2013

Absolute penalty and shrinkage estimation strategies in linear and partially linear models with correlated errors

Saber Fallahpour
University of Windsor

ABSOLUTE PENALTY AND SHRINKAGE ESTIMATION STRATEGIES IN LINEAR AND PARTIALLY LINEAR MODELS WITH CORRELATED ERRORS

by

Saber Fallahpour

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada

Your file Votre référence
ISBN: 978-0-494-79308-4

Our file Notre référence
ISBN: 978-0-494-79308-4

Absolute Penalty and Shrinkage Estimation Strategies in Linear and Partially Linear Models With Correlated Errors

by

Saber Fallahpour

APPROVED BY

Dr. Yulia Gel, External Examiner
University of Waterloo

Dr. L. Rueda
School of Computer Science

Dr. M. Hlynka
Department of Mathematics and Statistics

Dr. A. A. Hussein
Department of Mathematics and Statistics

Dr. S. E. Ahmed, Advisor
Department of Mathematics and Statistics

Dr. N. Ursel, Chair of Defense
Faculty of Graduate Studies

November 26, 2012

Declaration of Co-Authorship/ Previous Publication

I. Co-Authorship Declaration

I hereby declare that this thesis incorporates the outcome of joint research undertaken in collaboration with my supervisor, Professor S. Ejaz Ahmed. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contribution of co-author was primarily through the provision of some theoretical results.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-authors to include in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

II. Declaration of Previous Publication

This thesis includes two original papers that have been previously published, one paper which has been accepted, and another which has received invitation for submission.

Thesis Chapter	Publication title/ full citation	Publication Status
Chapter 2	High and Low Dimensional Data Analysis in Multiple Regression Models with Random Coefficient Autoregressive Errors. *Journal of Bernoulli*	Submitted
Chapter 3	L_1 Penalty and Shrinkage Estimation in Partially Linear Models with Random Coefficient Autoregressive Errors, *Journal of Applied Stochastic Models in Business and Industry*, 28(3), 236-250, 2012	Published
Chapter 4	Shrinkage Estimation Strategy in Quasi Likelihood Models. *Statistics & Probability Letters*, 82(12), 2170-2179, 2012	Published
Chapter 4	Variable Selection and Post-Estimation of regression Parameters Using Quasi-Likelihood Approach. *Proceeding of the 9th Tartu Conference on Multivariate Statistics*	Accepted

I certify that I have obtained written permission from the copyright owners to include the above published materials in my thesis. I certify that the above material describes work completed during my registration as graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained written permission from the copyright owner to include such material in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University of Institution.

Abstract

In this dissertation we propose shrinkage estimators and absolute penalty estimators (APEs) in linear models, partially linear models (PLM) and quasi-likelihood models. We study the asymptotic properties of shrinkage estimators both analytically and through simulation studies, and compare their performance with APEs.

In Chapter 2, we propose shrinkage estimators for a multiple linear regression with first order random coefficient autoregressive (RCAR(1)) error term. We also present two APEs for this models which are modified versions of lasso and adaptive lasso estimators. We compare the performance of shrinkage estimators and APEs through the mean squared error criterion. Monte Carlo studies were conducted to compare the estimators in two situations: when $p > n$ and when $p < n$. A data example is presented to illustrate the usefulness of the suggested methods.

In Chapter 3, we develop shrinkage estimators for a PLM with RCAR(1) error term. The nonparametric function is estimated using a kernel function. We also compare the performance of shrinkage estimators with a modified version of lasso for correlated data. Monte Carlo studies were conducted to compare the behavior of the proposed estimators. A data example is presented to illustrate the application of the suggested methods.

In Chapter 4, we propose pretest and shrinkage estimators for quasi-likelihood models. We investigate the asymptotic properties of these estimators both analytically and through simulation studies. We also apply a lasso estimator and compare its performance with the other proposed estimators.

Acknowledgements

It is a pleasure to thank the people who made this dissertation possible. Foremost, I would like to express my sincere gratitude to my advisor Prof S. Ejaz Ahmed for the continuous support of my doctoral studies and research. I am thankful to him for his patience and immense knowledge without which it would not have been possible to complete the work in time.

I would like to extend my thanks to external examiner Dr. Yulia Gel, and to the advisory committee members - Dr. Myron Hlynka, Dr. Abdul Hussein and Dr. Luis Rueda for reviewing the dissertation and providing with valuable suggestions that has improved my dissertation.

During my stay at University of Windsor, I had the opportunity to meet people whose friendship and association I will cherish all my life. The small encouraging talks with Dr. Paul, Dr. Nkurunziza, Dr. Barron and Dr. Caron and all other faculty members made the day-to-day travail enjoyable. I am also grateful to the department secretaries for assisting me in many different ways, Dina Labelle and Rose Spence. I acknowledge my debt to all of them for their affectionate concern about my work and well being. Thanks are also due to Tanvir Quadir and Enayetur Raheem for their excellent friendship during my studies at this university.

I would also like to thank my wonderful and lovely wife Mahsa. Just as this chapter is closing, you have come into my life with open arms welcoming me into the next and most important chapter of our lives together as husband and wife.

Lastly, and most importantly, my deepest gratitude goes to my mother Eftekhar Tofighi, my father Mohammadreza Fallahpour and my brother Sajed Fallahpour, for their love, encouragement, and continuous support. My parents and their expectations have been a constant source of inspirations throughout my life. They supported me spiritually all the time and no words of gratitude would be enough to acknowledge their contributions. I am extremely grateful to all your patience, support and prayers.

The research of this dissertation was further supported by the Ontario Graduate Scholarship (OGS), Ontario Graduate Scholarship in Science and Technology (OGSST), University of Windsor Doctoral Tuition Scholarships and graduate assistantships provided by Department of Mathematics and Statistics, University of Windsor.

Saber Fallahpour
Nov 26, 2012
Windsor, Ontario, Canada

Contents

List of Figures

List of Tables

Abbreviations

ADB	asymptotic distributional bias
ADR	asymptotic distributional risk
AL	adaptive lasso
APE	absolute penalty estimator/estimation
APEs	absolute penalty estimators
CV	cross validation
LAR	least angle regression
Lasso	least absolute shrinkage and selection operator
MSE	mean squared error
NSI	non-sample information
PLM	partially linear model
PMSE	predictive mean squared error
PSE	positive shrinkage estimator
PTE	pretest estimator
QL	quasi-likelihood
RCAR	random coefficient autoregressive
RE	restricted estimator
RMSE	relative mean squared error
RPMSE	relative predictive mean squared error
SE	shrinkage estimator
UE	unrestricted estimator
UPI	uncertain prior information

List of Symbols

$\boldsymbol{\beta}$	parameter vector
$\boldsymbol{\beta}_1$	parameter vector of main variables
$\boldsymbol{\beta}_2$	parameter vector of nuisance variables
p	the number of regression parameters
n	sample size
H_0	null hypothesis
T_n	test statistic
λ	tuning parameter
$\hat{\boldsymbol{\beta}}$	unrestricted estimator
$\tilde{\boldsymbol{\beta}}$	restricted estimator
$\hat{\boldsymbol{\beta}}^{S}$	shrinkage estimator
$\hat{\boldsymbol{\beta}}^{S+}$	positive shrinkage estimator
$\hat{\boldsymbol{\beta}}^{PT}$	pretest estimator
$\hat{\boldsymbol{\beta}}^{lasso}$	lasso estimator
$\hat{\boldsymbol{\beta}}^{AL}$	adaptive lasso estimator
$I(A)$	indicator function
$\boldsymbol{M}$	positive semi-definite weight matrix in the quadratic loss function
$\boldsymbol{\Gamma}(\cdot)$	asymptotic variance covariance matrix of an estimator
$R(\cdot)$	asymptotic distributional risk of an estimator
$\boldsymbol{F}$	$p \times q$ matrix
$\boldsymbol{d}$	$q \times 1$ vector of constants

$\mathcal{L}$	weighted loss function
K_n	local alternative hypothesis
$H_v(x;\Delta)$	noncentral chi-square distribution function with non-centrality parameter Δ and v degrees of freedom
$\boldsymbol{\omega}$	a fixed real valued vector in K_n
Δ	non-centrality parameter
$F(\boldsymbol{x})$	asymptotic distributional function
tr	trace of a matrix

Chapter 1

Background

1.1 Introduction

Statistical models are used to describe the relationships between the response(s) or dependent variable(s) and a set of explanatory variables or predictors. The basic form of these models can be written in the following form

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

where $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)$ is the vector of responses, $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p)$ are the predictors, $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)$ is an unknown vector of parameters, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)$ is the vector of unobservable random errors. If appropriate, such models can be used to predict the value of the response variable for a set of known values of the predictors and for any such prediction, estimation of the model parameters ($\boldsymbol{\beta}$), is essential. Estimation of parameters is also essential for performing statistical tests on any individual or set of model parameters. However, the commonly-used classical estimators of the unknown parameters of the statistical models are based on the sample information. In real life situations, researchers may have some prior information on the parameters either in the form of a prior distribution or as a constraint on some (all) of the parameters. The source of such prior information can be extracted from previous studies, empirical work or expert knowledge.

The prior distribution of a parameter is used in the Bayesian approach to statistical analysis. However, if the prior information about the parameters is available as values of parameters or relationships among them rather than as a distribution, the Bayesian approach cannot be implemented. There are however other estimation methods that use this kind of prior information in addition to the sample information. The inclusion of such additional information in the estimation process would result in a better estimator than using sample information alone. This additional information is called non-sample information (NSI) or uncertain prior information (UPI) and can be expressed in the form of a general linear constraint such as

$$H_0 : \boldsymbol{F}'\boldsymbol{\beta} = \boldsymbol{d}, \tag{1.2}$$

where $\boldsymbol{F}$ is a $p \times q$ full rank matrix with rank $q \leq p$, $\boldsymbol{d}$ is a given $q \times 1$ vector of constants and $\boldsymbol{\beta}$ is a $p \times 1$ vector of model parameters. A model with no restriction on parameters is called a full model and a model with restrictions given in (1.2) is a reduced model or sub-model.

As a specific form of the above null hypothesis, consider the case when $\boldsymbol{F}' = (\boldsymbol{0}, \boldsymbol{I})$ where $\boldsymbol{I}_{p_2 \times p_2}$ is the identity matrix, $\boldsymbol{0}_{p_2 \times p_1}$ is the matrix of 0s, and $\boldsymbol{d}_{p_2 \times 1} = \boldsymbol{0}$. In this case, the parameter vector $\boldsymbol{\beta}$ can be partitioned to $(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ where $\boldsymbol{\beta}_1$ is the vector of main effects and $\boldsymbol{\beta}_2$ relates to the nuisance parameters which can be excluded from the model. As we see, the general form (1.2) reduces to

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0}. \tag{1.3}$$

Here, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have dimensions of p_1 and p_2 respectively, with $p_1 + p_2 = p$. Now, the question arises as to how one incorporates this UPI into the estimation process of model parameters.

This dissertation deals with improved estimation strategy for the parameter estimation in some statistical models. This strategy is called James-Stein estimation strategy, also known as shrinkage estimation strategy inspired by Stein's result that shows, in a parameter dimension greater than two, efficient estimates can be obtained by shrinking full model estimates in the direction of reduced model estimates. We apply this method to three different models where sample as well as non-sample prior information about the model parameters are available.

In particular, we propose improved estimators for a multiple linear regression model and partially linear model (PLM) with random coefficient autoregressive (RCAR) errors when UPI is available in the form of (1.3). We also propose improved estimators for the quasi-likelihood (QL) models when UPI is given as (1.2). Furthermore, we consider absolute penalty estimators (APEs) for parameter estimation in the above models and for comparison purposes with our proposed Stein-type estimators.

1.1.1 Unrestricted and Restricted Estimators

When an estimator is solely based on sample information and not a function of UPI, it is called the unrestricted estimator (UE). Denote the UE of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$. This estimator can be achieved through different methods such as ordinary least square (OLS), generalized least square (GLS), maximum likelihood (ML), etc. However, when non-sample information on $\boldsymbol{\beta}$ exists as in (1.2), $\hat{\boldsymbol{\beta}}$ may not be efficient anymore. In order to take advantage of the available UPI, we define $\tilde{\boldsymbol{\beta}}$, the restricted estimator (RE) of $\boldsymbol{\beta}$. When UPI holds, $\tilde{\boldsymbol{\beta}}$ will be an unbiased estimator of $\boldsymbol{\beta}$ with smaller variance than the $\hat{\boldsymbol{\beta}}$. However, if the UPI is not true, then $\tilde{\boldsymbol{\beta}}$ will be a biased estimator and the $\hat{\boldsymbol{\beta}}$ will outperform $\tilde{\boldsymbol{\beta}}$.

Therefore, it is natural to combine the sample information as well as the UPI to define an improved estimator that may outperform both $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$, under certain conditions.

1.1.2 Pretest Estimator

Let T_n be the test statistic for the null hypothesis in (1.2) and $c_{q,\alpha}$ be the critical value of the distribution of T_n under H_0. The pretest estimator (PTE) is defined as follows:

$$\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}}\, I(T_n > c_{q,\alpha}) + \tilde{\boldsymbol{\beta}}\, I(T_n < c_{q,\alpha}),$$

where $I(A)$ is an indicator function of set A. If the researcher is uncertain of the accuracy of the UPI, then the procedure usually followed in practice is to pretest the validity of the UPI. If the outcome of the pretest suggests that the UPI is correct, then the parameters are estimated incorporating the restrictions, which leads to $\tilde{\boldsymbol{\beta}}$. However, if the pretest rejects the UPI, then the parameters are estimated from the sample information alone, which leads to $\hat{\boldsymbol{\beta}}$.

If the UPI is nearly correct, then $\hat{\boldsymbol{\beta}}^{PT}$ outperforms $\hat{\boldsymbol{\beta}}$. But for incorrect restrictions, $\hat{\boldsymbol{\beta}}^{PT}$ is a biased estimator since $\tilde{\boldsymbol{\beta}}$ is biased under incorrect restriction.

More useful discussions about this estimator can be found in Bancroft (1944), Giles and

Giles (1993), Albertson (1993), Ahmed (2001), Saleh (2006) and Ahmed and Liu (2009), among others.

To overcome the problem associated with $\hat{\boldsymbol{\beta}}^{PT}$, an improved estimation strategy, called James-Stein or shrinkage estimation strategy, is defined. This estimator, despite $\hat{\boldsymbol{\beta}}^{PT}$, is a continuous function of the test statistic and it performs better than $\hat{\boldsymbol{\beta}}^{PT}$ in the entire parameter space induced by UPI.

1.1.3 Shrinkage and Positive Shrinkage Estimators

The shrinkage estimator (SE) of the parameter vector $\boldsymbol{\beta}$ is defined as:

$$\hat{\boldsymbol{\beta}}^{S} = \tilde{\boldsymbol{\beta}} + (1 - c_{opt}T_n^{-1})(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}),$$

where c_{opt} is the optimal constant that minimizes the risk. As mentioned earlier, this estimator is a continuous function of the test statistic T_n; and the binary function of $I(A)$ in $\hat{\boldsymbol{\beta}}^{PT}$ is replaced by the continuous function $c_{opt}T_n^{-1}$. To avoid the over-shrinking problem in SE, we define the positive shrinkage estimator (PSE):

$$\hat{\boldsymbol{\beta}}^{S+} = \tilde{\boldsymbol{\beta}} + (1 - c_{opt}T_n^{-1})^{+}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}),$$

where $z^{+} = max(0, z)$. This estimator outperforms $\hat{\boldsymbol{\beta}}^{S}$ and will control the possible over-shrinking in $\hat{\boldsymbol{\beta}}^{S}$. Both $\hat{\boldsymbol{\beta}}^{S}$ and $\hat{\boldsymbol{\beta}}^{S+}$ uniformly dominate $\hat{\boldsymbol{\beta}}$.

More useful discussions about these estimators can be found in Stein (1956), James and Stein (1961), Ahmed and Saleh (1989), Ahmed (1997), Ahmed and Krzanowski (2004), Ahmed et al. (2007) and Ahmed et al. (2010), among others.

1.1.4 Absolute Penalty Estimator

Absolute Penalty Estimators (APEs) are a class of estimators in the penalized least square family and since the absolute value of the penalty term is considered for estimation process they are known as APE. One of the most commonly used class of APEs is the L_1 penalized least square estimator or the lasso (least absolute shrinkage and selection operator) proposed by Tibshirani (1996) which performs both variable selection and parameter estimation simultaneously. Lasso has become a popular model selection strategy since it shrinks some of the coefficients and sets some of them exactly equal to zero.

As we know, in regression models the parameters are estimated by minimizing the residual sums of squares:

$$\min_{\boldsymbol{\beta}} (\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}),$$

but lasso imposes an additional constraint on the coefficients,

$$\sum_{j=1}^{p} |\beta_j| \leq \tau,$$

where τ is the tuning parameter. When τ is large enough, the constraint has no effect and the solution will be the usual least square estimates; however, for small value of τ the solutions are shrunk estimates often with some of them equal to zero. Thus, choosing τ can be thought of as choosing the number of predictors to include in a regression model.

Note that the lasso results are similar to the shrinkage method by both shrinking and deleting coefficients. However, it is different from the shrinkage procedure in that it treats all the covariate coefficients equally.

Later, some other APEs were introduced by researchers. Fan and Li (2001) introduced the smoothly clipped absolute deviation (SCAD) approach. Efron et al. (2004) introduced the Least Angle Regression algorithm and its connection to lasso. Zou (2006) introduced adaptive lasso (AL) which uses a weighted L_1 penalty. Park and Hastie (2007) developed L_1 regularization paths for generalized linear models. Ahmed et al. (2007) proposed an APE for partially linear models (PLM), which is an extension of the lasso method for linear models. Nowadays, APEs are frequently being used for variable selection and estimation strategy in problems with fixed and high dimensional data. In our dissertation we also implement the APE and compare the results with the proposed shrinkage estimators.

1.2 Appraisal of the Estimators

1.2.1 Asymptotic Comparison

In this dissertation we study the performance of $\hat{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}^{S}$, $\hat{\boldsymbol{\beta}}^{S+}$ and $\hat{\boldsymbol{\beta}}^{PT}$ using the notion of asymptotic distributional bias (ADB) and asymptotic distributional risk (ADR). In general, it is not easy to achieve the finite sample properties of the shrinkage and pretest estimators for non-normal models. Asymptotic methods have overcome this difficulty (Ahmed, 1991, Ahmed, 2001, and others) which is related to convergence in distribution, but does not guarantee convergence in quadratic risk. By implementing the notion of asymptotic distributional risk (ADR), this technicality will be taken care of and it plays a useful role in asymptotic risk analysis.

For this aim, we consider the weighted quadratic loss function criterion to examine the performance of the estimators.

$$\mathcal{L}(\boldsymbol{\beta}^0, \boldsymbol{\beta}) = n(\boldsymbol{\beta}^0 - \boldsymbol{\beta})' \boldsymbol{M} (\boldsymbol{\beta}^0 - \boldsymbol{\beta}),$$

where $\boldsymbol{\beta}^0$ is any one of $\hat{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}^S$, $\hat{\boldsymbol{\beta}}^{S+}$ and $\hat{\boldsymbol{\beta}}^{PT}$ and $\boldsymbol{M}$ is a positive semidefinite matrix. Obviously, when $\boldsymbol{M} = \boldsymbol{I}$, we get the squared loss function. The expectation of the loss function as $n \to \infty$

$$E[\lim_{n\to\infty} \mathcal{L}(\boldsymbol{\beta}^0, \boldsymbol{\beta}); \boldsymbol{M}] = R[(\boldsymbol{\beta}^0, \boldsymbol{\beta}); \boldsymbol{M}],$$

is called the asymptotic risk (AR), which can be written as

$$\begin{aligned} R[(\boldsymbol{\beta}^0, \boldsymbol{\beta}); \boldsymbol{M}] &= E[\lim_{n\to\infty} \mathcal{L}(\boldsymbol{\beta}^0, \boldsymbol{\beta}); \boldsymbol{M}] \\ &= E[\lim_{n\to\infty} n(\boldsymbol{\beta}^0 - \boldsymbol{\beta})' \boldsymbol{M} (\boldsymbol{\beta}^0 - \boldsymbol{\beta})] \\ &= \mathrm{tr}[\boldsymbol{M} E\{\lim_{n\to\infty} n\, (\boldsymbol{\beta}^0 - \boldsymbol{\beta})(\boldsymbol{\beta}^0 - \boldsymbol{\beta})'\}] \\ &= \mathrm{tr}[\boldsymbol{M}\boldsymbol{\Gamma}], \end{aligned}$$

where $\boldsymbol{\Gamma}$ is the asymptotic covariance matrix of $\boldsymbol{\beta}^0$.

We can evaluate the performance of the estimators by comparing the AR with a suitable matrix $\boldsymbol{M}$. The smaller the AR, the better the estimator. If there exists another estimator $\boldsymbol{\beta}^{\diamond}$ such that

$$R[(\boldsymbol{\beta}^{\diamond}, \boldsymbol{\beta}); \boldsymbol{M}] \leq R[(\boldsymbol{\beta}^0, \boldsymbol{\beta}); \boldsymbol{M}] \quad \forall\, (\boldsymbol{\beta}, \boldsymbol{M}) \tag{1.4}$$

with strict inequality for some $\boldsymbol{\beta}$, then the estimator $\boldsymbol{\beta}^0$ will be called an inadmissible estimator. In such cases, we say that the estimator $\boldsymbol{\beta}^{\diamond}$ dominates $\boldsymbol{\beta}^0$.

Ahmed (1997) noted that since the statistic T_n is consistent against fixed alternative, the SE and PSE will be asymptotically equivalent in probability to UE, i.e., the asymptotic distribution of $\sqrt{n}(\boldsymbol{\beta}^0 - \boldsymbol{\beta})$, is equivalent to $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ as $n \to \infty$. Therefore, for large sample situations there is not much to investigate on the estimators. In this case, to obtain meaningful asymptotic results and to evaluate the behavior of the estimators in a neighborhood of the null hypothesis, a class of local alternatives, $\{K_n\}$, is considered, which is given by

$$K_n : \boldsymbol{F}'\boldsymbol{\beta} = \boldsymbol{d} + \frac{\boldsymbol{\omega}}{\sqrt{n}}, \quad \boldsymbol{\omega} \neq 0 \text{ fixed},$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_q) \in \Re^q$ is a real fixed vector. It is obvious that for all n when $\boldsymbol{\omega} = \mathbf{0}$, then $\boldsymbol{F}'\boldsymbol{\beta} = \boldsymbol{d}$. The expression in (1.4) is not easy to prove. An alternative is to consider the asymptotic distributional risk (ADR) for the sequence of local alternatives $\{K_n\}$.

Suppose that the asymptotic cumulative distribution function (cdf) of $\boldsymbol{\beta}^0$ under $\{K_n\}$ exists, and is defined as

$$F(\boldsymbol{x}) = \lim_{n\to\infty} P\{\sqrt{n}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) \leq \boldsymbol{x} | K_n\},$$

where $F(\boldsymbol{x})$ is nondegenerate. Then, the ADR of $\boldsymbol{\beta}^0$ is defined as

$$R(\boldsymbol{\beta}^0, \boldsymbol{M}) = \int \cdots \int \boldsymbol{x}'\boldsymbol{M}\boldsymbol{x}\; dF(\boldsymbol{x}) = \operatorname{tr}(\boldsymbol{M}\boldsymbol{\Gamma}),$$

where $\boldsymbol{M}$ is a positive semidefinite matrix and $\boldsymbol{\Gamma} = \int \cdots \int \boldsymbol{x}\boldsymbol{x}'\; dF(\boldsymbol{x})$ is the dispersion matrix obtained from $F(\boldsymbol{x})$.

We also compare the behavior of the estimators based on the asymptotic distributional bias (ADB). The ADB of the estimator $\boldsymbol{\beta}^0$ under K_n is defined as

$$ADB(\boldsymbol{\beta}^0) = E\{\lim_{n\to\infty} \sqrt{n}(\boldsymbol{\beta}^0 - \boldsymbol{\beta})\}.$$

1.2.2 Monte Carlo Comparison

Along with the bias and risk comparisons of the proposed estimators, we also carry out Monte Carlo simulation studies to investigate the achieved results for $\hat{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}^{S}$ and $\hat{\boldsymbol{\beta}}^{S+}$ and $\hat{\boldsymbol{\beta}}^{PT}$ and also to compare and examine the performance of the suggested estimators with absolute penalty estimators.

1.3 Review of Literature

In the following subsections we give an introduction and literature review of three models that were used in this dissertation.

1.3.1 Multiple Regression Models with Random Coefficient Autoregressive Errors

In Chapter 2 we consider the following multiple regression model:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.5}$$

where y_i's are responses, $\boldsymbol{x}_i$ are known $p \times 1$ vectors of covariates, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is an unknown $p \times 1$ vector of regression parameters, and ε_i's are unobservable random errors. In practice, it is plausible that the independent and identically distributed (i.i.d) assumption of ε_i's may be violated, especially for sequentially collected economic data that often exhibit evident dependence in the errors. One way to model dependence in the error is to use a linear stationary process, for instance, an AR process, an MA process, or an ARMA process.

However, in many fields of research including air pollution, image analysis, economics and finance, linear time series is not the best model to fit. As a result, various nonlinear time series models have been proposed, see, for instance, Tong (1990) and the references therein.

There are many papers that consider ordinary linear models with nonlinear time series error. For example, Weiss (1986) established the consistency and asymptotic normality of the maximum likelihood estimators for a regression model with autoregressive conditional heteroscedastic (ARCH) errors. Bera and Zuo (1996) developed a specification test for a linear regression model with ARCH errors and Dutta (1999) derived the Wald and score tests for additional linear regression parameters.

One of the nonlinear forms of time series models is the random coefficient autoregressive (RCAR) model. Specifically we assume that ε_i in model (1.5) is a first order random coefficient autoregressive process RCAR(1), which is a stationary solution of

$$\varepsilon_i = (\theta + z_i)\varepsilon_{i-1} + e_i, \quad i = 1, \ldots, n, \tag{1.6}$$

where θ is the autoregression parameter and $\{z_i\}$ and $\{e_i\}$ are zero mean independent processes each consisting of i.i.d random variables with finite second moments σ_z^2 and σ_e^2, respectively.

For complete background on this model, we refer to Nicholls and Quinn (1982). Later Akharif and Hallin (2003) introduced a test statistic for detecting randomness in the coefficients of an AR(p) model. Aue et al. (2006) proposed the quasi-maximum likelihood estimator for the parameters of model (1.6) and derived strong consistency and the asymptotic normality of the proposed estimator.

Hwang and Basawa (1993) considered parameter estimation of the model (1.5) with errors given in (1.6) and investigated the limit distribution of the regression and the autoregression parameters. Also, Hwang and Basawa (1997) established the local asymptotic normality for a class of generalized random coefficient autoregressive processes.

In Chapter 2, we use the model in Hwang and Basawa (1993) and propose improved estimation strategy for the parameter vector in the presence of UPI given in (1.3). We obtained restricted, shrinkage and positive shrinkage estimators and presented two absolute penalty estimators for this model which are modified versions of lasso and AL for the correlated errors.

1.3.2 Partially Linear Models with Random Coefficient Autoregressive Errors

In Chapter 3 we consider the following partially linear model (PLM):

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + g(t_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.7}$$

where y_i's are responses, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ and $t_i \in [0,1]$ are design points, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is an unknown parameter vector, $g(\cdot)$ is an unknown bounded real-valued smooth function defined on the compact subset $[0,1]$, and ε_i's are unobservable random errors with mean zero. When ε_i are i.i.d random variables, Heckman (1986), Rice (1986), Chen (1988), Robinson (1988), Speckman (1988), Eubank and Speckman (1990), Chen and Shiau (1991), Donald and Newey (1994), Hamilton and Truong (1997) and Liang and Härdle (1999) used various estimation methods, such as the kernel method, spline method, series estimation, local linear estimation, two-stage estimation and M-estimation to obtain estimators of the regression parameters in model (1.7) and discussed the asymptotic properties of these estimators. Further, Gao (1997) and González-Manteiga and Aneiros-Pérez (2003) discussed the problem of testing for model (1.7). For a more complete review, the reader is referred to the monograph by Härdle et al. (2000) and a book by Horowitz (2009).

The majority of the work done so far, including that mentioned above, assume that the errors are independent. However, the independence assumption is not always practical, specially in areas like economics and finance. Recently, there has been more attraction to model (1.7) with serially correlated errors.

When the error is an AR(1) process, Schick (1994) presented an estimator for the autocorrelation coefficient in the presence of partially linear regression trend. Schick (1996, 1998) further constructed efficient estimators of the regression coefficient and the autocorrelation coefficient, respectively. Gao (1995) considered the estimation problem of the model (1.7) with MA(∞) error process. Furthermore, Aneiros-Pérez and Quintela (2002) and Aneiros-Pérez et al. (2004), You and Chen (2007), among others, have studied the estimation problem of this model with serially correlated errors.

You and Chen (2002) considered the partially linear model with nonlinear time series error. They specifically assumed a RCAR(1) model for the errors and used kernel estimates of the nonparametric function $g(.)$ to investigate the estimation problem and the limit distribution of regression parameters and autocorrelation coefficient.

In chapter 3, we consider the model in You and Chen (2002) and obtain restricted, shrinkage and positive shrinkage estimators. We also present the absolute penalty estimator for this model which is the extended and modified version of lasso and compared its performance with the shrinkage estimators (Fallahpour et al., 2012).

1.3.3 Quasi-likelihood Models

Nelder and Wedderburn (1972) extended general linear models to generalized linear models (GLMs) by including the exponential family of error distribution along with the normal. The GLM requires full distributional assumptions. However, sometimes a full distributional assumption is not possible, especially in discrete data problems. To overcome this problem, Wedderburn (1974) introduced quasi-likelihood (QL) methodology. This model is based on only the first two moments of the response variable and is useful for estimating the mean or the regression parameters. Consider the uncorrelated data y_i with

$$E(y_i) = \mu_i, \quad var(y_i) = \phi V(\mu_i) \quad i = 1, \ldots, n,$$

and

$$g(\mu_i) = \sum_{r=0}^{p} \beta_r x_{ir}, \quad i = 1, \ldots, n,$$

where the link function $g(.)$ and variance function $V(.)$ are assumed known and the dispersion parameter ϕ may be unknown. The constant variance linear regression has $g(\mu) = \mu$, $V(\mu) = 1$ and dispersion parameter denoted by σ^2. Now the quasi-likelihood function is given by

$$Q(\boldsymbol{y}, \boldsymbol{\mu}) = \sum_{i=1}^{n} \left[\int_{y_i}^{\mu_i} \frac{(y_i - t)}{\phi V(t)} \, dt \right].$$

McCullagh (1983) examined the asymptotic properties of the QL function and showed that the estimators enjoy a certain asymptotic optimality property. Firth (1987) investigated the efficiency of the quasi-likelihood estimator under more general distributions than the exponential family. Nelder and Pregibon (1987) and Godambe and Thompson (1989) proposed extended quasi-likelihood (EQL) functions by introducing a normalizing factor to the QL. The EQL resembles a likelihood involving not only the mean (regression) parameters but also the variance parameter of the response variable. Also Lee and Nelder (1996) developed hierarchical likelihood (HL) methodology for joint estimation of the mean and the dispersion parameters in over dispersed models. The HL is based on full distributional assumptions. In order to avoid full distributional assumption at the stages of the hierarchy, Lee and Nelder (2001) introduced the double extended quasi-likelihood (DEQL) function for the joint estimation of the mean and the dispersion parameters.

In Chapter 4, we propose an improved estimation strategy for the QL models in the presence of UPI given in (1.2). We obtain shrinkage, positive shrinkage and pretest estimators. We also apply lasso to this model and compare the performance of all the estimators by simulation studies.

1.4 Highlights of Contributions

In this dissertation, we extend the concept of pretest and shrinkage estimation in three different models when UPI is available. We derive the asymptotic properties of these estimators and compare their performances with absolute penalty estimators. We also conduct extensive simulation studies for all three models and demonstrate the application of the proposed estimators in real life problems.

The highlights of our contributions in this dissertation are summarized as follows:

In Chapter 2, we apply the shrinkage and absolute penalty methodologies in multiple regression model with RCAR(1) errors. We divide Chapter 2 into two parts. In the first part we consider the parameter estimation in high dimensional case, i.e., when the sample size is less than the number of parameters ($n < p$). Here, we only provide absolute penalty estimators since shrinkage estimators do not exist when $n < p$. In particular, we present a modified version of lasso and AL in order to apply these techniques to the correlated observations. The simulation results show the superiority of AL over lasso.

In the second part of Chapter 2, we consider the fixed dimensional case, i.e., when $n > p$. In this case we obtain shrinkage generalized least square estimators and we derive the asymptotic bias and risk of the estimators. We also compare the performance of these estimators with APEs through simulation. The results show that when UPI is correct the restricted estimator is the best. However, for misspecified UPI, positive shrinkage has superior performance over the restricted, unrestricted, and shrinkage estimator. In comparing shrinkage and APE, positive shrinkage estimator is superior to APEs when there are large number of nuisance variables in the model with respect to significant variables, i.e., when the dimension in (1.3) is large. Finally, a real data analysis is presented to illustrate the results.

In Chapter 3, we consider the shrinkage and absolute penalty estimator in partially linear model with RCAR(1) errors. We investigate the asymptotic properties of shrinkage estimators and we show that these estimators dominate the unrestricted generalized least square estimator. The relative performance of the estimators is examined using asymptotic risk and bias. We also consider an absolute penalty estimator for partially linear model with correlated error. We conduct a Monte Carlo simulation study and the results show that the shrinkage method outperforms the absolute penalty estimator when the dimension in (1.3) is large.

In Chapter 4, we study the application of shrinkage and pretest estimation methods to the quasi-likelihood models. Asymptotic properties of the restricted, shrinkage, positive shrinkage, and pretest estimators are discussed and compared with the unrestricted quasi-maximum likelihood estimator. It is demonstrated that the positive shrinkage estimator is superior to the ordinary shrinkage estimator. Simulation results reveal that the shrinkage estimator outperforms the unrestricted quasi-maximum likelihood estimator in the entire parameter space and the pretest estimator dominates the unrestricted quasi-maximum likelihood estimator on a small part of the parameter space. We also apply the lasso estimator and compare its performance with the other proposed estimators.

In Chapter 5, we summarize the results of this dissertation and present an outline for future research.

Chapter 2

Estimation Strategies in Regression Models with Random Coefficient Autoregressive Errors

2.1 Introduction

A classic problem in statistical analysis is finding a reasonable relationship between a response variable and a set of regressor variables under certain assumptions on the random errors. The usual assumption is that the errors are independent, identically distributed (i.i.d) random variables. This has been later extended to many different cases when the errors are correlated. However, in many fields of research including economics, finance and biology, it is well known that not all correlated errors can be fitted well by linear time series errors. Therefore, much attention is now transferred to nonlinear time series models. Random coefficient time series models are one of the tools to handle the possible nonlinear features of real-life data. For complete background on this model, we refer the reader to Nicholls and Quinn (1982). Liu and Tiao (1980) applied the random coefficient first-order autoregressive model to panel data and they fitted this model to annual average hourly earnings in goods manufacturing in California. Also Singpurwalla and Soyer (1985) implemented this model in real life data on software failures. They used this model for describing and assessing the software reliability growth or decay.

In this chapter, we consider an improved estimation for the parameters in a multiple regression model with random coefficient autoregressive errors (RCAR). We consider methodologies for model selection and parameter estimation using shrinkage, lasso, and adaptive lasso strategies. Consider the following model:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where y_i's are responses, $\boldsymbol{x}_i$ are known $p \times 1$ vectors of covariates, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is an unknown $p \times 1$ vector of regression parameters, and ε_i are unobservable random errors. Here, we assume that ε_i is a first order random coefficient autoregressive (RCAR(1)) process, which is a stationary solution of

$$\varepsilon_i = (\theta + z_i)\varepsilon_{i-1} + e_i, \qquad i = 1, \ldots, n, \tag{2.2}$$

where θ is the autoregression parameter and $\{z_i\}$ and $\{e_i\}$ are zero mean independent processes each consisting of i.i.d random variables with finite second moments σ_z^2 and σ_e^2, respectively.

In this chapter, we propose shrinkage and absolute penalty estimation strategies for fixed dimensional $(p \leq n)$ and high dimensional $(p > n)$ data problems. In the case of fixed dimensionality, we consider shrinkage estimation strategy and propose shrinkage and positive shrinkage estimators for $\boldsymbol{\beta}_1$ when UPI is given in the form of $\boldsymbol{\beta}_2 = \mathbf{0}$. We study the properties of these estimators using the notion of asymptotic distributional bias and risk. For the case of high dimension data, we apply two variable selection methods such as lasso and adaptive lasso (AL). We also provide Monte Carlo simulation studies in both cases. The simulation experiments are conducted for each estimator in a different scenario and the performance of each estimator is evaluated in terms of simulated mean squared error. We also compare the relative performance of both lasso and AL estimation with the SE and PSE. A real data example is given to illustrate the methods.

