2013

Shrinkage and Penalty Estimation Strategies in Some Spatial Models

Marwan Al-Momani
University of Windsor

SHRINKAGE AND PENALTY ESTIMATION STRATEGIES IN SOME SPATIAL MODELS

by

Marwan Al-Momani

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

Shrinkage and Penalty Estimation Strategies in Some Spatial Models

by

Marwan Al-Momani

APPROVED BY:

Dr. S. Chenouri, External Examiner
University of Waterloo

Dr. R. Kent
School of Computer Science

Dr. M. Hlynka
Department of Mathematics and Statistics

Dr. S. Nkurunziza
Department of Mathematics and Statistics

Dr. A.A. Hussein, Co-Advisor
Department of Mathematics and Statistics

Dr. S. E. Ahmed, Co-Advisor
Brock University

12 September 2013

Declaration of Co-Authorship/ Previous Publication

I hereby declare that this thesis incorporates the outcome of joint research undertaken in collaboration with my supervisors, Professor S. Ejaz Ahmed and Abdulkadir Hussein. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contribution of co-authors 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.

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 study the asymptotic properties of pretest and shrinkage estimators of the large-scale effect in some spatial regression models, and compare their relative performance with respect to the classical *maximum likelihood estimator* (MLE) analytically and numerically through Monte Carlo experiments and real data examples. The shrinkage estimators were also numerically compared with three penalty estimators, namely, the LASSO, adaptive LASSO, and the SCAD penalty functions.

A linear model with conditional autoregressive errors was studied in Chapter 2. The asymptotic properties of the shrinkage estimators, under local alternatives, were established, including the derivations of the asymptotic distributional bias, asymptotic mean squared error matrix, and the asymptotic quadratic risk. These results showed the effectiveness of the suggested estimation technique. Monte Carlo experiments with two real data examples were conducted to demonstrate the superiority of the proposed shrinkage estimators over the MLE and the penalty estimators.

In Chapter 3, we consider another spatial case of a linear model with simultaneous autoregressive errors. We study the properties of the shrinkage estimators and compare their performance with the penalty estimators numerically through simulation studies and real data examples.

Chapter 4 contains a study of a general linear model with spatial moving average error terms. Asymptotic properties of the shrinkage estimators for the mean parameter vector are investigated. A numerical comparison is carried out and the relative performance of estimators is investigated.

Finally, we summarize the findings of the thesis in Chapter 5. Also, some problems for future research are outlined in Chapter 5.

Acknowledgements

All glorification and gratitude to Almighty who has given me the ability to accomplish this dissertation.

I sincerely thank my supervisor, Prof. S.E. Ahmed for his guidance, help, continuous encouragement and financial support. I am grateful to him for his advice and deep look during my study without which it would not have been possible to finish my thesis in time.

I would also like to thank my co-supervisor, Dr. A. A. Hussein for his guidance, caring and providing me with a solution to all problems I had faced in the numerical studies as well as some analytical results during the preparation of this work. I am thankful to him for his friendship and the financial support that I had got as a Research Assistantship (RA) many times during my doctoral study.

My thanks also go to the external examiner Dr. S. Chenouri and the advisory committee members, Prof. Myron Hlynka, Dr. S. Nkurunziza and Prof. Robert Kent for reading the dissertation and providing with useful suggestions and comments which have improved it. I am grateful to Dr. S. Nkurunziza for his financial support and the valuable mathematical suggestions in the proofs of the theorems in this dissertation.

My thanks extended to all faculty, staff members, and the graduate students in Department of Mathematics and Statistics. Also, I would like to acknowledge the University of Windsor for financially supporting my PhD study.

I must also thank my parents for their support, patience, love and prayers that have not stopped during all my life. All words of gratitude are unable to acknowledge their real contribution.

No achievement without sacrifice. My family has sacrificed the most. In spite of many restrictions, stress and hardship during my study, the love, sense, emotion and the continuous encouragement from my wife Lara Al Momani kept me on track. Big love and adore to my two little angles, my daughter Nour, and son Malik for giveing me joyful company in my busy graduate study time.

Marwan Al-Momani
September 12, 2013
Windsor, Ontario, Canada.

Contents

List of Figures

List of Tables

Abbreviations

ADB	Asymptotic distributional bias
AIC	Akaike information criterion
AMSEM	Asymptotic mean squared error matrix
AQR	Asymptotic quadratic risk
BIC	Bayesian information criterion
CAR	Conditional autoregressive
DCAR	Directional conditional autoregressive
iid	Identically and independently distributed
LARS	Least angle regression
LASSO	Least absolute shrinkage and selection operator
MLE	Maximum likelihood estimator
MSPE	Mean squared prediction error
NSI	Non-sample information
RMSPE	Relative efficiency of mean squared prediction error
SAR	Simultaneous autoregressive
SARMA	Spatial autoregressive moving average
SARMAX	Spatial autoregressive moving average regression
SCAD	Smoothly clipped absolute deviation
SMA	Spatial moving average
SMSE	Simulated mean squared error
SRE	Simulated relative efficiency
STCAR	Spatial temporal conditional autoregressive
UPI	Uncertain prior information

List of Symbols

$\boldsymbol{\beta}$	Large-Scale effect regression parameter vector
$\hat{\boldsymbol{\beta}}$	Unrestricted maximum likelihood estimator
$\hat{\boldsymbol{\beta}}^{R}$	Restricted estimator
$\hat{\boldsymbol{\beta}}^{PT}$	Pretest estimator
$\hat{\boldsymbol{\beta}}^{JS}$	James-Stein (Shrinkage) estimator
$\hat{\boldsymbol{\beta}}^{JS+}$	Positive shrinkage estimator
$\hat{\boldsymbol{\beta}}^{Lasso}$	LASSO estimator
$\hat{\boldsymbol{\beta}}^{A.Lasso}$	Adaptive LASSO estimator
$\hat{\boldsymbol{\beta}}^{SCAD}$	SCAD estimator
$ch_{min}(\boldsymbol{A})$	Smallest characteristic root of the matrix $\boldsymbol{A}$
$ch_{max}(\boldsymbol{A})$	Largest characteristic root of the matrix $\boldsymbol{A}$
$diag\{a_i\}$	Diagonal matrix with a_i in the main diagonal
Δ^2	Non-centrality parameter
$\boldsymbol{I}$	Identity matrix
$I(.)$	Indicator function
$\boldsymbol{I}(\boldsymbol{\beta})$	Expected Fisher information matrix
$G(x)$	Non degenerate asymptotic distribution function of x
$\boldsymbol{H}$	A $(q \times p)$ known matrix of rank q
$\boldsymbol{A}_0$	Null hypothesis
$\boldsymbol{A}_{(n)}$	A class of local alternatives
$H_q(a, \Delta^2)$	Non-central chi-square distribution function with $q-$degrees of freedom and non–centrality parameter Δ^2

$\boldsymbol{h}$	A $(q \times 1)$ vector of known constants
$L(.,.)$	Quadratic loss function
$\boldsymbol{M}$	A $(p \times p)$ positive definite matrix
n	Sample size
N	Grid size
s_i	A site or location
$\boldsymbol{s}$	A set of sites (locations)
$N(s_i)$	Neighborhood set of the site s_i
p	Number of parameters in the $\boldsymbol{\beta}$ vector under the full model
q	Number of parameters in the $\boldsymbol{\beta}$ vector under the submodel
$\mathbb{R}^d$	A $d-$dimensional real valued vector
$tr(\boldsymbol{A})$	Trace of the matrix $\boldsymbol{A}$
$cov(.,.)$	Covariance function
$var(.)$	Variance function
$\boldsymbol{W}$	Weighted adjacency, spatial neighborhood or proximity matrix
$\boldsymbol{W}^*$	Standardized proximity matrix
λ	Tuning parameter
Υ	Test statistic
$\boldsymbol{\xi}$	A $(q \times 1)$ fixed vector in $\mathbb{R}^q$
$\boldsymbol{\Sigma_Y}$	Variance covariance matrix of $\boldsymbol{Y}$
$\Vert.\Vert$	Euclidean norm

Chapter 1

Introduction and Literature Review

1.1 Introduction

In real life, we often need to make inferences about the dynamics of natural phenomena. In order to carry out a sensible inference, one would collect sample data, which we can call objective information, as well as subjective non-sample information about the natural phenomena of interest. A statistician's job is to find a model that is best consistent with the information provided and utilize the model for making inferences and predictions about the behavior of the phenomena. The subjective non-sample information is often called *uncertain prior information* (UPI). It is well known that Bayesian statistical methods were originated from the need of injecting UPIs into models fitted to the objective sample data and hence, account for the uncertainty brought in by both streams of information.

In the frequentist (non-Bayesian) literature, one of the earliest attempts to incorporate UPIs into the estimation of regression coefficients was made by Bancroft (1944) who proposed an estimation technique known as the pretest estimation method. Bancroft's idea was that, if the UPI states that some of the regression coefficients should be zero, then we can incorporate such information into the estimation procedure by testing the null hypothesis, implicit in the UPI, and according to the results of the test we choose either the full model containing all coefficients or the reduced model stated by the null hypothesis. Such a method, obviously, combines the data model (full model) and the UPI-based model (reduced) via binary weights.

Stein (1956) improved Bancroft's idea by suggesting the use of smooth weights in combining the two models, instead of the binary weights of Bancroft. The method of Stein was labeled as a shrinkage approach in the sense that one shrinks the least squares regression coefficient estimators towards a target value dictated by the UPI. The seminal work of Stein was then followed by Stein (1966) who proposed an improved version of Stein's estimator, known as positive part shrinkage estimator. These fundamental works opened the way to a large body of literature and development of shrinkage estimation in many diverse areas of statistical analysis.

Ahmed (1997b) developed the asymptotic properties of the positive part shrinkage, improved preliminary test, and shrinkage preliminary test R-estimators of regression coefficients when the errors are not necessarily normally distributed. Ahmed (1998) investigated the pretest estimators in nonparametric multivariate regression models. Khan and Ahmed (2003) considered the estimation of the regression coefficient vector when it is suspected that it may belong to a subspace given by a set of consistent equations. Ahmed et al. (2006) considered shrinkage M-estimators in linear models. For more details about the shrinkage estimators, the reader is referred to Ahmed

(1997a), Ahmed (2001), Saleh (2006), Ahmed et al. (2007), Nkurunziza and Ahmed (2011), Nkurunziza (2011), Raheem et al. (2012), Nkurunziza (2012b) and Fallahpour et al. (2012), among others.

Shrinkage estimation procedures are closely related to the so called model selection procedures. Recently, there has been a growing interest in a class of model selection and estimation procedures known as penalty estimation methods. These methods are based on constrained maximization of an objective function such as the log-likelihood function where the constraint is defined through a penalty function on the absolute value of the model parameters. These methods have become popular because of their ability to estimate parameters in high dimensional models where the number of parameters exceeds the available number of observations. Model selection procedures can thus serve as tools to find UPIs which then can be incorporated into the estimation procedure through shrinkage procedures.

The main objective of this dissertation is to propose shrinkage and penalty estimators in the context of spatial regression models. Therefore, in this chapter, we will give a brief review of shrinkage and penalty estimation procedures as well as an account of the spatial regression models to be considered in the thesis.

1.2 Efficient Estimation Strategies

To have a clear picture of the shrinkage estimation technique, we illustrate the case of a multiple linear regression model with *identically and independently distributed*

(iid) errors. The model in a matrix format can be written as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.1}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ is a $(n \times 1)$ response vector, $\boldsymbol{X} = \big(X_1, \ldots, X_p\big)'$ is a full rank $(n \times p)$ non-random design matrix consisting of the predictors $\big(X_1, \ldots, X_p\big)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a $(p \times 1)$ unknown vector of regression coefficients without an intercept, and $\boldsymbol{\epsilon} = \big(\epsilon_1, \ldots, \epsilon_n\big)'$ is a $(n \times 1)$ vector of random errors with $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

1.2.1 Full Model Estimation

The least squares and the maximum likelihood estimators of the parameter vector $\boldsymbol{\beta}$ for the model in (1.1) coincide and have the following form:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \tag{1.2}$$

It is well known that the maximum likelihood estimator MLE of $\boldsymbol{\beta}$ is the *best linear unbiased estimator* (BLUE), and it is normally distributed with mean $\boldsymbol{\beta}$ and variance-covariance matrix $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$. The MLE represents the *unrestricted estimator* (UE) of $\boldsymbol{\beta}$, in the sense that all possible and available predictors are included in the model, and the corresponding model is labeled as the full regression model. On the other hand, if a UPI is available, say in the form of a hypothesis stating that there is a set of linear restrictions on these coefficients, then the resulting estimated coefficients under such UPI are known as *restricted estimators* (RE) of the regression coefficients.

1.2.2 Uncertain Prior Information

Suppose that the restriction is formulated in the form of the null hypothesis:

$$H_0 : \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}, \tag{1.3}$$

where $\boldsymbol{H}$ is a $(q \times p)$ known matrix of rank (q) with $(q \leq p)$, $\boldsymbol{h}$ is a $(q \times 1)$ vector of known constants. For instance, if an investigator contemplated a UPI that the vector of regression coefficients can be partitioned into two parts, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, and that it is safe to conjecture $\boldsymbol{H_0} : \boldsymbol{\beta}_2 = \mathbf{0}$, then in this case $\boldsymbol{h} = \mathbf{0}$ while $\boldsymbol{H}$ is chosen as block matrix with zero everywhere except an identity matrix in the block corresponding to the component $\boldsymbol{\beta}_2$.

1.2.3 Submodel Estimation

Based on the restriction given by (1.3), the restricted estimator of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}^R$, is given by:

$$\hat{\boldsymbol{\beta}}^R = \hat{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}' \left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1} (\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}). \tag{1.4}$$

Obviously, the quality of $\hat{\boldsymbol{\beta}}^R$ depends on the quality of the UPI presented in (1.3). If UPI is correct then $\hat{\boldsymbol{\beta}}^R$ has better performance than the unrestricted estimator $\hat{\boldsymbol{\beta}}$, while such relative performance deteriorates as the quality of the UPI deteriorates.

1.2.4 Pretest Estimator

The pretest estimator, denoted by $\hat{\boldsymbol{\beta}}^{PT}$, is produced by combining the unrestricted with the restricted estimators as follows:

$$\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R}) I(\Upsilon \leq \Upsilon_{n,\alpha}). \tag{1.5}$$

Here, Υ is an appropriate test statistic for testing the null hypothesis in (1.3), $I(.)$ is an indicator function such that $I(A) = 1$ if the statement A is true and zero otherwise, and $\Upsilon_{n,\alpha}$ is the $\alpha-$level critical value of the exact distribution of the test statistics Υ.

The pretest estimator depends on the level of significance, and by the definition of the indicator function, it chooses the unrestricted estimator for large values of the test statistic, and the restricted estimator for small values. The performance of this estimator is considerable when the information provided by the null hypothesis is true, or approximately true.

1.2.5 Shrinkage Estimators

The pretest estimator produces either the unrestricted or the restricted estimator depending on whether the test statistic is above or below the $\alpha-$level critical value. A smoother weighting can be achieved via the James-Stein estimator, defined as follows:

$$\hat{\boldsymbol{\beta}}^{JS} = \hat{\boldsymbol{\beta}}^{R} + \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R}\right)\left\{1 - (q-2)\Upsilon^{-1}\right\}, \quad q \geq 3. \tag{1.6}$$

Thus, $\hat{\boldsymbol{\beta}}^{JS}$ overcomes the problem of binary choice inherent in the pretest estimator. However, $\hat{\boldsymbol{\beta}}^{JS}$ may suffer from a phenomenon known as over-shrinkage, whereby negative coordinates of $\hat{\boldsymbol{\beta}}$ are obtained whenever $(\Upsilon < q-2)$. This problem is solved by excluding the following modified version of the original Stein estimator,

$$\hat{\boldsymbol{\beta}}^{JS+} = \hat{\boldsymbol{\beta}} + \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R}\right)\left\{1 - (q-2)\Upsilon^{-1}\right\}^{+}, \tag{1.7}$$

where $a^{+} = \max\{0, a\}$. This estimator is known as the positive part James-Stein estimator.

Therefore, it is hopefully clear that the pretest and shrinkage methodologies provide a middle way between these two options, by combining $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{R}$ through appropriately chosen weights that are functions of the quality of the UPI.

1.3 Auxiliary Information

As we have seen in the previous sections, the class of shrinkage estimators is useful in incorporating UPIs into the estimation process. These estimators have demonstrated superior performance in large classes of statistical models which go beyond the usual regression model considered in the past section. Although the general restriction in (1.3), known also as candidate subspace restriction, accommodates a variety of prior non-sample information, it is sometimes possible that such subjective UPI is not available. In these cases, one could still resort to model selection procedures such as *Akaike's Information Criterion* (AIC), *Bayesian Information Criterion* (BIC) and penalty model selection approaches in order to formulate a candidate submodel which could then be formulated in the form of the restriction (1.3). It is well-known that in addition to giving biased estimators, model selection procedures cannot be taken as gold standard, since the resulting final models also have uncertainty inherited from the selection procedure. That is, the final model selected by a model selection procedure need not be the true data generating model. Therefore, it is still safer to resort to estimation methods which take into account the linear restriction induced by the model selection criteria along with the full model. Thus, here we clearly state that model selection criteria and shrinkage estimation procedures are not rivals, but rather complementary to each other.

Recently, there has been a growing literature on new model selection methods known as penalty estimation or selection methods. These methods are based on imposing a penalty on the model parameters. In the next few sections, we will review some of these penalty methods which are key in this thesis.

1.4 Penalty Estimators

Penalty estimators result from simultaneous model selection and parameter estimation procedures via imposition of penalty on the estimating equation used. Therefore, penalty methods are both estimation and model selection procedures. The first such method, known as *least absolute selection and shrinkage operator* (LASSO) was proposed by Tibshirani (1996) to overcome the problem of large dimensional data regressions where the number of parameters in the model exceeds the available number of independent observations, i.e., $n < p$. Such situations arise in many applications including micro-array data analysis in which the LASSO has demonstrated to give reasonably well-behaved estimators. A large body of literature following the work of Tibshirani was dedicated to improving the LASSO procedure. Among others, we mention here the *adaptive* LASSO (A.LASSO) of Zou (2006), and the *smoothly clipped absolute deviation* (SCAD) method of Fan and Li (2001). These three methods will be described in the next few sections.

1.4.1 LASSO

Tibshirani (1996) proposed a method for variable selection and parameter estimation in linear models known as LASSO. The LASSO algorithm uses the L_1- norm of the $\boldsymbol{\beta}$ vector in order to define a penalty term in the usual least squares estimation of regression coefficients. The LASSO estimators are defined as:

$$\hat{\boldsymbol{\beta}}^{Lasso} = \arg\min_{\boldsymbol{\beta}} \left[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\boldsymbol{\beta}_j| \right], \tag{1.8}$$

where $\boldsymbol{\beta}$ is assumed as a $(p \times 1)$ vector of regression coefficients, and $\lambda \geq 0$ is a tuning parameter. Here, $\arg\min_{\boldsymbol{\beta}}[.]$, stands for the argument of the minimum, that is the set of all points $\boldsymbol{\beta}$ for which the expression [.] attains minimum value. An efficient algorithm for calculating the LASSO estimators and computing an optimal value of the tuning parameter, known as the *least angle regression* (LARS), was introduced by Efron et al. (2004). Although the LASSO method was appealing, it had several shortcomings, one of which is that it does not enjoy a desirable property called the *Oracle property*. A variable selection procedure is said to have oracle property if it identifies the right subset of zero coefficients in the regression model under consideration and furthermore, the estimators of the remaining non-zero coefficients are consistent and asymptotically normal (Zou (2006)). Two procedures which enjoy the oracle property were introduced by Fan and Li (2001) and Zou (2006). In the next two sections, we define these procedures.

1.4.2 SCAD

The *smoothly clipped absolute deviation* (SCAD) estimator of Fan and Li (2001), which is an improved version of the LASSO, is defined by

$$\hat{\boldsymbol{\beta}}^{SCAD} = \arg\min_{\boldsymbol{\beta}} \left[\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta}) + n\sum_{j=1}^{p} P_{\lambda_S}(|\boldsymbol{\beta}_j|) \right], \tag{1.9}$$

where,

$$P_{\lambda_S}(t) = \begin{cases} \lambda_S|t| & , |t| \leq \lambda_S \\ -\frac{(t^2 - 2a\lambda_S|t| + \lambda_S^2)}{2(a-1)} & , \lambda_S < |t| < a\lambda_S \\ \frac{(a+1)\lambda_S^2}{2} & , |t| > a\lambda_S \end{cases}, \tag{1.10}$$

for some $a > 0$ and λ_S, a tuning parameter. Detailed discussions about the SCAD can be found in Leeb and Poetscher (2008) who also studied the distribution of the LASSO and SCAD estimators in both finite and large sample cases.

1.4.3 Adaptive LASSO

The idea behind the adaptive LASSO of Zou (2006) is to incorporate data driven tuning parameters in the original LASSO procedure. The adaptive LASSO estimator of the vector $\boldsymbol{\beta}$ is defined by

$$\hat{\boldsymbol{\beta}}^{A.LASSO} = \arg\min_{\boldsymbol{\beta}} \left[\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta}) + n\sum_{j=1}^{p} \lambda_j|\boldsymbol{\beta}_j| \right], \tag{1.11}$$

where $\{\lambda_j, j = 1, \ldots, p\}$ are coefficient-specific tuning parameters.

Adaptive LASSO estimators have been shown to posses the oracle property and they have been widely discussed, used and extended. Among others, Zhang and Lu (2007) studied adaptive LASSO for the Cox's proportional hazards model. They studied the consistency and rate of convergence of the estimators obtained. Huang et al. (2008) studied the asymptotic properties of the adaptive LASSO estimators in sparse, high-dimensional, linear regression models when the number of regression coefficients may increase with the sample size, and showed that it has the oracle property. Pötscher and Schneider (2009) considered the distribution of the adaptive LASSO in finite samples. Further, we refer to some recent work in this case, Kamarianakis et al. (2012), Guo et al. (2013), Evans and Forcina (2013), Ren and Zhang (2013), Qian and Yang (2013) and Lin et al. (2013).

1.4.4 Penalty and Shrinkage Estimators

Ahmed and Raheem (2012) compared shrinkage estimators in linear regression models to the LASSO, SCAD, and the adaptive LASSO estimators. Their comparison showed a superiority of the shrinkage over the penalty estimators in the sense of giving smaller average prediction errors. They also noted that the penalty estimators outperform shrinkage estimators as the dimension of zero coefficients becomes very large relative to the sample size.

An extension of the LASSO method for the semiparametric partially linear regression model was proposed by Ahmed et al. (2007). This class of LASSO estimators was compared with shrinkage estimators through prediction errors. Fallahpour et al. (2012) studied the LASSO and shrinkage estimation strategies in partially linear models with random coefficient autoregressive errors. Recent literature studying shrinkage and penalty methods and comparing them include Hossain et al. (2009), Raheem et al.

(2012), Ahmed and Fallahpour (2012) and Hossain and Ahmed (2012).

1.5 Spatial Data Regression Models

"*We believe that in order to answer the "why" question, Science should address the "where" and "when" questions...*" . Cressie and Wikle (2011).

The era of isolated marginal analysis of data is almost passing away as we move into a world of complex and massive data, collected in real-time over space and time. The era of conditional thinking has begun and at the frontier of this era is the analysis of spatio-temporal data.

In general, data collected over geographical space may exhibit some sort of dependence in the sense that closer observations are more alike than those far apart. Such behavior is modeled by including a covariance structure into the classical statistical models. In particular, spatial regression models which accommodate various types of spatial dependencies have been increasingly applied in epidemiology, geology, disease surveillance, urban planning, analysis and mapping of poverty indicators and others. In this section, we will give a brief introduction and literature review on three spatial regression models which will be the subject of study in this thesis. These models are: the *conditional autoregressive* (CAR) model, the *simultaneous autoregressive* (SAR) model and the *spatial moving average* (SMA) model.

1.5.1 Conditional Autoregressive Model

In time series analysis, autoregressive models represent the current data at time t (in temporally evolving data) as a linear combination of the most recent observations.

Likewise, in spatial framework, autoregressive models represent the data from a given spatial location as a function of data in neighboring locations. A geographical location on which data are collected is often called a site and the concept of neighborhood among sites is defined through a distance metric. That is, two sites are neighbors if they are close to each other according to a pre-specified closeness metric.

An important class of spatial regression models known as conditional autoregressive CAR, introduced by Besag (1974), exploits neighborhood structures. In order to describe the CAR model, suppose we have a set of n spatial sites denoted by $\boldsymbol{s} = \{s_1, \ldots, s_n\}$ forming a lattice, and suppose that a set of continuous measurements $\boldsymbol{Y}(\boldsymbol{s}) = (Y(s_1), ..., Y(s_n))$ is collected at these sites. Also, denote $N(s_i)$ a set of neighboring sites to the ith site. That is $N(s_i) = \{s_j : \forall j = 1, ..., n | s_j \text{ is neighbor to } s_i\}$. Assuming that the $Y(s_i)$ are Gaussian random processes, we have

$$f\Big(y(s_i)|\{y(s_j), j \neq i\}\Big) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\Big\{\frac{(y(s_i) - \mu_i)^2}{2\sigma_i^2}\Big\}, \tag{1.12}$$

where μ_i and σ_i^2 are respectively, the conditional mean and variance, given by

$$\begin{aligned} \mu_i &= E\Big(Y(s_i)|\{Y(s_j) : j \neq i\}\Big) \\ &= \mu(s_i) + \sum_{s_j \in N(s_i)} c_{ij}(Y(s_j) - \mu(s_j)) \qquad (1.13) \\ \sigma_i^2 &= var\Big(Y(s_i)|\{Y(s_j) : j \neq i\}\Big), \qquad (1.14) \end{aligned}$$

provided that $c_{ij}\sigma_j^2 = c_{ji}\sigma_i^2$, $c_{ii} = 0$ and $c_{ij} = 0$ if $j \notin N(s_i)$, $i, j = 1, \ldots, n$.

Besag (1974) proved that, the Gaussian conditional densities in (1.12) have Gaussian joint distribution given by

$$\boldsymbol{Y} \sim N(\boldsymbol{\mu}, (\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{M}), \tag{1.15}$$

provided that $(\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{M}$ is symmetric and positive definite, where $\boldsymbol{I}$ is the $(n \times n)$ identity matrix, $\boldsymbol{C} = \{c_{ij}\}_{i,j=1}^{n}$, $\boldsymbol{M} = diag\{\sigma_i^2\}_{i=1}^{n}$ and $\boldsymbol{\mu} = \big(\mu_1, \ldots, \mu_n\big)'$. Usually, $\boldsymbol{C}$, $\boldsymbol{M}$ and $\boldsymbol{\mu}$ are all unknown, and hence estimated from the data. In spatial regression context, the mean of the joint Gaussian distribution, $\boldsymbol{\mu}$ is called large-scale effect and often modeled as:

$$\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}, \tag{1.16}$$

where $\boldsymbol{X}$ is an $(n \times p)$ matrix of explanatory variables, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown parameters. The columns of the design matrix $\boldsymbol{X}$ are site-specific covariates. Using (1.15) and (1.16), we can write the conditional autoregressive CAR regression model as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.17}$$

with $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, (\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{M})$.

In practice, a simplified version of this model has been effectively employed by setting the covariance structure of this model to be $\sigma^2(\boldsymbol{I} - \rho\boldsymbol{W}^*)^{-1}\boldsymbol{D}$, where $\boldsymbol{W}^* = \{\frac{w_{ij}}{w_{i+}}\}_{i,j=1}^{n}$ with $w_{i+} = \sum_{j=1}^{n} w_{ij}$, is a known standardized proximity (neighborhood) matrix, $\boldsymbol{D} = \text{diag}\{\frac{1}{w_{1+}}, \frac{1}{w_{2+}}, \ldots, \frac{1}{w_{n+}}\}$ and $\sigma^2 > 0$. The proximity matrix often consists of elements $w_{ij} = 1$ if location j is neighbor to location i and $w_{ij} = 0$ otherwise.

1.5.2 Simultaneous Autoregressive Model

A second spatial regression model that will be studied in this thesis is the simultaneous autoregressive SAR model, proposed by Whittle (1954). To understand the idea behind this model, let $\boldsymbol{s} = \{s_1, \ldots, s_n\}'$ denote as before a lattice of spatial locations with associated responses $\boldsymbol{Y}(\boldsymbol{s}) = \big(Y(s_1), \ldots, Y(s_n)\big)'$ and associated $(n \times p)$ covariate matrix $\boldsymbol{X}(\boldsymbol{s}) = (\boldsymbol{X}(s_1), \ldots, \boldsymbol{X}(s_n))'$. The SAR approach models the response at the s_ith location as, (Waller and Gotway, 2004)

$$Y(s_i) = \boldsymbol{X}'(s_i)\boldsymbol{\beta} + \epsilon(s_i), \quad i = 1, \ldots, n, \tag{1.18}$$

where

$$\epsilon(s_i) = \sum_{j \neq i}^{n} \gamma_{ij}\epsilon(s_j) + e(s_i), \quad i = 1, \ldots, n, \tag{1.19}$$

where $\boldsymbol{\beta}$ is a $(p \times 1)$ unknown regression coefficients, $\boldsymbol{e}(\boldsymbol{s}) = \big(e(s_1), \ldots, e(s_n)\big)'$ are Gaussian errors with mean $\mathbf{0}$ and variance covariance matrix $\boldsymbol{\Lambda} = diag\{\sigma_i^2\}_{i=1}^n$. The parameters γ_{ij} with $\gamma_{ii} = 0$ are to model the spatial dependencies of the errors.

Ignoring the spatial indices s_i, this model can be re-written in a matrix format as follows

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{R}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{e}, \tag{1.20}$$

or, simply

$$(\boldsymbol{I} - \boldsymbol{R})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{e}, \tag{1.21}$$

where $\boldsymbol{R} = \{\gamma_{ij}\}_{i,j=1}^{n}$ and $\boldsymbol{I}$ is an $(n \times n)$ identity matrix. Yet, assuming that $(\boldsymbol{I} - \boldsymbol{R})$ is invertible, the SAR model can be re-written as

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X\beta} + \boldsymbol{u}, \\ \boldsymbol{u} &\sim N\left(\boldsymbol{0}, (\boldsymbol{I} - \boldsymbol{R})^{-1}\boldsymbol{\Lambda}(\boldsymbol{I} - \boldsymbol{R}')^{-1}\right). \end{aligned}$$

In many situations, nature exhibits sparsity, meaning that a small number of factors could capture most of the variability observed. This sparsity implies that we can express natural phenomena in the form of models with a relatively small number of parameters. Spatial regression models often exploit sparsity to imply simpler covariance structures and hence, less computational complexity. In the context of the SAR model, one of the most used covariance structures is obtained by setting $\boldsymbol{\Lambda} = \sigma^2\boldsymbol{I}$ and $\boldsymbol{R} = \rho\boldsymbol{W}$, where σ^2 is the variability of the pure noise component and $\boldsymbol{W}$ is a sparse and known proximity matrix as was described in the previous section.

The MLE estimators of the SAR model parameters have been extensively studied in the literature. A detailed treatment of the SAR model and its theoretical underpinning can be found in Kazar and Celik (2012). Durban et al. (2012) proposed a nonparamteric version of the SAR model whereby spline functions were used to model the large-scale spatial effects. Interpretations of the meaning of the SAR and CAR covariance structures can be found in Wall (2004). A review of spatial regression models in the context of ecology is found in Beale et al. (2010).

1.5.3 Spatial Moving Average Model

Yet another spatial regression model on lattices is the *spatial moving average* (SMA) model. As the name shows, this model imposes a moving average specification on the noise term as is the case in temporal time series processes. Let $\boldsymbol{s} = \{s_1, \ldots, s_n\}$ be a lattice of sites as before and $e(s_i)$ be the random error associated with site (s_i). The SMA model error specification is given by:

$$e(s_i) = \epsilon(s_i) + \sum_{j \neq i}^{n} g_{ij} \epsilon(s_j), \quad i, j = 1, \ldots, n, \tag{1.22}$$

where, $\{g_{ij}\}_{i,j=1}^{n}$ are unknown spatial dependence parameters with $g_{ii} = 0$, $\{\epsilon(s_i)\}_{i=1}^{n}$ are *iid* mean zero Gaussian errors. Thus, the SMA compiles the spatial regression model's error, associated with site s_i, as a linear combination of the random noises in the neighboring sites. Using matrix notation, the model in (1.22) can be written as:

$$\boldsymbol{e}(\boldsymbol{s}) = \boldsymbol{\epsilon}(\boldsymbol{s}) + \boldsymbol{G}\boldsymbol{\epsilon}(\boldsymbol{s}) = (\boldsymbol{I} + \boldsymbol{G})\boldsymbol{\epsilon}(\boldsymbol{s}), \tag{1.23}$$

where $\boldsymbol{G} = \{g_{ij}\}_{i,j=1}^{n}$, with $g_{ii} = 0$, $\boldsymbol{e}(\boldsymbol{s}) = (e(s_1), \ldots, e(s_n))'$, $\boldsymbol{\epsilon}(\boldsymbol{s}) = (\epsilon(s_1), \ldots, \epsilon(s_n))'$ and $\boldsymbol{I}$ is an $(n \times n)$ identity matrix.

By dropping the site index $\boldsymbol{s}$, the spatial response model can be formulated in terms of the joint Gaussian distribution

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} + \boldsymbol{G})\boldsymbol{\epsilon}, \tag{1.24}$$

where $\boldsymbol{Y} = \boldsymbol{Y}(\boldsymbol{s}) = (Y(s_1), \ldots, Y(s_n))'$ is the $(n \times 1)$ observed response vector at the lattice sites $(\boldsymbol{s})$, $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{s})$ the $(n \times p)$ fixed matrix of p explanatory variables and

$\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown regression parameters.

Again, sparsity can be exploited in the context of SMA by choosing the spatial dependence matrix $\boldsymbol{G}$ as $\rho\boldsymbol{W}$, where $\boldsymbol{W}$ is a sparse and known neighborhood matrix as before.

Assuming that the error term $\boldsymbol{\epsilon}$ follows a multivariate Gaussian distribution with mean $\mathbf{0}$, and variance covariance matrix $\sigma^2\boldsymbol{I}$, then the response vector $\boldsymbol{Y}$ is distributed as:

$$\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \sigma^2(\boldsymbol{I}+\boldsymbol{G})(\boldsymbol{I}+\boldsymbol{G}')). \tag{1.25}$$

1.5.4 Model Selection in Spatial Regressions

Most of the model selection methods in regression deal with independent data, while, in spatial regression it is a matter of challenge because the data are highly dependent. To be quite pessimistic, one could state that in spatial regressions we are basing our inferences on a single observation. Therefore, the literature on model selection procedures for the spatial regression models has been quite negligible. For example, Kashyap and Chellappa (1983) used the BIC selection criterion to choose the parameter of the SAR and CAR models. Florax et al. (2003) considered various search strategies in spatial econometric modeling by using the classical forward, stepwise, robust, hybrid, and Hendry's strategies, and compared these methods numerically. They observed that Hendry's strategy is dominated by other approaches with respect to detecting the spatial dependence. Hoeting et al. (2006) compared the selection performance of the independent AIC, which ignores the spatial correlation, with the AIC approach in spatial regressions. Their simulation results had shown the superiority of the spatial AIC in model selection as compared to the independence AIC. Kissling and Carl (2008) used various model selection procedures to select the best SAR model from a range of model specifications. They considered the spatial error model, spatial lagged model, and spatial mixed model. Their procedure was based on minimum residual spatial autocorrelation, maximum model fit R^2, and the AIC. Song and De Oliveira (2012) explained a Bayesian approach for model selection in Gaussian CAR and SAR models.

In the context of penalty selection and estimation methods, Wang and Zhu (2009) considered various penalty functions for variable selection and parameter estimation in spatial linear regressions. They considered L_q, hard thresholding and SCAD penalty functions, and established the oracle property of these methods under some regularity conditions, and conducted a numerical study to compare the estimators. These authors found that, the SCAD estimator outperformed the thresholding and the LASSO estimators in many cases. Zhu and Liu (2009) proposed a penalized likelihood to estimate the covariance matrix of spatial Gaussian Markov random field models with unspecified neighborhood structure. They used weighted L_1 regularization and showed that the LASSO type approach gives improved covariance estimators, measured by different criteria. Also, they derived the asymptotic properties of their proposed estimators. Zhu et al. (2010) developed a new methodology for simultaneous model selection and parameter estimation of spatial linear models via adaptive LASSO. The treatment of Zhu et al. (2010) was for general spatial linear models while the CAR and SAR models were studied as two special cases. The authors provided an efficient algorithm for obtaining an approximation of the penalized maximum likelihood and established the asymptotic properties.

Huang et al. (2010) considered a raster-based *geographic information systems*

(GIS), that organizes the spatial data in layers, and built a high-dimensional spatial regression model with unknown layers and neighborhoods. They proposed the use of LASSO that simultaneously selects variables, chooses neighborhoods and estimates the parameters via a generalized version of the LARS algorithm.

1.6 Thesis Organization and Highlights of Contributions

The problem of finding efficient estimators is of a central importance in all statistical models. When a large number of variables is to be related to a given set of responses, regression models are often the wise choice to pursue.

An important aspect of statistical inference is to select the correct set of variables that explain variations in the response. This leads us to model selection problem, which has been intensively studied in the context of regression models. Model selection procedures in the literature include the classical AIC, BIC procedures and the more recent penalty procedures. After a model is selected, we are still not completely certain of the validity of the selected model as including all important explanatory variables. Therefore, even after applying modern model selection procedures, the resulting submodel (often known as reduced or restricted model) can be considered as an uncertain information. Similarly, an investigator could have prior non-sample uncertain information concerning which of the predictor variables are important. In both cases, the resulting full and restricted models have their risks that are functions of the reliability of the information provided. In such a situation, investigators would like to take a middle way that protects them uniformly against high inefficiency at the expense of less efficiency in some subsets of the parameter space.

The class of shrinkage and pretest estimators provides such protection by combining the full and reduced model estimators, resulting in uniformly low risk estimators in the context of regression models. Therefore, in this dissertation, we propose the pretest, James-Stein, positive James-Stein estimators for the large-scale effects $\boldsymbol{\beta}$ in three spatial regression models. Namely, we propose these estimators for the conditional autoregressive CAR, simultaneous autoregressive SAR, and spatial moving average SMA models. In addition, we devise a simple procedure for computing penalty estimators, for the large-scale effects of these three regression models. Specifically, we will construct LASSO, Adaptive LASSO and SCAD estimators for regression parameters in the CAR, SAR and SMA models.

The proposed pretest and shrinkage estimators will be based on a general linear candidate subspace of the large-scale effects space, stemming from uncertain prior information. We will derive the asymptotic risks and biases of the proposed pretest and shrinkage estimators and compare them to those associated with the full space and candidate subspace parameter estimates. Also, we conduct numerical studies using simulated and real data examples to compare the performance of the proposed estimators with the absolute penalty estimators.

In Chapter 2, we consider the application of the pretest, shrinkage and penalty estimators in the conditional autoregressive CAR model. At the beginning of the chapter, we discuss the CAR model specifications, the maximum likelihood estimator MLE of the model parameters, and display the Mardia-Marshall Theorem, (Mardia and Marshall (1984)), which is at the core of the asymptotics in spatial lattice regression models. Based on the Mardia-Marshall Theorem, we derive the asymptotic distributional bias, mean squared error matrices and quadratic risk of the proposed estimators. We carry out analytical performance comparisons among these estimators and with respect to the restricted and unrestricted estimators. We propose a simple procedure for constructing penalty estimators and apply it in constructing LASSO, Adaptive LASSO and SCAD estimators for $\boldsymbol{\beta}$ in the CAR model. Numerical studies are then carried out to compare restricted and shrinkage estimators with the penalty estimators based on simulated as well as real data examples.

In Chapter 3, we propose pretest, shrinkage and penalty estimators for the large-scale effects of the SAR model and, following the structure of Chapter 2, we study numerically their relative performances. Finally, we consider in Chapter 4 the problem of constructing pretest, shrinkage and penalty estimators for the SMA model. The contributions in this dissertation are summarized as follows:

1. We propose the restricted, pretest, and shrinkage estimators for estimating the large-scale effect in the conditional autoregressive CAR, simultaneous autoregressive SAR and spatial moving average SMA models. This class of estimators is new for these spatial models and has never been considered in the literature. We indicate the importance of using the prior information in producing a sub-model, which carefully represents the data, and reduces the model complexity.

2. Analytical results on the risks and biases of the restricted, pretest, shrinkage and full model estimators are derived based on the concept of distributional biases and risks. Also, mean squared error matrices of these estimators are derived and compared analytically, taking the full model MLEs as a benchmark estimator for the comparison.

3. We introduce a simple algorithm for computing penalty estimators for the large-

scale effect parameters in the three spatial regression models, CAR, SAR and SMA. This algorithm exploits matrix decomposition and existing LARS algorithm for computing the LASSO, Adaptive LASSO and SCAD estimators of the large-scale effects in these three spatial models.

4. We carry out intensive empirical assessment of the above array of proposed estimators through two real data examples and Monte Carlo simulations. Specifically, we run large scale Monte Carlo simulations comparing the mean squared errors of the restricted, pretest, and shrinkage estimators proposed for the three spatial models with respect to the unrestricted MLEs. We also compare these estimators to the LASSO, Adaptive LASSO and SCAD estimators by using the *mean squared prediction error* (MSPE) as a measure of relative performance with respect to the benchmark estimator. Finally, we apply these estimation procedures to two data sets on housing prices and data on crime distribution. In this application to a real data set, we device a bootstrapping procedure for obtaining the mean squared prediction errors of the various estimators.

5. Finally, we appreciate the performance of the proposed estimators and give recommendations on which ones are safer to use in which situation and we propose important research topics for extending the results of this dissertation.

Chapter 2

Efficient Estimation for the Conditional Autoregressive Model

2.1 Introduction

In this chapter we will consider the CAR model and construct pretest, James-Stein and positive James-Stein shrinkage estimators for the so called, large-scale effects vector of parameters, $\boldsymbol{\beta}$. We postulate a general candidate subspace, $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$, where $\boldsymbol{H}$ is a known $q \times p$ real-valued matrix and $\boldsymbol{h}$ is a known q-dimensional vector of real numbers. Such a general restriction accommodates a variety of prior non-sample information about the parameters put forward by the investigator as well as restrictions stemming from model selection procedures such as AIC, BIC and penalty model selection approaches. For instance, if based on prior knowledge, the investigator believes that some of the large-scale effects are irrelevant, then $\boldsymbol{H}$ will be an appropriately defined contrast matrix and $\boldsymbol{h}$ could be set to zero. On the other hand, when there is no prior non-sample information, one could resort to model selection procedures and identify some of the components of $\boldsymbol{\beta}$ as being practically zero. In this latter situation, as the final model selected need not be the true model, it is still safer to resort to estimation methods which take into account the linear restriction induced by the model selection criteria.

Based on the postulated candidate subspace, we build restricted MLE estimators for the effects of the model reduced to the candidate subspace $\hat{\boldsymbol{\beta}}^R$. Consequently, we show that this vector is jointly and asymptotically multivariate normal with the vector of unrestricted MLE estimator of the large-scale effects, $\hat{\boldsymbol{\beta}}$. For completeness, we reiterate the marginal asymptotic normality of $\hat{\boldsymbol{\beta}}$ result due to Mardia and Marshall (1984). At this point, we define the shrinkage estimators as combinations of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^R$ and provide theoretical analysis of their risks and biases by comparing to the benchmark $\hat{\boldsymbol{\beta}}$ as well as to the restricted estimator $\hat{\boldsymbol{\beta}}^R$. A Monte Carlo simulation study is then undertaken in order to compare the small sample performance of this array of estimators.

The second objective of the chapter is to construct penalty estimators for $\boldsymbol{\beta}$ based on the LASSO, adaptive LASSO and SCAD penalty estimators. We devise a second Monte Carlo simulation to compare the shrinkage to penalized estimators in terms of risks and prediction errors. Finally, we apply the estimators to real data on Boston Housing Prices.

2.1.1 Chapter Organization

Section 2.2 discusses the conditional autoregressive (CAR) model and preliminaries. The unrestricted maximum likelihood estimation is discussed in detail in Section

2.3. In Section 2.4, we present estimation strategies based on shrinkage techniques. Asymptotic results are provided in Section 2.5. We present the asymptotic risk analysis in Section 2.6. In Section 2.7, we consider estimation of the parameter vector $\boldsymbol{\beta}$ using three penalty functions. Numerical studies to compare the performance of all estimators are illustrated in Section 2.8. We present a conclusion in Section 2.9.

2.2 The model and preliminaries

Following Cressie (1993), assume that there are n spatial sites (reference locations such as small geographical areas, pixels, etc..), the collection of which forms what is known as a lattice, denoted by $\boldsymbol{s} = \{s_1, s_2, \ldots, s_n\}$. For each one of these sites, a set of neighboring sites is defined by

$$N(s_i) = \{s_j : j = 1, \ldots, n \text{ is a neighbor of } i\}, \quad i = 1, \ldots, n,$$

where a site is neighbor to another if they are close to each other under a certain metric. The collection of observations at these sites is denoted by $\boldsymbol{Y}_n(\boldsymbol{s}) = \{Y(s_1), Y(s_2), \ldots, Y(s_n)\}$, while the set of covariates that comes with it, is denoted by $\boldsymbol{X}(s) = \boldsymbol{X_i} = (X_{1i}, X_{2i}, \ldots, X_{pi})'$. The effect of these covariates on $\boldsymbol{Y}_n(\boldsymbol{s})$ is a vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$, known also as the large-scale effects of the spatial model. When $Y(s_i)$ is continuous, it is often modeled as a Gaussian process with mean $\boldsymbol{X}(s_i)'\boldsymbol{\beta}$ and covariance matrix allowing for the spatial dependence among responses in a neighborhood. For simplicity of notation, we shall compact the covariate vectors for all sites into a design matrix $\boldsymbol{X}_n(s)$. The subscript n and the spatial location index s will be omitted sometimes, unless we need to display them explicitly. That is, we will simply refer to the data on the lattice $\boldsymbol{s}$ as $(\boldsymbol{Y}, \boldsymbol{X})$.

In this chapter we shall concentrate on a class of spatial models known as CAR (conditional autoregressive) models introduced by Besag et al. (1991). Dropping the index s referring to the site, the CAR model of Besag et al. (1991) can be defined as

$$\boldsymbol{Y} \sim N\left(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2(\boldsymbol{I} - \rho\boldsymbol{W}^*)^{-1}\boldsymbol{D}\right),$$

where $\boldsymbol{W}^* = \{\frac{w_{ij}}{w_{i+}}\}_{i,j=1}^n$ with $w_{i+} = \sum_{j=1}^n w_{ij}$ is a known standardized proximity (closeness) matrix, $\boldsymbol{D} = \text{diag}\{\frac{1}{w_{1+}}, \frac{1}{w_{2+}}, \ldots, \frac{1}{w_{n+}}\}$ and $\sigma^2 > 0$. The proximity matrix often consists of elements $w_{ij} = 1$ if location j is neighbor to location i and $w_{ij} = 0$ otherwise. Essentially, this model is a multivariate Gaussian model with only one observation, $\boldsymbol{Y(s)}$. The name conditional autoregressive comes from the fact that this model can be re-written in the following conditional form,

$$\begin{aligned} E\{Y(s_i) \mid Y(s_j), j \neq i\} &= \boldsymbol{X}'(s_i)\boldsymbol{\beta} + \rho\sum_{j=1}^n W_{ij}(Y(s_j) - \boldsymbol{X}'(s_j)\boldsymbol{\beta}) \\ var\{Y(s_i) \mid Y(s_j), j \neq i\} &= \sigma_i^2 = \frac{\sigma^2}{W_{i+}}. \end{aligned}$$

2.3 Unrestricted Maximum Likelihood Estimation

The maximum likelihood estimators (MLEs) of $\boldsymbol{\beta}$, σ^2, and ρ for the CAR model are usually obtained from the log-likelihood function given by

$$\begin{aligned} l = \log(L(\boldsymbol{\beta}, \sigma^2, \rho)) &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(|(\boldsymbol{I} - \rho\boldsymbol{W}^*)^{-1}\boldsymbol{D}|) \\ &- \frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{D}^{-1}(\boldsymbol{I} - \rho\boldsymbol{W}^*)(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), \end{aligned} \quad (2.1)$$

where $|\boldsymbol{A}|$ denotes the determinant of the matrix $\boldsymbol{A}$, through the following simple profiling approach (Cressie, 1993):

i) For a fixed ρ, solve the likelihood equations

$$\begin{aligned}\frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{1}{2\sigma^2}\Big\{\boldsymbol{X}'\boldsymbol{D}^{-1}(\boldsymbol{I}-\rho\boldsymbol{W}^*)(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) \\ &\quad +(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{D}^{-1}(\boldsymbol{I}-\rho\boldsymbol{W}^*)(\boldsymbol{X})\Big\} = 0 \\ \frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} - \frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{D}^{-1}(\boldsymbol{I}-\rho\boldsymbol{W}^*)(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) = 0\end{aligned}$$

to obtain

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\rho) &= \left(\boldsymbol{X}'\boldsymbol{D}^{-1}(\boldsymbol{I}-\rho\boldsymbol{W}^*)\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{D}^{-1}(\boldsymbol{I}-\rho\boldsymbol{W}^*)\boldsymbol{Y} \qquad (2.2) \\ \hat{\sigma}^2(\rho) &= \frac{(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}(\rho))'\boldsymbol{D}^{-1}(\boldsymbol{I}-\rho\boldsymbol{W}^*)(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}(\rho))}{n}\end{aligned}$$

ii) Plug $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ back in the log-likelihood function, and maximize the profile log-likelihood function

$$\begin{aligned}l^*(\rho) &= -\frac{n}{2}\log\left(\frac{(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}(\rho))'\boldsymbol{D}^{-1}(\boldsymbol{I}-\rho\boldsymbol{W}^*)(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}(\rho))}{n}\right) \\ &- \frac{1}{2}\log(|(\boldsymbol{I}-\rho\boldsymbol{W}^*)^{-1}\boldsymbol{D}|),\end{aligned}$$

with respect to ρ to obtain a maximum profile likelihood estimator, $\hat{\rho}$.

iii) Finally, replace $\hat{\rho}$ back into (2.2) to obtain the final estimators, which we shall call the *unrestricted maximum likelihood estimators* (UMLE) of $\boldsymbol{\beta}$ and σ^2, denoted by $\hat{\boldsymbol{\beta}}$, and $\hat{\sigma}^2$, respectively.

The consistency and the asymptotic normality of the vector $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\rho})$ follow directly from a general result due to Mardia and Marshall (1984) via an increasing domain asymptotic method. Generally, in increasing domain asymptotics, it is assumed that the number of sites is approaching infinity while number of observations at each site is held fixed. The asymptotic normality of the large-scale effects $\hat{\boldsymbol{\beta}}$ is a straightforward consequence of the Mardia-Marshall result.

Theorem 2.3.1. (Mardia and Marshall (1984)) As $n \longrightarrow \infty$, and under the conditions in the Appendix A, $\hat{\boldsymbol{\beta}} \stackrel{P}{\longrightarrow} \boldsymbol{\beta}$,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \stackrel{D}{\longrightarrow} N\left(\mathbf{0}, \sigma^2 \boldsymbol{C}_0^{-1}\right).$$

2.4 Improved Estimation Strategies

In this section, we propose four estimators for the large-scale effects of the CAR model. The first is an estimator restricted to a candidate subspace of the form $\boldsymbol{H\beta} = \boldsymbol{h}$, as discussed in the introduction of this chapter. The candidate subspace could come from uncertain prior information or from model selection methodologies such as the penalty estimation methods to be discussed in this chapter. The remaining estimators which will be proposed in this section are the pretest and shrinkage estimators.

