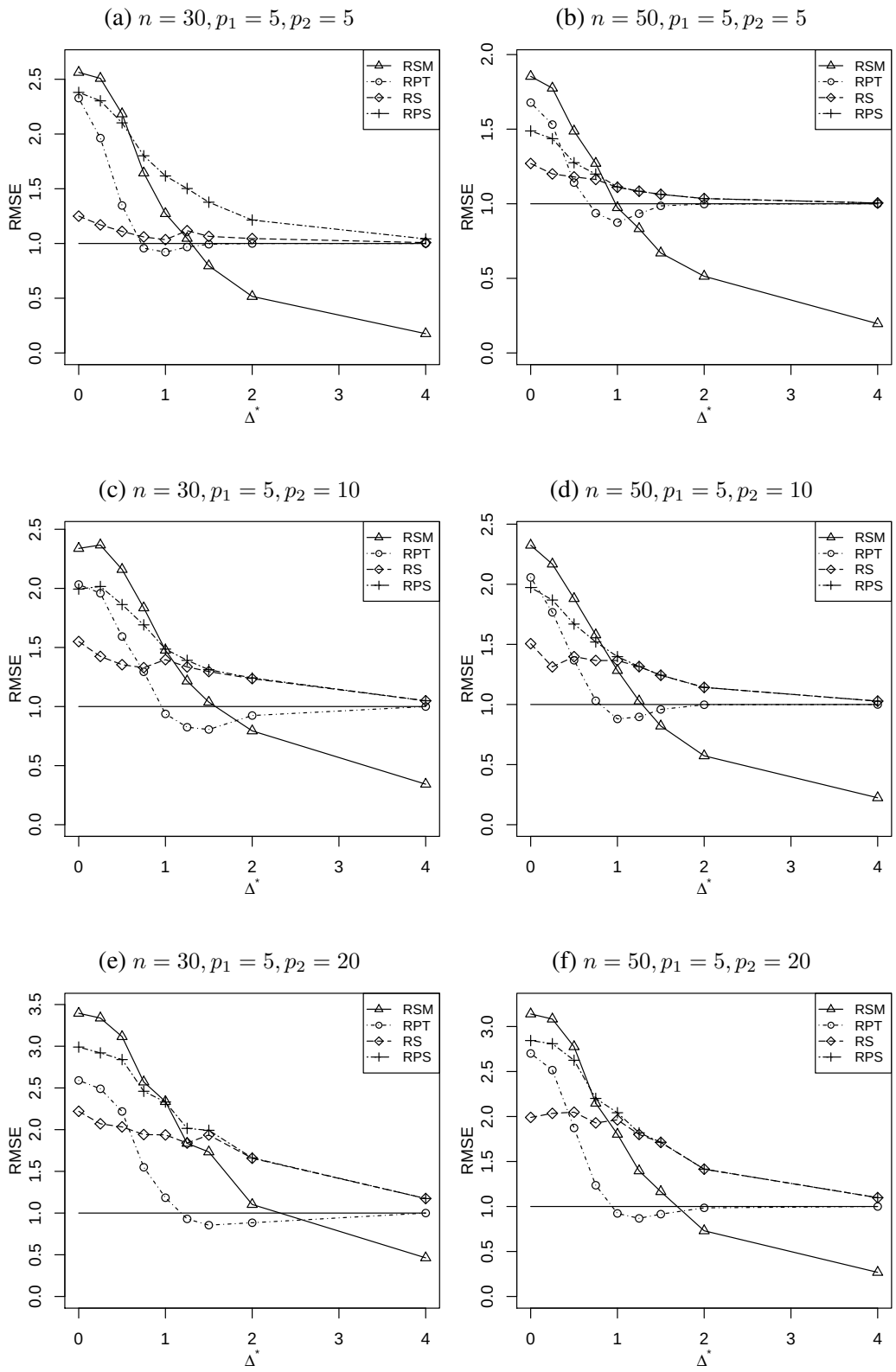


Figure 2.3: Relative efficiency of the estimators as a function of the non-centrality parameter Δ^* for various sample size $n = 30$ and 50, and $p_2 = 5, 10$ and 20 when $\rho = 0.75$.

2.7 Comparison Non-Penalty Estimators with Penalty Estimators

In this subsection, we compare shrinkage ridge regression estimators with penalty estimators. In the following Table 2.8, we present simulation results when $p_1 = 5$, $p_2 = 5, 7, 9, 11, 15, 20$, $n = 30, 40, 50, 75$, $\Delta^* = 0$ and $\rho = 0.25$.

Table 2.8: CNT and Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$ and some (n, p_2) values when $\Delta^* = 0$ and $\rho = 0.25$.

n	p_2	CNT	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPT}$	$\widehat{\boldsymbol{\beta}}_1^{RS}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$	$\widehat{\boldsymbol{\beta}}^{LASSO}$	$\widehat{\boldsymbol{\beta}}^{aLASSO}$	$\widehat{\boldsymbol{\beta}}^{SCAD}$
30	5	18.69	0.80	2.65	2.40	1.56	1.92	1.14	1.21	1.12
	7	27.69	0.72	3.32	2.29	1.91	2.24	1.19	1.32	1.05
	9	44.08	0.70	3.96	3.29	2.13	3.11	1.46	1.41	1.02
	11	68.70	0.56	4.72	3.72	2.71	3.24	1.44	1.52	1.05
	15	188.03	0.43	7.22	5.10	2.69	5.16	1.66	1.77	1.46
	20	981.08	0.19	8.75	5.76	4.14	5.97	1.78	2.04	1.35
40	5	13.86	0.83	2.48	1.89	1.30	1.74	1.17	1.29	1.19
	7	19.06	0.84	2.99	2.32	1.83	2.24	1.29	1.47	1.41
	9	26.60	0.79	3.61	2.64	2.39	2.70	1.40	1.56	1.46
	11	40.37	0.73	4.84	3.20	1.87	3.29	1.75	2.04	1.77
	15	69.20	0.60	5.23	3.99	3.06	4.01	1.91	2.12	1.78
	20	164.15	0.45	7.09	5.61	4.27	5.78	2.20	2.79	2.37
50	5	10.88	0.90	2.63	2.54	1.61	1.83	1.30	1.46	1.39
	7	15.85	0.86	3.24	2.21	1.71	2.13	1.38	1.59	1.43
	9	20.36	0.85	3.84	3.50	2.34	2.78	1.54	1.92	1.89
	11	26.02	0.81	3.81	3.02	1.90	2.86	1.64	1.96	1.80
	15	46.75	0.76	5.51	4.12	2.85	4.17	1.92	2.56	2.77
	20	90.66	0.65	7.32	5.36	4.24	5.44	2.25	3.18	3.14
75	5	8.90	0.94	2.20	1.97	1.53	1.64	1.25	1.53	1.48
	7	11.75	0.92	3.10	2.17	1.81	2.14	1.44	1.89	1.67
	9	15.12	0.91	3.44	2.96	2.25	2.68	1.60	2.12	2.09
	11	18.29	0.88	3.90	2.48	2.36	2.78	1.65	2.26	2.10
	15	28.11	0.85	5.54	3.26	3.63	3.88	2.05	3.13	2.99
	20	43.77	0.78	7.15	4.11	4.94	5.47	2.63	4.13	3.78

For a succinct comparison among suggested estimators, we plotted RMSEs, which are shown in Table 2.8, in a three-dimensional diagram. The horizontal axis represents n, the diagonal axis shows p_2, while the RMSEs are plotted on the vertical axis.

In the following Figure 2.4, we compared shrinkage ridge regression estimators when $\rho = 0.25$. As we see RPT has higher RMSE than the other estimators after RSM for large sample sizes and p_2. We also plotted RPS with penalty estimators in the following Figure 2.5. It safety can be concluded that RPS has highest RMSE for all n and p_2 values.

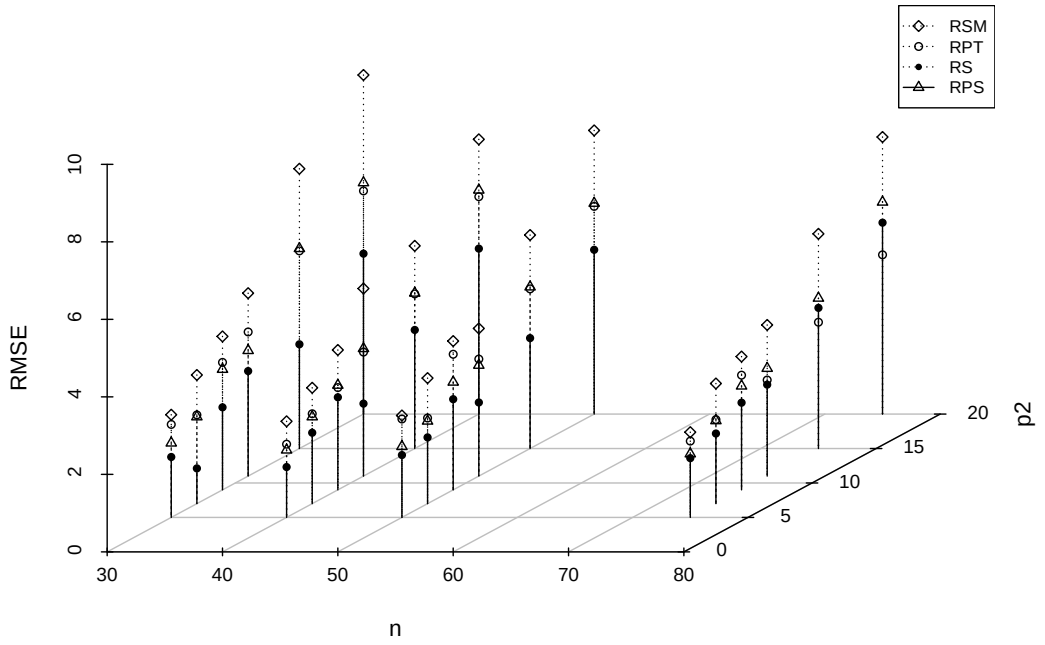


Figure 2.4: Three-dimensional plot of simulated RMSE against n and p_2 to compare non-penalty estimators in situation $p_1 = 5$ and $\rho = 0.25$.

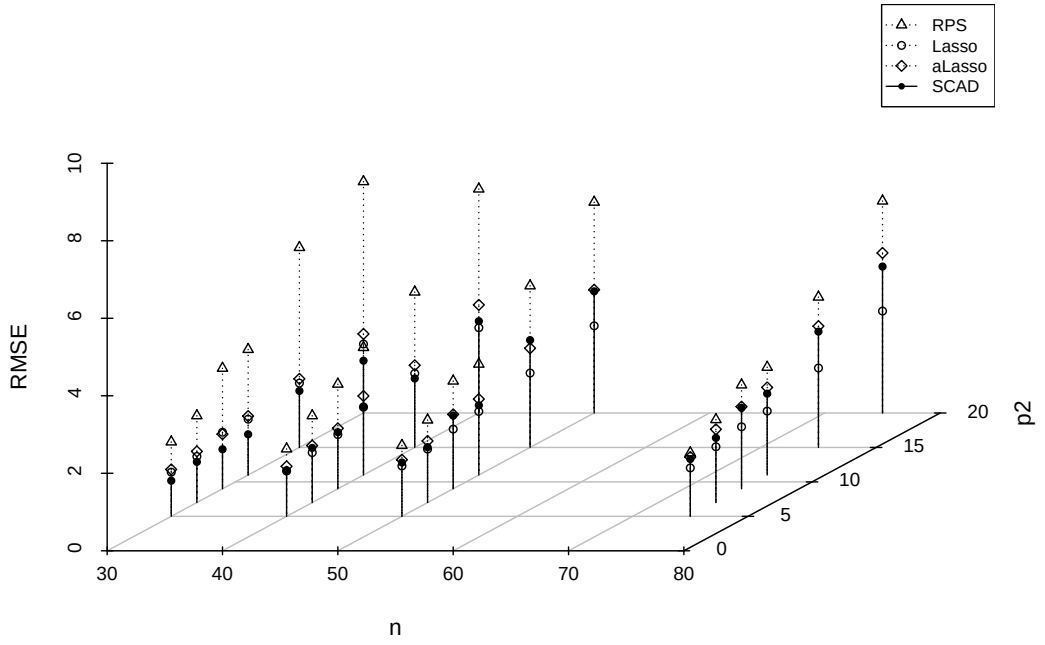


Figure 2.5: Three-dimensional plot of simulated RMSE against n and p_2 to compare penalty and non-penalty estimators in situation $p_1 = 5$ and $\rho = 0.25$.

In the following Table 2.9, we present simulation results when $p_1 = 5$, $p_2 = 5, 7, 9, 11, 15, 20$, $n = 30, 40, 50, 75$, $\Delta^* = 0$ and $\rho = 0.5$.

Table 2.9: CNT and Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$ and some (n, p_2) values when $\Delta^* = 0$ and $\rho = 0.5$.

n	p_2	CNT	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPT}$	$\widehat{\boldsymbol{\beta}}_1^{RS}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$	$\widehat{\boldsymbol{\beta}}^{LASSO}$	$\widehat{\boldsymbol{\beta}}^{aLASSO}$	$\widehat{\boldsymbol{\beta}}^{SCAD}$
30	5	46.13	0.70	3.06	2.43	1.28	1.98	1.01	0.93	0.68
	7	74.73	0.65	3.78	2.84	1.79	2.52	1.17	1.08	0.71
	9	118.11	0.55	5.09	3.70	2.23	3.26	1.29	1.19	0.71
	11	184.68	0.49	5.01	3.85	2.39	3.61	1.37	1.22	0.75
	15	491.27	0.33	6.89	3.93	3.35	4.54	1.31	1.22	0.75
	20	3398.80	0.14	10.55	6.20	4.31	7.09	1.48	1.55	0.92
40	5	34.54	0.79	2.82	2.27	1.55	1.97	1.16	1.12	0.84
	7	48.82	0.72	3.45	2.78	1.95	2.37	1.24	1.16	1.01
	9	76.31	0.70	4.09	3.14	1.31	2.87	1.29	1.25	1.12
	11	98.05	0.65	4.63	4.01	2.10	3.39	1.49	1.52	1.15
	15	191.96	0.52	6.40	4.21	2.86	4.34	1.62	1.73	1.13
	20	469.68	0.38	8.36	4.61	4.02	5.50	1.84	1.99	1.36
50	5	29.39	0.85	2.75	2.39	1.04	1.93	1.22	1.29	1.16
	7	39.03	0.80	3.58	2.73	1.77	2.42	1.28	1.44	1.22
	9	53.39	0.77	4.10	3.32	2.22	2.94	1.46	1.67	1.38
	11	75.87	0.73	4.80	2.88	2.56	3.21	1.56	1.69	1.34
	15	126.44	0.67	5.93	4.69	3.06	4.38	1.76	2.13	1.65
	20	245.60	0.54	8.07	5.77	4.13	5.70	1.99	2.38	1.94
75	5	22.77	0.88	2.59	2.11	1.45	1.79	1.25	1.45	1.19
	7	30.34	0.89	3.16	2.62	1.84	2.35	1.43	1.76	1.52
	9	38.33	0.86	4.03	2.95	2.32	2.73	1.56	1.96	1.59
	11	49.99	0.84	4.85	3.45	2.59	3.39	1.74	2.35	2.20
	15	77.16	0.78	5.79	4.34	3.42	4.35	1.97	2.71	2.45
	20	122.80	0.72	7.30	5.52	4.13	5.50	2.28	3.15	2.75

In the following three-dimensional Figure 2.6, we compared shrinkage ridge regression estimators when $\rho = 0.5$. As we see RPT has higher RMSE than the other estimators after RSM for large sample sizes and p_2. We also plotted RPS with penalty estimators in the following Figure 2.7. Again, it safety can be concluded that RPS has highest RMSE for all n and p_2 values.

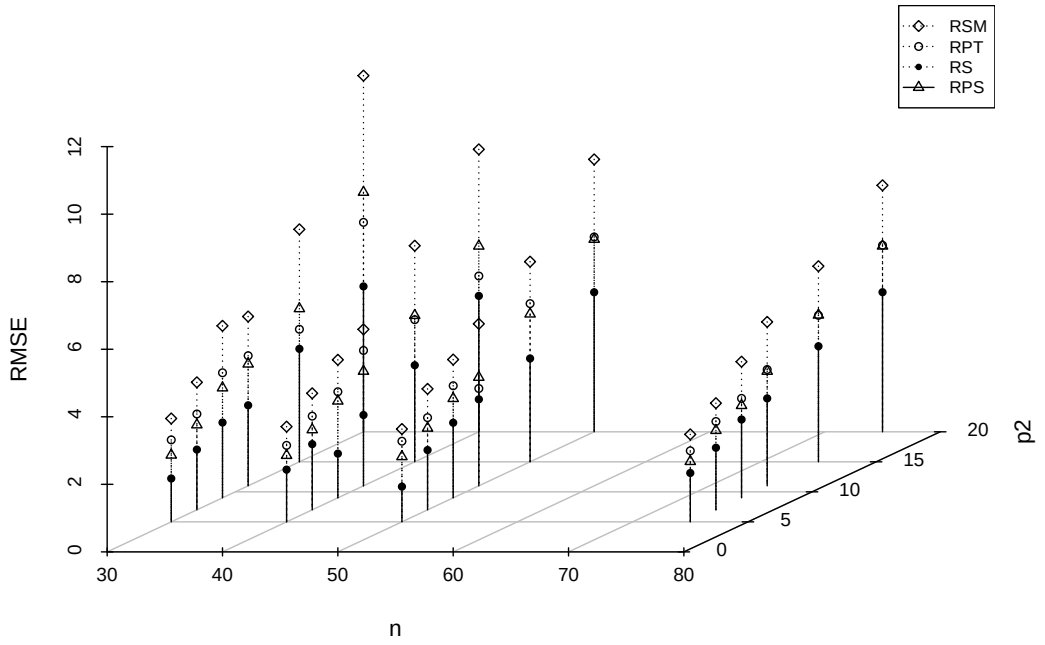


Figure 2.6: Three-dimensional plot of simulated RMSE against n and p_2 to compare non-penalty estimators in situation $p_1 = 5$ and $\rho = 0.5$.

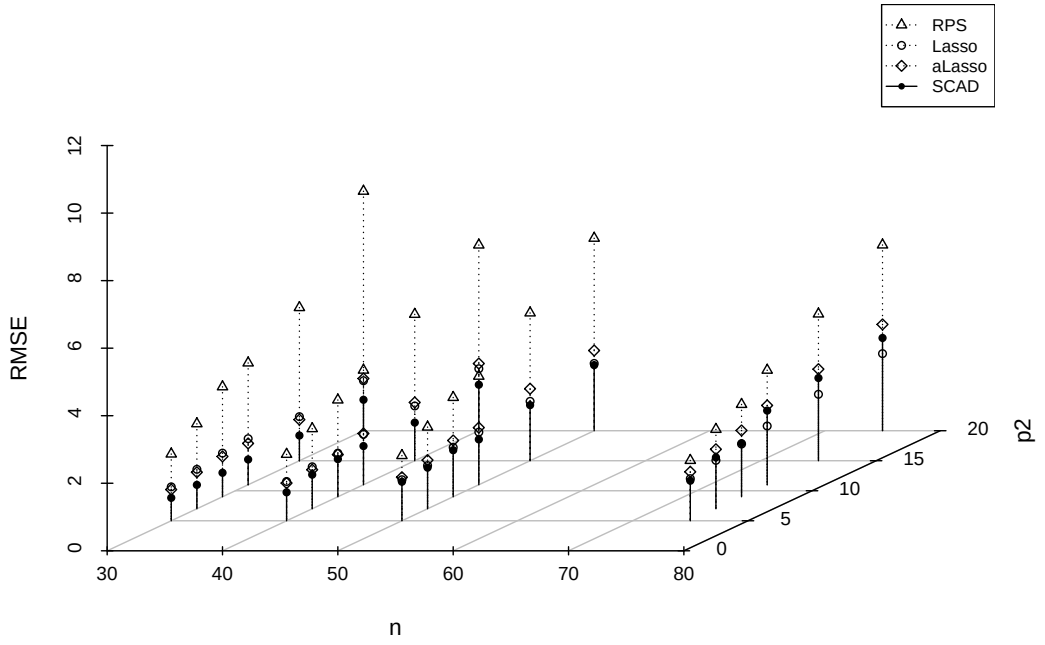


Figure 2.7: Three-dimensional plot of simulated RMSE against n and p_2 to compare penalty and non-penalty estimators in situation $p_1 = 5$ and $\rho = 0.5$.

In the following Table 2.10, we present simulation results when $p_1 = 5$, $p_2 = 5, 7, 9, 11, 15, 20$, $n = 30, 40, 50, 75$, $\Delta^* = 0$ and $\rho = 0.75$.

Table 2.10: CNT and Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$ and some (n, p_2) values when $\Delta^* = 0$ and $\rho = 0.75$.

n	p_2	CNT	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPT}$	$\widehat{\boldsymbol{\beta}}_1^{RS}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$	$\widehat{\boldsymbol{\beta}}^{LASSO}$	$\widehat{\boldsymbol{\beta}}^{aLASSO}$	$\widehat{\boldsymbol{\beta}}^{SCAD}$
30	5	128.71	0.51	3.47	2.83	0.93	2.10	0.81	0.66	0.48
	7	202.48	0.50	4.82	3.68	1.37	2.91	0.94	0.75	0.52
	9	334.47	0.40	5.67	3.55	1.77	3.35	0.98	0.75	0.47
	11	513.10	0.36	6.79	3.66	2.45	3.85	0.98	0.86	0.49
	15	1408.17	0.21	6.37	3.82	3.09	4.34	0.97	0.75	0.47
	20	9555.58	0.10	9.65	4.23	4.01	5.63	1.16	0.93	0.50
40	5	93.44	0.62	3.51	2.54	1.14	2.03	0.91	0.75	0.57
	7	144.03	0.57	4.53	2.91	1.42	2.61	1.00	0.80	0.58
	9	203.81	0.52	5.16	3.70	2.28	3.19	1.04	0.89	0.64
	11	293.29	0.48	6.00	4.54	2.62	3.79	1.17	0.96	0.66
	15	543.65	0.39	6.55	4.36	2.68	4.50	1.23	1.04	0.65
	20	1370.50	0.27	7.87	5.27	3.70	5.79	1.30	1.12	0.69
50	5	79.58	0.71	3.93	2.80	1.66	2.12	1.01	0.93	0.73
	7	116.11	0.66	4.02	2.84	1.56	2.52	1.09	0.98	0.70
	9	156.47	0.64	4.75	3.14	2.07	3.01	1.18	1.03	0.73
	11	219.45	0.57	6.04	3.94	2.06	3.76	1.30	1.13	0.78
	15	385.42	0.48	6.50	4.18	2.72	4.41	1.35	1.26	0.83
	20	718.83	0.39	8.94	4.64	3.40	5.91	1.50	1.25	0.83
75	5	65.35	0.80	3.21	2.63	1.33	2.02	1.13	1.11	0.89
	7	87.24	0.77	4.14	3.02	1.98	2.54	1.22	1.24	0.90
	9	113.78	0.76	5.27	3.67	2.30	3.32	1.42	1.52	1.17
	11	144.93	0.73	5.54	3.90	2.72	3.74	1.48	1.61	1.09
	15	225.06	0.66	6.81	4.24	3.80	4.68	1.61	1.71	1.20
	20	359.89	0.57	7.59	5.52	4.11	5.58	1.82	1.91	1.41

In the following three-dimensional Figure 2.8, we compared shrinkage ridge regression estimators when $\rho = 0.75$. As we see RPT has higher RMSE than the other estimators after RSM for large sample sizes and p_2. We also plotted RPS with penalty estimators in the following Figure 2.9. Again, it safety can be concluded that RPS has highest RMSE for all n and p_2 values.

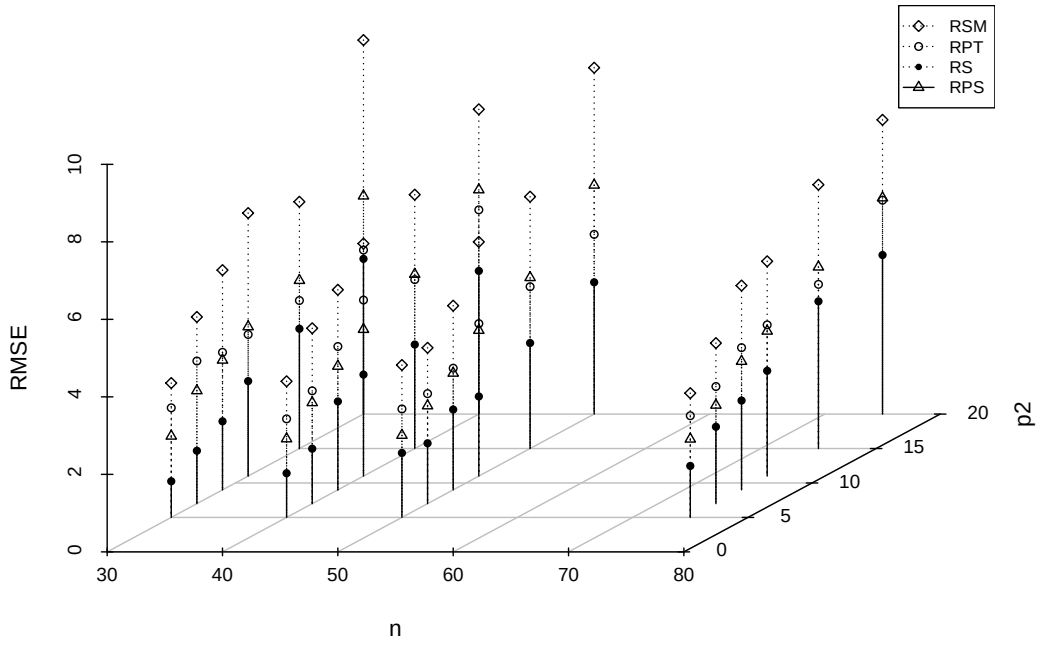


Figure 2.8: Three-dimensional plot of simulated RMSE against n and p_2 to compare non-penalty estimators in situation $p_1 = 5$ and $\rho = 0.75$.

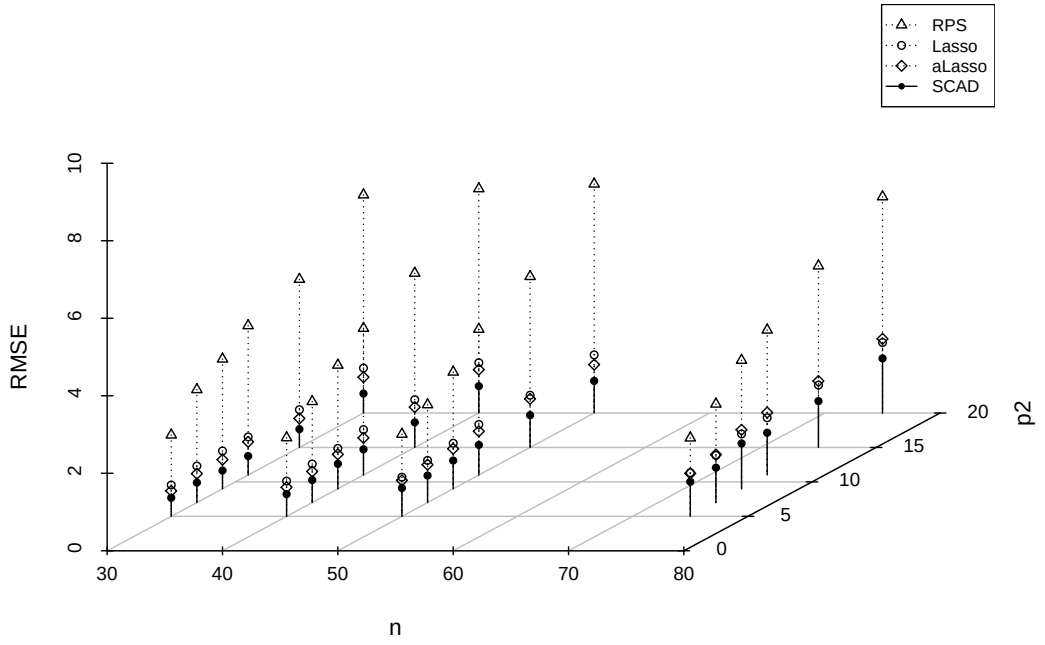


Figure 2.9: Three-dimensional plot of simulated RMSE against n and p_2 to compare penalty and non-penalty estimators in situation $p_1 = 5$ and $\rho = 0.75$.

2.8 Real Data Applications

We present three real data examples to illustrate the usefulness of the suggested strategies. In this part, we used R-package glmnet, R-package parcor and R-package ncvreg to obtain lasso estimates, adaptive lasso estimates and SCAD estimates, respectively.