2.1.1 Organization of the Chapter

The rest of this chapter is organized as the following. In Section 2.2, shrinkage and absolute penalty estimators are presented. Section 2.3 and 2.4 provide asymptotic results of shrinkage and positive shrinkage estimators. In Section 2.5, the performance of the proposed estimators are evaluated through simulation studies. Section 2.6 provides a real data example. Finally, in Section 2.7, we present our concluding thoughts.

2.2 Statistical Model and Estimation

Assume model (2.1) in the general form of $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{X}$ is the $n \times p$ matrix of covariates and $\boldsymbol{y}$ is the $n \times 1$ vector of responses. If θ is known, then the generalized least squares (GLS) estimator of $\boldsymbol{\beta}$ is

$$\check{\boldsymbol{\beta}} = (\boldsymbol{X}'\Omega^{-1}(\theta)\boldsymbol{X})^{-1}\boldsymbol{X}'\Omega^{-1}(\theta)\boldsymbol{y},$$

where $\Omega^{-1}(\theta)$ is a $n \times n$ matrix defined as

$$\Omega^{-1}(\theta) = \begin{pmatrix} 1 & -\theta & 0 & 0 & \dots & 0 \\ -\theta & 1+\theta^2 & -\theta & & & \vdots \\ 0 & -\theta & 1+\theta^2 & -\theta & & \\ \vdots & & & & & \\ 0 & & \dots & & -\theta & 1 \end{pmatrix}.$$

When θ is unknown, as is often the case in practice, $\Omega(\theta)$ is replaced by $\Omega(\hat{\theta})$ where $\hat{\theta}$ is a suitable estimator of θ. Noting that ε_i is unobservable, a reasonable estimator of θ is the least square estimator $\hat{\theta}$ based on the residuals $\hat{\varepsilon}_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n,\ \ i = 1, \dots, n$, and is given by $\hat{\theta} = \sum_{i=2}^{n} \hat{\varepsilon}_i \hat{\varepsilon}_{i-1} / \sum_{i=2}^{n} \hat{\varepsilon}_{i-1}^2$. Noting that $\hat{\boldsymbol{\beta}}_n = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ is the ordinary least square (OLS) estimate of $\boldsymbol{\beta}$. consequently, the estimated GLS $\hat{\boldsymbol{\beta}}$ can be written as

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\Omega^{-1}(\hat{\theta})\boldsymbol{X})^{-1}\boldsymbol{X}'\Omega^{-1}(\hat{\theta})\boldsymbol{y}.$$

The properties of $\hat{\boldsymbol{\beta}}$ were investigated in Hwang and Basawa (1993).

2.2.1 Fixed Dimensional Estimation ($n \geq p$)

The model (2.1) is generally regarded as a full model, which is built at the initial stage of modeling and contains all the possibly relevant variables.

Suppose that $\boldsymbol{\beta}$ can be partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, where sub-vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have dimensions p_1 and p_2 respectively, and $p_1 + p_2 = p$, $p_i \geq 0$ for $i = 1, 2$. Thus we can rewrite the full candidate model as follows:

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon},$$

where $\boldsymbol{X}_1$ is the $n \times p_1$ matrix of the first p_1 covariates and $\boldsymbol{X}_2$ is the $n \times p_2$ matrix of the second p_2 covariates. We are mainly interested in the estimation of $\boldsymbol{\beta}_1$ when based on some variable selection method or prior information from previous studies indicate it is plausible that $\boldsymbol{\beta}_2$ is close to some specified $\boldsymbol{\beta}_2^0$ which, without loss of generality, we may set to $\mathbf{0}$. Thus, by removing these insignificant variables, we have a candidate sub-model as

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}. \tag{2.3}$$

Our goal is to construct an efficient estimation for the regression parameter $\boldsymbol{\beta}_1$ when $\boldsymbol{\beta}_2$ may be equal to $\mathbf{0}$. For example, in the case of a multi-factor design, we maybe interested in the estimating of the main effects $\boldsymbol{\beta}_1$, while there is a question whether the vector of interaction effects $\boldsymbol{\beta}_2$ may be ignored. Now suppose $\boldsymbol{\beta}_1$ is the $p_1 \times 1$ coefficient vector for main effects and $\boldsymbol{\beta}_2$ is the $p_2 \times 1$ coefficient vector for nuisance effects and there is evidence that nuisance variables do not provide useful information, that is, $\boldsymbol{\beta}_2 = 0$.

Unrestricted and Restricted Estimators

According to the inverse matrix formula we have

$$A^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11.2}^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}A_{11.2}^{-1}A_{12}A_{22}^{-1} \end{pmatrix}, \tag{2.4}$$

where $A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$. Also, by partitioning $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2')'$ and $\boldsymbol{X} = [\boldsymbol{X}_1|\boldsymbol{X}_2]$ we have

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \left([\boldsymbol{X}_1|\boldsymbol{X}_2]'\boldsymbol{\Omega}^{-1}(\hat{\theta})[\boldsymbol{X}_1|\boldsymbol{X}_2]\right)^{-1}[\boldsymbol{X}_1|\boldsymbol{X}_2]'\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{y}$$

$$= \begin{pmatrix} \boldsymbol{X}_1'\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_2 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{X}_1'\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{y} \\ \boldsymbol{X}_2'\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{y} \end{pmatrix}$$

By using the inverse matrix formula (2.4), the unrestricted GLS estimator (UE) $\hat{\boldsymbol{\beta}}_1$ of $\boldsymbol{\beta}_1$ will be in the form of

$$\hat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_2}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_2}\boldsymbol{y},$$

where $\boldsymbol{X}_1$ is composed of the first p_1 column vectors of $\boldsymbol{X}$, $\boldsymbol{X}_2$ is composed of the last p_2 column vectors of $\boldsymbol{X}$ and

$$M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_2} = \boldsymbol{\Omega}^{-1}(\hat{\theta}) - \boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{\Omega}^{-1}(\hat{\theta}).$$

The restricted GLS estimator (RE) $\tilde{\boldsymbol{\beta}}_1$ of $\boldsymbol{\beta}_1$ for model (2.3) when $\boldsymbol{\beta}_2 = \boldsymbol{0}$ has the form

$$\tilde{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{\Omega}^{-1}(\theta^*)\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{\Omega}^{-1}(\theta^*)\boldsymbol{y},$$

where $\theta^* = \sum_{i=2}^{n}\varepsilon_i^*\varepsilon_{i-1}^*/\sum_{i=2}^{n}\varepsilon_{i-1}^{*2}$ with $\varepsilon_i^* = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_1^*$, $\boldsymbol{\beta}_1^* = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1\boldsymbol{y}$, $\boldsymbol{X}_1 = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p_1})$, $\boldsymbol{x}_j = (x_{j1}, \ldots, x_{jn})'$, $\quad j = 1, \cdots, p_1$.

Generally speaking, $\tilde{\boldsymbol{\beta}}_1$ performs better than $\hat{\boldsymbol{\beta}}_1$ when $\boldsymbol{\beta}_2$ is null vector (or very close to null vector). But for $\boldsymbol{\beta}_2$ away from the null vector, $\tilde{\boldsymbol{\beta}}_1$ may be considerably biased, inefficient, and even possibly inconsistent while the $\hat{\boldsymbol{\beta}}_1$ holds its performance characteristics steadily over the departure of $\boldsymbol{\beta}_2$ from null vector. Thus, in the face of uncertain prior information $\boldsymbol{\beta}_2 = \mathbf{0}$, we have two extreme estimators $\hat{\boldsymbol{\beta}}_1$ and $\tilde{\boldsymbol{\beta}}_1$ suited best for the full model and sub-model, respectively. One natural attempt is to strike a smooth compromise between $\hat{\boldsymbol{\beta}}_1$ and $\tilde{\boldsymbol{\beta}}_1$ so that the performance characteristic of the compromise estimator behaves reasonably well relative to $\hat{\boldsymbol{\beta}}_1$ as well as $\tilde{\boldsymbol{\beta}}_1$.

Shrinkage and Positive Shrinkage Estimators

The shrinkage estimator (SE) $\hat{\boldsymbol{\beta}}_1^S$ of $\boldsymbol{\beta}_1$ is defined as follows:

$$\hat{\boldsymbol{\beta}}_1^S = \tilde{\boldsymbol{\beta}}_1 + \{1 - c_{opt} T_n^{-1}\}(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1), \quad \text{where} \quad c_{opt} = p_2 - 2, \quad p_2 \geq 3$$

and

$$T_n = n\hat{\sigma}_n^{-2} \hat{\boldsymbol{\beta}}_2' \boldsymbol{X}_2' M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_1} \boldsymbol{X}_2 \hat{\boldsymbol{\beta}}_2,$$

where $\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n (y_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}})^2$ and $M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_1}$ has the same definition as $M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\boldsymbol{X}_2}$. It is clear that $\hat{\boldsymbol{\beta}}_1^S$ is a smooth compromise between $\hat{\boldsymbol{\beta}}_1$ and $\tilde{\boldsymbol{\beta}}_1$. It tends to $\hat{\boldsymbol{\beta}}_1$ as T_n tends to infinity and tends to $\tilde{\boldsymbol{\beta}}_1$ as $T_n \to p_2 - 2$. The problem with SE is that its components may have a different sign from the coordinates of $\hat{\boldsymbol{\beta}}_1$. This could happen if $c_{opt} T_n^{-1} > 1$. In this case the change in sign would affect its interpretability. In an attempt to overcome this difficulty, we define the positive shrinkage estimator (PSE) by using the positive part of the SE which will control the possible over-shrinking in SE. The PSE has the form

$$\hat{\boldsymbol{\beta}}_1^{S+} = \tilde{\boldsymbol{\beta}}_1 + \{1 - c_{opt} T_n^{-1}\}^+ (\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1), \;\; p_2 \geq 3,$$

where $z^+ = max(0,z)$. For the sake of computation, the PSE can be written in the following form

$$\hat{\boldsymbol{\beta}}_1^{S+} = \hat{\boldsymbol{\beta}}_1^{S} - [1-(p_2-2)T_n^{-1}]I(T_n < p_2-2)(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1), \;\; p_2 \geq 3,$$

where $I(.)$ is the indicator function.

2.2.2 High Dimensional Estimation ($n < p$)

In this section, we present two absolute penalty estimators, namely, lasso and adaptive lasso (AL) for the model (2.1) with the errors in (2.2).

Lasso Estimator

The least absolute shrinkage and selection operator (lasso) proposed by Tibshirani (1996) is a regularization technique for simultaneous estimation and variable selection. The lasso estimates are defined as

$$\hat{\boldsymbol{\beta}}_{lasso} = \underset{\boldsymbol{\beta}}{argmin}\{\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \lambda\sum_{j=1}^{p}|\beta_j|\}, \tag{2.5}$$

where λ is a nonnegative regularization parameter and the second term is the so-called L_1 penalty. This method has become a popular model selection procedure since it shrinks some coefficients and, because of its L_1 penalty, the method will set many of the coefficients exactly equal to 0. When λ is large enough, the solutions are shrunk versions of the least square estimates often with some of them equal to zero; however, for smaller values of $\lambda(\lambda \geq 0)$, the constraint may have no effect. A cross-validation method is mainly used to find the best value for λ.

Many well developed procedures can be used to find the solution for the above penalized functions. For example: quadratic programming (Tibshirani, 1996), the shooting algorithm (Fu, 1998), local quadratic approximation (Fan and Li, 2001), and least angle regression (LAR) algorithm (Efron et al., 2004). The latter exploits the special structure of the lasso problem, and provides an efficient way to compute the solutions simultaneously for all values of λ.

In recent years, there has been a vast amount of research devoted to regularization methods. Rosset and Zhu (2007) studied the piecewise linear regularized solution paths to differentiable and piecewise quadratic loss functions with L_1 penalty. Friedman et al. (2007), Wu and Lange (2008), and Friedman et al. (2010) developed the coordinate descent (CD) algorithm for penalized linear regression and generalized linear models. For a review, the reader is referred to Hesterberg et al. (2008), Zhang and Chai (2010), Ahmed and Raheem (2012) and references therein for an up-to-date comprehensive review on this topic.

Since in our model the errors are correlated, we cannot achieve the APEs by directly applying the L_1 penalty to the data. As suggested by Hastie, we transform the data first and then apply the L_1 penalty to the transformed data.

Based on Hwang and Basawa (1993) we have $Var(\boldsymbol{\varepsilon}) = \boldsymbol{\Upsilon} = \frac{(1-\theta^2)\sigma_\varepsilon^2}{1-\theta^2-\sigma_z^2}\boldsymbol{\Omega}(\theta)$. Using Cholesky decomposition, we factor $\boldsymbol{\Upsilon} = \boldsymbol{A}\boldsymbol{A}'$ where $\boldsymbol{\Upsilon}$ and $\boldsymbol{A}$ are $n \times n$ matrices. Now if we multiply both sides of the equation $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ by $\boldsymbol{A}^{-1}$, we then get the linear model $\boldsymbol{y}^\star = \boldsymbol{X}^\star\boldsymbol{\beta} + \boldsymbol{\varepsilon}^\star$, where $\boldsymbol{y}^\star = \boldsymbol{A}^{-1}\boldsymbol{y}$, $\boldsymbol{X}^\star = \boldsymbol{A}^{-1}\boldsymbol{X}$ and $\boldsymbol{\varepsilon}^\star = \boldsymbol{A}^{-1}\boldsymbol{\varepsilon}$. In this model, $Var(\boldsymbol{\varepsilon}^\star) = \boldsymbol{A}^{-1}\boldsymbol{\Upsilon}\boldsymbol{A}^{-1} = I$; therefore, the lasso method can be applied to these transformed data.

Since $\boldsymbol{\Upsilon}$ is unknown in most cases, we first estimate it by $\hat{\boldsymbol{\Upsilon}} = \hat{\sigma}_n^2\Omega(\hat{\theta})$ where $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_i^2 = n^{-1}\sum_{i=1}^n (y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n)^2$ and $\hat{\boldsymbol{\beta}}_n$ is the OLS estimate of $\boldsymbol{\beta}$. But in the high dimension case since $n < p$, the OLS estimate can not be achieved. Therefore the initial estimate for $\boldsymbol{\beta}$ is computed from a lasso fit on the original data. Once $\boldsymbol{\Upsilon}$ is estimated, we then factor $\hat{\boldsymbol{\Upsilon}} = \boldsymbol{C}\boldsymbol{C}'$ where $\boldsymbol{C}$ is also an $n \times n$ matrix and lasso coefficients are the solutions to the L_1 optimization problem:

$$\hat{\boldsymbol{\beta}}_{lasso} = \underset{\boldsymbol{\beta}}{argmin}\{(\boldsymbol{y}^\star - \boldsymbol{X}^\star\boldsymbol{\beta})'(\boldsymbol{y}^\star - \boldsymbol{X}^\star\boldsymbol{\beta}) + \lambda\sum_{j=1}^{p}|\beta_j|\},$$

where $\boldsymbol{y}^\star = \boldsymbol{C}^{-1}\boldsymbol{y}$, $\boldsymbol{X}^\star = \boldsymbol{C}^{-1}\boldsymbol{X}$.

Adaptive Lasso Estimator

The asymptotic setup in lasso is somewhat unfair, because it forces the coefficients to be equally penalized in the L_1 penalty. Also there are certain situations where the lasso is inconsistent for variable selection. Zou (2006) proposed a weighted lasso method called adaptive lasso (AL) where different weights are assigned to different coefficients. He showed that AL enjoys oracle properties; namely, it performs as well as if the true underlying model were given in advance. The AL method uses adaptive weights for penalizing different coefficients in the L_1 penalty and is defined as

$$\hat{\boldsymbol{\beta}}_{AL} = \underset{\boldsymbol{\beta}}{argmin}\{\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \lambda\sum_{j=1}^{p}w_j|\beta_j|\},$$

where λ is the tuning parameter, w_j is a known weight which is better to be data-dependent chosen, $\hat{w}_j = 1/|\hat{\beta}_j|^\gamma$ for $\gamma > 0$, and $\hat{\boldsymbol{\beta}}$ is a root n-consistent estimator to $\boldsymbol{\beta}$; for instance, ordinary least square estimator (OLS). But since in high-dimensional problems the OLS estimates do not exist, the lasso estimates can be used instead to compute the weights. Similar steps described for applying the lasso technique to correlated data can be followed to achieve AL estimators. The steps are given below:

Step 1: Fit lasso to the original data to get the estimator $\boldsymbol{\beta}^*$

Step 2: Calculate $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_i^2 = n^{-1}\sum_{i=1}^n (y_i - \boldsymbol{x}_i'\boldsymbol{\beta}^*)^2$, $\hat{\theta} = \sum_{i=2}^n \hat{\varepsilon}_i\hat{\varepsilon}_{i-1} / \sum_{i=2}^n \hat{\varepsilon}_{i-1}^2$

Step 3: Calculate $\hat{\Upsilon} = \hat{\sigma}_n^2\Omega(\hat{\theta})$ and then factor $\hat{\Upsilon} = \boldsymbol{C}\boldsymbol{C}'$ using Cholesky decomposition

Step 4: Transform $\boldsymbol{y}$ and $\boldsymbol{X}$ to $\boldsymbol{y}^\star = \boldsymbol{C}^{-1}\boldsymbol{y}$, $\boldsymbol{X}^\star = \boldsymbol{C}^{-1}\boldsymbol{X}$

Step 5: Solve the AL problem as

$$\hat{\boldsymbol{\beta}}_{AL} \quad = \quad \underset{\boldsymbol{\beta}}{argmin}\big\{(\boldsymbol{y}^\star - \boldsymbol{X}^\star\boldsymbol{\beta})'(\boldsymbol{y}^\star - \boldsymbol{X}^\star\boldsymbol{\beta}) + \lambda\sum_{j=1}^p w_j|\beta_j|\big\}.$$

In the next section we present the assumptions needed to provide asymptotic results for shrinkage estimators and compare their performances based on their asymptotic distributional bias (ADB) and asymptotic distributional risk (ADR).

2.3 First-Order Asymptotics

The following assumptions are needed to derive the main results in Section 2.4.

Assumption 2.3.1. For all $t,k = 1,\cdots,p$ and $h = 0,\pm1,\pm2,\ldots$, there exist $g_{ij}(h)$ such that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n-h} x_{it}x_{i+|h|,k}}{n} = g_{tk}(h).$$

Assumption 2.3.2.

$$E\varepsilon_1^4 < \infty, \text{ and } |\theta| < 1, \ \theta^2 + \sigma_z^2 < 1.$$

Theorem 2.3.1. Suppose that Assumptions 2.3.1 and 2.3.2 hold. Then we have that

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N\left(0, \frac{(1-\theta^2)\sigma_e^2}{1-\theta^2-\sigma_z^2}\boldsymbol{B}^{-1}\right), \text{ where } \lim_{n\to\infty}\left(\frac{1}{n}\boldsymbol{X}'\Omega^{-1}(\theta)\boldsymbol{X}\right) = \boldsymbol{B},$$

and $\xrightarrow{D}$ denotes convergence in distribution.

Proof. The proof can be found in Hwang and Basawa (1993). □

Lemma 2.3.1. Suppose that Assumptions 2.3.1 and 2.3.2 hold. Then we have that

$$\lim_{n\to\infty}\frac{1}{n}\boldsymbol{X}'\Omega^{-1}(\hat{\theta})\boldsymbol{X}=\boldsymbol{B}.$$

Proof. According to Theorem 4.4 in Hwang and Basawa (1993), we need to show that

$$\frac{1}{n}(\boldsymbol{X}'\Omega^{-1}(\hat{\theta})\boldsymbol{X}-\boldsymbol{X}'\Omega^{-1}(\theta)\boldsymbol{X})=o_p(1), \tag{2.6}$$

and

$$\frac{1}{\sqrt{n}}(\boldsymbol{X}'\Omega^{-1}(\hat{\theta})\varepsilon-\boldsymbol{X}'\Omega^{-1}(\theta)\varepsilon)=o_p(1). \tag{2.7}$$

Consider the (i,j) element of the term on the left-hand side of equation (2.6) given by

$$(\hat{\theta}^2-\theta^2)\frac{\sum_{i=2}^{n}x_{it}x_{ik}}{n}-2(\hat{\theta}-\theta)\frac{\sum_{i=1}^{n-1}x_{it}x_{i+1,k}}{n},$$

which implies (2.6) since $\sqrt{n}(\hat{\theta}-\theta)$ is bounded in probability by Theorem 4.3 in Hwang and Basawa (1993) and $n^{-1}\sum_{i=2}^{n}x_{it}x_{ik}$ and $n^{-1}\sum_{i=1}^{n-1}x_{it}x_{i+1,k}$ are bounded by Assumption 2.3.1. The proof of (2.7) follows on a similar way. □

Lemma 2.3.2. Suppose that Assumptions 2.3.1 and 2.3.2 hold. Then we have

$$\hat{\sigma}_n^2=(1-\theta^2-\sigma_z^2)^{-1}\{(1-\theta^2)\sigma_e^2\}+o_p\left(n^{-\frac{1}{2}}\right),\ \ \tilde{\boldsymbol{\beta}}_1=(\boldsymbol{I},\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12})\hat{\boldsymbol{\beta}}_G+o_p\left(n^{-\frac{1}{2}}\right),$$

and

$$T_n = n(1-\theta^2-\sigma_z^2)^{-1}\{(1-\theta^2)\sigma_e^2\}\hat{\boldsymbol{\beta}}_{2G}'\boldsymbol{B}_{22.1}\hat{\boldsymbol{\beta}}_{2G} + o_p(1),$$

where $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_i^2 = n^{-1}\sum_{i=1}^n (y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}})^2$.

Proof. According to Nicholls and Quinn (1982) we need to show

$$\frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^2 - \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 = o_p(n^{-\frac{1}{2}}).$$

It is easy to show that

$$\frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^2 - \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 = \frac{1}{n}\sum_{i=1}^n (\hat{\varepsilon}_i - \varepsilon_i)^2 + \frac{2}{n}\sum_{i=1}^n (\hat{\varepsilon}_i - \varepsilon_i)\varepsilon_i.$$

Since

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^2 \quad \text{and} \quad \hat{\varepsilon}_i = \varepsilon_i - \boldsymbol{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

it holds that

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 + \frac{1}{n}\sum_{i=1}^n (\boldsymbol{x}_i'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}))^2 + \frac{2}{n}\sum_{i=1}^n \varepsilon_i \boldsymbol{x}_i'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) = I_1 + I_2 + I_3,$$

where $I_1 = \sigma_\varepsilon^2 = Var(\varepsilon_1) = (1-\theta^2-\sigma_z^2)^{-1}\{(1-\theta^2)\sigma_e^2\}$. Also based on Assumption 2.3.1 and the results of Lemma 2.3.1, it can be shown that $I_i = o_p(n^{-\frac{1}{2}}) \quad i = 2,3$. The other results of Lemma 2.3.2 can be proved by combining Theorem 4.4 in Hwang and Basawa (1993) and Lemma 2.3.1. □

Lemma 2.3.3. In an effort to establish some important properties of proposed estimators, let $\boldsymbol{\eta}_1 = \sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)$, $\boldsymbol{\eta}_2 = \sqrt{n}(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)$, and $\boldsymbol{\eta}_3 = \sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1)$. Under the local alternative $\{K_n\}$ and the results of Lemma 2.3.1 and Lemma 2.3.2, the asymptotic joint distributions of listed estimators are given below:

$$\text{(i)} \quad \begin{pmatrix} \boldsymbol{\eta}_2 \\ \boldsymbol{\eta}_3 \end{pmatrix} \sim N_{2p_1}\left\{ \begin{pmatrix} -\boldsymbol{\gamma} \\ \boldsymbol{\gamma} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} - \boldsymbol{\Sigma}^* & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^* \end{pmatrix} \right\}$$

$$\text{(ii)} \quad \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_3 \end{pmatrix} \sim N_{2p_1}\left\{ \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\gamma} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}^* \\ \boldsymbol{\Sigma}^* & \boldsymbol{\Sigma}^* \end{pmatrix} \right\},$$

where $\boldsymbol{\Sigma} = \frac{(1-\theta^2)\sigma_e^2}{1-\theta^2-\sigma_z^2}\boldsymbol{B}_{11.2}^{-1}$, $\boldsymbol{\Sigma}^* = \frac{(1-\theta^2)\sigma_e^2}{1-\theta^2-\sigma_z^2}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22.1}^{-1}\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}$ and $\boldsymbol{\gamma} = \boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}$.

Lemma 2.3.4. Let $\boldsymbol{x}$ be a p-dimensional normal vector distributed as $N_p(\boldsymbol{\mu}_{\boldsymbol{x}}, \Sigma_p)$. Then, for a measurable function of ϕ, we have

$$\begin{aligned} E[\boldsymbol{x}\phi(\boldsymbol{x}'\boldsymbol{x})] &= \boldsymbol{\mu}_{\boldsymbol{x}} E[\phi(\chi^2_{p+2}(\Delta))], \\ E[\boldsymbol{x}\boldsymbol{x}'\phi(\boldsymbol{x}'\boldsymbol{x})] &= \Sigma_p E[\phi(\chi^2_{p+2}(\Delta))] + \boldsymbol{\mu}_{\boldsymbol{x}}\boldsymbol{\mu}'_{\boldsymbol{x}} E[\phi(\chi^2_{p+4}(\Delta))], \end{aligned}$$

where $\chi^2_\nu(\Delta)$ is a non-central chi-square distribution with ν degrees of freedom and non-centrality parameter $\Delta = \boldsymbol{\mu}'_{\boldsymbol{x}}\Sigma_p^{-1}\boldsymbol{\mu}_{\boldsymbol{x}}$.

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

Lemma 2.3.5. Let $\boldsymbol{A} = (\boldsymbol{A}'_1 \vdots \boldsymbol{A}'_2)'$ be a vector distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_p)$ with

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

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

$$\boldsymbol{\mu}_{11.2} = \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{a}_2 - \boldsymbol{\mu}_2)$$

and

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Proof. The proof can be found in Johnson and Wichern (2001). □

2.4 Asymptotic Properties of the Shrinkage Estimators

In this section, we investigate the performance of the UE, RE, SE and PSE of $\boldsymbol{\beta}_1$ using the notion of asymptotic distributional bias (ADB) and asymptotic distributional risk (ADR) under $\{K_n\}$. Consider the sequence of $\{K_n\}$ given by

$$K_n : \boldsymbol{\beta}_{2(n)} = n^{-\frac{1}{2}}\boldsymbol{\omega}, \quad \boldsymbol{\omega} \neq 0 \text{ fixed},$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_{p_2}) \in \Re^{p_2}$ is a real fixed vector.

Suppose that the asymptotic cumulative distribution function (cdf) of $\sqrt{n}(\boldsymbol{\beta}_1^0 - \boldsymbol{\beta}_1)$ under $\{K_n\}$ exists, and is defined as

$$F(\boldsymbol{x}) = P[\lim_{n\to\infty} \sqrt{n}(\boldsymbol{\beta}_1^0 - \boldsymbol{\beta}_1) \leq \boldsymbol{x}|K_n].$$

Further let

$$\boldsymbol{\Gamma} = \int \ldots \int \boldsymbol{x}\boldsymbol{x}' dF(\boldsymbol{x})$$

be the dispersion matrix obtained from cdf. Then the ADR is defined as

$$R[(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_1); \boldsymbol{M}] = \text{tr}(\boldsymbol{M}\boldsymbol{\Gamma}).$$

Also, the ADB of the estimator $\boldsymbol{\beta}_1^0$ under K_n is defined as

$$ADB(\boldsymbol{\beta}_1^0) = E\{\lim_{n\to\infty} \sqrt{n}(\boldsymbol{\beta}_1^0 - \boldsymbol{\beta}_1)\}.$$

In the following we present the ADB and ADR of the estimators. The results of the Lemma 2.3.4 and Lemma 2.3.5 will be extensively used to derive the asymptotic results of the estimators under K_n.

2.4.1 Asymptotic Distributional Bias (ADB)

Theorem 2.4.1. Suppose that Assumptions 2.3.1 and 2.3.2 hold. Then under $\{K_n\}$, as $n \to \infty$, the ADB of the estimators $\hat{\boldsymbol{\beta}}_1$, $\tilde{\boldsymbol{\beta}}_1$, $\hat{\boldsymbol{\beta}}_1^S$, and $\hat{\boldsymbol{\beta}}_1^{S+}$ are, respectively,

$$\begin{aligned}
\text{ADB}(\hat{\boldsymbol{\beta}}_1) &= \mathbf{0}, \\
\text{ADB}(\tilde{\boldsymbol{\beta}}_1) &= -\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}, \\
\text{ADB}(\hat{\boldsymbol{\beta}}_1^S) &= -(p_2-2)\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}E(\chi_{p_2+2,\alpha}^{-2};\Delta), \\
\text{ADB}(\hat{\boldsymbol{\beta}}_1^{S+}) &= \text{ADB}(\hat{\boldsymbol{\beta}}_1^S) - \boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}\big\{H_{p_2+2}(p_2-2;\Delta) \\
&- (p_2-2)E(\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) < p_2-2))\big\},
\end{aligned}$$

where $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix}$ is defined in Theorem 2.3.1, $\Delta = (\boldsymbol{\omega}'\boldsymbol{B}_{22.1}\boldsymbol{\omega})(1-\theta^2-\sigma_z^2)\{(1-\theta^2)\sigma_e^2\}^{-1}$, $\boldsymbol{B}_{22.1} = \boldsymbol{B}_{22} - \boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}$ and $H_v(x;\Delta)$ denotes the noncentral chi-square distribution function with non-centrality parameter Δ and v degrees of freedom and

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

Proof. Here, we provide the proof of bias expressions. Lemma 2.3.2 and Lemma 2.3.3 are used to prove the results.

$$
\begin{aligned}
ADB(\tilde{\boldsymbol{\beta}}_1) &= E\lim_{n\to\infty}\sqrt{n}(\tilde{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1)\\
&= E\lim_{n\to\infty}\sqrt{n}(\hat{\boldsymbol{\beta}}_1-\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\hat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_1)\\
&= E\lim_{n\to\infty}\sqrt{n}(\hat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1)-E\lim_{n\to\infty}\sqrt{n}(\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\hat{\boldsymbol{\beta}}_2)\\
&= -\sqrt{n}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\beta}_2=-\sqrt{n}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}n^{-\frac{1}{2}}\boldsymbol{\omega}=-\boldsymbol{\gamma},\\
ADB(\hat{\boldsymbol{\beta}}_{1G}^S) &= E\lim_{n\to\infty}\sqrt{n}(\hat{\boldsymbol{\beta}}_{1G}^S-\boldsymbol{\beta}_1)\\
&= E\lim_{n\to\infty}\sqrt{n}(\tilde{\boldsymbol{\beta}}_1+(1-(p_2-2)T_n^{-1})(\hat{\boldsymbol{\beta}}_1-\tilde{\boldsymbol{\beta}}_1)-\boldsymbol{\beta}_1)\\
&= E\lim_{n\to\infty}\sqrt{n}[\tilde{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1+(\hat{\boldsymbol{\beta}}_1-\tilde{\boldsymbol{\beta}}_1)-(p_2-2)T_n^{-1}(\hat{\boldsymbol{\beta}}_1-\tilde{\boldsymbol{\beta}}_1)]\\
&= E\lim_{n\to\infty}[\boldsymbol{\eta}_2+\boldsymbol{\eta}_3-(p_2-2)T_n^{-1}\boldsymbol{\eta}_3]\\
&= E\lim_{n\to\infty}[\boldsymbol{\eta}_1-(p_2-2)T_n^{-1}\boldsymbol{\eta}_3]=-(p_2-2)E\lim_{n\to\infty}[T_n^{-1}\boldsymbol{\eta}_3]\\
&= -(p_2-2)\boldsymbol{\gamma}E(\chi_{p_2+2}^{-2}(\Delta)),\\
ADB(\hat{\boldsymbol{\beta}}_1^{S+}) &= E\lim_{n\to\infty}\sqrt{n}(\hat{\boldsymbol{\beta}}_1^{S+}-\boldsymbol{\beta}_1)\\
&= E\lim_{n\to\infty}\sqrt{n}[\hat{\boldsymbol{\beta}}_1^S-\boldsymbol{\beta}_1-(1-(p_2-2)T_n^{-1})(\hat{\boldsymbol{\beta}}_1-\tilde{\boldsymbol{\beta}}_1)I(T_n<p_2-2)]\\
&= ADB(\hat{\boldsymbol{\beta}}_1^S)-E\lim_{n\to\infty}\sqrt{n}[(1-(p_2-2)T_n^{-1})(\hat{\boldsymbol{\beta}}_1-\tilde{\boldsymbol{\beta}}_1)I(T_n<p_2-2)]\\
&= ADB(\hat{\boldsymbol{\beta}}_1^S)-E\lim_{n\to\infty}[\boldsymbol{\eta}_3(1-(p_2-2)T_n^{-1})I(T_n<p_2-2)]\\
&= ADB(\hat{\boldsymbol{\beta}}_1^S)-E\lim_{n\to\infty}[\boldsymbol{\eta}_3I(T_n<p_2-2)]+E\lim_{n\to\infty}[(p_2-2)\boldsymbol{\eta}_3T_n^{-1}I(T_n<p_2-2)]\\
&= ADB(\hat{\boldsymbol{\beta}}_1^S)-\boldsymbol{\gamma}H_{p_2+2}(p_2-2;\Delta)+\boldsymbol{\gamma}(p_2-2)E(\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta)<p_2-2)).
\end{aligned}
$$

□

Since the bias expressions of all the estimators are not in the scalar form, we convert them to quadratic form. Thus, we define the asymptotic quadratic distributional bias (AQDB) of an estimator $\boldsymbol{\beta}_1^0$ of $\boldsymbol{\beta}_1$ by

$$\mathrm{AQDB}(\boldsymbol{\beta}_1^0) = [\mathrm{ADB}(\boldsymbol{\beta}_1^0)]'\boldsymbol{B}_{11.2}[\mathrm{ADB}(\boldsymbol{\beta}_1^0)],$$

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

Corollary 2.4.1. Suppose that conditions in Theorem 2.4.1 hold. Then under $\{K_n\}$, as $n \to \infty$, the AQDB of the estimators are given as follows:

$$\begin{aligned}
\mathrm{AQDB}(\hat{\boldsymbol{\beta}}_1) &= \boldsymbol{0}, \\
\mathrm{AQDB}(\tilde{\boldsymbol{\beta}}_1) &= \boldsymbol{\omega}'\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{11.2}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}, \\
\mathrm{AQDB}(\hat{\boldsymbol{\beta}}_1^S) &= (p_2-2)^2\boldsymbol{\omega}'\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{11.2}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}\left[E(\chi_{p_2,\alpha}^{-2};\Delta)\right]^2, \\
\mathrm{AQDB}(\hat{\boldsymbol{\beta}}_1^{S+}) &= \boldsymbol{\omega}'\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{11.2}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}\Big\{H_{p_2+2}(p_2-2;\Delta) \\
&+ (p_2-2)^2E(\chi_{p_2+2}^{-2}(\Delta)) - (p_2-2)^2E(\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) < p_2-2))\Big\}^2.
\end{aligned}$$

Proof. The expressions for quadratic biases are obtained by following the definition of AQDB. □

2.4.2 Asymptotic Distributional Risk (ADR)

Theorem 2.4.2. Suppose that conditions in Theorem 2.4.1 hold. Then under $\{K_n\}$, as $n \to \infty$, the asymptotic covariance matrices of the estimators are as follows:

$$
\begin{aligned}
\Gamma(\hat{\boldsymbol{\beta}}_1) &= \sigma^* \boldsymbol{B}_{11.2}^{-1} \quad \text{where} \quad \sigma^* = (1-\theta^2-\sigma_z^2)^{-1}\{(1-\theta^2)\sigma_e^2\}, \\
\Gamma(\tilde{\boldsymbol{\beta}}_1) &= \sigma^* \boldsymbol{B}_{11.2}^{-1} - \sigma^*(\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22.1}^{-1}\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}) + \boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}\boldsymbol{\omega}'\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}, \\
\Gamma(\hat{\boldsymbol{\beta}}_1^S) &= \sigma^* \boldsymbol{B}_{11.2}^{-1} \\
&- \sigma^*(p_2-2)\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22.1}^{-1}\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\Big\{2E(\chi_{p_2+2}^{-2}(\Delta)) - (p_2-2)E(\chi_{p_2+2}^{-4}(\Delta))\Big\} \\
&+ (p_2-2)\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}\boldsymbol{\omega}'\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\Big\{2E(\chi_{p_2+2}^{-2}(\Delta) + (p_2-2)E(\chi_{p_2+4}^{-4}(\Delta) \\
&- 2E(\chi_{p_2+4}^{-2}(\Delta))\Big\}, \\
\Gamma(\hat{\boldsymbol{\beta}}_1^{S+}) &= \Gamma(\hat{\boldsymbol{\beta}}_1^S) \\
&+ \sigma^*\Big\{(p_2-2)(\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22.1}^{-1}\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1})\{2E\left[\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\right] \\
&- (p_2-2)E\left[\chi_{p_2+2}^{-4}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\right] \\
&- \boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22.1}^{-1}\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}H_{p_2+2}(p_2-2;\Delta)\Big\} \\
&+ \boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}\boldsymbol{\omega}'\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\Big[2H_{p_2+2}(p_2-2;\Delta) - H_{p_2+4}(p_2-2;\Delta)\Big] \\
&- (p_2-2)\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}\boldsymbol{\omega}'\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\Big\{2E\left[\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\right] \\
&- 2E\left[\chi_{p_2+4}^{-2}(\Delta)I(\chi_{p_2+4}^{2}(\Delta) \le p_2-2)\right] \\
&+ (p_2-2)E\left[\chi_{p_2+4}^{-4}(\Delta)I(\chi_{p_2+4}^{2}(\Delta) \le p_2-2)\right]\Big\}.
\end{aligned}
$$