2.4.1 Restricted Estimator

Here we are interested in estimating the vector $\boldsymbol{\beta}$ when it is suspected that $\boldsymbol{\beta}$ may be restricted to a subspace defined by

$$A_0 : \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h} \tag{2.3}$$

where $\boldsymbol{H}$ is a $q \times p$ known matrix of rank $q(q \leq p)$, and $\boldsymbol{h}$ is a $q \times 1$ vector of known constants. By using Lagrange multipliers, it is straight forward to show that the MLE of $\boldsymbol{\beta}$, restricted to the candidate subspace $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$, is

$$\hat{\boldsymbol{\beta}}^R = \hat{\boldsymbol{\beta}} - (\boldsymbol{X}_n' \hat{C}_n^{-1} \boldsymbol{X}_n)^{-1} \boldsymbol{H}' (\boldsymbol{H} (\boldsymbol{X}_n' \hat{C}_n^{-1} \boldsymbol{X}_n)^{-1} \boldsymbol{H}')^{-1} (\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}), \tag{2.4}$$

while the version with known ρ is

$$\hat{\boldsymbol{\beta}}^R(\rho) = \hat{\boldsymbol{\beta}}(\rho) - (\boldsymbol{X_n' C_n^{-1} X_n})^{-1} \boldsymbol{H}' (\boldsymbol{H} (\boldsymbol{X_n' C_n^{-1} X_n})^{-1} \boldsymbol{H}')^{-1} (\boldsymbol{H}\hat{\boldsymbol{\beta}}(\rho) - \boldsymbol{h}). \tag{2.5}$$

It is obvious that $\hat{\boldsymbol{\beta}}^R$ is a biased estimator unless the restriction induced by the candidate subspace is correct. On the other hand, the UMLE of $\boldsymbol{\beta}$ obtained in the previous section is unbiased and more efficient than the restricted estimator if the true parameter vector $\boldsymbol{\beta}$ lives in its natural space, free of restrictions. This comparative analysis as well as the joint asymptotic normality of the vector $(\hat{\boldsymbol{\beta}}^R, \hat{\boldsymbol{\beta}})$ will be discussed in Section 2.5. Since the prior information leading to $\hat{\boldsymbol{\beta}}^R$ is uncertain, so is the quality of this estimator. Therefore, a way out of this dilemma is to construct pretest and shrinkage-type estimators which combine $\hat{\boldsymbol{\beta}}^R$ and $\hat{\boldsymbol{\beta}}$ in such a way that the uncertainty in the prior information is incorporated in the estimation process.

2.4.2 Pretest Estimator

The pretest estimator denoted by $\hat{\boldsymbol{\beta}}^{PT}$ is defined as follows

$$\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}} I(\mathcal{L}_n > \chi^2_{q,\alpha}) + \hat{\boldsymbol{\beta}}^R I(\mathcal{L}_n \leq \chi^2_{q,\alpha})$$

where $I(A)$ is an indicator function for the event A, $\mathcal{L}_n$ is the test statistic for testing the null hypothesis (2.3) and given by

$$\mathcal{L}_n = \frac{(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})}{s_e^2}, \tag{2.6}$$

$$s_e^2 = \frac{(\boldsymbol{Y}_n - \boldsymbol{X}_n\hat{\boldsymbol{\beta}})'\hat{C}_n^{-1}(\boldsymbol{Y}_n - \boldsymbol{X}_n\hat{\boldsymbol{\beta}})}{n-p}, \tag{2.7}$$

and $\hat{C}_n = (\boldsymbol{I} - \hat{\rho}\boldsymbol{W}^*)^{-1}\boldsymbol{D}$. Here, $\chi_{q,\alpha}$, is the α^{th} upper quantile of a central chi-square distribution with q degrees of freedom. This estimator can also be rewritten as

$$\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R) I(\mathcal{L}_n \leq \chi^2_{q,\alpha}). \tag{2.8}$$

If we look at the candidate subspace as the linear hypothesis, $\boldsymbol{A}_0 : \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$, it is obvious that the pretest estimator depends on whether or not the candidate subspace restriction is accepted at level α. Accordingly, the pretest estimator yields only two possibilities: either $\hat{\boldsymbol{\beta}}^R$ or $\hat{\boldsymbol{\beta}}$. Therefore, the statistic $\mathcal{L}_n$ serves as test statistic for $\boldsymbol{A}_0$ with associated level of significance α.

2.4.3 Shrinkage Estimators

Following Ahmed (2001), the James-Stein estimator of $\boldsymbol{\beta}$ can be defined as

$$\hat{\boldsymbol{\beta}}^{JS} = \hat{\boldsymbol{\beta}}^{R} + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})\{1 - (q-2)\mathcal{L}_n^{-1}\}. \tag{2.9}$$

Sometimes, the James-Stein estimator defined above suffers from a phenomenon known as over-shrinkage, whereby negative coordinates of $\hat{\boldsymbol{\beta}}$ are obtained whenever $(q-2)\mathcal{L}_n^{-1} > 1$. In order to avoid such eventuality, we consider the positive rule James-Stein estimator,

$$\hat{\boldsymbol{\beta}}^{JS+} = \hat{\boldsymbol{\beta}}^{R} + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})\{1 - (q-2)\mathcal{L}_n^{-1}\}^{+}, \tag{2.10}$$

where $u^+ = \max(0, u)$. Alternatively, $\hat{\boldsymbol{\beta}}^{JS+}$ can be written as

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{JS+} &= \hat{\boldsymbol{\beta}}^{R} + \{1 - (q-2)\mathcal{L}_n^{-1}\} I(\mathcal{L}_n > (q-2))(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R}), & (2.11) \\ &= \hat{\boldsymbol{\beta}}^{JS} - (1 - (q-2)\mathcal{L}_n^{-1}) I(\mathcal{L}_n < (q-2))(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R}). & (2.12) \end{aligned}$$

2.5 Asymptotic Results

In this section we study the asymptotic behavior of the various estimators, $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^{R}, \hat{\boldsymbol{\beta}}^{JS}, \hat{\boldsymbol{\beta}}^{JS+}$. Specifically, we show that the restricted and unrestricted estimators are jointly asymptotically normal. Secondly, we define and derive expressions for the *asymptotic distributional bias* (ADB), the *asymptotic mean squared error matrix* (AMSEM), and the *asymptotic quadratic risk* (AQR) of the estimators by using the joint normality of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{R}$. In particular, the AQR is a measure of the risk of the estimators based on quadratic loss function and hence, it can be used to compare the various estimators discussed in the previous sections. Such comparative study will be detailed in Section 2.6 below.

2.5.1 Joint Normality

In this section we prove a technical result which shows that the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^R$ are asymptotically jointly normal under the sequence of local alternatives,

$$A_{(n)}\colon \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h} + \frac{\boldsymbol{\xi}}{\sqrt{n}}, \tag{2.13}$$

where $\boldsymbol{\xi}$ is a $q\times 1$ fixed vector in $\mathbb{R}^q$. If we set $\boldsymbol{\xi} = \mathbf{0}$, then the local alternative becomes $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$, which is the linear hypothesis representing the candidate subspace. The main result of this subsection is in the following theorem.

Theorem 2.5.1. Under the local alternatives in (2.13) and the regularity conditions (i)-(v) in the Appendix A, we have

(i) $\boldsymbol{T}_n^{(1)} = \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{T}^{(1)} \sim N_p(0, \sigma^2\boldsymbol{C}_0^{-1})$

(ii) $\boldsymbol{T}_n^{(2)} = \sqrt{n}(\hat{\boldsymbol{\beta}}^R - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{T}^{(2)} \sim N_p(-\boldsymbol{\delta}, \sigma^2\boldsymbol{A}_0)$,

(iii) $\boldsymbol{T}_n^{(3)} = \sqrt{n}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R) \xrightarrow{D} \boldsymbol{T}^{(3)} \sim N_p(\boldsymbol{\delta}, \sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0))$

(iv) $\begin{pmatrix} \boldsymbol{T}_n^{(1)} \\ \boldsymbol{T}_n^{(3)} \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \boldsymbol{T}^{(1)} \\ \boldsymbol{T}^{(3)} \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\delta} \end{pmatrix}, \sigma^2 \begin{pmatrix} \boldsymbol{C}_0^{-1} & \boldsymbol{C}_0^{-1} - \boldsymbol{A}_0 \\ \boldsymbol{C}_0^{-1} - \boldsymbol{A}_0 & \boldsymbol{C}_0^{-1} - \boldsymbol{A}_0 \end{pmatrix}\right)$

(v) $\begin{pmatrix} \boldsymbol{T}_n^{(2)} \\ \boldsymbol{T}_n^{(3)} \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \boldsymbol{T}^{(2)} \\ \boldsymbol{T}^{(3)} \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} -\boldsymbol{\delta} \\ \boldsymbol{\delta} \end{pmatrix}, \sigma^2 \begin{pmatrix} \boldsymbol{A}_0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{C}_0^{-1} - \boldsymbol{A}_0 \end{pmatrix}\right)$, where, $\boldsymbol{A}_0 = \boldsymbol{C}_0^{-1} - \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}H\boldsymbol{C}_0^{-1}$, $\boldsymbol{\delta} = \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}$.

Proof:

(i) The proof follows from Mardia and Marshall (1984).

(ii)

$$
\begin{aligned}
\boldsymbol{T}_n^{(2)} &= \sqrt{n}(\hat{\boldsymbol{\beta}}^R - \boldsymbol{\beta}) \\
&= \sqrt{n}\Big\{\hat{\boldsymbol{\beta}} - (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1} \\
&\quad (\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}) - \boldsymbol{\beta}\Big\} \\
&= \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \sqrt{n}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1} \\
&\quad \boldsymbol{H}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \sqrt{n}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1} \\
&\quad (\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h}) \\
&= \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1} \\
&\quad \sqrt{n}\boldsymbol{H}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&- (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi} \\
&= \Big[\boldsymbol{I}_p - (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\Big] \\
&\quad \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi},
\end{aligned}
$$

which is a linear function of $\boldsymbol{T}_n^{(1)}$, so as $n \to \infty, \boldsymbol{T}_n^{(2)} \xrightarrow{D} \boldsymbol{T}^{(2)} \sim N_p(\boldsymbol{\mu}^{(R)}, \boldsymbol{\Sigma}^{(R)})$, where

$$
\begin{aligned}
\boldsymbol{\mu}^{(R)} &= -\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi} \\
&= -\boldsymbol{\delta}, \text{ and}
\end{aligned}
$$

$$\begin{aligned}
\boldsymbol{\Sigma}^{(R)} &= \Big[\boldsymbol{I}_p - \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\Big]\sigma^2\boldsymbol{C}_0^{-1}\Big[\boldsymbol{I}_p - \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1} \\
&\quad \boldsymbol{H}\Big] \\
&= \sigma^2\Big\{(\boldsymbol{C}_0^{-1} - \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1})(\boldsymbol{I}_p - \boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1} \\
&\quad \boldsymbol{H}\boldsymbol{C}_0^{-1})\Big\} \\
&= \sigma^2\Big\{\boldsymbol{C}_0^{-1} - \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1} - \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1} \\
&\quad \boldsymbol{H}\boldsymbol{C}_0^{-1} + \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1}\Big\} \\
&= \sigma^2\{\boldsymbol{C}_0^{-1} - \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1}\} \\
&= \sigma^2\boldsymbol{A}_0.
\end{aligned}$$

(iii)

$$\begin{aligned}
\boldsymbol{T}_n^{(3)} &= \sqrt{n}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R) \\
&= \sqrt{n}\Big\{(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})\Big\} \\
&= \sqrt{n}\Big\{(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}\big(\boldsymbol{H}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&+ \boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h}\big)\Big\} \\
&= (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&+ (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi},
\end{aligned}$$

which is also a linear function of $\boldsymbol{T}_n^{(1)}$. Therefore, as $n \to \infty$, we have: $\boldsymbol{T}_n^{(3)} \xrightarrow{D} \boldsymbol{T}^{(3)} \sim N_p(\boldsymbol{\mu}^{(3)}, \boldsymbol{\Sigma}^{(3)})$, where

$$\begin{aligned}
\boldsymbol{\mu}^{(3)} &= \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi} \\
&= \boldsymbol{\delta},
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Sigma}^{(3)} &= \big(\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\big)\sigma^2\boldsymbol{C}_0^{-1}\big(\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1}\big) \\
&= \sigma^2\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1} \\
&= \sigma^2\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1} \\
&= \sigma^2\left(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0\right).
\end{aligned}$$

(iv) From (iii) $\boldsymbol{T}_n^{(3)}$ can be written as

$$\begin{aligned}
\boldsymbol{T}_n^{(3)} &= (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{T}_n^{(1)} \\
&+ (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\begin{pmatrix} \boldsymbol{T}_n^{(1)} \\ \boldsymbol{T}_n^{(3)} \end{pmatrix} &= \begin{pmatrix} \boldsymbol{I}_p \\ \boldsymbol{F}_n \end{pmatrix}\boldsymbol{T}_n^{(1)} + \begin{pmatrix} \boldsymbol{0}_p \\ \boldsymbol{G}_n \end{pmatrix} \\
&= \boldsymbol{Q}_n\boldsymbol{T}_n^{(1)} + \boldsymbol{U}_n,
\end{aligned}$$

where $\boldsymbol{I}_p$ is a $p \times p$ identity matrix, $\boldsymbol{0}_p$ is a $p \times 1$ vector of zeros,
$\boldsymbol{F}_n = (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}$,
$\boldsymbol{G}_n = (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}$,
$\boldsymbol{Q}_n = \begin{pmatrix} \boldsymbol{I}_p \\ \boldsymbol{F}_n \end{pmatrix}$ and $\boldsymbol{U}_n = \begin{pmatrix} \boldsymbol{0}_p \\ \boldsymbol{G}_n \end{pmatrix}$.
As $n \to \infty$, $\boldsymbol{F}_n \xrightarrow{P} \boldsymbol{F}_0 = \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}$ and $\boldsymbol{G}_n \xrightarrow{P} \boldsymbol{\delta}$. Therfore,

$$\begin{pmatrix} \boldsymbol{T}_n^{(1)} \\ \boldsymbol{T}_n^{(3)} \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \boldsymbol{T}^{(1)} \\ \boldsymbol{T}^{(3)} \end{pmatrix} \sim N_{2p}(\boldsymbol{\mu}^{(4)}, \boldsymbol{\Sigma}^{(4)}), \text{ where}$$

$$\begin{aligned}
\boldsymbol{\mu}^{(4)} &= \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\delta} \end{pmatrix}, \\
\boldsymbol{\Sigma}^{(4)} &= \begin{pmatrix} \boldsymbol{I}_p \\ \boldsymbol{F}_0 \end{pmatrix} \sigma^2 \boldsymbol{C}_0^{-1} \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{F}_0' \end{bmatrix} \\
&= \sigma^2 \begin{pmatrix} \boldsymbol{C}_0^{-1} & \boldsymbol{C}_0^{-1}\boldsymbol{F}_0' \\ \boldsymbol{F}_0\boldsymbol{C}_0^{-1} & \boldsymbol{F}_0\boldsymbol{C}_0^{-1}\boldsymbol{F}_0' \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} \boldsymbol{C}_0^{-1} & \boldsymbol{C}_0^{-1} - \boldsymbol{A}_0 \\ \boldsymbol{C}_0^{-1} - \boldsymbol{A}_0 & \boldsymbol{C}_0^{-1} - \boldsymbol{A}_0 \end{pmatrix}.
\end{aligned}$$

(v) Also, note that $\begin{pmatrix} \boldsymbol{T}_n^{(2)} \\ \boldsymbol{T}_n^{(3)} \end{pmatrix}$ is a linear combination of $\boldsymbol{T}_n^{(1)}$, and can be written as

$$\begin{pmatrix} \boldsymbol{T}_n^{(2)} \\ \boldsymbol{T}_n^{(3)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_p - \boldsymbol{F}_n \\ \boldsymbol{F}_n \end{pmatrix} \boldsymbol{T}_n^{(1)} + \begin{pmatrix} -\boldsymbol{I}_p \\ \boldsymbol{I}_p \end{pmatrix} \boldsymbol{G}_n.$$

So, the proof follows using the same procedure as in (iv). □

2.5.2 Asymptotic Distributional Bias

In this section we define a measure of an estimator's bias known as the asymptotic distributional bias (ADB). In general, it is not easy to obtain the finite sample risk and bias of estimators in many practical situations. It is often resorted to asymptotic methods which essentially exploit convergence in distribution. However, convergence in distribution does not guarantee convergence in quadratic risk, needed for the analysis of risk and bias in the case of the shrinkage estimators. This difficulty has been overcome largely by introducing the concept of asymptotic distributional bias and risk, which, in turn, is based on the concept of local alternatives defined in the previous section.

For any given estimator $\hat{\boldsymbol{\beta}}^*$ of $\boldsymbol{\beta}$, let $\boldsymbol{G}(\boldsymbol{x})$ be the asymptotic distribution function of $\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})$,

$$\boldsymbol{G}(\boldsymbol{x}) = \lim_{n\to\infty} P_{A_{(n)}}(\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) \leq \boldsymbol{x}). \tag{2.14}$$

We define the ADB as

$$ADB(\hat{\boldsymbol{\beta}}^*) = \int \boldsymbol{x} \boldsymbol{d} \boldsymbol{G}(\boldsymbol{x}). \tag{2.15}$$

The following result, cited also in (Saleh, 2006, p32), whose proof can be found in Judge and Bock (1978), and has been generalized in Nkurunziza (2012a), is necessary for deriving ADB expressions for our estimators.

Theorem 2.5.2. Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_q)'$ be a $q-$dimensional normal vector distributed as $N_q(\boldsymbol{\mu}_y, \boldsymbol{I}_q)$. Then for any measurable function φ, we have
$E[\boldsymbol{y}\varphi(\boldsymbol{y}\boldsymbol{y}')] = \boldsymbol{\mu}_y E\{\varphi(\chi^2_{q+2}(\Delta^2))\}$, where $\Delta^2 = \frac{\boldsymbol{\mu}_y'\boldsymbol{\mu}_y}{2}$.

The ADB expressions of our estimators of $\boldsymbol{\beta}$ are given in the following Theorem.

Theorem 2.5.3. Under the assumptions of Theorem (2.5.1), we have

(i) $ADB(\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$

(ii) $ADB(\hat{\boldsymbol{\beta}}^R) = -\boldsymbol{\delta}$

(iii) $ADB(\hat{\boldsymbol{\beta}}^{PT}) = -\boldsymbol{\delta} H_{q+2}(\chi_q^2(\alpha); \Delta^2)$

(iv) $ADB(\hat{\boldsymbol{\beta}}^{JS}) = -(q-2)\boldsymbol{\delta} E(\chi_{q+2}^{-2}(\Delta^2))$

(v) $$\begin{aligned} ADB(\hat{\boldsymbol{\beta}}^{JS+}) &= ADB(\hat{\boldsymbol{\beta}}^{JS}) - \boldsymbol{\delta} E\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2))I(\chi_{q+2}^2(\Delta^2) < (q-2))\} \\ &= -\boldsymbol{\delta}\big[(q-2)E(\chi_{q+2}^{-2}(\Delta^2)) + E\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2))I(\chi_{q+2}^2(\Delta^2) < (q-2))\}\big] \end{aligned}$$

where $H_r(.)$ is the cumulative distribution function of non-central χ^2 random variable with r degrees of freedom and non centrality parameter $\Delta^2 = \frac{1}{\sigma^2}\boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi} = \frac{1}{\sigma^2}\boldsymbol{\delta}'\boldsymbol{C_0}\boldsymbol{\delta}$.

Proof:

(i) $ADB(\hat{\boldsymbol{\beta}}) = E\{\boldsymbol{T}^{(1)}\} = 0$, by Theorem 2.5.1(i).

(ii) $ADB(\hat{\boldsymbol{\beta}}^R) = E\{\boldsymbol{T}^{(2)}\} = -\boldsymbol{\delta}$, by Theorem 2.5.1(ii).

(iii) Note that,

$$\begin{aligned} \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta}\right) &= \sqrt{n}\left(\hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)I(\mathcal{L}_n \le \mathcal{L}_{n,\alpha}) - \boldsymbol{\beta}\right) \\ &= \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \sqrt{n}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)I(\mathcal{L}_n \le \mathcal{L}_{n,\alpha}) \\ &= \boldsymbol{T}_n^{(1)} - \Big\{\sqrt{n}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1} \\ &\quad (\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})I(\mathcal{L}_n \le \mathcal{L}_{n,\alpha})\Big\}. \end{aligned}$$

Now, as $n \to \infty$, with Slutsky's Theorem, we have $\mathcal{L}_n \xrightarrow{D} \mathcal{L} \sim \chi_q^2$, and $\mathcal{L}_{n,\alpha} \xrightarrow{D} \chi_{q;\alpha}^2$, the upper $\alpha-$quantile of the χ_q^2, and

$$\sqrt{n}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}) \xrightarrow{D} N_q(\boldsymbol{\xi}, \sigma^2(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')).$$

Thus, using Theorem 2.5.2,

$$\begin{aligned} ABD(\hat{\boldsymbol{\beta}}^{PT}) &= -\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\ &= -\boldsymbol{\delta}H_{q+2}(\chi_q^2(\alpha);\Delta^2). \end{aligned}$$

(iv) Note that,

$$\begin{aligned} \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta}\right) &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}-(q-2)(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})\mathcal{L}_n^{-1}-\boldsymbol{\beta}\right) \\ &= \sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})-(q-2)\sqrt{n}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})\mathcal{L}_n^{-1} \\ &= \boldsymbol{T}_n^{(1)}-(q-2)\left(\boldsymbol{T}_n^{(3)}\mathcal{L}_n^{-1}\right). \\ \text{Thus, } ADB(\hat{\boldsymbol{\beta}}^{JS}) &= -(q-2)E\{\boldsymbol{T}^{(3)}\mathcal{L}^{-1}\} \\ &= -(q-2)\boldsymbol{\delta}E(\chi_{q+2}^{-2}(\Delta^2)),\ \text{using Theorem 2.5.2.} \end{aligned}$$

(v) Note that,

$$\begin{aligned} \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{JS+}-\boldsymbol{\beta}\right) &= \sqrt{n}\Big(\hat{\boldsymbol{\beta}}^{JS}-(1-(q-2)\mathcal{L}_n^{-1})I(\mathcal{L}_n<(q-2)) \\ &\quad (\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})-\boldsymbol{\beta}\Big) \\ &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta}\right)-\Big\{\sqrt{n}((\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})(1-(q-2)\mathcal{L}_n^{-1}) \\ &\quad I(\mathcal{L}_n<(q-2)))\Big\} \\ &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta}\right)-\left\{\boldsymbol{T}_n^{(3)}(1-(q-2)\mathcal{L}_n^{-1})I(\mathcal{L}_n<(q-2))\right\} \end{aligned}$$

Therefore,

$$
\begin{aligned}
ADB(\hat{\boldsymbol{\beta}}^{JS+}) &= ADB(\hat{\boldsymbol{\beta}}^{JS}) - E\Big\{\boldsymbol{T}^{(3)}(1-(q-2)\mathcal{L}^{-1})I(\mathcal{L} < (q-2))\Big\} \\
&= -(q-2)\boldsymbol{\delta}E(\chi^{-2}_{q+2}(\Delta^2)) - \boldsymbol{\delta}E\Big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2)) \\
&\quad I(\chi^2_{q+2}(\Delta^2) < (q-2))\Big\} \\
&= -\boldsymbol{\delta}\Big[(q-2)E(\chi^{-2}_{q+2}(\Delta^2)) + E\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2)) \\
&\quad I(\chi^2_{q+2}(\Delta^2) < (q-2))\}\Big].
\end{aligned}
$$

□

2.5.3 Asymptotic Quadratic Risk

For any estimator $\hat{\boldsymbol{\beta}}^*$ of $\boldsymbol{\beta}$, define the quadratic loss as

$$
\begin{aligned}
L(\hat{\boldsymbol{\beta}}^*, \boldsymbol{\beta}) &= n(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})'\boldsymbol{M}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) \\
&= tr\Big\{\boldsymbol{M}\Big(n(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})'\Big)\Big\}, \qquad (2.16)
\end{aligned}
$$

where $\boldsymbol{M}$ is a $p \times p$ positive definite matrix. If $\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{T}^*$, then the asymptotic mean squared error matrix of $\hat{\boldsymbol{\beta}}^*$ is defined by

$$
\begin{aligned}
AMSEM(\hat{\boldsymbol{\beta}}^*) &= E\{\boldsymbol{T}^*\boldsymbol{T}^{*\prime}\} \\
&= \int \boldsymbol{x}\boldsymbol{x}'d\boldsymbol{G}(\boldsymbol{x}), \qquad (2.17)
\end{aligned}
$$

where $\boldsymbol{G}(\boldsymbol{x})$ is defined in (2.14).

Finally, define the asymptotic quadratic risk (AQR) as

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^*, \boldsymbol{M}) &= E\{\boldsymbol{T}^{*\prime}\boldsymbol{M}\boldsymbol{T}^*\} \\ &= \int (\boldsymbol{x}'\boldsymbol{M}\boldsymbol{x})\, d\boldsymbol{G}(\boldsymbol{x}) \\ &= tr\big\{\boldsymbol{M} AMSEM(\hat{\boldsymbol{\beta}}^*)\big\}. \end{aligned} \tag{2.18}$$

Again, the following result cited in (Saleh, 2006, p32), whose proof can be found in Judge and Bock (1978), and has been generalized in Nkurunziza (2012a), is necessary for deriving AMSEM and AQR expressions for our estimators.

Theorem 2.5.4. Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_q)'$ be a $q-$dimensional normal vector distributed as $N_q(\boldsymbol{\mu}_y, \boldsymbol{I}_q)$. Then for any measurable function φ, we have

$$E[\boldsymbol{y}\boldsymbol{y}'\varphi(\boldsymbol{y}\boldsymbol{y}')] = I_q E\{\varphi(\chi^2_{q+2}(\Delta^2))\} + \boldsymbol{\mu}_y\boldsymbol{\mu}_y' E\{\varphi(\chi^2_{q+4}(\Delta^2))\}.$$

Theorem 2.5.5. Under the assumptions of Theorem (2.5.1) and for $\boldsymbol{M}$ defined above, we have

(i) $AMSEM(\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{C}_0^{-1}$,

$AQR(\hat{\boldsymbol{\beta}}, M) = \sigma^2 tr(\boldsymbol{M}\boldsymbol{C}_0^{-1})$,

(ii) $AMSEM(\hat{\boldsymbol{\beta}}^R) = \sigma^2\boldsymbol{A}_0 + \boldsymbol{\delta}\boldsymbol{\delta}'$,

$AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) = \sigma^2 tr\left(\boldsymbol{M}\boldsymbol{C}_0^{-1}\right) - \sigma^2 tr(\boldsymbol{V}_{11}) + \boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1$,

(iii)

$$
\begin{aligned}
AMSEM(\hat{\boldsymbol{\beta}}^{PT}) &= \sigma^2\boldsymbol{C}_0^{-1} - \sigma^2\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}H\boldsymbol{C}_0^{-1} \\
& \quad H_{q+2}(\chi_q^2(\alpha);\Delta^2) + \boldsymbol{\delta}\boldsymbol{\delta}'\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\
&- H_{q+4}(\chi_q^2(\alpha);\Delta^2)\}, \\
AQR(\hat{\boldsymbol{\beta}}^{PT},\boldsymbol{M}) &= \sigma^2 tr(\boldsymbol{M}\boldsymbol{C}_0^{-1}) - \sigma^2 tr(\boldsymbol{V}_{11})H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\
&+ \boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2) - H_{q+4}(\chi_q^2(\alpha);\Delta^2)\},
\end{aligned}
$$

(iv)

$$
\begin{aligned}
AMSEM(\hat{\boldsymbol{\beta}}^{JS}) &= \sigma^2\boldsymbol{C}_0^{-1} - (q-2)\sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0)\big\{2E(\chi_{q+2}^{-2}(\Delta^2)) - \\
& \quad (q-2)E(\chi_{q+2}^{-4}(\Delta^2))\big\} + (q-2)(q+2)\boldsymbol{\delta}\boldsymbol{\delta}'E(\chi_{q+4}^{-4}(\Delta^2)), \\
AQR(\hat{\boldsymbol{\beta}}^{JS},\boldsymbol{M}) &= \sigma^2 tr(\boldsymbol{M}\boldsymbol{C}_0^{-1}) - \sigma^2(q-2)\big\{2E(\chi_{q+2}^{-2}(\Delta^2)) \\
&- (q-2)E(\chi_{q+2}^{-4}(\Delta^2))\big\}tr(\boldsymbol{V}_{11}) \\
&+ (q-2)(q+2)E(\chi_{q+4}^{-4}(\Delta^2))\boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1,
\end{aligned}
$$

(v)

$$
\begin{aligned}
AMSEM(\hat{\boldsymbol{\beta}}^{JS+}) &= AMSEM(\hat{\boldsymbol{\beta}}^{JS}) - \sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0) \\
& \quad E\big\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2))^2 I(\chi_{q+2}^2(\Delta^2) \le (q-2))\big\} \\
&- \boldsymbol{\delta}\boldsymbol{\delta}'E\big\{(1-(q-2)\chi_{q+4}^{-2}(\Delta^2))^2 I(\chi_{q+4}^2(\Delta^2) < (q-2))\big\} \\
&+ 2\boldsymbol{\delta}\boldsymbol{\delta}'E\big\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2))I(\chi_{q+2}^2(\Delta^2) < (q-2))\big\}, \\
AQR(\hat{\boldsymbol{\beta}}^{JS+},\boldsymbol{M}) &= AQR(\hat{\boldsymbol{\beta}}^{JS},\boldsymbol{M})
\end{aligned}
$$

$$
\begin{aligned}
- &\ \sigma^2\big\{E\left((1-(q-2)\chi^{-2}_{q+2}(\Delta^2))^2 I(\chi^2_{q+2}(\Delta^2)<(q-2))\right)\big\}tr(\boldsymbol{V}_{11})\\
- &\ E\big\{(1-(q-2)\chi^{-2}_{q+4}(\Delta^2))^2 I(\chi^2_{q+4}(\Delta^2)<(q-2))\big\}\boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1\\
+ &\ 2E\big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))I(\chi^2_{q+2}(\Delta^2)<(q-2))\big\}\boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1,
\end{aligned}
$$

where $\boldsymbol{A}_0=\boldsymbol{C}_0^{-1}-\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1}$, $\boldsymbol{\eta}=\begin{pmatrix}\boldsymbol{\eta}_1\\ \boldsymbol{\eta}_2\end{pmatrix}=\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}'$ $(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}$ and $\boldsymbol{\Gamma}$ a $p\times p$ orthogonal matrix such that

$$
\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'=\begin{pmatrix} I_q & \mathbf{0}_{q\times(p-q)}\\ \mathbf{0}_{(p-q)\times q} & \mathbf{0}_{(p-q)\times(p-q)}\end{pmatrix},
$$

and $\boldsymbol{\Gamma}C_0^{-1/2}\boldsymbol{M}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'=\begin{pmatrix}\boldsymbol{V}_{11} & \boldsymbol{V}_{12}\\ \boldsymbol{V}_{12}' & \boldsymbol{V}_{22}\end{pmatrix}$.

Proof:

(i) Note that,

$$
\begin{aligned}
n(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})' &= \boldsymbol{T}_n^{(1)}\boldsymbol{T}_n^{(1)\prime}.\text{ Therefore by Theorem 2.5.1(i), we have}\\
AMSEM(\hat{\boldsymbol{\beta}}) &= E\{\boldsymbol{T}^{(1)}\boldsymbol{T}^{(1)\prime}\}=\sigma^2\boldsymbol{C}_0^{-1},\\
AQR(\hat{\boldsymbol{\beta}},\boldsymbol{M}) &= tr\{\boldsymbol{M}AMSEM(\hat{\boldsymbol{\beta}})\}=tr\{\boldsymbol{M}(\sigma^2\boldsymbol{C}_0^{-1})\}=\sigma^2 tr(\boldsymbol{M}\boldsymbol{C}_0^{-1}).
\end{aligned}
$$

(ii) Note that,

$$\begin{aligned}
n(\hat{\boldsymbol{\beta}}^R-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^R-\boldsymbol{\beta})' &= \boldsymbol{T}_n^{(2)}\boldsymbol{T}_n^{(2)\prime}. \text{ Therefore by Theorem 2.5.1(ii), we have}\\
AMSEM(\hat{\boldsymbol{\beta}}^R) &= E\big\{\boldsymbol{T}^{(2)}\boldsymbol{T}^{(2)\prime}\big\}\\
&= \sigma^2\boldsymbol{A}_0+(-\boldsymbol{\delta})(-\boldsymbol{\delta}')=\sigma^2\boldsymbol{A}_0+\boldsymbol{\delta}\boldsymbol{\delta}',\\
AQR(\hat{\boldsymbol{\beta}}^R,\boldsymbol{M}) &= tr\{\boldsymbol{M}AMSEM(\hat{\boldsymbol{\beta}}^R)\}\\
&= tr\big\{\boldsymbol{M}(\sigma^2\boldsymbol{A}_0+\boldsymbol{\delta}\boldsymbol{\delta}')\big\}\\
&= tr\big\{\boldsymbol{M}\big[\sigma^2(\boldsymbol{C}_0^{-1}-\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1})\\
&+ \boldsymbol{\delta}\boldsymbol{\delta}'\big]\big\}\\
&= \sigma^2 tr(\boldsymbol{M}\boldsymbol{C}_0^{-1})-\sigma^2 tr\Big(\boldsymbol{M}\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\\
&\quad \boldsymbol{H}\boldsymbol{C}_0^{-1}\Big)+\boldsymbol{\delta}'\boldsymbol{M}\boldsymbol{\delta}.
\end{aligned}$$

Note that the matrix $\boldsymbol{C}_0^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1/2}$ is symmetric and idempotent of rank $q(q \le p)$, thus, there exists an orthogonal $p\times p$ matrix $\boldsymbol{\Gamma}$ such that

$$\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}' = \begin{pmatrix} \boldsymbol{I}_q & \boldsymbol{0}_{q\times(p-q)} \\ \boldsymbol{0}_{(p-q)\times q} & \boldsymbol{0}_{(p-q)\times(p-q)} \end{pmatrix}, \quad (2.19)$$

and

$$\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{M}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}' = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{12}' & \boldsymbol{V}_{22} \end{pmatrix}.$$

where $\boldsymbol{V}_{11}$ and $\boldsymbol{V}_{22}$ are square matrices of orders q and $(p-q)$, respectively.

Further, $\boldsymbol{\delta}'\boldsymbol{M}\boldsymbol{\delta}$ can be written as

$$\begin{aligned}
\boldsymbol{\delta}'\boldsymbol{M}\boldsymbol{\delta} &= \left[\boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1}\right]\boldsymbol{M}\left[\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}\right] \\
&= \left[\boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}H\boldsymbol{C}_0^{-1}\right]\boldsymbol{M} \\
&\quad \left[\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}\right] \\
&= \left[\boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\right. \\
&\quad \left.\boldsymbol{H}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\right]\boldsymbol{M}\left[\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}'\right. \\
&\quad \left.(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}\right] \\
&= \left[\boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'\right]\left[\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\right. \\
&\quad \left.\boldsymbol{H}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'\right]\left[\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{M}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'\right]\left[\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\right. \\
&\quad \left.\boldsymbol{H}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'\right]\left[\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}\right].
\end{aligned}$$

By letting $\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}$, $\boldsymbol{\delta}'\boldsymbol{M}\boldsymbol{\delta}$ can be written as

$$\begin{aligned}
\boldsymbol{\delta}'\boldsymbol{M}\boldsymbol{\delta} &= \boldsymbol{\eta}'\begin{pmatrix} \boldsymbol{I}_q & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{12}' & \boldsymbol{V}_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{I}_q & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{\eta} \\
&= \boldsymbol{\eta}'\begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{\eta} \\
&= \boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1.
\end{aligned}$$

Also,

$$\begin{aligned}
tr\big[\boldsymbol{M}\boldsymbol{C}_0^{-1}(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1}\big] &= tr\big[\boldsymbol{M}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}' \\
&\quad (\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}HC_0^{-1/2}C_0^{-1/2}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\big] \\
&= tr\Big[\big(\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{M}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'\big)\big(\boldsymbol{\Gamma}\boldsymbol{C}_0^{-1/2}\boldsymbol{H}' \\
&\quad (\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1/2}\boldsymbol{\Gamma}'\big)\Big] \\
&= tr\left(\begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{12}' & \boldsymbol{V}_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{I}_q & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\right) \\
&= tr(\boldsymbol{V}_{11}).
\end{aligned}$$

Thus,

$$AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) = \sigma^2 tr(\boldsymbol{M}\boldsymbol{C}_0^{-1}) - \sigma^2 tr(\boldsymbol{V}_{11}) + \boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1.$$

(iii) Note that,

$$\begin{aligned}
n(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta})' &= n\Big[\hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)I(\mathcal{L}_n \le \mathcal{L}_{n,\alpha}) - \boldsymbol{\beta}\Big]\Big[\hat{\boldsymbol{\beta}} - \\
&\quad (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)I(\mathcal{L}_n \le \mathcal{L}_{n,\alpha}) - \boldsymbol{\beta}\Big]' \\
&= n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' + n\big[(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R) \\
&\quad (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)'I^2(\mathcal{L}_n \le \mathcal{L}_{n,\alpha})\big] \\
&- 2n\big[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)'I(\mathcal{L}_n \le \mathcal{L}_{n,\alpha})\big] \\
&= \boldsymbol{T}_n^{(1)}\boldsymbol{T}_n^{(1)\prime} + \boldsymbol{T}_n^{(3)}\boldsymbol{T}_n^{(3)\prime}I^2(\mathcal{L}_n \le \mathcal{L}_{n,\alpha}) \\
&- 2\boldsymbol{T}_n^{(1)}\boldsymbol{T}_n^{(3)\prime}I(\mathcal{L}_n \le \mathcal{L}_{n,\alpha}).
\end{aligned}$$

Note that,

$$\begin{aligned}
\boldsymbol{T}_n^{(3)}\boldsymbol{T}_n^{(3)\prime}I^2(\mathcal{L}_n \le \mathcal{L}_{n,\alpha}) &= s_e^2(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}'(H(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1/2} \\
&\quad \left[s_e^2(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')\right]^{-1/2}\sqrt{n}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h}) \\
&\quad \left[s_e^2(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')\right]^{-1/2}\sqrt{n}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h})' \\
&\quad I(\mathcal{L}_n \le \mathcal{L}_{n,\alpha})(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')^{-1/2}\boldsymbol{H} \\
&\quad (\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}.
\end{aligned}$$

Now as $n \longrightarrow \infty$, we have $\sqrt{n}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h}) \xrightarrow{D} N_q(\boldsymbol{\xi}, \sigma^2(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}'))$,

$$\begin{aligned}
&\left[s_e^2(\boldsymbol{H}(\boldsymbol{X}_n'\hat{C}_n^{-1}\boldsymbol{X}_n)^{-1}\boldsymbol{H}')\right]^{-1/2}.\sqrt{n}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h}) \\
&\quad \xrightarrow{D} N_q\left((\sigma^2(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1})^{-1/2}\boldsymbol{\xi}, \boldsymbol{I}_q\right).
\end{aligned}$$

Also, $\boldsymbol{T}^{(1)} \mid \boldsymbol{T}^{(3)} \xrightarrow{D} N_p(\boldsymbol{T}^{(3)} - \boldsymbol{\delta}, \sigma^2\boldsymbol{A}_0)$. Therefore,

$$AMSEM(\hat{\boldsymbol{\beta}}^{PT}) = E_1 + E_2 + E_3,$$

where, by Theorem 2.5.1(i) E_1 is given by:,

$$E_1 = E\{\boldsymbol{T}^{(1)}\boldsymbol{T}^{(1)\prime}\} = \sigma^2\boldsymbol{C}_0^{-1},$$

$$E_2 = E\big\{\boldsymbol{T}^{(3)}\boldsymbol{T}^{(3)\prime}I^2(\mathcal{L} \le \chi_q^2(\alpha);\Delta^2)\big\}.$$

By using Theorem 2.5.4 E_2 is given by:

$$\begin{aligned} E_2 &= \sigma^2 \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1} H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\ &+ \boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi} \\ &\quad \boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1} H_{q+4}(\chi_q^2(\alpha);\Delta^2), \\ &= \sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0) H_{q+2}(\chi_q^2(\alpha);\Delta^2) + \boldsymbol{\delta}\boldsymbol{\delta}' H_{q+4}(\chi_q^2(\alpha);\Delta^2), \end{aligned}$$

$$E_3 = -2E\big\{\boldsymbol{T}^{(1)}\boldsymbol{T}^{(3)\prime} I(\mathcal{L} \le \chi_q^2(\alpha);\Delta^2)\big\},$$

by using Theorem 2.5.3(iii) E_3 is given by

$$\begin{aligned} E_3 &= -2E\Big\{E\big\{\boldsymbol{T}^{(1)}\boldsymbol{T}^{(3)\prime} I(\mathcal{L} \le \chi_q^2(\alpha);\Delta^2) \mid \boldsymbol{T}^{(3)}\big\}\Big\} \\ &= -2E\Big\{\big(\boldsymbol{T}^{(3)} - \boldsymbol{\delta}\big)\boldsymbol{T}^{(3)\prime} I(\mathcal{L} \le \chi_q^2(\alpha);\Delta^2)\Big\} \\ &= -2 \times (\text{Second term}) + 2\boldsymbol{\delta}\boldsymbol{\delta}' H_{q+2}(\chi_q^2(\alpha);\Delta^2). \end{aligned}$$

Finally,

$$\begin{aligned} AMSEM(\hat{\boldsymbol{\beta}}^{PT}) &= E_1 + E_2 + E_3 \\ &= \sigma^2\boldsymbol{C}_0^{-1} - \sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0) H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\ &- \boldsymbol{\delta}\boldsymbol{\delta}' H_{q+4}(\chi_q^2(\alpha);\Delta^2) + 2\boldsymbol{\delta}\boldsymbol{\delta}' H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\ &= \sigma^2\boldsymbol{C}_0^{-1} - \sigma^2\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1} \\ &\quad H_{q+2}(\chi_q^2(\alpha);\Delta^2) + \boldsymbol{\delta}\boldsymbol{\delta}'\Big\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\ &- H_{q+4}(\chi_q^2(\alpha);\Delta^2)\Big\}. \end{aligned}$$

Following the same procedure as in part (ii), the $AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M})$ will be

$$
\begin{aligned}
AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M}) &= tr(\boldsymbol{M} AMSEM(\hat{\boldsymbol{\beta}}^{PT})) \\
&= tr\Big\{\boldsymbol{M}\sigma^2\boldsymbol{C}_0^{-1} - \sigma^2\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1} \\
&\quad H_{q+2}(\chi_q^2(\alpha);\Delta^2) + \boldsymbol{\delta}\boldsymbol{\delta}'\big\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\
&- H_{q+4}(\chi_q^2(\alpha);\Delta^2)\big\}\Big\} \\
&= \sigma^2 tr(\boldsymbol{M}\boldsymbol{C}_0^{-1}) - \sigma^2 tr(\boldsymbol{M}\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{C}_0^{-1}) \\
&\quad H_{q+2}(\chi_q^2(\alpha);\Delta^2) + \boldsymbol{\delta}'\boldsymbol{M}\boldsymbol{\delta}\big\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\
&- H_{q+4}(\chi_q^2(\alpha);\Delta^2)\big\} \\
&= \sigma^2 tr(\boldsymbol{M}\boldsymbol{C}_0^{-1}) - \sigma^2 tr(\boldsymbol{V}_{11}) H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\
&+ \boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1\big\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2) - H_{q+4}(\chi_q^2(\alpha);\Delta^2)\big\}.
\end{aligned}
$$

(iv) Note that,

$$
\begin{aligned}
n(\hat{\boldsymbol{\beta}}^{JS} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{JS} - \boldsymbol{\beta})' &= n(\hat{\boldsymbol{\beta}}^{JS} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{JS} - \boldsymbol{\beta})' \\
&= n\left(\hat{\boldsymbol{\beta}} - (q-2)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})\mathcal{L}_n^{-1} - \boldsymbol{\beta}\right) \\
&\quad \left(\hat{\boldsymbol{\beta}} - (q-2)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})\mathcal{L}_n^{-1} - \boldsymbol{\beta}\right)' \\
&= n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \\
&+ n(q-2)^2(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})'\mathcal{L}_n^{-2} \\
&- 2n(q-2)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})'\mathcal{L}_n^{-1} \\
&= \boldsymbol{T}_n^{(1)}\boldsymbol{T}_n^{(1)\prime} + (q-2)^2\boldsymbol{T}_n^{(3)}\boldsymbol{T}_n^{(3)\prime}\mathcal{L}_n^{-2} \\
&- 2(q-2)\boldsymbol{T}_n^{(1)}\boldsymbol{T}_n^{(3)\prime}\mathcal{L}_n^{-1}.
\end{aligned}
$$

Therefore,

$$AMSEM(\hat{\boldsymbol{\beta}}^{JS}) = E_1 + E_2 + E_3,$$

where E_1, E_2 and E_3 are, respectively

$$\begin{aligned}
E_1 &= E\{\boldsymbol{T}^{(1)}\boldsymbol{T}^{(1)\prime}\} = \sigma^2\boldsymbol{C}_0^{-1}, \\
E_2 &= (q-2)^2E\big\{\boldsymbol{T}^{(3)}\boldsymbol{T}^{(3)\prime}\mathcal{L}^{-2}\big\} \\
&= (q-2)^2\sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0)E(\chi^{-4}_{q+2}(\Delta^2)) + (q-2)^2\boldsymbol{\delta\delta}'E(\chi^{-4}_{q+4}(\Delta^2)),
\end{aligned}$$

using the same technique as in the previous part, and

$$\begin{aligned}
E_3 &= -2(q-2)E\big\{\boldsymbol{T}^{(1)}\boldsymbol{T}^{(3)\prime}\mathcal{L}^{-1}\big\} \\
&= -2(q-2)E\Big\{E\big\{\boldsymbol{T}^{(1)}\boldsymbol{T}^{(3)\prime}\mathcal{L}^{-1} \mid \boldsymbol{T}^{(3)}\big\}\Big\} \\
&= -2(q-2)E\Big\{\left(\boldsymbol{T}^{(3)} - \boldsymbol{\delta}\right)\boldsymbol{T}^{(3)\prime}\mathcal{L}^{-1}\Big\} \\
&= -2(q-2)\Big\{E\big\{\boldsymbol{T}^{(3)}\boldsymbol{T}^{(3)\prime}\mathcal{L}^{-1}\big\} - \boldsymbol{\delta}E\big\{\boldsymbol{T}^{(3)\prime}\mathcal{L}^{-1}\big\}\Big\} \\
&= -2(q-2)\Big\{\big[\sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0)E(\chi^{-2}_{q+2}(\Delta^2)) + \boldsymbol{\delta\delta}'E(\chi^{-2}_{q+4}(\Delta^2))\big] \\
&- \boldsymbol{\delta\delta}'E(\chi^{-2}_{q+2}(\Delta^2))\Big\} \\
&= -2(q-2)\sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0)E(\chi^{-2}_{q+2}(\Delta^2)) - 2(q-2)\boldsymbol{\delta\delta}'\big\{E(\chi^{-2}_{q+4}(\Delta^2)) \\
&- E(\chi^{-2}_{q+2}(\Delta^2))\big\}.
\end{aligned}$$

By using the following result (Saleh, 2006, p33)

$$E(\chi^{-2}_{q+4}(\Delta^2)) = E(\chi^{-2}_{q+2}(\Delta^2)) - 2E(\chi^{-4}_{q+4}(\Delta^2)), \tag{2.20}$$

E_3 can be simplified to

$$E_3 = -2(q-2)\sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0)E(\chi_{q+2}^{-2}(\Delta^2)) + 4(q-2)\boldsymbol{\delta}\boldsymbol{\delta}' E(\chi_{q+4}^{-4}(\Delta^2)).$$

By combining E_1, E_2, and E_3 we have

$$\begin{aligned}
AMSEM(\hat{\boldsymbol{\beta}}^{JS}) &= \sigma^2\boldsymbol{C}_0^{-1} - (q-2)\sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0)\big\{2E(\chi_{q+2}^{-2}(\Delta^2)) \\
&- (q-2)E(\chi_{q+2}^{-4}(\Delta^2))\big\} + (q-2)(q+2)\boldsymbol{\delta}\boldsymbol{\delta}' E(\chi_{q+4}^{-4}(\Delta^2)), \\
AQR((\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) &= tr(\boldsymbol{M} AMSEM(\hat{\boldsymbol{\beta}}^{JS})) \\
&= tr\Big\{\boldsymbol{M}\Big[\sigma^2\boldsymbol{C}_0^{-1} - (q-2)\sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0)\Big\{2E(\chi_{q+2}^{-2}(\Delta^2)) \\
&- (q-2)E(\chi_{q+2}^{-4}(\Delta^2))\Big\} + (q-2)(q+2)\boldsymbol{\delta}\boldsymbol{\delta}' E(\chi_{q+4}^{-4}(\Delta^2))\Big]\Big\} \\
&= tr(\boldsymbol{M}\boldsymbol{C}_0^{-1}) - (q-2)\sigma^2 tr\Big(\boldsymbol{M}\boldsymbol{C}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{C}_0^{-1}\boldsymbol{H}')^{-1} \\
&\quad \boldsymbol{H}\boldsymbol{C}_0^{-1}\Big)\{2E(\chi_{q+2}^{-2}(\Delta^2)) - (q-2)E(\chi_{q+2}^{-4}(\Delta^2))\} \\
&+ (q-2)(q+2)E(\chi_{q+4}^{-4}(\Delta^2)) tr(\boldsymbol{\delta}'\boldsymbol{M}\boldsymbol{\delta}) \\
&= \sigma^2 tr(\boldsymbol{M}\boldsymbol{C}_0^{-1}) - (q-2)\sigma^2\Big\{2E(\chi_{q+2}^{-2}(\Delta^2)) \\
&- (q-2)E(\chi_{q+2}^{-4}(\Delta^2))\Big\} tr(\boldsymbol{V}_{11}) + (q-2)(q+2)E(\chi_{q+4}^{-4}(\Delta^2)) \\
&\quad \boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1.
\end{aligned}$$

(v) Note that,

$$\begin{aligned}
n(\hat{\boldsymbol{\beta}}^{JS+} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{JS+} - \boldsymbol{\beta})' &= n\big[\hat{\boldsymbol{\beta}}^{JS} - (1 - (q-2)\mathcal{L}_n^{-1})I(\mathcal{L}_n < (q-2)) \\
&\quad (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R}) - \boldsymbol{\beta}\big]\big[\hat{\boldsymbol{\beta}}^{JS} - (1 - (q-2)\mathcal{L}_n^{-1}) \\
&\quad I(\mathcal{L}_n < (q-2))(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R}) - \boldsymbol{\beta}\big]'
\end{aligned}$$

$$
\begin{aligned}
&= n(\hat{\boldsymbol{\beta}}^{JS} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{JS} - \boldsymbol{\beta})' \\
&+ n(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})'(1 - (q-2)\mathcal{L}_n^{-1})^2 I(\mathcal{L}_n < (q-2)) \\
&- 2n(\hat{\boldsymbol{\beta}}^{JS} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})'(1 - (q-2)\mathcal{L}_n^{-1}) I(\mathcal{L}_n < (q-2)). \qquad (2.21)
\end{aligned}
$$

Note that, the last term of (2.21) can be written as follows:

$$
\begin{aligned}
\text{Last term} &= -2n(\hat{\boldsymbol{\beta}}^{JS} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})'(1 - (q-2)\mathcal{L}_n^{-1}) I(\mathcal{L}_n < (q-2)) \\
&= -2n\Big(\hat{\boldsymbol{\beta}}^{R} + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})(1 - (q-2)\mathcal{L}_n^{-1}) I(\mathcal{L}_n < (q-2)) \\
&- \boldsymbol{\beta}\Big)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})'(1 - (q-2)\mathcal{L}_n^{-1}) I(\mathcal{L}_n < (q-2)) \\
&= -2n(\hat{\boldsymbol{\beta}}^{R} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})'(1 - (q-2)\mathcal{L}_n^{-1}) I(\mathcal{L}_n < (q-2)) \\
&- 2n(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})'(1 - (q-2)\mathcal{L}_n^{-1})^2 I(\mathcal{L}_n < (q-2)) \\
&= -2\boldsymbol{T}_n^{(2)}\boldsymbol{T}_n^{(3)\prime}(1 - (q-2)\mathcal{L}_n^{-1}) I(\mathcal{L}_n < (q-2)) \\
&- 2\boldsymbol{T}_n^{(3)}\boldsymbol{T}_n^{(3)\prime}(1 - (q-2)\mathcal{L}_n^{-1})^2 I(\mathcal{L}_n < (q-2)).
\end{aligned}
$$

Therefore, the $AMSEM(\hat{\boldsymbol{\beta}}^{JS+}) = E_1 + E_2 + E_3$, where

$$
\begin{aligned}
E_1 &= AMSEM(\hat{\boldsymbol{\beta}}^{JS}), \\
E_2 &= E\big\{\boldsymbol{T}^{(3)}\boldsymbol{T}^{(3)\prime}(1 - (q-2)\mathcal{L}^{-1})^2 I(\mathcal{L} < (q-2))\big\} \\
&= \sigma^2(\boldsymbol{C}_0^{-1} - \boldsymbol{A}_0)E\big\{(1 - (q-2)\chi_{q+2}^{-2}(\Delta^2))^2 I(\chi_{q+2}^{2}(\Delta^2) < (q-2))\big\} \\
&+ \boldsymbol{\delta}\boldsymbol{\delta}' E\big\{(1 - (q-2)\chi_{q+4}^{-2}(\Delta^2))^2 I(\chi_{q+4}^{2}(\Delta^2) < (q-2))\big\},
\end{aligned}
$$

using the same procedure as in part (iii), and

$$\begin{aligned}
E_3 &= -2E(\boldsymbol{T}^{(2)})E\left(\boldsymbol{T}^{(3)\prime}(1-(q-2)\mathcal{L}^{-1})I(\mathcal{L}<(q-2))\right) \\
&- 2E\left(\boldsymbol{T}^{(3)}\boldsymbol{T}^{(3)\prime}(1-(q-2)\mathcal{L}^{-1})^2 I(\mathcal{L}<(q-2))\right) \\
&= 2\boldsymbol{\delta}\boldsymbol{\delta}' E\Big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))I(\chi^2_{q+2}(\Delta^2)<(q-2))\Big\} \\
&- 2\sigma^2(\boldsymbol{C}_0^{-1}-\boldsymbol{A}_0)E\Big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))^2 I(\chi^2_{q+2}(\Delta^2)<(q-2))\Big\} \\
&- 2\boldsymbol{\delta}\boldsymbol{\delta}' E\Big\{(1-(q-2)\chi^{-2}_{q+4}(\Delta^2))^2 I(\chi^2_{q+4}(\Delta^2)<(q-2))\Big\},
\end{aligned}$$

where the last equality follows from Theorem 2.5.1(v). By combining E_1, E_2 and E_3, we have

$$\begin{aligned}
AMSEM(\hat{\boldsymbol{\beta}}^{JS+}) &= AMSEM(\hat{\boldsymbol{\beta}}^{JS}) - \sigma^2(\boldsymbol{C}_0^{-1}-\boldsymbol{A}_0) \\
&\quad E\Big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))^2 I(\chi^2_{q+2}(\Delta^2)<(q-2))\Big\} \\
&- \boldsymbol{\delta}\boldsymbol{\delta}' E\Big\{(1-(q-2)\chi^{-2}_{q+4}(\Delta^2))^2 I(\chi^2_{q+4}(\Delta^2)<(q-2))\Big\} \\
&+ 2\boldsymbol{\delta}\boldsymbol{\delta}' E\Big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))I(\chi^2_{q+2}(\Delta^2)<(q-2))\Big\}, \\
AQR(\hat{\boldsymbol{\beta}}^{JS+},\boldsymbol{M}) &= tr(\boldsymbol{M}AMSEM(\hat{\boldsymbol{\beta}}^{JS+})) \\
&= AQR(\hat{\boldsymbol{\beta}}^{JS}) \\
&- \sigma^2 E\Big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))^2 I(\chi^2_{q+2}(\Delta^2)<(q-2))\Big\} \\
&\quad tr(\boldsymbol{V}_{11}) \\
&- E\Big\{(1-(q-2)\chi^{-2}_{q+4}(\Delta^2))^2 I(\chi^2_{q+4}(\Delta^2)<(q-2))\Big\} \\
&\quad \boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1 \\
&+ 2E\Big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))I(\chi^2_{q+2}(\Delta^2)<(q-2))\Big\} \\
&\quad \boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1.
\end{aligned}$$

2.6 Risk Analysis

In this section, we will use the AQR measure constructed in the previous section in order to compare the estimators $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^R, \hat{\boldsymbol{\beta}}^{JS}, \hat{\boldsymbol{\beta}}^{JS+}$. As the AQR is a measure combining bias and variance of the estimators, we will limit ourself to comparisons in terms of AQR only and will not discuss comparisons in terms of ADB of the proposed estimators.