2.8.1 Prostate data

The data which is studied by (Stamey et al., 1989) which examined the correlation between the level of prostate specific antigen and a number of clinical measures in men who were about to receive a radical prostatectomy. The eight predictors are: log(cancer volume) (lcavol), log (prostate weight) (lweight), age (age),the logarithm of benign prostatic hyperplasia amount (lbph), log(capsular penetration) (lcp), seminal vesicle invasion (svi), Gleason score (gleason), and percent of Gleason scores 4 or 5 (pgg45). The responce is the logarithm of prostate-specific antigen (lpsa).

In the following Table 2.11, it is summarized the coefficient of correlation for prostate data. From this table, we can easily see that correlation between the predictors in the prostate data.

Table 2.11: Correlations of predictors in the prostate data

	lcavol	lweight	age	lbph	svi	lcp	gleason
lweight	0.281						
age	0.225	0.348					
lbph	0.027	0.442	0.350				
svi	0.539	0.155	0.118	-0.086			
lcp	0.675	0.165	0.128	-0.007	0.673		
gleason	0.432	0.057	0.269	0.078	0.320	0.515	
pgg45	0.434	0.107	0.276	0.078	0.458	0.632	0.752

For this data, we fit ridge regression model, and apply the shrinkage estimation method to obtain ridge shrinkage estimates of the regression parameters. Based on best-subset selection (BSS), Lasso, AIC and BIC we obtain four different submodels, in given in Table 2.12.

Table 2.12: Full and candidate sub-models for prostate data

Selection Criterion	Model: Response $\sim$ Covariates
full model	lpsa$\sim$lcavol+lweight+svi+lbph+age+lcp+gleason+pgg45
AIC	lpsa$\sim$lcavol+lweight+svi+lbph+age
Lasso	lpsa$\sim$lcavol+lweight+svi+lbph
BIC	lpsa$\sim$lcavol+lweight+svi
BSS	lpsa$\sim$lcavol+lweight

In Table 2.13, we obtain shrinkage ridge regression estimators based four variables selection methods. Therefore, we compute lasso, adaptive lasso and SCAD estimator to compare with shrinkage estimators. Prediction errors are obtained for each of the cases using 10-fold cross validation. From Table 2.13, it is concluded that RPS(Lasso) estimator has the smallest average prediction error compared to each of the absolute penalty estimators. Also, among the penalty estimators, Lasso estimator has the smallest average prediction error, while SCAD estimator has the biggest

Table 2.13: Average prediction errors repeated 2500 times for prostate data. Numbers in the brackets are the corresponding standard errors of the prediction errors.

Ridge Shrinkage Estimators	AIC	BIC	BSS	Lasso
RFM	0.953(0.0071)	0.519(0.0072)	0.567(0.0077)	0.509(0.0071)
RSM	0.459(0.0070)	0.488(0.0072)	0.542(0.0076)	0.471(0.0071)
RPS	0.693(0.0070)	0.496(0.0072)	0.551(0.0076)	0.477(0.0071)
Penalty Estimators	10-Fold Cross Validation			
Lasso	0.479(0.0069)			
aLasso	0.567(0.0071)			
SCAD	0.607(0.0069)			

Figure 2.10 shows average prediction errors when prediction errors based on RPS(AIC), RPS(BIC), RPS(BSS) and RPS(Lasso) models are compared. It is clearly shown that RPS(Lasso) has smaller overall prediction errors compared to the others RPS estimators.

Figure 2.11 shows comparison between RPS(LASSO) and the Lasso models, adaptive lasso and SCAD, respectively. Clearly, RPS(LASSO) has smaller prediction errors compared to penalty estimators. We can concluded from the analyses demonstrate that the positive shrinkage ridge regression estimators minimize the overall risk when we have some prior information about some of the covariates.

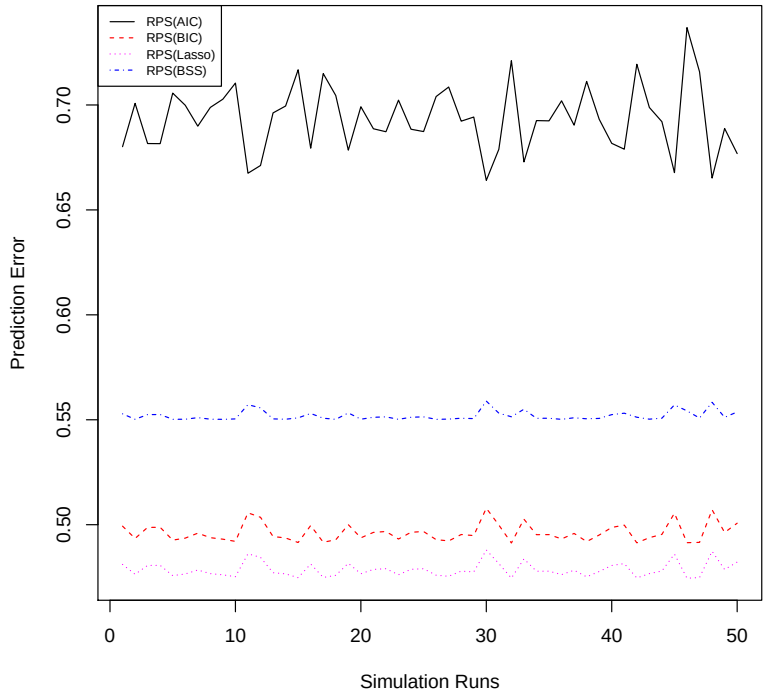


Figure 2.10: Comparison of average prediction error for positive shrinkage ridge regression estimators based on AIC, BIC, Lasso and BSS (only first 50 values).

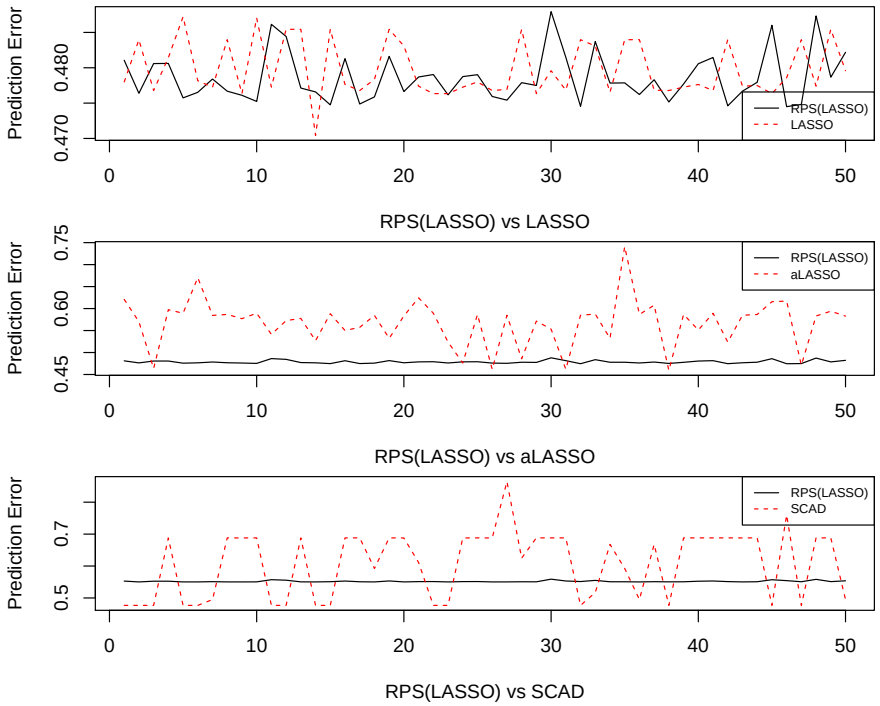


Figure 2.11: Comparison of average prediction error RPS based on Lasso with Lasso, aLasso and SCAD estimators using 10-fold cross validation (only first 50 values).

2.8.2 Account deficit data

The data set contains the cRFMnt account deficit, industrial production index, import and export rates and theirs ratio to gross domestic product, rate of exports meeting imports, balance of trade and trade openness. All the variables are obtained from International Financial Statistics database. We employ quarterly data from $1987 : Q1$ to $2013 : Q4$. The predictors are six economic measures : Industrial Production Index (IPI) is The Industrial Production Index is an index for India which details out the growth of various sectors in an economy such as mining, electricity and manufacturing, Import Ratio to Gross Domestic Product (M) is $(M/GDP) * 100$, Import Ratio to Gross Domestic Product (X) is $(X/GDP) * 100$, Import Ratio to Gross Domestic Product (X/M) is X/M, Balance of Trade (BT) is $(X - M)$ and Trade Openness (TO) is $(X + M)/GDP$. The response is CRFMnt Account Deficit (CA) which is a measurement of a country's trade in which the value of goods and services it imports exceeds the value of goods and services it exports.

In the following Table 2.14, it is summarized the coefficient of correlation for Account Deficit data. From this table, we can easily see that correlation between the predictors in the Account Deficit data. It is also shown that the coefficient of correlation between TO and X has the highest. Also, most of coefficients of correlation among the predictors of Account Deficit data are highly correlated.

Table 2.14: Correlations of predictors in Account Deficit Data

	IPI	M	X	(X/M)	BT
M	0.688				
X	0.814	0.854			
(X/M)	-0.234	0.298	-0.222		
BT	-0.666	-0.384	-0.803	0.743	
TO	0.794	0.944	0.975	-0.015	-0.659

The full model and the sub-models based on AIC, BIC, BSS and Lasso are shown in Table 2.15.

Table 2.15: Full and candidate sub-models for Account Deficit Data

Selection Criterion	Model: Response $\sim$ Covariates
full model	CA$\sim$IPI+M+X+(X/M)+BT+TO
AIC	CA$\sim$M+BT+TO
BIC	CA$\sim$M+BT
BSS	CA$\sim$M+(X/M)+BT
Lasso	CA$\sim$BT+TO

In the following Table 2.16, we summary ridge regression estimators of full model, sub-model, pretest, shrinkage and positive shrinkage with penalty estimators which are Lasso, adaptive lasso and SCAD for the Account Deficit Data. Average prediction errors along with their standard errors for related to estimators are also presented in Table 2.16.

Table 2.16: Average prediction errors repeated 2500 times for Account Deficit Data. Numbers in the brackets are the corresponding standard errors of the prediction errors.

Ridge Estimators	AIC	BIC	BSS	Lasso
RFM	0.00029(0.00016)	0.00029(0.00016)	0.00226(0.00018)	0.00447(0.00028)
RSM	0.00022(0.00014)	0.00023(0.00015)	0.00039(0.00019)	0.00043(0.00020)
RPT	0.00022(0.00014)	0.00023(0.00015)	0.00226(0.00018)	0.00447(0.00028)
RS	0.00235(0.00037)	0.01167(0.00077)	0.00218(0.00018)	0.00437(0.00028)
RPS	0.00024(0.00014)	0.00025(0.00015)	0.00213(0.00018)	0.00435(0.00028)

Penalty Estimators	10-Fold Cross Validation
Lasso	0.00206(0.00014)
aLasso	0.00295(0.00015)
SCAD	0.00126(0.00015)

2.8.3 Baseball hitter's data

The data are for 322 Major League Baseball regular and substitute hitters in 1986. The predictors are sixteen measures : Atbat is Number of times at bat in 1986, Hits are number of hits in 1986, Homer is number of home runs in 1986, Runs are number of runs in 1986, RBI are batted in during 1986, Walks are number of walks in 1986, Years are number of years in the major leagues, Atbatc is number of times at bat in his career, Hitsc is number of hits in career, Homerc is number of home runs in career, Runsc is number of runs in career, RBIc is number of Runs Batted In in career, Walksc is number of walks in career, Putouts are number of putouts in 1986, Assists are number of assists in 1986, Errors are number of errors

in 1986, Salary is annual salary (in thousands) on opening day 1987. The response is the logarithm of annual salary (in thousands) on opening day 1987 (logSal).

The data was originally published in 1988, ASA Statistical Graphics and Computing Data Exposition: $http://lib.stat.cmu.edu/data-expo/1988.html$. The salary data were taken from Sports Illustrated, April 20, 1987. The salary of any player not included in that article is listed as an NA. The 1986 and career statistics were taken from The 1987 Baseball Encyclopedia Update published by Collier Books, Macmillan Publishing Company, New York. The data is also analyzed by (Friendly, 2002).

In the following Table 2.17, it is summarized the coefficient of correlation for Baseball Hitter's Data.

Table 2.17: Correlations of predictors in Baseball Hitter's Data

	Atbat	Hits	Homer	Runs	RBI	Walks	Years	Atbatc	Hitsc	Homerc	Runsc	RBIc	Walksc	Putouts	Assists
Hits	0.964														
Homer	0.555	0.531													
Runs	0.900	0.911	0.631												
RBI	0.796	0.788	0.849	0.779											
Walks	0.624	0.587	0.440	0.697	0.570										
Years	0.013	0.019	0.113	-0.012	0.130	0.135									
Atbatc	0.207	0.207	0.217	0.172	0.278	0.269	0.916								
Hitsc	0.225	0.236	0.217	0.191	0.292	0.271	0.898	0.995							
Homerc	0.212	0.189	0.493	0.230	0.442	0.350	0.722	0.802	0.787						
Runsc	0.237	0.239	0.258	0.238	0.307	0.333	0.877	0.983	0.985	0.826					
RBIc	0.221	0.219	0.350	0.202	0.388	0.313	0.864	0.951	0.947	0.928	0.946				
Walksc	0.133	0.123	0.227	0.164	0.234	0.429	0.838	0.907	0.891	0.811	0.928	0.889			
Putouts	0.310	0.300	0.251	0.271	0.312	0.281	-0.020	0.053	0.067	0.094	0.059	0.095	0.058		
Assists	0.342	0.304	-0.162	0.179	0.063	0.103	-0.085	-0.008	-0.013	-0.189	-0.039	-0.097	-0.066	-0.043	
Errors	0.326	0.280	-0.010	0.193	0.150	0.082	-0.157	-0.070	-0.068	-0.165	-0.094	-0.115	-0.130	0.075	0.704

The full model and the sub-models based on Lasso, BIC and AIC are shown in Table 2.18.

Table 2.18: Full and candidate sub-models for Baseball Hitter's Data

Selection Criterion	Model: Response $\sim$ Covariates
full model	logSal $\sim$ Atbat + Hits + Homer + Runs + RBI + Walks + Years + Atbatc + Hitsc + Homerc + Runsc + RBIc + Walksc + Putouts + Assists + Errors
Lasso	logSal $\sim$ Atbat + Years + RBIc
BIC	logSal $\sim$ Runs + Years + Hitsc + Runsc + RBIc + Errors
AIC	logSal $\sim$ Runs + Years + Atbatc + Hitsc + Runsc + RBIc + Errors

As it is showed that in Table 2.18, Lasso gives a model with Atbat, Years and RBIc. In addition to Lasso, AIC selects model with Runs,Year, Hitsc, Runscs, RBIc and Error, even though BIC selects model with Runs, Years, Atbatc, Hitsc, Runsc and Error.

Table 2.19 shows average prediction errors and their standard deviations for full model ridge regression, sub-model ridge regression and positive shrinkage ridge regression estimators and penalty estimators based on 10-fold cross validation repeated 2500 times for the Baseball Hitter's Data.

Table 2.19: Average prediction errors repeated 2500 times for Baseball Hitter's Data. Numbers in the brackets are the corresponding standard errors of the prediction errors.

Ridge Estimators	AIC	BIC	Lasso
RFM	2.747(0.0015)	3.037(0.0039)	1.762(0.0017)
RSM	0.297(0.0020)	0.266(0.0019)	0.298(0.0020)
RPS	2.256(0.0014)	1.843(0.0034)	1.432(0.0017)
Penalty Estimators	**10-Fold Cross Validation**		
Lasso	3.705(0.0017)		
aLasso	3.730(0.0012)		
SCAD	3.729(0.0015)		

2.9 Shrinkage Estimators for High-Dimensional Data

We can use $\widehat{\boldsymbol{\beta}}^{RFM} = \left(\boldsymbol{X}^\top \boldsymbol{X} + \lambda^R \boldsymbol{I}_p\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}$ as a full model estimator when $p > n$.

Based on UIP or AI, if $\boldsymbol{\beta}_2 = \mathbf{0}$, then a sub-model estimator for $\boldsymbol{\beta}_1$ will be $\widehat{\boldsymbol{\beta}}_1^{RSM} = \left(\boldsymbol{X}_1^\top \boldsymbol{X}_1 + \lambda^R \boldsymbol{I}_{p_1}\right)^{-1} \boldsymbol{X}_1^\top \boldsymbol{y}$ and a full model estimator may be defined as $\widehat{\boldsymbol{\beta}}_1^{RFM} = \left(\boldsymbol{X}_1^\top \boldsymbol{M}_2^R \boldsymbol{X}_1 + \lambda^R \boldsymbol{I}_{p_1}\right)^{-1} \boldsymbol{X}_1^\top \boldsymbol{M}_2^R \boldsymbol{y}$, where $\boldsymbol{M}_2^R$ is defined as $\boldsymbol{I}_n - \boldsymbol{X}_2 \left(\boldsymbol{X}_2^\top \boldsymbol{X}_2 + \lambda^R \boldsymbol{I}_{p_2}\right)^{-1} \boldsymbol{X}_2^\top$.

To construct shrinkage estimators, we suggest following high dimensional test statistics

$$\mathscr{L}_h = \frac{1}{\widehat{\sigma}_{RE}^2} \left(\widehat{\boldsymbol{\beta}}_2^{RFM}\right)^\top \left(\boldsymbol{X}_2^\top \boldsymbol{M}_1^R \boldsymbol{X}_2\right) \widehat{\boldsymbol{\beta}}_2^{RFM},$$

where $\widehat{\sigma}_{RE}^2$ is estimated as follows:

$$\widehat{\sigma}_{RE}^2 = \frac{1}{(n - p_1)} (\boldsymbol{y} - \boldsymbol{X}_1 \widehat{\boldsymbol{\beta}}_1^{RSM})^\top (\boldsymbol{y} - \boldsymbol{X}_1 \widehat{\boldsymbol{\beta}}_1^{RSM}).$$

The shrinkage or stein-type ridge regression estimator of $\boldsymbol{\beta}_1$ defined by

$$\widehat{\boldsymbol{\beta}}_1^{RS} = \widehat{\boldsymbol{\beta}}_1^{RSM} + \left(\widehat{\boldsymbol{\beta}}_1^{RFM} - \widehat{\boldsymbol{\beta}}_1^{RSM}\right) \left(1 - (\widehat{\vartheta} - 2) \mathscr{L}_p^{-1}\right),$$

where $\widehat{\vartheta} = \frac{1}{2} \left\| \widehat{\boldsymbol{\beta}}_2^{RFM} \right\|_1^2$ and $\|\cdot\|_1$ indicates the one norm.

The positive shrinkage ridge regression estimator of $\boldsymbol{\beta}_1$ defined by

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

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

2.10 Real Data Applications for High-Dimensional Data

In this section, we present two real data examples to illustrate the usefulness of the suggested strategies for high-dimensional data.

2.10.1 Biscuit doughs data

This data set contains measurements from quantitative NIR spectroscopy. The example studied arises from an experiment done to test the feasibility of NIR spec-

troscopy to measure the composition of biscuit dough pieces (formed but unbaked biscuits). Two similar sample sets were made up, with the standard recipe varied to provide a large range for each of the four constituents under investigation: fat, sucrose, dry flour, and water. The calculated percentages of these four ingredients represent the 4 responses. There are 40 samples in the calibration or training set (with sample 23 being an outlier) and a further 32 samples in the separate prediction or validation set (with example 21 considered as an outlier). An NIR reflectance spectrum is available for each dough piece. The spectral data consist of 700 points measured from 1100 to 2498 nanometers (nm) in steps of 2 nm. A data frame of dimension 72×704. The first 700 columns correspond to the NIR reflectance spectrum, the last four columns correspond to the four constituents fat, sucrose, dry flour, and water. (Brown et al., 2001; Osborne et al., 2984).

Table 2.20: Average prediction errors repeated 1000 times for Biscuit Doughs Data. Numbers in the brackets are the corresponding standard errors of the prediction errors.

Dataset	(n, p_1, p_2)	RFM	RSM	RPS	Lasso	aLasso	SCAD
Biscuit Doughs (y is fat)	$(72, 22, 678)$	2.0998 (0.0203)	0.0533 (0.0032)	0.0904 (0.0145)	0.0855 (0.0155)	0.0871 (0.0042)	0.1545 (0.0054)
Biscuit Doughs (y is sucrose)	$(72, 25, 675)$	11.0199 (0.0464)	0.7335 (0.0120)	8.6631 (0.0457)	6.6631 (0.0457)	1.0758 (0.0145)	1.4597 (0.0166)
Biscuit Doughs (y is dry)	$(72, 16, 684)$	9.4245 (0.0423)	0.5358 (0.0102)	2.4062 (0.0189)	0.7699 (0.0123)	0.9365 (0.0135)	1.0792 (0.0145)
Biscuit Doughs (y is water)	$(72, 24, 676)$	1.8177 (0.0189)	0.0878 (0.0041)	0.4401 (0.0168)	0.1319 (0.0051)	0.1378 (0.0052)	0.1909 (0.0061)

In Table 2.20, it has beed shown that the average prediction errors and their standard deviations for proposed estimators based on 10-fold cross validation repeated 1000 times for the Biscuit Doughs Data. To get sub-model, we used Lasso variable selection method. As it can seen from Table 2.20 that Lasso shrinks 678, 675, 684 and 676 parameters to zero, respectively.

2.10.2 Riboflavin data

Dataset of riboflavin production by bacillus subtilis containing 71 observations of 4088 predictors (gene expressions) and a one-dimensional response (riboflavin production). In this data set, the response variable is Log-transformed riboflavin

production rate (original name: qRIBFLV) and the predictor variables are measuring the logarithm of the expression level of 4088 genes, (Bühlmann et al., 2013).

Table 2.21: Average prediction errors repeated 1000 times for Riboflavin Data. Numbers in the brackets are the corresponding standard errors of the prediction errors.

Dataset	(n, p_1, p_2)	RFM	RSM	RPS	Lasso	aLasso	SCAD
Riboflavin	$(71, 41, 4047)$	0.4224	0.0089	0.0166	0.0169	0.0609	0.0469
		(0.0071)	(0.0010)	(0.0010)	(0.0014)	(0.0027)	(0.0024)

In Table 2.21, it has beed shown that the average prediction errors and their standard deviations for proposed estimators based on 10-fold cross validation repeated 1000 times for the Riboflavin Data. As it can seen from Table 2.21 that Lasso shrinks 4047 parameters to zero. For this data, RPS outperform to penalty estimators.

2.11 Simulation Studies for High-Dimensional Data

We perform Monte Carlo simulation experiments to examine the quadratic risk performance of the proposed estimators. We simulate the response from the following model:

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

where $\boldsymbol{x}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_x)$ and ε_i are i.i.d. $N(0, 1)$. We also define $\boldsymbol{\Sigma}_x$ that is positive definite matrix with correlated. To generate $\boldsymbol{\Sigma}_x$, we define ρ which is coefficient of correlation $\boldsymbol{X}$. In simulation results are obtained for three kinds of correlation which are low($\rho = 0.25$), medium($\rho = .5$) and high($\rho = 0.75$) levels. We consider the hypothesis $H_0 : \beta_j = 0$, for $j = p_1 + 1, p_1 + 2, ..., p$, with $p = p_1 + p_2$. Hence, we partition the regression coefficients as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (\boldsymbol{\beta}_1, \mathbf{0})$ with $\boldsymbol{\beta}_1 = (1, 1, 1, 1, 1)$.

Each realization was repeated 1000 times to obtain sensible results, and for each realization, we calculated bias of the estimators. We define $\Delta^* = \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|$, where $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_1, \mathbf{0})$, and $\|\cdot\|$ is the Euclidean norm. In order to investigate the behaviour of the estimators for $\Delta^* > 0$, further samples were generated from those distributions under local alternative hypotheses

The performance of an estimator $\boldsymbol{\beta}_1$ was evaluated by calculating its mean

squared error (MSE) criterion. We numerically calculated the RMSE of $\widehat{\boldsymbol{\beta}}_1^{RSM}$, $\widehat{\boldsymbol{\beta}}_1^{RPT}$, $\widehat{\boldsymbol{\beta}}_1^{RS}$ and $\widehat{\boldsymbol{\beta}}_1^{RPS}$, with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$. The relative mean square efficiency(RMSE) of the $\boldsymbol{\beta}_1^{\blacktriangle}$ to the unrestricted least squares estimator $\widehat{\boldsymbol{\beta}}_1^{RFM}$ is indicated by

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

where $\boldsymbol{\beta}_1^{\blacktriangle}$ is one of the proposed estimators. The amount by which a RMSE is larger than one indicates the degree of superiority of the estimator $\boldsymbol{\beta}_1^{\blacktriangle}$ over $\widehat{\boldsymbol{\beta}}_1^{RFM}$.

We summarized the simulation results in the following Tables 2.22 and 2.23.

Table 2.22: In case of high-dimensional, simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $n = 30$ and $p_1 = 5$.

ρ	p_2	Δ^*	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$	p_2	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$	p_2	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$
		0.0	7.947	2.636		8.948	3.348		9.541	3.196
		0.5	5.739	2.414		7.488	3.140		7.577	3.174
		1.0	3.074	2.154		4.116	2.659		5.490	2.855
0.25	100	1.5	1.901	1.901	250	2.991	2.524	500	3.165	2.567
		2.0	1.396	1.832		1.880	2.317		2.008	2.414
		4.0	0.497	1.512		0.667	1.893		0.677	1.996
		8.0	0.161	1.277		0.193	1.415		0.216	1.574
		16.0	0.065	1.064		0.076	1.051		0.078	0.990
		0.0	5.059	1.643		6.210	2.140		7.498	2.259
		0.5	4.464	1.528		5.151	2.011		6.509	2.073
		1.0	3.120	1.461		3.440	1.898		3.920	2.159
0.5	100	1.5	1.872	1.444	250	2.332	1.818	500	2.137	1.890
		2.0	1.349	1.444		1.802	1.778		1.781	1.942
		4.0	0.453	1.382		0.512	1.745		0.576	1.862
		8.0	0.139	1.344		0.173	1.639		0.194	1.814
		16.0	0.044	1.278		0.053	1.445		0.069	1.817
		0.0	4.354	1.246		3.392	1.579		4.899	1.813
		0.5	3.328	1.233		3.579	1.476		3.799	1.686
		1.0	2.422	1.218		3.056	1.459		3.013	1.637
0.75	100	1.5	1.892	1.219	250	1.713	1.413	500	2.440	1.608
		2.0	1.281	1.216		1.288	1.397		1.499	1.611
		4.0	0.478	1.229		0.578	1.398		0.608	1.586
		8.0	0.137	1.259		0.186	1.437		0.196	1.607
		16.0	0.042	1.326		0.055	1.475		0.063	1.697

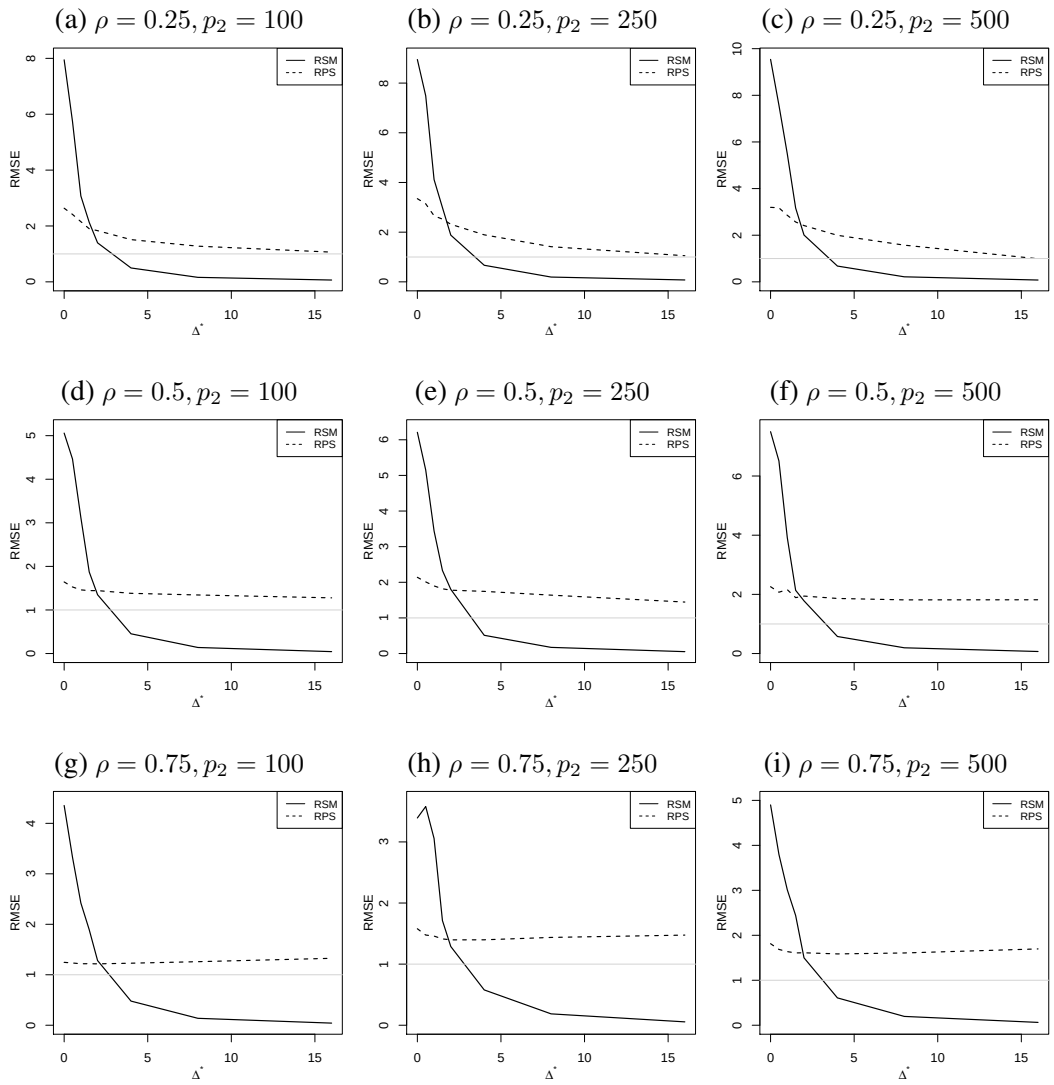


Figure 2.12: Relative efficiency of the estimators as a function of the non-centrality parameter Δ^* for $n = 30$, $p_1 = 5$, $p_2 = 100, 250, 500$ and $\rho = 0.25, 0.5, 0.75$.