Proof. The asymptotic covariance matrix of an estimator $\boldsymbol{\beta}_1^0$ is defined as follows:

$$\Gamma(\boldsymbol{\beta}_1^0) = E \lim_{n\to\infty}(n(\boldsymbol{\beta}_1^0 - \boldsymbol{\beta}_1)(\boldsymbol{\beta}_1^0 - \boldsymbol{\beta}_1)').$$

By definition,

$$
\begin{aligned}
\Gamma(\hat{\boldsymbol{\beta}}_1) &= E\lim_{n\to\infty}(n(\hat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1)') \\
&= E\lim_{n\to\infty}(\boldsymbol{\eta}_1\boldsymbol{\eta}_1{}') = Var(\boldsymbol{\eta}_1)+E(\boldsymbol{\eta}_1)E(\boldsymbol{\eta}_1)' \\
&= \boldsymbol{\Sigma}. \\
\Gamma(\tilde{\boldsymbol{\beta}}_1) &= E\lim_{n\to\infty}(n(\tilde{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1)(\tilde{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1)') \\
&= E\lim_{n\to\infty}(\boldsymbol{\eta}_2\boldsymbol{\eta}_2{}') = Var(\boldsymbol{\eta}_2)+E(\boldsymbol{\eta}_2)E(\boldsymbol{\eta}_2)' \\
&= \boldsymbol{\Sigma}-\boldsymbol{\Sigma}^*+\boldsymbol{\gamma}\boldsymbol{\gamma}'. \\
\Gamma(\hat{\boldsymbol{\beta}}_1^S) &= E\lim_{n\to\infty}(n(\hat{\boldsymbol{\beta}}_1^S-\boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_1^S-\boldsymbol{\beta}_1)') \\
&= E\lim_{n\to\infty}\sqrt{n}\Big((\tilde{\boldsymbol{\beta}}_1+(1-(p_2-2)T_n^{-1})(\hat{\boldsymbol{\beta}}_1-\tilde{\boldsymbol{\beta}}_1)-\boldsymbol{\beta}_1\Big) \\
&\times \sqrt{n}\Big(\tilde{\boldsymbol{\beta}}_1+(1-(p_2-2)T_n^{-1})(\hat{\boldsymbol{\beta}}_1-\tilde{\boldsymbol{\beta}}_1)-\boldsymbol{\beta}_1\Big)' \\
&= E\lim_{n\to\infty}[(\boldsymbol{\eta}_1-(p_2-2)T_n^{-1}\boldsymbol{\eta}_3)(\boldsymbol{\eta}_1-(p_2-2)T_n^{-1}\boldsymbol{\eta}_3)'] \\
&= E\lim_{n\to\infty}[\boldsymbol{\eta}_1\boldsymbol{\eta}_1{}'-(p_2-2)T_n^{-1}\boldsymbol{\eta}_1\boldsymbol{\eta}_3{}'-(p_2-2)T_n^{-1}\boldsymbol{\eta}_3\boldsymbol{\eta}_1{}'+(p_2-2)^2T_n^{-2}\boldsymbol{\eta}_3\boldsymbol{\eta}_3{}'] \\
&= Var(\boldsymbol{\eta}_1)-2(p_2-2)E\lim_{n\to\infty}(\boldsymbol{\eta}_3\boldsymbol{\eta}_1{}'T_n^{-1})+(p_2-2)^2E\lim_{n\to\infty}(T_n^{-2}\boldsymbol{\eta}_3\boldsymbol{\eta}_3{}').
\end{aligned}
$$

Note that, by using Lemma 2.3.2, Lemma 2.3.3 and the formula for a conditional mean of a multivariate normal, we have

$$
\begin{aligned}
E\lim_{n\to\infty}(\boldsymbol{\eta}_3\boldsymbol{\eta}_1{}'T_n^{-1}) &= E\lim_{n\to\infty}(E(\boldsymbol{\eta}_3\boldsymbol{\eta}_1{}'T_n^{-1}|\boldsymbol{\eta}_3)) \\
&= E\lim_{n\to\infty}(\boldsymbol{\eta}_3[E(\boldsymbol{\eta}_1)+\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{*-1}(\boldsymbol{\eta}_3-E(\boldsymbol{\eta}_3))]'T_n^{-1}) \\
&= E\lim_{n\to\infty}(\boldsymbol{\eta}_3[\boldsymbol{\eta}_3{}'-\boldsymbol{\gamma}']T_n^{-1}) \\
&= E\lim_{n\to\infty}(\boldsymbol{\eta}_3\boldsymbol{\eta}_3{}'T_n^{-1})-E\lim_{n\to\infty}(\boldsymbol{\eta}_3\boldsymbol{\gamma}'T_n^{-1}) \\
&= \boldsymbol{\Sigma}^*E(\chi^{-2}_{p_2+2}(\Delta))+\boldsymbol{\gamma}\boldsymbol{\gamma}'E(\chi^{-2}_{p_2+4}(\Delta))-\boldsymbol{\gamma}\boldsymbol{\gamma}'E(\chi^{-2}_{p_2+2}(\Delta)) \\
&= \boldsymbol{\Sigma}^*E(\chi^{-2}_{p_2+2}(\Delta))+\boldsymbol{\gamma}\boldsymbol{\gamma}'[E(\chi^{-2}_{p_2+4}(\Delta))-E(\chi^{-2}_{p_2+2}(\Delta))].
\end{aligned}
$$

Therefore,

$$\begin{aligned}
\Gamma(\hat{\boldsymbol{\beta}}_1^S) &= \boldsymbol{\Sigma} - 2(p_2-2)\boldsymbol{\Sigma}^* E(\chi^{-2}_{p_2+2}(\Delta)) - 2(p_2-2)\boldsymbol{\gamma}\boldsymbol{\gamma}'[E(\chi^{-2}_{p_2+4}(\Delta)) - E(\chi^{-2}_{p_2+2}(\Delta))] \\
&+ (p_2-2)^2[\boldsymbol{\Sigma}^* E(\chi^{-4}_{p_2+2}(\Delta)) + \boldsymbol{\gamma}\boldsymbol{\gamma}' E(\chi^{-4}_{p_2+4}(\Delta))] \\
&= \boldsymbol{\Sigma} + \boldsymbol{\Sigma}^*[-2(p_2-2)E(\chi^{-2}_{p_2+2}(\Delta)) + (p_2-2)^2 E(\chi^{-4}_{p_2+2}(\Delta))] \\
&+ \boldsymbol{\gamma}\boldsymbol{\gamma}'[-2(p_2-2)E(\chi^{-2}_{p_2+4}(\Delta)) + 2(p_2-2)E(\chi^{-2}_{p_2+2}(\Delta)) + (p_2-2)^2 E(\chi^{-4}_{p_2+4}(\Delta))] \\
&= \boldsymbol{\Sigma} + (p_2-2)\boldsymbol{\Sigma}^*[-2E(\chi^{-2}_{p_2+2}(\Delta)) + (p_2-2)E(\chi^{-4}_{p_2+2}(\Delta))] \\
&+ (p_2-2)\boldsymbol{\gamma}\boldsymbol{\gamma}'[-2E(\chi^{-2}_{p_2+4}(\Delta)) + 2E(\chi^{-2}_{p_2+2}(\Delta)) + (p_2-2)E(\chi^{-4}_{p_2+4}(\Delta))]
\end{aligned}$$

$$\begin{aligned}
\Gamma(\hat{\boldsymbol{\beta}}_1^{S+}) &= E \lim_{n\to\infty} (n(\hat{\boldsymbol{\beta}}_1^{S+} - \boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_1^{S+} - \boldsymbol{\beta}_1)') \\
&= E \lim_{n\to\infty} \sqrt{n}\left((\hat{\boldsymbol{\beta}}_1^S - (1-(p_2-2)T_n^{-1})I(T_n < p_2-2)(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1) - \boldsymbol{\beta}_1 \right) \\
&\times \sqrt{n}\left(\hat{\boldsymbol{\beta}}_1^S - (1-(p_2-2)T_n^{-1})I(T_n < p_2-2)(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1) - \boldsymbol{\beta}_1 \right)' \\
&= \Gamma(\hat{\boldsymbol{\beta}}_1^S) - 2E \lim_{n\to\infty} [\boldsymbol{\eta}_3 \boldsymbol{\eta}_2'(1-(p_2-2)T_n^{-1})I(T_n < p_2-2)] \\
&- 2E \lim_{n\to\infty} [\boldsymbol{\eta}_3 \boldsymbol{\eta}_3'(1-(p_2-2)T_n^{-1})^2 I(T_n < p_2-2)] \\
&+ E \lim_{n\to\infty} (\boldsymbol{\eta}_3 \boldsymbol{\eta}_3'(1-(p_2-2)T_n^{-1})^2 I(T_n < p_2-2)) \\
&= \Gamma(\hat{\boldsymbol{\beta}}_1^S) - 2E \lim_{n\to\infty} [\boldsymbol{\eta}_3 \boldsymbol{\eta}_2'(1-(p_2-2)T_n^{-1})I(T_n < p_2-2)] \\
&- E \lim_{n\to\infty} [\boldsymbol{\eta}_3 \boldsymbol{\eta}_3'(1-(p_2-2)T_n^{-1})^2 I(T_n < p_2-2)].
\end{aligned}$$

Now we have

$$
\begin{aligned}
&E\lim_{n\to\infty}[\boldsymbol{\eta}_3\boldsymbol{\eta}_2{}'(1-(p_2-2)T_n^{-1})I(T_n<p_2-2)] = \\
&E\lim_{n\to\infty}[\boldsymbol{\eta}_3 E(\boldsymbol{\eta}_2{}'(1-(p_2-2)T_n^{-1})I(T_n<p_2-2)|\boldsymbol{\eta}_3)] = \\
&E\lim_{n\to\infty}[\boldsymbol{\eta}_3(-\boldsymbol{\gamma}+\mathbf{0}\times\boldsymbol{\Sigma}^*(\boldsymbol{\eta}_3-\boldsymbol{\gamma}))'(1-(p_2-2)T_n^{-1})\times I(T_n<p_2-2)] = \\
&E\lim_{n\to\infty}[\boldsymbol{\eta}_3\boldsymbol{\gamma}'(1-(p_2-2)T_n^{-1})I(T_n<p_2-2)] - \\
&\boldsymbol{\gamma}\boldsymbol{\gamma}'E[(1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta))I(\chi^2_{p_2+2}(\Delta)<p_2-2)],
\end{aligned}
$$

and based on Lemma 2.3.3 we have

$$
\begin{aligned}
&E\lim_{n\to\infty}[\boldsymbol{\eta}_3\boldsymbol{\eta}_3{}'(1-(p_2-2)T_n^{-1})I(T_n<p_2-2)] = \\
&\boldsymbol{\Sigma}^*E[(1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta))^2 I(\chi^2_{p_2+2}(\Delta)<p_2-2)] \\
&+\boldsymbol{\gamma}\boldsymbol{\gamma}'E[(1-(p_2-2)\chi^{-2}_{p_2+4}(\Delta))^2 I(\chi^2_{p_2+4}(\Delta)<p_2-2)].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\Gamma(\hat{\boldsymbol{\beta}}_1^{S+}) &= \Gamma(\hat{\boldsymbol{\beta}}_1^{S}) + 2\boldsymbol{\gamma}\boldsymbol{\gamma}'E[(1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta))I(\chi^2_{p_2+2}(\Delta)<p_2-2)] \\
&- \boldsymbol{\Sigma}^*E[(1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta))^2 I(\chi^2_{p_2+2}(\Delta)<p_2-2)] \\
&- \boldsymbol{\gamma}\boldsymbol{\gamma}'E[(1-(p_2-2)\chi^{-2}_{p_2+4}(\Delta))^2 I(\chi^2_{p_2+4}(\Delta)<p_2-2)].
\end{aligned}
$$

□

The asymptotic risk expressions for the estimators are contained in the following corollary.

Corollary 2.4.2. Suppose that conditions in Theorem 2.4.1 hold. Then under $\{K_n\}$, as

$n \to \infty$, the ADR of the estimators $\hat{\boldsymbol{\beta}}_1$, $\tilde{\boldsymbol{\beta}}_1$, $\hat{\boldsymbol{\beta}}_1^S$ and $\hat{\boldsymbol{\beta}}_1^{S+}$ are respectively

$$
\begin{aligned}
ADR(\hat{\boldsymbol{\beta}}_1;\boldsymbol{M}) &= \sigma^* \operatorname{tr}(\boldsymbol{M}\boldsymbol{B}_{11.2}^{-1}),\\
ADR(\tilde{\boldsymbol{\beta}}_1;\boldsymbol{M}) &= \sigma^* \operatorname{tr}(\boldsymbol{M}\boldsymbol{B}_{11.2}^{-1} - \boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{M}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22.1}^{-1}) + \boldsymbol{\omega}'\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{M}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega},\\
ADR(\hat{\boldsymbol{\beta}}_1^S;\boldsymbol{M}) &= \sigma^* \operatorname{tr}(\boldsymbol{M}\boldsymbol{B}_{11.2}^{-1})\\
&- \sigma^*\Big[(p_2-2)\operatorname{tr}(\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{M}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22.1}^{-1})\{2E(\chi_{p_2+2,\alpha}^{-2}(\Delta))\\
&- (p_2-2)E(\chi_{p_2+2}^{-4}(\Delta))\}\Big] + (p_2-2)\boldsymbol{\omega}'\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{M}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}\\
&\times \Big\{2E(\chi_{p_2+2}^{-2}(\Delta) + (p_2-2)E(\chi_{p_2+4}^{-4}(\Delta) - 2E(\chi_{p_2+4}^{-2}(\Delta))\Big\},\\
ADR(\hat{\boldsymbol{\beta}}_1^{S+};\boldsymbol{M}) &= ADR(\hat{\boldsymbol{\beta}}_1^S;\boldsymbol{M})\\
&+ \sigma^*\Big\{(p_2-2)\operatorname{tr}(\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{M}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22.1}^{-1})\{2E\left[\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\right]\\
&- (p_2-2)E\left[\chi_{p_2+2}^{-4}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\right]\}\\
&- \operatorname{tr}(\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{M}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{B}_{22.1}^{-1})H_{p_2+2}(p_2-2;\Delta)\Big\}\\
&+ \boldsymbol{\omega}'\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{M}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}\Big[2H_{p_2+2}(p_2-2;\Delta) - H_{p_2+4}(p_2-2;\Delta)\Big]\\
&- (p_2-2)\boldsymbol{\omega}'\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{M}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\boldsymbol{\omega}\Big\{2E\left[\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\right]\\
&- 2E\left[\chi_{p_2+4}^{-2}(\Delta)I(\chi_{p_2+4}^{2}(\Delta) \le p_2-2)\right]\\
&+ (p_2-2)E\left[\chi_{p_2+4}^{-4}(\Delta)I(\chi_{p_2+4}^{2}(\Delta) \le p_2-2)\right]\Big\}.
\end{aligned}
$$

Proof. The expressions for risk are obtained by following the definition of ADR. □

2.4.3 Bias and Risk Comparison

According to Theorem 2.4.1 and 2.4.2, if $\boldsymbol{B}_{12} = 0$, then all the AQDB reduce to zero and all the ADR reduce to common value $(1-\theta^2-\sigma_z^2)^{-1}\{(1-\theta^2)\sigma_e^2\}\operatorname{tr}(\boldsymbol{M}\boldsymbol{B}_{11}^{-1})$ for all $\boldsymbol{\omega}$. Hence, in sequel we assume that $\boldsymbol{B}_{12} \neq 0$. By Theorem 2.4.1 we know the AQDB of

$\tilde{\beta}_1$ is an unbounded function of $\omega' B_{21} B_{11}^{-1} B_{11.2}\ B_{11}^{-1} B_{12} \omega$. By application of Courant's theorem, in order to investigate the AQDB($\hat{\beta}_1^S$) and AQDB($\hat{\beta}_1^{S+}$), we will have

$$\begin{aligned} \mathrm{ch}_{\min}(B_{21} B_{11}^{-1} B_{11.2} B_{11}^{-1} B_{12} B_{22.1}^{-1}) &\le \frac{\omega' B_{21} B_{11}^{-1} B_{11.2} B_{11}^{-1} B_{12} \omega}{\omega' B_{22.1} \omega} \\ &\le \mathrm{ch}_{\max}(B_{21} B_{11}^{-1} B_{11.2} B_{11}^{-1} B_{12} B_{22.1}^{-1}). \end{aligned}$$

Therefore, AQDB($\hat{\beta}_1^S$) starts from zero at $\omega' B_{21} B_{11}^{-1} B_{11.2} B_{11}^{-1} B_{12} \omega = 0$, increases to a point, then decreases towards zero due to $E(\chi_{p_2+2}^{-2}(\Delta))$ being a decreasing log-convex function of Δ.

Now, we provide ADR analysis. By comparing ADR($\hat{\beta}_1^S$) and ADR($\hat{\beta}_1$), we can see if $M \in M^D$ $\hat{\beta}_1^S$ dominates $\hat{\beta}_1$ for any ω in the sense of ADR where

$$M^D = \left\{ M : \frac{\mathrm{tr}(B_{21} B_{11}^{-1} M B_{11}^{-1} B_{12} B_{22.1}^{-1})}{\mathrm{ch}_{\max}(B_{21} B_{11}^{-1} M B_{11}^{-1} B_{12} B_{22.1}^{-1})} \ge \frac{p_2+2}{2} \right\}.$$

The behavior of $\hat{\beta}_1^{S+}$ is similar to $\hat{\beta}_1^S$; however, the quadratic bias curve of $\hat{\beta}_1^{S+}$ remains below the curve of $\hat{\beta}_1^S$ for all values of Δ. By comparing ADR($\hat{\beta}_1^{S+}$) with ADR($\hat{\beta}_1^S$), we observe $\hat{\beta}_1^{S+}$ dominates $\hat{\beta}_1^S$ for all the values of ω, with strict inequality for some ω. Further, the largest risk improvement of $\hat{\beta}_1^{S+}$ over $\hat{\beta}_1^S$ is near the null hypothesis. Therefore, the risk of $\hat{\beta}_1^{S+}$ is also smaller than the risk of $\hat{\beta}_1$ in the entire parameter space, and the upper limit is attained when Δ approaches ∞. It also clearly indicates the asymptotic inferiority of $\hat{\beta}_1^S$ and $\hat{\beta}_1$ compared to $\hat{\beta}_1^{S+}$ for $\Delta \in [0, \infty)$. ADR($\hat{\beta}_1^{S+}$) increases monotonically towards ADR($\hat{\beta}_1$) from below, as Δ moves away from 0. This implies that

$$\mathrm{ADR}(\hat{\beta}_1^{S+}) \le \mathrm{ADR}(\hat{\beta}_1^S) \le \mathrm{ADR}(\hat{\beta}_1) \text{ for any } M \in M^D \text{ and } \omega$$

with strict inequality for some ω. By the analysis above, we conclude that $\hat{\boldsymbol{\beta}}_1^S$ and $\hat{\boldsymbol{\beta}}_1^{S+}$ perform better than $\hat{\boldsymbol{\beta}}_1$ regardless of the correctness of UPI, although the gain in risk over $\hat{\boldsymbol{\beta}}_1$ is substantial when the UPI is nearly correct. Moreover, $\hat{\boldsymbol{\beta}}_1^{S+}$ is asymptotically superior to $\hat{\boldsymbol{\beta}}_1^S$. In the next section we present our simulation results and compare the performances of the proposed estimators.

2.5 Monte Carlo Simulation

In this section, we carry out a Monte Carlo simulation study to examine the relative performance of the proposed estimators. In Subsection 2.5.1, we consider the situation when $n < p$ by implementing APE strategy. In Subsection 2.5.2, we do the simulation for the case when $n \geq p$ and compare our proposed shrinkage estimators with APE, UE, and RE. In our simulation study, we use the following model:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{2.8}$$

with $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$. Moreover, $\varepsilon_i = (\theta + z_i)\varepsilon_{i-1} + e_i$, $\theta = 0.4$ with z_i and e_i being i.i.d N(0,1). Furthermore, $x_{si} = t_s + \nu_i$ with t_s being i.i.d $N(\frac{s}{2}, \frac{s}{2} + 0.1)$ and ν_i are i.i.d N(0,1) for all $s = 1, \ldots, p$ and $i = 1, \ldots, n$. For each n, we generate 5000 samples and x_i values are generated once for each n. We use this setup for simulation experiments in Subsections 2.5.1 and 2.5.2.

2.5.1 High Dimensional Estimation

In this section we consider the case when ($n < p$). In this situation, the GLS estimator cannot be achieved, which is why we perform the APE strategy. Consider the following cases:

Model 1: $p = 20$, $n = 15$ and $\boldsymbol{\beta} = (\mathbf{1}', \mathbf{0}')'$ where $\mathbf{1}$ and $\mathbf{0}$ are 10×1 vectors.

Model 2: $p = 20$, $n = 10$ and $\boldsymbol{\beta} = (\mathbf{1}', \mathbf{0}')'$ where $\mathbf{1}$ and $\mathbf{0}$ are 10×1 vectors.

Model 3: $p = 30$, $n = 20$ and $\boldsymbol{\beta} = (\mathbf{1}', \mathbf{0}')'$ where $\mathbf{1}$ and $\mathbf{0}$ are 15×1 vectors.

Model 4: $p = 30$, $n = 15$ and $\boldsymbol{\beta} = (\mathbf{1}', \mathbf{0}')'$ where $\mathbf{1}$ and $\mathbf{0}$ are 15×1 vectors.

Model 5: $p = 40$, $n = 20$ and $\boldsymbol{\beta} = (\mathbf{1}', \mathbf{0}')'$ where $\mathbf{1}$ and $\mathbf{0}$ are 20×1 vectors.

Model 6: $p = 40$, $n = 15$ and $\boldsymbol{\beta} = (\mathbf{1}', \mathbf{0}')'$ where $\mathbf{1}$ and $\mathbf{0}$ are 20×1 vectors.

We used the lars (Hastie and Efron, 2012) and parcor (Kraemer and Schaefer, 2010) package in R software (R Development Core Team, 2011) to achieve lasso and AL estimators, respectively. In parcor package, the ${w_j}'s$ are computed in terms of a lasso fit. Also, a 10-fold cross-validation method is used for choosing the best value for λ. We show the results of model 1 and 2 in Table 2.1 and Table 2.2. We compared the performance of these two estimators for the above models based on their total simulated MSE as shown in Table 2.3.

Table 2.1: APE estimators of Model 1

$\boldsymbol{\beta}$	1	1	1	1	1	1	1	1	1	1
$\hat{\boldsymbol{\beta}}_{AL}$	1.014	0.989	0.993	0.993	1.000	0.988	1.001	0.988	0.991	0.992
$\hat{\boldsymbol{\beta}}_{lasso}$	1.008	0.963	0.955	1.011	0.933	0.982	0.988	0.947	1.047	1.002
$\boldsymbol{\beta}$	0	0	0	0	0	0	0	0	0	0
$\hat{\boldsymbol{\beta}}_{AL}$	0.000	0.013	0.028	0.004	0.001	0.000	0.000	0.000	0.000	0.008
$\hat{\boldsymbol{\beta}}_{lasso}$	-0.009	0.000	0.001	0.033	0.000	0.011	0.083	-0.003	0.000	-0.002

Both lasso and AL methods applied to the transformed data, provide estimates close to the true parameters. In both models, the total MSE of the APE estimators are calculated. In

Table 2.2: APE estimators of Model 2

β	1	1	1	1	1	1	1	1	1	1
$\hat{\beta}_{AL}$	1.019	1.080	0.988	0.984	1.000	0.988	0.962	0.844	1.021	1.013
$\hat{\beta}_{lasso}$	0.984	0.991	0.986	0.977	0.986	0.967	0.999	0.985	1.000	0.997
β	1	0	0	0	0	0	0	0	0	0
$\hat{\beta}_{AL}$	0.006	0.000	0.003	-0.003	0.000	0.000	0.000	0.000	0.004	0.000
$\hat{\beta}_{lasso}$	0.000	0.000	-0.009	0.039	-0.039	0.000	0.000	0.000	0.000	-0.008

model 1, $MSE(\hat{\beta}_{lasso}) = 4.81$ whereas the AL estimators have a total MSE of $MSE(\hat{\beta}_{AL}) = 2.55$ (see Table 2.3). As we see, the AL estimators have smaller MSE than lasso and similar results are obtained for model 2, which indicate the AL method gives better performance than lasso.

Table 2.3: Simulated MSE of APE estimators

Model	$MSE(\hat{\beta}_{lasso})$	$MSE(\hat{\beta}_{AL})$
Model 1 ($n = 15, p = 20$)	4.81	2.55
Model 2 ($n = 10, p = 20$)	5.45	3.71
Model 3 ($n = 20, p = 30$)	3.75	3.17
Model 4 ($n = 15, p = 30$)	3.92	3.66
Model 5 ($n = 20, p = 40$)	4.85	4.76
Model 6 ($n = 15, p = 40$)	5.69	5.32

2.5.2 Fixed Dimension Estimation

In this section, we consider model (2.8) and compare the performance of different estimators when $n \geq p$. We set the regression coefficients $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ to $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \mathbf{0}')'$ for the following cases:

Case 1: $\boldsymbol{\beta}_1 = (1,1,1)'$ and $\boldsymbol{\beta}_2 = \mathbf{0}_{p_2\times 1}$ where $\boldsymbol{\beta}_2$ is a vector of 0 with dimensions $p_2 = 3,4,\cdots,8$,

Case 2: $\boldsymbol{\beta}_1 = (1,1,1,1,1)'$ and $\boldsymbol{\beta}_2 = \mathbf{0}_{p_2\times 1}$ where $p_2 = 3,4,\cdots,8$.

Now we define the parameter $\Delta^* = \|\boldsymbol{\beta} - \boldsymbol{\beta}^*\|^2$, where $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1', \mathbf{0}')'$ and $\| \, . \, \|$ is the Euclidian norm. Here Δ^* is the degree of deviation from the restriction $(\boldsymbol{\beta}_2 = \mathbf{0})$. The objective is to investigate the behavior of the proposed estimators of $\boldsymbol{\beta}_1$ under varying degrees of model misspecification, i.e., when $\Delta^* \geq 0$. In order to do this, further samples are generated from those distributions under local alternative hypotheses. Various Δ^* values between 0 and 2 are considered. To produce different values of Δ^*, different values of $\boldsymbol{\beta}_2$ are chosen.

To compare the performance of the proposed estimators of $\boldsymbol{\beta}_1$ we have numerically calculated their risk. Their performance was evaluated in terms of relative mean square error (RMSE). The simulated RMSE of $\hat{\boldsymbol{\beta}}_1^\star$ to the unrestricted estimator $\hat{\boldsymbol{\beta}}_1$ is defined by

$$RMSE(\hat{\boldsymbol{\beta}}_1 : \hat{\boldsymbol{\beta}}_1^\star) = \frac{MSE(\hat{\boldsymbol{\beta}}_1)}{MSE(\hat{\boldsymbol{\beta}}_1^\star)},$$

where $\hat{\boldsymbol{\beta}}_1^\star$ can be any of $\tilde{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_1^S, \hat{\boldsymbol{\beta}}_1^{S+}, \hat{\boldsymbol{\beta}}_1^{lasso}$ and $\hat{\boldsymbol{\beta}}_1^{AL}$. It is obvious that a RMSE larger than one indicates the degree of superiority of the estimator $\hat{\boldsymbol{\beta}}_1^\star$ over $\hat{\boldsymbol{\beta}}_1$.

We designed the simulation study for the sample sizes $n = 30, 50, 80$ and 100. Since the results were similar for different sample sizes, We report the results for $n = 30$ and 100 with $p_2 = 3, 4, \cdots, 8$. Comparative RMSEs for RE, UE, SE and PSE in Figures 2.1-2.4 portray the relative performance of the suggested estimators. The line at RMSE=1 indicates the UE as the $RMSE(\hat{\boldsymbol{\beta}}_1 : \hat{\boldsymbol{\beta}}_1) = 1$.

As we see in the figures, for all combinations of p_2 and n, the RE outperforms both SE and PSE at and near $\Delta^* = 0$, i.e., RE is the optimal choice as an estimator of $\boldsymbol{\beta}_1$ in this case. However if the sub-model is misspecified, i.e., Δ^* moves away from 0 and the restriction is not correctly specified, then the estimated risk of RE increases and becomes unbounded. Therefore its RMSE goes below the horizontal line at RMSE=1 and its efficiency converges to 0.

The figures also reveal that as Δ^* moves away from 0, the RMSE of both SE and PSE decreases, i.e., their risk increases but remains bounded for the worst case when $\Delta^* >> 0$. In this case their risk converges to the risk of UE irrespective of p_1, p_2 and n. Also in all the cases, the RMSE of SE remains below the RMSE of PSE which indicates the superiority and better performance of PSE over SE and that the risk of PSE is less than SE. Thus one can not go wrong by choosing the PSE even if the restriction or the sub-model is not correctly specified. In this case, the PSE has the highest risk, equal to the risk of UE.

Comparison of Shrinkage with Absolute Penalty Estimator

We also compare the performance of shrinkage estimators with APEs (lasso and AL) relative to $\hat{\boldsymbol{\beta}}_1$ based on the RMSE criterion. We used the 10-fold cross validation method to estimate the tuning parameter λ to compute APEs. For comparison purposes we considered again Case 1 and 2 for $p_2 = 3, 4, \cdots, 9$ and $n = 30, 100$. We compare the RMSE only at $\Delta^* = 0$, since according to Ahmed et al. (2007), APE does not take advantage of the fact that the parameter vector $\boldsymbol{\beta}$ is partitioned into main and nuisance parts, and is at a disadvantage when $\Delta^* > 0$. Simulated RMSE for both cases and different sample sizes are presented in Tables 2.4-2.7.

The results reveal that the AL estimator outperforms the lasso in all the cases. In the first case when $p_1 = 3$, we see that both APEs outperform shrinkage estimators when p_2 is small indicating that APEs have lower MSE compared to SE and PSE. But as p_2 increases, the RMSE of shrinkage estimators beat the RMSE of APEs. When $p_1 = 3$, the RMSE of PSE is higher than lasso when $p_2 \geq 3$ and is higher than AL when $p_2 \geq 4$. Also SE has higher RMSE than AL and lasso when $p_2 \geq 6$ and $p_2 \geq 7$, respectively. Similar results are

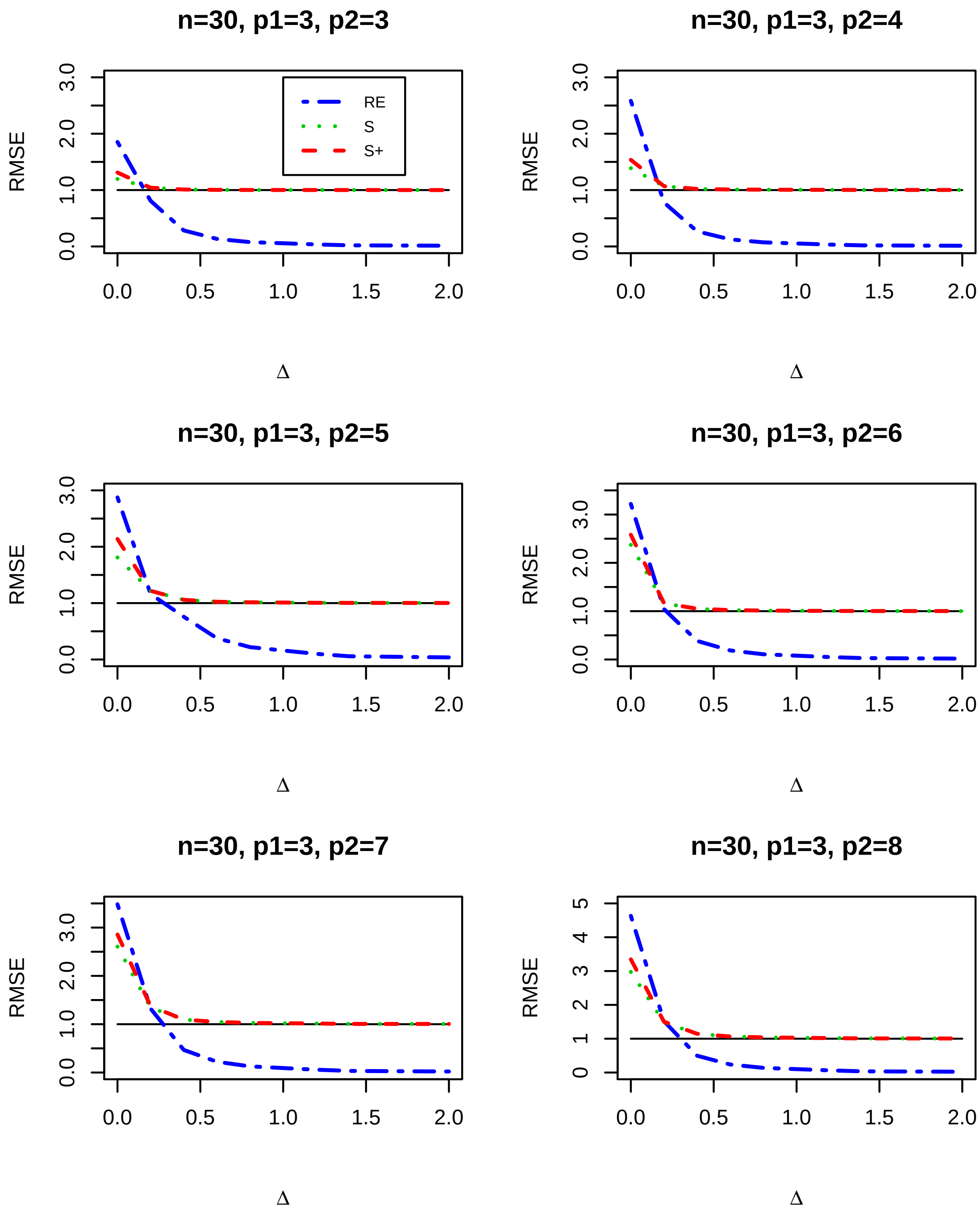

Figure 2.1: Relative MSE of the estimators for various p_2 when $p_1 = 3$ and $n = 30$. "- - -" denotes the positive shrinkage estimator, "$\cdots$" denotes the shrinkage estimator, "— - — -" denotes the restricted estimator and "—" denotes the unrestricted estimator.