Definition 1. Let $\mathbf{B}$ be the parameter space of $\boldsymbol{\beta}$. If two estimators $\hat{\boldsymbol{\beta}}^*, \hat{\boldsymbol{\beta}}^{**}$ are such that $AQR(\hat{\boldsymbol{\beta}}^*, \boldsymbol{M}) \leq AQR(\hat{\boldsymbol{\beta}}^{**}, \boldsymbol{M})$ for all values of $\boldsymbol{\beta} \in \mathbf{B}$, with strict inequality for some values of $\boldsymbol{\beta}$, we say that $\hat{\boldsymbol{\beta}}^*$ dominates $\hat{\boldsymbol{\beta}}^{**}$.

2.6.1 Comparing $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^R$

It is obvious from the expressions in Theorem 2.5.5 that the AQR of $\hat{\boldsymbol{\beta}}$ is a constant, while the AQR of $\hat{\boldsymbol{\beta}}^R$ can be re-written as

$$AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) = AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{V}_{11}) + \boldsymbol{\eta}_1' \boldsymbol{V}_{11} \boldsymbol{\eta}_1.$$

By using Courant Theorem (Saleh, 2006, p.39), we have

$$\sigma^2 \Delta^2 ch_{min}(\boldsymbol{V}_{11}) \leq \boldsymbol{\eta}_1' \boldsymbol{V}_{11} \boldsymbol{\eta}_1 \leq \sigma^2 \Delta^2 ch_{max}(\boldsymbol{V}_{11}),$$

where $ch_{min}(\boldsymbol{V}_{11})$ and $ch_{max}(\boldsymbol{V}_{11})$ are, respectively, the smallest and largest characteristic roots (eigenvalues) of the matrix $\boldsymbol{V}_{11}$, and $\Delta^2\sigma^2 = \boldsymbol{\eta}_1'\boldsymbol{\eta}_1$. It follows that

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{V}_{11}) + \sigma^2\Delta^2 ch_{min}(\boldsymbol{V}_{11}) &\leq AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) \\ &\leq AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{V}_{11}) + \sigma^2\Delta^2 ch_{max}(\boldsymbol{V}_{11}). \end{aligned} \tag{2.22}$$

When $\Delta^2 = 0$, the lower and upper bounds on the $AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M})$ in this latter expression are equal and hence $\hat{\boldsymbol{\beta}}^R$ dominates $\hat{\boldsymbol{\beta}}$. Also, when $0 < \Delta^2 \leq \frac{tr(\boldsymbol{V}_{11})}{ch_{max}(\boldsymbol{V}_{11})}$, from the second part of (2.22), we get

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) \leq AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) &- \sigma^2 tr(\boldsymbol{V}_{11}) + \sigma^2\Delta^2 ch_{max}(\boldsymbol{V}_{11}) \\ AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) &\leq -\sigma^2 tr(\boldsymbol{V}_{11}) + \sigma^2\Delta^2 ch_{max}(\boldsymbol{V}_{11}) \\ &\leq 0, \end{aligned}$$

which means that also in the above interval $\hat{\boldsymbol{\beta}}^R$ dominates $\hat{\boldsymbol{\beta}}$. Finally, if $\Delta^2 \geq \frac{tr(\boldsymbol{V}_{11})}{ch_{min}(\boldsymbol{V}_{11})}$, then from the first part of the inequality (2.22), we get

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) \geq AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) &- \sigma^2 tr(\boldsymbol{V}_{11}) + \sigma^2\Delta^2 ch_{min}(\boldsymbol{V}_{11}) \\ AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) &\geq -\sigma^2 tr(\boldsymbol{V}_{11}) + \sigma^2\Delta^2 ch_{min}(\boldsymbol{V}_{11}), \end{aligned}$$

and hence, $\hat{\boldsymbol{\beta}}$ performs better than $\hat{\boldsymbol{\beta}}^R$. In fact, the risk of $\hat{\boldsymbol{\beta}}^R$ becomes unbounded beyond $\frac{tr(\boldsymbol{V}_{11})}{ch_{min}(\boldsymbol{V}_{11})}$.

2.6.2 Comparing $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{PT}$

The asymptotic quadratic risk of $\hat{\boldsymbol{\beta}}^{PT}$ can be re-written in terms of $AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M})$ as

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M}) &= AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{V}_{11}) H_{q+2}(\chi_q^2(\alpha); \Delta^2) \\ &+ \boldsymbol{\eta}_1' \boldsymbol{V}_{11} \boldsymbol{\eta}_1 \{2H_{q+2}(\chi_q^2(\alpha); \Delta^2) - H_{q+4}(\chi_q^2(\alpha); \Delta^2)\}. \end{aligned}$$

Again, by using Courant's Theorem, we get

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) &- \sigma^2 tr(\boldsymbol{V}_{11}) H_{q+2}(\chi_q^2(\alpha); \Delta^2) \\ &+ \sigma^2 \Delta^2 ch_{min}(\boldsymbol{V}_{11}) \big\{2H_{q+2}(\chi_q^2(\alpha); \Delta^2) - H_{q+4}(\chi_q^2(\alpha); \Delta^2)\big\} \\ &\leq AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M}) \leq AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{V}_{11}) H_{q+2}(\chi_q^2(\alpha); \Delta^2) \\ &+ \sigma^2 \Delta^2 ch_{max}(\boldsymbol{V}_{11}) \big\{2H_{q+2}(\chi_q^2(\alpha); \Delta^2) - H_{q+4}(\chi_q^2(\alpha); \Delta^2)\big\}. \end{aligned} \tag{2.23}$$

Now, from the second part of (2.23), we have

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) &\leq \sigma^2 \Big(\Delta^2 ch_{max}(\boldsymbol{V}_{11}) \big\{2H_{q+2}(\chi_q^2(\alpha); \Delta^2) \\ &- H_{q+4}(\chi_q^2(\alpha); \Delta^2)\big\} - tr(\boldsymbol{V}_{11}) \\ &\quad H_{q+2}(\chi_q^2(\alpha); \Delta^2)\Big) \\ &\leq 0, \end{aligned}$$

whenever $\Delta^2 \leq \frac{tr(\boldsymbol{V}_{11}) H_{q+2}(\chi_q^2(\alpha); \Delta^2)}{ch_{max}(\boldsymbol{V}_{11}) \left\{2H_{q+2}(\chi_q^2(\alpha); \Delta^2) - H_{q+4}(\chi_q^2(\alpha); \Delta^2)\right\}}$. This means that $\hat{\boldsymbol{\beta}}^{PT}$ performs better than $\hat{\boldsymbol{\beta}}$ for all $\Delta^2 \in \left[0, \frac{tr(\boldsymbol{V}_{11}) H_{q+2}(\chi_q^2(\alpha); \Delta^2)}{ch_{max}(\boldsymbol{V}_{11}) \left\{2H_{q+2}(\chi_q^2(\alpha); \Delta^2) - H_{q+4}(\chi_q^2(\alpha); \Delta^2)\right\}}\right)$.

On the other hand, from the first part of (2.23), we have

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^{PT},\boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}},\boldsymbol{M}) &\geq \sigma^2\Big(\Delta^2 Ch_{min}(\boldsymbol{V}_{11}) \\ &\quad \big\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2) - H_{q+4}(\chi_q^2(\alpha);\Delta^2)\big\} \\ &- tr(\boldsymbol{V}_{11})H_{q+2}(\chi_q^2(\alpha);\Delta^2)\Big) \\ &\geq 0, \end{aligned}$$

whenever $\Delta^2 \geq \frac{tr(\boldsymbol{V}_{11})H_{q+2}(\chi_q^2(\alpha);\Delta^2)}{ch_{min}(\boldsymbol{V}_{11})\left\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2)-H_{q+4}(\chi_q^2(\alpha);\Delta^2)\right\}}$. That is $\hat{\boldsymbol{\beta}}$ performs better than $\hat{\boldsymbol{\beta}}^{PT}$ for $\Delta^2 \in \left[\frac{tr(\boldsymbol{V}_{11})H_{q+2}(\chi_q^2(\alpha);\Delta^2)}{ch_{min}(\boldsymbol{V}_{11})\left\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2)-H_{q+4}(\chi_q^2(\alpha);\Delta^2)\right\}}, \infty\right)$.

When $\Delta^2 = 0$, the lower and the upper bounds of $AQR(\hat{\boldsymbol{\beta}}^{PT},\boldsymbol{M})$ are equal, hence we get,

$$AQR(\hat{\boldsymbol{\beta}},\boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}^{PT},\boldsymbol{M}) = \sigma^2 tr(\boldsymbol{V}_{11})H_{q+2}(\chi_q^2(\alpha);\Delta^2) \geq 0.$$

Therefore, $\hat{\boldsymbol{\beta}}^{PT}$ performs better than $\hat{\boldsymbol{\beta}}$ at $\Delta^2 = 0$.

2.6.3 Comparing $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{JS}$

In order to compare the $AQR(\hat{\boldsymbol{\beta}}^{JS},\boldsymbol{M})$ with the $AQR(\hat{\boldsymbol{\beta}},\boldsymbol{M})$ we use the following identity

$$\Delta^2 E(\chi_{q+4}^{-4}(\Delta^2)) = E(\chi_{q+2}^{-2}(\Delta^2)) - (q-2)E(\chi_{q+2}^{-4}(\Delta^2)). \qquad (2.24)$$

Now, let us re-write the asymptotic quadratic risk of $\hat{\boldsymbol{\beta}}^{JS}$ in terms of $AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M})$,

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) &= AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{V}_{11})(q-2)\big\{2E(\chi^{-2}_{q+2}(\Delta^2)) \\ &- (q-2)E(\chi^{-4}_{q+2}(\Delta^2))\big\} + (q-2)(q+2)E(\chi^{-4}_{q+4}(\Delta^2))\boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1. \end{aligned}$$

Using (2.24), we get

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) &= AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{V}_{11})(q-2) \\ &\quad \big\{(q-2)E(\chi^{-4}_{q+2}(\Delta^2)) + 2\Delta^2 E(\chi^{-4}_{q+4}(\Delta^2))\big\} \\ &+ (q-2)(q+2)E(\chi^{-4}_{q+2}(\Delta^2))\boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1 \\ &= AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{V}_{11})(q-2)\Bigg\{(q-2)E(\chi^{-4}_{q+2}(\Delta^2)) \\ &+ 2\Delta^2 E(\chi^{-4}_{q+4}(\Delta^2)) - (q+2)E(\chi^{-4}_{q+4}(\Delta^2))\frac{\boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1}{\sigma^2 tr(\boldsymbol{V}_{11})}\Bigg\} \\ &= AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{V}_{11})(q-2)\Bigg\{(q-2)E(\chi^{-4}_{q+2}(\Delta^2)) \\ &+ 2\Delta^2 E(\chi^{-4}_{q+4}(\Delta^2))\left[1 - \frac{(q+2)\boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1}{2\Delta^2\sigma^2 tr(\boldsymbol{V}_{11})}\right]\Bigg\}. \end{aligned}$$

From these inequalities and from Courant's Theorem, we see that $AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) \leq AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M})$ for all $\Delta^2, \boldsymbol{M}$ if

$$1 - \frac{(q+2)\boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1}{2\Delta^2\sigma^2 tr(\boldsymbol{V}_{11})} \geq 0, \tag{2.25}$$

$$\frac{tr(\boldsymbol{V}_{11})}{ch_{max}(\boldsymbol{V}_{11})} \geq \frac{q+2}{2}, \quad q \geq 3. \tag{2.26}$$

Therefore, $AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M})$ is less than or equal to $AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M})$ in the whole parameter space, provided that the inequality (2.26) holds.

2.6.4 Comparing $\hat{\boldsymbol{\beta}}^{JS}$ and $\hat{\boldsymbol{\beta}}^{JS+}$

From part (v) of Theorem 2.5.5, we can re-write the asymptotic risk difference between $\hat{\boldsymbol{\beta}}^{JS+}$ and $\hat{\boldsymbol{\beta}}^{JS}$ as

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}^{JS+}, \boldsymbol{M}) &= \sigma^2 tr(\boldsymbol{V}_{11}) \\ & \quad E\Big((1-(q-2)\chi^{-2}_{q+2}(\Delta^2))^2 I(\chi^2_{q+2}(\Delta^2) < \\ & \quad (q-2))\Big) \\ &+ E\Big((1-(q-2)\chi^{-2}_{q+4}(\Delta^2))^2 I(\chi^2_{q+4}(\Delta^2) < \\ & \quad (q-2))\Big)\boldsymbol{\eta}_1' \boldsymbol{V}_{11}\boldsymbol{\eta}_1 \\ &- 2E\Big((1-(q-2)\chi^{-2}_{q+2}(\Delta^2)) I(\chi^2_{q+2}(\Delta^2) < \\ & \quad (q-2))\Big)\boldsymbol{\eta}_1' \boldsymbol{V}_{11}\boldsymbol{\eta}_1. \end{aligned}$$

Therefore,

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^{JS},\boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}^{JS+},\boldsymbol{M}) &= \sigma^2\Bigg\{ tr(\boldsymbol{V}_{11})E\Big[(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))^2 \\ & \quad I(\chi^2_{q+2}(\Delta^2) < q-2)\Big] \\ &+ \frac{1}{\sigma^2}E\Big[(1-(q-2)\chi^{-2}_{q+4}(\Delta^2))^2 \\ & \quad I(\chi^2_{q+4}(\Delta^2) < q-2)\Big]\boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1 \\ &- \frac{2}{\sigma^2}E\Big[(1-(q-2)\chi^{-2}_{q+2}(\Delta^2)) \\ & \quad I(\chi^2_{q+2}(\Delta^2) < q-2)\Big]\boldsymbol{\eta}_1'\boldsymbol{V}_{11}\boldsymbol{\eta}_1\Bigg\}. \qquad (2.27) \end{aligned}$$

Since,

$$\begin{aligned} \big((1-(q-2)\chi^{-2}_{q+2}(\Delta^2))I(\chi^2_{q+2}(\Delta^2) < q-2)\big) &\leq 0, \quad \text{and} \\ \big((1-(q-2)\chi^{-2}_{q+4}(\Delta^2))^2 I(\chi^2_{q+4}(\Delta^2) < q-2)\big) &\geq 0, \end{aligned}$$

the expected values appearing in (2.27) are always nonnegative. Therefore, for all $\Delta^2, \boldsymbol{M}$ and $q \geq 3$, the risk of $\hat{\boldsymbol{\beta}}^{JS+}$ is less than or equal to that of $\hat{\boldsymbol{\beta}}^{JS}$ which, in turn, is less than or equal to the risk of $\hat{\boldsymbol{\beta}}$ in the whole parameter space. Thus for all Δ^2, the following result holds

$$AQR(\hat{\boldsymbol{\beta}}^{JS+},\boldsymbol{M}) \leq AQR(\hat{\boldsymbol{\beta}}^{JS},\boldsymbol{M}) \leq AQR(\hat{\boldsymbol{\beta}},\boldsymbol{M}).$$

2.7 Penalty Estimators

In this section we will construct estimators of $\boldsymbol{\beta}$ by using three penalty functions, namely, the LASSO, the adaptive LASSO and the SCAD penalty functions for the CAR model.

Recall that the log-likelihood of the CAR model is

$$\begin{aligned} l = \log(L(\boldsymbol{\beta}, \sigma^2, \rho)) &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(\boldsymbol{C}) \\ &- \frac{1}{2\sigma^2}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{C}^{-1}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}), \end{aligned} \tag{2.28}$$

where $\boldsymbol{C} = (\boldsymbol{I} - \rho\boldsymbol{W}^*)^{-1}\boldsymbol{D}$.

Since neither σ^2 nor ρ is subject to any penalty in our study, maximizing the log-likelihood in (2.28) is equivalent to maximizing

$$l^{**}(\boldsymbol{\beta}, \sigma^2, \rho) \propto -\frac{1}{2}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{C}^{-1}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})$$

or simply, $\min_{\boldsymbol{\beta}} \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$.

A general form of the objective function based on a penalty $f(\lambda, \boldsymbol{\beta})$ in which λ serves as a regularization parameter is given by

$$Q(\boldsymbol{\beta}) = (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{C}^{-1}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}) + f(\lambda, \boldsymbol{\beta}).$$

In general, algorithms for computing penalized estimators for non-spatial regression models are based on the assumption of independent errors. However, the CAR model does not enjoy the condition of independent errors. In order to overcome this difficulty, for fixed σ^2 and ρ, we propose the use of transformation of the CAR model to obtain independent errors. In a related topic, Cressie (1993) discussed the use of transformations in bootstrapping or jackknifing spatial lattice models.

Since

$$\boldsymbol{C} = (\boldsymbol{I} - \rho \boldsymbol{W}^*)^{-1}\boldsymbol{D}, \tag{2.29}$$

is a $n \times n$ positive definite matrix, there exists an $n \times n$ upper triangular matrix $\boldsymbol{U}$ with positive diagonal elements such that $\boldsymbol{C} = \boldsymbol{U}'\boldsymbol{U}$ see (Seber, 2008, p.338). Multiplying both sides of (1.17) by $(\boldsymbol{U}^{-1})'$ we have

$$\begin{aligned} (\boldsymbol{U}^{-1})'\boldsymbol{Y} &= (\boldsymbol{U}^{-1})'\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{U}^{-1})'\boldsymbol{\epsilon} \\ \boldsymbol{Y}^* &= \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*, \end{aligned} \tag{2.30}$$

where $\boldsymbol{Y}^* = (\boldsymbol{U}^{-1})'\boldsymbol{Y}$, $\boldsymbol{X}^* = (\boldsymbol{U}^{-1})'\boldsymbol{X}$ and $\boldsymbol{\epsilon}^* = (\boldsymbol{U}^{-1})'\boldsymbol{\epsilon}$, which yields $\boldsymbol{\epsilon}^* \sim N(0, \sigma^2\boldsymbol{I})$. This transformed model will be used to construct penalty estimators in the next three subsections.

2.7.1 LASSO

As described in Chapter 1, the LASSO method uses L_1-type of penalty function and hence, for the CAR,

$$\hat{\boldsymbol{\beta}}^{LASSO} = \arg\min_{\boldsymbol{\beta}} \left[(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})'(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j| \right] \tag{2.31}$$

where $\lambda \geq 0$ is a tuning parameter which can be obtained through cross-validation techniques. Computationally, $\hat{\boldsymbol{\beta}}^{LASSO}$ can be calculated via the numerical algorithm known as the least angle regression (LARS) algorithm proposed in Efron et al. (2004).

The absolute penalty estimators were, originally, designed for high dimensional cases where $p \geq n$, but they also work well in the classical cases where $n > p$. The LASSO method does not satisfy a desirable property known as *Oracle Property*. A variable selection procedure is said to have oracle property if it identifies the right subset of zero coefficients in the regression model under consideration and furthermore, the estimators of the remaining non-zero coefficients are consistent and asymptotically normal, Zou (2006). Two procedures which posess the oracle property were introduced by Fan and Li (2001) and Zou (2006), in the next two sections, we define these two procedures.

2.7.2 SCAD

The smoothly clipped absolute deviation penalty (SCAD) variable selection procedure was originally introduced in Fan and Li (2001) in order to overcome the lack of oracle property in the LASSO method. For our CAR model, we define the SCAD estimator as

$$\hat{\boldsymbol{\beta}}^{SCAD} = \arg\min_{\boldsymbol{\beta}} \left[(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})'(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}) + \sum_{j=1}^{p} P_{\lambda_S}(|\beta_j|) \right] \tag{2.32}$$

where $P_{\lambda_S}(|\beta_j|)$ is the SCAD penalty function as defined in equation (1.10).

2.7.3 Adaptive LASSO

Another selection procedure which enjoys the oracle property is the adaptive LASSO of Zou (2006). For the CAR model, we define the adaptive LASSO estimator as

$$\hat{\boldsymbol{\beta}}^{A.LASSO} = \arg\min_{\boldsymbol{\beta}} \left[(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})'(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}) + n\sum_{j=1}^{p} \lambda_j |\beta_j| \right] \tag{2.33}$$

where $\{\lambda_j : j = 1, 2, \ldots, p\}$ are coefficient specific tuning parameters. Zou (2006) suggested that $\lambda_j = \frac{1}{\hat{\beta}_j}$, where $\hat{\beta}_j$ is an initial estimator, such as the least squares estimator in the case of linear regression. In the current CAR model, we will set $\hat{\beta}_j$ to be the unrestricted MLE of β_j for $j = 1, \ldots, p$.

2.8 Numerical Studies

In this section we will carry out two sets of Monte Carlo simulations. The first set of simulations aims at examining the relative performance of the restricted, pretest and shrinkage estimators, while appointing the unrestricted estimator as a benchmark. The results of this set of Monte Carlo simulations turned out to be consistent with our analytical comparisons in Section 2.6. Thus, the positive James-Stein estimator stands out in terms of overall risk performance. This leads us to the second set of Monte Carlo simulations which will restrict attention to the comparison between the positive James-Stein, restricted estimator and the class of penalty estimators of Section 2.7. Finally, we conclude the section by applying the proposed estimators to real data sets.

2.8.1 Simulated Efficiency Analysis

We use the Monte Carlo simulation experiments to compare the restricted, pretest and shrinkage estimators via their simulated mean squared errors. We consider $N \times N$ square lattices for $N = 6$ and 9 and corresponding sample sizes of $n = 36$ and 81, respectively. In this experiment, we simulate the response variable $\boldsymbol{Y}(s)$ from a multivariate normal distribution with mean $\boldsymbol{X\beta}$, where the design matrix $\boldsymbol{X}$ is generated from standard multivariate normal distribution, and the error term has the CAR model covariance, $\boldsymbol{\Sigma}_{CAR} = \sigma^2(\boldsymbol{I} - \rho\boldsymbol{W}^*)^{-1}\boldsymbol{D}$, with mean $\boldsymbol{0}$ and $\sigma^2 = 1$. We employed queen-based contiguity to define our neighborhood matrix, $\boldsymbol{W}^*$. Two sites are queen-based neighbors if they have common boundaries and common corners. For the spatial dependence parameter we considered five values, $\rho = (-0.9, -0.5, 0.0, 0.5, 0.9)$.

The vector of regression coefficients $\boldsymbol{\beta}$ was partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are respectively, $(p-q) \times 1$ and $q \times 1$ vectors. The candidate subspace was chosen to be $\boldsymbol{A}_0 : \beta_j = 0$ for $j = p-q+1, p-q+2, \ldots, p$. We chose, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (\boldsymbol{1}_{p-q}, \boldsymbol{0}_q)$, where $\boldsymbol{1}_{p-q}$ is a $(p-q) \times 1$ vector of ones, and $\boldsymbol{0}_q$ is a $q \times 1$ vector of zeroes. For simplicity, we defined the non-centrality parameter Δ^2, which is essentially a measure of how far away we go from the candidate subspace, as $\Delta^2 = \parallel \boldsymbol{\beta} - \boldsymbol{\beta}_0 \parallel$, where $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_1, \boldsymbol{0})$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{0} + \boldsymbol{\delta})$ and $\parallel . \parallel$ denotes the Euclidian norm. Thus, essentially, our $\Delta^2 = \parallel \boldsymbol{\delta} \parallel$, where this vector of alternative values was chosen to vary from 0 to 2 with steps of 0.1. Various choices of (p, q) were used in combination with configurations of ρ, n, Δ^2 and 2000 Monte Carlo runs. In each of these Monte Carlo runs, the restricted, unrestricted, pretest, shrinkage and positive shrinkage estimators were computed and their *simulated mean squared errors* (SMSE) were obtained from the empirical formula

$$SMSE(\hat{\boldsymbol{\beta}}^*) = \sum_{i=1}^{p} (\hat{\beta}_i^* - \beta_i)^2, \tag{2.34}$$

where $\hat{\beta}^*$ denotes any one of $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^R, \hat{\boldsymbol{\beta}}^{PT}, \hat{\boldsymbol{\beta}}^{JS}, \hat{\boldsymbol{\beta}}^{JS+}$. The *simulated relative efficiency* (SRE) was defined as

$$SRE(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^*) = \frac{SMSE(\hat{\boldsymbol{\beta}})}{SMSE(\hat{\boldsymbol{\beta}}^*)}, \tag{2.35}$$

where $\hat{\boldsymbol{\beta}}$ is the unrestricted estimator, appointed as benchmark. A value greater than one of the $SRE(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^*)$ indicates that $\hat{\boldsymbol{\beta}}^*$ performs better than $\hat{\boldsymbol{\beta}}$, and vice versa. Results of these simulations are reported in Figures 2.1 to 2.5 and in Tables 2.1 to 2.20.

Figure 2.1: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = -0.90$ for different values of (p, q) based on the CAR model

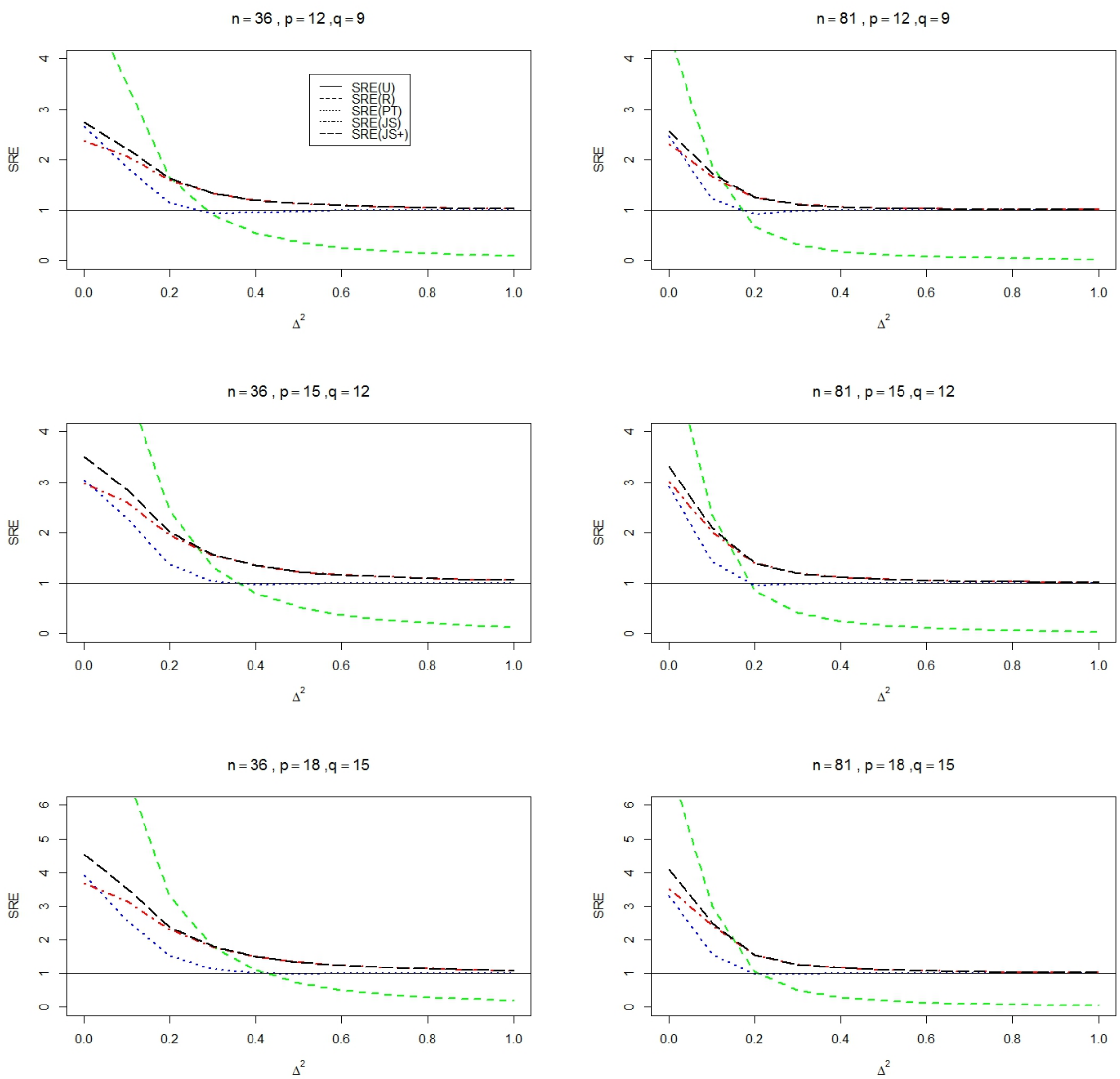

Figure 2.2: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = -0.50$ for different values of (p, q) based on the CAR model

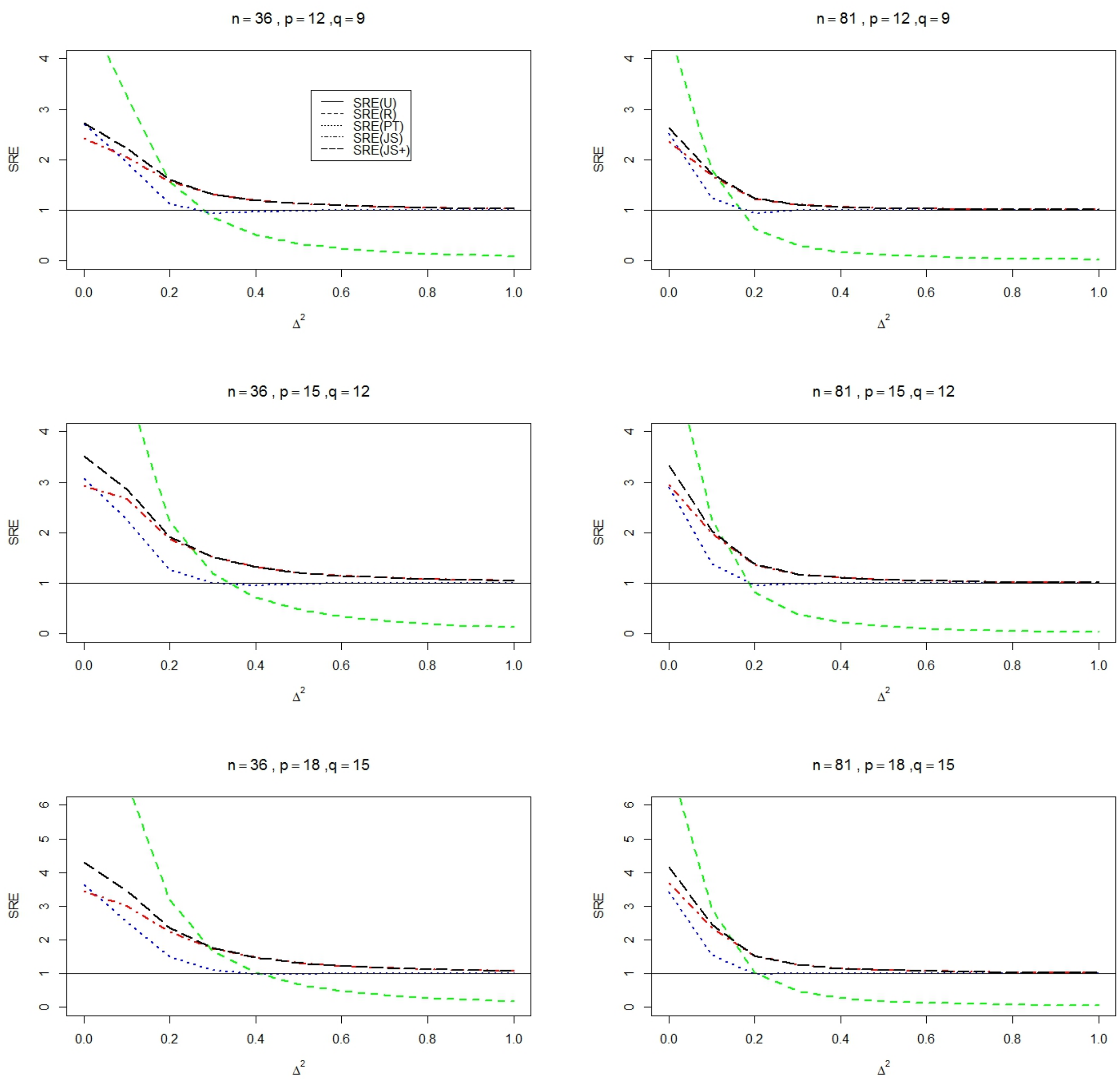

Figure 2.3: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = 0$ for different values of (p, q) based on the CAR model

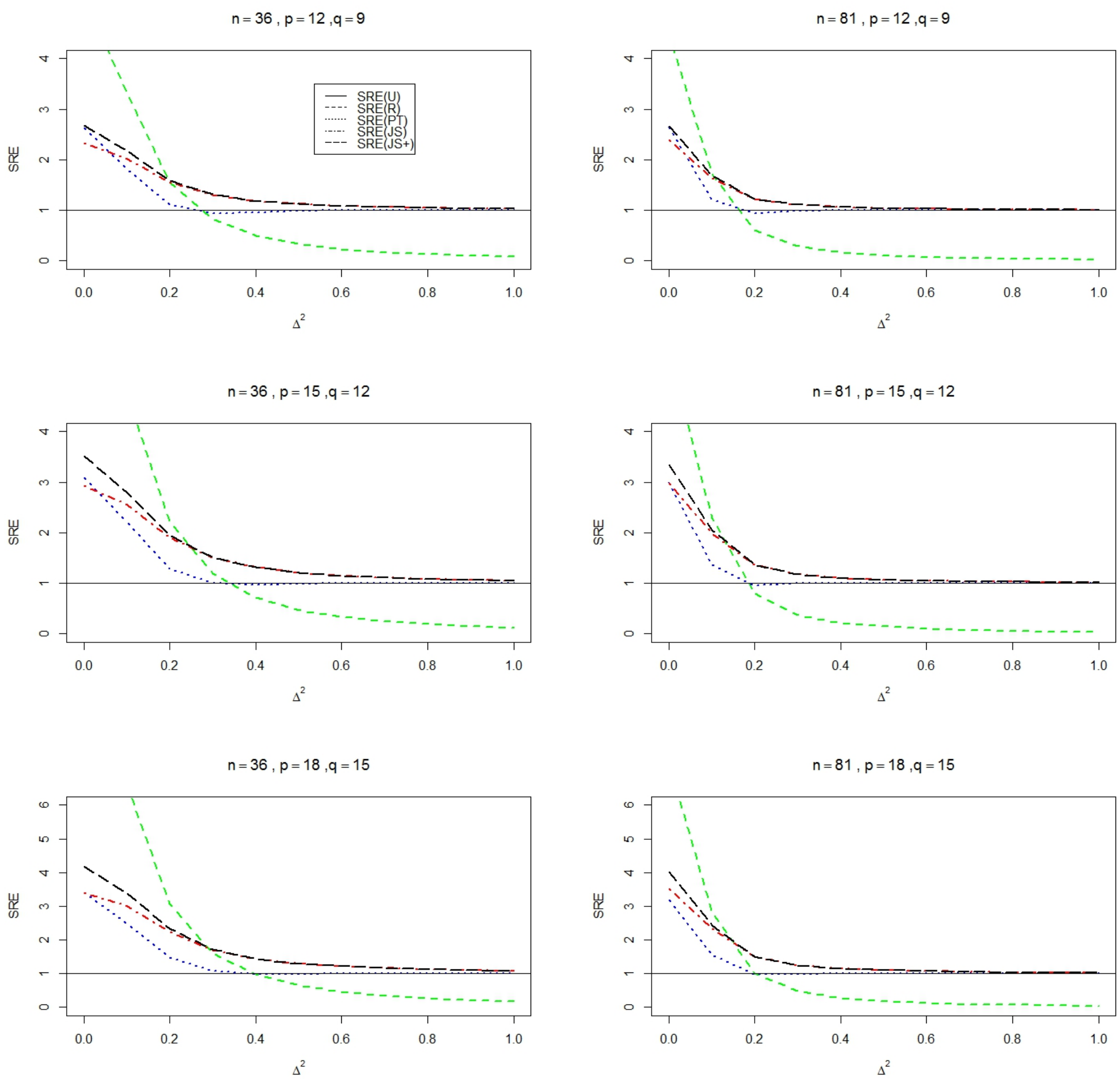

Figure 2.4: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = 0.50$ for different values of (p, q) based on the CAR model

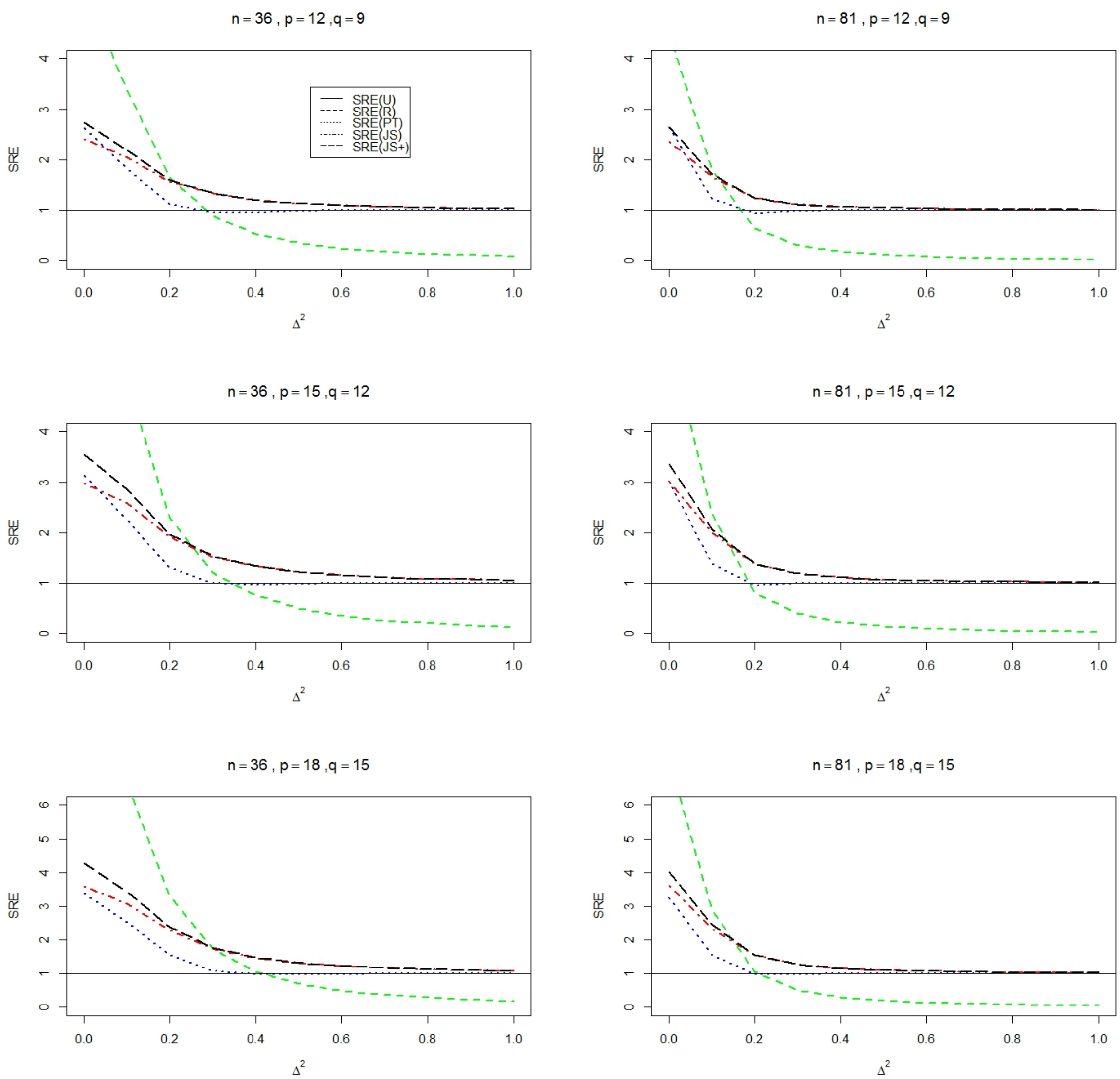

Figure 2.5: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = 0.90$ for different values of (p, q) based on the CAR model

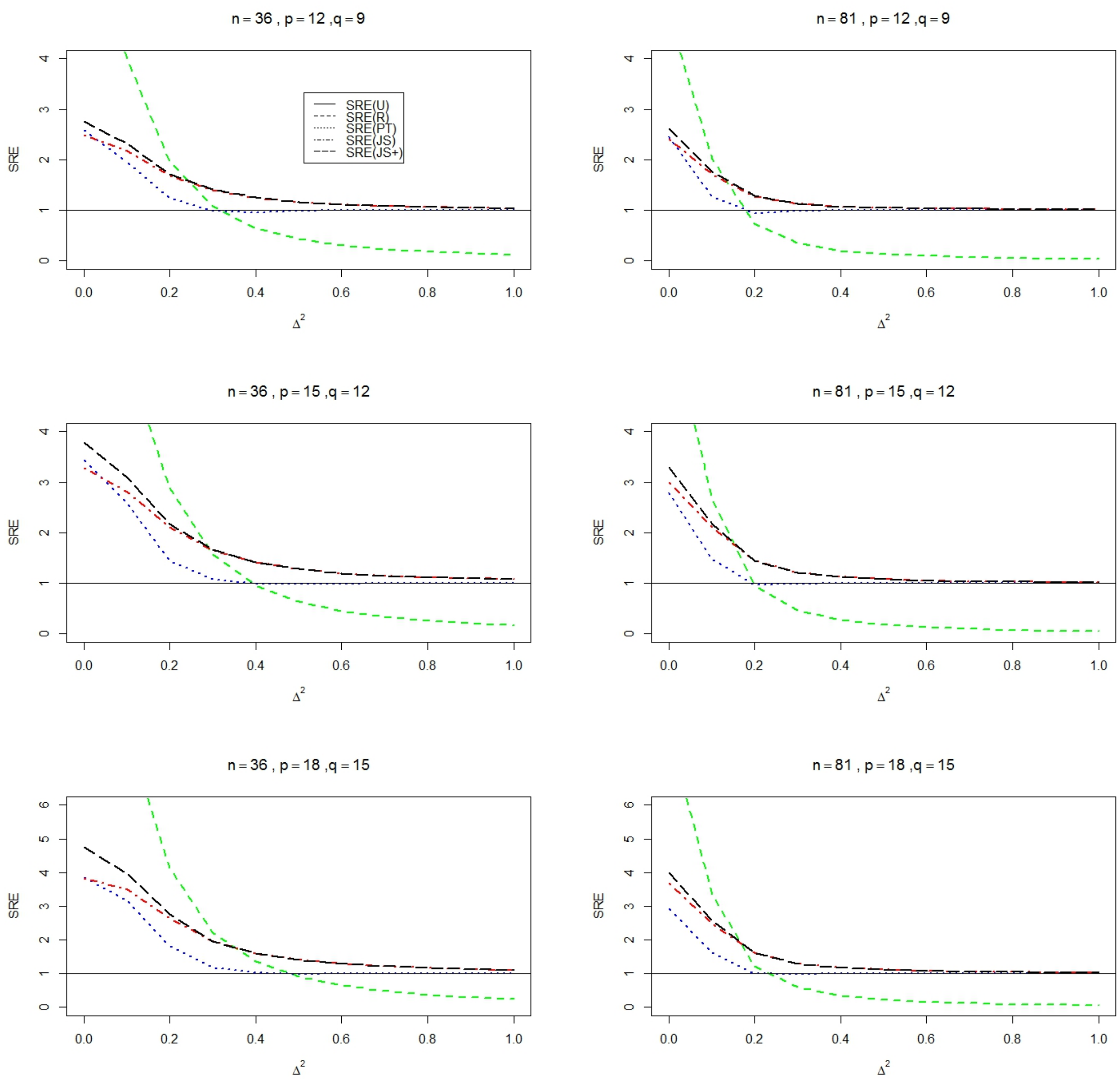

Table 2.1: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = -0.90$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1807	1.6427	1.1777	1.2634
0.1	1.3753	1.1168	1.0938	1.1496
0.3	0.3661	0.8759	1.0185	1.0185
0.5	0.1478	0.9956	1.0064	1.0064
0.7	0.0767	1.0000	1.0042	1.0042
0.9	0.0480	1.0000	1.0022	1.0022
1.1	0.0329	1.0000	1.0009	1.0009
1.3	0.0235	1.0000	1.0005	1.0005
1.5	0.0177	1.0000	1.0010	1.0010
1.7	0.0140	1.0000	1.0007	1.0007
2.0	0.0098	1.0000	1.0002	1.0002

Table 2.2: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = -0.90$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.6576	2.1784	1.7707	1.9939
0.1	2.3076	1.4826	1.5875	1.6548
0.3	0.6129	0.9108	1.1585	1.1600
0.5	0.2410	0.9860	1.0508	1.0508
0.7	0.1282	1.0000	1.0286	1.0286
0.9	0.0792	1.0000	1.0215	1.0215
1.1	0.0544	1.0000	1.0128	1.0128
1.3	0.0389	1.0000	1.0092	1.0092
1.5	0.0288	1.0000	1.0082	1.0082
1.7	0.0222	1.0000	1.0036	1.0036
2.0	0.0159	1.0000	1.0038	1.0038

Table 2.3: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = -0.50$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1923	1.6142	1.1226	1.2580
0.1	1.3800	1.1153	1.0815	1.1463
0.3	0.3541	0.8845	1.0131	1.0143
0.5	0.1412	0.9964	1.0051	1.0051
0.7	0.0741	1.0000	1.0035	1.0035
0.9	0.0458	1.0000	1.0012	1.0012
1.1	0.0306	1.0000	1.0010	1.0010
1.3	0.0216	1.0000	1.0007	1.0007
1.5	0.0163	1.0000	1.0007	1.0007
1.7	0.0126	1.0000	1.0003	1.0003
2.0	0.0090	1.0000	0.9997	0.9997

Table 2.4: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = -0.50$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.5402	2.1290	1.7893	1.9649
0.1	2.2816	1.5303	1.5680	1.6713
0.3	0.5812	0.8996	1.1428	1.1445
0.5	0.2321	0.9952	1.0573	1.0573
0.7	0.1188	1.0000	1.0289	1.0289
0.9	0.0740	1.0000	1.0186	1.0186
1.1	0.0493	1.0000	1.0102	1.0102
1.3	0.0358	1.0000	1.0079	1.0079
1.5	0.0267	1.0000	1.0057	1.0057
1.7	0.0206	1.0000	1.0045	1.0045
2.0	0.0153	1.0000	1.0044	1.0044

Table 2.5: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = 0$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1623	1.5912	1.1173	1.2620
0.1	1.3419	1.1133	1.1179	1.1457
0.3	0.3378	0.8901	1.0145	1.0147
0.5	0.1364	0.9937	1.0050	1.0050
0.7	0.0720	1.0000	1.0027	1.0027
0.9	0.0444	1.0000	1.0018	1.0018
1.1	0.0288	1.0000	1.0003	1.0003
1.3	0.0212	1.0000	1.0007	1.0007
1.5	0.0161	1.0000	1.0004	1.0004
1.7	0.0125	1.0000	1.0006	1.0006
2.0	0.0090	1.0000	1.0004	1.0004

Table 2.6: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = 0$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.6198	2.1612	1.7695	1.9737
0.1	2.2361	1.4418	1.5440	1.6225
0.3	0.5591	0.9005	1.1413	1.1440
0.5	0.2206	0.9944	1.0449	1.0449
0.7	0.1175	1.0000	1.0274	1.0274
0.9	0.0716	1.0000	1.0171	1.0171
1.1	0.0478	1.0000	1.0124	1.0124
1.3	0.0344	1.0000	1.0065	1.0065
1.5	0.0254	1.0000	1.0065	1.0065
1.7	0.0205	1.0000	1.0076	1.0076
2.0	0.0144	1.0000	1.0039	1.0039