In Table 2.22, the $\widehat{\boldsymbol{\beta}}_1^{RPS}$ estimator outperforms the $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for all Δ^* values. As $\Delta^* = 0$, not surprisingly, $\widehat{\boldsymbol{\beta}}_1^{RSM}$ has the biggest RMSE. However, $\widehat{\boldsymbol{\beta}}_1^{PRS}$ performs better than $\widehat{\boldsymbol{\beta}}_1^{RSM}$ when Δ^* larger than zero.

In the Figure 2.12, we plot the simulation results which is shown in Table 2.22. The RMSE curve of $\widehat{\boldsymbol{\beta}}_1^{RSM}$ declines and approaches zero step by step while the RMSE curve of $\widehat{\boldsymbol{\beta}}_1^{RPS}$ declines step by step and approaches one gradually as Δ^* is larger than zero.

Table 2.23: In case of high-dimensional, simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $n = 80$ and $p_1 = 5$.

ρ	p_2	Δ^*	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$	p_2	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$	p_2	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$
		0.0	9.192	3.846		17.351	11.482		23.027	16.243
		0.5	7.054	2.857		12.337	9.339		17.020	11.279
		1.0	4.040	2.295		7.077	5.919		9.852	7.605
0.25	100	1.5	2.312	1.885	250	4.390	4.115	500	6.311	5.565
		2.0	1.690	1.647		2.708	3.043		4.317	4.042
		4.0	0.580	1.365		0.943	2.135		1.054	2.695
		8.0	0.192	1.255		0.240	1.703		0.307	2.263
		16.0	0.068	1.145		0.079	1.324		0.090	1.469
		0.0	8.459	2.051		12.663	5.337		15.923	8.041
		0.5	6.474	1.804		9.487	4.288		12.648	6.168
		1.0	3.884	1.612		6.116	3.293		7.445	4.707
0.5	100	1.5	2.435	1.525	250	3.506	2.767	500	4.337	4.055
		2.0	1.665	1.446		2.323	2.372		3.102	3.595
		4.0	0.511	1.325		0.736	2.019		0.895	2.918
		8.0	0.159	1.277		0.200	1.794		0.251	2.706
		16.0	0.052	1.242		0.063	1.589		0.070	2.094
		0.0	7.535	1.425		8.480	2.729		9.407	4.756
		0.5	4.944	1.321		6.849	2.240		8.171	3.978
		1.0	3.967	1.274		4.523	1.975		5.014	3.058
0.75	100	1.5	2.741	1.245	250	3.156	1.840	500	3.670	2.810
		2.0	1.793	1.233		2.301	1.707		2.329	2.571
		4.0	0.594	1.202		0.709	1.540		0.807	2.348
		8.0	0.165	1.191		0.207	1.468		0.231	2.231
		16.0	0.044	1.204		0.057	1.414		0.059	1.881

According to the above table, the $\widehat{\boldsymbol{\beta}}_1^{RPS}$ estimator outshines the $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for all Δ^* values. In case of $\Delta^* = 0$, $\widehat{\boldsymbol{\beta}}_1^{RSM}$ has the biggest RMSE. On the other hand, $\widehat{\boldsymbol{\beta}}_1^{RSM}$ is getting worser than $\widehat{\boldsymbol{\beta}}_1^{PRS}$ as Δ^* is larger than zero.

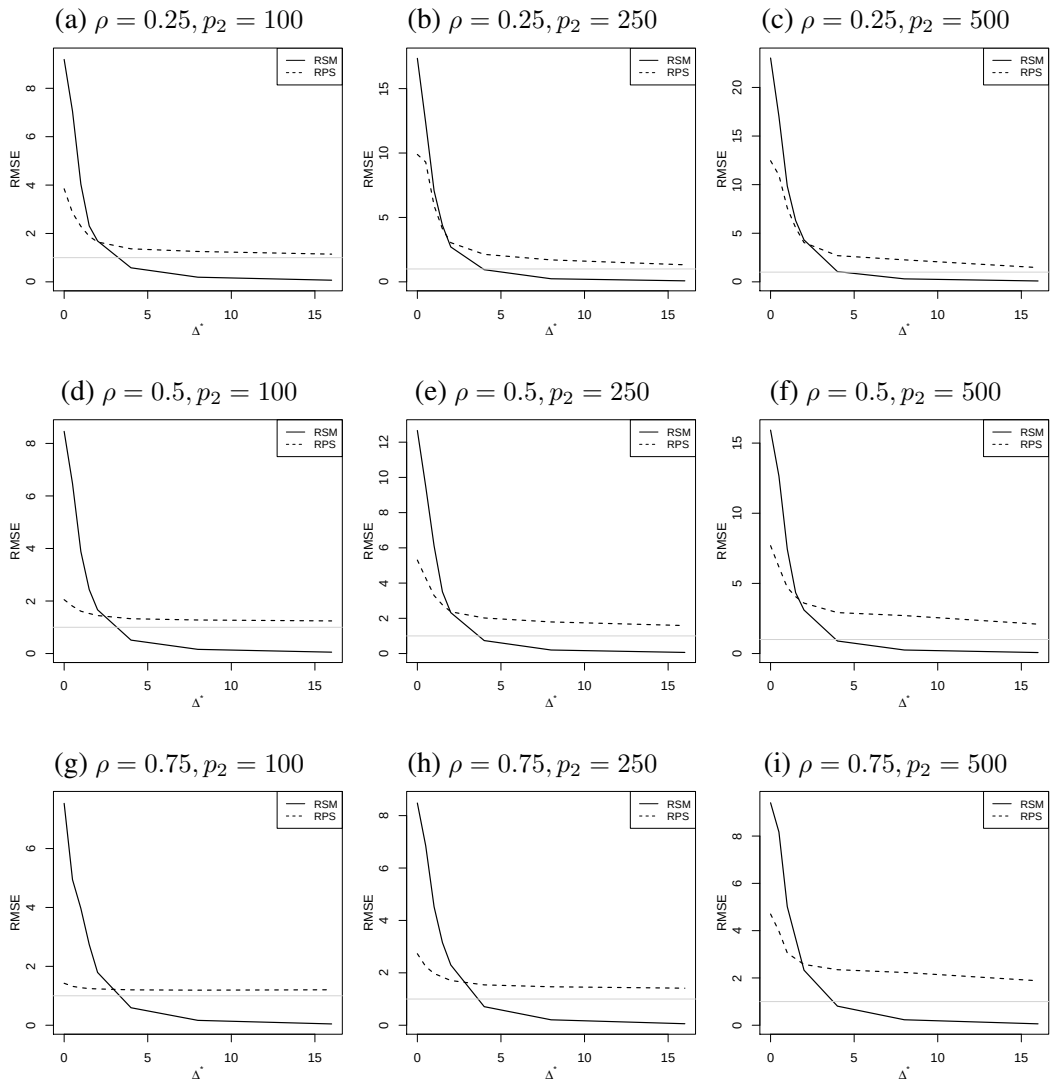


Figure 2.13: Relative efficiency of the estimators as a function of the non-centrality parameter Δ^* for $n = 80$, $p_1 = 5$, $p_2 = 100, 250, 500$ and $\rho = 0.25, 0.5, 0.75$.

We also plot the simulation results in the Figure 2.13. In summary, as Δ^* is further away from zero, the RMSE curve of $\widehat{\boldsymbol{\beta}}_1^{RSM}$ declines and approaches zero step by step, while the RMSE curve of $\widehat{\boldsymbol{\beta}}_1^{RPS}$ declines step by step and approaches one gradually.

2.12 HD Comparison Non-Penalty Estimators with Penalty Estimators

In this subsection, we compare positive shrinkage ridge regression estimator with penalty estimators. The results are summarized in the following Table 2.24.

Table 2.24: In case of High-Dimensional situation, simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for $p_1 = 5$ and some (n, p_2) values when $\Delta^* = 0$ and $\rho = 0.25$.

n	ρ	p_2	$\widehat{\boldsymbol{\beta}}_1^{RSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$	$\widehat{\boldsymbol{\beta}}^{LASSO}$	$\widehat{\boldsymbol{\beta}}^{aLASSO}$	$\widehat{\boldsymbol{\beta}}^{SCAD}$
30	0.25	100	6.635	2.517	1.613	1.298	0.685
		200	9.215	2.859	1.530	1.313	0.800
		400	9.699	3.197	1.284	1.011	0.761
		600	9.703	3.088	1.251	1.021	0.767
		1000	10.935	2.889	1.193	1.029	0.835
	0.5	100	4.755	1.622	1.331	1.007	0.649
		200	6.285	1.961	1.335	1.060	0.561
		400	6.295	2.080	1.244	0.995	0.569
		600	6.896	2.126	1.198	0.948	0.588
		1000	7.313	1.993	1.160	1.006	0.589
	0.75	100	3.531	1.241	1.190	0.799	0.486
		200	3.718	1.484	1.165	0.821	0.493
		400	3.799	1.611	1.090	0.837	0.545
		600	4.525	1.801	1.135	0.899	0.535
		1000	4.972	1.746	1.086	0.885	0.534
75	0.25	100	8.208	3.596	3.750	6.013	4.489
		200	14.504	7.407	4.978	9.970	6.607
		400	21.165	11.092	5.230	12.038	7.731
		600	24.800	12.248	5.458	12.270	8.499
		1000	25.067	8.526	4.591	11.103	7.888
	0.5	100	9.045	2.004	3.345	4.919	3.823
		200	10.880	3.740	3.564	5.185	3.684
		400	15.651	5.677	3.526	5.663	3.934
		600	17.820	5.762	3.555	5.181	3.368
		1000	20.151	4.260	3.362	5.356	3.995
	0.75	100	6.779	1.395	2.267	1.786	1.095
		200	7.343	2.085	2.298	1.767	1.017
		400	8.160	3.293	2.114	1.733	1.133
		600	9.156	3.699	1.934	1.490	1.081
		1000	10.562	3.509	1.834	1.458	1.064

We also plot the simulation results in the Figure 2.14 . In summary, the performance of RPS outperforms penalty estimators in situation both the larger sizes of the coefficient of multicollinearity and the larger values of p_2. We can also concluded that RFM dominates adaptive lasso and SCAD estimators when $n = 30$ and some p_2 values, indicated by the RMSE of aLASSO and SCAD being smaller than one.

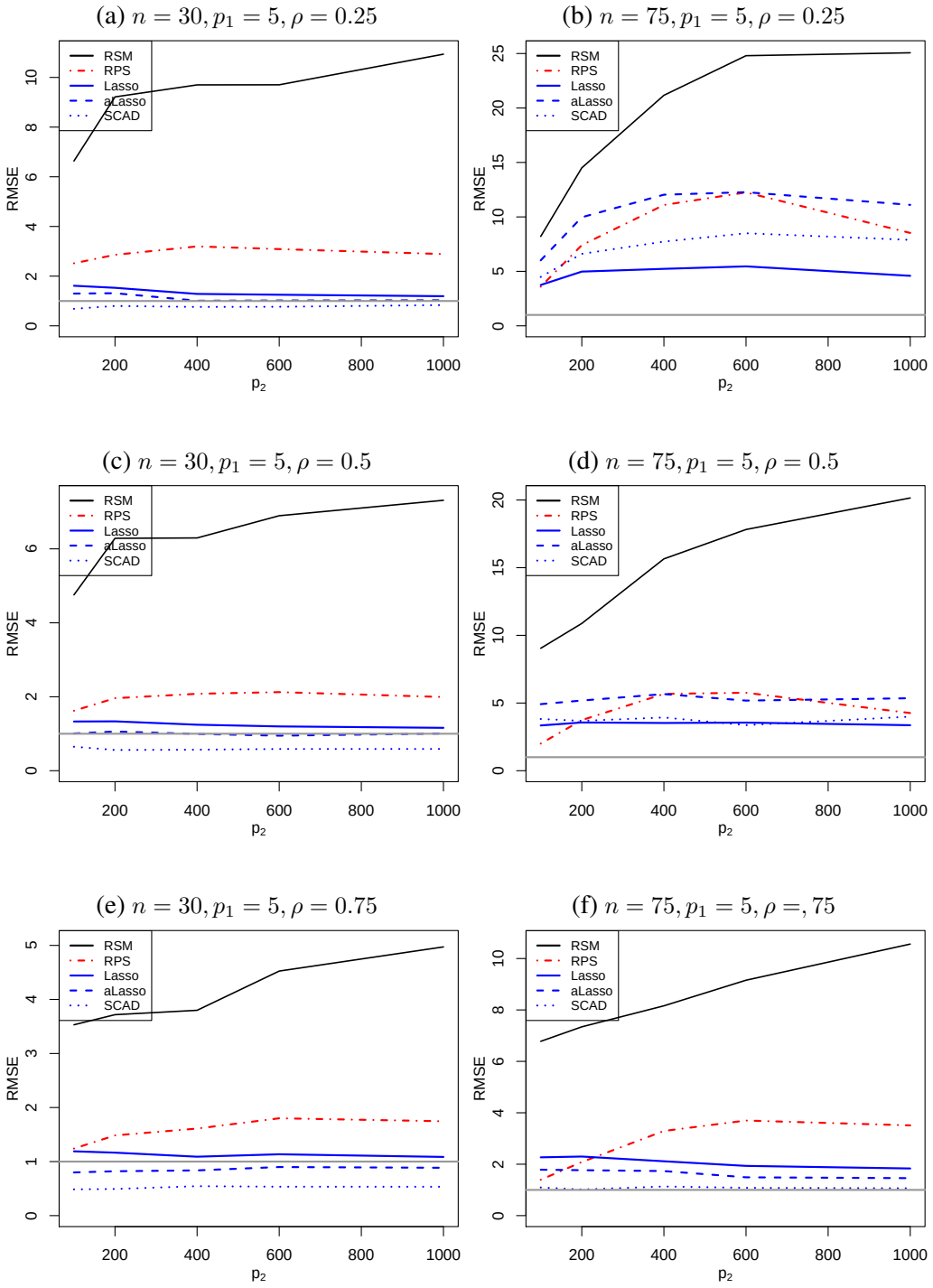


Figure 2.14: Relative efficiency of the estimators for $n = 30, 75$, $p_1 = 5$, $p_2 = 100, 200, 400, 600, 1000$ and $\rho = 0.25, 0.5, 0.75$.

2.13 Conclusion

In this chapter, we suggested pretest ridge regression estimator and shrinkage ridge regression estimators for linear models both in the existence of multicollinearity and in the situation high-dimensional data.

In the first half part of this chapter, we demonstrated shrinkage and pretest estimation based on ridge estimation in the context of a multiple linear regression. We also compared the performance of the suggested estimators with penalty estimators with three real data examples. To illustrate the methods, we used four different model selection criterion which are AIC, BIC, Lasso and BSS. For every data set, it has been obtained restricted ridge, pretest ridge, shrinkage ridge and positive shrinkage ridge estimators by using these model selection methods. Based on repeated cross validation estimate of the error rates show that the sub model ridge regression estimator has the smallest average prediction errors. Following this, among the suggested estimators, RPS estimator has superior risk performance compared to the full model ridge regression estimator and penalty estimators. Although the real data examples considered, this do not tell us how sensitive are the prediction errors under model misspecification. Further, in this chapter, we conduct Monte Carlo simulation to study the behaviour of proposed estimators under varying Δ^*, different sizes of the nuisance subsets and different magnitude of multicollinearity. For the purpose of comparison, we use the relative mean square error for all estimators with respect to full model estimator. Not surprisingly, the sub model ridge regression estimators outshines all other estimators as $\Delta^* = 0$. However, when Δ^* is further away from zero, the RMSE of RSM decrease and goes zero gradually. We also concluded from simulation results that the PRS outperforms RFM, RPT and PS when the nuisance subset is large. We also compared the suggested estimators with penalty estimators such as lasso, adaptive lasso and SCAD in the context of a multiple linear regression model. In summary, when $\Delta^* = 0$, pretest and shrinkage ridge regression estimators outshines penalty estimators. We also noticed that the performance of pretest and shrinkage ridge regression estimators increase compared to penalty while the magnitude of multicollinearity is increase. Asymptotic risk properties of the suggested estimators have been showed and their dominance over classical estimators demonstrated.

In the second half part of this chapter, we suggested shrinkage ridge regression estimators for high-dimensional data in the context of a multiple linear regression. As we did in the first half part of this chapter, for comparison, we applied two real data application. After that, we also conducted a Monte Carlo simulation. Not surprisingly, when $\Delta^* = 0$, $\widehat{\boldsymbol{\beta}}_1^{RSM}$ has the biggest RMSE. However, when Δ^* larger than zero, RSM loses its efficiency quickly, whereas RPS is less affected. Also, RPS outperform when the number of nuisance covariates gets extremely large relative to the sample size. We also compared shrinkage ridge regression estimator with penalty estimators. As a result of the simulation results, the RMSE of RSM increase when the number of nuisance covariates are large. As the size of multicollinearity increase, RPS dominates all penalty estimators.

3 PENALTY and NON-PENALTY ESTIMATION STRATEGIES in PARTIALLY LINEAR MODELS

3.1 Introduction

The models which combines both parametric and nonparametric models are called semiparametric models. In this chapter, for these models, we present non-penalty estimations based on ridge regression, and compare their performance with penalty estimators.

3.2 Organization of the Chapter

We demonstrate non-penalty estimation strategies when the number of experimental units is larger than the number of covariates ($n > p$) in partially linear models. To this end, we present full model estimation by using ridge regression with smoothing spline perspective, and sub-model estimation similarly in the following two subsection. Asymptotic properties of ridge regression full model and sub-model estimators, pretest ridge regression, shrinkage ridge regression and positive shrinkage ridge regression estimators are obtained in section 3.9. Monte Carlo simulation studies are obtained to compare non-penalty estimators with penalty estimators in section 3.10.

In section 3.11, we also suggest test statistic, and present penalty estimators when the number of experimental units is smaller than the number of covariates ($n < p$). In Section 3.12, performance of penalty and non-penalty is compared using a Monte Carlo experiment.

Let us consider the following partially linear model (PLM):

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + f(t_i) + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{3.1}$$

where $y_i's$ are observations, $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})^\top$ is known design points, $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^\top$ is an unknown $p-$dimensional vector of regression coefficients, $t_i \in [0, 1]$ are non-stochastic knot points of an explanatory variable t, $f(\cdot) \in C^2[0, 1]$ is an unknown smooth function, $\varepsilon_i's$ are unobservable random errors assumed to be iid $N(0, \sigma^2)$ distributed and the superscript $(^\top)$ denotes the transpose of a vector or matrix.

The model (3.1) also called a semiparametric model, and in vector–matrix form is written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{f} + \boldsymbol{\varepsilon}, \tag{3.2}$$

where $\boldsymbol{y} = (y_1, y_2, ..., y_n)^\top$, $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n)^\top$, $\boldsymbol{f} = (f(t_1), f(t_2), ..., f(t_n))^\top$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^\top$ is random vector with $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $Var(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$.

Model (3.1) has been widely used in various applications due to they allows easier interpretation of the effect of each variables and are preferable to a completely nonparametric model since the well-known reason "curse of dimensionality". In addition, the models are much more flexible than the standard linear model since they combine both parametric and nonparametric components.

Model (3.1) was the first applied by (Engle et al., 1986) in analyzing the relationship between weather and electricity sales. (Liang, 2006) studied this model by making numerical comparisons. The most popular approach for the partially linear models is based on the fact that the cubic spline is a linear estimator for the nonparametric regression problem. Hence, the nonparametric procedure can be naturally extended to handle the partially linear models. Among the most important approaches are smoothing splines applied by (Eubank et al., 1988), (Green et al., 1985), (Heckman, 1986), (Wahba, 1990), (Green and Silverman, 1994) and (Schimek, 2000), the smoothing kernel methods adopted by (Robinson, 1988; Speckman, 1988), the piecewise polynomial method discussed by (Chen, 1988), and local linear smoothing method examined by (Hamilton and Truong, 1997). In this paper, we consider the smoothing splines method adopted by (Green and Silverman, 1994) based on (Green et al., 1985).

For partially linear models, (Ahmed et al., 2007) considered a profile least squares approach based on using kernel estimates of $f(\cdot)$ to construct absolute penalty, shrinkage, and pretest estimators of $\boldsymbol{\beta}$ in the case where $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top$. Similarly, for partially linear models, the suitability of estimating the nonparametric component based on the B-spline basis function is explored by (Raheem et al., 2012).

In most application of regression analysis, the column vectors of the design matrix $\boldsymbol{X}$ in $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ linear model are highly intercorrelated is referred to as multicollinearity. If matrix $\boldsymbol{X}^\top \boldsymbol{X}$ is ill-conditioned, the precision of the least

squares estimators may be very poor, i.e., their variance may be very large. To overcome this, many researchers suggest different methods. Among them, in 1970s, ridge regression which is suggested by (Hoerl and Kennard, 1970a) is one of the most popular. Since then, many new versions of this method have been studied to extend Hoerl's and Kennard's original estimator.

For semiparametric models, (Roozbeh et al., 2011a) introduce a semiparametric ridge regression estimator for the vector-parameter which is also assumed that some additional artificial linear restrictions are imposed to the whole parameter space. (Roozbeh et al., 2011b) defined a generalized restricted difference-based ridge estimator for the vector parameter in a partial linear model when the errors are dependent. (Tabakan and Akdeniz, 2010) introduced a new difference-based ridge regression estimator of the regression parameters $\boldsymbol{\beta}$ in the partial linear model. (Roozbeh et al., 2012) propose semiparametric ridge and non ridge type estimators combining the restricted least squares methods in the model under study, also they used kernel smoothing and cross validation methods for estimating the nonparametric function. Lastly, (Roozbeh and Arashi, 2013) defined semiparametric ridge and non-ridge type estimators in a restricted manifold.

In this paper, we introduce estimations techniques based on ridge regression to be followed when the matrix $\boldsymbol{X}^\top \boldsymbol{X}$ appears to be ill-conditioned in the partially linear model using smoothing splines. Also, we consider that the coefficients $\boldsymbol{\beta}$ can be partitioned as $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ where $\boldsymbol{\beta}_1$ is the coefficient vector for main effects, and $\boldsymbol{\beta}_2$ is the vector for nuisance effects. We are essentially interested in the estimation of $\boldsymbol{\beta}_1$ when it is reasonable that $\boldsymbol{\beta}_2$ is close to zero. We suggest pretest ridge regression, shrinkage ridge regression and positive shrinkage ridge regression estimators for semiparametric models. We obtain results of a simulation study that includes a comparison of relative efficiency of proposed estimators with respect to unrestricted ridge regression estimator. Moreover, the asymptotic properties of the proposed estimators are obtained.

3.3 Full Model Estimation

Generally, the back-fitting algorithm is considered for the estimation of the model (3.2). In this thesis, we consider Speckman approach based on penalized

residual sum of squares method for estimation purpose. We estimate $\boldsymbol{\beta}$ and $\boldsymbol{f}$ by minimizing the following penalized sum of squares equation

$$SS(\boldsymbol{\beta}, f) = \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^{\top}\boldsymbol{\beta} + f(t_i) + \varepsilon_i\right)^2 + \lambda \int_a^b \left(f''(t)\right)^2 dt \quad (3.3)$$

$$= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{f})^{\top} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{f}) + \lambda \boldsymbol{f}^{\top}\boldsymbol{K}\boldsymbol{f}, \quad (3.4)$$

where $f \in C^2[a, b]$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{x}_i$ is the row of the matrix of $\boldsymbol{X}$ and $\boldsymbol{K}$ is positive definite penalty matrix with solution

$$\widehat{\boldsymbol{f}} = \boldsymbol{S}_{\lambda}\boldsymbol{y},$$

where $\boldsymbol{S}_{\lambda} = (\boldsymbol{I} - \lambda\boldsymbol{K})^{-1}$ is a well-known positive-definite (symmetrical) smoother matrix which depends on fixed smoothing parameter $\lambda > 0$, and the knot points $t_1, t_2, ..., t_n$. More specifically, $\boldsymbol{S}_{\lambda}$ transforms the vector of observations $\boldsymbol{y} = (y_1, y_2, ..., y_n)^{\top}$ in a model $(y_i = f(t_i) + \varepsilon_i)$ with $\boldsymbol{\beta} = \boldsymbol{0}$ into fitted values $\widehat{\boldsymbol{y}} = \boldsymbol{S}_{\lambda}\boldsymbol{y}$. Function $\widehat{f_{\lambda}}$, the estimator of function f, is obtained by cubic spline interpolation that rests on condition $\widehat{f}(t_i) = \left(\widehat{f}\right)_i, i = 1, 2, ..., n$. To gain better perspective on smoothing spline, (Eubank, 1999; Schimek, 2000; Green and Silverman, 1994; Wahba, 1990) state studied opinions.

The penalty $\boldsymbol{K}$ matrix in (3.4) is obtained by means of the knot points, and defied as following way:

$$\boldsymbol{K} = \boldsymbol{U}^{\top}\boldsymbol{R}^{-1}\boldsymbol{U},$$

where $h_j = t_{j+1} - t_j$, $j = 1, 2, ..., n-1$, $\boldsymbol{U}$ is tridiagonal $(n-2) \times n$ matrix with $\boldsymbol{U}_{jj} = 1/h_j$, $\boldsymbol{U}_{j,j+1} = (1/h_j + 1/h_{j+1})$, $\boldsymbol{U}_{j,j+2} = 1/h_{j+1}$ and $\boldsymbol{R}$ is symmetric tridiagonal matrix of order $(n-2)$ with $\boldsymbol{R}_{j-1,j} = \boldsymbol{R}_{j,j-1} = h_j/6$ and $\boldsymbol{R}_{jj} = (h_j + h_{j+1})/3$.

The first term in equation (3.3) denotes the residual sum of the squares and it penalizes the lack of fit. The second term which is weighted by λ denotes the roughness penalty and it imposes a penalty on roughness. In other words, it penalizes the curvature of the function f. The λ in (3.3) is recognized to be the smoothing parameter. As λ varies from 0 to $+\infty$, the solution varies from interpolation to a linear model. As $\lambda \to +\infty$, the roughness penalty dominates in (3.3) and the spline es-

timate is compelled to be a constant. As $\lambda \to 0$, the roughness penalty disappears in (3.3) and the spline estimation interpolates the data. Thus, the smoothing parameter λ plays a key role in controlling the trade-off between the goodness of fit and smoothness of the estimate .

The resulting estimator is called as partial spline (see Wahba, 1990). On the other hand, equation (3.3) is also known as the roughness penalty approach (Green and Silverman, 1994). This estimation concept is based on iterative solution of the normal equations. They propose to use back-fitting algorithm. (Rice, 1986) indicated that partial spline estimator is generally biased for the optimal choice when the components $\boldsymbol{X}$ of on t . This asymptotic bias can be larger than the standard error. Applying results due to (Speckman, 1988) this bias can be substantially reduced. In the following section, we present full model semiparametric estimation based on ridge regression.

3.3.1 Full model semiparametric ridge strategies

For a prespecified value of λ the corresponding estimators $\boldsymbol{\beta}$ and $\boldsymbol{f}$ for based on model (3.2) can be obtained by

$$\widehat{\boldsymbol{\beta}} = \left(\tilde{\boldsymbol{X}}^{\top}\tilde{\boldsymbol{X}}\right)^{-1}\tilde{\boldsymbol{X}}^{\top}\tilde{\boldsymbol{y}} \text{ and } \widehat{\boldsymbol{f}} = \boldsymbol{S}_{\lambda}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right),$$

where $\tilde{\boldsymbol{X}} = (\boldsymbol{I} - \boldsymbol{S}_{\lambda})\,\boldsymbol{X}$ and $\tilde{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{S}_{\lambda})\,\boldsymbol{y}$, respectively.