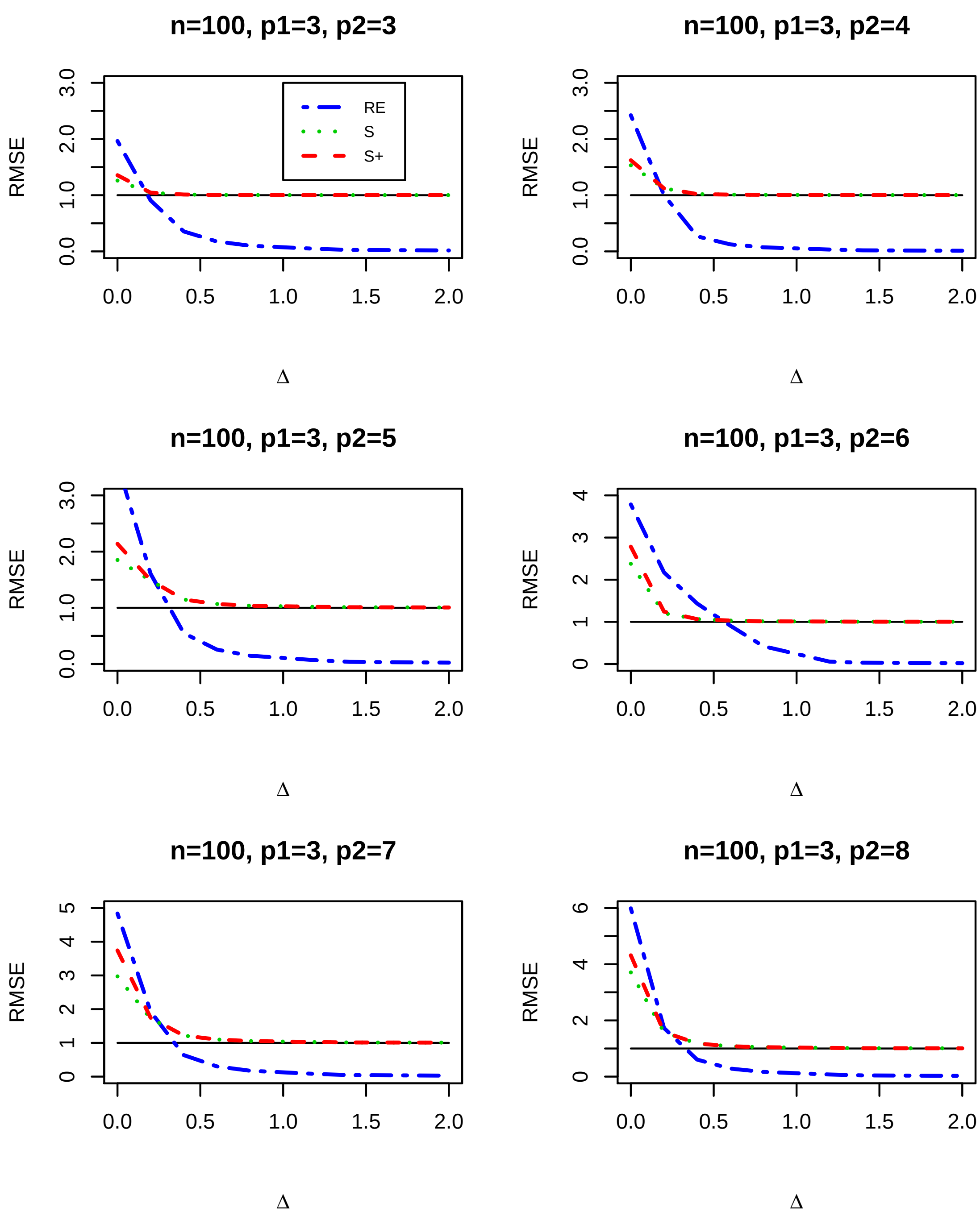

Figure 2.2: Relative MSE of the estimators for various p_2 when $p_1 = 3$ and $n = 100$. “- - -” denotes the positive shrinkage estimator, “$\cdots$” denotes the shrinkage estimator, “— - — -” denotes the restricted estimator and “—” denotes the unrestricted estimator.

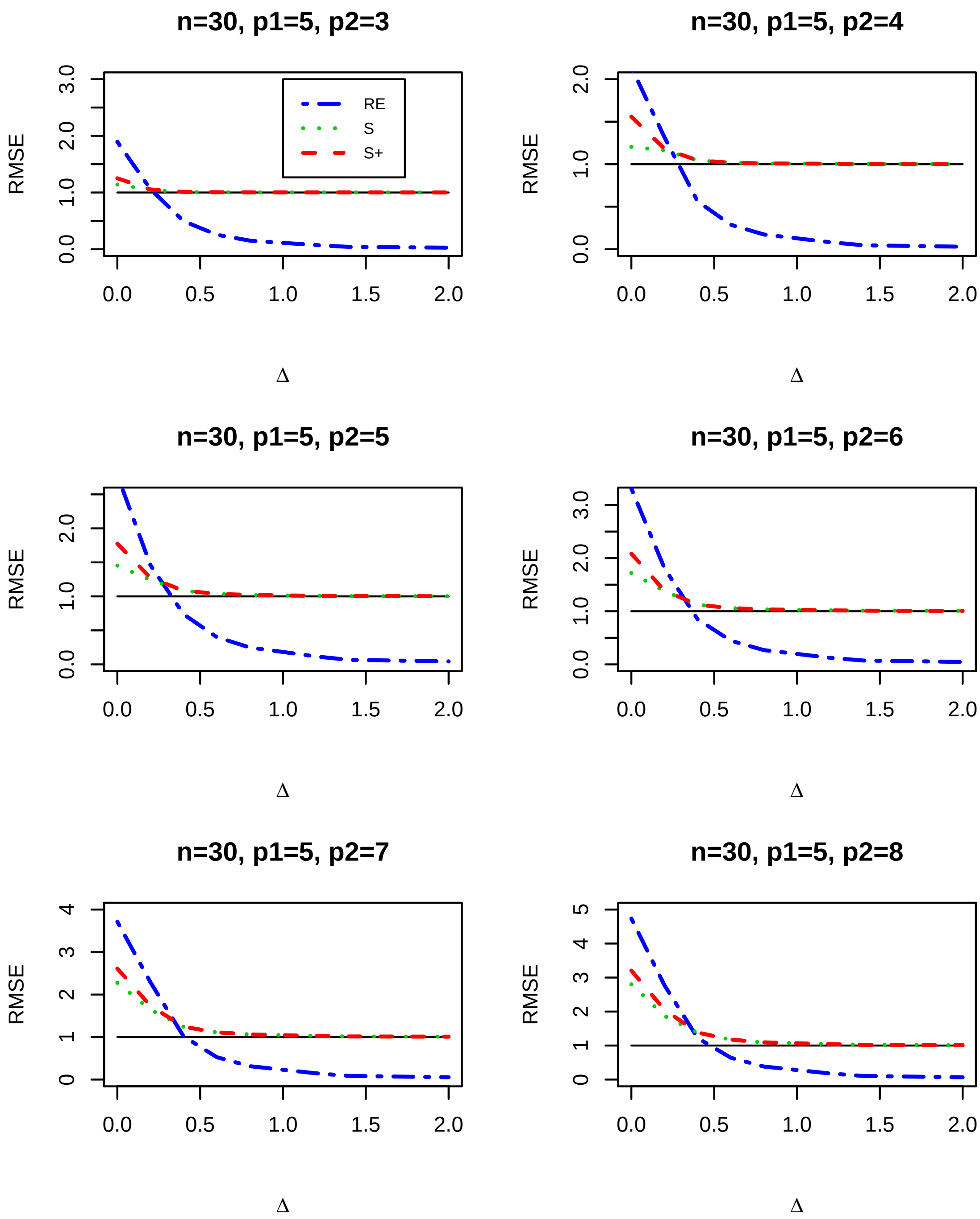

Figure 2.3: Relative MSE of the estimators for various p_2 when $p_1 = 5$ and $n = 30$. "- - -" denotes the positive shrinkage estimator, "···" denotes the shrinkage estimator, "— - — -" denotes the restricted estimator and "—" denotes the unrestricted estimator.

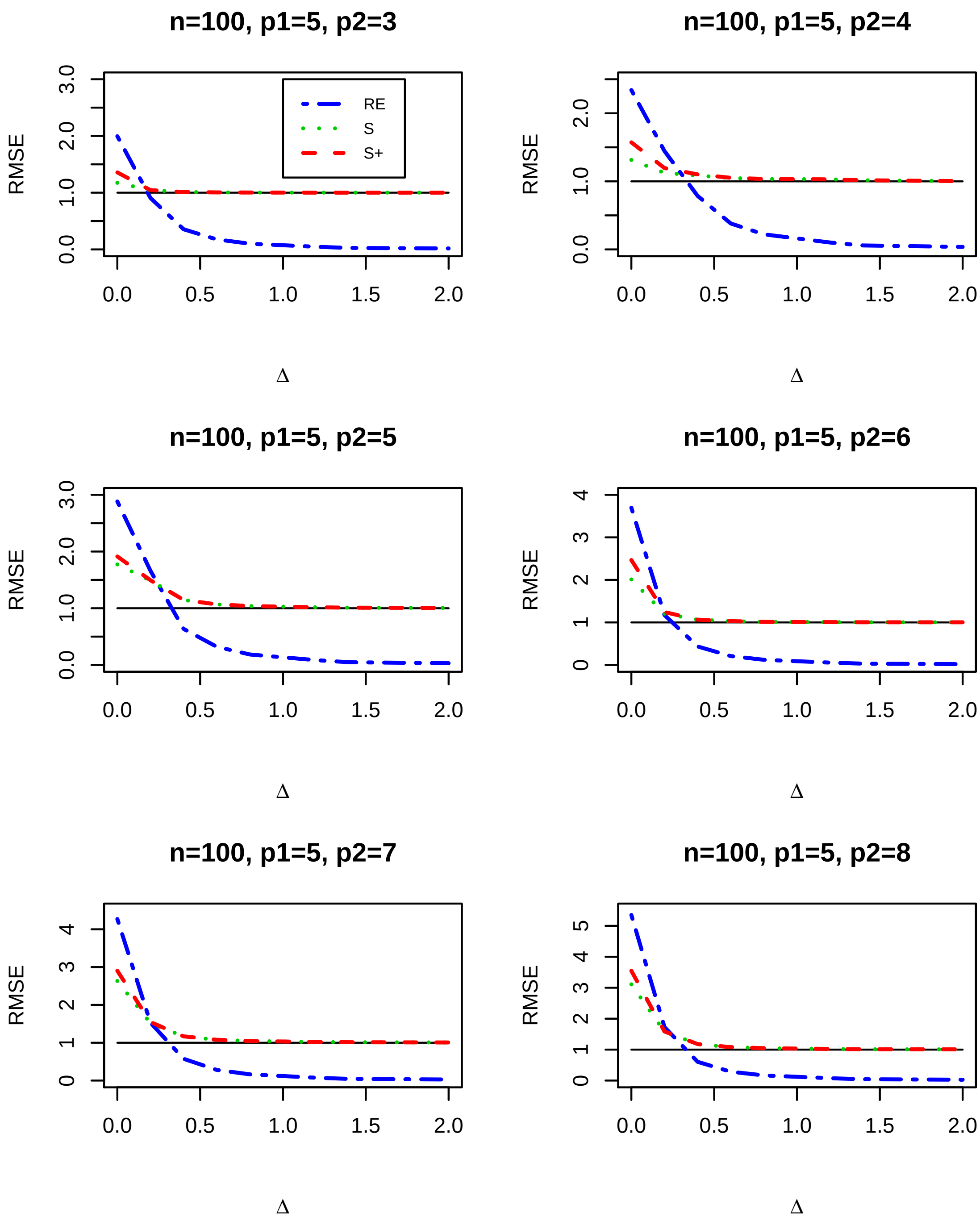

Figure 2.4: Relative MSE of the estimators for various p_2 when $p_1 = 5$ and $n = 100$. "- - -" denotes the positive shrinkage estimator, "···" denotes the shrinkage estimator, "— - — -" denotes the restricted estimator and "—" denotes the unrestricted estimator.

obtained when $p_1 = 5$ which indicates that shrinkage estimators perform better than APEs when there are many nuisance predictors in the model. The gain in efficiency depends on the value of p_2, that is, the larger p_2 is relative to p_1, the larger the gain in efficiency.

Table 2.4: Simulated RMSE of shrinkage and APE estimators with respect to $\hat{\boldsymbol{\beta}}_1$ when $p_1 = 3, n = 30$

p_2	$\tilde{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{S+}$	$\hat{\boldsymbol{\beta}}_{lasso}$	$\hat{\boldsymbol{\beta}}_{AL}$
3	1.853	1.197	1.311	1.388	1.475
4	2.583	1.388	1.617	1.536	1.743
5	2.875	1.810	2.144	1.822	2.122
6	3.219	2.355	2.481	2.103	2.372
7	3.482	2.606	2.756	2.227	2.536
8	4.632	2.974	3.016	2.562	2.825
9	5.216	3.346	3.566	2.982	3.251

Table 2.5: Simulated RMSE of shrinkage and APE estimators with respect to $\hat{\boldsymbol{\beta}}_1$ when $p_1 = 3, n = 100$

p_2	$\tilde{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_1^S$	$\hat{\boldsymbol{\beta}}_1^{S+}$	$\hat{\boldsymbol{\beta}}_{lasso}$	$\hat{\boldsymbol{\beta}}_{AL}$
3	1.964	1.260	1.356	1.423	1.647
4	2.422	1.532	1.693	1.622	1.859
5	3.562	1.851	2.165	1.912	2.138
6	3.786	2.478	2.684	2.257	2.516
7	4.837	2.973	3.742	2.604	2.911
8	5.987	3.709	4.316	3.241	3.542
9	6.324	4.379	5.161	4.001	4.212

Table 2.6: Simulated RMSE of shrinkage and APE estimators with respect to $\hat{\beta}_1$ when $p_1 = 5, n = 30$

p_2	$\tilde{\beta}_1$	$\hat{\beta}_1^S$	$\hat{\beta}_1^{S+}$	$\hat{\beta}_{lasso}$	$\hat{\beta}_{AL}$
3	1.895	1.140	1.251	1.395	1.467
4	2.141	1.203	1.559	1.602	1.710
5	2.783	1.452	1.775	1.830	1.967
6	3.308	1.719	2.083	2.055	2.198
7	3.712	2.275	2.611	2.216	2.443
8	4.735	2.803	3.206	2.576	2.795
9	5.223	3.125	3.615	2.818	3.082

Table 2.7: Simulated RMSE of shrinkage and APE estimators with respect to $\hat{\beta}_1$ when $p_1 = 5, n = 100$

p_2	$\tilde{\beta}_1$	$\hat{\beta}_1^S$	$\hat{\beta}_1^{S+}$	$\hat{\beta}_{lasso}$	$\hat{\beta}_{AL}$
3	1.996	1.174	1.358	1.461	1.552
4	2.341	1.315	1.573	1.693	1.818
5	2.883	1.773	1.913	1.935	2.204
6	3.697	2.011	2.466	2.233	2.485
7	4.271	2.634	2.902	2.477	2.709
8	5.352	3.110	3.524	2.931	3.085
9	6.033	3.547	4.256	3.271	3.499

2.6 Data Example

We now implement suggested strategies to quarterly macroeconomic time series data (United Kingdom, 1948-1956). The data can be found in Reinsel and Velu (1998, p. 233) and they were initially analyzed by Klein et al. (1961). In this data set, we consider the dependent variable y_i as the total exports and the explanatory variables x_{i1} , x_{i2} , x_{i3} , x_{i4} and x_{i5} are total labor force, weekly wage rates, price index of imports, price index of exports and price index of consumption, respectively. The sample size is $n = 36$. We first fit a multiple regression model to the data and plot the autocorrelation and partial-autocorrelation of the OLS residuals (Figure 2.5). We also computed the Durbin-Watson statistics for the regression model (Table 2.8). The results suggest an AR(1) process at the 0.05 significance level. We now consider a regression model with first order autoregressive errors given in the form of:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \varepsilon_i, \quad i = 1, 2, \ldots, 36, \tag{2.9}$$

where ε_i follows an AR(1) process. The model estimation is $\hat{y}_i = 0.49x_{1i} + 1.64x_{2i} + 1.47x_{2i} - 1.05x_{4i} - 1.45x_{5i}$ and the corresponding residual mean square error and AIC is 54.41 and 253.66 respectively. Now we consider the following linear regression model with random coefficient AR(1) error in (2.2):

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \varepsilon_i, \quad i = 1, 2, \ldots, 36. \tag{2.10}$$

The estimated model is $\hat{y}_i = 0.51x_{1i} + 1.63x_{2i} + 1.42x_{2i} - 1.01x_{4i} - 1.46x_{5i}$. The residual mean square error and the calculated AIC for this model is 48.05 and 239.41 respectively, which are less than those in model (2.9). Therefore we choose this model for further analysis. Now based on preliminary analysis (Table 2.9), we set

$$\beta_1 = \beta_4 = \beta_5 = 0.$$

Table 2.8: Durbin-Watson test statistic

Lag	Autocorrelation	D-W Statistic	$p - value$
1	0.349	1.173	0.004
2	0.174	1.441	0.054
3	0.031	1.693	0.358
4	-0.173	2.099	0.662
5	-0.058	1.839	0.990

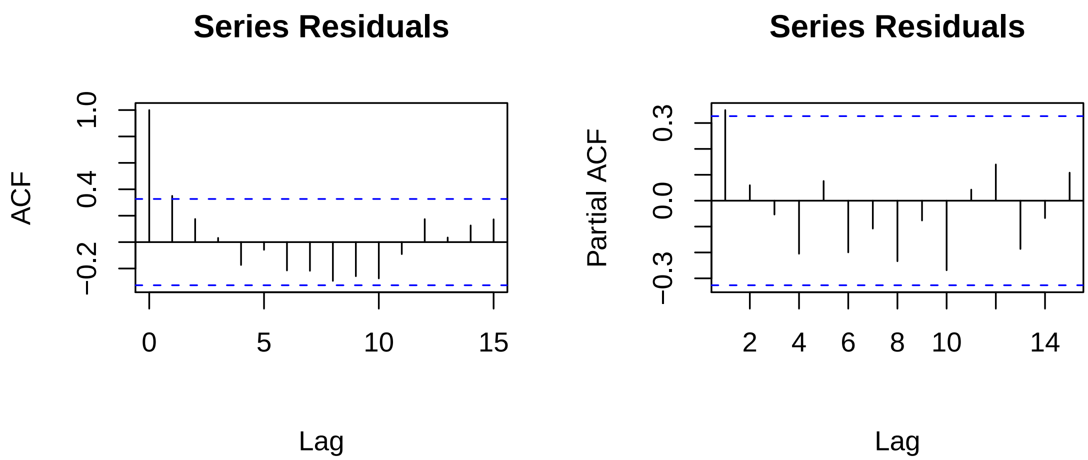

Figure 2.5: Autocorrelation and partial-autocorrelation function for the residuals from the OLS regression.

Table 2.9: Coefficients of model (2.9)

Covariate	Value	Std.Error	t-value	$p-value$
x_1	0.490	0.449	1.092	0.283
x_2	1.641	0.577	2.840	0.007*
x_3	1.473	0.522	2.821	0.008*
x_4	-1.059	0.676	-1.566	0.127
x_5	-1.450	1.190	-1.218	0.232

* Significant at 0.05

The estimation results are given in Table 2.10. The performance of the estimators are evaluated in terms of predictive MSE (PMSE). The PMSE of $\tilde{\beta}$, $\hat{\beta}^S, \hat{\beta}^{S+}, \hat{\beta}^{lasso}$ and $\hat{\beta}^{AL}$

Table 2.10: Estimated Coefficients

$Estimator$	β_1	β_2	β_3	β_4	β_5
$\hat{\beta}$	0.515	1.634	1.421	-1.009	-1.465
$\tilde{\beta}$	0	0.512	0.519	0	0
$\hat{\beta}^S$	-0.514	-1.630	-1.418	1.006	1.460
$\hat{\beta}^{S+}$	0	0.512	0.519	0	0
$\hat{\beta}_{lasso}$	0.424	0.602	0.375	0	0
$\hat{\beta}_{AL}$	0	0.441	0.618	0	0

relative to $\hat{\beta}$ is given by:

$$RPMSE(\hat{\beta} : \hat{\beta}^{*}) = \frac{PMSE(\hat{y}_i; \hat{\beta})}{PMSE(\hat{y}_i; \hat{\beta}^{*})},$$

where $\hat{\beta}^{*}$ can be any of the $\tilde{\beta}$, $\hat{\beta}^{S}, \hat{\beta}^{S+}, \hat{\beta}^{lasso}$ and $\hat{\beta}^{AL}$. The RPMSE of $\tilde{\beta}$, $\hat{\beta}^{S}, \hat{\beta}^{S+}, \hat{\beta}^{lasso}$ and $\hat{\beta}^{AL}$ is calculated and reported in Table 2.11. Table 2.11 reveals that RPMSE for PSE is larger than that of APE and lasso, which indicates the superiority of this estimator over AL and lasso. But the SE has smaller RPMSE compared to the both APEs. In fact this is because of over-shrinking problem in SE. We see this over-shrinking in the SE causes the estimations to have opposite sign than UE and the SE is being dominated by both APEs. However, the sub-model estimator under the assumption of the correctly specified model is always the best estimator and we see that the PSE has the same RPMSE as the RE with *RPMSE*=2.04.

Table 2.11: The Relative PMSE of the Estimators

$Estimators(\hat{\beta}^{*})$	$RPMSE(\hat{\beta} : \hat{\beta}^{*})$
$\tilde{\beta}$	2.04
$\hat{\beta}^{S}$	1.29
$\hat{\beta}^{S+}$	2.04
$\hat{\beta}^{lasso}$	1.65
$\hat{\beta}^{AL}$	1.72

2.7 Concluding Remarks

In this chapter, we considered multiple regression models with random coefficient autoregressive errors. We suggested estimation approaches for two scenarios: high and fixed dimension data analysis. For the case of high dimensionality, we proposed APEs including lasso and AL. For the case of fixed dimension data, we proposed shrinkage estimators. We compared the performance of these estimators via simulation. Further, we explored and compared the risk properties of shrinkage estimators, RE and UE based on the asymptotic distributional risk. The simulation results indicated that RE dominates the other estimators under a correctly specified model. Numerical results showed that the AL method performs better than lasso. Moreover, comparing the APE with shrinkage estimators demonstrated that the APE estimators are better than SE and PSE when there is a large number of predictors in the model with only a few of them being irrelevant. On the other hand, the shrinkage estimators perform well when p and the number of nuisance parameters (p_2) are relatively large. We demonstrated that, based on both analytical and numerical findings, PSE outperforms the UE and SE in the entire parameter space. When the restriction is true, RE is superior to all the other estimation rules; however, its MSE may become unbounded when such restrictions are incorrect.

Chapter 3

Estimation Strategies in Partially Linear Models with Random Coefficient Autoregressive Errors

3.1 Introduction

Many estimation problems involve an unknown function or unknown function with unknown finite-dimension parameter. Models and estimation problems that only involve an unknown function are called nonparametric whereas models with an unknown function and unknown finite-dimensional parameter are called semiparametric. Several semiparametric models have been proposed in the literature. Partially linear model, semiparametric single index models, and varying coefficient models are among the popular ones.

A partially linear model (PLM) can be written as the following:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + g(t_i) + \varepsilon_i, i = 1, \ldots, n, \tag{3.1}$$

where y_i's are responses, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ and $t_i \in [0,1]$ are design points, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is an unknown parameter vector, $g(\cdot)$ is an unknown bounded real-valued smooth function defined on the compact subset $[0,1]$ which will be estimated based on nonparametric methods using a kernel function, and ε_i's are unobservable random errors with mean zero. In nonparametric statistics, a kernel is a weighting function used in non-parametric estimation techniques. It is a non-negative real-valued integrable function K satisfying the following two requirements:

1) $\int_{-\infty}^{+\infty} K(u)du = 1$;

2) $K(-u) = K(u)$ for all values of u.

Let $(t_1, t_2, \cdots, t_n)$ be a sample drawn from some distribution with an unknown density f. We are interested in estimating the shape of this function f. Its kernel density estimator is

$$\hat{f}_h(t) = \frac{1}{n}\sum_{i=1}^{n} K_h(t - t_i) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{t - t_i}{h}),$$

where $K(.)$ is the kernel function and $h > 0$ is a smoothing parameter called the bandwidth. A range of kernel functions are commonly used: uniform, triangular, triweight, Gaussian, and others.

An advantage of the model (3.1) is that it allows dependence of the response variable y_i on some covariates in an unknown fashion and, hence, is more flexible than the conventional linear regression model. However, it is to be noted that the bulk of the research that has been done so far assumes that ε_i are i.i.d. random variables. In practice, this is not always the case, especially for sequentially collected economic data that often exhibit evident dependence in the errors.

One interesting case is to consider that the errors are modeled by a first order random coefficient autoregressive (RCAR(1)) process, that is, a stationary solution of

$$\varepsilon_i = (\theta + z_i)\varepsilon_{i-1} + e_i, \qquad i = 1, \ldots, n, \tag{3.2}$$

where θ is the autoregression parameter, $\{z_i\}$ and $\{e_i\}$ are zero mean independent processes each consisting of i.i.d. random variables with finite second moments σ_z^2 and σ_e^2 respectively. You and Chen (2002) considered estimation of the regression and autocorrelation parameters of model (3.1) with errors in (3.2) and investigated their properties.

For model (3.1) with independent errors, Ahmed et al. (2007) considered a profile least squares approach based on using kernel estimates of $g(.)$ to construct absolute penalty, shrinkage, and pretest estimators of the regression parameters $\boldsymbol{\beta}$. They also studied the relative performance of APE with shrinkage and positive shrinkage estimators through Monte Carlo simulation.

In this chapter, we extend their work with errors given in (3.2). We consider variable selection and parameter estimation when the parameter vector $\boldsymbol{\beta}$ can be partitioned to $(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have dimensions of p_1 and p_2 respectively, with $p_1 + p_2 = p$. We are essentially interested in estimation of $\boldsymbol{\beta}_1$ when it is plausible that $\boldsymbol{\beta}_2$ is close to the null vector.

In this situation we consider shrinkage estimation strategy for $\boldsymbol{\beta}_1$ based on kernel estimates of $g(.)$. We study the properties of these estimators using the notion of asymptotic distributional bias and risk. We also present an absolute penalty estimator, a modified version of lasso when errors are correlated. We provide simulation study for all the estimators to appraise their performances. A real data example is given to illustrate the methods.

3.1.1 Organization of the Chapter

Section 3.2 provides the proposed shrinkage estimators as well as an APE which is a modified version of lasso for correlated data. Section 3.3 and 3.4 provide asymptotic results of shrinkage and positive shrinkage estimators. In Section 3.5, we design and conduct a Monte Carlo experiment to study the performance of the proposed estimators and compare them with the APE. Section 3.6 provides a real data example. In Section 3.7, we present concluding thoughts.

3.2 Statistical Model and Estimation

Throughout this chapter we will assume that $\mathbf{1}_n = (1,\ldots,1)'$ is not in the space spanned by the column vectors of $\boldsymbol{X} = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)'$. As a result, according to Chen (1988) model (3.1) is identifiable. In addition, we assume the design points $\boldsymbol{x}_i$ and t_i are fixed for $i = 1,\ldots,n$, and errors ε_i have the form (3.2). Assume that $\{\boldsymbol{x}_i', t_i, y_i; i = 1,\ldots,n\}$ satisfy model (3.1). We first define a least square estimator (LSE) for the parameter vector $\boldsymbol{\beta}$ based on $g(\cdot)$ estimated using a general kernel smoothing. If $\boldsymbol{\beta}$ is known to be the true parameter, then by $E(\varepsilon_i) = 0$ we have $g(t_i) = E(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}), i = 1,\ldots,n$. Hence, a natural nonparametric estimator of $g(\cdot)$ given $\boldsymbol{\beta}$ is

$$\tilde{g}_n(t,\boldsymbol{\beta}) = \sum_{i=1}^{n} W_{ni}(t)(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}),$$

where $W_{ni}(t)$ are positive weight functions depending on $\boldsymbol{t}$ and will be defined in Assumption 3.3.3 and Remark 3 in Subsection 3.3. To estimate $\boldsymbol{\beta}$, we seek to minimize

$$SS(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \tilde{g}_n(t_i, \boldsymbol{\beta}))^2 = \sum_{i=1}^{n} (\hat{y}_i - \hat{\boldsymbol{x}}_i'\boldsymbol{\beta})^2. \quad (3.3)$$

The minimizer to (3.3) is found as $\hat{\boldsymbol{\beta}}_n = (\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1}\widehat{\boldsymbol{X}}'\hat{\boldsymbol{y}}$, where $\hat{y}_i = y_i - \sum_{j=1}^{n} W_{nj}(t_i) y_j$, $\hat{\boldsymbol{x}}_i = \boldsymbol{x}_i - \sum_{j=1}^{n} W_{nj}(t_i)\boldsymbol{x}_j$, for $i = 1, \ldots, n$, $\hat{\boldsymbol{y}} = (\hat{y}_1, \ldots, \hat{y}_n)'$, and $\widehat{\boldsymbol{X}} = (\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_n)'$. $\hat{\boldsymbol{\beta}}_n$ is called the LSE of $\boldsymbol{\beta}$.

Since the error is correlated, the LSE $\hat{\boldsymbol{\beta}}_n$ will not be asymptotically efficient. We use a weighted least squares estimator (WLSE) of $\boldsymbol{\beta}$ which is

$$\hat{\boldsymbol{\beta}}^{WLSE} = (\widehat{\boldsymbol{X}}'\Omega^{-1}(\theta)\widehat{\boldsymbol{X}})^{-1}\widehat{\boldsymbol{X}}'\Omega^{-1}(\theta)\hat{\boldsymbol{y}},$$

where

$$\Omega^{-1}(\theta) = \begin{pmatrix} 1 & -\theta & 0 & 0 & \ldots & 0 \\ -\theta & 1+\theta^2 & -\theta & & & \vdots \\ 0 & -\theta & 1+\theta^2 & -\theta & & \\ \vdots & & & & & \\ 0 & & \ldots & & -\theta & 1 \end{pmatrix}$$

and θ is unknown, as is often the case in practice, $\Omega(\theta)$ is replaced by $\Omega(\hat{\theta})$ where $\hat{\theta}$ is a suitable estimator of θ. Noting that ε_i is unobservable, a reasonable estimator of θ is the least squares estimator $\hat{\theta}_n$ based on the residuals $\hat{\varepsilon}_i = \hat{y}_i - \hat{\boldsymbol{x}}_i'\hat{\boldsymbol{\beta}}_n, \; i = 1, \ldots, n$, and is given by $\hat{\theta} = \sum_{i=2}^{n} \hat{\varepsilon}_i\hat{\varepsilon}_{i-1} / \sum_{i=2}^{n} \hat{\varepsilon}_{i-1}^2$. Consequently, the estimated WLSE of $\boldsymbol{\beta}$ can be written as

$$\hat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{X}}'\Omega^{-1}(\hat{\theta})\widehat{\boldsymbol{X}})^{-1}\widehat{\boldsymbol{X}}'\Omega^{-1}(\hat{\theta})\hat{\boldsymbol{y}}.$$

The properties of $\hat{\boldsymbol{\beta}}$ was investigated in You and Chen (2002).

Suppose that, $\boldsymbol{\beta}$ can be partitioned such that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$. Thus the model (3.1) may be written as follows:

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + g(\boldsymbol{t}) + \boldsymbol{\varepsilon}.$$

Without loss of generality, we suppose that some variable selection method or prior information suggests that $\boldsymbol{X}_2$ is relatively insignificant and thus is removed from the model (3.1). Then we obtain a candidate sub-model as

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + g(\boldsymbol{t}) + \boldsymbol{\varepsilon}. \tag{3.4}$$

This model is assumed to be a low dimensional working model. Thus, we study a working sub-model of a partially linear model determined by variable selection or UPI. This situation may arise when there is over-modeling and one wishes to cut down the irrelevant part from the model (3.1). Our goal is to construct an efficient estimation for the regression parameter $\boldsymbol{\beta}_1$.

3.2.1 Unrestricted and Restricted Estimators

Let $\hat{\boldsymbol{\beta}}_1$ be the unrestricted weighted least square estimator (UE) of $\boldsymbol{\beta}_1$. Using the same argument in Subsection 2.2.1 and the inverse matrix formula (2.4), $\hat{\boldsymbol{\beta}}_1$ will be in the form of

$$\hat{\boldsymbol{\beta}}_1 = (\widehat{\boldsymbol{X}}_1' M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\widehat{\boldsymbol{X}}_2} \widehat{\boldsymbol{X}}_1)^{-1} \widehat{\boldsymbol{X}}_1' M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\widehat{\boldsymbol{X}}_2} \hat{y},$$

where $\widehat{\boldsymbol{X}}_1$ is composed by the first p_1 column vectors of $\widehat{\boldsymbol{X}}$, $\widehat{\boldsymbol{X}}_2$ is composed by the last p_2 column vectors of $\widehat{\boldsymbol{X}}$ and

$$M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\widehat{\boldsymbol{X}}_2} = \boldsymbol{\Omega}^{-1}(\hat{\theta}) - \boldsymbol{\Omega}^{-1}(\hat{\theta})\widehat{\boldsymbol{X}}_2(\widehat{\boldsymbol{X}}_2'\boldsymbol{\Omega}^{-1}(\hat{\theta})\widehat{\boldsymbol{X}}_2)^{-1}\widehat{\boldsymbol{X}}_2'\boldsymbol{\Omega}^{-1}(\hat{\theta}).$$

Moreover, when $\boldsymbol{\beta}_2 = \mathbf{0}$, we have the restricted partially linear regression model given in (3.4).

Let $\tilde{\boldsymbol{\beta}}_1$ be the restricted weighted least square estimator (RE) of $\boldsymbol{\beta}_1$. Then $\tilde{\boldsymbol{\beta}}_1$ has the form

$$\tilde{\boldsymbol{\beta}}_1 = (\widehat{\boldsymbol{X}}_1'\boldsymbol{\Omega}^{-1}(\theta^*)\widehat{\boldsymbol{X}}_1)^{-1}\widehat{\boldsymbol{X}}_1'\boldsymbol{\Omega}^{-1}(\theta^*)\hat{\boldsymbol{y}},$$

where $\theta^* = \sum_{i=2}^n \varepsilon_i^*\varepsilon_{i-1}^* / \sum_{i=2}^n \varepsilon_{i-1}^{*2}$ with $\varepsilon_i^* = \hat{y}_i^* - \hat{\boldsymbol{x}}_i^{*'}\boldsymbol{\beta}_1^*$, $\boldsymbol{\beta}_1^* = (\widehat{\boldsymbol{X}}_1'\widehat{\boldsymbol{X}}_1)^{-1}\widehat{\boldsymbol{X}}_1\hat{\boldsymbol{y}}$, $\widehat{\boldsymbol{X}}_1 = (\hat{\boldsymbol{x}}_1^*, \ldots, \hat{\boldsymbol{x}}_n^*)'$, $\hat{\boldsymbol{x}}_i^* = (\hat{x}_{i1}, \ldots, \hat{x}_{ip_1})'$.

3.2.2 The Shrinkage Estimators

We define the shrinkage estimator (SE) $\hat{\boldsymbol{\beta}}_1^S$ as follows:

$$\hat{\boldsymbol{\beta}}_1^S = \tilde{\boldsymbol{\beta}}_1 + \{1 - c_{opt}T_n^{-1}\}(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1), \quad \text{where} \quad c_{opt} = p_2 - 2, \quad p_2 \geq 3$$

and

$$T_n = n\hat{\sigma}_n^{-2}\hat{\boldsymbol{\beta}}_2'\widehat{\boldsymbol{X}}_2'M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\widehat{\boldsymbol{X}}_1}\widehat{\boldsymbol{X}}_2\hat{\boldsymbol{\beta}}_2,$$

where $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n(\hat{y}_i - \hat{\boldsymbol{x}}_i'\hat{\boldsymbol{\beta}}_n)^2$ and $M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\widehat{\boldsymbol{X}}_1}$ has the same definition as $M_{\boldsymbol{\Omega}^{-1}(\hat{\theta})\widehat{\boldsymbol{X}}_2}$. Finally, we define the positive shrinkage estimator (PSE) $\hat{\boldsymbol{\beta}}_1^{S+}$. The PSE is defined as

$$\hat{\boldsymbol{\beta}}_1^{S+} = \tilde{\boldsymbol{\beta}}_1 + \{1 - c_{opt}T_n^{-1}\}^+(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1), \ \ p_2 \geq 3,$$

where $z^+ = max(0,z)$, or alternatively

$$\hat{\boldsymbol{\beta}}_1^{S+} = \hat{\boldsymbol{\beta}}_1^{S} - [1 - c_{opt} T_n^{-1}] I(T_n < c_{opt})(\hat{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_1), \;\; p_2 \geq 3.$$

3.2.3 Absolute Penalty Estimation

Although most penalized methods deal with a standard linear regression, there are some results for other models. Ahmed et al. (2007) extended the lasso to a partially linear model with independent errors by developing an absolute penalty estimator (APE) and compared its relative performance with the shrinkage estimators. Their APE coefficients are the solutions to the L_1 optimization problem

$$\hat{\boldsymbol{\beta}}_{lasso} = \underset{\boldsymbol{\beta}}{argmin} \Big\{ \sum_{i=1}^{n} (\hat{y}_i - \hat{\boldsymbol{x}}_i' \boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \Big\} \tag{3.5}$$

where $\hat{y}_i = y_i - \sum_{j=1}^{n} W_{nj}(t_i) y_j$, $\hat{\boldsymbol{x}}_i = \boldsymbol{x}_i - \sum_{j=1}^{n} W_{nj}(t_i) \boldsymbol{x}_j$, for $i = 1, \ldots, n$ and λ is the tuning parameter.

In this section we propose an APE for the partially linear model with RCAR(1) errors. In particular, this is an extension to (3.5). Since the errors are correlated, we transform the data and apply the L_1 penalty to the transformed data. We have $\boldsymbol{\Upsilon} = Var(\varepsilon) = \frac{(1-\theta^2)\sigma_e^2}{1-\theta^2-\sigma_z^2} \boldsymbol{\Omega}(\theta)$ and using Cholesky decomposition, we factor $\boldsymbol{\Upsilon} = \boldsymbol{A}\boldsymbol{A}'$ where $\boldsymbol{\Upsilon}$ and $\boldsymbol{A}$ are $n \times n$ matrices. Now, we follow the steps similar to those in Subsection 2.2.2 in Chapter 2, to achieve APE coefficients.