Table 2.7: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = 0.50$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.2368	1.6431	1.1647	1.2663
0.1	1.4014	1.1189	1.0850	1.1486
0.3	0.3592	0.8671	1.0135	1.0135
0.5	0.1398	0.9912	1.0032	1.0032
0.7	0.0756	1.0000	1.0028	1.0028
0.9	0.0462	1.0000	1.0016	1.0016
1.1	0.0329	1.0000	1.0008	1.0008
1.3	0.0221	1.0000	1.0007	1.0007
1.5	0.0167	1.0000	0.9999	0.9999
1.7	0.0130	1.0000	1.0006	1.0006
2.0	0.0092	1.0000	1.0004	1.0004

Table 2.8: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = 0.50$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.6138	2.1441	1.8068	1.9621
0.1	2.3468	1.4987	1.5688	1.6508
0.3	0.5854	0.8985	1.1477	1.1485
0.5	0.2312	0.9835	1.0541	1.0541
0.7	0.1246	1.0000	1.0281	1.0281
0.9	0.0769	1.0000	1.0166	1.0166
1.1	0.0499	1.0000	1.0146	1.0146
1.3	0.0372	1.0000	1.0099	1.0099
1.5	0.0270	1.0000	1.0061	1.0061
1.7	0.0216	1.0000	1.0051	1.0051
2.0	0.0156	1.0000	1.0035	1.0035

Table 2.9: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = 0.90$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.3014	1.6343	1.1914	1.2661
0.1	1.5034	1.1654	1.1370	1.1622
0.3	0.4150	0.8835	1.0249	1.0266
0.5	0.1699	0.9904	1.0076	1.0076
0.7	0.0889	1.0000	1.0032	1.0032
0.9	0.0567	1.0000	1.0032	1.0032
1.1	0.0386	1.0000	1.0014	1.0014
1.3	0.0279	1.0000	1.0015	1.0015
1.5	0.0218	1.0000	1.0012	1.0012
1.7	0.0174	1.0000	1.0009	1.0009
2.0	0.0127	1.0000	1.0008	1.0008

Table 2.10: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = 0.90$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	4.0530	2.2871	1.8593	2.0070
0.1	2.6353	1.5693	1.6359	1.7093
0.3	0.7077	0.9103	1.1854	1.1882
0.5	0.2918	0.9846	1.0693	1.0693
0.7	0.1511	1.0000	1.0364	1.0364
0.9	0.0948	1.0000	1.0206	1.0206
1.1	0.0674	1.0000	1.0163	1.0163
1.3	0.0460	1.0000	1.0111	1.0111
1.5	0.0356	1.0000	1.0065	1.0065
1.7	0.0273	1.0000	1.0065	1.0065
2.0	0.0208	1.0000	1.0038	1.0038

Table 2.11: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = -0.90$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.0914	1.6038	1.1226	1.2651
0.1	0.8604	0.8651	1.0665	1.0708
0.3	0.1542	0.9978	1.0072	1.0072
0.5	0.0559	1.0000	1.0021	1.0021
0.7	0.0292	1.0000	1.0016	1.0016
0.9	0.0184	1.0000	1.0008	1.0008
1.1	0.0123	1.0000	1.0009	1.0009
1.3	0.0086	1.0000	1.0002	1.0002
1.5	0.0068	1.0000	1.0003	1.0003
1.7	0.0052	1.0000	1.0005	1.0005
2.0	0.0037	1.0000	1.0001	1.0001

Table 2.12: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = -0.90$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.2840	2.0852	1.8013	1.9475
0.1	1.3388	1.0591	1.3451	1.3690
0.3	0.2314	0.9937	1.0515	1.0515
0.5	0.0872	1.0000	1.0165	1.0165
0.7	0.0462	1.0000	1.0113	1.0113
0.9	0.0274	1.0000	1.0078	1.0078
1.1	0.0191	1.0000	1.0043	1.0043
1.3	0.0137	1.0000	1.0037	1.0037
1.5	0.0101	1.0000	1.0024	1.0024
1.7	0.0079	1.0000	1.0019	1.0019
2.0	0.0058	1.0000	1.0007	1.0007

Table 2.13: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = -0.50$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.0675	1.5987	1.1866	1.2599
0.1	0.8227	0.8566	1.0568	1.0652
0.3	0.1452	0.9974	1.0068	1.0068
0.5	0.0543	1.0000	1.0014	1.0014
0.7	0.0283	1.0000	1.0009	1.0009
0.9	0.0172	1.0000	1.0006	1.0006
1.1	0.0113	1.0000	1.0008	1.0008
1.3	0.0083	1.0000	1.0001	1.0001
1.5	0.0062	1.0000	1.0001	1.0001
1.7	0.0049	1.0000	1.0002	1.0002
2.0	0.0034	1.0000	1.0002	1.0002

Table 2.14: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = -0.50$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.2516	2.1524	1.7774	1.9516
0.1	1.3018	1.0356	1.3337	1.3608
0.3	0.2194	0.9975	1.0514	1.0514
0.5	0.0835	1.0000	1.0184	1.0184
0.7	0.0420	1.0000	1.0080	1.0080
0.9	0.0258	1.0000	1.0050	1.0050
1.1	0.0175	1.0000	1.0033	1.0033
1.3	0.0126	1.0000	1.0016	1.0016
1.5	0.0094	1.0000	1.0027	1.0027
1.7	0.0074	1.0000	1.0013	1.0013
2.0	0.0053	1.0000	1.0004	1.0004

Table 2.15: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = 0$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.0551	1.6457	1.2294	1.2761
0.1	0.7990	0.8460	1.0576	1.0648
0.3	0.1359	0.9970	1.0051	1.0051
0.5	0.0515	1.0000	1.0025	1.0025
0.7	0.0265	1.0000	0.9999	0.9999
0.9	0.0158	1.0000	1.0010	1.0010
1.1	0.0107	1.0000	1.0001	1.0001
1.3	0.0076	1.0000	1.0005	1.0005
1.5	0.0060	1.0000	1.0001	1.0001
1.7	0.0046	1.0000	1.0001	1.0001
2.0	0.0034	1.0000	1.0001	1.0001

Table 2.16: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = 0$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.2795	2.2142	1.8308	1.9880
0.1	1.2556	1.0310	1.3286	1.3608
0.3	0.2127	0.9964	1.0464	1.0464
0.5	0.0793	1.0000	1.0174	1.0174
0.7	0.0412	1.0000	1.0068	1.0068
0.9	0.0246	1.0000	1.0059	1.0059
1.1	0.0168	1.0000	1.0034	1.0034
1.3	0.0116	1.0000	1.0017	1.0017
1.5	0.0090	1.0000	1.0025	1.0025
1.7	0.0071	1.0000	1.0023	1.0023
2.0	0.0051	1.0000	1.0014	1.0014

Table 2.17: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = 0.50$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1598	1.6506	1.1568	1.2796
0.1	0.8182	0.8449	1.0574	1.0664
0.3	0.1432	1.0000	1.0054	1.0054
0.5	0.0534	1.0000	1.0034	1.0034
0.7	0.0273	1.0000	1.0011	1.0011
0.9	0.0166	1.0000	1.0007	1.0007
1.1	0.0113	1.0000	1.0005	1.0005
1.3	0.0084	1.0000	1.0003	1.0003
1.5	0.0064	1.0000	1.0002	1.0002
1.7	0.0049	1.0000	1.0002	1.0002
2.0	0.0035	1.0000	1.0004	1.0004

Table 2.18: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = 0.50$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.3184	2.1979	1.8256	1.9856
0.1	1.2898	1.0414	1.3299	1.3600
0.3	0.2218	0.9989	1.0512	1.0512
0.5	0.0819	1.0000	1.0224	1.0224
0.7	0.0430	1.0000	1.0092	1.0092
0.9	0.0262	1.0000	1.0054	1.0054
1.1	0.0175	1.0000	1.0031	1.0031
1.3	0.0124	1.0000	1.0023	1.0023
1.5	0.0095	1.0000	1.0021	1.0021
1.7	0.0075	1.0000	1.0015	1.0015
2.0	0.0054	1.0000	1.0018	1.0018

Table 2.19: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = 0.90$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1553	1.6339	1.2068	1.2686
0.1	0.8987	0.8725	1.0523	1.0750
0.3	0.1591	1.0000	1.0079	1.0079
0.5	0.0621	1.0000	1.0014	1.0014
0.7	0.0315	1.0000	1.0008	1.0008
0.9	0.0199	1.0000	1.0007	1.0007
1.1	0.0133	1.0000	1.0009	1.0009
1.3	0.0100	1.0000	1.0003	1.0003
1.5	0.0074	1.0000	1.0008	1.0008
1.7	0.0062	1.0000	1.0000	1.0000
2.0	0.0045	1.0000	1.0003	1.0003

Table 2.20: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = 0.90$ based on the CAR model

Δ^2	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.5315	2.1117	1.8328	1.9691
0.1	1.4154	1.0792	1.3553	1.3930
0.3	0.2494	0.9952	1.0605	1.0605
0.5	0.0972	1.0000	1.0203	1.0203
0.7	0.0504	1.0000	1.0128	1.0128
0.9	0.0315	1.0000	1.0058	1.0058
1.1	0.0214	1.0000	1.0056	1.0056
1.3	0.0154	1.0000	1.0042	1.0042
1.5	0.0118	1.0000	1.0033	1.0033
1.7	0.0096	1.0000	1.0016	1.0016
2.0	0.0071	1.0000	1.0014	1.0014

The following conclusions can be drawn from the SRE results.

1. In general, for fixed n, p, q, varying the value of ρ does not affect much the SRE of the estimators, thus agreeing with Mardia and Marshall (1984) Theorem about the asymptotic independence of the large-scale variation $\boldsymbol{\beta}$ and the small-scale

variation ρ, σ^2.

2. For all values of n and (p, q), the restricted estimator $\hat{\boldsymbol{\beta}}^R$ is the best in terms of SRE when the candidate subspace is true ($\Delta^2 = 0$), but as Δ^2 moves away from 0, the SRE of $\hat{\boldsymbol{\beta}}^R$ approaches 0. That is, the SMSE of $\hat{\boldsymbol{\beta}}^R$ becomes unbounded while the SMSE of the remaining estimators approach that of the unrestricted estimator. This obviously agrees with the theoretical results of Section 2.5.

3. The positive shrinkage estimator always dominates the shrinkage estimator, and it dominates the pretest for all Δ^2 values that are away from 0.

These conclusions for the small sample performance of the proposed estimator are therefore, in line with the theoretical results obtained in Section 2.5.

Application to Columbus Crime Data

The Columbus crime data set was collected in 1980 and originally reported in Anselin (1988). The data set consists of observations for 49 contiguous planning neighborhoods in Columbus, Ohio. Neighborhoods correspond to census tracts, or aggregates of small number of census tracts. The outcome of interest was `CRIME`, the combined total of residential burglaries and vehicle thefts per thousand households. A number of covariates were also collected: income `INC`, housing values `HOVAL` in thousands of dollars, the variable `DISCBD` measuring the distance to the cental business district (CBD), open space in neighborhood `OPEN`, percentage of housing units without plumbing `PLUMB`. The data is also available in `spdep` `R-package` (Bivand et al., 2012).

Several authors used these data as an illustrative application example. Among others, Anselin (1988) fitted two separate regression curves to illustrate the presence of separate level of spatial dependence for the east and west sides of Columbus city using SAR spatial model (see Chapter 3 of this thesis). Kyung and Ghosh (2009) used these data to fit three different regression curves using a Bayesian version of the CAR model.

We apply our suggested estimation strategies to this data set. Following Kyung and Ghosh (2010) we first apply a variance stabilizing log-transformation to the response variable, `CRIME`. The transformed variable is denoted by `log(CRIME)` and a CAR model is fitted to this variable along with the complete set of covariates explained above via the `spdep` package, thus obtaining unrestricted estimators, $\hat{\boldsymbol{\beta}}$. A candidate subspace is then obtained by the AIC and BIC model selection criteria via the `R`-function `spautolm` in the `spdep` package. Consequently, the selected reduced model is used to compute the restricted, the pretest, and James-Stein estimators, according to the formulae given in this Chapter. The candidate subspace model and the full model are both listed in the following Table (2.21).

Table 2.21: Full and reduced models for the Columbus crime data

Selection Criterion	Model
Full	`log(CRIME) ~ HOVAL+PLUMB+INC+DISCBD+OPEN`
AIC/BIC	`log(CRIME)~ HOVAL+PLUMB`

In the above table, we have used the `R`-notation (`~`) to write the models.

To asses the performance of the estimators, we use mean squared prediction error based on a bootstrap method suggested by Hall (1985). The procedure can be summarized as follows:

1. For $k = 1, ..., B$ sample $Y_{k1}^*, ..., Y_{kn}^*$ with replacement from the original data $Y_1, ..., Y_n$.

2. For the k^{th} bootstrap sample, compute the estimator of interest, $\hat{\boldsymbol{\beta}}^*$, (which could be either of $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^R, \hat{\boldsymbol{\beta}}^{PT}, \hat{\boldsymbol{\beta}}^{JS}, \hat{\boldsymbol{\beta}}^{JS+}$).

3. Compute the *mean squared prediction error* (MSPE) for the bootstrap sample as follows:

$$MSPE_k = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}^*_{ki})^2, \tag{2.36}$$

where

$$\hat{Y}^*_{ki} = \boldsymbol{X_i}\hat{\boldsymbol{\beta}}^* + \hat{\rho}^*\sum_{j=1}^{n} W^*_{ij}(Y^*_{kj} - \boldsymbol{X_j}\hat{\boldsymbol{\beta}}^*). \tag{2.37}$$

4. Compute the average of the MSPE for $\hat{\boldsymbol{\beta}}^*$ over the B bootstrap samples as follows:
$$MSPE(\hat{\boldsymbol{\beta}}^*) = \sum_{k=1}^{B} MSPE(k)/B.$$

The *relative efficiency of the mean squared prediction error* (RMSPE) with respect to the benchmark $\hat{\boldsymbol{\beta}}$ is then computed for each one of the shrinkage type estimators as follows:

$$RMSPE(\hat{\boldsymbol{\beta}}^\diamond) = \frac{MSPE(\hat{\boldsymbol{\beta}})}{MSPE(\hat{\boldsymbol{\beta}}^\diamond)}, \tag{2.38}$$

where $\hat{\boldsymbol{\beta}}^\diamond$ belongs to the set $\{\hat{\boldsymbol{\beta}}^R, \hat{\boldsymbol{\beta}}^{PT}, \hat{\boldsymbol{\beta}}^{JS}, \hat{\boldsymbol{\beta}}^{JS+}\}$. The results of the RMSPEs based on $B = 2000$ are reported in Table 2.22.

From this table we can clearly see that all estimators are better than the benchmark and the best among them is the $\hat{\boldsymbol{\beta}}^R$, followed by $\hat{\boldsymbol{\beta}}^{PT}$, $\hat{\boldsymbol{\beta}}^{JS+}$. However, the main purpose of the shrinkage estimators is to provide a safe ground on which we do not

Table 2.22: RMSPE with respect to $\hat{\boldsymbol{\beta}}$ for Columbus crime data

Estimator	RMSPE
$\hat{\boldsymbol{\beta}}^{R}$	1.0905
$\hat{\boldsymbol{\beta}}^{PT}$	1.0619
$\hat{\boldsymbol{\beta}}^{JS+}$	1.0356
$\hat{\boldsymbol{\beta}}^{JS}$	1.0208

rely completely on the selected reduced model nor on the full unrestricted model, while keeping much of the efficiency of both. Overall, this is what we see from this data set, as the shrinkage estimators are not much less efficient than the restricted estimators.

2.8.2 Comparison of Non-Penalty and Penalty Estimators

Now we run the second set of Monte Carlo simulations aiming at comparing performances of the positive James-Stein estimator, the restricted, unrestricted and the class of penalty estimators of Section 2.7.

We consider an $N \times N$ square lattice where $N = 7, 8, 10$, with corresponding sample sizes of $n = 49, 64, 100$, respectively. We fix $\sigma^2 = 1$ and $p - q = 4$, $q = 5, 10, 15, 20, 25$ and $\rho = (-0.95, -0.50, 0.00, 0.50, 0.90)$, and the nonzero coefficients are set to 1 as before.

To obtain the penalty estimators, we first fit a full CAR model using the `spautolm` `R`-function. From the full CAR model we extract the (MLEs) of $\boldsymbol{\beta}$, σ^2, and ρ. The MLE of ρ, which is $\hat{\rho}$, is then used in the $\boldsymbol{C}$ matrix in (2.29) to obtain the $\boldsymbol{U}$ matrix used in transforming the response vector to independent data $\boldsymbol{Y}^*$, with the corresponding transformed design matrix $\boldsymbol{X}^*$. Consequently, a $10-$fold cross-validation was applied to the transformed data in order to select the optimal value of the tuning parameter $\hat{\lambda}_{LASSO}$ for the LASSO fit while the initial weights for the adaptive LASSO are computed based on the LASSO estimators. Thus, we used the LASSO as an initial starter for the adaptive LASSO procedure. This procedure can be performed using the `adalasso` `R`-function in `parcor` package; Kraemer and Schaefer (2010). For the SCAD penalty, a was fixed to be 3.7 as suggested by Fan and Li (2001). We used the function `cv.ncvreg` in the `ncvreg` `R`-package Breheny and Huang (2011) which performs a $k-$fold cross-validation to choose $\hat{\lambda}_{SCAD}$.

In order to carry out a fair comparison, we examine the relative performances of the estimators under the candidate subspace, that is when $\Delta^2 = 0$, as the penalty estimators do not depend on Δ^2. The *simulated relative efficiency* (SRE) based on the *simulated mean squared error* (SMSE) with respect to the benchmark estimator, $\hat{\boldsymbol{\beta}}$ as defined in the previous sections, are used as a performance measure.

Table 2.23: Simulated relative efficiency of the restricted, positive James-Stein and penalty estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 49$, $p - q = 4$ and $\Delta^2 = 0$ for different values of ρ and q based on the CAR model

ρ	q	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{JS+}$	$\hat{\boldsymbol{\beta}}^{LASSO}$	$\hat{\boldsymbol{\beta}}^{A.LASSO}$	$\hat{\boldsymbol{\beta}}^{SCAD}$
-0.95	5	2.4002	1.5982	1.0864	1.6371	1.5031
	10	4.1852	2.6458	1.2954	2.2081	1.9820
	15	6.4118	3.8246	1.6270	2.9278	2.7034
	20	9.9443	5.2107	2.0789	4.0984	3.6696
	25	15.0526	6.6719	2.6802	5.5746	5.2168
-0.50	5	2.3952	1.5985	1.0552	1.5947	1.4093
	10	4.3488	2.6457	1.2706	2.1924	1.9635
	15	6.6751	3.8193	1.5424	2.6884	2.4109
	20	10.1787	5.2720	1.9885	3.7958	3.4674
	25	15.3874	6.4454	2.6031	5.3502	5.0465
0.00	5	2.4100	1.6268	1.0379	1.5510	1.3778
	10	4.1541	2.6407	1.2359	2.1134	1.8351
	15	6.7886	3.8420	1.5645	2.8345	2.4847
	20	10.6979	5.3261	1.9821	3.8377	3.5279
	25	15.2230	6.2243	2.5674	5.3302	4.9059
0.50	5	2.5776	1.6288	1.1143	1.7013	1.4849
	10	4.5038	2.6927	1.3454	2.3413	2.0163
	15	6.7291	3.8123	1.6303	3.0033	2.6881
	20	10.5470	5.1218	2.0692	4.0855	3.6945
	25	15.7147	6.4360	2.7148	5.5084	4.9614
0.90	5	2.6527	1.6015	1.3852	2.0997	1.8752
	10	4.5794	2.6185	1.6216	2.7549	2.3694
	15	7.4124	3.9009	2.0960	3.7669	3.1920
	20	11.8901	5.6427	2.7095	5.2586	4.6352
	25	17.9607	7.8400	3.4701	7.2633	6.5204

Table 2.24: Simulated relative efficiency of the restricted, positive James-Stein and penalty estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 64$, $p - q = 4$ and $\Delta^2 = 0$ for different values of ρ and q based on the CAR model

ρ	q	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{JS+}$	$\hat{\boldsymbol{\beta}}^{LASSO}$	$\hat{\boldsymbol{\beta}}^{A.LASSO}$	$\hat{\boldsymbol{\beta}}^{SCAD}$
-0.95	5	2.3051	1.5708	1.0321	1.5953	1.3179
	10	3.9968	2.5347	1.1520	1.9103	1.5586
	15	5.6321	3.6063	1.3659	2.4380	2.0657
	20	8.1514	4.7615	1.6130	2.8766	2.5115
	25	11.1813	6.2885	1.9185	3.6213	3.1771
-0.50	5	2.3977	1.6040	0.9747	1.5406	1.2412
	10	4.0781	2.6297	1.1176	1.8802	1.5021
	15	6.0539	3.7408	1.3007	2.2584	1.8705
	20	8.2454	4.9361	1.5322	2.7780	2.4834
	25	11.6291	6.2906	1.8528	3.4227	2.9380
0.00	5	2.3458	1.6039	0.9522	1.4717	1.1944
	10	4.0894	2.6258	1.1002	1.8999	1.5222
	15	6.2356	3.7018	1.2697	2.2091	1.7539
	20	8.6303	4.8372	1.5232	2.6897	2.2732
	25	11.8719	6.2049	1.8590	3.4116	2.8723
0.50	5	2.4049	1.6084	0.9919	1.5660	1.2424
	10	4.1553	2.6344	1.1459	1.9113	1.5602
	15	6.1165	3.6585	1.3780	2.4287	1.9693
	20	8.5591	4.9969	1.5892	2.8705	2.4049
	25	11.9753	6.2183	1.9364	3.6719	3.2230
0.90	5	2.5291	1.5772	1.2217	1.8797	1.5486
	10	4.3628	2.5360	1.4292	2.4108	1.9368
	15	6.5771	3.6803	1.6983	2.9371	2.4058
	20	9.7069	4.9972	1.9792	3.6229	2.9859
	25	13.0738	6.6616	2.4282	4.4767	3.9300

Table 2.25: Simulated relative efficiency of the restricted, positive James-Stein and penalty estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 100$, $p - q = 4$ and $\Delta^2 = 0$ for different values of ρ and q based on the CAR model

ρ	q	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{JS+}$	$\hat{\boldsymbol{\beta}}^{LASSO}$	$\hat{\boldsymbol{\beta}}^{A.LASSO}$	$\hat{\boldsymbol{\beta}}^{SCAD}$
-0.95	5	2.2926	1.5660	1.1005	1.9881	1.4249
	10	3.6788	2.4935	1.2294	2.4862	1.7530
	15	5.1634	3.4215	1.4210	2.9173	2.2110
	20	6.9769	4.4252	1.5951	3.2607	2.5834
	25	8.6824	5.4691	1.8464	3.8328	3.2033
-0.50	5	2.3140	1.5954	1.0149	1.8302	1.2990
	10	3.8221	2.5749	1.1532	2.2980	1.6740
	15	5.3100	3.5021	1.3272	2.7467	2.0227
	20	7.4015	4.6831	1.5501	3.3110	2.5712
	25	9.1404	5.6321	1.7262	3.6417	2.8883
0.00	5	2.3644	1.6311	0.9697	1.7683	1.2337
	10	3.8354	2.5439	1.1289	2.2868	1.5971
	15	5.5753	3.5968	1.2846	2.6278	1.9286
	20	7.6565	4.7442	1.5067	3.0660	2.4247
	25	9.2551	5.6831	1.6990	3.5944	2.8202
0.50	5	2.3718	1.6121	1.0110	1.8383	1.2803
	10	3.9038	2.5516	1.1875	2.4234	1.6732
	15	5.7310	3.5972	1.3417	2.7847	1.9492
	20	7.6471	4.6469	1.5613	3.2695	2.4980
	25	9.5247	5.5812	1.7615	3.7110	2.9541
0.90	5	2.4853	1.6025	1.2358	2.2698	1.6019
	10	3.9386	2.4547	1.4382	2.8938	2.0401
	15	5.8313	3.4234	1.6395	3.4570	2.5209
	20	7.8693	4.3889	1.8978	3.9747	3.0520
	25	10.4245	5.5692	2.1638	4.5160	3.7376

From the results summarized in Tables 2.23-2.25, we can conclude the following:

1. The restricted estimator, $\hat{\boldsymbol{\beta}}^R$, outperforms all other estimators for all the cases that are considered in this simulation. This is expected, since we are working completely under the candidate true subspace. That is, the data generating model is under $\Delta^2 = 0$.

2. Changing the value of ρ does not have much impact on the SRE values for a fixed n and q.

3. In general, the positive James-Stein estimator, $\hat{\boldsymbol{\beta}}^{JS+}$, outperforms all the penalty estimators.

Application to Boston Housing Prices Data

Harrison and Rubinfeld (1978) studied several practical issues related to the use of housing market data for census tracts in the Boston Standard Metropolitan Statistical Area (SMSA) in 1970. Among others, Breiman and Friedman (1985), Lange and Ryan (1989), Pace (1993), Stine (2004) have used this data for illustration purposes. The major objective in all these works was to identify the relationship between a set of over 13 covariates and the median value of owner-occupied houses in Boston.

The data consist of 506 observations, each relating to one census tract. The data contain the following variables, the tract id number (`TRACT`), the median values of owner-occupied housing in (USD 1000's) (`MEDV`), the corrected median values of owner-occupied housing in (USD 1000's) (`CMEDV`), the proportions of residential land zoned for lots over 2500 sq.ft per town (constant for all Boston tracts) (`ZN`), the proportions of non-retail business areas per town (`INDUS`), average numbers of rooms per dwelling (`RM`), proportions of owner-occupied units built prior to 1940 (`AGE`), a dummy variable with two levels, 1 if tract border to Charles River; 0 otherwise (`CHAS`), levels of nitrogen oxides concentration (parts per 10 million) per town (`NOX`), crime rate per capita (`CRIM`), weighted distance to five employment centers (`DIS`), an index of accessibility to radial highway per town (constant for all Boston tracts) (`RAD`), percentage of lower status population (`LSTAT`), property tax rate per (USD 10,000) per town (constant for all Boston tracts) (`TAX`), pupil-teacher ratios per town (constant for all Boston tracts) (`PTRATIO`), the variable $1000(b - 0.63)^2$, where b is the proportion of blacks (`B`), the location of each tract in latitude (`LAT`) and longitude (`LON`), where the last two variables were added by Pace and Gilley (1997).

Following Pace and Gilley (1997), we predict the response variable `log(CMEDV)` using the available predictors assuming a Gaussian CAR. We fit a full CAR model, then three submodels are selected using a forward selection method based on AIC and BIC selection procedures. The first submodel contains the most important two predictors that have the smallest AIC and BIC values among all possible groups of two predictors, the second submodel contains the best three predictors that have the smallest values of AIC and BIC values among all possible groups of three predictors including the previous two in the first submodel. The third submodel was the final one selected by the AIC and BIC selection methods for which including any other predictors would not decrease the values of the AIC or BIC. Each of the selected models is considered as a candidate subspace model and the various proposed estimators are obtained based on such candidate submodel. The full and candidate submodels are summarized in Table 2.26.

Table 2.26: Full and submodels for the Boston Housing data.

Selection Criterion	Model
Full	`log(CMEDV) ~CRIM + I(RM^2) + log(LSTAT)+ TAX + CHAS + I(NOX^2)+ log(DIS) + log(RAD)+ B + PTRATIO + ZN + INDUS + AGE+LAT + LON`
Model 1	`log(CMEDV) ~ I(RM^2) + log(LSTAT)`
Model 2	`log(CMEDV) ~ CRIM + I(RM^2) + log(LSTAT)`
Model 3	`log(CMEDV) ~CRIM + I(RM^2) + log(LSTAT)+ TAX`

The LASSO, adaptive LASSO and SCAD estimators, are computed as in the previous data example using a 10−fold cross-validation procedure. We used the same bootstrap procedure explained in the data example of Section 2.8.1 to compute the *relative mean squared prediction error* (RMSPE) with respect to the full model estimator, defined in (2.38). The bootstrap sample size was set to $B = 1000$. Our findings are summarized in Table 2.27.

Table 2.27: RMSPE with respect to $\hat{\boldsymbol{\beta}}$ for Boston Housing data based on the CAR model

Model	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{JS+}$
Model 1	1.1112	1.0980
Model 2	1.1054	1.0914
Model 3	1.0982	1.0839
Penalty Estimators		
$\hat{\boldsymbol{\beta}}^{LASSO}$	$\hat{\boldsymbol{\beta}}^{A.LASSO}$	$\hat{\boldsymbol{\beta}}^{SCAD}$
0.9737	1.0606	1.0532

The following conclusions can be drawn from Table 2.27

1. The restricted estimator, $\hat{\boldsymbol{\beta}}^R$, outperforms all estimators regardless of the submodel chosen. Therefore, if the restriction given by the submodel is correct, then $\hat{\boldsymbol{\beta}}^R$ is optimum.

2. The positive James-Stein estimator dominates all penalty estimators in all can-

didate submodels.

3. The adaptive LASSO and the SCAD estimators have similar performance and both are better than the LASSO as expected.

4. The first submodel is recommended, because it gives the highest RMSPE with respect to $\hat{\boldsymbol{\beta}}$.

2.9 Conclusion

In this chapter, we suggested the pretest and James-Stein shrinkage estimators for the large-scale effects $\boldsymbol{\beta}$ in the conditional autoregressive model CAR. These estimators were based on uncertain prior information (UPI) in the form of a linear hypothesis $\boldsymbol{H\beta} = \boldsymbol{h}$ whereby a restricted estimator under this hypothesis and an unrestricted MLE were combined. Analytical formulae were derived to calculate the risks and biases of these estimators and their relative performances were examined via these formulae.

An algorithm for obtaining penalty estimators for the large-scale effects of the CAR model was also proposed and applied in computing LASSO, Adaptive LASSO and SCAD estimators. These arrays of estimators were then compared through Monte Carlo simulations and by means of real data sets on housing prices and crime distribution.

Our analytical and numerical studies showed that, in general, the class of the proposed shrinkage estimators, safeguard against the high risks associated with submodels when the validity of such submodels is questionable, while providing a higher efficiency than the full models. Also, the positive James-Stein estimator proved to be superior to the the class of penalty estimators in many of the situations considered in our simulations.

In any case, completely relying on models obtained through subjective information or through model selection procedures such the LASSO family of procedures is not a wise choice, given the risks attached to the validity of this information. Therefore, we recommend the use of the James-Stein estimators for the large-scale effects of the CAR model while model selection methods can be used to provide the prior uncertain information. For instance, $\hat{\boldsymbol{\beta}}^{JS+}$, will always result in a reasonably good performance relative to the estimators obtained via the reduced and full models as well as the penalty estimators regardless of the accuracy of the given restriction.

Chapter 3

Efficient estimation for the Simultaneous Autoregressive Spatial Model

3.1 Introduction

In this chapter we will consider the spatial regression model known as SAR which was introduced in Section 1.5.2. Following the structure laid down in Chapter 2, we will first review the existing results on the maximum likelihood estimators, $\hat{\boldsymbol{\beta}}$, for the large-effect parameters $\boldsymbol{\beta}$ of the SAR model in (1.18). Secondly, we will compute the restricted MLE, $\hat{\boldsymbol{\beta}}^R$, of these parameters under the general candidate subspace $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$ representing the uncertain prior information obtained either by subjective opinion or through model selection methodologies. Thirdly, we will construct pretest and shrinkage estimators for these large-effect parameters. Monte Carlo simulations are conducted to confirm these asymptotic relative risk performances. Finally, we apply these estimators to a real data set.

The second objective of the chapter is to reappraise three penalty estimators, namely, LASSO, Adaptive LASSO and SCAD estimators for the large-effect coefficients of the SAR model. The performance of these estimators are then compared with the restricted and positive shrinkage estimators by using Monte Carlo simulations via their simulated relative efficiency, and using a real data example via their relative mean squared prediction error as was done in Chapter 2 for the CAR model.

3.1.1 Chapter Organization

In Section 3.2, we discuss the model and preliminaries. The proposed estimation strategies using the unrestricted, restricted, pretest, and James-Stein estimators was discussed in 3.3. Numerical studies using simulation experiments and a real data example to confirm theoretical results are illustrated in Section 3.4. In Section 3.5, we consider estimating the mean vector $\boldsymbol{\beta}$ of the SAR model using penalty estimators, and illustrate numerical studies to compare their performance with both the restricted and positive James-Stein estimators. Conclusions are presented in Section 3.6.

3.2 The model and preliminaries

Recall the simultaneous autoregressive spatial model (SAR) introduced in Section 1.5.2 where $\boldsymbol{R} = \rho \boldsymbol{W}^*$ and $\boldsymbol{\Lambda} = \sigma^2 \boldsymbol{I}$ so that the vector of observations over a regular lattice has the following joint distribution

$$\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2(\boldsymbol{I} - \rho\boldsymbol{W}^*)^{-1}(\boldsymbol{I} - \rho\boldsymbol{W}^{*\prime})^{-1}). \tag{3.1}$$

As in Chapters 1& 2, $\boldsymbol{Y} = \boldsymbol{Y}(\boldsymbol{s}) = (Y(s_1), ..., Y(s_n))$ are observations at the spatial locations s_i which form a regular lattice, $\boldsymbol{W} = \{w_{ij}\}_{i,j=1}^n$ is a spatial proximity weight matrix, ρ is a parameter controlling the spatial dependence and chosen so that $(\boldsymbol{I}-\rho\boldsymbol{W})$ is nonsingular matrix. The quantity $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{s})$ is an $n\times p$ matrix containing location specific covariates and $\boldsymbol{\beta}$ is a p-dimensional unknown vector of large-effect parameters.

The SAR model was first introduced by Whittle (1954) who showed that the least squares estimator of $\boldsymbol{\beta}$ is inconsistent while Ord (1975) showed that the MLEs for such parameters are consistent estimators. Statistical inference of the SAR model appears mostly in economics literature. For example, Bell and Bockstael (2000) used the generalized-moments estimation technique for the SAR model in the context of micro level spatially correlated data. Lee and Yu (2010) established the asymptotic properties of the quasi-maximum likelihood estimator in economic panel data with fixed effects and SAR errors. Su (2012) proposed generalized method of moments (GMM) estimators for a semiparametric SAR model and derived their limiting distributions. Su and Jin (2010) proposed a profile quasi-maximum likelihood estimation of a partially linear SAR model and showed that such estimators are consistent at the usual $\sqrt{n}$ rate of convergence. An overview of the statistical inference for the SAR model and its variants can be found in Anselin (1988), Cressie (1993), Wall (2004) and Kazar and Celik (2012)

On the other hand, the literature on model selection and penalized estimation for the SAR model is in its infancy. For instance, Song and De Oliveira (2012) used Bayesian approach for model selection in Gaussian conditional autoregressive CAR and simultaneous autoregressive SAR models for spatial lattice data. More details about model selection for spatial regression models can be found in Chapter 2 of this thesis. Zhu and Liu (2009) proposed a penalized likelihood to estimate the covariance matrix of spatial Gaussian Markov random field models with unspecified neighborhood structure. They used weighted L_1 regularization, and showed that the LASSO type approach gives improved covariance estimators measured by different criteria. They also derived the asymptotic properties of their proposed estimator.

3.2.1 Unrestricted Maximum Likelihood Estimation

Often, a large model containing all available covariates is called full model or unrestricted model. The unrestricted maximum likelihood estimators of the SAR model parameters, $(\boldsymbol{\beta}, \sigma^2, \rho)$, can be obtained by following the procedure described in (i)-(iii) of Section 2.3. The only difference is that we replace the CAR covariance matrix in the log-likelihood function given (2.1) by the covariance matrix of the SAR model

$$\boldsymbol{\Sigma}_{SAR} = \sigma^2 \boldsymbol{Q} = \sigma^2 (\boldsymbol{I} - \rho \boldsymbol{W}^*)^{-1} (\boldsymbol{I} - \rho \boldsymbol{W}^{*\prime})^{-1},$$

to obtain

$$\hat{\boldsymbol{\beta}}(\rho) = (\boldsymbol{X}' (\boldsymbol{I} - \rho \boldsymbol{W}^{*\prime})(\boldsymbol{I} - \rho \boldsymbol{W}^{*\prime}) \boldsymbol{X})^{-1} \boldsymbol{X} (\boldsymbol{I} - \rho \boldsymbol{W}^{*\prime})(\boldsymbol{I} - \rho \boldsymbol{W}^{*\prime}) \boldsymbol{Y}, \quad (3.2)$$

$$\hat{\sigma}^2(\rho) = \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\rho}))' (\boldsymbol{I} - \rho \boldsymbol{W}^{*\prime})(\boldsymbol{I} - \rho \boldsymbol{W}^*)(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\rho}))}{n}. \quad (3.3)$$

By plugging these estimators of $\boldsymbol{\beta}$ and σ^2 into the log-likelihood, the maximum likelihood estimator of ρ can be obtained by maximizing the profile log-likelihood

$$
\begin{aligned}
l^*(\rho) &= -\frac{n}{2}\log\left(\frac{(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}(\rho))'(\boldsymbol{I}-\rho\boldsymbol{W}^*)^{-1}(\boldsymbol{I}-\rho\boldsymbol{W}^{*\prime})^{-1}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}(\rho))}{n}\right) \\
&- \frac{1}{2}\log(|(\boldsymbol{I}-\rho\boldsymbol{W}^*)^{-1}(\boldsymbol{I}-\rho\boldsymbol{W}^{*\prime})^{-1}|).
\end{aligned}
$$

Finally, the unrestricted MLEs of $\boldsymbol{\beta}$ and σ^2 are computed by plugging $\hat{\rho}$ in (3.2) and (3.3), respectively. Similar to Section 2.5.1, the consistence and asymptotic normality of the unrestricted MLEs of the SAR model will follow directly from the general result of Mardia and Marshall (1984). In order to adapt Theorem 2.3.1 for the SAR case, all we need is to replace $\hat{\boldsymbol{C}}$ therein by $\hat{\boldsymbol{Q}} = (\boldsymbol{I}-\hat{\rho}\boldsymbol{W}^*)^{-1}(\boldsymbol{I}-\hat{\rho}\boldsymbol{W}^{*\prime})^{-1}$.

3.3 The Proposed Estimation Strategies

Following the steps of Chapter 2, we consider again the UPI presented in the form of a general linear hypothesis,

$$
\boldsymbol{A}_0 : \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}, \tag{3.4}
$$

where $\boldsymbol{H}$ is a $p\times q$ known matrix of rank (q), and $\boldsymbol{h}$ is a $q\times 1$ known vector of constants. The construction of the restricted $\hat{\boldsymbol{\beta}}^R$, the pretest $\hat{\boldsymbol{\beta}}^{PT}$, the James-Stein $\hat{\boldsymbol{\beta}}^{JS}$ and the positive James-Stein $\hat{\boldsymbol{\beta}}^{JS+}$ estimators in SAR model is similar, mutatis mutandis, to that of the CAR model in Section 2.4. All that is needed to be changed is to replace the matrix $\hat{\boldsymbol{C}}$ in the expressions of Section 2.4 by $\hat{\boldsymbol{Q}} = (\boldsymbol{I}-\hat{\rho}\boldsymbol{W}^*)^{-1}(\boldsymbol{I}-\hat{\rho}\boldsymbol{W}^{*\prime})^{-1}$ and $\hat{\boldsymbol{\beta}}$ therein by the one described above in Section 3.2.1. Therefore, here, we only re-iterate these expressions for the SAR model.

The pretest estimator:

$$\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}} I\left(\Psi_n > \chi^2_{q,\alpha}\right) + \hat{\boldsymbol{\beta}}^R I(\Psi_n \leq \chi^2_{q,\alpha}) \tag{3.5}$$

where $I(A)$ denotes the indicator function for the event A,

$$\Psi_n = \frac{(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'(\boldsymbol{H}(\boldsymbol{X_n'Q_n^{-1}X_n})^{-1}\boldsymbol{H'})^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})}{s_e^2}, \tag{3.6}$$

$$s_e^2 = \frac{(\boldsymbol{Y}_n - \boldsymbol{X}_n\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{Q}}^{-1}(\boldsymbol{Y}_n - \boldsymbol{X}_n\hat{\boldsymbol{\beta}})}{n-p}, \tag{3.7}$$

and $\hat{\boldsymbol{Q}} = (\boldsymbol{I} - \hat{\rho}\boldsymbol{W}^*)^{-1}(\boldsymbol{I} - \hat{\rho}\boldsymbol{W}^{*\prime})^{-1}$, $\chi^2_{q,\alpha}$ is the α^{th} upper quantile of a central chi-square distribution with q degrees of freedom.

The James-Stein estimator:

$$\hat{\boldsymbol{\beta}}^{JS} = \hat{\boldsymbol{\beta}}^R + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)\{1 - (q-2)\Psi_n^{-1}\}. \tag{3.8}$$

The positive rule James-Stein estimator:

$$\hat{\boldsymbol{\beta}}^{JS+} = \hat{\boldsymbol{\beta}}^R + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)\{1 - (q-2)\Psi_n^{-1}\}^+, \tag{3.9}$$

where $u^+ = \max(0, u)$.

Also, the asymptotic distributional quadratic risk (AQR), mean squared error matrix (AMSEM) and bias results and performance conclusions are also same, mutatis mutandis. Therefore, we will not repeat risk expressions and performance analysis here, but rather, we will proceed directly to the numerical studies using Monte Carlo simulations as well as application to real data.

Similarly, the construction of penalty estimators for the SAR model parameters, $\boldsymbol{\beta}$, follows the same lines as in Section 2.7 with replacement of $\boldsymbol{C}$ in (2.28) by $\boldsymbol{Q} = (\boldsymbol{I} - \rho \boldsymbol{W}^*)^{-1}(\boldsymbol{I} - \rho \boldsymbol{W}^{*\prime})^{-1}$.

3.4 Numerical Studies

As in Chapter 2, in this section we carry out two sets of Monte Carlo simulations. The first is to compare the performances of the restricted, pretest, James-Stein and positive James-Stein estimators relative to the unrestricted estimator as a benchmark. The second set of simulations restricts attention to comparisons between the positive James-Stein estimator, the restricted estimator and the class of penalty estimators. We apply the methods to the Boston housing and Columbus crime data sets as was done in Chapter 2 and prescribe a bootstrap procedure for estimating the prediction errors of the estimators.

3.4.1 Comparing the Unrestricted with Shrinkage Estimators

In this simulation study, we used $\sigma^2 = 1$, a queen-based proximity matrix $\boldsymbol{W}^*$ with the rest of the parameters being exactly the same as in Chapter 2. The results of the simulations are reported in Tables 3.1- 3.20 and Figures 3.1-3.5.

Figure 3.1: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = -0.90$ for different values of (p, q) based on the SAR model.

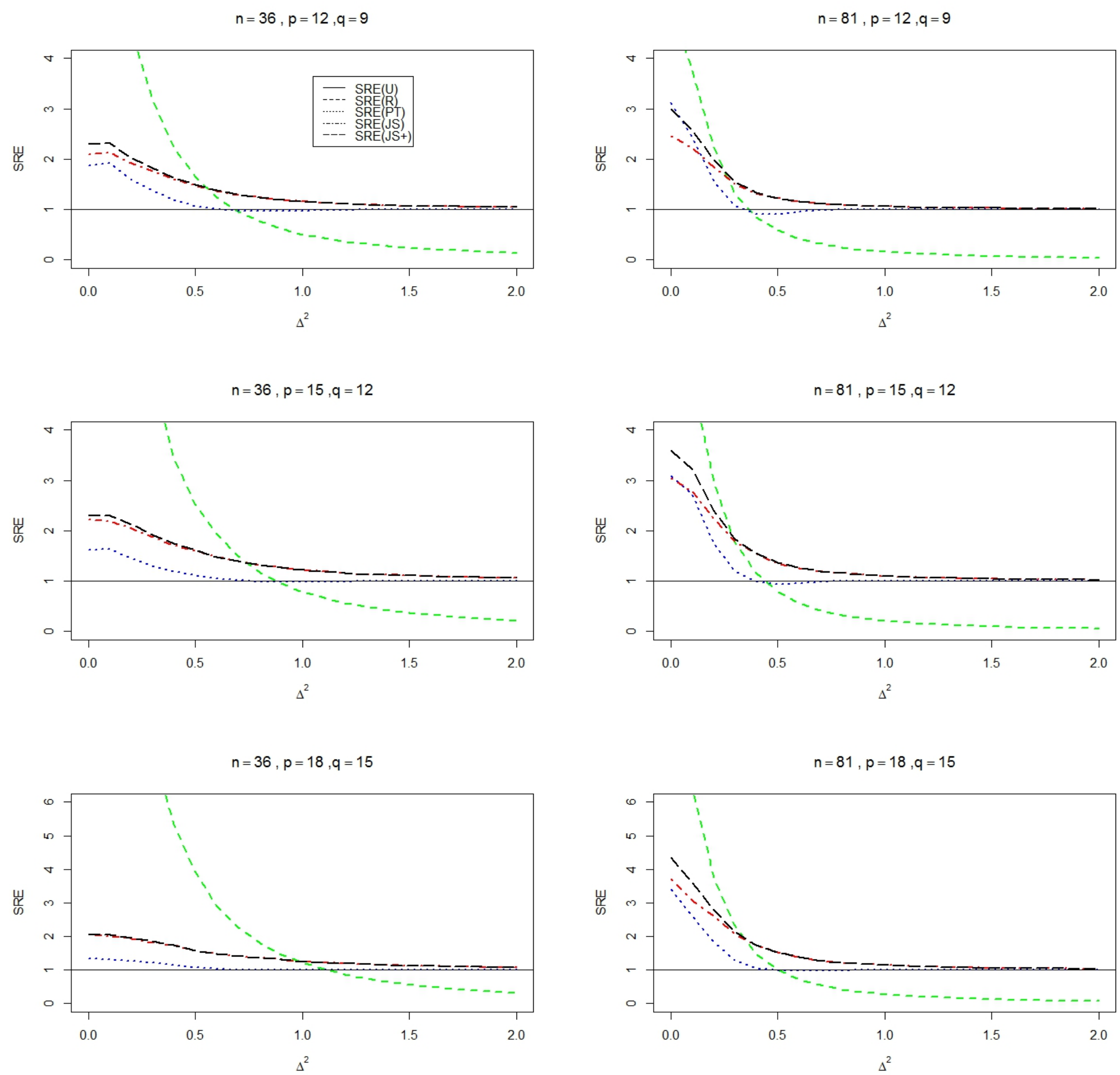

Figure 3.2: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = -0.50$ for different values of (p, q) based on the SAR model.

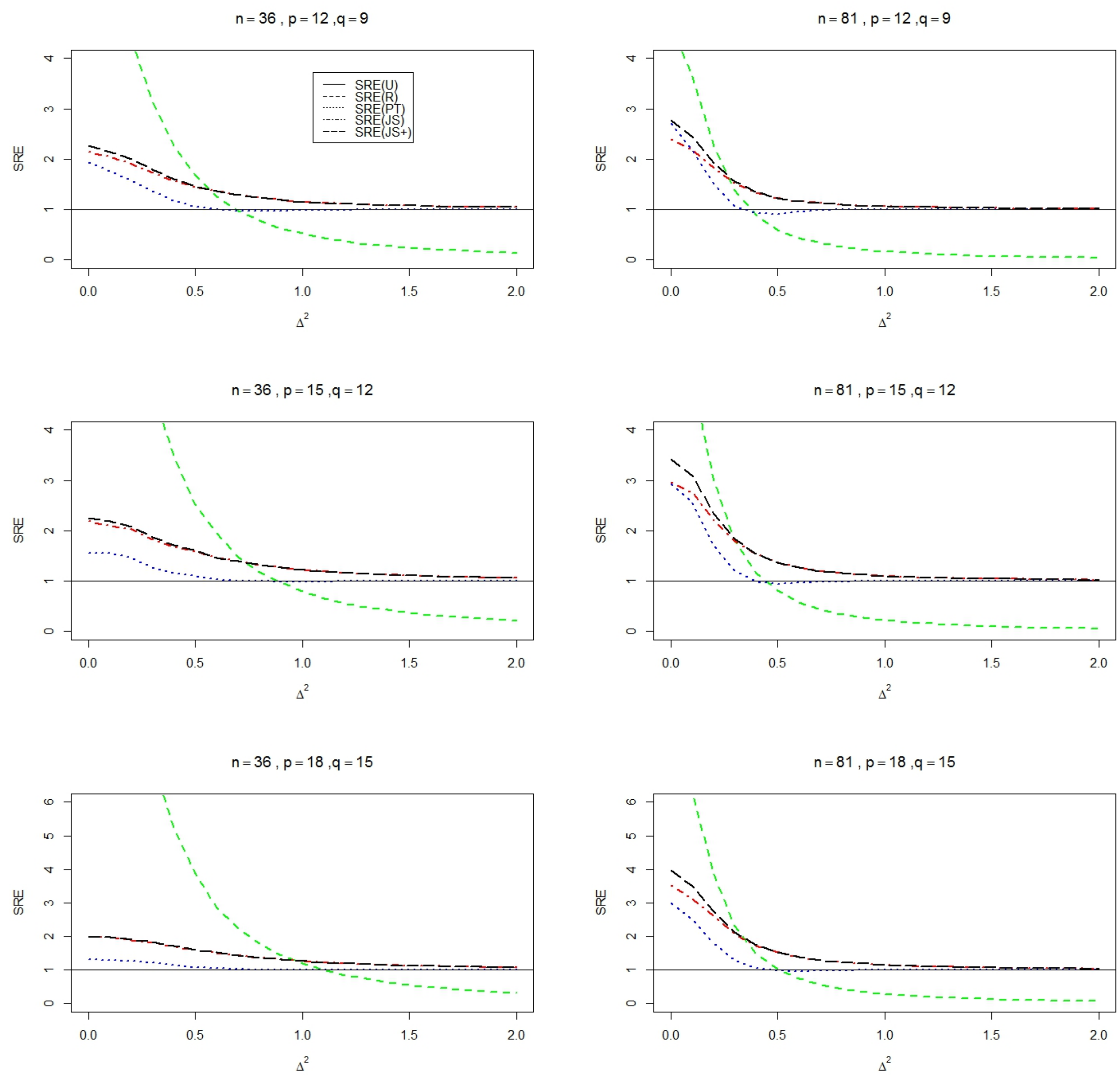

Figure 3.3: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = 0$ for different values of (p, q) based on the SAR model.