By multiplying both sides of model 3.2 with $(\boldsymbol{I} - \boldsymbol{S}_{\lambda})$, it is obtained by

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}, \tag{3.5}$$

where $\tilde{\boldsymbol{f}} = (\boldsymbol{I} - \boldsymbol{S}_{\lambda})\,\boldsymbol{f}$, $\tilde{\boldsymbol{\varepsilon}} = \tilde{\boldsymbol{f}} + \boldsymbol{\varepsilon}^{*}$ and $\boldsymbol{\varepsilon}^{*} = (\boldsymbol{I} - \boldsymbol{S}_{\lambda})\,\boldsymbol{\varepsilon}$.

Therefore, model (3.5) is transformed into an optimal problem to estimate semiparametric estimator. We now consider model (3.5) with ridge penalty to estimate semiparametric ridge estimator. We formulate this with the following equation 3.6,

$$\arg\min_{\boldsymbol{\beta}} \left(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right)^{\top}\left(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right) + \lambda^{R}\boldsymbol{\beta}^{\top}\boldsymbol{\beta}. \tag{3.6}$$

By solving 3.6, we get full model semiparametric ridge regression estimator of

$\boldsymbol{\beta}_1$ as follows:

$$\widehat{\boldsymbol{\beta}}_1^{SRFM} = \left(\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}} + \lambda^R \boldsymbol{I}_p\right)^{-1} \tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{y}}. \tag{3.7}$$

3.4 Sub-Model Semiparametric Ridge Strategies

In fact, the coefficient parameter $\boldsymbol{\beta}$ can be regarded as a vector in p dimensions space. If $\boldsymbol{X}^\top \boldsymbol{X}$ is ill conditioned, to estimate the coefficients $\boldsymbol{\beta}$, one of best methods is to use additional conditions to restrict the coefficients in some directions of p dimensions space

In this thesis, we consider a restriction on the parameters $\boldsymbol{R\beta} = \boldsymbol{r}$, and adding ridge penalty function on model (3.1),

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + f\left(t_i\right) + \varepsilon_i \quad \text{subject to } \boldsymbol{\beta}^\top \boldsymbol{\beta} \leq \phi^2 \text{ and } \boldsymbol{R\beta} = \boldsymbol{r}, \tag{3.8}$$

where $\boldsymbol{R}$ is an $s \times p$ restriction matrix, and $\boldsymbol{r}$ is an $s \times 1$ vector of constants. In this study, we let $\boldsymbol{X} = (\boldsymbol{X_1}, \boldsymbol{X_2})$, where $\boldsymbol{X_1}$ is an $n \times p_1$ sub-matrix containing the regressors of interest and $\boldsymbol{X_2}$ is an $n \times p_2$ sub-matrix that may or may not be relevant in the analysis of the main regressors. Similarly, $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top$ be the vector of parameters, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have dimensions p_1 and p_2, respectively, with $p_1 + p_2 = p$, $p_i \geq 0$ for $i = 1, 2$. We are constitutively interested in the estimation of $\boldsymbol{\beta}_1$ when it is suspected that $\boldsymbol{\beta}_2$ is close to $\mathbf{0}$. As a special case, we consider $\boldsymbol{R} = [\mathbf{0}, \boldsymbol{I}]$, and $\boldsymbol{r} = \mathbf{0}$, where $\mathbf{0}$ is a $p_2 \times p_1$ matrix of zeroes and $\boldsymbol{I}$ is the identity matrix. Thus, we have

$$\boldsymbol{\beta}_2 = \mathbf{0}. \tag{3.9}$$

Let $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ be the semiparametric unrestricted or full model ridge estimator of $\boldsymbol{\beta}_1$.

Now, the semiparametric full model ridge estimator $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ of $\boldsymbol{\beta}_1$ is

$$\widehat{\boldsymbol{\beta}}_1^{SRFM} = \left(\tilde{\boldsymbol{X}}_1^\top \tilde{\boldsymbol{M}}_2^R \tilde{\boldsymbol{X}}_1 + \lambda^R \boldsymbol{I}_{p_1}\right)^{-1} \tilde{\boldsymbol{X}}_1^\top \tilde{\boldsymbol{M}}_2^R \tilde{\boldsymbol{y}},$$

where $\tilde{\boldsymbol{M}}_2^R = \boldsymbol{I}_n - \tilde{\boldsymbol{X}}_2 \left(\tilde{\boldsymbol{X}}_2^\top \tilde{\boldsymbol{X}}_2 + \lambda^R \boldsymbol{I}_{p_2}\right)^{-1} \tilde{\boldsymbol{X}}_2^\top$ and $\tilde{\boldsymbol{X}}_i = \left(\boldsymbol{I}_n - \boldsymbol{S}_\lambda\right) \boldsymbol{X}_i$, $i = 1, 2$.

On the other hand, if $\boldsymbol{\beta}_2 = \mathbf{0}$, then we have the following restricted partially

linear model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{f} \quad \text{subject to } \lambda^R\boldsymbol{\beta}_1^\top. \tag{3.10}$$

Let us denote $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ the semiparametric sub-model or restricted ridge estimator of $\boldsymbol{\beta}_1$ as defined subsequently.

Generally speaking, $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ performs better than $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ when $\boldsymbol{\beta}_2$ close to zero. However, for $\boldsymbol{\beta}_2$ away from the zero, $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ can be inefficient. But, the estimate $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ is unaffected for departure of $\boldsymbol{\beta}_2$ from zero. Thus, we have two estimators for model the model at hand.

Also, the semiparametric sub-model ridge estimator $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ of $\boldsymbol{\beta}_1$ has the form

$$\widehat{\boldsymbol{\beta}}_1^{SRSM} = \left(\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{X}}_1 + \lambda^R\boldsymbol{I}_{p_1}\right)^{-1}\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{y}}.$$

3.5 Test Statistics

We define test statistics for testing $H : \boldsymbol{\beta}_2 = \boldsymbol{0}$ as follows:

$$\mathcal{T}_n = \frac{n}{\widehat{\sigma}_{UE}^2}\left(\widehat{\boldsymbol{\beta}}_2^{FM}\right)^\top\left(\tilde{\boldsymbol{X}}_2^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{X}}_2\right)\left(\widehat{\boldsymbol{\beta}}_2^{FM}\right),$$

where

$$\widehat{\sigma}_{UE}^2 = \frac{1}{n}.\frac{\|(\boldsymbol{I}_n - \boldsymbol{H}_\lambda)\,\tilde{\boldsymbol{y}}\|^2}{\text{tr}\,(\boldsymbol{I}_n - \boldsymbol{H}_\lambda)^\top(\boldsymbol{I}_n - \boldsymbol{H}_\lambda)},$$

and

$$\widehat{\boldsymbol{\beta}}_2^{FM} = \left(\tilde{\boldsymbol{X}}_2^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{X}}_2\right)^{-1}\tilde{\boldsymbol{X}}_2^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{y}},$$

with $\tilde{\boldsymbol{M}}_1 = \boldsymbol{I}_n - \tilde{\boldsymbol{X}}_1\left(\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{X}}_1\right)^{-1}\tilde{\boldsymbol{X}}_1^\top$ and $\boldsymbol{H}_\lambda$ is called as smoother matrix for the model (3.1).

This matrix is given by

$$\begin{aligned}
\widehat{\boldsymbol{y}} &= \boldsymbol{X}\widehat{\boldsymbol{\beta}}_1^{SRFM} + \widehat{\boldsymbol{f}} \\
&= \boldsymbol{X}\left(\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}} + \lambda^R\boldsymbol{I}_p\right)^{-1}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{y}} + \boldsymbol{S}_\lambda\left(\boldsymbol{y} - \boldsymbol{X}\left(\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}} + \lambda^R\boldsymbol{I}_p\right)^{-1}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{y}}\right) \\
&= \boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{S}_\lambda\right)\left(\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}} + \lambda^R\boldsymbol{I}_p\right)^{-1}\tilde{\boldsymbol{X}}^\top\boldsymbol{y} \\
&\quad + \boldsymbol{S}_\lambda\left(\boldsymbol{y} - \boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{S}_\lambda\right)\left(\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}} + \lambda^R\boldsymbol{I}_p\right)^{-1}\tilde{\boldsymbol{X}}^\top\boldsymbol{y}\right) \\
&= \tilde{\boldsymbol{X}}\left(\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}} + \lambda^R\boldsymbol{I}_p\right)^{-1}\tilde{\boldsymbol{X}}^\top\boldsymbol{y} + \boldsymbol{S}_\lambda\left(\boldsymbol{y} - \tilde{\boldsymbol{X}}\left(\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}} + \lambda^R\boldsymbol{I}_p\right)^{-1}\tilde{\boldsymbol{X}}^\top\boldsymbol{y}\right) \\
&= \boldsymbol{H}\boldsymbol{y} + \boldsymbol{S}_\lambda\boldsymbol{y} - \boldsymbol{S}_\lambda\boldsymbol{H}\boldsymbol{y} \\
&= \left(\boldsymbol{S}_\lambda + \left(\boldsymbol{I}_n - \boldsymbol{S}_\lambda\right)\boldsymbol{H}\right)\boldsymbol{y} \\
&= \boldsymbol{H}_\lambda\boldsymbol{y},
\end{aligned}$$

where $\boldsymbol{H} = \tilde{\boldsymbol{X}}\left(\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}} + \lambda^R\boldsymbol{I}_p\right)^{-1}\tilde{\boldsymbol{X}}^\top$. Thus, the mentioned smoother matrix is

$$\boldsymbol{H}_\lambda = \left(\boldsymbol{S}_\lambda + \left(\boldsymbol{I}_n - \boldsymbol{S}_\lambda\right)\boldsymbol{H}\right). \tag{3.11}$$

Under H_0, the test statistic $\mathcal{T}_n$ follows chi-square distribution with p_2 degrees of freedom for large n values. Then, we can choose an $\alpha-$level critical value $c_{n,\alpha}$.

3.6 Pretest Estimation Strategies

The pretest test estimation which was introduced by (Bancroft, 1944). This estimator is a combination of $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ and $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ via an indicator function $I\left(\mathcal{T}_n \leq c_{n,\alpha}\right)$, where $\mathcal{T}_n$ is an appropriate test statistic to test $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_A : \boldsymbol{\beta}_2 \neq \mathbf{0}$. Moreover, $c_{n,\alpha}$ is an $\alpha-$level critical value using the distribution of $\mathcal{T}_n$.

The semiparametric ridge pretest estimator $\widehat{\boldsymbol{\beta}}_1^{SRPT}$ of $\boldsymbol{\beta}_1$ defined by

$$\widehat{\boldsymbol{\beta}}_1^{SRPT} = \widehat{\boldsymbol{\beta}}_1^{SRFM} - \left(\widehat{\boldsymbol{\beta}}_1^{SRFM} - \widehat{\boldsymbol{\beta}}_1^{SRSM}\right) I\left(\mathcal{T}_n \leq c_{n,\alpha}\right).$$

Hence, $\widehat{\boldsymbol{\beta}}_1^{SRPT}$ choose $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ when H_0 is tenable; otherwise, $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ is chosen.

3.7 Shrinkage Estimation Strategies

Shrinkage estimator for a partial linear model was introduced by (Ahmed et al., 2007). This shrinkage estimator $\widehat{\boldsymbol{\beta}}_1^{SRS}$ is a smooth function of the test statistic.

The semiparametric ridge shrinkage or stein-type estimator of $\boldsymbol{\beta}_1$ defined by

$$\widehat{\boldsymbol{\beta}}_1^{SRS} = \widehat{\boldsymbol{\beta}}_1^{SRSM} + \left(\widehat{\boldsymbol{\beta}}_1^{SRFM} - \widehat{\boldsymbol{\beta}}_1^{SRSM}\right)\left(1 - (p_2 - 2)\mathcal{T}_n^{-1}\right), p_2 \geq 3.$$

The estimator $\widehat{\boldsymbol{\beta}}_1^{SRS}$ is general form of the Stein-rule family of estimators.

The positive part of the semiparametric ridge shrinkage estimator of $\boldsymbol{\beta}_1$ defined by

$$\widehat{\boldsymbol{\beta}}_1^{SRPS} = \widehat{\boldsymbol{\beta}}_1^{SRSM} + \left(\widehat{\boldsymbol{\beta}}_1^{SRFM} - \widehat{\boldsymbol{\beta}}_1^{SRSM}\right)\left(1 - (p_2 - 2)\mathcal{T}_n^{-1}\right)^+,$$

where $z^+ = max\,(0, z)$

3.8 Semiparametric Penalty Strategy

In this subsection, we propose the semiparamtric penalty estimators by using the smoothing spline method. For a given penalty function $\pi\,(\cdot)$, and regularization parameter λ, the general form Semiparametric Penalty Estimators can be written as

$$\sum\nolimits_{i=1}^{n}\left(\tilde{y}_i - \tilde{\boldsymbol{x}}_i^{\top}\boldsymbol{\beta}\right)^2 + \lambda\pi\,(\cdot),$$

where $\tilde{y}_i$ is the ith observation of $\tilde{\boldsymbol{y}}$, $\tilde{\boldsymbol{x}}_i^{\top}$ is the ith row of $\tilde{\boldsymbol{X}}$ and $\pi\,(\cdot) = \sum_{i=1}^{p}|\beta_i|^{\iota}$, $\iota > 0$.

if $\iota = 2$, then the semiparametric ridge regression estimator can be written

$$\widehat{\boldsymbol{\beta}}^{ridge} = \arg\min_{\boldsymbol{\beta}}\left\{\sum\nolimits_{i=1}^{n}\left(\tilde{y}_i - \tilde{\boldsymbol{x}}_i^{\top}\boldsymbol{\beta}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|^2\right\}.$$

For the special case when $\iota = 1$ is related to semiparametric Least Absolute Shrinkage and Selection Operator for partially linear models, that is,

$$\widehat{\boldsymbol{\beta}}^{lasso} = \arg\min_{\boldsymbol{\beta}}\left\{\sum\nolimits_{i=1}^{n}\left(\tilde{y}_i - \tilde{\boldsymbol{x}}_i^{\top}\boldsymbol{\beta}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\}.$$

For partially linear models, adaptive lasso estimator $\boldsymbol{\beta}^{alasso}$ is defined as

$$\widehat{\boldsymbol{\beta}}^{aLasso} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(\tilde{y}_i - \tilde{\boldsymbol{x}}_i^{\top} \boldsymbol{\beta} \right)^2 + \lambda \sum_{j=1}^{p} \widehat{\zeta}_j \left| \beta_j \right| \right\}, \tag{3.12}$$

where the wight function is

$$\widehat{\boldsymbol{\zeta}} = \frac{1}{|\boldsymbol{\beta}^*|^{\iota}};\ \iota > 0,$$

where $\boldsymbol{\beta}^*$ is also a consistent estimator of $\boldsymbol{\beta}$.

For partially linear models, SCAD estimator $\widehat{\boldsymbol{\beta}}^{SCAD}$ is defined as

$$\widehat{\boldsymbol{\beta}}^{SCAD} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(\tilde{y}_i - \tilde{\boldsymbol{x}}_i^{\top} \boldsymbol{\beta} \right)^2 + \lambda \sum_{j=1}^{p} J_{\alpha,\lambda} \left\| \beta_j \right\|_1 \right\},$$

where

$$J_{\alpha,\lambda}(x) = \lambda \left\{ I\left(|x| \leq \lambda \right) + \frac{(\alpha\lambda - |x|)_+}{(\alpha - 1)\lambda} I\left(|x| > \lambda \right) \right\},\ \ x \geq 0. \tag{3.13}$$

3.9 Asymptotic Analysis

In this subsection, we define expressions for asymptotic distributional bias(ADB), quadratic distributional bias(AQDB), covariance matrices and distributional risks(ADR) of the proposed estimators. Our main aim is the performance of these estimators when $\boldsymbol{\beta}_2$ is close to $\mathbf{0}$, and we consider a sequence $\{K_n\}$ given by

$$K_n : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{2(n)} = \frac{\boldsymbol{w}}{\sqrt{n}}, \quad \boldsymbol{w} = (w_1, w_2, ..., w_{p_2})^{\top} \in \mathbb{R}^{p_2},$$

thus, for $\boldsymbol{w} = \mathbf{0}$ it reduces to H_0.

To study the asymptotic quadratic risks of $\widehat{\boldsymbol{\beta}}_1^{SRFM}, \widehat{\boldsymbol{\beta}}_1^{SRSM}, \widehat{\boldsymbol{\beta}}_1^{SRPT}, \widehat{\boldsymbol{\beta}}_1^{SRS}$ and $\widehat{\boldsymbol{\beta}}_1^{SRPS}$, we may define a quadratic loss function using a positive definite matrix (p.d.m) $\boldsymbol{W}$, by

$$L\left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1 \right) = n \left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1 \right)^{\top} \boldsymbol{W} \left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1 \right),$$

where $\boldsymbol{\beta}_1^*$ is anyone of suggested estimators. Now, under $\{K_n\}$, we can write the

asymptotic distribution function of $\boldsymbol{\beta}_1^*$ as

$$F(\boldsymbol{x}) = \lim_{n\to\infty} P\left(\sqrt{n}\left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1\right) \leq \boldsymbol{x} | K_n\right),$$

where $F(\boldsymbol{x})$ is non degenerate. Then ADR of $\boldsymbol{\beta}_1^*$ is defined as follows:

$$R(\boldsymbol{\beta}_1^*) = \text{tr}\left(\boldsymbol{W}\int_{\mathbb{R}^{p_1}}\int \boldsymbol{x}\boldsymbol{x}^\top dF(\boldsymbol{x})\right) = \text{tr}(\boldsymbol{W}\boldsymbol{V}),$$

where $\boldsymbol{V}$ is the dispersion matrix for the distribution $F(\boldsymbol{x})$.

Asymptotic distributional bias of an estimator $\boldsymbol{\beta}_1^*$ is defined as

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

First, let $H_v(x, \Delta)$ be the cumulative distribution function of the non-central chi-squared distribution with non-centrality parameter Δ and v degree of freedom,then

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

Assumption 9. We make the following two regularity conditions

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

(ii) $\tilde{\boldsymbol{Q}}_n = \frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}} \to \tilde{\boldsymbol{Q}}$.

We assume that the sub-matrices of $\tilde{\boldsymbol{Q}}$ satisfy the following equations

$$\lim_{n\to\infty}\frac{1}{n}\left(\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{X}}_1 + \lambda^R\boldsymbol{I}_{p_1}\right) = \lim_{n\to\infty}\frac{1}{n}\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{X}}_1 = \tilde{\boldsymbol{Q}}_{11}, \tag{3.14}$$

$$\lim_{n\to\infty}\frac{1}{n}\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{X}}_2 = \tilde{\boldsymbol{Q}}_{12}, \tag{3.15}$$

$$\lim_{n\to\infty}\frac{1}{n}\tilde{\boldsymbol{X}}_2^\top\tilde{\boldsymbol{X}}_1 = \tilde{\boldsymbol{Q}}_{21}, \tag{3.16}$$

$$\lim_{n\to\infty}\frac{1}{n}\left(\tilde{\boldsymbol{X}}_2^\top\tilde{\boldsymbol{X}}_2 + \lambda^R\boldsymbol{I}_{p_2}\right) = \lim_{n\to\infty}\frac{1}{n}\tilde{\boldsymbol{X}}_2^\top\tilde{\boldsymbol{X}}_2 = \tilde{\boldsymbol{Q}}_{22}, \tag{3.17}$$

where $\tilde{\boldsymbol{Q}} = \begin{pmatrix} \tilde{\boldsymbol{Q}}_{11} & \tilde{\boldsymbol{Q}}_{12} \\ \tilde{\boldsymbol{Q}}_{21} & \tilde{\boldsymbol{Q}}_{22} \end{pmatrix}$.

Theorem 10. If $\lambda^R/\sqrt{n} \to \lambda_0 \geq 0$ and $\tilde{\boldsymbol{Q}}$ is non-singular, then

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{RFM} - \boldsymbol{\beta}\right) \xrightarrow{\mathscr{D}} \left(-\lambda_0^{-1}\tilde{\boldsymbol{Q}}^{-1}\boldsymbol{\beta}, \sigma^2\tilde{\boldsymbol{Q}}^{-1}\right).$$

Proof. It can be proof this theorem by following the same way of proof of (Knight and Fu, 2000). □

Now, under the assumed regularity conditions, theorems and equations above, and under $\{K_n\}$, the $ADBs$ of the estimators are given as follows:

Theorem 11.

$$\begin{aligned}
ADB\left(\widehat{\boldsymbol{\beta}}_1^{SRFM}\right) &= -\boldsymbol{\eta}_{11.2}, \\
ADB\left(\widehat{\boldsymbol{\beta}}_1^{SRSM}\right) &= -\boldsymbol{\xi}, \\
ADB\left(\widehat{\boldsymbol{\beta}}_1^{SRPT}\right) &= -\boldsymbol{\eta}_{11.2} - \tilde{\boldsymbol{\delta}} H_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right), \\
ADB\left(\widehat{\boldsymbol{\beta}}_1^{SRS}\right) &= -\boldsymbol{\eta}_{11.2} - (p_2-2)\tilde{\boldsymbol{\delta}} E\left(\chi^{-2}_{p_2+2}(\Delta)\right), \\
ADB\left(\widehat{\boldsymbol{\beta}}_1^{SRPS}\right) &= -\boldsymbol{\eta}_{11.2} - \tilde{\boldsymbol{\delta}} H_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right), \\
&\quad -(p_2-2)\tilde{\boldsymbol{\delta}} E\left\{\chi^{-2}_{p_2+2}(\Delta) I\left(\chi^2_{p_2+2}(\Delta) > p_2-2\right)\right\},
\end{aligned}$$

where $\Delta = \left(\boldsymbol{w}^{\intercal}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\boldsymbol{w}\right)\sigma^{-2}$, $\tilde{\boldsymbol{Q}}_{22.1} = \tilde{\boldsymbol{Q}}_{22} - \tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}$, $\boldsymbol{\eta} = -\lambda_0^{-1}\tilde{\boldsymbol{Q}}^{-1}\boldsymbol{\beta}$, $\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{pmatrix}$, $\boldsymbol{\eta}_{11.2} = \boldsymbol{\eta}_1 - \tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22}^{-1}((\boldsymbol{\beta}_2 - \boldsymbol{w}) - \boldsymbol{\eta}_2)$, $\boldsymbol{\xi} = \boldsymbol{\eta}_{11.2} + \tilde{\boldsymbol{\delta}}$ and $\tilde{\boldsymbol{\delta}} = \tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\boldsymbol{\omega}$.

Proof. See Appendix B. □

Since the bias expressions for all the estimators are not in scaler form, we also convert them to quadratic forms. Thus, we define $AQDB$ of an estimator $\boldsymbol{\beta}_1^*$ is defined as

$$AQDB\left(\boldsymbol{\beta}_1^*\right) = \left(ADB\left(\boldsymbol{\beta}_1^*\right)\right)^{\top}\tilde{\boldsymbol{Q}}_{11.2}\left(ADB\left(\boldsymbol{\beta}_1^*\right)\right), \tag{3.18}$$

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

Considering equation (3.18), we present the $AQDB$ of the estimators, under $\{K_n\}$, in the following theorem.

Theorem 12.

$$
\begin{aligned}
AQDB\left(\widehat{\boldsymbol{\beta}}_1^{SRFM}\right) &= \boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2}, \\
AQDB\left(\widehat{\boldsymbol{\beta}}_1^{SRSM}\right) &= \boldsymbol{\xi}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\xi}, \\
AQDB\left(\widehat{\boldsymbol{\beta}}_1^{SRPT}\right) &= \boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2} + \boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\tilde{\boldsymbol{\delta}}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&\quad +\tilde{\boldsymbol{\delta}}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&\quad +\tilde{\boldsymbol{\delta}}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\tilde{\boldsymbol{\delta}}H^2_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right), \\
AQDB\left(\widehat{\boldsymbol{\beta}}_1^{SRS}\right) &= \boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2} + (p_2-2)\boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\tilde{\boldsymbol{\delta}}E\left(\chi^{-2}_{p_2+2}(\Delta)\right) \\
&\quad +(p_2-2)\tilde{\boldsymbol{\delta}}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2}E\left(\chi^{-2}_{p_2+2}(\Delta)\right) \\
&\quad +(p_2-2)^2\tilde{\boldsymbol{\delta}}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\tilde{\boldsymbol{\delta}}\left(E\left(\chi^{-2}_{p_2+2}(\Delta)\right)\right)^2, \\
AQDB\left(\widehat{\boldsymbol{\beta}}_1^{SRPS}\right) &= \boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2} + \left(\tilde{\boldsymbol{\delta}}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2} + \boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\tilde{\boldsymbol{\delta}}\right) \\
&\quad \cdot\left[H_{p_2+2}\left((p_2-2);\Delta\right)\right. \\
&\quad \left.+(p_2-2)E\left\{\chi^{-2}_{p_2+2}(\Delta)\,I\left(\chi^{-2}_{p_2+2}(\Delta)>p_2-2\right)\right\}\right] \\
&\quad +\tilde{\boldsymbol{\delta}}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\tilde{\boldsymbol{\delta}}\left[H_{p_2+2}\left((p_2-2);\Delta\right)\right. \\
&\quad \left.+(p_2-2)E\left\{\chi^{-2}_{p_2+2}(\Delta)\,I\left(\chi^{-2}_{p_2+2}(\Delta)>p_2-2\right)\right\}\right]^2.
\end{aligned}
$$

Proof. See Appendix B. □

In the results of the above Theorem, for $\tilde{\boldsymbol{Q}}_{12}=0$, because of $\tilde{\boldsymbol{\delta}}=0$, $\boldsymbol{\xi}=\boldsymbol{\eta}_{11.2}$ and $\tilde{\boldsymbol{Q}}_{11.2}=\tilde{\boldsymbol{Q}}_{11}$, all the $AQDBs$ reduce to common value $\boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{Q}}_{11}\boldsymbol{\eta}_{11.2}$ for all $\boldsymbol{\omega}$. Following this, if we suppose that $\tilde{\boldsymbol{Q}}_{12}\neq 0$, it is given following results

(i) The $AQDB$ of $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ is an unbounded function of $\boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2}$.

(ii) The $AQDB$ of $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ is an unbounded function of $\boldsymbol{\xi}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\xi}$.

(iii) The $AQDB$ of $\widehat{\boldsymbol{\beta}}_1^{SRPT}$ begins from $\boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2}$ at $\Delta=0$. For $\Delta>0$, it increases to a maximum and then decreases towards 0.

(iv) Similarly, the $AQDB$ of $\widehat{\boldsymbol{\beta}}_1^{SRS}$ starts from $\boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2}$ at $\Delta=0$, and it increases to a point and then decreases towards zero for non-zero Δ values because of $E\left(\chi^{-2}_{p_2+2}(\Delta)\right)$ being a non increasing log convex function of Δ. Lastly, for all Δ values, the behaviour of the $AQDB$ of $\widehat{\boldsymbol{\beta}}_1^{SRPS}$ is almost the same $\widehat{\boldsymbol{\beta}}_1^{SRS}$, but the quadratic bias curve of $\widehat{\boldsymbol{\beta}}_1^{SRPS}$ remains on below the curve of $\widehat{\boldsymbol{\beta}}_1^{SRS}$.

Under a sequence $\{K_n\}$, the asymptotic covariance matrices of the estimators are given by following theorem.