Step 1: Estimate $\hat{\boldsymbol{\beta}}_n = (\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1} \widehat{\boldsymbol{X}}' \hat{\boldsymbol{y}}$ defined in Section 3.2

Step 2: Calculate $\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = n^{-1} \sum_{i=1}^{n} (\hat{y}_i - \hat{\boldsymbol{x}}_i' \hat{\boldsymbol{\beta}}_n)^2$

Step 3: Calculate $\hat{\theta} = \sum_{i=2}^{n} \hat{\varepsilon}_i \hat{\varepsilon}_{i-1} / \sum_{i=2}^{n} \hat{\varepsilon}_{i-1}^2$

Step 4: Calculate $\hat{\Upsilon} = \hat{\sigma}_n^2\Omega(\hat{\theta})$ and then factor $\hat{\Upsilon} = CC'$ using Cholesky decomposition

Step 5: Transform $\boldsymbol{y}$ and $\boldsymbol{X}$ to $\hat{\boldsymbol{y}}^\star = \boldsymbol{C}^{-1}\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{X}}^\star = \boldsymbol{C}^{-1}\hat{\boldsymbol{X}}$

Step 6: Solve the APE problem as

$$\hat{\beta}_{lasso} = \underset{\boldsymbol{\beta}}{argmin}\left\{(\hat{\boldsymbol{y}}^\star - \hat{\boldsymbol{X}}^\star\boldsymbol{\beta})'(\hat{\boldsymbol{y}}^\star - \hat{\boldsymbol{X}}^\star\boldsymbol{\beta}) + \lambda\sum_{j=1}^{p}|\beta_j|\right\}. \tag{3.6}$$

In the next section we present the assumptions needed to provide the asymptotic results for shrinkage estimators and compare their performances based on their asymptotic distributional bias (ADB) and asymptotic distributional (ADR).

3.3 First-Order Asymptotics

We now present the following assumptions required to derive the asymptotic results in Section 3.4.

Assumption 3.3.1. There exist bounded functions $h_j(\cdot)$ over $[0,1]$ such that

$$x_{ij} = h_j(t_i) + u_{ij},\ i = 1,\ldots,n,\ j = 1,\ldots,p, \tag{3.7}$$

where $u_i = (u_{i1},\ldots,u_{ip})'$ are real vectors satisfying

$$\lim_{n\to\infty}\frac{\sum_{k=1}^{n-h} u_{ki}u_{k+|h|,j}}{n} = b_{hij},\ \text{ for } h = 0,\pm 1,\pm 2,\ldots,\ i = 1,\ldots,p,\ j = 1,\ldots,p, \tag{3.8}$$

and the matrix $B = (b_{0ij})$ is nonsingular. Moreover, for any permutation $(j_1,\ldots,j_n)$ of the integers $(1,\ldots,n)$,

$$\limsup_{n\to\infty}\frac{1}{\sqrt{n}\log n}\max_{1\le m\le n}\left\|\sum_{i=1}^{m}u_{j_i}\right\|<\infty, \tag{3.9}$$

where $||\cdot||$ denotes the Euclidean norm.

Assumption 3.3.2. The functions $g(\cdot)$ and $h_j(\cdot)$ satisfy the Lipschitz condition of order 1 on $[0,1]$.

Assumption 3.3.3. The probability weight functions $W_{ni}(\cdot)$ satisfy

(i) $\max_{1\le i\le n}\sum_{j=1}^{n}W_{ni}(t_j)=O(1)$,

(ii) $\max_{1\le i,j\le n}W_{ni}(t_j)=O(b_n)$,

(iii) $\max_{1\le j\le n}\sum_{i=1}^{n}W_{nk}(t_j)I(|t_j-t_i|>c_n)=O(d_n)$,

where $b_n=o[n^{-1/2}(\log n)^{-2}]$, c_n satisfies $\limsup_{n\to\infty}nc_n^4\log n<\infty$, d_n satisfies $\limsup_{n\to\infty}n$ $d_n^4\log n<\infty$, and $I(A)$ is the indicator function of a set A.

Assumption 3.3.4. $\{z_i\}$ is a Gaussian process and we further assume that

$$Ee_0^4<\infty,\;\;|\theta|<1,\;\;\theta^2+\sigma_z^2<1\quad\text{and}\quad\theta^4+6\theta^2\sigma_z^2+3\sigma_z^4<1.$$

Remark 1. In Assumption 3.3.1, the u_{ij} behave like zero mean, uncorrelated random variables, and the $h_j(t_i)$ is like the regression of x_{ij} on t_i. If the design points (x_i,t_i) are i.i.d. random variables, then by the law of large numbers and the law of iterated logarithm, (3.8) and (3.9) hold with probability 1 for $h_j(t_i)=E(x_{ij}|t_i)$, $u_{ij}=x_{ij}-h_j(t_i)$, and $B=E(u_iu_i')$. Conditions (3.8) and (3.9) are assumed in Härdle et al. (2000), and Gao (1995), among others.

Remark 2. Assumption 3.3.2 is very mild. The usual polynomial and trigonometric functions satisfy this assumption.

Remark 3. The following two weight functions satisfy Assumption 3.3.3:

$$W_{ni}^{(1)}(t) = \frac{1}{h_n}\int_{s_{i-1}}^{s_i} K\left(\frac{t-s}{h_n}\right) ds, \qquad W_{ni}^{(2)}(t) = K\left(\frac{t-t_i}{h_n}\right)\left[\sum_{j=1}^{n} K\left(\frac{t-t_j}{h_n}\right)\right]^{-1},$$

where $s_i = (t_i + t_{i-1})/2$, $i = 1, \ldots, n-1, s_0 = 0, s_n = 1$, $K(\cdot)$ is the Parzen-Rosenblatt kernel function, and h_n is a bandwidth parameter verifying suitable conditions.

Theorem 3.3.1. Suppose that Assumptions 3.3.1 to 3.3.4 hold. Then the limiting distribution of $\hat{\boldsymbol{\beta}}$ is given by

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N\left(0, \frac{(1-\theta^2)\sigma_e^2}{1-\theta^2-\sigma_z^2}\boldsymbol{G}^{-1}\right),$$

where $\xrightarrow{D}$ denotes convergence in distribution, $\boldsymbol{G} = \lim_{n\to\infty} n^{-1}\boldsymbol{U}'\Omega^{-1}(\theta)\boldsymbol{U}$ provided it exists and $\boldsymbol{U} = (u_1, \ldots, u_n)'$ is defined in Assumption 3.3.1.

Proof. The proof can be found in You and Chen (2002). □

Lemma 3.3.1. (i) Suppose that Assumptions 3.3.2 and 3.3.3 (iii) hold. Then as $n \to \infty$

$$\max_{0\le s\le p}\max_{1\le i\le n}\left|G_s(t_i) - \sum_{j=1}^{n} W_{nj}(t_i)G_s(t_j)\right| = O(c_n) + O(d_n),$$

where $G_0(\cdot) = g(\cdot)$ and $G_s(\cdot) = h_s(\cdot),\ s = 1, \ldots, p$.

(ii) Suppose that Assumptions 3.3.1 to 3.3.3 hold. Then as $n \to \infty$

$$\max_{1\le s\le p}\max_{1\le i\le n}\left|\hat{h}_{ns}(t_i) - h_s(t_i)\right| = O(c_n) + O(d_n) + O(a_n b_n),$$

where $\hat{h}_{ns}(t_i) = \sum_{j=1}^{n} W_{nj}(t_i)x_{js}$ and $a_n = \sqrt{n}\log n$.

Proof. The proof can be found in Gao (1995). □

Lemma 3.3.2. Suppose that Assumptions 3.3.1 to 3.3.3 hold. Then we have that

$$\lim_{n\to\infty} \frac{1}{n} \widehat{\boldsymbol{X}}' \Omega^{-1}(\hat{\theta}) \widehat{\boldsymbol{X}} = \boldsymbol{G}.$$

Proof. According to Lemma 2 of You and Chen (2002), we just need to show that

$$\frac{1}{n}(\widehat{\boldsymbol{X}}' \Omega^{-1}(\hat{\theta}) \widehat{\boldsymbol{X}} - \widehat{\boldsymbol{X}}' \Omega^{-1}(\theta) \widehat{\boldsymbol{X}}) = o_p(1). \tag{3.10}$$

Consider the (i, j) element of the term on the left-hand side of above equation given by

$$(\hat{\theta}^2 - \theta^2) \frac{\sum_{k=2}^{n-1} \hat{x}_{ki} \hat{x}_{kj}}{n} - 2(\hat{\theta} - \theta) \frac{\sum_{k=1}^{n-1} \hat{x}_{ki} \hat{x}_{k+1,j}}{n},$$

which implies (3.10), since $\hat{\theta} - \theta = O_p(n^{-\frac{1}{2}})$ by You and Chen (2002), and $n^{-1} \sum_{k=2}^{n-1} \hat{x}_{ki} \hat{x}_{kj}$ and $n^{-1} \sum_{k=1}^{n-1} \hat{x}_{ki} \hat{x}_{k+1,j}$ are bounded by Lemma 3.3.1 (i). □

Lemma 3.3.3. Suppose that Assumptions 3.3.1 to 3.3.4 hold. Then we have that

$$\hat{\sigma}_n^2 = (1 - \theta^2 - \sigma_z^2)^{-1} \{(1 - \theta^2) \sigma_e^2\} + O_p\left(n^{-\frac{1}{2}}\right), \;\; \tilde{\boldsymbol{\beta}}_1 = (\boldsymbol{I}, \boldsymbol{G}_{11}^{-1} \boldsymbol{G}_{12}) \hat{\boldsymbol{\beta}}_w + o_p\left(n^{-\frac{1}{2}}\right),$$

and

$$T_n = n(1 - \theta^2 - \sigma_z^2)^{-1} \{(1 - \theta^2) \sigma_e^2\} \hat{\boldsymbol{\beta}}_2' \boldsymbol{G}_{22.1} \hat{\boldsymbol{\beta}}_2 + o_p(1),$$

where $\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i^2 = n^{-1} \sum_{i=1}^n (\hat{y}_i - \hat{\boldsymbol{x}}_i' \hat{\boldsymbol{\beta}})^2$. Moreover, we have

$$\lim_{n\to\infty} P\{T_n \le x | K_n\} = H_{p_2}(x; \Delta) \;\; \text{where} \;\; \Delta = (\boldsymbol{\omega}' \boldsymbol{G}_{22.1} \boldsymbol{\omega})(1 - \theta^2 - \sigma_z^2) \{(1 - \theta^2) \sigma_e^2\}^{-1}.$$

Proof. According to Nicholls and Quinn (1982) we just need to show

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2-\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2=o_p(n^{-\frac{1}{2}}).$$

It is easy to see that

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2-\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2=\frac{1}{n}\sum_{i=1}^{n}(\hat{\varepsilon}_i-\varepsilon_i)^2+\frac{2}{n}\sum_{i=1}^{n}(\hat{\varepsilon}_i-\varepsilon_i)\varepsilon_i=I_1+I_2,\ \text{say.}$$

Since

$$\hat{\varepsilon}_i=\varepsilon_i-\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})-(\hat{g}(t_i)-g(t_i)),$$

where $\hat{g}(t_i)=\sum_{j=1}^{n}W_{nj}(t_i)(Y_j-\boldsymbol{x}'_j\hat{\boldsymbol{\beta}}_n)$, it holds that

$$\begin{aligned}I_2 &= \frac{1}{n}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_n)'\left\{\sum_{i=1}^{n}\left[\boldsymbol{x}_i-\sum_{j=1}^{n}W_{nj}(t_i)\boldsymbol{x}_j\right]\right\}\varepsilon_i\\ &\quad+\frac{1}{n}\sum_{i=1}^{n}\left[g(t_i)-\sum_{j=1}^{n}W_{nj}(t_i)g(t_j)\right]\varepsilon_i-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{nj}(t_i)\varepsilon_j\varepsilon_i=I_{21}+I_{22}-I_{23},\ \text{say.}\end{aligned}$$

By Assumption 3.3.1, I_{21} can be decomposed into a sum of two terms as follows,

$$\begin{aligned}I_{21} &= \frac{1}{n}\sum_{s=1}^{p}(\beta_s-\hat{\beta}_{ns})\sum_{i=1}^{n}\left[h_s(t_i)-\sum_{j=1}^{n}W_{nj}(t_i)x_{js}\right]\varepsilon_i+\frac{1}{n}\sum_{s=1}^{p}(\beta_s-\hat{\beta}_{ns})\sum_{i=1}^{n}u_{is}\varepsilon_i\\ &= I_{211}+I_{212},\ \text{say.}\end{aligned}$$

It follows from Lemma 3.3.1, Theorem 3.3.1, and the strict stationarity and ergodicity of

$\{|\varepsilon_i|\}$ that

$$|I_{211}| = O_p(n^{-\frac{1}{2}}) \cdot \max_{1\le s\le p} \max_{1\le k\le n} \left| h_s(t_k) - \sum_{j=1}^{n} W_{nj}(t_k) x_{js} \right| \cdot \frac{1}{n} \sum_{k=1}^{n-h} |\varepsilon_k|$$
$$= O_p(n^{-\frac{1}{2}}) \cdot [O(c_n) + O(d_n) + O(a_n b_n)] = o_p(n^{-\frac{1}{2}}).$$

According to Theorem 3.3.1, it follows that $I_{212} = O_p(n^{-1}) = o_p(n^{-\frac{1}{2}})$. Therefore, $I_{21} = o_p(n^{-\frac{1}{2}})$. Moreover, we have

$$EI_{22}^2 \le \frac{c}{n^2} \sum_{i=1}^{n} \left[g(t_i) - \sum_{j=1}^{n} W_{nj}(t_i) g(t_j) \right]^2 = n^{-1}[O(c_n) + O(d_n)]^2 = o(n^{-1}),$$

which implies that $I_{22} = o_p(n^{-\frac{1}{2}})$. On the other hand, by Assumptions 3.3.3 and 3.3.4, we have $I_{23} = o_p(n^{-\frac{1}{2}})$. Hence, $I_2 = o_p(n^{-\frac{1}{2}})$. By the same reason, it follows that $I_1 = o_p(n^{-\frac{1}{2}})$. The first result holds. Moreover, by combining Theorem 3.3.1 and Lemma 3.3.2, it is easy to prove the other results. According to Lemma 3.3.3, we demonstrate that $\sqrt{n}(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)$ and T_n are asymptotically independent under K_n. Therefore,

$\lim_{n\to\infty} P\left\{\sqrt{n}(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \le \boldsymbol{x}, T_n \le \chi^2_{p_2,\alpha} | K_n\right\} =$

$\Phi_{p_1}(\boldsymbol{x} + \boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}; 0, (1-\theta^2-\sigma_z^2)^{-1}\{(1-\theta^2)\sigma_e^2\}\boldsymbol{G}_{11}^{-1}) H_{p_2}(\chi^2_{p_2,\alpha}; \Delta)$.

Moreover, combining Lemmas 3.3.2 and 3.3.3, by Saleh (2006)

$\lim_{n\to\infty} P\left\{\sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \le \boldsymbol{x}, T_n \ge \chi^2_{p_2,\alpha} | K_n\right\} =$

$\int_{E(\boldsymbol{\omega})} \Phi_{p_1}(\boldsymbol{x} - \boldsymbol{D}_{12}^{-1}\boldsymbol{D}_{22}\boldsymbol{z}; 0, (1-\theta^2-\sigma_z^2)^{-1}\{(1-\theta^2)\sigma_e^2\}\boldsymbol{D}_{11.2}^{-1}) d\Phi_{p_2}(\boldsymbol{z}; 0, (1-\theta^2-\sigma_z^2)^{-1}$

$\cdot \{(1-\theta^2)\sigma_e^2\}\boldsymbol{D}_{22})$,

where $E(\boldsymbol{\omega}) = \{\boldsymbol{z} : (1-\theta^2-\sigma_z^2)\{(1-\theta^2)\sigma_e^2\}^{-1}(\boldsymbol{z}+\boldsymbol{\omega})'\boldsymbol{G}_{22.1}(\boldsymbol{z}+\boldsymbol{\omega}) \ge \chi^2_{p_2,\alpha}\}$ and $\boldsymbol{D} = \boldsymbol{G}^{-1} = \begin{pmatrix} \boldsymbol{D}_{11} & \boldsymbol{D}_{12} \\ \boldsymbol{D}_{21} & \boldsymbol{D}_{22} \end{pmatrix}$.

Therefore under $\{K_n\}$, $\sqrt{n}(\hat{\boldsymbol{\beta}}_1^S - \boldsymbol{\beta}_1)$ has the same distribution as

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_1^S - \boldsymbol{\beta}_1) \to_D \boldsymbol{D}_1\boldsymbol{U} + (p_2-2)\frac{(1-\theta^2-\sigma_z^2)^{-1}\{(1-\theta^2)\sigma_e^2\}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}(\boldsymbol{D}_1\boldsymbol{U}+\boldsymbol{\omega})}{(\boldsymbol{D}_2\boldsymbol{U}+\boldsymbol{\omega})'\boldsymbol{G}_{22.1}(\boldsymbol{D}_2\boldsymbol{U}+\boldsymbol{\omega})},$$

when n tends to infinity, where $\boldsymbol{U} \sim N_p(\mathbf{0}, (1-\theta^2-\sigma_z^2)^{-1}\{(1-\theta^2)\sigma_e^2\}\boldsymbol{G}^{-1})$. Now, combining Lemma 3.3.3, the proofs of Theorems 3.4.1 and 3.4.2 in the following section, follow by direct computation and definitions of the estimators. □

3.4 Asymptotic Properties of the Estimators

In this section, we provide the asymptotic distributional bias (ADB) and asymptotic distributional risk (ADR) of the UE, RE, SE and PSE. Since the main concern is the performance of these four estimators when $\boldsymbol{\beta}_2$ is close to the null vector, namely $\boldsymbol{\beta}_2 = 0$, we consider a sequence of local alternatives

$$K_n : \boldsymbol{\beta}_{2(n)} = n^{-\frac{1}{2}}\boldsymbol{\omega}, \quad \boldsymbol{\omega} \neq 0 \text{ fixed}$$

to establish the asymptotic properties of these estimators. Consider the following loss function

$$\mathcal{L}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_1) = n(\boldsymbol{\beta}_1^0 - \boldsymbol{\beta}_1)'\boldsymbol{M}(\boldsymbol{\beta}_1^0 - \boldsymbol{\beta}_1),$$

where $\boldsymbol{\beta}_1^0$ is any one of $\hat{\boldsymbol{\beta}}_1$, $\tilde{\boldsymbol{\beta}}_1$, $\hat{\boldsymbol{\beta}}_1^S$ and $\hat{\boldsymbol{\beta}}_1^{S+}$. The cumulative distribution function of $\boldsymbol{\beta}_1^0$ under $\{K_n\}$ can be denoted as

$$F(\boldsymbol{x}) = \lim_{n\to\infty} P\{\sqrt{n}(\boldsymbol{\beta}_1^0 - \boldsymbol{\beta}_1) \leq \boldsymbol{x} | K_n\},$$

where $F(\boldsymbol{x})$ is nondegenerate. Then, the ADR of $\boldsymbol{\beta}_1^0$ is defined as

$$R(\boldsymbol{\beta}_1^0, \boldsymbol{M}) = \text{tr}\left\{\boldsymbol{M}\int_{\mathcal{R}^{p_1}}\int \boldsymbol{x}\boldsymbol{x}' dF(\boldsymbol{x})\right\} = \text{tr}(\boldsymbol{M}\boldsymbol{\Gamma}),$$

where $\boldsymbol{\Gamma}$ is the dispersion matrix obtained from $F(\boldsymbol{x})$.

First we present the expression for the asymptotic distribution bias (ADB) of the proposed estimators. The ADB of an estimator $\boldsymbol{\beta}_1^0$ is defined as

$$\text{ADB}(\boldsymbol{\beta}_1^0) = \lim_{n\to\infty} E\left\{n^{\frac{1}{2}}(\boldsymbol{\beta}_1^0 - \boldsymbol{\beta}_1)\right\}.$$

In the following we present the ADB and ADR of the estimators. The result of Theorem 3.3.1 along with Lemma 2.3.4 and 2.3.5 in Chapter 2, will be used to derive the asymptotic results of the estimators under K_n.

3.4.1 Asymptotic Distributional Bias (ADB)

Theorem 3.4.1. Suppose that Assumptions 3.3.1 to 3.3.4 hold. Then under $\{K_n\}$, as $n \to \infty$, the ADB of the estimators $\hat{\boldsymbol{\beta}}_1$, $\tilde{\boldsymbol{\beta}}_1$, $\hat{\boldsymbol{\beta}}_1^S$, and $\hat{\boldsymbol{\beta}}_1^{S+}$ are respectively

$$\begin{aligned}
\text{ADB}(\hat{\boldsymbol{\beta}}_1) &= \mathbf{0}, \\
\text{ADB}(\tilde{\boldsymbol{\beta}}_1) &= -\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}, \\
\text{ADB}(\hat{\boldsymbol{\beta}}_1^S) &= -(p_2-2)\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}E(\chi_{p_2+2}^{-2}(\Delta)), \\
\text{ADB}(\hat{\boldsymbol{\beta}}_1^{S+}) &= \text{ADB}(\hat{\boldsymbol{\beta}}_1^S) - \boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}\{H_{p_2+2}(p_2-2;\Delta)\} \\
&+ \boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}\{(p_2-2)E(\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) < p_2-2))\},
\end{aligned}$$

where $\boldsymbol{G} = \begin{pmatrix} \boldsymbol{G}_{11} & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{pmatrix}$ is defined in Theorem 3.3.1, $\Delta = (\boldsymbol{\omega}'\boldsymbol{G}_{22.1}\boldsymbol{\omega})(1-\theta^2-\sigma_z^2)\{(1-\theta^2)\sigma_e^2\}^{-1}$, $\boldsymbol{G}_{22.1} = \boldsymbol{G}_{22} - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}$, and $H_v(x;\Delta)$ denotes the noncentral chi-square distribution function with noncentrality parameter Δ and v degrees of freedom and

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

Proof. See Section 3.3. □

Since the bias expression of all the estimators are not in the scalar form, we therefore take the recourse by converting them into the quadratic form. Thus, let us define the asymptotic quadratic distributional bias (AQDB) of an estimator $\boldsymbol{\beta}_1^0$ of $\boldsymbol{\beta}_1$ by

$$\text{AQDB}(\boldsymbol{\beta}_1^0) = [\text{ADB}(\boldsymbol{\beta}_1^0)]'\boldsymbol{G}_{11.2}[\text{ADB}(\boldsymbol{\beta}_1^0)],$$

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

Corollary 3.4.1. Suppose that conditions in Theorem 3.4.1 hold. Then under $\{K_n\}$, as $n \to \infty$, the AQDB of the estimators are given as follows:

$$\begin{aligned}
\text{AQDB}(\hat{\boldsymbol{\beta}}_1) &= \mathbf{0}, \\
\text{AQDB}(\tilde{\boldsymbol{\beta}}_1) &= \boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{11.2}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}, \\
\text{AQDB}(\hat{\boldsymbol{\beta}}_1^S) &= (p_2-2)^2\boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{11.2}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}\left[E(\chi_{p_2,\alpha}^{-2};\Delta)\right]^2, \\
\text{AQDB}(\hat{\boldsymbol{\beta}}_1^{S+}) &= \boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{11.2}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}\cdot\Big\{H_{p_2+2}(p_2-2;\Delta) \\
&+ (p_2-2)^2E(\chi_{p_2+2}^{-2}(\Delta)) - (p_2-2)^2E(\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^2(\Delta) < p_2-2))\Big\}^2.
\end{aligned}$$

Proof. The expressions for quadratic biases are obtained by following the definition of AQDB. □

3.4.2 Asymptotic Distributional Risk (ADR)

Theorem 3.4.2. Suppose that conditions in Theorem 3.4.1 hold. Then under $\{K_n\}$, as $n \to \infty$, the asymptotic covariance matrices of the estimators are as follows:

$$
\begin{aligned}
\Gamma(\hat{\boldsymbol{\beta}}_1) &= \sigma^* \boldsymbol{G}_{11.2}^{-1}, \quad \text{where} \quad \sigma^* = (1-\theta^2-\sigma_z^2)^{-1}\{(1-\theta^2)\sigma_e^2\}, \\
\Gamma(\tilde{\boldsymbol{\beta}}_1) &= \sigma^* \boldsymbol{G}_{11.2}^{-1} - \sigma^*(\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1}\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}) + \boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}\boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}, \\
\Gamma(\hat{\boldsymbol{\beta}}_1^S) &= \sigma^* \boldsymbol{G}_{11.2}^{-1} \\
&- \sigma^*(p_2-2)\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1}\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\Big\{2E(\chi_{p_2+2,\alpha}^{-2}(\Delta)) \\
&- (p_2-2)E(\chi_{p_2+2}^{-4}(\Delta))\Big\} + (p_2-2)\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}\boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\Big\{2E(\chi_{p_2+2}^{-2}(\Delta)) \\
&+ (p_2-2)E(\chi_{p_2+4}^{-4}(\Delta)) - 2E(\chi_{p_2+4}^{-2}(\Delta))\Big\}, \\
\Gamma(\hat{\boldsymbol{\beta}}_1^{S+}) &= \Gamma(\hat{\boldsymbol{\beta}}_1^S) \\
&+ \sigma^*\Big\{(p_2-2)(\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1}\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1})\{2E\left[\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\right] \\
&- (p_2-2)E\left[\chi_{p_2+2}^{-4}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\right] \\
&- \boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1}\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}H_{p_2+2}(p_2-2;\Delta)\Big\} \\
&+ \boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}\boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\Big[2H_{p_2+2}(p_2-2;\Delta) - H_{p_2+4}(p_2-2;\Delta)\Big] \\
&- (p_2-2)\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}\boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\Big\{2E\left[\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\right] \\
&- 2E\left[\chi_{p_2+4}^{-2}(\Delta)I(\chi_{p_2+4}^{2}(\Delta) \le p_2-2)\right] \\
&+ (p_2-2)E\left[\chi_{p_2+4}^{-4}(\Delta)I(\chi_{p_2+4}^{2}(\Delta) \le p_2-2)\right]\Big\}.
\end{aligned}
$$

Proof. See Section 3.3. □

The ADR expressions for the estimators are contained in the following corollary.

Corollary 3.4.2. Suppose that conditions in Theorem 3.4.1 hold. Then under $\{K_n\}$, as $n \to \infty$, the ADR of the estimators $\hat{\boldsymbol{\beta}}_1$, $\tilde{\boldsymbol{\beta}}_1$, $\hat{\boldsymbol{\beta}}_1^S$ and $\hat{\boldsymbol{\beta}}_1^{S+}$ are respectively

$$
\begin{aligned}
ADR(\hat{\boldsymbol{\beta}}_1;\boldsymbol{M}) &= \sigma^* \operatorname{tr}(\boldsymbol{M}\boldsymbol{G}_{11.2}^{-1}),\\
ADR(\tilde{\boldsymbol{\beta}}_1;\boldsymbol{M}) &= \sigma^* \operatorname{tr}(\boldsymbol{M}\boldsymbol{G}_{11.2}^{-1} - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{M}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1}) + \boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{M}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega},\\
ADR(\hat{\boldsymbol{\beta}}_{1G}^S;\boldsymbol{M}) &= \sigma^* \operatorname{tr}(\boldsymbol{M}\boldsymbol{G}_{11.2}^{-1})\\
&- \sigma^*\Big[(p_2-2)\operatorname{tr}(\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{M}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1})\big\{2E(\chi_{p_2+2,\alpha}^{-2}(\Delta))\\
&- (p_2-2)E(\chi_{p_2+2}^{-4}(\Delta))\big\}\Big] + (p_2-2)\boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{M}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}\\
&\times \Big\{2E(\chi_{p_2+2}^{-2}(\Delta))(p_2-2)E(\chi_{p_2+4}^{-4}(\Delta)) - 2E(\chi_{p_2+4}^{-2}(\Delta))\Big\},\\
ADR(\hat{\boldsymbol{\beta}}_1^{S+};\boldsymbol{M}) &= ADR(\hat{\boldsymbol{\beta}}_1^S;\boldsymbol{M})\\
&+ \sigma^*\Big\{(p_2-2)\operatorname{tr}(\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{M}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1})\big\{2E\big[\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\big]\\
&- (p_2-2)E\big[\chi_{p_2+2}^{-4}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\big]\\
&- \operatorname{tr}(\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{M}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1})H_{p_2+2}(p_2-2;\Delta)\Big\}\\
&+ \boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{M}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}\Big[2H_{p_2+2}(p_2-2;\Delta) - H_{p_2+4}(p_2-2;\Delta)\Big]\\
&- (p_2-2)\boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{M}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}\Big\{2E\big[\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^{2}(\Delta) \le p_2-2)\big]\\
&- 2E\big[\chi_{p_2+4}^{-2}(\Delta)I(\chi_{p_2+4}^{2}(\Delta) \le p_2-2)\big]\\
&+ (p_2-2)E\big[\chi_{p_2+4}^{-4}(\Delta)I(\chi_{p_2+4}^{2}(\Delta) \le p_2-2)\big]\Big\}.
\end{aligned}
$$

Proof. The expressions for risk are obtained by following the definition of ADR. □

3.4.3 Bias and Risk Comparisons

In this section, we compare the ADB and ADR of the proposed estimators. Using the results of Theorem 3.4.1 and 3.4.2, we will have all the AQDB reduced to null vector and all the ADR reduced to $(1-\theta^2-\sigma_z^2)^{-1}\{(1-\theta^2)\sigma_e^2\}\mathrm{tr}(\boldsymbol{M}\boldsymbol{G}_{11}^{-1})$ for all $\boldsymbol{\omega}$, when $\boldsymbol{G}_{12}=0$. Therefore, we assume that $\boldsymbol{G}_{12}\neq 0$. Based on Theorem 3.4.2, the AQDB of $\tilde{\boldsymbol{\beta}}_1$ is an unbounded function of $\boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{11.2}\ \boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}$. In order to investigate the AQDB($\hat{\boldsymbol{\beta}}_1^S$) and AQDB($\hat{\boldsymbol{\beta}}_1^{S+}$) we will have

$$\begin{aligned}\mathrm{ch}_{\min}(\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{11.2}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1}) &\leq \frac{\boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{11.2}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}}{\boldsymbol{\omega}'\boldsymbol{G}_{22.1}\boldsymbol{\omega}}\\ &\leq \mathrm{ch}_{\max}(\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{11.2}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1}).\end{aligned}$$

Therefore, AQDB($\hat{\boldsymbol{\beta}}_1^S$) starts from zero at $\boldsymbol{\omega}'\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{11.2}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{\omega}=0$, increases to a point, then decreases towards zero due to $E(\chi_{p_2+2}^{-2}(\Delta))$ being a decreasing log-convex function of Δ. The behavior of $\hat{\boldsymbol{\beta}}_1^{S+}$ is similar to $\hat{\boldsymbol{\beta}}_1^S$. However, the quadratic bias curve of $\hat{\boldsymbol{\beta}}_1^{S+}$ remains below the curve of $\hat{\boldsymbol{\beta}}_1^S$ for all values of Δ.

Consider the ADR($\hat{\boldsymbol{\beta}}_1^S$) and ADR($\hat{\boldsymbol{\beta}}_1$). We can see if $M\in M^D$, $\hat{\boldsymbol{\beta}}_1^S$ dominates $\hat{\boldsymbol{\beta}}_1$ for any $\boldsymbol{\omega}$ in the sense of ADR where

$$\boldsymbol{M}^D=\left\{\boldsymbol{M}:\frac{\mathrm{tr}(\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{M}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1})}{\mathrm{ch}_{\max}(\boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{M}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}\boldsymbol{G}_{22.1}^{-1})}\geq\frac{p_2+2}{2}\right\}.$$

Moreover, by comparing ADR($\hat{\boldsymbol{\beta}}_1^{S+}$) with ADR($\hat{\boldsymbol{\beta}}_1^S$) we observe $\hat{\boldsymbol{\beta}}_1^{S+}$ dominates $\hat{\boldsymbol{\beta}}_1^S$ for all the values of $\boldsymbol{\omega}$, with strict inequality holds for some $\boldsymbol{\omega}$ and the largest risk improvement of $\hat{\boldsymbol{\beta}}_1^{S+}$ over $\hat{\boldsymbol{\beta}}_1^S$ is near the null hypothesis. Thus, the $ADR(\hat{\boldsymbol{\beta}}_1^{S+})\leq ADR(\hat{\boldsymbol{\beta}}_1)$ in the entire parameter space, and the upper limit is attained when Δ approaches ∞. We can see that, ADR($\hat{\boldsymbol{\beta}}_1^{S+}$) increases monotonically towards ADR($\hat{\boldsymbol{\beta}}_1$) from below, as Δ moves away from

0. This implies that

$$\text{ADR}(\hat{\boldsymbol{\beta}}_1^{S+}) \leq \text{ADR}(\hat{\boldsymbol{\beta}}_1^{S}) \leq \text{ADR}(\hat{\boldsymbol{\beta}}_1) \text{ for any } \boldsymbol{M} \in \boldsymbol{M}^D \text{ and } \boldsymbol{\omega}$$

with strict inequality holds for some $\boldsymbol{\omega}$. Finally, we conclude that $\hat{\boldsymbol{\beta}}_1^{S}$ and $\hat{\boldsymbol{\beta}}_1^{S+}$ outperform the $\hat{\boldsymbol{\beta}}_1$ in the entire parameter space induced by Δ. The gain in risk over $\hat{\boldsymbol{\beta}}_1$ is substantial when $\Delta = 0$ or near.

3.5 Monte Carlo Simulation studies

In this section, we carry out a Monte Carlo simulation study to examine the performance of the proposed estimators. In this study, we use the model (3.1) in which

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

where $\varepsilon_i = (\theta + z_i)\varepsilon_{i-1} + e_i$, $\theta = 0.1$ and z_i and e_i being i.i.d N(0,1). Furthermore, $g(t_i) = \sin(2\pi t_i)$, t_i are i.i.d U(0,1) and $x_{si} = (\zeta_{si}^{(1)})^2 + \zeta_i^{(2)}$ with $\zeta_{si}^{(1)}$ and $\zeta_i^{(2)}$ being i.i.d U(0,1) and N(0,1), respectively, for all $s = 1, \ldots, p$ and $i = 1, \ldots, n$. Our sampling experiment consists of various combinations of sample sizes, i.e., $n = 50$ and 100. For each n, we generate 5000 samples using the above model, also the x_i and t_i values are generated once for each n. We set $\beta_j = 0$, for $j = p_1 + 1, \ldots, p$ with $p = p_1 + p_2$. We set the regression coefficients $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ to $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \mathbf{0}')'$ for different cases.

Case 1: $\boldsymbol{\beta}_1 = (1.5, 1.7, 1.1)'$

Case 2: $\boldsymbol{\beta}_1 = (1.5, 1.7, 1.1, 0.8, 0.2)'$

Case 3: $\boldsymbol{\beta}_1 = (1.5, 1.7, 1.1, 0.8, 0.2, 2.5, 3)'$.

Our aim is to estimate $\boldsymbol{\beta}_1$ based on proposed strategies when $\boldsymbol{\beta}_2$ does not provide useful information ($\boldsymbol{\beta}_2 = \mathbf{0}$). For each case we provide detailed results for $p_2 = 3, 4, \ldots, 8$. The estimator in Priestley and Chao (1972) with a gaussian kernel is used for the weight function $W_{ni}(t_j)$

$$W_{ni}(t_j) = \frac{1}{nh_n} K(\frac{t_i - t_j}{h_n}) = \frac{1}{nh_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t_i - t_j)^2}{2h_n^2}}.$$

The cross-validation (CV) method (Bowman and Azzalini, 1997) is used to select the optimal bandwidth h_n, which minimizes the CV function:

$$CV(h_n) = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}^{-i} - \hat{\beta}_{1n}^{-i} \hat{x}_1^{-i} - \hat{\beta}_{2n}^{-i} \hat{x}_2^{-i} - \ldots - \hat{\beta}_{pn}^{-i} \hat{x}_p^{-i})^2,$$

where $(\hat{\beta}_{1n}^{-i}, \hat{\beta}_{2n}^{-i}, \ldots, \hat{\beta}_{pn}^{-i})' = (\widehat{\boldsymbol{X}}'^{-i} \widehat{\boldsymbol{X}}^{-i})^{-1} \boldsymbol{X}'^{-i} \hat{\boldsymbol{y}}'^{-i}$, $\widehat{\boldsymbol{X}}^{-i} = (\hat{x}_{sk}^{-i})', 1 \leq k \leq n, 1 \leq s \leq p, \hat{\boldsymbol{y}}^{-i} = (\hat{y}_1^{-i}, \hat{y}_2^{-i}, \ldots, \hat{y}_n^{-i}), \hat{x}_{sk}^{-i} = x_{sk} - \sum_{j \neq i}^{n} W_{nj}(t_i) x_{sj}, \hat{y}_k^{-i} = y_k - \sum_{j \neq i}^{n} W_{nj}(t_i) y_j$. Noting that $\hat{y}^{-i}$ is the predicted value of $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ at $\boldsymbol{x} = (x_{1i}, x_{2i}, \ldots, x_{pi})$ with y_i and x_i left out of the estimation of the βs.