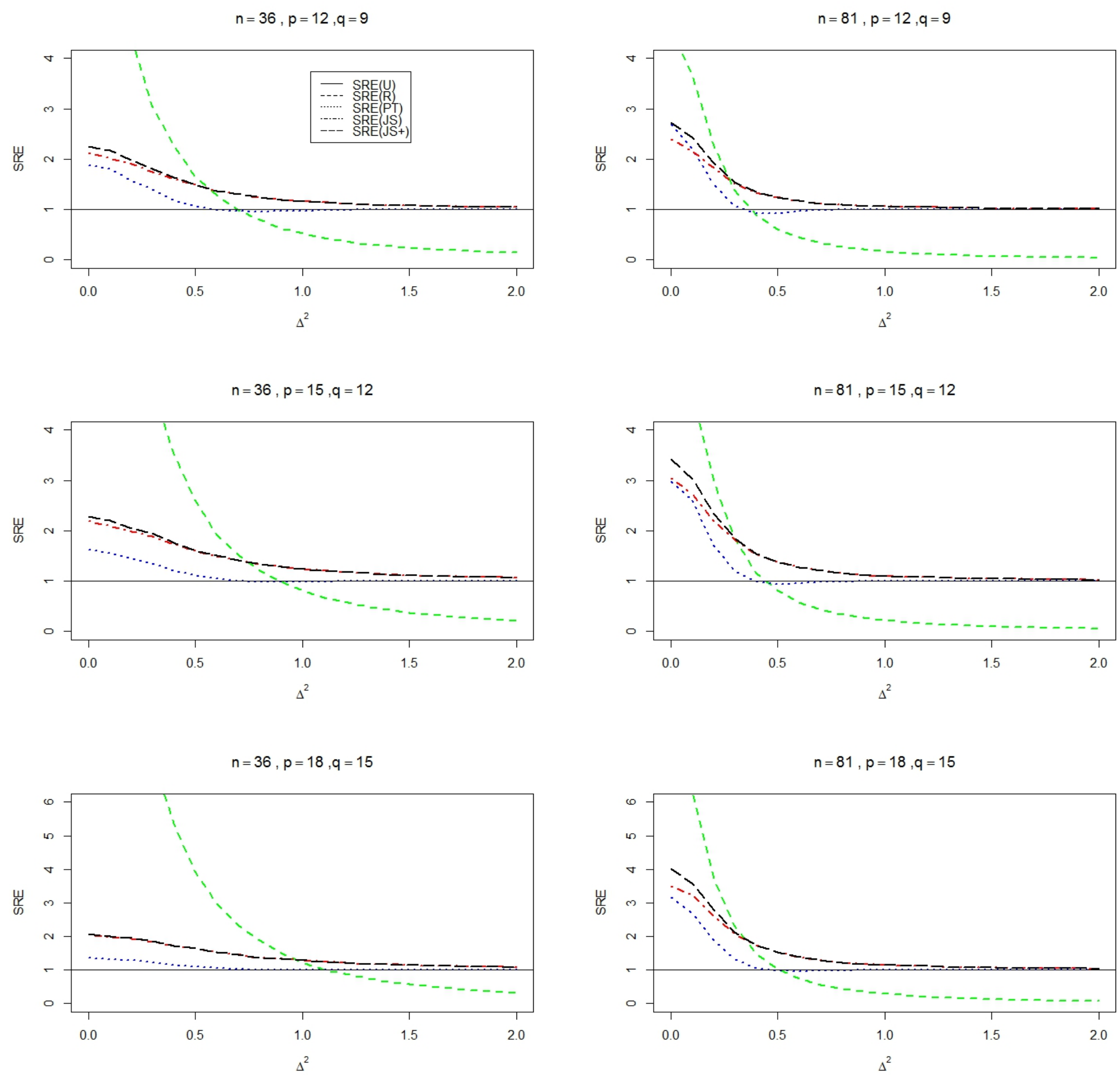

Figure 3.4: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = 0.50$ for different values of (p, q) based on the SAR model.

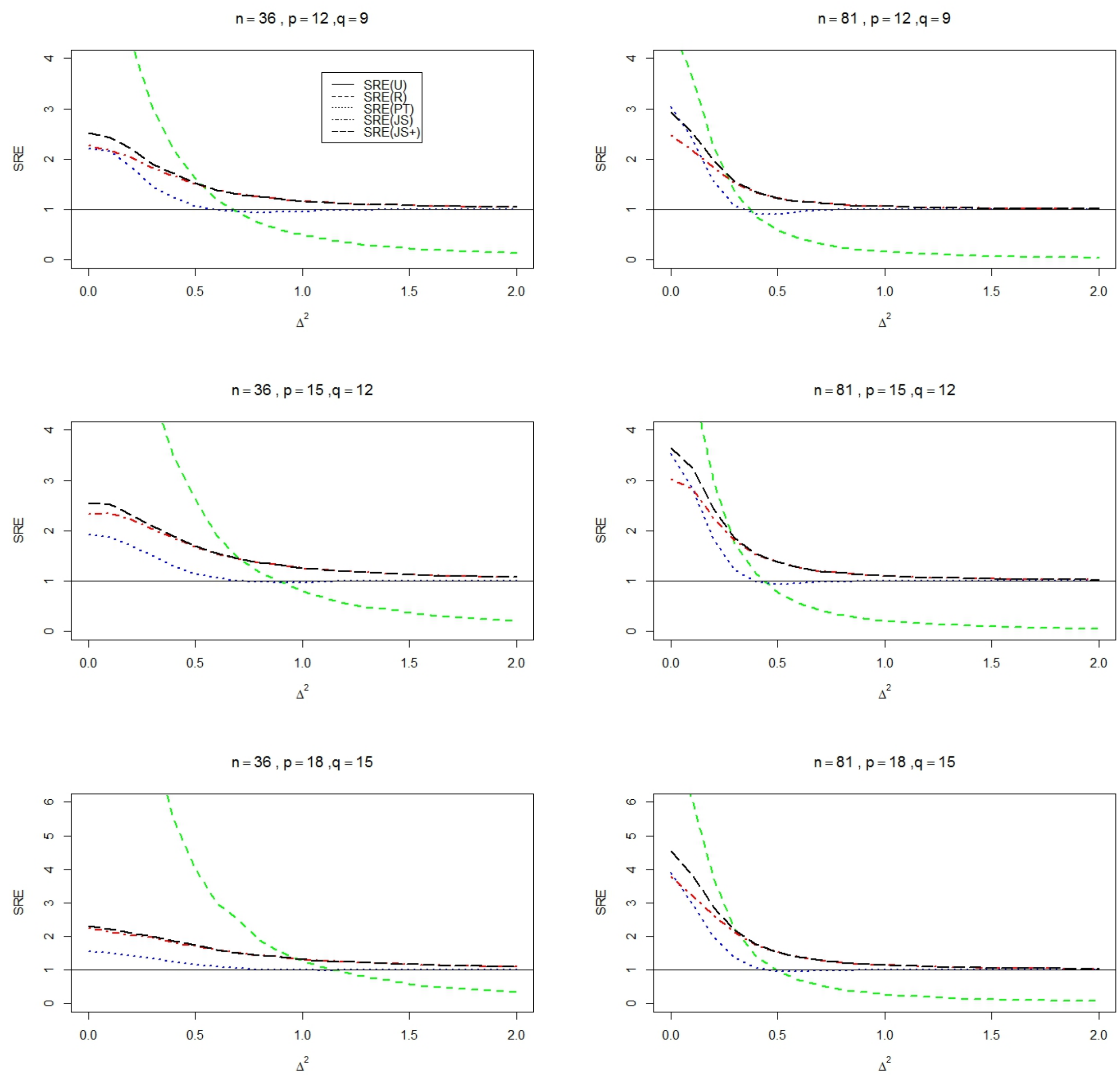

Figure 3.5: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = 0.90$ for different values of (p, q) based on the SAR model.

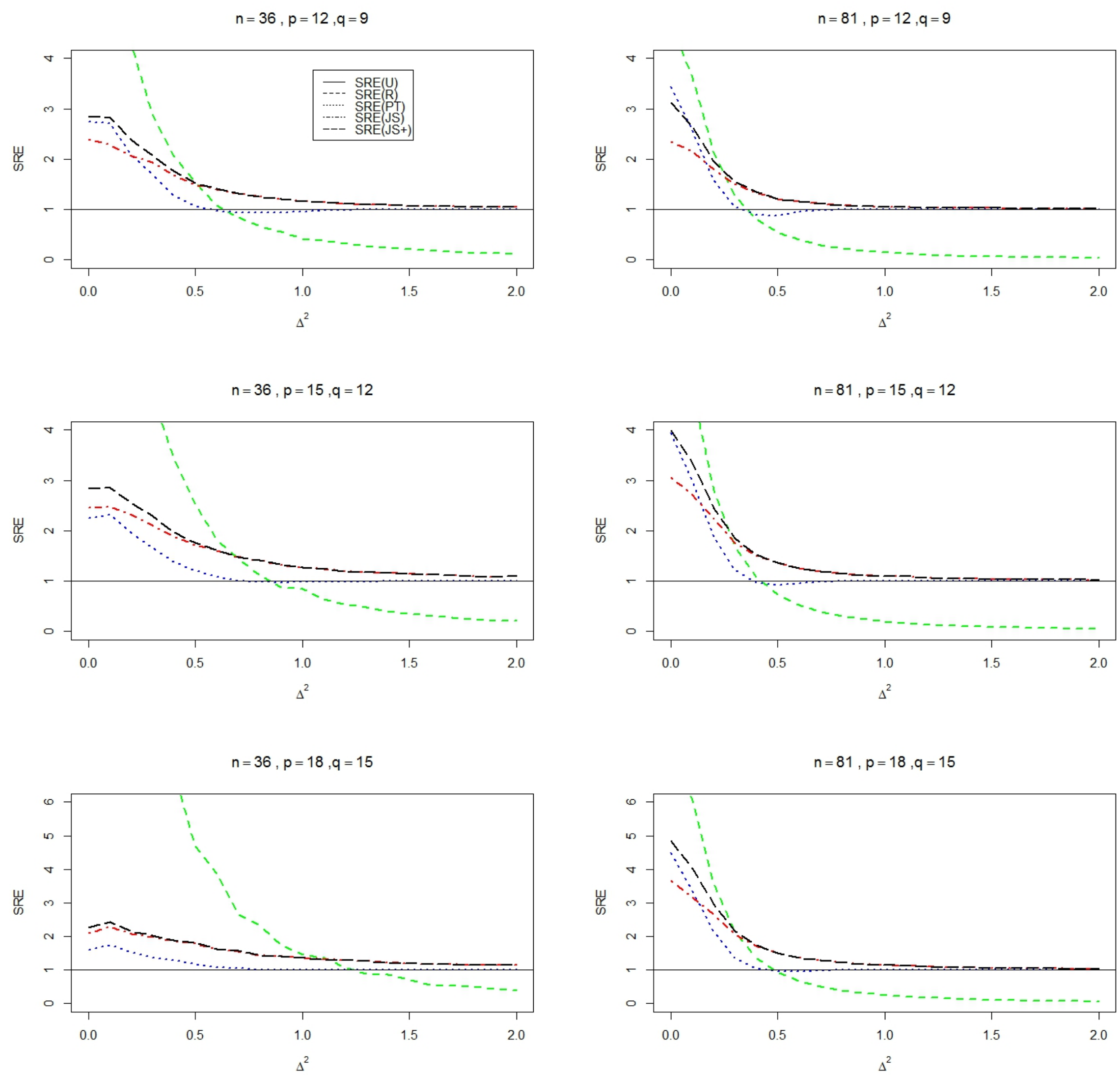

Table 3.1: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = -0.90$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.3946	1.7614	1.2359	1.3132
0.1	2.1520	1.5739	1.2083	1.2658
0.3	1.0803	0.9711	1.0971	1.1171
0.5	0.5820	0.7929	1.0387	1.0426
0.7	0.3317	0.8601	1.0130	1.0143
0.9	0.2134	0.9572	1.0127	1.0127
1.1	0.1482	0.9875	1.0057	1.0057
1.3	0.1103	0.9968	1.0043	1.0043
1.5	0.0811	1.0000	1.0026	1.0026
1.7	0.0639	1.0000	1.0025	1.0025
2.0	0.0462	1.0000	1.0029	1.0029

Table 3.2: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = -0.90$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	4.2391	2.1070	1.7579	1.9932
0.1	3.7940	1.9111	1.7416	1.9170
0.3	1.9690	1.2743	1.4703	1.5264
0.5	1.0421	0.9568	1.2506	1.2648
0.7	0.6035	0.9226	1.1493	1.1521
0.9	0.3875	0.9464	1.0947	1.0949
1.1	0.2668	0.9769	1.0600	1.0600
1.3	0.1945	0.9977	1.0471	1.0471
1.5	0.1464	0.9972	1.0347	1.0347
1.7	0.1179	1.0000	1.0268	1.0268
2.0	0.0856	1.0000	1.0214	1.0214

Table 3.3: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = -0.50$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.3180	1.7109	1.1769	1.2892
0.1	2.0006	1.4879	1.2001	1.2526
0.3	1.1336	0.9834	1.0820	1.1116
0.5	0.5975	0.8119	1.0322	1.0364
0.7	0.3544	0.8922	1.0203	1.0211
0.9	0.2252	0.9511	1.0098	1.0098
1.1	0.1563	0.9955	1.0088	1.0088
1.3	0.1133	1.0000	1.0054	1.0054
1.5	0.0861	1.0000	1.0045	1.0045
1.7	0.0679	1.0000	1.0027	1.0027
2.0	0.0503	1.0000	1.0026	1.0026

Table 3.4: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = -0.50$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	4.0927	2.0340	1.7804	1.9039
0.1	3.4740	1.8928	1.7210	1.8381
0.3	1.9874	1.2511	1.4621	1.5083
0.5	1.0401	0.9586	1.2559	1.2650
0.7	0.6192	0.9197	1.1491	1.1492
0.9	0.3982	0.9602	1.0978	1.0978
1.1	0.2786	0.9916	1.0649	1.0649
1.3	0.1956	0.9939	1.0457	1.0457
1.5	0.1548	1.0000	1.0381	1.0381
1.7	0.1181	1.0000	1.0285	1.0285
2.0	0.0863	1.0000	1.0185	1.0185

Table 3.5: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = 0$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.2733	1.6644	1.1967	1.2761
0.1	2.0276	1.4776	1.2010	1.2436
0.3	1.1396	0.9876	1.0912	1.1123
0.5	0.5986	0.8084	1.0359	1.0392
0.7	0.3621	0.8757	1.0195	1.0197
0.9	0.2259	0.9575	1.0099	1.0099
1.1	0.1540	0.9930	1.0067	1.0067
1.3	0.1140	0.9907	1.0063	1.0063
1.5	0.0869	1.0000	1.0027	1.0027
1.7	0.0690	1.0000	1.0044	1.0044
2.0	0.0504	1.0000	1.0020	1.0020

Table 3.6: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = 0$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.8909	1.9435	1.7887	1.8923
0.1	3.6014	1.7993	1.6810	1.8138
0.3	1.9699	1.2525	1.4413	1.4955
0.5	1.0375	0.9586	1.2582	1.2712
0.7	0.6208	0.9128	1.1506	1.1511
0.9	0.3978	0.9498	1.0891	1.0891
1.1	0.2796	0.9848	1.0624	1.0628
1.3	0.2018	0.9974	1.0459	1.0459
1.5	0.1504	1.0000	1.0340	1.0340
1.7	0.1203	1.0000	1.0281	1.0281
2.0	0.0893	1.0000	1.0183	1.0183

Table 3.7: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = 0.50$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.2625	1.7122	1.2177	1.3017
0.1	2.0114	1.5412	1.2109	1.2680
0.3	1.0903	0.9772	1.0982	1.1174
0.5	0.5758	0.7826	1.0317	1.0359
0.7	0.3322	0.8782	1.0152	1.0176
0.9	0.2126	0.9513	1.0083	1.0083
1.1	0.1478	0.9941	1.0066	1.0066
1.3	0.1092	1.0000	1.0019	1.0019
1.5	0.0827	1.0000	1.0031	1.0031
1.7	0.0641	1.0000	1.0050	1.0050
2.0	0.0466	1.0000	1.0031	1.0031

Table 3.8: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = 0.50$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	4.0745	2.2484	1.8266	2.0504
0.1	3.4015	2.0832	1.7438	1.9394
0.3	1.8984	1.2966	1.4385	1.5215
0.5	0.9988	0.9445	1.2606	1.2725
0.7	0.5826	0.8946	1.1421	1.1442
0.9	0.3727	0.9371	1.0901	1.0902
1.1	0.2582	0.9762	1.0594	1.0594
1.3	0.1898	0.9923	1.0402	1.0402
1.5	0.1433	0.9979	1.0334	1.0334
1.7	0.1121	1.0000	1.0278	1.0278
2.0	0.0810	1.0000	1.0205	1.0205

Table 3.9: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = 0.90$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^{R}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.2404	1.8612	1.2720	1.3543
0.1	2.0145	1.6023	1.2149	1.2984
0.3	1.0518	0.9451	1.0533	1.1173
0.5	0.5142	0.7649	1.0301	1.0341
0.7	0.3101	0.8323	1.0067	1.0113
0.9	0.1948	0.9474	1.0094	1.0094
1.1	0.1333	0.9890	1.0041	1.0041
1.3	0.1003	1.0000	1.0038	1.0038
1.5	0.0744	1.0000	1.0031	1.0031
1.7	0.0575	1.0000	1.0031	1.0031
2.0	0.0417	1.0000	1.0017	1.0017

Table 3.10: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = 0.90$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^{R}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.9870	2.5775	1.7814	2.2395
0.1	3.3952	2.3104	1.7250	2.0933
0.3	1.8755	1.3083	1.4283	1.5517
0.5	0.9204	0.9126	1.2561	1.2689
0.7	0.5276	0.8484	1.1281	1.1348
0.9	0.3482	0.9494	1.0908	1.0908
1.1	0.2387	0.9778	1.0642	1.0642
1.3	0.1729	0.9925	1.0406	1.0406
1.5	0.1274	1.0000	1.0300	1.0300
1.7	0.1025	1.0000	1.0217	1.0217
2.0	0.0726	1.0000	1.0204	1.0204

Table 3.11: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = -0.90$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1239	1.8077	1.2162	1.3241
0.1	1.6552	1.3894	1.1622	1.2406
0.3	0.5767	0.7600	1.0353	1.0401
0.5	0.2576	0.9238	1.0108	1.0110
0.7	0.1412	0.9975	1.0048	1.0048
0.9	0.0886	1.0000	1.0032	1.0032
1.1	0.0601	1.0000	1.0018	1.0018
1.3	0.0433	1.0000	1.0014	1.0014
1.5	0.0332	1.0000	1.0015	1.0015
1.7	0.0255	1.0000	1.0004	1.0004
2.0	0.0182	1.0000	1.0004	1.0004

Table 3.12: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = -0.90$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.3544	2.5017	1.7997	2.1403
0.1	2.6386	1.9034	1.7011	1.8933
0.3	0.9378	0.9002	1.2404	1.2638
0.5	0.4117	0.8910	1.0856	1.0867
0.7	0.2230	0.9943	1.0511	1.0511
0.9	0.1393	1.0000	1.0322	1.0322
1.1	0.0946	1.0000	1.0222	1.0222
1.3	0.0678	1.0000	1.0142	1.0142
1.5	0.0522	1.0000	1.0090	1.0090
1.7	0.0399	1.0000	1.0086	1.0086
2.0	0.0289	1.0000	1.0088	1.0088

Table 3.13: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = -0.50$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.0973	1.7649	1.2278	1.3071
0.1	1.6931	1.3781	1.1724	1.2237
0.3	0.6107	0.7720	1.0362	1.0412
0.5	0.2691	0.9120	1.0110	1.0110
0.7	0.1500	0.9979	1.0061	1.0061
0.9	0.0924	1.0000	1.0042	1.0042
1.1	0.0616	1.0000	1.0021	1.0021
1.3	0.0458	1.0000	1.0003	1.0003
1.5	0.0342	1.0000	1.0004	1.0004
1.7	0.0270	1.0000	1.0017	1.0017
2.0	0.0197	1.0000	1.0010	1.0010

Table 3.14: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = -0.50$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.2932	2.3870	1.8402	2.0741
0.1	2.6815	1.8723	1.6772	1.8604
0.3	0.9589	0.9081	1.2479	1.2705
0.5	0.4260	0.9129	1.1013	1.1020
0.7	0.2314	0.9874	1.0502	1.0502
0.9	0.1460	1.0000	1.0317	1.0317
1.1	0.0988	1.0000	1.0193	1.0193
1.3	0.0715	1.0000	1.0132	1.0132
1.5	0.0550	1.0000	1.0126	1.0126
1.7	0.0421	1.0000	1.0109	1.0109
2.0	0.0306	1.0000	1.0062	1.0062

Table 3.15: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = 0$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.0200	1.6994	1.2038	1.2911
0.1	1.6470	1.3478	1.1729	1.2156
0.3	0.6169	0.7860	1.0354	1.0409
0.5	0.2746	0.9272	1.0097	1.0100
0.7	0.1499	0.9983	1.0056	1.0056
0.9	0.0923	1.0000	1.0045	1.0045
1.1	0.0636	1.0000	1.0022	1.0022
1.3	0.0463	1.0000	1.0012	1.0012
1.5	0.0357	1.0000	1.0017	1.0017
1.7	0.0271	1.0000	1.0004	1.0004
2.0	0.0196	1.0000	1.0014	1.0014

Table 3.16: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = 0$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.2799	2.3262	1.8065	2.0342
0.1	2.5585	1.9189	1.6530	1.8446
0.3	0.9934	0.9239	1.2559	1.2680
0.5	0.4357	0.8912	1.0929	1.0956
0.7	0.2373	0.9856	1.0543	1.0543
0.9	0.1452	1.0000	1.0344	1.0344
1.1	0.1028	1.0000	1.0238	1.0238
1.3	0.0728	1.0000	1.0178	1.0178
1.5	0.0552	1.0000	1.0125	1.0125
1.7	0.0437	1.0000	1.0088	1.0088
2.0	0.0319	1.0000	1.0061	1.0061

Table 3.17: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = 0.50$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1258	1.7604	1.1915	1.3107
0.1	1.6204	1.3493	1.1318	1.2209
0.3	0.6149	0.7734	1.0332	1.0377
0.5	0.2644	0.9308	1.0114	1.0114
0.7	0.1459	0.9960	1.0050	1.0050
0.9	0.0911	1.0000	1.0038	1.0038
1.1	0.0612	1.0000	1.0026	1.0026
1.3	0.0434	1.0000	1.0012	1.0012
1.5	0.0346	1.0000	1.0009	1.0009
1.7	0.0261	1.0000	1.0011	1.0011
2.0	0.0186	1.0000	1.0009	1.0009

Table 3.18: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = 0.50$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.3500	2.4012	1.8522	2.1197
0.1	2.5436	1.9451	1.6674	1.8657
0.3	0.9623	0.9058	1.2489	1.2695
0.5	0.4196	0.9021	1.1006	1.1009
0.7	0.2297	0.9987	1.0487	1.0487
0.9	0.1417	1.0000	1.0377	1.0377
1.1	0.0996	1.0000	1.0191	1.0191
1.3	0.0696	1.0000	1.0151	1.0151
1.5	0.0524	1.0000	1.0104	1.0104
1.7	0.0414	1.0000	1.0069	1.0069
2.0	0.0297	1.0000	1.0056	1.0056

Table 3.19: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = 0.90$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1127	1.7931	1.2216	1.3393
0.1	1.6130	1.3735	1.1696	1.2384
0.3	0.5735	0.7445	1.0304	1.0378
0.5	0.2471	0.9145	1.0085	1.0085
0.7	0.1318	0.9923	1.0032	1.0032
0.9	0.0840	1.0000	1.0030	1.0030
1.1	0.0563	1.0000	1.0023	1.0023
1.3	0.0401	1.0000	1.0026	1.0026
1.5	0.0300	1.0000	1.0005	1.0005
1.7	0.0237	1.0000	0.9998	0.9998
2.0	0.0170	1.0000	1.0006	1.0006

Table 3.20: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = 0.90$ based on the SAR model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.3427	2.7123	1.7904	2.2242
0.1	2.5278	2.0246	1.6523	1.9299
0.3	0.9002	0.8894	1.2328	1.2568
0.5	0.3952	0.8933	1.0988	1.1007
0.7	0.2109	0.9936	1.0499	1.0499
0.9	0.1332	1.0000	1.0324	1.0324
1.1	0.0888	1.0000	1.0214	1.0214
1.3	0.0652	1.0000	1.0120	1.0120
1.5	0.0496	1.0000	1.0087	1.0087
1.7	0.0386	1.0000	1.0083	1.0083
2.0	0.0277	1.0000	1.0049	1.0049

The findings from these simulations results can be summarized as follows:

1. In general, changing the value of ρ does not have a significant effect in the SRE results for the same n and (p, q).

2. In all cases, the restricted estimator, $\hat{\boldsymbol{\beta}}^R$, dominates all other estimators when

$\Delta^2 = 0$, but the SMSE of $\hat{\boldsymbol{\beta}}^R$ decreases as Δ^2 moves away from 0, and becomes unbounded, while the SRE of other estimators becomes closer and closer to 1; that is, the SMSE values of the other estimators approaches the value of SMSE of the unrestricted estimator.

3. The SRE of $\hat{\boldsymbol{\beta}}^{JS+}$ increases from 1.31 to more than 2 when $n = 36$, and from 1.30 to more than 4 when $n = 81$ keeping $p - q = 3$ and $\Delta^2 = 0$, that is $\hat{\boldsymbol{\beta}}^{JS+}$ performs better for $n = 81$, while $\hat{\boldsymbol{\beta}}^R$ performs much better for $n = 36$ than $n = 81$.

3.4.2 Columbus Crime Data Analysis

This data set was described in Chapter 2 in the context of CAR model. Anselin (1988) fitted two separate regression models to illustrate the presence of separate levels of spatial dependence for the east and west sides of Columbus city by using SAR model. Also, Griffith (2000) and Lee and Yu (2013) used SAR models to fit this data set.

Recently, Li et al. (2012) considered a one-step estimation of spatial dependence parameter as an alternative method for the maximum likelihood estimation. The authors solved a one-step approximate profile likelihood (APLE) estimating equation which had a closed form. They explored the finite sample and asymptotic properties of the APLE for the SAR model and developed exploratory spatial data analysis tools. They illustrated their methods by using the Columbus crime data set.

Here, we fit a full SAR model based on all the available covariates to predict the log-transformed response variable `log(CRIME)`. A reduced model is then searched for, based on AIC and BIC selection criteria. The resulting models are reported in Table

3.21.

Table 3.21: Full and reduced SAR models for the Columbus crime data

Selection Criterion	Model
Full	`log(CRIME) ~ HOVAL+PLUMB+INC+DISCBD+OPEN`
AIC/BIC	`log(CRIME)~ HOVAL+PLUMB`

Using these models, we then compute the pretest, James-Stein and positive James-Stein estimators. We then compare the performance of the estimators via their relative mean squared prediction errors (RMSPE) based on 2000 bootstrap samples according to the procedure laid down in Section 2.8.1. The results are reported in Table 3.22.

Table 3.22: RMSPE with respect to $\hat{\boldsymbol{\beta}}$ based on SAR model for Columbus crime data

Estimator	RMSPE
$\hat{\boldsymbol{\beta}}^{R}$	1.0895
$\hat{\boldsymbol{\beta}}^{PT}$	1.0557
$\hat{\boldsymbol{\beta}}^{JS+}$	1.0306
$\hat{\boldsymbol{\beta}}^{JS}$	1.0279

It is clear that $\hat{\boldsymbol{\beta}}^{R}$ outperforms all other estimators, which indicates that it is optimum if the null hypothesis is correct. $\hat{\boldsymbol{\beta}}^{PT}$ comes the second, then $\hat{\boldsymbol{\beta}}^{JS+}$. $\hat{\boldsymbol{\beta}}^{JS}$ performs better than $\hat{\boldsymbol{\beta}}$, even though it was the worst among the other estimators.

3.5 Comparison of Penalty and Non-Penalty Estimators

Again, we carry out Monte Carlo simulations with the same exact parameter values as in Section 2.8.2 to compare the performance of $\hat{\boldsymbol{\beta}}^{R}, \hat{\boldsymbol{\beta}}^{JS+}, \hat{\boldsymbol{\beta}}^{Lasso}, \hat{\boldsymbol{\beta}}^{A.Lasso}, \hat{\boldsymbol{\beta}}^{SCAD}$

relative to the benchmark unrestricted estimator, $\hat{\boldsymbol{\beta}}$.

The penalty estimators are computed according to the same procedure given in Section 2.8.2 by using $k = 10$-fold cross validation. In each configuration of the parameters, we used 2000 Monte Carlo runs. The simulated relative efficiency results are reported in Tables 3.23 to 3.25.

Table 3.23: Simulated relative efficiency of the restricted, positive James-Stein and penalty estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 49$, $p - q = 4$ and $\Delta^2 = 0$ for different values of ρ and q based on the SAR model

ρ	q	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{JS+}$	$\hat{\boldsymbol{\beta}}^{Lasso}_{SAR}$	$\hat{\boldsymbol{\beta}}^{A.Lasso}_{SAR}$	$\hat{\boldsymbol{\beta}}^{SCAD}_{SAR}$
-0.95	5	2.6961	1.7635	1.1101	1.4154	1.4488
	10	5.0142	2.9690	1.5131	2.1855	2.2423
	15	8.5343	4.4283	2.1137	3.2113	3.2295
	20	13.9370	5.8915	2.9065	4.7969	4.8001
	25	23.6113	7.6766	4.3543	7.4049	7.1793
-0.50	5	2.6573	1.7179	1.1872	1.5104	1.6164
	10	4.9192	2.9392	1.6464	2.3813	2.5808
	15	8.3743	4.2474	2.3668	3.7411	3.9014
	20	14.2132	6.3728	3.4287	5.7566	6.1516
	25	23.6329	9.5879	5.1867	8.9466	9.4595
0.00	5	2.5972	1.7141	1.2208	1.5477	1.6194
	10	4.9956	2.8972	1.7200	2.5342	2.6841
	15	8.2686	4.2256	2.3359	3.6658	3.9478
	20	14.0684	6.2427	3.3935	5.6686	5.8585
	25	22.8300	10.6935	5.2689	9.1042	9.6564
0.50	5	2.5450	1.6893	1.1688	1.5036	1.5874
	10	4.9849	2.8858	1.6079	2.3830	2.4979
	15	8.4155	4.2560	2.3768	3.7448	4.1170
	20	14.4793	6.2687	3.6147	6.0404	6.3423
	25	26.4566	10.5100	5.6806	9.9734	10.2143
0.90	5	2.6502	1.7506	1.1509	1.5047	1.5768
	10	5.0335	2.9900	1.6019	2.4123	2.5467
	15	8.3611	4.3084	2.2035	3.6471	3.7998
	20	15.5653	6.2167	3.5210	6.0747	6.4696
	25	37.2105	12.2344	7.7071	13.2181	14.3314

Table 3.24: Simulated relative efficiency of the restricted, positive James-Stein and penalty estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 64$, $p - q = 4$ and $\Delta^2 = 0$ for different values of ρ and q based on the SAR model

ρ	q	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{JS+}$	$\hat{\boldsymbol{\beta}}_{SAR}^{Lasso}$	$\hat{\boldsymbol{\beta}}_{SAR}^{A.Lasso}$	$\hat{\boldsymbol{\beta}}_{SAR}^{SCAD}$
-0.95	5	2.5967	1.7408	1.0772	1.4350	1.5061
	10	4.5948	2.9071	1.4447	2.1362	2.2062
	15	7.0526	4.2472	1.8764	3.0749	3.1289
	20	10.7990	5.5033	2.4188	4.2171	4.2667
	25	16.1002	7.1381	3.1748	5.6535	5.6840
-0.50	5	2.5820	1.7027	1.1417	1.5166	1.6063
	10	4.4099	2.7699	1.4996	2.2228	2.3738
	15	7.1507	4.0023	2.0028	3.2589	3.5616
	20	10.5353	5.3575	2.6682	4.7114	5.0499
	25	15.7094	7.1851	3.4982	6.5778	7.1995
0.00	5	2.4686	1.7025	1.1405	1.5403	1.6501
	10	4.4121	2.7949	1.5177	2.2674	2.4649
	15	6.9735	3.9630	2.0186	3.2686	3.7296
	20	10.3991	5.3797	2.6849	4.7350	5.3797
	25	15.4484	7.0098	3.6614	6.5008	7.4750
0.50	5	2.4746	1.7126	1.1309	1.5196	1.6207
	10	4.4069	2.8610	1.5070	2.2847	2.4922
	15	6.5812	4.0144	1.9357	3.2566	3.4348
	20	10.3217	5.4170	2.5610	4.5374	4.9922
	25	14.7141	7.0012	3.3408	6.1540	6.6966
0.90	5	2.5870	1.7556	1.1340	1.5400	1.6070
	10	4.4620	2.9350	1.4774	2.3120	2.4252
	15	7.0513	4.2196	1.9145	3.1973	3.4121
	20	10.4262	5.6425	2.4925	4.4396	5.1000
	25	15.0268	7.1556	3.3193	6.2766	7.0547

Table 3.25: Simulated relative efficiency of the restricted, positive James-Stein and penalty estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 100$, $p - q = 4$ and $\Delta^2 = 0$ for different values of ρ and q based on the SAR model

ρ	q	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{JS+}$	$\hat{\boldsymbol{\beta}}^{Lasso}_{SAR}$	$\hat{\boldsymbol{\beta}}^{A.Lasso}_{SAR}$	$\hat{\boldsymbol{\beta}}^{SCAD}_{SAR}$
-0.95	5	2.5197	1.7422	1.0302	1.4576	1.4735
	10	4.2305	2.8889	1.3364	2.1179	2.1763
	15	6.0886	4.0556	1.6254	2.7140	2.9463
	20	8.3070	5.3419	1.9756	3.4896	3.8036
	25	11.5286	6.8243	2.3902	4.5449	4.8465
-0.50	5	2.3917	1.6949	1.0716	1.5072	1.5555
	10	4.0523	2.7865	1.3773	2.1536	2.3292
	15	5.9626	3.9636	1.7102	2.9255	3.3385
	20	8.0740	4.9426	2.0705	3.6349	4.0654
	25	10.6174	6.1888	2.4797	4.6703	5.5258
0.00	5	2.4094	1.6815	1.0978	1.5326	1.6060
	10	3.9606	2.7516	1.3878	2.1697	2.3621
	15	5.9571	3.9193	1.6976	2.8780	3.3908
	20	7.9465	5.0605	2.0366	3.7062	4.2954
	25	10.1456	6.0192	2.4881	4.6022	5.4337
0.50	5	2.3837	1.6869	1.0798	1.5145	1.5308
	10	4.0483	2.7496	1.3480	2.1382	2.2856
	15	5.8140	3.9090	1.6754	2.8278	3.1710
	20	8.1139	5.1800	2.0277	3.7030	4.1931
	25	10.5099	6.2243	2.4103	4.5004	5.1981
0.90	5	2.4133	1.7118	1.0670	1.5159	1.5341
	10	4.0890	2.8350	1.3305	2.1380	2.2401
	15	5.8289	4.0125	1.6717	2.8759	3.0660
	20	8.1820	5.2785	2.0037	3.6655	4.1478
	25	10.8982	6.6536	2.4162	4.5709	5.2662

The following conclusions can be drawn:

1. For all the cases, the restricted estimator, $\hat{\boldsymbol{\beta}}^{R}$, outperforms all other estimators in this simulation study.

2. The performance of $\hat{\boldsymbol{\beta}}^{R}$ increases as the number of zero parameters (q) increases regardless of the values of ρ and n.

3. For a fixed ρ, $\hat{\boldsymbol{\beta}}^{R}$ performs much better when $n = 49$ than $n = 64$ or 100.

4. The value of ρ does not have a significant effect on the SRE results for a fixed n and q.

5. In general, the positive rule James-Stein estimator, $\hat{\boldsymbol{\beta}}^{JS+}$, dominates all penalty estimators, and its performance improves as q increases.

3.5.1 Application to Boston Housing Prices Data

Here we use the Boston housing prices data described in Section 2.8.2. The fitted model is considered as a full SAR model. The MLE of ρ is used in the variance covariance matrix of the SAR model to transform the response and the design matrix data that will be used for the penalty estimator algorithms.

Several selection methods were employed on these data to choose the submodel on which the restricted and the positive James-Stein estimators were based. We use forward selection, backward elimination, and the adaptive LASSO algorithm to select four different submodels. Both the forward and backward methods selected the same set of three predictors. Then a second model was selected using forward selection, a third model was selected by backward elimination, and finally a fourth model was obtained by adaptive LASSO. The full and submodels are reported in Table 3.26.

Table 3.26: Full and Submodels for the Boston Housing data based on SAR model

Selection Criterion	Model
Full	`log(CMEDV) ~ log(LSTAT) +I(RM^2) + TAX +B +CRIM + PTRATIO +log(RAD)+ CHAS + I(NOX^2) + log(DIS) + ZN+ INDUS + AGE + LAT+ LON`
Forward/Backward	`log(CMEDV) ~ log(LSTAT)+ I(RM^2)+ LON`
Forward	`log(CMEDV) ~ log(LSTAT)+ I(RM^2)+ LON+ CRIM`
Backward	`log(CMEDV) ~ log(LSTAT)+ I(RM^2)+ LON + TAX`
Adaptive LASSO	`log(CMEDV) ~ log(LSTAT)+ I(RM^2)+ TAX+ B + CRIM+ PTRATIO`

To compare the restricted and positive James-Stein with the penalty estimators, we use the relative mean squared prediction error (RMSPE) in (2.38) as an evaluation method. In each case, we select 1000 samples of size 506 each with replacement using Hall's (1985) method. Table 3.27 summarizes the results.

Table 3.27: RMSPE with respect to $\hat{\boldsymbol{\beta}}$ for Boston Housing data based on the SAR model

Model	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{JS+}$
Forward/Backword	1.0864	1.0741
Forward	1.0844	1.0713
Backward	1.0841	1.0702
Adaptive LASSO	1.0782	1.0629
	Penalty Estimators	
$\hat{\boldsymbol{\beta}}^{Lasso}$	$\hat{\boldsymbol{\beta}}^{A.Lasso}$	$\hat{\boldsymbol{\beta}}^{SCAD}$
0.9046	1.0409	1.0437

The following conclusions can be drawn from the above Table.

1. The restricted estimator, $\hat{\boldsymbol{\beta}}^R$, outperforms all estimators, which indicates that if the restriction is correct, then $\hat{\boldsymbol{\beta}}^R$ is optimum.

2. The positive rule James-Stein estimator dominates all penalty estimators in all suggested submodels.

3. Both SCAD and adaptive LASSO perform better than LASSO.

4. The LASSO estimator has less MSPE than the unrestricted estimator, but this is not true in general.

5. The variable `LON`, the longitude location of each tract, was the most important variable in explaining the logarithm of the corrected median values of the owner-occupied housing `CMEDV`. This indicates that the study made by Harrison and Rubinfeld (1978) that measured the demand for clean air was affected more by `LON` than the levels of nitrogen oxides concentration `NOX` variable, as this variable did not appear in any submodel selected using the three selection criteria applied to this data example. In addition, `LON` was the first one included in the forward selection, and the last one left in the backward elimination method.

3.6 Conclusion

In this chapter, we considered shrinkage and penalty estimation for a regression model with simultaneous autoregressive SAR error specification. All the estimators were compared numerically by using simulated and real data examples. The simulation results were similar to those in Chapter 2 for the CAR model.

We used the Boston housing prices data to study the performance of the $\hat{\boldsymbol{\beta}}^{R}$, $\hat{\boldsymbol{\beta}}^{JS+}$ and the penalty estimators. Four submodels were selected using forward selection, backward elimination, and the adaptive LASSO algorithm. In all the cases, $\hat{\boldsymbol{\beta}}^{R}$ outperformed all others, $\hat{\boldsymbol{\beta}}^{JS+}$ dominated all the penalty estimators in terms of relative mean squared prediction error.

Chapter 4

Spatial Moving Average Model

4.1 Introduction

In this chapter we consider another spatial regression model known as *spatial moving average* (SMA) model. This model was studied in Huang (1984) who analyzed spatial interaction by using a two-dimensional autoregressive moving average model. He extended the familiar ARMA(p, q) model to accommodate spatial autocorrelations and called it SARMA(p, q). Haining (1978) considered a two-dimensional moving average model to study spatial interactions, and derived the likelihood ratio statistic for the model to test the moving average parameters.

Mur (1999) studied the problem of testing the spatial autocorrelation parameters and clarified the difference between a spatial autoregressive and a spatial moving average, which in fact differ in their variance structure. Anselin and Florax (1995) used eight tests for testing the spatial dependence parameter using Monte Carlo experiments in regression models for both regular and irregular lattices based on small sample sizes. More information about SMA model can be found in Anselin (1988), Bailey and Gatrell (1995), Cliff and Ord (1981) and Cressie and Wikle (2011).

Following the layout of the previous two chapters, here we establish pretest and shrinkage estimators of the large-scale effects, $\boldsymbol{\beta}$, in SMA. Furthermore, we provide algorithms for computing penalty estimators for $\boldsymbol{\beta}$.

The chapter is organized as follows. In Section 4.2, we discuss the spatial moving average model, and describe the maximum likelihood estimator of the large-scale effects vector $\boldsymbol{\beta}$. Estimation strategies using the restricted, pretest, shrinkage and penalty estimators are presented in Section 4.3. Some asymptotic results of the restricted and unrestricted estimators are provided in Section 4.4. The asymptotic distributional bias, mean squared error matrix, and quadratic risk of the pretest and shrinkage estimators are derived in Section 4.5. Some analytical risk comparisons are made in Section 4.6. In Section 4.7, we use Monte Carlo experiments and a real data example to study the performance of the proposed pretest and shrinkage estimators. We develop a technique to obtain the penalty estimators in Section 4.8, and provide numerical comparison results of the relative performance of the restricted, positive James-Stein and penalty estimators through Monte Carlo experiments and a real data example.

4.2 The SMA Model and Preliminaries

Recall that in Chapter 1, the spatial moving average model, SMA, was defined as

$$\boldsymbol{Y}(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})\boldsymbol{\beta} + (\boldsymbol{I} + \boldsymbol{G}(\boldsymbol{s}))\boldsymbol{\epsilon}(\boldsymbol{s}), \tag{4.1}$$

where $\boldsymbol{Y}(\boldsymbol{s}) = (Y(s_1), \ldots, Y(s_n))'$ is the $(n \times 1)$ observed response vector at the lattice sites $\boldsymbol{s}$, $\boldsymbol{X}(\boldsymbol{s})$ the $(n \times p)$ fixed matrix of p explanatory variables, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown regression parameters and $\boldsymbol{G} = \{g_{ij}\}_{i,j=1}^{n}$, is a matrix of spatial dependence parameters, with $g_{ii} = 0$, $\boldsymbol{\epsilon}(\boldsymbol{s}) = (\epsilon(s_1), \ldots, \epsilon(s_n))'$ and $\boldsymbol{I}$ is an $(n \times n)$ identity matrix.

If the spatial dependence matrix $\boldsymbol{G}$ is completely unstructured, the model becomes computationally cumbersome. Fortunately, nature is often sparse, a fact that can be exploited in the context of SMA by choosing the spatial dependence matrix $\boldsymbol{G}$ as $\rho\boldsymbol{W}$, where $\boldsymbol{W}$ is a sparse and known proximity matrix as before.

Dropping the lattice symbol $\boldsymbol{s}$, and assuming that the error vector term $\boldsymbol{\epsilon}$ follows a multivariate Gaussian distribution with mean $\mathbf{0}$, and variance covariance matrix $\sigma^2\boldsymbol{I}$, then the response vector $\boldsymbol{Y}$ will be distributed as (Cressie and Wikle, 2011):

$$\boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{X\beta}, \boldsymbol{\sigma^2}(\boldsymbol{I} + \boldsymbol{\rho W})(\boldsymbol{I} + \boldsymbol{\rho W'})). \tag{4.2}$$

4.2.1 Unrestricted Maximum Likelihood Estimation

Following Huang (1984), the log-likelihood of the SMA model parameters is written as

$$\begin{aligned} l &= \log(L(\boldsymbol{\beta}, \sigma^2, \rho)) \\ &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \log|\boldsymbol{I} + \boldsymbol{\rho W}| \\ &- \frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X\beta})'\left((\boldsymbol{I} + \boldsymbol{\rho W})(\boldsymbol{I} + \boldsymbol{\rho W'})\right)^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta}). \end{aligned} \tag{4.3}$$

Since $\boldsymbol{W}$ is symmetric, there exists an orthogonal matrix, say $\boldsymbol{T}$, such that $\boldsymbol{W} = \boldsymbol{T'\Lambda T}$, where $\boldsymbol{\Lambda} = \text{diag}\{\lambda_i\}_{i=1}^n$, and $\{\lambda_i : i = 1, \ldots, n\}$ are the eigenvalues with corresponding eigenvectors $\boldsymbol{T}$. Accordingly, transforming the responses, $\boldsymbol{Y}$ as well as the covariate matrix, $\boldsymbol{X}$ in the SMA model, the log-likelihood function in (4.3) becomes

$$\begin{aligned} l &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n}\log(1+\rho\lambda_i) \\ &- \frac{1}{2\sigma^2}\boldsymbol{(Z - X_0\beta)'V_n^{-1}(Z - X_0\beta)}, \end{aligned} \tag{4.4}$$

where $\boldsymbol{Z = TY}$, $\boldsymbol{X_0 = TX}$, and $var(\boldsymbol{Z}) = \sigma^2\boldsymbol{V_n} = \sigma^2(\boldsymbol{I} + \rho\boldsymbol{\Lambda})(\boldsymbol{I} + \rho\boldsymbol{\Lambda'})$.

By maximizing this log-likelihood function with respect to $\boldsymbol{\beta}, \sigma^2$, we obtain the following estimators of $\boldsymbol{\beta}$ and σ^2, respectively

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\rho) &= \left(\boldsymbol{X_0'V_n}^{-1}(\rho)\boldsymbol{X_0}\right)^{-1}\boldsymbol{X_0'V_n}^{-1}(\rho)\boldsymbol{Z} \\ \hat{\sigma^2}(\rho) &= \frac{(\boldsymbol{Z} - \boldsymbol{X_0}\hat{\boldsymbol{\beta}}(\rho))'\boldsymbol{V_n}^{-1}(\rho)(\boldsymbol{Z} - \boldsymbol{X_0}\hat{\boldsymbol{\beta}}(\rho))}{n}. \end{aligned} \tag{4.5}$$

Now, plugging these expressions back into the the log-likelihood function, we get the profile log-likelihood, a function of ρ only,

$$\begin{aligned} l^*(\rho) &\approx -\frac{n}{2}\log\left\{(\boldsymbol{Z} - \boldsymbol{X_0}\hat{\boldsymbol{\beta}}(\rho))'\boldsymbol{V_n}^{-1}(\boldsymbol{Z} - \boldsymbol{X_0}\hat{\boldsymbol{\beta}}(\rho))\right\} \\ &- \sum_{i=1}^{n}\log(1+\rho\lambda_i), \end{aligned}$$

which can be maximized to obtain $\hat{\rho}$, the (MLE) of ρ. Finally, plugging $\hat{\rho}$ in (4.5), we obtain the final MLEs $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma^2}$ of $\boldsymbol{\beta}, \sigma^2$, respectively. In the sequel, these estimators will be called the unrestricted MLEs.

The consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}$ follow from results in Huang (1984). Here we summarize such results in the following theorem.

Theorem 4.2.1. (Huang (1984)) As $n \to \infty$, and under the assumption

$$\frac{\left(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0}\right)}{n} \xrightarrow{P} \boldsymbol{V}_0, \tag{4.6}$$

where $\hat{\boldsymbol{V_n}} = (\boldsymbol{I} + \hat{\rho}\boldsymbol{W})(\boldsymbol{I} + \hat{\rho}\boldsymbol{W'})$, and $\boldsymbol{V}_0$ is a $p \times p$ finite and positive definite matrix, we have:

$$\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}, \text{ and } \sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{D} N\left(\boldsymbol{0}, \sigma^2\boldsymbol{V_0^{-1}}\right).$$

4.3 Improved Estimation Strategies

In the following subsections, we present the restricted, pretest, and positive shrinkage estimators for the SMA model's large-scale effect parameters.

4.3.1 Restricted Estimator

Here, we are interested in estimating the large-scale effect parameter $\boldsymbol{\beta}$ in the SMA model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} + \rho\boldsymbol{W})\boldsymbol{\epsilon}, \tag{4.7}$$

subject to the restriction given in a form of null hypothesis by:

$$A_0 : \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}, \tag{4.8}$$

where $\boldsymbol{X}$ is a $n \times p$ fixed design matrix of rank p, $\boldsymbol{H}$ is a $p \times q$ known matrix of rank $(q \leq p)$, and $\boldsymbol{h}$ is a $q \times 1$ vector of known constants.

By using Lagrange multipliers, it is easy to show that the estimator $\hat{\boldsymbol{\beta}}^R(\rho)$ of $\boldsymbol{\beta}$ under the above null hypothesis is given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}^R(\rho) &= \hat{\boldsymbol{\beta}}(\rho) - (\boldsymbol{X_0'}\boldsymbol{V_n}^{-1}(\rho)\boldsymbol{X_0})^{-1}\boldsymbol{H}' \\ &\quad (\boldsymbol{H}(\boldsymbol{X_0'}\boldsymbol{V_n}^{-1}(\rho)\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}(\rho) - h), \end{aligned} \tag{4.9}$$

while the plug-in version of it, which we will call the restricted MLE, is given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}^R &= \hat{\boldsymbol{\beta}}(\hat{\rho}) - (\boldsymbol{X_0'}\boldsymbol{V_n}^{-1}(\hat{\rho})\boldsymbol{X_0})^{-1}\boldsymbol{H}' \\ &\quad (\boldsymbol{H}(\boldsymbol{X_0'}\boldsymbol{V_n}^{-1}(\hat{\rho})\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}(\hat{\rho}) - h). \end{aligned} \tag{4.10}$$

The restricted estimator is a biased estimator of $\boldsymbol{\beta}$ unless the restriction given by (4.8) is correct.

4.3.2 Pretest Estimator

The pretest estimator is obtained by combining the unrestricted, $\hat{\boldsymbol{\beta}}$, and the restricted estimator, $\hat{\boldsymbol{\beta}}^R$, as follows:

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{PT} &= \hat{\boldsymbol{\beta}} I(\Phi_n > \Phi_{n,\alpha}) + \hat{\boldsymbol{\beta}}^R I(\Phi_n \leq \Phi_{n,\alpha}) \\ &= \hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R) I(\Phi_n \leq \Phi_{n,\alpha}), \end{aligned} \tag{4.11}$$

where

$$\Phi_n = \frac{(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H'})^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})}{s_e^2},$$

and

$$s_e^2 = \frac{(\boldsymbol{Z} - \boldsymbol{X_0}\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{V_n}}^{-1}(\boldsymbol{Z} - \boldsymbol{X_0}\hat{\boldsymbol{\beta}})}{n-p},$$

is a consistent estimator of σ^2. The quantity $\Phi_{n,\alpha}$ is the $\alpha-$level critical value of the exact distribution of the test statistic Φ_n, and $I(.)$ is an indicator function.

The statistic Φ_n given above follows asymptotically a central chi-square distribution as $n \to \infty$ with $q-$degrees of freedom under the null hypothesis $\boldsymbol{A_0} : \boldsymbol{H\beta} = \boldsymbol{h}$

It is clear that if the null hypothesis is rejected at level α, then the pretest estimator will be $\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}}^{R}$ otherwise. That is, $\hat{\boldsymbol{\beta}}^{PT}$ is a discrete function of the unrestricted and restricted estimators, and when it picks $\hat{\boldsymbol{\beta}}^{R}$, we may fall in a type II error.

As was explained in earlier chapters, it may be better in many cases to consider a smooth function of the test statistic Φ_n as opposed to the binary choice function. Answers to this desire are the James-Stein, Stein (1956), and the positive James-Stein estimators, which will be defined in the next subsection.

4.3.3 Shrinkage Estimators

Following Ahmed (2001) the shrinkage estimator of $\boldsymbol{\beta}$ is given by

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{JS} &= \hat{\boldsymbol{\beta}}^{R} + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})\{1 - (q-2)\Phi_n^{-1}\} \\ &= \hat{\boldsymbol{\beta}} - (q-2)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})\Phi_n^{-1}, \quad q \geq 3. \end{aligned} \tag{4.12}$$

On the other hand, to avoid over shrinkage phenomena, (see Chapter 2, p.32), which could happen in the James-Stein estimator whenever $(q-2)\Phi_n^{-1} > 1$, we define the positive shrinkage estimator as follows:

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{JS+} &= \hat{\boldsymbol{\beta}}^{R} + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})\{1 - (q-2)\Phi_n^{-1}\}^{+} \\ &= \hat{\boldsymbol{\beta}}^{JS} - (1 - (q-2)\Phi_n^{-1})I(\Phi_n < (q-2))(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R}), \end{aligned} \tag{4.13}$$

where $a^{+} = \max\{0, a\}$.

In the following, we estimate the regression parameter vector $\boldsymbol{\beta}$ for the spatial moving average SMA model using three penalty estimators. Further computational details will be given in the numerical studies section at the end of the chapter.