Theorem 13.

$$
\begin{aligned}
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{SRFM}\right) &= \sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1}+\boldsymbol{\eta}_{11.2}\boldsymbol{\eta}_{11.2}^{\top},\\
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{SRSM}\right) &= \sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1}+\boldsymbol{\xi}\boldsymbol{\xi}^{\top},\\
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{SRPT}\right) &= \sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1}+\boldsymbol{\eta}_{11.2}\boldsymbol{\eta}_{11.2}^{\top}+2\boldsymbol{\eta}_{11.2}^{\top}\tilde{\boldsymbol{\delta}}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad+\sigma^2\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad+\tilde{\boldsymbol{\delta}}\tilde{\boldsymbol{\delta}}^{\top}\left[2H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)-H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\right],\\
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{SRS}\right) &= \sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1}+\boldsymbol{\eta}_{11.2}\boldsymbol{\eta}_{11.2}^{\top}+2(p_2-2)\tilde{\boldsymbol{\delta}}\boldsymbol{\eta}_{11.2}^{\top}E\left(\chi^{-2}_{p_2+2}(\Delta)\right)\\
&\quad-(p_2-2)\sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\{2E\left(\chi^{-2}_{p_2+2}(\Delta)\right)\\
&\quad-(p_2-2)E\left(\chi^{-4}_{p_2+2}(\Delta)\right)\}\\
&\quad+(p_2-2)\tilde{\boldsymbol{\delta}}\tilde{\boldsymbol{\delta}}^{\top}\{2E\left(\chi^{-2}_{p_2+2}(\Delta)\right)\\
&\quad-2E\left(\chi^{-2}_{p_2+4}(\Delta)\right)-(p_2-2)E\left(\chi^{-4}_{p_2+4}(\Delta)\right)\},\\
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{SRPS}\right) &= \boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{SRS}\right)\\
&\quad-2\tilde{\boldsymbol{\delta}}\boldsymbol{\eta}_{11.2}^{\top}E\left(\left\{1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta)\right\}I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\\
&\quad+(p_2-2)\sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\\
&\quad\cdot\left[2E\left(\chi^{-2}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\right.\\
&\quad\left.-(p_2-2)E\left(\chi^{-4}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\right]\\
&\quad-\sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}H_{p_2+2}\left((p_2-2);\Delta\right)\\
&\quad+\tilde{\boldsymbol{\delta}}\tilde{\boldsymbol{\delta}}^{\top}\left[2H_{p_2+2}\left((p_2-2);\Delta\right)-H_{p_2+4}\left((p_2-2);\Delta\right)\right]\\
&\quad-(p_2-2)\tilde{\boldsymbol{\delta}}\tilde{\boldsymbol{\delta}}^{\top}\left[2E\left(\chi^{-2}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\right.\\
&\quad-2E\left(\chi^{-2}_{p_2+4}(\Delta)I\left(\chi^2_{p_2+4}(\Delta)\le p_2-2\right)\right)\\
&\quad\left.+(p_2-2)E\left(\chi^{-4}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\right].
\end{aligned}
$$

Proof. See Appendix B. □

Finally, we obtain the ADRs of the estimators under $\{K_n\}$ as follows:

Theorem 14.

$$
\begin{aligned}
R\left(\widehat{\boldsymbol{\beta}}_1^{SRFM}\right) &= \sigma^2 \text{tr}\left(\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11.2}^{-1}\right) + \boldsymbol{\eta}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\eta}_{11.2}, \\
R\left(\widehat{\boldsymbol{\beta}}_1^{SRSM}\right) &= \sigma^2 \text{tr}\left(\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\right) + \boldsymbol{\xi}^{\top}\boldsymbol{W}\boldsymbol{\xi}, \\
R\left(\widehat{\boldsymbol{\beta}}_1^{SRPT}\right) &= R\left(\widehat{\boldsymbol{\beta}}_1^{SRFM}\right) - 2\boldsymbol{\eta}_{11.2}^{\top}\boldsymbol{W}\tilde{\boldsymbol{\delta}}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&\quad -\sigma^2 \text{tr}\left(\boldsymbol{W}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\right) H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&\quad +\tilde{\boldsymbol{\delta}}^{\top}\boldsymbol{W}\tilde{\boldsymbol{\delta}}\left\{2H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) - H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\right\}, \\
R\left(\widehat{\boldsymbol{\beta}}_1^{SRS}\right) &= R\left(\widehat{\boldsymbol{\beta}}_1^{SRFM}\right) + 2(p_2-2)\boldsymbol{\eta}_{11.2}^{\top}\boldsymbol{W}\tilde{\boldsymbol{\delta}}E\left(\chi^{-2}_{p_2+2}(\Delta)\right) \\
&\quad -(p_2-2)\sigma^2 \text{tr}\left(\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\right)\left\{2E\left(\chi^{-2}_{p_2+2}(\Delta)\right)\right. \\
&\quad \left.-(p_2-2)E\left(\chi^{-4}_{p_2+2}(\Delta)\right)\right\} \\
&\quad +(p_2-2)\tilde{\boldsymbol{\delta}}^{\top}\boldsymbol{W}\tilde{\boldsymbol{\delta}}\{2E\left(\chi^{-2}_{p_2+2}(\Delta)\right) \\
&\quad -2E\left(\chi^{-2}_{p_2+4}(\Delta)\right) - (p_2-2)E\left(\chi^{-4}_{p_2+4}(\Delta)\right)\}, \\
R\left(\widehat{\boldsymbol{\beta}}_1^{SRPS}\right) &= R\left(\widehat{\boldsymbol{\beta}}_1^{SRS}\right) \\
&\quad -2\boldsymbol{\eta}_{11.2}^{\top}\boldsymbol{W}\tilde{\boldsymbol{\delta}}E\left(\left\{1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta)\right\}I\left(\chi^2_{p_2+2}(\Delta)\leq p_2-2\right)\right) \\
&\quad +(p_2-2)\sigma^2 \text{tr}\left(\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\right) \\
&\quad \cdot\left[2E\left(\chi^{-2}_{p_2+2}(\Delta)\,I\left(\chi^2_{p_2+2}(\Delta)\leq p_2-2\right)\right)\right. \\
&\quad \left.-(p_2-2)E\left(\chi^{-4}_{p_2+2}(\Delta)\,I\left(\chi^2_{p_2+2}(\Delta)\leq p_2-2\right)\right)\right] \\
&\quad -\sigma^2 \text{tr}\left(\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\right) H_{p_2+2}\left((p_2-2);\Delta\right) \\
&\quad +\tilde{\boldsymbol{\delta}}^{\top}\boldsymbol{W}\tilde{\boldsymbol{\delta}}\left[2H_{p_2+2}\left((p_2-2);\Delta\right) - H_{p_2+4}\left((p_2-2);\Delta\right)\right] \\
&\quad -(p_2-2)\tilde{\boldsymbol{\delta}}^{\top}\boldsymbol{W}\tilde{\boldsymbol{\delta}}\left[2E\left(\chi^{-2}_{p_2+2}(\Delta)\,I\left(\chi^2_{p_2+2}(\Delta)\leq p_2-2\right)\right)\right. \\
&\quad -2E\left(\chi^{-2}_{p_2+4}(\Delta)\,I\left(\chi^2_{p_2+4}(\Delta)\leq p_2-2\right)\right) \\
&\quad \left.+(p_2-2)E\left(\chi^{-4}_{p_2+2}(\Delta)\,I\left(\chi^2_{p_2+2}(\Delta)\leq p_2-2\right)\right)\right].
\end{aligned}
$$

Proof. See Appendix B. □

In the results of the above Theorem, if $\tilde{\boldsymbol{Q}}_{12} = 0$, then $\tilde{\boldsymbol{\delta}} = 0$, $\boldsymbol{\xi} = \boldsymbol{\eta}_{11.2}$ and $\tilde{\boldsymbol{Q}}_{11.2} = \tilde{\boldsymbol{Q}}_{11}$, all the $ADRs$ reduce to common value $\sigma^2 \text{tr}\left(\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\right) + \boldsymbol{\eta}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\eta}_{11.2}$ for all $\boldsymbol{\omega}$. On the other hand, if we suppose that $\tilde{\boldsymbol{Q}}_{12} \neq 0$, then it is given following results:

(i) As Δ moves away from 0, the risk of $R\left(\widehat{\boldsymbol{\beta}}_1^{SRSM}\right)$ becomes unbounded. Furthermore, the risk of $\widehat{\boldsymbol{\beta}}_1^{RPT}$ perform better than $\widehat{\boldsymbol{\beta}}_1^{RFM}$ for all values of $\Delta \geq 0$, that is $R\left(\widehat{\boldsymbol{\beta}}_1^{SRPT}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_1^{SRFM}\right)$.

(ii) Under $\{K_n\}$, for all $\boldsymbol{W}$ and $\boldsymbol{\omega}$, it can be shown by using the following equation that $R\left(\widehat{\boldsymbol{\beta}}_1^{SRS}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_1^{SRFM}\right)$.

$$\frac{\text{tr}\left(\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\right)}{ch_{max}\left(\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\right)} \geq \frac{p_2+2}{2},$$

where $ch_{max}(.)$ is the maximum characteristic root.

(iii) To compare $\widehat{\boldsymbol{\beta}}_1^{SRPS}$ and $\widehat{\boldsymbol{\beta}}_1^{SRS}$, we observe that the the risk of $\widehat{\boldsymbol{\beta}}_1^{SRPS}$ overshadows $\widehat{\boldsymbol{\beta}}_1^{SRS}$ for all the values of $\boldsymbol{\omega}$. Moreover, with result (ii), we have $R\left(\widehat{\boldsymbol{\beta}}_1^{SRPS}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_1^{SRS}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_1^{SRFM}\right)$.

To illustrate the properties of the theoretical results, we conducted a simulation study and reported in the next subsection, to compare the performance of the proposed estimators.

3.10 Simulation Studies

In this subsection, we consider a Monte Carlo simulation study designed to evaluate the quadratic risk performance of the proposed estimators introduced in this section. All calculations were carried out in (R Development Core Team , 2010). For the simulations we considered the partially linear model which was defined in the model (3.1), the model described here is

$$y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + ... + x_{pi}\beta_p + f\left(t_i\right) + \varepsilon_i\,,\ i = 1, 2, ..., n, \tag{3.19}$$

where $\boldsymbol{x}_i \sim N_p\left(\boldsymbol{0}, \boldsymbol{\Sigma}_x\right)$ and ε_i are i.i.d. $N\left(0, 1\right)$. We also define $\boldsymbol{\Sigma}_x$ that is positive definite covariance matrix. To generate $\boldsymbol{\Sigma}_x$, we define ρ which is coefficient of correlation $\boldsymbol{X}$. We use three values of correlation which are low($\rho = .25$), medium($\rho = .5$) and high($\rho = .75$). The condition number test (CNT) is performed to test the multicollinearity, which is defined as the ratio of the largest eigenvalue to the smallest eigenvalue of matrix $\boldsymbol{C}$. If CNT is larger than 30, then it implies the existence of multicollinearity in the data set. We also get $\alpha = 0.05$

and $t_i = (i - 0.5)/n$. Furthermore, we consider the hypothesis $H_0 : \beta_j = 0$, for $j = p_1 + 1, p_1 + 2, ..., p$, with $p = p_1 + p_2$. Hence, we partition the regression coefficients as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (\boldsymbol{\beta}_1, \mathbf{0})$ with $\boldsymbol{\beta}_1 = (1, 1, 1, 1, 1)$. In (3.19), we consider two different the nonparametric functions $f_1(t_i) = \sqrt{t_i(1 - t_i)} \sin\left(\frac{2.1\pi}{t_i + 0.05}\right)$ and $f_2(t_i) = 0.5 \sin(4\pi t_i)$ to generate response y_i.

We use Generalized Cross-Validation (GCV) which is is a modified form of the Cross-Validation (CV) to select the optimal λ^R value. GCV is a modified form of the CV which is a conventional method for choosing the smoothing parameter. The GCV score which is constructed by analogy to CV score can be obtained from the ordinary residuals by dividing by the factors $1 - (\boldsymbol{H}_\lambda)_{ii}$. The underlying design of GCV is to replace the factors $1 - (\boldsymbol{H}_\lambda)_{ii}$ in equation (3.11) with the average score $1 - n^{-1}\text{tr}(\boldsymbol{H}_\lambda)$. Thus, by summing the squared corrected residual and factor $\{1 - n^{-1}\text{tr}(\boldsymbol{H}_\lambda)\}^2$, by the analogy ordinary cross-validation, the GCV score function can be procured as follow (see Craven and Wahba, 1979; Wahba, 1990).

$$\text{GCV}\left(\lambda^R\right) = \frac{1}{n} \frac{\sum_{i=1}^{n} \left\{y_i - \widehat{f}(t_i)\right\}^2}{\left\{1 - n^{-1}\text{tr}(\boldsymbol{H}_\lambda)\right\}^2} = \frac{n \left\|(\boldsymbol{I}_n - \boldsymbol{H}_\lambda)\boldsymbol{y}\right\|^2}{\left\{\text{tr}(\boldsymbol{I}_n - \boldsymbol{H}_\lambda)\right\}^2}.$$

The parametric component in model (3.19) is $p-$dimensional whereas nonparametric component is one-dimensional. The number of simulations were varied in the beginning. Finally, each reliaziation was repeated 5000 times to obtain sensible results. For each realization, we calculated bias of the estimators. We define $\Delta^* = \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|$, where $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_1, \mathbf{0})$, and $\|\cdot\|$ is the Euclidean norm. In order to investigate of the behaviour of the estimators for $\Delta^* > 0$, further datasets were generated from those distributions under local alternative hypothesis.

The performance of an estimator $\boldsymbol{\beta}_1$ was evaluated by calculating its mean squared error criterion. We numerically calculated the relative MSE of $\widehat{\boldsymbol{\beta}}_1^{SRSM}$, $\widehat{\boldsymbol{\beta}}_1^{SRPT}$, $\widehat{\boldsymbol{\beta}}_1^{SRS}$ and $\widehat{\boldsymbol{\beta}}_1^{SRPS}$, with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$. Therelative mean square efficiency (RMSE) of the $\boldsymbol{\beta}_1^{\blacktriangledown}$ to the unrestricted least squares estimator $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ is indicated by

$$RMSE\left(\widehat{\boldsymbol{\beta}}_1^{SRFM} : \boldsymbol{\beta}_1^{\blacktriangledown}\right) = \frac{MSE\left(\widehat{\boldsymbol{\beta}}_1^{SRFM}\right)}{MSE\left(\boldsymbol{\beta}_1^{\blacktriangledown}\right)},$$

where $\boldsymbol{\beta}_1^{\blacktriangledown}$ is one of the proposed estimators. The amount by which a RMSE is larger

than one indicates the degree of superiority of the estimator $\boldsymbol{\beta}_1^{\blacktriangledown}$ over $\widehat{\boldsymbol{\beta}}_1^{SRFM}$. In this chapter, we consider low-dimensional and high-dimensional data. Our methods were applied to several times simulated data sets.

We present the simulation results in Tables 3.1–3.6.

Table 3.1: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ for $p_1 = 5$, $n = 50$, $p_2 = 10, 15, 20$ values when $\rho = 0.25$.

(n, p_2)	Δ^*	CNT	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPT}$	$\widehat{\boldsymbol{\beta}}_1^{SRS}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$
	0.00	31.850	0.924	1.442	1.350	1.216	1.340
	0.25	19.553	0.869	1.369	1.212	1.165	1.296
	0.50	29.829	0.920	1.223	0.989	1.263	1.271
	0.75	26.153	0.903	1.040	0.928	1.176	1.176
(50,10)	1.00	15.875	0.957	0.887	1.000	1.154	1.154
	1.25	33.807	0.875	0.618	1.000	1.097	1.097
	1.50	22.263	0.866	0.506	1.000	1.060	1.060
	2.00	17.311	0.851	0.329	1.000	1.032	1.032
	4.00	28.652	0.935	0.128	1.000	1.007	1.007
	0.00	49.959	0.892	1.602	1.280	1.356	1.486
	0.25	38.657	0.890	1.564	1.122	1.428	1.463
	0.50	41.441	0.880	1.420	1.032	1.345	1.398
	0.75	47.093	0.893	1.060	0.970	1.297	1.297
(50,15)	1.00	57.004	0.828	0.820	0.994	1.224	1.224
	1.25	45.117	0.798	0.586	1.000	1.139	1.139
	1.50	74.438	0.841	0.570	1.000	1.127	1.127
	2.00	38.320	0.860	0.390	1.000	1.051	1.051
	4.00	37.689	0.904	0.123	1.000	1.014	1.014
	0.00	64.850	0.846	1.862	1.467	1.642	1.746
	0.25	118.234	0.923	1.926	1.490	1.677	1.850
	0.50	82.994	0.784	1.684	1.089	1.695	1.704
	0.75	77.496	0.861	1.102	0.923	1.399	1.399
(50,20)	1.00	133.267	0.850	0.905	0.975	1.355	1.355
	1.25	65.301	0.767	0.840	0.990	1.284	1.284
	1.50	113.894	0.841	0.577	1.000	1.173	1.173
	2.00	99.294	0.770	0.437	1.000	1.107	1.107
	4.00	85.431	0.854	0.148	1.000	1.021	1.021

In generally, the findings can be summarized as follows:

(i) When $\Delta^* = 0$, SRSM outperforms all the other estimators, because $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ is the biggest RMSE. On the other hand, after the small interval near $\Delta^* = 0$, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ decreases and goes to one.

Table 3.2: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ for $p_1 = 5$, $n = 50$, $p_2 = 10, 15, 20$ values when $\rho = 0.5$.

(n, p_2)	Δ^*	CNT	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPT}$	$\widehat{\boldsymbol{\beta}}_1^{SRS}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$
	0.00	88.632	0.891	1.791	1.579	1.457	1.609
	0.25	86.056	0.886	1.658	1.370	1.344	1.474
	0.50	57.550	0.827	1.464	1.047	1.370	1.387
	0.75	63.961	0.832	1.210	0.933	1.281	1.287
(50,10)	1.00	72.646	0.817	0.986	0.963	1.210	1.210
	1.25	51.278	0.795	0.699	0.989	1.129	1.129
	1.50	40.377	0.793	0.574	1.000	1.078	1.078
	2.00	70.227	0.836	0.359	1.000	1.057	1.057
	4.00	64.188	0.874	0.145	1.000	1.008	1.008
	0.00	175.698	0.877	2.024	1.536	1.717	1.853
	0.25	110.824	0.837	1.668	1.360	1.538	1.619
	0.50	151.451	0.796	1.449	1.125	1.463	1.504
	0.75	193.836	0.794	1.171	0.935	1.384	1.400
(50,15)	1.00	166.628	0.803	0.906	0.967	1.310	1.310
	1.25	124.040	0.769	0.765	0.973	1.246	1.246
	1.50	132.680	0.786	0.685	1.000	1.246	1.246
	2.00	161.094	0.765	0.463	1.000	1.118	1.118
	4.00	176.410	0.862	0.169	1.000	1.021	1.021
	0.00	327.152	0.807	2.564	2.161	2.112	2.328
	0.25	503.995	0.726	2.495	1.925	2.008	2.301
	0.50	201.653	0.771	2.159	1.247	1.928	2.025
	0.75	234.745	0.837	1.654	0.920	1.754	1.804
(50,20)	1.00	201.376	0.754	1.203	0.925	1.522	1.526
	1.25	256.446	0.719	1.051	0.990	1.449	1.449
	1.50	255.424	0.685	0.771	0.967	1.314	1.314
	2.00	238.655	0.686	0.537	1.000	1.191	1.191
	4.00	224.349	0.797	0.186	1.000	1.047	1.047

(ii) For large p_2 values in situation fix p_1 and n values, as the CNT increase, whereas the RMSE of $\widehat{\boldsymbol{\beta}}_1^{FM}$ decrease, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ increase.

(iii) The SRPT outperforms shrinkage ridge regression estimators as $\Delta^* = 0$ and values which are p_1 and p_2 close to each other. But, for large p_2 values in situation fix p_1 and n values, $\widehat{\boldsymbol{\beta}}_1^{SRPS}$ has biggest RMSE, after that the RMSE of $\widehat{\boldsymbol{\beta}}_1^{SRPT}$ is following that, and finally the RMSE of $\widehat{\boldsymbol{\beta}}_1^{SRS}$ is smaller than the other two estimators. As Δ^* is larger than zero, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{SRPT}$ decreases, and it staies on below 1 as Δ^* change around 0.5 to 1, after that it increases and approaches one as Δ^* is larger than 1.

Table 3.3: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ for $p_1 = 5$, $n = 50$, $p_2 = 10, 15, 20$ values when $\rho = 0.75$.

(n, p_2)	Δ^*	CNT	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPT}$	$\widehat{\boldsymbol{\beta}}_1^{SRS}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$
	0.00	225.936	0.687	2.273	1.873	1.605	1.838
	0.25	166.131	0.700	2.254	1.678	1.690	1.866
	0.50	145.911	0.667	1.656	1.137	1.353	1.507
	0.75	160.995	0.664	1.641	0.994	1.396	1.489
(50,10)	1.00	248.180	0.680	1.328	0.921	1.440	1.440
	1.25	196.610	0.609	1.035	0.911	1.252	1.266
	1.50	176.992	0.662	0.892	0.964	1.237	1.237
	2.00	168.403	0.647	0.640	0.980	1.115	1.115
	4.00	219.168	0.803	0.246	1.000	1.036	1.036
	0.00	314.292	0.701	2.678	2.180	1.788	2.233
	0.25	316.339	0.709	2.505	1.787	1.833	2.140
	0.50	277.221	0.686	1.935	1.313	1.629	1.816
	0.75	526.006	0.626	1.841	1.100	1.687	1.718
(50,15)	1.00	689.631	0.659	1.537	0.934	1.652	1.692
	1.25	620.821	0.567	1.318	0.902	1.536	1.536
	1.50	477.222	0.612	0.973	0.900	1.380	1.381
	2.00	439.955	0.623	0.686	0.980	1.232	1.234
	4.00	431.478	0.719	0.250	1.000	1.052	1.052
	0.00	596.460	0.763	3.180	2.648	2.288	2.676
	0.25	871.495	0.665	3.083	2.206	2.281	2.724
	0.50	445.554	0.574	2.641	1.463	2.046	2.345
	0.75	673.650	0.632	2.379	1.321	2.031	2.198
(50,20)	1.00	686.046	0.582	1.580	0.894	1.813	1.898
	1.25	655.975	0.581	1.288	0.820	1.671	1.749
	1.50	419.215	0.588	1.370	0.894	1.725	1.725
	2.00	1073.161	0.590	0.910	1.000	1.442	1.442
	4.00	1686.734	0.625	0.274	1.000	1.090	1.090

(iv) Comparing the SRS and the SRPS shows that the RMSE of $\widehat{\boldsymbol{\beta}}_1^{SRS}$ is smaller than the RMSE of $\widehat{\boldsymbol{\beta}}_1^{SRPS}$, which implies that the performance of SRS is the worser than the performance of SRPS.

We also plot the above results in the following Figure 3.1.

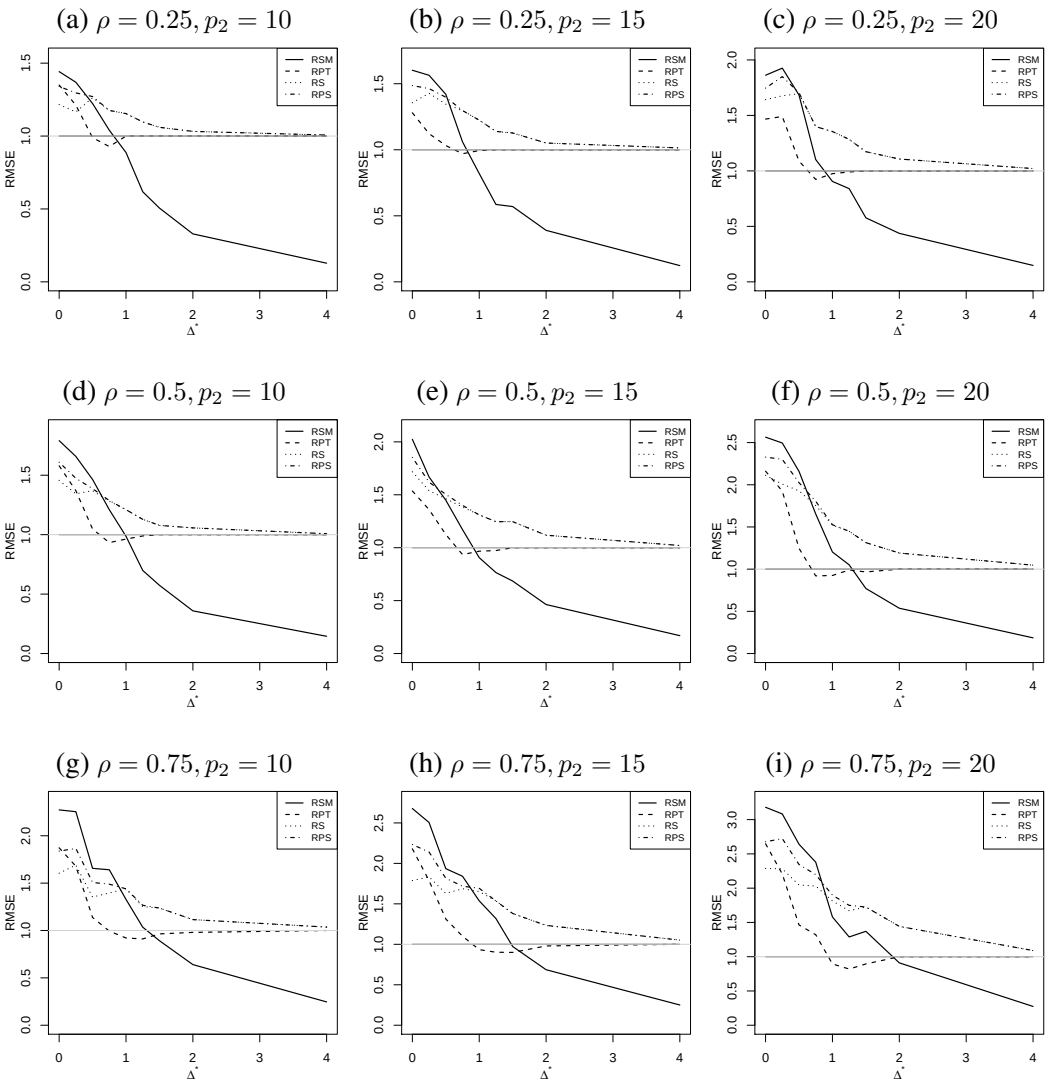


Figure 3.1: Relative efficiency of the estimators as a function of the non-centrality parameter Δ^* in the cases of $n = 50$, $p_1 = 5$, $p_2 = 10, 15, 20$ and $\rho = 0.25, 0.5, 0.75$.

Similarly, we give the simulation results for $p_1 = 5$, $n = 100$, $p_2 = 10, 15, 20$ values when $\rho = 0.25, 0.5, 0.75$ in Tables 3.4–3.6.

Table 3.4: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ for $p_1 = 5$, $n = 100$, $p_2 = 10, 15, 20$ values when $\rho = 0.25$.

(n, p_2)	Δ^*	CNT	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPT}$	$\widehat{\boldsymbol{\beta}}_1^{SRS}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$
	0.00	16.573	0.959	1.215	1.132	1.147	1.163
	0.25	16.403	0.963	1.135	1.001	1.118	1.129
	0.50	12.508	0.976	0.917	0.956	1.086	1.086
	0.75	15.723	0.953	0.735	1.000	1.051	1.051
(100,10)	1.00	17.276	0.982	0.568	1.000	1.048	1.048
	1.25	14.052	0.948	0.441	1.000	1.024	1.024
	1.50	15.487	0.947	0.354	1.000	1.029	1.029
	2.00	13.873	0.951	0.228	1.000	1.016	1.016
	4.00	19.841	0.964	0.064	1.000	1.003	1.003
	0.00	59.961	0.925	1.742	1.463	1.458	1.642
	0.25	30.183	0.951	1.787	1.371	1.470	1.639
	0.50	46.626	0.866	1.515	0.993	1.389	1.430
	0.75	79.166	0.865	1.121	0.949	1.341	1.341
(100,15)	1.00	55.726	0.898	0.961	1.000	1.225	1.225
	1.25	45.016	0.831	0.727	1.000	1.145	1.145
	1.50	41.805	0.847	0.540	1.000	1.126	1.126
	2.00	55.517	0.854	0.400	1.000	1.049	1.049
	4.00	59.849	0.868	0.148	1.000	1.008	1.008
	0.00	108.266	0.983	2.413	2.110	1.975	2.276
	0.25	87.594	0.907	2.518	1.616	1.948	2.165
	0.50	144.025	0.935	1.809	1.033	1.624	1.683
	0.75	98.665	0.867	1.490	0.974	1.524	1.554
(100,20)	1.00	84.118	0.821	1.080	0.963	1.322	1.328
	1.25	106.487	0.876	0.841	1.000	1.274	1.274
	1.50	93.385	0.831	0.675	1.000	1.192	1.192
	2.00	51.335	0.814	0.521	1.000	1.127	1.127
	4.00	74.168	0.833	0.176	1.000	1.020	1.020

Table 3.5: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ for $p_1 = 5$, $n = 100$, $p_2 = 10, 15, 20$ values when $\rho = 0.5$.