We define the parameter $\Delta^* = \|\boldsymbol{\beta} - \boldsymbol{\beta}^*\|^2$, where $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1', \mathbf{0}')'$ and $\| \,.\, \|$ is the Euclidian norm. In order to produce values of Δ^* between 0 to 2, different values of $\boldsymbol{\beta}_2$ were chosen. The criterion for comparing the performance of the estimators of $\boldsymbol{\beta}_1$ is based on the mean squared error (MSE). The relative MSE of $\tilde{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_1^S$ and $\hat{\boldsymbol{\beta}}_1^{S+}$ have been numerically calculated with respect to $\hat{\boldsymbol{\beta}}_1$. The relative mean squared error (RMSE) of the other estimators to the unrestricted estimator $\hat{\boldsymbol{\beta}}_1$ is defined by:

$$RMSE(\hat{\boldsymbol{\beta}}_1 : \hat{\boldsymbol{\beta}}_1^\star) = \frac{MSE(\hat{\boldsymbol{\beta}}_1)}{MSE(\hat{\boldsymbol{\beta}}_1^\star)},$$

where $\hat{\boldsymbol{\beta}}_1^\star$ can be any of $\tilde{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_1^S$ and $\hat{\boldsymbol{\beta}}_1^{S+}$. Comparative RMSEs for RE, UE, SE and PSE are illustrated in Figures 3.1-3.3 to portray the relative performance of the estimators. The horizontal line of RMSE=1 facilitates a comparison among the other estimators. It is obvious that a RMSE larger than one indicates the degree of superiority of the estimator $\hat{\beta}_1^{\star}$ over $\hat{\beta}_1$.

We summarize the findings of our simulation study as follows:

(i) Simulation studies show that in all the cases the maximum efficiency of all the other estimators relative to UE occurred when $\Delta^* = 0$. The RE dominates all the estimators when the UPI is true ($\Delta^* = 0$). On the contrary, the risk of RE explodes as Δ^* increases, i.e., as the restriction moves away from $\Delta^* = 0$, the risk of RE goes below the horizontal line and becomes an inefficient estimator.

(ii) In all combinations of p_2, p_1 and n, departure from the restriction has less impact on shrinkage estimators risks which is consistent with the theory and their RMSE approaches one as we move away from the restriction.

(iii) The SE in all cases is dominated by the PSE (see the RMSE curve of SE which is lower than the PSE curve). This indicates that in the event of imprecise UPI (i.e., when $\beta_2 \neq \mathbf{0}$), the PSE has the smallest risk among the other estimators which makes it an ideal choice for real-life problems.

In summary, the simulation results are in agreement with our asymptotic results.

Comparison of Shrinkage with Absolute Penalty Estimator

Here, we compare the performance of shrinkage estimators with an APE (lasso). We used the 10-fold cross validation method to estimate the tuning parameter λ in (3.6) to compute APE.

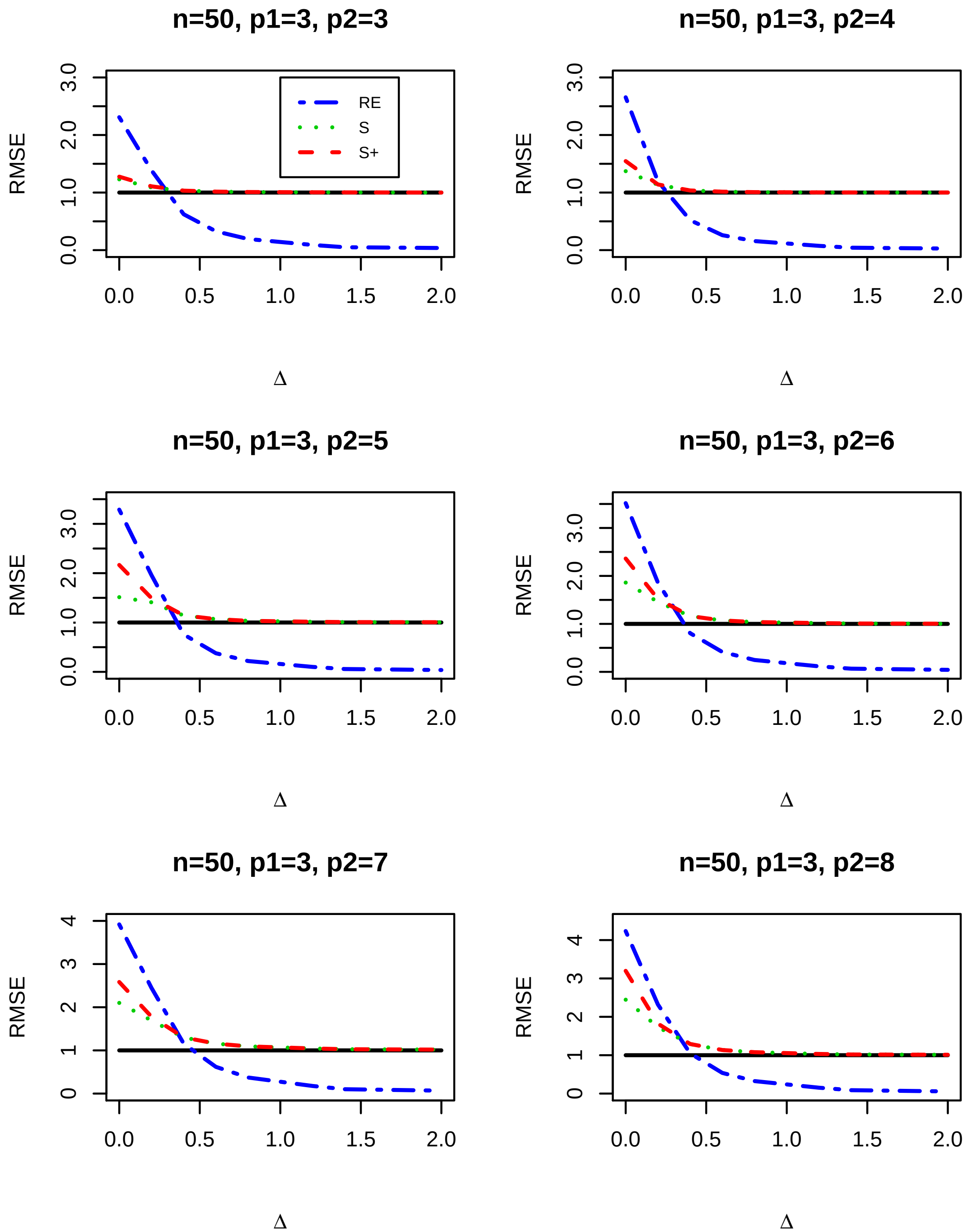

Figure 3.1: Relative MSE of the estimators for various p_2 when $p_1 = 3$ and $n = 50$. "- - -" denotes the PSE, "$\cdots$" denotes the SE, "– · – ·" denotes the RE, and "—" denotes the UE.

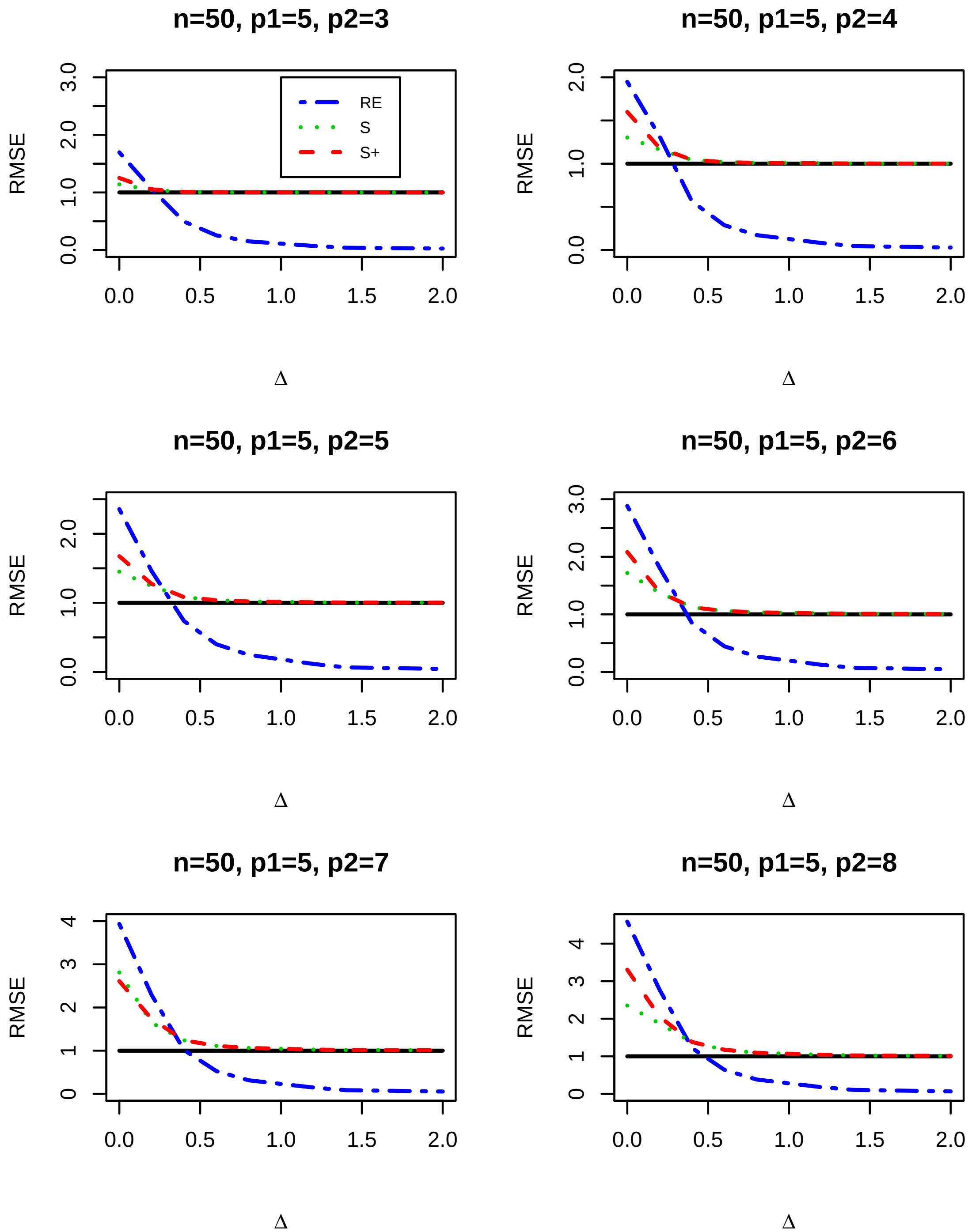

Figure 3.2: Relative MSE of the estimators for various p_2 when $p_1 = 5$ and $n = 50$. "- - -" denotes the PSE, "$\cdots$" denotes the SE, "– · – ·" denotes the RE, and "—" denotes the UE.

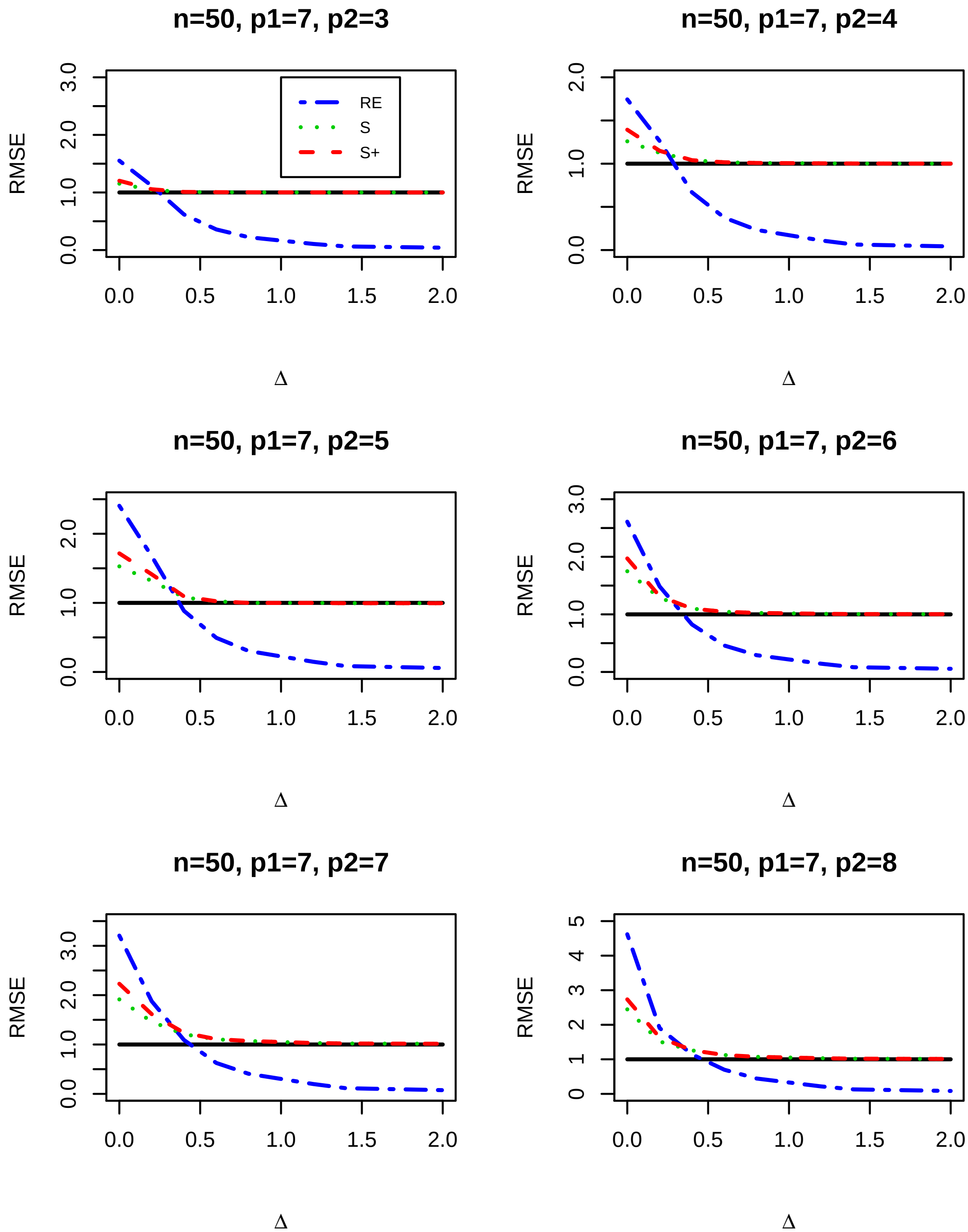

Figure 3.3: Relative MSE of the estimators for various p_2 when $p_1 = 7$ and $n = 50$. "- - -" denotes the PSE, "$\cdots$" denotes the SE, "– · – ·" denotes the RE, and "—" denotes the UE.

Table 3.1: Relative MSE of estimators with respect to UE when $n = 50$ and $\boldsymbol{\beta}_1 = (1.5, 1.7, 1.1)$

p_2	lasso	S	S+	RE
3	1.373	1.230	1.276	2.309
4	1.612	1.372	1.546	2.665
5	1.779	1.514	2.166	3.288
6	1.835	1.861	2.362	3.517
7	1.971	2.100	2.584	3.919
8	2.368	2.446	3.198	4.235
9	2.626	2.854	3.577	5.550
10	3.284	3.468	4.175	6.286

The criterion RMSE was used to compare the performance of the estimators

$$RMSE(\hat{\boldsymbol{\beta}}_1 : \hat{\boldsymbol{\beta}}_1^{\star}) = \frac{MSE(\hat{\boldsymbol{\beta}}_1)}{MSE(\hat{\boldsymbol{\beta}}_1^{\star})},$$

where $\hat{\boldsymbol{\beta}}_1^{\star}$ is one the $\tilde{\boldsymbol{\beta}}_1$, $\hat{\boldsymbol{\beta}}_1^{S}$, $\hat{\boldsymbol{\beta}}_1^{S+}$ and $\hat{\boldsymbol{\beta}}_1^{lasso}$. We performed the simulation for all 3 cases for $n = 50$ and 100 with $p_2 =, 4, \cdots, 10$. The simulated risk of the APE and shrinkage estimators when $\Delta^* = 0$ are shown in Tables 3.1 - 3.6.

In Table 3.1 and 3.2 when $p_1 = 3$ and $n = 50$ and 100, the APE outperforms PSE and SE when $p_2 = 5$ and $p_2 = 6$, respectively. On the other hand when $p_2 > 5$ the PSE dominates the APE and when $p_2 > 6$ the SE outperforms the APE. In Table 3.3 and 3.4 when $p_1 = 5$, we observe that again SE and PSE have less simulated risk than APE when $p_2 = 7$ and $p_2 = 8$, respectively. Also when $p_1 = 7$ (Table 3.5 and 3.6), SE and PSE have less risk than APE when $p_2 = 8$ and $p_2 = 9$, respectively. Thus as p_2 increases the shrinkage estimators start to perform better than the APE and we recommend shrinkage method for the large p_2. Not surprisingly, the RE is the best estimator when $\Delta^* = 0$, compared with all other estimators.

Table 3.2: Relative MSE of estimators with respect to UE when $n = 100$ and $\boldsymbol{\beta}_1 = (1.5, 1.7, 1.1)$

p_2	lasso	S	S+	RE
3	1.550	1.268	1.356	1.953
4	1.742	1.488	1.677	2.344
5	1.944	1.803	2.008	3.875
6	2.129	2.153	2.337	3.409
7	2.242	2.315	2.725	3.712
8	2.509	2.538	3.016	4.211
9	2.920	2.956	3.545	4.962
10	3.260	3.401	4.053	5.567

Table 3.3: Relative MSE of estimators with respect to UE when $n = 50$ and $\boldsymbol{\beta}_1 = (1.5, 1.7, 1.1, 0.8, 0.2)$

p_2	lasso	S	S+	RE
3	1.341	1.140	1.251	1.696
4	1.544	1.303	1.459	1.947
5	1.908	1.452	1.675	2.356
6	2.139	1.719	2.083	2.883
7	2.340	2.104	2.510	3.933
8	2.525	2.862	3.306	4.586
9	2.835	3.485	3.628	5.823
10	3.482	3.864	4.036	7.253

Table 3.4: Relative MSE of estimators with respect to UE when $n = 100$ and $\boldsymbol{\beta}_1 = (1.5, 1.7, 1.1, 0.8, 0.2)$

p_2	lasso	S	S+	RE
3	1.437	1.148	1.202	1.541
4	1.691	1.274	1.394	1.789
5	1.954	1.589	1.826	2.605
6	2.261	1.899	2.134	3.190
7	2.402	2.282	2.591	3.384
8	2.611	2.648	2.855	3.841
9	2.909	3.036	3.352	4.538
10	3.157	3.388	3.945	5.612

Table 3.5: Relative MSE of estimators with respect to UE when $n = 50$ and $\boldsymbol{\beta}_1 = (1.5, 1.7, 1.1, 0.8, 0.2, 2.5, 3)$

p_2	lasso	S	S+	RE
3	1.343	1.153	1.203	1.552
4	1.551	1.259	1.393	1.744
5	1.764	1.528	1.716	2.405
6	2.103	1.749	1.971	2.609
7	2.397	1.914	2.230	3.206
8	2.652	2.446	2.734	4.619
9	2.954	2.974	3.187	5.158
10	3.488	3.501	3.753	6.486

Table 3.6: Relative MSE of estimators with respect to UE when $n = 100$ and $\boldsymbol{\beta}_1 = (1.5, 1.7, 1.1, 0.8, 0.2, 2.5, 3)$

p_2	lasso	S	S+	RE
3	1.414	1.106	1.216	1.591
4	1.686	1.228	1.341	1.643
5	1.818	1.409	1.553	1.984
6	1.923	1.518	1.676	2.058
7	2.154	1.762	1.866	2.684
8	2.296	1.936	2.343	2.919
9	2.595	2.622	2.943	4.562
10	3.012	3.342	3.588	5.523

3.6 Data Example

We now implement suggested strategies to quarterly macroeconomic time series data (United Kingdom, 1948-1956). The data can be found in Reinsel and Velu (1998, p. 233) and they were initially analyzed by Klein et al. (1961). In this data set, we consider the dependent variable y_i as the total exports and the explanatory variables x_{i1} , x_{i2} , x_{i3} , x_{i4} and x_{i5} are total labor force, weekly wage rates, price index of imports, price index of exports and price index of consumption, respectively and the sample size is $n = 36$. By applying the Durbin-Watson d test, it can be shown that the errors are autocorrelated with order one, with an estimated auto-coefficient -0.349. Thus, we first consider a regression model with first order autoregressive errors given in the form of:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \varepsilon_i, \quad i = 1, 2, \ldots, 36, \tag{3.11}$$

where ε_i follows an AR(1) process. The model estimation is $\hat{y}_i = 0.49x_{1i} + 1.64x_{2i} + 1.47x_{2i} - 1.05x_{4i} - 1.45x_{5i}$ and the corresponding residual mean square error is 54.41. Now we consider the following partially linear regression model with random coefficient AR(1) error in (3.2):

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + g(t_i) + \varepsilon_i, \quad i = 1, 2, \ldots, 36, \tag{3.12}$$

where $g(t_i)$ is an unknown function and $t_i = i/36$. The estimated model is $\hat{y}_i = 0.53x_{1i} + 1.56x_{2i} + 1.35x_{2i} - 1.01x_{4i} - 1.33x_{5i} + \hat{g}(t_i)$. The bandwidth is 0.02 and the residual mean square error for this model is 4.95, which is less than that in model (3.11). Moreover, the lower panel of Figure 3.4 shows that our model in (3.12) is adequate. Now based on preliminary analysis using autoreg procedure in SAS software with first order autoregressive

Table 3.7: Estimated Coefficients

$Estimator$	β_1	β_2	β_3	β_4	β_5
$\hat{\beta}$	0.53	1.56	1.35	-1.01	-1.33
$\tilde{\beta}$	0	1.54	1.49	0	0
$\hat{\beta}^S$	0.16	0.71	0.40	-0.86	-1.14
$\hat{\beta}^{S+}$	0.16	0.71	0.40	-0.86	-1.14
$\hat{\beta}_{lasso}$	0.33	1.62	1.59	0	0

error, we realize that 3 of the variables are not significant and we consider the following hypothesis:

$$\beta_1 = \beta_4 = \beta_5 = 0.$$

The UE, RE, SE, PSE and lasso estimators are given in Table 3.7.

The performance of the estimators are evaluated in terms of their model predictive MSE (PMSE). The PMSE of $\tilde{\beta}$, $\hat{\beta}^S, \hat{\beta}^{S+}$ and $\hat{\beta}_{lasso}$ relative to $\hat{\beta}$ is given by: $RPMSE(\hat{\beta} : \hat{\beta}^*) = \frac{PMSE(\hat{y}_i;\hat{\beta})}{PMSE(\hat{y}_i;\hat{\beta}^*)}$, where $\hat{\beta}^*$ can be any of the $\tilde{\beta}$, $\hat{\beta}^S, \hat{\beta}^{S+}$ and $\hat{\beta}_{lasso}$. The RPMSE of $\tilde{\beta}$, $\hat{\beta}^S, \hat{\beta}^{S+}$ and $\hat{\beta}_{lasso}$ is calculated and reported in Table 3.8. The Table 3.8. reveals that RPMSE for shrinkage estimators are larger than that of $\hat{\beta}_{lasso}$. Since in this model p is relatively small, it is expected that lasso will do better. However, the sub-model estimator under the assumption of correctly specified model is the best one with $RPMSE$=2.15.

Table 3.8: The Relative PMSE of the Estimators

$Estimators(\hat{\beta}^*)$	$RPMSE(\hat{\beta} : \hat{\beta}^*)$
$\tilde{\beta}$	2.15
$\hat{\beta}^S$	1.47
$\hat{\beta}^{S+}$	1.47
$\hat{\beta}_{lasso}$	1.69

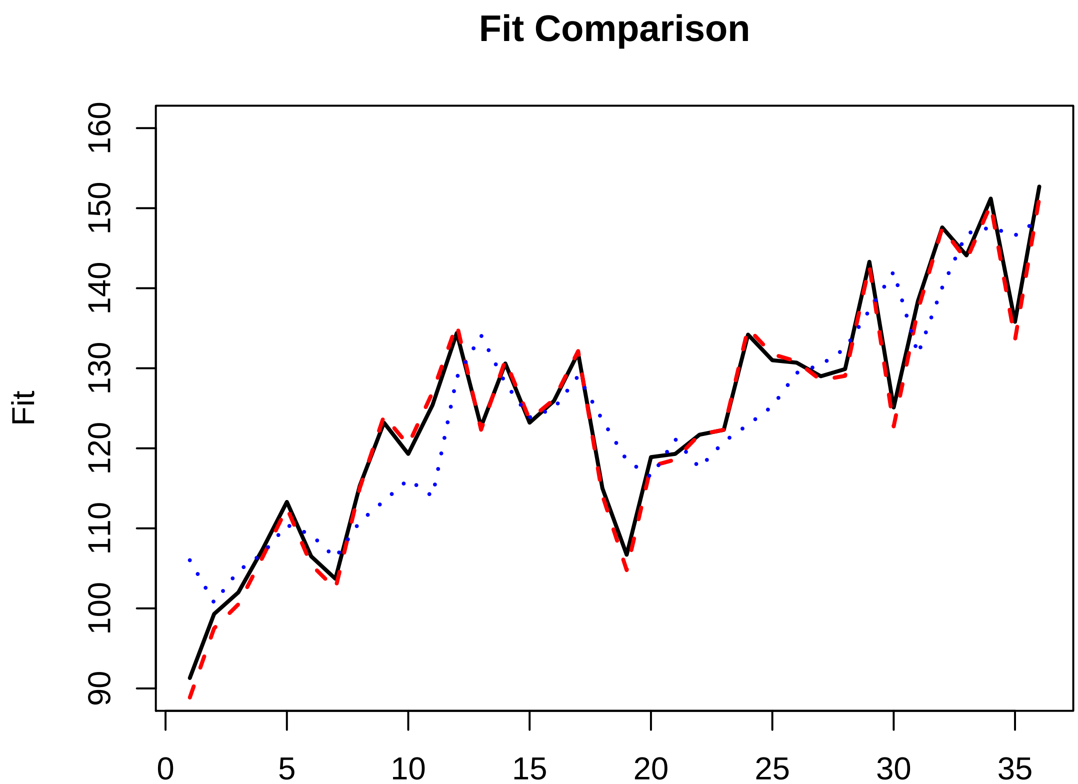

Figure 3.4: Comparison of fits from model (3.11) and model (3.12). "—" denotes the actual observations, "· · ·" denotes the linear regression model, and "– – –" denotes the the partial linear model.

3.7 Concluding Remarks

In this chapter, we considered partial linear models with random coefficient autoregressive errors. We suggested a unified estimation approach including a candidate sub-model estimator, lasso, and shrinkage estimators. We appraised the relative performance of these estimators. We demonstrated via simulation that our suggested methodology has sound finite sample properties and can be useful in practical applications. Our findings here are consistent with that of Ahmed et al. (2007), in which lasso is competitive to SE and PSE. The shrinkage estimators will perform better than the lasso estimator. Further, we reappraised the properties of shrinkage and restricted estimators for PLM with an autoregressive error. We demonstrated that, based on both analytical and numerical findings, the PSE outperforms the usual SE and the unrestricted estimator in the entire parameter space. When the restriction is true the RE is superior to all the other estimation rules; however, the MSE of RE may become unbounded when such restrictions are incorrect. Also the risk of SE and PSE is always smaller or equal to the risk of the UE.

Chapter 4

Estimation Strategies in Quasi-likelihood Models

4.1 Introduction

In real life problems, sometimes a full distributional assumption on response variables is not possible, specially in discrete data. Therefore the frequently used models such as generalized linear models (GLMs) or any model assuming a distribution on observed data can not be implemented. In order to be able to model a response variable y_i based on some existing covariates when no distributional assumption on y_i is assumed, Wedderburn (1974) introduced the term quasi-likelihood (QL) function. This term is also used as QL models. In this dissertation, we use "QL function" and "QL model" interchangeably.

The QL function has similar properties to the log-likelihood function, except that a QL function is not the log-likelihood function corresponding to any actual probability distribution. Instead of specifying a probability distribution for the data, only a relationship between the mean and the variance is specified in the form of a variance function when given the variance as a function of the mean. Thus, QL is based on the assumption of only the first two moments of the response variable.

Over the last two decades, QL estimation strategy has been widely used in different areas. For example, Lin et al. (2009) used the QL model to develop an estimating equation for the analysis of spatially correlated binary data. Alzghool et al. (2010) proposed an asymptotic QL approach for parameter estimation in multivariate heteroscedastic models with an unspecific correlation. Aue and Horváth (2011) proposed a QL procedure for estimating the unknown parameters of a first-order random coefficient autoregressive model, among other research, which indicates the importance of this model in broad areas.

In this chapter, we provide and compare shrinkage and pretest estimation strategy for QL models when UPI is in the general form given in (1.2), i.e., $\boldsymbol{F}'\boldsymbol{\beta} = \boldsymbol{d}$. We study the properties of these estimators using the notion of asymptotic distributional bias and risk. We also apply a penalty estimation strategy and compare the relative performance with shrinkage and pretest estimators through simulation studies.

4.1.1 Organization of the Chapter

The rest of this chapter is organized as follows. In Section 4.2, we propose an estimation strategy as well as an APE. Section 4.3 provides the asymptotic results of shrinkage and pretest estimators. In Section 4.4, we design and conduct a Monte Carlo experiment to study the performance of the proposed estimators and compare them with an APE. In Section 4.5, we present our concluding thoughts.

4.2 Statistical Model and Estimation

Consider the uncorrelated data y_i with $E(y_i) = \mu_i$ and $var(y_i) = \phi V(\mu_i)$, where μ_i is to be modeled in terms of a p-vector of parameters $\boldsymbol{\beta}$, the variance function $V(.)$ is assumed a known function of μ_i, and ϕ is a multiplicative factor known as the dispersion parameter or scale parameter that is estimated from the data. Suppose that for each observation y_i , the QL function $Q(y_i;\mu_i)$ is given by

$$Q(y_i;\mu_i) = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi V(t)} dt.$$

Furthermore, for each observation, the quasi-score function $U(y_i,\mu_i)$ is defined by the relation

$$U(y_i,\mu_i) = \frac{\partial Q}{\partial \mu_i} = \frac{y_i - \mu_i}{\phi V(\mu_i)}.$$

$U(y_i,\mu_i)$ has the following properties which are in common with a log likelihood derivatives:

$$\begin{aligned} E(U) &= 0 \\ var(U) &= \frac{1}{\phi V(\mu)} \\ -E(\frac{\partial U}{\partial \mu}) &= \frac{1}{\phi V(\mu)}. \end{aligned}$$

Now consider n independent observations $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ with a set of predictor values $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$. In the generalized linear form we have

$$E(y_i) = \mu_i, \qquad g(\mu_i) = \sum_{r=0}^{p} \beta_r x_{ir} \qquad i = 1, \ldots, n,$$

with the generalized form of variance

$$var(y_i) = \phi V(\mu_i) \qquad i = 1, \ldots, n,$$

where $g(.)$ is the link function which connects the random component $\boldsymbol{y}$ to the systematic components $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p$. It is obvious that μ_i is a function of $\boldsymbol{\beta}$ since $\mu_i = g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta})$, so we can rewrite $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$.

4.2.1 Unrestricted and Restricted Estimators

Our concern is to estimate the regression parameters $\beta_1, \beta_2, \ldots, \beta_p$. Since the observations are independent by assumption, the QL for the complete data is the sum of the individual quasi-likelihoods:

$$Q(\boldsymbol{y}, \boldsymbol{\mu}) = \sum_{i=1}^{n} Q(y_i, \mu_i).$$

Therefore, the estimation of the regression parameters $\boldsymbol{\beta}$ is obtained by differentiating $Q(\boldsymbol{y}, \boldsymbol{\mu})$ with respect to $\boldsymbol{\beta}$, which may be written in the form of $\boldsymbol{U}(\hat{\boldsymbol{\beta}}) = 0$, where

$$\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{D}'\boldsymbol{V}^{-1}(\boldsymbol{\mu})(\boldsymbol{y} - \boldsymbol{\mu})/\phi$$

is called the quasi-score function and $\hat{\boldsymbol{\beta}}$ is the unrestricted quasi-maximum likelihood estimator (UE) of $\boldsymbol{\beta}$. Here, $\boldsymbol{D}$ is a $n \times p$ matrix and the components

$$D_{ir} = \frac{\partial \mu_i}{\partial \beta_r} \qquad i = 1, 2, \ldots, n \qquad r = 1, 2, \ldots, p,$$

are the derivatives of $\boldsymbol{\mu}(\boldsymbol{\beta})$ with respect to the parameters. Since the data are independent, $\boldsymbol{V}(\boldsymbol{\mu})$ can be considered in the form of a diagonal matrix $\boldsymbol{V}(\boldsymbol{\mu}) =$

$diag\{V_1(\mu_1),\ldots,V_n(\mu_n)\}$, where $V_i(\mu_i)$ is a known function depending only on the i^{th} component of the mean vector $\boldsymbol{\mu}$. Wedderburn (1974) and McCullagh (1983) show that quasi-likelihoods and their corresponding quasi-maximum likelihood estimates have many properties similar to those of likelihoods and their corresponding maximum likelihood estimates.

We wish to estimate $\boldsymbol{\beta}$ when it is plausible that $\boldsymbol{\beta}$ lies in the subspace

$$\boldsymbol{F}'\boldsymbol{\beta} = \boldsymbol{d}.$$

Hence, the UPI is $\boldsymbol{F}'\boldsymbol{\beta} = \boldsymbol{d}$, where $\boldsymbol{F}$ is a $p \times q$ full rank matrix with rank $q \leq p$ and $\boldsymbol{d}$ is a given $q \times 1$ vector of constants. Based on Heyde (1997), the restricted quasi-maximum likelihood estimator (RE) $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ under UPI can be written as

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}(\boldsymbol{F}'\hat{\boldsymbol{\beta}} - \boldsymbol{d}),$$

where $\boldsymbol{\Sigma} = \boldsymbol{D}'\boldsymbol{V}^{-1}(\boldsymbol{\mu})\boldsymbol{D}$ and $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$.

4.2.2 Pretest Estimation

The pretest estimator (PTE) based on the UE and RE is defined as

$$\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}}\, I(T_n > c_{q,\alpha}) + \tilde{\boldsymbol{\beta}}\, I(T_n < c_{q,\alpha}),$$

where $c_{q,\alpha}$ is the upper α-level critical value of the χ^2 distribution with q degrees of freedom, $I(A)$ is the indicator function of the set A, and T_n is the test-statistic to test the null-hypothesis $H_0 : \boldsymbol{F}'\boldsymbol{\beta} - \boldsymbol{d} = \boldsymbol{0}$ defined as

$$T_n = \hat{\phi}^{-1}\boldsymbol{\eta}_3{}'\boldsymbol{F}(\boldsymbol{F}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\eta}_3;$$

where $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{D}}'\boldsymbol{V}^{-1}(\hat{\boldsymbol{\mu}})\hat{\boldsymbol{D}}$, $\hat{\phi} = \frac{1}{n-p}\sum_i (y_i - \hat{\mu}_i)^2/V_i(\hat{\mu}_i)$, $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}(\hat{\boldsymbol{\beta}})$ and $\boldsymbol{\eta}_3$ is defined in Proposition 4.3.1. Thus, $\hat{\boldsymbol{\beta}}^{PT}$ chooses $\tilde{\boldsymbol{\beta}}$ when H_0 is tenable, otherwise $\hat{\boldsymbol{\beta}}$. For some useful discussions on pretest estimation strategy, we refer to Ahmed and Liu (2009), among others.

4.2.3 Shrinkage Estimation

The shrinkage estimator (SE) based on the UE and RE is defined as

$$\hat{\boldsymbol{\beta}}^{S} = \tilde{\boldsymbol{\beta}} + \{1 - c_{opt}T_n^{-1}\}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \quad \text{where} \quad c_{opt} = q-2, \quad q \geq 3.$$

To avoid the over shrinking inherent in SE, we define the PSE as follows:

$$\hat{\boldsymbol{\beta}}^{S+} = \tilde{\boldsymbol{\beta}} + \{1 - c_{opt}T_n^{-1}\}^{+}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \qquad q \geq 3,$$

where $z^+ = max(0,z)$. For the sake of computation, the PSE can be rewritten in the following form

$$\hat{\boldsymbol{\beta}}^{S+} = \hat{\boldsymbol{\beta}}^{S} - [1 - c_{opt}T_n^{-1}]I(T_n < c_{opt})(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \qquad q \geq 3.$$

4.2.4 Absolute Penalty Estimator

Absolute penalty estimators (APEs) of regression coefficients are the solutions to the L_1 optimization problem

$$\hat{\boldsymbol{\beta}}_{lasso} = \underset{\boldsymbol{\beta}}{argmin}\Big\{\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \lambda\sum_{j=1}^{p}|\beta_j|\Big\},$$

where λ is the tuning parameter. Park and Hastie (2007) proposed an L_1 regularization procedure for fitting generalized linear models. It is similar to the lasso procedure, in which the loss function is replaced by the negative log-likelihood of any distribution in the exponential family; i.e.,

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{argmin}\Big\{-\ell(\boldsymbol{\beta}) + \lambda\sum_{j=1}^{p}|\beta_j|\Big\},$$

where $\ell(\boldsymbol{\beta})$ is the log-likelihood of the underlying GLM. For a review on other available techniques we refer the reader to Friedman et al (2010) and references therein.