4.3.4 Penalty Estimators

By dropping the subscript n from the variance matrix $\boldsymbol{V_n}$, the log-likelihood function for the transformed SMA model is

$$\begin{aligned} l = \log(L(\boldsymbol{\beta}, \sigma^2, \rho)) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n}\log(1+\rho\lambda_i) \\ &- \frac{1}{2\sigma^2}(\boldsymbol{Z} - \boldsymbol{X_0\beta})'\boldsymbol{V^{-1}}(\boldsymbol{Z} - \boldsymbol{X_0\beta}). \end{aligned} \quad (4.14)$$

As our interest has been in the estimation of large-scale effects, neither σ^2 nor ρ will be subjected to any penalty. Therefore, we consider the following objective function:

$$F(\lambda, \boldsymbol{\beta}) = (\boldsymbol{Z} - \boldsymbol{X_0\beta})'\boldsymbol{V^{-1}}(\boldsymbol{Z} - \boldsymbol{X_0\beta}) + f(\lambda, \boldsymbol{\beta}),$$

where λ is a tuning (regularization) parameter. However, in order for us to use the existing computational algorithms, such as LARS, for computing the penalty estimators, we need to bring the SMA model errors into an *iid* setup. Therefore, we employ a transformation of the form $\boldsymbol{Y^*} = \boldsymbol{(I + \rho\Lambda)^{-1}Z}$, $\boldsymbol{X^*} = \boldsymbol{(I + \rho\Lambda)^{-1}X_0}$. The log-likelihood function of the model in (4.14) now becomes

$$\begin{aligned} l = \log(L\left(\boldsymbol{\beta}, \sigma^2, \rho\right)) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2} \\ &\quad \boldsymbol{(Y^* - X^*\beta)'(Y^* - X^*\beta)}. \end{aligned} \quad (4.15)$$

Now the objective function for computing the penalty estimators is

$$F(\boldsymbol{\beta}) = \boldsymbol{(Y^* - X^*\beta)'(Y^* - X^*\beta)} + f(\lambda, \boldsymbol{\beta}).$$

In the following, we re-iterate the definitions of the LASSO, SCAD, and adaptive LASSO estimators of the parameter vector $\boldsymbol{\beta}$.

LASSO

The LASSO estimator of $\boldsymbol{\beta}$ for the SMA model, denoted by $\hat{\boldsymbol{\beta}}^{Lasso}$, is defined as

$$\arg\min_{\boldsymbol{\beta}} \Big[(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})'(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j| \Big],$$

where $\lambda \geq 0$ is a tuning parameter to be estimated. $\hat{\boldsymbol{\beta}}^{Lasso}$ can be computed using the LARS algorithm of Efron et al. (2004).

SCAD

For the SMA model, the SCAD estimator of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}^{SCAD}$, is defined as

$$\arg\min_{\boldsymbol{\beta}} \Big[(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})'(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}) + n \sum_{j=1}^{p} P_{\lambda_s}(|\beta_j|) \Big],$$

where λ_s is the regularization parameter for the SCAD penalty function $P_{\lambda_s}(|.|)$.

Adaptive LASSO

For the SMA model, the Adaptive LASSO estimator, denoted by $\hat{\boldsymbol{\beta}}^{A.Lasso}$, is defined as

$$\arg\min_{\boldsymbol{\beta}} \Big[(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})'(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}) + n \sum_{j=1}^{p} \lambda_j |\beta_j| \Big],$$

where $\{\lambda_j : j = 1, \ldots, p\}$ are the coefficient specific regularization parameters.

4.4 Asymptotic Results

In this section we study the asymptotic behavior of the estimators proposed earlier, namely $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^R, \hat{\boldsymbol{\beta}}^{JS}, \hat{\boldsymbol{\beta}}^{JS+}$. Specifically, we will prove that the restricted and unrestricted, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^R$ are jointly asymptotically normal. We define and derive expressions for the asymptotic distributional bias (ADB), the asymptotic mean squared error matrix (AMSEM), and the asymptotic quadratic risk (AQR) of the estimators $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^R, \hat{\boldsymbol{\beta}}^{JS}, \hat{\boldsymbol{\beta}}^{JS+}$ by using the joint normality of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^R$.

As explained in Chapter 2, the AQR is a measure of the risk of estimators based on quadratic loss function and it can be used to compare the relative performance of the various estimators proposed.

4.4.1 Joint Asymptotic Normality of Unrestricted and Restricted Estimators

To study the asymptotic properties of the shrinkage estimators we need the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^R$, since all other estimators are represented as functions of these two. To this end, let us define a sequence of local alternatives,

$$\boldsymbol{A}_{(n)}\colon \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h} + \frac{\boldsymbol{\xi}}{\sqrt{n}}, \tag{4.16}$$

where $\boldsymbol{\xi}$ is a $q \times 1$ fixed vector in $\mathbb{R}^q$. If we set $\boldsymbol{\xi} = \mathbf{0}$, then the local alternative becomes $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$, which is the linear hypothesis representing the candidate subspace. The main result of this subsection is the following theorem.

Theorem 4.4.1. Under the assumptions of Theorem 4.2.1, and local alternatives given by (4.16), we have

(i) $\boldsymbol{I}_n^{(1)} = \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{I}^{(1)} \sim N_p(\boldsymbol{0}, \sigma^2 \boldsymbol{V}_0^{-1})$

(ii) $\boldsymbol{I}_n^{(2)} = \sqrt{n}(\hat{\boldsymbol{\beta}}^R - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{I}^{(2)} \sim N_p(-\boldsymbol{\nu}, \sigma^2 \boldsymbol{E_0})$,

(iii) $\boldsymbol{I}_n^{(3)} = \sqrt{n}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R) \xrightarrow{D} \boldsymbol{I}^{(3)} \sim N_p\left(\boldsymbol{\nu}, \sigma^2(\boldsymbol{V}_0^{-1} - \boldsymbol{E_0})\right)$

(iv) $\begin{pmatrix} \boldsymbol{I}_n^{(1)} \\ \boldsymbol{I}_n^{(3)} \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \boldsymbol{I}^{(1)} \\ \boldsymbol{I}^{(3)} \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{\nu} \end{pmatrix}, \sigma^2 \begin{pmatrix} \boldsymbol{V}_0^{-1} & \boldsymbol{V}_0^{-1} - \boldsymbol{E_0} \\ \boldsymbol{V}_0^{-1} - \boldsymbol{E_0} & \boldsymbol{V}_0^{-1} - \boldsymbol{E_0} \end{pmatrix}\right)$

(v) $\begin{pmatrix} \boldsymbol{I}_n^{(2)} \\ \boldsymbol{I}_n^{(3)} \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \boldsymbol{I}^{(2)} \\ \boldsymbol{I}^{(3)} \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} -\boldsymbol{\nu} \\ \boldsymbol{\nu} \end{pmatrix}, \sigma^2 \begin{pmatrix} \boldsymbol{E_0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_0^{-1} - \boldsymbol{E_0} \end{pmatrix}\right)$,

where

$\boldsymbol{E_0} = \boldsymbol{V}_0^{-1} - \boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}$, $\boldsymbol{\nu} = \boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}$.

Proof:

(i) The proof follows from Theorem 4.2.1.

(ii) Note that $\boldsymbol{I}_n^{(2)} = \sqrt{n}(\hat{\boldsymbol{\beta}}^R - \boldsymbol{\beta})$ can be written as follows:

$$\begin{aligned} \boldsymbol{I}_n^{(2)} &= \sqrt{n}\Big\{\hat{\boldsymbol{\beta}} - (\boldsymbol{X_0}'\hat{\boldsymbol{V}}_{\boldsymbol{n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0}'\hat{\boldsymbol{V}}_{\boldsymbol{n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1} \\ &\quad (\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}) - \boldsymbol{\beta}\Big\} \\ &= \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \Big\{(\boldsymbol{X_0}'\hat{\boldsymbol{V}}_{\boldsymbol{n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0}'\hat{\boldsymbol{V}}_{\boldsymbol{n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1} \\ &\quad \big[\boldsymbol{H}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h}\big]\Big\} \end{aligned}$$

$$\begin{aligned}
&= \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \sqrt{n}\Big\{(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1} \\
&\quad \boldsymbol{H}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\Big\} - (\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi} \\
&= \big[\boldsymbol{I_p} - (\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\big] \\
&\quad \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - (\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi},
\end{aligned}$$

which is a linear combination of $\boldsymbol{I}_n^{(1)}$. Therefore, as $n \longrightarrow \infty$,

$$\boldsymbol{I}_n^{(2)} \xrightarrow{D} \boldsymbol{I}^{(2)} \sim N_P(\boldsymbol{\mu_{(2)}}, \boldsymbol{\Sigma_{(2)}}),$$

where

$$\begin{aligned}
\boldsymbol{\mu_{(2)}} &= -\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi} \\
&= -\boldsymbol{\nu}, \\
\boldsymbol{\Sigma_{(2)}} &= \big[\boldsymbol{I_p} - \boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\big].\sigma^2\boldsymbol{V}_0^{-1} \\
&\quad \big[\boldsymbol{I_p} - \boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}\big] \\
&= \sigma^2\Big\{\boldsymbol{V}_0^{-1} - \boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}\Big\} \\
&= \sigma^2\boldsymbol{E_0}, \text{ with } \boldsymbol{E_0} = \boldsymbol{V}_0^{-1} - \boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}.
\end{aligned}$$

(iii) $\boldsymbol{I}_n^{(3)}$ can be rewritten as

$$\begin{aligned}
\boldsymbol{I}_n^{(3)} &= \sqrt{n}\Big\{(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})\Big\} \\
&= \sqrt{n}\Big\{(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}\big[\boldsymbol{H}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&\quad +\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h}\big]\Big\} \\
&= (\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&+ (\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi},
\end{aligned}$$

which is also a linear combination of $\boldsymbol{I}_n^{(1)}$. So, as $n \longrightarrow \infty$,

$$\boldsymbol{I}_n^{(3)} \xrightarrow{D} \boldsymbol{I}^{(3)} \sim N_p(\boldsymbol{\mu_{(3)}}, \boldsymbol{\Sigma_{(3)}}),$$

where

$$\begin{aligned}
\boldsymbol{\mu_{(3)}} &= \boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi} \\
&= \boldsymbol{\nu}, \\
\boldsymbol{\Sigma_{(3)}} &= \left[\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\right].\sigma^2\boldsymbol{V}_0^{-1}\left[\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}\right] \\
&= \sigma^2\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1} \\
&= \sigma^2\left(\boldsymbol{V}_0^{-1} - \boldsymbol{E_0}\right).
\end{aligned}$$

(iv) By using (iii),

$$\begin{aligned}
\begin{pmatrix} \boldsymbol{I}_n^{(1)} \\ \boldsymbol{I}_n^{(3)} \end{pmatrix} &= \begin{pmatrix} \boldsymbol{I_p} \\ \boldsymbol{F_{n(1)}} \end{pmatrix} \boldsymbol{I}_n^{(1)} + \begin{pmatrix} \boldsymbol{0_p} \\ \boldsymbol{G_{n(1)}} \end{pmatrix} \\
&= \boldsymbol{Q_{n(1)}}\boldsymbol{I}_n^{(1)} + \boldsymbol{U_{n(1)}},
\end{aligned}$$

where $\boldsymbol{I_p}$ is a $p \times p$ identity matrix, $\boldsymbol{0_p}$ is a $p \times 1$ vector of zeros,

$$\begin{aligned}
\boldsymbol{F_{n(1)}} &= (\boldsymbol{X_0'}\hat{\boldsymbol{V}}_{\boldsymbol{n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V}}_{\boldsymbol{n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}, \\
\boldsymbol{G_{n(1)}} &= (\boldsymbol{X_0'}\hat{\boldsymbol{V}}_{\boldsymbol{n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V}}_{\boldsymbol{n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}.
\end{aligned}$$

Therefore, as $n \longrightarrow \infty$, we have

$$\boldsymbol{F_{n(1)}} \xrightarrow{P} \boldsymbol{F_0} = \boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}$$
$$\boldsymbol{G_{n(1)}} \xrightarrow{P} \boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi} = \boldsymbol{\nu}.$$

Hence,

$$\begin{pmatrix} \boldsymbol{I}_n^{(1)} \\ \boldsymbol{I}_n^{(3)} \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \boldsymbol{I}^{(1)} \\ \boldsymbol{I}^{(3)} \end{pmatrix} \sim N_{2p}\left(\boldsymbol{\mu_{(4)}}, \boldsymbol{\Sigma_{(4)}}\right), \quad \text{where}$$

$$\begin{aligned}
\boldsymbol{\mu_{(4)}} &= \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{\nu} \end{pmatrix}, \\
\boldsymbol{\Sigma_{(4)}} &= \begin{pmatrix} \boldsymbol{I_p} \\ \boldsymbol{F_0} \end{pmatrix} \sigma^2 \boldsymbol{V}_0^{-1} \begin{pmatrix} \boldsymbol{I_p} & \boldsymbol{F_0'} \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} \boldsymbol{V}_0^{-1} & \boldsymbol{V}_0^{-1}\boldsymbol{F_0'} \\ \boldsymbol{F_0}\boldsymbol{V}_0^{-1} & \boldsymbol{F_0}\boldsymbol{V}_0^{-1}\boldsymbol{F_0'} \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} \boldsymbol{V}_0^{-1} & \boldsymbol{V}_0^{-1} - \boldsymbol{E_0} \\ \boldsymbol{V}_0^{-1} - \boldsymbol{E_0} & \boldsymbol{V}_0^{-1} - \boldsymbol{E_0} \end{pmatrix}.
\end{aligned}$$

(v) Similarly, $\boldsymbol{I}_n^{(2)}$ and $\boldsymbol{I}_n^{(3)}$ can be rewritten in terms of $\boldsymbol{I}_n^{(1)}$, $\boldsymbol{F_{n(1)}}$, and $\boldsymbol{G_{n(1)}}$ as follows

$$\begin{aligned}
\boldsymbol{I}_n^{(2)} &= \left(\boldsymbol{I_p} - \boldsymbol{F_{n(1)}}\right)\boldsymbol{I}_n^{(1)} - \boldsymbol{G_{n(1)}} \\
\boldsymbol{I}_n^{(3)} &= \boldsymbol{F_{n(1)}}\boldsymbol{I}_n^{(1)} + \boldsymbol{G_{n(1)}}.
\end{aligned}$$

Alternatively,

$$\begin{pmatrix} \boldsymbol{I}_n^{(2)} \\ \boldsymbol{I}_n^{(3)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I_p} - \boldsymbol{F_{n(1)}} \\ \boldsymbol{F_{n(1)}} \end{pmatrix} \boldsymbol{I}_n^{(1)} + \begin{pmatrix} -\boldsymbol{I_p} \\ \boldsymbol{I_p} \end{pmatrix} \boldsymbol{G_{n(1)}}$$
$$= \boldsymbol{Q_{n(2)}} \boldsymbol{I}_n^{(1)} + \boldsymbol{U_{(2)}} \boldsymbol{G_{n(1)}},$$

$$\boldsymbol{Q_{n(2)}} = \begin{pmatrix} \boldsymbol{I_p} - \boldsymbol{F_{n(1)}} \\ \boldsymbol{F_{n(1)}} \end{pmatrix}, \boldsymbol{U_{(2)}} = \begin{pmatrix} -\boldsymbol{I_p} \\ \boldsymbol{I_p} \end{pmatrix}.$$

Therefore, as $n \longrightarrow \infty$, $\boldsymbol{F_{n(1)}} \xrightarrow{P} \boldsymbol{F_0}$, and $\boldsymbol{G_{n(1)}} \xrightarrow{P} \boldsymbol{\nu}$. So

$$\begin{pmatrix} \boldsymbol{I}_n^{(2)} \\ \boldsymbol{I}_n^{(3)} \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \boldsymbol{I}^{(2)} \\ \boldsymbol{I}^{(3)} \end{pmatrix} \sim N_{2p}\left(\boldsymbol{\mu_{(5)}}, \boldsymbol{\Sigma_{(5)}}\right),$$

where

$$\boldsymbol{\mu_{(5)}} = \begin{pmatrix} -\boldsymbol{\nu} \\ \boldsymbol{\nu} \end{pmatrix},$$
$$\boldsymbol{\Sigma_{(5)}} = \begin{pmatrix} \boldsymbol{I_p} - \boldsymbol{F_0} \\ \boldsymbol{F_0} \end{pmatrix} \sigma^2 \boldsymbol{V}_0^{-1} \begin{pmatrix} \boldsymbol{I_p} - \boldsymbol{F_0'} & \boldsymbol{F_0'} \end{pmatrix}$$
$$= \sigma^2 \begin{pmatrix} \boldsymbol{E_0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_0^{-1} - \boldsymbol{E_0} \end{pmatrix}.$$

□

4.5 Asymptotic Analysis of Bias and Risk

In the following subsections, the performance of the restricted, pretest, James-Stein and positive James-Stein estimators will be examined asymptotically under local alternatives given by (4.16).

4.5.1 Asymptotic Distributional Bias (ADB)

The concept of asymptotic distributional bias is defined in Chapter 2. The ADBs are convenient tools for deriving comparative picture of the listed estimators in terms of their estimation biases. For the purpose of completeness, we recall here the definition of the ADB as follows. Let $G(\boldsymbol{x})$ be the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})$, where $\hat{\boldsymbol{\beta}}^*$ is any of the listed estimators, and

$$G(\boldsymbol{x}) = \lim_{n\to\infty} P_{A_{(n)}}\big\{\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) \le \boldsymbol{x}\big\}.$$

If and when the limit exists. If $\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{I^*}$, then the asymptotic distributional bias of $\hat{\boldsymbol{\beta}}^*$ is defined by

$$\begin{aligned} ADB(\hat{\boldsymbol{\beta}}^*) &= E\big\{\boldsymbol{I^*}\big\} \\ &= \int \boldsymbol{x d G(x)}. \end{aligned} \tag{4.17}$$

The following theorem gives us expressions for the ADBs of $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^R, \hat{\boldsymbol{\beta}}^{JS}, \hat{\boldsymbol{\beta}}^{JS+}$.

Theorem 4.5.1. Under local alternatives given in (4.16), we have

(i) $ADB(\hat{\boldsymbol{\beta}}) = \mathbf{0}$.

(ii) $ADB(\hat{\boldsymbol{\beta}}^{R}) = -\boldsymbol{\nu}$.

(iii) $ADB(\hat{\boldsymbol{\beta}}^{PT}) = -\boldsymbol{\nu} H_{q+2}(\chi^2_q(\alpha); \Delta^2)$.

(iv) $ADB(\hat{\boldsymbol{\beta}}^{JS}) = -(q-2)\boldsymbol{\nu} E(\chi^{-2}_{q+2}(\Delta^2))$.

(v) $$\begin{aligned} ADB(\hat{\boldsymbol{\beta}}^{JS+}) &= ADB(\hat{\boldsymbol{\beta}}^{JS}) - \boldsymbol{\nu} E\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))I(\chi^2_{q+2}(\Delta^2) < (q-2))\} \\ &= -\boldsymbol{\nu}\Big[(q-2)E(\chi^{-2}_{q+2}(\Delta^2)) + E\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))I(\chi^2_{q+2}(\Delta^2) < (q-2))\}\Big], \end{aligned}$$

where $\Delta^2 = \frac{1}{\sigma^2}\boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi} = \frac{1}{\sigma^2}\boldsymbol{\nu}'\boldsymbol{V_0}\boldsymbol{\nu}$.

Proof:

(i) By Theorem 4.4.1(i), we have

$$ADB(\hat{\boldsymbol{\beta}}) = E\{\boldsymbol{I}^{(1)}\} = 0.$$

(ii) Also, by Theorem 4.4.1(ii), we have

$$ADB(\hat{\boldsymbol{\beta}}^{R}) = E\{\boldsymbol{I}^{(2)}\} = -\boldsymbol{\nu}.$$

(iii) Note that,

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta}) &= \sqrt{n}\Big((\hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})I(\Phi_n \le \Phi_{n,\alpha})) - \boldsymbol{\beta}\Big) \\ &= \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \sqrt{n}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{R})I(\Phi_n \le \Phi_{n,\alpha}) \\ &= \boldsymbol{I}_n^{(1)} - \sqrt{n}(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0'}\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1} \\ &\quad (\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})I(\Phi_n \le \Phi_{n,\alpha})). \end{aligned}$$

Now, as $n \longrightarrow \infty$, and by Slutsky's Theorem, we get $\Phi_n \xrightarrow{D} \Phi \sim \chi^2_q$, and $\Phi_{n,\alpha} \xrightarrow{D} \chi^2_{q;\alpha}$, where $\chi^2_{q;\alpha}$ is the upper $\alpha-$quantile of χ^2_q random variable, and furthermore,

$$\begin{aligned}\sqrt{n}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h}) &= \sqrt{n}\left(\boldsymbol{H}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}+\boldsymbol{\beta})-\boldsymbol{h}\right)\\ &= \sqrt{n}\boldsymbol{H}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})+\boldsymbol{\xi},\ \text{under local alternatives,}\\ &= \boldsymbol{H}\boldsymbol{I}_n^{(1)}+\boldsymbol{\xi}.\end{aligned}$$

Therefore,

$$\sqrt{n}(\boldsymbol{H}\hat{\boldsymbol{\beta}}-\boldsymbol{h}) \xrightarrow{D} N_q\left(\boldsymbol{\xi},\sigma^2(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\right).$$

Therefore, by using Theorem 2.5.2, we have

$$\begin{aligned}ADB(\hat{\boldsymbol{\beta}}^{PT}) &= -\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}H_{q+2}(\chi_q^2(\alpha);\Delta^2)\\ &= -\boldsymbol{\nu}H_{q+2}(\chi_q^2(\alpha);\Delta^2).\end{aligned}$$

(iv) Note that,

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta}) &= \sqrt{n}\left(\hat{\boldsymbol{\beta}}-(q-2)(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^R)\Phi_n^{-1}-\boldsymbol{\beta}\right)\\ &= \sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})-(q-2)\sqrt{n}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^R)\Phi_n^{-1}\\ &= \boldsymbol{I}_n^{(1)}-(q-2)\boldsymbol{I}_n^{(3)}\Phi_n^{-1}\end{aligned}$$

Therefore, by using Theorem 2.5.2, we have

$$\begin{aligned}ADB(\hat{\boldsymbol{\beta}}^{JS}) &= 0-(q-2)\boldsymbol{\nu}E\left(\chi_{q+2}^{-2}(\Delta^2)\right)\\ &= -(q-2)\boldsymbol{\nu}E\left(\chi_{q+2}^{-2}(\Delta^2)\right).\end{aligned}$$

(v) Similar to (iv),

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}}^{JS+}-\boldsymbol{\beta}) &= \sqrt{n}\Big\{\hat{\boldsymbol{\beta}}^{JS}-(1-(q-2)\Phi_n^{-1})I(\Phi_n<(q-2))(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})-\boldsymbol{\beta}\Big\} \\
&= \sqrt{n}(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})-\sqrt{n}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})(1-(q-2)\Phi_n^{-1}) \\
&\quad I(\Phi_n<(q-2)) \\
&= \sqrt{n}(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})-\big\{\boldsymbol{I}_n^{(3)}(1-(q-2)\Phi_n^{-1})I(\Phi_n<(q-2))\big\}. \text{ So,} \\
ADB(\hat{\boldsymbol{\beta}}^{JS+}) &= ADB(\hat{\boldsymbol{\beta}}^{JS}) \\
&- \boldsymbol{\nu}E\Big\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2))I(\chi_{q+2}^{2}(\Delta^2)<(q-2))\Big\} \\
&= -(q-2)\boldsymbol{\nu}E\left(\chi_{q+2}^{-2}(\Delta^2)\right) \\
&- \boldsymbol{\nu}E\Big\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2))I(\chi_{q+2}^{2}(\Delta^2)<(q-2))\Big\} \\
&= -\boldsymbol{\nu}\Big\{(q-2)E\left(\chi_{q+2}^{-2}(\Delta^2)\right) \\
&+ E\big\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2))I(\chi_{q+2}^{2}(\Delta^2)<(q-2))\big\}\Big\}.
\end{aligned}$$

□

4.5.2 Risk Analysis

Following the AMSEM and AQR concepts introduced in Chapter 2, we have for any estimator $\hat{\boldsymbol{\beta}}^*$, if $\sqrt{n}\left(\hat{\boldsymbol{\beta}}^*-\boldsymbol{\beta}\right)\xrightarrow{D}\boldsymbol{I}^*$, then the $AMSEM(\hat{\boldsymbol{\beta}}^*)$ is defined as:

$$\begin{aligned}
AMSEM(\hat{\boldsymbol{\beta}}^*) &= E\big\{\boldsymbol{I}^*\boldsymbol{I}^{*\prime}\big\} \\
&= \int \boldsymbol{x}\boldsymbol{x}'d\boldsymbol{G}(\boldsymbol{x}),
\end{aligned}$$

and the asymptotic quadratic risk (AQR) of $\hat{\boldsymbol{\beta}}^*$ is defined as:

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^*, \boldsymbol{M}) &= E\big\{\boldsymbol{I^{*\prime}MI^*}\big\} \\ &= \int (\boldsymbol{x'Mx})\, \boldsymbol{dG(x)} \\ &= tr\big\{\boldsymbol{M}AMSEM(\hat{\boldsymbol{\beta}}^*)\big\}, \end{aligned}$$

where $\boldsymbol{M}$ is a $p \times p$ positive definite matrix, and $tr(\boldsymbol{M})$ is the trace of the matrix $\boldsymbol{M}$. The AMSEM and AQR expressions for $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^R, \hat{\boldsymbol{\beta}}^{JS}, \hat{\boldsymbol{\beta}}^{JS+}$ are driven in the following theorem.

Theorem 4.5.2. Suppose that $\boldsymbol{M}$ is a $p \times p$ positive definite matrix, then under the assumption (4.6) and local alternatives (4.16), we have

(i) $AMSEM(\hat{\boldsymbol{\beta}}) = \sigma^2 \boldsymbol{V}_0^{-1}$,

$AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) = \sigma^2 tr(\boldsymbol{M}\boldsymbol{V}_0^{-1})$,

(ii) $AMSEM(\hat{\boldsymbol{\beta}}^R) = \sigma^2 \boldsymbol{E_0} + \boldsymbol{\nu\nu'}$,

$AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) = \sigma^2 tr(\boldsymbol{M}\boldsymbol{V}_0^{-1}) - \sigma^2 tr(\boldsymbol{U_{11}}) + \boldsymbol{u_1' U_{11} u_1}$,

(iii)

$$\begin{aligned} AMSEM(\hat{\boldsymbol{\beta}}^{PT}) &= \sigma^2 \boldsymbol{V}_0^{-1} \\ &- \sigma^2 \boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1} H_{q+2}(\chi^2_q(\alpha); \Delta^2) \\ &+ \boldsymbol{\nu\nu'}\big\{2H_{q+2}(\chi^2_q(\alpha); \Delta^2) - H_{q+4}(\chi^2_q(\alpha); \Delta^2)\big\}, \\ AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M}) &= \sigma^2 tr(\boldsymbol{M}\boldsymbol{V}_0^{-1}) \\ &- \sigma^2 tr(\boldsymbol{U_{11}}) H_{q+2}(\chi^2_q(\alpha); \Delta^2) \\ &+ \boldsymbol{u_1' U_{11} u_1}\big\{2H_{q+2}(\chi^2_q(\alpha); \Delta^2) - H_{q+4}(\chi^2_q(\alpha); \Delta^2)\big\}, \end{aligned}$$

(iv)

$$\begin{aligned}
AMSEM(\hat{\boldsymbol{\beta}}^{JS}) &= \sigma^2 \boldsymbol{V}_0^{-1} \\
&- (q-2)\sigma^2 \left(\boldsymbol{V}_0^{-1} - \boldsymbol{E_0}\right) \\
&\quad \Big\{2E(\chi_{q+2}^{-2}(\Delta^2)) - (q-2)E(\chi_{q+2}^{-4}(\Delta^2))\Big\} \\
&+ (q-2)(q+2)\boldsymbol{\nu}\boldsymbol{\nu}' E(\chi_{q+4}^{-4}(\Delta^2)), \\
AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) &= \sigma^2 tr(\boldsymbol{M}\boldsymbol{V}_0^{-1}) \\
&- \sigma^2(q-2)\Big\{2E(\chi_{q+2}^{-2}(\Delta^2)) - (q-2)E(\chi_{q+2}^{-4}(\Delta^2))\Big\} tr(\boldsymbol{U_{11}}) \\
&+ (q-2)(q+2)E(\chi_{q+4}^{-4}(\Delta^2))\boldsymbol{u_1'}\boldsymbol{U_{11}}\boldsymbol{u_1},
\end{aligned}$$

(v)

$$\begin{aligned}
AMSEM(\hat{\boldsymbol{\beta}}^{JS+}) &= AMSEM(\hat{\boldsymbol{\beta}}^{JS}) \\
&- \sigma^2 \left(\boldsymbol{V}_0^{-1} - \boldsymbol{E_0}\right) \\
&\quad E\Big\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2))^2 I(\chi_{q+2}^{2}(\Delta^2) < (q-2))\Big\} \\
&- \boldsymbol{\nu}\boldsymbol{\nu}' E\Big\{(1-(q-2)\chi_{q+4}^{-2}(\Delta^2))^2 I(\chi_{q+4}^{2}(\Delta^2) < (q-2))\Big\} \\
&+ 2\boldsymbol{\nu}\boldsymbol{\nu}' E\Big\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2)) I(\chi_{q+2}^{2}(\Delta^2) < (q-2))\Big\},
\end{aligned}$$

$$\begin{aligned}
AQR(\hat{\boldsymbol{\beta}}^{JS+}, \boldsymbol{M}) &= AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) \\
&- \sigma^2 E\Big\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2))^2 I(\chi_{q+2}^{2}(\Delta^2) < (q-2))\Big\} \\
&\quad tr(\boldsymbol{U_{11}})
\end{aligned}$$

$$
\begin{aligned}
&- \quad E\Big\{(1-(q-2)\chi_{q+4}^{-2}(\Delta^2))^2 I(\chi_{q+4}^2(\Delta^2) < (q-2))\Big\} \\
&\qquad \boldsymbol{u}_1' \boldsymbol{U}_{11} \boldsymbol{u}_1 \\
&+ \quad 2E\Big\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2)) I(\chi_{q+2}^2(\Delta^2) < (q-2))\Big\} \\
&\qquad \boldsymbol{u}_1' \boldsymbol{U}_{11} \boldsymbol{u}_1,
\end{aligned}
$$

where $\boldsymbol{E_0}$ is defined as in Theorem 4.4.1,

$$tr(\boldsymbol{U_{11}}) = tr\left(\boldsymbol{M}\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}\right),$$

$$\boldsymbol{u} = \boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi},$$

and $\boldsymbol{\Gamma}$ is a $p \times p$ orthogonal matrix such that

$$\boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V_0}^{-1/2}\boldsymbol{\Gamma}' = \begin{pmatrix} \boldsymbol{I_q} & \boldsymbol{0_{q\times(p-q)}} \\ \boldsymbol{0_{(p-q)\times q}} & \boldsymbol{0_{(p-q)\times(p-q)}} \end{pmatrix}, \tag{4.18}$$

and

$$\boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{M}\boldsymbol{V_0}^{-1/2}\boldsymbol{\Gamma}' = \begin{pmatrix} \boldsymbol{U_{11}} & \boldsymbol{U_{12}} \\ \boldsymbol{U_{12}'} & \boldsymbol{U_{22}} \end{pmatrix}, \boldsymbol{u} = \begin{pmatrix} \boldsymbol{u_1} \\ \boldsymbol{u_2} \end{pmatrix}.$$

Proof:

(i) Note that,

$$\begin{aligned}
n(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})' &= \boldsymbol{I}_n^{(1)}\boldsymbol{I}_n^{(1)\prime}. \text{ Therefore,}\\
AMSEM(\hat{\boldsymbol{\beta}}) &= E\big\{\boldsymbol{I}^{(1)}\boldsymbol{I}^{(1)\prime}\big\}\\
&= \sigma^2\boldsymbol{V}_0^{-1}, \text{ by Theorem 4.4.1(i),}\\
AQR(\hat{\boldsymbol{\beta}},\boldsymbol{M}) &= tr\left(\boldsymbol{M}\{AMSEM(\hat{\boldsymbol{\beta}})\}\right)\\
&= tr(\boldsymbol{M}\sigma^2\boldsymbol{V}_0^{-1})\\
&= \sigma^2 tr(\boldsymbol{M}\boldsymbol{V}_0^{-1}).
\end{aligned}$$

(ii) Similarly,

$$\begin{aligned}
n(\hat{\boldsymbol{\beta}}^R-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^R-\boldsymbol{\beta})' &= \boldsymbol{I}_n^{(2)}\boldsymbol{I}_n^{(2)\prime}. \text{ Therefore,}\\
AMSEM(\hat{\boldsymbol{\beta}}^R) &= E\big\{\boldsymbol{I}^{(2)}\boldsymbol{I}^{(2)\prime}\big\}\\
&= \sigma^2\boldsymbol{E_0}+(-\boldsymbol{\nu})(-\boldsymbol{\nu}')\\
&= \sigma^2\boldsymbol{E_0}+\boldsymbol{\nu}\boldsymbol{\nu}', \text{ using Theorem 4.4.1(ii),}
\end{aligned}$$

$$\begin{aligned}
AQR(\hat{\boldsymbol{\beta}}^R,\boldsymbol{M}) &= tr\left(\boldsymbol{M}\{AMSEM(\hat{\boldsymbol{\beta}}^R)\}\right)\\
&= tr\big\{\boldsymbol{M}\big[\sigma^2\boldsymbol{E_0}+\boldsymbol{\nu}\boldsymbol{\nu}'\big]\big\}\\
&= tr\Big\{\boldsymbol{M}\big[\sigma^2\left(\boldsymbol{V}_0^{-1}-\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\right)+\boldsymbol{\nu}\boldsymbol{\nu}'\big]\Big\}\\
&= tr\left(\boldsymbol{M}\boldsymbol{V}_0^{-1}\right)\\
&- \sigma^2 tr\big\{\boldsymbol{M}\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}\big\}+\boldsymbol{\nu}'\boldsymbol{M}\boldsymbol{\nu}.
\end{aligned}$$

The $p\times p$ matrix $\boldsymbol{V_0}^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V_0}^{-1/2}$ is symmetric and idempotent of rank $q(q \leq p)$, therefore, there exists an orthogonal $p \times p$ matrix $\boldsymbol{\Gamma}$ such that $\boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V_0}^{-1/2}\boldsymbol{\Gamma}' = \begin{pmatrix} \boldsymbol{I_q} & \boldsymbol{0_{q\times(p-q)}} \\ \boldsymbol{0_{(p-q)\times q}} & \boldsymbol{0_{(p-q)\times(p-q)}} \end{pmatrix}$, and $\boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{M}\boldsymbol{V_0}^{-1/2}\boldsymbol{\Gamma}' = \begin{pmatrix} \boldsymbol{U_{11}} & \boldsymbol{U_{12}} \\ \boldsymbol{U'_{12}} & \boldsymbol{U_{22}} \end{pmatrix}$, where $\boldsymbol{U_{11}}$ and $\boldsymbol{U_{22}}$ are square matrices of orders q and $(p-q)$, respectively. Thus,

$$\begin{aligned}
& tr\{\boldsymbol{M}\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}\} = \\
& tr\Big\{(\boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{M}\boldsymbol{V_0}^{-1/2}\boldsymbol{\Gamma}')(\boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{H}' \\
& (\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V_0}^{-1/2}\boldsymbol{\Gamma}')\Big\} \\
= \; & tr\left(\begin{pmatrix} \boldsymbol{U_{11}} & \boldsymbol{U_{12}} \\ \boldsymbol{U'_{12}} & \boldsymbol{U_{22}} \end{pmatrix}\begin{pmatrix} \boldsymbol{I_q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\right) \\
= \; & tr\,(\boldsymbol{U_{11}}).
\end{aligned}$$

Moreover, $\boldsymbol{\nu}'\boldsymbol{M}\boldsymbol{\nu}$ can be written as follows

$$\begin{aligned}
\boldsymbol{\nu}'\boldsymbol{M}\boldsymbol{\nu} &= \left[\boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}\right]\boldsymbol{M}\left[\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}\right] \\
&= \left[\boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}\right]\boldsymbol{M} \\
&\quad \left[\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}\right] \\
&= \left[\boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V_0}^{-1/2}\boldsymbol{\Gamma}'\right]\left[\boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1} \right. \\
&\quad \left. \boldsymbol{H}\boldsymbol{V_0}^{-1/2}\boldsymbol{\Gamma}'\right]\left[\boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{M}\boldsymbol{V_0}^{-1/2}\boldsymbol{\Gamma}'\right]\left[\boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{H}' \right. \\
&\quad \left. (\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V_0}^{-1/2}\boldsymbol{\Gamma}'\right]\left[\boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}\right].
\end{aligned}$$

Let $\boldsymbol{u} = \boldsymbol{\Gamma}\boldsymbol{V_0}^{-1/2}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}$, then we have

$$
\begin{aligned}
\boldsymbol{\nu}'\boldsymbol{M}\boldsymbol{\nu} &= \boldsymbol{u}'\begin{pmatrix} \boldsymbol{I_q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{U_{11}} & \boldsymbol{U_{12}} \\ \boldsymbol{U_{12}'} & \boldsymbol{U_{22}} \end{pmatrix}\begin{pmatrix} \boldsymbol{I_q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{u} \\
&= \boldsymbol{u}'\begin{pmatrix} \boldsymbol{U_{11}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{u} \\
&= \boldsymbol{u_1'}\boldsymbol{U_{11}}\boldsymbol{u_1}, \quad \text{where} \quad \boldsymbol{u} = \begin{pmatrix} \boldsymbol{u_1} \\ \boldsymbol{u_2} \end{pmatrix}.
\end{aligned}
$$

Therefore,

$$
AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) = \sigma^2 tr\left(\boldsymbol{M}\boldsymbol{V}_0^{-1}\right) - \sigma^2 tr(\boldsymbol{U_{11}}) + \boldsymbol{u_1'}\boldsymbol{U_{11}}\boldsymbol{u_1}.
$$

(iii) Also, note that

$$
\begin{aligned}
n(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta})' &= n\left(\hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)I(\Phi_n \le \Phi_{n,\alpha}) - \boldsymbol{\beta}\right) \\
&\quad \left(\hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)I(\Phi_n \le \Phi_{n,\alpha}) - \boldsymbol{\beta}\right)' \\
&= n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \\
&+ n(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)'I^2(\Phi_n \le \Phi_{n,\alpha}) \\
&- 2n(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'I(\Phi_n \le \Phi_{n,\alpha}) \\
&= \boldsymbol{I}_n^{(1)}\boldsymbol{I}_n^{(1)\prime} + \boldsymbol{I}_n^{(3)}\boldsymbol{I}_n^{(3)\prime}I^2(\Phi_n \le \Phi_{n,\alpha}) \\
&- 2\boldsymbol{I}_n^{(3)}\boldsymbol{I}_n^{(1)\prime}I(\Phi_n \le \Phi_{n,\alpha}).
\end{aligned}
$$

Now, the second term of the last equality can be written as:

$$\begin{aligned}
\boldsymbol{I}_n^{(3)}\boldsymbol{I}_n^{(3)\prime}I^2(\Phi_n \le \Phi_{n,\alpha}) &= n(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)'I^2(\Phi_n \le \Phi_{n,\alpha}) \\
&= n\Big[(\boldsymbol{X_0}'\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}'(\boldsymbol{H}(\boldsymbol{X_0}'\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1} \\
&\quad (\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})\Big]\Big[(\boldsymbol{X_0}'\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}' \\
&\quad (\boldsymbol{H}(\boldsymbol{X_0}'\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h})\Big]' I(\Phi_n \le \Phi_{n,\alpha}).
\end{aligned}$$

By Theorem 4.5.1(iii) we have

$$\sqrt{n}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}) \xrightarrow{D} N_q\left(\boldsymbol{\xi}, \sigma^2(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\right).$$

Therefore,

$$\begin{aligned}
&\left[s_e^2(\boldsymbol{H}(\boldsymbol{X_0}'\hat{\boldsymbol{V_n}}^{-1}\boldsymbol{X_0})^{-1}\boldsymbol{H}')^{-1}\right]^{-1/2}\sqrt{n}(\boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{h}) \\
&\xrightarrow{D} N_q\left((\sigma^2(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1})^{-1/2}\boldsymbol{\xi}, \boldsymbol{I_q}\right).
\end{aligned}$$

So,

$$AMSEM(\hat{\boldsymbol{\beta}}^{PT}) = E_1 + E_2 + E_3,$$

where the three terms in the right-hand side of the last equality can be manipulated as follows:

$$\begin{aligned}
E_1 &= E\big\{\boldsymbol{I}^{(1)}\boldsymbol{I}^{(1)\prime}\big\} \\
&= \sigma^2\boldsymbol{V}_0^{-1} \text{ by Theorem 4.4.1(i).} \\
E_2 &= E\big\{\boldsymbol{I}^{(3)}\boldsymbol{I}^{(3)\prime}I^2(\Phi \le \chi_q^2(\alpha); \Delta^2)\big\} \\
E_3 &= -2E\big\{\boldsymbol{I}^{(3)}\boldsymbol{I}^{(1)\prime}I(\Phi \le \chi_q^2(\alpha); \Delta^2)\big\}.
\end{aligned}$$

Hence, by Theorem 4.4.1(i), Theorem 2.5.2, and Theorem 2.5.4 E_2 is given by

$$\begin{aligned} E_2 &= \sigma^2 \boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\ &+ \boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{\xi}\boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1} \\ &\quad H_{q+4}(\chi_q^2(\alpha);\Delta^2) \\ &= \sigma^2\left(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0}\right)H_{q+2}(\chi_q^2(\alpha);\Delta^2)+\boldsymbol{\nu}\boldsymbol{\nu}'H_{q+4}(\chi_q^2(\alpha);\Delta^2), \text{ and} \end{aligned}$$

$$\begin{aligned} E_3 &= -2E\Big\{E\Big[\boldsymbol{I}^{(3)}\boldsymbol{I}^{(1)\prime}I(\Phi\le\chi_q^2(\alpha);\Delta^2)|\boldsymbol{I}^{(3)}\Big]\Big\} \\ &= -2E\Big\{\boldsymbol{I}^{(3)}E\Big[\boldsymbol{I}^{(1)}+(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0})(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0})^{-1} \\ &\quad (\boldsymbol{I}^{(3)}-\boldsymbol{\nu})\Big]'I(\Phi\le\chi_q^2(\alpha);\Delta^2)\Big\} \\ &= -2E\Big\{\boldsymbol{I}^{(3)}\left(\boldsymbol{I}^{(3)}-\boldsymbol{\nu}\right)'I(\Phi\le\chi_q^2(\alpha);\Delta^2)\Big\} \\ &= -2\times(\text{Second term})+2\boldsymbol{\nu}\boldsymbol{\nu}'H_{q+2}(\chi_q^2(\alpha);\Delta^2), \text{ using 2.5.2.} \end{aligned}$$

By combining the three terms, we have

$$\begin{aligned} AMSEM(\hat{\boldsymbol{\beta}}^{PT}) &= E_1+E_2+E_3 \\ &= \sigma^2\boldsymbol{V}_0^{-1} \\ &- \sigma^2(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0})H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\ &- \boldsymbol{\nu}\boldsymbol{\nu}'H_{q+4}(\chi_q^2(\alpha);\Delta^2)+2\boldsymbol{\nu}\boldsymbol{\nu}'H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\ &= \sigma^2\boldsymbol{V}_0^{-1} \\ &- \sigma^2\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}'\boldsymbol{V}_0^{-1}H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\ &+ \boldsymbol{\nu}\boldsymbol{\nu}'\big\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2)-H_{q+4}(\chi_q^2(\alpha);\Delta^2)\big\}, \end{aligned}$$

$$
\begin{aligned}
AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M}) &= tr\Big(\boldsymbol{M}\big\{AMSEM(\hat{\boldsymbol{\beta}}^{PT})\big\}\Big) \\
&= tr\Big\{\boldsymbol{M}\Big[\sigma^2\boldsymbol{V}_0^{-1} - \sigma^2\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1} \\
&\quad H_{q+2}(\chi_q^2(\alpha);\Delta^2) + \boldsymbol{\nu}\boldsymbol{\nu}'\big\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\
&\quad -H_{q+4}(\chi_q^2(\alpha);\Delta^2)\big\}\Big]\Big\} \\
&= \sigma^2 tr(\boldsymbol{M}\boldsymbol{V}_0^{-1}) \\
&- \sigma^2 tr\left(\boldsymbol{M}\boldsymbol{V}_0^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}_0^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{V}_0^{-1}\right) H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\
&+ \boldsymbol{\nu}'\boldsymbol{M}\boldsymbol{\nu}\big\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2) - H_{q+4}(\chi_q^2(\alpha);\Delta^2)\big\} \\
&= \sigma^2 tr(\boldsymbol{M}\boldsymbol{V}_0^{-1}) \\
&- \sigma^2 tr(\boldsymbol{U}_{\mathbf{11}})H_{q+2}(\chi_q^2(\alpha);\Delta^2) \\
&+ \boldsymbol{u}_{\mathbf{1}}'\boldsymbol{U}_{\mathbf{11}}\boldsymbol{u}_{\mathbf{1}}\big\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2) - H_{q+4}(\chi_q^2(\alpha);\Delta^2)\big\},
\end{aligned}
$$

as in part (ii).

(iv) Also, note that

$$
\begin{aligned}
n(\hat{\boldsymbol{\beta}}^{JS} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{JS} - \boldsymbol{\beta})' &= n\Big[\hat{\boldsymbol{\beta}} - (q-2)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)\Phi_n^{-1} - \boldsymbol{\beta}\Big] \\
&\quad \Big[\hat{\boldsymbol{\beta}} - (q-2)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)\Phi_n^{-1} - \boldsymbol{\beta}\Big]' \\
&= n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \\
&+ n(q-2)^2(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)'\Phi_n^{-2} \\
&- 2n(q-2)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^R)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\Phi_n^{-1} \\
&= \boldsymbol{I}_n^{(1)}\boldsymbol{I}_n^{(1)\prime} + (q-2)^2\boldsymbol{I}_n^{(3)}\boldsymbol{I}_n^{(3)\prime}\Phi_n^{-2} \\
&- 2(q-2)\boldsymbol{I}_n^{(3)}\boldsymbol{I}_n^{(1)\prime}\Phi_n^{-1}.
\end{aligned}
$$

Therefore, the $AMSEM(\hat{\boldsymbol{\beta}}^{JS}) = E_1 + E_2 + E_3$, where these three terms can be manipulated as follows:

By Theorem 4.4.1(i),

$$\begin{aligned} E_1 &= E\big\{\boldsymbol{I}^{(1)}\boldsymbol{I}^{(1)\prime}\big\} \\ &= \sigma^2\boldsymbol{V}_0^{-1}, \end{aligned}$$

by Theorem 2.5.4, the second term can be simplified to

$$\begin{aligned} E_2 &= (q-2)^2 E\big\{\boldsymbol{I}^{(3)}\boldsymbol{I}^{(3)\prime}\Phi^{-2}\big\} \\ &= (q-2)^2\sigma^2\left(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0}\right)E\left(\chi^{-4}_{q+2}(\Delta^2)\right) \\ &+ (q-2)^2\boldsymbol{\nu}\boldsymbol{\nu}'E\left(\chi^{-4}_{q+4}(\Delta^2)\right), \end{aligned}$$

$$\begin{aligned} E_3 &= -2(q-2)E\big\{\boldsymbol{I}^{(3)}\boldsymbol{I}^{(1)\prime}\Phi^{-1}\big\} \\ &= -2(q-2)E\Big\{E\big\{\boldsymbol{I}^{(3)}\boldsymbol{I}^{(1)\prime}\Phi^{-1}|\boldsymbol{I}^{(3)}\big\}\Big\} \\ &= -2(q-2)E\Big\{\left(\boldsymbol{I}^{(3)}-\boldsymbol{\nu}\right)\boldsymbol{I}^{(3)\prime}\Phi^{-1}\Big\} \\ &= -2(q-2)\Big\{E\big\{\boldsymbol{I}^{(3)}\boldsymbol{I}^{(3)\prime}\Phi^{-1}\big\}-\boldsymbol{\nu}E\big\{\boldsymbol{I}^{(3)\prime}\Phi^{-1}\big\}\Big\} \\ &= -2(q-2)\Big\{\sigma^2\left(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0}\right)E(\chi^{-2}_{q+2}(\Delta^2))+\boldsymbol{\nu}\boldsymbol{\nu}'E(\chi^{-2}_{q+4}(\Delta^2)) \\ &- \boldsymbol{\nu}\boldsymbol{\nu}'E(\chi^{-2}_{q+2}(\Delta^2))\Big\} \\ &= -2(q-2)\sigma^2\left(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0}\right)E(\chi^{-2}_{q+2}(\Delta^2)) \\ &- 2(q-2)\boldsymbol{\nu}\boldsymbol{\nu}'\big\{E(\chi^{-2}_{q+4}(\Delta^2))-E(\chi^{-2}_{q+2}(\Delta^2))\big\}. \end{aligned}$$

By (2.20), the third term becomes

$$\begin{aligned}
E_3 &= -2(q-2)\sigma^2\left(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0}\right)E(\chi_{q+2}^{-2}(\Delta^2)) \\
&- 2(q-2)\boldsymbol{\nu\nu'}\big\{E(\chi_{q+2}^{-2}(\Delta^2))-2E(\chi_{q+4}^{-4}(\Delta^2))-E(\chi_{q+2}^{-2}(\Delta^2))\big\} \\
&= -2(q-2)\sigma^2\left(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0}\right)E(\chi_{q+2}^{-2}(\Delta^2))+4(q-2)\boldsymbol{\nu\nu'}E(\chi_{q+4}^{-4}(\Delta^2)).
\end{aligned}$$

Now, combining the three terms, we have

$$\begin{aligned}
AMSEM(\hat{\boldsymbol{\beta}}^{JS}) &= \sigma^2\boldsymbol{V}_0^{-1} \\
&- (q-2)\sigma^2\left(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0}\right)\Big\{2E(\chi_{q+2}^{-2}(\Delta^2))-(q-2) \\
&\quad E(\chi_{q+2}^{-4}(\Delta^2))\Big\}+(q-2)(q+2)\boldsymbol{\nu\nu'}E(\chi_{q+4}^{-4}(\Delta^2)),\ \text{and}
\end{aligned}$$

$$\begin{aligned}
AQR(\hat{\boldsymbol{\beta}}^{JS},\boldsymbol{M}) &= tr\left(\boldsymbol{M}\big\{AMSEM(\hat{\boldsymbol{\beta}}^{JS})\big\}\right) \\
&= tr\Big\{\boldsymbol{M}\Big[\sigma^2\boldsymbol{V}_0^{-1}-(q-2)\sigma^2\left(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0}\right)\big\{2E(\chi_{q+2}^{-2}(\Delta^2)) \\
&\quad -(q-2)E(\chi_{q+2}^{-4}(\Delta^2))\big\}+(q-2)(q+2)\boldsymbol{\nu\nu'}E(\chi_{q+4}^{-4}(\Delta^2))\Big]\Big\} \\
&= \sigma^2 tr\left(\boldsymbol{M}\boldsymbol{V}_0^{-1}\right) \\
&- (q-2)\sigma^2\big\{2E(\chi_{q+2}^{-2}(\Delta^2))-(q-2)E(\chi_{q+2}^{-4}(\Delta^2))\big\}tr(\boldsymbol{U_{11}}) \\
&+ (q-2)(q+2)E(\chi_{q+4}^{-4}(\Delta^2))\boldsymbol{u_1'}\boldsymbol{U_{11}}\boldsymbol{u_1}.
\end{aligned}$$

(v) Finally,

$$\begin{aligned}
n(\hat{\boldsymbol{\beta}}^{JS+}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{JS+}-\boldsymbol{\beta})' &= n\Big[\hat{\boldsymbol{\beta}}^{JS}-(1-(q-2)\Phi_n^{-1})I(\Phi_n<(q-2)) \\
&\quad (\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})-\boldsymbol{\beta}\Big]\Big[\hat{\boldsymbol{\beta}}^{JS}-(1-(q-2)\Phi_n^{-1}) \\
&\quad I(\Phi_n<(q-2))(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})-\boldsymbol{\beta}\Big]' \\
&= n(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})' \\
&+ n(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})'(1-(q-2)\Phi_n^{-1})^2 \\
&\quad I^2(\Phi_n<(q-2))-2n(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})' \\
&\quad (1-(q-2)\Phi_n^{-1})I(\Phi_n<(q-2)) \\
&= n(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})' \\
&+ \boldsymbol{I}_n^{(3)}\boldsymbol{I}_n^{(3)\prime}(1-(q-2)\Phi_n^{-1})^2I^2(\Phi_n<(q-2)) \\
&- 2\sqrt{n}(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})\boldsymbol{I}_n^{(3)\prime}(1-(q-2)\Phi_n^{-1}) \\
&\quad I(\Phi_n<(q-2)). \qquad (4.19)
\end{aligned}$$

Note that the third term of (4.19) can be written as:

$$\begin{aligned}
\text{Third term} &= -2n(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})'(1-(q-2)\Phi_n^{-1})I(\Phi_n<(q-2)) \\
&= -2n\left(\hat{\boldsymbol{\beta}}^{R}+(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})(1-(q-2)\Phi_n^{-1})-\boldsymbol{\beta}\right)(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})' \\
&\quad (1-(q-2)\Phi_n^{-1})I(\Phi_n<(q-2)) \\
&= -2n(\hat{\boldsymbol{\beta}}^{R}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})'(1-(q-2)\Phi_n^{-1})I(\Phi_n<(q-2)) \\
&- 2n(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}^{R})'(1-(q-2)\Phi_n^{-1})^2I(\Phi_n<(q-2)) \\
&= -2\boldsymbol{I}_n^{(2)}\boldsymbol{I}_n^{(3)\prime}(1-(q-2)\Phi_n^{-1})I(\Phi_n<(q-2)) \\
&- 2\boldsymbol{I}_n^{(3)}\boldsymbol{I}_n^{(3)\prime}(1-(q-2)\Phi_n^{-1})^2I(\Phi_n<(q-2)).
\end{aligned}$$

Therefore, $AMSEM(\hat{\boldsymbol{\beta}}^{JS+}) = E_1 + E_2 + E_3$, where the three terms of the $AMSEM(\hat{\boldsymbol{\beta}}^{JS+})$ can be worked out as follows:

$$
\begin{aligned}
E_1 &= AMSEM(\hat{\boldsymbol{\beta}}^{JS}),\\
E_2 &= E\Big\{\boldsymbol{I}^{(3)}\boldsymbol{I}^{(3)\prime}(1-(q-2)\Phi^{-1})^2 I^2(\Phi<(q-2))\Big\}\\
&= \sigma^2\left(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0}\right)E\Big\{\left(1-\chi_{q+2}^{-2}(\Delta^2)\right)^2 I(\chi_{q+2}^2(\Delta^2)<(q-2))\Big\}\\
&+ 2\boldsymbol{\nu}\boldsymbol{\nu}'E\Big\{\left(1-(q-2)\chi_{q+4}^{-2}(\Delta^2)\right)^2 I(\chi_{q+4}^2(\Delta^2)<(q-2))\Big\},
\end{aligned}
$$

Using Theorem 4.4.1(v), we have

$$
\begin{aligned}
E_3 &= -2E\Big\{\boldsymbol{I}^{(2)}\boldsymbol{I}^{(3)\prime}(1-(q-2)\Phi^{-1})I(\Phi<(q-2))\Big\}\\
&- 2E\Big\{\boldsymbol{I}^{(3)}\boldsymbol{I}^{(3)\prime}(1-(q-2)\Phi^{-1})^2 I(\Phi<(q-2))\Big\}\\
&= 2\boldsymbol{\nu}\boldsymbol{\nu}'E\Big\{(1-(q-2)\chi_{q+2}^{-2}(\Delta^2))I(\chi_{q+2}^2(\Delta^2)<(q-2))\Big\}\\
&- 2\times(\text{Second term}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
AMSEM(\hat{\boldsymbol{\beta}}^{JS+}) &= AMSEM(\hat{\boldsymbol{\beta}}^{JS})\\
&- \sigma^2\left(\boldsymbol{V}_0^{-1}-\boldsymbol{E_0}\right)\\
&\quad E\Big\{\left(1-\chi_{q+2}^{-2}(\Delta^2)\right)^2 I(\chi_{q+2}^2(\Delta^2)<(q-2))\Big\}\\
&- \boldsymbol{\nu}\boldsymbol{\nu}'E\Big\{\left(1-\chi_{q+4}^{-2}(\Delta^2)\right)^2 I(\chi_{q+4}^2(\Delta^2)<(q-2))\Big\}\\
&+ 2\boldsymbol{\nu}\boldsymbol{\nu}'E\Big\{\left(1-\chi_{q+2}^{-2}(\Delta^2)\right) I(\chi_{q+2}^2(\Delta^2)<(q-2))\Big\},
\end{aligned}
$$

$$
\begin{aligned}
AQR(\hat{\boldsymbol{\beta}}^{JS+}, \boldsymbol{M}) &= tr\Big(\boldsymbol{M}\Big\{AMSEM(\hat{\boldsymbol{\beta}}^{JS+})\Big\}\Big) \\
&= AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) \\
&- \sigma^2 E\Big\{\big(1-(q-2)\chi^{-2}_{q+2}(\Delta^2)\big)^2 I(\chi^2_{q+2}(\Delta^2) < (q-2))\Big\} \\
&\quad tr(\boldsymbol{U_{11}}) - E\Big\{\big(1-(q-2)\chi^{-2}_{q+4}(\Delta^2)\big)^2 \\
&\quad I(\chi^2_{q+4}(\Delta^2) < (q-2))\Big\}\boldsymbol{u_1'U_{11}u_1} \\
&+ 2E\Big\{\big(1-(q-2)\chi^{-2}_{q+2}(\Delta^2)\big) I(\chi^2_{q+2}(\Delta^2) < (q-2))\Big\} \\
&\quad \boldsymbol{u_1'U_{11}u_1}.
\end{aligned}
$$

□

4.6 Risk Comparisons

In the following, we compare analytically the asymptotic quadratic risk for the listed estimators with respect to the unrestricted as a benchmark estimator.