(n, p_2)	Δ^*	CNT	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPT}$	$\widehat{\boldsymbol{\beta}}_1^{SRS}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$
	0.00	57.214	0.840	1.770	1.620	1.349	1.586
	0.25	106.757	0.887	1.744	1.499	1.488	1.577
	0.50	66.295	0.806	1.488	1.038	1.292	1.381
	0.75	76.053	0.835	1.189	0.926	1.271	1.284
(100,10)	1.00	45.005	0.788	1.053	0.959	1.257	1.257
	1.25	78.790	0.835	0.795	0.978	1.180	1.180
	1.50	53.546	0.764	0.604	1.000	1.097	1.097
	2.00	39.744	0.792	0.423	1.000	1.054	1.054
	4.00	54.362	0.880	0.164	1.000	1.019	1.019
	0.00	117.076	0.807	2.119	1.730	1.764	1.890
	0.25	114.218	0.825	1.894	1.605	1.551	1.741
	0.50	171.543	0.857	1.723	1.113	1.517	1.628
	0.75	121.424	0.768	1.349	0.888	1.368	1.389
(100,15)	1.00	150.357	0.796	1.101	0.965	1.382	1.382
	1.25	194.160	0.782	0.855	0.989	1.255	1.255
	1.50	100.695	0.752	0.700	0.993	1.209	1.209
	2.00	118.501	0.747	0.489	1.000	1.125	1.125
	4.00	178.740	0.866	0.165	1.000	1.027	1.027
	0.00	420.612	0.820	2.754	2.018	2.056	2.321
	0.25	318.659	0.798	2.557	1.681	2.084	2.273
	0.50	529.710	0.756	1.954	1.192	1.743	1.838
	0.75	212.033	0.806	1.598	0.913	1.660	1.683
(100,20)	1.00	198.209	0.733	1.427	0.928	1.642	1.645
	1.25	183.612	0.783	1.236	0.984	1.535	1.535
	1.50	190.198	0.762	0.839	1.000	1.345	1.345
	2.00	292.300	0.700	0.557	1.000	1.218	1.218
	4.00	250.063	0.757	0.219	1.000	1.050	1.050

Table 3.6: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ for $p_1 = 5$, $n = 100$, $p_2 = 10, 15, 20$ values when $\rho = 0.75$.

(n, p_2)	Δ^*	CNT	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPT}$	$\widehat{\boldsymbol{\beta}}_1^{SRS}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$
	0.00	408.531	0.757	2.059	1.752	1.396	1.798
	0.25	261.824	0.654	1.730	1.368	1.384	1.545
	0.50	266.820	0.693	1.604	1.220	1.375	1.524
	0.75	353.797	0.687	1.323	0.956	1.358	1.398
(100,10)	1.00	133.519	0.618	1.091	0.846	1.247	1.250
	1.25	191.746	0.687	0.890	0.914	1.216	1.216
	1.50	249.561	0.656	0.748	0.929	1.176	1.176
	2.00	156.192	0.655	0.558	0.972	1.106	1.106
	4.00	202.087	0.803	0.220	1.000	1.017	1.017
	0.00	454.488	0.725	2.486	2.051	1.812	2.173
	0.25	544.863	0.683	2.380	1.725	1.805	2.145
	0.50	477.780	0.638	1.982	1.333	1.697	1.828
	0.75	600.107	0.661	1.843	1.050	1.709	1.825
(100,15)	1.00	429.514	0.687	1.653	0.913	1.722	1.723
	1.25	311.913	0.637	1.100	0.866	1.452	1.472
	1.50	808.337	0.615	0.967	0.920	1.439	1.439
	2.00	260.787	0.657	0.708	1.000	1.278	1.278
	4.00	432.728	0.689	0.280	1.000	1.064	1.064
	0.00	651.691	0.688	3.075	2.077	1.948	2.542
	0.25	916.529	0.656	2.897	1.977	2.076	2.497
	0.50	1102.479	0.573	2.307	1.418	1.550	2.150
	0.75	650.099	0.579	2.068	1.060	1.977	2.044
(100,20)	1.00	1036.395	0.661	1.662	0.912	1.799	1.913
	1.25	838.582	0.570	1.355	0.920	1.730	1.730
	1.50	523.026	0.524	1.076	0.931	1.580	1.580
	2.00	549.680	0.580	0.766	0.992	1.364	1.364
	4.00	838.293	0.642	0.246	1.000	1.060	1.060

We plot the above results in the following Figure 3.2.

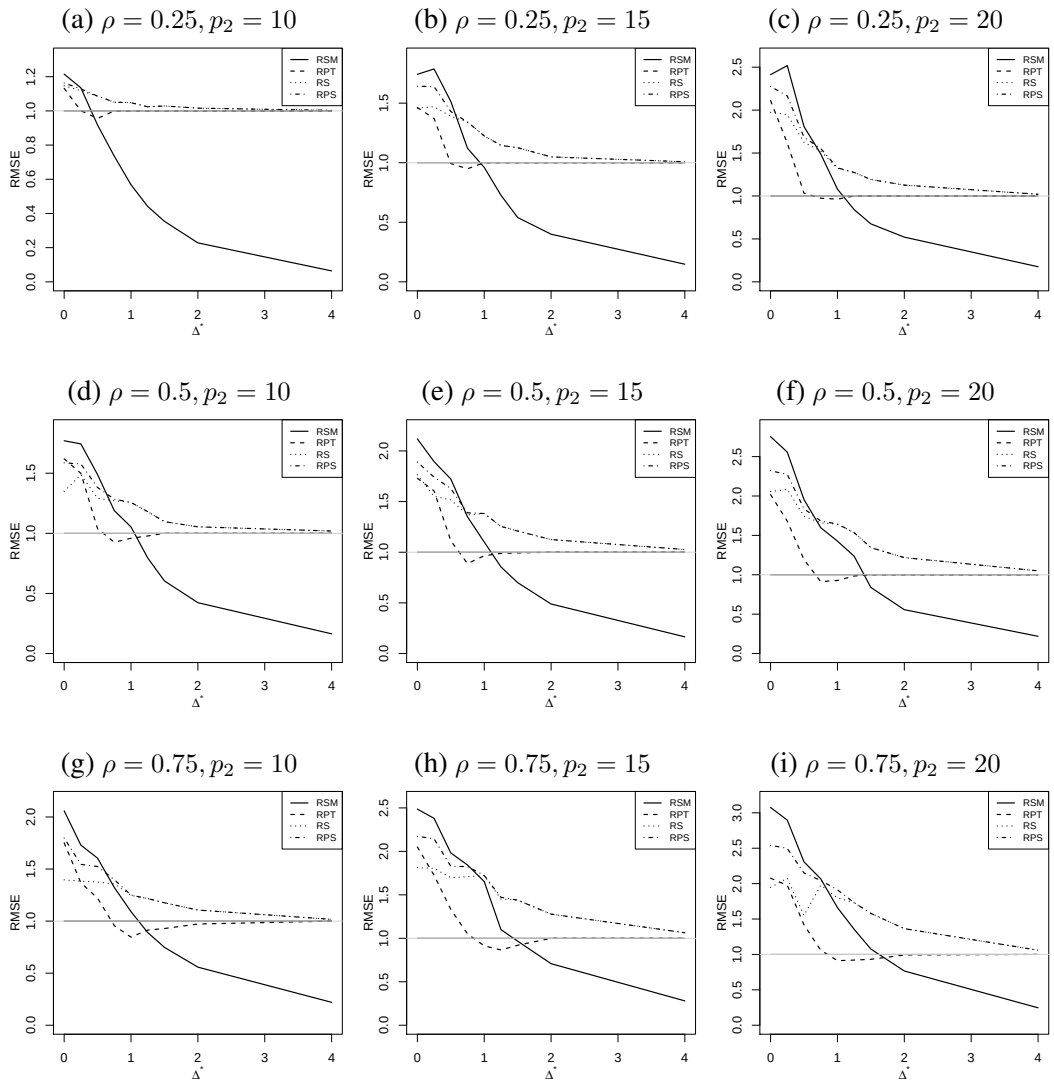
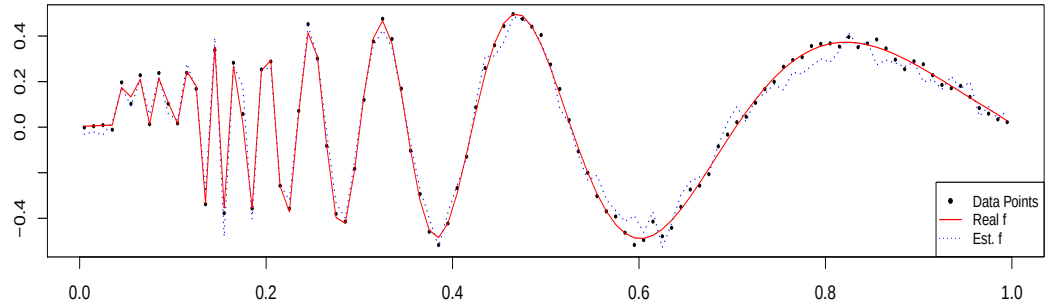


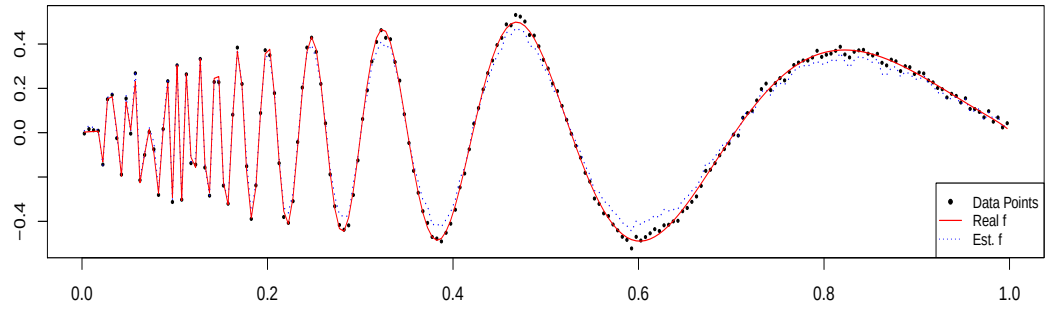
Figure 3.2: Relative efficiency of the estimators as a function of the non-centrality parameter Δ^* in the cases of $n = 100$, $p_1 = 5$, $p_2 = 10, 15, 20$ and $\rho = 0.25, 0.5, 0.75$.

(3.1) based on smoothing spline for different sample size n. The residuals which obtained after estimation of the linear component of the model (3.1) and real regression function with together fitted regression function correspond to the smoothing spline estimates using GCV. Similarly, the residuals and real function and its estimates for $n = 100, 200$ and the nonparametric functions f_1 and f_2 are represented in Figure 3.3. As shown in outcomes from the Figure 3.3, the curves estimated by smoothing spline denoted a similar behaviours to real functions. These plots suggest that real functions are quite good estimated, especially for larger sample size.

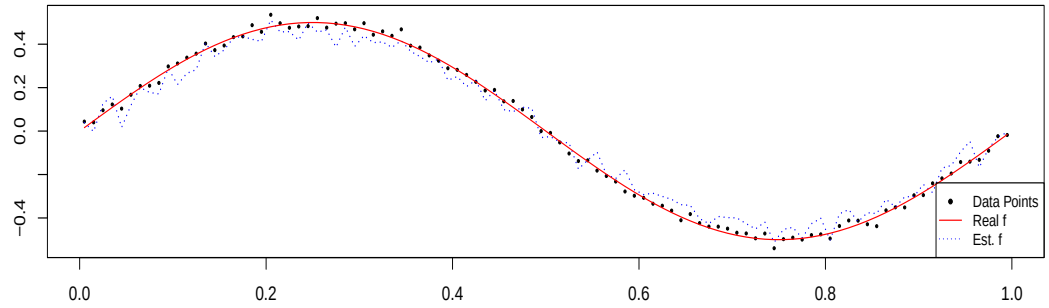
(a) f_1 when $n = 100$



(b) f_1 when $n = 200$



(c) f_2 when $n = 100$



(d) f_2 when $n = 200$

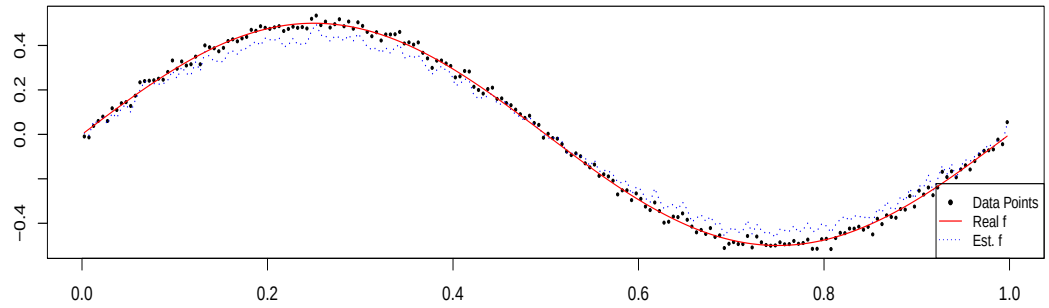


Figure 3.3: Estimation of f_1 and f_2 when $n = 100, 200$, $p_1 = 5$ and $p_2 = 10$. The data points are the residuals.

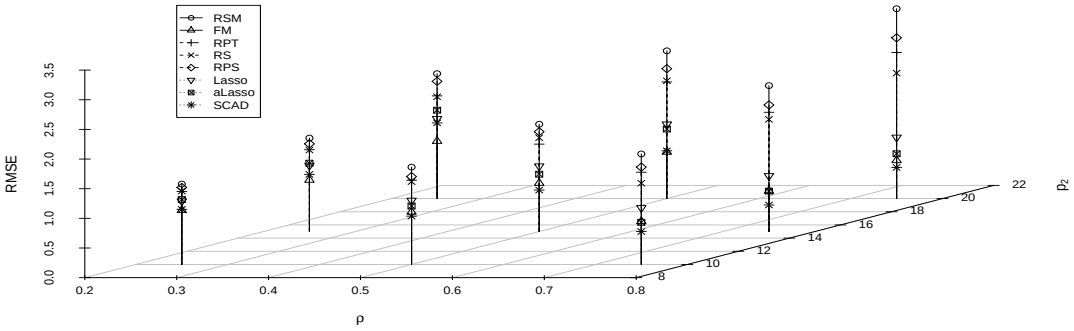
In the following Table 3.7, we show the results the comparison the suggested estimators with penalty estimators.

Table 3.7: CNT and Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ for $p_1 = 5, 10$, $p_2 = 10, 15, 20$ and $n = 50, 100$ values when $\Delta^* = 0$ and $\rho = 0.25, 0.5, 0.75$.

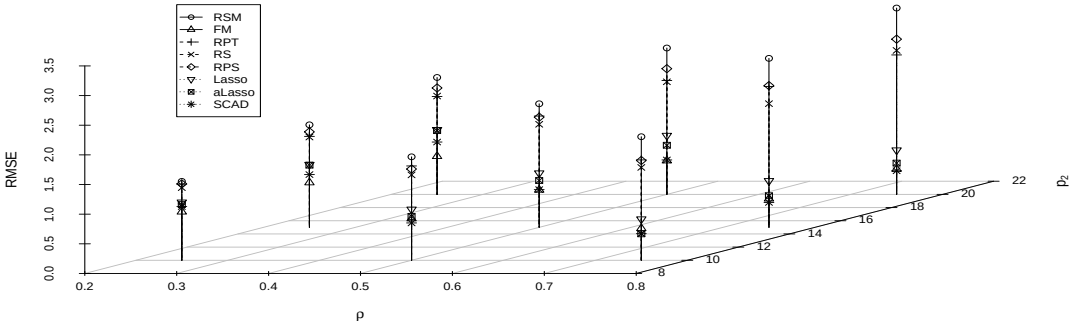
$(n, p1)$	ρ	p_2	CNT	$\widehat{\boldsymbol{\beta}}_1^{FM}$	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPT}$	$\widehat{\boldsymbol{\beta}}_1^{SRS}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$	$\widehat{\boldsymbol{\beta}}^{LASSO}$	$\widehat{\boldsymbol{\beta}}^{aLASSO}$	$\widehat{\boldsymbol{\beta}}^{SCAD}$
(50,5)	0.25	10	32.244	0.914	1.356	1.230	1.238	1.299	1.068	1.095	0.930
		15	64.630	0.869	1.576	1.380	1.375	1.479	1.113	1.153	0.964
		20	73.416	0.969	2.107	1.732	1.717	1.977	1.345	1.491	1.276
	0.5	10	44.713	0.887	1.642	1.421	1.397	1.482	1.080	0.986	0.815
		15	141.643	0.817	1.809	1.474	1.582	1.683	1.102	0.967	0.699
		20	286.533	0.786	2.492	1.967	1.987	2.191	1.254	1.173	0.806
	0.75	10	140.853	0.727	1.864	1.556	1.371	1.642	0.955	0.711	0.555
		15	345.621	0.684	2.461	2.011	1.892	2.134	0.938	0.682	0.448
		20	1181.337	0.644	3.202	2.464	2.116	2.711	1.032	0.756	0.525
(50,10)	0.25	10	43.232	0.818	1.331	1.284	1.225	1.289	0.978	0.940	0.906
		15	97.064	0.755	1.729	1.534	1.523	1.610	1.058	1.049	0.890
		20	199.648	0.643	1.972	1.651	1.659	1.795	1.090	1.075	0.883
	0.5	10	237.503	0.703	1.745	1.591	1.440	1.539	0.859	0.739	0.630
		15	187.336	0.631	2.084	1.877	1.738	1.858	0.914	0.790	0.636
		20	624.443	0.569	2.469	1.921	1.899	2.120	0.992	0.829	0.581
	0.75	10	410.867	0.538	2.084	1.693	1.566	1.683	0.694	0.446	0.449
		15	491.891	0.482	2.850	2.376	2.086	2.390	0.784	0.533	0.420
		20	1586.046	0.427	3.143	2.352	2.428	2.618	0.746	0.528	0.398
(100,5)	0.25	10	16.906	0.979	1.293	1.255	1.219	1.249	1.001	1.101	0.957
		15	21.330	0.982	1.435	1.376	1.283	1.366	1.012	1.171	1.001
		20	32.509	1.092	1.534	1.460	1.412	1.517	1.119	1.348	1.215
	0.5	10	24.492	0.917	1.091	1.107	1.123	1.157	1.098	1.076	0.849
		15	51.180	1.028	1.687	1.570	1.531	1.602	1.086	1.152	0.881
		20	118.785	1.071	1.880	1.651	1.626	1.760	1.121	1.230	0.998
	0.75	10	103.700	0.870	1.876	1.738	1.444	1.637	0.942	0.835	0.657
		15	129.756	0.974	2.096	1.755	1.712	1.839	0.999	0.875	0.704
		20	226.138	0.999	2.591	2.097	1.880	2.232	1.095	0.970	0.734
(100,10)	0.25	10	19.494	0.916	1.150	1.104	1.114	1.122	0.992	0.995	0.926
		15	42.708	0.931	1.357	1.258	1.250	1.291	0.996	1.042	0.979
		20	46.569	0.941	1.387	1.336	1.342	1.361	1.044	1.122	1.060
	0.5	10	56.776	0.879	1.453	1.388	1.333	1.336	0.989	0.914	0.818
		15	106.365	0.901	1.553	1.405	1.434	1.447	1.032	0.995	0.810
		20	129.272	0.870	1.648	1.496	1.521	1.541	1.053	1.051	0.801
	0.75	10	241.912	0.755	1.955	1.833	1.704	1.677	0.843	0.524	0.756
		15	331.234	0.783	2.244	1.980	1.871	1.919	0.867	0.595	0.727
		20	455.707	0.804	2.462	2.035	2.069	2.101	0.901	0.647	0.714

In the following Figure 3.4, we plot three-dimensional comparison of estimators.

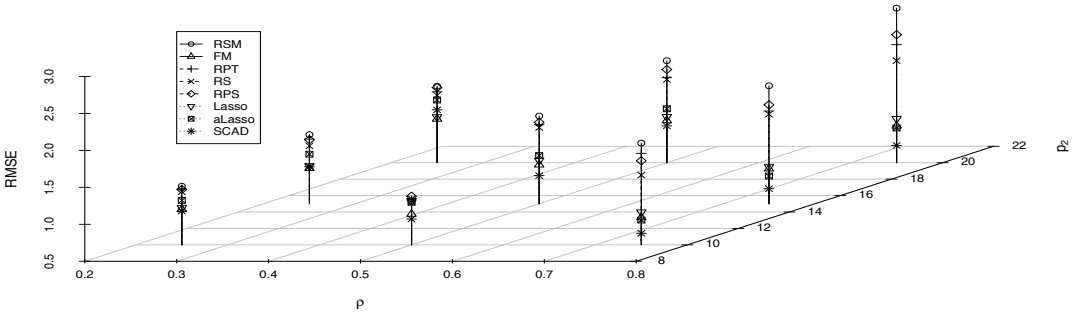
(a) $n = 50, p_1 = 5$



(b) $n = 50, p_1 = 10$



(c) $n = 100, p_1 = 5$



(d) $n = 100, p_1 = 10$

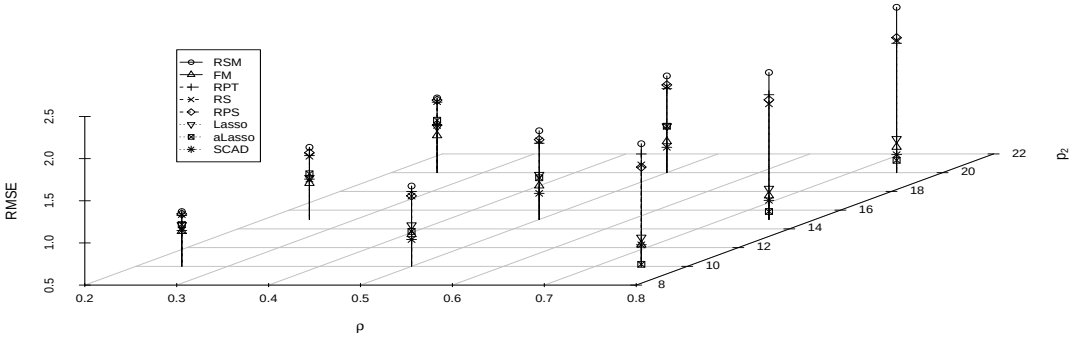


Figure 3.4: Three-dimensional plot of simulated RMSE against n and p_2 to compare efficiency of the estimators for $n = 50, 100$, $p_1 = 5, 10$, $p_2 = 10, 15, 20$ and $\rho = 0.25, 0.5, 0.75$.

3.11 Shrinkage Estimators for High-Dimensional Data

We can use $\widehat{\boldsymbol{\beta}}_1^{SRFM} = \left(\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}} + \lambda^R \boldsymbol{I}_p\right)^{-1} \tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{y}}$ as a full-model estimator when $p > n$.

Based on UIP or AI, if $\boldsymbol{\beta}_2 = \mathbf{0}$, then a sub-model estimator for $\boldsymbol{\beta}_1$ will be $\widehat{\boldsymbol{\beta}}_1^{SRSM} = \left(\tilde{\boldsymbol{X}}_1^\top \tilde{\boldsymbol{X}}_1 + \lambda^R \boldsymbol{I}_{p_1}\right)^{-1} \tilde{\boldsymbol{X}}_1^\top \tilde{\boldsymbol{y}}$ and a full-model estimator may be defined as $\widehat{\boldsymbol{\beta}}_1^{SRFM} = \left(\tilde{\boldsymbol{X}}_1^\top \boldsymbol{M}_2^R \tilde{\boldsymbol{X}}_1 + \lambda^R \boldsymbol{I}_{p_1}\right)^{-1} \tilde{\boldsymbol{X}}_{12}^\top \tilde{\boldsymbol{M}}_2^R \tilde{\boldsymbol{y}}$, where $\tilde{\boldsymbol{M}}_2^R$ is defined as $\boldsymbol{I}_n - \tilde{\boldsymbol{X}}_2 \left(\tilde{\boldsymbol{X}}_2^\top \tilde{\boldsymbol{X}}_2 + \lambda^R \boldsymbol{I}_{p_2}\right)^{-1} \tilde{\boldsymbol{X}}_2^\top$.

To construct semiparemetric shrinkage estimators, we suggest following high dimensional test statistics

$$\mathcal{T}_h = \frac{1}{\widehat{\sigma}_{UE}^2} \left(\widehat{\boldsymbol{\beta}}_2^{RFM}\right)^\top \left(\tilde{\boldsymbol{X}}_2^\top \tilde{\boldsymbol{M}}_1^R \tilde{\boldsymbol{X}}_2\right) \left(\widehat{\boldsymbol{\beta}}_2^{RFM}\right),$$

where $\widehat{\sigma}_{UE}^2$ is estimated as follows:

$$\widehat{\sigma}_{UE}^2 = \frac{1}{n}.\frac{\left\|\left(\boldsymbol{I}_n - \boldsymbol{H}_\lambda\right)\boldsymbol{y}\right\|^2}{\text{tr}\left(\boldsymbol{I}_n - \boldsymbol{H}_\lambda\right)^\top \left(\boldsymbol{I}_n - \boldsymbol{H}_\lambda\right)}.$$

Then, we can introduce semiparametric shrinkage ridge regression estimator of $\boldsymbol{\beta}_1$ defined by

$$\widehat{\boldsymbol{\beta}}_1^{SRS} = \widehat{\boldsymbol{\beta}}_1^{SRSM} + \left(\widehat{\boldsymbol{\beta}}_1^{SRFM} - \widehat{\boldsymbol{\beta}}_1^{SRSM}\right)\left(1 - \left(\widehat{\vartheta} - 2\right)\mathcal{T}_h^{-1}\right),$$

where $\widehat{\vartheta} = \frac{1}{2}\left\|\widehat{\boldsymbol{\beta}}_2^{SRFM}\right\|_1^2$ and $\|\cdot\|_1$ indicates the one norm.

The semiparametric positive shrinkage ridge regression estimator of $\boldsymbol{\beta}_1$ defined by

$$\widehat{\boldsymbol{\beta}}_1^{SRPS} = \widehat{\boldsymbol{\beta}}_1^{SRSM} + \left(\widehat{\boldsymbol{\beta}}_1^{SRFM} - \widehat{\boldsymbol{\beta}}_1^{SRSM}\right)\left(1 - \left(\widehat{\vartheta} - 2\right)\mathcal{T}_h^{-1}\right)^+,$$

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

3.12 Simulation studies for High-Dimensional Data

We perform Monte Carlo simulation experiments to examine the quadratic risk performance of the proposed estimators. We simulate the response from the follow-

ing model:

$$y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + ... + x_{pi}\beta_p + f\left(t_i\right) + \varepsilon_i\,,\ i = 1, 2, ..., n,$$

where $\boldsymbol{x}_i \sim N_p\left(\mathbf{0}, \boldsymbol{\Sigma}_x\right)$ and ε_i are i.i.d. $N\left(0, 1\right)$. We also define $\boldsymbol{\Sigma}_x$ that is positive definite covariance matrix. To generate $\boldsymbol{\Sigma}_x$, we define ρ which is coefficient of correlation $\boldsymbol{X}$. We use three values of correlation which are low($\rho = .25$), medium($\rho = .5$) and high($\rho = .75$). We also get $\alpha = 0.05$ and $t_i = \left(i - 0.5\right)/n$. Furthermore, we consider the hypothesis $H_0 : \beta_j = 0$, for $j = p_1 + 1, p_1 + 2, ..., p$, with $p = p_1 + p_2$. Hence, we partition the regression coefficients as $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\right) = \left(\boldsymbol{\beta}_1, \mathbf{0}\right)$ with $\boldsymbol{\beta}_1 = \left(1, 1, 1, 1, 1\right)$. We consider two different the nonparametric functions $f_1\left(t_i\right) = \sqrt{t_i\left(1 - t_i\right)}\sin\left(\frac{2.1\pi}{t_i + 0.05}\right)$ and $f_2\left(t_i\right) = 0.5\sin\left(4\pi t_i\right)$ to generate response y_i.

Each realization was repeated 1000 times to obtain sensible results, and for each realization, we calculated bias of the estimators. We define $\Delta^* = \left\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\right\|$, where $\boldsymbol{\beta}_0 = \left(\boldsymbol{\beta}_1, \mathbf{0}\right)$, and $\left\|\cdot\right\|$ is the Euclidean norm. Further, to investigate the performance of the estimators for $\Delta^* > 0$, more datasets were generated from those distributions.

The performance of an estimator $\boldsymbol{\beta}_1$ was evaluated by calculating its mean squared error (MSE) criterion. We numerically calculated the RMSE of $\widehat{\boldsymbol{\beta}}_1^{SRSM}$, $\widehat{\boldsymbol{\beta}}_1^{SRPT}$, $\widehat{\boldsymbol{\beta}}_1^{SRS}$ and $\widehat{\boldsymbol{\beta}}_1^{SRPS}$, with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$. The relative mean square efficiency(RMSE) of the $\boldsymbol{\beta}_1^{\blacktriangledown}$ to the unrestricted least squares estimator $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ is indicated by

$$RMSE\left(\widehat{\boldsymbol{\beta}}_1^{SRFM} : \boldsymbol{\beta}_1^{\blacktriangledown}\right) = \frac{MSE\left(\widehat{\boldsymbol{\beta}}_1^{SRFM}\right)}{MSE\left(\boldsymbol{\beta}_1^{\blacktriangledown}\right)},$$

where $\boldsymbol{\beta}_1^{\blacktriangledown}$ is one of the proposed estimators. The amount by which a RMSE is larger than one indicates the degree of superiority of the estimator $\boldsymbol{\beta}_1^{\blacktriangledown}$ over $\widehat{\boldsymbol{\beta}}_1^{SRFM}$.

We summarized the simulation results in the following Tables 3.8 and 3.9.

Table 3.8: In case of high-dimensional, simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ for $n = 50$ and $p_1 = 5$.