In order to apply the L_1 penalty in the QL model, we first generated observations from the quasi-Poisson model. In a personal communication with Trevor Hastie, he suggested to use the glmnet package (Friedman et al., 2009) in R software in order to obtain the parameter estimates of the quasi-Poisson model based on L_1 penalty. The results are shown in Section 4.4.

4.3 Asymptotic Properties

In this section, we derive the asymptotic properties of the estimators. For this aim, we consider a sequence of local alternatives $\{K_n\}$ given by

$$K_n : \boldsymbol{F}'\boldsymbol{\beta} = \boldsymbol{d} + \frac{\boldsymbol{\omega}}{\sqrt{n}},$$

where $\boldsymbol{\omega}$ is a fixed q-column vector. Under the local alternative, we compute ADB and ADR of the estimators for fixed $\boldsymbol{\beta}$.

Theorem 4.3.1. Under the following regularity conditions:

(1) Weak conditions on the third derivative of $E(\boldsymbol{y}) = \boldsymbol{\mu}(\boldsymbol{\beta})$ and the third moments of $\boldsymbol{y}$ are finite;

(2) Assuming $\lim_{n\to\infty} n^{-1}\boldsymbol{D}_n'\boldsymbol{V}^{-1}(\mu)\boldsymbol{D}_n = \boldsymbol{\Sigma}$, finite and positive definite matrix,

we will have:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N_p(\boldsymbol{0}, \phi\, \boldsymbol{\Sigma}^{-1}),$$

where $\hat{\boldsymbol{\beta}}$ is the UE of $\boldsymbol{\beta}$.

Proof. The proof of the theorem can be found in McCullagh (1983). □

Proposition 4.3.1. If the regularity conditions (1) and (2) in Theorem 4.3.1 hold, then under local alternative $\{K_n\}$, as $n \to \infty$, we have

$$\begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_3 \end{pmatrix} \sim N_{2p}\left\{ \begin{pmatrix} \boldsymbol{0} \\ \gamma \end{pmatrix}, \phi \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{*} \\ \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{*} & \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{*} \end{pmatrix} \right\}$$

$$\begin{pmatrix} \boldsymbol{\eta}_2 \\ \boldsymbol{\eta}_3 \end{pmatrix} \sim N_{2p}\left\{ \begin{pmatrix} -\gamma \\ \gamma \end{pmatrix}, \phi \begin{pmatrix} \boldsymbol{\Sigma}^{*} & 0 \\ 0 & \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{*} \end{pmatrix} \right\},$$

where $\boldsymbol{\eta}_1 = \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, $\boldsymbol{\eta}_2 = \sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$, $\boldsymbol{\eta}_3 = \sqrt{n}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$, $\boldsymbol{\gamma} = \boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{\omega}$ and $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}$.

Proof. Since $\boldsymbol{\eta}_2$ and $\boldsymbol{\eta}_3$ are linear functions of $\hat{\boldsymbol{\beta}}$, they are also asymptotically normally distributed.

$$
\begin{aligned}
E(\boldsymbol{\eta}_2) &= E \lim_{n\to\infty} \sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&= E \lim_{n\to\infty} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}(\boldsymbol{F}'\hat{\boldsymbol{\beta}} - \boldsymbol{d}) - \boldsymbol{\beta}) \\
&= E \lim_{n\to\infty} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} - \boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}(\boldsymbol{F}'\hat{\boldsymbol{\beta}} - \boldsymbol{F}'\boldsymbol{\beta} + \boldsymbol{F}'\boldsymbol{\beta} - \boldsymbol{d}) \\
&= E \lim_{n\to\infty} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - E \lim_{n\to\infty} \sqrt{n}\left\{\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}[\boldsymbol{F}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \frac{\boldsymbol{\omega}}{\sqrt{n}}]\right\} \\
&= E(\lim_{n\to\infty} \boldsymbol{\eta}_1) - \boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}(E \lim_{n\to\infty}(\boldsymbol{F}'\boldsymbol{\eta}_1) + \boldsymbol{\omega}) \\
&= -\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{\omega} = -\boldsymbol{\gamma} \\
V(\boldsymbol{\eta}_2) &= Var(\sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})) = Var(\boldsymbol{\eta}_1 - \boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\eta}_1) \\
&= \phi\left\{\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\right. \\
&- \left. 2\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\right\} \\
&= \phi\left\{\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\right\}.
\end{aligned}
$$

In a similar way, one can achieve the asymptotic results of $\boldsymbol{\eta}_3$. Now the joint distribution of $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ and $(\boldsymbol{\eta}_2, \boldsymbol{\eta}_3)$ will be asymptotically normal as well. □

Proposition 4.3.2. If Proposition 4.3.1 holds, then $\phi^{-1}\boldsymbol{\eta}_3{}'\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\eta}_3 \xrightarrow[n\to\infty]{D} \chi^2_q(\Delta)$, where $\chi^2_q(\Delta)$ is a non-central chi-square distribution with q degrees of freedom and non-centrality parameter $\Delta = \phi^{-1}\boldsymbol{\omega}'(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{\omega}$. Note that the covariance matrix $\boldsymbol{\Sigma}$ and ϕ can be estimated using $\hat{\boldsymbol{\beta}}$

$$
\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{D}}'\boldsymbol{V}^{-1}(\hat{\boldsymbol{\mu}})\hat{\boldsymbol{D}}, \qquad \hat{\phi} = \frac{1}{n-p}\sum_i (y_i - \hat{\mu}_i)^2 / V_i(\hat{\mu}_i),
$$

where $\hat{\mu}_i = \mu_i(\hat{\boldsymbol{\beta}})$. Since $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, thus, by Slutsky's theorem we will have $D_n = \hat{\phi}^{-1}\boldsymbol{\eta}_3{}'\boldsymbol{F}(\boldsymbol{F}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\eta}_3 \xrightarrow[n\to\infty]{D} \chi^2_q(\Delta)$.

Proof. By using Proposition 4.3.1, under local alternative, we have $\boldsymbol{\eta}_3 \xrightarrow[n\to\infty]{D} N_p(\boldsymbol{\gamma}, \phi\, \boldsymbol{C})$ where $\boldsymbol{C} = \phi(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^*) = \phi\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}$. Consider $\boldsymbol{A} = \phi^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'$ which is a symmetric matrix, one can verify that the following conditions hold:

1) $(\boldsymbol{A}\boldsymbol{C})^2 = \boldsymbol{A}\boldsymbol{C}$,

2) $\boldsymbol{\gamma}'(\boldsymbol{A}\boldsymbol{C})^2 = \boldsymbol{\gamma}'\boldsymbol{A}\boldsymbol{C}$,

3) $\boldsymbol{\gamma}'\, \boldsymbol{A}\boldsymbol{C}\boldsymbol{A}\boldsymbol{\gamma} = \boldsymbol{\gamma}'\, \boldsymbol{A}\boldsymbol{\gamma}$,

4) $r(\boldsymbol{A}\boldsymbol{C}) = q$

under the regularity conditions in Theorem 4.3.1 and using Theorem 4 in Styan (1970), we get

$\phi^{-1}\boldsymbol{\eta}_3{}'\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\eta}_3 \sim \chi^2_q(\Delta)$ where $\Delta = \phi^{-1}\boldsymbol{\gamma}'\boldsymbol{A}\boldsymbol{\gamma} = \phi^{-1}\boldsymbol{\omega}'(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{\omega}$. □

Based on the above results, the UE and RE of $\boldsymbol{\beta}$ are consistent and they are asymptotically normal under the local alternative. In the next section we present the ADB and ADR of the estimators. The above results along with Theorems 2.3.4 and 2.3.5 in Chapter 2 will be used to derive the asymptotic results of the estimators under K_n.

4.3.1 Asymptotic Distributional Bias (ADB)

Theorem 4.3.1. If the regularity conditions (1) and (2) in Theorem 4.3.1 hold, then under $\{K_n\}$, as $n \to \infty$, the ADB of the estimators are respectively

$$\begin{aligned}
ADB(\hat{\boldsymbol{\beta}}) &= \mathbf{0}, \\
ADB(\tilde{\boldsymbol{\beta}}) &= -\boldsymbol{\gamma}, \\
ADB(\hat{\boldsymbol{\beta}}^{PT}) &= -\boldsymbol{\gamma} H_{q+2}(\chi_q^2(\alpha);\Delta), \\
ADB(\hat{\boldsymbol{\beta}}^{S}) &= -(q-2)\,\boldsymbol{\gamma}\, E[\chi_{q+2}^{-2}(\Delta)], \\
ADB(\hat{\boldsymbol{\beta}}^{S+}) &= ADB(\hat{\boldsymbol{\beta}}^{S}) - \boldsymbol{\gamma}\big\{H_{q+2}(q-2;\Delta) - (q-2)E(\chi_{q+2}^{-2}(\Delta) I(\chi_{q+2}^{2}(\Delta) < q-2))\big\}.
\end{aligned}$$

where $\Delta = \phi^{-1}\boldsymbol{\omega}'(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{\omega}$, $H_\nu(x;\Delta)$ is the distribution function of a non-central chi-square with ν degrees of freedom and non-centrality parameter Δ, and

$$E\big(\chi_\nu^{-2j}(\Delta)\big) = \int_0^\infty \chi^{-2j} dH_\nu(x;\Delta).$$

Proof. From Proposition 4.3.1, we get directly the statements $ADB(\hat{\boldsymbol{\beta}}) = \mathbf{0}$, and $ADB(\hat{\boldsymbol{\beta}}) = -\boldsymbol{\gamma}$. The ADB of the shrinkage estimators are as follows:

$$\begin{aligned}
ADB(\hat{\boldsymbol{\beta}}^S) &= E\lim_{n\to\infty}\sqrt{n}(\hat{\boldsymbol{\beta}}^S-\boldsymbol{\beta}) = E\lim_{n\to\infty}\sqrt{n}(\tilde{\boldsymbol{\beta}}+(1-(q-2)T_n^{-1})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})-\boldsymbol{\beta})\\
&= E\lim_{n\to\infty}\sqrt{n}[\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}+(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})-(q-2)T_n^{-1}(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})]\\
&= E\lim_{n\to\infty}[\boldsymbol{\eta}_2+\boldsymbol{\eta}_3-(q-2)T_n^{-1}\boldsymbol{\eta}_3]\\
&= E\lim_{n\to\infty}[\boldsymbol{\eta}_1-(q-2)T_n^{-1}\boldsymbol{\eta}_3] = -(q-2)E\lim_{n\to\infty}[T_n^{-1}\boldsymbol{\eta}_3]\\
&= -(q-2)\boldsymbol{\gamma}E(\chi_{q+2}^{-2}(\Delta)),\\
ADB(\hat{\boldsymbol{\beta}}^{S+}) &= E\lim_{n\to\infty}\sqrt{n}(\hat{\boldsymbol{\beta}}^{S+}-\boldsymbol{\beta})\\
&= E\lim_{n\to\infty}\sqrt{n}[\hat{\boldsymbol{\beta}}^S-\boldsymbol{\beta}-(1-(q-2)T_n^{-1})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})I(T_n<q-2)]\\
&= ADB(\hat{\boldsymbol{\beta}}^S)-E\lim_{n\to\infty}\sqrt{n}[(1-(q-2)T_n^{-1})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})I(T_n<q-2)]\\
&= ADB(\hat{\boldsymbol{\beta}}^S)-E\lim_{n\to\infty}[\boldsymbol{\eta}_3(1-(q-2)T_n^{-1})I(T_n<q-2)]\\
&= ADB(\hat{\boldsymbol{\beta}}^S)-E\lim_{n\to\infty}[\boldsymbol{\eta}_3 I(T_n<q-2)]+E\lim_{n\to\infty}[(q-2)\boldsymbol{\eta}_3 T_n^{-1}I(T_n<q-2)]\\
&= ADB(\hat{\boldsymbol{\beta}}^S)-\boldsymbol{\gamma}H_{q+2}(q-2;\Delta)+\boldsymbol{\gamma}(q-2)E(\chi_{q+2}^{-2}(\Delta)I(\chi_{q+2}^{2}(\Delta)<q-2)).
\end{aligned}$$

□

Since the bias expressions are not in scalar form, we convert them to quadratic form. The asymptotic quadratic distributional bias (AQDB) of an estimator as follows

$$AQDB(\boldsymbol{\beta}^0) = (ADB(\boldsymbol{\beta}^0))'\phi^{-1}\boldsymbol{\Sigma}(ADB(\boldsymbol{\beta}^0))$$

Corollary 4.3.1. Suppose that the assumptions of Theorem 4.3.1 hold. Then under $\{K_n\}$, as $n \to \infty$, the AQDB of the estimators are

$$\begin{aligned}
AQDB(\hat{\boldsymbol{\beta}}) &= \mathbf{0}, \\
AQDB(\tilde{\boldsymbol{\beta}}) &= \phi^{-1}\gamma'\boldsymbol{\Sigma}\gamma = \Delta, \\
AQDB(\hat{\boldsymbol{\beta}}^{PT}) &= \Delta\big(H_{q+2}(\chi_q^2(\alpha);\Delta)\big)^2, \\
AQDB(\hat{\boldsymbol{\beta}}^{S}) &= \Delta(q-2)^2\big(E[\chi_{q+2}^{-2}(\Delta)]\big)^2, \\
AQDB(\hat{\boldsymbol{\beta}}^{S+}) &= \Delta\Big((q-2)E[\chi_{q+2}^{-2}(\Delta)] - H_{q+2}(q-2;\Delta) \\
&+ (q-2)E(\chi_{q+2}^{-2}(\Delta)I(\chi_{q+2}^{2}(\Delta) < q-2))\Big)^2.
\end{aligned}$$

Proof. The expressions for quadratic biases are obtained by following the definition of AQDB. □

4.3.2 Asymptotic Distributional Risk (ADR)

Theorem 4.3.2. Suppose that the assumptions of Theorem 4.3.1 hold. Then under $\{K_n\}$, as $n \to \infty$, the asymptotic covariance matrices of the estimators are

$$\begin{aligned}
\Gamma(\hat{\boldsymbol{\beta}}) &= \phi\boldsymbol{\Sigma}^{-1}, \\
\Gamma(\tilde{\boldsymbol{\beta}}) &= \phi\Sigma^{*} + \gamma\gamma', \\
\Gamma(\hat{\boldsymbol{\beta}}^{PT}) &= \phi\boldsymbol{\Sigma}^{-1} - \phi(\boldsymbol{\Sigma}^{-1} - \Sigma^{*})H_{q+2}(\chi^2_q(\alpha);\Delta) + \gamma\gamma'\left\{2H_{q+2}(\chi^2_q(\alpha);\Delta) - H_{q+4}(\chi^2_q(\alpha);\Delta)\right\}, \\
\Gamma(\hat{\boldsymbol{\beta}}^{S}) &= \phi\boldsymbol{\Sigma}^{-1} + (q-2)\phi(\boldsymbol{\Sigma}^{-1} - \Sigma^{*})[-2E(\chi^{-2}_{q+2}(\Delta)) + (q-2)E(\chi^{-4}_{q+2}(\Delta))] \\
&+ (q-2)\gamma\gamma'[-2E(\chi^{-2}_{q+4}(\Delta)) + 2E(\chi^{-2}_{q+2}(\Delta)) + (q-2)E(\chi^{-4}_{q+4}(\Delta))], \\
\Gamma(\hat{\boldsymbol{\beta}}^{S+}) &= \Gamma(\hat{\boldsymbol{\beta}}^{S}) + 2\gamma\gamma' E[(1-(q-2)\chi^{-2}_{q+2}(\Delta))I(\chi^2_{q+2}(\Delta) < q-2)] \\
&- \gamma\gamma' E[(1-(q-2)\chi^{-2}_{q+4}(\Delta))^2 I(\chi^2_{q+4}(\Delta) < q-2)] \\
&- \phi(\boldsymbol{\Sigma}^{-1} - \Sigma^{*})E[(1-(q-2)\chi^{-2}_{q+2}(\Delta))^2 I(\chi^2_{q+2}(\Delta) < q-2)].
\end{aligned}$$

Proof. The asymptotic covariance of an estimator $\boldsymbol{\beta}^*$ is defined as follows:

$$\Gamma(\boldsymbol{\beta}^0) = E\lim_{n\to\infty}(n(\boldsymbol{\beta}^0 - \boldsymbol{\beta})(\boldsymbol{\beta}^0 - \boldsymbol{\beta})').$$

Therefore,

$$
\begin{aligned}
\Gamma(\hat{\boldsymbol{\beta}}) &= E\lim_{n\to\infty}(n(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})') = E\lim_{n\to\infty}(\boldsymbol{\eta}_1\boldsymbol{\eta}_1{}')\\
&= Var(\boldsymbol{\eta}_1)+E(\boldsymbol{\eta}_1)E(\boldsymbol{\eta}_1)' = \phi\boldsymbol{\Sigma}^{-1},\\
\Gamma(\tilde{\boldsymbol{\beta}}) &= E\lim_{n\to\infty}(n(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})') = E\lim_{n\to\infty}(\boldsymbol{\eta}_2\boldsymbol{\eta}_2{}')\\
&= Var(\boldsymbol{\eta}_2)+E(\boldsymbol{\eta}_2)E(\boldsymbol{\eta}_2)' = \phi\boldsymbol{\Sigma}^{*}+\boldsymbol{\gamma}\boldsymbol{\gamma}',\\
\Gamma(\hat{\boldsymbol{\beta}}^S) &= E\lim_{n\to\infty}(n(\hat{\boldsymbol{\beta}}^S-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^S-\boldsymbol{\beta})')\\
&= E\lim_{n\to\infty}\sqrt{n}\left((\tilde{\boldsymbol{\beta}}+(1-(q-2)T_n^{-1})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})-\boldsymbol{\beta}\right)\\
&\times\ \sqrt{n}\left(\tilde{\boldsymbol{\beta}}+(1-(q-2)T_n^{-1})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})-\boldsymbol{\beta}\right)'\\
&= E\lim_{n\to\infty}[(\boldsymbol{\eta}_1-(q-2)T_n^{-1}\boldsymbol{\eta}_3)(\boldsymbol{\eta}_1-(q-2)T_n^{-1}\boldsymbol{\eta}_3)']\\
&= E\lim_{n\to\infty}[\boldsymbol{\eta}_1\boldsymbol{\eta}_1{}'-(q-2)T_n^{-1}\boldsymbol{\eta}_1\boldsymbol{\eta}_3{}'-(q-2)T_n^{-1}\boldsymbol{\eta}_3\boldsymbol{\eta}_1{}'+(q-2)^2T_n^{-2}\boldsymbol{\eta}_3\boldsymbol{\eta}_3{}']\\
&= Var(\boldsymbol{\eta}_1)-2(q-2)E\lim_{n\to\infty}(\boldsymbol{\eta}_3\boldsymbol{\eta}_1{}'T_n^{-1})+(q-2)^2E\lim_{n\to\infty}(T_n^{-2}\boldsymbol{\eta}_3\boldsymbol{\eta}_3{}').
\end{aligned}
$$

Note that, by using Lemmas 2.3.4 and 2.3.5 in Chapter 2, we have

$$
\begin{aligned}
E\lim_{n\to\infty}(\boldsymbol{\eta}_3\boldsymbol{\eta}_1{}'T_n^{-1}) &= E\lim_{n\to\infty}(E(\boldsymbol{\eta}_3\boldsymbol{\eta}_1{}'T_n^{-1}|\boldsymbol{\eta}_3))\\
&= E\lim_{n\to\infty}(\boldsymbol{\eta}_3[E(\boldsymbol{\eta}_1)+\phi(\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{*})\phi^{-1}(\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{*})^{-1}(\boldsymbol{\eta}_3-E(\boldsymbol{\eta}_3))]'T_n^{-1})\\
&= E\lim_{n\to\infty}(\boldsymbol{\eta}_3[\boldsymbol{\eta}_3{}'-\boldsymbol{\gamma}']T_n^{-1})\\
&= E\lim_{n\to\infty}(\boldsymbol{\eta}_3\boldsymbol{\eta}_3{}'T_n^{-1})-E\lim_{n\to\infty}(\boldsymbol{\eta}_3\boldsymbol{\gamma}'T_n^{-1})\\
&= \phi(\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{*})E(\chi^{-2}_{p_2+2}(\Delta))+\boldsymbol{\gamma}\boldsymbol{\gamma}'E(\chi^{-2}_{p_2+4}(\Delta))-\boldsymbol{\gamma}\boldsymbol{\gamma}'E(\chi^{-2}_{p_2+2}(\Delta))\\
&= \phi(\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{*})E(\chi^{-2}_{p_2+2}(\Delta))+\boldsymbol{\gamma}\boldsymbol{\gamma}'[E(\chi^{-2}_{p_2+4}(\Delta))-E(\chi^{-2}_{p_2+2}(\Delta))].
\end{aligned}
$$

Therefore,

$$\begin{aligned}
\Gamma(\hat{\boldsymbol{\beta}}^{S}) &= \phi\boldsymbol{\Sigma}^{-1} - 2(q-2)\phi(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{*})E(\chi^{-2}_{p_2+2}(\Delta)) \\
&- 2(q-2)\boldsymbol{\gamma}\boldsymbol{\gamma}'[E(\chi^{-2}_{p_2+4}(\Delta)) - E(\chi^{-2}_{p_2+2}(\Delta))] \\
&+ (q-2)^2[\phi(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{*})E(\chi^{-4}_{p_2+2}(\Delta)) + \boldsymbol{\gamma}\boldsymbol{\gamma}'E(\chi^{-4}_{p_2+4}(\Delta))] \\
&= \phi\boldsymbol{\Sigma}^{-1} + (q-2)\phi(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{*})[-2E(\chi^{-2}_{p_2+2}(\Delta)) + (q-2)E(\chi^{-4}_{p_2+2}(\Delta))] \\
&+ (q-2)\boldsymbol{\gamma}\boldsymbol{\gamma}'[-2E(\chi^{-2}_{p_2+4}(\Delta)) + 2E(\chi^{-2}_{p_2+2}(\Delta)) + (q-2)E(\chi^{-4}_{p_2+4}(\Delta))] \\
\Gamma(\hat{\boldsymbol{\beta}}^{S+}) &= E\lim_{n\to\infty}(n(\hat{\boldsymbol{\beta}}^{S+} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{S+} - \boldsymbol{\beta})') \\
&= E\lim_{n\to\infty}\sqrt{n}\Big((\hat{\boldsymbol{\beta}}^{S} - (1-(q-2)T_n^{-1})I(T_n < q-2)(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}\Big) \\
&\times \sqrt{n}\Big(\hat{\boldsymbol{\beta}}^{S} - (1-(q-2)T_n^{-1})I(T_n < q-2)(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}\Big)' \\
&= \Gamma(\hat{\boldsymbol{\beta}}^{S}) - 2E\lim_{n\to\infty}[\boldsymbol{\eta}_3\boldsymbol{\eta}_2'(1-(q-2)T_n^{-1})I(T_n < q-2)] \\
&- 2E\lim_{n\to\infty}[\boldsymbol{\eta}_3\boldsymbol{\eta}_3'(1-(q-2)T_n^{-1})^2 I(T_n < q-2)] \\
&+ E\lim_{n\to\infty}(\boldsymbol{\eta}_3\boldsymbol{\eta}_3'(1-(q-2)T_n^{-1})^2 I(T_n < q-2)) \\
&= \Gamma(\hat{\boldsymbol{\beta}}^{S}) - 2E\lim_{n\to\infty}[\boldsymbol{\eta}_3\boldsymbol{\eta}_2'(1-(q-2)T_n^{-1})I(T_n < q-2)] \\
&- E\lim_{n\to\infty}[\boldsymbol{\eta}_3\boldsymbol{\eta}_3'(1-(q-2)T_n^{-1})^2 I(T_n < q-2)].
\end{aligned}$$

Now we have

$$\begin{aligned}
&E\lim_{n\to\infty}[\boldsymbol{\eta}_3\boldsymbol{\eta}_2'(1-(q-2)T_n^{-1})I(T_n < q-2)] \\
&= E\lim_{n\to\infty}[\boldsymbol{\eta}_3 E(\boldsymbol{\eta}_2'(1-(q-2)T_n^{-1})I(T_n < q-2)|\boldsymbol{\eta}_3)] \\
&= E\lim_{n\to\infty}[\boldsymbol{\eta}_3(-\boldsymbol{\gamma} + \mathbf{0}\times\phi(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{*})^{-1}(\boldsymbol{\eta}_3 - \boldsymbol{\gamma}))'(1-(q-2)T_n^{-1})I(T_n < q-2)] \\
&= -E\lim_{n\to\infty}[\boldsymbol{\eta}_3\boldsymbol{\gamma}'(1-(q-2)T_n^{-1})I(T_n < q-2)] \\
&= -\boldsymbol{\gamma}\boldsymbol{\gamma}'E[(1-(q-2)\chi^{-2}_{p_2+2}(\Delta))I(\chi^{2}_{p_2+2}(\Delta) < q-2)],
\end{aligned}$$

and based on Lemmas 2.3.4 in Chapter 2, we have

$$
\begin{aligned}
& E\lim_{n\to\infty}[\boldsymbol{\eta}_3\boldsymbol{\eta}_3{}'(1-(q-2)T_n^{-1})^2 I(T_n<q-2)] \\
& = \phi(\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{*})E[(1-(q-2)\chi^{-2}_{p_2+2}(\Delta))^2 I(\chi^2_{p_2+2}(\Delta)<q-2)] \\
& +\boldsymbol{\gamma\gamma}'E[(1-(q-2)\chi^{-2}_{p_2+4}(\Delta))^2 I(\chi^2_{p_2+4}(\Delta)<q-2)].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\Gamma(\hat{\boldsymbol{\beta}}^{S+}) &= \Gamma(\hat{\boldsymbol{\beta}}^{S})+2\boldsymbol{\gamma\gamma}'E[(1-(q-2)\chi^{-2}_{p_2+2}(\Delta))I(\chi^2_{p_2+2}(\Delta)<q-2)] \\
&- \boldsymbol{\gamma\gamma}'E[(1-(q-2)\chi^{-2}_{p_2+4}(\Delta))^2 I(\chi^2_{p_2+4}(\Delta)<q-2)] \\
&- \phi(\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{*})E[(1-(q-2)\chi^{-2}_{p_2+2}(\Delta))^2 I(\chi^2_{p_2+2}(\Delta)<q-2)].
\end{aligned}
$$

□

Corollary 4.3.2. If the assumptions of Theorem 4.3.1 hold, then under $\{K_n\}$, as $n\to\infty$, the ADR of the estimators are

$$
\begin{aligned}
ADR(\hat{\boldsymbol{\beta}};\boldsymbol{M}) &= \phi\,\mathrm{tr}(\boldsymbol{M}\boldsymbol{\Sigma}^{-1}), \\
ADR(\tilde{\boldsymbol{\beta}};\boldsymbol{M}) &= \phi\,\mathrm{tr}(\boldsymbol{M}\boldsymbol{\Sigma}^{-1})-\phi\mathrm{tr}(\boldsymbol{A}_{11})+\boldsymbol{\gamma}'\boldsymbol{M}\boldsymbol{\gamma}, \\
ADR(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{M}) &= \phi\,\mathrm{tr}(\boldsymbol{M}\boldsymbol{\Sigma}^{-1})-\phi\,\mathrm{tr}(\boldsymbol{A}_{11})H_{q+2}(\chi^2_q(\alpha);\Delta) \\
&+ \boldsymbol{\gamma}'\boldsymbol{M}\boldsymbol{\gamma}\Big\{2H_{q+2}(\chi^2_q(\alpha);\Delta)-H_{q+4}(\chi^2_q(\alpha);\Delta)\Big\}, \\
ADR(\hat{\boldsymbol{\beta}}^{S};\boldsymbol{M}) &= \phi\,\mathrm{tr}(\boldsymbol{M}\boldsymbol{\Sigma}^{-1})+(q-2)\phi\,\mathrm{tr}(\boldsymbol{A}_{11})\Big\{-2E(\chi^{-2}_{q+2}(\Delta))+(q-2)E(\chi^{-4}_{q+2}(\Delta))\Big\} \\
&+ (q-2)\boldsymbol{\gamma}'\boldsymbol{M}\boldsymbol{\gamma}[-2E(\chi^{-2}_{p_2+4}(\Delta)1)+2E(\chi^{-2}_{q+2}(\Delta))+(q-2)E(\chi^{-4}_{q+4}(\Delta))], \\
ADR(\hat{\boldsymbol{\beta}}^{S+}) &= ADR(\hat{\boldsymbol{\beta}}^{S})+2\,\boldsymbol{\gamma}'\boldsymbol{M}\boldsymbol{\gamma}\,E[(1-(q-2)\chi^{-2}_{q+2}(\Delta))I(\chi^2_{q+2}(\Delta)<q-2)] \\
&- \boldsymbol{\gamma}'\boldsymbol{M}\boldsymbol{\gamma}\,E[(1-(q-2)\chi^{-2}_{q+4}(\Delta))^2 I(\chi^2_{q+4}(\Delta)<q-2)] \\
&- \phi\,\mathrm{tr}(\boldsymbol{A}_{11})\,E[(1-(q-2)\chi^{-2}_{q+2}(\Delta))^2 I(\chi^2_{q+2}(\Delta)<q-2)].
\end{aligned}
$$

where $\boldsymbol{A}_{11} = \boldsymbol{M}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}$.

Proof. The expressions for risk are obtained by following the definition of ADR. □

4.3.3 Bias and Risk Comparisons

Clearly, the quadratic biases of the estimators are functions of $\boldsymbol{\omega}$. Therefore, for comparison, it suffices to compare the scalar factor Δ only. Under H_0, i.e., when $\Delta = 0$, all the estimators are unbiased. However, as Δ moves away from 0, the $AQDB(\tilde{\boldsymbol{\beta}})$ becomes unbounded function of Δ and $AQDB(\hat{\boldsymbol{\beta}}^{PT})$ will be less than $AQDB(\hat{\boldsymbol{\beta}}^{S})$ for all values of Δ since $H_{q+2}(\chi^2_q(\alpha);\Delta)$ lies between 0 and 1. The $AQDB(\hat{\boldsymbol{\beta}}^{S+})$ and $AQDB(\hat{\boldsymbol{\beta}}^{S})$ start from the origin at $\Delta = 0$, and, as Δ increases, they increase to a maximum and then decrease towards 0. It can be shown that the $AQDB(\hat{\boldsymbol{\beta}}^{S+}) \leq AQDB(\hat{\boldsymbol{\beta}}^{S})$; thus, we have $AQDB(\hat{\boldsymbol{\beta}}^{S+}) \leq AQDB(\hat{\boldsymbol{\beta}}^{S}) \leq AQDB(\tilde{\boldsymbol{\beta}})$.

By comparing the risk of the estimators, we see that, as Δ moves away from 0, the risk of $\tilde{\boldsymbol{\beta}}$ becomes unbounded and the risk of $\hat{\boldsymbol{\beta}}^{S+}$ is asymptotically superior to $\hat{\boldsymbol{\beta}}^{S}$ for all values of $\Delta \geq 0$. Thus, not only does $\hat{\boldsymbol{\beta}}^{S+}$ confirm the inadmissibility of $\hat{\boldsymbol{\beta}}^{S}$, but it also provides a simple superior estimator. Also, by comparing the risk of $\hat{\boldsymbol{\beta}}^{S}$ and $\hat{\boldsymbol{\beta}}$, it can be easily shown that, under certain conditions $ADR(\hat{\boldsymbol{\beta}}^{S}) \leq ADR(\hat{\boldsymbol{\beta}}) = \phi \ \mathrm{tr}(\boldsymbol{M}\boldsymbol{\Sigma}^{-1})$ for all $\Delta \geq 0$. Hence, the PSE dominates the UE and we have $ADR(\hat{\boldsymbol{\beta}}^{S+}) \leq ADR(\hat{\boldsymbol{\beta}}^{S}) \leq ADR(\hat{\boldsymbol{\beta}})$. By comparing the risk of $\boldsymbol{\beta}^{PT}$ and $\hat{\boldsymbol{\beta}}$, it can be shown that as Δ increases, the $ADR(\hat{\boldsymbol{\beta}}^{PT})$ will increase and reaches the $ADR(\hat{\boldsymbol{\beta}})$ from below. Furthermore, beyond small values of $\Delta(\Delta \in [0, c])$, the risk of RE is higher than the other estimators, however, under the null hypothesis, i.e., for $\Delta = 0$:

$$ADR(\tilde{\boldsymbol{\beta}}) \leq ADR(\hat{\boldsymbol{\beta}}^{PT}) \leq ADR(\hat{\boldsymbol{\beta}}^{S+}) \leq ADR(\hat{\boldsymbol{\beta}}^{S}) \leq ADR(\hat{\boldsymbol{\beta}}).$$

4.4 Monte Carlo Simulation

In this section, we provide a Monte Carlo simulation study to investigate the performance of the proposed estimators with different numbers of explanatory variables. Our sampling experiment consists of different combination of sample sizes, i.e., $n = 30, 50, 100$. In this study, we considered the following model:

$$log(\lambda_i) = \boldsymbol{x}_i'\boldsymbol{\beta}, \quad i = 1, 2, \ldots, n,$$

where $\boldsymbol{x}_i' = (x_{1i}, x_{2i}, \ldots, x_{pi})$, $\lambda_i = E(y_i|\boldsymbol{x}_i)$, and $y_i's$ are observations from an over-dispersed Poisson model, and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$ is the vector of coefficients. In order to generate over-dispersed Poisson observations with mean λ_i and variance $\phi\lambda_i$, we considered a Negative Binomial distribution $NB(r_i, p)$ with

$$r_i = \frac{\lambda_i}{\phi - 1} \text{ and } p = \frac{1}{\phi}, \quad i = 1, 2, \ldots, n,$$

where $\lambda_i = e^{\boldsymbol{x}_i'\boldsymbol{\beta}}$ and $x_{si} = t_s^2 + \nu_i$ with t_s and ν_i being i.i.d $N(0, 1)$ for all $s = 1, \ldots, p$ and $i = 1, \ldots, n$. Also, in the simulation we considered $\phi = 2$. Our sampling experiment consists of various combinations of sample sizes, i.e., $n = 50$ and 100. For each n, we generate 5000 samples using the above model. We also use the 10-fold cross validation method to estimate the tuning parameter λ to compute lasso. Furthermore, we use the aod-package (Lesnoff and Lancelot, 2012) in R statistical software to fit the above model to account for the over-dispersed Poisson model. In our simulation, we consider the UPI in the following format: $\boldsymbol{F}' = (\mathbf{0}, \boldsymbol{I})$ where $\boldsymbol{I}_{p_2 \times p_2}$ is the identity matrix, and $\mathbf{0}_{p_2 \times p_1}$ is the matrix of 0s and $\boldsymbol{d}_{p_2 \times 1} = \mathbf{0}$. Also, we set the regression coefficients of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ to $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \mathbf{0}')'$ with $\beta_j = 0$, for $j = p_1 + 1, \ldots, p$ with $p = p_1 + p_2$ for the following cases:

Case 1: $\boldsymbol{\beta}_1 = (1,1,1)'$ and $\boldsymbol{\beta}_2 = \mathbf{0}_{p_2 \times 1}$, where $p_2 = 3, 4, \cdots, 8$.

Case 2: $\boldsymbol{\beta}_1 = (1,1,1,1,1)'$ and $\boldsymbol{\beta}_2 = \mathbf{0}_{p_2 \times 1}$, where $p_2 = 3, 4, \cdots, 8$.

Now we define the parameter $\Delta^* = \|\boldsymbol{\beta} - \boldsymbol{\beta}^*\|^2$, where $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1', \mathbf{0}')'$ and $\| \ . \ \|$ is the Euclidian norm. The objective is to investigate the behavior of the estimators for $\Delta^* \geq 0$. In order to do this, further samples are generated from those distributions (i.e. for different Δ^* between 0 and 2). To produce different values of Δ^*, different values of $\boldsymbol{\beta}_2$ are chosen. The relative MSE of the estimators $\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^S$, and $\hat{\boldsymbol{\beta}}^{S+}$ have been numerically calculated with respect to $\hat{\boldsymbol{\beta}}$ using the *R* statistical software. The relative mean squared error (RMSE) of the other estimators to the unrestricted estimator $\hat{\boldsymbol{\beta}}$ is defined by

$$RMSE(\hat{\boldsymbol{\beta}} : \hat{\boldsymbol{\beta}}^{\star}) = \frac{MSE(\hat{\boldsymbol{\beta}})}{MSE(\hat{\boldsymbol{\beta}}^{\star})},$$

where $\hat{\boldsymbol{\beta}}^{\star}$ can be any of $\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^S, \hat{\boldsymbol{\beta}}^{S+}, \boldsymbol{\beta}^{PT}$ and $\hat{\boldsymbol{\beta}}_{lasso}$. It is obvious that a RMSE larger than one indicates the degree of superiority of the estimator $\hat{\boldsymbol{\beta}}^{\star}$ over $\hat{\boldsymbol{\beta}}$.