4.6.1 Comparing $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^R$

It is clear from Theorem 4.5.2 that the asymptotic quadratic risk of the $\hat{\boldsymbol{\beta}}$ is a constant. In the contrary, the AQR of the restricted estimator depends on $\boldsymbol{u_1'U_{11}u_1}$, and performs better than $\hat{\boldsymbol{\beta}}$ at and near the null hypothesis, $\boldsymbol{A}_0$.

For a given $p \times p$ positive definite matrix $\boldsymbol{M}$, the asymptotic quadratic risk of $\hat{\boldsymbol{\beta}}^R$ can be rewritten as

$$AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) = AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{U_{11}}) + \boldsymbol{u_1'U_{11}u_1}.$$

Using Courant Theorem (Saleh, 2006, p.39), we have

$$ch_{\min}(\boldsymbol{U_{11}}) \leq \frac{\boldsymbol{u_1'U_{11}u_1}}{\boldsymbol{u_1'u_1}} \leq ch_{\max}(\boldsymbol{U_{11}}), \tag{4.20}$$

where $ch_{\min}(\boldsymbol{U_{11}})$, and $ch_{\max}(\boldsymbol{U_{11}})$ are the smallest and largest characteristic roots of $\boldsymbol{U_{11}}$, respectively. With $\boldsymbol{u_1'u_1} = \sigma^2\Delta^2$, the inequality given by (4.20) becomes

$$\sigma^2\Delta^2 ch_{\min}(\boldsymbol{U_{11}}) \leq \boldsymbol{u_1'U_{11}u_1} \leq \sigma^2\Delta^2 ch_{\max}(\boldsymbol{U_{11}}),$$

and hence

$$\begin{aligned} &AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{U_{11}}) + \sigma^2\Delta^2 ch_{\min}(\boldsymbol{U_{11}}) \\ &\leq AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) \leq \\ &AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{U_{11}}) + \sigma^2\Delta^2 ch_{\max}(\boldsymbol{U_{11}}). \end{aligned} \tag{4.21}$$

We can conclude the following results:

1. If $\Delta^2 = 0$, the lower and upper bounds of $AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M})$ are equal, the local alternatives (4.16) reduce to (4.8), and thus, $\hat{\boldsymbol{\beta}}^R$ has asymptotic quadratic risk less than or equal to that of $\hat{\boldsymbol{\beta}}$. That is, if the restriction given by (4.8) is correct, the restricted estimator, $\hat{\boldsymbol{\beta}}^R$, always dominates $\hat{\boldsymbol{\beta}}$.

2. When $\Delta^2 > 0$, then from the first part of the inequality (4.21), we have

$$AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) \geq -\sigma^2 tr(\boldsymbol{U_{11}}) + \sigma^2 ch_{\min}(\boldsymbol{U_{11}}).$$

   The difference in the left hand side of the above inequality is non-negative whenever $\Delta^2 \geq \frac{tr(\boldsymbol{U_{11}})}{ch_{\min}(\boldsymbol{U_{11}})}$. That is, $\hat{\boldsymbol{\beta}}$ performs better than $\hat{\boldsymbol{\beta}}^R$ for all $\Delta^2 \in$

$\left[\frac{tr(\boldsymbol{U_{11}})}{ch_{\min}(\boldsymbol{U_{11}})}, \infty\right)$. In fact, the asymptotic quadratic risk of $\hat{\boldsymbol{\beta}}^R$ increases without bound beyond $\frac{tr(\boldsymbol{U_{11}})}{ch_{\min}(\boldsymbol{U_{11}})}$.

Further, from the last part of the inequality (4.21), we have

$$AQR(\hat{\boldsymbol{\beta}}^R, \boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) \leq -\sigma^2 tr(\boldsymbol{U_{11}}) + \sigma^2\Delta^2 ch_{\max}(\boldsymbol{U_{11}}).$$

The difference in the left hand side of the above inequality is negative whenever $\Delta^2 \leq \frac{tr(\boldsymbol{U_{11}})}{ch_{\max}(\boldsymbol{U_{11}})}$, thus $\hat{\boldsymbol{\beta}}^R$ performs better than $\hat{\boldsymbol{\beta}}$ for all $\Delta^2 \in \left[0, \frac{tr(\boldsymbol{U_{11}})}{ch_{\max}(\boldsymbol{U_{11}})}\right]$. Moreover, the asymptotic quadratic risk of both estimators are equal when $ch_{\min}(\boldsymbol{U_{11}}) = ch_{\max}(\boldsymbol{U_{11}})$ regardless of σ^2 and Δ^2, one possible case for such equality occurs when $\boldsymbol{U_{11}} = \boldsymbol{I_q}$.

4.6.2 Comparing $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{PT}$

Note that,

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M}) &= AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{U_{11}}) H_{q+2}(\chi_q^2(\alpha); \Delta^2) \\ &+ \left\{2H_{q+2}(\chi_q^2(\alpha); \Delta^2) - H_{q+4}(\chi_q^2(\alpha); \Delta^2)\right\} \boldsymbol{u_1'} \boldsymbol{U_{11}} \boldsymbol{u_1}. \end{aligned}$$

Now, by Courant Theorem (Saleh, 2006, p.39),

$$\begin{aligned} & AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{U_{11}}) H_{q+2}(\chi_q^2(\alpha); \Delta^2) \\ + \ & \sigma^2\Delta^2 ch_{\min}(\boldsymbol{U_{11}}) \left\{2H_{q+2}(\chi_q^2(\alpha); \Delta^2) - H_{q+4}(\chi_q^2(\alpha); \Delta^2)\right\} \\ & \quad \leq AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M}) \leq \\ & AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2 tr(\boldsymbol{U_{11}}) H_{q+2}(\chi_q^2(\alpha); \Delta^2) \\ + \ & \sigma^2\Delta^2 ch_{\max}(\boldsymbol{U_{11}}) \left\{2H_{q+2}(\chi_q^2(\alpha); \Delta^2) - H_{q+4}(\chi_q^2(\alpha); \Delta^2)\right\}. \end{aligned} \quad (4.22)$$

From the above inequality, and following the same procedure in the previous subsection, we can conclude the following.

1. When $\Delta^2 = 0$, the lower and upper bounds of the inequality (4.22) are equal with

$$AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M}) = \sigma^2 tr(\boldsymbol{U_{11}}) H_{q+2}(\chi^2_q(\alpha); \Delta^2) \quad \geq 0,$$

which indicates that $\hat{\boldsymbol{\beta}}^{PT}$ has less risk than $\hat{\boldsymbol{\beta}}$.

2. When $\Delta^2 > 0$, there are two cases

   a. From the first part of the inequality (4.22) we have,

$$\begin{aligned} & AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) \leq -\sigma^2 tr(\boldsymbol{U_{11}}) \\ + \quad & \sigma^2 \Delta^2 ch_{\min}(\boldsymbol{U_{11}}) \big\{ 2H_{q+2}(\chi^2_q(\alpha); \Delta^2) - H_{q+4}(\chi^2_q(\alpha); \Delta^2) \big\}. \end{aligned}$$

   This difference is greater than or equal to zero whenever $\Delta^2 \geq \frac{tr(\boldsymbol{U_{11}}) H_{q+2}(\chi^2_q(\alpha);\Delta^2)}{ch_{\min}(\boldsymbol{U_{11}}) \left\{ 2H_{q+2}(\chi^2_q(\alpha);\Delta^2) - H_{q+4}(\chi^2_q(\alpha);\Delta^2) \right\}}$,

   and hence, $\hat{\boldsymbol{\beta}}$ performs better than $\hat{\boldsymbol{\beta}}^{PT}$ for all $\Delta^2 \in \left[\frac{tr(\boldsymbol{U_{11}}) H_{q+2}(\chi^2_q(\alpha);\Delta^2)}{ch_{\min}(\boldsymbol{U_{11}}) \left\{ 2H_{q+2}(\chi^2_q(\alpha);\Delta^2) - H_{q+4}(\chi^2_q(\alpha);\Delta^2) \right\}}, \infty \right)$.

   b. From the last part of the inequality (4.22) we have,

$$\begin{aligned} & AQR(\hat{\boldsymbol{\beta}}^{PT}, \boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) \leq -\sigma^2 tr(\boldsymbol{U_{11}}) \\ + \quad & \sigma^2 \Delta^2 ch_{\max}(\boldsymbol{U_{11}}) \big\{ 2H_{q+2}(\chi^2_q(\alpha); \Delta^2) - H_{q+4}(\chi^2_q(\alpha); \Delta^2) \big\}. \end{aligned}$$

The above difference is less than or equal to zero whenever
$\Delta^2 \leq \frac{tr(\boldsymbol{U_{11}})H_{q+2}(\chi_q^2(\alpha);\Delta^2)}{ch_{\max}(\boldsymbol{U_{11}})\left\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2)-H_{q+4}(\chi_q^2(\alpha);\Delta^2)\right\}}$,

that is, $\hat{\boldsymbol{\beta}}^{PT}$ performs better than $\hat{\boldsymbol{\beta}}^{R}$ for all
$\Delta^2 \in \left[0, \frac{tr(\boldsymbol{U_{11}})H_{q+2}(\chi_q^2(\alpha);\Delta^2)}{ch_{\max}(\boldsymbol{U_{11}})\left\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2)-H_{q+4}(\chi_q^2(\alpha);\Delta^2)\right\}}\right]$.

Moreover, the asymptotic quadratic risk of the two estimators are equal for all Δ^2 in the interval $\left[L, U\right]$, where L and U are, respectively,

$$\begin{aligned} L &= \frac{tr(\boldsymbol{U_{11}})H_{q+2}(\chi_q^2(\alpha);\Delta^2)}{ch_{\max}(\boldsymbol{U_{11}})\left\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2)-H_{q+4}(\chi_q^2(\alpha);\Delta^2)\right\}}, \\ U &= \frac{tr(\boldsymbol{U_{11}})H_{q+2}(\chi_q^2(\alpha);\Delta^2)}{ch_{\min}(\boldsymbol{U_{11}})\left\{2H_{q+2}(\chi_q^2(\alpha);\Delta^2)-H_{q+4}(\chi_q^2(\alpha);\Delta^2)\right\}}. \end{aligned}$$

Also, the lower and upper limits of the above interval may be equal when $\boldsymbol{U_{11}} = \boldsymbol{I_q}$. So if the test statistic fails to reject the null hypothesis, then $\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}}^{R}$, and hence, the equality of the asymptotic quadratic risks holds when $\boldsymbol{U_{11}} = \boldsymbol{I_q}$.

4.6.3 Comparing $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{JS}$

Note that,

$$\begin{aligned} AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) &= AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) - \sigma^2(q-2)\left\{2E(\chi_{q+2}^{-2}(\Delta^2)) - (q-2)E(\chi_{q+2}^{-4}(\Delta^2))\right\} \\ & \quad tr(\boldsymbol{U_{11}}) + (q-2)(q+2)E(\chi_{q+4}^{-4}(\Delta^2))\boldsymbol{u_1'}\boldsymbol{U_{11}}\boldsymbol{u_1}. \end{aligned}$$

By using the results in (Saleh, 2006, p.32), we get

$$\begin{aligned}
AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) &= AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) \\
&- \sigma^2(q-2)tr(\boldsymbol{U_{11}})\Big\{(q-2)E\left(\chi_{q+2}^{-4}(\Delta^2)\right) + 2\Delta^2 E\left(\chi_{q+4}^{-4}(\Delta^2)\right)\Big\} \\
&+(q-2)(q+2)E\left(\chi_{q+4}^{-4}(\Delta^2)\right)\boldsymbol{u_1'U_{11}u_1}
\end{aligned}$$

$$\begin{aligned}
AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) &= AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) \\
&- \sigma^2 tr(\boldsymbol{U_{11}})(q-2)\Bigg\{(q-2)E\left(\chi_{q+2}^{-4}(\Delta^2)\right) + 2\Delta^2 E\left(\chi_{q+4}^{-4}(\Delta^2)\right) \\
&-(q+2)E\left(\chi_{q+4}^{-4}(\Delta^2)\right)\frac{\boldsymbol{u_1'U_{11}u_1}}{\sigma^2 tr(\boldsymbol{U_{11}})}\Bigg\} \\
&= AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}) \\
&- \sigma^2 tr(\boldsymbol{U_{11}})(q-2)\Bigg\{(q-2)E\left(\chi_{q+2}^{-4}(\Delta^2)\right) \\
&+ 2\Delta^2 E\left(\chi_{q+4}^{-4}(\Delta^2)\right)\left[1 - \frac{(q+2)\boldsymbol{u_1'U_{11}u_1}}{2\Delta^2\sigma^2 tr(\boldsymbol{U_{11}})}\right]\Bigg\}.
\end{aligned}$$

Therefore, $AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) \leq AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M})$ for all Δ^2, $\boldsymbol{M}$, and $q \geq 3$ when

$$\begin{aligned}
1 - \frac{(q+2)\boldsymbol{u_1'U_{11}u_1}}{2\Delta^2\sigma^2 tr(\boldsymbol{U_{11}})} &\geq 0 \\
\text{or} \quad \frac{(q+2)\boldsymbol{u_1'U_{11}u_1}}{2\Delta^2\sigma^2 tr(\boldsymbol{U_{11}})} &\leq 1.
\end{aligned}$$

Using Courant Theorem (Saleh, 2006, p.32), the above inequality holds whenever

$$\begin{aligned}\frac{(q+2)ch_{\max}(\boldsymbol{U_{11}})}{2tr(\boldsymbol{U_{11}})} &\leq 1\\ \frac{tr(\boldsymbol{U_{11}})}{ch_{\max}(\boldsymbol{U_{11}})} &\geq \frac{q+2}{2}, \qquad q\geq 3. \end{aligned} \tag{4.23}$$

That is, the $AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M})$ is less than $AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M})$ in the whole parameter space provided (4.23) holds, with an upper limit achieved when $\Delta^2 \longrightarrow \infty$.

4.6.4 Comparing $\hat{\boldsymbol{\beta}}^{JS}$ and $\hat{\boldsymbol{\beta}}^{JS+}$

From Theorem 4.5.2(v), we get:

$$\begin{aligned}&AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) - AQR(\hat{\boldsymbol{\beta}}^{JS+}, \boldsymbol{M}) =\\ &\sigma^2\Bigg\{E\Big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))^2 I(\chi^2_{q+2}(\Delta^2)<(q-2))\Big\}tr(\boldsymbol{U_{11}})\\ +\;&\frac{1}{\sigma^2}E\Big\{(1-(q-2)\chi^{-2}_{q+4}(\Delta^2))^2 I(\chi^2_{q+4}(\Delta^2)<(q-2))\Big\}\boldsymbol{u_1'U_{11}u_1}\\ -\;&\frac{2}{\sigma^2}E\Big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2)) I(\chi^2_{q+2}(\Delta^2)<(q-2))\Big\}\boldsymbol{u_1'U_{11}u_1}\Bigg\}.\end{aligned}$$

All the expected value expressions in the above risk difference are nonnegative, since

$$\Big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2))^2 I(\chi^2_{q+2}(\Delta^2)<(q-2))\Big\} \geq 0,$$

$$\Big\{(1-(q-2)\chi^{-2}_{q+4}(\Delta^2))^2 I(\chi^2_{q+4}(\Delta^2)<(q-2))\Big\} \geq 0,$$

$$\Big\{(1-(q-2)\chi^{-2}_{q+2}(\Delta^2)) I(\chi^2_{q+2}(\Delta^2)<(q-2))\Big\} \leq 0.$$

Therefore, for all values of Δ^2, positive definite matrix $\boldsymbol{M}$, with $q \geq 3$, the $AQR(\hat{\boldsymbol{\beta}}^{JS+}, \boldsymbol{M})$ is smaller than the $AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M})$, with an upper limit achieved as $\Delta^2 \longrightarrow \infty$. Consequently, we have the following result,

$$AQR(\hat{\boldsymbol{\beta}}^{JS+}, \boldsymbol{M}) \leq AQR(\hat{\boldsymbol{\beta}}^{JS}, \boldsymbol{M}) \leq AQR(\hat{\boldsymbol{\beta}}, \boldsymbol{M}), \text{ for all values of } \Delta^2.$$

4.7 Numerical Studies

In this section, we use Monte Carlo experiments and two real examples to compare the array of estimators proposed in the past sections. In the first part we will compare the restricted, pretest, and shrinkage estimators with respect to the benchmark unrestricted estimator. These results have shown clearly that for the SMA model, the positive shrinkage estimator has the best performance among the non-penalty competitors. Therefore, in the second part of the numerical studies we will restrict attention to the comparison of the penalty, the restricted and positive shrinkage estimators only with respect to the unrestricted estimator.

4.7.1 Relative Performance of the Estimators

In this simulation study, we consider $N \times N$ regular lattices with $N = 6$ and 9, with corresponding sample sizes of $n = 36$ and 81, respectively. For the spatial moving average SMA regression model given by (4.7), we generate design matrix $\boldsymbol{X}$ of dimension $n \times p$ from a standard multivariate normal distribution. We fix $\sigma^2 = 1$, and consider different values of $\rho \in \{-0.90, -0.50, 0, 0.50, 0.90\}$. The regression coefficient $\boldsymbol{\beta}$ is partitioned as $\boldsymbol{\beta} = \big(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\big)$, where $\boldsymbol{\beta}_1$ is a $(p-q) \times 1$ vector of ones, and $\boldsymbol{\beta}_2$ is a $q \times 1$ vector of zeros in order to test the null hypothesis

$$\boldsymbol{A_0}: \beta_j = 0, \quad \text{for} j = p - q + 1, \ldots, p.$$

Finally, the spatial response is generated from the SMA model in (4.2) with a rook based contiguity matrix $\boldsymbol{W}$. Accordingly, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{0})$, and hence the response variable is obtained as in (4.7), that is

$$\boldsymbol{Y} = \boldsymbol{X\beta} + (\boldsymbol{I} + \boldsymbol{\rho W})\boldsymbol{\epsilon}.$$

As in Chapter 2, we used $\alpha = 0.05$ as our level of significance for the test statistic, and for simplicity, we defined the non-centrality parameter Δ^2, which is essentially a measure of how far away we go from the candidate subspace, as $\Delta^2 = \parallel \boldsymbol{\beta} - \boldsymbol{\beta}_{(0)} \parallel$, where $\boldsymbol{\beta}_{(0)} = (\boldsymbol{\beta}_1, \boldsymbol{0})$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{0} + \boldsymbol{\delta})$ and $\parallel . \parallel$ denotes the Euclidian norm. Thus, $\Delta^2 = \parallel \boldsymbol{\delta} \parallel$, where this vector of alternative values was chosen to vary from 0 to 2 with steps of 0.1. Various choices of (p, q) were used in combination with configurations of $\rho \in \{-0.90, -0.50, 0, 0.50, 0.90\}$, $n = 36, 81$ and 2000 Monte Carlo runs for each scenario. In each of these Monte Carlo runs, the restricted, unrestricted, pretest and shrinkage estimators were computed and their *simulated relative efficiency* (SRE) with respect to the benchmark estimator were computed as follows:

$$SRE(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^*) = \frac{SMSE(\hat{\boldsymbol{\beta}})}{SMSE(\hat{\boldsymbol{\beta}}^*)}, \tag{4.24}$$

where

$$SMSE(\hat{\boldsymbol{\beta}}^*) = \sum_{i=1}^{p} (\hat{\beta}_i^* - \beta_i)^2,$$

is the simulated mean squared errors of $\boldsymbol{\beta}^*$, representing any of the estimators of interest. The SRE results are reported in Tables 4.1 to 4.20 for $(p, q) \in \{(6, 3), (9, 6)\}$, and in Figures 4.1 to 4.5 for $(p, q) \in \{(12, 9), (15, 12), (18, 15)\}$.

Figure 4.1: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = -0.90$ for different values of (p, q) based on the SMA model.

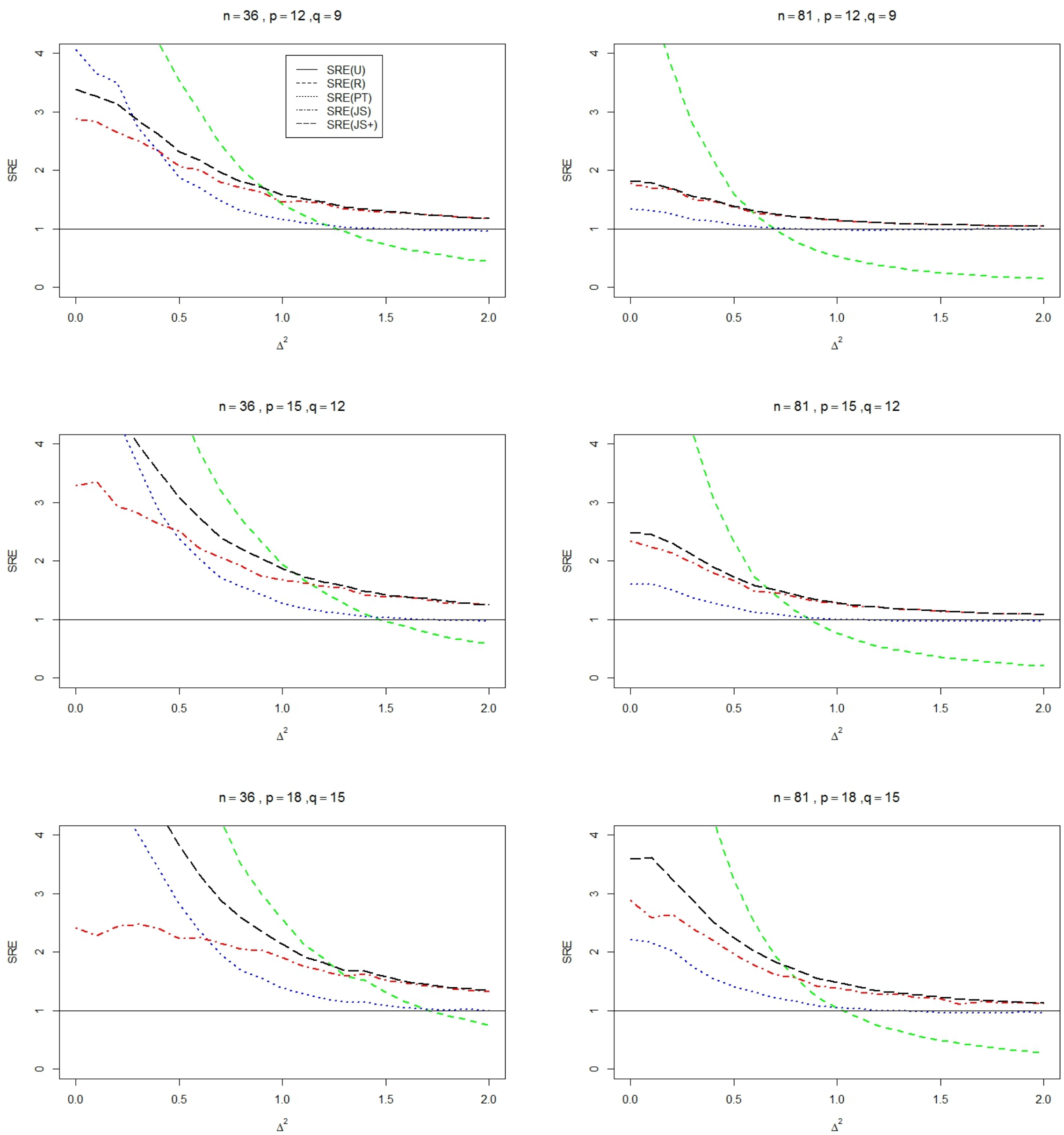

Figure 4.2: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = -0.50$ for different values of (p, q) based on SMA model.

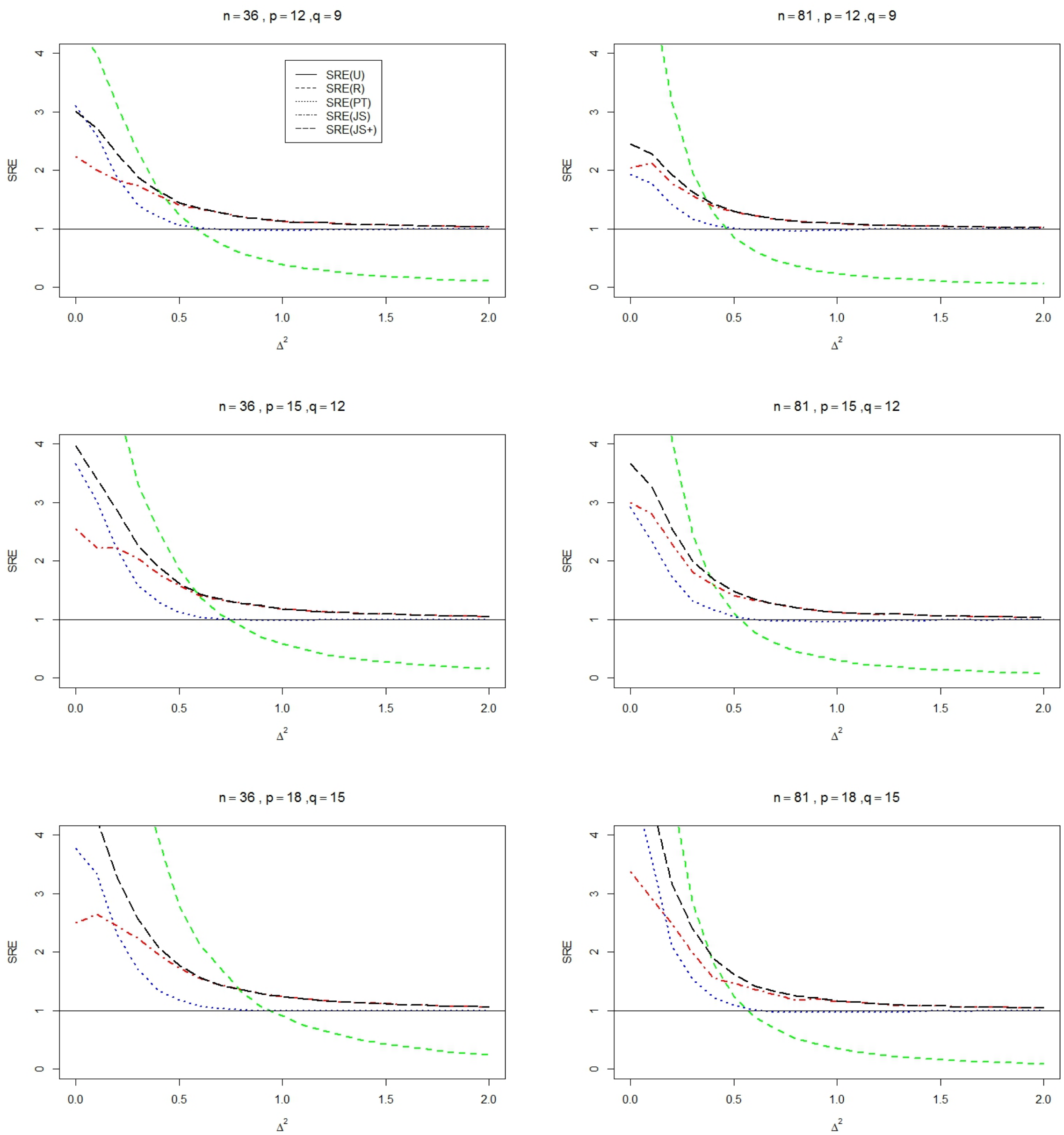

Figure 4.3: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = 0$ for different values of (p, q) based on SMA model.

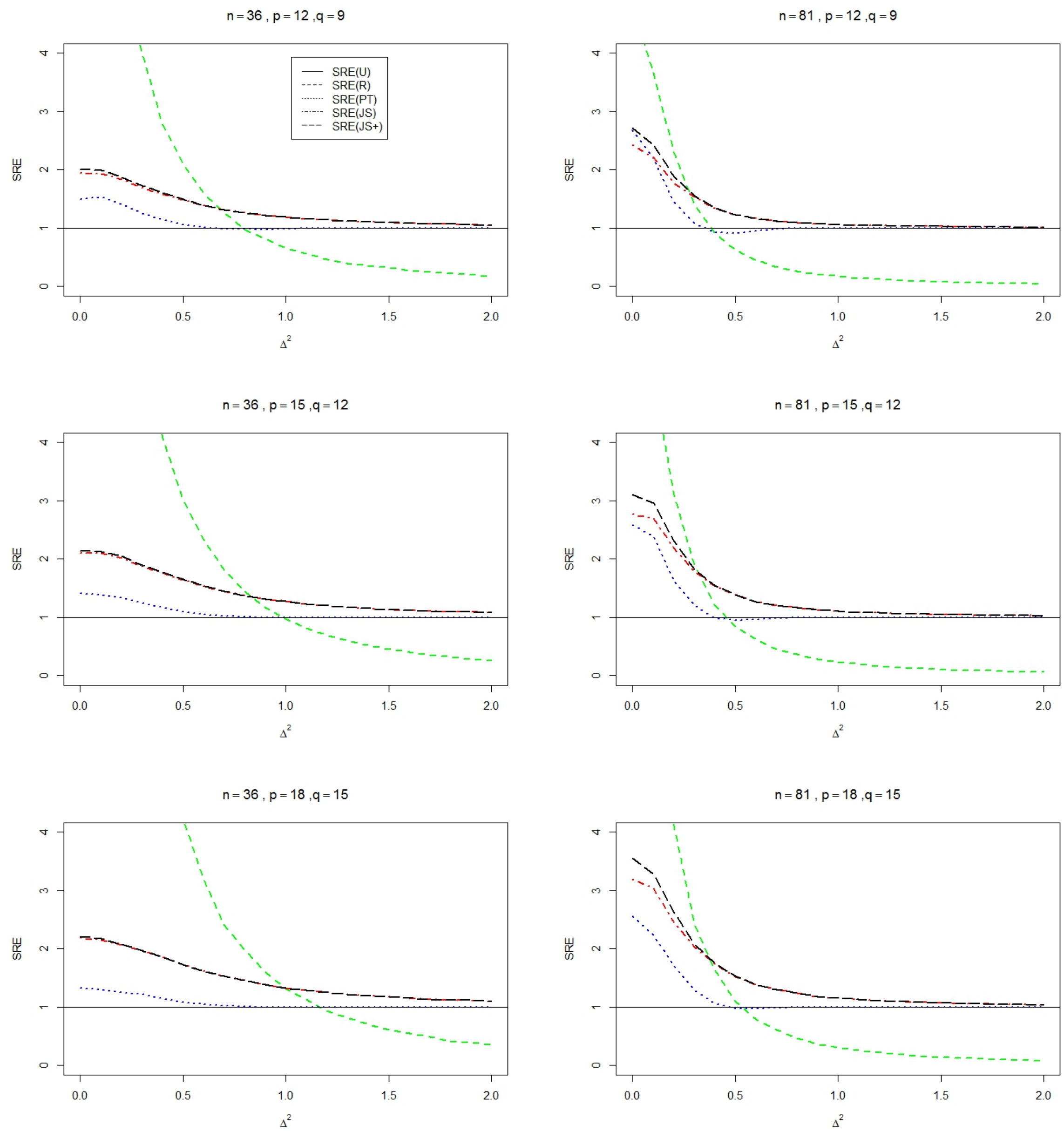

Figure 4.4: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = 0.50$ for different values of (p, q) based on SMA model.

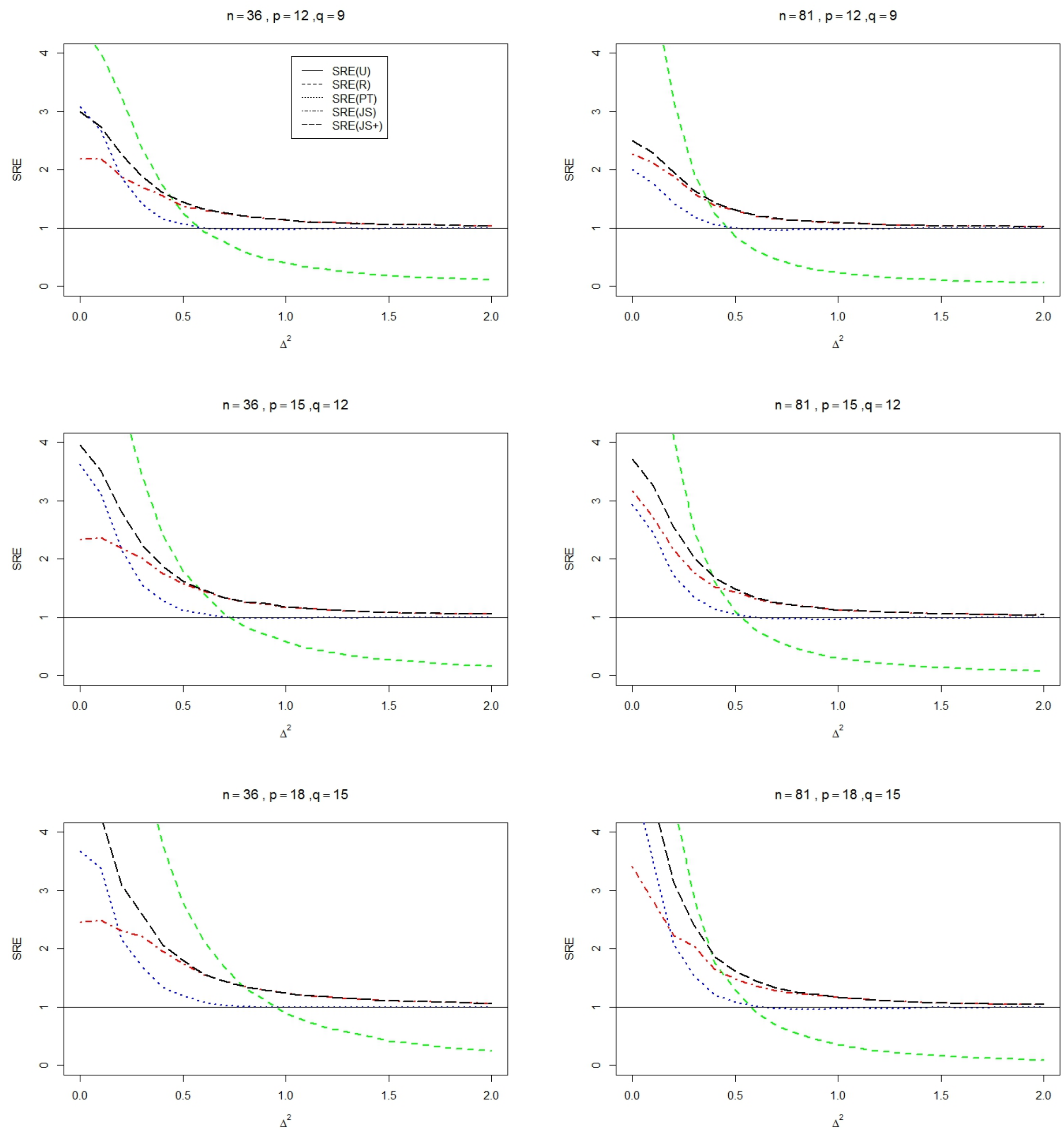

Figure 4.5: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, 81$ and $\rho = 0.90$ for different values of (p, q) based on SMA model.

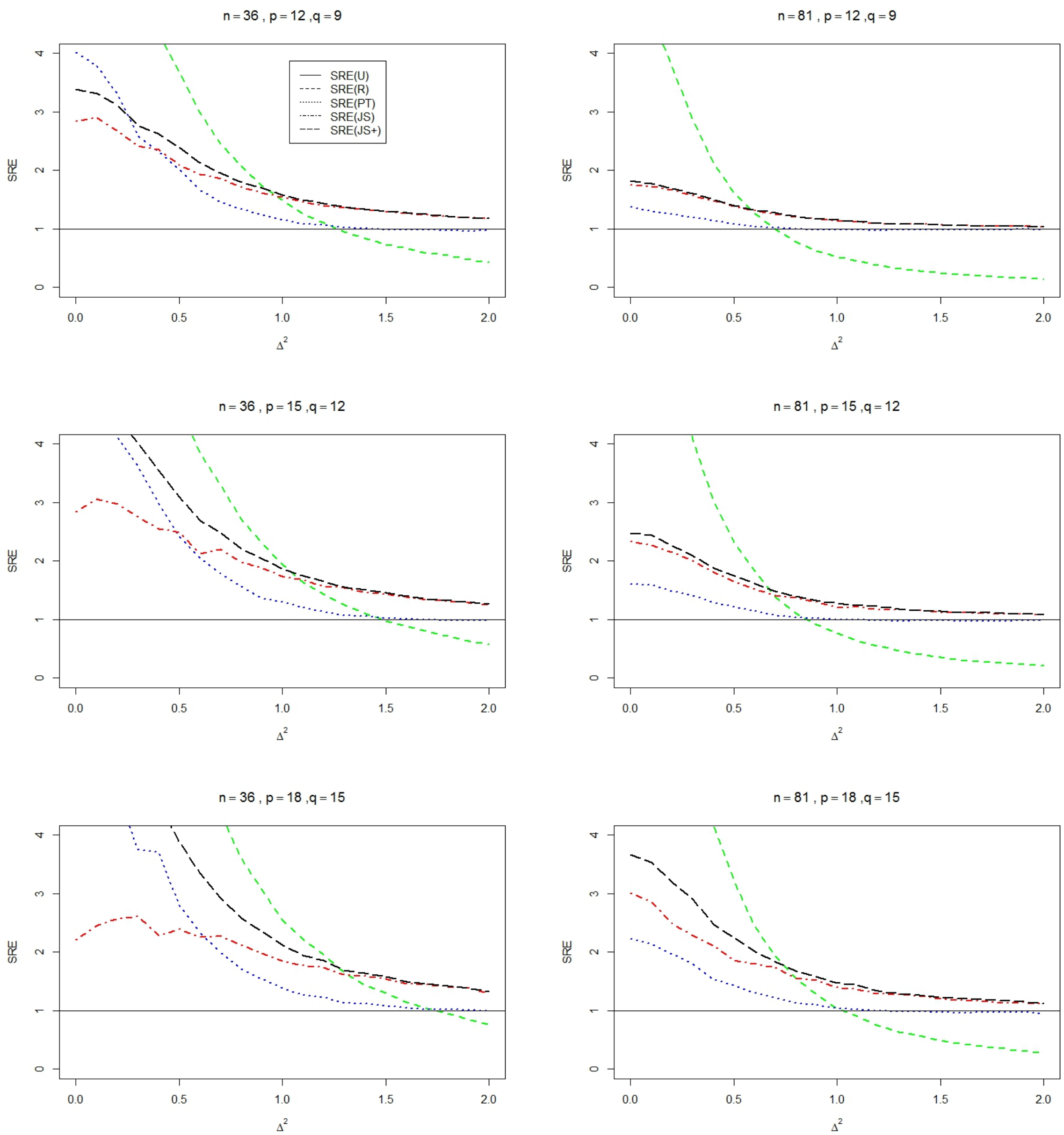

Table 4.1: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = -0.90$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.6179	1.4234	1.1784	1.2007
0.1	2.4983	1.3886	1.1851	1.1995
0.3	1.9078	1.2182	1.1418	1.1540
0.5	1.3234	1.1096	1.1081	1.1134
0.7	0.9141	0.9979	1.0780	1.0803
0.9	0.6435	0.9691	1.0465	1.0540
1.1	0.4773	0.9565	1.0256	1.0381
1.3	0.3657	0.9662	1.0297	1.0297
1.5	0.2804	0.9644	1.0208	1.0208
1.7	0.2183	0.9626	1.0156	1.0156
2.0	0.1630	0.9780	1.0125	1.0125

Table 4.2: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = -0.90$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	4.6106	2.2575	1.9267	2.0102
0.1	4.4516	2.1731	1.9177	1.9874
0.3	3.4899	1.7728	1.7613	1.8145
0.5	2.4386	1.3729	1.5580	1.6153
0.7	1.6174	1.1904	1.4207	1.4499
0.9	1.1946	1.0868	1.3390	1.3551
1.1	0.8490	1.0045	1.2350	1.2430
1.3	0.6390	0.9814	1.1842	1.1923
1.5	0.4991	0.9725	1.1408	1.1499
1.7	0.3909	0.9636	1.1135	1.1136
2.0	0.2991	0.9660	1.0952	1.0952

Table 4.3: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = -0.50$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1959	1.5027	1.1600	1.2562
0.1	2.1494	1.4815	1.1368	1.2541
0.3	1.1810	1.0699	1.0623	1.1358
0.5	0.6568	0.9195	1.0543	1.0671
0.7	0.3798	0.9301	1.0372	1.0439
0.9	0.2488	0.9300	1.0235	1.0235
1.1	0.1787	0.9663	1.0193	1.0193
1.3	0.1282	0.9791	1.0116	1.0116
1.5	0.0957	0.9922	1.0102	1.0102
1.7	0.0769	1.0000	1.0070	1.0070
2.0	0.0546	1.0000	1.0070	1.0070

Table 4.4: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = -0.50$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.3869	2.4671	1.7663	2.1485
0.1	3.1151	2.0463	1.6588	1.9691
0.3	1.7085	1.2227	1.4724	1.5243
0.5	0.9521	0.9873	1.2606	1.2848
0.7	0.5588	0.9492	1.1693	1.1719
0.9	0.3580	0.9504	1.0976	1.0990
1.1	0.2480	0.9646	1.0767	1.0767
1.3	0.1824	0.9862	1.0512	1.0512
1.5	0.1376	0.9934	1.0400	1.0400
1.7	0.1111	1.0000	1.0306	1.0306
2.0	0.0788	0.9950	1.0238	1.0238

Table 4.5: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = 0$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.4696	1.5562	1.2042	1.2564
0.1	2.2344	1.4784	1.1762	1.2344
0.3	1.2649	1.0031	1.1056	1.1189
0.5	0.6740	0.8480	1.0167	1.0499
0.7	0.3980	0.9080	1.0258	1.0268
0.9	0.2561	0.9597	1.0156	1.0156
1.1	0.1745	0.9926	1.0080	1.0082
1.3	0.1292	0.9973	1.0062	1.0062
1.5	0.0975	1.0000	1.0057	1.0057
1.7	0.0741	1.0000	1.0035	1.0035
2.0	0.0554	1.0000	1.0025	1.0025

Table 4.6: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = 0$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	4.6928	1.6689	1.6440	1.7585
0.1	4.1720	1.5606	1.6122	1.7014
0.3	2.4152	1.2053	1.4254	1.4708
0.5	1.3114	0.9845	1.2752	1.2796
0.7	0.7485	0.9409	1.1733	1.1734
0.9	0.4813	0.9625	1.1086	1.1086
1.1	0.3408	0.9912	1.0777	1.0777
1.3	0.2463	0.9973	1.0538	1.0538
1.5	0.1851	1.0000	1.0448	1.0448
1.7	0.1521	1.0000	1.0365	1.0365
2.0	0.1080	1.0000	1.0238	1.0238

Table 4.7: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = 0.50$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.3508	1.5665	1.2382	1.2730
0.1	2.0759	1.4302	1.1562	1.2312
0.3	1.1988	1.0420	1.0932	1.1343
0.5	0.6481	0.9273	1.0640	1.0677
0.7	0.3787	0.9132	1.0303	1.0382
0.9	0.2473	0.9465	1.0255	1.0255
1.1	0.1683	0.9672	1.0157	1.0157
1.3	0.1257	0.9865	1.0112	1.0121
1.5	0.0958	1.0000	1.0100	1.0100
1.7	0.0756	0.9960	1.0088	1.0088
2.0	0.0545	1.0000	1.0046	1.0046

Table 4.8: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = 0.50$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.4139	2.5328	1.7761	2.1420
0.1	2.9762	2.0615	1.7119	1.9733
0.3	1.7274	1.2671	1.4482	1.5389
0.5	0.9209	0.9897	1.2499	1.2789
0.7	0.5494	0.9557	1.1540	1.1607
0.9	0.3546	0.9544	1.1087	1.1094
1.1	0.2489	0.9797	1.0749	1.0749
1.3	0.1818	0.9830	1.0494	1.0495
1.5	0.1370	0.9878	1.0368	1.0368
1.7	0.1072	0.9967	1.0300	1.0300
2.0	0.0798	1.0000	1.0232	1.0232

Table 4.9: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (6, 3)$ and $\rho = 0.90$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.6254	1.3802	1.1835	1.1971
0.1	2.5114	1.3566	1.1783	1.1908
0.3	1.9867	1.2303	1.1469	1.1631
0.5	1.3491	1.0931	1.1046	1.1163
0.7	0.9230	1.0066	1.0752	1.0801
0.9	0.6291	0.9595	1.0447	1.0489
1.1	0.4662	0.9475	1.0331	1.0341
1.3	0.3609	0.9658	1.0285	1.0301
1.5	0.2784	0.9676	1.0228	1.0228
1.7	0.2156	0.9785	1.0167	1.0167
2.0	0.1649	0.9768	1.0127	1.0127

Table 4.10: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 36, (p, q) = (9, 6)$ and $\rho = 0.90$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	4.8694	2.1639	1.9283	2.0065
0.1	4.3093	2.1085	1.8832	1.9688
0.3	3.6889	1.7719	1.7743	1.8380
0.5	2.4021	1.4083	1.5873	1.6194
0.7	1.6629	1.1903	1.4231	1.4661
0.9	1.1743	1.0738	1.3296	1.3421
1.1	0.8740	1.0174	1.2461	1.2549
1.3	0.6256	0.9762	1.1790	1.1858
1.5	0.5027	0.9604	1.1441	1.1454
1.7	0.3953	0.9609	1.1178	1.1221
2.0	0.2982	0.9630	1.0886	1.0889

Table 4.11: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = -0.90$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1539	1.0962	1.0891	1.0907
0.1	1.9227	1.0859	1.0787	1.0831
0.3	1.2013	1.0160	1.0479	1.0497
0.5	0.6725	0.9761	1.0232	1.0239
0.7	0.4074	0.9812	1.0129	1.0129
0.9	0.2664	0.9869	1.0056	1.0056
1.1	0.1843	0.9936	1.0030	1.0030
1.3	0.1357	0.9951	1.0025	1.0025
1.5	0.1056	1.0000	1.0029	1.0029
1.7	0.0805	1.0000	1.0023	1.0023
2.0	0.0604	1.0000	1.0016	1.0016

Table 4.12: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = -0.90$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.3459	1.1821	1.3668	1.3896
0.1	3.2387	1.1554	1.3591	1.3686
0.3	1.9297	1.0758	1.2476	1.2602
0.5	1.1012	1.0078	1.1546	1.1638
0.7	0.6692	0.9872	1.1059	1.1060
0.9	0.4237	0.9828	1.0679	1.0679
1.1	0.2970	0.9833	1.0430	1.0439
1.3	0.2194	0.9878	1.0308	1.0312
1.5	0.1726	0.9948	1.0277	1.0277
1.7	0.1340	1.0000	1.0216	1.0216
2.0	0.0966	1.0000	1.0150	1.0150

Table 4.13: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = -0.50$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.3302	1.2378	1.1461	1.1493
0.1	1.8596	1.1556	1.1117	1.1268
0.3	0.7054	0.9764	1.0343	1.0447
0.5	0.3065	0.9777	1.0127	1.0162
0.7	0.1671	0.9878	1.0073	1.0073
0.9	0.1071	1.0000	1.0065	1.0065
1.1	0.0732	1.0000	1.0043	1.0043
1.3	0.0513	1.0000	1.0015	1.0015
1.5	0.0393	1.0000	1.0013	1.0013
1.7	0.0309	1.0000	1.0025	1.0025
1.9	0.0251	1.0000	1.0013	1.0013
2.0	0.0222	1.0000	1.0010	1.0010

Table 4.14: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = -0.50$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	4.2996	1.4525	1.6412	1.6701
0.1	3.4034	1.3577	1.5523	1.5723
0.3	1.2638	1.0660	1.2770	1.2865
0.5	0.5670	0.9827	1.1359	1.1376
0.7	0.3112	0.9792	1.0781	1.0793
0.9	0.1925	0.9865	1.0495	1.0495
1.1	0.1337	0.9912	1.0338	1.0338
1.3	0.0955	1.0000	1.0247	1.0247
1.5	0.0716	1.0000	1.0195	1.0195
1.7	0.0565	1.0000	1.0148	1.0148
2.0	0.0402	1.0000	1.0089	1.0089

Table 4.15: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = 0$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1077	1.6699	0.9155	1.2970
0.1	1.6563	1.3586	1.1529	1.2181
0.3	0.6344	0.7868	1.0357	1.0406
0.5	0.2778	0.9281	1.0108	1.0109
0.7	0.1537	0.9946	1.0071	1.0071
0.9	0.0939	1.0000	1.0053	1.0053
1.1	0.0640	1.0000	1.0027	1.0027
1.3	0.0468	1.0000	1.0021	1.0021
1.5	0.0353	1.0000	1.0016	1.0016
1.7	0.0280	1.0000	1.0005	1.0005
1.9	0.0221	1.0000	1.0006	1.0006
2.0	0.0203	1.0000	1.0012	1.0012

Table 4.16: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = 0$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.3134	2.3122	1.8473	2.0484
0.1	2.6916	1.8830	1.6932	1.8451
0.3	0.9893	0.9160	1.2486	1.2668
0.5	0.4495	0.9149	1.1106	1.1106
0.7	0.2407	0.9869	1.0551	1.0551
0.9	0.1497	1.0000	1.0290	1.0290
1.1	0.1017	1.0000	1.0166	1.0166
1.3	0.0736	1.0000	1.0170	1.0170
1.5	0.0565	1.0000	1.0132	1.0132
1.7	0.0435	1.0000	1.0092	1.0092
2.0	0.0317	1.0000	1.0076	1.0076

Table 4.17: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = 0.50$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.3345	1.2232	1.1361	1.1454
0.1	1.7755	1.1393	1.1202	1.1248
0.3	0.6973	0.9710	1.0420	1.0438
0.5	0.3113	0.9681	1.0173	1.0174
0.7	0.1659	0.9914	1.0091	1.0091
0.9	0.1062	1.0000	1.0064	1.0064
1.1	0.0728	1.0000	1.0046	1.0046
1.3	0.0536	1.0000	1.0033	1.0033
1.5	0.0393	1.0000	1.0025	1.0025
1.7	0.0314	1.0000	1.0020	1.0020
2.0	0.0219	1.0000	1.0007	1.0007

Table 4.18: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = 0.50$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	4.4010	1.4545	1.6401	1.6678
0.1	3.4652	1.3641	1.5642	1.6067
0.3	1.2837	1.0653	1.2681	1.2902
0.5	0.5661	0.9815	1.1244	1.1356
0.7	0.3238	0.9814	1.0768	1.0839
0.9	0.1906	0.9909	1.0509	1.0509
1.1	0.1292	0.9887	1.0291	1.0291
1.3	0.0948	0.9963	1.0270	1.0270
1.5	0.0727	1.0000	1.0211	1.0211
1.7	0.0569	1.0000	1.0170	1.0170
2.0	0.0415	1.0000	1.0124	1.0124

Table 4.19: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (6, 3)$ and $\rho = 0.90$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.1122	1.0829	1.0818	1.0854
0.1	1.9677	1.0841	1.0815	1.0844
0.3	1.1984	1.0135	1.0455	1.0516
0.5	0.6629	0.9715	1.0191	1.0234
0.7	0.4058	0.9795	1.0111	1.0111
0.9	0.2681	0.9906	1.0068	1.0068
1.1	0.1867	0.9975	1.0043	1.0043
1.3	0.1356	1.0000	1.0034	1.0034
1.5	0.1050	1.0000	1.0031	1.0031
1.7	0.0835	1.0000	1.0017	1.0017
1.9	0.0658	1.0000	1.0009	1.0009
2.0	0.0585	1.0000	1.0012	1.0012

Table 4.20: Simulated relative efficiency of the restricted, pretest and shrinkage estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 81, (p, q) = (9, 6)$ and $\rho = 0.90$ based on SMA model.