ρ	p_2	Δ^*	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$	p_2	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$	p_2	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$
		0.0	7.592	3.315		12.123	5.547		14.399	6.436
		0.5	5.936	2.950		10.257	4.796		11.042	5.284
		1.0	3.594	2.422		5.544	3.801		6.863	4.653
0.25	100	1.5	2.343	2.049	250	3.428	3.065	500	4.325	3.684
		2.0	1.610	1.819		2.407	2.732		2.825	3.165
		4.0	0.547	1.482		0.730	2.035		0.787	2.318
		8.0	0.179	1.307		0.232	1.709		0.255	1.900
		16.0	0.068	1.162		0.082	1.347		0.080	1.187
		0.0	5.812	1.890		8.730	3.210		9.330	3.642
		0.5	5.244	1.728		7.180	2.775		7.871	3.282
		1.0	3.570	1.603		4.837	2.422		4.984	2.818
0.5	100	1.5	2.314	1.520	250	3.005	2.159	500	2.991	2.578
		2.0	1.658	1.479		1.939	2.142		2.134	2.459
		4.0	0.523	1.383		0.668	1.913		0.693	2.241
		8.0	0.147	1.302		0.198	1.753		0.184	2.056
		16.0	0.047	1.262		0.058	1.615		0.054	1.623
		0.0	3.830	1.316		4.672	2.007		5.210	2.627
		0.5	3.592	1.261		4.355	1.735		5.014	2.325
		1.0	2.617	1.230		3.333	1.616		4.107	2.195
0.75	100	1.5	1.925	1.221	250	2.486	1.598	500	2.822	2.057
		2.0	1.493	1.213		1.793	1.538		2.001	1.933
		4.0	0.484	1.204		0.599	1.457		0.635	1.901
		8.0	0.141	1.207		0.192	1.450		0.197	1.883
		16.0	0.037	1.221		0.050	1.403		0.059	1.843

According to the above table, $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ has the biggest RMSE as $\Delta^* = 0$. On the other hand, $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ is getting worser than $\widehat{\boldsymbol{\beta}}_1^{SPRS}$ as Δ^* larger than zero. Also, it can be concluded that the performance of SRPS increase when the number of nuisance parameters increase.

In the following Figure 3.5, we plot the simulation results in Table 3.8.

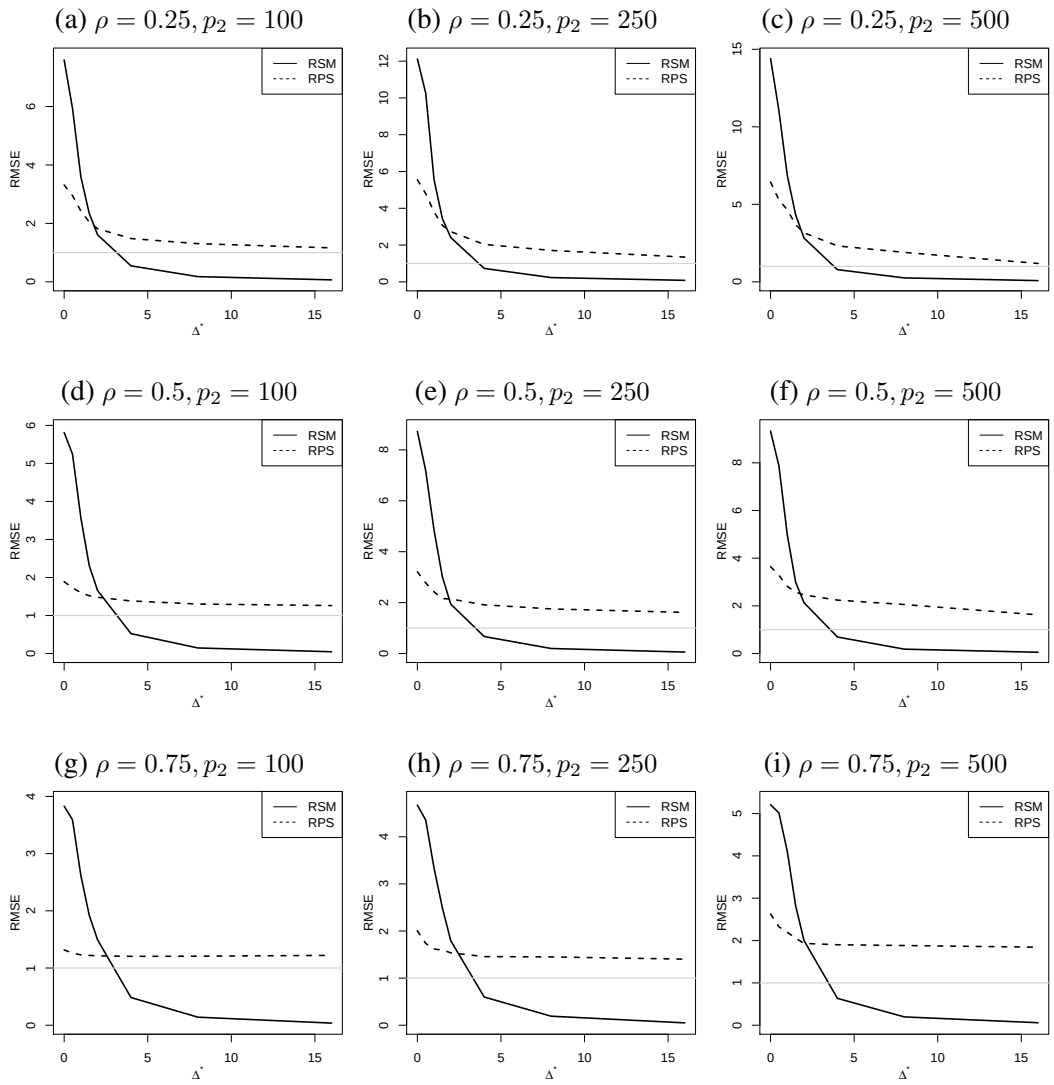


Figure 3.5: Relative efficiency of the estimators as a function of the non-centrality parameter Δ^* for $n = 50$, $p_1 = 5$, $p_2 = 100, 250, 500$ and $\rho = 0.25, 0.5, 0.75$.

In the following Table 3.9, we give simulated relative efficiency of the estimators for $n = 100$.

Table 3.9: In case of high-dimensional, simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ for $n = 100$ and $p_1 = 5$.

ρ	p_2	Δ^*	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$	p_2	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{RPS}$	p_2	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$
		0.0	6.642	3.185		16.306	13.606		22.871	20.224
		0.5	5.863	2.619		13.172	10.404		21.389	16.112
		1.0	2.961	2.102		6.780	6.134		10.693	9.144
0.25	100	1.5	1.970	1.714	250	4.434	4.171	500	6.037	6.204
		2.0	1.413	1.516		2.588	3.001		3.818	4.402
		4.0	0.546	1.313		0.898	2.051		1.056	2.765
		8.0	0.187	1.236		0.273	1.719		0.340	2.240
		16.0	0.068	1.118		0.082	1.515		0.099	1.698
		0.0	7.905	2.111		14.076	7.210		19.204	12.201
		0.5	6.295	1.909		10.767	5.042		13.019	9.048
		1.0	3.462	1.664		5.966	3.979		8.012	6.717
0.5	100	1.5	2.188	1.535	250	3.470	3.364	500	5.275	5.338
		2.0	1.457	1.432		2.412	2.727		2.962	4.286
		4.0	0.505	1.298		0.713	2.065		0.913	3.304
		8.0	0.153	1.277		0.212	1.834		0.253	2.916
		16.0	0.050	1.213		0.062	1.671		0.069	2.290
		0.0	6.567	1.448		9.098	3.490		8.656	6.257
		0.5	6.129	1.353		6.892	2.747		8.206	5.094
		1.0	3.893	1.297		5.078	2.276		5.640	4.493
0.75	100	1.5	2.566	1.259	250	3.167	2.015	500	3.666	3.598
		2.0	1.754	1.238		2.191	1.831		2.575	3.129
		4.0	0.581	1.197		0.694	1.634		0.780	2.578
		8.0	0.171	1.191		0.212	1.505		0.230	2.388
		16.0	0.043	1.189		0.057	1.422		0.064	2.062

In the following Figure 3.6, we plot the simulation results in Table 3.9.

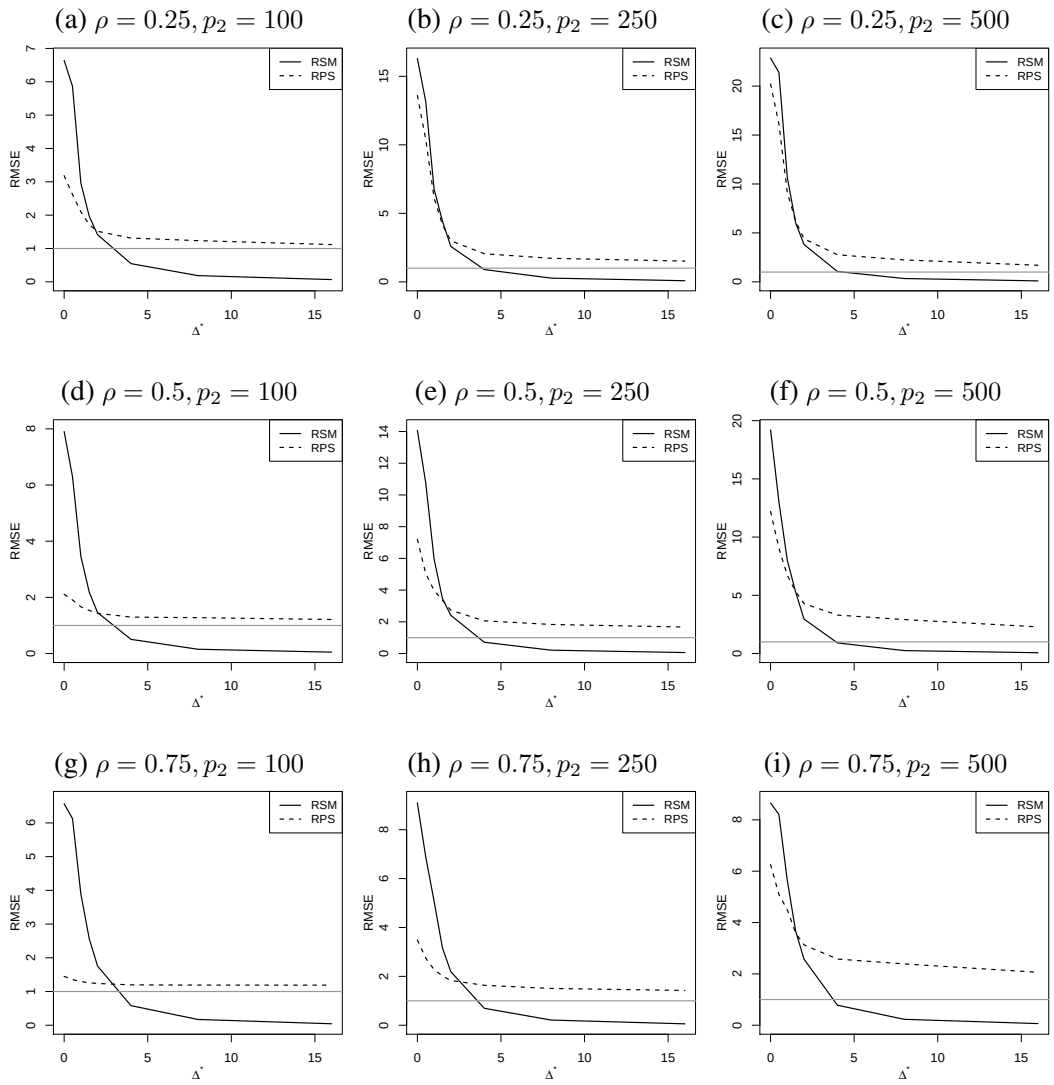


Figure 3.6: Relative efficiency of the estimators as a function of the non-centrality parameter Δ^* for $n = 100$, $p_1 = 5$, $p_2 = 100, 250, 500$ and $\rho = 0.25, 0.5, 0.75$.

In the following Table 3.10, we present the simulation results of comparison shrinkage ridge regression estimators with penalty estimators. We can concluded that SRPS outperforms penalty estimators both as ρ is large and as p_2 is large. Also, SRFM dominates scad estimator when $\rho = 0.75$, indicated by the RMSE of the estimators remain below one.

Table 3.10: Simulated relative efficiency with respect to $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ for $p_1 = 5$, $p_2 = 100, 200, 400, 600, 1000$ and $n = 50, 100$ values when $\Delta^* = 0$ and $\rho = 0.25, 0.5, 0.75$.

n	ρ	p_2	$\widehat{\boldsymbol{\beta}}_1^{SRSM}$	$\widehat{\boldsymbol{\beta}}_1^{SRPS}$	$\widehat{\boldsymbol{\beta}}^{LASSO}$	$\widehat{\boldsymbol{\beta}}^{aLASSO}$	$\widehat{\boldsymbol{\beta}}^{SCAD}$
50	0.25	100	7.694	4.509	2.852	3.726	2.121
		200	10.930	6.753	2.930	4.033	2.542
		400	13.450	8.467	2.657	3.060	1.737
		600	14.134	8.851	2.457	2.685	1.513
		1000	14.536	10.182	2.110	2.178	1.266
	0.50	100	6.489	2.654	2.295	2.001	1.374
		200	8.124	3.672	2.012	1.726	1.175
		400	9.516	4.732	2.051	1.935	1.108
		600	10.231	5.760	1.766	1.532	0.996
		1000	12.405	6.633	1.626	1.437	0.877
	0.75	100	4.404	1.701	1.572	1.031	0.662
		200	4.782	2.352	1.518	1.070	0.627
		400	5.217	3.202	1.406	1.015	0.690
		600	5.647	3.870	1.338	1.026	0.741
		1000	5.940	3.981	1.267	0.971	0.702
100	0.25	100	7.575	3.773	3.492	5.888	4.496
		200	13.699	12.886	5.398	10.458	7.337
		400	21.582	20.053	6.564	15.802	10.809
		600	25.756	23.664	6.901	18.095	12.145
		1000	31.287	26.361	6.834	20.247	13.317
	0.50	100	7.505	2.733	3.418	5.208	3.861
		200	11.566	7.321	3.931	6.764	5.317
		400	15.228	13.665	4.820	9.194	5.544
		600	20.748	16.984	4.920	10.839	6.683
		1000	21.655	17.805	4.655	9.908	6.047
	0.75	100	6.128	1.733	2.680	2.322	0.889
		200	8.256	3.501	2.623	2.431	1.065
		400	9.637	7.504	2.469	2.254	1.344
		600	10.899	9.102	2.626	2.495	1.454
		1000	10.430	9.394	2.382	2.058	1.515

In the following Figure 3.7, we plot the simulation results in Table 3.10.

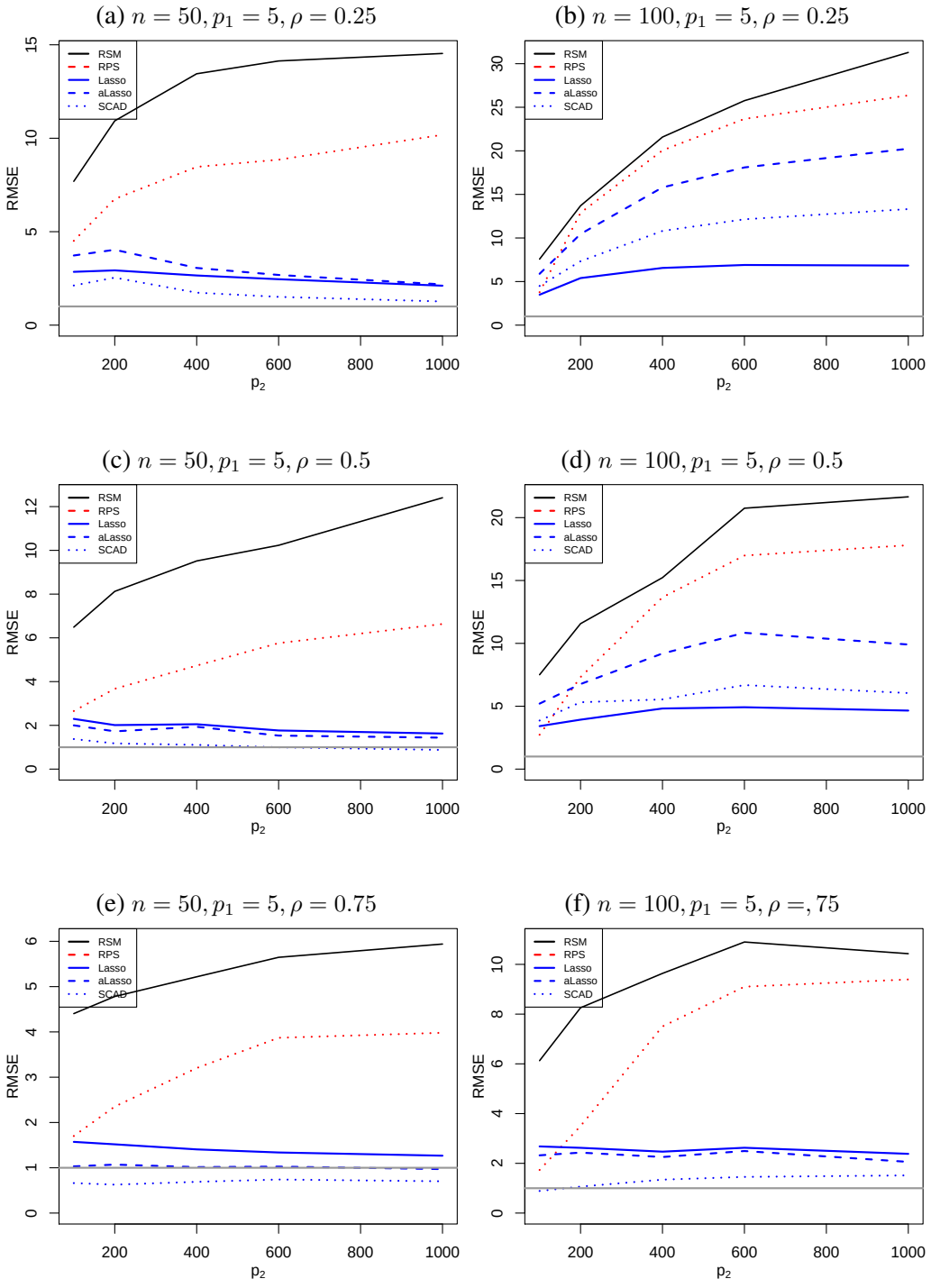
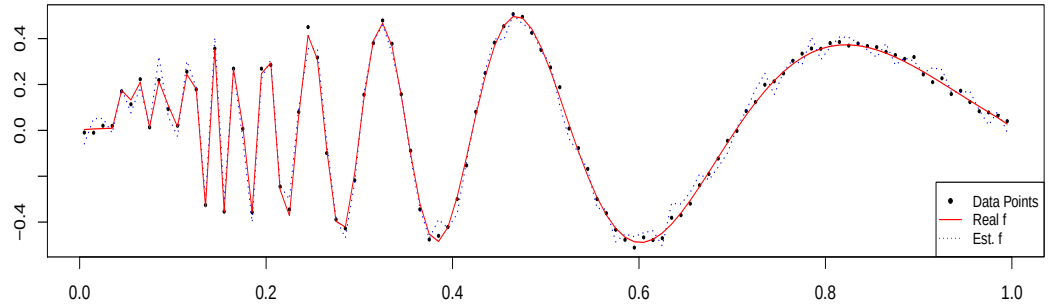


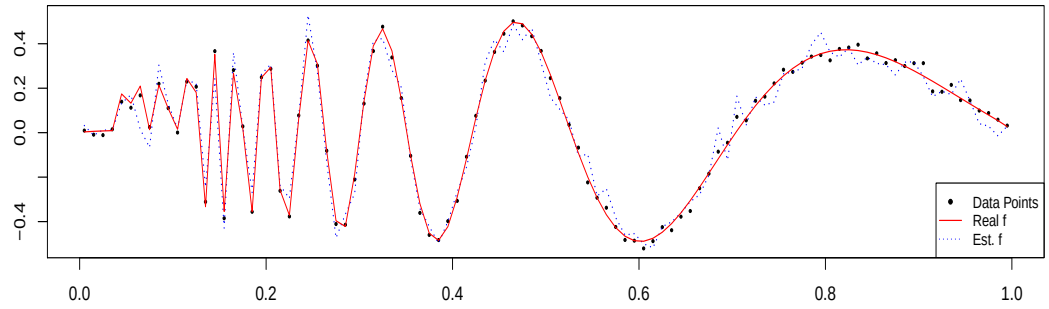
Figure 3.7: Relative efficiency of the estimators for $n = 50, 100$, $p_1 = 5$, $p_2 = 100, 200, 400, 600, 1000$ and $\rho = 0.25, 0.5, 0.75$.

In the following Figure 3.8, we plot the estimations of f_1 and f_2 functions in the case of high-dimensional.

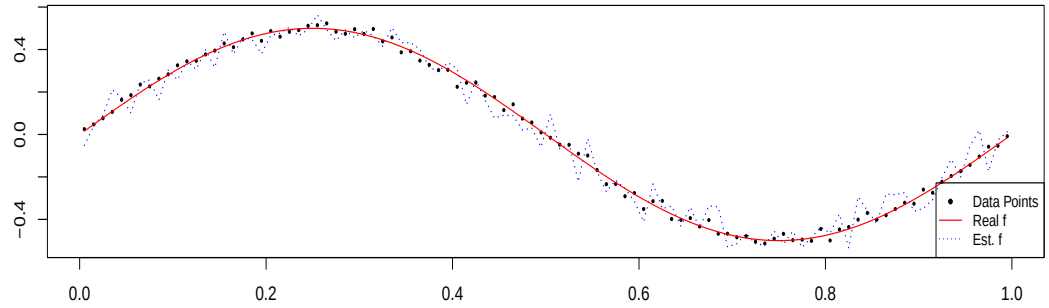
(a) f_1 when $p_2 = 200$



(b) f_1 when $p_2 = 500$



(c) f_2 when $p_2 = 200$



(d) f_2 when $p_2 = 500$

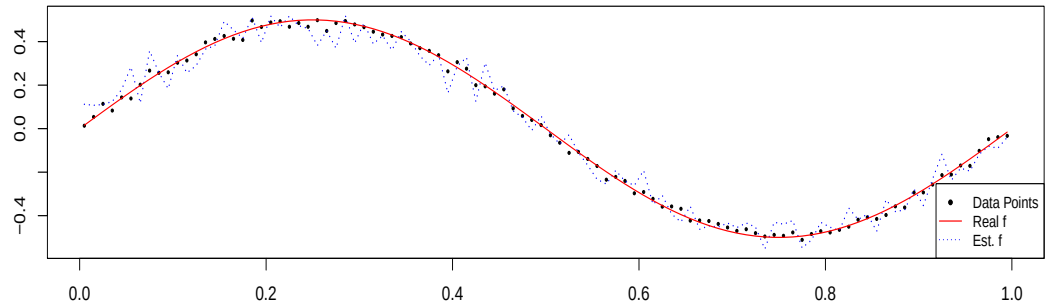


Figure 3.8: Estimation of f_1 and f_2 when $n = 100$, $p_1 = 5$ and $p_2 = 200, 500$. The data points are the residuals.

3.13 Conclusion

In this chapter, we suggested semiparametric pretest ridge regression estimator and semiparametric shrinkage ridge regression estimators for partially linear models.

In the first half part of this chapter, we obtained semiparametric pretest ridge regression and semiparametric shrinkage ridge regression estimators for semiparametric linear models which includes both parametric and nonparametric components. While the parametric components is estimated by using ridge regression approach, the nonparametric components is estimated by using Speckman approach based on penalized residual sum of squares method. We conduct Monte Carlo simulation to study the behaviour of proposed estimators under varying Δ^*, different magnitude of multicollinearity and different sizes of the nuisance subsets. For comparison, we use the relative mean square error for all estimators with respect to semiparametric full model estimator. As a results of the simulations, as $\Delta^* = 0$, the semiparametric sub model ridge regression estimator outperforms all other estimators for all the cases considered in this study. Also, under this condition, SPRS outshines all penalty estimators. However, as the restriction moves away from $\Delta^* = 0$, the RMSE of SRSM sharply goes below one, and the RMSE of SPRS approaches to one at the slowed rate. We also compared the suggested estimators with penalty estimators in simulations. In summary, when $\Delta^* = 0$, semiparametric pretest and shrinkage ridge regression estimators outshines penalty estimators. We also noticed that the performance of semiparametric pretest and shrinkage ridge regression estimators increase compared to penalty while the magnitude of multicollinearity is increase. Asymptotic risk properties of the proposed estimators have been demonstrated and their dominance over classical estimators showed.

In the second half part of this chapter, we suggested semiparametric shrinkage ridge regression estimators for high-dimensional data in the context of a partially linear models. The nonparametric components is estimated by using Speckman approach based on penalized residual sum of squares method for HD. We conducted a Monte Carlo simulation. Not surprisingly, when $\Delta^* = 0$, $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ has the biggest RMSE. However, when Δ^* larger than zero, SRSM loses its efficiency quickly, whereas SRPS is less affected. We also compare semiparametric shrinkage esti-

mators with penalty estimators which are Lasso, adaptive Lasso and SCAD. As a results of simulations, SRPS outperforms penalty estimators both when ρ is large and when p_2 is large. Even SRFM dominates penalty estimators for some values of ρ and p_2.

4 CONCLUSIONS and FUTURE WORK

4.1 Conclusions

In this thesis, two main problems which are multicollinearity and high dimensional estimation have been studied. To this end, we suggested pretest and shrinkage estimations based ridge regression for linear and partially models. Some real data application of the proposed estimators have been carried out for both models. Furthermore, Monte Carlo studies are conducted to compare the estimators. Finally, asymptotic properties of the proposed are obtained.

In the first half part of Chapter 2, we obtained pretest ridge regression, shrinkage and positive shrinkage ridge regression estimations in the existence of multicollinearity for linear models, and took into considering their applicability using real data examples. We numerically showed sub-model ridge regression estimator has superior risk performance compared to all other estimators as the underlying model is correctly specified. On the other hand, under model misspecification, positive shrinkage ridge regression estimator showed superior performance from the point of minimizing prediction errors. We also conducted Monte Carlo simulation to study the behaviour of proposed estimators under varying Δ^*, different sizes of the nuisance subsets and different magnitude of multicollinearity. Not surprisingly, the sub model ridge regression estimator outperforms pretest ridge regression, shrinkage and positive shrinkage ridge regression estimators when the restriction is at and near $\Delta^* = 0$. However, as Δ^* is larger than zero, the RMSE the sub-model ridge regression estimator increases and becomes unbounded, while the RMSEs of all other proposed estimators remain bounded and approach one. We also showed that non-penalty ridge regression estimators overshadow ordinary least square estimators in the existence of mullticollinearity. We also compared the proposed estimators with penalty estimators. In a nutshell, when $\Delta^* = 0$, pretest and shrinkage ridge regression estimators outshines penalty estimators. We also noticed that the performance of pretest and shrinkage ridge regression estimators outshine penalty while the magnitude of multicollinearity is increase. Asymptotic risk properties of the suggested estimators have been demonstrated and their dominance over classical estimators demonstrated.

In the second half part of Chapter 2, we proposed shrinkage ridge regression estimators for high-dimensional data in the context of a multiple linear regression. For comparison, we applied two real data application. After that, we also conducted a Monte Carlo simulation. Not surprisingly, when $\Delta^* = 0$, $\widehat{\boldsymbol{\beta}}_1^{RSM}$ has the biggest RMSE. However, when Δ^* larger than zero, RSM loses its efficiency quickly, whereas RPS is less affected. Also, RPS outperform when the number of nuisance covariates gets extremely large relative to the sample size. Furthermore, we compared shrinkage ridge regression estimator with penalty estimators. Hence, the RMSE of RSM increase when the number of nuisance covariates are large. As the size of multicollinearity increase, RPS dominates all penalty estimators.

In the first half part of Chapter 3, we obtained non-penalty estimators based on ridge regression for semiparametric linear models which includes both parametric and nonparametric components. In order to estimate the nonparametric components, it has been used Speckman approach based on penalized residual sum of squares method. In this chapter, we conducted Monte Carlo simulation to study the behaviour of proposed estimators under varying Δ^*. For comparison, we used the relative mean square error for all estimators with respect to semiparametric full model estimator. As a results of the simulations, as $\Delta^* = 0$, the semiparametric sub-model ridge regression estimator dominates all other estimators for all the cases considered in this study. Also, under this condition, SPRS outshines all penalty estimators. However, as the restriction moves away from $\Delta^* = 0$, the RMSE of SRSM sharply goes below one, and the RMSE of SPRS approaches to one at the slowed rate. We also compared the non-penalty estimators with penalty estimators. As $\Delta^* = 0$, semiparametric non-penalty estimators based on ridge regression outperform penalty estimators. We also noticed that the performance of semiparametric pretest and shrinkage ridge regression estimators increase compared to penalty while the magnitude of multicollinearity is increase. Asymptotic risk properties of the proposed estimators have been demonstrated and their dominance over classical estimators showed.