The performance of lasso is independent of the parameter Δ^*. This estimator does not take advantage of the fact that the regression parameter lies in a subspace and is at a disadvantage when $\Delta^* > 0$. Therefore, only $\Delta^* = 0$ was considered to compare the MSE of lasso with MSE of other estimators. Figures 4.1 to 4.4 portray the relative performance of the suggested estimators excluding lasso and Tables 4.1 to 4.4 show the relative performance of lasso compared to the other estimators when $\Delta^* = 0$. We summarize our findings as follows:

(i) The lasso estimator outperforms the UE. In all simulation cases, the RMSE of lasso is higher than that in the UE, indicating that this estimator has a lower MSE compared to the UE.

(ii) Shrinkage estimators perform better than the UE. Comparing the SE and the PSE shows that the PSE has a higher RMSE than the SE, which shows the better performance of the PSE over the SE.

(iii) Shrinkage estimators perform better than lasso only when there are many nuisance predictors in the model. The gain in efficiency depends on the value of p_2: the larger p_2 is relative to p_1, the larger the gain in efficiency.

(iv) For smaller values of Δ^*, the RE has less risk than PTE; however, beyond the small interval near the null hypothesis ($\Delta^* = 0$), the PTE performs better than the RE. The PTE outperforms shrinkage and lasso at $\Delta^* = 0$ for all values of p_2; however after small intervals near $\Delta^* = 0$, both SE and PSE dominate PTE and then they all reach the risk of UE.

(v) The RE performs best only when Δ^* is small. For large values of Δ^*, it becomes very inconsistent and its efficiency converges to 0. Again, if a sub-model is nearly correctly specified, then RE is optimal one.

Table 4.1: Relative MSE of estimators with respect to $\hat{\beta}_1$ when $p_1 = 3$ and $n = 50$

p_2	$\tilde{\beta}$	$\hat{\beta}^S$	$\hat{\beta}^{S+}$	$\hat{\beta}^{PT}$	$\hat{\beta}_{lasso}$
3	2.309	1.230	1.276	1.929	1.407
5	2.175	1.514	1.766	2.023	1.723
7	3.919	2.100	2.584	2.943	1.985

Table 4.2: Relative MSE of estimators with respect to $\hat{\beta}_1$ when $p_1 = 3$ and $n = 100$

p_2	$\tilde{\beta}$	$\hat{\beta}^S$	$\hat{\beta}^{S+}$	$\hat{\beta}^{PT}$	$\hat{\beta}_{lasso}$
3	2.225	1.260	1.356	1.877	1.496
5	3.482	1.810	2.137	2.756	1.823
7	3.715	2.115	2.414	2.780	2.041

Table 4.3: Relative MSE of estimators with respect to $\hat{\beta}_1$ when $p_1 = 5$ and $n = 50$

p_2	$\tilde{\beta}$	$\hat{\beta}^S$	$\hat{\beta}^{S+}$	$\hat{\beta}^{PT}$	$\hat{\beta}_{lasso}$
3	1.696	1.140	1.251	1.544	1.345
5	2.156	1.452	1.675	1.942	1.453
7	3.933	2.104	2.610	3.039	1.759

Table 4.4: Relative MSE of estimators with respect to $\hat{\beta}_1$ when $p_1 = 5$ and $n = 100$

p_2	$\tilde{\beta}$	$\hat{\beta}^S$	$\hat{\beta}^{S+}$	$\hat{\beta}^{PT}$	$\hat{\beta}_{lasso}$
3	1.551	1.148	1.202	1.432	1.392
5	2.605	1.584	1.826	2.213	1.577
7	3.384	1.979	2.391	2.745	1.841

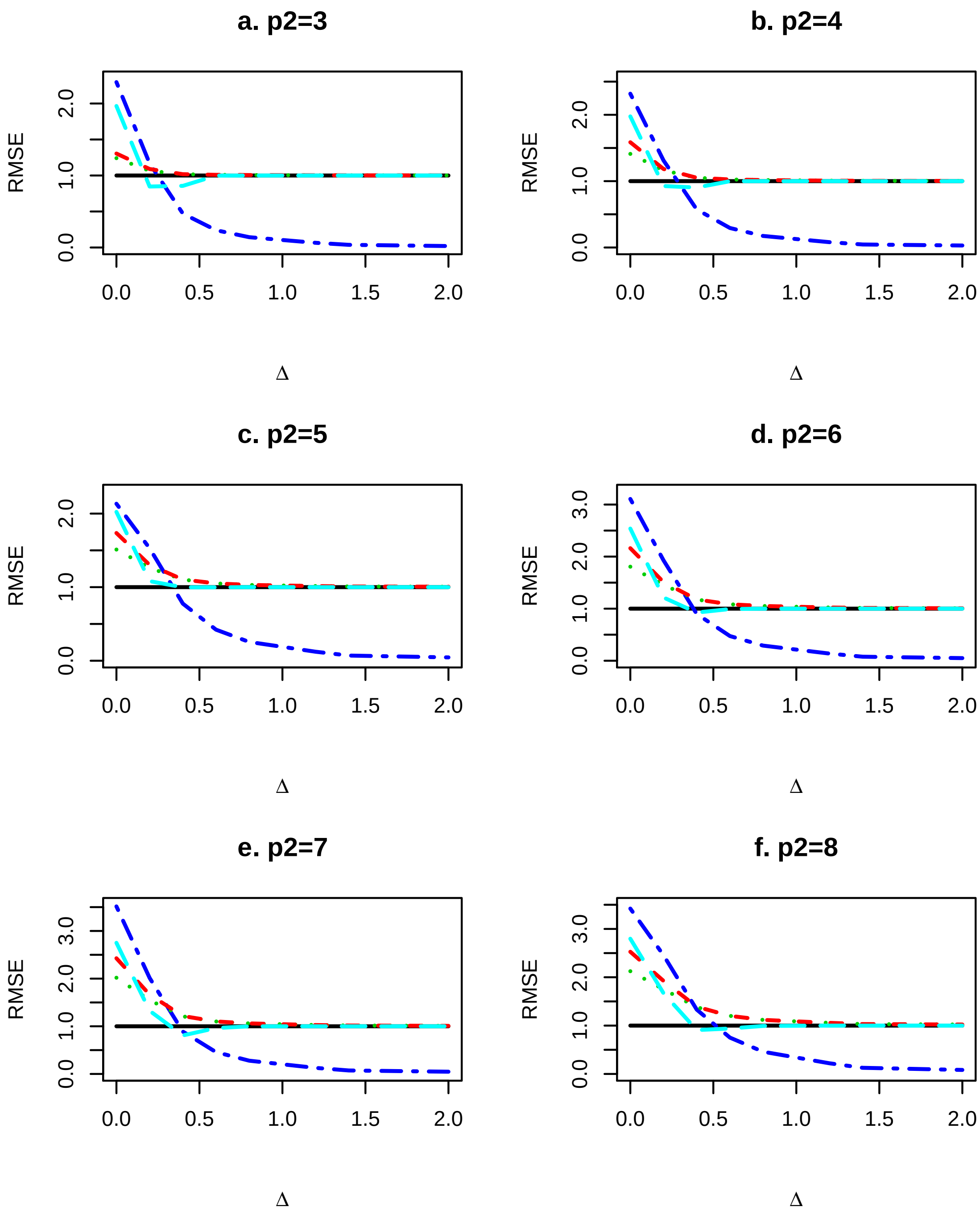

Figure 4.1: Relative MSE of the estimators for various p_2 when $p_1 = 3$ and $n = 50$. "- - -" denotes the PSE, "$\cdots$" denotes the SE, "– · – ·" denotes the RE, "—" denotes the UE, and "– – –" denotes the PTE.

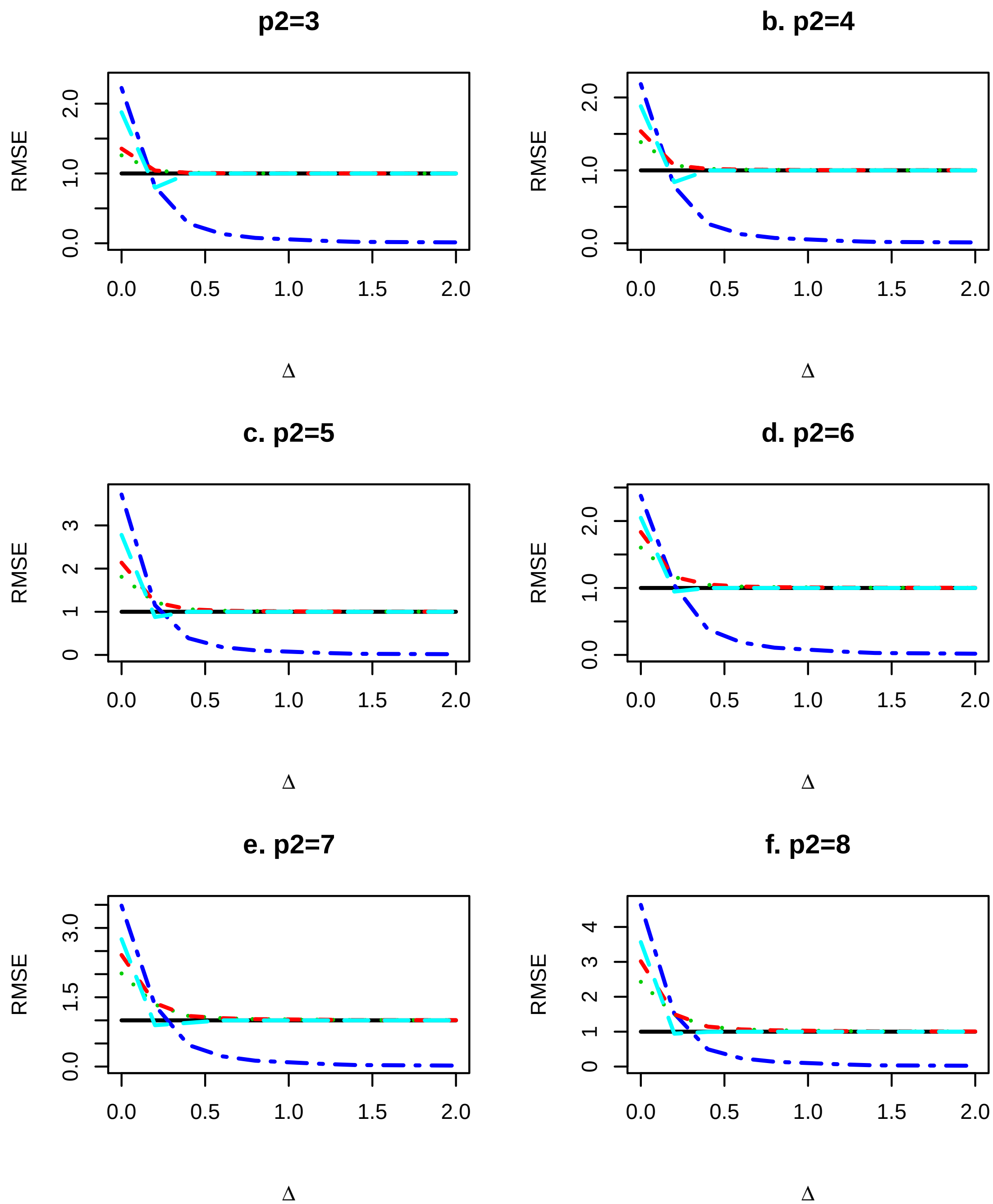

Figure 4.2: Relative MSE of the estimators for various p_2 when $p_1 = 3$ and $n = 100$. "- - -" denotes the PSE, "···" denotes the SE, "– · – ·" denotes the RE, "—" denotes the UE, and "– – –" denotes the PTE.

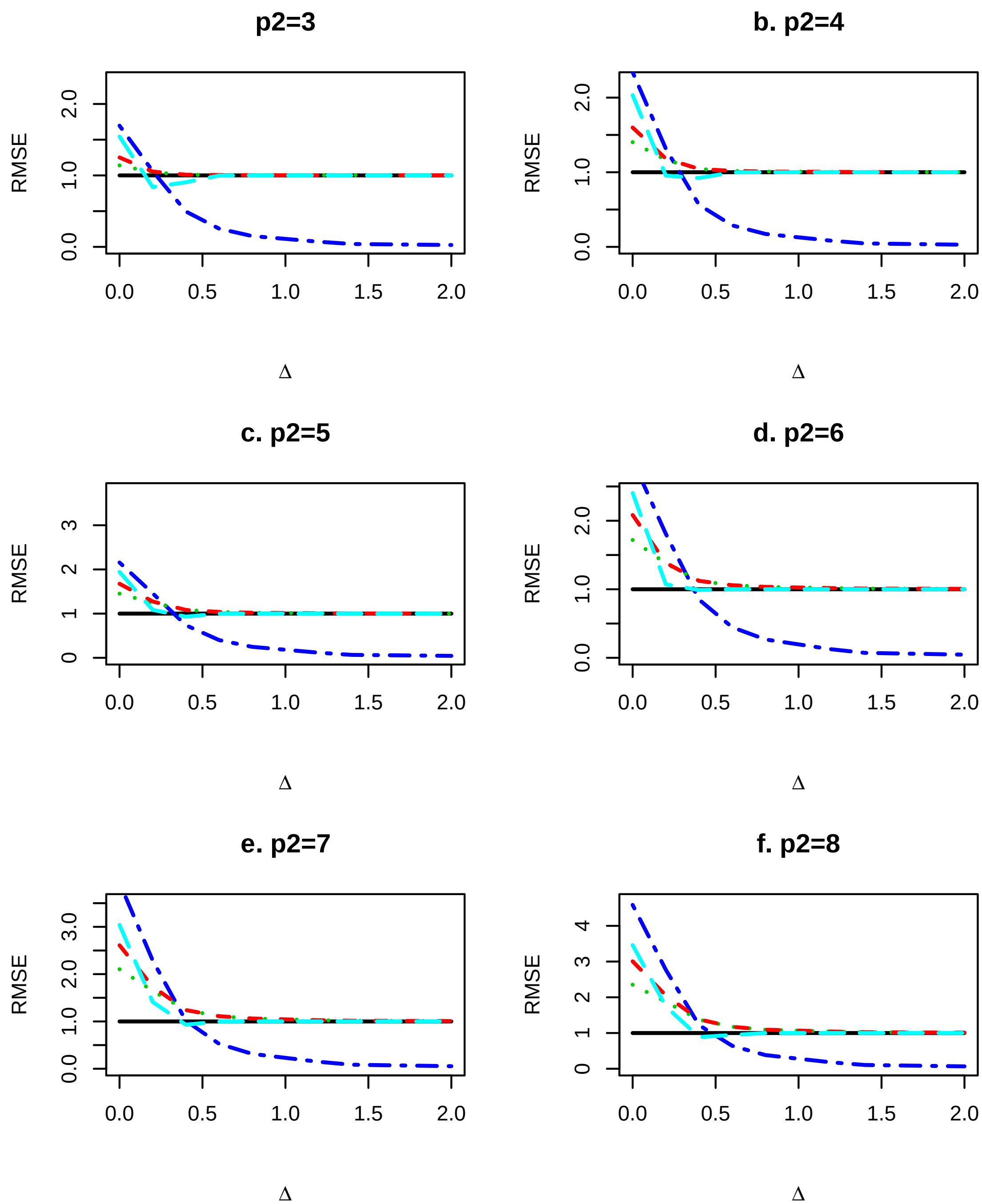

Figure 4.3: Relative MSE of the estimators for various p_2 when $p_1 = 5$ and $n = 50$. "- - -" denotes the PSE, "···" denotes the SE, "– · – ·" denotes the RE, "—" denotes the UE, and "– – –" denotes the PTE.

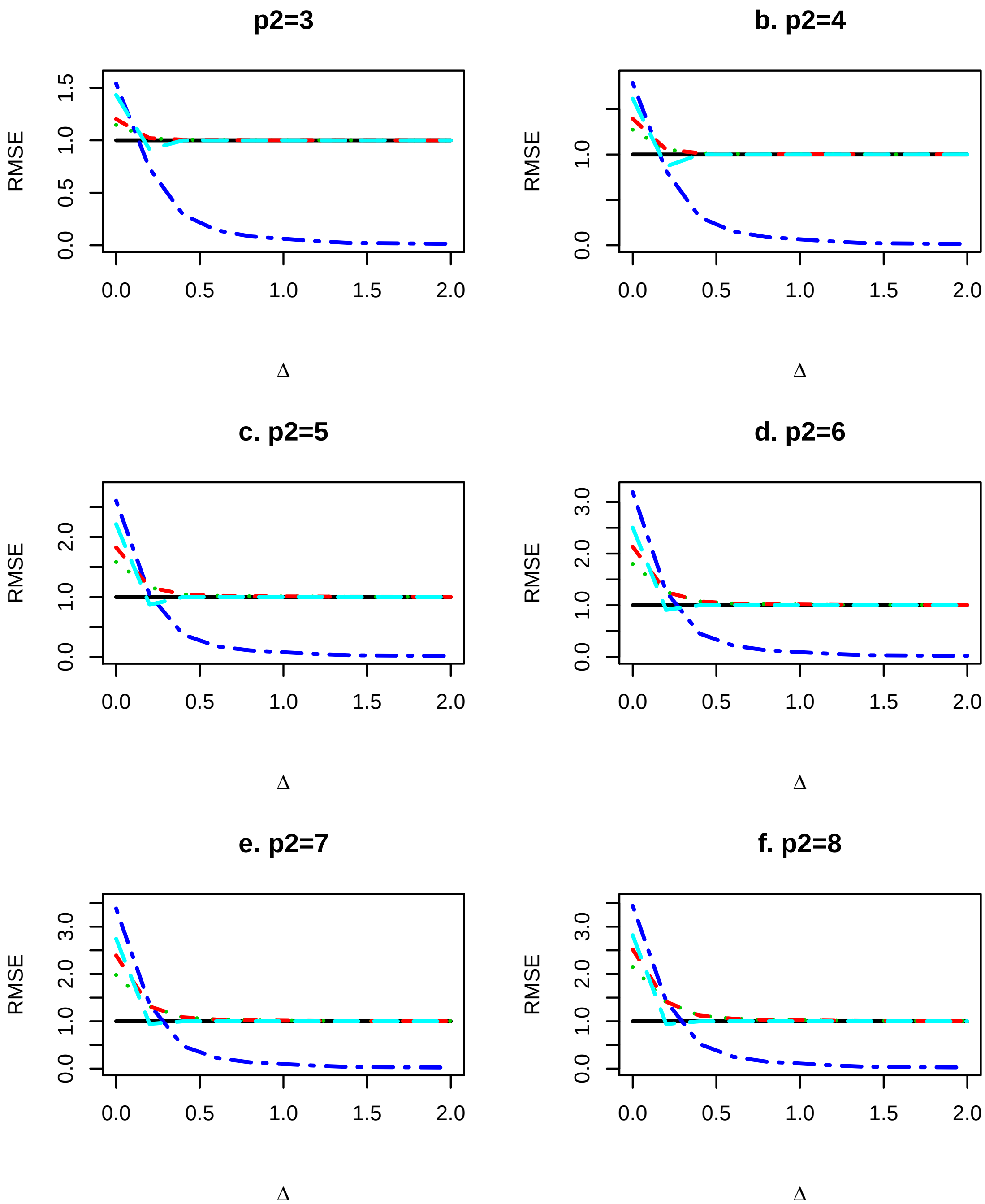

Figure 4.4: Relative MSE of the estimators for various p_2 when $p_1 = 5$ and $n = 100$. “- - -” denotes the PSE, “· · ·” denotes the SE, “– · – ·” denotes the RE, “—” denotes the UE, and “– – –” denotes the PTE.

4.5 Concluding Remarks

In this chapter, we suggested an estimation strategy for QL models. We proposed shrinkage estimators, explored the risk properties of these estimators, and compared the performance of these estimators with the UE, RE, and the lasso estimator via simulation. The simulation results indicated that the RE dominates the other estimators under a correctly specified model. Numerical results demonstrated that the lasso estimator is better than the SE and the PSE when there is a large number of predictors in the model and when only a few of them are irrelevant. On the other hand, the shrinkage estimators perform well when p and the number of nuisance parameters p_2 are relatively large. We demonstrated that, based on both analytical and numerical findings, the PSE outperforms the UE and SE in the entire parameter space. When the restriction is true, the RE is superior to all the other estimation rules; however, its MSE may become unbounded when such restrictions are incorrect.

Chapter 5

Conclusions and Future Research

In this dissertation, we studied different estimation strategies for linear and partially linear models with first order random coefficient autoregressive errors (RCAR(1)) and quasi-likelihood models.

The following estimation procedures are discussed in this dissertation

(i) Application and comparison of shrinkage and absolute penalty estimation in multiple linear regression model with RCAR(1) errors.

(ii) Application and comparison of shrinkage and absolute penalty estimation in partially linear models with RCAR(1) errors using kernel function.

(iii) Shrinkage, pretest and absolute penalty estimation in quasi-likelihood models.

We applied the above estimation procedures to improve the performance of existing estimators when non-sample information is available. The shrinkage estimators perform uniformly better than the unrestricted estimator. The estimator produced by the pretest procedure is superior to the estimators based on sample data only in some part of the parameter space induced by non-sample prior information. The absolute penalty estimators perform better than the shrinkage estimators when the number of restriction on the parameter space is small.

The weighted quadratic loss function was used to calculate the risk. The relative mean square error was used as a criterion for comparison of the performance of the proposed shrinkage estimators. The dominance of proposed shrinkage estimators over the unrestricted estimator is investigated analytically and computationally. In the following we summarize our findings:

We divided Chapter 2 into two parts, namely, low dimensional and high dimensional data problems. In the first part of this chapter, we considered the high dimensional case, i.e., when $n < p$, and we proposed absolute penalty estimators which are the modified version of lasso and adaptive lasso estimation technique for the correlated data. We conducted Monte Carlo simulation studies for different scenarios and compared their performances based on the simulated relative mean square error of the estimators. In all the situations adaptive lasso estimates showed superior performance over the lasso estimates.

In the second part of Chapter 2 we considered the low dimensional case, i.e., when $n > p$, and proposed shrinkage estimation strategy. We investigated statistical properties of these estimators analytically and numerically. The simulations results support our theoretical findings. Based on relative mean square error, our simulation study concluded that the shrinkage and positive shrinkage estimators outperform the classical unrestricted estimator. We also compared the performance of shrinkage estimators with lasso and adaptive lasso numerically. The numerical results showed that APEs perform well when the number of parameters p_2 in the nuisance parameter vector $\boldsymbol{\beta}_2$ is small relative to p_1, but the shrinkage estimators perform best when p_2 is large relative to p_1. For all p_2, the positive shrinkage estimator dominates the shrinkage estimator and they both perform well relative to the classical unrestricted estimator. However, when the restriction is correctly specified, the restricted estimator is the best, but as it departs from the restriction, the risk of the restricted estimator increases and becomes unbounded.

In Chapter 3, we compared the performance of shrinkage, positive shrinkage, absolute penalty-type and weighted semiparametric least squares estimator in the context of partially linear models with RCAR(1) errors. A kernel-based method was to estimate the nonparametric component in the model. This work is an extension of Ahmed et al. (2007). A numerical example based on real life data is used for illustration of proposed estimators. The risk performance of the estimators is investigated through asymptotic distributional risk and Monte Carlo experiments. We found that shrinkage estimators outperform the full model estimator uniformly. The lasso-type estimator performs well when the number of parameters p_2 in the nuisance parameter vector $\boldsymbol{\beta}_2$ is small relative to p_1, but the shrinkage estimators perform best when p_2 is large. For all p_2, the positive shrinkage estimator dominates the usual shrinkage estimator and they both perform better than the classical full model weighted semiparametric least squares estimator in the entire parameter space. On the other hand the performance of the restricted estimator heavily depends on the quality of the UPI.

In Chapter 4 we considered shrinkage, pretest and absolute penalty estimators of parameter $\boldsymbol{\beta}$ for quasi-likelihood models. It is concluded through numerical simulation that the positive shrinkage estimator dominates the usual shrinkage estimator and they both dominate the unrestricted maximum quasi-likelihood estimator in terms of asymptotic distributional risk in the entire parameter space. Under the null hypothesis, the pretest estimator dominates the shrinkage estimators and absolute penalty estimator. However, beyond small intervals near the null hypothesis, the shrinkage estimators dominate the pretest estimator and the risk of the pretest estimator keeps increasing, crosses the risk of unrestricted maximum quasi-likelihood estimator, reaches a maximum, then decreases monotonically to the risk of the unrestricted maximum quasi-likelihood estimator. The absolute penalty estimator performs better than the shrinkage and pretest estimators when the number of restrictions on the parameter space is small and the opposite conclusion holds when it is large.

Future Research

There are possibilities of extending our works in the following ways. In Chapter 2, we compared shrinkage estimation strategies with absolute penalty estimators (APEs), such as, lasso and adaptive lasso. We found that the shrinkage estimators perform better than APEs when the number of nuisance variable (p_2) in the model is high compared to the number of main effects (p_1). This study can be extended to investigate if there exists a ratio of p_2 to p_1 when shrinkage estimators outperform APEs uniformly.

For our future research in Chapter 3, we will consider comparing the performance of kernel with B-spline and penalized spline (P-spline) to estimate the nonparametric part of the model and propose spline based shrinkage estimators when errors are RCAR(1). Also we can consider other types of APEs such as: minimax concave penalty (MCP), smoothly clipped absolute deviation (SCAD) and penalized linear unbiased selection (PLUS) algorithms for comparison purposes with shrinkage estimators.

In Chapter 4, we considered a QL model with independent observations. We can extend this work to a QL model with dependent observations. We can also consider extended quasi-likelihood (EQL), hierarchical likelihood (HL) and double extended quasi-likelihood (DEQL) models to propose pretest and shrinkage estimators.

Bibliography

Ahmed, S. E. (1991). To pool or not to pool: The discrete data. *Statistics and Probability Letters*, 11:233–237.

Ahmed, S. E. (1997). Asymptotic shrinkage estimation: the regression case. *Applied Statistical Science II*, pages 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*, Springer-Verlag: New York, 148:103–120. Springer–Verlag, New York.

Ahmed, S. E., Doksum, K. A., Hossain, S., and You, J. (2007). Shrinkage, pretest and absolute penalty estimators in partially linear models. *Australian & New Zealand Journal of Statistics*, 49:435–454.

Ahmed, S. E., Hussein, A., and Nkurunziza, S. (2010). Robust inference strategy in the presence of measurements error. *Statistics and Probability Letters*, 80:726–732.

Ahmed, S. E. and Krzanowski, W. J. (2004). Biased estimation in a simple multivariate regression model. *Computational Statistics and Data Analysis*, 45:689–696.

Ahmed, S. E. and Liu, S. (2009). Asymptotic theory of simultaneous estimation of Poisson means. *Linear Algebra and its Applications*, 430:2734–2748.

Ahmed, S. E. and Raheem, S. E. (2012). Shrinkage and absolute penalty estimation in linear models. *WIREs Computational Statistics*, to appear.

Akharif, A. and Hallin, M. (2003). Efficient detection of random coefficients in autoregressive models. *The Annals of Statistics*, 31:675–704.

Albertson, K. V. (1993). Pre-test estimation in a regression model with a misspecified error covariance matrix. Department of Economics, University of Canterbury, New Zealand.

Alzghool, R., Lin, Y. X., and Chen, S. X. (2010). Asymptotic quasi-likelihood based on kernel smoothing for multivariate heteroskedastic models with correlation. *American Journal of Mathematical and Management Sciences*, 30:147–177.

Aneiros-Pérez, G., González-Manteiga, W., and Vieu, P. (2004). Estimation and testing in a partial linear regression model under long-memory dependence. *Bernoulli*, 10:49–78.

Aneiros-Pérez, G. and Quintela, A. (2002). Plug-in bandwidth choice in partial linear models with autoregressive errors. *Journal of Statistical Planning and Inference*, 100:23–48.

Aue, A. and Horváth, L. (2011). Quasi-likelihood estimation in stationary and nonstationary autoregressive models with random coefficients. *Statistica Sinica*, 21:973–999.

Aue, A., Horváth, L., and Steinebach, J. (2006). Estimation in random coefficient autoregressive models. *Journal of Time Series Analysis*, 27:61–76.

Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary test of significance. *The Annals of Mathematical Statistics*, 15:190–240.

Bera, A. K. and Zuo, X. L. (1996). Specification test for a linear regression model with ARCH process. *Journal of Statistical Planning and Inference*, 50:283308.

Bowman, A. W. and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: The Kernel Approach with S-Plus Illustrations*, Oxford: Oxford University Press.

Chen, H. (1988). Convergence rates for parametric components in a partially linear model. *The Annals of Statistics*, 16:136–147.

Chen, H. and Shiau, J. (1991). A two-stage spline smoothing method for partially linear models. *Journal of Statistics Planning and Inference*, 27:187–202.

Donald, G. and Newey, K. (1994). Series estimation of semilinear models. *Journal of Multivariate Analysis*, 50:30–40.

Dutta, H. (1999). Large sample tests for a regression model with autoregressive conditional heteroscedastic errors. *Communications in Statistics: Theory & Methods*, 28:105117.

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

Eubank, R. and Speckman, P. (1990). Trigonometric series regression estimators with an application to partially linear models. *Journal of Multivariate Analysis*, 32:7085.

Fallahpour, S., Ahmed, S. E., and Doksum, K. A. (2012). L_1 penalty and shrinkage estimation in partially linear models with random coefficient autoregressive errors. *Journal of Applied Stochastic Models in Business and Industry*, 28:236–250.

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

Firth, D. (1987). On the efficiency of quasi-likelihood estimation. *Biometrika*, 74:233–245.

Friedman, J., Hastie, T., Höfling, H., and Tibshirani, R. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1:302–332.

Friedman, J., Hastie, T., and Tibshirani, R. (2009). *glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models*, R package version 1.1-4, URL http://CRAN.R-project.org/package=glmnet.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33:1–22.

Fu, W. J. (1998). Penalized regressions: The bridge versus the lasso. *Journal of Computational and Graphical Statistics*, 7:397–416.

Gao, J. T. (1995). Asymptotic theory for partially linear models. *Communications in Statistics: Theory & Methods*, 24:1985–2009.

Gao, J. T. (1997). Adaptive parametric test in a semiparametric regression model. *Communications in Statistics: Theory & Methods*, 26:787–800.

Giles, J. A. and Giles, D. E. A. (1993). Preliminary test estimation of the regression scale parameter when the loss function is asymmetric. *Communications in Statistics: Theory & Methods*, 22:1709–1734.

Godambe, V. P. and Thompson, M. E. (1989). An extension of quasi-likelihood estimation. *Journal of Statistical Planning and Inference*, 22:137–152.

González-Manteiga, W. and Aneiros-Pérez, G. (2003). Testing in partial linear regression models with dependent errors. *Journal of Nonparametric Statistics*, 15:93–111.

Hamilton, A. and Truong, K. (1997). Local linear estimation in partially linear models. *Journal of Multivariate Analysis*, 60:1–19.

Härdle, W., Liang, H., and Gao, J. (2000). *Partially Linear Models*, Physica-Verlag, Heidelberg.

Hastie, T. and Efron, B. (2012). *lars: least angle regression, lasso and forward stagewise*, R package version 1.1, URL http://CRAN.R-project.org/package=lars.

Heckman, N. (1986). Spline smoothing in a partially linear model. *Journal of the Royal Statistical Society: Series B*, 48:244–248.

Hesterberg, T., Choi, N. H., Meier, L., and Fraley, C. (2008). Least angle and ℓ_1 penalized regression: A review. *Statistics Surveys*, 2:61–93.

Heyde, C. C. (1997). *Quasi-Likelihood And Its Application: A General Approach to Optimal Parameter Estimation*, Springer-Verlag, New York.

Horowitz, J. L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*, Springer, New York.

Hwang, S. Y. and Basawa, I. V. (1993). Parameter estimation in a regression model with random coefficient autoregressive errors. *Journal of Statistics Planning and Inference*, 36:57–67.

Hwang, S. Y. and Basawa, I. V. (1997). The local asymptotic normality of a class of generalized random coefficient autoregressive processes. *Statistics and Probability Letters*, 34:165–170.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berekley, CA.

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 Implication of Pre-test and Stein-rule Estimators in Econometrics*, North-Holland, Amsterdam.

Klein, H. R., Ball, R. J., Hazlewood, A., and Vandome, P. (1961). *An Econometric Model of the United Kingdom*, Oxford: Blackwell.

Kraemer, N. and Schaefer, J. (2010). *parcor: Regularized estimation of partial correlation matrices.*, R package version 0.2-2, URL http://CRAN.R-project.org/package=parcor.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society: Series B*, 58:619–678.

Lee, Y. and Nelder, J. A. (2001). Hierarchical generalized linear models: a synthesis of generalized linear models, random effect models and structured dispersions. *Biometrika*, 88:987–1006.

Lesnoff, M. and Lancelot, R. (2012). *aod: Analysis of Over-dispersed Data*, R package version 1.1-4, URL http://CRAN.R-project.org/package=aod.

Liang, H. and Härdle, W. (1999). Large sample theory of the estimation of the error distribution for a semiparametric model. *Communications in Statistics: Theory & Methods*, 28:2025–2037.

Lin, P. S., Lee, H. Y., and Clayton, M. (2009). A comparison of efficiencies between quasi-likelihood and pseudo-likelihood estimates in non-separable spatial-temporal binary data. *Journal of Statistical Planning and Inference*, 139:3310–3318.

Liu, L. and Tiao, G. C. (1980). Random coefficient first-order autoregressive models. *Annals of Econometrics*, 13:305–325.

McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics*, 11:59–67.

Nelder, J. A. and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, 74:221–232.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series B*, 58:619–678.

Nicholls, D. F. and Quinn, B. G. (1982). Random coefficient autoregressive models: An introduction. *Lecture Notes in Statistics*, 11:Springer–Verlag: New York.

Park, M. Y. and Hastie, T. (2007). An l_1 regularization-path algorithm for generalized linear models. *Journal of the Royal Statistical Society: Series B*, 69:659–677.

Priestley, M. B. and Chao, M. T. (1972). Non-parametric function fitting. *Journal of the Royal Statistical Society: Series B*, 34:385392.

R Development Core Team (2011). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. ISBN 3-900051-07-0, url http://www.r-project.org/.

Reinsel, G. C. and Velu, R. P. (1998). Multivariate reduced-rank regression: theory and applications. *Lecture Notes in Statistics*, 136:Springer–Verlag, New York.

Rice, J. (1986). Convergence rates for partially splined models. *Statistics & Probability Letters*, 4:203–208.

Robinson, P. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 56:931–954.

Rosset, S. and Zhu, J. (2007). Piecewise linear regularized solution paths. *The Annals of Statistics*, 35:1012–1030.

Saleh, A. K. M. E. (2006). *Theory of Preliminary Test and Stein-Type Estimation with Applications*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Schick, A. (1994). Estimation of the autocorrelation coefficient in the presence of a regression trend. *Statistics & Probability Letters*, 21:371–380.

Schick, A. (1996). Efficient estimation in a semiparametric additive regression model with autoregressive errors. *Stochastic Processes and their Applications*, 61:339–361.

Schick, A. (1998). An adaptive estimator of the autocorrelation coefficient in regression models with autoregressive errors. *Journal of Time Series Analysis*, 19:575–589.

Singpurwalla, N. D. and Soyer, R. (1985). Assessing (software) reliability growth using a random coefficient autoregressive process and its remifications. *IEEE Transactions on Software Engineering*, 11:1456–1465.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society: Series B*, 50:413–437.

Stein, C. (1956). Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. *Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berekley, CA:197–206.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58:267–288.

Tong, H. (1990). *Non-linear Time Series*, Oxford University Press: Oxford:96–121.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61:439–447.

Weiss, A. A. (1986). Asymptotic theory for arch models: estimation and testing. *Econometric Theoy*, 2:107–131.

Wu, T. T. and Lange, K. (2008). Coordinate descent algorithms for lasso penalized regression. *The Annals of Applied Statistics*, 1:224–244.

You, J. and Chen, G. (2002). Parameter estimation on a partly linear regression model with random coefficient autoregressive errors. *Communications in Statistics: Theory & Methods*, 31:1137–1158.

You, J. and Chen, G. (2007). Semiparametric generalized least squares estimation in partially linear regression models with correlated errors. *Journal of Statistical Planning and Inference*, 137:117–132.

Zhang, C. M., J. Y. and Chai, Y. (2010). Penalized bergman divergence for large dimensional regression and classification. *Biometrika*, 97:551–566.

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

Vita Auctoris

NAME:	Saber Fallahpour
PLACE OF BIRTH:	Iran
EDUCATION	Allameh Tabatabai University, Tehran, Iran
	2000-2004 B.Sc.
	Allameh Tabatabai University, Tehran, Iran
	2004-2007 M.Sc.
	University of Windsor, Windsor, Canada
	2008-2012 Ph.D.