Δ^2	$\boldsymbol{\beta}^R$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.4038	1.1787	1.3686	1.3862
0.1	3.1443	1.1600	1.3621	1.3707
0.3	1.9271	1.0816	1.2554	1.2656
0.5	1.0800	1.0180	1.1637	1.1671
0.7	0.6714	0.9936	1.1142	1.1146
0.9	0.4318	0.9796	1.0662	1.0662
1.1	0.3000	0.9887	1.0483	1.0484
1.3	0.2264	0.9919	1.0313	1.0313
1.5	0.1696	0.9968	1.0252	1.0252
1.7	0.1323	0.9976	1.0196	1.0198
2.0	0.0950	1.0000	1.0140	1.0140

From these results, we can draw the following conclusions:

1. For fixed n and (p, q), varying ρ, the spatial dependence parameter does not affect much the SRE.

2. When the null hypothesis is true, $\Delta^2 = 0$, the restricted estimator $\hat{\boldsymbol{\beta}}^R$ outperforms all other estimators, but as we move away from the null hypothesis, that is, as Δ^2 moves away from zero, the SRE of $\hat{\boldsymbol{\beta}}^R$ approaches zero. In other words, the $SMSE(\hat{\boldsymbol{\beta}}^R)$ becomes unbounded. The SRE of all other estimators gets closer to one as Δ^2 moves away from zero.

3. For fixed value of $p - q = 3$, and when $\Delta^2 = 0$, the SRE of $\hat{\boldsymbol{\beta}}^{JS+}$ increases from 1.26 to more than 4 as q increases from 3 to 15, in addition, it performs better when $n = 36$ than 81.

4. The pretest estimator performs better than the shrinkage estimators near $\Delta^2 = 0$.

Thus, as Δ^2 moves away from zero, the performance of the positive James-Stein estimator is uniformly better than all other estimators. This leads us again to the conclusion that the positive James-Stein shrinkage estimator is a safer way to go when there are full and candidate competing submodels.

4.7.2 Application to Columbus Crime Data

The data was explained in Chapter 2. Here, we will follow Kyung and Ghosh (2010) and will use the log-transformation to restore some normality and to stabilize the variance. We assume a SMA error model with a Gaussian distribution, then we fit a full SMA model using the available regressors to predict the log(`CRIME`). A reduced model is then selected and fitted based on AIC and BIC selection criteria. Consequently, the pretest, James-Stein, and positive James-Stein estimators are computed based on full and reduced models. The reduced model obtained through the AIC/BIC selection criteria is reported in Table 4.21.

Table 4.21: Full and reduced SMA models for the Columbus crime data

Selection Criterion	Model
Full	`log(CRIME) ~ HOVAL+PLUMB+INC+DISCBD+OPEN`
AIC/BIC	`log(CRIME)~ HOVAL+PLUMB`

To compare the performance of the estimators on the crime data, we use their relative mean squared prediction error (RMSPE) with respect to the benchmark, $\hat{\boldsymbol{\beta}}$, following the same bootstrap procedure as laid down in Section 2.8.1.

Table 4.22: RMSPE with respect to $\hat{\boldsymbol{\beta}}$ based on SMA model for Columbus crime data

Estimator	RMSPE
$\hat{\boldsymbol{\beta}}^{R}$	1.1006
$\hat{\boldsymbol{\beta}}^{PT}$	1.0552
$\hat{\boldsymbol{\beta}}^{JS+}$	1.0301
$\hat{\boldsymbol{\beta}}^{JS}$	1.0297

From Table 4.22, we can see that the restricted estimator $\hat{\boldsymbol{\beta}}^{R}$ performs the best in terms of RMSPE followed by the pretest estimator $\hat{\boldsymbol{\beta}}^{PT}$ and then by the positive James-Stein. This may be an indication that the AIC/BIC selection criteria worked quite well on this data set.

4.8 Numerical Study For the Penalty Estimators

In this section, we present empirical studies to compare the penalty estimators with the restricted and positive James-Stein estimators via Monte Carlo simulations as well as through application to housing prices data.

4.8.1 Monte Carlo Simulations

Here we restrict attention to the comparison of the penalty, the restricted and the positive James-Stein estimators with respect to the benchmark unrestricted estimator. We use the simulated relative efficiency as measure of relative performance, as was done in Section 4.7.1, but under the candidate subspace, i.e., when $\Delta^2 = 0$. This will make the comparison fair as the penalty estimators do not depend on the value of Δ^2. In this simulation, we will consider square lattices with $N = 7$, 8, and 10, with corresponding sample sizes of $n = 49$, 64, and 100, respectively.

We fix $\sigma^2 = 1$, $p - q = 4$ (number of non zero parameters) with $q = \{5, 10, 15, 20, 25\}$, and $\rho = \{-0.95, -0.50, 0, 0.50, 0.95\}$. First, we fit a full and reduced SMA models to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^R$ estimators, then $\hat{\boldsymbol{\beta}}^{JS+}$ is obtained as in the previous section.

For computing the penalty estimators, we first extract the MLEs of ρ, σ^2, say $\hat{\rho}, \hat{\sigma^2}$ and use them to obtain the transformation in equation (4.15).

A 10−fold cross validation is used for $(\boldsymbol{Y}^*, \boldsymbol{X}^*)$ to select the optimum value of $\hat{\lambda}_{LASSO}$ for the LASSO fit, which is then used as initial weights (coefficient-specific regularization parameters) for the adaptive LASSO to obtain $\hat{\boldsymbol{\beta}}^{A.Lasso}$ via the `R`-function `adalasso` in `parcor` package (Kraemer and Schaefer, 2010). For the SCAD penalty function, we choose $a = 3.7$, as suggested by Fan and Li (2001), and perform a 10−fold cross-validation to obtain $\hat{\lambda}_{SCAD}$ using the `R`-function `cv.ncvreg` in the `ncvreg` package (Breheny and Huang, 2011). The SRE results are reported in Tables 4.23 to 4.25.

Table 4.23: Simulated relative efficiency of the restricted, positive James-Stein and penalty estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 49$, $p - q = 4$ and $\Delta^2 = 0$ for different values of ρ and q based on SMA model

ρ	q	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{JS+}$	$\hat{\boldsymbol{\beta}}^{Lasso}$	$\hat{\boldsymbol{\beta}}^{A.Lasso}$	$\hat{\boldsymbol{\beta}}^{SCAD}$
-0.95	5	4.0341	2.1241	1.0225	1.4432	1.2668
	10	9.1573	4.0540	1.5010	2.4868	2.0378
	15	16.5486	6.2364	2.1149	3.8801	3.0838
	20	33.4581	9.1395	3.3557	6.5302	4.4450
	25	67.7421	14.3414	5.4929	11.7382	7.3205
-0.50	5	2.9100	1.6757	1.4019	1.7656	1.7741
	10	4.7298	2.7099	1.7886	2.4884	2.4777
	15	8.1725	4.1547	2.4638	3.7706	3.8323
	20	13.1909	6.6220	3.6637	5.9903	5.9967
	25	22.5393	11.1292	5.4823	9.3836	9.5000
0.00	5	2.8231	1.6937	1.2603	1.5980	1.6545
	10	5.7844	2.8170	1.8699	2.6523	2.7520
	15	9.3916	4.1679	2.5667	3.8871	3.7748
	20	15.0166	6.6661	3.6596	5.8485	5.8570
	25	24.4345	10.8120	4.9206	8.0107	8.0193
0.50	5	2.7494	1.7050	1.2083	1.5338	1.5980
	10	5.8748	2.9235	1.8203	2.5913	2.6465
	15	9.9913	4.3876	2.6152	3.9519	4.1164
	20	15.6802	6.7174	3.5104	5.7011	5.9531
	25	24.2069	10.2452	4.7149	7.8391	7.6848
0.95	5	3.0027	1.7335	1.3668	1.7653	1.8184
	10	6.2542	2.9614	2.0345	2.9753	2.9840
	15	10.1222	4.4995	2.6253	4.0560	4.0376
	20	15.6824	6.3323	3.4726	5.6246	5.3575
	25	24.2740	8.4491	4.5684	8.0935	7.9708

Table 4.24: Simulated relative efficiency of the restricted, positive James-Stein and penalty estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 64$, $p - q = 4$ and $\Delta^2 = 0$ for different values of ρ and q based on SMA model

ρ	q	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{JS+}$	$\hat{\boldsymbol{\beta}}^{Lasso}$	$\hat{\boldsymbol{\beta}}^{A.Lasso}$	$\hat{\boldsymbol{\beta}}^{SCAD}$
-0.95	5	3.8106	2.0981	0.9647	1.4036	1.1468
	10	8.1229	4.1367	1.3452	2.2232	1.7111
	15	14.4330	5.9722	1.8258	3.1479	2.3049
	20	22.7448	8.0071	2.3616	4.4297	3.2192
	25	35.9866	10.2039	3.2867	6.3208	4.4977
-0.50	5	2.9438	1.6717	1.3238	1.7003	1.7159
	10	5.0695	2.7208	1.7489	2.5163	2.5784
	15	7.3697	3.8841	2.1492	3.2323	3.3386
	20	10.3082	5.3604	2.7011	4.4072	4.5905
	25	15.4137	7.2617	3.4376	5.9378	5.9148
0.00	5	2.6039	1.7033	1.1621	1.5015	1.5841
	10	4.8030	2.7239	1.5243	2.2290	2.3628
	15	7.9032	3.8728	2.1522	3.3954	3.6596
	20	11.8012	5.2703	2.7699	4.6025	4.7678
	25	17.9109	7.5420	3.7548	6.2601	6.7438
0.50	5	2.5603	1.7020	1.0928	1.4451	1.4947
	10	4.7799	2.7513	1.4939	2.1496	2.2221
	15	8.1815	3.9406	2.0517	3.2786	3.3743
	20	12.9093	5.3899	2.8185	4.6167	4.9842
	25	19.0960	7.3974	3.7405	6.3002	6.6010
0.95	5	2.7048	1.7104	1.1194	1.4574	1.4055
	10	5.4470	2.7735	1.5608	2.2104	2.2317
	15	9.0122	4.0444	2.1269	3.2458	3.1851
	20	13.4988	5.6737	2.6441	4.0898	3.9078
	25	18.6272	7.6109	3.1868	5.1346	5.0891

Table 4.25: Simulated relative efficiency of the restricted, positive James-Stein and penalty estimators with respect to $\hat{\boldsymbol{\beta}}$ when $n = 100$, $p-q = 4$ and $\Delta^2 = 0$ for different values of ρ and q based on SMA model

ρ	q	$\hat{\boldsymbol{\beta}}^{R}$	$\hat{\boldsymbol{\beta}}^{JS+}$	$\hat{\boldsymbol{\beta}}^{Lasso}$	$\hat{\boldsymbol{\beta}}^{A.Lasso}$	$\hat{\boldsymbol{\beta}}^{SCAD}$
-0.95	5	3.2393	2.0064	1.0271	1.6210	1.1831
	10	6.6126	3.8175	1.4364	2.7623	1.8455
	15	11.4987	5.8604	1.8727	3.9771	2.4442
	20	17.9655	8.1352	2.3439	5.3015	3.5426
	25	24.9062	9.9798	2.9342	6.7137	4.4651
-0.50	5	2.6876	1.6746	1.0385	1.3825	1.3209
	10	5.0449	2.5740	1.4030	2.0519	1.9109
	15	7.4354	3.5389	1.7282	2.5437	2.5155
	20	10.1443	4.7061	1.9594	2.9805	2.8748
	25	12.1422	5.9156	2.2092	3.4301	3.4859
0.00	5	2.3561	1.6744	0.9701	1.3222	1.3221
	10	4.0476	2.6876	1.2040	1.8044	1.8201
	15	6.1059	3.8248	1.5082	2.3632	2.4735
	20	8.7330	4.7064	1.9052	3.1921	3.3944
	25	11.6248	5.6481	2.2859	3.8072	4.1446
0.50	5	2.4333	1.7104	0.9746	1.3355	1.3270
	10	4.0514	2.7560	1.1578	1.7483	1.7270
	15	5.9224	3.7835	1.3911	2.1793	2.1643
	20	8.2712	4.7700	1.7798	2.8456	2.9118
	25	11.8818	5.9175	2.2144	3.7751	3.9972
0.95	5	2.4257	1.6717	0.9531	1.2987	1.2276
	10	4.3794	2.7497	1.1850	1.7623	1.6904
	15	6.7940	3.7312	1.4937	2.2603	2.1188
	20	10.2289	4.7003	1.9014	2.9551	2.8791
	25	13.8147	5.9497	2.2698	3.6787	3.5552

The following conclusions may be drawn from these Tables.

1. In this simulation, the restricted estimator, $\hat{\boldsymbol{\beta}}^{R}$, outperforms all other estimators for all the cases, and more so for small sample sizes. This is expected as we are working under $\boldsymbol{H}_0$.

2. As q increases, the relative efficiency of $\hat{\boldsymbol{\beta}}^{R}$ increases regardless of the values of ρ and n.

3. The positive rule James-Stein estimator, $\hat{\boldsymbol{\beta}}^{JS+}$, dominates all the penalty estimators.

4. The performance of the adaptive LASSO and the SCAD estimators is comparable, and both performed better than the LASSO estimator.

5. The performance of all penalty estimators increases as q increases, and in general, it does not depend on the value of ρ.

4.8.2 Application to Baltimore House Sale Prices Data

The Baltimore housing sale prices data was used, for instance, in Dubin (1992) in the context of hedonic regression, (Dunse and Jones (1998)). The authors proposed a method which excludes the variables that represent the neighborhood and accessibility, and then modeled the autocorrelation of the residuals, taking the spatial relationship explicitly into account.

The Baltimore housing prices data consist of 211 observations of each of the following variables, the selling price of houses in (USD 1000's) as the dependent variable (`PRICE`); a dummy variable with 1 if the unit is detached, and 0 otherwise (`DWELL`); number of bathrooms (`NBATH`); a dummy variable with 1 if the dwelling is located in Baltimore, 0 otherwise (`CITCOU`); number of rooms in a house (`NROOM`), a dummy variable with 1 if there is a basement, 0 otherwise (`BMENT`); the lot size in hundreds of square feet (`LOTSZ`); a dummy variable with 1 if fireplace, 0 otherwise (`FIREPL`); the age of dwelling in years (`AGE`); a dummy variable with 1 if a house contains air conditioning, 0 otherwise (`AC`); the living area in hundreds of square feet (`SQFT`); number of stories (`NSTOR`); a dummy variable with 1 if patio, 0 otherwise (`PATIO`); the X- and

Y-coordinates of the house (X) and (Y), respectively. In our analysis, we set the age of the house as one year if it is less than a year.

We use the available explanatory variables to predict the log(PRICE) assuming the errors follow a spatial moving average SMA structure to fit a full model. To set up a submodel, we apply the forward and backward selection methods on the available predictor variables. If the number of variables allowed in the model is set to two, the forward selection chooses the Y coordinate and the (DWELL) variables, while the backward elimination ends with the number of bathrooms (NBATH), and (CITCOU) variables. The two models are considered as the first and second candidate submodels in our study. When the number of important variables is restricted to only three, both forward and backward methods selected (DWELL), the number of bathrooms (NBATH), and the dummy variable (CITCOU). This model is considered as the third submodel. The full and suggested candidate submodels are reported in Table 4.26.

Table 4.26: Full and submodels for the Baltimore House prices data based on SMA error structure Model

Selection Criterion	Model
Full	log(PRICE) ~ NBATH + CITCOU+ DWELL + Y + BMENT + FIREPL + AC + NROOM + LOTSZ + PATIO + log(AGE)+ X + log(SQFT) + NSTOR + GAR
Forward	log(PRICE) ~ DWELL + Y
Backward	log(PRICE) ~ NBATH + CITCOU
Forward/Backward	log(PRICE) ~ NBATH + CITCOU + DWELL

To obtain the penalty estimators, we firstly extract the MLE of the spatial dependence parameter ρ from the full SMA model, and use it in constructing the necessary transformation matrix explained in Section 2.8.1. We compute the penalty estimators using 10-fold cross validation as detailed in the previous sections.

The restricted, positive James-Stein, and penalty estimators were compared by using the relative mean squared prediction error (RMSPE) criteria. This quantity was computed as described in Chapter 2 via a bootstrapping approach with 2000 bootstrap samples. The RMSPE results are summarized in Table 4.27.

Table 4.27: RMSPE with respect to $\hat{\boldsymbol{\beta}}$ for Baltimore House prices data based on SMA model

Model	$\boldsymbol{\beta}^{R}$	$\hat{\boldsymbol{\beta}}^{JS+}$
Forward	1.1033	1.0856
Backward	1.0999	1.0843
Forward/Backward	1.0870	1.0728
	Penalty Estimators	
$\hat{\boldsymbol{\beta}}^{LASSO}$	$\hat{\boldsymbol{\beta}}^{A.LASSO}$	$\hat{\boldsymbol{\beta}}^{SCAD}$
1.0391	1.0607	1.0315

We may draw the following conclusions from Table 4.27:

1. As expected, the restricted estimator $\hat{\boldsymbol{\beta}}^{R}$ has the highest RMSPE values among all estimators. So, if the suggested submodel is correct, then $\hat{\boldsymbol{\beta}}^{R}$ is optimum.

2. The RMSPE values of the positive James-Stein estimator $\hat{\boldsymbol{\beta}}^{JS+}$ are better than those of the penalty estimators.

3. All penalty estimators performed better than the classical estimator. In addition, for this data set, the adaptive LASSO performs better than the LASSO and the SCAD estimators.

4. The first submodel is recommended since it contains only two predictors in addition to the intercept and produces the highest values of the RMSPE for the restricted and positive James-Stein estimators.

4.9 Conclusion

In this chapter, we proposed shrinkage and pretest estimators in the context of the spatial moving average regression. We have shown that the full model MLE (unrestricted) and the restricted (candidate submodel) estimators are jointly multivariate normal. Using this key result, we presented some analytical results to compare the asymptotic biases and quadratic risks of the pretest, the shrinkage, the restricted and the unrestricted estimators. Consequently, we came to the conclusion that the positive James-Stein estimator followed by the pretest should be a safer way of estimating the SMA large-scale effects when uncertain prior information is available. We used Monte carlo simulations to confirm these findings and applied these methods to real data set for illustration purposes.

Also, in this Chapter, we developed a numerical technique for computing penalty estimators in general, and applied to the special cases of computing LASSO, Adaptive LASSO, and the SCAD estimators for the SMA model's large-scale effect parameter $\boldsymbol{\beta}$. This was followed by Monte Carlo comparative study and an application to Baltimore housing price data. These numerical studies confirmed the dominance property of the restricted estimator over all penalty estimators when the submodel is correct as well as when the submodel comes from the usual AIC/BIC selection criteria. Moreover, the study confirms that the positive James-Stein estimator outperforms all penalty estimators in terms of simulated relative efficiency SRE.

Chapter 5

Conclusions and Future Research

In this dissertation, we have studied three important spatial regression models and developed efficient estimation strategies. More specifically, we considered the conditional autoregressive, simultaneous autoregressive and moving average spatial regression models, respectively. We constructed pretest ($\hat{\boldsymbol{\beta}}^{PT}$), James-Stein shrinkage ($\hat{\boldsymbol{\beta}}^{JS}$) and positive James-Stein shrinkage ($\hat{\boldsymbol{\beta}}^{JS+}$) estimators for the regression coefficients of these models. These three estimators are well-known to be efficient in incorporating prior uncertain information (UPI) into the estimation of model parameters. UPIs accommodate a wide range of possibilities such as an expert's opinion that some of the regression coefficients are irrelevant as well as information in the form of submodels selected by model selection procedures. For that reason, we formulated a very general UPI in the form of linear restriction on the regression coefficients and thence, we obtained restricted (to the sub-space defined by the linear restriction) estimator, $\hat{\boldsymbol{\beta}}^{R}$.

The asymptotic distribution of the restricted and the unrestricted estimators were derived and used in obtaining expressions for the risks and biases of our proposed estimators. Analytical comparisons were undertaken based on the risk and bias expressions of the pretest, James-Stein, positive James-Stein, and the restricted estimators of the regression coefficients for the conditional autoregressive, simultaneous autoregressive and spatial moving average models with respect to the unrestricted estimator. We also devised procedures for obtaining penalty estimators for these three spatial models. The proposed penalty estimators exploit existing penalty estimation algorithms such as LARS algorithm of Efron et al. (2004).

Overall, the following topics have been discussed in this dissertation in the context of three spatial regression models, CAR, SAR and SMA:

(1) Unrestricted, restricted, pretest and shrinkage estimators.

(2) A class of penalty estimators: LASSO, Adaptive LASSO and SCAD.

(3) Analytical and numerical comparisons of the pretest, shrinkage, restricted and unrestricted estimators.

(4) Numerical comparisons of the restricted and positive shrinkage with the penalty estimators through simulation experiments and real data examples.

We summarize the findings as follows: In Chapter 2, we proposed the restricted, the pretest and the shrinkage estimators of the large-scale effect parameter vector, $\boldsymbol{\beta}$, in the CAR model under a general linear restriction, $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$. We derived the joint asymptotic distribution of the unrestricted and restricted via Mardia-Marshall (Mardia and Marshall (1984)) Theorem and consequently, obtained the asymptotic quadratic risks and biases of the proposed estimators. We examined analytically the relative dominance picture of these four estimators with respect to the unrestricted estimator of $\boldsymbol{\beta}$. We also carried out an intensive Monte Carlo simulation study to compare these estimators in terms of their relative mean squared errors. Real data set on Boston crime statistics were employed to illustrate the actual implementation of these estimators. We indicated that the submodel, required for building the pretest and shrinkage estimators, can be practically obtained via any model selection procedure. In the application to Boston crime data, we selected submodels via stepwise selection procedures based on the AIC and BIC criteria. We concluded that among the proposed estimators, the positive shrinkage estimator performs the best in the sense of giving the smallest mean squared prediction error in most of the parameter configurations considered.

The second important contribution of Chapter 2 was the construction of penalty estimators via model transformations along with the existing penalty estimation algorithms. Intensive numerical comparisons were undertaken to contrast the restricted, positive shrinkage and the penalty estimators via Monte Carlo simulations as well as application to real data set. When discussing the practical implementation of the penalty estimators through the real data set, we also illustrated how prediction errors of these estimators as well as those of the restricted, unrestricted and positive shrinkage can be computed by using a bootstrapping procedure. This strengthens the viewpoint that these estimators are not just for mathematical exercise but rather implementable efficient choices of estimation in spatial regression models.

In a nutshell, in the CAR spatial model, the large-scale effect, $\boldsymbol{\beta}$ can be efficiently estimated using the positive shrinkage which outperforms all other estimators considered in the chapter by giving the smallest mean squared error. Such estimator can be based, in practice, on submodels chosen through model selection criteria.

In Chapters 3 and 4, the ideas and the milestones of Chapter 2 were then extended to other spatial regression models. Specifically, in Chapter 3 we considered the estimation of $\boldsymbol{\beta}$ in the simultaneous autoregressive regression model. We proposed the restricted, pretest, shrinkage as well as the penalty estimators for estimating the large-scale effects, $\boldsymbol{\beta}$, of the SAR model under a linear candidate submodel. The estimators were compared via their relative risks and biases by using Monte Carlo simulations and applications to the Boston housing prices data. The final conclusions in this Chapter were essentially same as those reached in Chapter 2. A similar study was undertaken in Chapter 4 by constructing the restricted, pretest, shrinkage as well as the penalty estimators for a spatial regression model known as the spatial moving average model. The conclusions in Chapter 4 were in line with those of Chapters 2 and 3.

5.1 Future Research

The topic of spatial regressions, which was the focus of this dissertation, is surely one that has been gaining momentum in the past few years due to the availability of large and complex spatio-temporal data and due to the need of Governments to utilize such data for policy making. The areas of application for spatio-temporal methodologies are ever widening and include, but not limited to, epidemiology and disease mapping, estimation and mapping of geographically distributed resources such as water, oil etc, climate related issues and much more.

This dissertation proposed and studied some efficient estimation strategies for three spatial regression models. This is just a scratch on the surface of the spatial regression models. There are endless opportunities of extending the efficient estimation strategies developed in this dissertation to many other spatial regression models.

For instance, our estimation strategies can be extended to conditional autoregressive models for discrete spatial data types. The generalized linear models with covariances structured as in the CAR model can serve as a vehicle for the analysis of discrete spatial data (Gotway and Stroup, 1997). On the other hand, Banerjee et al. (2003), introduced and studied a Bayesian version of the Cox's proportional hazards (PH) model for spatial survival data and illustrated the methodology by estimating and mapping infant mortality in the state of Minnesota. These authors introduced a multivariate frailty in the PH model and imposed CAR, SAR and geostatistical covariance structures on the frailties.

Based on reviewed literature, there has not been any study investigating frequentist's estimation approaches for such spatial PH models. That is, there is a need for constructing MLEs for the spatial PH model under various covariance structures and then applying the efficient estimation procedures developed in this dissertation. Aalen's additive survival models are flexible alternatives to the PH model. Spatial accelerated failure time models similar to the models of Banerjee et al. (2003) have been studied and applied to mapping of prostate cancer survival by Zhang and Lawson (2011). Hussein et al. (2013) studied shrinkage estimation strategies for non-spatial Aalen's model. There is an opportunity in incorporating spatial covariance structures into the Aalen's additive model and proposing efficient estimation strategies.

Appendix A

Mardia and Marshall Theorem

Based on increasing-domain asymptotic, Mardia and Marshall (1984) proved the asymptotic normality of the MLEs for spatial regression models. The authors considered a real valued Gaussian process $\{\boldsymbol{Y}(\boldsymbol{s}) : \boldsymbol{s} \in \boldsymbol{S}\}$, where $\boldsymbol{S}$ is an index set that satisfies for all $\boldsymbol{s} \in \boldsymbol{S}, E(\boldsymbol{Y}(\boldsymbol{s})) = \boldsymbol{X}(\boldsymbol{s})'\boldsymbol{\beta}$, where $\boldsymbol{X}(\boldsymbol{s}) = \{X_1(s), X_2(s), \ldots, X_p(s)\}'$ is a $p \times 1$ vector of fixed regressors, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a parameter vector in $\boldsymbol{R}^p$. As an example, $\boldsymbol{S} = \boldsymbol{Z}^k$ describes a k-dimensional lattice process, and $\boldsymbol{S} = \boldsymbol{R}^k$ describes a continuous parameter process or $\boldsymbol{S}$ may be a collection of spatial counties or regions. They assumed that $cov(\boldsymbol{Y}(s_i), \boldsymbol{Y}(s_j)) = \boldsymbol{\Sigma}(s_i, s_j; \boldsymbol{\gamma})$ is twice differentiable with respect to $\boldsymbol{\gamma} \in \boldsymbol{R}^k$ and $s_i, s_j \in \boldsymbol{S}$ is a positive definite matrix in the sense that for every finite subset, $\boldsymbol{S}_n = \{s_1, s_2, \ldots, s_n\}$ of $\boldsymbol{S}$, the variance covariance matrix $\boldsymbol{\Sigma}_n = \{\boldsymbol{\Sigma}(s_i, s_j; \boldsymbol{\gamma})\}, i, j = 1, \ldots, n$ is positive definite. The vector γ is to model the spatial dependence structure.

In our case $\boldsymbol{\gamma}$ is a 2×1 vector where $\gamma_1 = \sigma^2$ and $\gamma_2 = \rho$. Suppose that $\boldsymbol{Y} = \{Y(s_1), Y(s_2), \ldots, Y(s_n)\}'$ is the data at each point in $\boldsymbol{S}_n = \{s_1, s_2, \ldots, s_n\} \subset \boldsymbol{S}$, and $\boldsymbol{X}_n$ is an $n \times p$ matrix of regressors of rank p where the i^{th}column of $\boldsymbol{X}_n$ is $X_i(\boldsymbol{s}) \equiv \boldsymbol{X}_i = \{X_i(s_1), X_i(s_2), \ldots, X_i(s_n)\}$ for $i = 1, 2, \ldots, p$. For simplicity, we drop the indices $\boldsymbol{s}$ and n, and let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ denote the $(p+2) \times 1$ parameter vector, the variance covariance matrix by $\boldsymbol{\Sigma}$. The log-likelihood for $\boldsymbol{\theta}$ is given by:

$$L(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta}), \tag{A.1}$$

with the first order partial derivatives,

$$\begin{aligned} L^{(1)} &= \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \left(L'_{\boldsymbol{\beta}}, L'_{\boldsymbol{\gamma}}\right)', \end{aligned} \tag{A.2}$$

$$\begin{aligned} L'_{\boldsymbol{\beta}} &= \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ &= -\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X\beta} + \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}, \end{aligned} \tag{A.3}$$

and the i^{th} element of $L_{\boldsymbol{\gamma}}$ is

$$\begin{aligned} (L_{\boldsymbol{\gamma})_i} &= \frac{\partial L(\boldsymbol{\theta})}{\partial \gamma_i} \\ &= \frac{1}{2}tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\right) + \frac{1}{2}\boldsymbol{e}'\boldsymbol{\Sigma}^i\boldsymbol{e}, \quad i = 1, 2 \end{aligned} \tag{A.4}$$

where $tr(\boldsymbol{A})$ means the trace of the matrix $\boldsymbol{A}$, $\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X\beta}, \boldsymbol{\Sigma}_i = \frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_i}$, and $\boldsymbol{\Sigma}^i = \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \gamma_i} = -\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}$ for $i = 1, 2$.

The second order partial derivative of L is given by:

$$\begin{aligned} L^{(2)} &= \frac{\partial^{(2)} L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \\ &= \begin{bmatrix} L_{\boldsymbol{\beta\beta}} & L_{\boldsymbol{\beta\gamma}} \\ L'_{\boldsymbol{\beta\gamma}} & L_{\boldsymbol{\gamma\gamma}} \end{bmatrix}, \end{aligned} \tag{A.5}$$

where

$$L_{\boldsymbol{\beta\beta}} = -\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}, \tag{A.6}$$

the i^{th} column of $L_{\boldsymbol{\beta\gamma}}$ is

$$-\boldsymbol{X}'\boldsymbol{\Sigma}^{i}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{X}'\boldsymbol{\Sigma}^{i}\boldsymbol{Y}, \quad i = 1, 2, \tag{A.7}$$

and the $(i, j)^{th}$ element of $L_{\boldsymbol{\gamma\gamma}}$ is

$$(L_{\boldsymbol{\gamma\gamma}})_{ij} = \frac{1}{2}\{tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{ij} + \boldsymbol{\Sigma}^{i}\boldsymbol{\Sigma}_{j}) + \boldsymbol{e}'\boldsymbol{\Sigma}^{ij}\boldsymbol{e}, \tag{A.8}$$

where $\boldsymbol{\Sigma}_{ij} = \frac{\partial^2 \boldsymbol{\Sigma}}{\partial\gamma_i\partial\gamma_j}$, and $\boldsymbol{\Sigma}^{ij} = \frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial\gamma_i\partial\gamma_j} = \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_{ij}\right)\boldsymbol{\Sigma}^{-1}$. The expected information matrix $-(E(L^{(2)}))$ is given by:

$$-E(L^{(2)}) = J = \begin{bmatrix} J_{\boldsymbol{\beta}} & \mathbf{0} \\ \mathbf{0} & J_{\boldsymbol{\gamma}} \end{bmatrix}, \tag{A.9}$$

where $J_{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}X$, and the $(i, j)^{th}$ element of $J_{\boldsymbol{\gamma}}$ is $\frac{1}{2}b_{ij}$ with

$$b_{ij} = tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j\right) = tr\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{i}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{j}\right) \quad i, j = 1, 2. \tag{A.10}$$

General results of the MLEs $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}$ are given by Magnus (1978) and Sweeting (1980), and based on Sweeting's result, Mardia and Marshall proved the following theorem.

Theorem A.0.1. (Mardia and Marshall, 1984): Suppose that $\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed but unknown mean parameters, and $\boldsymbol{\Sigma}$ is a function of $\boldsymbol{\gamma}$, a $k \times 1$ vector of unknown spatial-dependence parameters. Let $\lambda_1 \leq \lambda_2 \leq \ldots \lambda_n$ be the eigenvalues of $\boldsymbol{\Sigma}$, and let those for $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Sigma}_{ij}$ be $\{\lambda_l^i : l = 1, 2, \ldots, n\}$ and $\{\lambda_l^{ij} : l = 1, 2, \ldots, n\}$ respectively, with $|\lambda_1^i| \leq |\lambda_2^i| \leq \cdots \leq |\lambda_n^i|$ and $|\lambda_1^{ij}| \leq |\lambda_2^{ij}| \leq \cdots \leq |\lambda_n^{ij}|$ for $i, j = 1, 2, \ldots, k$. Suppose that

(i) $\lim_{n\to\infty} \lambda_n = e < \infty$, $\lim_{n\to\infty} |\lambda_n^i| = e_i < \infty$, $\lim_{n\to\infty} |\lambda_n^{ij}| = e_{ij} < \infty$ for all $i, j = 1, 2, \ldots, k$

(ii) $\|\boldsymbol{\Sigma}_i\|^{-2} = O\left(n^{-\frac{1}{2}-\delta}\right)$, for some $\delta > 0$, $i = 1, 2, \ldots, k$ where $\|.\|$ denotes the Euclidean matrix norm.

(iii) For all $i, j = 1, 2, \ldots, k$, $\lim_{n\to\infty} \left\{ \frac{b_{ij}}{\sqrt{b_{ii} b_{jj}}} \right\} = a_{ij}$, where b_{ij} are given in (A.10) and $\boldsymbol{A} = \{a_{ij}\}_{i,j=1}^n$ is nonsingular matrix.

(iv) $\lim_{n\to\infty} (\boldsymbol{X}'\boldsymbol{X})^{-1} = 0$.

Then the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ satisfies the following:

1. $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} \xrightarrow{P} \boldsymbol{\theta}$, where $\xrightarrow{P}$ means convergence in probability.
2. $\boldsymbol{J}^{\mathbf{1/2}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{I})$, where $\xrightarrow{D}$ means convergence in distribution, and $\boldsymbol{J}$ is the expected information matrix given in (A.9).

From the above theorem, it is clear that $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ are asymptotically independent, therefore:

$$\boldsymbol{J}_{\boldsymbol{\beta}}^{\mathbf{1/2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{I}_p).$$

(v) $\frac{\left(\boldsymbol{X}_n' \hat{C}_n^{-1} \boldsymbol{X}_n\right)}{n} \xrightarrow{P} \boldsymbol{C}_0$, and $\hat{C}_n = (\boldsymbol{I} - \hat{\rho}\boldsymbol{W}^*)^{-1}\boldsymbol{D}$, where $\boldsymbol{C}_{\mathbf{0}}$ is a $p \times p$ positive definite matrix.

Bibliography

Ahmed, S. E. (1997a). Asymptotic shrinkage estimation: the regression case. In *Applied statistical science, II (Malang, 1996)*, pages 113–143. Nova Sci. Publ., Commack, NY.

Ahmed, S. E. (1997b). Improved *R*-estimation of regression coefficients. *J. Statist. Res.*, 31(1):53–73.

Ahmed, S. E. (1998). Improved pretest nonparametric estimation in a multivariate regression model. *Comm. Statist. Theory Methods*, 27(10):2391–2421.

Ahmed, S. E. (2001). Shrinkage estimation of regression coefficients from censored data with multiple observations. In *Empirical Bayes and likelihood inference (Montreal, QC, 1997)*, volume 148 of *Lecture Notes in Statist.*, pages 103–120. Springer, 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(4):435–454.

Ahmed, S. E. and Fallahpour, S. (2012). Shrinkage estimation strategy in quasi-likelihood models. *Statistics & Probability Letters*, 82(12):2170–2179.

Ahmed, S. E., Hussein, A. A., and Sen, P. K. (2006). Risk comparison of some shrinkage M-estimators in linear models. *J. Nonparametr. Stat.*, 18(4-6):401–415.

Ahmed, S. E. and Raheem, S. E. (2012). Shrinkage and absolute penalty estimation in linear regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(6):541–553.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht.

Anselin, L. and Florax, R. (1995). Small Sample Properties Of Tests For Spatial Dependence In Regression Models: Some Further Results. *New directions in spatial econometrics*, pages 21–74.

Bailey, T. and Gatrell, A. (1995). *Interactive Spatial Data Analysis*. Longman Scientific & Technical Essex.

Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Statistics*, 15:190–204.

Banerjee, S., Wall, M. M., and Carlin, B. P. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, 4(1):123–142.

Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., and Elston, D. A. (2010). Regression analysis of spatial data. *Ecology Letters*, 13(2):246–264.

Bell, K. P. and Bockstael, N. E. (2000). Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data. *The Review of Economics and Statistics*, 82(1):72–82.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B*, 36:192–236. With discussion by D. R. Cox, A. G. Hawkes, P. Clifford, P. Whittle, K. Ord, R. Mead, J. M. Hammersley, and M. S. Bartlett and with a reply by the author.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, 43(1):1–59. With discussion and a reply by Besag.

Breheny, P. and Huang, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Annals of Applied Statistics*, 5(1):232–253.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.*, 80(391):580–619. With discussion and with a reply by the authors.

Cliff, A. D. and Ord, J. K. (1981). *Spatial processes: models & applications.* Pion Ltd., London.

Cressie, N. A. C. (1993). *Statistics for spatial data.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York. Revised reprint of the 1991 edition, A Wiley-Interscience Publication.

Cressie, N. A. C. and Wikle, C. K. (2011). *Statistics for spatio-temporal data.* Wiley Series in Probability and Statistics. John Wiley & Sons Inc., Hoboken, NJ.

Dubin, R. A. (1992). Spatial autocorrelation and neighborhood quality. *Regional Science and Urban Economics*, 22(3):433–452.

Dunse, N. and Jones, C. (1998). A hedonic price model of office rents. *Journal of Property Valuation and Investment*, 16(3):297–312.

Durban, M., Minguez, R., and Montero, J. (2012). SAR models with nonparametric spatial trends. A P-spline approach. *Spanish Statistics*, 54(177):89–111.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.*, 32(2):407–499. With discussion, and a rejoinder by the authors.

Evans, R. and Forcina, A. (2013). Two algorithms for fitting constrained marginal models. *Comutational Statistics & Data Analysis*, 66:1 – 7.

Fallahpour, S., Ahmed, S. E., and Doksum, K. A. (2012). L1 penalty and shrinkage estimation in partially linear models with random coefficient autoregressive errors. *Applied Stochastic Models in Business and Industry*, 28(3):236–250.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, 96(456):1348–1360.

Florax, R. J., Folmer, H., and Rey, S. (2003). Specification searches in spatial econometrics: the relevance of Hendry's methodology. *Regional Science and Urban Economics*, 33(5):557–579.

Griffith, D. A. (2000). A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems*, 2(2).

Guo, J., Tang, M., Tian, M., and Zhu, K. (2013). Variable selection in high-dimensional partially linear additive models for composite quantile regression. *Computational Statistics & Data Analysis*, 65:56–67.

Haining, R. P. (1978). The Moving Average Model for Spatial Interaction. *Transactions of the Institute of British Geographers*, 3(2):pp. 202–225.

Hall, P. (1985). Resampling a coverage pattern. *Stochastic Process. Appl.*, 20(2):231–246.

Harrison, D. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5(1):81–102.

Hoeting, J. A., Davis, R. A., Merton, A. A., and Thompson, S. E. (2006). Model Selection for Geostatistical Models. *Ecological Applications*, 16(1).

Hossain, S. and Ahmed, S. E. (2012). Shrinkage and Penalty Estimators of a Poisson Regression Model. *Australian & New Zealand Journal of Statistics*, 54(3):359–373.

Hossain, S., Doksum, K. A., and Ahmed, S. E. (2009). Positive shrinkage, improved pretest and absolute penalty estimators in partially linear models. *Linear Algebra Appl.*, 430(10):2749–2761.

Huang, H.-C., Hsu, N.-J., Theobald, D. M., and Breidt, F. J. (2010). Spatial lasso with applications to GIS model selection. *J. Comput. Graph. Statist.*, 19(4):963–983. Supplementary materials available online.

Huang, J., Ma, S., and Zhang, C.-H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statist. Sinica*, 18(4):1603–1618.

Huang, J. S. (1984). The autoregressive moving average model for spatial analysis. *Austral. J. Statist.*, 26(2):169–178.

Hussein, A., Nkurunziza, S., and Tomanelli, K. (2013). Shrinkage estimation strategies for Aalen's model (Under review). *Australian & New Zealand Journal of Statistics*.

Judge, G. G. and Bock, M. E. (1978). *The statistical implications of pre-test and Stein-rule estimators in econometrics*. North-Holland Publishing Co., Amsterdam. Studies in Mathematical and Managerial Economics, Vol. 25.

Kamarianakis, Y., Shen, W., and Wynter, L. (2012). Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. *Applied Stochastic Models in Business and Industry*, 28(4):297 – 315.

Kashyap, R. and Chellappa, R. (1983). Estimation and choice of neighbors in spatial-interaction models of images. *IEEE Trans. Inf. Theor.*, 29(1):60–72.

Kazar, B. and Celik, M. (2012). *Spatial AutoRegression (SAR) Model: Parameter Estimation Techniques*. Springer US.

Khan, B. U. and Ahmed, S. E. (2003). Improved estimation of coefficient vector in a regression model. *Comm. Statist. Simulation Comput.*, 32(3):747–769.

Kissling, W. D. and Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, 17(1):59–71.

Kraemer, N. and Schaefer, J. (2010). *parcor: Regularized estimation of partial correlation matrices*. R package version 0.2-3.

Kyung, M. and Ghosh, S. K. (2009). Bayesian inference for directional conditionally autoregressive models. *Bayesian Anal.*, 4(4):675–706.

Kyung, M. and Ghosh, S. K. (2010). Maximum likelihood estimation for directional conditionally autoregressive models. *J. Statist. Plann. Inference*, 140(11):3160–3179.

Lange, N. and Ryan, L. (1989). Assessing normality in random effects models. *Ann. Statist.*, 17(2):624–642.

Lee, L.-F. and Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *J. Econometrics*, 154(2):165–185.

Lee, L.-F. and Yu, J. (2013). Near Unit Root in the Spatial Autoregressive Model. *Spatial Economic Analysis*, pages 1–38.

Leeb, H. and Poetscher, B. M. (2008). Sparse estimators and the oracle property, or the return of hodges estimator. *Journal of Econometrics*, 142(1):201 – 211.

Li, H., Calder, C. A., and Cressie, N. (2012). One-step estimation of spatial dependence parameters: Properties and extensions of the APLE statistic. *J. Multivariate Analysis*, 105(1):68–84.

Lin, Y., Wang, S., and Chappell, R. J. (2013). Lasso tree for cancer staging with survival data. *Biostatistics*, 14(2):327–339.

Magnus, J. R. (1978). Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. *Journal of Econometrics*, 7(3):281 – 312.

Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71(1):135–146.

Mur, J. (1999). Testing for spatial autocorrelation: moving average versus autoregresive processes. *Environment and Planning A*, 31(8):1371–1382.

Nkurunziza, S. (2011). Shrinkage strategy in stratified random sample subject to measurement error. *Statistics & Probability Letters*, 81(2):317 – 325.

Nkurunziza, S. (2012a). The risk of pretest and shrinkage estimators. *Statistics*, 46(3):305–312.

Nkurunziza, S. (2012b). Shrinkage strategies in some multiple multi-factor dynamical systems. *ESAIM: Probability and Statistics*, 16:139–150.

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

Ord, K. (1975). Estimation methods for models of spatial interaction. *J. Amer. Statist. Assoc.*, 70:120–126.

Pace, R. K. (1993). Nonparametric Methods with Applications to Hedonic Models. *The Journal of Real Estate Finance and Economics*, 7(3):185–204.

Pace, R. K. and Gilley, O. W. (1997). Using the Spatial Configuration of the Data to Improve Estimation. *The Journal of Real Estate Finance and Economics*, 14(3):333–340.

Pötscher, B. M. and Schneider, U. (2009). On the distribution of the adaptive LASSO estimator. *J. Statist. Plann. Inference*, 139(8):2775–2790.

Qian, W. and Yang, Y. (2013). Model selection via standard error adjusted adaptive lasso. *Annals of the Institute of Statistical Mathematics*, 65(2):295–318.

Raheem, S. M. E., Ahmed, S. E., and Doksum, K. A. (2012). Absolute penalty and shrinkage estimation in partially linear models. *Comput. Statist. Data Anal.*, 56(4):874–891.

Ren, Y. and Zhang, X. (2013). Model selection for vector autoregressive processes via adaptive lasso. *Communications in Statistics - Theory and Methods*, 42(13):2423–2436.

Saleh, A. K. M. E. (2006). *Theory of preliminary test and Stein-type estimation with applications*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.

Seber, G. A. F. (2008). *A matrix handbook for statisticians*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.

Song, J. J. and De Oliveira, V. (2012). Bayesian model selection in spatial lattice models. *Stat. Methodol.*, 9(1-2):228–238.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume I, pages 197–206, Berkeley and Los Angeles. University of California Press.

Stein, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. *Research Papers in Statistics*, pages 351–366.

Stine, R. A. (2004). Model selection using information theory and the MDL principle. *Sociol. Methods Res.*, 33(2):230–260.

Su, L. (2012). Semiparametric GMM estimation of spatial autoregressive models. *J. Econometrics*, 167(2):543–560.

Su, L. and Jin, S. (2010). Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *J. Econometrics*, 157(1):18–33.

Sweeting, T. J. (1980). Uniform Asymptotic Normality of the Maximum Likelihood Estimator. *The Annals of Statistics*, 8(6):pp. 1375–1381.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121(2):311–324.

Wang, H. and Zhu, J. (2009). Variable selection in spatial regression via penalized least squares. *Canad. J. Statist.*, 37(4):607–624.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41:434–449.

Zhang, H. H. and Lu, W. (2007). Adaptive Lasso for Cox's proportional hazards model. *Biometrika*, 94(3):691–703.

Zhang, J. and Lawson, A. B. (2011). Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. *J. Appl. Stat.*, 38(3):591–603.

Zhu, J., Huang, H.-C., and Reyes, P. E. (2010). On selection of spatial linear models for lattice data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):389–402.

Zhu, Z. and Liu, Y. (2009). Estimating spatial covariance using penalised likelihood with weighted L_1 penalty. *J. Nonparametr. Stat.*, 21(7):925–942.

Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, 101(476):1418–1429.

Vita Auctoris

Marwan Al-Momani was born in 1973 in Irbid, Jordan. He obtained his B.Sc. and M.Sc. in Statistics in 1996 and 2000 respectively from Yarmouk University, Jordan. In 2002, he joined in King Fahd University of Petroleum & Minerals as a lecturer and worked there before coming to the University of Windsor in 2008. He is currently a candidate for the degree of Doctor of Philosophy in Statistics and hopes to graduate in summer 2013.