In the second half part Chapter 3, we proposed semiparametric shrinkage ridge regression estimators for high-dimensional data in the context of a partially linear models. We conducted a Monte Carlo simulation. As $\Delta^* = 0$, $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ has the

biggest RMSE. On the other hand, as Δ^* larger than zero, SRSM loses its efficiency quickly, whereas SRPS is less affected. We also compare semiparametric shrinkage estimators with penalty estimators which are Lasso, adaptive Lasso and SCAD. As a results of simulations, SRPS outperforms penalty estimators both when ρ is large and when p_2 is large. Futhermore, SRFM dominates penalty estimators for some values of ρ and p_2.

4.2 Future Work

Our research goal is considering an estimation of the nonparametric part of suggested estimation.

REFERENCES

Ahmed, S. E. (1997). Asymptotic shrinkage estimation: the regression case. *Applied Statistical Science II*, 113 – 139.

Ahmed, S. E. (2001). Shrinkage estimation of regression coefficients from censored data with multiple observations. In Ahmed, S. E. and Reid, N., editors, Empirical Bayes and Likelihood Inference, Lecture Notes in Statistics, **148**, 103-120. Springer-Verlag, New York.

Ahmed, S. E. (2002). Simultaneous estimation of coefficient of variations. *Journal of Statistical Planning and Inference*, **104**, 31-51.

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

Ahmed, S. E., Chitsaz, S. and Fallahpour, S. (2010). Shrinkage Preliminary Test Estimation. *International Encyclopaedia of Statistical Science*, M. Lovric (editor), Springer.

Ahmed, S. E. and Raheem, S.M.E. (2012). Shrinkage and absolute penalty estimation in linear models, *WIREs Computational Statistics*, **4**, 541–553.

Ahmed, S. E., Doksum, K. A., Hossain, S. and You, J. (2007). Shrinkage, Pretest and Absolute Penalty Estimators in Partially Linear Models. *Aust. N. Z. J. Stat*. **49**(4), 435 – 454.

Akaike, H. 1977. "On entropy maximization principle". In: Krishnaiah, P.R. (Editor). Applications of Statistics, North-Holland, Amsterdam, 27 – 41.

Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary tests of significances. *Annals of Mathematical Statistics*, **15**, 190 – 204.

Bickel, P.J. (1984). Parametric robustness: small biases can be worthwhile. *Ann. Statist*. **12**, 864–879.

Brown, P.J., Fearn, T. and Vannucci M. (2001) Bayesian Wavelet Regression on Curves with Applications to a Spectroscopic Calibration Problem. *Journal of the American Statistical Association*. **96**, 398–408.

Bühlmann, P., Kalisch, M. and Meier, L. (2013). High-dimensional statistics with a view towards applications in biology. *Annual Review of Statistics and its Applications.*

Burhnam, K.P. and Anderson, D.R. (2002). Model selection and multimodel inference, New York: Springer-Verlag.

Castner, L. A. and Schirm, A. L. (2003). Empirical bayes shrinkage estimates of state food stamp participation rates for 1998-2000. Technical report, Mathematica Policy Research, Inc

Chen, H. (1988). Convergence rates for parametric components in a partially linear models, *Annals of Statistics*, **16**, 136 – 146.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Num. Math.*, **31**, 377 – 403.

Donoho, D. L. and Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, **26**(3), 879-921.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, **32**, 407 – 499.

Engle, R. F., Granger, C.W.J., Rice, C.A. and Weiss A. (1986). Semi-parametric estimates of the relation between weather and electricity sales. *Journal of Amer. Statist., Assoc.* **81**, 310 – 320.

Eubank, R. L. (1999). *Nonparametric Regression and Smoothing Spline*, Marcel Dekker Inc., New York.

Eubank, R.L., Kamboura, E.L., Kimb, J.T., Klipplea, K., Reesea, C.S. and Schimekc. M. (1998) Estimation in partially linear models. *Computational Statistics & Data Analysis*. **29**, 27 – 34.

Fallahpour, S. and Ahmed, S.E. (2013). Variable selection and post-estimation of regression parameters using qusi-likelihood approach. Multivariate Statistics, T. Kollo (editor), 1-13, World Scientific:USA.

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

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

Friendly, M. 2002. Corrgrams: Exploratory Displays for Correlation Matrices. *The American Statistician*. **56**, 316 – 324.

Furnival, G. and Wilson, R. (1974). Regression by leaps and bounds, *Technometrics*. **16**, 499 – 511

Gibbons, D. G. (1981) A simulation study of some ridge estimators. *J. Amer.Stat. ASSOC.*, **76**,131 – 139.

Green P., J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Model*, Chapman &Hall.

Green, P. J., Jennison, C. and Seheult, A., (1985). Analysis of field experiments by least square smoothing , *J. Roy. Statis. Soc. B*, **47**, 299 – 315.

Gruber, M. H.J. (1998). *Improving Effciency by Shrinkage: The James-Stein and Ridge Regression Estimators*. Marcel Dekker, New York.

Hamilton, S. A. and Truong, Y. K. (1997). Local linear estimation in partially linear models, *J. Statist. Plan. Inference*, **80**, 37 – 57.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer.

Heckman, N. E. (1986). Smoothing spine in partially linear models. *J. Roy. Statist. Soc. Ser. B*. **48**, 24 – 248.

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

Hoerl, A.E., Kennard, R.. W. and Baldwin, K.F. (1970b). Ridge regression: Applications to Nonorthogonal Problems. *Technometrics*, **12**(1),69 – 82.

Hoeting, J.A., Madigan, D., Raffery, A.E. and Volinsky, C.T. (1999). Bayesian model averaging: A tutorial. *Statist. Sci*. **14**, 382–401.

Hoeting, J.A., Raffery, A.E. and Madigan, D. (2002). Bayesian variable and transformation selection in linear regression averaging: A tutorial. *J. Comput. Graph. Statist*. **11**, 485–507.

Inoue, T. (2001). Improving the "HKB" ordinary type ridge estimators. *J. Jpn. Stat. SOC.***31**(1), 67 – 83.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. on Math. Statist. and Probability*, University of California, **1**1961, 361–379.

Johnson, R. A. and Wichern, D. W. (2001). Applied Multivariate Statistical Analysis, 3rd Ed. Prentice-Hall.

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

Kaçıranlar, S., Sakallioglu, S., Ozkale, M., R. and Guler, H. (2011). More on the restricted ridge regression estimation. *Journal of Statistical Computation and Simulation*. **81**(11), 1433 – 1448

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

Khan, B. U. and Ahmed, S. E. (2003). Improved estimation of coefficient vector in a regression model. *Communications in Statistics - Simulation and Computation*. **32**(3), 747 – 769.

Kibria, B. M. G. (1996a). On preliminary test ridge regression estimator for the restricted linear model with non-normal disturbances. *Commun. Stat. Theory Meth*. **25**, 2349 – 2369.

Kibria, B. M. G. (1996b). On shrinkage ridge regression estimators for restricted linear models with multivariate t disturbances Students. **1**(3), 177 – 188.

Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Commun. Statist. Simul. Comput.* **B32**, 429 – 435.

Kibria, B. M. G. (2004).Performance of the shrinkage preliminary test ridge regression estimators based on the conflicting of W, LR and LM tests. *Journal of Statistical Computation and Simulation*. **74**(11), 793 – 810.

Kibria, B. M. G. and Saleh, A. K. M. E. (1993). Performance of shrinkage preliminary test estimator in regression analysis. *Jahangirnagar Review*.**A17**,133 –148.

Kibria, B.M. G. and Saleh, A.K. M. E. (2003). Effect of W, LR, and LM tests on the performance of preliminary test ridge regression estimators. *J. Jpn. Stat. SOC.***33**,119 – 136.

Kibria, B. M. G. and Saleh, A. K. M. E. (2004a). Preliminary test ridge regression estimators with Student's t errors and conflicting test-statistics. *Metrica***59**, 105 – 124.

Kibria, B. M. G. and Saleh, A. K. M. E. (2004b). Performance of Positive Rule Ridge Regression Estimators for the Ill-Conditioned Gaussian Regression Models. *Calcutta Stat. Assoc. Bull*. **55**, 219 – 241.

Liang, H., (2006). Estimation partially linear models and numerical comparison. *Computational Statistics &Data Analysis*, **50**, 675 – 687.

Montgomery, D. C., Peck, E. A. and Vining, G. G. (2001). *Introduction to Linear Regression Analysis*. 3rd ed. Wiley, New York.

Najarian, S., Arashi, M. and Kibria, B. M. (2013). A Simulation Stduy on Some Restricted Ridge Regression Estimators. *Commun. Statist. Simul. Comput.* **42**, 871 – 890.

Nkurunziza, S. and Ahmed, S. E. (2011). Estimation strategies for the regression coefficient parameter matrix in multivariate multiple regression. *Statistica Neerlandica*, **65**(4), 387-406.

Osborne, B.G., Fearn, T., Miller, A.R. and Douglas, S. (1984). Application of Near-Infrared Reflectance Spectroscopy to Compositional Analysis of Biscuits and Biscuit Dough. *Journal of the Science of Food and Agriculture*, **35**, 99 –105.

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

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

Rice, J. (1986). Convergence rates for partially spline models, *Statis. Prob. Lett.* **4**, 203 – 208.

Robinson, M. P. (1988). Root-n-consistent semi-parametric regression. *Econometrica*. **56**(4), 931 – 954.

Roozbeh, M. and Arashi, M. (2013). Feasible ridge estimator in partially linear models. *Journal of Multivariate Analysis*. **116**, 35 – 44.

Roozbeh, M., Arashi, M. and Niroumanda, H.A. (2011a).Semiparametric Ridge Regression Approach in Partially Linear Models. *Communications in Statistics-Simulation and Computation*. **39**, 449 – 460.

Roozbeh, M., Arashi, M. and Niroumanda, H.A. (2011b). Ridge regression methodology in partial linear models with correlated errors. *Journal of Statistical Computation and Simulation*. **81**(4), 517 – 528.

Roozbeh, M., Arashi, M., Gasparini, M. (2012). Seemingly Unrelated Ridge Regression in Semiparametric Models, *Communications in Statistics-Theory and Methods*. **41**, 1364 – 1386.

Saleh, A. K. M. E. (2006). *Theory of Preliminary Test and Stein-Type Estimation with Applications*. Wiley

Saleh, A.K.M.E. and Kibria, B.M.G. (1993). Performances of some new preliminary test ridge regression estimators and their properties. *Commun. Stat. Theory Meth.* **22**, 2747 – 2764.

Sarkar, N. (1992). A new estimator combining the ridge regression and the restricted least squares methods of estimation. *Commun. Statist. Theory Meth*. **21**,1987 – 2000.

Schimek G. M. (2000). *Smoothing and Regression: Approaches, Computation, and Application*. John Willey&Sons, Inc. USA, 19 – 39.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*. **6**(2), 461 – 464.

Scolve, S.L., Morris, C. and Radhakrishnan, R. (1972). Optimality of preliminary test estimation for the multinormal mean. *Ann. Math. Statist*. **43**, 1481–1490.

Singh, S. and Tracy, D. S. (1999). Ridge-regression using scrambled responses. *Metrika*. 147 – 157.

Speckman, P. (1988). Kernel smoothing in partially linear model, *J. Royal Statist., Soc. B*. **50**, 413 – 436.

Stamey, T., Kabalin, J., McNeal, J., Johnstone, I., Freiha, F., Redwine, E. and Yang, N. (1989) Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. II. Radical prostatectomy treated patients. J*ournal of Urology*. **16**, 1076 – 1083.

Stein, C. (1956). The admissibility of hotelling's t^2-test. *Mathematical Statistics*. **27**, 616 – 623.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*. 267 – 288.

Tabakan, G. and Akdeniz, F. (2010). Difference-based ridge estimator of parameters in partial linear model, *Stat Papers*, **51**, 357 – 368.

Tabatabaey, S. M. (1995). *Preliminary test approach estimation: regression model with spherically symmetric errors*. Ph.D. thesis. Carleton University, Ottawa.

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regular- ization method. Soviet Math Dok– English translation of Dokl Akad Nauk SSSR 151, 1963, 501-504, 4:1035–1038.

van Houwelingen, J. C. (2001). Shrinkage and penalized likelihood as methods to improve predictive accuracy. *Statistica Neerlandica,* **55**(1), 17–34

Vinod, H. D. and Ullah, A. (1981). *Recent advances in regression methods*. Marcel Dekker, New York.

Wahba, G. (1990). *Spline Model For Observational Data*. Siam, Philadelphia, PA: SIAM.

Wencheko, E. (2000). Estimation of the signal-t-noise in the linear regression model. *Statist. Papers*. **41**, 327 – 343.

Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H. and Hattersley, A. T. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, **316**(5829), 1336-1341.

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

APPENDIX A

The distributions of $\widehat{\boldsymbol{\beta}}_1^{RFM}$ and $\widehat{\boldsymbol{\beta}}_1^{RSM}$ are given by

$$\vartheta_1 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{RFM} - \boldsymbol{\beta}_1\right) \xrightarrow{\mathscr{D}} N_{p_1}\left(-\boldsymbol{\mu}_{11.2}, \sigma^2\boldsymbol{Q}_{11.2}^{-1}\right) \text{ and} \tag{4.1}$$

$$\vartheta_2 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{RSM} - \boldsymbol{\beta}_1\right) \xrightarrow{\mathscr{D}} N_{p_1}\left(-\boldsymbol{\gamma}, \sigma^2\boldsymbol{Q}_{11}^{-1}\right), \tag{4.2}$$

where $\boldsymbol{\gamma} = \boldsymbol{\mu}_{11.2} + \boldsymbol{\delta}$ and $\boldsymbol{\delta} = \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\omega}$.

To obtain the relationship between sub-model and full model estimators of $\boldsymbol{\beta}_1$, we use following equation by using $\widetilde{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2^{RFM}$

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1^{RFM} &= \underset{\boldsymbol{\beta_1}}{\arg\min}\left\{\|\widetilde{\boldsymbol{y}} - \boldsymbol{X}_1\boldsymbol{\beta}_1\| + \lambda^R\|\boldsymbol{\beta}_1\|^2\right\} \\
&= \left(\boldsymbol{X}_1^\top\boldsymbol{X}_1 + \lambda^R\boldsymbol{I}_{p_1}\right)^{-1}\boldsymbol{X}_1^\top\widetilde{\boldsymbol{y}} \\
&= \left(\boldsymbol{X}_1^\top\boldsymbol{X}_1 + \lambda^R\boldsymbol{I}_{p_1}\right)^{-1}\boldsymbol{X}_1^\top\boldsymbol{y} - \left(\boldsymbol{X}_1^\top\boldsymbol{X}_1 + \lambda^R\boldsymbol{I}_{p_1}\right)^{-1}\boldsymbol{X}_1^\top\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2^{RFM} \\
&= \widehat{\boldsymbol{\beta}}_1^{RSM} - \left(\boldsymbol{X}_1^\top\boldsymbol{X}_1 + \lambda^R\boldsymbol{I}_{p_1}\right)^{-1}\boldsymbol{X}_1^\top\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2^{RFM}.
\end{aligned} \tag{4.3}$$

Under local alternative $\{K_n\}$ and with the equations (2.7)-(2.8), one can obtain the following distribution result by using the equation 4.3,

$$\vartheta_3 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{RFM} - \widehat{\boldsymbol{\beta}}_1^{RSM}\right) \xrightarrow{\mathscr{D}} N_{p_1}\left(\boldsymbol{\delta}, \boldsymbol{\Phi}_*\right), \tag{4.4}$$

where $\boldsymbol{\Phi}_* = -\sigma^2\boldsymbol{Q}_{12}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}$.

Now, it can be easly written the following proposition under favour of the equations 4.1 – 4.4 .

Proposition 15. Under local alternative $\{K_n\}$ as $n \to \infty$ we have

$$\begin{pmatrix} \vartheta_1 \\ \vartheta_3 \end{pmatrix} \sim N_{p_1+p_2}\left[\begin{pmatrix} -\boldsymbol{\mu}_{11.2} \\ \boldsymbol{\delta} \end{pmatrix}, \begin{pmatrix} \sigma^2\boldsymbol{Q}_{11.2}^{-1} & \boldsymbol{\Phi}_* \\ \boldsymbol{\Phi}_* & \boldsymbol{\Phi}_* \end{pmatrix}\right],$$

$$\begin{pmatrix} \vartheta_3 \\ \vartheta_2 \end{pmatrix} \sim N_{p_1+p_2}\left[\begin{pmatrix} \boldsymbol{\delta} \\ -\boldsymbol{\gamma} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Phi}_* & \boldsymbol{0} \\ \boldsymbol{0} & \sigma^2\boldsymbol{Q}_{11}^{-1} \end{pmatrix}\right].$$

Using the equation 4.3, we can also obtain $\boldsymbol{\Phi}_*$ as follows:

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

To present the proofs of Theorem 12, we use the definition of asymptotic bias for given an estimator $\boldsymbol{\beta}_1^*$ is as follows:

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

Proof of Theorem 12. Here, we provide the proof of bias expressions by using Proposition 16 as following:

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

To verify the asymptotic bias of $\widehat{\boldsymbol{\beta}}_1^{RSM}$, we use the equations (4.3) and (4.5) . Hence, it can be written as follows:

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

To obtain the asymptotic bias of $\widehat{\boldsymbol{\beta}}_1^{RPT}$, we use definitions of ADB and the pretest estimator. So, it can be written as follows:

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

Similarly, we get the asymptotic bias of $\widehat{\boldsymbol{\beta}}_1^{RS}$ as follows:

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

Lastly, we verify the asymptotic bias of $\widehat{\boldsymbol{\beta}}_1^{RPS}$ as follows:

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

In this part, we present how to get the asymptotic covariances of the estimators. The asymptotic covariance of an estimator $\boldsymbol{\beta}_1^*$ is defined as follows:

$$\Gamma\left(\boldsymbol{\beta}_1^*\right) \;=\; E\left\{\lim_{n\to\infty} n\left(\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_1\right)\left(\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_1\right)^\top\right\}. \tag{4.6}$$

In the following proof, we use the equation (4.6) .

Proof of Theorem 13. Firstly, the asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{RFM}$ is given by

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

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

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

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

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

Now, by using the following formula for a conditional mean of a bivariate normal, we have

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

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

then,

$$
\begin{aligned}
&\Gamma\left(\widehat{\boldsymbol{\beta}}_1^{RPT}\right)\\
&= \boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^{\top}+2\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad \sigma^2\boldsymbol{Q}_{11.2}^{-1}-\boldsymbol{\Phi}_{*}H_{p_2+2}\left(\chi^2_{p_2,\alpha};(\Delta)\right)-\boldsymbol{\delta}\boldsymbol{\delta}^{\top}H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad +2\boldsymbol{\delta}\boldsymbol{\delta}^{\top}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&= \sigma^2\boldsymbol{Q}_{11.2}^{-1}+\boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^{\top}+2\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad +\sigma^2\boldsymbol{Q}_{12}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad +\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\left[2H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)-H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\right].
\end{aligned}
$$

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

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

Now, by using the following formula for a conditional mean of a bivariate normal,

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

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

we have,

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

Finally, the asymptotic covariance matrix of positive shrinkage ridge regression estimator is derived as follows:

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

From the definition of $\widehat{\boldsymbol{\beta}}_1^{RPS}$, we have

$$
\begin{aligned}
& \Gamma\left(\widehat{\boldsymbol{\beta}}_1^{RPS}\right) \\
= \; & \Gamma\left(\widehat{\boldsymbol{\beta}}_1^{RS}\right) - 2E\left\{\vartheta_3\vartheta_1^{\top}\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\}I\left(\mathscr{L}_n \leq p_2-2\right)\right\} \\
& +2E\left\{\vartheta_3\vartheta_3^{\top}(p_2-2)\mathscr{L}_n^{-1}I\left(\mathscr{L}_n \leq p_2-2\right)\right\} \\
& -2E\left\{\vartheta_3\vartheta_3^{\top}(p_2-2)^2\mathscr{L}_n^{-2}I\left(\mathscr{L}_n \leq p_2-2\right)\right\} \\
& +E\left\{\vartheta_3\vartheta_3^{\top}I\left(\mathscr{L}_n \leq p_2-2\right)\right\} \\
& -2E\left\{\vartheta_3\vartheta_3^{\top}(p_2-2)\mathscr{L}_n^{-1}I\left(\mathscr{L}_n \leq p_2-2\right)\right\} \\
& +E\left\{\vartheta_3\vartheta_3^{\top}(p_2-2)^2\mathscr{L}_n^{-2}I\left(\mathscr{L}_n \leq p_2-2\right)\right\} \\
= \; & \Gamma\left(\widehat{\boldsymbol{\beta}}_1^{RS}\right) - 2E\left\{\vartheta_3\vartheta_1^{\top}\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\}I\left(\mathscr{L}_n \leq p_2-2\right)\right\} \\
& -E\left\{\vartheta_3\vartheta_3^{\top}(p_2-2)^2\mathscr{L}_n^{-2}I\left(\mathscr{L}_n \leq p_2-2\right)\right\} \\
& +E\left\{\vartheta_3\vartheta_3^{\top}I\left(\mathscr{L}_n \leq p_2-2\right)\right\}.
\end{aligned}
$$

Now, by using the following formula for a conditional mean of a bivariate normal,

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

we have

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

It can be easily derived the asymptotic risks of the estimators by following the definition of ADR

$$
\begin{aligned}
R\left(\boldsymbol{\beta}_1^*\right) &= nE\left[\left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1\right)^{\top} \boldsymbol{W}\left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1\right)\right] \\
&= n\mathrm{tr}\left[\boldsymbol{W} E\left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1\right)\left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1\right)^{\top}\right] \\
&= \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{\Gamma}^*\right),
\end{aligned} \tag{4.7}
$$

where $\boldsymbol{\Gamma}^*$ is the covariance matrix of $\boldsymbol{\beta}_1$.

Proof of Theorem 14. By using the equation (4.7) .

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

APPENDIX B

The distributions of $\widehat{\boldsymbol{\beta}}_1^{SRFM}$ and $\widehat{\boldsymbol{\beta}}_1^{SRSM}$ are given by

$$\vartheta_1 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{SRFM} - \boldsymbol{\beta}_1\right) \xrightarrow{\mathscr{D}} N_{p_1}\left(-\boldsymbol{\eta}_{11.2}, \sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1}\right) \text{ and} \tag{4.8}$$

$$\vartheta_2 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{SRSM} - \boldsymbol{\beta}_1\right) \xrightarrow{\mathscr{D}} N_{p_1}\left(-\boldsymbol{\xi}, \sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1}\right) \tag{4.9}$$

where $\boldsymbol{\xi} = \boldsymbol{\eta}_{11.2} + \tilde{\boldsymbol{\delta}}$ and $\tilde{\boldsymbol{\delta}} = \tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\boldsymbol{\omega}$.

To obtain the relationship between sub-model and full model estimators of $\boldsymbol{\beta}_1$, we use following equation by using $\boldsymbol{y}^* = \tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}_2\widehat{\boldsymbol{\beta}}_2^{SRFM}$

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1^{SRFM} &= \underset{\boldsymbol{\beta_1}}{\arg\min}\left\{\left\|\boldsymbol{y}^* - \tilde{\boldsymbol{X}}_1\boldsymbol{\beta}_1\right\| + \lambda^R\left\|\boldsymbol{\beta}_1\right\|^2\right\} \\
&= \left(\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{X}}_1 + \lambda^R\boldsymbol{I}_{p_1}\right)^{-1}\tilde{\boldsymbol{X}}_1^\top\boldsymbol{y}^* \\
&= \left(\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{X}}_1 + \lambda^R\boldsymbol{I}_{p_1}\right)^{-1}\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{y}} - \left(\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{X}}_1 + \lambda^R\boldsymbol{I}_{p_1}\right)^{-1}\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{X}}_2\widehat{\boldsymbol{\beta}}_2^{SRFM} \\
&= \widehat{\boldsymbol{\beta}}_1^{SRSM} - \left(\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{X}}_1 + \lambda^R\boldsymbol{I}_{p_1}\right)^{-1}\tilde{\boldsymbol{X}}_1^\top\tilde{\boldsymbol{X}}_2\widehat{\boldsymbol{\beta}}_2^{SRFM}.
\end{aligned} \tag{4.10}$$

Under local alternative $\{K_n\}$ and with the equations (2.7)-(2.8), one can obtain the following distribution result by using the equation 4.10,

$$\vartheta_3 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{SRFM} - \widehat{\boldsymbol{\beta}}_1^{SRSM}\right) \xrightarrow{\mathscr{D}} N_{p_1}\left(\tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\Phi}}_*\right), \tag{4.11}$$

where $\tilde{\boldsymbol{\Phi}}_* = -\sigma^2\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}$.

Now, it can be easly written the following proposition under favour of the equations 4.8 – 4.11.

Proposition 16. Under local alternative $\{K_n\}$ as $n \to \infty$ we have

$$\begin{pmatrix} \vartheta_1 \\ \vartheta_3 \end{pmatrix} \sim N_{p_1+p_2}\left[\begin{pmatrix} -\boldsymbol{\eta}_{11.2} \\ \tilde{\boldsymbol{\delta}} \end{pmatrix}, \begin{pmatrix} \sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1} & \tilde{\boldsymbol{\Phi}}_* \\ \tilde{\boldsymbol{\Phi}}_* & \tilde{\boldsymbol{\Phi}}_* \end{pmatrix}\right]$$

$$\begin{pmatrix} \vartheta_3 \\ \vartheta_2 \end{pmatrix} \sim N_{p_1+p_2}\left[\begin{pmatrix} \tilde{\boldsymbol{\delta}} \\ -\boldsymbol{\xi} \end{pmatrix}, \begin{pmatrix} \tilde{\boldsymbol{\Phi}}_* & \boldsymbol{0} \\ \boldsymbol{0} & \sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1} \end{pmatrix}\right]$$

Using the equation 4.10, we can also obtain $\tilde{\boldsymbol{\Phi}}_*$ as follows:

$$\begin{aligned}
\tilde{\boldsymbol{\Phi}}_* &= Cov\left(\widehat{\boldsymbol{\beta}}_1^{SRFM} - \widehat{\boldsymbol{\beta}}_1^{SRSM}\right) \\
&= E\left[\left(\widehat{\boldsymbol{\beta}}_1^{SRFM} - \widehat{\boldsymbol{\beta}}_1^{SRSM}\right)\left(\widehat{\boldsymbol{\beta}}_1^{SRFM} - \widehat{\boldsymbol{\beta}}_1^{SRSM}\right)^\top\right] \\
&= E\left[\left(\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\widehat{\boldsymbol{\beta}}_2^{SRFM}\right)\left(\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\widehat{\boldsymbol{\beta}}_2^{SRFM}\right)^\top\right] \\
&= \tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}E\left[\widehat{\boldsymbol{\beta}}_2^{SRFM}\left(\widehat{\boldsymbol{\beta}}_2^{SRFM}\right)^\top\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\right] \\
&= \sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}.
\end{aligned}$$

In the light of all these above equations, the proofs of theorems can be shown by using similar the way of proofs in Appendix A.

CURRICULUM VITAE

Name and Surname	:	Bahadır Yüzbaşı
Place and Date of Birth	:	Elbistan – 1985
Address	:	Department of Econometrics, İnönü University, Malatya, Turkey
E-Mail	:	bahadir.yuzbasi@inonu.edu.tr
Bachelor of Science	:	Department of Mathematics, Firat University, Elazig, Turkey
Master of Science	:	Department of Mathematics, Graduate School of Natural and Applied Sciences, Fırat University, Elazığ, Turkey

Academic Experience : In 2008, he completed his BSc in Department of Mathematics, Fırat University, Elazığ, Turkey. In 2009, he has worked as a research assistant in Department of Econometrics, İnönü University, Malatya, Turkey. In 2010, he completed his MSc in Department of Mathematics at Fırat University, Elazığ, Turkey. Between years 2013 and 2014, he was assigned as research assistant in Department of Mathematics and Statistics, Brock University, Canada.

Reward : Tubitak 2214/A – International Doctorate Research Fellowship Program. Between January 2013 – January 2014, at the Department of Mathematics and Statistics, Brock University, Canada.

Articles published in international refereed journals :

Güngör, M., Bulut, Y. and **Yüzbaşı, B**. (2012). On joint distributions of order statistics for a discrete case. *Journal of Inequalities and Applications* **2012.1**, 1–12. (SCI)

Kayhan, S., Bayat, T. and **Yüzbaşı, B**. (2013). Government expenditures and trade deficits in Turkey: Time domain and frequency domain analyses. *Economic Modelling*, **35**, 153–158. (SSCI)

Yüzbaşı, B. and Güngör, M. (2013). On Joint Distributions of Order Statistics from Nonidentically Distributed Discrete Vectors, *Journal of Computational Analysis and Applications*, **15**(5), 917–927. (SCI)